Material on Enclosed CD-ROM

The table of contents is hyperlinked. To go to any section click on its corresponding listing in this table of contents.

Case Studies

A problem well defined is a problem half solved.

—Charles Kettering, 1876–1958

W e provide three case studies to illustrate the use of the Statistical Engineering variation reduction algorithm. We hope that the case studies help you to:

- Better understand the stages of the algorithm and how to move among them.

- Understand the considerations in the choice of a working approach.

- Select appropriate investigations and analysis tools in the search for a dominant cause.

- Assess the feasibility of an approach and implement it.

As W. Edwards Deming somewhat paradoxically said, "You cannot learn by example." We recognize that there are no other processes or problems for which the case studies are perfect models. We recommend, when reading the case studies, that you think about where your own processes and problems are similar to and different from those in the cases.
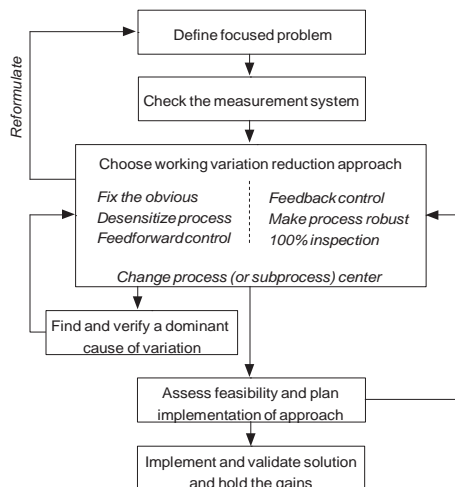
Statistical Engineering variation reduction algorithm.

Case Study I

Brake Rotor Balance

The value of an idea lies in using it.

—Thomas Edison, 1847–1931

An iron foundry produced veined brake rotors (Figure I.1) that were machined at a separate location. The machining plant 100% inspected the rotors for balance and welded a weight into the veins if the imbalance was too severe. We call a rotor needing added weight a balance reject.

The historic rate of balance rejects was approximately 25%. The foundry initiated the project because the reject rate jumped to 50%. This increase in rework coincided with a move from a four-cavity to a six-cavity (called a four-gang and six-gang) core mold to increase productivity in the foundry. The cores were set in the mold to create the veins when the rotor was cast.

Figure I.1   Brake rotor.
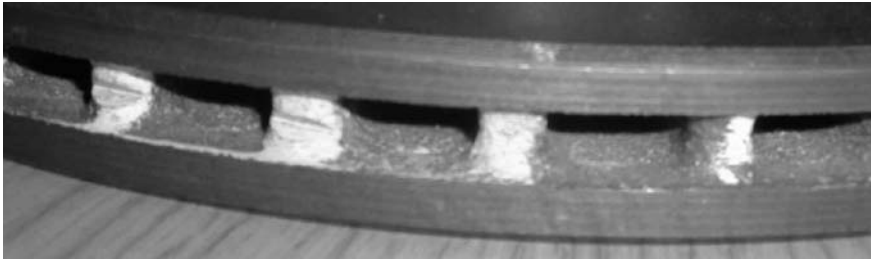
Figure I.2   Side view of a brake rotor showing veins.

The foundry was convinced that the change to the six-cavity mold was not the cause of the increase in balance rejects. Their confidence was based on their previous experience and because a full analysis of the six-cavity mold had shown all dimensions well within specification.

The increased reject rate could not be explained by any other changes made at either the foundry or the machining operation. As it stood, each party blamed the other. To address the increased rework, the machining operation planned to add another rework station. The foundry formed a team with the goal of reducing the reject rate to at most the historical level. To determine imbalance, the machining plant measured the center of gravity (a distance and direction from the rotor center), which was then translated into a weight (in ounces) and orientation needed to balance the rotor. If needed, the weights were welded to the veins on the inside of the rotor, as seen in Figure I.2. A balance reject was any rotor needing weight greater than 0.5 ounces. To focus the problem, the team selected the balance weight as the output. They knew that if they could reduce the weight, they could eliminate the rework, regardless of the orientation.

To establish the baseline in terms of the balance weight, the team selected 300 machined rotors spread out over the previous week's production. The data are given in *brake rotor balance baseline*. The baseline histogram and run chart are given in Figure I.3. In the baseline, 46% of the rotors required rework. The run chart of the balance weights suggests no obvious pattern over time.
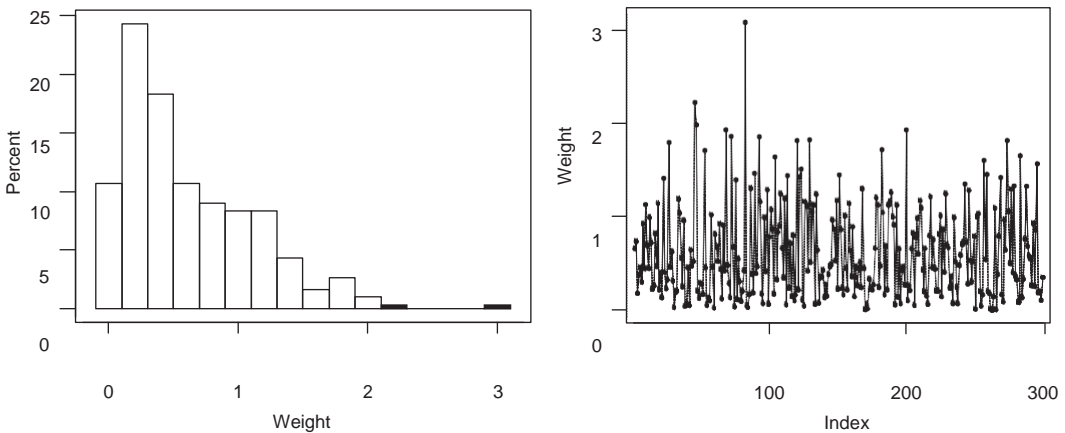
Figure I.3   Histogram and run chart of balance weights in the baseline investigation.

Summarizing the baseline results we have:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
weight	300	0.6215	0.4700	0.5813	0.5079	0.0293

Variable	Minimum	Maximum	Q1	Q3
weight	0.0100	3.0900	0.2100	0.9625

There was one outlier in the baseline sample. If we exclude this rotor, the standard deviation is reduced slightly to 0.49. Ignoring this casting, the full extent of variation of the imbalance weight is about 0 to 2.25. The team set the goal to reduce variation in the balance weight so that at least 75% of the rotors had weight less than 0.5.

The team moved on to the Check the Measurement System stage. They wanted to ensure they had a reliable way to measure balance. There were three gages in parallel used to measure balance weight. See the process map in Figure I.4.

For the investigation, the team selected three rotors with initial measured weights of 0.10, 0.54, and 1.12. They measured the three rotors twice using each of the three gages on three separate days. There was little operator effect since the gages were mostly automated. There were 54 measurements in total. The data are given in the file *brake rotor balance measurement*.

We see in Figure I.5 that the measurement system can easily distinguish the three rotors.

We calculate the summary measures by rotor (part).

Variable	rotor	N	Mean	Median	TrMean	StDev
weight	1	18	0.1083	0.1150	0.1088	0.0685
	2	18	0.5028	0.5050	0.4994	0.0697
	3	18	1.0172	1.0350	1.0175	0.0584

Variable	rotor	SE Mean	Minimum	Maximum	Q1	Q3
weight	1	0.0162	-0.0100	0.2200	0.0500	0.1700
	2	0.0164	0.4000	0.6600	0.4275	0.5600
	3	0.0138	0.9100	1.1200	0.9700	1.0525
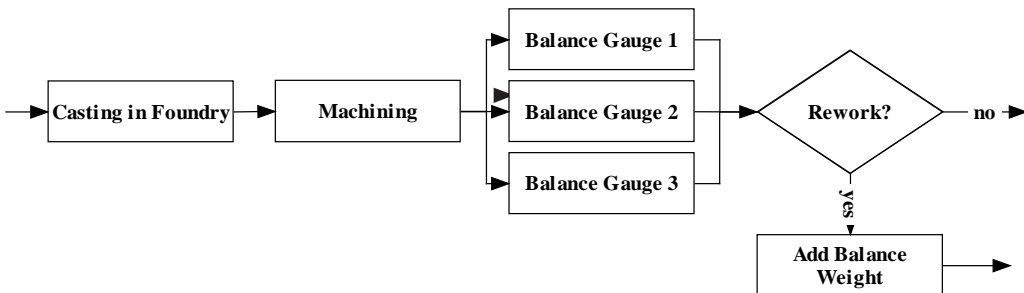
Figure I.4   Brake rotor process map.
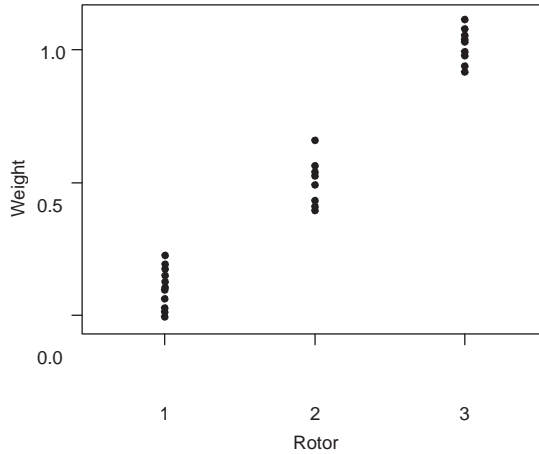
Figure I.5   Brake rotor balance measurement investigation results.

There are no obvious outliers in these data. We estimate the standard deviation of the measurement system as

$$\sqrt{\frac{0.0685^2 + 0.0697^2 + 0.0584^2}{3}} = 0.06$$

The measurement system was judged to be acceptable since this is much less than the overall standard deviation, estimated in the baseline investigation to be 0.51. The discrimination ratio D is

$$\frac{\sqrt{0.51^2 - 0.06^2}}{0.06} = 8.4$$

Next, the team had to select a working approach. They first considered the non-cause-based approaches. They ruled out 100% inspection, since that was the current approach and was too costly. They also eliminated feedback control, since there is no strong pattern in the variation over time in the baseline data and they had no idea of how to adjust the weight. Robustness or Move the Process Center (equivalent approaches in this case) were possibilities but, without more process knowledge, were not likely to succeed. The team decided to search for the dominant cause of variation in the balance weight.

The team first looked at easily available data to see what causes could be eliminated. They recorded, on a defect concentration diagram, the location (in increments of 30°) of the welded rework weight for the 140 balance rejects from the baseline investigation. The dots on the part schematic in Figure I.6 show the nonsymmetrical pattern of balance weights observed. Since the machining process is rotationally symmetric and the casting process is oriented, the team eliminated all causes in the machining operation. With this simple investigation the team made tremendous progress with little cost and in a short time.

The team next planned a group comparison. They selected 30 balance rejects (average weight 1.02) and 30 balanced brake rotors (average weight 0.15) for the comparison. They then measured 26 foundry-determined characteristics on each machined rotor. Note that
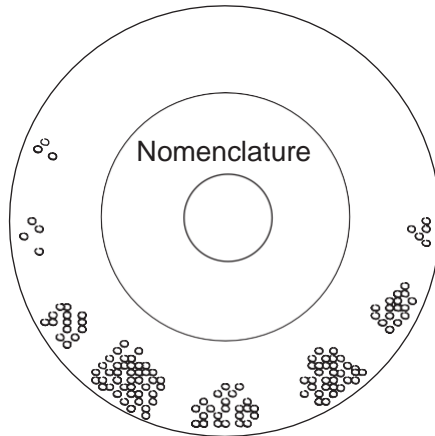
Figure I.6   Concentration diagram of weight locations.

machining destroyed many foundry characteristics. The input characteristics included vein thickness at eight different locations (four orientations at inner and outer edges of the rotor), three derived thickness variation measures (variation at inner and outer locations, and the overall variation), core offset at two orientations, four diameters related to the core size, and five hole diameters (on the inner part of the rotor, see Figure I.1). The data are given in the file *brake rotor balance comparison*.

The team identified two input characteristics, thickness variation and core position (offset), that were substantially different for balanced and unbalanced rotors. See Figure I.7. The results for thickness variation were more compelling than for core position. However, based on engineering knowledge, both inputs were plausible dominant causes of imbalance variation.

The team decided to verify these suspects hoping that they could then reformulate the problem, since the suspects could be measured in the foundry. This would save time and effort in future investigations since they would no longer need to trace rotors between the
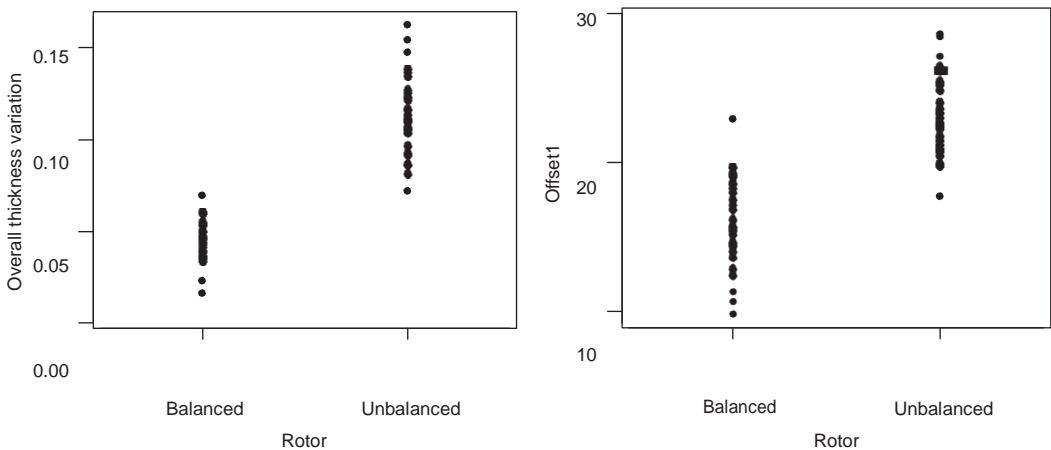
Figure I.7   Plot of thickness variation and offset 1 in balanced and unbalanced rotors.

Table I.1   Brake rotor verification experiment plan and results.

Treatment	Tooling	Core position	Thickness variation	Run order	Average balance weight
1	4-gang	Offset	30-thousandths	8	0.56
2	4-gang	Offset	Nominal	1	0.17
3	4-gang	Nominal	30-thousandths	3	0.44
4	4-gang	Nominal	Nominal	7	0.08
5	6-gang	Offset	30-thousandths	2	1.52
6	6-gang	Offset	Nominal	5	0.37
7	6-gang	Nominal	30-thousandths	4	1.34
8	6-gang	Nominal	Nominal	6	0.03

foundry and the machining operation. They planned and conducted a verification experiment to confirm that core thickness variation and core position were substantive causes of the balance weight variation and that the six-cavity mold was not. They used two levels for each input and a full factorial design. The team selected the nonnominal levels for core position and thickness variation at the high end of their normal range of variation. We give the eight treatments and input levels in Table I.1. For each run, they planned to produce eight castings. The order of the treatments was randomized.

The 64 experimental castings were tagged and shipped to the machining plant to be processed and measured. We give the data in the file *brake rotor balance verification* and the average weight over the eight castings for each run in Table I.1.

We plot the weights by treatment in Figure I.8. We see that some treatments have very little variation and result in a low average weight. We also see roughly the full extent of the variation in weight given by the dashed line on Figure I.8.
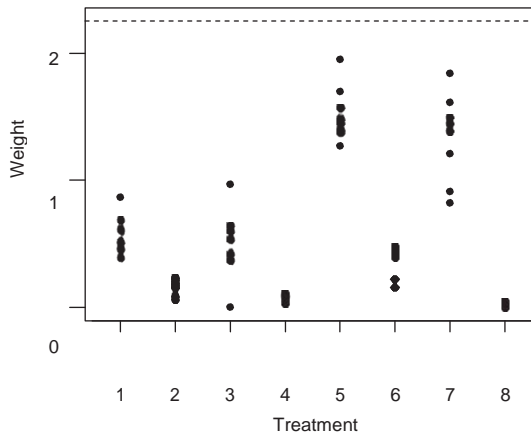
Figure I.8   Plot of weight by treatment.
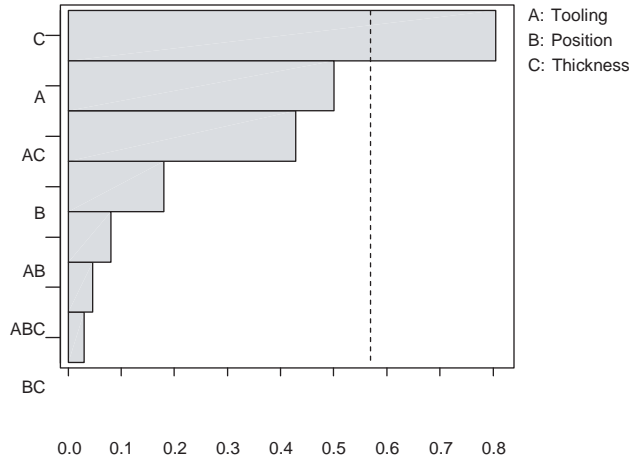
Figure I.9   Pareto chart of the effects for brake rotor verification experiment.

From the Pareto plot of the effects, given in Figure I.9, we see the effect of core position is small, so we eliminate it as a suspect. Since the tooling and thickness variation have a relatively large interaction, we look at the effect of these two suspects simultaneously.

We give the main and interaction effects plots in figures I.10 and I.11, respectively.

From the interaction plot, the team concluded that low thickness variation using the four-cavity mold produced the optimal results (the weights required were so small that the balance specification was met without rework). Thus, the dominant cause of the imbalance problem was in the core molding process.

This search for the dominant cause is summarized in the diagnostic tree shown in Figure I.12.

The team made the obvious fix and recommended that the foundry go back to the original four-cavity core mold. When this change was implemented, the rate of balance rejects immediately dropped to its historical levels. The team had met the original project goal.

The major lesson learned in the project was the effect of the thickness variation on the balance weight. The verification experiment showed that thickness variation in the cores was a dominant cause of balance weight variation in the *original process* that used the four-cavity
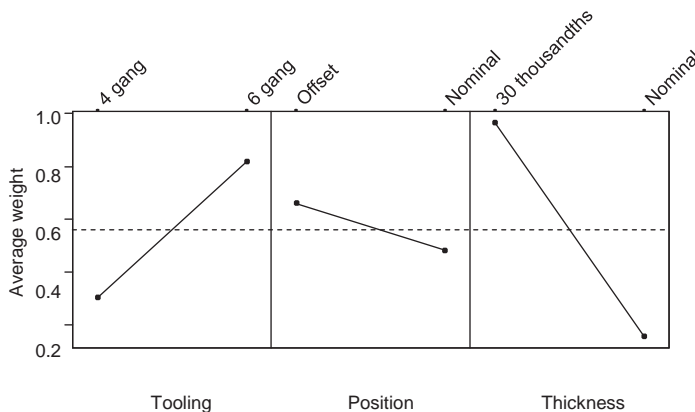
Figure I.10   Main effects plot for brake rotor verification experiment.
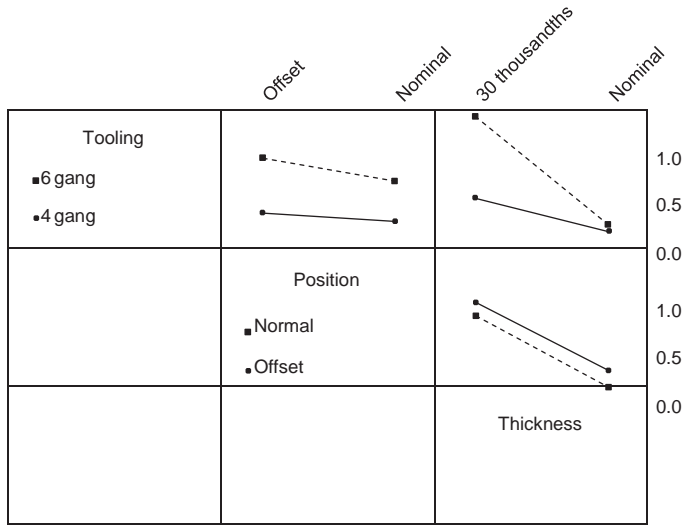
Figure I.11   Interaction plot for brake rotor verification experiment.

mold. The team was puzzled about the interaction between the thickness variation and the number of cavities in the mold. Knowledge of a dominant cause provided the opportunity to improve the process further. There was no immediately known method for adjusting the core-making process to reduce thickness variation. The team considered looking for an adjuster in the core-making process but rejected this approach since they did not expect to be successful.
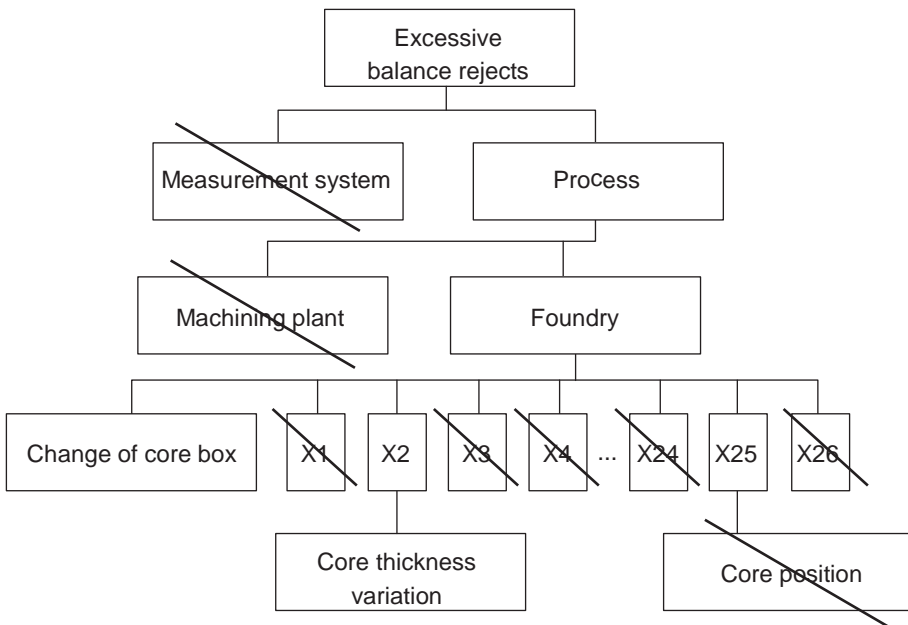
Figure I.12   Summary of the method of elimination for brake rotor example.

The team had the opportunity to implement a new core-making process for the veined rotor. The equipment was already available in the plant but not in use. The team knew that the cold box process was dimensionally stable, and they expected much less thickness variation with this process. This is the Fix the Obvious approach. With the implementation of the cold box method, the process was greatly improved. Over the next four months the rate of balance rejects dropped to 0.2%, a large reduction from the 50% at the start of the project. The machining plant eliminated the expensive rework stations and scrapped the few balance rejects in the new process.

Highlights

Some strengths and weaknesses of this case are:

- The use of the available data together with the knowledge of the symmetry of the machining operation to eliminate all causes in the machining plant.

- In the group comparison, the team could have selected two smaller sets of rotors that were more extreme with respect to the balance weights.

- The carefully planned and conducted verification experiment.

- The application of the knowledge gained about thickness variation (the dominant cause) to select the cold box process.

Case Study II

Rod Thickness

The ancestor of every action is a thought.

—Ralph Waldo Emerson, 1803–1882

A plant manufactured approximately 12,000 connecting rods per day for use in an engine assembled in the plant. The rod, illustrated in Figure II.1, connects the piston (at the small or pin end of the rod) to the crankshaft (at the large or crank end of the rod). The plant received forged blanks and machined the rods in a large number of process steps.

Management identified the rod line for a variation reduction project because the overall scrap cost was greater than budget. The yearly scrap cost was in excess of $400,000, and the scrap rate was 3.2% over the previous four months. Management set a goal to reduce the scrap rate to 1.6% in its annual business plan. The rod line production manager assigned a team to the project.

Looking at scrap records, the team found that scrap occurred at several stages in the process and for several reasons. To focus the problem, they used Pareto analysis on the records for one month. The results, in Figure II.2, showed that 65% of the scrap occurred at a grinding operation. At this operation, the team discovered that about 90% of the scrap was due to rods with their crank end thickness less than specification. The team focused their attention on reducing variation in rod thickness.
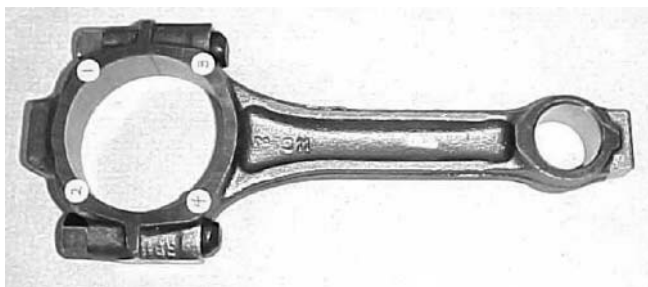
Figure II.1   A connecting rod.

Defect	Grind	Bore	Broach	Assembly	Others
Count	85	24	14	6	2
Percent	64.9	18.3	10.7	4.6	1.5
Cum%	64.9	83.2	93.9	98.5	100.0
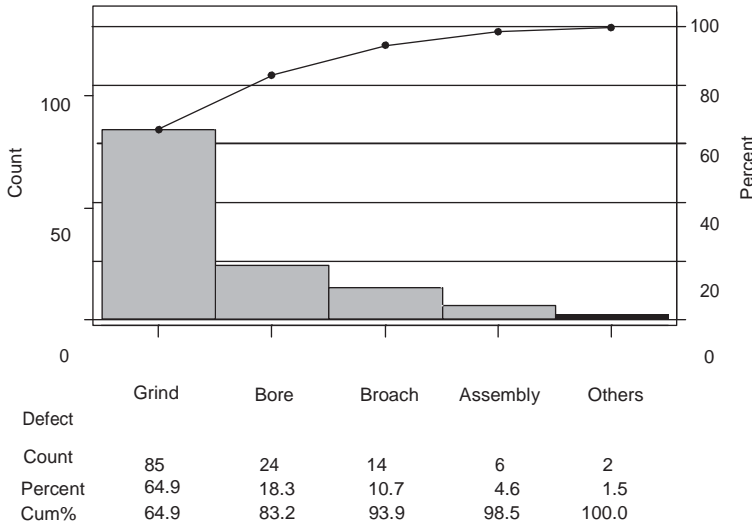
Figure II.2   Pareto chart for scrap by operation for the rod line.

At the grinder, the final thickness of the rod was set in a two-pass operation. An operator loaded the parts into the machine. In the first pass, one side of the rod was ground in three steps. The rod was then turned over by the machine operator, and the second side was ground. The grinder had a rotary table with 20 pallets that passed under four different grinding wheels. An internal control system automatically adjusted the grinding wheels based on thickness measurements taken by a series of gages internal to the grinder.

After grinding, an in-line gage measured the thickness of every rod at four positions (given by the white circles on the crank end of the rod, also faintly numbered 1 through 4, as shown on Figure II.1). The specifications were 0.91 to 0.96 inches at each location. The gage automatically rejected to a rework station a rod with any of the four thickness measurements not meeting the specification. At the rework station, an operator remeasured the rejects using a different gage and scrapped undersized rods.

The rod line ran on three shifts, five days per week. To establish the baseline, the team sampled 200 rods chosen in two batches of 20 rods for each of five days on the day shift only. They recorded the thickness measurements for the four positions using the in-line gage that did not normally store the data. We give the data in the file *rod thickness baseline*. Thickness is given as a deviation from 0.9 in thousandths of an inch.

Figure II.3 gives a histogram of thickness, where the dashed vertical lines are the specification limits. Of the 200 rods, 10 rods had thickness values less than the lower specification limit for at least one of the positions. This was somewhat higher than expected, given the historical scrap rate. There are no obvious outliers. A numerical summary of the baseline data (across all positions) is:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
thickness	800	34.575	36.000	34.840	11.023	0.390

Variable	Minimum	Maximum	Q1	Q3
thickness	2.000	59.000	28.000	43.000
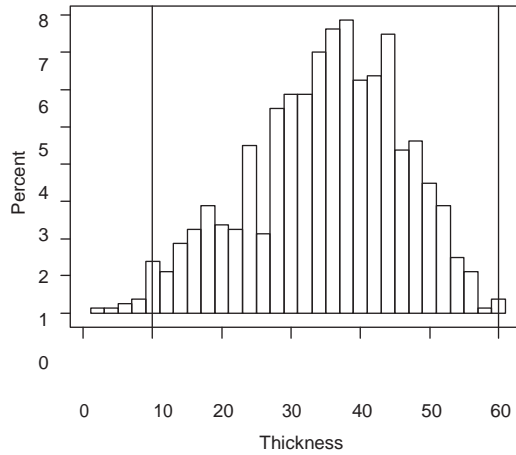
Figure II.3   Baseline histogram of rod thickness (dashed vertical lines give the specification limits).

The team set the objective to produce all rods within the thickness specification. If this ambitious goal could be achieved, they would eliminate 90% of the scrap at the grinder or 58% of the total rod line scrap, hence meeting the project goal.

The process was well centered, so to meet the goal the team needed to reduce the standard deviation to 8.5 from around 11 thousandths of an inch, while keeping the process centered on target. The full extent of the thickness variation was 2 to 59 thousandths. In Figure II.4 we show the baseline data over time (batch), which suggests there was no systematic drift in the process.

The next step was to assess the measurement system. The in-line gage used four sets of transducers to measure rod thickness at the four positions. In effect there were four gages. The team selected three rods to roughly cover the full extent of the thickness variation.

The team was concerned about a relative bias among the four gages. Since it was impossible to measure the same position on the different gages, they sent the three rods to
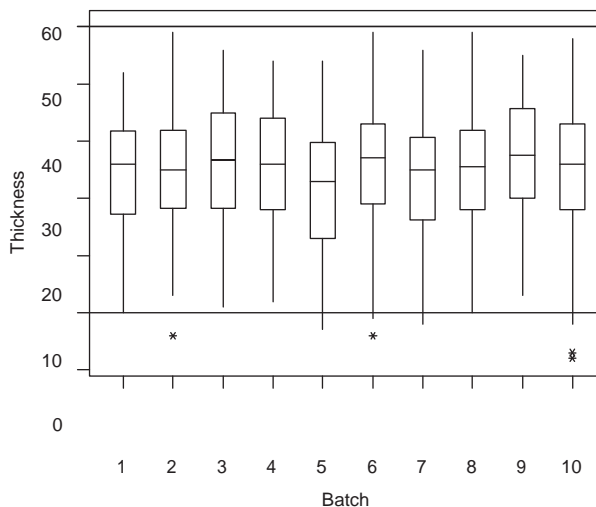
Figure II.4   Box plot of rod thickness by batch for baseline investigation (dashed horizontal lines give the specification limits).

Table II.1   True values of rod thickness by position.

Rod	Position 1	Position 2	Position 3	Position 4
1	54.6	44.7	32.4	59.0
2	22.6	33.8	11.7	22.9
3	38.3	40.6	21.4	41.8

a precision laboratory to have the thicknesses determined with little measurement bias and variation. The precision lab values are given in Table II.1.

The team chose two operators, one from the day shift and one on the afternoon shift. For three days, each operator measured the three rods three times at each of the four positions. In total, there were 216 measurements, 56 per gage. Since we know the true values, we give the measurement errors in the file *rod thickness measurement*.

A summary of the measurement errors for each position (gage) are:

Variable	position	N	Mean	Median	TrMean	StDev
error	1	54	0.037	0.400	0.021	1.564
	2	54	–0.033	0.250	–0.013	1.384
	3	54	–0.019	–0.400	–0.013	1.326
	4	54	0.081	0.100	0.075	1.775

Variable	position	SE Mean	Minimum	Maximum	Q1	Q3
error	1	0.213	–3.300	5.400	–0.775	1.400
	2	0.188	–3.700	3.300	–0.800	0.400
	3	0.180	–3.400	2.600	–0.700	0.775
	4	0.242	–3.800	4.000	–1.000	1.125

There is no evidence of substantial bias in any of the four gages, since the average errors are close to zero.

Because there is no relative bias among the gages and we have the measurement errors, we estimate the measurement system variation by the standard deviation of the 216 measurement errors. As given, we obtain the estimate 1.512.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
error	216	0.017	0.200	0.012	1.512	0.103

Variable	Minimum	Maximum	Q1	Q3
error	–3.800	5.400	–0.800	1.100

The baseline variation was 11.023. We can estimate the process variation, using Equation (7.2), by $\sqrt{11.023^2 - 1.512^2} = 10.92$ and the discrimination ratio as $D = 10.92 \div 1.512 = 7.2$. The team decided that the in-line gage was not a dominant cause of the variation and was adequate to move to the next stage of the Statistical Engineering algorithm.
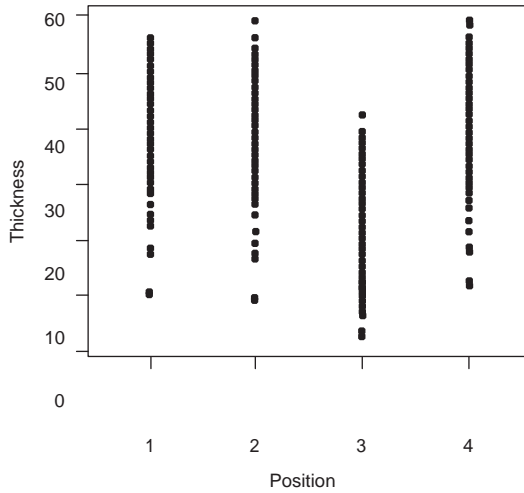
Figure II.5   Box plot of thickness by position from the baseline data.

When considering the choice of a working variation reduction approach, the team quickly rejected the approaches that did not require knowledge of a dominant cause. The process was already well targeted, there was no evidence of an exploitable time pattern in thickness, 100% inspection was already present, and making the process robust seemed difficult. The team decided to search for a dominant cause.

The team first looked at the available data from the baseline investigation to see what family of causes could be eliminated. We show a box plot of thickness by position in Figure II.5. There is a difference between position 3 and the other positions. However, the variation within positions was close to the full extent of variation, so position is not the dominant cause. All but one of the scrapped rods in the baseline sample was undersized at position 3. The team did not know of an obvious fix to move the thickness center for position 3.

The team decided to carry out a multivari investigation to compare pallet-to-pallet, position-to-position, day-to-day, and part-to-part families. They planned to select three consecutive rods from six different pallets (pallets occur in pairs, so three pairs of pallets were chosen) on four different days. They measured thickness at all four positions to confirm the findings from the baseline investigation. In total, they sampled 72 rods and made 288 measurements. The data are given in the file *rod thickness multivari*.

We show the results of the multivari analysis for the pallet-to-pallet, position-to-position, and day-to-day families in Figure II.6. We see that none of these families is home to a dominant cause, although again the thickness was lowest at position 3. To examine the part-to-part family, we form a new input group corresponding to all 96 combinations of day, position, and pallet. We use ANOVA to isolate the variation within the groups. The results are:

Analysis of Variance for thickness

Source	DF	SS	MS	F	P
group	95	14444.32	152.05	3.93	0.000
Error	192	7433.33	38.72		
Total	287	21877.65			

We estimate the within group variation due to causes in the part-to-part family as $\sqrt{38.72} = 6.22$, a substantial component of the baseline variation 11.023. The team concluded:

- There was substantial part-to-part variation within the same pallet, position, and day.

- Position 3 was systematically thinner than the other positions.

However, later, when the team examined the histogram of the multivari data in Figure II.6, they noticed that the range of variation in thickness from the multivari did not cover the full extent of variation. There were no undersized rods with thickness values below the lower specification limit of 10. As a result, they decided that their conclusions from the multivari were premature.

This was an unexpected result. The plan for the multivari investigation matched the baseline investigation quite closely. By sampling rods over a number of days, the team had hoped that they would observe undersized rods. The lack of undersized rods suggested that either the dominant cause acts in pallets not chosen for the multivari or that it acts only occasionally in time. Based on this thinking, the team decided to conduct another multivari investigation.
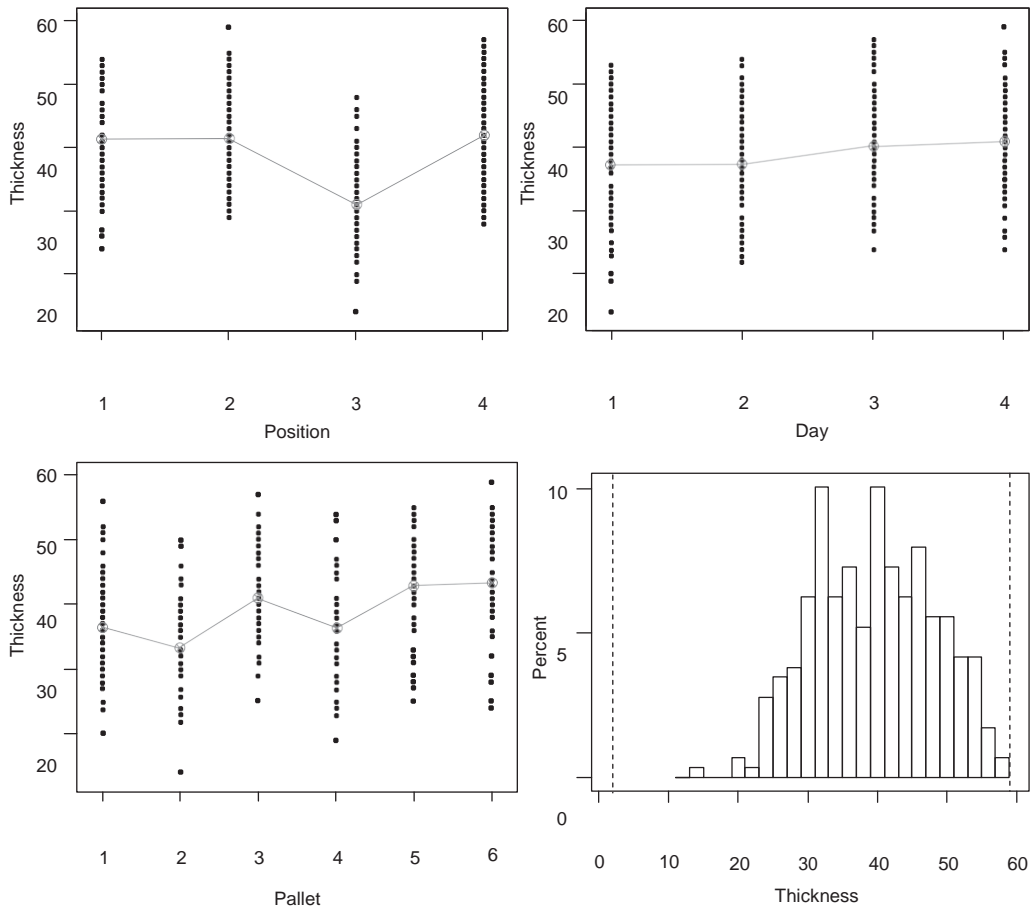
Figure II.6   Summary of the data from the first multivari investigation (vertical dashed lines give the full extent of variation).

In the second multivari, they planned to sample three consecutive rods from a different set of six pallets than those used in the first multivari investigation. This time they repeated the sampling scheme five times within a single shift. This shorter time frame seemed sufficient from the baseline results, as shown in Figure II.4. In total, the team sampled 90 rods and made 360 measurements. The data are given in the file *rod thickness multivari2*.

The team first checked that their new sampling plan generated the full extent of thickness variation. A histogram of the new multivari data, shown in Figure II.7, shows that the full extent of variation was captured. The dominant cause of thickness variation must have acted during the course of the investigation.

We can see the effects of the families of variation for time, pallet, and position in the one-input-at-a-time multivari charts in Figure II.8. We see that a dominant cause of thickness variation is acting time to time. Within each time period, the variation is substantially less than the full extent of variation. All of the undersized rods occur at time 3. The differences among the positions is smaller than previously.

We expect causes in the rod-to-rod (that is part-to-part) family to act haphazardly. To examine the rod-to-rod family, we define a new input group that uniquely numbers the 120 combinations of position, time, and pallet. The within-group standard deviation (Pooled StDev in the MINITAB one-way ANOVA) is 4.29. The rod-to-rod family is not dominant. During the multivari investigation, 40 undersized rods were produced in a matter of two minutes at time 3. The tool setter suspected that there was a problem with the feedback control scheme that automatically adjusted the grinding wheels. For the finishing wheel, the controller used the measurements from an internal gage and the following rules:

- Measure every part at all four positions.

- If the thickness at any position for two consecutive parts exceeds 50-thousandths of an inch, lower the finishing grinding wheel for three seconds at a fixed rate.

- Ignore the measurements from the next three parts until the adjustment has taken effect.
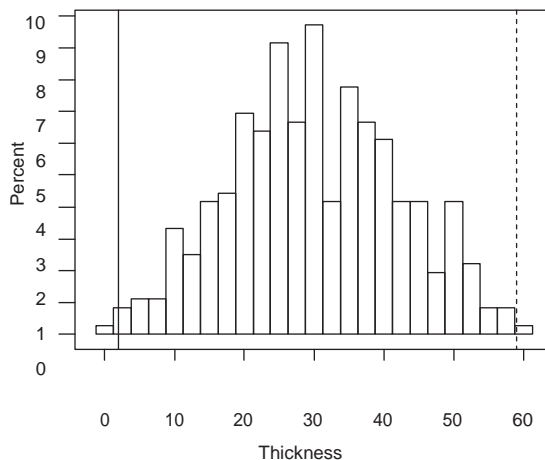
Figure II.7　Histogram of data from the second multivari investigation.
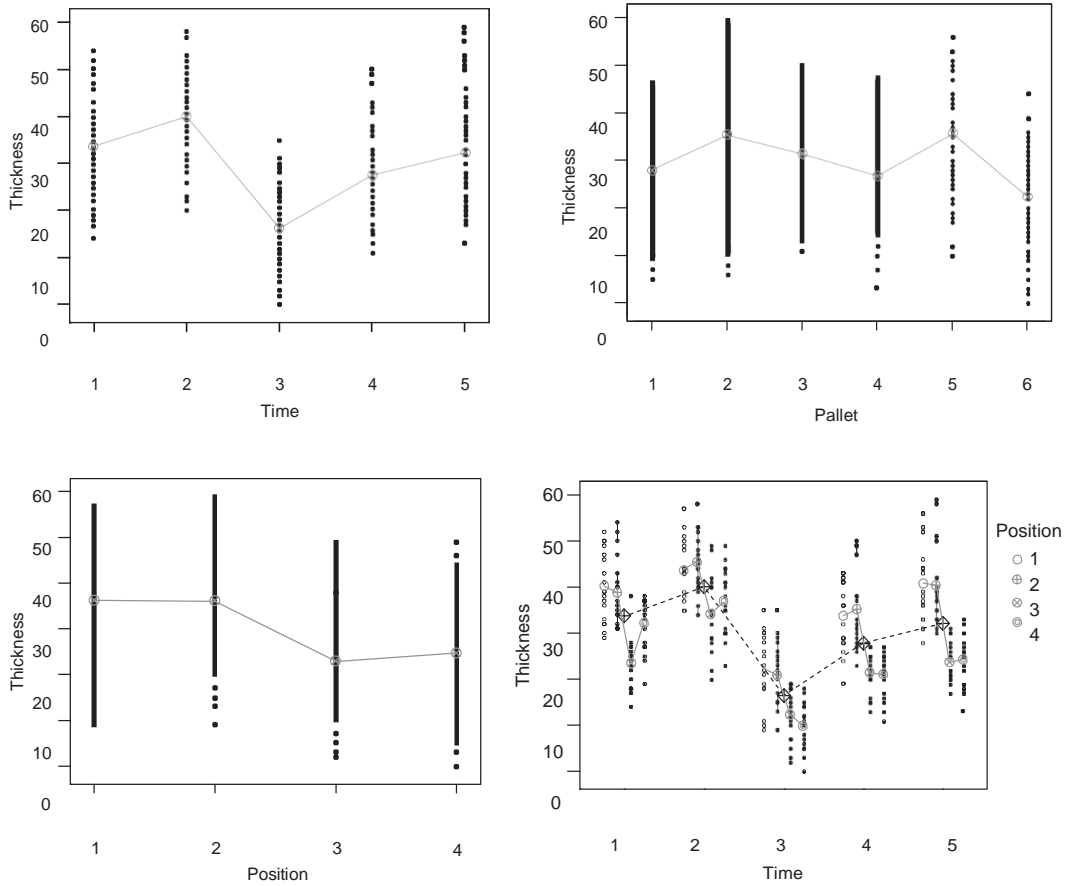
Figure II.8   Multivari charts using data from the second multivari investigation.

The last rule was necessary because of the time delay between the grinding by the finishing wheel and the measurement of thickness by the internal gage. The rule was based on the feedback timer and did not count parts. Upon closer inspection, the team discovered that, occasionally, the last rod ground before compensation was treated as the first rod after the wheel was lowered, thereby fooling the equipment into thinking a second compensation was needed. Combined with the systematic difference among the positions, the double compensation was a dominant cause of undersized rods.

The team adopted the Fix the Obvious approach. They adjusted the logic of the controller to count the parts processed after compensation instead of using the timer. This change prevented the double compensation. The team also looked for ways to adjust the process to better center the four positions. They initiated maintenance on the alignment of the grinding wheels to reduce the position-to-position differences. Since they expected the equipment to deteriorate over time, they set up a monitoring procedure to detect when the position-to-position differences became large. Each day, the operator recorded the thicknesses for five consecutive parts and plotted the position averages on a run chart. The process engineer checked the chart on a regular basis to see if there were systematic differences among the four positions.

In the validation stage, the team assessed the effect of the process changes by comparing performance before and after the changes. We give a histogram of thickness values observed when the plan for the original baseline investigation was repeated in Figure II.9.
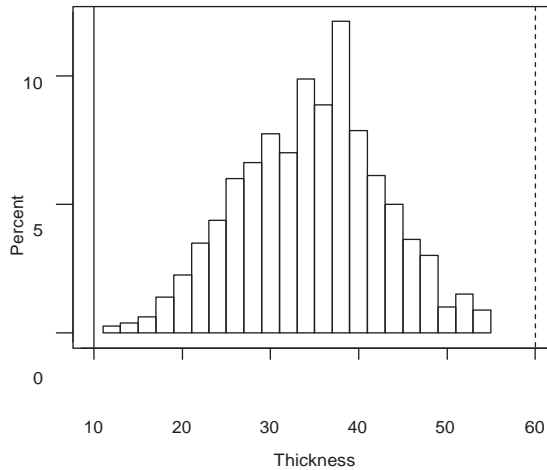
Figure II.9   Histogram of rod thickness from validation investigation.

The data are given in the file *rod thickness validation*. No undersized rods were observed in the validation investigation.

   The scrap rate substantially decreased. In the first month after eliminating the double compensation, the scrap rate was 1.7%. The project team fell short of the initial goal, but the project was judged to be successful by the rod line management.

Highlights

Some strengths and weaknesses of this case are:

- The team did a good job of focusing the problem, linking the problem goal "eliminate undersized rods" to the project goal "reduce rod line scrap by at least half."

- The team adopted a good solution to the assessment of the four gages in the measurement system investigation. In this case, they could not measure the same part/position on each gage, so there was no way to tell if there were relative biases among the gages without measuring rods with known thicknesses. Note that a relative bias could have explained why the average thickness for position 3 was smaller than the other positions.

- The team did not make the best use of the information from the baseline investigation. Given that the full extent of thickness variation occurred within a single day (actually a shift), they should have planned the first multivari within a single shift. The time-to-time variation is captured by repeatedly sampling within the shift, as in the second multivari investigation.

- The team made the mistake of not initially checking that the first multivari investigation captured the full extent of thickness variation.

- The team was wise to change the control plan to monitor the process for position-to-position differences, since these were expected to recur, given the nature of the grinding process.

Case Study III

Crankshaft Main Diameter

The true creator is necessity, which is the mother of our invention.

—Plato, 427–347 B.C.

An engine plant machined approximately 1500 V8 crankshafts per day on three shifts. There was 100% inspection at an automated final gage that measured more than 60 characteristics to ensure that the customer, the engine assembly operation, received a high-quality product. We show a schematic of the crankshaft in Figure III.1.

At the beginning of this project, the monthly scrap rate ranged between 6 and 8%, averaging 7.2% over the previous four months. The first-time-through rate, the ratio of the number of parts that were accepted by the final gage to the total number of parts processed, was highly variable, and sometimes as low as 20%. Parts rejected at the final gage were scrapped or reworked and remeasured using an off-line gage.

The goals of this project were to reduce the overall scrap rate to 4.5% or less and increase the first-time-through rate at the final gage to at least 75%. Management hoped to achieve these goals without any substantial capital expenditure.
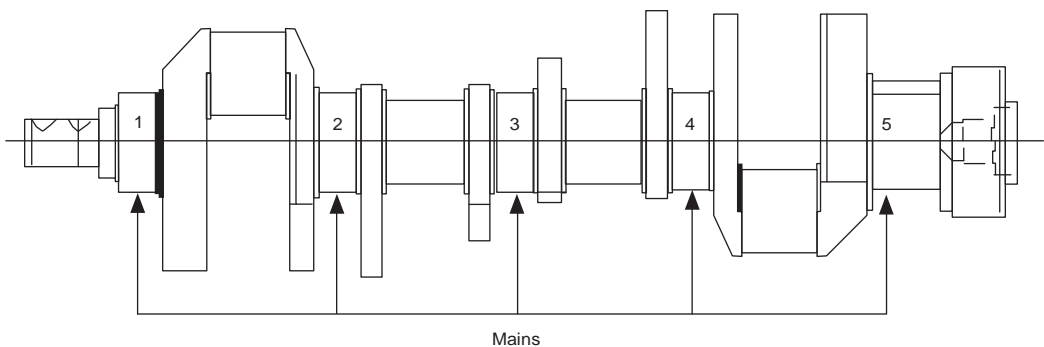
Figure III.1   V8 crankshaft showing five mains.

The team began by determining a more specific goal that could support the overall project goal. The reason was recorded for every crankshaft rejected by the final gage. The team used Pareto analysis to determine that about 85% of rejects and 73% of the scrap were related to the main diameters. The five main diameters are numbered in Figure III.1.

The final gage measured the diameters at three positions—front, center, and rear—on each of the five mains. The specification limits for diameter were ±4.0 thousandths of an inch measured from nominal. The rejects related to main diameter were due to undersized, oversized, and excess taper from the front to the rear of the main. A crankshaft with any main diameter less than –4.0 was scrapped. Taper was the difference between front and rear diameters on each main. The taper specifications were ±2.0 thousandths of an inch. Most parts rejected for taper could be reworked.

To achieve the project goal, the team decided to concentrate efforts on reducing variation in main diameter to eliminate scrap due to undersized diameters and to reduce rework due to taper and oversized diameters.

The team extracted baseline data on main journal diameters for 2000 crankshafts over four days (500 per day selected haphazardly throughout the day) from the final gage. The data are given in the file *crankshaft main diameter baseline*. Because of the large number of measurements, the data are stored in three different formats on the same worksheet:

- In the first 21 columns by crankshaft
- In the next three columns by individual measurements for diameter
- In the last two columns by individual measurements for taper

In the baseline investigation, the scrap rate was 6.9%, with 53% rework, of which 40% was due to excess taper. We show the histograms of all measured diameters and calculated taper values for the 2000 crankshafts in Figure III.2.
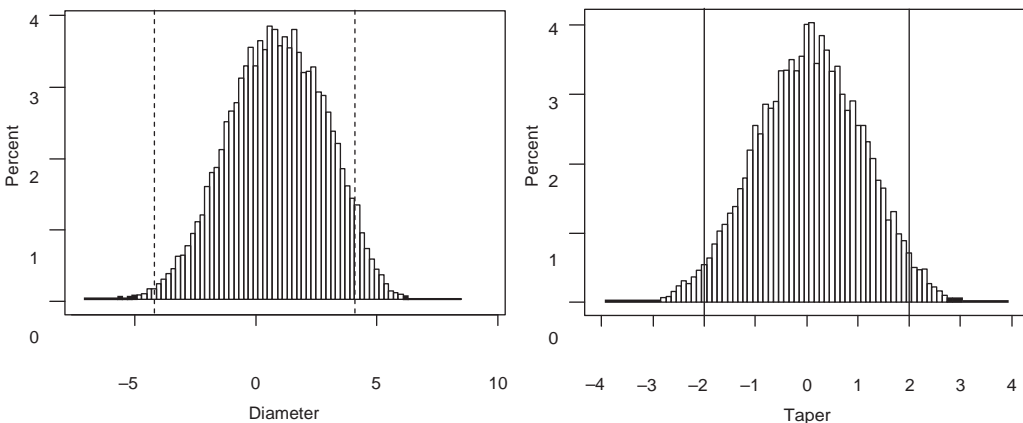
Figure III.2   Baseline histograms of main diameters and taper over all mains and positions (dashed vertical lines show the specification limits).

There are many diameters and taper values outside the specification limits. Numerical summaries of the baseline data are:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
diameter	30000	0.8810	0.9037	0.9005	1.9933	0.0115
taper	10000	0.0619	0.0686	0.0652	1.0367	0.0104

Variable	Minimum	Maximum	Q1	Q3
diameter	–6.9285	8.1401	–0.5168	2.3502
taper	–3.9438	3.5072	–0.6595	0.7870

The team saw that if the process was centered on target, they needed to reduce the standard deviation of diameter to less than 1.30 to meet their goal of producing no diameters out of specification. The full extent of diameter variation was about –5.1 to 6.9 (0.88 ±3*1.99) thousandths of an inch.

There are so many diameter observations that it is difficult to look for patterns over time. We give the box plot of diameter by day in Figure III.3. The team also looked at plots of diameter over time for each of the individual positions. There are no day-to-day patterns. The next step was to assess the final measurement system. There were 15 different gages and the team could not measure the same part on each of them; for example, they could not measure the number 1 main front diameter on the number 3 rear gage. These no-contact gages used a common air pressure system. The team selected six different crankshafts to be measured three times each on three days. The team chose six crankshafts to roughly cover the full extent of variation because here they were simultaneously investigating 15 different measurement systems. It was not worth the effort to find three crankshafts that gave the full extent of diameter variation for all 15 positions simultaneously. There was no operator effect because the measurement process was automated. There were a total of 54 measurements on the front, center, and rear positions for each of the five mains. The data are given in the file *crankshaft main diameter measurement*.
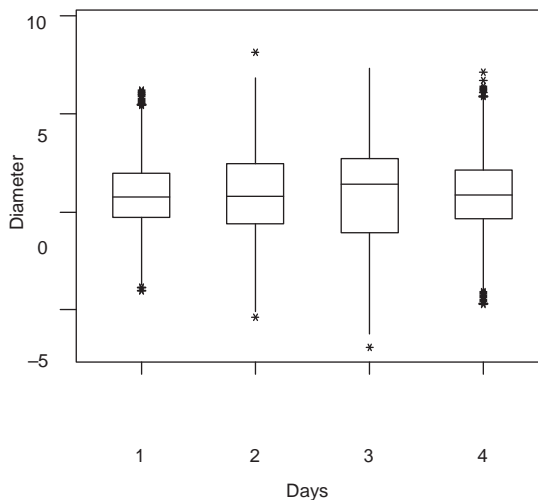
Figure III.3   Box plots of final diameter by day.

The estimates of the measurement variation (standard deviation) of the 15 gages ranged from 0.28 to 1.06. The MINITAB ANOVA results for the two extreme cases are:

Analysis of Variance for 1rear

Source	DF	SS	MS	F	P
crankshaft	5	125.43	25.09	22.37	0.000
Error	48	53.82	1.12		
Total	53	179.25			

```
                                     Individual 95% CIs For Mean
                                     Based on Pooled StDev
Level     N      Mean    StDev    ---+---------+---------+---------+---
1         9     4.450    0.990                           (----*---)
2         9     1.697    1.148              (----*---)
3         9     0.265    1.035    (----*---)
4         9     4.326    1.055                         (---*---)
5         9     2.237    1.050                (---*---)
6         9     3.763    1.070                    (----*---)
                                     ---+---------+---------+---------+---
Pooled StDev =    1.059              0.0       1.6       3.2       4.8
```
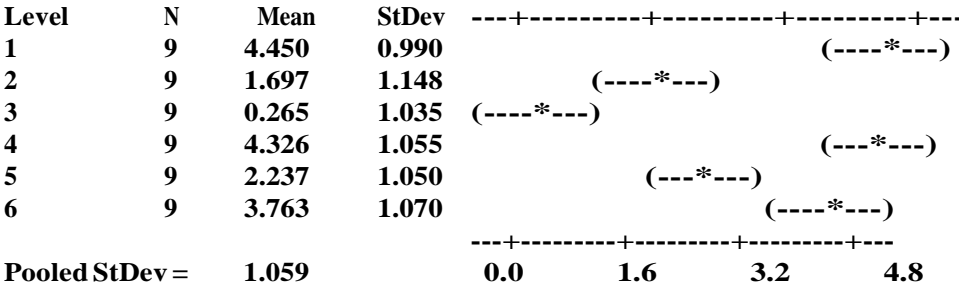
Analysis of Variance for 4front

Source	DF	SS	MS	F	P
crankshaft	5	59.9397	11.9879	153.84	0.000
Error	48	3.7403	0.0779		
Total	53	63.6800			

```
                                     Individual 95% CIs For Mean
                                     Based on Pooled StDev
Level     N      Mean    StDev    -+---------+---------+---------+----- 1    9
               -0.1608  0.2674    (–*)
2         9    –2.1253  0.1886                  (*–)
3         9    –0.8184  0.2653                           (*–)
4         9    –1.4794  0.3258                     (–*)
5         9    –1.9434  0.3204                  (–*)
6         9    –3.4972  0.2850    (–*)
                                     -+---------+---------+---------+-----
Pooled StDev =    0.2791           –3.6      –2.4      –1.2       0.0
```
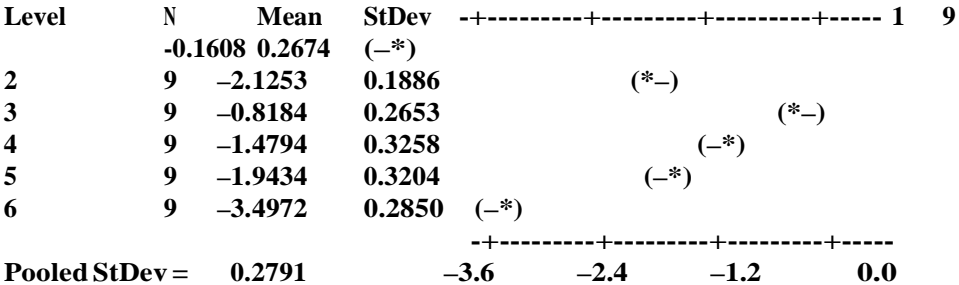
Because there were 15 gages, there may have been relative biases among the gages. These biases would contribute to the variation in diameter across all mains as determined in the baseline. We cannot assess these biases with the data from the measurement system investigation. The team was concerned about how to determine if the measurement system was a dominant cause of the overall variation. They looked at the box plot of diameter by position from the baseline investigation, given in Figure III.4, and noted that there were systematic position-to-position differences.

These differences could be due to the process or the gages. The team decided to assess each gage against the baseline variation within its own position. We give the results in Table III.1.
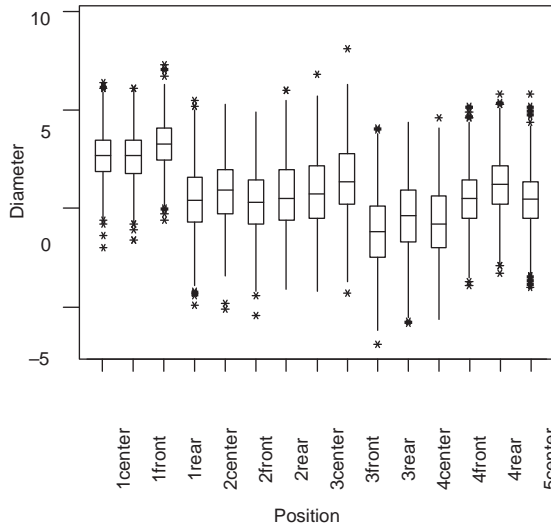
Figure III.4   Diameter by position from baseline data.

Table III.1 Measurement discrimination ratios by position.

Position	Baseline standard deviation	Measurement system variation	Discrimination ratio
1 center	1.17	0.96	0.71
1 front	1.20	0.97	0.72
1 rear	1.18	1.06	0.51
2 center	1.65	0.86	1.65
2 front	1.64	0.76	1.89
2 rear	1.65	0.83	1.70
3 center	1.80	0.86	1.84
3 front	1.78	0.92	1.67
3 rear	1.78	0.85	1.84
4 center	1.78	0.28	6.18
4 front	1.77	0.33	5.28
4 rear	1.78	0.29	5.95
5 center	1.46	0.69	1.88
5 front	1.45	0.56	2.37
5 rear	1.45	0.74	1.69

For some positions, the measurement system was a dominant cause of the variation. This was a shocking discovery, since the gage was thought to be reliable based on the required gage R&R investigations. The team needed to improve the measurement system immediately. The small discrimination ratios helped to explain why some operators remeasured scrapped crankshafts a second time. Given the observed measurement variation, it was not surprising that these crankshafts were sometimes found to be acceptable. Surprisingly, no one found the need to remeasure crankshafts that were first time passes through the measurement system.

To improve the final gage, the team decided to look for a dominant cause of measurement variation. They started with the existing data from the measurement investigation. They examined the data from positions 4 front (one of the best gages) and 1 rear (one of the worst) in more detail. In Figure III.5, we give plots of the estimated measurement errors (measurements minus the position average) by day. We see for 1 rear that the measurement error is substantially different from one day to the next. The dominant cause of the measurement variation for position 1 rear acted in the day-to-day family.

To explore this behavior further, the team measured the same crankshaft once per day for 19 days. The data are given in the file *crankshaft main diameter measurement stability*. The results were striking and surprising. We present results only for the front position for all five mains here. The results were similar for the other positions. In Figure III.6, and in the subsequent numerical results, we see that the final gage was unstable. The day-to-day variation in the gage was much greater than expected based on the existing short-term R&R results.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
1front	19	3.380	3.381	3.351	0.943	0.216
2front	19	1.599	1.309	1.583	1.255	0.288
3front	19	–1.319	–1.205	–1.305	0.778	0.178
4front	19	–1.553	–1.814	–1.573	0.712	0.163
5front	19	2.790	2.721	2.794	0.794	0.182
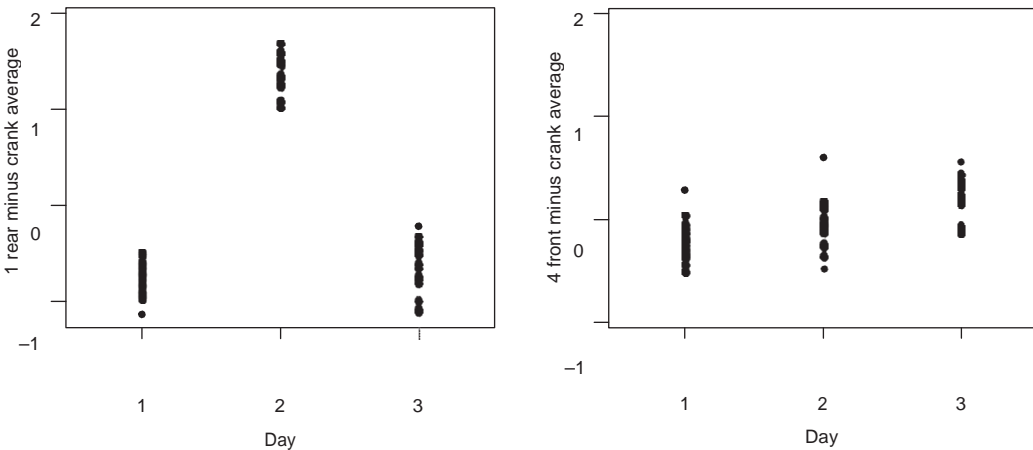
Figure III.5   Diameter minus main average by day.

Variable	Minimum	Maximum	Q1	Q3
1front	1.946	5.301	2.656	3.985
2front	−0.428	3.915	0.730	2.511
3front	−2.905	0.018	−1.827	−0.730
4front	−2.402	−0.355	−2.188	−0.858
5front	1.438	4.071	2.184	3.485

The team decided to address the measurement system instability with feedback control. The controller was feasible because:

- Quick adjustment of the measurement system was available by changing an offset.

- The shifts in the measurement process were persistent, as seen in Figure III.6.

The team developed a procedure to monitor the stability of the gages and a reaction plan. Each day, an operator measured a reference part and plotted the results on a control chart, one for each of the 15 positions. The centerlines and adjustment limits were based on the within-day measurement variation (that is, short-term variation) from the initial measurement investigation. If a plotted point fell outside the control (adjustment) limits, the gage was cleaned and remastered. The control charts provide an ongoing record of the performance of the gages. If the gages performed consistently over time on the reference part, the team had confidence in using data from the gages to make process decisions. After the feedback system was implemented, the team validated the improvement by repeating the initial measurement investigation. They found that discrimination ratios, calculated as in Table III.1, all exceeded 3.0.

The team was ready to proceed to the next step of the Statistical Engineering algorithm, confident in the measurement system and knowing they had already improved the process substantially. They were tracking the first-time-through rate at the final gage and saw a marked increase. The team considered the possible variation reduction approaches and decided to look for a dominant cause of the variation in main diameter.
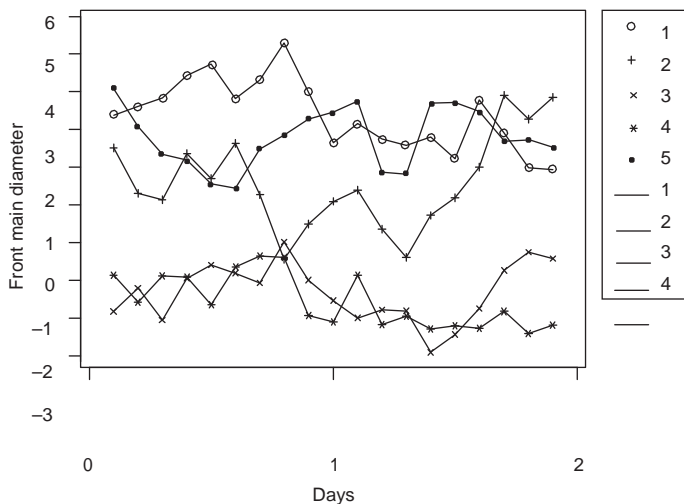
Figure III.6 Front position diameter for the five mains when measuring the same crankshaft.

Because of the faulty measurement system, they debated whether they should repeat the baseline investigation. They decided to proceed without doing so. They continued to use the original full extent of variation, –5.1 to 6.9, recognizing that this was likely too wide given the improvements they had made.

We show a simplified flowchart of the crankshaft production system in Figure III.7, highlighting the operations that affected the main diameters. The team believed initially that most of the variation in main diameters at the final gage was caused by differences in the four journal grinders. There were little data to support this belief.

The team decided to explore the families of variation defined by the process steps using a variation transmission investigation. The main diameters were measured with an in-process contact gage between the grinding and lapping operations. The team planned to measure a sample of parts with the in-process gage and then track the parts through the process and remeasure them at the final gage. They could use this variation transmission investigation to eliminate the lapper or the upstream process as the home of the dominant cause.

Based on their experience with the final gage, the team decided to next investigate the in-process gage. They carried out an investigation of this system with the same plan used for the final gage. We do not give the data here. They found the measurement variation to be relatively small (estimated standard deviation 0.31) and no evidence of instability.

The team knew that both the in-process and final measurement systems were used in an informal way to control the grinders. For that reason, the team next checked the relative bias of the two systems. They measured six crankshafts on the final gage. Then, they returned the parts to the in-process gage and remeasured the diameter. Using the same six parts, they repeated this process two more times. In the data, the intermediate (in-process) diameters have been increased by 2.25 units to reflect the fact that the nominal diameter is different at the two gages by 2.25 thousandths. The data are given in the file *crankshaft main diameter relative bias*.
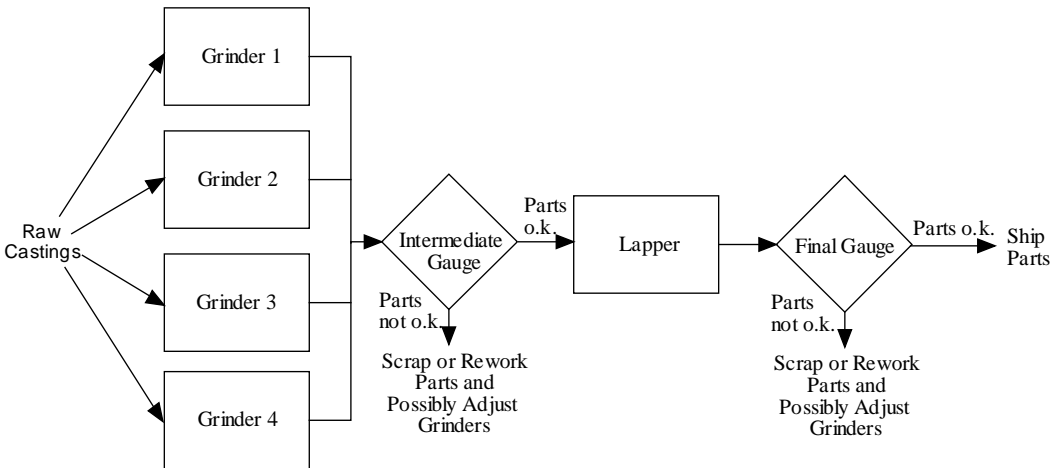
Figure III.7   Crankshaft production process map.
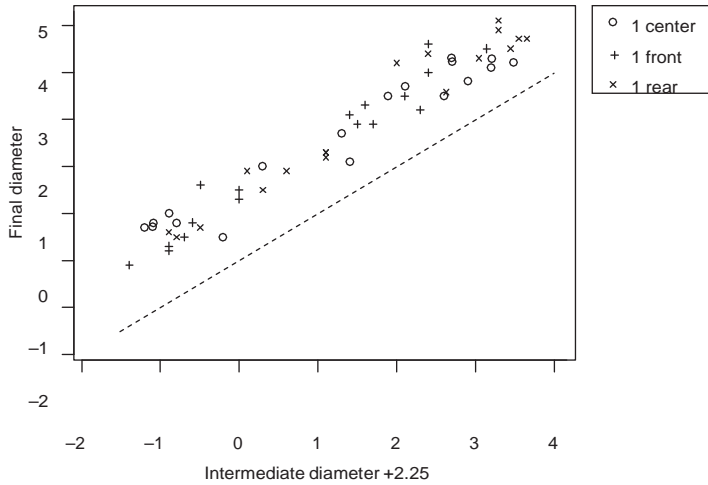
Figure III.8   Scatter plot of final versus adjusted intermediate diameter for the first main.

From the scatter plot for the first main shown in Figure III.8, we see that there is a bias between the two gages. There are similar patterns on the other mains. The team changed some of the offsets on the in-process gage to remove the bias.

The team now had confidence in the two measurement systems and recognized that they had made another improvement in the process. The first-time-through rate increased again. They proceeded with the planned variation transmission investigation to separate the effects of the lapper from those of upstream operations.

The team selected four crankshafts from each of the four grinders on four different days. Using the two gages, they measured the diameters of the 64 parts before and after lapping. The data are given in the file *crankshaft main diameter variation transmission*.

The final diameters varied between –4.3 and 5.5. This is somewhat less than the full extent of variation but, given that the process had been improved, the team was confident that the dominant cause had acted during the investigation.

In Figure III.9, we plot the final versus intermediate (in-process) diameters across all mains and positions. We see that that the intermediate diameter is a dominant cause of variation in the final diameter. The lapper transmits the upstream variation.

The team was not surprised because the lapper was a so-called dumb machine. It had no gauging or compensation and lapped for a fixed number of rotations on each main. The team expected the lapper to remove about 2.25 thousandths from the diameter. A numerical summary of the difference in diameters is:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
diameter difference	960	–4.1755	–4.1000	–4.1815	0.5121	0.0165

Variable	Minimum	Maximum	Q1	Q3
diameter difference	–5.6000	–2.4000	–4.5000	–3.9000
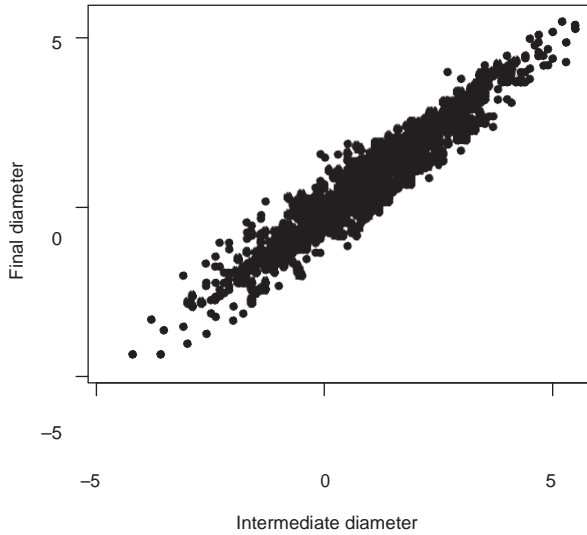
Figure III.9   Final diameter versus intermediate diameter over all positions.

Because of the offset between the two gages, the average change in diameter is –4.17 + 2.25 = –1.92, close to what was expected. The standard deviation of the change in diameter, 0.51, is small.

In Figure III.10, we look at the change in diameter by main and, more critically, by position within main. We see that the lapper removed systematically different amounts of material within mains, especially for mains 2, 4, and 5. The lapper is home to a dominant cause of the taper variation.

All parts met the taper specification ±2.0 thousandths before the lapping but not after. A summary of the before and after lapper taper is:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
int. taper	320	–0.2440	–0.1800	–0.2313	0.7449	0.0416
final taper	320	0.1996	0.2900	0.2162	0.9421	0.0527

Variable	Minimum	Maximum	Q1	Q3
Int. taper	–1.9600	1.2500	–0.8675	0.3900
final taper	–2.6500	2.3000	–0.5650	0.9375

Using the results from Figure III.10, the team arranged for maintenance on the lapper to balance all shoes so that, on average, the change in diameter was consistent from main to main and from position to position within each main. They also changed the control plan for the lapper. Once per day, an operator measured a crankshaft before and after lapping, using the final gage, and plotted the change in taper for each main on a control chart. Any changes in lapper performance could be quickly identified and remedied as required.

The team next returned to the problem of excess diameter variation. They had identified the intermediate diameter as a dominant cause of the variation. The dominant cause was not verified since the conclusion made physical sense and seemed clear-cut. To select a working
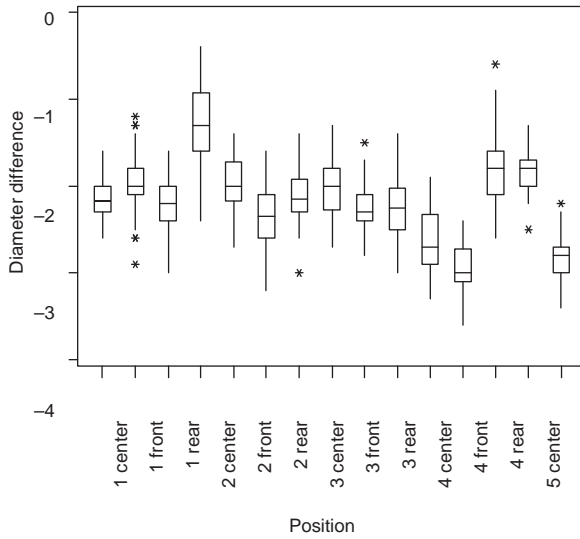
Figure III.10   Diameter difference (final minus intermediate) by position.

approach, the team briefly considered using the lapper to adjust the final diameter based on the observed intermediate diameter. They rejected feedforward control because of the cost. They would need a smart lapper that could remove a varying amount of material on each main after receiving input from the intermediate gage. They also rejected desensitization, since they did not believe any change to the lapper settings could mitigate the variation in the intermediate diameter. They decided to reformulate the problem in terms of the dominant cause.

The team used the data from the variation transmission investigation to set the baseline for the intermediate diameter. We give a numerical summary as follows and the histogram in Figure III.11.
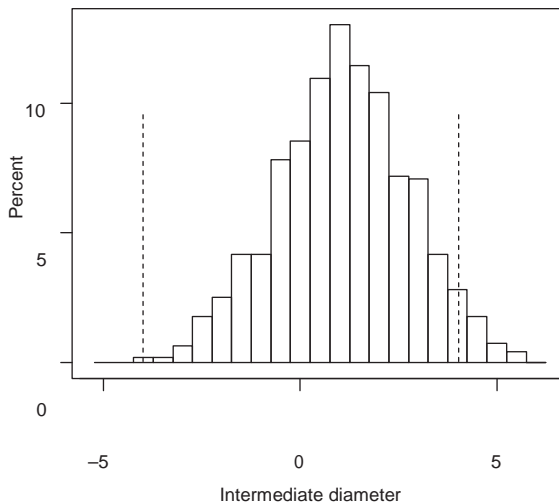
Figure III.11   Histogram of intermediate diameter (dashed lines are the specification limits ±4.0).

Variable	N	Mean	Median	TrMean	StDev	SE Mean
intermediate diameter	960	1.0716	1.1000	1.0802	1.6808	0.0542

Variable	Minimum	Maximum	Q1	Q3
intermediate diameter	–4.2000	5.5000	–0.1000	2.2000

The team made several observations:

- The within-process specification limits ±4.0 thousandths were the same as the final diameter specifications. There was no allowance for the addition of any variation at the lapper.

- The grinder operators ran the process on the high side to avoid scrap at the intermediate gage.

- The baseline standard deviation at the intermediate grinder was 1.68 and the full extent of variation of the intermediate diameter was –4.2 to 5.5 thousandths.

To set the goal for the reformulated problem, the team used the data from the variation transmission investigation to fit a regression model relating the final diameter to the intermediate diameter across all mains. From a portion of the MINITAB results, we have

The regression equation is
final diameter = –0.145 + 0.968 intermediate diameter

Predictor	Coef	SE Coef	T	P
Constant	–0.14496	0.01953	–7.42	0.000
intermed	0.968367	0.009803	98.79	0.000

S = 0.5102   R-Sq = 91.1%   R-Sq(adj) = 91.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2540.6	2540.6	9758.54	0.000
Residual Error	958	249.4	0.3		
Total	959	2790.0			

The residual standard deviation is 0.51. If we could hold the intermediate diameter fixed, this is the estimate of the standard deviation in the final diameter. The slope of the regression equation is 0.968. This gives the expected change in average final diameter for each unit change in intermediate diameter.

The original goal was to reduce the standard deviation of the final diameter to about 1.30. Using the formula from Chapter 2, we have the equation for the required standard deviation at the intermediate diameter

$$1.30 = \sqrt{0.986^2 \left(\text{required stdev} \right)^2 + 0.51^2}$$

Solving, we find that the required standard deviation for the intermediate diameter is 1.24. The team set a goal to reduce the standard deviation of the intermediate diameters to 1.25 or less.

The team had already completed the Check the Measurement System stage for the in-process gage, so they proceeded to consider a working approach. The team expected that they could find a dominant cause of the variation. However, they first considered tightening the within-process specifications, that is, using 100% inspection. The current practice was to scrap parts if any diameter was less than –4.5 and to process all others with the hope that the lapper might bring the part back into specification. The team considered reducing the specifications to ±3.0 and insisting that any part outside of specification be scrapped or reworked at the intermediate gage. The team believed that the operators at the intermediate gage and the grinding process could meet these specifications on an ongoing basis. The suggestion was met with hostile resistance from the operators. Without further variation reduction effort, the team saw that by tightening the specifications, they would transfer the scrap from the final gage to the intermediate gage with little savings.

The team next considered feedback control. In the current informal system, each of the four grinder operators used their judgment plus the data from the intermediate gage. The final gage operators also informed the grinder operators when there were a large number of crankshafts out of specification. The team suspected that grinder adjustments were made only when scrap parts were produced.

The operator could quickly make an adjustment that affected all mains simultaneously but required the help of a tool setter and considerable downtime to adjust the diameter for a single main. The team decided to adopt feedback control as a working approach and to investigate further how the diameters varied over time and position.

Since taper at the intermediate gage was small and the adjuster applied simultaneously across all 15 positions, the team concentrated on the center diameter for each main. They also decided to look at one grinder only since all grinders had the same adjustment mechanism.

The team looked back at the data from the variation transmission investigation for the five center positions. We see in Figure III.12 that there were small average differences from day to day and main to main. Within each day, we see most of the full extent of variation of the diameter. The team decided to investigate the time-to-time variation within a single shift.

The team sampled and measured three consecutive crankshafts from a single grinder every 20 minutes across the eight-hour shift. The sample was about half the crankshafts produced by that grinder. They recorded the diameters from the five center positions only. During the investigation, the grinder operator was encouraged to use his usual ad hoc feedback control method. There were no adjustments during the shift. We give the data for the
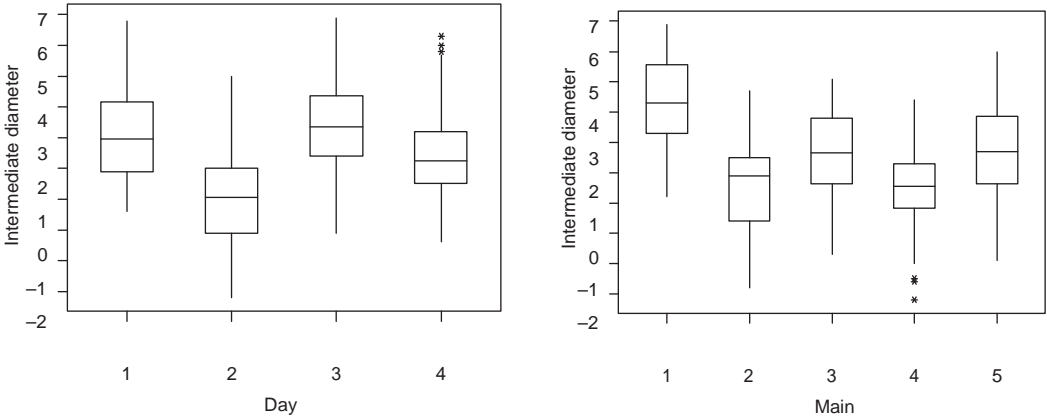
Figure III.12   Box plots of intermediate center diameter by day and main.

63 crankshafts (21 time points and three crankshafts per time) in the file *crankshaft main diameter feedback*.

The overall standard deviation of the diameters was 1.45, greater than the target value 1.25 but less than 1.68, the standard deviation of the diameter before lapping in the variation transmission study. The team used ANOVA to assess the effects of time and mains:

Analysis of Variance for diameter

Source	DF	SS	MS	F	P
main	4	91.187	22.797	15.96	0.000
time	20	155.538	7.777	5.44	0.000
Error	290	414.203	1.428		
Total	314	660.927			

The residual standard deviation is $\sqrt{1.428} \approx 1.19$. If the time-to-time and main-to-main differences could be eliminated, the within-grinder variation could be reduced to less than the target 1.25. Since the easy adjustment applied simultaneously to all mains, the team looked at the variation of the average diameter (averaged over the five mains) versus time using a multivari chart. See Figure III.13.

There was a strong trend that dominated the part-to-part variation in average diameter suggesting that feedback control would be effective. For each main, the team also plotted the average diameter (over each time period) versus time as shown in Figure III.14. There were persistent differences among the mains with main 3 being large and main 4 being small. The average diameters over all parts by main are:

main	average diameter
1	0.5048
2	0.4762
3	1.1587
4	–0.4968
5	0.6635
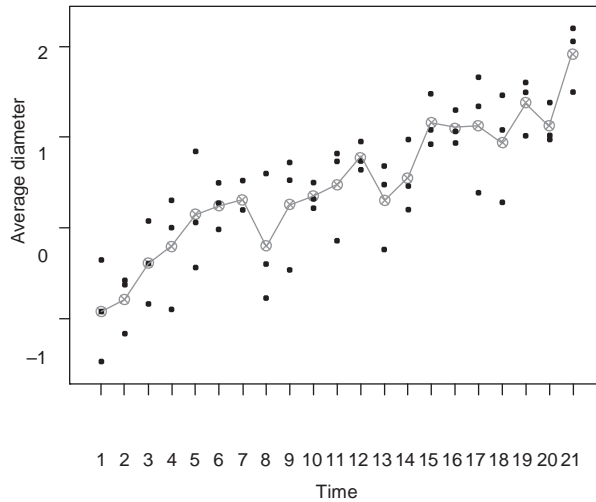
Figure III.13   Multivari chart of crankshaft average diameter versus time.
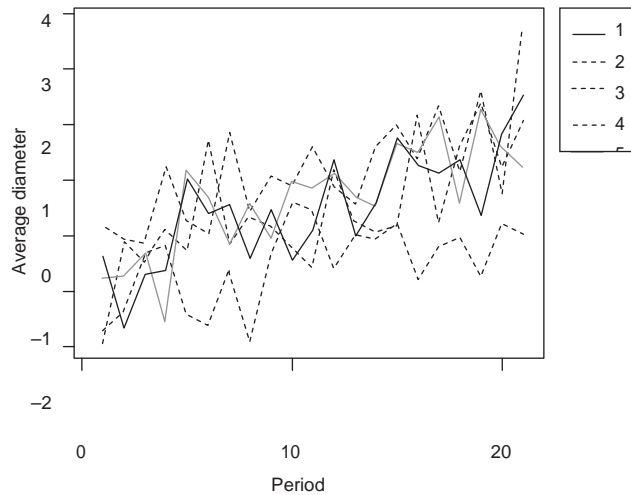
Figure III.14   Average diameter at each time period versus time period stratified by main.

The team decided to build a feedback controller using the average across all mains as the output. Separately for each grinder, they plotted a run chart of three part averages every hour. The operators were trained to look at the chart and make an adjustment if the average fell outside the range ±2.5. For political reasons, the team did not change the in-process specifications. Since the diameters drifted upwards, the target for the adjusted process was –2.0. The team carried out a small experiment on each grinder to verify that changing the adjuster produced the desired effect.

To deal with the problem of systematic differences among the mains, the team changed the section of the control plan that dealt with the setup of the grinding wheels. The tool set-ter had to ensure that the main averages of the first five parts after setup differed by no more than one-thousandth of an inch.

Once the changes were in place, the team sampled 100 crankshafts, 25 from each grinder, over one day from the intermediate gage. The standard deviation of the center diameters was 1.28, close to the goal of 1.25.

In summary, the team added the following changes to the process control plan:

- Monitor the stability of the final gage using a reference part and a control chart. Clean and master the gage as necessary.

- Monitor the performance of the lapper on a daily basis by measuring the change in diameter (before minus after lapping) on one part for all mains and record the results on control charts, one chart for each main. Watch for evidence of taper change. Rebalance the lapper shoes if taper becomes an issue.

- Adjust each grinder based on a feedback control scheme that uses the average across the five center positions and three consecutive parts at the in-process gage.

- Change the grinder setup procedure to ensure that all main averages are close to equal after setup.

The changes provided the process management with ongoing, timely information to better manage the process. The process engineer reacted to trends and anticipated problems before they occurred.

There were many benefits to this project. The taper rework was virtually eliminated by maintenance on the lapper. The scrap rate was reduced from 7.2% to 2.0% as measured over a one-month period. Each percent reduction in scrap saved approximately $250,000 per annum. The first-time-through rate at the final gage was held at over 90%. The crankshaft line had a large gain in productivity and a large reduction in scrap costs.

Highlights

Some strengths and weaknesses of this case are:

- Addressing both taper and diameter scrap and rework problem at the same time resulted in a complicated problem. The team needed to look simultaneously at 20 outputs (15 diameters and 5 tapers).

- The team wisely checked the stability of both the final and intermediate gages.

- To prevent recurrence of the problem, the team implemented feedback control using reference parts for ongoing monitoring of the gages and lapper.

- Not establishing a new baseline after the team made substantial improvements to the final gage was risky since it was harder to tell if a dominant cause acted in later investigations.

Exercises

Exercises

CHAPTER 1—NO EXERCISES

CHAPTER 2

2.1 The word *variation* is used in other contexts to describe a difference between a realized and target value such as in budget variation. How does this use compare to variation as discussed in Chapter 2?

2.2 We have heard the following comment many times from manufacturing engineers: "The cause of the variation is the product design—what can you expect me to do?" Discuss the comment in light of the definition of cause in Section 2.2.

2.3 Profile A is a measure of deviation of the actual from the ideal shape of a camshaft lobe over one region (A) of the lobe. The target value is zero and the upper specification limit is 250 microns. Use the data in the file *camshaft lobe runout baseline* to summarize the variation in this output. Do all lobes exhibit the same variation? Is there any time pattern in the variation?

2.4 Construct histograms and run charts for output 1 and output 2 given in the data file *chapter 2 exercise 4*. Find the average and standard deviation for each output. Assume the target value and upper specification limit for these lower-is-better outputs are 0 and 35.
a. Is the variation the same for each output?
b. Is the nature of the variation over time the same for each output?

2.5 You may convince yourself that the formulas for combining means and standard deviations given in Section 2.4 are true with the following numerical demonstration you can conduct in MINITAB. Generate two columns of 100 values sampled from some model (in MINITAB: Calc → Random Data → your choice of model and

parameters, for example, Normal with mean and standard deviation 0 and 1, respectively). Then, calculate two new data columns. Let one column be the sum of the original two columns and the other the difference.

   a. Find the standard deviation and average for each of the four columns.

   b. Calculate the sum and differences of the averages for the first two columns. How do these compare to the average of the other two columns respectively?

   c. Calculate the standard deviation for the sum and difference using the "square root of sum of squares" formula given by 2.1. How do the results compare to the standard deviations for the last two columns?

2.6 At a project review, the team presented the following summary of their investigation based on standard deviations.

Source of variation	Percent of total
Measurement system	30
Identified cause	50
Unidentified causes	81

   a. The reviewing manager questioned the numbers in the second column of the table because they did not add to 100. Is there an error? Explain.

   b. By what percentage can the process standard deviation be reduced by eliminating the contribution of the identified cause?

   c. Is the identified cause a dominant cause?

2.7 In Chapter 1, we discussed a project to reduce variation in pull, an alignment characteristic of light trucks. Recall that

$$\text{Pull} = 0.23*(\text{right caster} - \text{left caster}) + 0.13*(\text{right camber} - \text{left camber})$$

and that the data for two months' production are stored in the file *truck pull baseline*. The data are summarized in the following table.

Output	Average	Standard deviation
Left camber	0.257	0.129
Right camber	0.249	0.130
Left caster	3.519	0.224
Right caster	4.519	0.243
Pull	0.231	0.082

   a. Use the formula for pull and the results for how averages and standard deviations combine to predict the average and standard deviation for pull given by the last row in the table indirectly from the component averages and standard deviations.

b. Suppose you had the resources to reduce the variation in one of the alignment angles by 50%. Which angle would you choose? By how much, approximately, would the pull standard deviation be reduced?

CHAPTER 3

3.1 For a problem of interest to you speculate about the likely costs and feasibility of implementing each of the possible variation reduction approaches.

3.2 Variation in the location of a drilled hole in a machined casting can cause poor fits when the part is bolted to an engine. To reduce this variation, an engineer considers a variety of possible approaches.
   a. A vision system is available that can measure location on 100% of the parts and reject those that it judges to be out of specification. What are the advantages and disadvantages of such an approach?
   b. Institute a feedback controller by measuring two parts every hour. If hole location on either part is outside of specification, stop and adjust the process. When is such a scheme likely to be effective?
   c. A third choice is to find a dominant cause of the variation. What are the advantages and disadvantages of this strategy?
   d. If a dominant cause can be discovered, what options does the engineer have?

CHAPTER 4—NO EXERCISES

CHAPTER 5

5.1 Briefly discuss the advantages and disadvantages of the following—be sure to think of potential errors as described within the QPDAC framework.
   a. To estimate the baseline performance of a grinding process, 100 consecutive pistons were sampled and the diameters were measured.
   b. To investigate a proposed change to a chemical process, the investigators tried the change in a pilot process rather than the production process.

5.2 In the camshaft lobe BC runout problem described in Chapter 1, the team selected 50 parts (10 per day over 5 days) and measured the BC runout for each of the 12 lobes on each camshaft to quantify the baseline. The 600 runout measurements are stored in the file *camshaft lobe runout baseline original*. Conduct an analysis of these data. Are your conclusions different from those in Chapter 1? Why?

5.3 To assess a measurement system used to check the diameter of an engine bore, an investigator plans to repeatedly measure the same four (of the eight) bores on five blocks sampled from a shift of production.

a. Discuss the advantages and disadvantages of using 10 rather than 5 blocks.
b. In the investigation, all the blocks produced over one shift were available for study. Give two considerations that the investigators should take into account in making the choice of available blocks.
c. The plan was to make all measurements in a single day. Discuss the advantages and disadvantages of making the measurements over a longer time period.
d. When would the investigator be better off devoting the available resources to measuring all eight bores on fewer engine blocks?

5.4 You are a manager with the responsibility to decide if you should change the supplier for a tooling insert. You receive a report from your process engineer who has conducted an investigation into a new insert. He gives you the following verbal report and recommendation:

> Our current insert has an average life of 1105 parts. To assess the performance of the new supplier, we asked them to supply 10 inserts. We checked them the inserts out on one of our machines last week and got an average of 1300 pieces. Since the cost is the same, I think we should switch to the new inserts.

Using the QPDAC framework, think of five questions you would ask about the conduct of the investigation before you might accept the recommendation.

CHAPTER 6

6.1 In Chapter 1, we described a problem in terms of the lobe geometry of camshafts. The data are given in the file *camshaft lobe runout baseline*. Quantify the problem baseline for the following output.
a. BC runout
b. Angle error

6.2 Many programs such as Excel cannot easily handle missing observations. MINITAB is an exception. Missing values are often stored using a special numerical code (–99 is common). These special codes can result in much confusion and lead to incorrect conclusions. Consider the data *rod thickness baseline with missing observation*. In the file, there are two outputs. The output thickness_–99 uses a numerical code of –99 for missing observations, while thickness_missing uses the MINITAB missing observation symbol (*). Quantify the baseline for these two outputs. Which data summaries show the missing observation and which do not?

6.3 The baseline investigation for the V6 piston diameter example was described in Chapter 5. The data are given in the file *V6 piston diameter baseline*. Suppose the data were collected so that all the pistons from a given hour were collected at the start of the hour. Now the data come in subgroups as defined by hour. What summaries used in the baseline analysis are affected by the subgrouping? When taking

the subgrouping into account are the conclusions any different than those derived in Chapter 5?

6.4 Based on customer complaints concerning installation difficulties, a team investigated variation of a key fascia dimension. To establish a baseline, they measured the dimension on 147 fascias sampled from one month's production. The data are given in the file *fascia dimension baseline*. Using appropriate summaries of the data, quantify the baseline. Are there any concerns?

CHAPTER 7

7.1 In a process improvement problem to improve the quality of a roof panel, the measurement system (specially designed for the project) counted the number of updings on each panel. To assess the measurement system, the number of updings on 20 bad panels and 20 good panels was counted twice. The data are given in *roof panel updings measurement*.
   a. Can this investigation be used to assess the measurement variation of the counting process? Explain.
   b. Can this investigation be used to assess the bias of the counting process? Explain.
   c. The same operator counted all panels. Does the order in which he makes the counts matter? It is most convenient to count the same panel twice in a row. Is this a good idea?
   d. A scatter plot of the first versus the second measurement is given as follows. Note that some plotting symbols correspond to more than one pair of measurements? What does the scatter plot tell you about the counting process?

   e. This investigation was conducted over one hour. What are the advantages and disadvantages of spreading the two measurements on each panel over two days?
   f. Can the counting process discriminate between good and bad panels?

7.2 To monitor the process that produces engine blocks, piston bore diameters are measured on every block because they are key characteristics. Each engine block has

eight bores. The bore diameter is measured at three different heights in each bore (bottom, middle, and top) and at two different orientations at each height. Because the measurement process is automated, there are no operators. A measurement investigation was conducted over a day where the diameter of every bore on four blocks was measured four times each. The main concern was out-of-round, given by 10,000 times the difference of the two diameters at a particular height. The data are given in the file *block bore diameter measurement*. From a baseline investigation the out-of-round standard deviation was 22.8.

    a. Determine the discrimination ratio. Is the measurement system adequate?

    b. What would have been the advantage and disadvantage of conducting the measurement investigation over a longer time period?

7.3 The following MINITAB results and graphs arise from a measurement system investigation in which two different operators measured five parts three times each. The five parts were selected with initial measured values spread out over the full extent of variation, 0 to 8. The data are given in the file *chapter 7 exercise 3*. The two operators worked different shifts so the parts were saved from one shift to the next. The results include an edited ANOVA analysis as suggested in the supplement to Chapter 7 and the default gage R&R analysis in MINITAB.

Analysis of Variance for measurement

Source	DF	SS	MS	F	P
part	4	230.819	57.705	81.25	0.000
Error	25	17.754	0.710		
Total	29	248.573			

Pooled StDev =   0.8427

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.900	8.62
Repeatability	0.425	4.07
Reproducibility	0.475	4.55
operator	0.475	4.55
Part-To-Part	9.547	91.38
Total Variation	10.447	100.00

Source	StDev (SD)	Study Var (5.15*SD)	%Study Var (%SV)
Total Gage R&R	0.94876	4.8861	29.35
Repeatability	0.65207	3.3582	20.17
Reproducibility	0.68917	3.5492	21.32
operator	0.68917	3.5492	21.32
Part-To-Part	3.08975	15.9122	95.59
Total Variation	3.23214	16.6455	100.00

Number of Distinct Categories = 5

Gage name:
Date of study:
Reported by:
Tolerance:
Misc:

Gage R&R (ANOVA) for measurement

a. What do the given results tell us about the bias and variation of the measurement system?
b. In the gage R&R results, the $\overline{X}$ chart by operator is out of control. What does this mean?
c. In the gage R&R results, why is the sum of the % study variation column not 100%?
d. What is the discrimination ratio (D) for this system? How does the part selection procedure influence this ratio?
e. The gage R&R is about 29%, yet *D* is small. Why?
f. The results suggest a small operator-to-operator difference. This observed difference may be due to a difference in method or a drift of the system over the two shifts. How can you separate these two possibilities?

7.4 To assess the variation in the system designed to measure camshaft lobe geometry over time, the same camshaft was measured daily for a month. At each measurement, a number of characteristics (for example, angle error, BC runout, taper, and so on) on each lobe were determined. The data are given in the file *camshaft lobe runout measurement stability*. Is there evidence of time-to-time variation in this measurement system?

7.5 In a process that produced V8 pistons, problems occurred when pistons in inventory were remeasured (for an audit) and found to be out of specification. Since the process used 100% final inspection, this could only occur if there was a problem with the measurement system. This was puzzling because a recent gage R&R investigation at the final gage had concluded that the measurement system was acceptable. As a result, the team decided to conduct a long-term measurement investigation.

Two pistons were chosen to span the range of diameter normally seen. Each piston was measured four times a day (spread out over the day) for 12 days. During that time the regular gage calibration was performed every four hours. The data are given in the file *V8 piston diameter measurement stability*.

   a. Does the measurement system drift over time?

   b. What effect does the regular gage calibration have?

7.6 Consider the brake rotor balance example described in the case studies. In the measurement investigation, three rotors were specially selected: one well balanced, another poorly balanced, and the final rotor requiring weight near the specification limit of 0.5. The three rotors were measured twice by each of the three gages on three separate days. There is no operator effect since the gages are automated. The 54 measurements are given in *brake rotor balance measurement*. The analysis given in the case study focuses on the measurement of the weight needed to balance the rotor. However, the location (or orientation) of the weight needed to move the rotor's center of gravity is also important. Can the measurement system consistently determine the orientation of the required balance weight? From the baseline investigation, the orientation of the weight was roughly uniform from 0° to 360°.

7.7 If necessary, measurement variation can be reduced by applying the Statistical Engineering algorithm. Describe how each of the seven variation reduction approaches might be used to improve a measurement system.

CHAPTER 8—NO

EXERCISES CHAPTER 9

9.1 Think of a process and problem you know well. Define various families of causes.

9.2 The following plot shows the results of a process investigation aimed at finding a dominant cause. The dashed lines give the full extent of variation in the output as defined by the problem baseline. Can the input be ruled out as a dominant cause of variation?
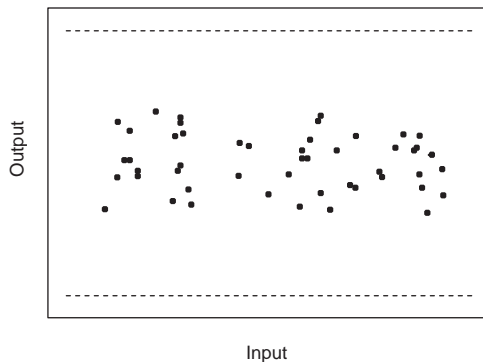
CHAPTER 10

10.1 The flow chart that follows shows the major steps in an assembly process to set the wheel alignment of a truck.

The characteristic of interest is right camber with specification $0.5 \pm 0.5°$. Camber is measured on every truck by one of the four gages (aligners). The process performance for right camber is shown as follows based on about 6200 consecutive trucks.

a. Based on this histogram, can the measurement system be eliminated as a dominant cause of the camber variation?

b. What data could you collect to demonstrate that a dominant cause does not act in the measurement system?

c. How could you rule out the assembly operation as the home of a dominant cause?

d. How could you eliminate differences in the suppliers of the upper control arm as the home of a dominant cause?

e. The plot that follows shows the process behavior over three shifts. What family of causes can be eliminated based on these data?

f. The following plot shows the camber variation for the first nine trucks in the data set. What families can be ruled out as the home of a dominant cause using these data?

g. In a special study, one key characteristic of the lower control arm was measured for 30 trucks. The other components were specially selected to ensure that they were well within specification. Based on the plot that follows, is the lower control arm characteristic a dominant cause of right camber variation? Explain.

10.2 Consider again the camshaft lobe runout problem introduced in Chapter 1. Each camshaft has 12 lobes with a number of characteristics of interest. In a search for a dominant cause, we may compare the lobe-to-lobe and camshaft-to-camshaft families of variation. Using the problem baseline data given in the file *camshaft lobe runout baseline,* explore the relative sizes of the two families for the following characteristics and decide which family, if any, can be eliminated.
   a. Profile *A*
   b. Profile *B*
   c. Profile *C*

10.3 In the manufacture of an injection molded part, a key crossbar dimension exhibited excess variation. The problem baseline estimated the standard deviation of the crossbar dimension as 0.46 with full extent of variation –0.3 to 2.0. The goal was to reduce the standard deviation to less than 0.25. An investigation showed the measurement system to be highly capable.

   Next the team conducted a multivari investigation where five consecutive parts were sampled every 30 minutes for four hours. Analyze the data given in *crossbar dimension multivari*. Which family of variation can be eliminated as the home of the dominant cause?

10.4 As described in Chapter 7, in a process that placed labels on bottles, the team searched for an acceptable measurement system. The file *label height measurement* contains the data from an investigation in which three operators using a hand feeler gage measured three specially chosen bottles three times on two different days. The bottles were chosen to roughly cover the range of label height values seen in the process. From a baseline investigation an estimate of the overall standard deviation was 0.022. The results of a one-way ANOVA are:

Analysis of Variance for height

Source	DF	SS	MS	F	P
part	2	0.0413263	0.0206631	263.10	0.000
Error	51	0.0040054	0.0000785		
Total	53	0.0453317			

				Individual 95% CIs For Mean Based on Pooled StDev
Level	N	Mean	StDev	----+---------+---------+---------+--
1	18	0.06966	0.00965	(-*-)
2	18	0.10950	0.00930	(-*)
3	18	0.13705	0.00749	(-*)
				----+---------+---------+---------+--
Pooled StDev = 0.00886				0.075 0.100 0.125 0.150

We have stdev(due to measurement) = 0.00886, and thus

$$stdev(\text{due to process}) = 0.0204 \left(\sqrt{(0.022)^2 - (0.00886)^2} \right)$$

and an estimated measurement discrimination ratio of 2.3. The team decided to improve the measurement system before addressing the original label height

variation problem. Reanalyze the measurement investigation results to eliminate families of possible dominant causes of measurement variation.

10.5 A process improvement problem was initiated to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. The team discovered that the dominant cause was contamination before the forming process step. In an investigation, the team measured the particle count on coils directly after the arrival from steel supplier and again after blanking and stamping (before the forming process). They measured at the tail, middle, and head of four different coils. The data are given in the file *roof panel updings variation transmission.* What does the following scatter plot tell us about the dominant cause? The plotting symbols correspond to the four different coils.

10.6 In the engine block porosity example discussed in Chapter 10, the team found the occurrence of high porosity coincided with production directly after breaks. To explore this clue further, they conducted another investigation in which the porosity of 50 consecutive blocks was measured. The first six blocks were taken from directly before the lunch break, and the next 44 blocks were the first ones produced after the break. The data are given in the file *engine block porosity run chart*. What does the observed pattern in the run chart tell us about the dominant cause?

10.7 High silicon concentration in cast iron is undesirable as it was found be a dominant cause of fluidity variation. However, measuring the silicon level can be difficult. The measurement process consisted of sampling the molten iron by pouring sample coins for testing. The coins are then machined and polished before being spectrochemically analyzed. The full extent of variation in percent silicon as measured in the current process was 1 to 4%. The measurement system was investigated by repeatedly measuring three different coins that roughly covered full extent of variation in the observed percent silicon. Two operators measured each of the three coins three times on each of two days. The data are given in the file *iron silicon concentration measurement*. Analysis of the measurement results estimated the measurement standard deviation as 0.33. The corresponding estimate of the process standard deviation was 0.5; thus the discrimination is too small at around 1.5. The team needs to improve the measurement system. Using the existing measurement investigation data, are there any clues about the dominant cause of the measurement variation?

CHAPTER 11

11.1 In a multivari investigation, two consecutive pieces are selected from each of three pallets once per hour. Sketch the appearance of the multivari chart that shows all three families at the same time if a dominant cause lies in the following family. Use the following multivari chart template in which the dashed lines indicate the full extent of variation.
   a. Pallet-to-pallet family
   b. Part-to-part family
   c. Hour-to-hour family
   d. An interaction between the part-to-part and pallet-to-pallet families

11.2 In the engine block leakers example, introduced in Chapter 1, the baseline defect rate was 2–3%. The team conducted a multivari investigation where three consecutive blocks were taken at twelve different times throughout the day. The investigation continued for three production days giving a total of 108 castings. Each block

was tested for leaks. The data are given in the file *engine block leaks multivari*. What can you conclude?

11.3 At an intermediate operation the team planned a multivari investigation in which three consecutive parts were taken from each of two machines operating in parallel once every hour for two days. Consider two different processes. In the first process, the order of the parts coming from upstream is preserved, while in the second process the order is jumbled. When interpreting the resulting multivari chart (think specifically about the part-to-part family), what difference does it make which process we are observing?

11.4 In a multivari investigation, the diameter of a transmission shaft was measured at four positions (left and right side at two different orientations) for three consecutively sampled shafts each hour. The data are available in the file *transmission shaft diameter multivari*.

a. What conclusion can you draw from the multivari charts that follow?
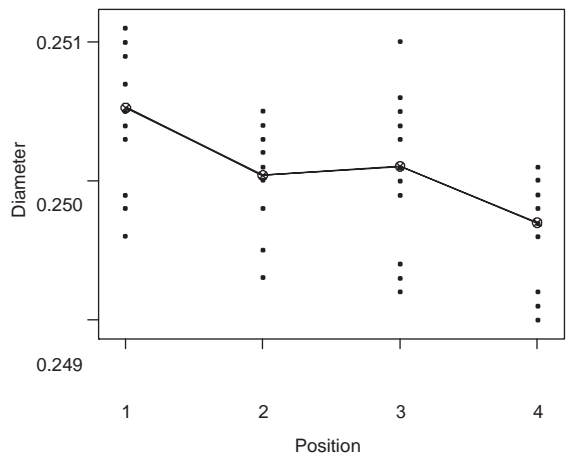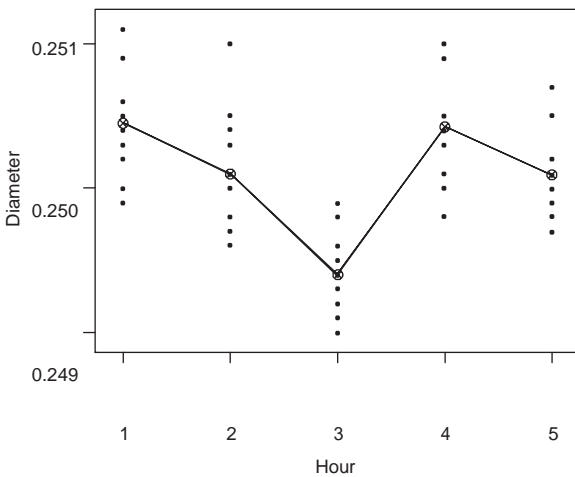
b. Using the data assess whether the dominant cause acts in the shaft-to-shaft family.

11.5 In the production of engine blocks, bore diameters are key characteristics. Bore diameter is measured at three heights and two orientations in each of the eight bores in each block. The team used Statistical Engineering to address a problem of excess bore diameter variation. The baseline investigation found a standard deviation of 3.04 and the full extent of variation of –9 to 9 as measured from nominal in microns. There were no strong differences between the different bores, heights, or positions. Another investigation concluded that the measurement process was adequate. To isolate the processing step where the dominant cause acts, the team selected 30 engine blocks haphazardly from a day's production. In the investigation the bore diameter (measured from nominal at that processing step) in the first bore at the top position and first orientation was measured at each of five processing steps in the machining part of the process. The data are given in the file *block bore diameter variation transmission*. Which processing step is home to the dominant cause?

11.6 In the paint film build example described in Chapter 3, a baseline investigation found the standard deviation was 0.315, with an average of 16.2 thousandths of an inch. The full extent of variation was 15.2 to 18.5. To search for a dominant cause,

the team conducted a multivari investigation where they measured the film build at five positions on five cars in a row every hour for two shifts (16 hours). This resulted in a total of 400 film build measurements. The data are given in the file *paint film build multivari*. Based on the plots that follow, what conclusions can you draw? We define group as (hour – 1) · 5+ position.
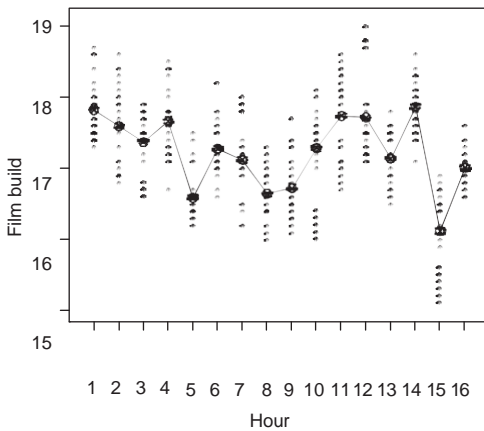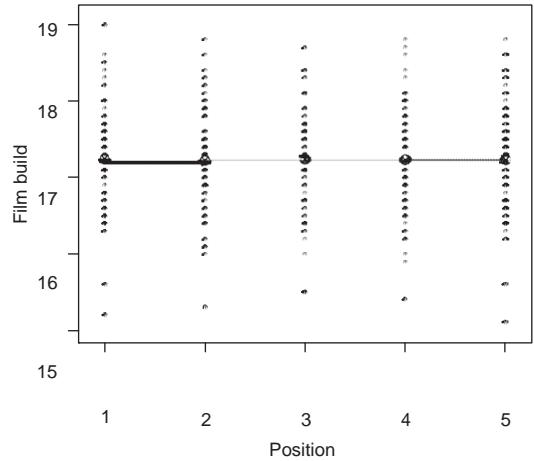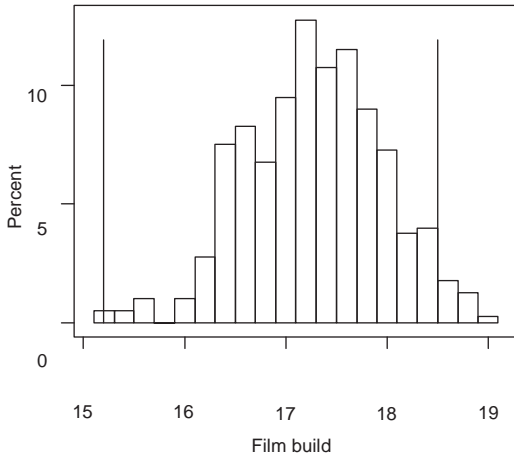
Multivari chart for film build by group

11.7 A team wanted to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. A baseline investigation found that the total number of updings in 20 consecutive panels ranged between 5 and 438. To search for a dominant cause the team conducted a multivari investigation where the number of updings was counted for 20 consecutive roof panels from three sections of seven different pallets of steel sheets. Originally, the plan was to repeat this data collection over two separate days. However, the team found the full extent of variation from the baseline was observed on the first day so they stopped collecting data. The data are given in the file *roof panel updings multivari*.

   a. Analyze the data using multivari charts and draw conclusions.
   b. When the number of updings was counted they were classified into one of the nine locations as numbered in the schematic that follows. Analyze the multivari data using a concentration diagram based on the given schematic.

Roof panel

Front

11.8 The baseline investigation for the sand core example discussed in Chapter 1 involved taking five samples over a single day of five consecutive shots of four cavities each. The data are given in *sand core strength baseline*. What conclusions can you draw?

CHAPTER 12

12.1 Vehicle plant and warranty complaints for loose spark plug wires at the spark plug end prompted an improvement project. As a result of several investigations, the family of causes related to push forces on the wires was the home of a dominant cause. A further investigation then compared eight loose and eight good connections. For each of the 16 connections, the team measured the terminal position of wires and terminal runout of the spark plug in millimeters. The data are given in the file *spark plug connection comparison*. What do the data tell us about the dominant cause?

12.2 A sunroof installation process suffered from a 90% rework rate due to a lack of flushness. Flushness is defined as the difference in height between the sunroof seal and the metal roof. It is measured using digital calipers at six points (three at the front and three at the back). A baseline investigation showed that flushness problems were most pronounced at the two front corners with full extent of variation between –3.5 to 4 mm and standard deviation 1.25 millimeters. A goal of reducing the front corner flushness variation to 0.5 and a range of –2 to 1 millimeters was established. Based on engineering knowledge, the team felt that only two types of characteristics could lead to flushness variation, namely roof crown height and attachment pad height. When the roof is adapted to allow installation of a sunroof, six installation pads are added. Based on this knowledge, the team selected six vehicles with large positive flushness and six vehicles with large negative flushness on both front corners. The sunroof modules were removed and the six attachment pad heights and

roof crown height were measured at the front and back. The data are given in the file *sunroof flushness input-output*. What conclusions can you draw?

12.3 An example related to sand defects in manifolds was discussed in Chapter 12. Before the problem mentioned in Chapter 12, the team carefully monitored the process for a shift. Each of 970 manifolds was classified as either scrap (due to sand issues) or not scrap. In addition many inputs relating to each manifold, including some discrete inputs such as mold number and continuous inputs such as pour time were recorded. In some cases the linkages were difficult to establish, and the team did the best they could. The data are given in the file *manifold sand scrap comparison*. What conclusions can you draw?

CHAPTER 13

13.1 In a verification experiment there were two suspects at two levels. The low and high levels for each suspect were chosen based on the extremes from historical variation. The results of the first three runs of the experiment are shown in the following table.

Input *A*	Input *B*	Order	Output
Low	Low	2	13
Low	High	3	16
High	Low	1	17
High	High	4	?

Given that the full extent of output variation is 12 to 30, what conclusions can you draw about the dominant cause?

13.2 In the engine block porosity example discussed in the text and exercises of Chapter 10, a dominant cause of porosity acted immediately following scheduled breaks in production. Based on this evidence, the team identified two suspects: iron pouring temperature and the addition of ladle wash. During work stoppages, iron that remained in the six pouring ladles cooled off because there was no external heat source. At the start of the break, ladle wash was added to the ladles to protect the refractory (surface). The team could not easily manipulate the pouring temperature, but they could change the amount of ladle wash. They conducted a verification experiment in which they added the normal amount of wash to ladles 1, 3, and 5 and half the normal amount to the other three ladles over two lunch breaks. At each break, they measured the porosity of the first 30 blocks poured (five from each ladle). The data are given in the file *engine block porosity verification*.
   a. What have we learned about the identity of the dominant cause of porosity?
   b. Explain how the effects of ladle number and the presence or absence of ladle wash are confounded. Does this matter?

c. Is it a problem that in this verification experiment we have not observed the behavior of the process before lunch breaks?

13.3 The manufacture of a tube assembly required a protective nylon sleeve to be positioned and bonded to a tube. The bond strength of this tube assembly was occasionally tested using a destructive test where the sleeve was subject to increased tensile shear load until failure. In the current process, the average pull-off force was around 15 pounds, but roughly 8% of assemblies tested had a pull-off force less than the desired minimum of five pounds. The team decided to try to solve the problem by reducing the variation in pull-off force rather than by increasing the average pull-off force. A number of investigations were conducted to find the dominant cause. A multivari investigation suggested that the dominant family of causes was tube-to-tube. At this point, the team decided to conduct an experiment to search for a dominant cause using the limited process information they had gathered. They planned a factorial experiment with three suspects—clearance between the sleeve and tube, amount of adhesive, and cure time—all consistent with the tube-to-tube family clue. The team chose the low and high levels of each suspect to roughly match their range in regular production. The levels of clearance were achieved by sorting sleeves and tubes. There were two replicates of each treatment, and the run order was randomized. The data are given in the file *nylon bond strength verification* and summarized in the following table:

Treatment	Order	Clearance	Adhesive	Cure time	Bond strength
1	6, 7	Low	Low	Low	26, 28
2	14, 8	High	Low	Low	10, 10
3	2, 3	Low	High	Low	25, 26
4	12, 5	High	High	Low	7, 9
5	16, 4	Low	Low	High	24, 27
6	13, 10	High	Low	High	12, 13
7	15, 11	Low	High	High	23, 21
8	1, 9	High	High	High	7, 7

What do the results tell us about the dominant cause?

13.4 Steering knuckles are produced in a gray iron casting process. Around 2% of castings were scrapped because the percent nodularity was too small. In this example the team did not clearly establish a problem baseline. The team thought the cause must be related to the inoculation of the molten iron using a silicon-based alloy. The inoculant was added as the iron was poured to increase nodularity (and thus casting strength). Based on observing the process, the team noticed that the amount of inoc-

ulant added by the automated delivery system seemed to vary. The desired amount of inoculant was obtained by slowly shaking the inoculant onto a plate. The plate was designed to tip automatically when the required weight of inoculant was present. The team saw that the location on the plate where the inoculant fell varied, and they thought that this might influence when the plate tipped and thus how much inoculant was delivered. The team decided to verify inoculant amount as the dominant cause of nodularity variation. In the verification experiment, they produced a total of 20 castings at each of two levels of inoculant amount, 12.3 and 13.5 grams. For the experiment the inoculant was carefully weighed and added by hand. The experiment consisted of eight runs of five castings each. The order of the eight runs (four at each level) was randomized. For each of the 40 castings the percent nodularity was determined. The data are given in the file *steering knuckle strength verification* and are summarized in the table that follows:

Run	Inoculant amount	Order	Percent nodularity
1	12.3	2	81.8, 79.4, 80.3, 80.6, 79.3
2	12.3	3	79.8, 77.0, 77.8, 79.3, 78.7
3	12.3	8	80.9, 82.0, 80.6, 80.6, 81.1
4	12.3	4	81.0, 79.4, 77.0, 80.6, 80.2
5	13.5	5	82.5, 86.1, 82.3, 83.5, 85.2
6	13.5	7	82.1, 84.6, 83.9, 85.0, 85.6
7	13.5	6	85.0, 87.8, 83.1, 84.0, 84.4
8	13.5	1	85.0, 84.3, 86.3, 83.8, 82.9

    a. What considerations should the team have used in choosing the two levels for inoculant?

    b. Why was randomizing the order of the runs important?

    c. Has the team verified the dominant cause of nodularity variation?

CHAPTER 14

14.1 In the camshaft lobe runout example, the team searched for a dominant cause of variation. As discussed in Chapter 10, they conducted a variation transmission investigation where runout was measured directly before heat treatment and after the final step of the process, on the same 32 parts selected over the course of one day. In the investigation the grinder (one of eight) and heat treatment spindles (one of four) used were also recorded. The data are given in the file *camshaft lobe runout variation transmission*. They found that a dominant cause of variation was the

BC runout just after heat treatment and, more specifically, as shown in the plot that follows, that heat treatment spindle was a dominant cause.

In this example, the team decided not to reformulate the problem but to look for a more specific cause.

a. Discuss the advantages and disadvantages of the decision not to reformulate.

b. Suppose the team had reformulated the problem based on heat treatment spindle and that the original goal was to reduce the final runout standard deviation to less than 4.5. Using the results from a one-way ANOVA model based on heat treatment spindles, derive a goal for the new problem based on differences among spindle averages.

14.2 In Chapter 11, the team found that the piston diameter directly after operation 270 was a dominant cause of final V6 piston diameter variation. The relationship is illustrated in the scatter plot that follows. The data are given in the file *V6 piston diameter variation transmission*.
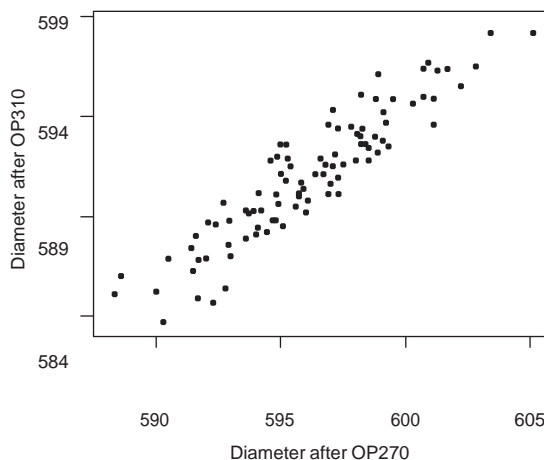
The team decided to look further for a more specific dominant cause. Suppose, however, they had wanted to reformulate the problem in terms of the operation 270 diameter. Determine an appropriate goal for the reformulated problem. Recall that the goal for the original problem was to reduce the final diameter standard deviation to less than 2.0.

CHAPTER 15

15.1 Based on customer complaints about assembly difficulty, a team investigated fascia dimension variation. A baseline investigation found that some fascias were too large. The team felt that reducing the average size of the fascias could solve the problem (that is, they adopted the Move the Center approach). They planned a (full) factorial experiment with two candidates, cycle time and cure time, each at two levels to look for an adjuster. They chose the levels for each candidate based on engineering judgment. The results of the experiment are given in the file *fascia dimension move center* and in the following table. For each treatment, the team conducted four runs producing 10 fascias for each run. The order of the 16 runs was randomized over a day. In the data, we give only the average fascia dimension from each run and not the individual values.

Treatment	Run order	Cycle time (minutes)	Cure time (hours)	Average fascia size (from nominal)
1	8, 10, 1, 14	85	5	4.50, 5.23, 5.75, 6.51
2	6, 15, 12, 2	113	5	7.12, 8.25, 9.06, 9.28
3	11, 3 , 9, 16	85	19	3.65, 3.75, 4.27, 5.34
4	13, 5, 4, 7	113	19	4.24, 6.31, 7.15, 8.22

   a. Can cycle time or cure time be used as an adjuster?
   b. Suppose the goal was to reduce the average fascia size to 3.0. What do you recommend?
   c. What is the advantage of looking at the dimensions for all the fascias within a run, rather than the averages?

15.2 An experiment was carried out to investigate four candidates to search for an adjuster of the formability safety margin of galvanized sheet metal trunk lids. The purpose was to increase the average safety margin from the baseline value of 10.7. In the experiment, each candidate was tested at two levels, selected to be near the edge of what was physically possible—see the table that follows. Note that none of the treatments corresponded to the existing process settings.

Treatment	Run order	Tonnage	Lubrication	Blank size		Prebending	Safety margin
1	6	310	Unlubricated	949	1494_mm	No	8
2	16	375	Unlubricated	949	1494_mm	No	11
3	3	310	Lubricated	949	1494_mm	No	12
4	11	375	Lubricated	949	1494_mm	No	0
5	7	310	Unlubricated	965	1500_mm	No	13
6	15	375	Unlubricated	965	1500_mm	No	6
7	4	310	Lubricated	965	1500_mm	No	11
8	13	375	Lubricated	965	1500_mm	No	1
9	1	310	Unlubricated	949	1494_mm	Yes	18
10	12	375	Unlubricated	949	1494_mm	Yes	17
11	5	310	Lubricated	949	1494_mm	Yes	18
12	14	375	Lubricated	949	1494_mm	Yes	10
13	2	310	Unlubricated	965	1500_mm	Yes	8
14	10	375	Unlubricated	965	1500_mm	Yes	12
15	8	310	Lubricated	965	1500_mm	Yes	16
16	9	375	Lubricated	965	1500_mm	Yes	8

Press tonnage was very difficult to change so all eight runs with low press tonnage were carried out first. Within each group of eight runs, the order was randomized. The data are given in the file *sheet metal move center*.

   a. Analyze the experimental data to see if any of the candidates is an adjuster.
   b. Does the restriction on randomization required for this experiment make any difference to the conclusions we can draw?

15.3 In the sand core strength example introduced in Chapter 1, too many cores were breaking during handling. A suggested solution was to increase the core strength (and thereby reduce core breakage) by increasing the resin concentration. It was known that increasing the resin would result in a stronger core. However, the precise nature of the relationship—that is, how much the core strength increases for a given change in resin concentration—was not known. An experimental investigation was planned to quantify the relationship. Three levels of resin concentration (1.3, 1.6, 1.9% by weight) were chosen based on engineering judgment. In the experiment, 40 cores for each level of resin were produced; 15 were measured for strength (using a destructive test) and the remaining 25 were processed to look for casting problems.

The experiment consisted of three runs with 15 repeats. The order of the runs was not randomized. The data are given in the file *sand core strength move center*.

a. What can you conclude about the relationship between resin concentration and core strength?

b. The team used only three runs with 15 repeats for each run. Discuss the advantages and disadvantages of this plan compared with using five replicates for each treatment with three repeats each.

CHAPTER 16

16.1 In a sonic welding operation, problems arose due to poor weld strength, measured as pull-off force. The goal was to reduce the variation and increase the average pull-off force. The second goal is not addressed here. From the baseline, the full extent of variation for pull-off force was 0.9 to 3.0. The team discovered that the dominant cause acted in the time-to-time family. While they could not be more specific, the team felt that the dominant cause was related to material hardness, which was outside their control. They decided to try to desensitize the process to variation in the dominant cause.

The team planned a fractional factorial experiment with four candidates at two levels each in eight treatments. Using the results of regular process monitoring, they identified three time periods when weld strength was low, medium, and high relative to the baseline. In each period, they randomized the order and then produced a part with each of the eight treatments. The pull-off force data and plan are given in the file *sonic weld desensitization* and the table that follows. The three values in the columns Order and Pull-off force correspond to the three different time periods. The original settings of the candidates correspond to treatment 2.

Treatment	Order	A	B	C	D	Pull-off force
1	7, 2, 6	−1	−1	−1	−1	1.8, 2.1, 2.3
2	4, 6, 1	−1	−1	1	1	0.9, 2.0, 2.7
3	6, 3, 4	−1	1	−1	1	2.0, 2.4, 2.1
4	8, 1, 2	−1	1	1	−1	0.6, 1.5, 3.0
5	3, 7, 8	1	−1	−1	1	2.8, 2.9, 3.0
6	2, 4, 3	1	−1	1	−1	2.1, 3.0, 3.7
7	1, 8, 7	1	1	−1	−1	2.9, 3.2, 3.1
8	5, 5, 5	1	1	1	1	2.3, 2.5, 4.0

a. Explain why the team believed the dominant cause acted over the three runs for each treatment.

b. What levels of the candidates do you recommend to reduce the variation in pull-off force?

c. Another way to assess the results of this experiment is to summarize the output across each treatment using log standard deviation. Using this performance measure, do your conclusions differ from part b?

16.2 In the crossbar dimension example discussed in Chapter 12, the team found that the dominant cause of dimension variation was barrel temperature. Because it was hard to control in regular operation, the team decided to try to make the process less sensitive to barrel temperature variation. In the current process, barrel temperatures ranged over roughly 4°C. The team planned a half fraction factorial experiment with three candidates—target barrel temperature, injection pressure, and material—at two levels each, as shown in the following table. The current injection pressure and target barrel temperature were 1000 and 75, respectively. Note that although the variation in barrel temperature was the dominant cause, the target barrel temperature is a fixed input. Five crossbars were produced and measured in each run. For each treatment, there were two runs, one at the target barrel temperature plus 2°C and the other at the target barrel temperature minus 2°C. The data are given in the file *crossbar dimension desensitization* and in the table as follows.

Treatment	Target barrel temperature	Injection pressure	Material	Dimensions at barrel temperature −2°C	Dimensions at barrel temperature +2°C
1	75	1000	Old	0, −0.1, 0.1, −0.1, −0.2	0.5, 1.1, 0.8, 0.9, 0.7
2	75	1200	New	1.1, 0.6, 1.0, 1.4, 1.1	1.5, 1.8, 1.5, 1.4, 1.3
3	79	1000	New	1.1, 1.0, 1.3, 0.9, 0.8	1.0, 1.1, 0.8, 0.9, 1.0
4	79	1200	Old	1.2, 1.8, 1.8, 1.7, 1.9	2.3, 2.1, 2.4, 2.1, 1.9

Since the average dimension can be easily adjusted, we focus the analysis on finding a way to make the process less sensitive to barrel temperature variation.

a. What levels of the candidates do you recommend?

b. Injection pressure and material were chosen as candidates based on engineering judgment. Looking again at the results presented in Chapter 12, what motivates the choice of target barrel temperature as a possible candidate?

16.3 In Chapter 16, we describe a desensitization experiment for the refrigerator frost buildup example where each refrigerator is subjected to only two extreme levels of environmental causes. Here we consider a hypothetical experiment in which each

refrigerator is exposed to a number of environmental conditions to ensure that any chosen new design works well under any conditions, not just extreme conditions.

The experimental design for the four candidates—D1, D2, D3, and D4—is the same as in Chapter 16. Here we plan to test each of the eight refrigerators (treatments) under all eight possible combinations of the usage or environmental inputs as given in the following table:

Varying input	Cause combination							
	1	2	3	4	5	6	7	8
Ambient temperature (°C)	26	26	26	26	32	32	32	32
Relative humidity (%)	70	70	90	90	70	70	90	90
Door openings per hour	4	8	4	8	4	8	4	8

The experimental plan had 64 runs. To conduct the experiment, all eight refrigerators were simultaneously placed in a test chamber and exposed to each cause combination in the given order. The cooling plate temperatures are given in the file *refrigerator frost buildup desensitization2* and in the following table:

Treatment	Candidates				Cooling plate temperatures (in cause combination)							
	D1	D2	D3	D4	1	2	3	4	5	6	7	8
1	N	O	O	N	3.6	3.9	4.6	1.0	4.4	0.1	4.4	0.7
2	N	O	N	O	5.1	4.7	4.3	2.9	4.2	4.1	7.1	5.1
3	N	N	O	O	4.6	4.6	4.3	4.9	2.4	5.0	5.5	16.0
4	N	N	N	N	3.8	12.8	6.9	6.9	7.1	6.7	3.0	15.7
5	O	O	O	O	2.9	0.2	−0.2	−0.2	−0.2	−0.2	−0.2	16.0
6	O	O	N	N	0.1	1.9	0.8	1.3	5.9	5.1	0.0	14.7
7	O	N	O	N	0.7	0.8	0.1	0.1	0.4	0.2	0.1	−0.1
8	O	N	N	O	0.2	3.4	0.3	1.0	4.0	0.2	5.2	16.0

In the table, we have coded the new and original settings for the candidates as N and O, respectively. What conclusions can you draw? Remember, the goal is to desensitize cooling plate temperature to changes in the environmental conditions.

16.4 There were excessive failures in the accelerated life testing of electric motors. Using a group comparison investigation, the team found that unevenness in the commutator shaft surface was a dominant cause of these failures. The team next reformulated the problem to one of reducing the unevenness in the commutator shaft. The surface

unevenness is measured on a scale of 1 (smooth) to 10 (rough). With further investigation, the team determined that the dominant cause of the variation in the (final) smoothness was the shaft profile before machining. The team adopted the Desensitization approach. They decided to conduct a fractional factorial experiment with eight treatments using four candidates. For each of the eight treatments there were two runs, one that used a shaft with a premachined smooth profile, and a second that used a rough profile. The experimental plan and data are given in the file *electric motor failure desensitization* and the table that follows. The order of the runs was randomized.

| Treatment | Depth | Grind time | Rotational speed | Feed rate | Order | Smoothness | |
						Smooth profile	Rough profile
1	Shallow	Short	1800	Slow	4, 5	2	7
2	Deep	Short	1800	Fast	6, 11	3	8
3	Shallow	Long	1800	Fast	1, 14	1	9
4	Deep	Long	1800	Slow	16, 12	2	8
5	Shallow	Short	2400	Fast	13, 9	3	2
6	Deep	Short	2400	Slow	10, 8	1	4
7	Shallow	Long	2400	Slow	3, 7	2	3
8	Deep	Long	2400	Fast	15, 2	3	5

a. What is the confounding structure of the design? What limitations does this introduce?
b. What conclusions can you draw?
c. What would be the advantages and disadvantages of measuring the time to failure using the accelerated life test for each run rather than judging the smoothness of the commutator surface after machining?

CHAPTER 17

17.1 In an investigation, 100 trucks were selected from regular production over two weeks. The frame geometry as given by four summaries (left and right front, left and right rear) and the alignment outputs left and right camber and caster were determined for all 100 trucks. The data are given in the file *truck pull feedforward*. In Chapter 17 an analysis determined that feedforward control based on frame geometry was feasible for left

caster. Repeat the analysis for the other outputs: right caster, left camber, and right camber.

17.2 Engine assembly problems occurred due to a poor fit between the pistons and the engine bore. The dominant cause of poor fit was found to be variation in the clearance, the difference between the (minimum) bore diameter and the (maximum) piston diameter. To solve this problem, the team thought about using the feedforward (selective fitting) approach. The idea was to measure each piston diameter and place them into bins of similar diameter. Then, after each bore diameter was measured, a piston would be selected from the appropriate bin. To assess this proposal the diameter measurements for 469 pistons and bores, as measured from nominal, are given in the file *block bore diameter feedforward*. Quantify the expected reduction in clearance variation when using one (that is, no selective fitting), two, three, or four bins of pistons. A suggestion is to define the bins by dividing the range in piston and bore diameters (roughly –10 to 10 microns) into equal widths.

17.3 In the V6 piston diameter example discussed in Chapter 11, the team found that piston diameter after Operation 270 was a dominant cause of the final diameter. The data are given in the file *V6 piston diameter variation transmission*. This suggested that feedforward control might be a feasible approach.
   a. What are the requirements for feedforward to be feasible in this context?
   b. If feedforward were feasible, assess the potential benefit using the results of the variation transmission investigation.
   c. Could the team also use the diameter after Operation 200, rather than the diameter after Operation 270, as the input to a feedforward controller?

CHAPTER 18

18.1 The bias of the system used to measure camshaft journal diameters tended to increase over time. The cause of this increase was not determined. Instead, the team introduced a feedback controller. At the start of each shift, a master journal with known diameter was measured. If the measured diameter deviated from the known value by a large amount, the measurement system was recalibrated.
   a. How could we sensibly define a large deviation in this context?
   b. What would happen to the measurement variation if the measurement device were recalibrated every time the master journal was measured, rather than only when the deviation from the known dimension was large?

18.2 In a machining process, the dominant cause of dimension variation acted in the setup family. That is, the dimension variation within a particular setup was small relative to the variation from one setup to the next. The existing control plan called for a complete process adjustment back to the target based on the first observation after each setup. There were more than 200 parts machined between setups. The baseline dimension standard deviation was 0.31. The team decided to explore a new feedback

control scheme based on the average for the first five observations after each setup. In an offline investigation, they carried out 10 setups and produced 20 parts after each setup without any adjustment. The dimension data, scaled so that the target is zero, are given in the file *machining dimension feedback*.

a. Use a one-way ANOVA to estimate the standard deviation if the process could be adjusted so that the dimension averages across all setups were equal.

b. Use simulation to compare the performance of the existing feedback controller with the proposed controller that makes a complete adjustment based on the average for the first five observations after each setup.

c. In general, we may design a feedback controller by averaging the output from the first *n* observations after each setup. What considerations help you decide how many observations should be used to estimate the process average after each setup?

18.3 In a machining process, there was excess variation in the diameter of a precision ground shaft. The shaft diameter was measured for all shafts using a complex automated gage (that also measured other outputs). Upon investigation, the team discovered that the dominant cause acted in the measurement family. In particular, the measurement bias changed from day to day, consistent with the pattern observed in the baseline. To explore this bias change further the team planned an investigation where the diameter of the same shaft was measured each hour for four days. A total of 32 diameter measurements were made. The data are given in the file *precision shaft diameter feedback*, with the output being the diameter measured from nominal. The results show a gradual drift. The team speculated that the drift was caused by changes in some (unidentified) environmental conditions. They decided to reduce the measurement variation using a feedback controller.

a. What type of feedback controller (that is, what prediction equation and what adjustment rule) would you recommend in this application?

b. Suppose the team decided to use a feedback controller based on EWMA forecasts with the smoothing parameter alpha equal to 0.4. What kind of a reduction in the measurement variation could they expect?

CHAPTER 19

19.1 In the paint film build example introduced in Chapter 3, the baseline standard deviation in film build (paint thickness) was 0.67 thousandths of an inch. With this variation, to ensure a minimum film build of 15-thousandths of an inch, the process was centered at 17. The goal was to reduce the standard deviation to 0.35, thereby allowing for a reduction in the average film build.

The dominant cause of film build variation was found using a multivari investigation to act in the car-to-car family. Despite further effort, the dominant cause was not found. The team decided to adopt the process robustness approach. Based on process experience, candidates and their corresponding levels were chosen as follows:

Candidate	Low level	High level
Anode dimension	3.1	3.9
Conductivity of paint	Low	High
Temperature	30	50
Zone X voltage	450	475
Zone Y voltage	500	525

The team selected a fractional factorial resolution V experiment with the 16 treatments given as follows. To reduce the cost of the experiment, panels were used rather than cars. With this choice there was a risk of study error.

Treatment	Anode dimension	Conductivity of paint	Temperature	X voltage	Z voltage
1	3.1	Low	30	450	500
2	3.9	Low	30	450	525
3	3.1	High	30	450	525
4	3.9	High	30	450	500
5	3.1	Low	50	450	525
6	3.9	Low	50	450	500
7	3.1	High	50	450	500
8	3.9	High	50	450	525
9	3.1	Low	30	475	525
10	3.9	Low	30	475	500
11	3.1	High	30	475	500
12	3.9	High	30	475	525
13	3.1	Low	50	475	500
14	3.9	Low	50	475	525
15	3.1	High	50	475	525
16	3.9	High	50	475	500

For each run, five panels were painted. The order of the treatments was randomized. Since the dominant cause acted car to car, the team believed the unknown dominant cause would act within each run. Film build was measured at five locations on each panel. The data for one location are given in the file *paint film build robustness* and in the table that follows.

Treatment	Order	Film build	Average	Log(s)
1	14	15.6, 15.3, 15.9, 15.2, 15.8	15.56	−0.51
2	5	16.0, 16.3, 17.3, 16.2, 16.6	16.47	−0.31
3	6	15.0, 14.8, 14.9, 15.3, 16.1	15.22	−0.28
4	2	16.1, 17.6, 17.2, 16.3, 16.1	16.69	−0.16
5	9	15.7, 15.6, 15.2, 15.2, 15.7	15.49	−0.57
6	12	17.3, 17.6, 16.8, 17.5, 17.3	17.28	−0.49
7	13	16.2, 14.4, 15.4, 14.5, 15.9	15.30	−0.09
8	4	17.3, 16.6, 16.6, 16.4, 17.8	16.94	−0.25
9	7	16.1, 14.7, 16.2, 14.7, 16.2	15.59	−0.09
10	16	17.2, 15.8, 16.4, 16.0, 15.8	16.23	−0.24
11	15	15.4, 15.2, 15.4, 15.3, 15.2	15.29	−1.06
12	1	16.6, 16.4, 16.4, 16.5, 16.4	16.48	−1.00
13	3	15.1, 15.4, 15.4, 15.0, 14.4	15.05	−0.41
14	10	16.8, 16.9, 17.0, 17.3, 16.3	16.89	−0.42
15	11	15.0, 15.1, 15.0, 14.9, 14.8	14.97	−0.86
16	8	16.6, 16.7, 16.3, 16.5, 16.3	16.48	−0.79

a. Analyze the data using the standard deviation of film build over the five consecutive panels to measure performance. Is it possible to make the process robust to noise variation? What levels of the candidates do you suggest?

b. The team had a way to adjust the process center. However, we can also use the robustness experiment to look for an adjuster. Analyze the data using the average film build over the five consecutive panels to measure performance. Are any of the candidates adjusters?

c. In the experiment, the film build at a particular location on five consecutive cars (panels) was used to define a run. Suppose, instead, that the five observations came from five fixed locations on a single door. What, if any, changes are needed in the analysis presented in part a?

19.2 In a trim plant, customer complaints about seat appearance prompted management to assign a team the task of reducing shirring variation. The team proceeded without using Statistical Engineering and made a number of mistakes. Seat cover shirring was scored on a scale of 1 to 6 using boundary samples by how much cloth was gathered by the parallel stitching. Shirring scores of 1 to 4 were acceptable with 1 being the best. Scores of 5 or 6 resulted from either too much or too little shirring. A review of historical data suggested that the observed shirring score over a week cov-

ered all six possible values. Next, the team informally checked the measurement system. They found the measurement system added little variation. The team decided not to look for a dominant cause. Rather they moved directly to assessing the feasibility of making the process robust. They used brainstorming to select six candidates with two levels each as follows:

Candidate	Low level	High level
Leather thickness	0.8	1.2
Leather toughness	Pliable (soft)	Stiff (tough)
Seam width	9 mm	11 mm
Material feed	Top up	Bottom up
Steam to skin bun	Used	Not used
Bun thickness	+5 mm	−5 mm

The team planned a resolution III fractional factorial experiment with 16 runs (one for each treatment) as follows:

Treatment	Leather thickness	Seam width	Leather toughness	Machine feed	Steam	Bun thickness
1	High	Low	Tough	Top up	Yes	High
2	High	Low	Soft	Bottom up	No	Low
3	Low	High	Tough	Top up	No	Low
4	Low	High	Soft	Bottom up	Yes	High
5	High	High	Tough	Bottom up	Yes	Low
6	High	High	Soft	Top up	No	High
7	Low	Low	Tough	Bottom up	No	High
8	Low	Low	Soft	Top up	Yes	Low
9	High	High	Tough	Bottom up	No	Low
10	High	High	Soft	Top up	Yes	High
11	Low	Low	Tough	Bottom up	Yes	High
12	Low	Low	Soft	Top up	No	Low
13	High	Low	Tough	Top up	No	High
14	High	Low	Soft	Bottom up	Yes	Low
15	Low	High	Tough	Top up	Yes	Low
16	Low	High	Soft	Bottom up	No	High

Each run consisted of three seats (repeats). The runs were conducted in the treatment order given in the table. The data are given in the file *seat cover shirring robustness* and reproduced as follows:

Treatment	Order	Seat 1	Seat 2	Seat 3	Average score
1	13	3	1	2	2.0
2	16	1	2	1	1.3
3	7	2	2	2	2.0
4	6	2	2	2	2.0
5	10	2	1	1	1.3
6	1	3	1	3	2.3
7	11	4	2	1	2.3
8	15	2	2	4	2.7
9	5	1	2	2	1.7
10	3	4	5	2	3.7
11	14	3	3	2	2.7
12	9	2	3	3	2.7
13	8	1	2	2	1.7
14	2	2	2	3	2.3
15	4	1	4	2	2.3
16	12	2	3	1	2.0

a. Explain why choosing the process output as a measure of variation (that is, high scores come from either too much or too little shirring) was a poor one.
b. The goal is to find process settings that lower the average shirring score. Can we use any of the candidates to achieve the goal?
c. Each run consisted of three seats. Discuss this choice in the context of a robustness experiment.

For the last two parts of this question, suppose the first three candidates (leather thickness, leather toughness, and seam width) used in the robustness experiment were normally varying rather than fixed inputs.
d. How should the levels of the first three inputs have been chosen?
e. Discuss changes you would make to the analysis you conducted in part b.

19.3 Torsional rigidity of the weather stripping was the dominant cause of door assembly problems. Management set a goal of reducing standard deviation in torsional rigidity to 0.3. A baseline investigation found the variation in torsional rigidity was

roughly 0.55 mm and that the dominant cause acted over the short term and certainly within any half hour. The team looked briefly for a dominant cause of rigidity variation without success. Next, they planned a robustness experiment with four candidates at two levels each, chosen based on engineering judgment. The candidates and levels are:

Candidate	Low level (−1)	High level (+1)
Heat (pre)	100	700
Extruder RPM	22	26
Tension (pre)	1	5
Water flow	2	6

The team planned a full factorial experiment with 16 runs, one for each treatment. The correspondence between treatments and candidate levels is given in the table that follows.

Treatment	Heat	Extruder RPM	Tension	Water flow
1	−1	−1	−1	−1
2	−1	−1	−1	1
3	−1	−1	1	−1
4	−1	−1	1	1
5	−1	1	−1	−1
6	−1	1	−1	1
7	−1	1	1	−1
8	−1	1	1	1
9	1	−1	−1	−1
10	1	−1	−1	1
11	1	−1	1	−1
12	1	−1	1	1
13	1	1	−1	−1
14	1	1	−1	1
15	1	1	1	−1
16	1	1	1	1

Each run consisted of running the process for half an hour after the candidate levels had been reached. Within each run, 10 weather-strip samples were selected spread out over the half hour. The order of the runs was randomized. The torsion rigidity of each of the 10 weather-strip samples for each treatment is given in columns s1 to s10 of the table that follows and in the file *weatherstrip torsional rigidity robustness*.

Treatment	Order	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
1	13	10.3	13.0	11.5	11.8	10.7	9.9	10.7	11.5	11.0	11.1
2	6	11.5	13.0	10.4	11.1	10.9	10.6	12.0	9.3	9.2	9.3
3	9	11.6	13.0	10.4	16.0	10.3	10.8	11.5	11.0	11.3	10.9
4	1	11.5	11.7	10.4	11.7	14.0	11.7	10.4	11.7	10.4	10.4
5	3	14.0	11.7	11.7	19.0	11.9	11.7	12.1	13.0	11.1	11.0
6	11	22.0	15.0	18.3	11.7	20.3	21.0	12.6	13.6	14.7	15.1
7	5	9.1	9.6	10.2	9.8	9.0	9.7	10.0	12.0	9.0	8.8
8	14	10.0	9.1	10.6	10.4	10.8	11.0	11.1	10.8	10.5	10.8
9	2	11.7	12.5	11.9	11.7	20.0	14.0	10.4	11.5	11.7	20.0
10	10	10.3	11.6	10.5	10.6	13.0	14.0	11.7	10.3	15.0	11.8
11	7	10.3	10.5	11.0	11.4	9.8	10.4	11.7	11.8	11.5	11.9
12	15	11.6	11.0	11.4	11.3	12.0	10.6	10.9	10.7	10.7	10.7
13	16	10.6	10.7	11.6	10.6	10.7	22.0	11.0	10.4	10.4	23.0
14	8	9.1	10.4	10.6	11.4	10.9	10.4	10.8	10.9	11.0	11.6
15	12	10.3	11.0	12.0	12.1	10.5	10.7	11.3	11.4	10.8	10.9
16	4	10.4	10.4	10.4	10.5	10.9	11.4	9.0	9.6	9.8	10.2

a. To analyze the results of this robustness experiment, what performance measure(s) do you recommend and why?
b. Analyze the experimental results using your chosen performance measure(s). What can you conclude?

CHAPTER 20—NO EXERCISES

CHAPTER 21

21.1 Discuss whether lessons learned are properly maintained in corporate memory in your organization. What could be done to improve the situation?

21.2 In the paint film build example described in Chapter 19, the team found new process settings that resulted in reduced car-to-car variation in film build. To validate the proposed solution, 80 cars were painted over one day with the settings given in the following table. These were the best settings found in the robustness investigation. The film build values from five specific positions on one door for each of the cars are available in the file *paint film build validation*.

Candidate	Setting
Anode dimension	3.5 (midpoint)
Conductivity of paint	High
Temperature	30
Zone X voltage	475
Zone Y voltage	500

   a. The baseline film build standard deviation was 0.68. The problem goal was to reduce the standard deviation to 0.35, and the robustness experiment results suggested that changing settings would reduce the standard deviation to about 0.37. Has the solution been validated?

   b. What, if anything, do the validation results tell us about the home of the dominant cause in the remaining variation?

21.3 In the truck pull example described in Chapter 17 and Exercise 17.1, a feedforward controller was implemented to compensate for the effect of truck-frame geometry on pull. After the feedforward system had been operating successfully for some time, management decided to review its operation. The four frame geometry measurements and left and right caster and camber were recorded for roughly a month of production consisting of over 6600 trucks. The data are given in the file *truck pull validation*.

   a. The standard deviations for caster and camber before implementation of the feedforward controller can be estimated from the 100-truck investigation described in Chapter 17. From the same investigation, the team predicted the possible reduction in standard deviation using a feedforward controller. A summary is given in the following table.

Characteristic	Baseline standard deviation	Predicted reduction in standard deviation
Left caster	0.90	0.18
Right caster	0.83	0.20
Left camber	0.51	0.13
Right camber	0.41	0.10

Do the results of the investigation validate the reduction in left and right caster variation due to the feedforward controller?

b. For each of the two caster characteristics, conduct a regression analysis to see if the feedforward controller can be improved. Recall that the feedforward controller should be compensating for variation in the frame geometry.

c. Repeat the analysis in parts a and b for left and right camber.

Exercise Solutions

CHAPTER 1—NO EXERCISES

CHAPTER 2

2.1 The word *variation* is used in other contexts to describe a difference between a realized and target value such as in budget variation. How does this use compare to variation as discussed in Chapter 2?

We define two types of *variation,* an off-target component and a part-to-part component. *Budget variation* only describes the off-target component of variation.

2.2 We have heard the following comment many times from manufacturing engineers: "The cause of the variation is the product design—what can you expect me to do?" Discuss the comment in light of the definition of cause in Section 2.2.

Product design cannot be the cause of part-to-part variation because it does not change from part to part. It is, however, possible that changes to the product design may solve the problem, though in most cases, the process produces good parts. This suggests improvement is possible with the existing design.

2.3 Profile A is a measure of deviation of the actual from the ideal shape of a camshaft lobe over one region (A) of the lobe. The target value is zero and the upper specification limit is 250 microns. Use the data in the file *camshaft lobe runout baseline* to summarize the variation in this output. Do all lobes exhibit the same variation? Is there any time pattern in the variation?

We use a combination of plots and numerical summaries. Some typical plots follow. The histogram (and the subsequent numerical summary) suggests the full extent of variation is 59 to 292.
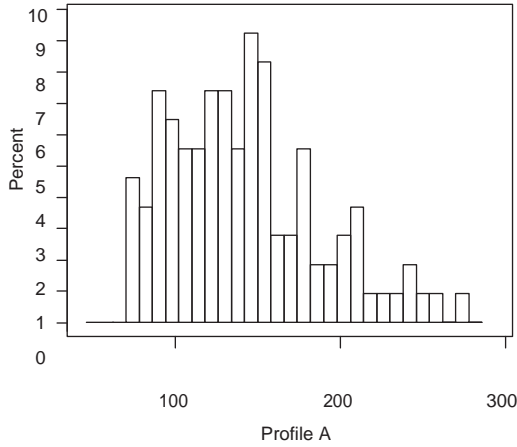
Variable	N	Mean	Median	TrMean	StDev	SE Mean
profile A	1296	136.32	130.00	134.54	39.83	1.11

Variable	Minimum	Maximum	Q1	Q3
profile A	59.00	292.00	104.00	164.00

Stratifying by lobe we see the average and variation in profile A is roughly the same across all 12 lobes.
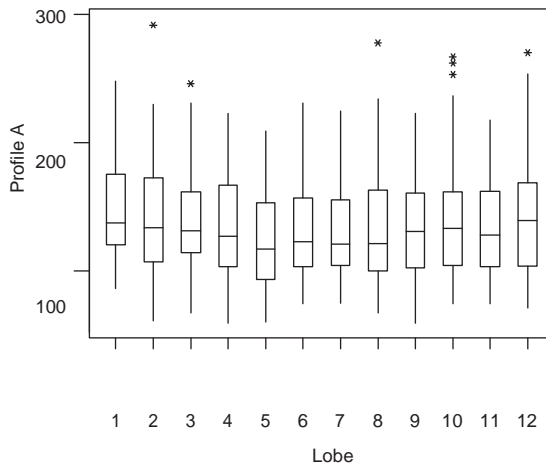
We need to be careful looking at the data by time. As stored, the 1296 profile A values come from 108 camshafts. The camshafts are given in production sequence.

We can plot profile A for each of the 12 lobes by camshaft in order of production. The plot shows the 12 values for each camshaft. There are no strong time effects that influence all lobes in the same way.
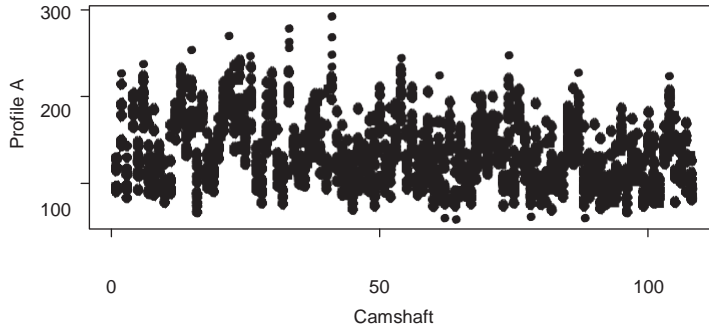
To determine if there is a time effect on individual lobes, we can look at run charts of profile A values for individual lobes. For example, for lobe 12 we get:
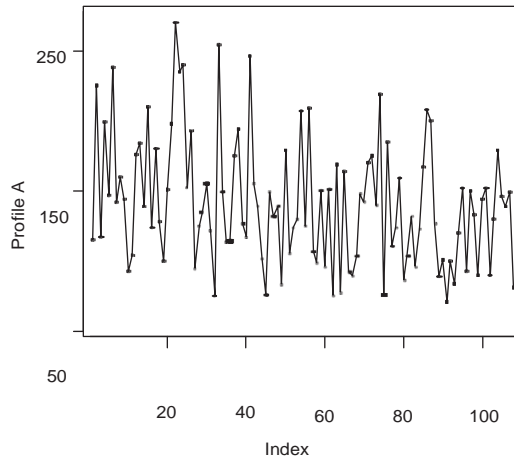
There are no time patterns evident for the different lobes.

A common mistake is to create the run chart of all 1296 profile A values as follows:
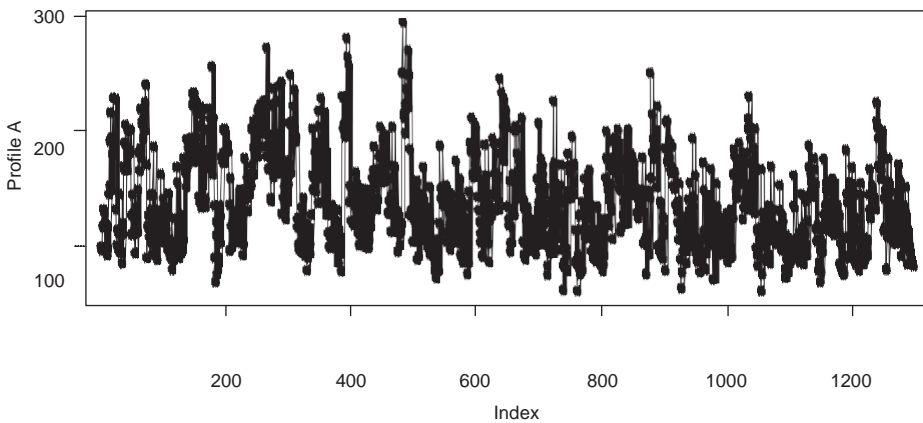
The observations are not ordered by time because each group of 12 values comes from the 12 lobes on a single camshaft. In this plot, large lobe-to-lobe effects could be misinterpreted.

2.4 Construct histograms and run charts for output 1 and output 2 given in the data file *chapter 2 exercise 4.* **Find the average and standard deviation for each output. Assume the target value and upper specification limit for these lower-is-better outputs are 0 and 35.**

    a. Is the variation the same for each output?
    b. Is the nature of the variation over time the same for each output?

    The results are:

Descriptive Statistics: output1, output2

Variable	N	Mean	Median	TrMean	StDev	SE Mean
output1	100	11.634	10.713	11.364	4.882	0.488
output2	100	11.634	10.713	11.364	4.882	0.488

Variable	Minimum	Maximum	Q1	Q3
output1	3.017	34.488	8.539	14.041
output2	3.017	34.488	8.539	14.041
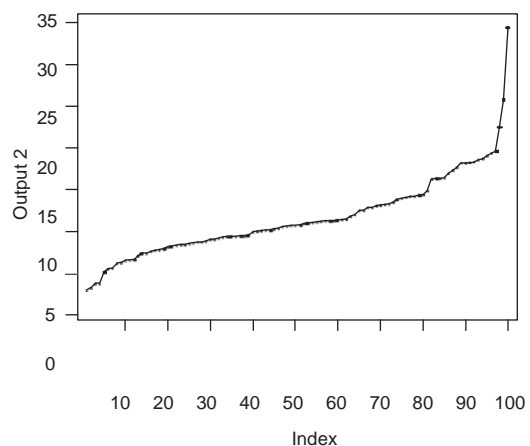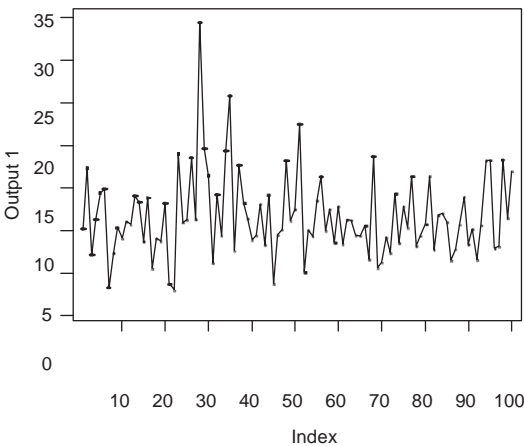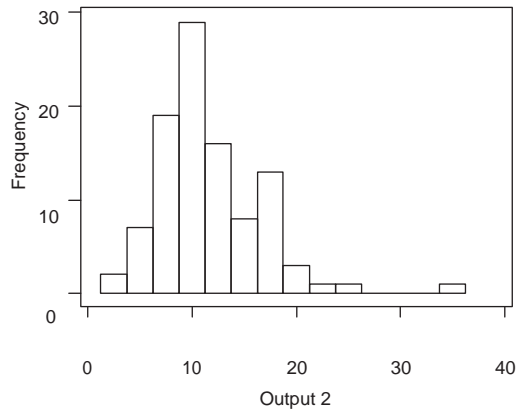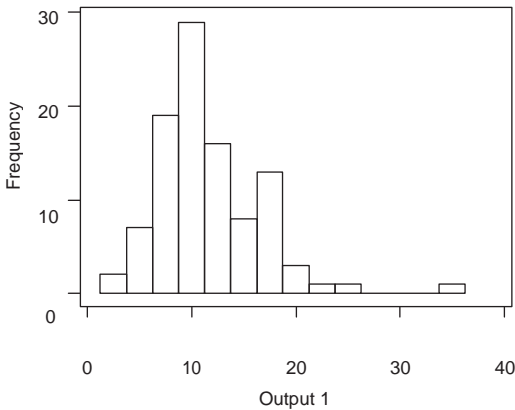
    a. The standard deviations and histograms are the same for both outputs.
    b. Output 2 is ordered from smallest to largest over time.

2.5 **You may convince yourself that the formulas for combining means and standard deviations given in Section 2.4 are true with the following numerical demonstration you can conduct in MINITAB. Generate two columns of 100 values sampled from some model (in MINITAB: Calc → Random Data → your choice of model and parameters, for example, Normal with mean and standard deviation 0 and 1, respectively). Then, calculate two new data columns. Let one column be the sum of the original two columns and the other the difference.**

   a. Find the standard deviation and average for each of the four columns.
   b. Calculate the sum and differences of the averages for the first two columns. How do these compare to the average of the other two columns respectively?
   c. Calculate the standard deviation for the sum and difference using the "square root of sum of squares" formula given by 2.1. How do the results compare to the standard deviations for the last two columns?

   a. An example of the results you will get is:

Descriptive Statistics: C1, C2, sum, diff

Variable	N	Mean	Median	TrMean	StDev	SE Mean
C1	100	0.061	0.016	0.055	1.126	0.113
C2	100	–0.0976	–0.0853	–0.0904	0.9145	0.0915
sum	100	–0.037	–0.226	–0.048	1.571	0.157
diff	100	0.158	0.222	0.167	1.320	0.132

Variable	Minimum	Maximum	Q1	Q3
C1	–3.214	2.801	–0.787	0.802
C2	–2.3400	2.3015	–0.6814	0.5067
sum	–3.482	4.000	–1.181	1.221
diff	-3.134	3.308	-0.746	1.115

   b. The sum and difference of the averages for the first two columns equal the average of the sum and difference columns, respectively.

   c. For the sum (and difference) of C1 and C2 we have

$$stdev(C1 \pm C2) = \sqrt{stdev(C1)^2 + stdev(C2)^2}$$

In the example data $\sqrt{1.126^2 + 0.9145^2} \approx 1.45$, which closely matches the standard deviation of the last two columns.

2.6 **At a project review, the team presented the following summary of their investigation based on standard deviations.**

Source of variation	Percent of total
Measurement system	30
Identified cause	50
Unidentified causes	81

   a. **The reviewing manager questioned the numbers in the second column of the table because they did not add to 100. Is there an error? Explain.**
   b. **By what percentage can the process standard deviation be reduced by eliminating the contribution of the identified cause?**
   c. **Is the identified cause a dominant cause?**

   a. There is no error. The percentages are calculated on the standard deviation scale. Recall that standard deviations combine using the square root formula as illustrated by Equation (2.1). The total is given by $\sqrt{0.30^2 \boxed{}0.50^2 \boxed{}0.81^2 \boxed{}}$.
   b. If we eliminate the contribution of the identified cause the remaining variation is given by $\sqrt{0.30^2 \boxed{}0.81^2 \boxed{}}0.85$. So the process standard deviation could be reduced by roughly 15%.
   c. No.

2.7 **In Chapter 1, we discussed a project to reduce variation in pull, an alignment characteristic of light trucks. Recall that**

> **Pull = 0.23*(right caster – left caster) + 0.13*(right camber – left camber)**

and that the data for two months' production are stored in the file *truck pull baseline*. The data are summarized in the following table.

Output	Average	Standard deviation
Left camber	0.257	0.129
Right camber	0.249	0.130
Left caster	3.519	0.224
Right caster	4.519	0.243
Pull	0.231	0.082

   a. **Use the formula for pull and the results for how averages and standard deviations combine to predict the average and standard deviation for pull given by the last row in the table indirectly from the component averages and standard deviations.**
   b. **Suppose you had the resources to reduce the variation in one of the alignment angles by 50%. Which angle would you choose? By how much, approximately, would the pull standard deviation be reduced?**

   a. The derived standard deviation for pull is

$$\sqrt{0.23^2 * \boxed{0}224^2 \boxed{}0.243^2 \boxed{}0.13^2} \qquad \boxed{}0.08$$

The derived standard deviation for pull is not exactly 0.082, because the alignment angles do not vary independently; there is a small correlation. Note that the averages of the components play no role in the standard deviation.

The derived average is:

$$0.23(4.519+3.519)+0.13(0.249+0.257)=0.229$$

There is some rounding error, so this value does not match the pull average in the table.

b. The largest reduction in overall standard deviation would be achieved by reducing the variation in the right caster. Reducing the variation in right caster by 50% would reduce the variation in pull to roughly

$$\sqrt{0.13^2 *(0.13^2+0.129^2)+0.23^2 *(0.224^2+0.121^2)}=0.063$$

This corresponds to approximately a 20% reduction.

CHAPTER 3

3.1 For a problem of interest to you speculate about the likely costs and feasibility of implementing each of the possible variation reduction approaches.

The solution is dependent on the chosen problem.

3.2 Variation in the location of a drilled hole in a machined casting can cause poor fits when the part is bolted to an engine. To reduce this variation, an engineer considers a variety of possible approaches.

    a. A vision system is available that can measure location on 100% of the parts and reject those that it judges to be out of specification. What are the advantages and disadvantages of such an approach?

    b. Institute a feedback controller by measuring two parts every hour. If hole location on either part is outside of specification, stop and adjust the process. When is such a scheme likely to be effective?

    c. A third choice is to find a dominant cause of the variation. What are the advantages and disadvantages of this strategy?

    d. If a dominant cause can be discovered, what options does the engineer have?

a. 100% inspection would ensure that no out-of-specification parts were shipped to the customer, assuming there were no measurement errors. The inspection system may be expensive to install and run. Also, we need to specify how to handle rejected parts.

b. For feedback control to be effective, the short-term variation in hole location must be substantially smaller than the hour-to-hour variation. There must also be a way to adjust the center of the process.

c. Finding a dominant cause of hole location variation would be valuable information that may lead directly to a low-cost solution. However, finding a dominant cause may be difficult or expensive. Also, the dominant cause may be outside the control of local management or difficult to control.

d. With knowledge of a dominant cause, the engineer can consider the variation reduction approaches that require knowledge of a dominant cause, namely: fixing the obvious, desensitizing the process, and feedforward control. Any of the non-cause-based approaches are also still options.

CHAPTER 4—NO

EXERCISES CHAPTER 5

5.1 Briefly discuss the advantages and disadvantages of the following—be sure to think of potential errors as described within the QPDAC framework.

a. To estimate the baseline performance of a grinding process, 100 consecutive pistons were sampled and the diameters were measured.

b. To investigate a proposed change to a chemical process, the investigators tried the change in a pilot process rather than the production process.

a. The given plan results in quick and easy data collection. However, there is a danger that process variation will be underestimated since 100 consecutive pistons may be more similar than 100 pistons chosen over a longer time frame (study error). Also, 100 pistons is a relatively small sample. See Table S6.1 in the supplement to Chapter 6 to get a better idea of the uncertainty in the estimate of the process standard deviation from a small sample size (sample error).

b. An investigation on the pilot process would be cheaper and easier than using the production process. The main concern is whether results from the pilot process can be scaled up to the regular process (study error.)

5.2 In the camshaft lobe BC runout problem described in Chapter 1, the team selected 50 parts (10 per day over 5 days) and measured the BC runout for each of the 12 lobes on each camshaft to quantify the baseline. The 600 runout measurements are stored in the file *camshaft lobe runout baseline original*. Conduct an analysis of these data. Are your conclusions different from those in Chapter 1? Why?

From the MINITAB results that follow, we see that the variation in BC runout as measured by the standard deviation is now much larger at 31.7. However, looking at the numerical or graphical summary, we see that the data contain a large outlier. In particular, observation number 74 is 1130. A transcription error put the decimal point in the wrong place. The value should have been 11.3. This mistake was readily identified and corrected before proceeding with the rest of the analysis.

Descriptive Statistics: `BC runout`

Variable	N	Mean	Median	TrMean	StDev	SE Mean
BC runout	1296	13.51	11.10	12.28	31.69	0.88

Variable	Minimum	Maximum	Q1	Q3
BC runout	2.60	1130.00	7.43	17.00

5.3 To assess a measurement system used to check the diameter of an engine bore, an investigator plans to repeatedly measure the same four (of the eight) bores on five blocks sampled from a shift of production.

  a. Discuss the advantages and disadvantages of using 10 rather than 5 blocks.
  b. In the investigation, all the blocks produced over one shift were available for study. Give two considerations that the investigators should take into account in making the choice of available blocks.
  c. The plan was to make all measurements in a single day. Discuss the advantages and disadvantages of making the measurements over a longer time period.
  d. When would the investigator be better off devoting the available resources to measuring all eight bores on fewer engine blocks?

  a. Using 5 blocks rather than 10 would be cheaper and quicker. However, with fewer blocks there is a greater chance of sample error. The performance of the measurement system on the chosen 5 blocks may be different (better or worse) than its performance on other blocks.
  b. The investigator needs to trade off cost and convenience with the risk of study error. The key question is whether the performance of the measurement system is likely to be different when examining blocks from other days. It may be, for example, that engine bores from a single day have similar dimensions and that the measurement system works well only for some dimensions.

c. Again the investigator needs to trade off cost and convenience with the risk of study error. Here the key question is whether the properties of the measurement system change substantially over time. Measurement systems, like other processes, change over time unless properly maintained. As we shall see in Chapter 7, we recommend, if at all possible, that measurement investigations compare the measurements over a longer time period.

d. Having fewer engines makes the logistics of the investigation easier. In the measurement investigation to repeatedly measure the bores on the same block, the team had to move the block back into the process before the gage. Measuring all the bores on fewer blocks would be a good idea if we expected the measurement system to be sensitive to the bore number (though if this were known, another option would be to focus on the known problem bores). If, on the other hand, the dimensions of all the bores on a particular block were similar, using fewer blocks might prevent us from seeing the full range of bore dimensions in the measurement investigations.

5.4 **You are a manager with the responsibility to decide if you should change the supplier for a tooling insert. You receive a report from your process engineer who has conducted an investigation into a new insert. He gives you the following verbal report and recommendation:**

Our current insert has an average life of 1105 parts. To assess the performance of the new supplier, we asked them to supply 10 inserts. We checked the inserts out on one of our machines last week and got an average of 1300 pieces. Since the cost is the same, I think we should switch to the new inserts.

Using the QPDAC framework, think of five questions you would ask about the conduct of the investigation before you might accept the recommendation.

You should ask questions like:

Is the machine used for the investigation typical of all machines we use?

Did the manufacturer supply 10 specially good inserts? How can we tell?

How much variation was there in the number of pieces cut by the 10 new inserts?

What data were used to estimate the average life of the current inserts?

Are the environmental and production conditions in the week of the investigation of the new inserts typical?

CHAPTER 6

6.1 In Chapter 1, we described a problem in terms of the lobe geometry of camshafts. The data are given in the file *camshaft lobe runout baseline*. Quantify the problem baseline for the following output.

   a. BC runout
   b. Angle error

   a. We summarize the data using a histogram and a run chart as follows. The run chart plots the 12 lobe BC runout values for the 108 camshafts in production order.
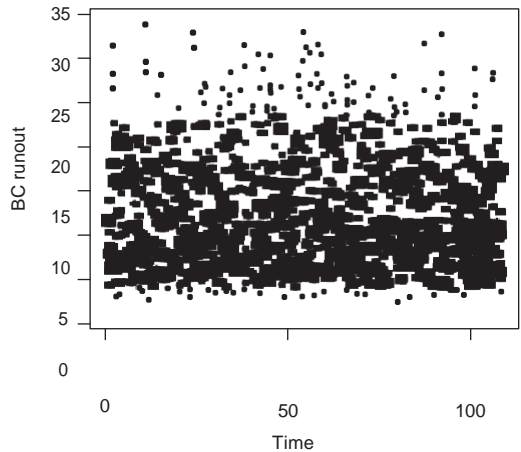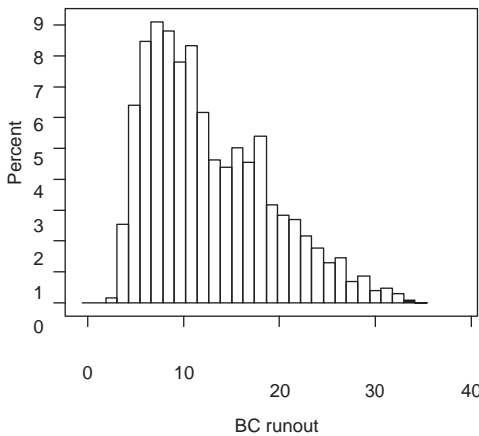
Variable	N	Mean	Median	TrMean	StDev	SE Mean
BC runout	1296	12.643	11.100	12.271	6.389	0.177

Variable	Minimum	Maximum	Q1	Q3
BC runout	2.600	33.900	7.425	17.000

   The baseline standard deviation is 6.4, and the full extent of variation is 2.6 to 33.9 microns. We see no obvious patterns over time. We can also look at run charts for each lobe separately.

   b. Looking at angle error we use the same summaries as in part a.
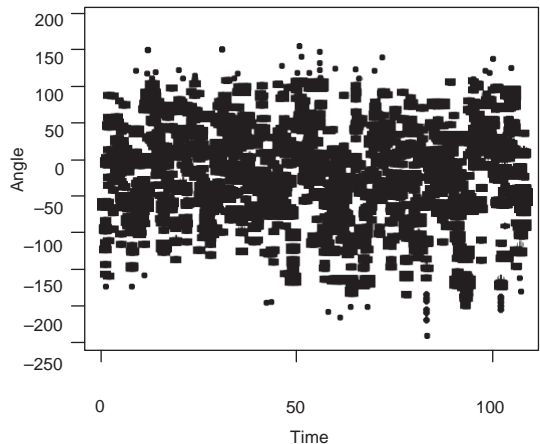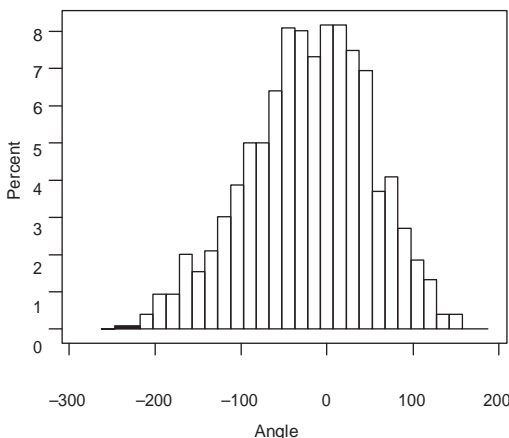
Variable	N	Mean	Median	TrMean	StDev	SE Mean
angle	1296	–21.30	–18.00	–20.04	71.50	1.99

Variable	Minimum	Maximum	Q1	Q3
angle	–241.00	155.00	–67.00	30.00

The baseline standard deviation is 71.5 and the full extent of variation is roughly –240 to 155. There is a weak pattern over time.

6.2 Many programs such as Excel cannot easily handle missing observations. MINITAB is an exception. Missing values are often stored using a special numerical code (–99 is common). These special codes can result in much confusion and lead to incorrect conclusions. Consider the data *rod thickness baseline with missing observation*. In the file, there are two outputs. The output thickness_–99 uses a numerical code of –99 for missing observations, while thickness_missing uses the MINITAB missing observation symbol (*). Quantify the baseline for these two outputs. Which data summaries show the missing observation and which do not?

If we forget we are using a special code for a missing observation, it is not readily apparent in the numerical summary and the –99 inflates the standard deviation by about 9%. The code for the missing observation is easily identified in the graphical display if the code is extreme relative to the usual values. Using a missing observation symbol is much preferred over a special numerical code.

Variable	N	N*	Mean	Median	TrMean	StDev
Thickness_–99	800	0	34.426	36.000	34.832	11.981
Thickness_miss	799	1	34.593	36.000	34.861	11.018

Variable	SE Mean	Minimum	Maximum	Q1	Q3
Thickness_–99	0.424	–99.000	59.000	28.000	43.000
Thickness_miss	0.390	2.000	59.000	28.000	43.000
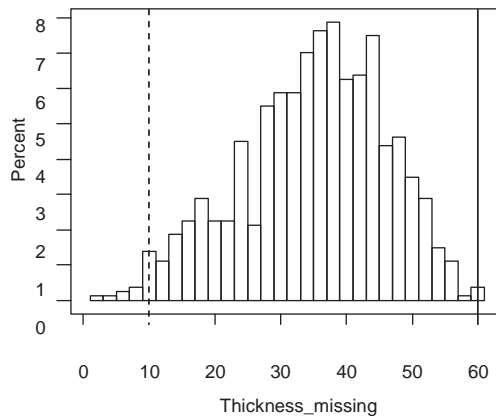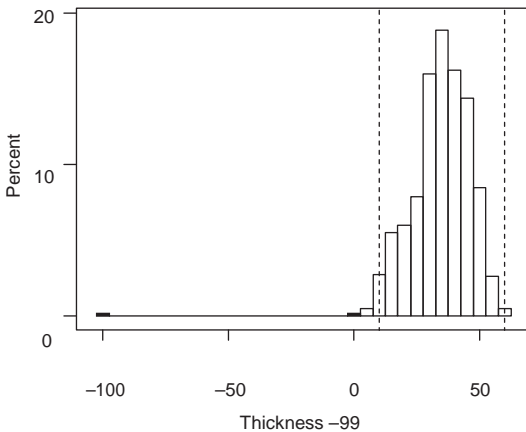
6.3 The baseline investigation for the V6 piston diameter example was described in Chapter 5. The data are given in the file *V6 piston diameter baseline*. Suppose the data were collected so that all the pistons from a given hour were collected at the start of the hour. Now the data come in subgroups as defined by hour. What summaries used in the baseline analysis are affected by the subgrouping? When taking the subgrouping into account are the conclusions any different than those derived in Chapter 5?

The subgrouping in the data affects the interpretation of the run chart. The run chart (as given in Chapter 5) assumes the observations are equally spaced in time. Since the data are collected so that the five observations in each subgroup are much closer together in time then the observations from subsequent hours a better display would plot the diameters by hour (time).
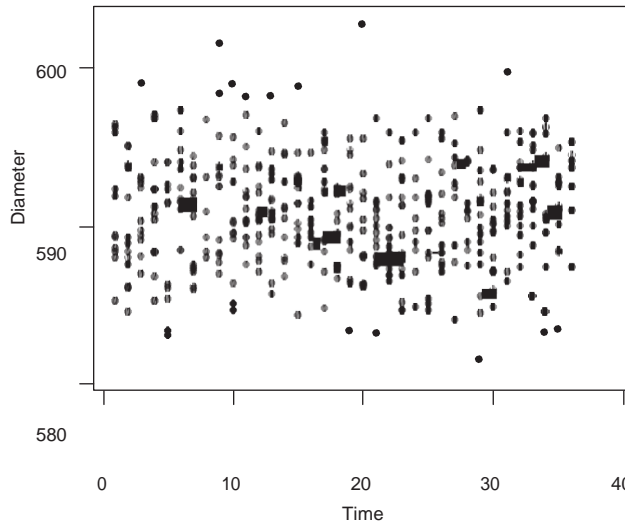
There appears to be no special pattern over time. The conclusions for the baseline would be unchanged.

6.4 Based on customer complaints concerning installation difficulties, a team investigated variation of a key fascia dimension. To establish a baseline, they measured the dimension on 147 fascias sampled from one month's production. The data are given in the file *fascia dimension baseline*. Using appropriate summaries of the data, quantify the baseline. Are there any concerns?

We summarize the baseline data numerically and using a histogram and run chart.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
dimension	147	7.303	7.000	7.256	2.261	0.186

Variable	Minimum	Maximum	Q1	Q3
dimension	2.000	12.000	6.000	9.000
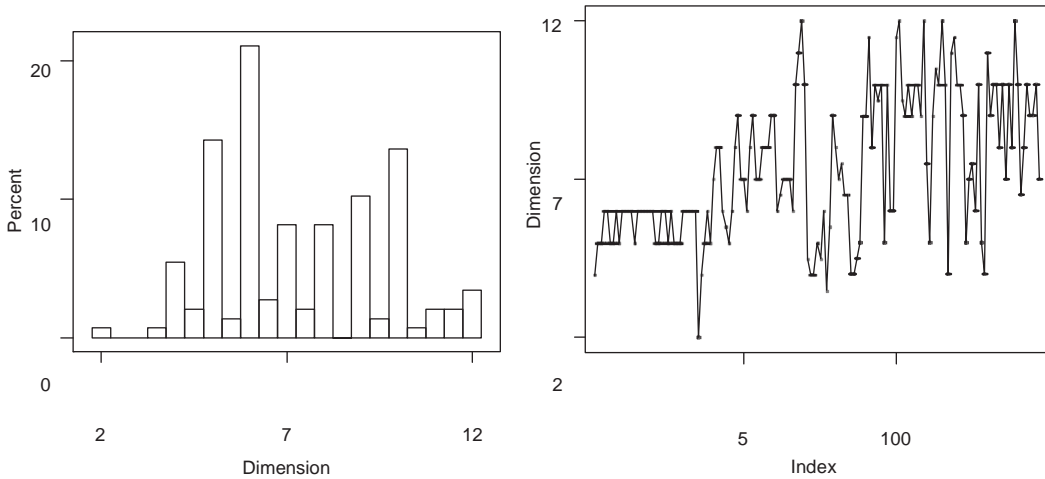
There is some evidence of an increase in the fascia dimension over the month of the baseline investigation. This suggests the investigation was not conducted over a long enough period of time. We need to plan a new baseline investigation.

CHAPTER 7

7.1 In a process improvement problem to improve the quality of a roof panel, the measurement system (specially designed for the project) counted the number of updings on each panel. To assess the measurement system, the number of updings on 20 bad panels and 20 good panels was counted twice. The data are given in *roof panel updings measurement*.

    a. Can this investigation be used to assess the measurement variation of the counting process? Explain.

    b. Can this investigation be used to assess the bias of the counting process? Explain.

    c. The same operator counted all panels. Does the order in which he makes the counts matter? It is most convenient to count the same panel twice in a row. Is this a good idea?

    d. A scatter plot of the first versus the second measurement is given as follows. Note that some plotting symbols correspond to more than one pair of measurements? What does the scatter plot tell you about the counting process?

e. **This investigation was conducted over one hour. What are the advantages and disadvantages of spreading the two measurements on each panel over two days?**

f. **Can the counting process discriminate between good and bad panels?**

a. Yes. By measuring each panel twice we can assess measurement variation. However, we may underestimate the variation if important causes of measurement variation do not vary sufficiently during the investigation.

b. No. We cannot assess bias since we do not know the true upding count for each panel.

c. It is best to randomize the order of the panels. The danger with measuring the same panel twice in row is that on the second count the operator will remember and be influenced by the first count.

d. The points lie close to the 45° line and there is a clear separation between good and bad panels. This suggests the measurement system is adequate.

e. Using two days for the investigation reduces the risk of study error but takes longer.

f. Yes. See part d.

7.2 **To monitor the process that produces engine blocks, piston bore diameters are measured on every block because they are key characteristics. Each engine block has eight bores. The bore diameter is measured at three different heights in each bore (bottom, middle, and top) and at two different orientations at each height. Because the measurement process is automated, there are no operators. A measurement investigation was conducted over a day where the diameter of every bore on four blocks was measured four times each. The main concern was out-of-round, given by 10,000 times the difference of the two diameters at a particular height. The data are given in the file *block bore diameter measurement*. From a baseline investigation the out-of-round standard deviation was 22.8.**

> **a. Determine the discrimination ratio. Is the measurement system adequate?**
> **b. What would have been the advantage and disadvantage of conducting the measurement investigation over a longer time period?**

a. Out-of-round is measured four times at each of 96 locations (4 blocks by 8 bores by 3 heights). To analyze the data, we define a new characteristic (called "block_bore_height") to uniquely identify each of the 96 locations. Plotting out-of-round by this new characteristic we get
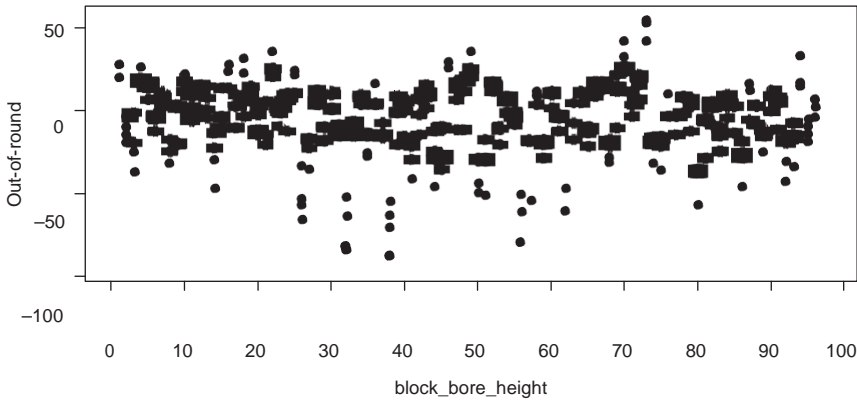
The measurement variation is roughly comparable across the different blocks, bores, and heights. To estimate the measurement variation from this data we use a one-way ANOVA analysis; part of the results follow.

Analysis of Variance for out-of-round

Source	DF	SS	MS	F	P
block_bore_height	95	177032.7	1863.5	24.94	0.000
Error	288	21516.3	74.7		
Total	383	198549.0			

Pooled StDev =   8.643

The variation due to the measurement system is estimated as 8.6. We cannot assess the stability of the measurement system here. The order of measurements was not preserved and, in any case, all the measurements were conducted over a single day.

Using the baseline standard deviation we can solve for an estimate of the standard deviation attributable to the process. We get 21.1 ($= \sqrt{22.8^2 - 8.643^2}$). The measurement system discrimination ratio is thus 2.44, and we conclude that the measurement system is not the dominant cause but that the measurement system should be improved before proceeding with the Statistical Engineering algorithm.

b. Spreading the measurements over a longer time period would have delayed reaching a conclusion about the adequacy of the measurement system. An advantage is that the longer time would have allowed slowly changing causes to act. From the

baseline investigation, the team knew that the dominant cause acted within a day. Thus, a single day was long enough for the measurement investigation.

7.3 **The following MINITAB results and graphs arise from a measurement system investigation in which two different operators measured five parts three times each. The five parts were selected with initial measured values spread out over the full extent of variation, 0 to 8. The data are given in the file** *chapter 7 exercise 3*. **The two operators worked different shifts so the parts were saved from one shift to the next. The results include an edited ANOVA analysis as suggested in the supplement to Chapter 7 and the default gage R&R analysis in MINITAB.**

Analysis of Variance for measurement

Source	DF	SS	MS	F	P
part	4	230.819	57.705	81.25	0.000
Error	25	17.754	0.710		
Total	29	248.573			

Pooled StDev =   0.8427

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.900	8.62
Repeatability	0.425	4.07
Reproducibility	0.475	4.55
operator	0.475	4.55
Part-To-Part	9.547	91.38
Total Variation	10.447	100.00

Source	StDev (SD)	Study Var (5.15*SD)	%Study Var (%SV)
Total Gage R&R	0.94876	4.8861	29.35
Repeatability	0.65207	3.3582	20.17
Reproducibility	0.68917	3.5492	21.32
operator	0.68917	3.5492	21.32
Part-To-Part	3.08975	15.9122	95.59
Total Variation	3.23214	16.6455	100.00

Number of Distinct Categories = 5

Gage name:
Date of study:
Reported by:
Tolerance:
Misc:

Gage R&R (ANOVA) for measurement

a. **What do the given results tell us about the bias and variation of the measurement system?**

b. **In the gage R&R results, the $\bar{X}$ chart by operator is out of control. What does this mean?**

c. **In the gage R&R results, why is the sum of the % study variation column not 100%?**

d. **What is the discrimination ratio (D) for this system? How does the part selection procedure influence this ratio?**

e. **The gage R&R is about 29%, yet D is small. Why?**

f. **The results suggest a small operator-to-operator difference. This observed difference may be due to a difference in method or a drift of the system over the two shifts. How can you separate these two possibilities?**

a. We estimate the standard deviation due to the measurement system as 0.84, the pooled standard deviation from the ANOVA. This calculation corresponds to the formula given in Section 7.2. We cannot assess measurement bias because the true dimensions are not known. As an aside, it is always a good idea to also plot the measured values versus part number as follows.
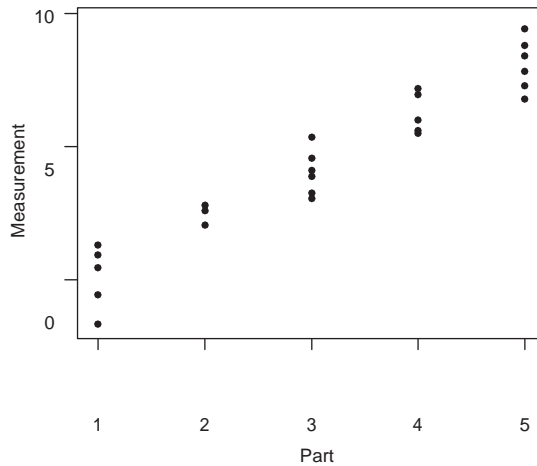
b. The points plotted on the $\bar{X}$ chart are the average measurement for a particular part and operator. The control limits are determined by the variation of the measurements within the 10 part and operator combinations. If the measurement system is able to distinguish among the parts, there should be many points outside the control limits. Here, out-of-control signals indicate a good measurement system.

c. The column partitions the percent variation due to the measurement system and the parts with the overall standard deviation as the divisor. To combine standard deviations we need to square and add. We see that $29.35^2 + 95.59^2 = 100^2$.

d. To calculate the discrimination ratio, we can estimate the overall standard deviation from the full extent of variation as $8 / 6 = 1.33$ (we are assuming a bell-shaped histogram in this calculation). Thus we can estimate the standard deviation due to the process as $1.03 = (\sqrt{1.33^2 - 0.84^2})$. The corresponding estimate for the discrimination ratio D is 1.22 ($= 1.03 / 0.84$). The selection of the parts plays no role in the calculation of D. We assume that the variation within each part is the same regardless of true size.

e. The denominator of the gage R&R calculation is the variation observed in the investigation. In this case, the overall standard deviation is 2.93, much larger than the estimated baseline variation, which decreases the R&R. The reason for the large overall variation is the part selection procedure.

f. With the plan, there is no way to distinguish the operator and shift effects (we say the effects are confounded). If we want to be able to separate the effects, we need another investigation in which both operators measure parts at the same time over several shifts.

7.4 **To assess the variation in the system designed to measure camshaft lobe geometry over time, the same camshaft was measured daily for a month. At each measurement, a number of characteristics (for example, angle error, BC runout, taper, and so on) on each lobe were determined. The data are given in the file** *camshaft lobe runout measurement stability*. **Is there evidence of time-to-time variation in this measurement system?**

Checking for time-to-time variation is difficult due to the large number of output characteristics and lobes. We plot each output versus day, stratifying by lobe number.
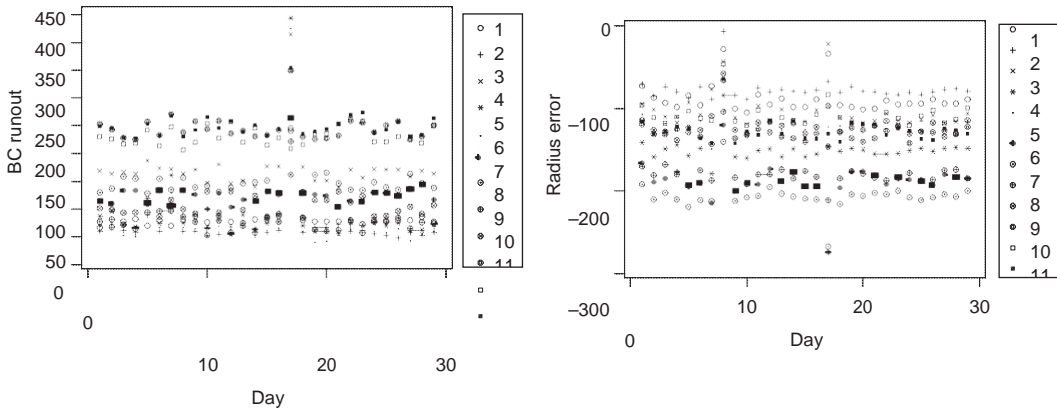
Examining a run chart of any characteristic versus the order in the MINITAB worksheet may be misleading since much of the observed pattern may be due to differences between lobes.

Most of the characteristics show a pattern similar to the one observed for BC runout, where the gage appears stable except for the results on day 17. Radius error has unusual results for day 8.

7.5 In a process that produced V8 pistons, problems occurred when pistons in inventory were remeasured (for an audit) and found to be out of specification. Since the process used 100% final inspection, this could only occur if there was a problem with the measurement system. This was puzzling because a recent gage R&R investigation at the final gage had concluded that the measurement system was acceptable. As a result, the team decided to conduct a long-term measurement investigation. Two pistons were chosen to span the range of diameter normally seen. Each piston was measured four times a day (spread out over the day) for 12 days. During that time the regular gage calibration was performed every four hours. The data are given in the file *V8 piston diameter measurement stability*.

a. Does the measurement system drift over time?
b. What effect does the regular gage calibration have?

a. To assess drift we plot the diameter for each piston over time.
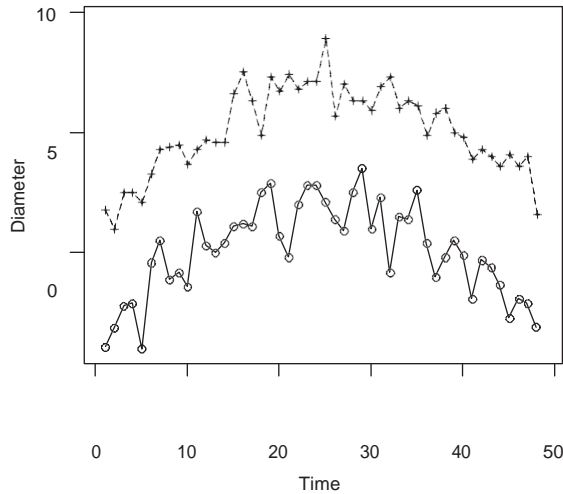
There is clear evidence of a drift over time. Note that since the pistons were chosen to span the normal range of diameters, the amount of drift seen is substantial.

b. The gage was recalibrated a number of times during the investigation. However, looking at the run chart, the regular calibration appears to have no effect.

7.6 **Consider the brake rotor balance example described in the case studies. In the measurement investigation, three rotors were specially selected: one well balanced, another poorly balanced, and the final rotor requiring weight near the specification limit of 0.5. The three rotors were measured twice by each of the three gages on three separate days. There is no operator effect since the gages are automated. The 54 measurements are given in *brake rotor balance measurement*. The analysis given in the case study focuses on the measurement of the weight needed to balance the rotor. However, the location (or orientation) of the weight needed to move the rotor's center of gravity is also important. Can the measurement system consistently determine the orientation of the required balance weight? From the baseline investigation, the orientation of the weight was roughly uniform from 0° to 360°.**

With an output like orientation measured in degrees we must be careful when analyzing the data. The data are circular since 0° is the same as 360°. Plotting the data for orientation by the three rotors we see that for this data we can ignore the circular nature of the output since none of the measured orientations are near 0°.
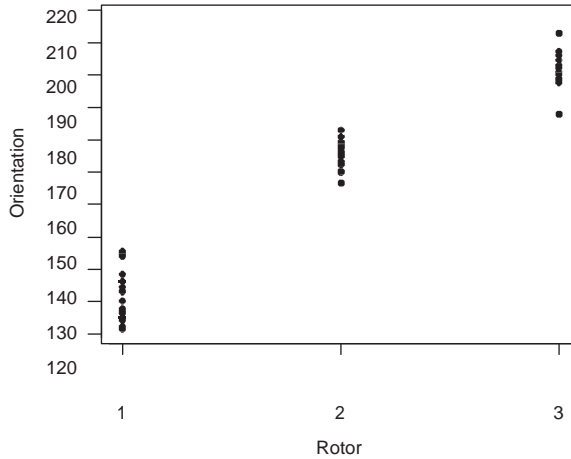
A one-way ANOVA analysis gives:

Analysis of Variance for orientation

Source	DF	SS	MS	F	P
rotor	2	43340.3	21670.1	589.70	0.000
Error	51	1874.1	36.7		
Total	53	45214.4			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	---------+---------+---------+-------
1	18	132.57	8.11	(*)
2	18	175.47	4.08	(*)
3	18	201.26	5.27	(-*)

Pooled StDev =   6.06   150   175   200

The measurement variation is estimated as roughly 6°, the pooled standard deviation. The team decided that the measurement system was adequate, since the measurement variation was small relative to the full extent of variation.

7.7 **If necessary, measurement variation can be reduced by applying the Statistical Engineering algorithm. Describe how each of the seven variation reduction approaches might be used to improve a measurement system.**

Fix the Obvious

If there are substantial operator-to-operator or gage-to-gage differences, we may reduce measurement variation by using fewer operators or gages or by training the operators.

Desensitization

With desensitization, we need to change some input in the measurement system that makes it less sensitive to variation in a known dominant cause of measurement variation. Desensitization was the chosen approach to improve eddy current measurements

of rotor hardness. The team explored changing a number of settings to see if it was possible to make the measurement system less sensitive to variation in the amount of dirt on the casting, the dominant cause of measurement variation. Ultimately, in this example, the approach failed (see the discussion in Chapter 16).

Feedforward Control

The use of feedforward control to improve a measurement system is difficult to imagine. One idea is to use different gages depending on the class of part we are to measure. This would be effective in reducing measurement variation if each measurement system performed better for one of the classes of parts. For instance, suppose we wish to measure the thickness of foam. Say the current measurement system (hand calipers) worked well for narrow foam, but not very well for thick foam. Then, to implement feedforward, when foam of different grades required measurement we would choose the best gage for the job.

Feedback Control

We can use a feedback control scheme if a measurement system drifts. To set up the controller, we select a reference part and measure it repeatedly in the short term to determine the average and the standard deviation. Then to implement, we measure the same part on a periodic basis and adjust the measurement system if the characteristic value is materially different from the average. We use the short-term variation to determine if the change is material.

Robustness

Maintenance on a measurement system can be thought of as the Robustness approach. We hope to reduce the measurement variation without knowledge of the dominant cause. Another idea to reduce measurement variation that uses the robustness idea is to take two or more measurements on each part and average the obtained results. This will reduce the measurement variation, at the expense of increased measurement costs, if the repeated measurements are independent.

Inspection

Inspection is not an option to improve a measurement process. We have no way to determine which measurements should be rejected.

Move the Process Center

Moving the process center is trivial for a measurement process. We just add some value to all measured values. The Move Process Center approach is appropriate to eliminate a known and persistent measurement bias.

CHAPTER 8—NO

EXERCISES CHAPTER 9

9.1 Think of a process and problem you know well. Define various families of causes.

The solution depends on the process and problem chosen.

9.2 The following plot shows the results of a process investigation aimed at finding a dominant cause. The dashed lines give the full extent of variation in the output as defined by the problem baseline. Can the input be ruled out as a dominant cause of variation?
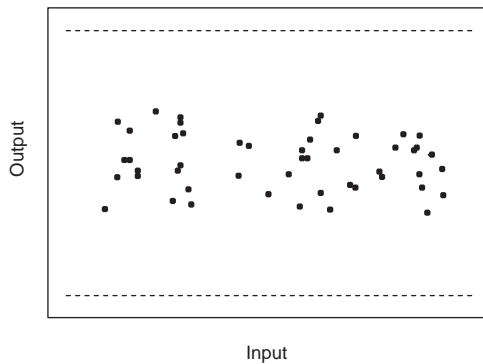
The simple answer is no, because we have not seen the full extent of variation. However,

• If the full range of variation in the input is bigger than that observed, the input may be a dominant cause.

• If the full range of variation in the input has been observed, the input is not a dominant cause on its own but may be dominant in combination with another input (one that has not varied over its full range of variation in the investigation).

For instance, the input/output relationship may look like the plot that follows, where different plotting symbols are used to distinguish between two possible values for another input. In the original investigation, the level of the other input was restricted to the values given by solid circles.

CHAPTER
10

10.1 The flow chart that follows shows the major steps in an assembly process to set the wheel alignment of a truck.

Components   Measurement

   The characteristic of interest is right camber with specification $0.5 \pm 0.5°$. Camber is measured on every truck by one of the four gages (aligners). The process performance for right camber is shown as follows based on about 6200 consecutive trucks.

a. **Based on this histogram, can the measurement system be eliminated as a dominant cause of the camber variation?**
b. **What data could you collect to demonstrate that a dominant cause does not act in the measurement system?**
c. **How could you rule out the assembly operation as the home of a dominant cause?**
d. **How could you eliminate differences in the suppliers of the upper control arm as the home of a dominant cause?**
e. **The plot that follows shows the process behavior over three shifts. What family of causes can be eliminated based on these data?**

f. **The following plot shows the camber variation for the first nine trucks in the data set. What families can be ruled out as the home of a dominant cause using these data?**

g. **In a special study, one key characteristic of the lower control arm was measured for 30 trucks. The other components were specially selected to ensure that they were well within specification. Based on the plot that follows, is the lower control arm characteristic a dominant cause of right camber variation? Explain.**

a. No. The measurement variation contributes to the range of values in the histogram but there is no way to tell from this investigation how much.

b. We need to conduct an investigation of the measurement system where right camber is measured more than once on a number of trucks.

c. If we are able to disassemble and reassemble the alignment components repeatedly without damage, we can assess the assembly operation. If disassembly/ reassembly does not change the right camber values much, we eliminate the assembly operation as the home of a dominant cause. Repeated disassembly and reassembly using the production assembly process is difficult in this application.

d. If we were able to track which trucks contained the upper control arms from the different suppliers, we could stratify the right camber values by control arm supplier. If the control arm supplier was a dominant cause, the average right camber values in the two groups would be quite different, relative to the full extent of variation.

e. We are seeing roughly the full extent of variation in the output (right camber) within a short time frame. With this evidence we eliminate the family of causes that acts over the long term.

f. We have not observed the full extent of variation in right camber. It is premature to draw conclusions about the home of the dominant cause.

g. While the relationship between the lower control arm characteristic and right camber appears strong, we have not observed the full extent of variation in the right camber. We may ask if the range of lower control arm characteristic values observed in the investigation is typical of the process.

10.2 Consider again the camshaft lobe runout problem introduced in Chapter 1. Each camshaft has 12 lobes with a number of characteristics of interest. In a search for a dominant cause, we may compare the lobe-to-lobe and camshaft-to-camshaft families of variation. Using the problem baseline data given in the file *camshaft lobe runout baseline,* **explore the relative sizes of the two families for the following characteristics and decide which family, if any, can be eliminated.**

a. **Profile** *A*
b. **Profile** *B*
c. **Profile** *C*

We can graphically compare the lobe-to-lobe and camshaft-to-camshaft families of causes by stratifying the baseline data by lobes using box plots. The observations for each lobe show the effect of the camshaft-to-camshaft family, while comparing the box plots for different lobes shows the effect of the lobe-to-lobe family.

a. For profile *A,* the camshaft-to-camshaft family is the home of a dominant cause. There is very little difference from lobe to lobe in either the average or variation of profile *A.*
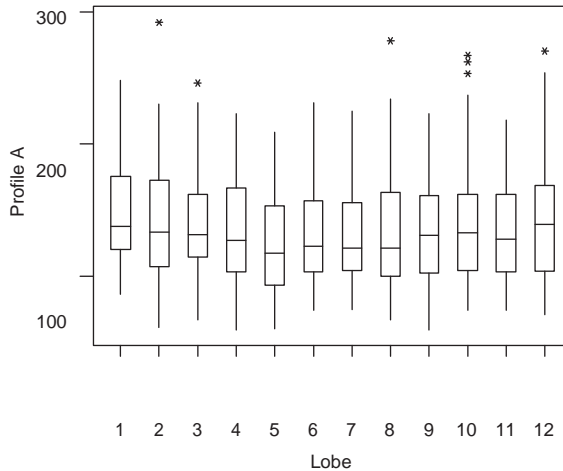
b. For profile *B,* we conclude the dominant cause acts in the camshaft-to-camshaft family. There is, however, more variation (though still small) attributable to the lobe-to-lobe family than for profile *A.*
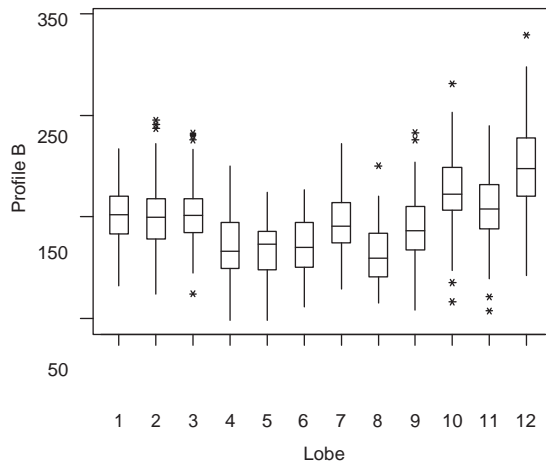
c. For profile *C,* the middle lobes exhibits less variation (and lower average) than for lobes on the end. The dominant cause is an interaction between lobe number and some, as yet, unidentified cause that acts in the lobe-to-lobe family. Note that ANOVA results corresponding to this plot will not reflect the differences in variation by lobe. The ANOVA analysis focuses on differences between the profile *C* averages across lobes.
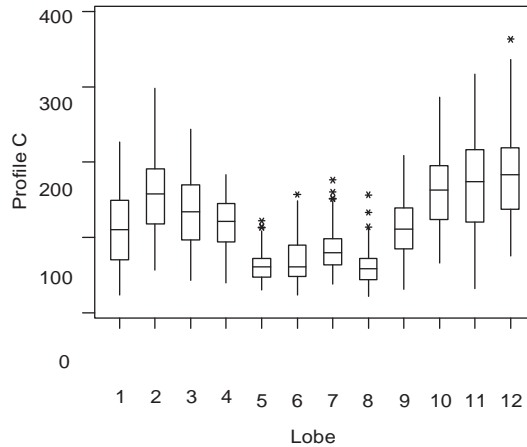
10.3 **In the manufacture of an injection molded part, a key crossbar dimension exhibited excess variation. The problem baseline estimated the standard deviation of the crossbar dimension as 0.46 with full extent of variation –0.3 to 2.0. The goal was to reduce the standard deviation to less than 0.25. An investigation showed the measurement system to be highly capable.**

   Next the team conducted a multivari investigation where five consecutive parts were sampled every 30 minutes for four hours. Analyze the data given in *crossbar dimension multivari*. **Which family of variation can be eliminated as the home of the dominant cause?**

From the plot of crossbar dimension by time (given as follows) we see the time-to-time family contains the dominant cause of variation. The dominant cause is acting over the 30-minute time periods, but not over consecutive parts, that is, not in the part-to-part family. We can eliminate the part-to-part family.

10.4 **As described in Chapter 7, in a process that placed labels on bottles, the team searched for an acceptable measurement system. The file** *label height measurement* **contains the data from an investigation in which three operators using a hand feeler gage measured three specially chosen bottles three times on two**

> **different days. The bottles were chosen to roughly cover the range of label height values seen in the process. From a baseline investigation an estimate of the overall standard deviation was 0.022. The results of a one-way ANOVA are:**

Analysis of Variance for height

Source	DF	SS	MS	F	P
part	2	0.0413263	0.0206631	263.10	0.000
Error	51	0.0040054	0.0000785		
Total	53	0.0453317			

```
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level   N      Mean     StDev    ----+---------+---------+---------+--
1       18    0.06966   0.00965  (-*-)
2       18    0.10950   0.00930                  (-*)
3       18    0.13705   0.00749                             (-*)
                                 ----+---------+---------+---------+--
Pooled StDev = 0.00886              0.075     0.100     0.125     0.150
```

> **We have stdev(due to measurement) = 0.00886, and thus**

$$stdev\left(\text{due to process}\right) = 0.0204 \left(= \sqrt{0.022^2 - 0.00886^2}\right)$$

> **and an estimated measurement discrimination ratio of 2.3. The team decided to improve the measurement system before addressing the original label height variation problem. Reanalyze the measurement investigation results to eliminate families of possible dominant causes of measurement variation.**

We start our analysis by calculating the heights minus the part average since we want to look for sources of variation other than part. Plotting height minus part average versus operator and day we get the plots given as follows:
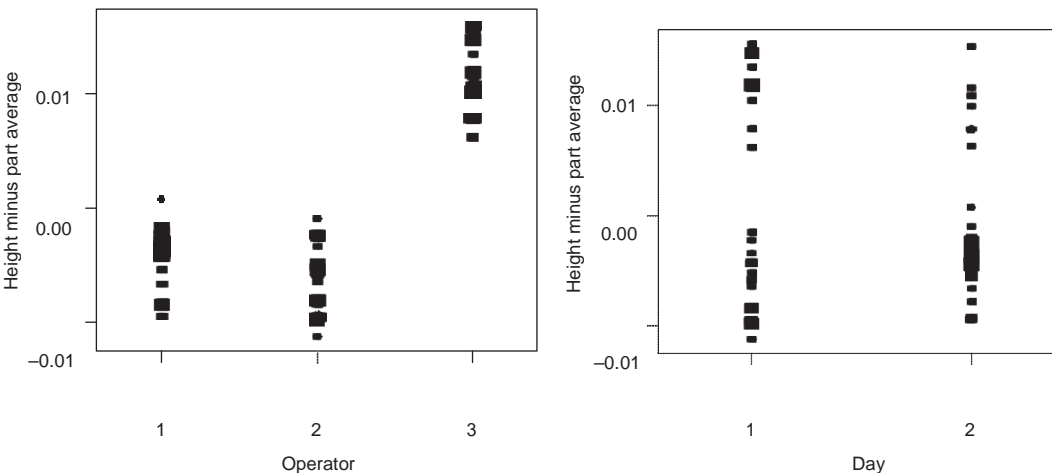
We see a clear difference in average height between operators and no difference between days. Operator three consistently gives larger values than the other two operators. This relative bias was found to be due to differences in procedure used by the three operators. With additional training the relative bias between operators was eliminated, and the measurement discrimination ratio was increased to around seven.

10.5 **A process improvement problem was initiated to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. The team discovered that the dominant cause was contamination before the forming process step. In an investigation, the team measured the particle count on coils directly after the arrival from steel supplier and again after blanking and stamping (before the forming process). They measured at the tail, middle, and head of four different coils. The data are given in the file** *roof panel updings variation transmission*. **What does the following scatter plot tell us about the dominant cause? The plotting symbols correspond to the four different coils.**

Assuming the investigation has seen the full extent of variation in the output (which we are not given in the exercise description), we conclude the dominant cause acts in the raw blanks. The blanking, stamping, and forming process steps all transmit but do not add variation.

10.6 **In the engine block porosity example discussed in Chapter 10, the team found the occurrence of high porosity coincided with production directly after breaks. To explore this clue further, they conducted another investigation in which the porosity of 50 consecutive blocks was measured. The first six blocks were taken from directly before the lunch break, and the next 44 blocks were the first ones produced after the break. The data are given in the file** *engine block porosity run chart*. **What does the observed pattern in the run chart tell us about the dominant cause?**
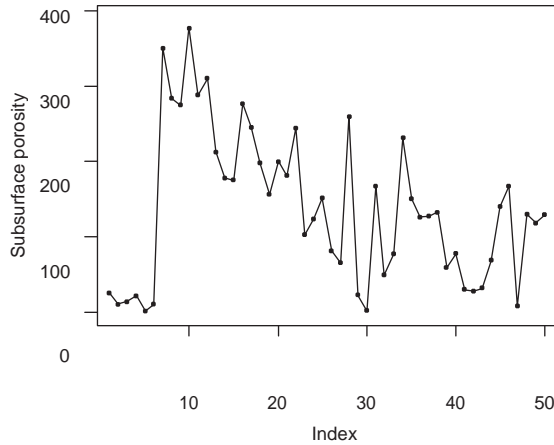
All six blocks from before lunch have extremely low porosity. Immediately follow-ing the break, the first six blocks have the highest porosity. The porosity gradually trends downward. This pattern implies the dominant cause is some input that changes abruptly during the breaks and then gradually returns to the original level. The return time is greater than 44 blocks. The team concluded that pouring temper-ature was a suspect dominant cause. The temperature of the iron decreased during breaks because there was no external heat source in the pouring ladles. The pouring temperature gradually increased after a break since the pouring ladles are frequently replenished with hot iron.

10.7 **High silicon concentration in cast iron is undesirable as it was found be a dom-inant cause of fluidity variation. However, measuring the silicon level can be difficult. The measurement process consisted of sampling the molten iron by pouring sample coins for testing. The coins are then machined and polished before being spectrochemically analyzed. The full extent of variation in percent silicon as measured in the current process was 1 to 4%. The measurement sys-tem was investigated by repeatedly measuring three different coins that roughly covered full extent of variation in the observed percent silicon. Two operators measured each of the three coins three times on each of two days. The data are given in the file *iron silicon concentration measurement*. Analysis of the meas-urement results estimated the measurement standard deviation as 0.33. The corresponding estimate of the process standard deviation was 0.5; thus the dis-crimination is too small at around 1.5. The team needs to improve the measure-ment system. Using the existing measurement investigation data, are there any clues about the dominant cause of the measurement variation?**

We start our analysis by calculating the silicon concentration minus the average for each coin since we want to look for sources of measurement variation other than dif-ferences from coin to coin. Some plots of the data are given as follows.
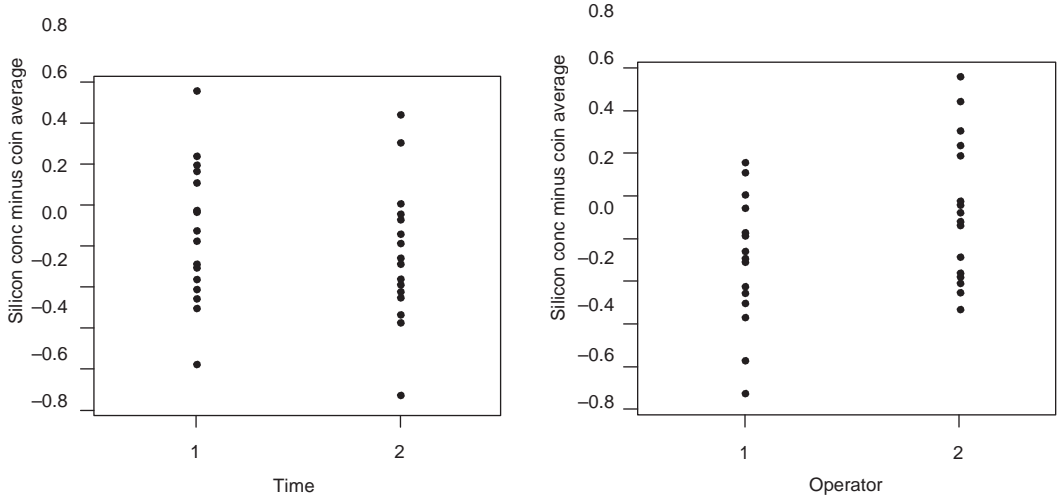
We see a small difference between the two operators, but no dominant cause. The dominant cause is acting over the short term. Using other investigations, the team was unable to determine the dominant cause of measurement variation. In Chapter 19, we describe a robustness investigation used to look for changes in this measurement process that would result in less variation.

CHAPTER 11

11.1 In a multivari investigation, two consecutive pieces are selected from each of three pallets once per hour. Sketch the appearance of the multivari chart that shows all three families at the same time if a dominant cause lies in the following family. Use the following multivari chart template in which the dashed lines indicate the full extent of variation.

    a. Pallet-to-pallet family
    b. Part-to-part family
    c. Hour-to-hour family
    d. An interaction between the part-to-part and pallet-to-pallet families
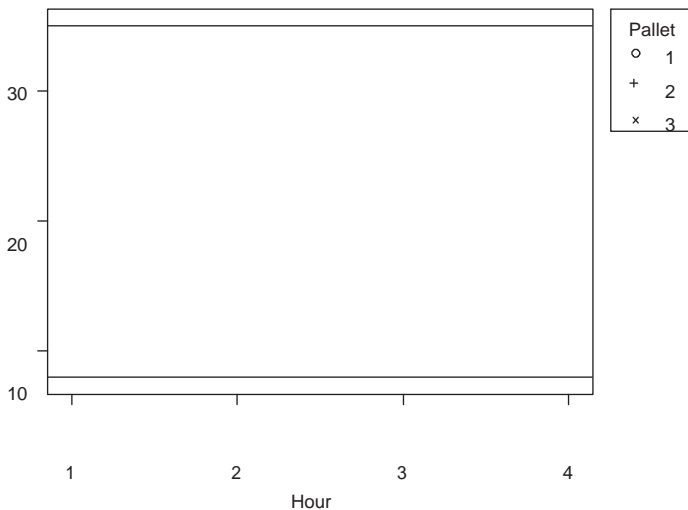
We give examples of the possible multivari charts, where the output name matches the question part. We suppose the full extent of variation is 8 to 35 units.
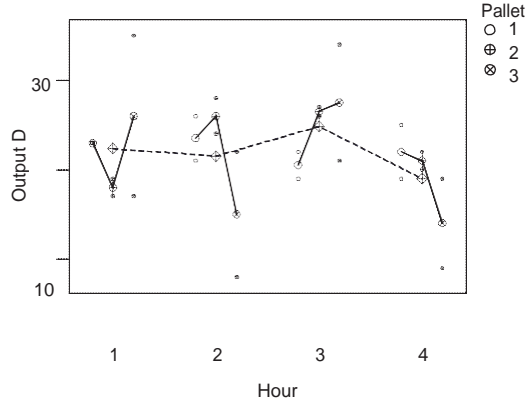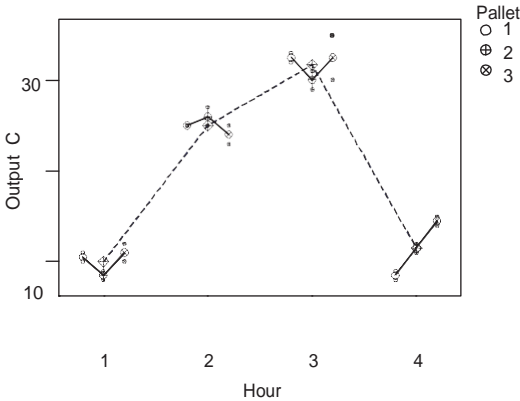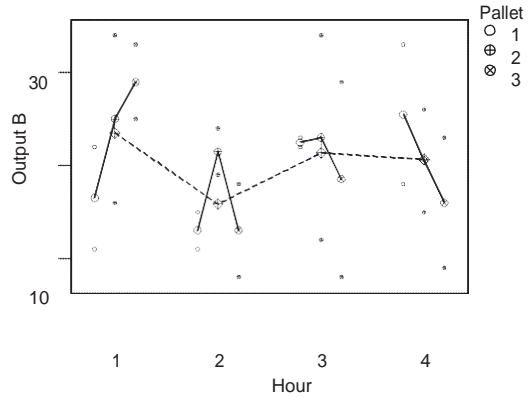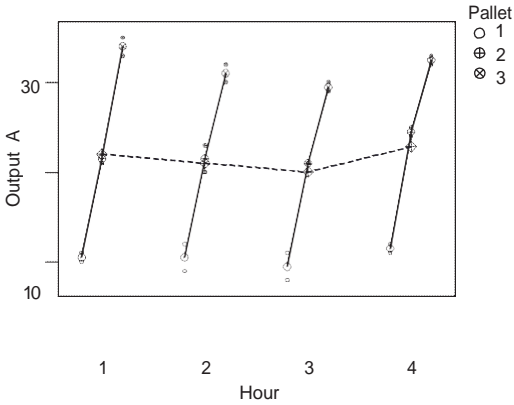
Note in the multivari chart for output *D,* the output has the most variation for pallet 3.

11.2 In the engine block leakers example, introduced in Chapter 1, the baseline defect rate was 2–3%. The team conducted a multivari investigation where three consecutive blocks were taken at twelve different times throughout the day. The investigation continued for three production days giving a total of 108 castings. Each block was tested for leaks. The data are given in the file *engine block leaks multivari*. What can you conclude?

The resulting multivari chart is:
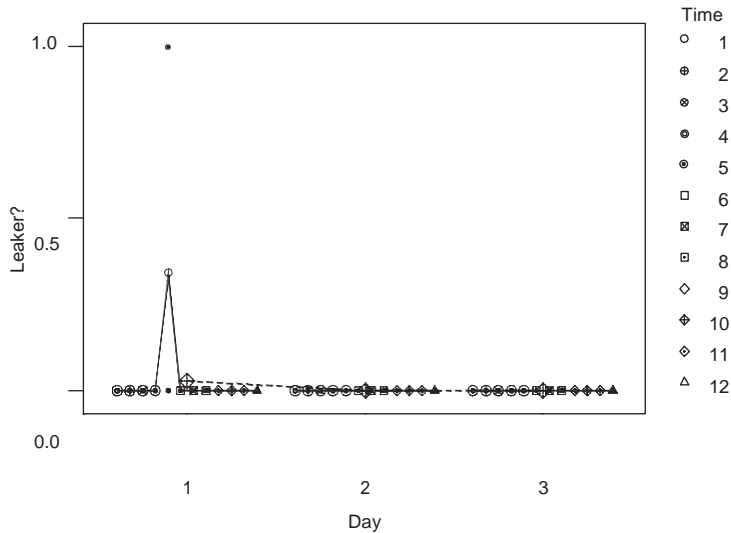
There is only a single leaker in the 108 blocks. This was not surprising consider-ing the low baseline leak rate. This plot does not provide much information. In gen-eral, multivari investigations are not effective for binary output characteristics unless large sample sizes are used and the proportion defective is plotted.

11.3 **At an intermediate operation the team planned a multivari investigation in which three consecutive parts were taken from each of two machines operating in parallel once every hour for two days. Consider two different processes. In the first process, the order of the parts coming from upstream is preserved, while in the second process the order is jumbled. When interpreting the result-ing multivari chart (think specifically about the part-to-part family), what dif-ference does it make which process we are observing?**

In the first process, the part-to-part family includes inputs that act from part-to-part upstream and in the intermediate operation. In the second process, the part-to-part family includes inputs that act part to part in the intermediate operation and all other upstream causes. For example, causes that acted in the time-to-time family upstream appear to act in the part-to-part family because of the jumbling.

11.4 **In a multivari investigation, the diameter of a transmission shaft was measured at four positions (left and right side at two different orientations) for three con-secutively sampled shafts each hour. The data are available in the file** *transmission shaft diameter multivari*.

a. What conclusion can you draw from the multivari charts that follow?
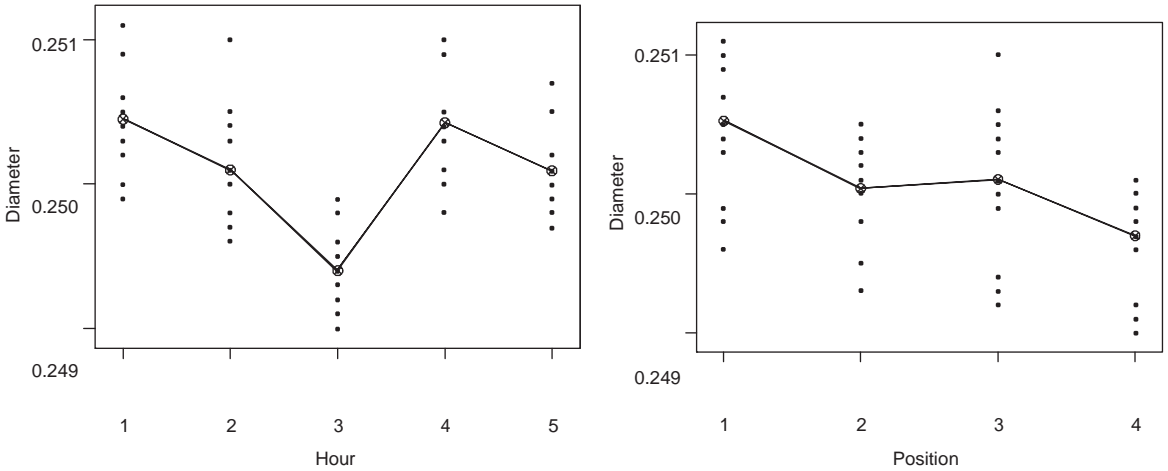
b. Using the data assess whether the dominant cause acts in the shaft-to-shaft family.

a. From the two multivari charts, we see that a dominant cause acts in the hour-to-hour family.

b. To assess the shaft-to-shaft family we define a new variate group to uniquely identify the 20 groups of three shafts that are consecutive. The multivari chart using group (=[hour – 1] · 4 + position) is:
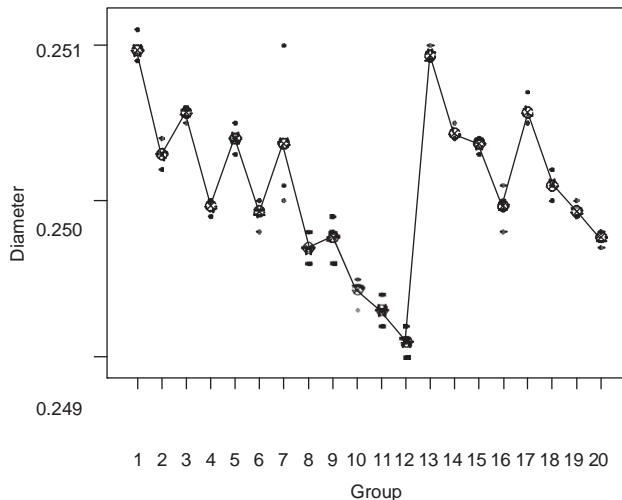
The variation from shaft to shaft inside each group is small, so the shaft-to-shaft family does not contain a dominant cause.

11.5 In the production of engine blocks, bore diameters are key characteristics. Bore diameter is measured at three heights and two orientations in each of the eight bores in each block. The team used Statistical Engineering to address a problem

of excess bore diameter variation. The baseline investigation found a standard deviation of 3.04 and the full extent of variation of –9 to 9 as measured from nominal in microns. There were no strong differences between the different bores, heights, or positions. Another investigation concluded that the measurement process was adequate. To isolate the processing step where the dominant cause acts, the team selected 30 engine blocks haphazardly from a day's production. In the investigation the bore diameter (measured from nominal at that processing step) in the first bore at the top position and first orientation was measured at each of five processing steps in the machining part of the process. The data are given in the file *block bore diameter variation transmission*. Which processing step is home to the dominant cause?

The plots showing incoming and outgoing diameters (from nominal) are given as follows. The output is labeled based on the operation after which it is measured.
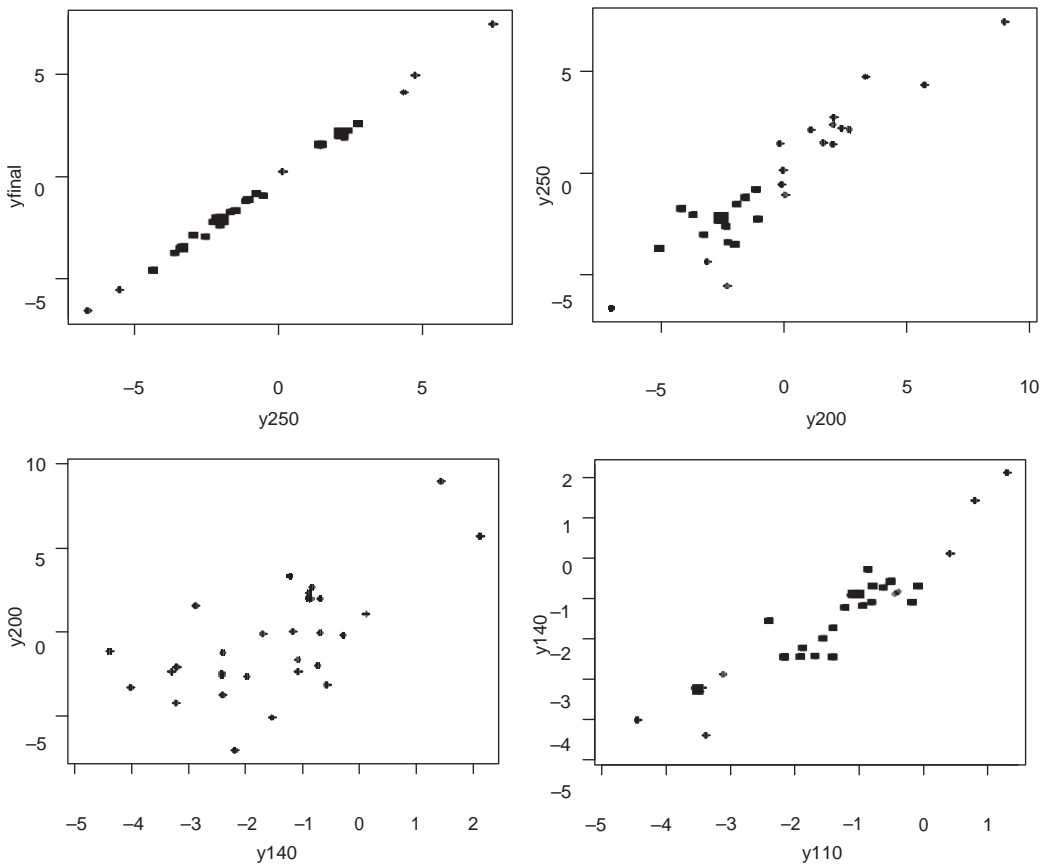
Starting from the end of the process, we see that little variation is added between operation 250 and the final diameter measurement, and also little variation is added between operations 200 and 250. Based on the plot in the lower left hand side, the dominant source of variation acts between operations 140 and 200 (including possibly operation 200 itself).

11.6 **In the paint film build example described in Chapter 3, a baseline investigation found the standard deviation was 0.315, with an average of 16.2 thousandths of an inch. The full extent of variation was 15.2 to 18.5. To search for a dominant cause, the team conducted a multivari investigation where they measured the film build at five positions on five cars in a row every hour for two shifts (16 hours). This resulted in a total of 400 film build measurements. The data are given in the file *paint film build multivari*. Based on the plots that follow, what conclusions can you draw? We define group as (hour – 1) · 5 + position.**
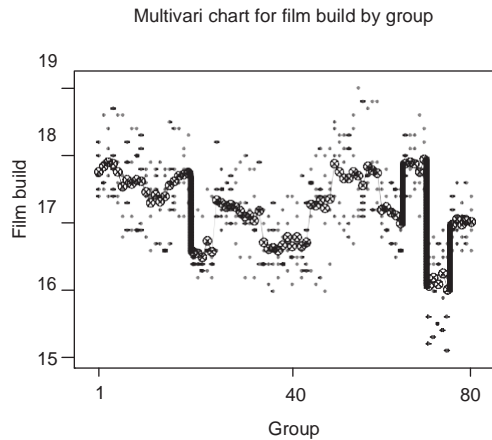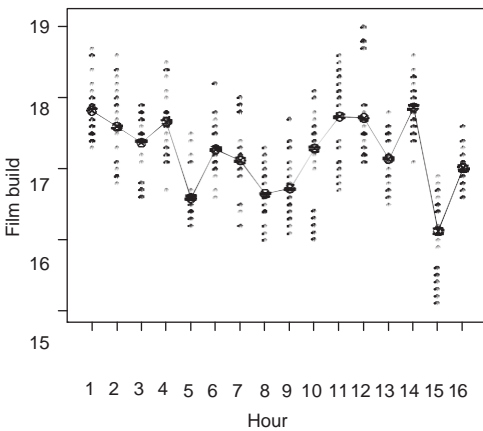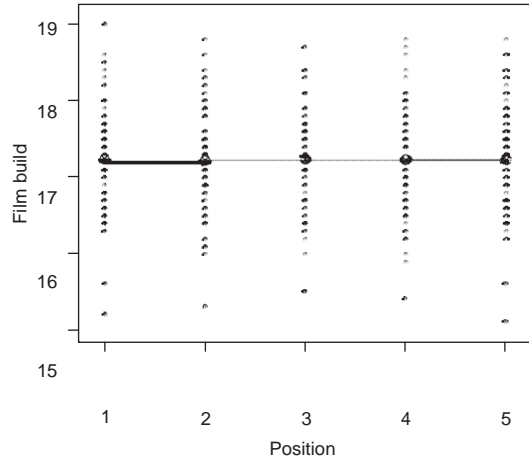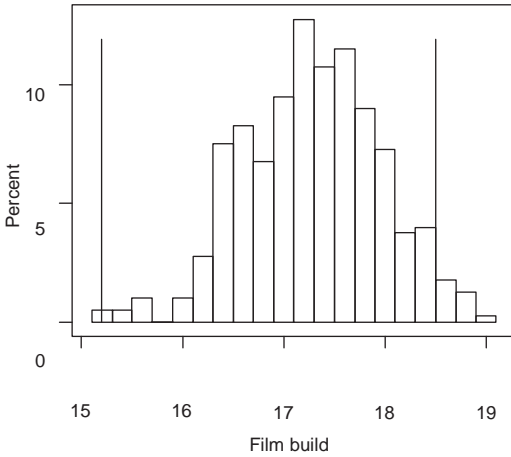
The histogram of the multivari data covers more than the full extent of variation, so we know the dominant cause has acted. From the multivari charts, we conclude that the dominant cause acts in the hour-to-hour family. We can use a one-way ANOVA to estimate the within-hour standard deviation as 0.47, much less than the baseline standard deviation 0.68. In other words, if we could make the average thickness the same at each hour, we could reduce the variation substantially.

11.7 **A team wanted to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. A baseline**

investigation found that the total number of updings in 20 consecutive panels ranged between 5 and 438. To search for a dominant cause the team conducted a multivari investigation where the number of updings was counted for 20 consecutive roof panels from three sections of seven different pallets of steel sheets. Originally, the plan was to repeat this data collection over two separate days. However, the team found the full extent of variation from the baseline was observed on the first day so they stopped collecting data. The data are given in the file *roof panel updings multivari*.

a. Analyze the data using multivari charts and draw conclusions.
b. When the number of updings was counted they were classified into one of the nine locations as numbered in the schematic that follows. Analyze the multivari data using a concentration diagram based on the given schematic.

Roof panel

7	8	9
4	5	6
1	2	3

Front

a. The multivari data covers the full extent of updings variation. The multivari charts suggest a dominant cause acts in the pallet-to-pallet family of causes. There is no section-to-section effect and no evidence of an interaction between the pallet and section families.
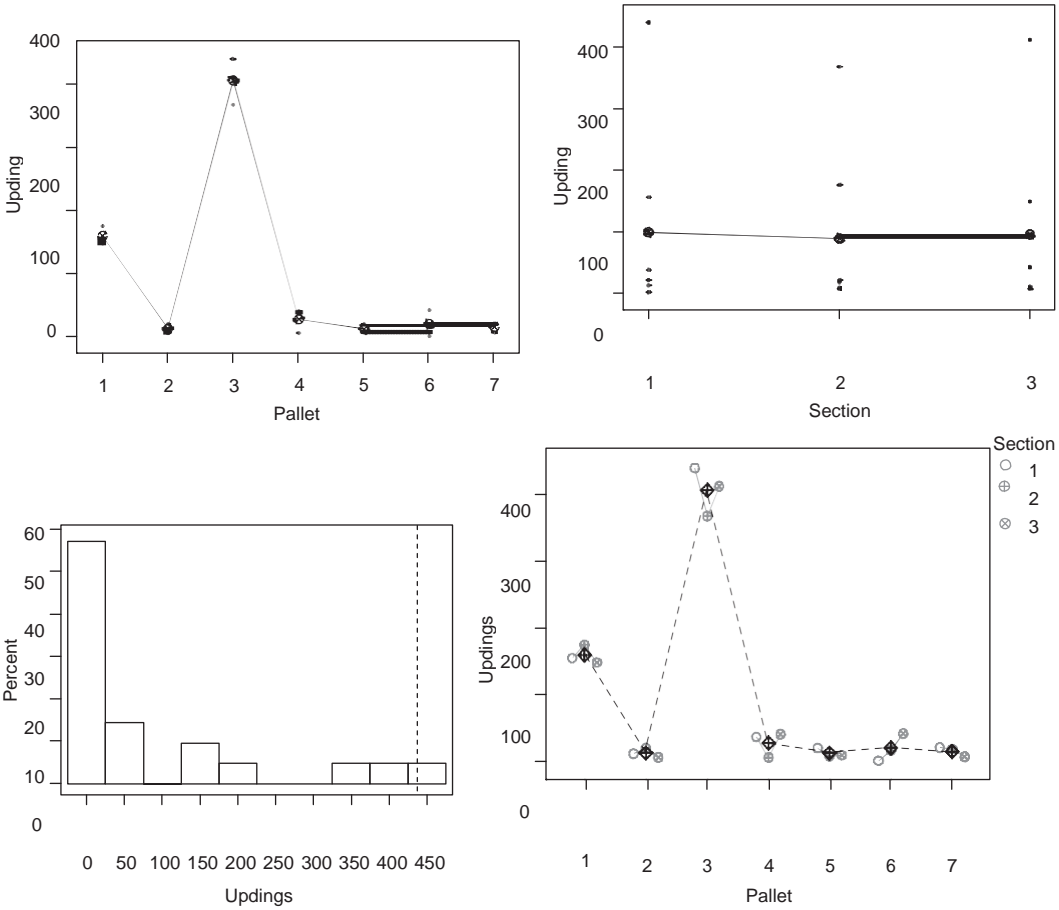
b. Plotting the number of updings by location using the concentration diagram does not show any clear patterns. We can eliminate the location-to-location family.
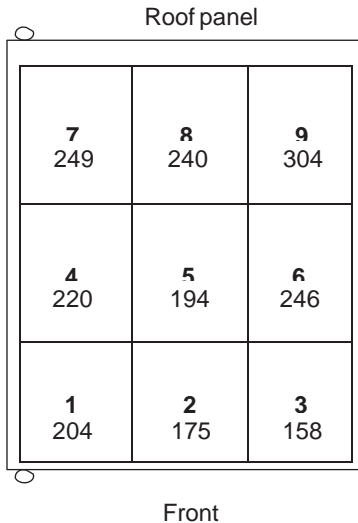
Roof panel

7	8	9
249	240	304
4	5	6
220	194	246
1	2	3
204	175	158

Front

11.8 The baseline investigation for the sand core example discussed in Chapter 1 involved taking five samples over a single day of five consecutive shots of four

cavities each. The data are given in *sand core strength baseline*. **What conclusions can you draw?**

Multivari charts using one input at a time follow. Since we are analyzing the baseline data, the core strength values by definition cover the full extent of variation. We see no strong cavity, time, or shot effects. However, there is much greater variation at time 4 than the other times. The core-to-core family is a large source of variation only at time 4.
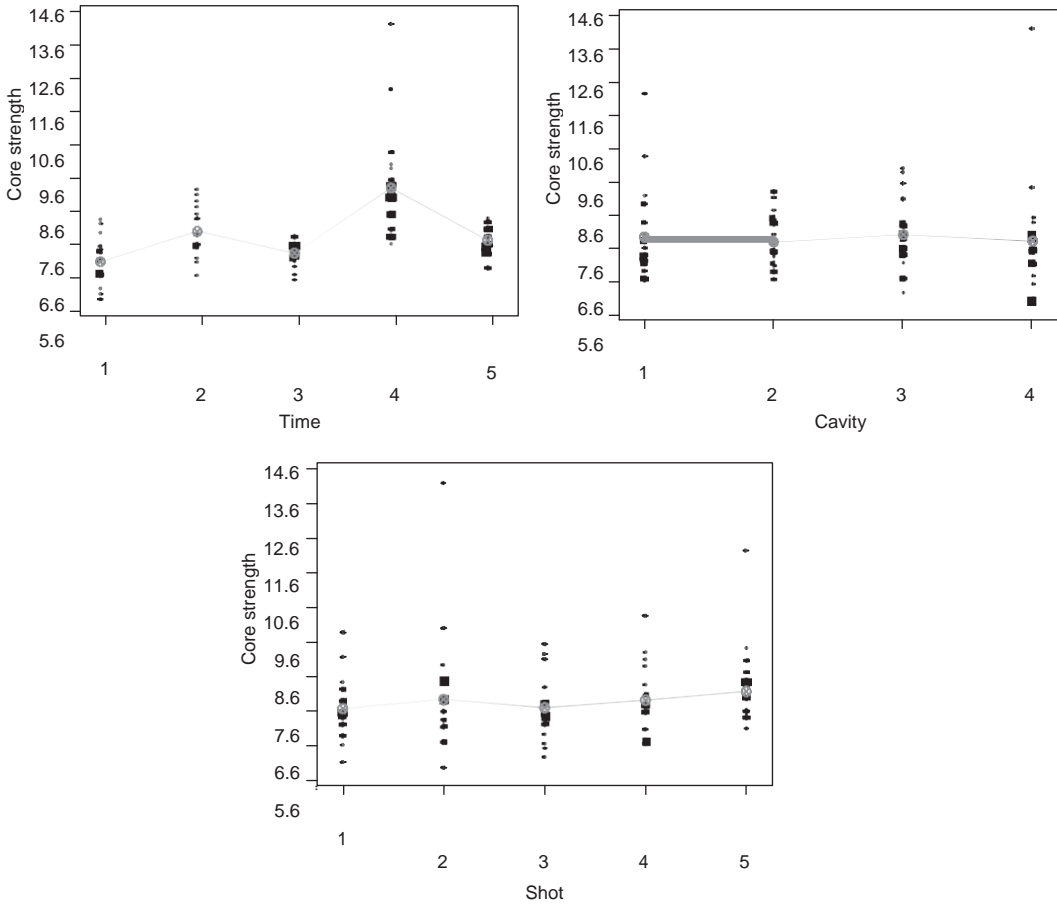
Exploring the data further, we make multivari plots by time together with shot and cavity in turn. As there is no clear pattern, the shot and cavity families can be eliminated.
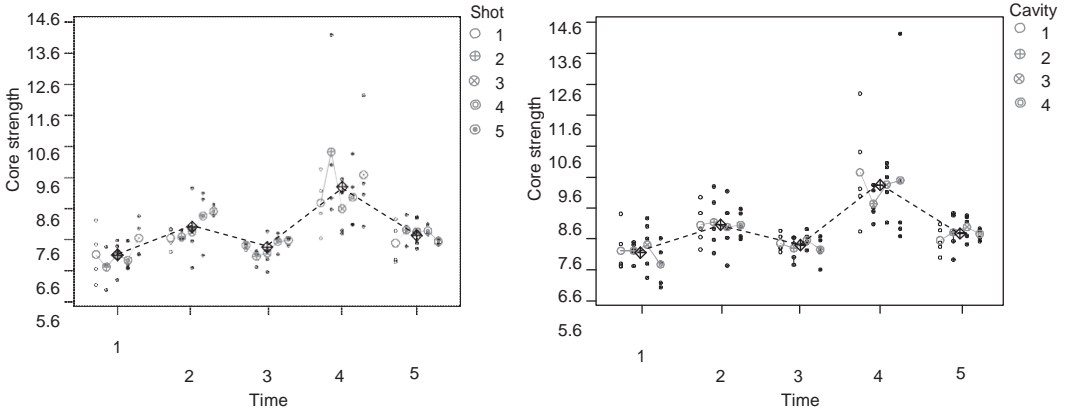
We need to look for a dominant cause consistent with the extra core-to-core variation at time 4.

CHAPTER 12

12.1 Vehicle plant and warranty complaints for loose spark plug wires at the spark plug end prompted an improvement project. As a result of several investigations, the family of causes related to push forces on the wires was the home of a dominant cause. A further investigation then compared eight loose and eight good connections. For each of the 16 connections, the team measured the terminal position of wires and terminal runout of the spark plug in millimeters. The data are given in the file *spark plug connection comparison*. What do the data tell us about the dominant cause?

Plotting the terminal runout and wire position by connection quality, as given, suggests wire position is a dominant cause. Note that there is one value in each group where wire position does not explain the connection quality. It is possible there is also a secondary cause. Before proceeding the team should verify that wire position is a dominant cause.
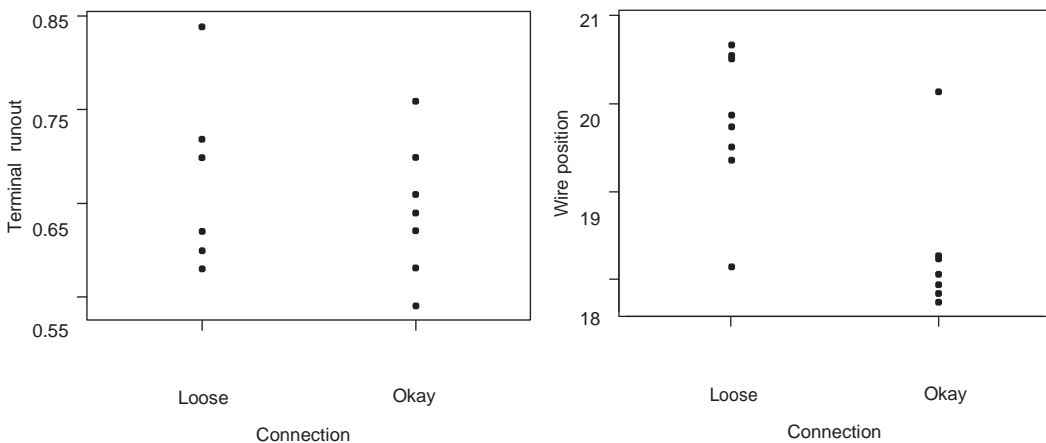
12.2 **A sunroof installation process suffered from a 90% rework rate due to a lack of flushness. Flushness is defined as the difference in height between the sunroof seal and the metal roof. It is measured using digital calipers at six points (three at the front and three at the back). A baseline investigation showed that flushness problems were most pronounced at the two front corners with full extent of variation –3.5 to 4 mm and standard deviation 1.25 millimeters. A goal of reducing the front corner flushness variation to 0.5 and a range of –2 to 1 millimeters was established. Based on engineering knowledge, the team felt that only two types of characteristics could lead to flushness variation, namely roof crown height and attachment pad height. When the roof is adapted to allow installation of a sunroof, six installation pads are added. Based on this knowledge, the team selected six vehicles with large positive flushness and six vehicles with large negative flushness on both front corners. The sunroof modules were removed and the six attachment pad heights and roof crown height were measured at the front and back. The data are given in the file *sunroof flushness input-output*. What conclusions can you draw?**

To start the analysis, we fit a multiple regression for both left front flushness and right front flushness with all the eight inputs. The residual variation for the two regression models was 0.35 (left) and 0.32 (right). Since both the residual variations are less than the target of reducing flushness variation to 0.5, the results appear promising. Plotting the data shows that a dominant cause of the right front flushness variation is the right front pad height. Similarly, a dominant cause of the left front flushness variation is the left front pad height.
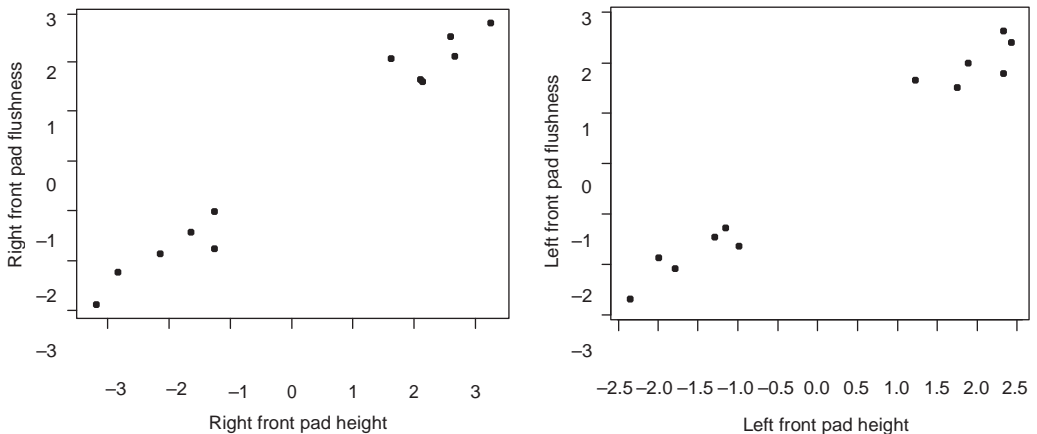
The regression results given for left front flushness show that by eliminating the dominant cause we could reduce the variation in flushness variation to 0.30 for the left side and 0.35 for the right side.

The regression equation is
left front flushness =   −0.117 + 1.05 left front pad height

Predictor	Coef	SE Coef	T	P
Constant	−0.11745	0.08689	−1.35	0.206
left front	1.05208	0.04677	22.49	0.000

S = 0.2996   R-Sq = 98.1%   R-Sq(adj) = 97.9%

12.3 **An example related to sand defects in manifolds was discussed in Chapter 12. Before the problem mentioned in Chapter 12, the team carefully monitored the process for a shift. Each of 970 manifolds was classified as either scrap (due to sand issues) or not scrap. In addition many inputs relating to each manifold, including some discrete inputs such as mold number and continuous inputs such as pour time were recorded. In some cases the linkages were difficult to establish, and the team did the best they could. The data are given in the file *manifold sand scrap comparison*. What conclusions can you draw?**

There are no clear differences in the distributions of any of the input values when we stratify by scrap or not scrap. The box plots that follow are typical.
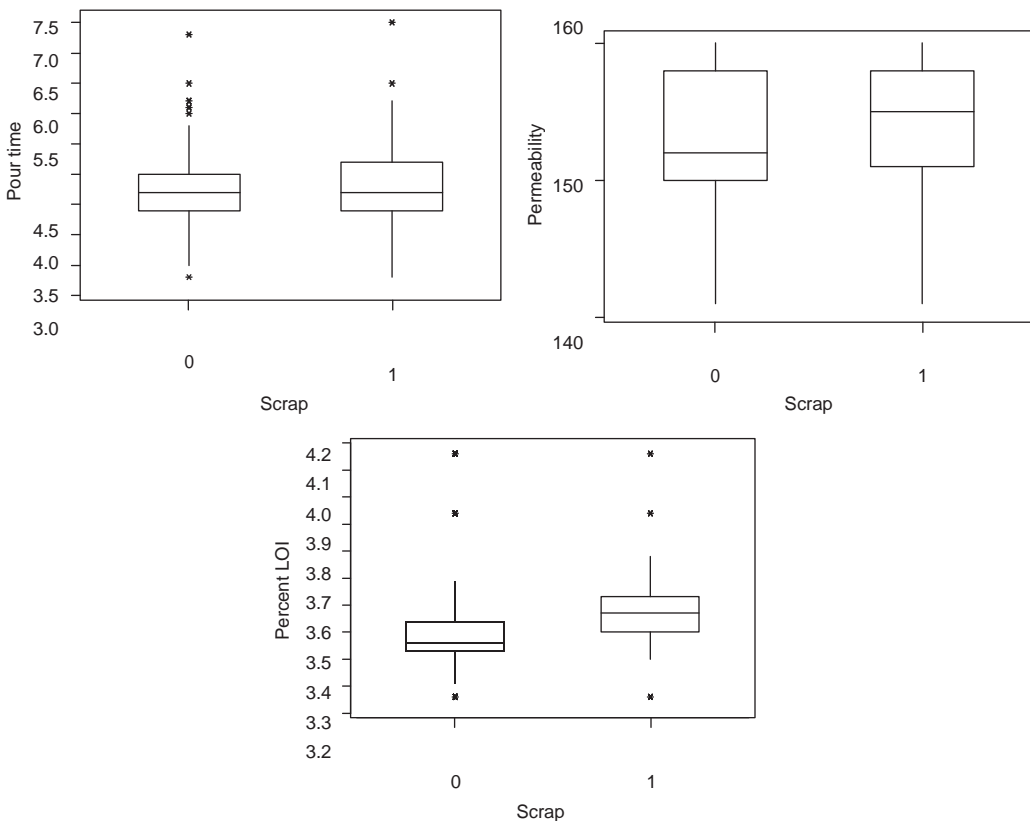
The team concluded that linking input values to individual manifolds was too difficult. Many of the inputs are only measured a few times each shift and there is great uncertainty about the time linkages between the condition of each manifold and the

measured input values. The team next conducted the investigation described in Chapter 12.

CHAPTER 13

13.1 In a verification experiment there were two suspects at two levels. The low and high levels for each suspect were chosen based on the extremes from historical variation. The results of the first three runs of the experiment are shown in the following table.

Input *A*	Input *B*	Order	Output
Low	Low	2	13
Low	High	3	16
High	Low	1	17
High	High	4	?

Given that the full extent of output variation is 12 to 30, what conclusions can you draw about the dominant cause?

Suspects *A* and *B* alone are not dominant causes. However, without making the last run, we cannot tell if there is a dominant cause that involves both inputs. If the run with both *A* and *B* at the high level gives an output value near 30, the dominant cause involves both inputs. If the last run gives an output value near 17, then neither suspect is a dominant cause.

13.2 In the engine block porosity example discussed in the text and exercises of Chapter 10, a dominant cause of porosity acted immediately following scheduled breaks in production. Based on this evidence, the team identified two suspects: iron pouring temperature and the addition of ladle wash. During work stoppages, iron that remained in the six pouring ladles cooled off because there was no external heat source. At the start of the break, ladle wash was added to the ladles to protect the refractory (surface). The team could not easily manipulate the pouring temperature, but they could change the amount of ladle wash. They conducted a verification experiment in which they added the normal amount of wash to ladles 1, 3, and 5 and half the normal amount to the other three ladles over two lunch breaks. At each break, they measured the porosity of the first 30 blocks poured (five from each ladle). The data are given in the file *engine block porosity verification*.

    a. What have we learned about the identity of the dominant cause of porosity?
    b. Explain how the effects of ladle number and the presence or absence of ladle wash are confounded. Does this matter?

c. **Is it a problem that in this verification experiment we have not observed the behavior of the process before lunch breaks?**

a. We use a box plot to stratify the porosity by the amount of ladle wash. We also plot porosity against block number for the two levels of ladle wash.
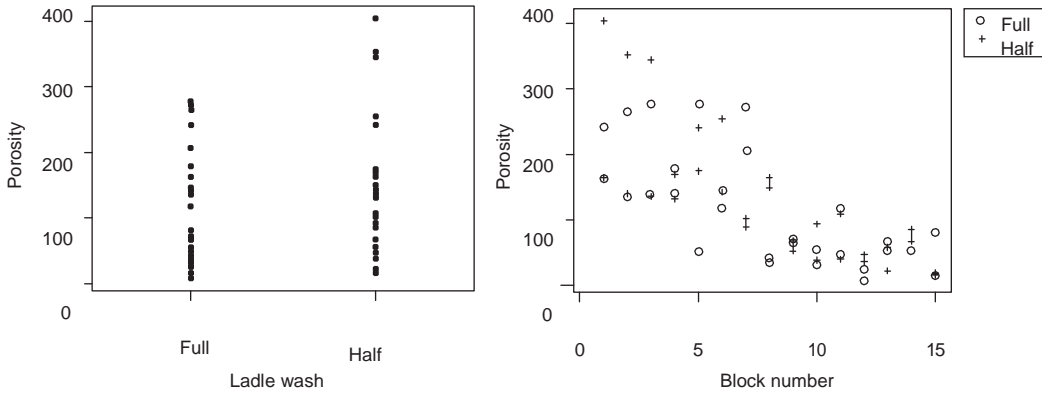
Ladle wash is not a dominant cause of porosity. Because the porosity decreased as the block number increased, the team concluded that pouring temperature was a dominant cause. They accepted the risk that some other unknown cause, confounded with pouring temperature, was actually responsible for the change in porosity.

b. In the experiment, the effects of ladle number and the presence or absence of ladle wash are confounded, since ladles 1, 3, and 5 always had full ladle wash and ladles 2, 4, and 6 always had half ladle wash. As shown in the following plot, the porosity scores were consistent across the odd- and even-numbered ladles. Based on engineering knowledge, the team believed there was no other cause that matched the change in the amount of ladle wash. Thus, in this case the confounding does not limit the conclusions in any meaningful way.
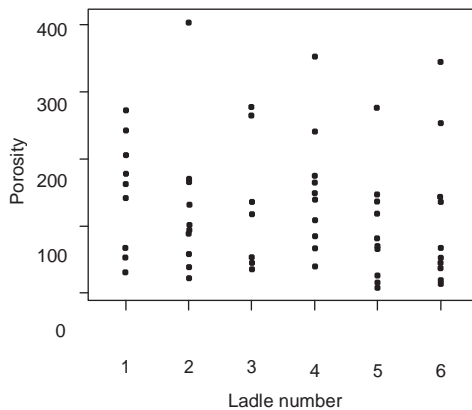
c. Observing the process before breaks was not necessary because the team was trying to determine if wash or temperature was the dominant cause of the porosity. These suspects changed at, or after, the break.

13.3 **The manufacture of a tube assembly required a protective nylon sleeve to be positioned and bonded to a tube. The bond strength of this tube assembly was occasionally tested using a destructive test where the sleeve was subject to increased tensile shear load until failure. In the current process, the average pull-off force was around 15 pounds, but roughly 8% of assemblies tested had a pull-off force less than the desired minimum of five pounds. The team decided to try to solve the problem by reducing the variation in pull-off force rather than by increasing the average pull-off force. A number of investigations were conducted to find the dominant cause. A multivari investigation suggested that the dominant family of causes was tube-to-tube. At this point, the team decided to conduct an experiment to search for a dominant cause using the limited process information they had gathered. They planned a factorial experiment with three suspects—clearance between the sleeve and tube, amount of adhesive, and cure time—all consistent with the tube-to-tube family clue. The team chose the low and high levels of each suspect to roughly match their range in regular production. The levels of clearance were achieved by sorting sleeves and tubes. There were two replicates of each treatment, and the run order was randomized. The data are given in the file *nylon bond strength verification* and summarized in the following table:**

Treatment	Order	Clearance	Adhesive	Cure time	Bond strength
1	6, 7	Low	Low	Low	26, 28
2	14, 8	High	Low	Low	10, 10
3	2, 3	Low	High	Low	25, 26
4	12, 5	High	High	Low	7, 9
5	16, 4	Low	Low	High	24, 27
6	13, 10	High	Low	High	12, 13
7	15, 11	Low	High	High	23, 21
8	1, 9	High	High	High	7, 7

What do the results tell us about the dominant cause?

Plotting bond strength by treatment and creating a Pareto plot of the effects based on a full model gives the following plots.
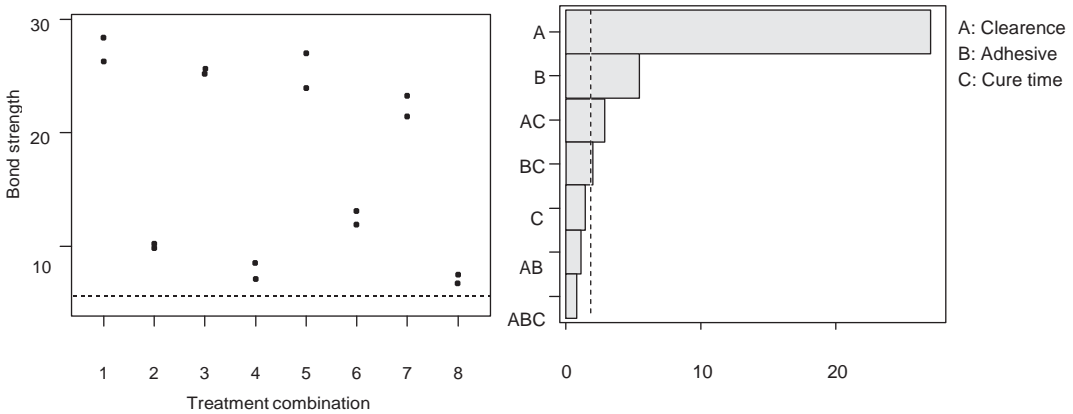
We see that input *A* (clearance) has by far the largest effect. A plot of bond strength by clearance (input *A*), given as follows, clearly shows that clearance is a dominant cause of bond strength variation, and that low clearance gives higher average bond strength.

13.4 Steering knuckles are produced in a gray iron casting process. Around 2% of castings were scrapped because the percent nodularity was too small. In this example the team did not clearly establish a problem baseline. The team thought the cause must be related to the inoculation of the molten iron using a silicon-based alloy. The inoculant was added as the iron was poured to increase nodularity (and thus casting strength). Based on observing the process, the team noticed that the amount of inoculant added by the automated delivery system seemed to vary. The desired amount of inoculant was obtained by slowly shaking the inoculant onto a plate. The plate was designed to tip automatically when the required weight of inoculant was present. The team saw that the location on the plate where the inoculant fell varied, and they thought that this might influence when the plate tipped and thus how much inoculant was delivered. The team decided to verify inoculant amount as the dominant cause of nodularity variation. In the verification experiment, they produced a total of 20 castings at each of two levels of inoculant amount, 12.3 and 13.5 grams. For the

experiment the inoculant was carefully weighed and added by hand. The exper-
iment consisted of eight runs of five castings each. The order of the eight runs
(four at each level) was randomized. For each of the 40 castings the percent
nodularity was determined. The data are given in the file *steering knuckle
strength verification* and are summarized in the table that follows:

Run	Inoculant amount	Order	Percent nodularity
1	12.3	2	81.8, 79.4, 80.3, 80.6, 79.3
2	12.3	3	79.8, 77.0, 77.8, 79.3, 78.7
3	12.3	8	80.9, 82.0, 80.6, 80.6, 81.1
4	12.3	4	81.0, 79.4, 77.0, 80.6, 80.2
5	13.5	5	82.5, 86.1, 82.3, 83.5, 85.2
6	13.5	7	82.1, 84.6, 83.9, 85.0, 85.6
7	13.5	6	85.0, 87.8, 83.1, 84.0, 84.4
8	13.5	1	85.0, 84.3, 86.3, 83.8, 82.9

a. **What considerations should the team have used in choosing the two levels
   for inoculant?**
b. **Why was randomizing the order of the runs important?**
c. **Has the team verified the dominant cause of nodularity variation?**

a. The team chose the levels to cover the range of inoculant values seen in regular
   production. Over a single shift, the team determined the range 12.3 and 13.5
   grams by weighing the amount of inoculant obtained using the automatic deliv-
   ery system on a separate scale.
b. The team chose an experiment with four replicates of each level of the suspect and
   five repeats for each run. Here, since the team had no information about the time
   family that contains the dominant cause, the randomization is extremely impor-
   tant. There is a danger that some unknown input may change in the same way as
   the suspect. Four replicates may not be sufficient. The repeats do not help.
c. Plotting the percent nodularity by amount of inoculant gives:
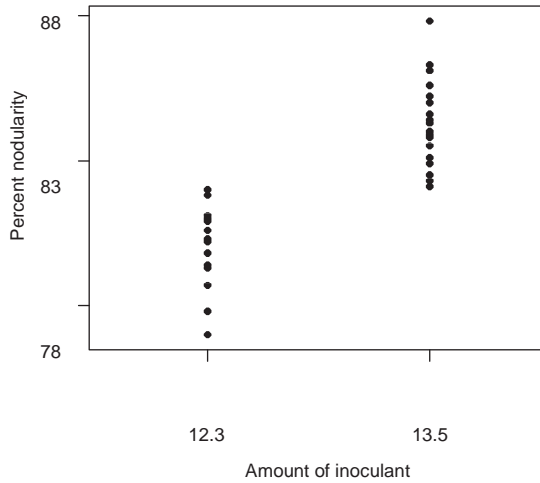
Because there was no baseline, we cannot tell if we have seen the full extent of nodularity variation in the experiment. However, since all of the castings with percent nodularity less than 80 would be scrapped, the team concluded that the amount of inoculant was the dominant cause. An obvious fix was to replace the plate with a cone. With the plate, not all the inoculant collected at the center. Using a cone, all the inoculant collects at the bottom, and the cone does not tip too early.

CHAPTER 14

14.1 In the camshaft lobe runout example, the team searched for a dominant cause of variation. As discussed in Chapter 10, they conducted a variation transmission investigation where runout was measured directly before heat treatment and after the final step of the process, on the same 32 parts selected over the course of one day. In the investigation the grinder (one of eight) and heat treatment spindles (one of four) used were also recorded. The data are given in the file *camshaft lobe runout variation transmission*. They found that a dominant cause of variation was the BC runout just after heat treatment and, more specifically, as shown in the plot that follows, that heat treatment spindle was a dominant cause.

In this example, the team decided not to reformulate the problem but to look for a more specific cause.

a. **Discuss the advantages and disadvantages of the decision not to reformulate.**
b. **Suppose the team had reformulated the problem based on heat treatment spindle and that the original goal was to reduce the final runout standard deviation to less than 4.5. Using the results from a one-way ANOVA model based on heat treatment spindles, derive a goal for the new problem based on differences among spindle averages.**

a. The advantages of not reformulating the problem are mainly that a new baseline for the runout after heat treatment (that is, the dominant cause, or the output in the new problem) does not need to be established, and we need not determine a goal for the new problem. The main disadvantage of not reformulating is that we still need to measure the final runout in future investigations. Had we reformulated, we would use the runout after heat treatment. In this case, it was not cheaper or easier to measure runout after heat treatment.

b. We use a one-way ANOVA model to assess the possible reduction in runout if we could align all the heat treatment spindles perfectly. The appropriate ANOVA results are:

Analysis of Variance for final BC

Source	DF	SS	MS	F	P
heat treatment	3	894.7	298.2	17.53	0.000
Error	28	476.3	17.0		
Total	31	1371.0			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	----------+---------+---------+------
1	8	15.538	2.819	(----*----)
2	8	16.688	5.491	(----*----)
3	8	28.900	2.751	(----*----)
4	8	18.675	4.729	(----*----)

----------+---------+---------+------

Pooled StDev =   4.124   18.0   24.0   30.0

By perfectly aligning the heat treatment averages, we predict the standard deviation could be reduced to 4.1. Hence, to meet the goal we need to ensure that the heat treatment averages are close to equal.

14.2 **In Chapter 11, the team found that the piston diameter directly after operation 270 was a dominant cause of final V6 piston diameter variation. The relationship is illustrated in the scatter plot that follows. The data are given in the file *V6 piston diameter variation transmission*.**
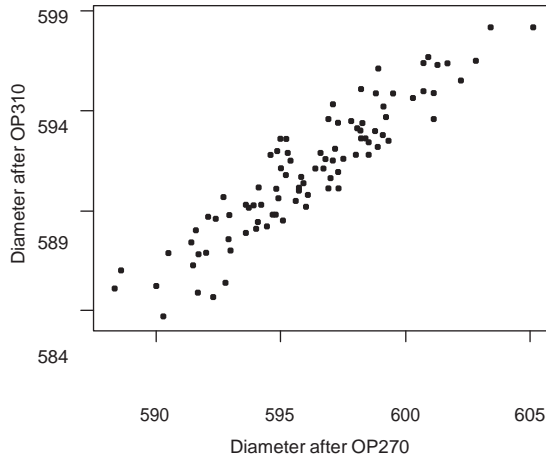
The team decided to look further for a more specific dominant cause. Suppose, however, they had wanted to reformulate the problem in terms of the operation 270 diameter. Determine an appropriate goal for the reformulated problem. Recall that the goal for the original problem was to reduce the final diameter standard deviation to less than 2.0.

Fitting a linear model to the variation transmission investigation results gives:

The regression equation is
diameter after OP310 = 64.3 + 0.884 diameter after OP270

Predictor	Coef	SE Coef	T	P
Constant	64.27	22.29	2.88	0.005
diameter	0.88358	0.03739	23.63	0.000

S = 1.224   R-Sq = 85.6%   R-Sq(adj) = 85.4%

Using the regression results, we estimate that

$$stdev(\text{final diameter}) \approx \sqrt{0.884^2\, stdev(\text{OP270 diameter})^2 + 1.224^2}$$

Thus, to meet the goal of reducing the final diameter variation to less than 2.0, we solve to get stdev(OP270 diameter) < 1.58. Thus, a reasonable goal for the reformulated problem is to reduce the diameter variation at operation 270 to less than 1.6. Note that in the variation transmission investigation, the diameter variation at OP270 was 3.4, so the goal requires more than a 50% reduction in diameter variation.

CHAPTER 15

15.1 Based on customer complaints about assembly difficulty, a team investigated fascia dimension variation. A baseline investigation found that some fascias were too large. The team felt that reducing the average size of the fascias could

solve the problem (that is, they adopted the Move the Center approach). They planned a (full) factorial experiment with two candidates, cycle time and cure time, each at two levels to look for an adjuster. They chose the levels for each candidate based on engineering judgment. The results of the experiment are given in the file *fascia dimension move center* and in the following table. For each treatment, the team conducted four runs producing 10 fascias for each run. The order of the 16 runs was randomized over a day. In the data, we give only the average fascia dimension from each run and not the individual values.

Treatment	Run order	Cycle time (minutes)	Cure time (hours)	Average fascia size (from nominal)
1	8, 10, 1, 14	85	5	4.50, 5.23, 5.75, 6.51
2	6, 15, 12, 2	113	5	7.12, 8.25, 9.06, 9.28
3	11, 3 , 9, 16	85	19	3.65, 3.75, 4.27, 5.34
4	13, 5, 4, 7	113	19	4.24, 6.31, 7.15, 8.22

a. Can cycle time or cure time be used as an adjuster?
b. Suppose the goal was to reduce the average fascia size to 3.0. What do you recommend?
c. What is the advantage of looking at the dimensions for all the fascias within a run, rather than the averages?

a. We first plot dimension by treatment. We see that there are differences between the average dimensions for different treatments. In particular, treatment 3, with low cycle time and high cure time, leads to smaller average dimensions.
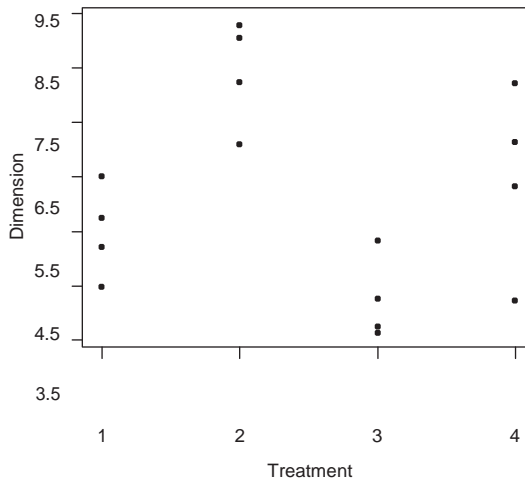
Next, we fit a full model and look at the Pareto plot of the effects.
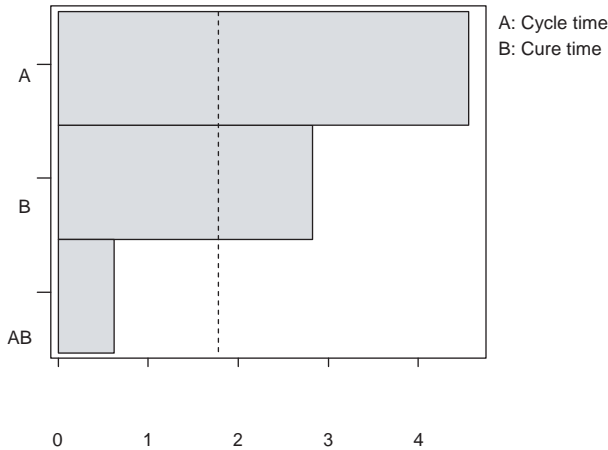
Both cycle time and cure time have large effects; hence, both are potential adjusters. The interaction between these two candidates is small. From the following main effect plots, we see that lowering cycle time and increasing cure time will reduce the average dimension.
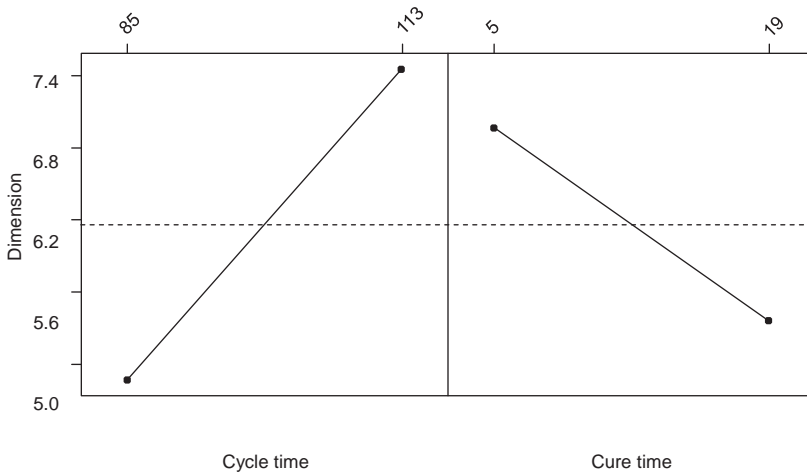
b. None of the treatments used in the experiment gave an average dimension as small as 3. Shorter cycle time and longer cure result in smaller dimensions. The team should consider another investigation to calibrate the effects of changing both adjusters.

c. With the individual dimensions for all 160 fascias, we could look at the variation within each run. This would allow us to determine whether changing the candidate settings changed the short-term fascia-to-fascia variation.

15.2 An experiment was carried out to investigate four candidates to search for an adjuster of the formability safety margin of galvanized sheet metal trunk lids. The purpose was to increase the average safety margin from the baseline value

of 10.7. In the experiment, each candidate was tested at two levels, selected to be near the edge of what was physically possible—see the table that follows. Note that none of the treatments corresponded to the existing process settings.

Treatment	Run order	Tonnage	Lubrication	Blank size		Prebending	Safety margin
1	6	310	Unlubricated	949	1494_m m	No	8
2	16	375	Unlubricated	949	1494_mm	No	11
3	3	310	Lubricated	949	1494_mm	No	12
4	11	375	Lubricated	949	1494_mm	No	0
5	7	310	Unlubricated	965	1500_mm	No	13
6	15	375	Unlubricated	965	1500_mm	No	6
7	4	310	Lubricated	965	1500_mm	No	11
8	13	375	Lubricated	965	1500_mm	No	1
9	1	310	Unlubricated	949	1494_mm	Yes	18
10	12	375	Unlubricated	949	1494_mm	Yes	17
11	5	310	Lubricated	949	1494_mm	Yes	18
12	14	375	Lubricated	949	1494_mm	Yes	10
13	2	310	Unlubricated	965	1500_mm	Yes	8
14	10	375	Unlubricated	965	1500_mm	Yes	12
15	8	310	Lubricated	965	1500_mm	Yes	16
16	9	375	Lubricated	965	1500_mm	Yes	8

Press tonnage was very difficult to change so all eight runs with low press tonnage were carried out first. Within each group of eight runs, the order was randomized. The data are given in the file *sheet metal move center*.

a. Analyze the experimental data to see if any of the candidates is an adjuster.
b. Does the restriction on randomization required for this experiment make any difference to the conclusions we can draw?

a. We start the analysis by looking at a plot of the safety margin by treatment. In this case, since there is only a single output for each treatment, it is difficult to draw conclusions from the plot. We do see that there are treatments with very different values for the safety margin.
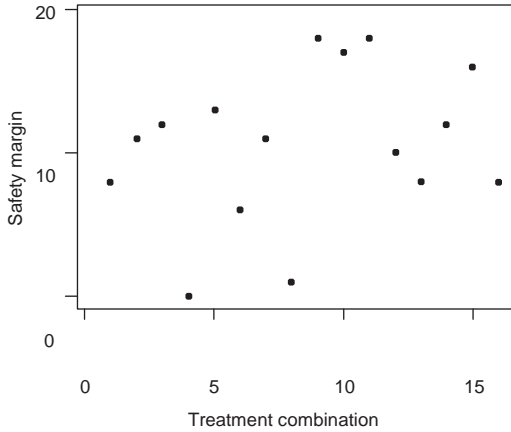
Next, we fit a full model and examine a Pareto plot of the effects.
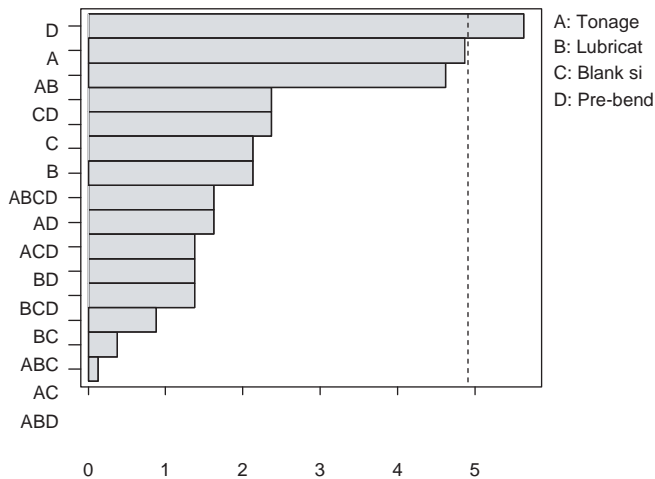
D has a large main effect and *AB* a large interaction effect. We follow up with plots of safety margin by prebending (input *D*) and an interaction plot of tonnage by lubrication.
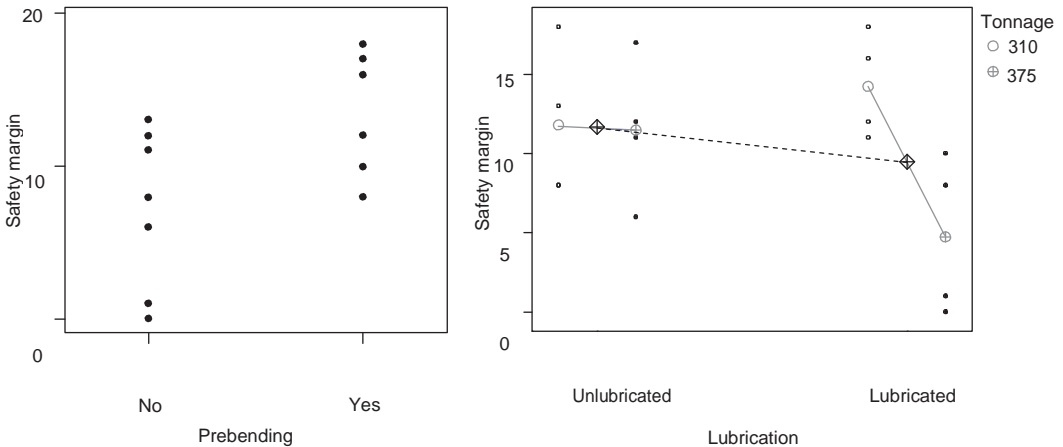
We see that prebending leads to higher safety margin on average and that the safety margin is much more sensitive to the presence or absence of lubrication at the high level of tonnage than at the low level. The average safety margin for the

two runs with prebending, low press tonnage, and lubrication was $(18 + 16)/2 = 17$, considerably higher than the baseline average 10.7. This finding needs to be confirmed. See the discussion in part b.

b. Without the randomization it is more likely that the effect attributed to press tonnage is due to some other (unknown) input. This could occur if the level of the unknown cause happened to change at (or near) the moment in time when press tonnage was changed. When interpreting the experimental results we cannot tell if this occurred, or not.

15.3 In the sand core strength example introduced in Chapter 1, too many cores were breaking during handling. A suggested solution was to increase the core strength (and thereby reduce core breakage) by increasing the resin concentration. It was known that increasing the resin would result in a stronger core. However, the precise nature of the relationship—that is, how much the core strength increases for a given change in resin concentration—was not known. An experimental investigation was planned to quantify the relationship. Three levels of resin concentration (1.3, 1.6, 1.9% by weight) were chosen based on engineering judgment. In the experiment, 40 cores for each level of resin were produced; 15 were measured for strength (using a destructive test) and the remaining 25 were processed to look for casting problems. The experiment consisted of three runs with 15 repeats. The order of the runs was not randomized. The data are given in the file *sand core strength move center*.

a. What can you conclude about the relationship between resin concentration and core strength?

b. The team used only three runs with 15 repeats for each run. Discuss the advantages and disadvantages of this plan compared with using five replicates for each treatment with three repeats each.

a. We plot strength by resin level as follows:
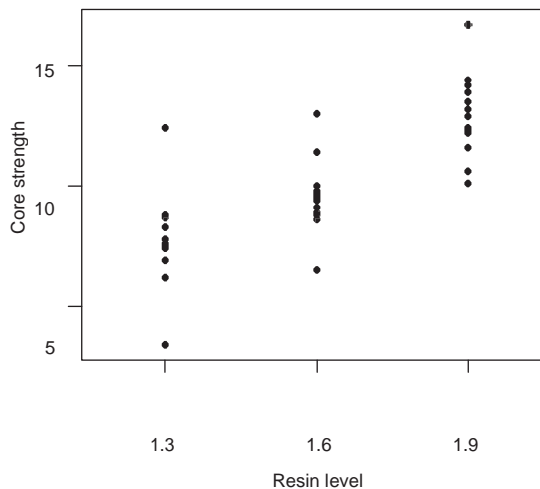
Because the increase in average strength is roughly linear as a function of the resin concentration, to quantify the relationship, we fit a regression model:

The regression equation is
core strength =   –3.76 + 8.61 resin level

Predictor	Coef	SE Coef	T	P
Constant	–3.762	1.609	–2.34	0.024
resin	8.6111	0.9943	8.66	0.000

S = 1.634   R-Sq = 63.6%   R-Sq(adj) = 62.7%

> We expect, on average, a 0.86 unit increase in core strength for each 0.1 increase in the percent resin. The team had quantified the effect of resin concentration. At the 2% level of resin, one of the 25 cores led to a defective casting, so the team knew they should not raise the resin concentration to this level. As a result, although the team had now quantified how to move the process center, the approach was abandoned because of the fear of increased scrap due to core-related defects.
>
> b. Using five replicates for each treatment would allow randomization of the order. This would help protect against some unknown input changing in the same way as resin in the experiment. A disadvantage is that it may be more difficult and expensive to change the resin level more often.

CHAPTER 16

16.1 In a sonic welding operation, problems arose due to poor weld strength, measured as pull-off force. The goal was to reduce the variation and increase the average pull-off force. The second goal is not addressed here. From the baseline, the full extent of variation for pull-off force was 0.9 to 3.0. The team discovered that the dominant cause acted in the time-to-time family. While they could not be more specific, the team felt that the dominant cause was related to material hardness, which was outside their control. They decided to try to desensitize the process to variation in the dominant cause.

The team planned a fractional factorial experiment with four candidates at two levels each in eight treatments. Using the results of regular process monitoring, they identified three time periods when weld strength was low, medium, and high relative to the baseline. In each period, they randomized the order and then produced a part with each of the eight treatments. The pull-off force data and plan are given in the file *sonic weld desensitization* and the table that follows. The three values in the columns Order and Pull-off force correspond to the three different time periods. The original settings of the candidates correspond to treatment 2.

Treatment	Order	A	B	C	D	Pull-off force
1	7, 2, 6	−1	−1	−1	−1	1.8, 2.1, 2.3
2	4, 6, 1	−1	−1	1	1	0.9, 2.0, 2.7
3	6, 3, 4	−1	1	−1	1	2.0, 2.4, 2.1
4	8, 1, 2	−1	1	1	−1	0.6, 1.5, 3.0
5	3, 7, 8	1	−1	−1	1	2.8, 2.9, 3.0
6	2, 4, 3	1	−1	1	−1	2.1, 3.0, 3.7
7	1, 8, 7	1	1	−1	−1	2.9, 3.2, 3.1
8	5, 5, 5	1	1	1	1	2.3, 2.5, 4.0

a. **Explain why the team believed the dominant cause acted over the three runs for each treatment.**
b. **What levels of the candidates do you recommend to reduce the variation in pull-off force?**
c. **Another way to assess the results of this experiment is to summarize the output across each treatment using log standard deviation. Using this performance measure, do your conclusions differ from part b?**

a. The team knew that the dominant cause acted time to time. Since they selected periods with low, medium, and high weld strength under the current conditions, they were confident that the dominant cause acted over the three periods. In the experiment, this assumption was verified since the range of pull-off force values for treatment 2, the current process settings, covered the full extent of variation.
b. To start the analysis we look at a plot of the pull-off force by treatment.
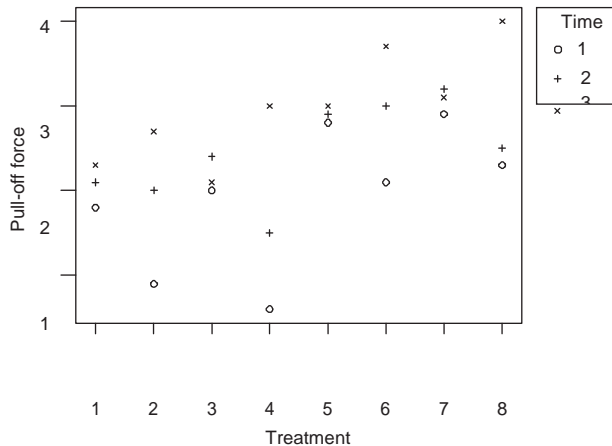
There are several promising treatments with little variation in pull-off force. To desensitize the process we are looking for a special kind of interaction between time (a surrogate for the dominant cause) and the candidates. The interaction plots for time versus the candidates (using the average pull-off force as the response) are
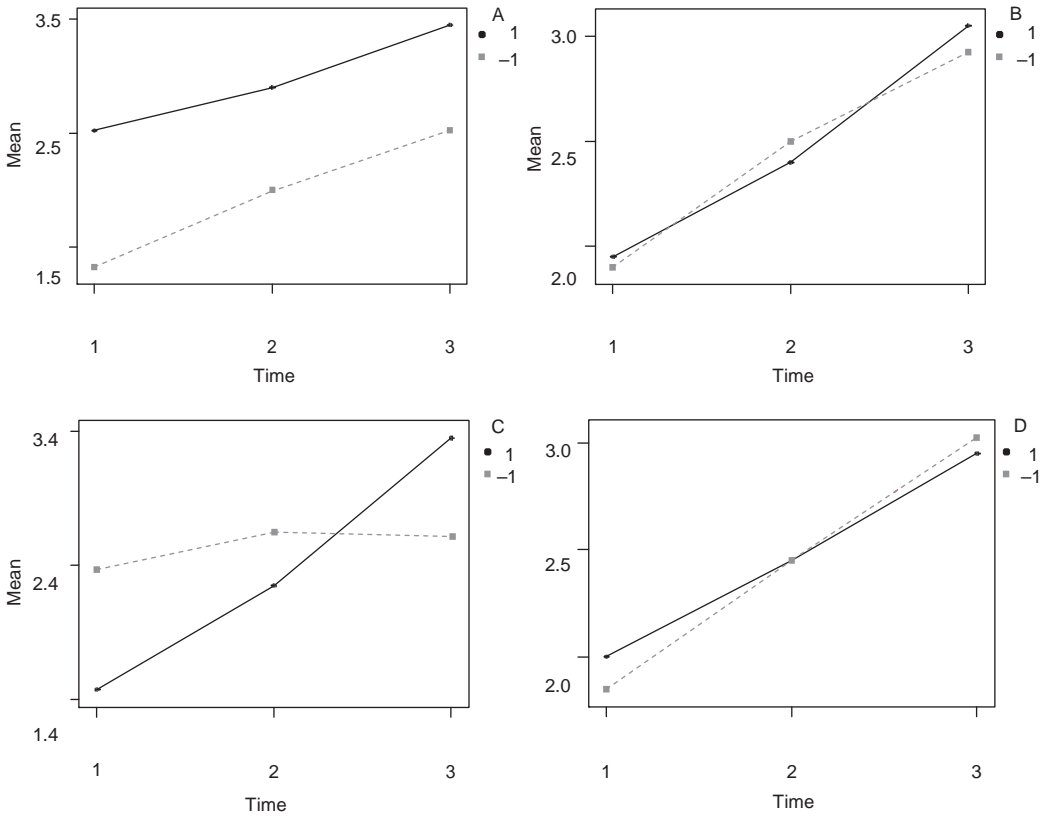
We see that changing the level of candidate *C* flattens the relationship between the pull-off force and time. More specifically, with the low level of candidate *C*, the time-to-time variation in pull-off force is much reduced. Changing the other candidates does not help. We recommend changing to the low level of candidate *C*, and leaving the remaining candidates at their current levels.

c. Fitting a full model in the four candidates and analyzing the results using the performance measure log(s) gives the Pareto effects plot that follows. The plot indicates that candidate *C* has the largest effect.
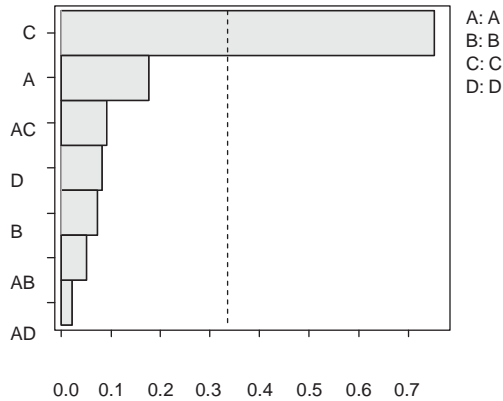
From the main effect plot that follows, we see that the low level of *C* reduces the variation in pull-off force. This is the same conclusion as in part b. In most cases the conclusions obtained with the two analysis methods will be the same. We prefer the analysis that looks directly at the candidate by cause interaction plots rather than the analysis based on log(s).

16.2 In the crossbar dimension example discussed in Chapter 12, the team found that the dominant cause of dimension variation was barrel temperature. Because it was hard to control in regular operation, the team decided to try to make the process less sensitive to barrel temperature variation. In the current process, barrel temperatures ranged over roughly 4°C. The team planned a half fraction factorial experiment with three candidates—target barrel temperature, injection pressure, and material—at two levels each, as shown in the following table. The current injection pressure and target barrel temperature were 1000 and 75, respectively. Note that although the variation in barrel temperature was the dominant cause, the target barrel temperature is a fixed input. Five crossbars were produced and measured in each run. For each treatment, there were two runs, one at the target barrel temperature plus 2°C and the other at the target barrel temperature minus 2°C. The data are given in the file *crossbar dimension desensitization* and in the table as follows.

Treatment	Target barrel temperature	Injection pressure	Material	Dimensions at barrel temperature −2°C	Dimensions at barrel temperature +2°C
1	75	1000	Old	0, −0.1, 0.1, −0.1, −0.2	0.5, 1.1, 0.8, 0.9, 0.7
2	75	1200	New	1.1, 0.6, 1.0, 1.4, 1.1	1.5, 1.8, 1.5, 1.4, 1.3
3	79	1000	New	1.1, 1.0, 1.3, 0.9, 0.8	1.0, 1.1, 0.8, 0.9, 1.0
4	79	1200	Old	1.2, 1.8, 1.8, 1.7, 1.9	2.3, 2.1, 2.4, 2.1, 1.9

Since the average dimension can be easily adjusted, we focus the analysis on finding a way to make the process less sensitive to barrel temperature variation.

a. What levels of the candidates do you recommend?
b. Injection pressure and material were chosen as candidates based on engineering judgment. Looking again at the results presented in Chapter 12, what motivates the choice of target barrel temperature as a possible candidate?

a. We begin the analysis by plotting dimension by treatment with a different plotting symbol for each of the two levels of the dominant cause. Treatment 3 looks most promising. Changing to the higher target barrel temperature, using the new material, and keeping the injection pressure the same is expected to make the process less sensitive to variation in the barrel temperature.
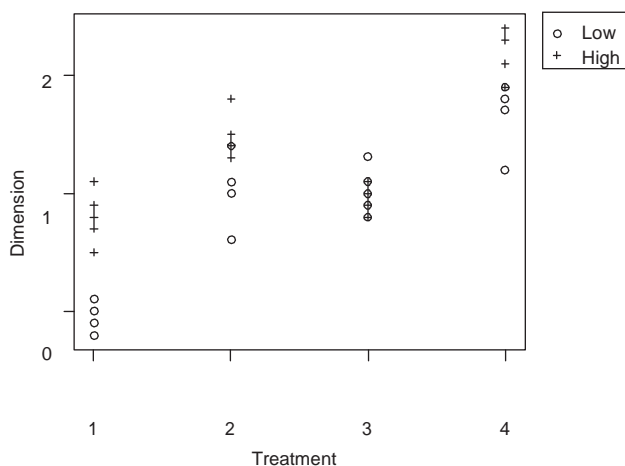
We examine the nature of the interaction between the candidates and the dominant cause with the following plots (where the response is the average dimension):
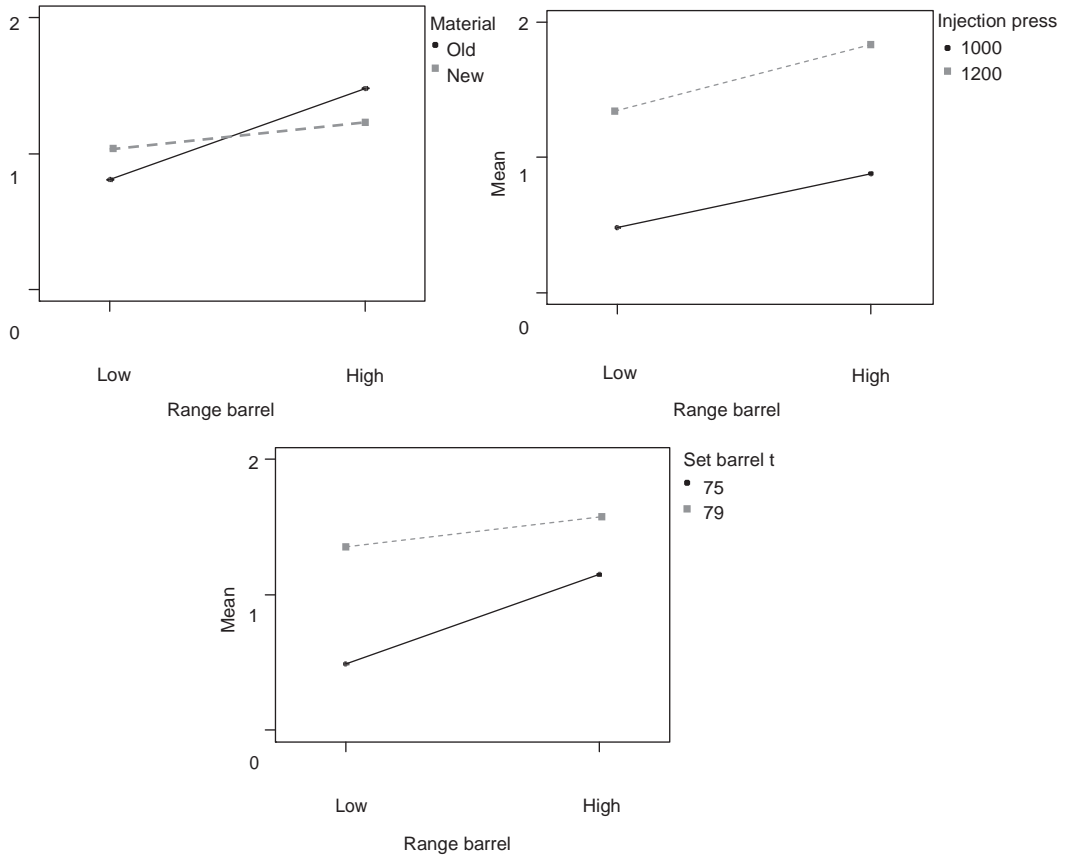
If we increase the target barrel temperature and, to a lesser degree, use the new material, we see that the relationship between crossbar dimension and barrel temperature is flatter. In other words, dimension is less sensitive to barrel temperature variation at the higher target barrel temperature. Changing the injection pressure has no effect. The benefits of changing both the target barrel temperature and the material were later validated with another investigation. Note that the observed variation in treatment 3 has likely overestimated the variation when actually running the process at these settings, since we have forced barrel temperature to near its expected extremes with the new target temperature.

b. The nonlinear relationship between barrel temperature and dimension is shown in the left panel of Figure 12.3. In the plot, it appears there is less dimension variation for high than for low barrel temperatures.

16.3 **In Chapter 16, we describe a desensitization experiment for the refrigerator frost buildup example where each refrigerator is subjected to only two extreme levels of environmental causes. Here we consider a hypothetical experiment in which each refrigerator is exposed to a number of environmental conditions to ensure that any chosen new design works well under any conditions, not just extreme conditions.**

The experimental design for the four candidates—D1, D2, D3, and D4—is the same as in Chapter 16. Here we plan to test each of the eight refrigerators (treatments) under all eight possible combinations of the usage or environmental inputs as given in the following table:

Varying input	Cause combination							
	1	2	3	4	5	6	7	8
Ambient temperature (°C)	26	26	26	26	32	32	32	32
Relative humidity (%)	70	70	90	90	70	70	90	90
Door openings per hour	4	8	4	8	4	8	4	8

The experimental plan had 64 runs. To conduct the experiment, all eight refrigerators were simultaneously placed in a test chamber and exposed to each cause combination in the given order. The cooling plate temperatures are given in the file *refrigerator frost buildup desensitization2* and in the following table:

Treatment	Candidates				Cooling plate temperatures (in cause combination)							
	D1	D2	D3	D4	1	2	3	4	5	6	7	8
1	N	O	O	N	3.6	3.9	4.6	1.0	4.4	0.1	4.4	0.7
2	N	O	N	O	5.1	4.7	4.3	2.9	4.2	4.1	7.1	5.1
3	N	N	O	O	4.6	4.6	4.3	4.9	2.4	5.0	5.5	16.0
4	N	N	N	N	3.8	12.8	6.9	6.9	7.1	6.7	3.0	15.7
5	O	O	O	O	2.9	0.2	−0.2	−0.2	−0.2	−0.2	−0.2	16.0
6	O	O	N	N	0.1	1.9	0.8	1.3	5.9	5.1	0.0	14.7
7	O	N	O	N	0.7	0.8	0.1	0.1	0.4	0.2	0.1	−0.1
8	O	N	N	O	0.2	3.4	0.3	1.0	4.0	0.2	5.2	16.0

In the table, we have coded the new and original settings for the candidates as N and O, respectively. What conclusions can you draw? Remember, the goal is to desensitize cooling plate temperature to changes in the environmental conditions.

We start the analysis by plotting the cooling plate temperatures for all of the eight environmental conditions (cause combinations) by treatment.
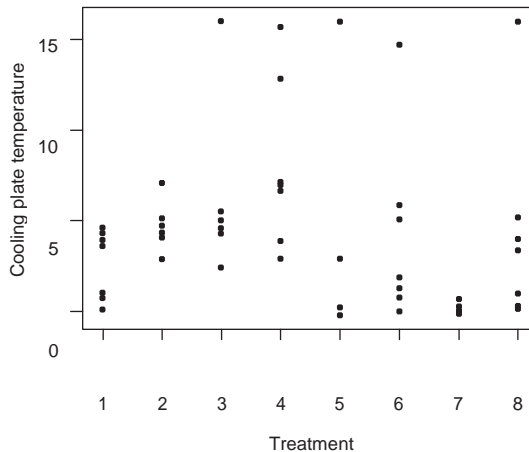
Treatment 7 gives consistently low cooling plate temperatures over all combinations of the environmental inputs. Treatment 7 corresponds to using the new levels for D2 and D4 and leaving D1 and D3 at their original settings. To examine how changing each candidate flattens the relationship between the changing environmental conditions and cooling plate temperature, we examine the appropriate interaction plots. To create an interaction plot of the environmental conditions versus D1 and D3, for example, we create a new column D1D3 with the four levels for the pair D1 and D3.
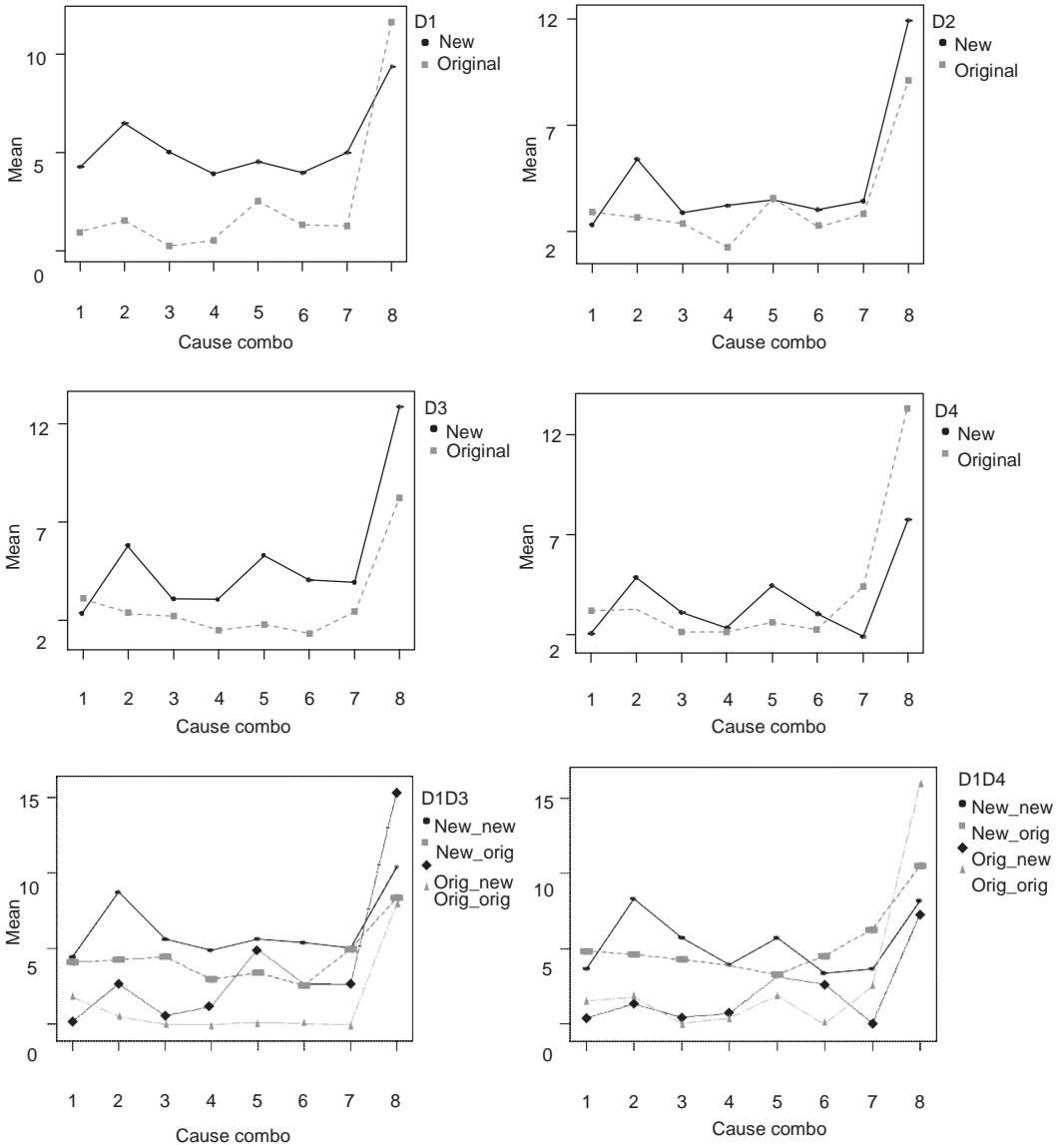
Not all possible interaction plots are given. The other plots give similar conclusions. From the interaction plots, we see that changing a single candidate or a pair of candidates does not reduce the variation in average cooling plate temperature. In particular, all combinations of the candidates have a relatively large cooling plate

temperature for cause combination 8 (high temperature, humidity, and number of door openings per hour). The good properties of treatment 7 are due to a high-order interaction between the candidates that we cannot explore due to the confounding in the design.

16.4 **There were excessive failures in the accelerated life testing of electric motors. Using a group comparison investigation, the team found that unevenness in the commutator shaft surface was a dominant cause of these failures. The team next reformulated the problem to one of reducing the unevenness in the commutator shaft. The surface unevenness is measured on a scale of 1 (smooth) to 10 (rough). With further investigation, the team determined that the dominant cause of the variation in the (final) smoothness was the shaft profile before machining. The team adopted the Desensitization approach. They decided to conduct a fractional factorial experiment with eight treatments using four candidates. For each of the eight treatments there were two runs, one that used a shaft with a premachined smooth profile, and a second that used a rough profile. The experimental plan and data are given in the file *electric motor failure desensitization* and the table that follows. The order of the runs was randomized.**

						Smoothness	
Treatment	**Depth**	**Grind time**	**Rotational speed**	**Feed rate**	**Order**	**Smooth profile**	**Rough profile**
1	Shallow	Short	1800	Slow	4, 5	2	7
2	Deep	Short	1800	Fast	6, 11	3	8
3	Shallow	Long	1800	Fast	1, 14	1	9
4	Deep	Long	1800	Slow	16, 12	2	8
5	Shallow	Short	2400	Fast	13, 9	3	2
6	Deep	Short	2400	Slow	10, 8	1	4
7	Shallow	Long	2400	Slow	3, 7	2	3
8	Deep	Long	2400	Fast	15, 2	3	5

    a. What is the confounding structure of the design? What limitations does this introduce?

    b. What conclusions can you draw?

    c. What would be the advantages and disadvantages of measuring the time to failure using the accelerated life test for each run rather than judging the smoothness of the commutator surface after machining?

    a. The confounding structure including two- and three-input interactions in candidates and the dominant cause (profile) is given by MINITAB as

Alias Structure (up to order 3)

I
Depth + Grind*Rotation*Feed
Grind + Depth*Rotation*Feed
Rotation + Depth*Grind*Feed
Feed + Depth*Grind*Rotation
Profile
Depth*Grind + Rotation*Feed
Depth*Rotation + Grind*Feed
Depth*Feed + Grind*Rotation
Depth*Profile
Grind*Profile
Rotation*Profile
Feed*Profile
Depth*Grind*Profile + Rotation*Feed*Profile
Depth*Rotation*Profile + Grind*Feed*Profile
Depth*Feed*Profile + Grind*Rotation*Profile

The design is resolution IV in the candidates; hence, in the crossed design we can separately estimate the interactions between each candidate and the premachined profile (the dominant cause). For example, in the alias structure we see the interaction depth by profile is in a row by itself.

b. We start the analysis by plotting smoothness by treatment with a different plotting symbol for each profile. Treatments 5 to 8, all with high rotational speed, look promising.
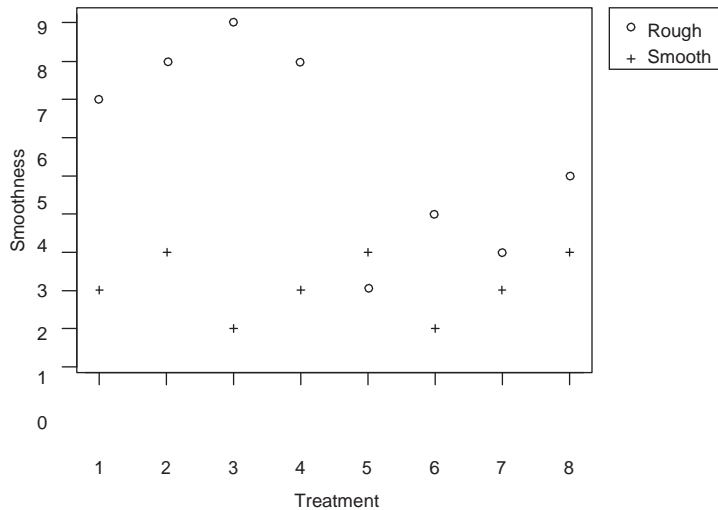
The interaction plots for profile versus each of the four candidates follow. Only rotational speed flattens the relationship between smoothness and the initial shaft profile.
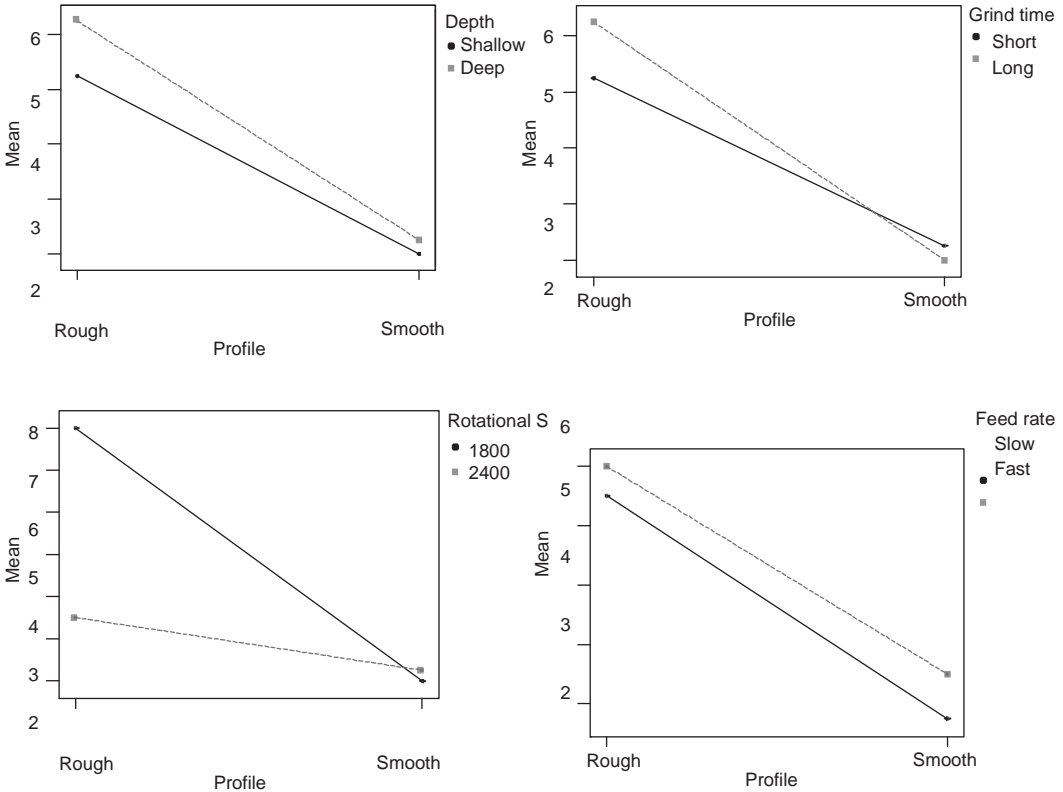
We conclude that setting the rotational speed to 2400 will desensitize smoothness to changes in the initial shaft profile.

c. With the existing experiment there is a risk that the treatment that results in a smoother commutator surface may not translate into better performance in the accelerated test. However, the chance of this occurring is small, since the team verified that the commutator surface unevenness was a dominant cause of electric motor failures. Using the accelerated test for the experiment would have been more expensive.

CHAPTER 17

17.1 **In an investigation, 100 trucks were selected from regular production over two weeks. The frame geometry as given by four summaries (left and right front, left and right rear) and the alignment outputs left and right camber and caster were determined for all 100 trucks. The data are given in the file *truck pull feedforward*. In Chapter 17 an analysis determined that feedforward control based on frame geometry was feasible for left caster. Repeat the analysis for the other outputs: right caster, left camber, and right camber.**

A numerical summary of the three outputs is

Variable	N	Mean	Median	TrMean	StDev	SE Mean
right caster	100	4.0106	3.9645	4.0213	0.8298	0.0830
left camber	100	0.3922	0.4130	0.3970	0.5077	0.0508
right camber	100	0.4811	0.4160	0.4836	0.4146	0.0415

Variable	Minimum	Maximum	Q1	Q3
right caster	1.7820	5.8710	3.5450	4.5588
left camber	-0.8490	1.4890	0.0078	0.7065
right camber	-0.5860	1.3650	0.2232	0.7630

The results from fitting a linear regression model for each of the three outputs with all four possible inputs are:

The regression equation is
right caster = −19.0 + 0.557 left front + 1.07 right front + 0.218 left rear + 0.399 right rear

Predictor	Coef	SE Coef	T	P
Constant	−19.0241	0.6798	−27.99	0.000
left front	0.55657	0.04478	12.43	0.000
right front	1.06849	0.03994	26.75	0.000
left rear	0.21845	0.03079	7.09	0.000
right rear	0.39874	0.04638	8.60	0.000

S = 0.1962   R-Sq = 94.6%   R-Sq(adj) = 94.4%

The regression equation is
left camber = −12.3 + 0.609 left front + 0.360 right front + 0.172 left rear + 0.113 right rear

Predictor	Coef	SE Coef	T	P
Constant	−12.3225	0.4321	−28.52	0.000
left front	0.60906	0.02846	21.40	0.000
right front	0.36025	0.02539	14.19	0.000
left rear	0.17192	0.01957	8.78	0.000
right rear	0.11300	0.02948	3.83	0.000

S = 0.1247   R-Sq = 94.2%   R-Sq(adj) = 94.0%

The regression equation is

right camber = –10.2 + 0.344 left front + 0.556 right front + 0.0136 left rear
   + 0.121 right rear

Predictor	Coef	SE Coef	T	P
Constant	–10.2119	0.3540	–28.84	0.000
left front	0.34438	0.02332	14.77	0.000
right front	0.55571	0.02080	26.71	0.000
left rear	0.01364	0.01604	0.85	0.397
right rear	0.12114	0.02416	5.01	0.000

S = 0.1022   R-Sq = 94.2%   R-Sq(adj) = 93.9%

In each case, the residual variation (given by *s* in the regression results) is much smaller than the variation for the given output in the 100 trucks. We assume this closely matches the baseline variation since the data for the feedforward investigation were collected over a relatively long time. Feedforward control using the frame geometry has the potential to greatly reduce pull variation. As described in Chapter 17, the team was able to use a model to predict camber and caster based on the truck frame geometry inputs and compensate for the effect if necessary.

17.2 **Engine assembly problems occurred due to a poor fit between the pistons and the engine bore. The dominant cause of poor fit was found to be variation in the clearance, the difference between the (minimum) bore diameter and the (maximum) piston diameter. To solve this problem, the team thought about using the feedforward (selective fitting) approach. The idea was to measure each piston diameter and place them into bins of similar diameter. Then, after each bore diameter was measured, a piston would be selected from the appropriate bin. To assess this proposal the diameter measurements for 469 pistons and bores, as measured from nominal, are given in the file *block bore diameter feedforward*. Quantify the expected reduction in clearance variation when using one (that is, no selective fitting), two, three, or four bins of pistons. A suggestion is to define the bins by dividing the range in piston and bore diameters (roughly –10 to 10 microns) into equal widths.**

A numerical summary of the data is:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
piston	469	–0.124	0.000	–0.147	3.219	0.149
bore	469	–0.366	–0.500	–0.366	2.899	0.134
clearance	469	–0.242	0.000	–0.190	4.316	0.199

Variable	Minimum	Maximum	Q1	Q3
piston	–9.000	11.000	–2.000	2.000
bore	–10.100	9.200	–2.200	1.500
clearance	–15.100	12.800	–2.750	2.500

The range in piston diameters roughly matches –10 to 10, and the variation in bore and piston diameters is very similar. The standard deviation of the clearance for the given random pairing between piston and bore is 4.3. This provides an estimate of the process performance without selective fitting.

To quantify the expected variation in clearance using selective fitting we use simulation. We have two options: we can sample from the existing data to simulate the effect of selective fitting or we can model the piston and bore diameter variation and sample from the model. In this case, since both the piston and bore diameters have a bell-curve shape, as shown by the histograms that follow, they are well modeled by Gaussian random variables.
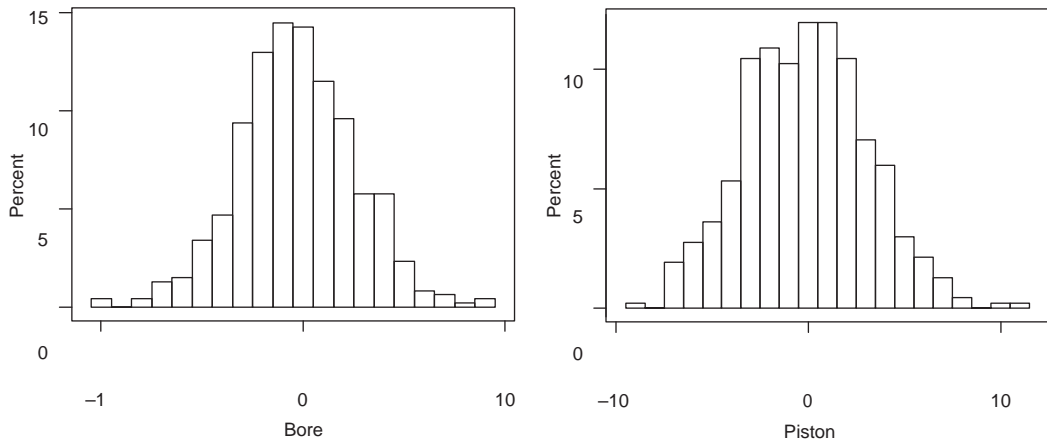
To implement the first simulation option where we sample from the existing data, we wrote a MINITAB macro that randomly selected a number of bores and corresponding pistons assuming a given number of bins, and calculated the clearance. See Appendix A for more information on writing MINITAB macros. The MINITAB macro that assessed the effect of using two bins, defined as pistons and bores with negative or positive diameter for this example, follows.

```
MACRO
#with the data file open call the MACRO from the command line as below
# %'macro2piston.txt' 'piston' 'bore'

macro2piston piston bore

mconstant I I2 cpiston borenow pnow test clearan
mcolumn piston bore borelist allvals temp

let cpiston=1
let  I2=1
let temp=1
let  allvals=0
```

```
#repeat 50 times to get good estimate of the standard deviation
while I2<=50
Sample 200 bore borelist      #get a list of 200 bore diameters at random
let I=1
while I<=200
let borenow=borelist(I)     #look at the next bore

if borenow>0    #try to find a piston that is also bigger than nominal
  let   test=0
  while  test<=0
    let  pnow=piston(cpiston)     #look at the next piston
    let cpiston=cpiston+1
    if cpiston>469  #end of list, start again
      let cpiston=1
      endif
    if pnow>0
      let test=1  #found appropriate match
      endif
    endwhile
  let clearan=borenow-pnow
  endif
if borenow<=0    #try to find a piston that is also smaller than nominal
  let  test=0
  while  test<=0
    let  pnow=piston(cpiston)     #look at the next piston
    let cpiston=cpiston+1
    if cpiston>469    #end of list, start again
      let cpiston=1
      endif
    if pnow<=0
      let  test=1     #found appropriate match
      endif
    endwhile
  let clearan=borenow-pnow
  endif
  if (I2=1) AND (I=1)
        let allvals=clearan
      else
    Stack allvals clearan allvals.    #store all the clearance values
  endif

  let  I=I+1
  endwhile
let I2=I2+1
endwhile

let test = MEAN(allvals)
print  test
Let test = STDEV(allvals)
print  test
ENDMACRO
```

Alternatively, we can simulate choosing random bores and finding the matching pistons from the Gaussian models in some programming language. We next give Matlab (see http://www.mathworks.com) code that allows us to explore the effect of any number of bins.

```
function []=selfitkbins(k,num)
%simulate the effect of using k bins for the piston/bore selective %fitting
    example
%num equals the number of simulated bore/piston combinations

%the estimated distribution of piston and bore diameters are Gaussian %with
    mean and standard deviation
pm=-0.12; ps=3.2;
bm=-.37; bs=2.9;

%we use k bins of pistons, divide the natural range -10 to 10 for the
    %pistons into intervals of equal width
% a better choice would be to use bins of roughly equal frequency
%we use the same bins definition for both pistons and bores
%this works because here the distributions are close to the same

%bin intervals bins=[-inf,-
10+(20/k):(20/k):10-(20/k),inf];

bores=normrnd(bm,bs,num,1);     %generate a random sample of bore diameters
%generate more pistons so we have enough in each bin
pistons=normrnd(pm,ps,num*2,1);
%match the bores with the appropriate pistons
allclear=[];   %keep track of all the clearance values

for i=1:k,  %find the bores in each group
    bind=find(bores<bins(k+1));
    bind2=find(bores(bind)>=bins(k));   %all the bores that match the group
    pind=find(pistons<bins(k+1));
    pind2=find(pistons(pind)>=bins(k));     %all pistons in appropriate bin
    len=length(bind2);  %number of bores to match
    if len>0,  %there are some bores in the bin
        %find some pistons that would match
        temp=pistons(pind2);  %pistons that match
        clearance=bores(bind2)-temp(1:len);
        allclear=[allclear;clearance]; %keep track of all clearance values
    end;
end;
hist(allclear)
mean(allclear)
std(allclear)
```

Conducting the simulation using either method, the expected reduction in clearance variation is:

Number of bins	Expected clearance standard deviation
1	4.3
2	2.6
3	1.8
4	1.6

The benefits of selective fitting are large even for only two bins. The team needed to decide whether the substantial costs and logistical difficulties of implementing feedforward control warranted adopting the approach. To implement selective fitting, all piston and bore diameters must be measured. In addition, pistons need to be placed in the appropriate bin. Logistical problems can occur if one of the bins runs out. In this example, the team decided that while feedforward was technically feasible other approaches should be considered due to concerns about the high cost of implementing selective fitting.

17.3 **In the V6 piston diameter example discussed in Chapter 11, the team found that piston diameter after Operation 270 was a dominant cause of the final diameter. The data are given in the file *V6 piston diameter variation transmission*. This suggested that feedforward control might be a feasible approach.**

    a. **What are the requirements for feedforward to be feasible in this context?**
    b. **If feedforward were feasible, assess the potential benefit using the results of the variation transmission investigation.**
    c. **Could the team also use the diameter after Operation 200, rather than the diameter after Operation 270, as the input to a feedforward controller?**

    a. For feedforward to be effective, there must be a way to quickly and cheaply measure the diameter after Operation 270 and then to adjust the process to compensate for large deviations between the predicted final diameter and the target. The prediction comes from the regression equation found in the analysis of the variation transmission investigation. Since there was no way to make the compensating adjustment, feedforward was not feasible in this example.
    b. Using the results of the variation transmission investigation, we fit a regression equation relating the final diameter (after Operation 310) and the diameter after Operation 270. We get:

The regression equation is
diameter after OP310 = 64.3 + 0.884 diameter after OP270

Predictor	Coef	SE Coef	T	P
Constant	64.27	22.29	2.88	0.005
diameter	0.88358	0.03739	23.63	0.000

S = 1.224   **R-Sq = 85.6%**   **R-Sq(adj) = 85.4%**

The strong relationship is also shown in the plot that follows:
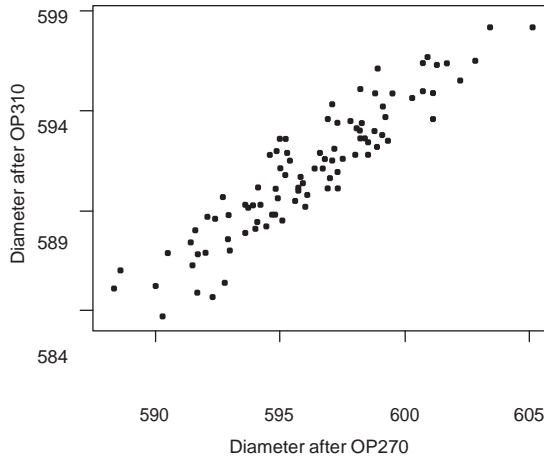
Based on the regression model, using feedforward control and assuming perfect compensation would reduce the standard deviation in the final diameter to 1.2. This is a large reduction from the baseline standard deviation 3.32.

c. The diameter after Operation 200 is not a dominant cause of the final diameter variation. We can see this clearly in the regression results and plot given as follows:

The regression equation is
diameter after OP310 = 225 + 0.325 diameter after OP200

Predictor	Coef	SE Coef	T	P
Constant	225.12	72.72	3.10	0.003
diameter	0.32542	0.06467	5.03	0.000

S = 2.862   R-Sq = 21.2%   R-Sq(adj) = 20.4%
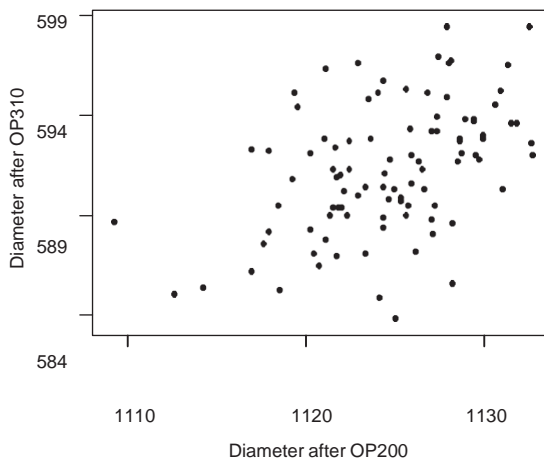
Feedforward control based on the diameter after Operation 200 would not work. The Operation 200 diameter does not provide a good prediction of the final diameter. The regression results suggest that if we implement perfect feedforward control based on the Operation 200 diameter, the final diameter variation

could be reduced to 2.86 from 3.32 in the baseline. One advantage over using the Operation 270 diameter is that there is more opportunity to find an adjuster.

CHAPTER 18

18.1 The bias of the system used to measure camshaft journal diameters tended to increase over time. The cause of this increase was not determined. Instead, the team introduced a feedback controller. At the start of each shift, a master journal with known diameter was measured. If the measured diameter deviated from the known value by a large amount, the measurement system was recalibrated.

a. How could we sensibly define a large deviation in this context?
b. What would happen to the measurement variation if the measurement device were recalibrated every time the master journal was measured, rather than only when the deviation from the known dimension was large?

a. We define large by comparing the measurement error to the short-term measurement variation. For instance, we may decide to recalibrate if the difference between the measured value and the known value differ (in absolute value) by more two times the standard deviation of the measurement system over the short term.

b. If the change in measurement bias were slow relative to the short-term measurement variation, adjusting for any deviation from the known value would increase diameter variation. This is tampering as defined by Deming (1986).

18.2 In a machining process, the dominant cause of dimension variation acted in the setup family. That is, the dimension variation within a particular setup was small relative to the variation from one setup to the next. The existing control plan called for a complete process adjustment back to the target based on the first observation after each setup. There were more than 200 parts machined between setups. The baseline dimension standard deviation was 0.31. The team decided to explore a new feedback control scheme based on the average for the first five observations after each setup. In an offline investigation, they carried out 10 setups and produced 20 parts after each setup without any adjustment. The dimension data, scaled so that the target is zero, are given in the file *machining dimension feedback***.**

a. Use a one-way ANOVA to estimate the standard deviation if the process could be adjusted so that the dimension averages across all setups were equal.
b. Use simulation to compare the performance of the existing feedback controller with the proposed controller that makes a complete adjustment based on the average for the first five observations after each setup.
c. In general, we may design a feedback controller by averaging the output from the first *n* observations after each setup. What considerations help you decide how many observations should be used to estimate the process average after each setup?

a. Plotting dimension by setup we get:
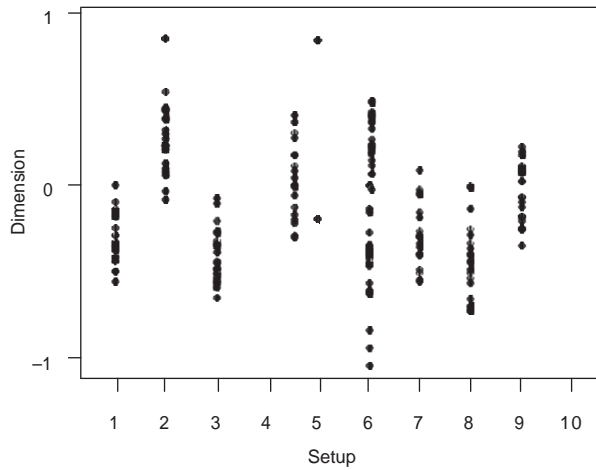
From the plot we can see the setup-to-setup variation. Using a one-way ANOVA we can partition the overall variation into the within-setup component and the variation between setups. The edited results of the ANOVA from MINITAB are:

One-way ANOVA: dimension versus setup

Analysis of Variance for dimension

Source	DF	SS	MS	F	P
setup	9	15.7998	1.7555	43.56	0.000
Error	190	7.6573	0.0403		
Total	199	23.4572			

Pooled StDev = 0.2008

The pooled standard deviation 0.2008 estimates the within-setup variation, the dimension standard deviation if we could adjust the process to keep the setup averages equal. This is substantially lower than the baseline value 0.31.

b. We can simulate the performance of the two adjustment schemes by:

Scheme 1: For each setup, subtract the first observation from the remaining 19 to model the adjustment. Find the standard deviation of the 20x19 adjusted values to estimate the process performance.

Scheme 2: For each setup, calculate the average of the first five observations and subtract the average from the remaining 15 to model the adjustment. Find the standard deviation of the 20x15 adjusted values to estimate the process performance.

We get:

Adjustment scheme	Estimated process standard deviation
None	0.34
First observation	0.29
Average of first five	0.22

By using the average of five parts, we can nearly reach the estimated minimum value 0.20. These comparisons are valid if the number of parts produced between setups is much greater than 20 pieces. If the number of parts was only 20 pieces, we should include the unadjusted parts (first one or first five parts) in the estimated standard deviation.

c. We need to ask questions like:

How much variation is there within each setup? If there is little variation, we need fewer observations.

How many parts are machined between setups? If there are only a small number of parts (or if the cost of poor parts is high), we may wish to make an initial adjustment based on a small number of observations and then adjust again once a few more observations are available.

What are the adjustment costs? If the adjustment cost is large, we may decide to adjust only if the deviation from target is large.

18.3 In a machining process, there was excess variation in the diameter of a precision ground shaft. The shaft diameter was measured for all shafts using a complex automated gage (that also measured other outputs). Upon investigation, the team discovered that the dominant cause acted in the measurement family. In particular, the measurement bias changed from day to day, consistent with the pattern observed in the baseline. To explore this bias change further the team planned an investigation where the diameter of the same shaft was measured each hour for four days. A total of 32 diameter measurements were made. The data are given in the file *precision shaft diameter feedback*, with the output being the diameter measured from nominal. The results show a gradual drift. The team speculated that the drift was caused by changes in some (unidentified) environmental conditions. They decided to reduce the measurement variation using a feedback controller.

a. **What type of feedback controller (that is, what prediction equation and what adjustment rule) would you recommend in this application?**

b. **Suppose the team decided to use a feedback controller based on EWMA forecasts with the smoothing parameter alpha equal to 0.4. What kind of a reduction in the measurement variation could they expect?**

a. The change in the diameter is gradual (that is, does not have sudden shifts). As a result, a feedback controller based on the EWMA forecasts is reasonable. The adjustments should be made so that if the one-step-ahead forecast is correct and the adjustment is perfect, the next value will be zero (that is, diameter equals the nominal value).

b. Using simple exponential smoothing with alpha equal to 0.4 gives:
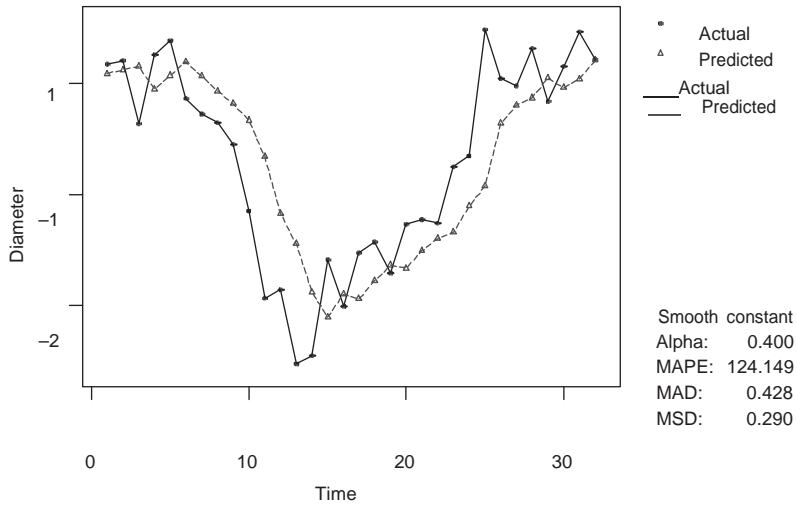
From these results, we predict that with the feedback controller, the measurement standard deviation would be reduced to roughly 0.54 ($\sqrt{0.29}$). Note that given the small sample size, we are not confident of this prediction.

If instead we try to optimize the value of alpha, MINITAB gives:
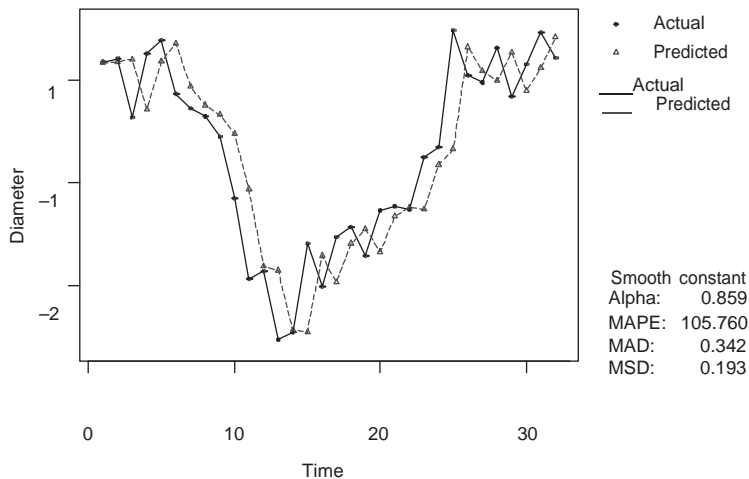
However, because of the small sample size, we are not confident that alpha = 0.86 would be substantially better than alpha equal to 0.4 when applied to the actual process.

CHAPTER 19

19.1 In the paint film build example introduced in Chapter 3, the baseline standard deviation in film build (paint thickness) was 0.67 thousandths of an inch. With this variation, to ensure a minimum film build of 15-thousandths of an inch, the process was centered at 17. The goal was to reduce the standard deviation to 0.35, thereby allowing for a reduction in the average film build.

The dominant cause of film build variation was found using a multivari investigation to act in the car-to-car family. Despite further effort, the dominant cause was not found. The team decided to adopt the process robustness approach. Based on process experience, candidates and their corresponding levels were chosen as follows:

Candidate	Low level	High level
Anode dimension	3.1	3.9
Conductivity of paint	Low	High
Temperature	30	50
Zone X voltage	450	475
Zone Y voltage	500	525

The team selected a fractional factorial resolution V experiment with the 16 treatments given as follows. To reduce the cost of the experiment, panels were used rather than cars. With this choice there was a risk of study error.

Treatment	Anode dimension	Conductivity of paint	Temperature	X voltage	Z voltage
1	3.1	Low	30	450	500
2	3.9	Low	30	450	525
3	3.1	High	30	450	525
4	3.9	High	30	450	500
5	3.1	Low	50	450	525
6	3.9	Low	50	450	500
7	3.1	High	50	450	500
8	3.9	High	50	450	525
9	3.1	Low	30	475	525
10	3.9	Low	30	475	500
11	3.1	High	30	475	500
12	3.9	High	30	475	525
13	3.1	Low	50	475	500
14	3.9	Low	50	475	525
15	3.1	High	50	475	525
16	3.9	High	50	475	500

For each run, five panels were painted. The order of the treatments was randomized. Since the dominant cause acted car to car, the team believed the unknown dominant cause would act within each run. Film build was measured at five locations on each panel. The data for one location are given in the file *paint film build robustness* and in the table that follows.

Treatment	Order	Film build	Average	Log(s)
1	14	15.6, 15.3, 15.9, 15.2, 15.8	15.56	−0.51
2	5	16.0, 16.3, 17.3, 16.2, 16.6	16.47	−0.31
3	6	15.0, 14.8, 14.9, 15.3, 16.1	15.22	−0.28
4	2	16.1, 17.6, 17.2, 16.3, 16.1	16.69	−0.16
5	9	15.7, 15.6, 15.2, 15.2, 15.7	15.49	−0.57
6	12	17.3, 17.6, 16.8, 17.5, 17.3	17.28	−0.49
7	13	16.2, 14.4, 15.4, 14.5, 15.9	15.30	−0.09
8	4	17.3, 16.6, 16.6, 16.4, 17.8	16.94	−0.25
9	7	16.1, 14.7, 16.2, 14.7, 16.2	15.59	−0.09
10	16	17.2, 15.8, 16.4, 16.0, 15.8	16.23	−0.24
11	15	15.4, 15.2, 15.4, 15.3, 15.2	15.29	−1.06
12	1	16.6, 16.4, 16.4, 16.5, 16.4	16.48	−1.00
13	3	15.1, 15.4, 15.4, 15.0, 14.4	15.05	−0.41
14	10	16.8, 16.9, 17.0, 17.3, 16.3	16.89	−0.42
15	11	15.0, 15.1, 15.0, 14.9, 14.8	14.97	−0.86
16	8	16.6, 16.7, 16.3, 16.5, 16.3	16.48	−0.79

a. **Analyze the data using the standard deviation of film build over the five consecutive panels to measure performance. Is it possible to make the process robust to noise variation? What levels of the candidates do you suggest?**

b. **The team had a way to adjust the process center. However, we can also use the robustness experiment to look for an adjuster. Analyze the data using the average film build over the five consecutive panels to measure performance. Are any of the candidates adjusters?**

c. **In the experiment, the film build at a particular location on five consecutive cars (panels) was used to define a run. Suppose, instead, that the five observations came from five fixed locations on a single door. What, if any, changes are needed in the analysis presented in part a?**

a. Plotting the film build by treatment suggests that treatments 11, 12, 15, and 16 are promising.
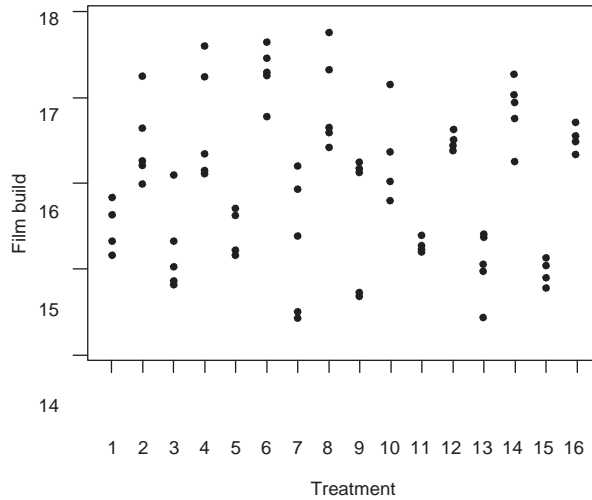
Using log standard deviation as the response, in MINITAB we get a Pareto chart of the effects, given as follows, that shows there are large main effects due to conductivity, zone *X* voltage, and large interactions between conductivity and zone *X* voltage and between conductivity and temperature.
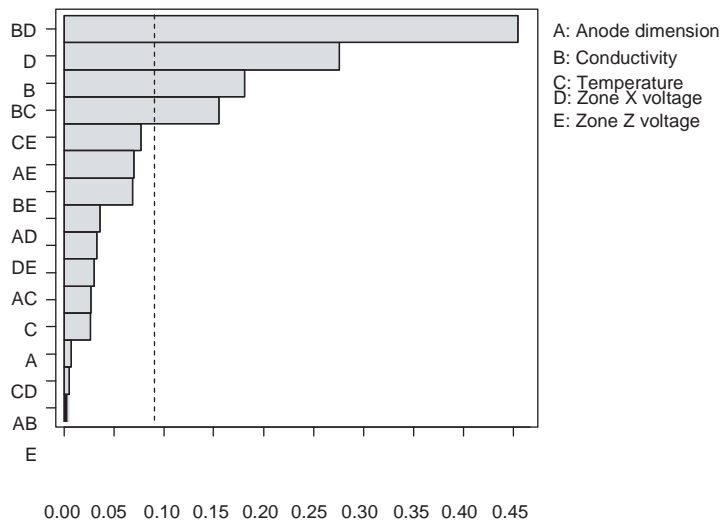
To draw conclusions we also look at the following main effects and interaction plots.
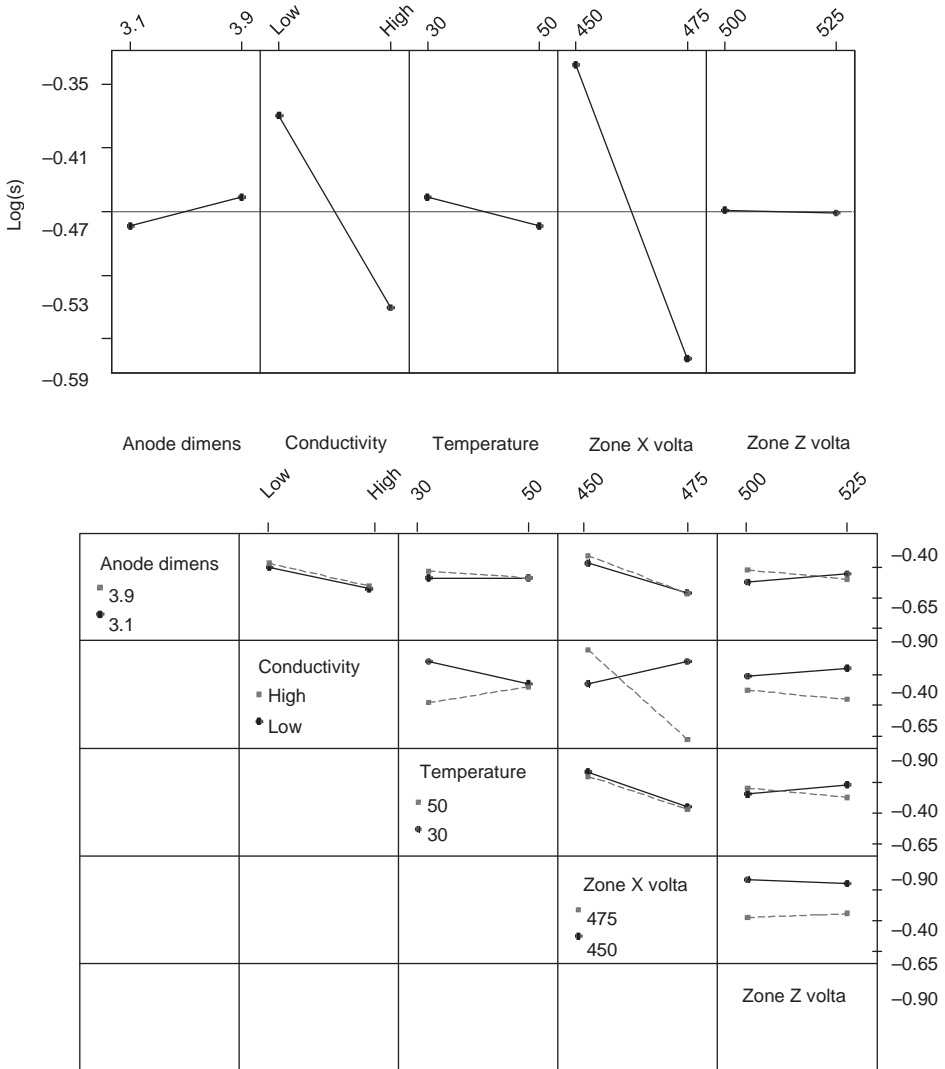
Recall that smaller log(*s*) is better. From the interaction plots, the combination of high zone *X* voltage, high conductivity, and low temperature is best. We are fortunate that high conductivity is best in both the large interactions. The high level of zone *X* voltage and conductivity, and low level for temperature corresponds to treatments 11 and 12. Using the new process settings is expected to reduce the baseline standard deviation of 0.67 to 0.09, since the average log(*s*) for treatments 11 and 12 is -1.03 and $10^{-1.03} = 0.09$.

With the new process settings, the average film build is about 15.9. The team had a way to adjust the film build average. The reduction in variation suggested a reduction in the average film build to around 16.25 from the current 17. This would translate into a 4% reduction in paint volume used. After implementing this solution the dominant cause acted within a door. There is ongoing effort to address this source of variation.

b. Looking for an adjuster, we analyze the experimental results using film build (or average film build in each run) as the output. The Pareto plot of the effects is:

A: Anode dimension
B: Conductivity
C: Temperature
D: Zone X voltage
E: Zone Z voltage
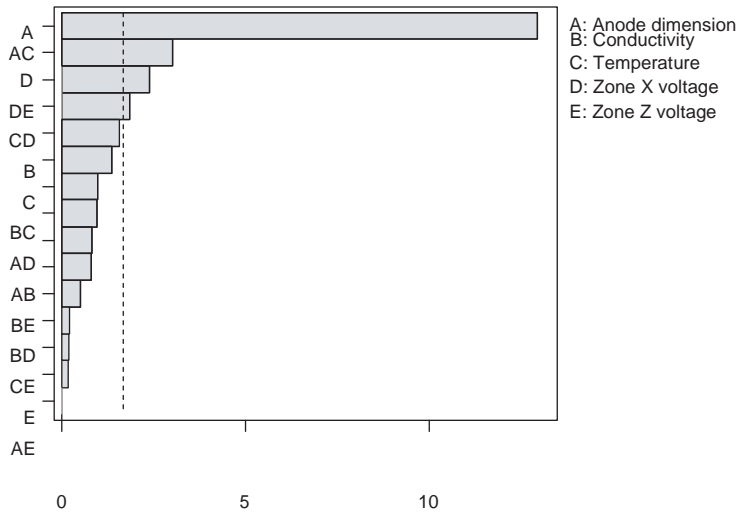
   Input *A* (anode dimension) is an adjuster. From the plot of film build by anode dimension, given as follows, we see that lowering the average dimension lowers the average film build.
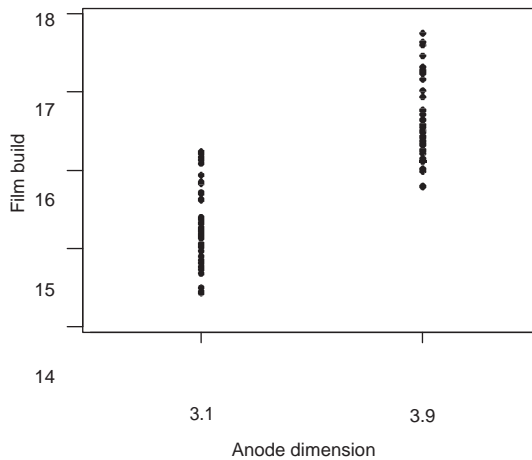
c. When measuring the same five locations on each door, we may expect a systematic difference between the door locations. In the analysis we would want to keep track of the location and use an analysis as in Chapter 16. Using a performance measure like log(s) is not advised.

19.2 **In a trim plant, customer complaints about seat appearance prompted management to assign a team the task of reducing shirring variation. The team proceeded without using Statistical Engineering and made a number of mistakes. Seat cover shirring was scored on a scale of 1 to 6 using boundary samples by how much cloth was gathered by the parallel stitching. Shirring scores of 1 to 4 were acceptable with 1 being the best. Scores of 5 or 6 resulted from either too much or too little shirring. A review of historical data suggested that the observed shirring score over a week covered all six possible values. Next, the team informally checked the measurement system. They found the measurement system added little variation. The team decided not to look for a dominant cause. Rather they moved directly to assessing the feasibility of making the process robust. They used brainstorming to select six candidates with two levels each as follows:**

Candidate	Low level	High level
Leather thickness	0.8	1.2
Leather toughness	Pliable (soft)	Stiff (tough)
Seam width	9 mm	11 mm
Material feed	Top up	Bottom up
Steam to skin bun	Used	Not used
Bun thickness	+5 mm	–5 mm

The team planned a resolution III fractional factorial experiment with 16 runs (one for each treatment) as follows:

Treatment	Leather thickness	Seam width	Leather toughness	Machine feed	Steam	Bun thickness
1	High	Low	Tough	Top up	Yes	High
2	High	Low	Soft	Bottom up	No	Low
3	Low	High	Tough	Top up	No	Low
4	Low	High	Soft	Bottom up	Yes	High
5	High	High	Tough	Bottom up	Yes	Low
6	High	High	Soft	Top up	No	High
7	Low	Low	Tough	Bottom up	No	High
8	Low	Low	Soft	Top up	Yes	Low
9	High	High	Tough	Bottom up	No	Low
10	High	High	Soft	Top up	Yes	High
11	Low	Low	Tough	Bottom up	Yes	High
12	Low	Low	Soft	Top up	No	Low
13	High	Low	Tough	Top up	No	High
14	High	Low	Soft	Bottom up	Yes	Low
15	Low	High	Tough	Top up	Yes	Low
16	Low	High	Soft	Bottom up	No	High

Each run consisted of three seats (repeats). The runs were conducted in the treatment order given in the table. The data are given in the file *seat cover shirring robustness* and reproduced as follows:

Treatment	Order	Seat 1	Seat 2	Seat 3	Average score
1	13	3	1	2	2.0
2	16	1	2	1	1.3
3	7	2	2	2	2.0
4	6	2	2	2	2.0
5	10	2	1	1	1.3
6	1	3	1	3	2.3
7	11	4	2	1	2.3
8	15	2	2	4	2.7
9	5	1	2	2	1.7
10	3	4	5	2	3.7
11	14	3	3	2	2.7
12	9	2	3	3	2.7
13	8	1	2	2	1.7
14	2	2	2	3	2.3
15	4	1	4	2	2.3
16	12	2	3	1	2.0

a. **Explain why choosing the process output as a measure of variation (that is, high scores come from either too much or too little shirring) was a poor one.**
b. **The goal is to find process settings that lower the average shirring score. Can we use any of the candidates to achieve the goal?**
c. **Each run consisted of three seats. Discuss this choice in the context of a robustness experiment.**

For the last two parts of this question, suppose the first three candidates (leather thickness, leather toughness, and seam width) used in the robustness experiment were normally varying rather than fixed inputs.

d. **How should the levels of the first three inputs have been chosen?**
e. **Discuss changes you would make to the analysis you conducted in part b.**

a. We should avoid defining the output in terms of a measure of variation if possible. With the shirring score as defined, it is difficult to find a dominant cause. Low or high values of the dominant cause both lead to high scores. The team was forced into the Make the Process Robust approach. A better scoring system would have given low scores for too little shirring and high scores for too much.

b. We first plot the shirring values versus the treatment number. In the plot that fol-
lows, we have added some jitter in the vertical direction (see Appendix C) to
eliminate the problem of overplotting.
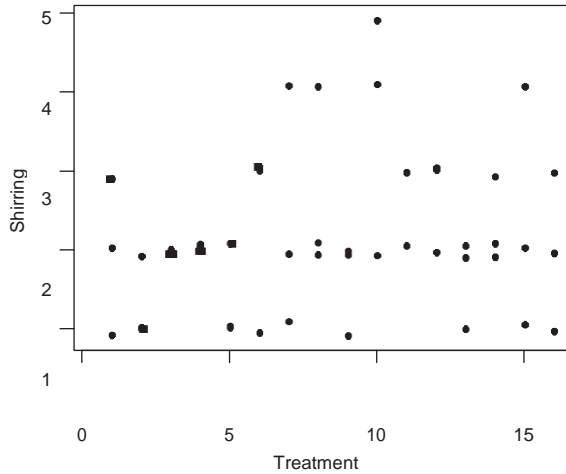
There are some promising treatments that result in a low average shirring
score. Next, we fit a model with all possible terms and present the results using a
Pareto plot of the effects.
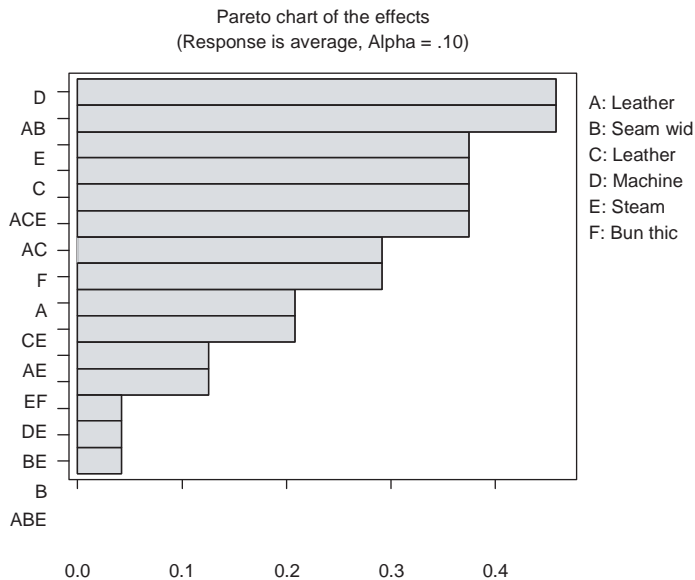
There are no large effects. We may be tempted to choose one of the treatments
that resulted in a low average shirring score as the new process settings. However,
consider the answer to part c.

As an aside the particular fractional factorial design selected has the confound-
ing structure given as follows. The inputs are labeled *A, B,* and so on, using the
same order as in the preceding tables. The design is resolution III. This is a poor
choice, since with 16 runs in six candidates a resolution IV design is possible.

Alias Structure (up to order 3)

I – A*D*F – B*C*F
A – D*F + B*C*D
B – C*F + A*C*D
C – B*F + A*B*D
D – A*F + A*B*C
E
F – A*D – B*C
A*B + C*D – A*C*F – B*D*F
A*C + B*D – A*B*F – C*D*F
A*E – D*E*F
B*E – C*E*F
C*E – B*E*F
D*E – A*E*F
E*F – A*D*E – B*C*E
A*B*E + C*D*E
A*C*E + B*D*E

    c. In a robustness experiment, we do not know the dominant cause, and we define a run over sufficient time so that the unknown dominant cause has time to act. In this example, the team did not know the time family of the dominant cause. It seems unlikely that three repeats (seats) for each run is sufficient. This means that the dominant cause has likely not acted in (most of) the treatments. With this choice of run, the experiment will not be able to identify process settings that are robust. The experiment was doomed to failure because of poor planning.

    d. The levels for each suspect should be selected at the extremes of the normal range of values. This requires measuring the value of the three suspects over a sufficient time span to see the full range before planning the robustness (desensitization) experiment.

    e. The goal of the experiment would change to finding settings of the last three inputs that result in a lower average shirring score. It no longer makes sense to find the best settings of the suspects, since they are varying in the normal process. We would refer to only the last three inputs as candidates. The appropriate performance measure would be the average shirring score across all runs (and repeats) with a given combination of the three candidates. There are now only eight treatments in terms of the candidates. For example, we would average the shirring scores across runs 1 and 10. There is still no guarantee that the experiment will provide useful results. We do not know if the three suspects include important, let alone dominant, causes. This sort of an experiment, where both candidates and suspects are mixed together, is a common mistake when teams know about designed experiments but do not follow a structured problem-solving algorithm.

19.3 Torsional rigidity of the weather stripping was the dominant cause of door assembly problems. Management set a goal of reducing standard deviation in torsional rigidity to 0.3. A baseline investigation found the variation in torsional rigidity was roughly 0.55 mm and that the dominant cause acted over the short

term and certainly within any half hour. The team looked briefly for a dominant cause of rigidity variation without success. Next, they planned a robustness experiment with four candidates at two levels each, chosen based on engineering judgment. The candidates and levels are:

Candidate	Low level (−1)	High level (+1)
Heat (pre)	100	700
Extruder RPM	22	26
Tension (pre)	1	5
Water flow	2	6

The team planned a full factorial experiment with 16 runs, one for each treatment. The correspondence between treatments and candidate levels is given in the table that follows.

Treatment	Heat	Extruder RPM	Tension	Water flow
1	−1	−1	−1	−1
2	−1	−1	−1	1
3	−1	−1	1	−1
4	−1	−1	1	1
5	−1	1	−1	−1
6	−1	1	−1	1
7	−1	1	1	−1
8	−1	1	1	1
9	1	−1	−1	−1
10	1	−1	−1	1
11	1	−1	1	−1
12	1	−1	1	1
13	1	1	−1	−1
14	1	1	−1	1
15	1	1	1	−1
16	1	1	1	1

Each run consisted of running the process for half an hour after the candidate levels had been reached. Within each run, 10 weather-strip samples were

selected spread out over the half hour. The order of the runs was randomized.
The torsion rigidity of each of the 10 weather-strip samples for each treatment
is given in columns s1 to s10 of the table that follows and in the file *weatherstrip
torsional rigidity robustness*.

Treatment	Order	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
1	13	10.3	13.0	11.5	11.8	10.7	9.9	10.7	11.5	11.0	11.1
2	6	11.5	13.0	10.4	11.1	10.9	10.6	12.0	9.3	9.2	9.3
3	9	11.6	13.0	10.4	16.0	10.3	10.8	11.5	11.0	11.3	10.9
4	1	11.5	11.7	10.4	11.7	14.0	11.7	10.4	11.7	10.4	10.4
5	3	14.0	11.7	11.7	19.0	11.9	11.7	12.1	13.0	11.1	11.0
6	11	22.0	15.0	18.3	11.7	20.3	21.0	12.6	13.6	14.7	15.1
7	5	9.1	9.6	10.2	9.8	9.0	9.7	10.0	12.0	9.0	8.8
8	14	10.0	9.1	10.6	10.4	10.8	11.0	11.1	10.8	10.5	10.8
9	2	11.7	12.5	11.9	11.7	20.0	14.0	10.4	11.5	11.7	20.0
10	10	10.3	11.6	10.5	10.6	13.0	14.0	11.7	10.3	15.0	11.8
11	7	10.3	10.5	11.0	11.4	9.8	10.4	11.7	11.8	11.5	11.9
12	15	11.6	11.0	11.4	11.3	12.0	10.6	10.9	10.7	10.7	10.7
13	16	10.6	10.7	11.6	10.6	10.7	22.0	11.0	10.4	10.4	23.0
14	8	9.1	10.4	10.6	11.4	10.9	10.4	10.8	10.9	11.0	11.6
15	12	10.3	11.0	12.0	12.1	10.5	10.7	11.3	11.4	10.8	10.9
16	4	10.4	10.4	10.4	10.5	10.9	11.4	9.0	9.6	9.8	10.2

   a. **To analyze the results of this robustness experiment, what performance
measure(s) do you recommend and why?**

   b. **Analyze the experimental results using your chosen performance measure(s).
What can you conclude?**

a. The goal of the experiment is to find process settings that results in less variation
in torsional rigidity (as the unknown dominant cause acts). The performance
measure log(s), where s is the standard deviation of the rigidity values across the
10 repeats within each run, is appropriate. A secondary performance measure is
the average rigidity across each run. If we find process settings that result in less
rigidity variation, we may need to adjust the process center back to the target.

b. Plotting the torsional rigidity values by treatment, we get:
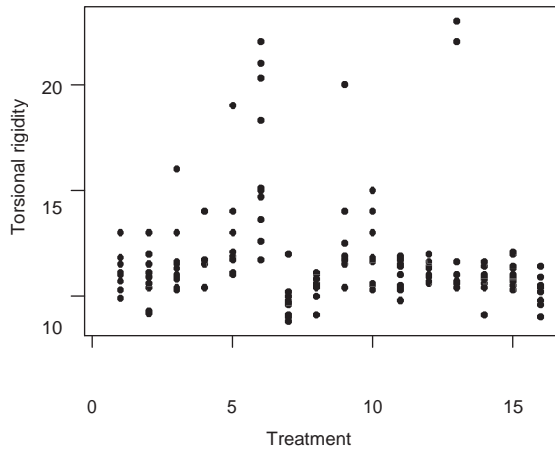
There are some treatments that have much less rigidity variation than others. Fitting a complete model (using the log(s) performance measure) a Pareto plot of the effects gives:
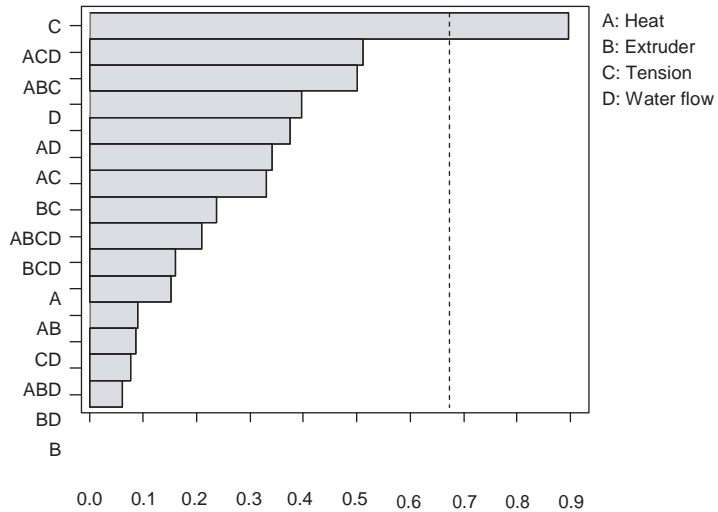
Candidate *C* (tension) has a large effect. Looking at the main effects plots suggests that the high level of tension gives substantially lower rigidity variation.
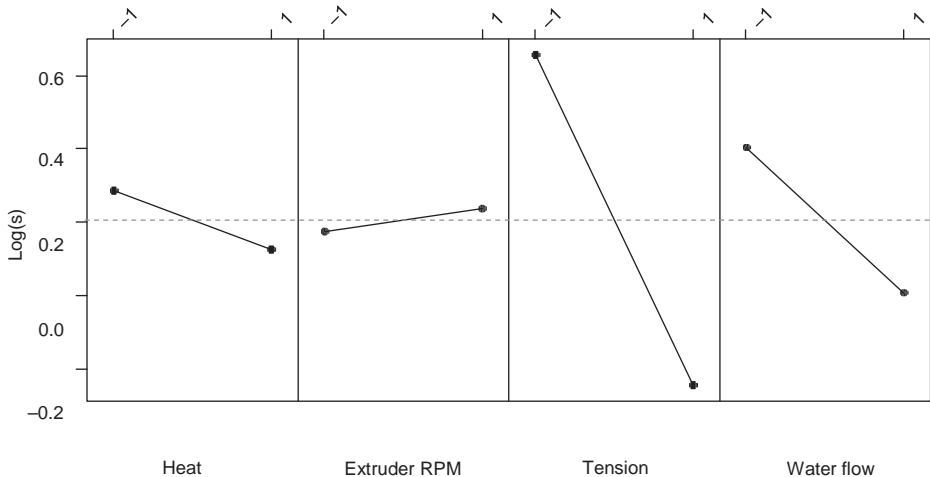
However, the average torsional rigidity variation at the high level of tension is 0.8 ($e^{-0.2}$), much higher than the baseline variation of 0.55. In fact, even the best treatment, number 12, has a variation of 0.47, only marginally better than the baseline. The experiment has identified a candidate (tension) that can be changed to *increase* rigidity variation. The experiment failed to find more robust process settings. The team was curious about this result and decided to investigate even higher tension levels. This investigation also failed since high tension led to other negative side effects and little reduction in rigidity variation. The approach was abandoned.

CHAPTER 20—NO

EXERCISES CHAPTER 21

21.1 Discuss whether lessons learned are properly maintained in corporate memory in your organization. What could be done to improve the situation?

The answer depends on the organization.

21.2 In the paint film build example described in Chapter 19, the team found new process settings that resulted in reduced car-to-car variation in film build. To validate the proposed solution, 80 cars were painted over one day with the settings given in the following table. These were the best settings found in the robustness investigation. The film build values from five specific positions on one door for each of the cars are available in the file *paint film build validation*.

Candidate	Setting
Anode dimension	3.5 (midpoint)
Conductivity of paint	High
Temperature	30
Zone X voltage	475
Zone Y voltage	500

    a. The baseline film build standard deviation was 0.68. The problem goal was to reduce the standard deviation to 0.35, and the robustness experiment results suggested that changing settings would reduce the standard deviation to about 0.37. Has the solution been validated?

    b. What, if anything, do the validation results tell us about the home of the dominant cause in the remaining variation?

a. We analyze the data as in the baseline investigation with the following numerical and graphical summaries.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
film build	400	16.208	16.222	16.212	0.302	0.015

Variable	Minimum	Maximum	Q1	Q3
film build	15.411	16.880	15.987	16.422
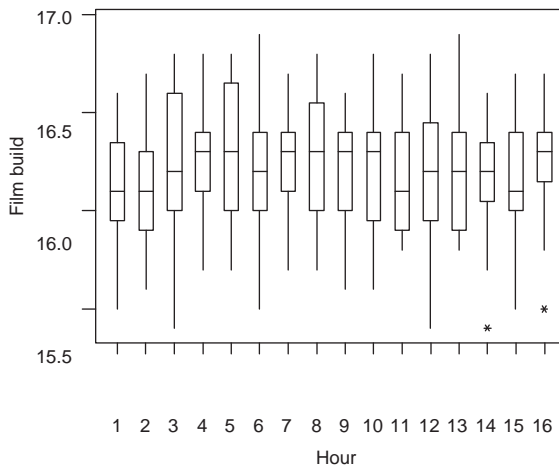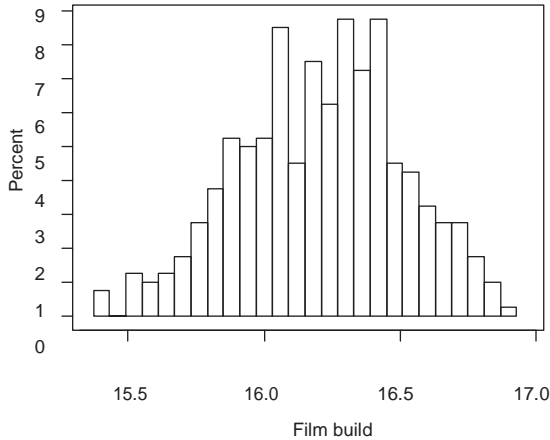
The overall standard deviation has been reduced to 0.302, a substantial reduction from the baseline standard deviation 0.68, and exceeding both the problem goal of 0.35 and the expected improvement. The average film build is 16.2. Further efforts were made to bring the average to 16, a little more than $3 \cdot 0.30$ above the minimum acceptable film build of 15 units. With these changes, the ultimate paint cost savings were about 6%.

b. As suggested by the plot that follows, the team could now look for further improvement by reducing the position-to-position variation in film build.
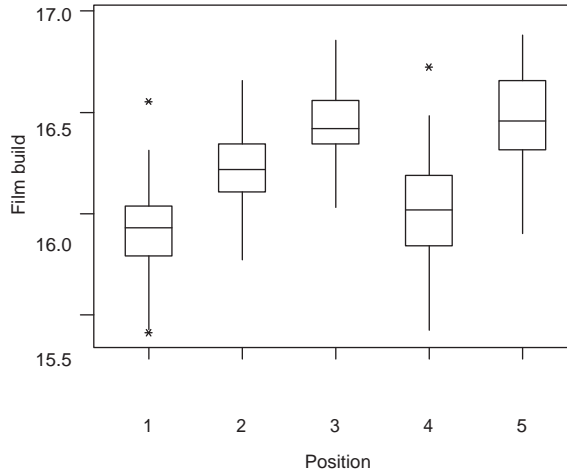
Looking at the results from a one-way ANOVA, we estimate that if we could eliminate the position-to-position variation, the film build variation would be further reduced to 0.21.

Analysis of Variance for film build

Source	DF	SS	MS	F	P
position	4	19.7020	4.9255	113.79	0.000
Error	395	17.0980	0.0433		
Total	399	36.8000			

Pooled StDev =   0.208

21.3 **In the truck pull example described in Chapter 17 and Exercise 17.1, a feedforward controller was implemented to compensate for the effect of truck-frame geometry on pull. After the feedforward system had been operating successfully for some time, management decided to review its operation. The four frame geometry measurements and left and right caster and camber were recorded for roughly a month of production consisting of over 6600 trucks. The data are given in the file *truck pull validation*.**

a. The standard deviations for caster and camber before implementation of the feedforward controller can be estimated from the 100-truck investigation described in Chapter 17. From the same investigation, the team predicted the possible reduction in standard deviation using a feedforward controller. A summary is given in the following table.

Characteristic	Baseline standard deviation	Predicted reduction in standard deviation
Left caster	**0.90**	**0.18**
Right caster	**0.83**	**0.20**
Left camber	**0.51**	**0.13**
Right camber	**0.41**	**0.10**

> **Do the results of the investigation validate the reduction in left and right caster variation due to the feedforward controller?**
> **b. For each of the two caster characteristics, conduct a regression analysis to see if the feedforward controller can be improved. Recall that the feedforward controller should be compensating for variation in the frame geometry.**
> **c. Repeat the analysis in parts a and b for left and right camber.**

a. To see if we can validate the process improvement, we summarize the caster characteristics from the validation investigation numerically and graphically.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
lcast	6632	3.5225	3.5120	3.5193	0.2461	0.0030
rcast	6632	4.0273	4.0260	4.0269	0.2314	0.0028

Variable	Minimum	Maximum	Q1	Q3
lcast	2.6630	4.8560	3.3570	3.6830
rcast	3.0650	4.9640	3.8680	4.1820
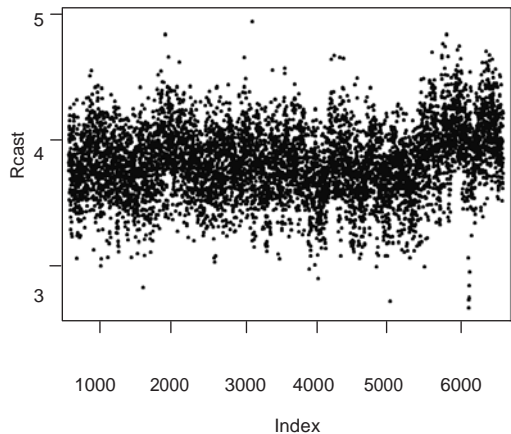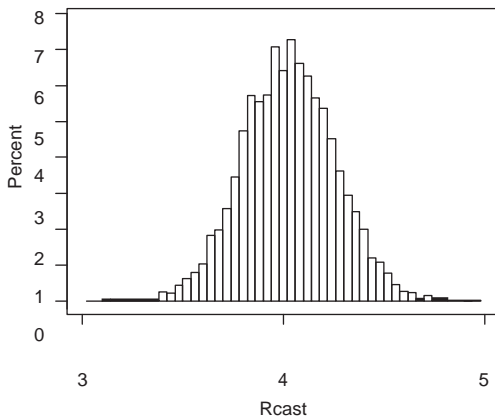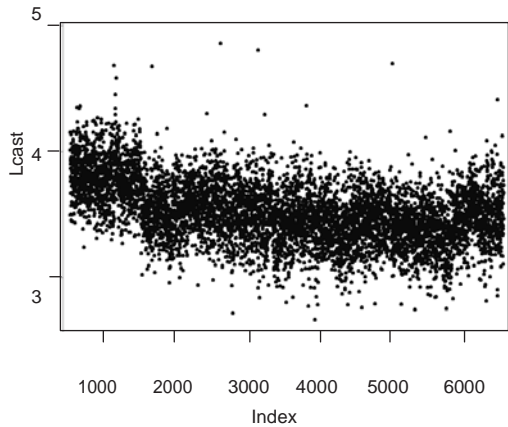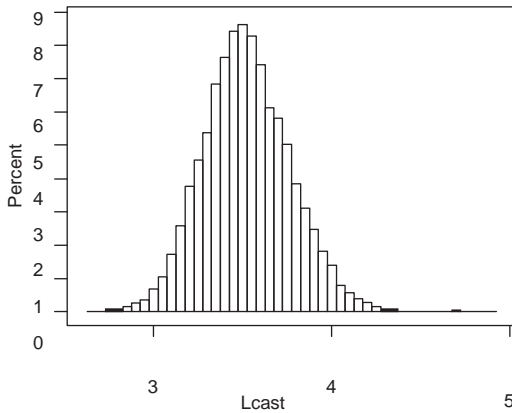
The standard deviations for left and right caster (0.25 and 0.23, respectively) are much smaller than they were before implementation of the feedforward controller.

The reduction in standard deviations did not completely meet the predictions based on the overly optimistic assumption that perfect compensation was possible. The feedforward controller was a great success.

b. To check the operation of the feedforward controller for the caster characteristics, we fit a regression model for left caster (and right caster) as a function of the truck frame characteristics—left front, right front, left rear, and right rear. If the feedforward controller is working properly, there will be at most a weak relationship between caster and the frame geometry characteristics.

The regression equation is
lcast = 3.86 + 0.00623 left front + 0.0732 right front - 0.0213 left rear
**   - 0.0978 right rear**

Predictor	Coef	SE Coef	T	P
Constant	3.85986	0.09292	41.54	0.000
left fro	0.006230	0.004943	1.26	0.208
right fr	0.073226	0.004031	18.16	0.000
left rea	-0.021255	0.004744	-4.48	0.000
right re	-0.097787	0.005149	-18.99	0.000

S = 0.2351   R-Sq = 8.8%   R-Sq(adj) = 8.7%

The regression equation is
rcast = 3.76 - 0.0718 left front + 0.0840 right front + 0.00547 left rear
**   + 0.00966 right rear**

Predictor	Coef	SE Coef	T	P
Constant	3.75977	0.08599	43.73	0.000
left fro	-0.071805	0.004574	-15.70	0.000
right fr	0.084034	0.003731	22.53	0.000
left rea	0.005466	0.004390	1.25	0.213
right re	0.009664	0.004765	2.03	0.043

S = 0.2176   R-Sq = 11.7%   R-Sq(adj) = 11.6%

From the solution for part a, the standard deviations for left and right caster are 0.2461 and 0.2314 when using the feedforward controller. The corresponding residual standard deviations in the regression analysis are 0.2351 and 0.2176. The residual standard deviations estimate the process standard deviation if we could hold the frame geometry fixed or totally compensate for its effect. Because the residual standard deviations are so close to the caster standard deviations with the existing feedforward controller, the effects of frame geometry on the outputs are very small. In other words, allowing for the inevitable small measurement and adjustment errors, the existing feedforward controller cannot be improved further using the frame geometry characteristics.

c. The summary of the current process performance, in terms of camber, is

Variable	N	Mean	Median	TrMean	StDev	SE Mean
lcmb	6632	0.49288	0.49000	0.49239	0.16324	0.00200
rcmb	6632	0.47708	0.48200	0.47883	0.16222	0.00199

Variable	Minimum	Maximum	Q1	Q3
lcmb	-0.42600	1.45600	0.38200	0.60100
rcmb	-0.39300	2.05400	0.37600	0.58400

To check the operation of the feedforward controller for camber, we fit a separate regression model for left camber and right camber, as a function of the truck frame characteristics.

The regression equation is
lcmb = - 1.69 + 0.0761 left front + 0.00141 right front + 0.0727 left rear + 0.0669 right rear

Predictor	Coef	SE Coef	T	P
Constant	-1.69356	0.05600	-30.24	0.000
left fro	0.076120	0.002979	25.55	0.000
right fr	0.001407	0.002430	0.58	0.563
left rea	0.072739	0.002859	25.44	0.000
right re	0.066909	0.003103	21.56	0.000

S = 0.1417   R-Sq = 24.7%   R-Sq(adj) = 24.7%

The regression equation is
rcmb = 0.761 - 0.0482 left front + 0.0192 right front - 0.0242 left rear + 0.0275 right rear

Predictor	Coef	SE Coef	T	P
Constant	0.76091	0.05994	12.69	0.000
left fro	-0.048213	0.003189	-15.12	0.000
right fr	0.019163	0.002601	7.37	0.000
left rea	-0.024170	0.003060	-7.90	0.000
right re	0.027540	0.003321	8.29	0.000

S = 0.1517   R-Sq = 12.7%   R-Sq(adj) = 12.6%

The current standard deviations for left and right camber are 0.1632 and 0.1622, respectively. The corresponding residual standard deviations from the regression analysis are 0.1417 and 0.1517. The team was surprised by the left camber results and found a small error in the mathematical model used to predict left camber from the frame geometry. This error was corrected; however, because camber has a small effect on pull, there was no noticeable reduction in the pull standard deviation.

Chapter Supplements

Chapter 2 Supplement

Describing Processes

S2.1 PARETO PRINCIPLE

The Pareto principle was introduced by Vilfedo Pareto (1848–1923) to describe the distribution of wealth. Pareto found that in most societies, the majority of the wealth was controlled by a small number of people. Over time, the principle found application in many areas. J. M. Juran was the first to apply the idea to manufacturing processes. He coined the terms *vital few* and *trivial many* to describe problems (Juran et al., 1979). The Pareto idea is sometimes summarized as the *80/20 rule,* meaning, for example in a manufacturing context, that 80% of the process problems are attributable to 20% of the process issues.

   Pareto analysis is best presented in graphical form using a bar graph to illustrate the relative importance of the various categories. For example, Figure S2.1 shows a Pareto chart of one month's records of rod scrap data (discussed further in Chapter 6). The Pareto plot shows that 62% of the scrap was found at a grinding operation. This suggests focusing on grind scrap.

Defect	Grind	Bore	Broach	Assembly	Others
Count	85	24	14	6	2
Percent	64.9	18.3	10.7	4.6	1.5
Cum%	64.9	83.2	93.9	98.5	100.0
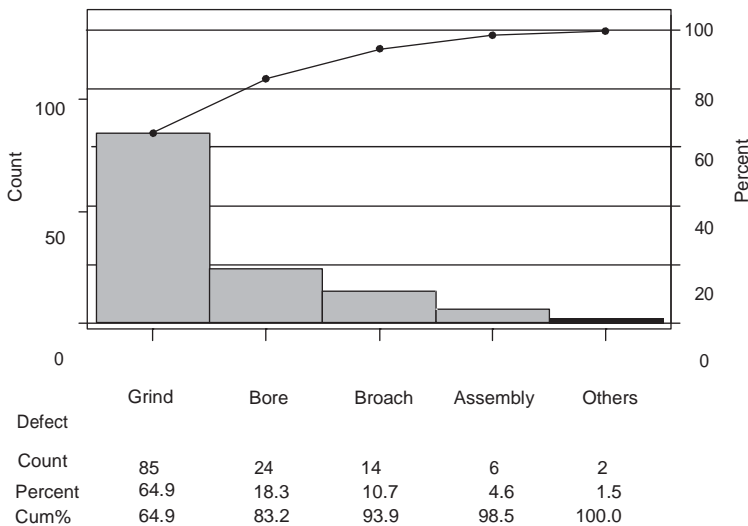
Figure S2.1   Pareto chart for scrap by operation for the rod line.

We use the Pareto principle extensively to choose projects and focus the problem. The Pareto principle implies that, for any problem, there are one or two dominant causes.

S2.2 DEFINING DOMINANT CAUSE

To define dominant cause we use the well-known conditional variance formula

$$Var(Y) = Var_x(E(Y|X)) + E(Var(Y|X))$$

See, for example, Kalbfleisch (1985). Replacing Y with the output and X with the input, we have

$$sd\left(output\right) = \sqrt{sd\left(E\left(output|input\right)\right)^2 + E\left(sd\left(output|input\right)^2\right)}$$

The first term under the square root sign can be thought of as the variation in the output explained by the input. The second term is the residual variation that would remain if we could hold the input constant. The input is a dominant cause if the first term is large compared to the second.

Formally, we call a cause *dominant* if it explains more than half the variation. This implies the residual standard deviation must be less than $1/\sqrt{2} = 0.71$, or roughly 70% of the overall standard deviation, sd(*output*).

A formal derivation of dominant cause when the output is discrete or binary is difficult. For a binary output, we use an informal definition and say an input is dominant if for some level of the input the proportion defective is substantially reduced.

S2.3 DOMINANT CAUSE INVOLVING TWO (OR MORE) INPUTS

For many problems we can find a dominant cause that involves a single input. However, for some problems the situation is more complex. Possibilities include:

1. The dominant cause involves a single input.

2. There are two (or more) large causes (that is, no dominant cause).

3. The dominant cause involves two (or more) inputs.

Later, we see that it is important to understand the difference between these three cases, especially when considering possible solutions. We illustrate using a simple situation where the output varies between 0 and 10, and there are two inputs that we examine at two levels each (low and high, coded –1 and +1). Figure S2.2 shows the four values as circles and also plots the average of the two values at each the low and high levels of input 1 using a diamond.
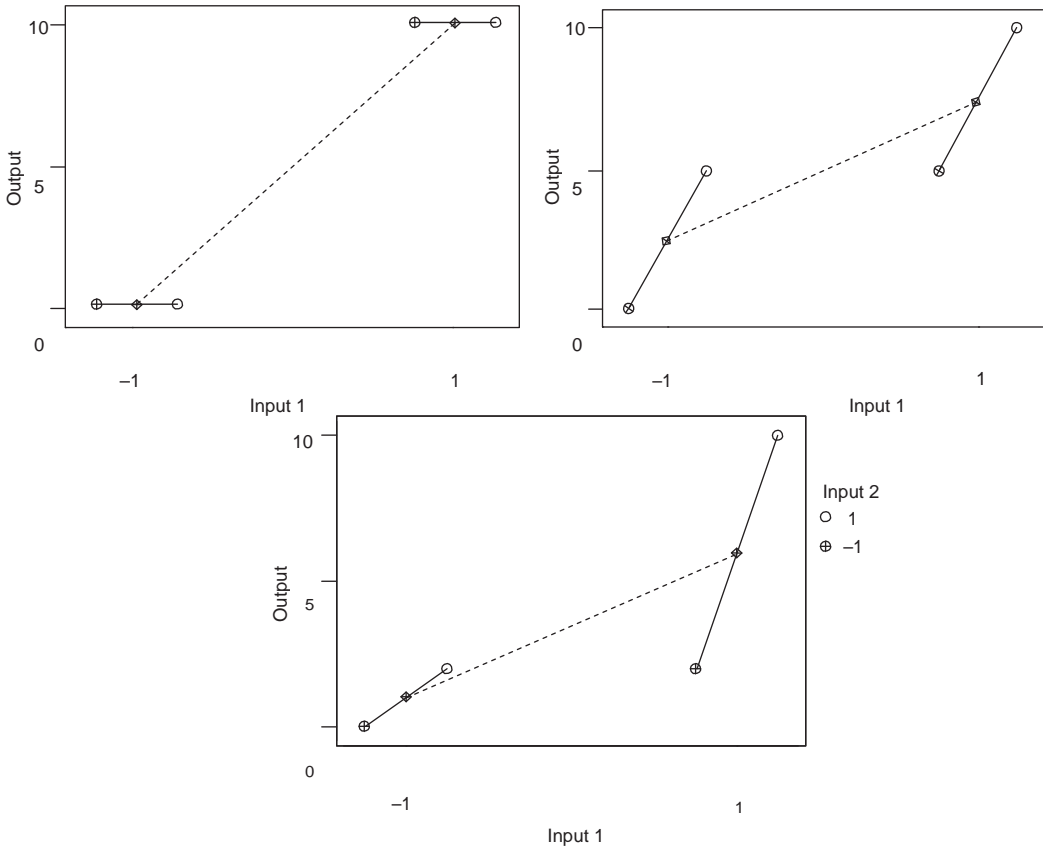
Figure S2.2   Plots illustrating the three cases for a dominant cause.

Case one is illustrated in the top left panel of Figure S2.2, where there is a single input (input 1) that explains the variation in the output. Here the second input has no effect on the output. In case two, as given by the top right panel of Figure S2.2, there are two large causes (inputs), but the effect of each input does not depend on the level of the other input (we say the effects of the two inputs on the output are additive). A dominant cause involving two inputs is presented in the bottom panel of Figure S2.2. This last case is the most complicated. We say a dominant cause involves two inputs if, for *both* inputs, we can observe close to the full extent of output values as the input changes for *some value of the other input*. In the plot, for the high level of input 1, changing the level of input 2 results in a large difference in the output, while similarly, for the high level of input 2, changing the level of input 1 results in a large difference in the output. There are gray areas among the three cases.

A dominant cause involving two or more inputs is related to the idea of an interaction. We say there is an *interaction* between input 1 and input 2 *in their effect on the output* if the relationship between input 1 and the output depends on the level input 2 and vice versa. Note that there can be an interaction between input 1 and 2 in their effect on the output even if there is no correlation between input 1 and 2 (that is, if input 1 and 2 vary independently of one another). We see the interaction between input 1 and input 2 clearly in the bottom panel of Figure S2.2, since the effect of changing from the low to high level of input 2 is much greater when input 1 is at its high level than when it is at its low level.

We return to the issue of the consequences of a dominant cause involving two (or more) inputs in Chapter 10 and its supplement, where we discuss finding such dominant causes, and in Chapter 14, where we consider possible solutions for problems with such a dominant cause.

S2.4 CLASSIFYING CAUSES OF VARIATION

There are several ways to classify causes of variation. For our purposes, we are interested in identifying inputs that make a large contribution to the variation of the output. We call these *dominant causes.* In the application of statistical process control (SPC), causes are classified as *common* or *special.* Taguchi (1986) uses the term *noise factor* and considers internal, external, and unit-to-unit noise. Here, we discuss the relationships among these classifications.

There is considerable confusion in SPC about the definition of special and common causes that are sometimes called assignable and chance causes. This classification depends on the control charting procedure in use and on how subgroup data are collected. A dominant cause may be either special or common. Suppose that subgroups for the control chart are based on five units produced consecutively by the process. If the dominant cause changes from part to part (that is, within the subgroups), it will be a common cause. If the dominant cause changes more slowly (that is, between subgroups), then it will be special.

We find that control charts are not useful in the identification of dominant causes. The main difficulties are that when the chart signals, this means that some input has changed but there is no clue given as to which input. The effect of the input is large with respect to the within-subgroup variation but may not be a dominant cause of the overall variation of the output. Furthermore, it is not predictable when the chart will signal and it is difficult to organize resources to look immediately for the cause. If the search is postponed, then the input may have changed again by the time it is scrutinized and the information about its effect will be lost.

Taguchi's noise factors are varying inputs that affect the output. In other words, noise factors are what we call *causes.* What Taguchi calls *control factors* are what we call *fixed inputs* that can be changed only by deliberate intervention. The use of the word *factor* comes from the language of experimental design, which plays a large role in Taguchi's variation reduction approach. He classifies causes as internal, external, and unit-to-unit. *Internal causes* are inputs that change over the life of the product. *External causes* are inputs that change in the usage environment, and *unit-to-unit causes* change from part to part at the time of production. In the refrigerator example discussed in Chapter 1, the identified external causes of frost buildup were ambient temperature and humidity, the amount of food added to the refrigerator, and the number of times the door was opened. An internal cause, not identified, was the deterioration of the door seals over time. The difference in compressor performance from one refrigerator to the next was a unit-to-unit cause. We discuss Taguchi's approach to variation reduction in chapters 16 and 19.

In general, we do not believe that using an experiment is an efficient way to *identify* a dominant cause of variation. Instead, we propose to use much simpler process investigations that do not involve changing fixed inputs or controlling varying inputs to search for such a cause. Within the Statistical Engineering algorithm, experimental design is needed to verify the contribution of a potential dominant cause, and in determining the feasibility of some of the variation reduction approaches.

S2.5 PROCESS CAPABILITY

There is an alphabet soup of capability ratios, such as P_{pk}, C_{pk}, and so on, that are used to quantify the relationship between the process center and variation, and the specification limits. See Kotz and Johnson (2002) and their discussion for a review of the merits and liabilities of various capability ratios. We give a brief discussion of P_{pk} since it is sometimes used to set the goal for a process improvement problem.

The index P_{pk} is defined as

$$P_{pk} = \frac{\text{distance from the process average to the closest specification limit}}{3 \cdot stdev}$$

where *stdev* is the standard deviation of the process output. Larger values of P_{pk} indicate smaller variation. The index is increased by better locating the average near the middle of the specification limits and/or by reducing the variation among the parts (that is, decreasing *stdev*). Note how P_{pk} is a function of both kinds of variation (off-target and part-to-part variation) and gets bigger as one or another kind of variation gets smaller.

For the angle error data discussed in the context of the camshaft lobe runout example in Chapter 2, the average and standard deviation are given as –21.3 and 71.5, respectively. The specification limits for angle error are ±400; thus the specification limit closest to the average is –400. Hence, the value of the capability index P_{pk} is

$$P_{pk} = \frac{-21.3 - (-400)}{3 \cdot 71.5} = 1.77$$

Such a large value of P_{pk} indicates that the closest specification limit is far from the average relative to the process variation, as described by three times the output standard deviation. Since the histogram for angle error data is bell shaped, this means that it is highly unlikely that any of the observed angle errors will be outside of the specification limits. This can be clearly seen in the histogram of angle error given in Figure S2.3, where we added dashed lines to show the specifications.
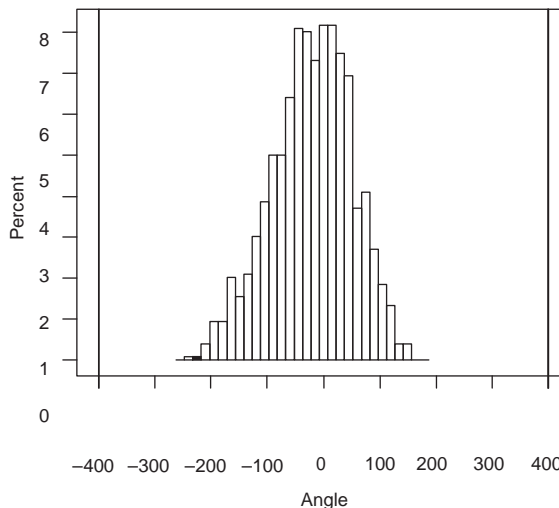
Figure S2.3   Histogram of angle error with specification limits.

Like all other summary measures of process performance, capability ratios depend on how the data are collected. If the sample is collected over a short time or when conditions are exceptional, the calculated capability ratio may be very misleading about the process performance in general.

S2.6 RELATING THE TWO KINDS OF VARIATION

We have described two kinds of variation, deviation of the average from the target and variation among parts as measured by the standard deviation. Suppose we have a set of n values for an output given by y_1, y_2,..., y_n. We provide a formula that relates the two kinds of variation to the variation from the process target T. The average and standard deviation are

$$avg \ \square \ \frac{y_1 + ... y_n}{n} \qquad stdev \ \square \ \sqrt{\frac{(y_1 - avg)^2 + ... + (y_n - avg)^2}{n-1}}$$

We define the root mean squared deviation (*RMSD*) of the data from the target by

$$RMSD = \sqrt{\frac{(y_1 - T)^2 + (y_2 - T)^2 + ... + (y_n - T)^2}{n}}$$

Note that *RMSD* is a measure of the average variation from the target. To relate *RMSD* to the two kinds of variation, we need the following result:

$$(y_1 - avg) + ... + (y_n - avg) = 0$$

That is, the sum of deviations of a set of numbers from their average is zero. We can show this result by noting that we can add the pieces before subtracting. That is,

$$(y_1 - avg) + ... + (y_n - avg) = y_1 + ... + y_n - (avg + ... + avg)$$

$$= n * avg - n * avg$$

$$\square \ \square \ 0$$

Now to decompose the *RMSD* into the two kinds of variation, we split each term in the numerator of the sum into three pieces.

$$(y_i - T)^2 = \left[(y_i - avg) + (avg - T) \right]^2$$

$$= (y_i - avg)^2 + 2(y_i - avg)(avg - T) + (avg - T)^2$$

We can add the three pieces separately. The sum of the first pieces is

$$\left(y_1 - avg \right)^2 + ... + \left(y_n - avg \right)^2$$

Note that in the second and third pieces $(avg - T)$ is a constant that is the same for all terms in the sum (i.e., it is the same for all i). The sum of the second pieces is

$$2 \left(avg - T \right) \left[\left(y_1 - avg \right) + ... + \left(y_n - avg \right) \right] = 0$$

The third piece is the sum of n constants and is $n(avg - T)^2$. Combining the pieces we have

$$\begin{aligned} RMSD &= \sqrt{\frac{\left(y_1 - avg \right)^2 + ... + \left(y_n - avg \right)^2}{n} + \left(avg - T \right)^2} \\ &= \sqrt{\frac{n-1}{n} stdev^2 + \left(avg - T \right)^2} \\ &\approx \sqrt{stdev^2 + \left(avg - T \right)^2} \end{aligned}$$

In words, the squared average deviation from the target is approximately the square root of the sum of squares of the measures of the two kinds of variation. The relative contributions of the two kinds of variation can be assessed from this formula. For example, if the *stdev* is large and the *avg* is close to target, we can make very little gain by moving the process average closer to the target.

If the target T is an ideal value for the output, then there is likely to be some cost to the supplier or loss to the customer when the actual output deviates from the target. The *RMSD* has the property that it is approximately linear in the deviation from target $\left| y_i - T \right|$ when y_i is relatively far from T (relatively here means compared to the other output values). Close to the target, the contribution of $\left| y_i - T \right|$ to the *RMSD* is small. In other words, *RMSD* is a reasonable surrogate for the cost associated with a deviation from target.

S2.7 VARIATION WITHIN GROUPS AND GROUP TO GROUP

In the camshaft lobe runout example, we showed how aligning the average angle error for each lobe could reduce the overall standard deviation. Using the angle error data as an example, we give a general formula to connect the overall standard deviation to the variation group to group and within groups. Here a group is defined by the lobe position.

As in the previous section, we need a bit of algebra to demonstrate the decomposition. Recall that there are 108 camshafts, each with 12 lobes. Let y_{ij} represent the angle error for lobe position j (= 1, 2, ..., 12), and camshaft i (= 1, 2, ..., 108). Then the overall standard deviation is

$$stdev = \sqrt{\frac{\left(y_{1,1} - avg \right)^2 + ... + \left(y_{108,12} - avg \right)^2}{108 \times 12 - 1}}$$

where *avg* is the average of the 1296 values. For any term in the sum, if avg_j is the average of the angle errors at lobe position *j*, we can write the squared deviation from the average as three pieces:

$$\left(y_{ij} - avg\right)^2 = \left[\left(y_{ij} - avg_j\right) + \left(avg_j - avg\right)\right]^2$$

$$= \left(y_{ij} - avg_j\right)^2 + 2\left(y_{ij} - avg_j\right)\left(avg_j - avg\right) + \left(avg_j - avg\right)^2$$

For each lobe position, we can add the pieces separately. Note that the sum of the middle piece is zero because it is a constant times the sum of deviations from the *j*th position average. Looking at the sum for position *j,* we have

$$\left(y_{1,j} - avg\right)^2 + ... + \left(y_{108,j} - avg\right)^2 = \left(y_{1,j} - avg_j\right)^2 + ... + \left(y_{108,j} - avg_j\right)^2 + 108(avg_j - avg)^2$$

In this equation, the first part of the sum on the right is directly related to the standard deviation of the output within position *j*, denoted $stdev_j$. That is,

$$\left(y_{1,j} - avg_j\right)^2 + ... + \left(y_{108,j} - avg_j\right)^2 = 107\, stdev_j^2$$

Hence, adding over all positions, we have

$$\left(y_{1,1} - avg\right)^2 + ... + \left(y_{108,12} - avg\right)^2$$

$$= 107(stdev_1^2 + ... + stdev_{12}^2) + 108\left[\left(avg_1 - avg\right)^2 + ... + \left(avg_{12} - avg\right)^2\right]$$

$$= 107\left(stdev_1^2 + ... + stdev_{12}^2\right) + 108 \times 11\, stdev(avg_1, ..., avg_{12})^2$$

Dividing both sides by 1295 (= 108 · 12 – 1) and taking the square root we get

$$stdev = \sqrt{\frac{107\left[stdev_1^2 + ... + stdev_{12}^2\right]}{1295} + \frac{108 \times 11\, stdev(avg_1, ..., avg_{12})^2}{1295}}$$

$$\approx \sqrt{avg(stdev_1^2, ..., stdev_{12}^2) + stdev(avg_1, ..., avg_{12})^2}$$

In words, we see that the overall standard deviation is determined approximately by two pieces. The first is the average squared within-position standard deviation. The second is the squared standard deviation of the within-position averages. One way to reduce the overall standard deviation is to ensure the second term is zero. That is, we can make sure that there are no differences in the averages for the 12 lobes.

In the camshaft lobe runout example, the average and standard deviation of angle error for each position are:

Lobe	*avg*	*stdev*
1	17.76	61.16
2	45.31	67.62
3	−19.37	57.62
4	19.69	63.49
5	−30.92	59.16
6	−43.14	60.65
7	−20.58	60.03
8	−30.43	59.22
9	−28.61	61.10
10	−73.96	56.83
11	−1.46	67.42
12	−89.85	63.53

If we could control the process so that the average angle error was constant for each position, then the overall standard deviation would be reduced from 71.5 to

$$\sqrt{avg(stdev_1^2,...,stdev_{12}^2)} = 61.6$$

In general, we can partition the overall standard deviation into variation within group and group to group. Aligning the group averages reduces the standard deviation by eliminating the variation due to group-to-group differences.

The division of the standard deviation into parts is the basis for the analysis of variance (ANOVA), a powerful numerical analysis method used in the book and discussed in Appendix D.

S2.8 GAUSSIAN MODEL

The Gaussian model is widely applicable. We describe some of its key properties. The Gaussian model describes the output of a process with a symmetric bell-shaped histogram. We repeat the form of the idealized curve in Figure S2.4.
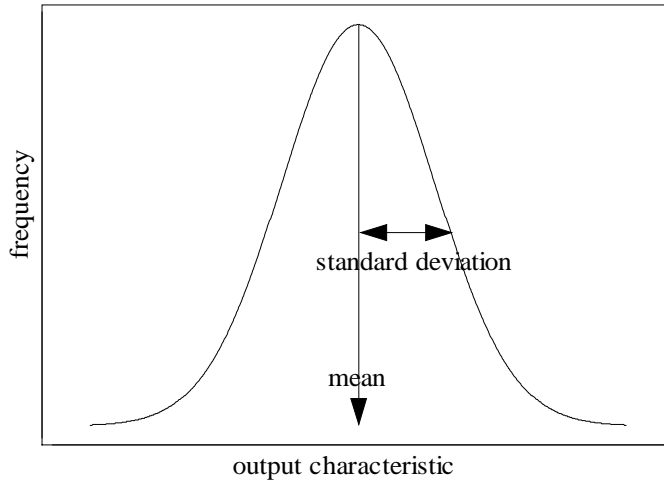
Figure S2.4   A Gaussian output model.

We prefer the name *Gaussian curve* to the more usual *Normal curve* because of the baggage that the word *normal* carries. We have been questioned many times by process engineers and production managers who look at the output from some statistical software that indicates that the process data is not "normal" and, hence, believe that something is wrong in the process. There is nothing abnormal or unusual about a process with a histogram of a different shape.

The area under the curve over any interval describes what proportion of the output values fall in that interval. We can specify the curve by two parameters, the mean (or center of the symmetric curve) and the standard deviation, a measure of the variation or spread of the values. The mean and standard deviation associated with the model are directly analogous to the *avg* and *stdev* of the histogram.

The Gaussian model has the beautiful property that it predicts that a fixed percentage of the output will fall within the interval (mean $\pm c$ standard deviation) for any constant c. For example, the percentages are approximately:

Within one standard deviation of the mean   68%

Within two standard deviations of the mean   95%

Within three standard deviations of the mean   almost 100%

There is a strong connection between the interpretation of capability measures such as P_{pk} and these percentages. If the process average is on target and we can describe the output by a Gaussian model, we can directly relate the magnitude of P_{pk} and the proportion of output that is out of specification:

P_{pk}	Parts per million out of specification
1.00	2700
1.33	64
1.67	0.6

These numbers should not be taken too seriously because a Gaussian model will never describe the process output perfectly.

In the book we rarely rely on formal statistical analysis. However, much of the formal analysis that we do present depends on the assumption that we can apply a Gaussian model. We often assume we can capture the effects of measured inputs in the mean of the model. The effects of all other varying inputs are lumped together in the standard deviation. For example, to formally analyze the position-to-position differences of the angle errors in the camshaft lobes, we assume that a Gaussian model can describe the output of the process where each position has a separate mean to explain the effect of position. We also assume that the variation within each position is the same since this variation captures the effect of all other varying inputs. Thus we assume that the standard deviation within each position is the same. We can describe the model pictorially using Figure S2.5 where we show the models for only three of the lobes to make it easier to interpret.
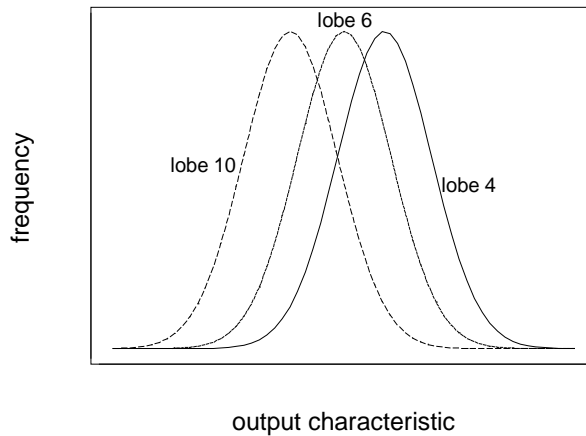
Figure S2.5   Graphical view of a model for lobe position differences.

Chapter 4 Supplement

An Algorithm for Reducing Variation

S4.1 COMPARISON OF VARIATION REDUCTION ALGORITHMS

There are many competing algorithms for solving chronic manufacturing problems and improving process performance. Some manufacturing organizations have invented or adapted an algorithm specialized to their own processes. Consulting companies have created their own versions, complete with acronyms and specialized language, driven partly by marketing considerations. We are perhaps also guilty of this sin. We have partially derived our algorithm from the one proposed by Shainin (1992, 1993).

Six Sigma (Harry and Schoeder, 2000) is a popular example. The algorithm is known by its acronym DMAIC (Define, Measure, Analyze, Improve, Control). The Define stage is sometimes omitted. The stages are divided into substages in a variety of ways that depend on which version you examine. See, for example, Breyfogle (1999).

The more detailed 12-step Breakthrough Cookbook associated with Six Sigma is described in Harry (1997, p. 21.19). A Six Sigma program is much more than the algorithm and the associated statistical methods and tools. It includes planning for implementation and ongoing management. For example, Six Sigma programs involve the training, certification, and deployment of specialized experts (Black Belts) in the use of the algorithm.

Scholtes (1988, pp. 5–19) provides a second example. The main stages of this algorithm are:

Understand the process $\longrightarrow$ Eliminate errors $\longrightarrow$ Remove slack $\longrightarrow$ Reduce variation $\longrightarrow$ Plan for continuous improvement
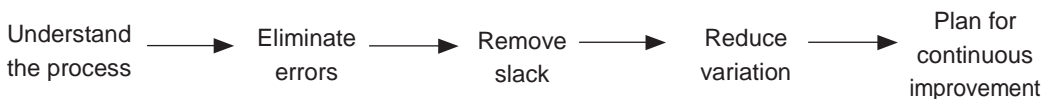
Each stage is further divided into a number of substages, used in completing the stage. Scholtes provides many other processes to support the algorithm, especially to help with the problem selection and the team operation.

Both the Six Sigma and Scholtes algorithms are designed to apply to processes in general and to many different kinds of problems. They are not specialized to reducing variation in high- to medium-volume manufacturing processes, as is Statistical Engineering.

In our view, these algorithms (including the one we propose) share two common elements:

- They are based on the diagnostic and remedial journeys described by Juran and Gryna (1980) and Juran (1988), as given in the figure that follows. The idea is that if we know the cause of the problem, we are more likely to find efficient and effective remedies or solutions.

Diagnostic Journey   Remedial Journey

Define the Problem → Find the Cause → Implement a Solution
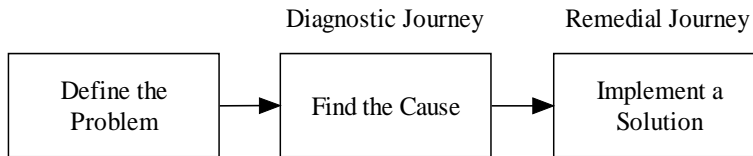
- They make heavy use of empirical (statistical) methods to increase knowledge about process behavior, to identify causes, and to validate remedies.

The differences among the algorithms lie in the detailed elaboration of the three boxes in the diagnostic and remedial journey. For example, we have explicitly included a stage in which we assess the measurement system for the output to ensure that it is not the dominant cause. The DMAIC and Scholtes algorithms include this assessment within a stage and, hence, give it a lower profile. We have chosen to highlight measurement system assessment because, in our experience, there are often serious problems with the measurement systems in high- to medium-volume manufacturing processes.

The explicit focus on a dominant cause of variation, as defined in Chapter 2, is another important differences between the Statistical Engineering algorithm and the Six Sigma and Scholtes algorithms. We also emphasize the use of the method of elimination and families of variation to help find the cause in an economical and timely manner. We give more detail on families of variation and the method of elimination in Chapter 9.

One unique feature of the Statistical Engineering algorithm is the idea that the team should consider potential remedies and select a working variation reduction approach before deciding if they need to find the dominant cause of the variation. The algorithm allows for improvement without knowledge of a dominant cause. We think that this consideration will increase the efficiency of the process improvement and lead to better remedies.

See De Mast (2003 and 2004), De Mast et al. (2000), and Logothetis (1990) for a detailed methodological comparison of the Shainin, Six Sigma, and Taguchi strategies for quality improvement. Here we do not compare or contrast Statistical Engineering with the Taguchi three-phase program. See, for instance, Ross (1988) and the modification given in Taylor (1991). A direct comparison is difficult, since the Taguchi program is focused on the design of a new product or process, while the aim of Statistical Engineering is improvement in an existing process. Note, however, that two of the variation reduction approaches (desensitization and robustness—see chapters 16 and 19) use ideas from Taguchi's program.

Because the algorithms discussed here have a common basis, we believe that the choice of algorithm is far less important to success than is the organizational discipline required for routine implementation. The tools and strategies from one algorithm can, and in fact should, be incorporated into another as appropriate.

Chapter 5 Supplement

Obtaining Process Knowledge Empirically

S5.1 ATTRIBUTES

In applying QPDAC to plan investigations, we specify many numerical attributes other than averages, standard deviations, and proportions. Here we give some examples that can apply to the target and study population or the corresponding sample.

Stratification

The first complication is stratification. For example, we may have a process with distinct streams, such as two suppliers of the same component, three production teams (one for each shift), four different parallel operations within a machining operation, and so on. In the truck alignment process described in Chapter 1, there are four gages operating in parallel to measure caster and other alignment characteristics (see Figure S5.1).

We can stratify the output by gage and define attributes such as the average and standard deviation for each gage. If the averages are different, we can reduce the overall standard deviation by improving the calibration process for the gages. If the standard deviation within one gage is much larger than within the others, we can look within that gage to understand why it is behaving differently. We can estimate these attributes by applying QPDAC.
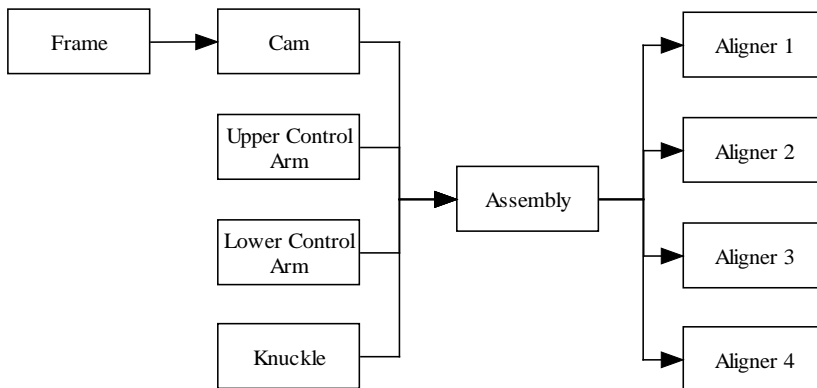
Figure S5.1   Truck alignment process.

We exploit the idea of stratification and specifying attributes by strata to generate clues about the dominant cause. See chapters 9 and 10.

Combined Attributes

Sometimes we express a question in terms of an attribute defined as a combination of two or more attributes. For example, in the supplement to Chapter 2, we defined the process performance measure P_{pk} as

$$P_{pk} = \frac{\text{distance from the process average to the closest specification limit}}{3*\text{stdev}}$$

P_{pk} is a function of the population average and standard deviation. Another example is the capability index C_{pk}, where

$$C_{pk} = \frac{\text{distance from the process average to the closest specification limit}}{3*\text{ within subgroup stdev}}$$

C_{pk} is a function of the average and a different attribute, the within-subgroup standard deviation.

To define within-subgroup standard deviation, we specify what we mean by a subgroup. We can choose as we please; often we define a subgroup as five consecutive parts from the process or all parts within a specified (short) period of time. In each case, we are *stratifying* the population of units into a large number of groups with no overlap. If all of the subgroups have the same size, then we define the within-subgroup standard deviation as

$$\text{within subgroup stdev} = \sqrt{average\left(stdev_1^2, stdev_2^2, ... \right)}$$

where $stdev_1, stdev_2,...$ are the standard deviations within subgroups 1, 2 ..., and so on. We average the squares and then take the square root to match the model behavior as discussed in Chapter 2.

We trust that you have noticed that P_{pk} and C_{pk} are different attributes because their denominators are not the same. Hence questions that we ask in terms of P_{pk} and C_{pk} are also different. We can see the connection between the denominators by noting the following result that we demonstrated in the supplement to Chapter 2.

$$stdev \approx \sqrt{(\text{within subgroup } stdev)^2 + stdev^2(\text{subgroup}}$$

Here we have stratified the population into subgroups. We have produced a very general formula, because the definition of a subgroup was up to us. We interpret the result by noting that the overall standard deviation is made up of a within-subgroup and a subgroup-to-subgroup component. In looking for a dominant cause of variation, we use this interpretation to eliminate from consideration causes associated with small components of the overall variation.

Note that the overall standard deviation and the within-subgroup standard deviation are almost equal when the standard deviation of the subgroup averages is small, that is, when all of the averages are about the same. In other words, the two attributes P_{pk} and C_{pk} will be almost the same when most of the variation in process output occurs within the subgroups, not subgroup to subgroup.

Scatter Plots and Fitted Lines

There is a set of useful numerical attributes associated with scatter plots. Suppose, in a population of units, we have an output characteristic *y* and an input characteristic *x* associated with each part. We do not expect *y* to be a simple function of *x* because there are many other inputs to the process. However, as shown in Figure S5.2, we often see a strong relationship in the scatter plot. In such cases, we can add a *fitted line* to describe the relationship numerically.

We fit the line using *least squares,* a procedure that minimizes the sum of squares of the vertical distances between the plotted points and the fitted line. Algebraically, if $y = a + bx$ represents any straight line, then we select the fitted line to minimize the quantity

$$\sum \left(y_i - a - bx_i \right)^2$$

where the sum is overall units in the population. We denote the fitted line in the population by the equation

$$y = \alpha + \beta x$$

The Greek letters α (alpha) and β (beta) are numerical attributes of the population. The slope β describes how much, on average, *y* will increase for a unit increase in *x*. Given a sample of parts from the study population, we estimate these attributes with the corresponding sample quantities. See Appendix E for more details on how to fit the line and create plots like Figure S5.2 for the data in the sample.
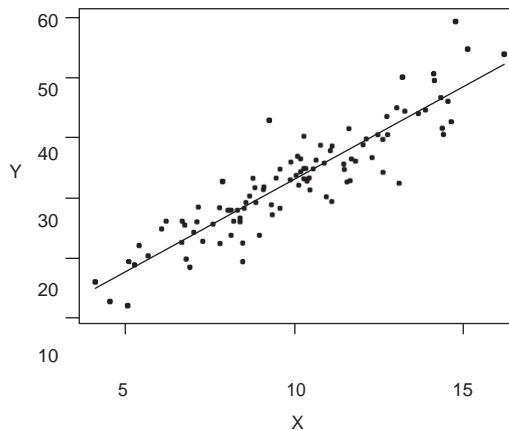
Figure S5.2   Scatter plot and fitted line.

Looking at Figure S5.2, we see that if x were held constant, the variation in y would be much smaller than when x is allowed to vary. For example, if x is held near 10, then from the scatterplot, we see that y varies from about 25 to 40. Overall, y varies from about 10 to 60. We use this idea to quantify the contribution of a single cause to the overall variation.

We can partition the standard deviation of y into two components:

$$stdev(y) \square \sqrt{\beta^2 stdev(x)^2 + stdev(rest)^2}$$

The second component under the square root sign describes the variation that would remain if x were held fixed. We can picture this attribute on the scatterplot as the variation around the fitted line. The first component describes the contribution of the input x to the variation in the output y. If x is a dominant cause of variation in the output, $stdev(rest)$ is small compared to $stdev(y)$.

Given a sample from the population, we can estimate $stdev(y)$ and the component contributions using MINITAB. Then we can decide if x is a major contributor to the variation in y.

S5.2 SAMPLING PROTOCOLS AND THEIR EFFECTS

In selecting the sample, the goal is to control:

- Sample error by matching the attributes of interest in the sample and the study population

- Sampling cost

There are some general considerations. First, to avoid sample error, we make sure that the sampling protocol will select units broadly across the study population. For example, if we define the study population to be one week's production, we should not entertain a sampling protocol that is limited to a single day. Second, if we choose a complex sampling scheme, we are likely to have added cost and a greater opportunity for the protocol to go wrong in the Data step. Third, we will generally incur a larger cost for larger samples, everything else being equal.

We make wide use of systematic sampling in which we deliberately select the sample over time and location. For example, we may select the next 300 units to assess process capability or we may select five consecutive pieces at the start of every two hours of production in a multivari investigation (see Chapter 11). Similarly, we may ensure that the sample has an equal number of parts from all process streams when assessing process performance. In assessing a measurement system, we select the parts to be measured repeatedly to cover the range of normal output values. We specifically sample both large and small parts. Systematic sampling has the major advantage that it is easy to organize and execute.

Random sampling is the most famous sampling protocol but has relatively few uses in process improvement. It is difficult and costly to implement. Suppose we want a sample of 100 parts selected at random from the study population, the week's production of 10,000 parts. We suppose that the units all have a unique identification code; if they do not, then we must assign such a code, which can be a daunting task. The first step is to associate each of the identification codes with a digit between 1 and 10,000. Then, using MINITAB, we generate a random sample of 100 values from the possible 10,000 numbers without replacement.

This is the easy part. Now we must find the parts corresponding to the sampled identification codes. This can be difficult if the process is complex and parts do not appear in the expected order. Finding the sampled units can take a long time and be expensive.

We once consulted on a project (by telephone only) where we naively recommended selecting a random sample of 50 water heaters from a lot of 2200 heaters. Each heater had a unique serial number. The heaters were stored in boxes in a warehouse. We provided a list of 50 serial numbers corresponding to a random sample of heaters. The sampling protocol fell apart immediately because it was not possible to locate the serial number without opening the box, and furthermore, it was very difficult to get at certain heaters due to the method of storage. In the end, a convenience sample was selected from locations spread throughout the warehouse. Note that when we discuss experimental plans in more detail, we will use random assignment extensively. This is a different use of randomization not to be confused with random sampling.

Sample size is a major issue. The most frequent question asked of a statistical consultant is, "How large a sample should I use?" The answer, like that to all statistical questions, is, "It depends!" For a given protocol, the larger the sample, the greater the chance of small sample error. However, even with a very large sample, we may be unlucky and have substantial sample error. To determine sample size, the first consideration is how the sampling fits into the rest of QPDAC. For instance, if there are likely to be large study and measurement errors, it makes little sense to try to get very small sample error. Second, there are cost and time constraints that often outweigh any consideration of sample error. We can often answer the sample size question based on what the team can afford.

There are some formal procedures that can be used to determine sample sizes if we have random sampling (or another sampling protocol where we pretend that the sampling will be random) and simple attributes in the Question step. We do not discuss these formally here. See Odeh and Fox (1975), Nelson (1985), and Neter et al. (1996). Also MINITAB has a "power and sample size" function. To understand how these methods work, consider the following example.

We were once asked how large a sample would be needed to determine if there were any defective rails in a suspect lot of 5700 rails. The defect was a vertical crack in the head of the rail that might cause derailment of a train. To detect the defect, the railway used a destructive test. The management wanted to be 99% confident that there were no defective rails in the lot. To answer the question, from a statistical perspective, the situation requiring the largest sample size would occur if there were exactly one defective rail in the lot. Suppose this was the case, that is, that there was one defective in the lot and we plan to choose a sample of n rails at random. We want to be 99% sure that we find the defective rail. Simple calculations show that we need to sample 99% of the rails. This was completely infeasible; sampling could not provide the information required in this case. The only feasible solutions were to scrap the lot of rails or to accept the risk of an accident.

In many QPDAC applications, we will use data that are already available; that is, data that have been collected for some other purpose. For example, we may examine scrap records or measurements taken as part of the process control plan. These data can be helpful but it is wise to think about how they were collected. Are we likely to be misled because of large study or sample error?

In summary, we most often recommend systematic sampling to cover the study population. Systematic sampling is relatively easy to implement. Choose a sample size as large as you can afford remembering that there are likely to be more investigations before the problem solving is complete.

S5.3 STUDY, SAMPLE, AND MEASUREMENT ERRORS

To help understand Figure 5.1 more fully, Deming (1986) distinguishes between descriptive, enumerative, and analytic investigations (see also Liberatore, 2001). In a descriptive investigation, the target population, the study population, and the sample are all the same. The only potential error is due to the measurement system. In an enumerative investigation, the target and study populations are the same. In this case, study error is not a concern. In an analytic investigation, the target population, the study population, and the sample are all different, so we must consider all types of error.

Deming used these distinctions to point out that investigations of the different types are subject to different types of uncertainty due to the possibility of the various types of errors. In our context, we are most interested in analytic investigations since the target population usually extends into the future. In analytic investigations, we must face concerns about possible study error and the trade-off with time and cost considerations.

For chronic problems involving production processes, the study process is often a time-limited version of the target process. However, in some circumstances, the study population may be generated from a different process altogether. For instance, we may draw conclusions about a proposed production process from empirical investigations of a pilot process. This difference increases the potential for study error since the pilot process may not be representative (in terms of the attribute of interest) of the production process.

We can use formal statistical methods (see Box et al., 1978) to quantify sample and measurement error. For a familiar example, think about the statement at the bottom of most public opinion polls:

A survey of this size is accurate to within 3 percentage points 19 times out of 20.

This is a statement about the likely size of sample and measurement error. For analytic investigations, Deming's point is that such a statement does not apply to study error. Our message is that we can control study error only by careful specification of the study and target populations.

S5.4 OUTLIERS

An *outlier* is an unexpectedly large or small value. The occurrence of outliers in any empirical investigation can have a large effect on the conclusions. We try to identify outliers and control their impact.

An outlier can have a large influence on the estimate of the standard deviation, one of the attributes we use most often. We use an artificial example to illustrate. Suppose we take a sample of 100 parts from a process and measure an output characteristic (denoted y). We present the data summary from MINITAB and the process histogram in Figure S5.3.
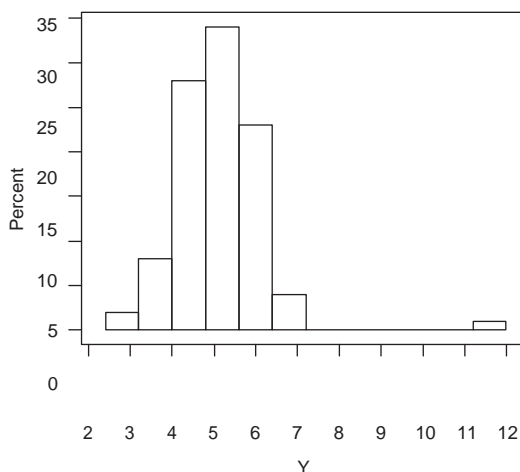
Figure S5.3   Histogram of 100 output values.

Descriptive Statistics: y

Variable	N	Mean	Median	TrMean	StDev	SE Mean
y	100	5.081	5.000	5.049	1.059	0.106

Variable	Minimum	Maximum	Q1	Q3
y	2.500	11.400	4.425	5.600

The first value in the data set is 11.4, an apparent outlier. We can study the effect of the outlier on the sample standard deviation by replacing 11.4 by a number of values between 1 and 20 (denoted x) and then recalculating the sample attribute. We plot the results in Figure S5.4.
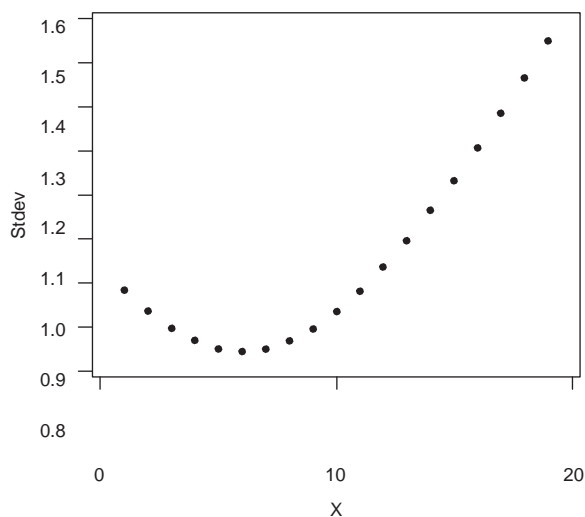
Figure S5.4   Effect of changing a single value on the sample standard deviation.

We see that the standard deviation is very sensitive to changes in a single value in a sample of size 100. If the first value in the sample had been 5 (close to the average) rather than 11.4, the standard deviation would have been reduced from 1.06 to 0.85, a 20% reduction.

What should we do if we find an outlier? We first look at the data collection procedure to ensure that no blunder occurred, such as incorrectly entering the data into the spreadsheet. If a blunder occurred and it cannot be easily corrected, we have a number of options. We can remove the observation from the analysis. This is the preferred approach for investigations such as estimating the problem baseline (see Chapter 6) where there are many observations under similar conditions. For other investigations such as designed experiments, where we make relatively few measurements under a number of combinations of inputs, we can repeat the observation or replace the outlier by an estimate of its value made from the remaining observations.

If we cannot attribute the outlier to a blunder, we have a much more difficult decision. First, if eliminating or changing the outlier has no effect on conclusions then we may ignore its presence. If the outlier does have an effect, the key question is whether the dominant cause of the outlier is the same as the cause of the problem we are trying to address. This question is unfortunately not answerable without considerable process knowledge. It is possible that the outlier comes from a different, perhaps rare, failure mechanism that we are not trying to address. On the other hand, the outlier could be an extreme example of the action of the dominant cause of the current problem. In the former case, the presence of the outlier will make finding the cause and solving the problem more difficult, since the outlier will cloud the effects we are looking for. In the latter case, the outlier may provide a great deal of useful process knowledge. Consideration of outliers is especially critical when using the idea of leverage, as defined in Chapter 9.

Chapter 6 Supplement

Defining a Focused Problem

S6.1 CONFIDENCE INTERVALS FOR THE PROCESS BASELINE ATTRIBUTES

We can use *confidence intervals* to assess uncertainty due to sampling and measurement error. A confidence interval is a range of plausible values for an attribute in the study population. The interval is based on the estimate of the attribute from the measured values in the sample. By default, we use 95% confidence intervals so that we are confident that the unknown attribute falls within the given interval.

For example, consider the V6 piston diameter data discussed in Chapter 5. The team collected diameter data for 469 pistons over a one-week period. The data are found in the file *V6 piston diameter baseline*. The attribute of interest is the standard deviation in the target population. Using MINITAB we produced the following numerical summaries:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
diameter	469	590.85	590.90	590.83	3.32	0.15

Variable	Minimum	Maximum	Q1	Q3
diameter	581.50	602.80	588.40	593.05

The estimated standard deviation is 3.32 microns. We can get an approximate 95% confidence interval for the study population standard deviation of the form

$$(c_1* \text{ estimated standard deviation}, c_2* \text{ estimated standard deviation})$$

where c_1 and c_2 are found in Table S6.1. The given constants can be derived using a Gaussian assumption (as discussed in Chapter 2) to model the distribution of output values. The degrees of freedom are related to the sample size. Technically, in the previous calculation, the degrees of freedom are one less than the sample size. In general, for large sample sizes (greater than 30), you can replace the degrees of freedom by the sample size.

In the V6 piston diameter example, the confidence interval for the study population standard deviation is (3.08, 3.59) microns. We can be confident that the standard deviation in the study population is in this range. Even though we used a large sample size, we can only be confident that the attribute is within about ±7% of the estimate. In general, from Table S6.1 we note that the process standard deviation is poorly estimated when the sample size is small.

Table S6.1   Constants for confidence intervals for standard deviation.

Degrees of freedom	c_1	c_2
10	0.69	1.75
20	0.77	1.44
30	0.80	1.34
40	0.82	1.28
50	0.84	1.24
60	0.85	1.22
80	0.87	1.18
100	0.88	1.16
200	0.91	1.11
300	0.93	1.09
500	0.94	1.07
1000	0.96	1.05

We use the estimated standard deviation as a measure of baseline performance. In the example, if we change the process and re-estimate the standard deviation, we hope to see an estimate substantially less than 3.08 (the lower endpoint of the confidence interval) in order to be confident that the change has produced a positive benefit.

We can use the confidence interval to assess the likely size of sampling and measurement errors. The range of values for the attribute captures these uncertainties. However, the confidence interval tells us nothing about study error.

Sometimes we may want a confidence interval for an average in the study population. In the V6 piston diameter example, MINITAB has done most of the work. The estimated attribute is the sample average 590.85. To derive a confidence interval for the average, we look for the SE (denotes standard error) of the mean. In the example, the standard error of the estimate for the average (denoted "SE Mean" in the MINITAB results) is 0.15. For large sample sizes (sample sizes greater than 30), the 95% confidence interval is (approximately)

$$\text{sample average} \pm 2* \text{ (standard error of the average)}$$

or, in this case, 590.85 ± 0.30. In other words, we can be confident that the study population average diameter falls in the range (590.55, 591.15) microns.

For many other attributes, the form of the confidence interval is the same as for the average, that is,

$$\text{estimated attribute} \pm 2* \text{ (standard error of the estimate)}$$

We can often find both the estimated attribute and the standard error of the estimate in the MINITAB summary.

For estimating a small proportion, such as the proportion defective, we get a rough idea of the precision of the estimate from a confidence interval of the form

$$\text{estimated proportion} \ \pm \ 2\sqrt{\left(\text{estimated proportion}\right)/\left(\text{sample size}\right)}$$

This formula is rough because we do not expect the underlying assumptions to hold. However, we use the formula to give some guidance about sample size. For example, if we expect the baseline defect rate to be around 0.01, then with a sample size of 1000, the confidence interval for the proportion defective is about 0.010 ± 0.006. We see there is large relative error in the estimate due to the sampling error.

We use confidence intervals when it seems necessary to quantify the uncertainty in estimated attributes. We often ignore this uncertainty, especially in the searching for the dominant cause. There are many reference books that deal with confidence intervals and other formal statistical procedures. See, for example, Box, Hunter, and Hunter (1978).

S6.2 PROCESS STABILITY AND QUANTIFYING BASELINE PERFORMANCE

There is a myth that we cannot improve process performance with Statistical Engineering methods until we have established a stable process in the sense of Statistical Process Control (SPC). We refer to this property as stability. For example, see section 6 of the AIAG QS-9000 SPC Manual (1995b).

It is true (see the discussion in Chapter 4) that we need a controlled process before it makes sense to apply Statistical Engineering. The main reason is that huge improvements can be made cheaply by ensuring that standard practices and a control plan are followed, that routine maintenance and housekeeping are carried out, that process personnel are trained, and so forth. We do not need to use the sledgehammer of Statistical Engineering to achieve the large reduction in variation that comes from fixing obvious problems.

What role does stability play in quantifying baseline performance? Consider, for example, the truck pull variation reduction problem discussed in Chapter 6. The study population was a two-month period and the sample was all trucks produced in that period. Here there

is no sample error. The baseline performance measure was the pull standard deviation calculated over the two-month period. We assumed that this two-month period was long enough so that the standard deviation over this study population would describe how the process would perform into the future if no changes were made. In the language of QPDAC, we assumed that there was little study error.

For the purpose of discussion, suppose that a control chart had been in place at the four gages at the start of the project. There are many sampling protocols that could be used to generate control charts. The process may be stable using some of these protocols and unstable using others. In other words, stability is not solely a property of the process; it also depends on the sampling and charting procedure used. Suppose that in our example, the chart was based on pull measured on five consecutive trucks, once per shift. In Figure S6.1, we show the $\bar{X}$ chart based on the first five trucks produced after 2:00 A.M., 10:00 A.M., and 4:00 P.M. sampled from the data set *truck pull baseline* for all days.

The process is unstable with respect to this charting procedure. That is, there are causes (varying inputs) that change from shift to shift *within* the two-month period that produce the systematic patterns seen on the chart. The effects of these causes contribute to the overall standard deviation. Our assumption about study error is that in the future, the control chart would show similar patterns of instability. This is an assumption and the control chart here provides no help in deciding if the assumption is reasonable or not. If we were concerned with the assumption and if the data were available, we could examine a sequence of two-month periods in the past to see if such a time frame captures most of the process variation.

In summary, the selection of a study population with a time frame long enough to well represent the future is a key issue in establishing a baseline. Stability within the selected study population is not the issue. We emphasize the word *within* because the baseline standard deviation describes the whole two-month period.

Whether the process is stable or not, we may use control chart data to quantify the baseline and provide clues about the dominant cause.
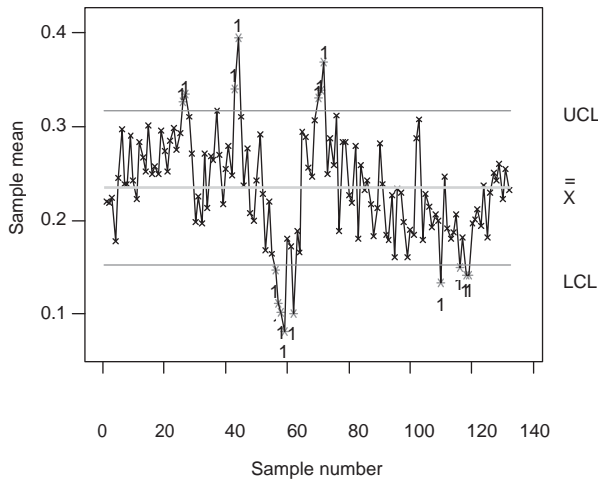
Figure S6.1 $\bar{X}$ chart for truck pull.

Chapter 7 Supplement

Checking the Measurement System

S7.1 ASSESSING A BINARY MEASUREMENT SYSTEM

With a binary measurement system, there are two outcomes we call *pass* and *fail*. We suppose here that the characteristic being measured is also binary, either good or bad. Many binary measurement systems first measure one or more continuous characteristics and then classify the part as pass or fail. If we can measure these underlying continuous characteristics, we can assess the properties of the measurement system (before discretization) using the plans proposed in Chapter 7.

Problems with a binary output may be specified in terms of excess scrap and rework or in terms of customer complaints about receiving bad parts. In the first case, the baseline performance measure is *P*(*pass*), the proportion of parts passed by the measurement system. In the second case, we use *P*(*good*|*pass*), the proportion of parts sent to the customer that are good. Note the | in the proportion indicates a conditional proportion—in this case, the proportion of good parts among all those passed by the measurement system. We can write this conditional proportion as

$$P\big(good\,\big|\,pass\big) = \frac{P\big(good \text{ and pass}\big)}{P\big(pass\big)} \tag{S7.1}$$

From the baseline investigation we have a good estimate of *P*(*pass*). This is analogous to obtaining an estimate of *stdev*(*total*) from the baseline in a problem with a continuous output. The key formula relating the properties of the measurement system and the manufacturing process is

$$P(pass) = P(pass\,|\,good)P(good) + P(pass\,|\,bad)P(bad) \tag{S7.2}$$

The proportions *P*(*good*) and *P*(*bad*) = 1 − *P*(*good*) are properties of the manufacturing process. The misclassification rates *P*(*fail*|*good*) = 1 − *P*(*pass*|*good*) and *P*(*pass*|*bad*) are properties of the measurement system. For a binary output, Equation (S7.2) corresponds to

the partition of the overall standard deviation into components due to the measurement system and the manufacturing process for a continuous output.

We can derive Equation (S7.2) by noting that

$$P\left(pass \middle| good\right) = \frac{P\left(good \text{ and } pass\right)}{P\left(good\right)} \text{ or } P\left(good \text{ \& } pass\right) = P\left(pass \middle| good\right)P\left(good\right)$$

and

$$P\left(pass \middle| bad\right) = \frac{P\left(bad \text{ and } pass\right)}{P\left(bad\right)} \text{ or } P\left(bad \text{ \& } pass\right) = P\left(pass \middle| bad\right)P\left(bad\right)$$

We finally note that the proportion of passed parts is

$$P(pass) = P(pass \text{ and } good) + P(pass \text{ and } bad)$$

$$P(pass) = P(pass \mid good)P(good) + P(pass \mid bad)P(bad)$$

If instead the baseline performance is specified by $P(good \mid pass)$, we can rewrite Equation (S7.1) as

$$P\left(good \middle| pass\right) = \frac{P\left(pass \middle| good\right)P\left(good\right)}{P\left(pass\right)}$$

to separate the effects of the measurement system from those of the rest of the process.

In summary, to assess the binary measurement system we estimate the misclassification rates $P(pass \mid bad)$ and $P(fail \mid good)$, attributes of the measurement system in the population of all future measurements. If these proportions are too high, then the measurement system is not adequate and must be improved before proceeding to the next stage of the Statistical Engineering algorithm.

We illustrate these ideas using an example in which credit cards were classified *pass* or *fail* by an automated visual inspection system. The aim was to find visual defects such as missing parts of the intended design, surface scratches, bleeding of the colors, fuzzy letters and numbers, and so on. The measurement system used a digital image of the front of each card to calculate hundreds of summary measures based on comparing the picture to a template of the ideal card. If none of the summary measures fell outside the prespecified ranges, the measurement system passed the card. In the baseline investigation, the team monitored the measurement system for more than one week and found an 89.5% pass rate. About 30,000 cards were measured each hour. The goal was to reduce the proportion of bad cards being shipped to the customer $P(pass \mid bad)$ without substantially decreasing the pass rate $P(pass)$.

Estimating the Misclassification Rates

To estimate the misclassification rates, we must determine whether a number of parts are good or bad. This corresponds directly to the need to know true values in the determination of bias in a continuous measurement system. We illustrate two different plans to estimate the misclassification rates using the credit card example.

Plan 1

We start with equal-sized groups of passed and failed parts. We select each group over a period of time and over a range of conditions so that they represent the long-run populations of passed and failed parts. Then we determine the true state (good or bad) of each part.

In the credit card example, the team selected 40 passed and failed cards per day for five days. Then human inspectors classified each card as good or bad. The data are summarized in the following table.

	Good	Bad	Total
Pass	195	5	200
Fail	16	184	200

From these data, we estimate $P(good \mid pass) = 195 / 200 = 0.975$ and $P(good \mid fail) = 16 / 200 = 0.080$.

With the a derivation used to obtain Equation (S7.2), we can express the proportion of good parts as

$$P(good) = P(good \mid pass)P(pass) + P(good \mid fail)P(fail)$$

Substituting the estimates for $P(good \mid pass)$ and $P(good \mid fail)$ and the overall pass rate 0.895, we estimate the proportion of good cards in the population as

$$P(good) = 0.975 \cdot 0.895 + 0.080 \cdot 0.105 = 0.881$$

and the first misclassification rate by

$$P(fail \mid good) = 1 - P(pass \mid good) = 1 - \frac{P(good \mid pass)P(pass)}{P(good)} = 1 - \frac{0.975 \cdot 0.895}{0.881} = 0.010$$

Similarly, the estimate of the second misclassification rate is

$$P\left(pass \mid bad\right) = \frac{P\left(bad \mid pass\right)P\left(pass\right)}{P\left(bad\right)} = \frac{0.25 * 0.895}{1 - 0.881} = 0.188$$

The team decided that the measurement system needed improvement because the estimate for $P(pass | bad)$ was large. If the team could reduce this misclassification rate, the proportion of bad cards shipped to the customer would decrease from 0.025.

Plan 2

Here we start with equal-sized groups of good and bad parts. We select each group over a period of time and over a range of conditions so that they represent the long-run populations of good and bad parts. Using the credit card example, the operators examined cards until they had 200 good cards and 200 bad cards. Next, the cards were classified by the measurement system with the following results.

	Pass	Fail	Total
Good	190	10	200
Bad	8	192	200

From these data, we can directly estimate the misclassification rates as $P(fail | good) = 10/200 = 0.05$ and $P(pass | bad) = 8/200 = 0.04$. Since we also know $P(pass) = 0.895$, we can estimate $P(good)$ using Equation (S7.2):

$$P(pass) = P(pass | good)P(good) + P(pass | bad)P(bad)$$

Substituting the estimates, we have $0.895 = 0.95P(good) + 0.04(1 - P(good))$, and solving, we find

$$P(good) = \frac{0.895 - 0.04}{0.95 - 0.04} = 0.940$$

Comments

The second plan gives direct estimates of the misclassification rates and an indirect estimate of $P(bad | pass)$, whereas the first plan provides indirect estimates of the misclassification rates and direct estimates of $P(bad | pass)$. Which plan is preferred depends on costs and the way the baseline is specified. The second plan will not be feasible in many cases because it will be difficult to get representative samples of good and bad parts.

If we decide that the misclassification rates are too high, we can investigate families of causes such as operator-to-operator or time-to-time by repeating the investigation over different time periods or with different operators.

The estimates produced by these plans will be imprecise unless we have large samples. In the first plan using the credit card example, we can be confident that the estimate $P(pass|bad) = 0.188$ is within about ±0.06 of the actual misclassification rate. To increase the precision of the estimate, we need to determine the characteristic (good or bad) of more than 200 passing and failing cards.

See Boyles (2001) and AIAG (1995a) for more details on alternative ways of assessing variation due to binary measurement systems. The AIAG method is not feasible in the credit card example because it assumes a measurable continuous output has been discretized to give the binary output.

S7.2 ASSESSING A DESTRUCTIVE MEASUREMENT SYSTEM

Here we discuss the planning and analysis of an investigation to assess a destructive measurement system for a continuous characteristic. A measurement system is destructive if we change the true value of the characteristic by making the measurement. Tensile strength is a good example. We cannot make repeated measurements on the same part when the system is destructive.

To assess measurement variation, we select parts that we hope have identical or very similar characteristic values. If we have many similar units, we recommend the plan and analysis in sections 7.2 and 7.3. If there are only pairs of similar units, we can use an investigation like an Isoplot (see Section S7.4).

The following example of assessing a destructive measurement system for the tensile strength of tin-plated steel is motivated by Spiers (1989). In the baseline investigation, the team found that the overall variation in strength was 2.5 KSI (thousands of pounds per square inch) relative to the tolerance ±6.0 KSI measured from nominal.

The next step was to check the measurement system. To determine tensile strength, the operator:

1. Cut a standard sized sample from a sheet of steel

2. Calculated the cross sectional area of the sample with a micrometer

3. Pulled the sample apart in a tensometer

The tensile strength is the ratio of the maximum load to the cross-sectional area.

In the measurement investigation, 30 standard-sized pieces were cut from three different sheets of steel, parallel to the rolling direction and one-quarter distance from the edge of the sheet to eliminate variation in thickness and tensile strength due to edge or crown effects. The sheets were chosen to cover the full extent of variation in strength seen in the baseline investigation. The 30 pieces from each sheet were then randomly divided into three groups of 10. The three appraisers each measured the tensile strength of all 30 pieces (10 pieces from each sheet). We give the data in the file *tin plate strength measurement*.

In the analysis, we assume that all the pieces cut from the same sheet have the same tensile strength. This is analogous to measuring the same part a number of times in the
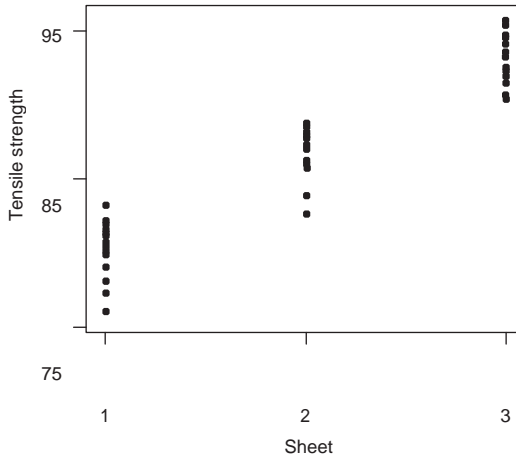
Figure S7.1 Plot of tensile strength by sheet.

nondestructive measurement case. Figure S7.1 shows the measured tensile strength versus the sheet number. A numerical summary is:

Descriptive Statistics: tensile strength by sheet

Variable	sheet	N	Mean	Median	TrMean	StDev
tensile	1	30	80.620	80.900	80.762	1.500
	2	30	87.177	87.450	87.354	1.413
	3	30	93.477	93.550	93.531	1.406

Variable	sheet	SE Mean	Minimum	Maximum	Q1	Q3
tensile	1	0.274	76.100	83.200	80.000	81.600
	2	0.258	82.600	88.800	86.300	88.300
	3	0.257	90.600	95.800	92.425	94.700

From the MINITAB results, we estimate *stdev*(due to measurement) as

$$\sqrt{\left(1.5^2 + 1.413^2 + 1.406^2\right)/3} = 1.44$$

Using the estimate *stdev* (total) of 2.50, obtained from the baseline investigation and Equation (7.2), we estimate

$$stdev(\text{process}) \square \sqrt{stdev(\text{total})^2 + stdev(\text{due to measurement})^2} \square \square 2.04$$

so the estimated discrimination ratio *D* is 1.42.

We need to be careful in the interpretation of *D,* because of the destructive nature of the measurement system. The term *stdev*(due to measurement) includes two components, one due to the measurement variation and one from the variation in true values of the samples from each sheet. If the ratio is large—that is, *D* > 3—we can conclude that the measurement system is acceptable because the combined effect of the two components is small. If the
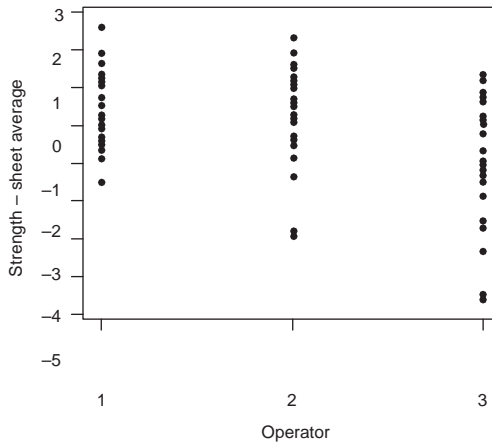
Figure S7.2 Tensile strength minus sheet average by operator.

ratio is small—that is, $D < 3$—we cannot be sure if the fault lies in the measurement system or if there is large variation within the samples taken from the same sheet.

In the example, the team concluded that they should improve the measurement system before continuing further with the Statistical Engineering algorithm. In further analysis of the measurement results, as shown in Figure S7.2, the team found that some of the measurement variation was due to the differences between the operators, with operator three exhibiting more variation than the other two operators.

To estimate measurement bias with a destructive measurement system, it is necessary to use parts with known characteristic value. These standard parts are difficult to obtain unless there is some other (better) nondestructive measurement system.

S7.3 REPEATABILITY, REPRODUCIBILITY, LINEARITY, AND STABILITY

For much of the automotive industry, as part of QS-9000 (Automotive Industry Action Group [AIAG], 1998), suppliers are required to periodically assess measurement systems used for process control and product inspection. The suppliers routinely consider *repeatability* and *reproducibility.* Less frequently, they examine *linearity* and *stability.* Here we define these attributes of the measurement system and relate them to measurement variation and bias as described in Section 7.1. Other than stability (which we incorporate into our definition of measurement variation and bias), we feel these additional attributes are only useful if we need to improve a measurement system.

From the Automotive Industry Action Group (AIAG) *Measurement Systems Analysis* manual (1995, pp. 17–18), we quote the following definitions:

Repeatability: the variation in measurements obtained with one measurement instrument when used several times by one appraiser while measuring the identical characteristic on the same part.

Reproducibility: the variation in the average of measurements made by different appraisers using the same measuring instruments when measuring identical characteristics on the same part.

Stability: the total variation in the measurements obtained with a measurement system on the same master or part when measuring a single characteristic over an extended time period.

Linearity: the difference in bias values through the expected operating range of the gage.

There is some ambiguity in these definitions that is evident if we try to create the corresponding attributes. The target population is all measurements to be made in the future with the measurement system over a wide range of parts, time, operators, and environmental conditions.

The output is the measurement error for each act of measuring. We assume that the measurement system has a single gage.

Repeatability is the variation in the measurement errors when a single operator measures the same part. The definition makes no mention of which operator, which part, over what time frame, and under what conditions. One way to make the definition more precise is to consider repeatability a short-term measure of the variation in the measurement errors when the operator, part, and conditions are held fixed. That is, we think of repeatability as the average over time, operators, parts, and conditions of the standard deviation of the measurement errors in the consecutive measurement of the same part.

Reproducibility is variation from operator to operator. We can specify reproducibility more precisely as the variation of bias from operator to operator in the system. Bias for a particular operator is the average measurement error for that operator over a wide range of times, parts, and environmental conditions. We think of reproducibility as the standard deviation of the operator biases.

Stability as defined by AIAG is the variation in the measurement errors if the same part is measured repeatedly over a long period. To make this notion precise, we have to average these standard deviations over different parts, operators, and environmental conditions. Stability is to the long term what repeatability is to the short term. Stability is a badly chosen name, since it is not connected to the Statistical Process Control (SPC) use of the word as in a stable process.

Linearity is a measure of the variation in bias over part size. More precisely, for each part, we define the bias as the average measurement error over operators, time, and environmental conditions. Linearity is the standard deviation over all parts of these biases. It is not clear why the definition of linearity does not include possible changes in measurement variation over part size.

The four attributes are a strange collection. Taken together, we cannot use them to estimate the overall variation in the measurement system as defined in Chapter 7. Each attribute is one component of the overall variation, but there are many other components. If there is no time, part, or environmental component to the overall variation of the measurement errors, then we can combine repeatability and reproducibility to estimate the overall variation of the system.

We can assess these attributes (except linearity as defined earlier in terms of bias) using the plan for assessing the overall variation described in Section 7.2. Consider the camshaft example in which there were three parts, three operators, and two weeks. On one day in the week, each operator measured every part three times. We assess repeatability by the average variation within each part, operator, and week and reproducibility by the variation in the operator averages over all parts, weeks, and repeated measurements. We assess stability graphically by comparing the within-week variation over the two weeks.

To estimate repeatability we use a one-way ANOVA model. See Appendix D for further discussion of ANOVA models. The relevant MINITAB ANOVA results are:

One-way ANOVA: diameter versus repeat

Analysis of Variance for diameter

Source	DF	SS	MS	F	P
repeat	17	5024.511	295.559	841.83	0.000
Error	36	12.639	0.351		
Total	53	5037.151			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	------+---------+---------+---------+
1	3	–11.658	0.449	*)
2	3	1.289	0.150	(*
3	3	11.940	1.020	(*)
4	3	–10.886	0.561	*)
5	3	0.024	0.242	(*)
6	3	12.581	0.792	(*)
7	3	–10.778	0.876	(*
8	3	1.533	0.872	(*)
9	3	13.058	0.141	(*)
10	3	–10.566	0.547	(*)
11	3	1.124	1.156	*)
12	3	12.494	0.245	(*
13	3	–11.219	0.251	(*)
14	3	1.358	0.410	(*)
15	3	12.196	0.571	(*)
16	3	–10.300	0.274	(*)
17	3	1.895	0.150	*)
18	3	13.843	0.397	(*)
				------+---------+---------+---------+
Pooled StDev =		**0.593**		-8.0 0.0 8.0 16.0

We estimate the repeatability as 0.593. Since the estimate of the overall measurement standard deviation was found earlier to be 0.756, the repeatability is a large component.

We give the operator averages in the following MINITAB summary:

Descriptive Statistics: diameter by operator

Variable	operator	N	Mean	Median	TrMean	StDev
diameter	1	18	0.76	1.20	0.81	9.84
	2	18	0.67	0.65	0.64	9.86
	3	18	1.55	1.80	1.58	10.10

Variable	operator	SE Mean	Minimum	Maximum	Q1	Q3
diameter	1	2.32	–12.20	12.90	–10.80	12.05
	2	2.32	–11.50	13.30	–10.95	11.85
	3	2.38	–11.50	14.10	–10.07	13.13
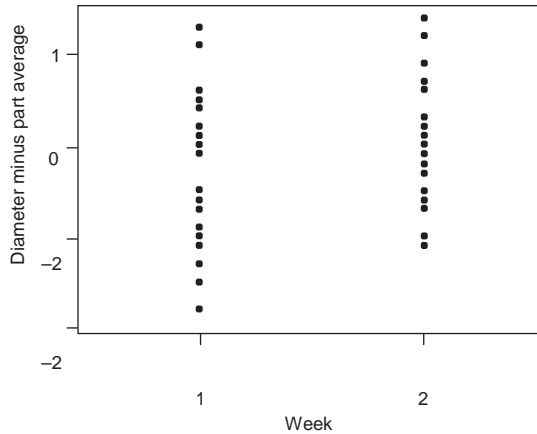
Figure S7.3   Deviation from part average versus week.

The standard deviation of the operator averages is 0.484. To estimate reproducibility, we calculate

$$\sqrt{0.484^2 - \left(0.593\right)^2 \big/ 18} \quad \boxed{} \, 0.463$$

also a large component of the overall variation.

   We can assess stability by plotting the deviations from part averages versus week as in Figure S7.3.

   If we know the true values of the parts being measured, as in Section 7.3, we can assess linearity by plotting the average measurement error (estimated bias) for each part, where the average is calculated over all days, operators, and repeats, versus the part size.

   In summary, the repeatability, reproducibility, stability, and linearity of a measurement system do not by themselves play an important role in the assessment of the system for use in the Statistical Engineering algorithm. What we need is an estimate of the overall measurement variation (and perhaps the bias) to determine if the measurement system is the home of the dominant cause of deviation.

S7.4 GAGE R&R AND ISOPLOT INVESTIGATIONS TO ASSESS MEASUREMENT VARIATION

In sections 7.2 and 7.3, we recommended a plan and an analysis to estimate the variation and bias of a measurement system. The plan involved repeatedly measuring three specially selected parts using different operators, and possibly different gages, over both the short term and long term. Here we discuss two alternatives for estimating the measurement variation: gage R&R and Isoplot investigations.

Gage Reproducibility and Repeatability (R&R)

To execute a gage R&R investigation, as described by AIAG (1995a) and Farnum (1994), we:

   • Choose 10 parts from regular production
   • Select two or three production operators
   • Have each operator measure each part two or three times

We then use an analysis of ranges to estimate the repeatability, the reproducibility, and the R&R, the combined effect of the two sources of variation. The R&R is an estimate of the overall measurement variation only if the time-to-time component is negligible. In the standard analysis, we also use the same data to estimate *stdev*(due to process).

This plan and the basic analysis have several deficiencies:

- There is no time-to-time component in the plan. Most R&R studies are carried out as quickly as possible and miss the day-to-day or week-to-week component of the measurement variation. The R&R likely underestimates the overall measurement system variation.

- The use of 10 parts selected over a short time is not sufficient to produce a reliable estimate of the variation of the true part dimensions in the process. Rather, we should use the available baseline information.

- The analysis of ranges is inefficient and should be replaced by the corresponding analysis of variance. The ANOVA analysis is the default in MINITAB. See Wheeler (1992).

- The numerical summaries should be augmented by appropriate graphs—see the MINITAB results that follow. This allows useful checks for outliers and interaction effects.

To illustrate, we use the camshaft journal diameter context for demonstration purposes. Data generated to be realistic are given in the data file *camshaft journal diameter measurement gageRR*. We simulated the use of three operators who measure each of the 10 parts twice for a total of 60 measurements.

The plan we recommended in Section 7.2 requires the same number of measurements in total as the typical R&R investigation but uses fewer parts and takes longer. We can add multiple gages and time periods to the R&R plan—see the AIAG *Measurement System Analysis* manual (1995)—but then the standard MINITAB gage R&R routine cannot be used.

In any case, using the traditional gage R&R analysis *stdev*(due to process) is poorly estimated. In Table 6.3, the relative precision for estimating a standard deviation with a sample size of 50 is about ±20% For a sample size of 10, relative precision is approximately ±50%. In other words, there is a huge uncertainty.

Alternately, we recommend estimating the overall variation in the baseline investigation that uses hundreds of parts. We can then calculate the variation in the true values, *stdev*(due to process), using the square root formula

$$\textit{stdev}(\text{due to process}) \square \sqrt{\textit{stdev}(\text{total})^2 + \textit{stdev}(\text{due to measurement})^2}$$

To use this idea in the MINITAB gage R&R analysis, we need to enter the historical standard deviation. The detailed MINITAB results for the gage R&R are:

Gage R&R Study - ANOVA Method
Gage R&R for measurement

Two-Way ANOVA Table With Interaction

Source	DF	SS	MS	F	P
part	9	2569.27	285.474	668.967	0.00000
operator	2	12.98	6.490	15.209	0.00014
operator*part	18	7.68	0.427	0.887	0.59664
Repeatability	30	14.43	0.481		
Total	59	2604.36			

Two-Way ANOVA Table Without Interaction

Source	DF	SS	MS	F	P
part	9	2569.27	285.474	619.608	0.00000
operator	2	12.98	6.490	14.087	0.00002
Repeatability	48	22.12	0.461		
Total	59	2604.36			

Gage R&R

Source	VarComp
Total Gage R&R	0.762
Repeatability	0.461
Reproducibility	0.301
operator	0.301
Part-To-Part	47.502
Total Variation	48.264

Source	StDev (SD)	Study Var (5.15*SD)	%Tolerance (SV/Toler)	%Process (SV/Proc)
Total Gage R&R	0.87304	4.4962	17.98	14.42
Repeatability	0.67877	3.4957	13.98	11.21
Reproducibility	0.54906	2.8277	11.31	9.07
operator	0.54906	2.8277	11.31	9.07
Part-To-Part	6.89219	35.4948	141.98	113.83
Total Variation	6.94726	35.7784	143.11	114.74

Number of Distinct Categories = 11

From this investigation, we obtain an estimate for *stdev*(due to measurement) of 0.87. The most critical result is given in the row labeled "Total Gage R&R" and the column labeled "%Process (SV/Proc)" in the final ANOVA table. This value estimates 100 * *stdev*(due to measurement)/*stdev*(due to process), that is, $100/D$ as defined by Equation (7.2).

In the example we have $100/D = 14.42$, or $D = 6.9$. Thus, we conclude that the measurement system is adequate. Note that by providing an estimate for *stdev*(overall) we have not used the estimate for *stdev*(due to process) derived just from the gage R&R (6.89 in the example).

The estimate for *stdev*(due to measurement) derived from the gage R&R results differs from the 0.756 we obtained in Section 7.2 based on the same data. The gage R&R results are determined using a *random effects* ANOVA model, while the Section 7.2 results come from a *fixed effects* ANOVA model with different assumptions. See Feder (1974) for a discussion of the difference between fixed and random effects models. For our purposes, neither model is ideal. However, since we are looking for large effects, either model would suffice.

The AIAG standard for measurement systems is that the measurement variation must be less than 30% of the overall variation. This criterion corresponds to a discrimination ratio *D*, as defined by Equation (7.2), greater than 3.2. We recommend a more liberal cutoff point for problem solving, since by using the baseline data we have a better estimate for *stdev*(process). The more stringent AIAG standard applies to measurement systems used for process control and part inspections.

The gage R&R analysis should be supplemented by graphical displays to check assumptions underlying the numerical calculations. The useful default plots automatically provided by MINITAB are given in Figure S7.4.

We can examine the various graphs looking for evidence of outliers and interactions between operators and parts. In the example, we see no concerns.
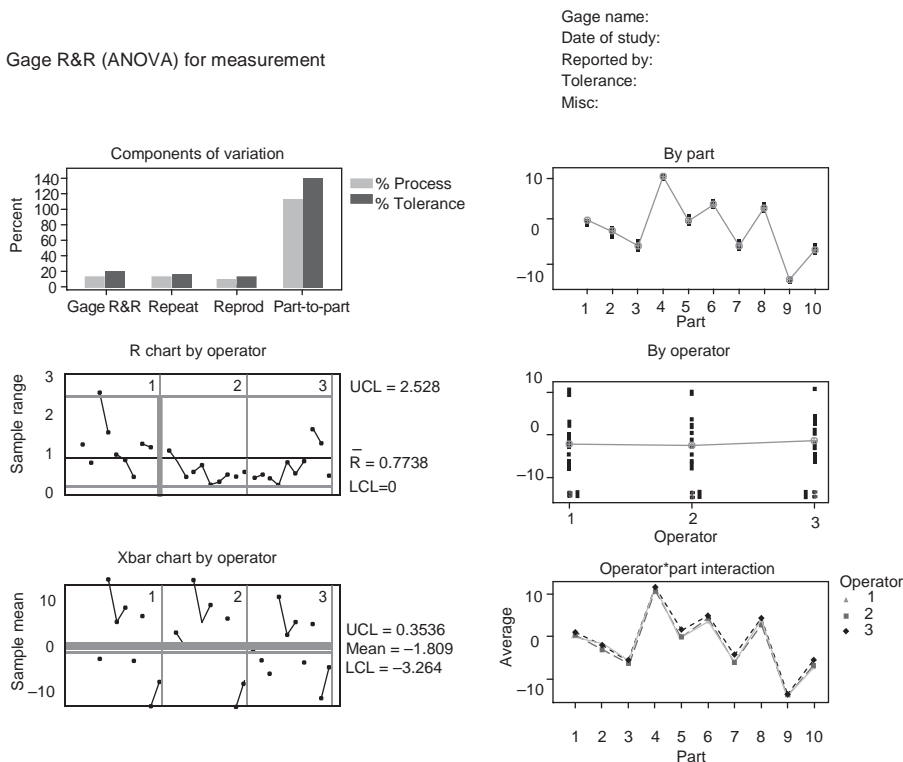
Figure S7.4 Default MINITAB graphical gage R&R output.

Isoplot

A second method for assessing a measurement system is an Isoplot investigation (Traver, 1995). An Isoplot investigation has a simple plan:

- Select 30 parts from production to reflect the baseline variation

- Measure each part twice using different operators, gages, and time periods

Note that if we use the same operator and gage to measure each part over a short time, we are assessing repeatability only. We analyze the data using ANOVA and a scatter plot. We give Isoplot data generated to be realistic for the camshaft journal diameter example in the data file *camshaft journal diameter measurement isoplot*.

 The (edited) ANOVA results for the Isoplot investigation are as follows. We use only a *part* term in the ANOVA model.

One-way ANOVA: diameter versus part

Analysis of Variance for diameter

Source	DF	SS	MS	F	P
part	29	2551.910	87.997	128.14	0.000
Error	30	20.602	0.687		
Total	59	2572.512			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	------+---------+---------+---------+
1	2	–10.998	0.030	(*-)
2	2	–8.322	0.168	(-*)
- - -				
29	2	10.150	0.976	(-*)
30	2	–2.352	0.508	(*-)

```
                                    ------+---------+---------+---------+
Pooled StDev =    0.829               –8.0      0.0      8.0     16.0
```

From the ANOVA results, we estimate *stdev*(due to measurement) as 0.829 and we can estimate *stdev*(due to process) as $6.55\sqrt{6.603^2 - 0.829^2}$. The plan uses 30 parts to estimate the process variation. This is an improvement over the gage R&R. With only 30 parts, the relative precision for estimating a standard deviation is approximately ±25%. We prefer to derive an estimate for *stdev*(due to process) by combining the results of the baseline and measurement investigations.

 We also plot the first versus second measurement on each of the selected parts as in Figure S7.5. This scatter plot is sometimes called an Isoplot. We can see the process variation by looking at the spread of values on either axis. Note that a point on the 45° line corresponds to the two measurements on that part being identical. If all the plotted points are clustered tightly around the 45° line, then we know that the measurement system standard deviation (or the appropriate component, depending on the plan) is small relative to the process variation.
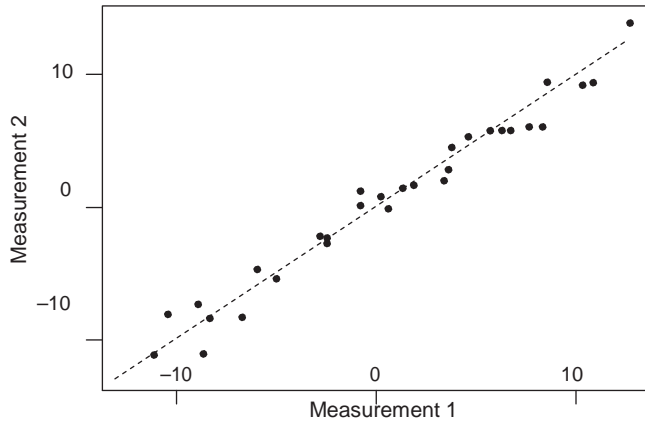
Figure S7.5   Scatter plot of the measurement results (dashed line represents "measurement 1" = "measurement 2," the 45° line).

The scatter plot is valuable because we can easily see the comparison of the measurement variation and process variation. We can also identify any outliers on the plot.

The Isoplot method has several deficiencies:

- There is no time-to-time component in the plan. Most Isoplot investigations are carried out as quickly as possible and miss the day-to-day or week-to-week component of measurement variation.

- Since all second measurements are conducted under identical conditions, only one family of causes that influence measurement variation is assessed at a time.

- The use of 30 parts and 30 repeated measurements is not enough to get good estimates of the process and measurement variation, respectively.

Bhote and Bhote (2000) recommended the use of the Isoplot and a minimum measurement discrimination ratio *D* of 5 before proceeding to the next stage of their process-improvement algorithm. We believe this is too conservative.

Comparison to Proposed Measurement Assessment Plan in Chapter 7

In comparison to both gage R&R and Isoplot, our proposed measurement assessment plan uses fewer parts with more measurements per part. In this way, the investigation focuses on estimating just the measurement variation (rather than the variation due to the process as well). The available baseline information gives an estimate of the total variation, and using Equation (7.2), we can solve for the variation due to the process. In this way we obtain more precise estimates of the variation both due to the measurement system and due to the process. This allows better decisions regarding the adequacy of the measurement system.

S7.5 INTERPRETING MEASUREMENT VARIATION

In the camshaft diameter example in Section 7.2, the estimated measurement variation is 0.756 microns. What does this tell us about the measurement system?

   If we know the bias is close to zero, then a series of measured values on the same part vary about the true value with standard deviation close to 0.756 microns. In other words, we would be surprised to find a measured value more than two standard deviations—that is, 1.63 microns—away from the true value.

   When the bias is not known, we cannot interpret the measurement variation in terms of how close the measured value will likely be to the true value. Instead we consider the difference in two measurements on the same part. Using the results from the models described in Chapter 2, the difference in two measurements on the same part will have mean zero (differencing eliminates the unknown bias and true value) and standard deviation close to $\sqrt{2} \times 0.756 = 1.07$. In other words, we would be surprised to see two diameter measurements on the same part differ by more than two standard deviations—that is, $2 \cdot 1.07 = 2.14$ microns.

S7.6 EFFECT OF MEASUREMENT VARIATION

We have argued that if the ratio of the process-to-measurement system standard deviations (that is, the discrimination ratio, *D*) falls between 2 and 3, we should improve the measurement system before we proceed, even though it is not a dominant cause of the overall output variation. Here we justify this recommendation.

   Measurement variation makes finding a dominant cause more difficult. For instance, measurement errors will influence estimates of the contributions of causes to the total variation. The importance of measurement variation depends on the approach and is of greater concern for approaches whose implementation requires ongoing measurement, such as feedforward control, feedback control, and 100% inspection.

   The effect of measurement variation depends on the goal of the process investigation. To keep the discussion simple, we formally compare the effect of measurement variation on our ability to detect differences in mean output level between two groups of parts. We may compare groups when trying to find or verify a dominant cause of variation and when assessing the proposed variation reduction approaches.

   As measurement variation increases, it becomes more difficult to determine if any observed difference between the groups is due to chance or due to a real difference in average level. We quantify the effect of measurement variation based on the increase in sample size (of the two groups) needed to detect the same size difference in the average level with the same power or probability. Our ability to detect a particular size difference in the average level depends on the inherent variation (*stdev*(total)) in the dimensions that here, for simplicity, we assume is the same in the two groups. Part of the inherent variation is due to measurement variation. We compare the case where there is no measurement error—that is, *stdev*(total) = *stdev*(due to process)—to the case where there is measurement error—that is, *stdev*(total), is given by Equation (7.1).

To detect differences in the two groups, we compare the observed sample averages in the two groups. Under conceptual repeated sampling and measuring, the standard error of an average decreases by a factor $1/\sqrt{n}$ as the sample size n increases. As a result, the sample size needed to maintain the desired power is greater by a factor

$$\left(stdev\left(\text{due to process}\right)^2 + stdev\left(\text{due to measurement}\right)^2 \right) \Big/ stdev\left(\text{due to process}\right)^2$$

if measurement variation is present.

Thus, using $D = stdev$(due to process)/$stdev$(due to measurement), measurement variation results in a $100/D^2$ percent increase in the sample size requirements over the case with no measurement error. For example, if $stdev$(due to measurement) = 0.3 $stdev$(due to process), that is, $D = 3.3$, the sample sizes needed to detect average differences are roughly 9% larger than if there were no measurement error. This seems a fairly small effect. To be somewhat conservative, and acknowledging that there may be some error in our of D, we set the minimum value at 3.

S7.7 FINDING A DOMINANT CAUSE OF MEASUREMENT VARIATION

There are many other attributes of the measurement process that may be of interest, especially if we judge the system to be inadequate. The basic idea is to divide the measurement bias and variation into components that can be assigned to inputs such as operators, parts, time, and the environment. We demonstrate the decomposition using operators.

Suppose that the measurement system includes two or more operators. We have defined the measurement variation and bias in terms of measurement errors based on a broad target population of measurements made on many parts over a long period of time. We can stratify this population according to which operator makes the measurements and, *for any particular operator,* we define the measurement bias and variation of the measurement system as the average and standard deviation of the measurement errors if that operator made the measurements.

If we assume that each operator makes the same proportion of the measurements, we can decompose the idealized histogram of the measurement errors into histograms for each operator such as the one shown in Figure S7.6.

In general (and as seen in Figure S7.6) it is true that:

- The measurement system bias is the average of the individual operator biases.

- The measurement variation depends on the measurement variation for individual operators and the relative biases among the operators.
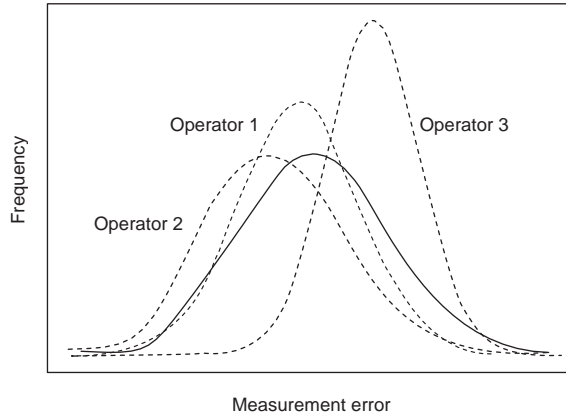
Figure S7.6   Decomposition of measurement errors by operator (solid line shows
distribution of overall measurement error).

The overall system bias can be zero even if each operator bias is not. Also, the differences in operator bias contribute to measurement variation. This is another example of the combination of group-to-group and within-group variation that we discussed in Chapter 2 and its supplement.

In looking for opportunities to improve a measurement system, we can stratify by operator. Suppose that we identify large differences among the operator-specific attributes. Then, we know that these differences arise because of differences in the methods among the operators. If we can standardize the method, we will see substantial improvement in the overall measurement system attributes.

We use the camshaft journal diameter data from the plan described in Section 7.2 to look at differences among operators. In this plan, each operator measured each part six times. Since we do not know the true value of the diameters, we calculate

$$\text{deviation} = \text{measured value} - \text{part average}$$

for each of the 54 measured diameters. We show these data plotted by operator in Figure S7.7.

The three operators have roughly the same measurement variation, but some (a small component) of the total variation comes from differences in relative bias between the operators, since the average "diameter minus part average" differs between operators.
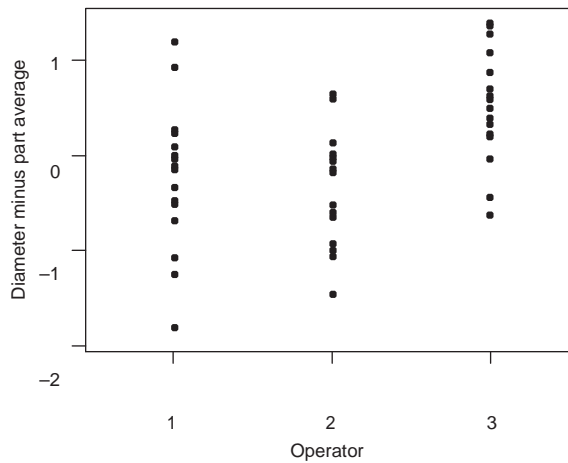
Figure S7.7   Diameter (deviation from part average) by operator.

We may see many different patterns in plots such as Figure S7.7 that will lead us to different actions such as:

- No differences among operator measurement bias and variation—look elsewhere for opportunities to improve the system.

- Different measurement variation or different relative biases for each operator— look for systematic differences among the methods used by the different operators.

We can do the same analysis with other inputs to the measurement process such as gages, environmental inputs, and time.

Chapter 9 Supplement

Finding a Dominant Cause Using the Method of Elimination

S9.1 COMPARISON OF STRATEGIES FOR FINDING A DOMINANT CAUSE

A standard strategy to find the (dominant) cause of a problem is to:

- Create a list of all possible causes, using people experienced with the process.

- Prioritize the list of causes in terms of the likelihood that each has a large effect.

- Investigate the causes in order of priority, either singly or in groups.

The first step often involves brainstorming (Evans and Lindsay, 1993), a formal process designed to elicit ideas from a group of people. We then use a cause-and-effect diagram (Ishikawa, 1982) to categorize the causes into what we call families (main branches) and subfamilies (twigs). We can prioritize the causes using multivoting (Scholtes, 1988), a nominal group technique (Brassard, 1988), or some other process, democratic or otherwise. We then have relatively few causes left to investigate. We can use available data and experimental or observational investigations to determine the effects of the top-ranked causes.

In Figure S9.1, we give a cause-and-effect diagram that was constructed by a team trying to reduce variation in the diameter runout on a transmission output shaft. The diagram is unusual in that it does not use the standard main branches (man, machine, material, method, measurement, and environment). Given this list of potential dominant causes, the team struggled with what to do next. In the end, they abandoned the idea of finding the dominant cause and instead looked to make the process more robust by changing a number of fixed inputs in an experimental investigation. This investigation was a failure.
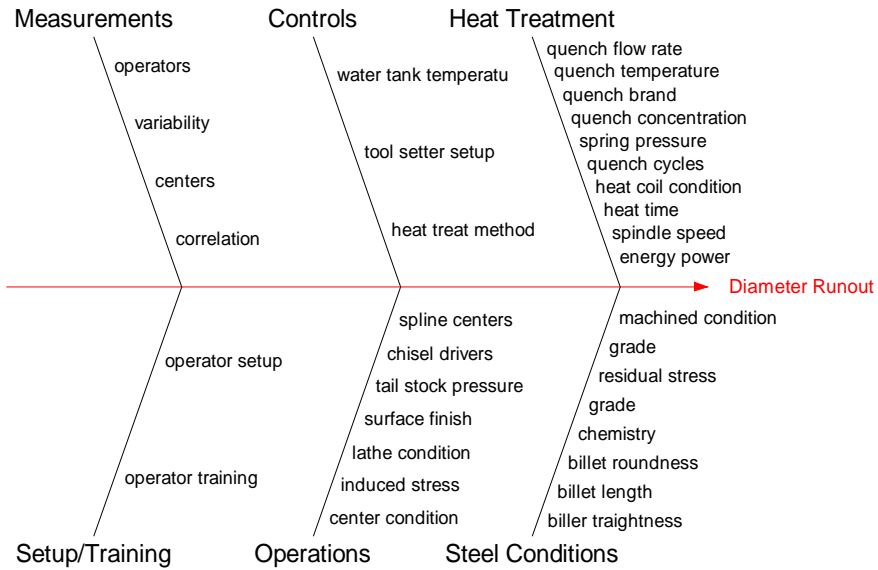
Figure S9.1   Example of conventional brainstorming to list all potential dominant causes.

We have recommended a different strategy, the method of elimination, based on forming and eliminating families of causes using available data and observational plans. The strategy involves:

- Dividing the causes into two or more families
- Using simple process investigations to rule out all but one family
- Iterating until one suspect (or a few at most) remains as a possible dominant cause
- Verifying the suspect(s) as the dominant cause

We expect that we will require several iterations to rule out most causes and home in on the dominant cause. At each iteration, we split the remaining causes into families based on process knowledge and the ease with which we can investigate the process. We never list causes in families that have been eliminated.

There are two major differences between the two strategies.

The cause-and-effect diagram is a hindrance for the method of elimination. The families are not constructed with the idea of elimination. In the output shaft example, we cannot think of an investigation that would eliminate all but one of the five families. If the main branches are considered families, there are too many families to start and time is wasted getting the detailed list of causes within each main branch. We have been involved in projects where considerable time was spent arguing about the labels for the main branches of the cause-and-effect diagram and which causes went where.

The second difference is the method of prioritizing the causes. With the method of elimination, at each step there are only a few families, and we can let the process do the voting by carrying out an appropriate investigation. On the other hand, with a cause-and-effect diagram and multivoting, for example, there is no guarantee that the dominant cause is anywhere near the top or even on the final list.

S9.2 NO SINGLE DOMINANT CAUSE

In the method of elimination, we assume there is a single dominant cause of the baseline variation. This assumption can be wrong in two ways:

- There are two (or perhaps three) causes, each with a large effect.

- There is no dominant cause, only a large number of causes, each with a relatively small effect.

If there are two (or three) dominant causes, each with a large effect, that live in different families, we cannot rule out either of those families. We then investigate the families separately to search for the dominant causes. Once we isolate the dominant causes, we may need to choose a different variation reduction approach to address each of them.

The second case in which there is no dominant cause contradicts the Pareto Principle. We believe this to be a rare occurrence. We may find a dominant family but no single cause within that family. We may rule out any of these families and then decide, in the next split, that there are two dominant families. As we partition the causes further, we will never be able to explain the full extent of variation found in the baseline investigation with a family of only a few causes.

In our experience, the search for the dominant cause is often abandoned in these cases. The consequence is that none of the caused-based variation reduction approaches is an available remedy. We must select a working approach that does not require the identification of the cause.

Chapter 10 Supplement

Investigations to Compare Two Families of Variation

S10.1 ANALYSIS OF VARIANCE

To find the dominant cause, we need to identify the family that explains a large component of the overall variation. In the examples presented in Chapter 10, we relied on graphical displays (for example, scatter plots, box plots, and multivari charts). If needed, we can use a one-way analysis of variance (ANOVA) to estimate the components of variation due to each family. We describe ANOVA in more detail in Appendix D.

Consider the rod thickness example. The data are given in the file *rod thickness base-line*. The ANOVA results correspond to Figure 10.3.

```
Analysis of Variance for thickness
Source      DF        SS        MS         F         P
position     3    46256.5   15418.8    241.45     0.000
Error      796    50831.0      63.9
Total      799    97087.5
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level        N      Mean     StDev   ------+---------+---------+---------+
1          200    37.820     8.040                           (-*-)
2          200    37.975     8.286                           (-*-)
3          200    21.580     7.559   (-*-)
4          200    40.925     8.062                                (-*-)
                                     ------+---------+---------+---------+
Pooled StDev =    7.991                24.0      30.0      36.0      42.0
```

The important result in the ANOVA output is the Pooled StDev, which is an estimate of the process standard deviation if all of the position averages are equal. In this example, the Pooled Stdev equals

$$\sqrt{\frac{stdev(\text{due to position } 1)^2 \ \square.\square\ stdev(\text{due to position } 4)^2\ 4}{}}$$

Pooling the standard deviations within each position is the method that we used to estimate the measurement system variation in Chapter 7, where we pooled the standard deviations within each part.

Since the baseline standard deviation for rod thickness was 11.023 (see Chapter 6), we can make substantial improvement by aligning the position averages. From the plots presented earlier, we need to increase the thickness center for position 3. We can adopt the Move the Process Center approach to learn how to make the change. We can use the ANOVA results to assess the maximum benefit from this process change.

We consider a second example to illustrate some further complexities in interpreting ANOVA results.

Block Porosity

Consider the block porosity example in Chapter 10. The data are given in the file *engine block porosity multivari*. Recall that porosity was measured on five blocks molded consecutively, once per hour over one shift. The goal was to identify which of the two families, hour-to-hour or mold-to-mold, was the home of the dominant cause. We apply a one-way ANOVA with porosity as the output and hour as the only term in the model. The results are:

One-way ANOVA: porosity versus hour

Analysis of Variance for porosity

Source	DF	SS	MS	F	P
hour	7	282168	40310	7.72	0.000
Error	32	167133	5223		
Total	39	449301			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	------+---------+---------+---------+
1	5	20.80	26.34	(------*------)
2	5	102.60	48.86	(-----*------)
3	5	7.40	5.22	(------*-----)
4	5	8.40	3.71	(------*-----)
5	5	224.80	127.12	(-----*------)
6	5	8.40	6.99	(------*-----)
7	5	14.20	8.96	(-----*------)
8	5	190.00	149.57	(------*------)

```
                                    ------+---------+---------+---------+
Pooled StDev =    72.27               0       100       200       300
```

The Pooled StDev, 72.3, at the bottom of the MINITAB results gives an estimate of the variation due to causes that act in the mold-to-mold family.

Using the overall standard deviation normally available from the baseline investigation, we estimate the component of the variation due to the hour-to-hour family as

$$stdev(\text{due to hour-to-hour}) \square \sqrt{stdev(\text{total})^2 + stdev(\text{due to mold-to-mold})^2}$$

Here the team did not conduct a baseline investigation for the new measure of porosity. They judged that they had seen the full extent of variation during the one-day investigation, because the proportion of scrapped blocks matched the historical level. As a result, we estimate the total variation by looking at the variation in porosity over the course of the investigation. We can use MINITAB to produce this estimate *stdev*(total) = 107.3:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
porosity	40	72.1	16.5	58.7	107.3	17.0

Variable	Minimum	Maximum	Q1	Q3
porosity	0.0	418.0	9.0	102.0

An alternative is to take into account the special structure in the data. We can indirectly estimate the total variation using the available data by specifying hour as a random effect in the one-way ANOVA. Here the MINITAB results are:

General Linear Model: porosity versus hour

Factor	Type	Levels	Values
hour	random	8	1 2 3 4 5 6 7 8

Analysis of Variance for porosity, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
hour	7	282168	282168	40310	7.72	0.000
Error	32	167133	167133	5223		
Total	39	449301				

Variance Components, using Adjusted SS

Source	Estimated Value
hour	7017
Error	5223

Based on this result, we estimate the variation associated with the hour-to-hour family as $\sqrt{7017} = 83.8$, and of the total variation as

$$stdev\left(\text{total}\right) = \sqrt{stdev\left(\text{due to hour-to-hour}\right)^2 + stdev\left(\text{due to mold-to-mold}\right)^2} = 110.6$$

In this example, the two methods give similar estimates for the total variation.

Here, neither family is dominant. This is not surprising since we only resort to ANOVA when the conclusion from the plots is not obvious. However, we see that if we could eliminate the mold-to-mold variation (or at least a substantial portion of it), we could substantially reduce the overall variation. The major advantage of using ANOVA when there is no obvious dominant family is that we can quantify the contributions of the various families. We can then decide which family we want to pursue. We expect there to be a dominant cause within each family.

Note also that ANOVA Pooled StDev does not capture the interaction between causes in the two families, which is apparent in Figure 10.9. This interaction was the most important discovery in the investigation. We can see the interaction in the ANOVA results since the StDev values for each hour are very different (for example, 26.3 for hour 1 and 149.6 for hour 8). The calculated Pooled StDev masks these differences.

S10.2 POSSIBLE EXPLANATIONS FOR OBSERVED CHANGES IN VARIATION

In some investigations, we observe a change in both process center and variation when the output is stratified by location or time. The engine block porosity problem provides an example. From Figure 10.9 we see that both the average level and variation of porosity change from hour to hour. We learn a lot about the nature of the dominant cause from this observed pattern.

Here we explore possible reasons for a change in both average and variation. We give two simple explanations in Figure S10.1, where the relationship between the dominant cause and the output is roughly linear. The horizontal and vertical axes give the full range of variation for the cause and the output.
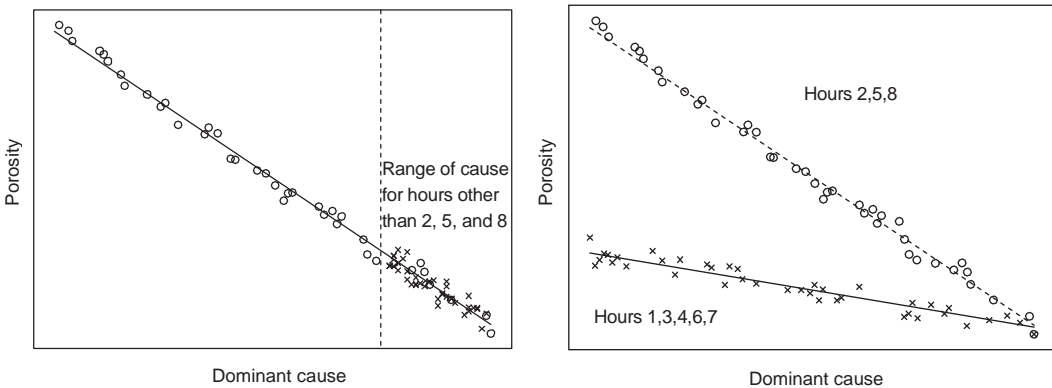
Figure S10.1   Possible explanations for engine block porosity results (*o*—observations for hours 2, 5, and 8; *x*—observations for other hours).

In the left-hand panel, the range of values for the dominant cause for hours 2, 5, and 8 is different than the range for the other hours. In the right-hand panel, the relationship between the cause and porosity depends on the time period, and in hours 2, 5, and 8 the process is much more sensitive to variation in the dominant cause. This pattern in the right panel indicates there is a strong interaction involving the plotted cause and some other input that changes from hour to hour.

In the example, the range of the dominant cause was different in hours 2, 5, and 8. At these times, the pouring temperature was lower and varied substantially, since during breaks the iron cooled.

There is a similar explanation for a change in variation between groups when there is no corresponding change in the process center, as in the hypothetical V6 piston diameter illustrated in Figure 10.17. The two possibilities are illustrated in Figure S10.2, where the horizontal and vertical axes give the full range of variation for the dominant cause and the output. We have:

- Left panel: The variation in the dominant cause is different for streams *A* and *B*

- Right panel: The effect of changes in the dominant cause is much greater for
  stream *B* than stream *A*

We can also have a combination of the two scenarios.

If we observe such patterns, we cannot rule out the within-strata or strata-to-strata families. However, we do get a strong clue about the nature of the dominant cause. To proceed, for example in the V6 piston investigation, we consider causes within stream *B* and look for differences in causes or effects between the two streams. We have a big advantage because we know it is possible to run process (that is, stream *A*) in a way that results in substantially less output variation.
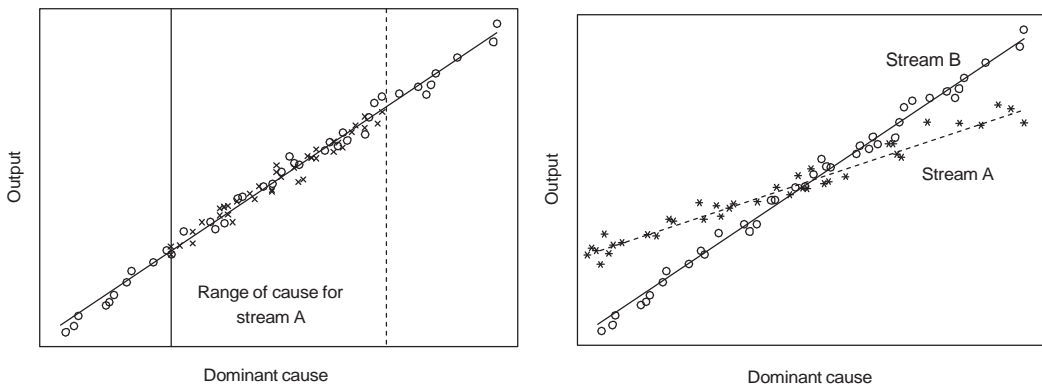
Figure S10.2   Possible explanations for pattern in Figure 10.17 (observation from streams *A* and *B* are denoted with *x* and *o,* respectively).

S10.3 WHY FORMAL HYPOTHESIS TESTS ARE NOT RECOMMENDED

Formal hypothesis tests, such as *t*-tests and the *F*-test given in ANOVA output, are used to determine if there is a *statistically significant* difference between the output averages when stratified by the value of an input. In this context, by a *statistically significant* difference we mean that we cannot explain the difference in strata averages by the variation within the strata. Consider the rod thickness example discussed in sections 10.1 and S10.1. We give again the ANOVA results that correspond to Figure 10.1.

One-way ANOVA: thickness versus position

Analysis of Variance for thickness

Source	DF	SS	MS	F	P
position	3	46256.5	15418.8	241.45	0.000
Error	796	50831.0	63.9		
Total	799	97087.5			

				Individual 95% CIs For Mean Based on Pooled StDev
Level	N	Mean	StDev	------+---------+---------+---------+
1	200	37.820	8.040	(-*-)
2	200	37.975	8.286	(-*-)
3	200	21.580	7.559	(-*-)
4	200	40.925	8.062	(-*-)
				------+---------+---------+---------+
Pooled StDev =	7.991			24.0 30.0 36.0 42.0

We see that:

- There are statistically significant differences among the position average thicknesses, since the *p*-value 0.000 in the *F*-test is so small

- Some of the 95% confidence intervals for the position means do not overlap in the plot

We conclude that there are position-to-position differences among the averages. We cannot conclude, however, that position is a dominant cause.

We see the same difficulties, as illustrated here for ANOVA, with nonparametric tests based on end-counts (due to Tukey, 1959) as recommended by Bhote and Bhote (2000) and others. The tests can identify groups with statistically significant differences but do not tell us if the dominant cause acts in the group-to-group family.

In general, with large sample sizes, a formal hypothesis test will show that there are statistically significant differences among group averages. The small difference may have no *practical significance,* however. We are searching for a dominant cause that explains a large part of the variation in the output, not for causes with small but statistically significant

effects. The dominant cause will produce significantly different averages, but the converse is not necessarily true. For this reason, we do not rely on hypothesis tests to help in the search for a dominant cause.

S10.4 OTHER PLANS FOR COMPARING UPSTREAM AND DOWNSTREAM FAMILIES

In Chapter 10, we looked at a plan to compare upstream and downstream families by measuring the output at an intermediate operation and at the end of the process for the *same* parts. Additional references concerning variation transmission investigations include Agrawal (1997), Lawless et al. (1999), and Agrawal et al. (1999).

Here we consider two other plans for comparing upstream and downstream families that do not require measuring the output mid-process. These plans, operations swap and randomized sequencing, have limited applicability. There are more complex versions if we want to compare several processing steps simultaneously.

Operations Swap

Operations swap (Ingram, 1993) is feasible if:

- Differences between two parallel processes are a dominant cause of variation.

- We can swap parts between the two parallel processes at some intermediate processing step.

The idea is simple. We swap the production path of parts moving through the process at a given processing step (for example, halfway through the process). Figure S10.3 illustrates the plan with two parallel lines and two operation steps. The solid lines show the usual process path through lines 1 and 2. The dashed lines show how we propose to swap the path of parts between operations *A* and *B* temporarily for the purposes of the investigation.

We keep track of the parts that are swapped. Then, we compare the output of the two streams of swapped parts with the performance of the regular processes. We know there was a substantial difference in performance between the two production lines before the swapping. If the performance of the swapped parts that went through operation *A* in line 1 and operation *B* in line 2 matches the performance of line 1 before the swapping
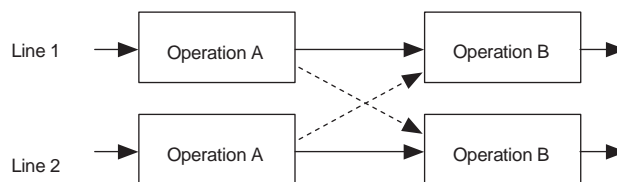
Figure S10.3   Operations swap illustration.

investigation, and similarly for line 2, then we can conclude that the dominant cause of the difference between the two lines is in operation *A* or earlier.

An operations swap investigation can be logistically difficult. We need to continue swapping, tracking swapped parts, and measuring finished parts (though it is not necessary to swap all parts in the given time period) until we see the full extent of variation in the output.

Swapping parts from line 1 to line 2 *and* vice versa is not necessary. We could draw conclusions just by swapping in one direction. However, the double swap is useful to ensure nothing else has changed in the process and, for logistical reasons, to balance the load on the two production lines.

Randomized Sequencing

Randomized sequencing (Asher, 1987) is feasible if:

- The process output exhibits a known pattern of variation—for example, a drift— or sustains shifts.

- We can change the processing order of parts at an intermediate operation and later identify the order used.

Suppose the output drifts with a predictable pattern over each day. To conduct the randomized sequencing plan, we:

- Select a sample of parts spread out over one day after operation *A*.

- Record the time for each sampled part and set it aside.

- Randomize the order of these parts.

- Process and track the parts through operation *B*.

- Measure the output.

This plan is illustrated in Figure S10.4.

If the dominant cause lives in operation *A* or upstream, we will see the predicted drift when we plot the output against the time of processing through operation *A*. If we do not see the predicted drift, then we know the dominant cause lies in operation *B*.
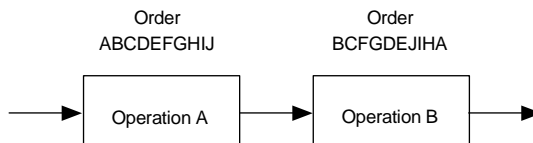
Figure S10.4   Randomized sequencing illustration.

S10.5 APPLYING LEVERAGE IN COMPARING ASSEMBLY AND COMPONENT FAMILIES

To compare the assembly and component families for an assembled product, we recommend the following plan:

- Select two parts with output at the extremes of the baseline distribution.

- Disassemble and reassemble each part at least three times and measure the output each time.

- Plot the results and look for changes in the output.

Large changes in the performance of the product after disassembly and reassembly indicate that the assembly family is the home of the dominant cause. It is surprising that we can make such a claim with such a small sample size. This is a good example of leverage, where we ensured that we would see the full extent of variation in the sample.

We can describe the output for any assembly as a function of a contribution from the components and a contribution from the assembly. The standard deviation of the output can be expressed as

$$stdev(\text{output}) \cong \sqrt{stdev(\text{due to components})^2 + stdev(\text{due to assembly})^2}$$

Suppose, for the moment, that the assembly variation is dominant—that is, *stdev*(due to assembly) is much larger than *stdev*(due to components). When we select two parts with extreme output values, we can be confident that the reason the values are extreme is that the contributing values from the assembly are also extreme. This must be true since, by our hypothesis, the contributions from the components are not highly variable and hence cannot produce the extreme values.

When we disassemble and reassemble the parts, the contribution of the components does not change. However, the assembly contributions will change, and since we started with extreme values, we expect that there will be large changes to the output.

On the other hand, if the dominant cause lives with the components—that is, *stdev*(due to components) is much larger than *stdev*(due to assembly)—then the reason for the extreme values is a large component contribution. When we disassemble and reassemble such parts, we will continue to see extreme values for the output since the component contribution is not changed.

Chapter 11 Supplement

Investigations to Compare Three or More Families of Variation

S11.1 ANALYSIS OF VARIANCE FOR MULTIVARI INVESTIGATIONS

We use multivari investigations to examine the effects of a number of families of causes simultaneously. If there are three or more families, our strategy is to examine several multivari charts to eliminate families as the home of a dominant cause. If there are many families, or, more important, no single dominant family, we can use analysis of variance to quantify the contribution of each family to the overall variation.

We illustrate the use of ANOVA with three examples where the additional analysis is warranted.

Casting Thickness

At a foundry, a team was assigned the task of reducing variation in thickness, specified at four locations on each casting. A baseline investigation found that the full extent of variation in thickness (as measured from nominal) was –20 to 35 thousandths of an inch. The team found that the measurement system was acceptable and decided to look for a dominant cause of thickness variation.

They conducted a multivari investigation in which they measured thickness at the four positions for three consecutive castings from each of the six cavities in the mold. They sampled each hour for six hours on two different days to get a total of 864 thickness measurements on 216 castings. The data are available in the file *casting thickness multivari*. This investigation will allow us to look at the following families and their interactions:

- Casting-to-casting
- Position-to-position
- Cavity-to-cavity
- Hour-to-hour
- Day-to-day

Plotting a histogram of the multivari data, as shown in Figure S11.1, we see close to the full extent of variation from the baseline.

We give a number of multivari charts in Figure S11.2 looking at the individual families. We have defined a new input *time* that sequentially numbers the 12 time periods representing all the possible combinations of hour and day and a new input *group* to uniquely identify the sets of three consecutive thickness measurements within a position at the different time periods. We used group = (time – 1)*24 + (cavity – 1)*4 + position.

The multivari charts are not easy to interpret, since there is no single dominant cause. We can attribute a large amount of variation to the position-to-position, cavity-to-cavity, and casting-to-casting families (as shown by the multivari chart by group). In this example, we can quantify the components of variation that can be attributed to the various families. Using the model from Chapter 2, if the causes in separate families act independently, we have

$$sd\ \text{total} \ = \sqrt{sd\left(\text{due to family1}\right)^2 + sd\left(\text{due to family2}\right)^2 + \dots + sd\left(\text{due to rest of causes}\right)^2}$$

In the ANOVA calculation, the total sum of squares is similarly partitioned into sums of squares associated with the various families. See Neter et al. (1996) and Box et al. (1978) for a general background on ANOVA. See also de Mast et al. (2001) for a more formal discussion of the connection between multivari charts and ANOVA.

We propose to fit an ANOVA model with all possible terms involving the inputs whose effects are expected to be systematic. For example, we might expect a systematic difference (in the average thickness) among the four positions. We would not expect systematic variation among the measurements taken at one position from consecutive castings. That is, we would be surprised if the average of the first measurement over all cavities and times was substantially different from the average of the second measurement.
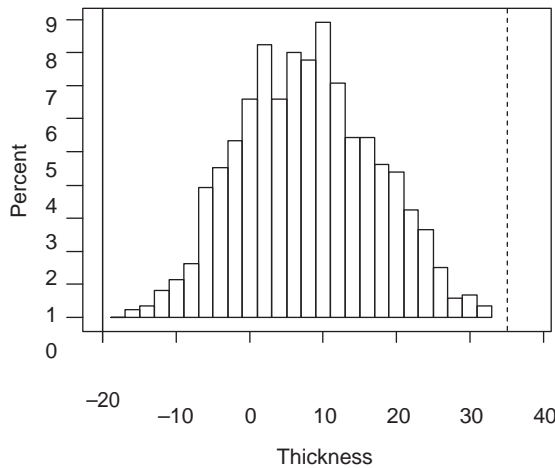
Figure S11.1   Histogram of casting thickness multivari data.
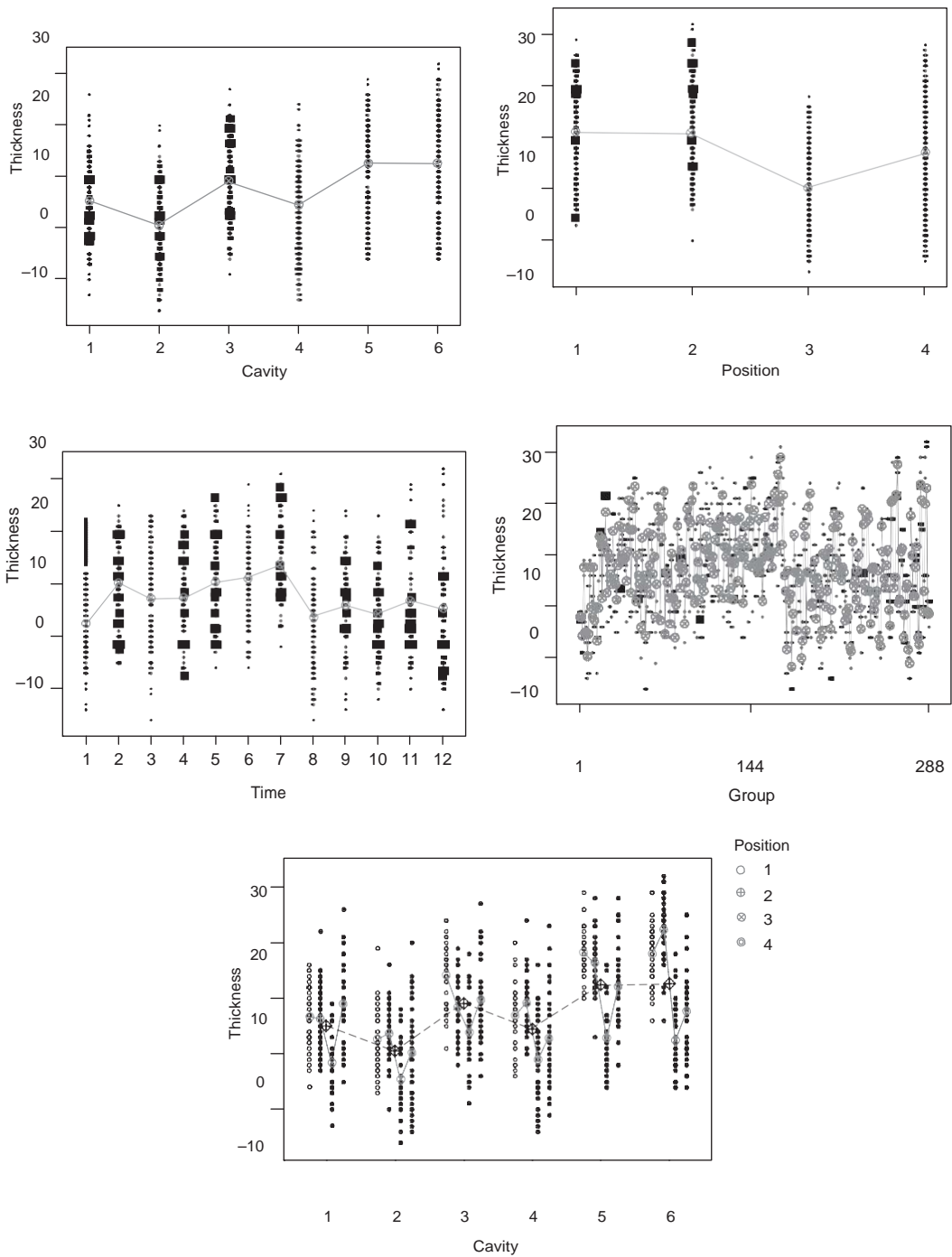
Figure S11.2   Multivari charts for casting thickness.

See Appendix D for details on how to fit the ANOVA model in MINITAB. We get the following results when we fit an ANOVA model with time, cavity, and position (the three inputs expected to have systematic effect) and all interactions among these three inputs.

ANOVA: thickness versus time, cavity, position

Factor	Type	Levels	Values						
time	fixed	12	1	2	3	4	5	6	7
			8	9	10	11	12		
cavity	fixed	6	1	2	3	4	5	6	
position	fixed	4	1	2	3	4			

Analysis of Variance for thickness

Source	DF	SS	MS	F	P
time	11	9008.82	818.98	31.81	0.000
cavity	5	16994.99	3399.00	132.00	0.000
position	3	16697.24	5565.75	216.15	0.000
time*cavity	55	1544.02	28.07	1.09	0.311
time*position	33	7580.19	229.70	8.92	0.000
cavity*position	15	5363.43	357.56	13.89	0.000
time*cavity*position	165	1634.06	9.90	0.38	1.000
Error	576	14832.00	25.75		
Total	863	73654.75			

To compare the relative sizes of the families, we look at the sum of squares (SS) column. We see large, roughly equal-sized effects due to position, cavity, and error. This matches our conclusion from the multivari charts. The calculated sum of squares tell us approximately how much we could expect to reduce the total sum of squares if we could eliminate all variation due to the given family.

It is not easy to translate these sums of squares into estimates for the standard deviation attributable to each family, but there is a rough correspondence. Recall that we are looking for dominant causes. For example, if we eliminate cavity-to-cavity variation, the total sum of squares is reduced by 16994.99 to 56659.76 and the estimated overall standard deviation is reduced from 9.24 ($\sqrt{73654.75 / 863}$) to 8.10 ($\sqrt{56659.76 / 863}$).

The error sum of squares is relatively large at 14832. This sum of squares includes the effect of the casting-to-casting family (recall that the casting-to-casting family is expected to have haphazard effect) and all interactions between the casting-to-casting family and the other families. We look more at this family in the next section.

As there was no single dominant family, the team proceeded by trying to address each of the three large families separately.

Block Bore Diameter

We can also use an ANOVA when the volume of data in a multivari investigation is large. With large amounts of data, there are so many points on the multivari charts that they are difficult to interpret, especially when we are looking at interactions.

In an investigation to try to find the dominant cause of engine block bore diameter variation, the team had access to a large volume of data that was automatically collected on every block. The team sampled from the complete database to isolate the effects of several families. They selected diameter measurements from:

- Three consecutive blocks selected from the process each hour for 80 hours

- All eight bores for each block

- Three heights and two orientations at each height for each bore

We give the 11,520 observations in the file *block bore diameter multivari*. To reduce the number of families, we define a new input, called *position*, with six values, which labels the two different orientations at three heights. If the position-to-position variation is found to be large, we can look more closely at this family.

We find multivari charts, such as Figure S11.3, difficult to interpret because of the large number of time periods used in the investigation. The chart does show that there are several odd values about a third of the way through the data set that we should investigate further.
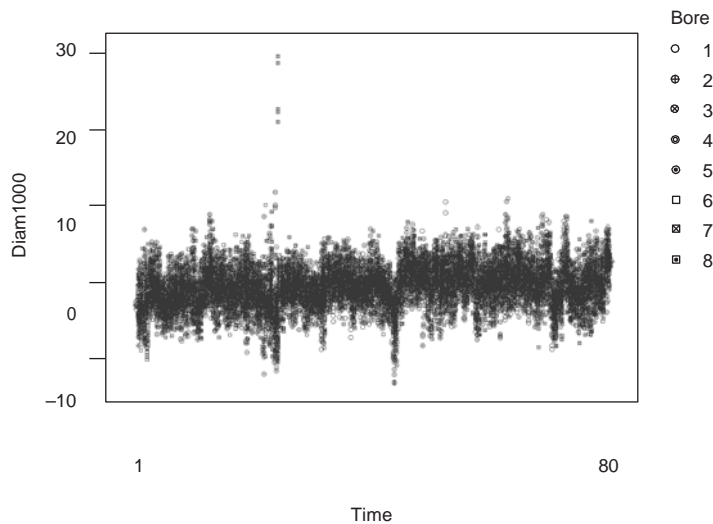
Figure S11.3   Multivari chart for bore by time.

To try to understand the results, we fit a full ANOVA model using all families (except block-to-block) and their interactions:

ANOVA: diam1000 versus time, bore, position

Factor	Type	Levels	Values						
time	fixed	80	1	2	3	4	5	6	7
			8	9	10	11	12	13	14
			15	16	17	18	19	20	21
			22	23	24	25	26	27	28
			29	30	31	32	33	34	35
			36	37	38	39	40	41	42
			43	44	45	46	47	48	49
			50	51	52	53	54	55	56
			57	58	59	60	61	62	63
			64	65	66	67	68	69	70
			71	72	73	74	75	76	77
			78	79	80				
bore	fixed	8	1	2	3	4	5	6	7
			8						
position	fixed	6	1	2	3	4	5	6	

Analysis of Variance for diam1000

Source	DF	SS	MS	F	P
time	79	24602.93	311.43	56.34	0.000
bore	7	951.65	135.95	24.60	0.000
position	5	10614.80	2122.96	384.09	0.000
time*bore	553	15568.13	28.15	5.09	0.000
time*position	395	5344.71	13.53	2.45	0.000
bore*position	35	4417.75	126.22	22.84	0.000
time*bore*position	2765	2935.01	1.06	0.19	1.000
Error	7680	42449.46	5.53		
Total	11519	106884.44			

The two largest sums of squares correspond to Error and time. We conclude that dominant causes of variation act in the block-to-block family (or in an interaction involving the block-to-block family) and the time-to-time family. The team proceeded to look for dominant causes within each of these families.

Camshaft Journal Diameter

In Chapter 11, we found a large grinder-to-grinder effect in the camshaft journal diameters. Looking at the part-to-part family with the multivari charts given in Figure 11.13 is difficult due to the large number of possible values for the input group that indexes the 192 sampling points. Here we fit an ANOVA model to the diameter data with inputs grinder,

batch, hour, and position, and all possible interactions among these inputs. This model allows us to quantitatively assess the relative sizes of the effects due to each of the families and interactions among the families.

ANOVA: diameter versus grinder, batch, position, hour

Factor	Type	Levels	Values						
grinder	fixed	2	A	B					
batch	fixed	3	1	2	3				
position	fixed	8	1f	1r	2f	2r	3f	3r	4f
			4r						
hour	fixed	4	1	3	5	7			

Analysis of Variance for diameter

Source	DF	SS	MS	F	P
grinder	1	11855.39	11855.39	1230.72	0.000
batch	2	2.15	1.07	0.11	0.894
position	7	298.55	42.65	4.43	0.000
hour	3	12.40	4.13	0.43	0.732
grinder*batch	2	22.54	11.27	1.17	0.311
grinder*position	7	47.14	6.73	0.70	0.673
grinder*hour	3	53.57	17.86	1.85	0.136
batch*position	14	223.45	15.96	1.66	0.060
batch*hour	6	85.25	14.21	1.48	0.184
position*hour	21	238.67	11.37	1.18	0.261
grinder*batch*position	14	92.57	6.61	0.69	0.789
grinder*batch*hour	6	112.93	18.82	1.95	0.070
grinder*position*hour	21	153.74	7.32	0.76	0.770
batch*position*hour	42	487.06	11.60	1.20	0.180
grinder*batch*position*hour	42	300.26	7.15	0.74	0.886
Error	768	7398.07	9.63		
Total	959	21383.75			

From the ANOVA results, we see (as expected) that the largest sum of squares is associated with the family of causes that differ between the two grinders. The sums of squares for any other systematic family (that is, batch, hour, position, and interactions) are much smaller. The error sum of squares, representing the part-to-part family, is also fairly large. We explore the part-to-part family further in the next section.

Comments

We can use a random-effects ANOVA model that ignores the systematic nature of the effects as an alternative analysis. See Feder (1974) for a background on random-effects models. However, it is incorrect to assume a random effect for a family like position in the casting thickness example since there are only four positions in total. The random-effects

ANOVA model provides estimates of the variance components attributable to the various families. In other words, the random-effects model provides estimates of *stdev*(due to a particular family).

In the casting thickness example, the results from the ANOVA model and the multivari charts match closely. The ANOVA has the advantage of quantifying the contributions of the various families and, in one step, allowing us to look at interactions among the families. However, note that ANOVA models are looking for variation caused by differences in average output for different level of the inputs. They are not good at detecting other patterns such as changes in output variation across different levels of the inputs or outliers. We recommend first looking at graphical summaries of the data such as multivari charts before fitting an ANOVA model.

If there are no families included in the multivari investigation that can be expected to have a haphazard effect, the proposed analysis strategy where we fit an ANOVA model with all inputs and interactions will not work in MINITAB. Although the appropriate sums of squares (which we use to draw conclusions) could still be calculated, MINITAB will not proceed because there are no degrees of freedom left to estimate the error sum of squares. If there are no haphazard families, we recommend leaving out one of the families in the ANOVA model. The error sum of squares is then attributable to the left-out family plus all interactions between the left-out family and other families in the model. If the error sum of squares is large, we redo the analysis leaving out a different family, continuing until we find the home of a dominant cause.

S11.2 HANDLING FAMILIES EXPECTED TO HAVE A HAPHAZARD EFFECT

For large data sets or multivari investigations with many families, we have difficulty displaying the effect of families, such as part-to-part, that are expected to have a haphazard rather than systematic effect. To explore the variation due to a haphazard family, we define a new input that uniquely numbers the groups of consecutive parts. For instance, in the cylinder head example from Chapter 11 we defined group as (time – 1)*4 + pattern. Similarly, in the fascia cratering example we used group = (time – 1)*2 + (machine – 1).

We illustrate the analysis to explore haphazard families with two examples.

Casting Thickness

The casting thickness example was introduced in the previous section. We expect the effect of the casting-to-casting family to be haphazard. We define a new input group that numbers the 288 different sets of consecutive thickness measurements from each different combination of time, cavity, and position—that is, we use group = (time – 1)*24 + (cavity – 1)*4 + position. Fitting a one-way ANOVA model with group gives:

One-way ANOVA: thickness versus group

Analysis of Variance for thickness

Source	DF	SS	MS	F	P
group	287	58822.7	205.0	7.96	0.000
Error	576	14832.0	25.7		
Total	863	73654.7			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	
				--+---------+---------+---------+----
1	3	-2.000	1.000	(---*---)
2	3	-2.000	1.000	(---*---)
3	3	-5.667	3.512	(---*---)
.				
.				
.				
286	3	23.000	14.731	(---*---)
287	3	-1.000	1.732	(---*---)
288	3	-1.333	2.517	(---*---)
				--+---------+---------+---------+----

Pooled StDev =   5.074

The ANOVA divides the overall variation (as measured by the total standard deviation $\sqrt{73654.7/863} = 9.24$) into two components.

Pooled StDev is an estimate of the variation attributable to the casting-to-casting family and all interactions between the casting-to-casting family and other families. In other words, if we could eliminate all group-to-group differences due to changes in time, cavity, and position, we estimate that the process standard deviation would be reduced to 5.074. That is, there is a dominant cause acting in the group-to-group family. We examined the components of this family (time, cavity, and position) in the first section of this supplement.

Here, the pooled standard deviation is relatively small. In other cases, there may be a dominant cause in the part-to-part family. To demonstrate the methodology, suppose there is a dominant cause in the casting-to-casting family and we explore the casting-to-casting family further. We define a new output characteristic, the within-group standard deviation for each group. We then repeat the multivari analysis using the new output *group stdev*. To do this we need to define new input characteristics that correspond to the levels of the other families associated with each group. In the example, we define *stime*, *scavity*, and *sposition*, which correspond to the time, cavity, and position associated with each group, respectively. See the discussion on multivari charts in Appendix C for details on how to do this easily with MINITAB.

In the multivari chart, we are looking for any systematic change in the casting-to-casting variation over the other families. Note that these changes reflect an interaction with the casting-to-casting family. Looking at the multivari charts in figures S11.4 and S11.5 we conclude that the largest effect is due to a time-by-cavity interaction. At some particular
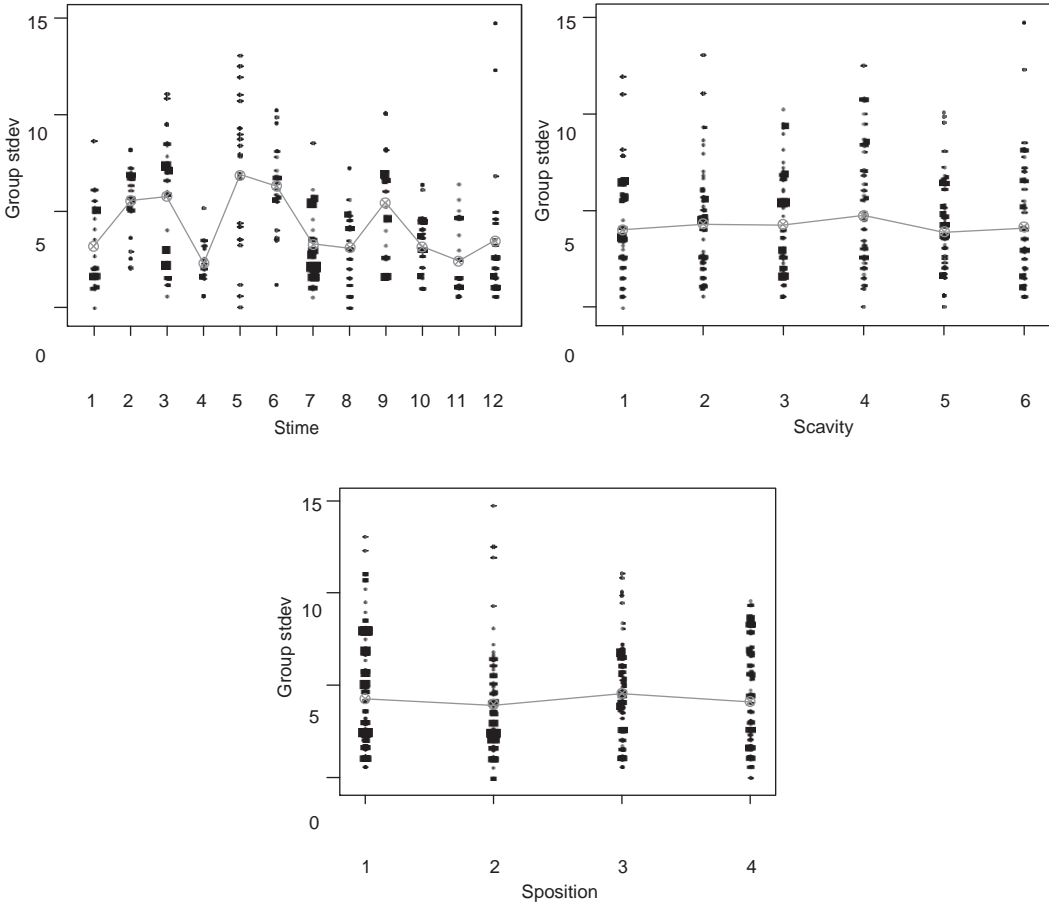
Figure S11.4   Multivari charts for casting thickness group standard deviation.
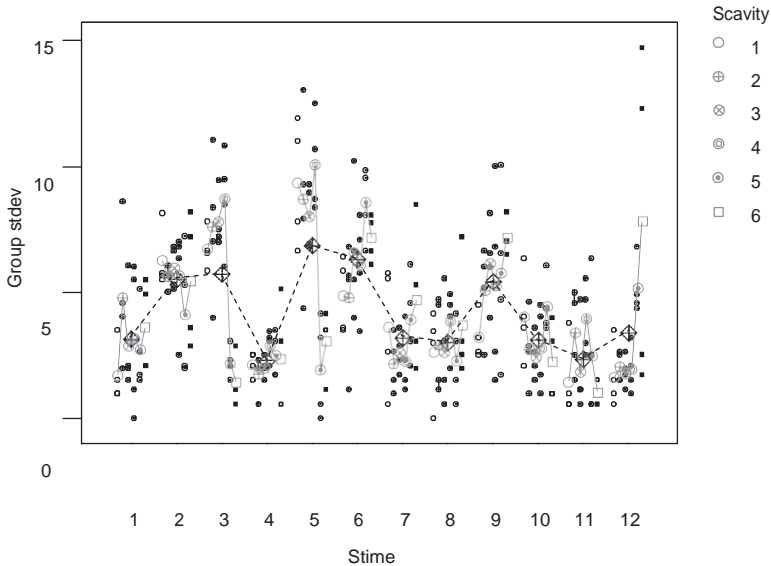
Figure S11.5   Multivari chart for *group stdev*, showing cavity by time interaction.

times and for some cavities, the variation in thickness for three consecutive castings changes substantially. We are looking for a cause that can explain such process behavior.

The interpretation of the results is difficult because the output is already a measure of variation. We see that for some times (for example, time 4) the variation in each group is small relative to other times. Also, for some times (for example, times 3, 5, and 12) the variation within a group varies for different cavities, while at all other times the variation within a group is similar for all cavities. From the interaction plot in Figure S11.5, we see that at some times, cavities 5 and 6 result in either more or less variation within a group than the other cavities. This matches the conclusion from fitting an ANOVA model with all but the three-input interaction term, as given:

ANOVA: group stdev versus stime, scavity, sposition

Factor	Type	Levels	Values						
stime	fixed	12	1	2	3	4	5	6	7
			8	9	10	11	12		
scavity	fixed	6	1	2	3	4	5	6	
sposition	fixed	4	1	2	3	4			

Analysis of Variance for group stdev

Source	DF	SS	MS	F	P
stime	11	705.022	64.093	22.19	0.000
scavity	5	22.070	4.414	1.53	0.184
sposition	3	14.738	4.913	1.70	0.169
stime*scavity	55	715.179	13.003	4.50	0.000
stime*sposition	33	288.226	8.734	3.02	0.000
scavity*sposition	15	70.297	4.686	1.62	0.073
Error	165	476.570	2.888		
Total	287	2292.102			

The conclusion is that the dominant cause acts casting to casting, but its effect varies with cavity and time. The team now must ask, "What causes vary in a way that matches this pattern?"

Camshaft Journal Diameter

The camshaft journal diameter example was introduced in Chapter 11 and also discussed in the previous section of this supplement. In the multivari analysis, we found the part-to-part variation was large, though not as large as the grinder-to-grinder variation. Here, we explore the part-to-part variation further by calculating the standard deviation within each group of parts by position, grinder, and batch. We define group = (grinder number – 1)*96 + (batch – 1)*32 + (hour–1)*8/2, since grinders *A* and *B* are coded 0 and 1, there are three batches, and the possible values for hour are 1, 3, 5, and 7. We then plot multivari charts using the within-group standard deviation as the output.
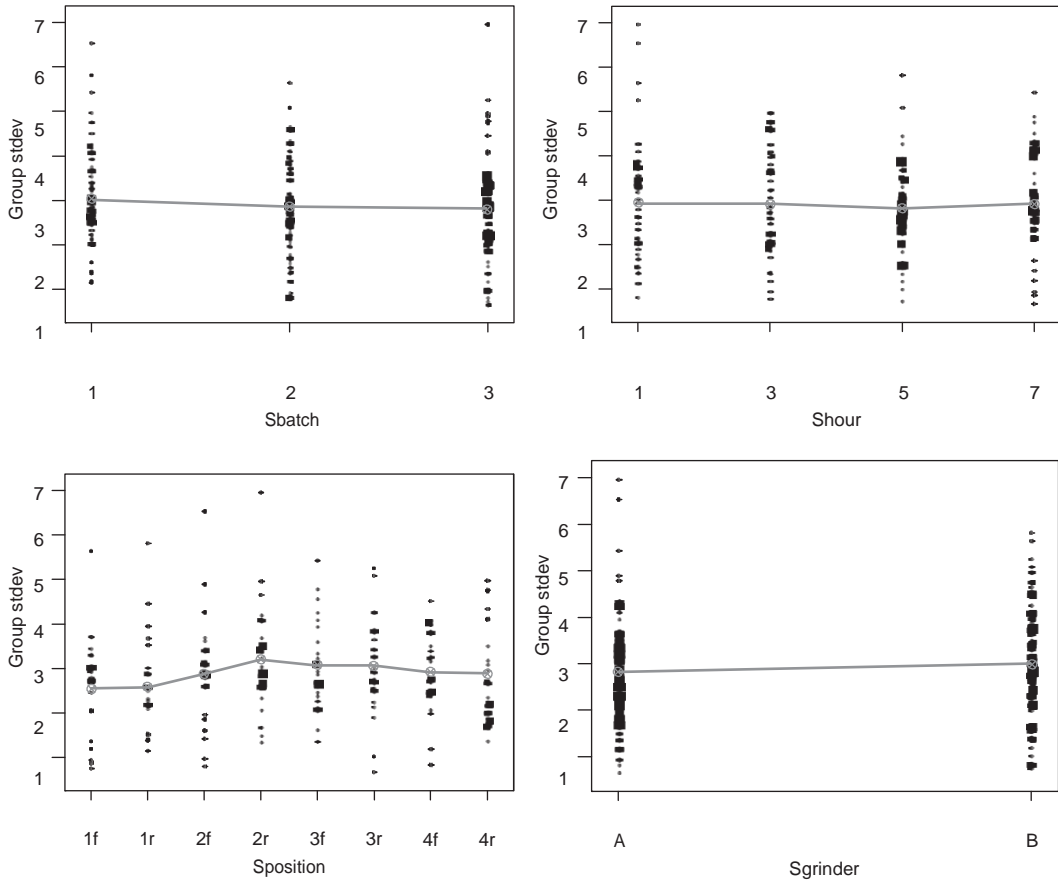
Figure S11.6   Multivari chart for *group stdev* for camshaft journal diameter multivari.

We conclude from Figure S11.6 that there is no evidence of any effects that change the within-group standard deviation. The dominant cause of the part-to-part variation acts in a similar way over all grinders, batches, and times.

S11.3 REGRESSION ANALYSIS FOR VARIATION TRANSMISSION INVESTIGATIONS

In Chapter 11, we analyzed the results of a variation transmission investigation using scatter plots. In most cases we find this graphical analysis sufficient. We can, however, supplement the graphical analysis with a numerical analysis based on regression models. Depending on the observed pattern in the scatter plots, we can use a variety of forms for the model. If the relationship appears roughly linear, as in Figure S11.7 which shows the V6 piston diameters after Operations 270 and 310, we can use a model of the form:

$$\text{diameter after OP310} = b_0 + b_1 * \text{diameter after OP270} + \text{residual},$$

where b_0 and b_1 are unknown model parameters that represent the *y*-intercept and slope,

respectively, of a straight line that summarizes the relationship between the two diameter measurements. The parameters b_0 and b_1 are estimated based on the available data.

Fitting a regression model (see Appendix E) to the data gives:

Regression Analysis: diameter after OP310 versus diameter after OP270

The regression equation is
diameter after OP310 = 64.3 + 0.884 diameter after OP270

Predictor	Coef	SE Coef	T	P
Constant	64.27	22.29	2.88	0.005
diameter	0.88358	0.03739	23.63	0.000

S = 1.224   R-Sq = 85.6%   R-Sq(adj) = 85.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	836.67	836.67	558.58	0.000
Residual Error	94	140.80	1.50		
Total	95	977.47			

The estimated model parameters for b_0 and b_1 are given as 64.3 and 0.884, respectively. The estimated slope parameter (0.884) can be interpreted as the change in the average diameter after Operation 310 for every unit change in the diameter after Operation 270. Often, as in this example, the *y*-intercept parameter estimate will have no useful interpretation, since it tells us the average diameter after Operation 310 if the diameter after Operation 270 is zero. The regression equation is a straight line that best summarizes the relationship between the piston diameters after Operation 270 and Operation 310. As we see in Figure S11.7, in this case the relationship is strong. The value of $s = 1.224$ from the MINITAB output is an estimate of the standard deviation in the final diameter if we hold the diameter after

Figure S11.7   Plot of final diameter versus diameter after Operation 270.
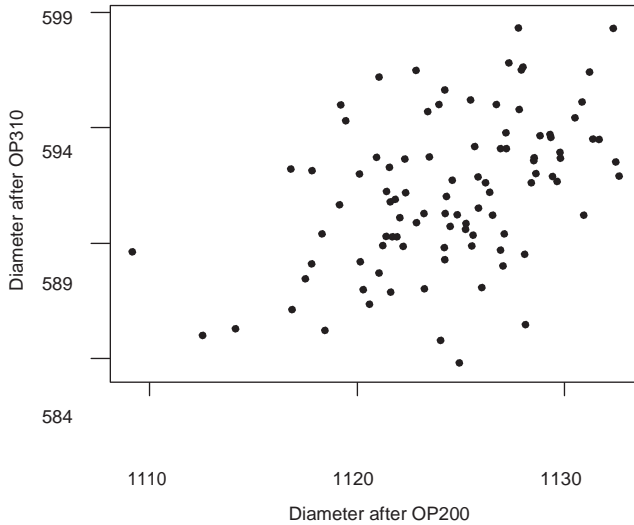
Figure S11.8   Plot of final diameter versus diameter after Operation 200.

Operation 270 fixed. We can compare this standard deviation to the baseline value 3.32 to determine that the input (in this case the diameter after Operation 270) is a dominant cause.

In the same example, we can also fit a regression model using the diameter after Operation 200 as the input. From MINITAB, we get the plot in Figure S11.8 and the following numerical results:

Regression Analysis: diameter after OP310 versus diameter after OP200

The regression equation is
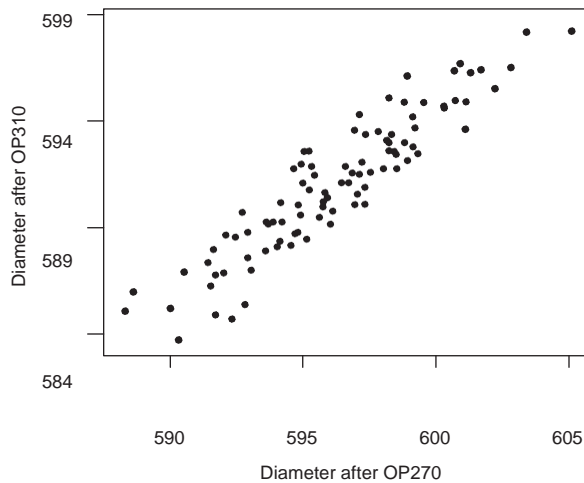**diameter after OP310 = 225 + 0.325 diameter after OP200**

Predictor	Coef	SE Coef	T	P
Constant	225.12	72.72	3.10	0.003
diameter	0.32542	0.06467	5.03	0.000

S = 2.862   R-Sq = 21.2%   R-Sq(adj) = 20.4%

From these results, it is clear that the relationship between the diameter after Operation 200 and the final diameter (after Operation 310) is much weaker than between the diameter after Operation 270 and the final diameter. In the second regression model, the estimate of the residual variation, $s = 2.862$, is close to the baseline standard deviation 3.32. If we were to hold the diameter at Operation 200 fixed, we would see little change in the variability after Operation 310. We can rule out all causes that act up to and including Operation 200 as possible dominant causes of final diameter variation.

Comments

In the context of variation transmission investigations, we do not recommend fitting a regression model to try to explain the final output as a function of two or more intermediate output measurements. Strong correlation between the inputs in the regression model

(that is, the intermediate diameter measurements) makes interpretation of the results difficult. The residual variation (given by *s* in the MINITAB results) would estimate the residual standard deviation if we could hold all the inputs fixed.

We use regression models in Chapter 12 and discuss them further in the supplement to Chapter 12.

S11.4 OTHER COMPONENT-SWAPPING INVESTIGATIONS

In Chapter 11, we propose what we believe is the preferred plan, in most circumstances, for comparing component families. Dorian Shainin first proposed an alternative component-swapping plan in 1956. More recently the method has been promoted and discussed by Shainin and Shainin (1988a), Amster and Tsui (1993), Ingram (1993), and Bhote and Bhote (2000). Cotter (1979) provides some motivation for the Shainin type of component-swap experiments.

We consider the four-stage Bhote and Bhote (2000) version here. The four stages are confirmation, elimination, capping, and analysis. There are complicated rules for dealing with all possible eventualities that we do not repeat here.

To start, we choose two units with extreme and opposite performance. In the confirmation stage we check that repeated disassembly and reassembly does not affect performance. We use this stage to confirm that a dominant cause lies in the component rather than the assembly family of causes. This matches the procedure described in Section 10.4.

We start the elimination stage by ranking the components in descending order of likely importance. Then, we swap components between the two extreme parts, one at a time, starting with the most likely. After each swap, we determine the performance of the two assemblies. If the low and high output assemblies remain unchanged, we swap back the particular component and move on to swap the next most likely component. If the performance of the two units follows the swapped component, we have found the component family that is the home of a dominant cause and we can move to the final stage. If the performance partially moves with swapped component, we continue swapping with that particular component identified as part of a dominant cause involving two or more inputs.

When a dominant cause involving two components has been tentatively identified, we move to the capping stage. The capping run involves simultaneously swapping the two components identified as important. If the performance follows the double swap, the dominant cause is an interaction between two inputs with one in each component.

The final stage is called the analysis stage. In the analysis stage, no more swaps are conducted. The existing results are summarized graphically in $2 \cdot 2$ tables to determine the magnitude and direction of the effects.

Comments

With the Bhote and Bhote (2000) procedure, the length of the search depends on engineering judgment regarding which components are likely to be most important. Also, Amster and Tsui (1993) provide some examples of when the Bhote and Bhote method would yield incorrect conclusions.

As suggested by Tippett (1934), Taguchi (1987), and Parmet and Steinberg (2001), we could also conduct a component-swapping experiment without using the idea of leveraging; that is, by choosing components from regular production. This strategy requires considerably more pairs of parts and a greater number of disassemblies and reassemblies.

The Bhote and Bhote (2000) component swap procedure matches the factorial component swap experiment as described in the next section, where we do not conduct all swaps once we believe the results are clear.

We believe iteratively forming two groups of components and using the method of elimination is a more efficient strategy (see Section 11.3).

S11.5 COMPONENT-SWAPPING WITH THREE GROUPS OR SUBASSEMBLIES

Here we look at component swap investigations with three groups. This may be useful if the assembly naturally divides into three subassemblies. We suppose for the moment that the three groups make up the entire assembly and that we have not yet eliminated any components as the home of the dominant cause.

We consider a component swap investigation as a factorial experiment, regardless of the number of groups. See Chapter 13 for a full discussion of factorial exeriments.We label the groups of components based on whether they come from the assembly that originally gave a low or high output. For instance, a component swap investigation with two groups, as presented in Chapter 11, corresponds to a (full) factorial experiment with two inputs, two levels for each input, and four treatments:

Treatment	G1	G2
1	Low	Low
2	Low	High
3	High	Low
4	High	High

Treatments 1 and 4 correspond to the two original assemblies and treatments 2 and 3 correspond to swapping the components in G1 (or G2) between the two assemblies.

With three groups of components, we disassemble and reassemble to get all eight possible combinations:

Treatment	G1	G2	G3
1	Low	Low	Low
2	Low	Low	High
3	Low	High	Low
4	Low	High	High
5	High	Low	Low
6	High	Low	High
7	High	High	Low
8	High	High	High

Treatments 1 and 8 correspond to the original assemblies. Since a swap of components corresponds to two runs of the experiment, the proposed plan is equivalent to conducting three swaps.

We can summarize the results of component swap experiments using two $2 \cdot 2$ tables, similar to those shown in Chapter 11. The eight cells correspond to the eight treatments.

Had we used three groups in the power window buzz example described in Chapter 11, we would have to start:

		G2	
G1	**High**	**Low**
High	7	
Low		
	G3 High	

		G2	
G1	**High**	**Low**
High		
Low		1
	G3 Low	

We fill the other six cells of the table with the results from three swaps. We might get:

		G2	
G1	**High**	**Low**
High	7	1
Low	7	2
	G3 High	

		G2	
G1	**High**	**Low**
High	7	1
Low	6	1
	G3 Low	

Here, since the results in the two tables are similar, we can eliminate all the components in G3 and interpret the result in either table as in Chapter 11. In this example, the dominant cause acts in G2.

To help us make the correct interpretation, we can generate three pairs of tables similar to the previous set using a different group to stratify the pair of tables. If none of the three pairs of tables gives similar results across the two $2 \cdot 2$ tables, the dominant cause involves a component (or more than one) in each of the three groups. This is unlikely.

Chapter 12 Supplement

Investigations Based on Single Causes

S12.1 MATRIX SCATTER PLOTS AND DRAFTSMAN PLOTS

To analyze the results of a group comparison investigation, we create plots of each of the inputs stratified by the two possible values of the output. This can be done using a series of box plots, where we select the option to plot the individual data points. To look simultaneously at two inputs, we create scatter plots of one input versus the second using a different plotting symbol for each of the two groups as defined by the output value. The number of such plots can be large.

In examples such as the window leak problem described in Chapter 12, where the number of inputs is less than 10, we can use a matrix scatter plot (see Appendix C) of all pairs of the (continuous) inputs. We give such a plot in Figure S12.1, where leakers are denoted by circles and nonleakers by plus signs. The plot in the right panel of Figure 12.2 is in the lower left corner of the matrix scatter plot in Figure S12.1. If the number of plots in the matrix scatter plot is too large, we find it difficult to see any patterns.

In input/output investigations, we want to examine scatter plots of the (continuous) output versus each of the inputs. In MINITAB, we can use a draftsman plot (see Appendix C) to automatically create all the desired scatterplots. Consider the crossbar dimension example described in Chapter 12. Figure S12.2 gives the draftsman plot with crossbar dimension and the five inputs. The left panel of Figure 12.3 is given in the middle of the draftsman plot.

With a large number of inputs, the aspect ratio of the draftsman plot makes interpretation difficult.
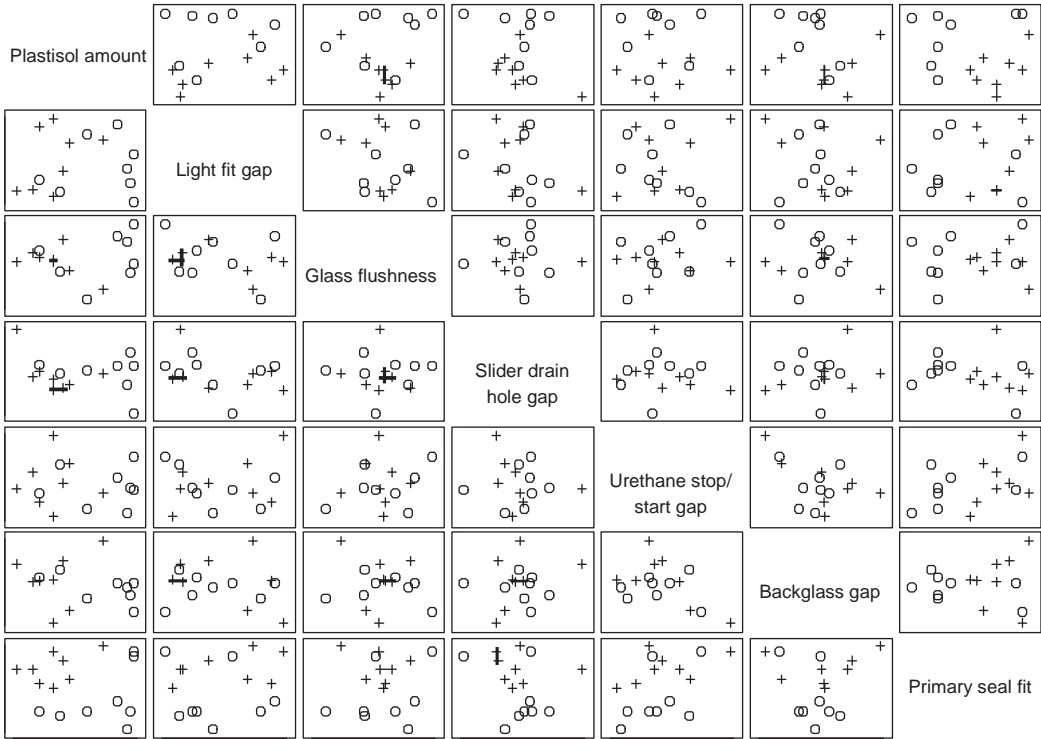
Figure S12.1   Matrix scatter plot for window leaks group comparison (leakers are denoted by o and nonleakers by +).

In output/input investigations with large numbers of inputs, examining all the scatter plots can be tedious, even when using a draftsman plot. In this context, another useful tool in MINITAB is best subsets regression (see Appendix E). With best subsets regression, we can ask MINITAB to fit all possible regression models involving a single input (or larger numbers of inputs, but that is not needed in this context). The results rank the possible models by the estimated residual standard deviation *s*. For illustration, results from using the best subsets regression routine for the crossbar dimension example are:

Best Subsets Regression: dimension versus die temp, nozzle temp, ...

Response is dimension

					d	n	b	h	c
					i	o	a	y	a
					e	z	r	d	v
						z	r	r	i
					t	l	e	a	t
					e	e	l	u	y
					m			l	
Vars	**R-Sq**	**R-Sq(adj)**	**C-p**	**S**	p	t	t	i	p

1	78.9	78.3	2.9	0.25439	X			
1	14.8	12.6	121.0	0.51123			X	
1	8.2	5.8	133.1	0.53061			X	
1	0.8	0.0	146.9	0.55174		X		
1	0.1	0.0	148.1	0.55353	X			

We see that barrel temperature is a dominant cause, because the residual standard deviation for the model with barrel temperature is so small relative to the baseline value of 0.46.

In the crossbar dimension example, there are only five inputs. Using best subsets regression is unnecessary, since we get the same information from Figure S12.2. One danger with relying on the best subsets regression is that the models are ranked according to how well a linear model fits the data. As a result, nonlinear relationships may be missed. We strongly recommend that you look at all of the scatter plots of the output versus the selected inputs.
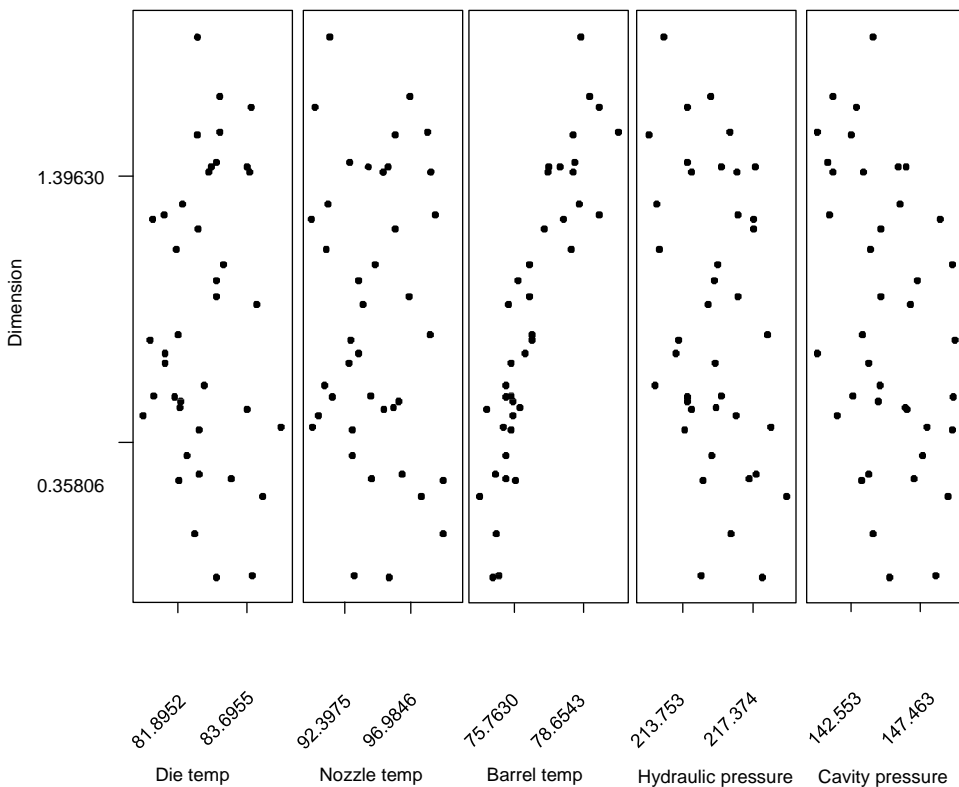
Figure S12.2   Draftsman plot for crossbar dimension regression investigation.

S12.2 GROUP COMPARISON VERSUS PAIRED COMPARISON

Bhote and Bhote (2000) promote a plan and analysis tool they call *paired comparison,* sometimes referred to as *group comparison.* The tool is similar to the proposed group comparison method discussed in Section 12.1. Like group comparison, the goal of paired comparison is to compare the effects of a number of inputs for parts with binary output.

There are some important differences in the two methods. First, in the paired comparison, a hypothesis test based on end-count is recommended to compare the two groups of parts. As discussed in the supplement to Chapter 10, we do not believe that formal hypothesis tests are necessary or appropriate when searching for a dominant cause of variation. Significant differences do not (necessarily) correspond to dominant causes.

A second difference is that paired comparison involves an arbitrary pairing of the parts, one from each group. In the search for the dominant cause, pairing makes little sense. Since the causes of variation are unknown, determining how to pair is problematic. In addition, the goal of the comparing extreme parts is to help identify a dominant cause of variation in the output. If we pair in such a way that a dominant cause is held constant within the pairs, the analysis will fail to identify the dominant cause. With artificial pairing, the analysis results depend on the way pairs are produced.

We do not recommend paired comparison as an alternative to group comparison.

S12.3 REGRESSION EXTENSIONS

Regression is a flexible analysis tool that can be extended and generalized in many ways. We consider a few useful extensions here.

Regression models can be used to analyze the designed experiments we encounter in Chapter 13. They are especially useful if, for example, we lose a run and the resulting experimental data are not balanced. The general linear model ANOVA option in MINITAB is based on fitting a regression model. Regression models can also be useful when assessing the feasibility or implementing a feedforward control.

Multiple Regression Model

We can fit regression models with more than one input, although in a search for a dominant cause, we do not, as a general rule, look at inputs simultaneously. If we think there is a dominant cause that involves two or more inputs, we may fit a model with all inputs or all pairs of inputs using best subsets regression. We may also add terms to a model if we suspect there is an nonlinear input/output relationship, as in the manifold sand scrap example in Chapter 12.

We are interested in understanding the remaining output variation if the input or inputs in the regression model are held fixed. This is given by the residual standard deviation. Whenever we consider including more than one input in a regression model, we need to consider the form of the model. We call this *model building.* An experienced analyst should conduct regression model building. Modeling assumptions can be checked using residual plots based on the estimated residuals from a regression model (Montgomery et al., 2001; Box et al., 1978). The estimated residuals are defined as the difference between the observed output value and the value predicted by the regression model.

Categorical Inputs and Indicator Variables

Suppose we have a categorical input *x,* such as a machine with three values: A, B, C or 0, 1, 2. We cannot use this input directly in a regression model. Instead, we replace *x* with two indicator variables. Let $x_B = 1$ if the part is from machine B; otherwise, let $x_B = 0$. Also, let $x_C = 1$ if the part is from machine C; otehwise, let $x_C = 0$. We know that if $x_B = 0$ and $x_C = 0$, then the part is from machine A. Now we include x_B and x_C in the regression model.

In general, we use $k - 1$ indicator variables if there are k categories. (See Montgomery et al., 2001, for more detailed discussion.)

Logistic Regression for Binary Outputs

Regression analysis can be relatively easily extended to handle binary outputs. Logistic regression is one example (Hosmer and Lemeshow, 2000). See the MINITAB regression menu. Logistic regression can be a useful alternative or complement to the graphical group comparison analysis described in Chapter 12. Logistic regression analysis may be helpful when the group sizes are large.

Other Regression Extensions

Another useful extension is regression with count data (Cameron and Trivedi, 1998). Both logistic regression and regression with count data are special cases of generalized linear models (McCullagh and Nelder, 1989; Hamada and Nelder, 1997).

Chapter 13 Supplement

Verifying a Dominant Cause

S13.1 UNDERSTANDING REPEATS AND REPLICATES

There is great confusion about the difference between replicates and repeats. Replicates are different runs with the same treatment. Repeats are different parts within a run. The measured output values on repeats vary due to causes that act within a run. Between replicates, the measured output values vary because of causes that act both within runs and from run to run.

In traditional applications of experimental design, the experimenters do not know the within-treatment variation and hence need an estimate from within the experiment itself. If there are no replicates and they calculate an estimate based on the variation of the repeats within each run, they are almost certain to underestimate the within-treatment variation. As a consequence, they are likely to conclude a suspect cause is dominant when it is not.

The same problem occurs when we use an experiment to assess the effects of changing one or more fixed inputs in Chapter 15, where we search for an adjuster to move the process center. We start the analysis in Chapter 13 (and in later uses of factorial experiments) by plot- ting the output values by treatment. To make this plot, we find it convenient to store the data with one row for each repeat. The plot includes output values from both replicates and repeats. We use the plots to look for large effects and promising treatments. We include horizontal lines showing the full extent of variation on the plot to get an assessment of the performance of the process with each treatment.

In the analysis for full factorial designs (and later fractional factorial designs—see Chapter 15), we use ANOVA to calculate, plot, and rank the effects of the inputs and their interactions. These calculations and plots are correct as long as the experimental design is (close to) balanced in terms of replicates and repeats (that is, there are the same number of runs per treatment and the same number of repeats per run) as we recommend. However, because of the way we have stored the data, the ANOVA program cannot separate replicates from repeats and hence the internal estimate of the within-treatment variation, derived from the residual variation, is too small. As a consequence, the formal hypothesis tests given in the output may to be misleading.

S13.2 RANDOMIZATION, REPLICATION, AND BLOCKING

Suppose we plan to verify a single suspect as a dominant cause. In making the decision to verify, we are implicitly or explicitly concerned about some other cause, identified or not, that acts in the same family as the suspect. We want to conduct the experiment to confirm the suspect and rule out any other possibility.

We say that other causes in the same family as the suspect are *confounded* with the suspect. That is, the pattern of behavior that we have seen in the output (in the observational investigations used in the search for a dominant cause) can be explained by the suspect or by some other member in the same family. In the verification experiment, we want to eliminate confounding.

Blocking is one tool that we can use to prevent confounding. If we hold the other members of the family fixed and change only the suspect, and if the output varies over its full extent, we know that the suspect is the dominant cause. For example, suppose the suspect is in the hour-to-hour family; that is, it changes relatively slowly. We can rule out other members of the family by changing the suspect quickly, say from one part to the next. From part to part we know that the other members of the hour-to-hour family are not changing much. This is blocking. By holding other possible causes fixed and varying the suspect, we can verify that the suspect is a dominant cause. Note that in other contexts blocking is used to increase the precision of the conclusions.

Sometimes we need more than blocking because we do not know how to hold all of the other causes in the suspect family fixed. For example, in the truck water leak example, a suspect was identified as an interaction between a gap and the plastisol application. These suspect inputs vary from truck to truck in a haphazard pattern. Many other inputs associated with any of the body components, the window, and the assembly vary in the same way. These causes had not been ruled out in the group comparison discussed in Chapter 12. We can use blocking to hold the body components fixed. That is, we use one truck and change the plastisol application and the gap with a different window. Using a single truck keeps all of the body components fixed. However, the causes in the assembly and window family change as we change the gap and plastisol levels.

We use randomization and replication to reduce the risk of confounding. For example, we select a number of trucks, say 10, and for each truck we change the plastisol and gap as described previously. Now we have 10 blocks. For each truck, we apply the four treatments in a random order. This random ordering over the 10 trucks will, with high probability, break the link between changes in the treatments and changes in other causes in the assembly family.

We use 10 different pairs of windows (one with a large gap and one with a small gap) to change the gap. There are other causes in the window family that match changes in the gap. We may be fooled in identifying the gap as a dominant cause when, in fact, there is some other characteristic of the window that is the cause. We accept this risk here because it is difficult to change the gap except by selecting the appropriate windows.

Randomization is likely to be effective only if there is sufficient replication—we recommend at least three runs at each level of the suspect and preferably five or more runs. Repeats are not helpful in reducing the risk of confounding, since their order cannot be randomized across runs.

S13.3　VARIABLES SEARCH

Bhote and Bhote (2000) highly recommend variables search for finding a dominant cause if there are five or more suspects. With fewer suspects, they recommend a full factorial experiment. Variables search competes with fractional factorial designs—see Chapter 15 and its supplement (also Shainin and Shainin, 1988).

　　With variables search, we conduct an experimental investigation using a procedure similar to component swap, as described in Chapter 11. We start with:

- A list of suspects ranked in order of expected importance

- Two levels for each suspect chosen at the extremes of their usual range of values

- Two treatment combinations that produce output that spans most of the full extent of variation

　　In the next stages, we swap levels of the suspects in a series of experimental runs starting with the highest-ranked suspect. The variables search method focuses on finding high-order interactions.

　　We cannot recommend variable search because:

- The length of the search depends on how well we order the suspects.

- All suspects not yet ruled out must be held fixed at one of the two levels for each run of the experiment. With a large number of suspects, all of which normally vary, this can be a daunting task.

- Substantial effort is required to find the initial two treatments that produce output that spans the full extent of variation. No direction is given on how to find these treatments.

　　See Ledolter and Swersey (1997a) for a critical view of variables search. We believe there is no good experimental plan when the list of suspects is long. Rather, we recommend continuing to use the method of elimination until the number of suspects is small.

Chapter 15 Supplement

Moving the Process Center

S15.1 FRACTIONAL FACTORIAL EXPERIMENTS

We give a brief description of the planning and analysis of fractional factorial experiments. We discuss only two-level fractional factorial designs and rely heavily on MINITAB. For a more complete description see Montgomery (2001); Box, Hunter, and Hunter (1978); or Wu and Hamada (2000). See Appendix F for details on how to set up the designs and conduct the analysis in MINITAB.

We use fractional factorial designs to estimate the effects of a number of inputs (usually four or more) simultaneously when the number of experimental runs is limited. In the language of factorial designs, inputs are called *factors* and the values used in the experiment are called the *levels* of the inputs. A *treatment* or a *treatment combination* is a set of particular levels, one for each input, that can be used to run the process.

We find it helpful to present the experimental plan using a code of –1 and +1 for the two levels of each input. To illustrate, consider the brake rotor balance verification experiment discussed in Chapter 13. This experiment was a full factorial design but we use it here to show how the coding works. We denote the inputs by the letters *A, B,* and *C,* as shown in Table S15.1.

We can represent any treatment combination using the code for each input. We list the eight possible treatments in Table S15.2. We call this plan a 2^3 design, because there are three inputs each at two levels and there are $2^3 = 8$ treatments used in the experiment.

Table S15.1   Input levels and coding for brake rotor verification experiment.

Input (factor)	Label	Low level (–1)	High level (+1)
Tooling	A	Old tooling	New tooling
Core position	B	Offset	Nominal
Thickness variation	C	30 thousandths	Nominal

Table S15.2   A list of all possible treatments in a 2^3 design.

Treatment	A	B	C
1	−1	−1	−1
2	−1	−1	+1
3	−1	+1	−1
4	−1	+1	+1
5	+1	−1	−1
6	+1	−1	+1
7	+1	+1	−1
8	+1	+1	+1

We generate the additional columns in Table S15.3 by multiplication. First we look at how this works and then we provide an interpretation. To get the AB column, for example, we multiply the corresponding plus or minus ones for each row in the table using the *A* and *B* columns. We repeat this calculation to get four additional columns corresponding to the products AB, AC, BC, and ABC. Note that we have put the columns in a standard order. We call the seven columns in Table S15.3 the *contrast matrix* for the 2^3 design. Taguchi (1987) calls this matrix an orthogonal array and labels it L_8.

We can calculate all the effects using the contrast matrix and the output from the experiment. In Table S15.4, we add one extra column that gives the output value for each treatment in the experiment. Here we assume there is a single run of the experiment corresponding to each treatment. With replication (that is, more runs), we replicate the rows in the contrast matrix. Note that the treatments in the contrast matrix are written in a specific order and the runs in the experiment likely occur in a different random order.

Table S15.3   Contrast matrix for the 2^3 design.

Treatment	A	B	AB	C	AC	BC	ABC
1	−1	−1	+1	−1	+1	+1	−1
2	−1	−1	+1	+1	−1	−1	+1
3	−1	+1	−1	−1	+1	−1	+1
4	−1	+1	−1	+1	−1	+1	−1
5	+1	−1	−1	−1	−1	+1	+1
6	+1	−1	−1	+1	+1	−1	−1
7	+1	+1	+1	−1	−1	−1	−1
8	+1	+1	+1	+1	+1	+1	+1

Table S15.4 Contrast matrix for the 2^3 design with output data.

Treatment	A	B	AB	C	AC	BC	ABC	Weight
1	−1	−1	+1	−1	+1	+1	−1	0.56
2	−1	−1	+1	+1	−1	−1	+1	0.17
3	−1	+1	−1	−1	+1	−1	+1	0.44
4	−1	+1	−1	+1	−1	+1	−1	0.08
5	+1	−1	−1	−1	−1	+1	+1	1.52
6	+1	−1	−1	+1	+1	−1	−1	0.37
7	+1	+1	+1	−1	−1	−1	−1	1.34
8	+1	+1	+1	+1	+1	+1	+1	0.03

Recall from Chapter 13 that the *main effect* of a particular input is the difference between the average output over runs at the high level and the average output over runs at the low level of the input. In the example, the main effect of *A* is

$$\frac{1.52 + 0.37 + 1.34 + 0.03}{4} - \frac{0.56 + 0.17 + 0.44 + 0.08}{4} \tag{S15.1}$$

$$= \frac{-0.56 - 0.17 - 0.44 - 0.08 + 1.52 + 0.37 + 1.34 + 0.03}{4} = 0.5025$$

We have changed the order of the terms in the sum of the second expression in this calculation to match the treatment order in Table S15.4. In terms of the contrast matrix, to get the numerator of Equation (S15.1), we apply the signs from column *A* to the data (that is, multiply the two columns element by element) and add.

To find an *interaction effect,* we compare the main effects for one input at the two different levels of a second input. For example, to look at the AC interaction, we have

$$\text{when C} = +1, \text{ the effect of A is } \frac{0.37 + 0.03}{2} - \frac{0.17 + 0.08}{2}, \text{ and}$$

$$\text{when C} = -1, \text{ the effect of A is } \frac{1.52 + 1.34}{2} - \frac{0.56 + 0.44}{2}$$

The interaction effect is half the difference. That is,

$$\left[\frac{0.37 + 0.03}{2} - \frac{0.17 + 0.08}{2} \right] - \left[\frac{1.52 + 1.34}{2} - \frac{0.56 + 0.44}{2} \right]$$

$$= \frac{0.56 - 0.17 + 0.44 - 0.08 - 1.52 + 0.37 - 1.34 + 0.03}{2} = -0.4275 \tag{S15.2}$$

**Table S15.5** Contrast matrix for the 2^3 design with estimated effects.

Treatment	A	B	AB	C	AC	BC	ABC	Weight
1	−1	−1	+1	−1	+1	+1	−1	0.56
2	−1	−1	+1	+1	−1	−1	+1	0.17
3	−1	+1	−1	−1	+1	−1	+1	0.44
4	−1	+1	−1	+1	−1	+1	−1	0.08
5	+1	−1	−1	−1	−1	+1	+1	1.52
6	+1	−1	−1	+1	+1	−1	−1	0.37
7	+1	+1	+1	−1	−1	−1	−1	1.34
8	+1	+1	+1	+1	+1	+1	+1	0.03
Effect	0.503	−0.182	−0.078	−0.803	−0.428	−0.033	−0.048	

Again we have reordered the terms in the numerator of the second expression in Equation (S15.2) and you can see that to get this sum, we apply the signs from the AC column of the contrast matrix to the data and add.

All main effects and interactions can be found by applying the signs from the appropriate column of the contrast matrix to the data, adding, and then dividing by half the number of runs. Every effect is the difference of two averages, where each average includes exactly half the data. We show all of the effects in Table S15.5, where we add an extra row to Table S15.4. In the analysis, we look for large (positive or negative) effects. For two-level factorial and fractional factorial designs, MINITAB will calculate and rank the absolute value of the effects in a Pareto diagram. We can use this diagram to isolate large effects. The effects in the example are plotted in Figure S15.1. We see that the main effects of *C* and *A* are relatively large, as is the interaction effect AC. Since there is evidence of interaction, we look at the effects of *A* and *C* simultaneously.

We now use the contrast matrix to explain fractional factorial designs. As an example, we start with the contrast matrix for a 2^4 design, that is, a design with four inputs each at two levels. We give the contrast matrix in Table S15.6. Note the order of the columns.
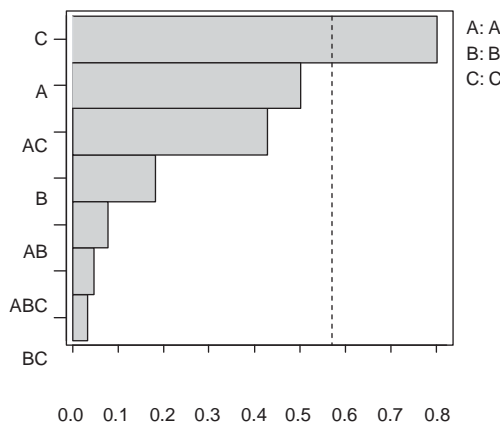
Figure S15.1 Pareto chart of effects in rotor balance experiment.

Table S15.6   Contrast matrix for a 2^4 design.

Treatment	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
1	−1	−1	+1	−1	+1	+1	−1	−1	+1	+1	−1	+1	−1	−1	+1
2	−1	−1	+1	−1	+1	+1	−1	+1	−1	−1	+1	−1	+1	+1	−1
3	−1	−1	+1	+1	−1	−1	+1	−1	+1	+1	−1	−1	+1	+1	−1
4	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1
5	−1	+1	−1	−1	+1	−1	+1	−1	+1	−1	+1	+1	−1	+1	−1
6	−1	+1	−1	−1	+1	−1	+1	+1	−1	+1	−1	−1	+1	−1	+1
7	−1	+1	−1	+1	−1	+1	−1	−1	+1	−1	+1	−1	+1	−1	+1
8	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1	+1	−1
9	+1	−1	−1	−1	−1	+1	+1	−1	−1	+1	+1	+1	+1	−1	−1
10	+1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1	+1	+1
11	+1	−1	−1	+1	+1	−1	−1	−1	−1	+1	+1	−1	−1	+1	+1
12	+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1	+1	+1	−1	−1
13	+1	+1	+1	−1	−1	−1	−1	−1	−1	−1	−1	+1	+1	+1	+1
14	+1	+1	+1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1
15	+1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	−1	−1	−1
16	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

Suppose we want to conduct an experiment and we can afford only 16 runs. With four inputs, *A, B, C,* and *D,* we can run all treatments. The four columns labeled *A, B, C,* and *D* give the levels of the four inputs for each treatment. If, however, we have a fifth input, *E,* there are 32 possible treatments and we have the resources to conduct an experiment using only half of the possible treatments. The question is, which half? The resulting design is often referred to as a 2^{5-1} design since we are using the half fraction (2^{-1}) design with five inputs at two levels each.

One possibility is to assign *E* to the column denoted ABCD in the contrast matrix in Table S15.6. Then, we can read the 16 treatments we will use in the experiment by looking at the five columns labeled *A, B, C, D,* and *E.* For example, with treatment 7, we have the levels *A:* –1, *B:* 1, *C:* 1, *D:* –1, *E:* 1. With this choice for *E,* exactly half the 16 runs have *E* at the high level, that is, $E = +1$, and the other half have *E* at the low level, that is, $E = –1$. The same is true for the other four inputs.

To see the consequences of this choice, suppose we carry out the experiment and measure the output for each of the selected 16 treatments. We can calculate the main effects of each of the five inputs by applying the appropriate column of plus and minus ones to the data, adding and dividing by eight. We carry out the same calculation for every column.

To calculate the effect of the four-input interaction ABCD and the main effect for *E,* we use the same column of the contrast matrix and we say these two effects are *confounded.* If the calculated effect is large, this may be due to the change in input *E* or the four-input interaction involving inputs *A, B, C,* and *D.* We cannot separate these two effects using the data. However, we usually assume that three- and four-input interaction effects are small and so, in this case, we would attribute the large effect to the main effect of *E.*

You should be wondering by now where to find the interactions involving *E.* To find the two-input interaction AE, we multiply the *A* and *E* columns as before. We found the column of signs for *E* by multiplying the corresponding columns for *A, B, C,* and *D.* We use the convenient notation $E = ABCD$. Hence we have

$$AE = A \cdot ABCD = BCD$$

since, if we multiply column *A* by itself, we get a column of plus ones, which has no effect on the overall product. In other words, the interaction effect AE is confounded with the three-input interaction BCD. If this column produces a large effect, we cannot tell if this is due to AE or BCD. Again we will attribute the effect to the lower-order interaction—in this case, the two-input interaction AE.

You can quickly find all other two-input interactions involving *E.* Every column of the contrast matrix corresponds to two effects (since we used a half fraction design). MINITAB will produce a list of the confounded effects for the design. For the example, the list of confounded effects (MINITAB calls these effects aliases) is:

Design Generators: E = ABCD

Alias   Structure

I + ABCDE

A + BCDE
B + ACDE
C + ABDE
D + ABCE
E + ABCD
AB + CDE
AC + BDE
AD + BCE
AE + BCD
BC + ADE
BD + ACE
BE + ACD
CD + ABE
CE + ABD
DE + ABC

Any effects in the same row, linked by a + or −, are confounded. See Appendix F for instructions on how to create this list. You might also wonder what happens if we start by assigning *E* to a different column, say *E = ABC*. Since the ABC column gives the level of *E* for each run, we now have a different set of 16 treatments. We use MINITAB to produce the list of confounded effects. This time, we ask that four- and five-input interactions be suppressed. The confounded effects with this design are:

Design Generators: E = ABC

Alias Structure (up to order 3)

I
A + BCE
B + ACE
C + ABE
D
E + ABC
AB + CE
AC + BE
AD
AE + BC
BD
CD
DE
ABD + CDE
ACD + BDE
ADE + BCD

**Table S15.7**   Design resolution interpretation.

Resolution	Meaning
III	Main effects are confounded with two-input interaction effects
IV	Main effects are confounded with three-input or higher-order interactions, and two-input interaction effects are confounded with other two-input interaction effects
V	Main and two-input interaction effects are confounded only with three-input or higher-order interactions

With this choice, we see that main effects are confounded with three-input interactions and, more important, pairs of two-input interactions such AB and CE are confounded. This plan is less desirable than the design based on assigning *E* to ABCD.

We say that the design with *E* = ABC has resolution IV, because at least one pair of two-input interactions is confounded. If we assign *E* to column AB, then a main effect is confounded with a two-input interaction and the design has resolution III. The first design discussed earlier with *E* = ABCD has resolution V. The higher the resolution, the less likely it is that important effects will be confounded. We summarize the meaning of *resolution* in Table S15.7.

For a fixed number of runs, the greater the number of inputs, the lower the highest possible resolution (see Figure S15.2).

In using MINITAB, we normally assign the letters A, B, C, and so on to the candidates and +1 and –1 to the two levels for each. It does not matter which letter is assigned to each candidate. Then we ask MINITAB to generate the design, a worksheet of the treatments selected for the experiment. Sometimes we want to ensure that one particular treatment is included in the experiment—for example, the current process levels of the candidates. In
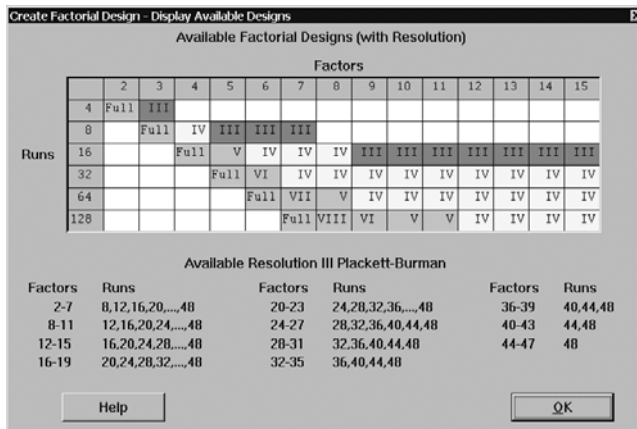
Figure S15.2   Two-level factorial experiments available in MINITAB.

this case, we generate the design first and then assign the letters and levels so that one of the treatments in the design has the required levels.

We recommend an analysis based on a plot of the output by treatment number, the Pareto plot of the effects, and main effect and interaction plots. MINITAB will produce the table of confounded effects to help interpret the important effects.

Despite the confounding, factorial experiments are useful because of the *scarcity of effects principle,* which states that when there are many inputs, there are likely only a few large effects. These effects are commonly the main effects and low-order interactions.

Resolving Confounded Effects

In fractional factorial experiments, all effects are confounded with at least one other effect. In some circumstances, to strengthen the conclusions from the experiment, we conduct additional runs with new treatments to break the confounding for important effects. This is especially necessary if the confounded effects are interactions of the same order. Simple follow-up experiments that use *fold-over* are covered in Montgomery (2001). A more sophis- ticated method is given in Meyer et al. (1996). Typical follow-up experiments involve 8 or 16 runs, depending on the number of inputs and degree of fractionation used in the initial design.

Use of Center Points

If the inputs are quantitative (that is, speed, voltage, amount, and so on), we have the option of adding center points to the design. The level of each input at the center point is the average of the high and low values. We illustrate the center point for a design with two inputs at two levels in Figure S15.3. MINITAB allows the addition of center points to factorial or fractional factorial designs with quantitative inputs.

When we add a center point to the design, we can check to see if there is curvature in the main effect of one or more of the inputs. The curvature may be important if we plan to use the input as an adjuster. The addition of the center point does not change the confounding of the effects in a fractional design. For more detail, see Montgomery (2001).
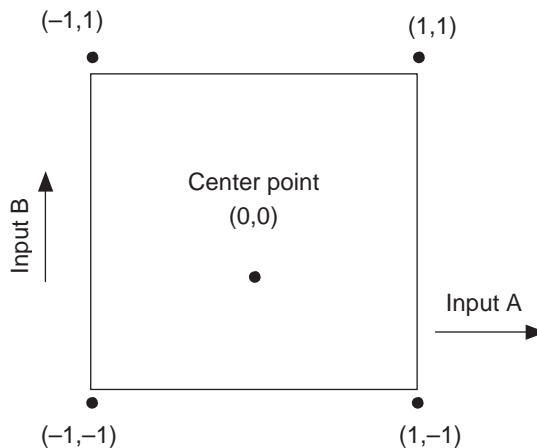
Figure S15.3   A 2^2 design with center point.

Summary

Most issues in the planning of fractional factorial experiments are the same as for full factorial experiments. We need to

- Choose the k inputs and their levels

- Define a run

- Use MINITAB to select the design (that is, choose the treatments, a fraction of all of the possibilities), mindful of the tradeoff between resolution and number of runs

- Randomize the run order

The choice of fraction is based on cost considerations and concerns about confounding. In fractional factorial experiments, every effect will be confounded with one or more other effects. We choose the design with the highest possible resolution to minimize problems associated with confounded effects. That is, we choose a design so that main effects and two-input interactions are confounded with high-order interactions as much as possible. If we find large effects, we typically attribute the effect to the main effect or two-input interaction.

MINITAB can help plan and analyze fractional factorial experiments. The complete confounding structure for any fractional factorial design is given by MINITAB. Your task is to assign the inputs and levels to the generic letters *A, B, C,* and so on, and the codes –1 and +1 to produce the design.

Chapter 16 Supplement

Desensitizing a Process to Variation in a Dominant Cause

S16.1 MATHEMATICAL REPRESENTATION OF PROCESS DESENSITIZATION

We can demonstrate process desensitization using a regression model. In the model given by Equation (S16.1), Y is the value of the output, X is the level of the dominant cause, c is the level of the desensitization candidate (here assumed to be quantitative), and the term R represents the (small, since X is a dominant cause) residual effect of all other varying inputs.

$$Y = b_0 + b_1X + b_2c + b_3Xc + R$$

which can be rewritten as

$$Y = b_0 + (b_1 + b_3c)X + b_2c + R \qquad \text{(S16.1)}$$

The levels of other fixed inputs determine the coefficient b_0. For a given level c of the candidate, the coefficient $b_1 + b_3c$ represents the effect of the cause X on the output. The standard deviation of the output is

$$sd(Y) = \sqrt{(b_1 + b_3c)^2 \, sd(X)^2 + sd(R)^2} \qquad \text{(S16.2)}$$

assuming that the effect of the other causes, R, varies independently of the dominant cause. In Equation (S16.1), we have modeled the interaction between the cause X and the candidate

by the product term b_3Xc. We see from Equation (S16.2) that the standard deviation of the output *sd(Y)* is sensitive to the level of the candidate if b_3 is not equal to zero. If we can set $c = -b_1 / b_3$, then the effect of the dominant cause *sd(X)* is completely eliminated and *sd(Y)* is reduced to *sd(R)*.

In practice, such a simple model rarely describes the relationship amongst the inputs and output exactly. As well, we must estimate the coefficients b_1 and b_3 without error before we can achieve the optimal reduction of variation.

If the candidate is binary (for example, supplier 1 or 2), then *c* in Equation (S16.1) can take only one of two possible values—for example, –1 or +1—and we choose the level of the candidate with the smaller value of $(b_1 + b_3 c)^2$.

S16.2 FRACTIONAL FACTORIAL EXPERIMENTS FOR DESENSITIZATION

In this section, we give a brief description of fractional factorial experiments specifically designed for desensitization of the output to the effects of a known dominant cause. See the supplement to Chapter 15 for more general information about fractional factorial designs.

We have two types of inputs: a dominant cause and the candidates. Taguchi calls these the noise and control factors, respectively. The main goal of the experiment is to look for interactions between the cause and the candidates.

We recommend that the two levels for the dominant cause be set at the extremes of the normal range of variation. If there is more than one dominant cause, then we suggest creating a pseudo-input with two levels to generate the full extent of variation in the output when the candidates are set at their original values. There is little need to use a full factorial design for the causes since we already know their effects. See the refrigerator frost buildup case in Chapter 16 for an example.

If there are four or more candidate inputs, we can use a two-level fractional factorial design to determine the treatments for the candidates. To combine the two types of inputs (cause and candidates) into a single design, we construct a *crossed design* with runs at both levels of the cause for each combination of the candidates.

As an example, suppose we have five candidates labeled *A* to *E* and a single dominant cause *X,* each at two levels. The total number of runs is the product of the number of combinations of the candidates times the number of levels of the dominant cause. Suppose we can carry out 16 runs in total. Since the dominant cause has two levels, we can have eight treatments. We use MINITAB to select the quarter fraction design of resolution III for the five candidates. The combinations are shown in Table S16.1 in a randomized order.

For the crossed design, we have two runs, one at the low level and one at the high level of *X* for each treatment, as shown in Table S16.2, where the asterisks correspond to the 16 runs of the experiment.

Table S16.1   Selected treatment combinations for a 2^{5-2} design.

Treatment	A	B	C	D	E
1	−1	+1	−1	−1	+1
2	−1	−1	+1	+1	−1
3	−1	+1	+1	−1	−1
4	+1	+1	+1	+1	+1
5	+1	+1	−1	+1	−1
6	+1	−1	−1	−1	−1
7	+1	−1	+1	−1	+1
8	−1	−1	−1	+1	+1

Table S16.2   Crossed design with 16 runs.

						Level of X	
Treatment	A	B	C	D	E	Low	High
1	−1	+1	−1	−1	+1	*	*
2	−1	−1	+1	+1	−1	*	*
3	−1	+1	+1	−1	−1	*	*
4	+1	+1	+1	+1	+1	*	*
5	+1	+1	−1	+1	−1	*	*
6	+1	−1	−1	−1	−1	*	*
7	+1	−1	+1	−1	+1	*	*
8	−1	−1	−1	+1	+1	*	*

In MINITAB, we need each row to correspond to a single run. We create the design by:

- Pasting a copy of the eight treatments into the next eight rows of the spreadsheet

- Adding a column for the dominant cause, denoted *X,* with eight –1s followed by eight +1s

We get the columns as shown in Table S16.3.

Table S16.3 Crossed design with 16 runs displayed in rows.

A	B	C	D	E	X
−1	+1	−1	−1	+1	−1
−1	−1	+1	+1	−1	−1
−1	+1	+1	−1	−1	−1
+1	+1	+1	+1	+1	−1
+1	+1	−1	+1	−1	−1
+1	−1	−1	−1	−1	−1
+1	−1	+1	−1	+1	−1
−1	−1	−1	+1	+1	−1
−1	+1	−1	−1	+1	+1
−1	−1	+1	+1	−1	+1
−1	+1	+1	−1	−1	+1
+1	+1	+1	+1	+1	+1
+1	+1	−1	+1	−1	+1
+1	−1	−1	−1	−1	+1
+1	−1	+1	−1	+1	+1
−1	−1	−1	+1	+1	+1

Next we use MINITAB to create a custom factorial design (see Appendix F) with the six inputs. We can get the confounding structure for this design before we collect the data by using a column of dummy output. Any set of 16 numbers will do. The given design has the following confounding structure, where we show only main effects and two- and three-input interactions:

Alias Structure (up to order 3)

I + ABD + ACE
A + BD + CE
B + AD + CDE
C + AE + BDE
D + AB + BCE
E + AC + BCD
X
AX + BDX + CEX
BC + DE + ABE + ACD
BE + CD + ABC + ADE
BX + ADX

CX + AEX
DX + ABX
EX + ACX
BCX + DEX
BEX + CDX

The crossed design has the advantage that all of the two-input interactions involving the dominant cause X and the candidates are not confounded with any other two-input interactions. This is a powerful motive for using the crossed design given that the goal of the experiment is to examine these interactions.

We recommend using a crossed design for desensitization experiments. There are other possible designs, perhaps using fewer runs. See Hamada and Wu (2000).

S16.3 FURTHER ANALYSIS FOR THE EDDY CURRENT MEASUREMENT EXAMPLE

In the eddy current example discussed in Chapter 16, the team concluded the measurement system was not reliable. They came to this conclusion by looking at plots of the eddy current hardness measurement versus the Brinell hardness for all the treatments. We gave an example plot in Figure 16.11. Had the results looked more promising, the team could have conducted further analysis. For instance, they could have fit a regression model of the form:

$$eddy\ current\ hardness = a + b\,(Brinell\ hardness) + residual$$

for each of the eight treatments defined by the candidates. In the model, the residual represents the variation in the eddy current measurements not explained by Brinell hardness. A good treatment would have a large estimated slope b and a small residual standard deviation (given as s in the MINITAB regression results—see Appendix E). We use $|b/s|$ as the performance measure. To illustrate the analysis, for treatment 7 we get:

The regression equation is
Eddy Current Hardness = 16.3 – 2.60 Brinell

Predictor	Coef	SE Coef	T	P
Constant	16.347	2.041	8.01	0.000
Brinell	–2.5951	0.4544	–5.71	0.000

S = 0.2172   R-Sq = 34.5%   R-Sq(adj) = 33.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.5394	1.5394	32.62	0.000
Residual Error	62	2.9261	0.0472		
Total	63	4.4656			

Table S16.4   Eddy current desensitization experiment results.

Treatment	Frequency	Temp	Gain	Slope *b*	Residual variation *s*	Performance |*b/s*|
1	200	35	30	−0.545	0.104	5.2404
2	200	35	40	−0.948	0.188	5.0426
3	200	65	30	0.435	0.110	3.9545
4	200	65	40	−1.010	0.143	7.0629
5	350	35	30	−0.871	0.195	4.4667
6	350	35	40	−0.038	0.071	0.5352
7	350	65	30	−2.600	0.217	11.9816
8	350	65	40	−0.929	0.110	8.4455

The performance measure for treatment 7 is $|{-}2.5951/0.2172| = 12.0$. The data with the calculated performance measures are given in Table S16.4.

Treatment 7 gives the best performance. The Pareto plot of the effects given in Figure S16.1 summarizes the results across all treatments. The temperature main effect and the temperature by frequency interaction are the largest effects, but nothing stands out.
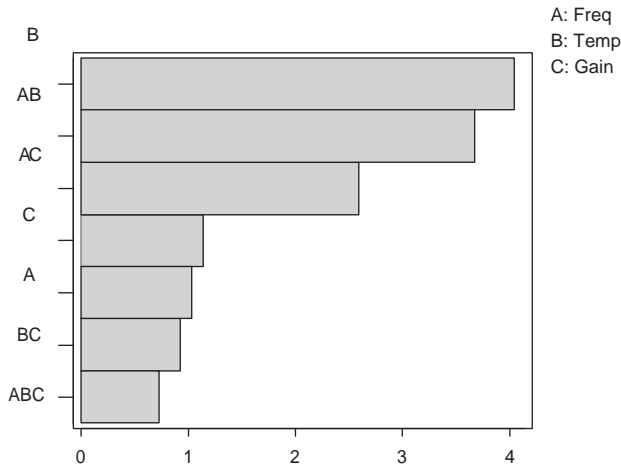

Figure S16.1   Pareto plot of the effects based on the performance measure.

We can also look at promising treatments in more detail by plotting the eddy current by Brinell hardness stratified by the four levels of the cause as defined by:

Cause combination	Day	Cleaning time
1	1	5
2	2	5
3	1	19
4	2	19

From Figure S16.2 for treatment 7, we see that the eddy current measurement system gives inconsistent output across the different levels of the cause. In addition, it does not work well for any cause combination, even well cleaned parts. We also see that the average Brinell hardness of the parts differed by day. This suggests the chemistry was indeed different across the two days.
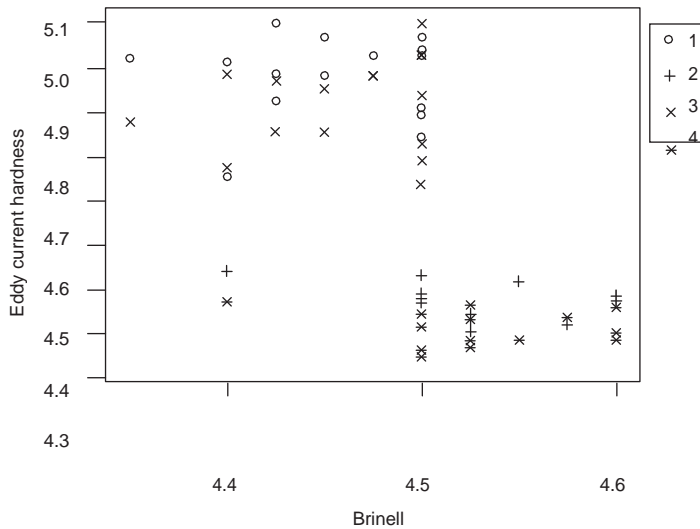
Figure S16.2   Scatter plot of eddy current hardness versus Brinell hardness for treatment 7 (plotting symbols represent the four values of the cause).

Chapter 17 Supplement

Feedforward Control Based on a Dominant Cause

S17.1 SIMULATING THE BENEFIT OF SELECTIVE FITTING

When considering implementation of selective fitting in the steering vibration example, the team needed to choose the number of bins and the location of the bin boundaries. They simulated the effect of two and three bins to determine their choice. The team chose the bin boundaries so that roughly equal numbers of components would fall into each bin. For example, with two bins they used the median center of gravity for part 2 (that is, 0.75) to define the boundary between the bins. We recommend choosing bin boundaries that result in roughly equal frequencies of components. For more advanced considerations, see Mease (2004).

We simulate the proposed selective fitting scheme using weights for the components sampled in some earlier investigation from the existing process. We start by specifying the number of bins, the bin boundaries, and the protocol for the assembly operation. In the example, we used a MINITAB macro that randomly selected a number of component pairs and calculated the distance from the center of gravity to the axis of rotation. See Appendix A for more information on writing MINITAB macros.

In the following, we give a MINITAB macro that assesses the effect of using two bins for part 2, called large and small, with bin boundary 0.75. The assembly protocol is to measure part 1. If the measured value exceeds the median for part 1, then select a part 2 from the large bin. In the simulation, the part 1 distances are sampled randomly from the baseline distribution.

```
MACRO
selectivefit2  part1  part2

mconstant I I2 cpart2 part1now part2now pnow test distc
mcolumn part2 part1 p1list allvals temp testcol

let cpart2=1
let   I2=1
let temp=1
let allvals=0

#repeat 50 times to get a reasonable estimate of the standard deviation
while I2<=50
 Sample 100 part1 p1list #get a list of 100 distances for part1 at random
  let I=1
   while I<=100
     let part1now=p1list(I)     #look at the next part1
     if part1now>0.75   #try to find a part2 that is also bigger than 0.75
        let  test=0
      while  test<=0
         let part2now=part2(cpart2)     #look at the next part2
         let cpart2=cpart2+1
         if cpart2>100  #end of list, start again
            let cpart2=1
         endif
         if part2now>0.75
           let test=1  #found appropriate match
         endif
         endwhile
      let distc=abso(part1now-part2now)  #vector
      endif
if part1now<=0.75    #try to find a part2 that is also smaller than 0.75
   let   test=0
   while  test<=0
     let part2now=part2(cpart2)     #look at the next part2
     let cpart2=cpart2+1
     if cpart2>100    #end of list, start again
       let cpart2=1
       endif
     if part2now<=0.75
       let   test=1    #found appropriate match
       endif
     endwhile
   let   distc=abso(part1now-part2now)      #vector components
```

```
      endif
    if (I2=1) AND (I=1)
        let allvals=distc
    else
        Stack  allvals  distc  allvals.     #store all the distance values
    endif
    let  I=I+1
    endwhile
let I2=I2+1
endwhile

let test = MEAN(allvals)      #some summaries of the combine distances
print   test
Let test = STDEV(allvals)
print   test

Code  (-100:2)  0  (2:100)  1  allvals  testcol
let test=sum(testcol)/(50*100)  #determine proportion bigger than 2
print  test
ENDMACRO
```

To call the macro, open the MINITAB file *steering wheel vibration feedforward* and copy the following command into the command line, replacing "filelocation" with the location of the macro on your system:

% filelocation\selectivefit2.txt' 'part1' 'part2'

Another way to simulate the effect of selective fitting is to first build models that describe the centers of gravity for the two components. Then, in the simulation, rather than drawing samples from the existing data, we draw samples from the models. We need to be careful that the models are appropriate. For an example of the modeling option, see the Chapter 17 exercises and solutions.

S17.2 MORE ON MAKING PREDICTIONS

There are many ways to create models to predict the output characteristic from the values of the dominant cause(s). Model building, which includes choosing the appropriate inputs and their form in the model, can be complicated. Residual plots can be helpful. Complex prediction models with many inputs should be validated to check for overfitting and other modeling errors. Validation involves building the prediction equation using one set of data and checking how well it works using a separate set (Neter et al., 1996).

Multiple Regression Models

In some circumstances, the prediction of output values can be improved by including other causes in addition to a dominant cause in the regression model. In other words, we may try to fit a model of the form:

$$output = b_0 + b_0 cause_1 + ... + b_k cause_k + residual,$$

where the residual is the variation due to all other causes not explicitly included in the model. This is called a *multiple regression model.* For further discussion of regression models, see the Chapter 12 supplement and Appendix E.

We want to avoid adding nondominant causes to the model because:

- Measuring additional inputs can be expensive.

- The inclusion of these inputs in the model can lead to overadjustment and, thus, to increased output variation.

Remember there are measurement errors, prediction errors, and adjustment errors. It does not make sense to improve the prediction marginally.

Smoothers

A good smoother is the LOcally WEighted Scatterplot Smoother (LOWESS) available in MINITAB (Cleveland, 1979). A LOWESS smoother models the cause/output relationship with a smooth curve. Predictions for the output can then be determined for any input value within the usual range. This prediction does not assume a linear relationship between the cause and the output.

Time Series Models

Time series prediction models (Box et al., 1994; Abraham and Ledholter, 1983) use the current and previous values of the dominant cause to help improve the prediction of the next output value. A time series model of this sort is called a *transfer function.* Transfer function models can be useful, for instance, if there is some time-to-time variation in the dominant cause or if the cause has a delayed reaction on the output characteristic. Box et al. (1994) call a feedforward controller dynamic if the prediction of the output is based on a transfer function.

Chapter 18 Supplement

Feedback Control

S18.1 ALTERNATIVE SIMPLE FEEDBACK CONTROLLERS

There are many informal ways to implement a feedback control.

**Grubbs's Rule for Feedback in a Setup-Dominated Process**

Grubbs proposed a method for the adjustment of a machine to center the process output at startup. See Grubbs (1954) and Del Castillo (2002) for details. The adjustment scheme is:

- Measure the first part after setup and make an adjustment equal to the difference between observed measurement and the process target.
- Measure the second part and make half the indicated adjustment.
- Measure the third part and make a third of the indicated adjustment.
- Stop after a set number of parts.

This scheme is designed for processes where the dominant cause of variation is in the setup. That is, once the process has been centered after startup, it does not drift or suffer sudden shifts.

Precontrol and Control Charts

Precontrol (or *stoplight control*) is a feedback control scheme where the possible output values are divided into regions based on the specification limits as shown in Figure S18.1. The green zone is the middle 50% of the specification range.
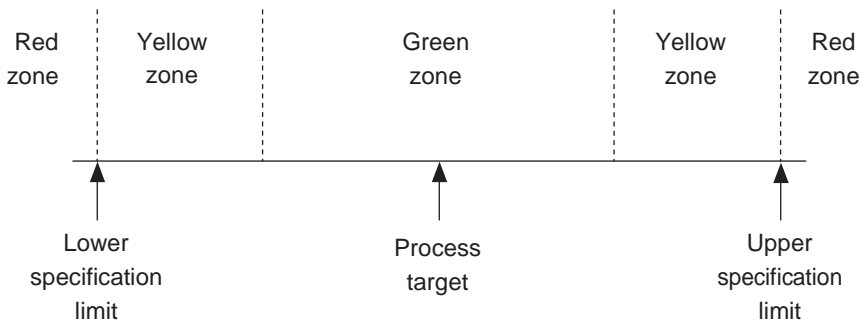
Figure S18.1   Precontrol division of the specification range.

To implement one of many versions of the rules, measure the output on one part at fixed time periods, and:

- If the measured output falls in the green zone, make no adjustment.

- If the measured output falls in the red zone, make a full adjustment and use a check (five consecutive parts in the green zone, for example) to ensure the process is properly centered.

- If the measured output falls in the yellow zone, measure another part. If the second measured value falls in the yellow (on the same side of the target) or red zones, make an adjustment. Otherwise, continue with no adjustment.

Satterthwaite (1954) first introduced Precontrol, which has received considerable attention in the research literature. For example, see Traver (1985), Shainin and Shainin (1989), Mack-ertich (1990), and Ledolter and Swersey (1997b). For Precontrol to be successful, the drift in the process must be relatively slow compared to the frequency of sampling and small relative to the specification range. Note that we need to add a rule to determine the size of the adjustment.

As another alternative, we can use a control chart to monitor the process and signal the need for adjustment. In the simplest version, we periodically measure one or more output values and plot the measured values on $\overline{X}$ and R charts. We make an adjustment if there is an out-of-control point on either chart. Again, we need a rule to decide on the size of the adjustment. To be successful, the sampling must be frequent enough to quickly detect changes in the process center.

Unlike Precontrol, we can also use control charts for process monitoring and for detecting the action of a cause of variation. A good comparison of process monitoring and feedback control is given in Box and Kramer (1992).

S18.2   SIMULATING   THE   BENEFIT   OF   FEEDBACK CONTROL

To quantify the potential benefit of a proposed feedback control scheme before implementation, we can simulate the effect by applying the scheme retrospectively to some historical (baseline or other) data from the existing process. The sampling plans in the proposed scheme and the historical data must match.

In the simulation, we keep track of the cumulative adjustment made to the series. Once the feedback control scheme is in operation, we observe only the adjusted series. To fix the notation we define

- y_t as the unadjusted output at time t

- y_t^* as the adjusted (observed) output at time t when using the feedback controller

- a_t as the adjustment made at time t (note the adjustment is not felt in the process until time $t + 1$)

- c_t as the cumulative adjustment up to and including time t

Then, we have:

- $c_t = \sum_{i=1}^{t} a_i = c_{t-1} + a_t$, that is, the cumulative adjustment is a sum of all previous adjustments

- $y_t^* = y_t + c_{t-1}$, that is, the observed output is the original (unadjusted) output plus the cumulative adjustment

As long as we keep track of the cumulative adjustments in the simulation, we can use these equations to go back and forth between the original and adjusted series.

To simulate the effect of applying the feedback controller, we start at time 1 and calculate adjustment a_1 and the cumulative adjustment $c_1 = a_1$. Then, at time 2, we observe the output $y_2^* = y_2 + c_1$. Next, we apply the feedback control rules to the observed series y_1, y_2 to get the adjustment at time 2, a_2, and the cumulative adjustment $c_2 = a_1 + a_2 = c_1 + a_2$. And so on.

The Matlab (see http://www.mathworks.com) code that follows illustrates the retrospective application of a feedback control scheme to the flow rate example from Chapter 18. The simulation is also straightforward to implement in Microsoft Excel. We show both the simulation of a controller based on the EWMA forecast with parameter alpha (as suggested in Chapter 18) and a simpler feedback controller that adjusts back to target if the observed flow rate is outside the range [99, 107].

```
function  []=feedbacksim(flowrate,alpha,dev)
%simulate the effect of using feedback on the fascia film build example
%original flow rate data given in vector "flowrate"
%example function call: feedbacksim(flowrate,0.2,4)

%try out the feedback controller based on the EWMA forecast of unadjusted %series,
  and based on partial adjustment of deviation from target 103

cumadj1=0; %cumulative adjustment for EWMA
cumadj2=0;  %cumulative adjustment for simple controller
z=flowrate-103*ones(1,180); %deviation from target for UNADJUSTED series

%do not make an changes for time 1
aseries1=z(1);      aseries2=z(1);        %initialize
```

```
for t=1:179,   %look at all flow rate values
    %try EWMA controller
    cumadj1=cumadj1-alpha*aseries1(t);
  aseries1(t+1)=z(t+1)+cumadj1;
  %try simple controller
    if (aseries2(t)<-dev) | (aseries2(t)>dev), %on standardized scale
    cumadj2=cumadj2-aseries2(t);
  end;
  aseries2(t+1)=z(t+1)+cumadj2;
end;

std(aseries1)      %stdev of series using EWMA based controller
std(aseries2)      %stdev of series using simple feedback controller

%plot of adjusted and unadjusted series
plot([1:180],z,'o-');     hold  on
plot([1:180],aseries1,'x-');
plot([1:180],aseries2,'x--');
hold off
```

Note that when using an exponential smoother to predict the next output value (as with the flow rate data shown in Figure 18.6) and making the full adjustment, MINITAB automatically gives the standard deviation of the adjusted series as the square root of the MSD, so there is no need to simulate.

Simulating the implementation of the proposed feedback controller using historical data will most likely overestimate the potential benefit since:

- We use the same historical data to model the data and develop a prediction equation.

- The simulation assumes there is no adjustment error.

An alternative way to assess the potential benefit of feedback is to fit a time series model to the historical data and simulate new output data. With a model for the output we need to check that the simulated series without any adjustment seems reasonable when compared to any historical data we have from the existing process. An advantage of having a model is that we can repeatedly simulate the effect of the proposed feedback controller.

S18.3 PROPERTIES OF THE EXPONENTIAL WEIGHTED MOVING AVERAGE

There are many options to obtain a forecast of future output values. Suppose we have a sampling protocol that looks at parts at equally spaced times labeled 1, 2, ..., t, and at each time point, we measure the output. The goal is to predict the output at time $t + 1$ using the observed values up to time t. Mathematically, we may view this as predicting the next

output y_{t+1}, using all previous outputs y_t, y_{t-1}, y_{t-2}, ..., where y_t denotes the value of the output measured at time t. For the moment, we assume there have been no adjustments. Denoting the one-step-ahead forecast as $\hat{y}_{t+1}$, two possible predictors are:

* Moving average: the average of the last k outputs; that is,

$$\hat{y}_{t+1} = \left(y_t + y_{t-1} + ... + y_{t-k+1} \right) \big/ k$$

* EWMA: a weighted average of all previous outputs with exponentially decaying weights; that is,

$$\hat{y}_{t+1} = \alpha \left[y_t + (1-\alpha) y_{t-1} + (1-\alpha)^2 y_{t-2} + (1-\alpha)^3 y_{t-3} + ... \right] \qquad \text{(S18.1)}$$

where α (the Greek letter *alpha*) is a constant and $0 < \alpha \leq 1$.

A special case of both options is using the last output value; that is, setting $\hat{y}_{t+1} = y_t$. Figure S18.2 illustrates the difference between the two forecasting options in terms of the relative weights given to past observations.

In the EWMA, a large value of α is best when the original series has small short-term variation, and thus recent observations provide a good prediction of the output value of the next unit. In situations where the process mean changes gradually, a good value for α often lies between 0.2 and 0.4 (Box and Luceno, 1997).

In any application, the best choice of forecast depends on the how the process changes over time when no adjustments occur. If there are sudden shifts, we want to use the moving average with k small. If the process shifts gradually, the EWMA is preferred. We need to determine k if we select the moving average and α if we pick the EWMA. The choice is best made empirically by simulating the performance of the possible schemes using some historical data. See Section S18.2.
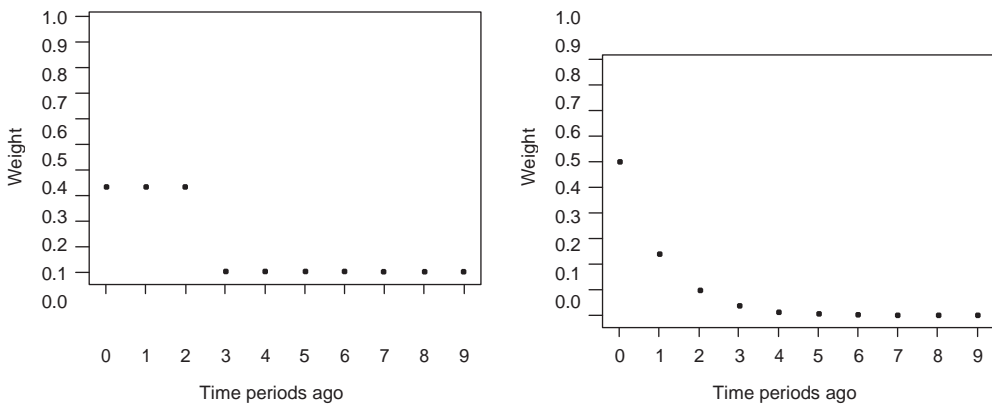
Figure S18.2   Possible weights (left panel moving average with $k = 3$, right panel EWMA with $\alpha = 0.6$).

Adjusting a Controlled Process

Once a controller is in operation, we no longer observe the original unadjusted series, and it is not obvious how to predict the next observation or what the appropriate adjustment is at each time point. In this section, we show that if we:

- Use the EWMA forecast and

- Assume perfect and complete adjustment after each observation,

then the correct adjustment is $\langle (y_t^* - T)$ toward the target, where y_t^* is the observed output (that is adjusted series) at time t, and T is the target value for the process center. We use the same notation as in the previous section to define the adjustment at time t, a_t, and the cumulative adjustment at time t, c_t.

Without loss of generality, we assume y_t and y_t^* have been scaled so that the target value is zero. Now suppose at each time period we make an adjustment to partially compensate for the difference between the observed output y_t and the process target; that is, we set

$$a_t = -\alpha \left(y_t^* - 0 \right)$$

Then, the cumulative adjustment at time t is given by

$$c_t = c_{t-1} + a_t$$
$$c_t = c_{t-1} - \alpha y_t^*$$
$$c_t = c_{t-1} - \alpha \left(y_t + c_{t-1} \right)$$
$$c_t = -\alpha y_t + (1-\alpha)c_{t-1}$$
$$c_t = -\alpha y_t + (1-\alpha)\left(c_{t-2} + a_{t-1} \right)$$
$$c_t = -\alpha y_t + (1-\alpha)\left(c_{t-2} - \alpha \left(y_{t-1} + c_{t-2} \right) \right)$$
$$c_t = -\alpha \left(y_t + (1-\alpha) y_{t-1} \right) + (1-\alpha)^2 c_{t-2}$$
$$c_t = -\alpha \left(y_t + (1-\alpha) y_{t-1} \right) + (1-\alpha)^2 \left(c_{t-3} - \alpha \left(y_{t-2} + c_{t-3} \right) \right)$$
$$c_t = -\alpha \left(y_t + (1-\alpha) y_{t-1} + (1-\alpha)^2 y_{t-2} \right) + (1-\alpha)^3 c_{t-3}$$

...

So $c_t = -\hat{y}_{t+1}$

Hence the cumulative adjustment is given by Equation (18.1); that is, we fully compensate for the deviation between the EWMA forecast for y_{t+1} made at time t and the target. Since, for the adjusted series, the predicted value at time $t+1$ equals $\hat{y}_{t+1}^* = \hat{y}_{t+1} + c_t$, the predicted output at time $t + 1$ will be on target.

Partial Adjustment

In discussing a feedback controller based on an EWMA forecast, we assume a full adjustment is made after each observation. Due to cost or other considerations, we may decide to make an adjustment only if the predicted deviation from target is large.

We can easily assess the effect of occasional adjustments by keeping track of the cumulative adjustment made up to time t (that is, c_{t-1}) so that we can recreate the series we would have observed had no adjustments been made ($y_t = y_t^* - c_{t-1}$). Using this unadjusted (and unobserved) series, we can determine an EWMA or any other forecast for the next time period. Then, translating back to the observed series, we compare the predicted observed output (that is, $\hat{y}_{t+1} + c_{t-1}$) to the process target to decide if further adjustment is needed.

S18.4 CONNECTION TO PID CONTROLLERS

There is extensive literature on feedback control in the engineering field. See for instance Aström (1970), Aström and Wittenmark (1989), and Prett and Gracia (1988). This literature does not, for the most part, address the important question of how to estimate model parameters from data. The EWMA-based controller is a discrete analog of a proportional, integral, derivative (PID) controller. A comprehensive treatment of feedback control from a statistical perspective is given in Box and Luceno (1997).

Chapter 19 Supplement
Making a Process Robust

S19.1 TAGUCHI'S METHODOLOGY

There is considerable controversy about the methods, experimental designs, and analysis suggested by Genichi Taguchi (1987); for instance, see Nair (1992) and Phadke (1989). Generally, at least in the statistical literature, it is felt that the fundamental ideas are valuable but that the specific designs and analyses can be improved upon. We briefly look at a number of ideas promoted by Taguchi.

In contrast to Statistical Engineering, with Taguchi's methodology:

- There is no recommended algorithm to reduce variation in existing processes

- There is little emphasis on understanding the patterns and causes of variation in the process output

- There is no explicit recognition of or search for a dominant cause

- There is heavy use of the robustness approach

To summarize the lack of emphasis put on identifying the cause, we quote Shin Taguchi from Nair (1992):

> *Notice that the objective of parameter design is very different from a pure scientific study. The goal in parameter design is not to characterize the system but to achieve robust function. Pure science strives to discover the causal relationships and to understand the mechanics of how things happen. Engineering, however, strives to achieve the results needed to satisfy the customer. Moreover, cost and time are very important issues for engineers. Science is to explain nature while engineering is to utilize nature.* (p. 130)

Taguchi considers using experimental design in both the product development process and to improve an existing process. This is a much broader view of improvement than is taken in Statistical Engineering, where we focus on improving an existing process.

To improve an existing process, Taguchi recommends:

- The process desensitization approach if there is a suspect cause

- The process robustness approach otherwise

In our experience, we have seen the desensitization approach fail because the cause included in the experiment was not dominant. There is little value in making the process insensitive to a cause that makes a small contribution to the overall variation.

Taguchi calls performance measures *signal-to-noise (S/N) ratios.* There is a large variety of signal-to-noise ratios that can be applied in different situations. Most often, Taguchi recommends S/N ratios of the form

$$10\log\left(\overline{y}^2/s^2\right) \text{ or} -10\log\left(s^2\right)$$

where $\overline{y}$ is the average output in a run and s is the standard deviation of the output values measured within a run. Note that $-10 \log(s^2) = -20 \log(s)$, so that Taguchi S/N ratio $-10 \log(s^2)$ is a rescaled version of the performance measure we recommend to analyze robustness experiments where the goal is to reduce the within-run variation. We get identical conclusions from the analysis with either performance measure.

In some problems, the goal is to lower or increase the process center. Taguchi calls these *smaller/larger-is-better* problems and recommends the S/N ratios:

$$\text{Smaller is better: } -10\log[(y_1^2 +...+ y_r^2)/r]$$

$$\text{Larger is better: } -10\log[\frac{1}{r}(\frac{1}{y_1^2} +...+ \frac{1}{y_r^2})]$$

where $y_1,..., y_r$ are the r measured values of the output characteristic on the repeats within each run. The idea is to calculate the performance measure (S/N ratio) for each run and then analyze these measures as the response in the experiment. In all cases, S/N ratios are defined so that larger values are better in terms of the goal of the problem. See Box (1988) for further discussion of signal-to-noise ratios. We recommend using two performance measures, the within-run average and log standard deviation, rather than a single S/N ratio.

Many of Taguchi's designs and methods of analysis are available in MINITAB. We do not recommend them.

APPENDICES

Using MINITAB

In the appendices we show how to create the graphs and statistical analyses described in this book using the statistical software package MINITAB. Throughout we assume the reader is familiar with a Windows-based environment.

MINITAB is a general-purpose statistical software package. See http://www.minitab.com. MINITAB is a leader in the area of quality improvement and is widely used in industry. There are other suitable packages such as SAS, Splus, Statgraphics, JMP, and Systat. We do not recommend Microsoft Excel because we find it cumbersome and inefficient for producing the required analyses. In this book we illustrate MINITAB release 13.30 for Windows.

MINITAB is easy to use since its interface is based on pull-down menus and dialog boxes. The capabilities of MINITAB are well explained in the manuals (MINITAB *User's Guide 1* and *2,* 2000a and 2000b) and by the online help. In particular, in each dialog box there is an option for context-sensitive help that usually includes an example. A good reference book for MINITAB is Ryan et al. (2000).

In these appendices, we show how to select a MINITAB command through its pull-down menu structure using words and arrows. For example, a multivari chart is obtained in MINITAB using the menu selection:

Stat → Quality Tools → Multi-Vari Chart

This corresponds to menu navigation.
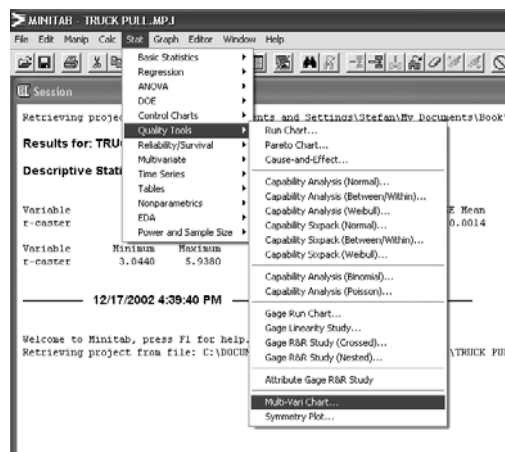
Finding a MINITAB command.

Appendix A

Data Storage and Manipulation

The first step is to get the data into the MINITAB worksheet. To illustrate, we use an investigation where the goal was to compare the performance of three different measurement devices—a feeler gage, a height gage, and a scale gage—for measuring the distance between a bottle lip and a label. The data are given in the file *label height measurement*.

To open a MINITAB worksheet (a file with the extension *.mtw)* we use:

File → Open Worksheet

To open an existing MINITAB project, which includes a worksheet with the data and all numerical and graphical summaries of the data we previously produced, use

File → Open Project

A.1 ROW/COLUMN DATA STORAGE FORMAT

In MINITAB, data are stored in a worksheet similar in appearance to a Microsoft Excel worksheet. See Figure A.1 for an example. There are, however, some important differences between Excel and MINITAB worksheets. With MINITAB, the data must be entered in the row/column format, where each column in the worksheet represents a different characteristic

	C1	C2	C3	C4	C5	C6	C7	C8	C9
	part	operator	feeler_low	feeler_high	height_low	height_high	scale_low	scale_high	
1	1	1	0.0620	0.0800	6.579	6.587	0.063000	0.094000	
2	1	1	0.0620	0.0830	6.580	6.588	0.094000	0.125000	
3	1	2	0.0590	0.0770	6.566	6.581	0.078130	0.093750	
4	1	2	0.0610	0.0860	6.574	6.598	0.062500	0.093750	
5	2	1	0.0170	0.0690	6.587	6.635	0.031000	0.094000	
6	2	1	0.0130	0.0650	6.586	6.626	0.031000	0.094000	
7	2	2	0.0200	0.0600	6.593	6.636	0.031250	0.062500	
8	2	2	0.0230	0.0640	6.588	6.631	0.031250	0.062500	
9	3	1	0.0410	0.0690	6.581	6.600	0.063000	0.094000	
10	3	1	0.0400	0.0710	6.577	6.605	0.063000	0.094000	
11	3	2	0.0430	0.0740	6.581	6.593	0.046875	0.078125	
12	3	2	0.0440	0.0740	6.582	6.612	0.031250	0.078125	
13	4	1	0.0550	0.0960	6.550	6.585	0.063000	0.094000	
14	4	1	0.0470	0.0870	6.575	6.618	0.063000	0.094000	
15	4	2	0.0530	0.0820	6.581	6.620	0.046875	0.078125	
16	4	2	0.0500	0.0870	6.576	6.624	0.046875	0.093750	

Figure A.1   Data in the MINITAB worksheet.

Microsoft Excel - feeler.wk1

File Edit View Insert Format Tools Data Window Help

A1 = Bott A

	A	B	C	D	E	F	G	H	I
1	Bott A	Trial #1	Trial #2	Average		Bott A	Trial #1	Trial #2	Average
2	Lowest	0.062	0.062	0.062		Lowest	0.059	0.061	0.06
3	Highest	0.08	0.083	0.0815		Highest	0.077	0.086	0.0815
4	Range	0.018	0.021	0.0195		Range	0.018	0.025	0.0215
5									
6	Bott B	Trial #1	Trial #2	Average		Bott B	Trial #1	Trial #2	Average
7	Lowest	0.017	0.013	0.015		Lowest	0.02	0.023	0.0215
8	Highest	0.069	0.065	0.067		Highest	0.06	0.064	0.062
9	Range	0.052	0.052	0.052		Range	0.04	0.041	0.0405
10									
11	Bott C	Trial #1	Trial #2	Average		Bott C	Trial #1	Trial #2	Average
12	Lowest	0.041	0.04	0.0405		Lowest	0.043	0.044	0.0435
13	Highest	0.069	0.071	0.07		Highest	0.074	0.074	0.074
14	Range	0.028	0.031	0.0295		Range	0.031	0.03	0.0305
15									
16	Bott D	Trial #1	Trial #2	Average		Bott D	Trial #1	Trial #2	Average
17	Lowest	0.055	0.047	0.051		Lowest	0.053	0.05	0.0515
18	Highest	0.096	0.087	0.0915		Highest	0.082	0.087	0.0845
19	Range	0.041	0.04	0.0405		Range	0.029	0.037	0.033
20									
21	Bott E	Trial #1	Trial #2	Average		Bott E	Trial #1	Trial #2	Average
22	Lowest	0.043	0.054	0.0485		Lowest	0.049	0.048	0.0485
23	Highest	0.077	0.076	0.0765		Highest	0.07	0.0755	0.07275
24	Range	0.034	0.022	0.028		Range	0.021	0.0275	0.02425
25									
26	Bott F	Trial #1	Trial #2	Average		Bott F	Trial #1	Trial #2	Average
27	Lowest	0.045	0.05	0.0475		Lowest	0.042	0.046	0.044
28	Highest	0.062	0.066	0.064		Highest	0.061	0.067	0.064
29	Range	0.017	0.016	0.0165		Range	0.019	0.021	0.02
30									
31	Bott G	Trial #1	Trial #2	Average		Bott G	Trial #1	Trial #2	Average
32	Lowest	0.055	0.057	0.056		Lowest	0.0535	0.053	0.05325
33	Highest	0.089	0.093	0.091		Highest	0.091	0.091	0.091
34	Range	0.034	0.036	0.035		Range	0.0375	0.038	0.03775
35									
36	Bott H	Trial #1	Trial #2	Average		Bott H	Trial #1	Trial #2	Average
37	Lowest	0.055	0.064	0.0595		Lowest	0.057	0.056	0.0595
38	Highest	0.084	0.079	0.0815		Highest	0.0845	0.088	0.0695
39	Range	0.029	0.015	0.022		Range	0.0275	0.032	0.02975

Figure A.2   Poorly stored data in an Excel spreadsheet.

(output or input), and each row represents a different observation. This format is not necessarily intuitive for people who commonly use Excel. Also, in MINITAB every column (i.e. characteristic) can be given a descriptive name.

To illustrate why the row/column format is recommended, see Figure A.2, which shows how the data for the measurement investigation were initially recorded. The Excel worksheet in Figure A.2 gives the results only for the feeler gage. The results for the height and scale gages were stored in separate worksheets. In Figure A.2, the meaning of each data value depends not only on its location in the worksheet, but also on the location of various labels. For example, the value in cell B2 (0.062) gives the lowest value of trial 1 of the first operator for bottle *A* using the feeler gage. This way of storing data is not convenient when doing analysis, since determining the meaning of the various data values is difficult.

Figure A.1 shows how the same data for all three gages were stored in MINITAB. The label above each column provides the name of the stored characteristic, and each row represents a different observation, or in this case, a measurement. The 0.062 value discussed previously is now stored in the first row of column C3 (labeled *feeler_low*). We can identify the corresponding part (bottle 1) and operator *(A)* by examining the values of the other characteristics in the same row.

The more efficient row/column format approach to storing data identifies the relationships among the collected values. We can also store data in Excel in the row/column format. Translating from the original data format, shown in Figure A.2, to the row/column format, shown in Figure A.1, is a tedious process.

Another important advantage of a MINITAB worksheet is its ability to easily handle missing values. Any blank spaces in the data file become stars (symbol for *missing*) in MINITAB. All analyses in MINITAB properly deal with missing values. It is a poor idea to use special numerical codes like –99 to signify missing values. Using numerical codes for missing observations will adversely affect many data summaries.

A.2 MAKING PATTERNED DATA

It is often useful to create patterned data. The patterned data commands are shown in Figure A.3 and accessed via the MINITAB menu selection:

Calc → Make Patterned Data

In the label height measurement example, to create the column corresponding to part numbers, we fill in the dialog window as in Figure A.4. Each of 12 parts is measured four times (twice each by two different operators). Applying the command creates four 1s, four 2s, and so on in column C1. If preferred, text values (for example, *A*, *B*, *C*, and so on) can be used to represent the different bottles.
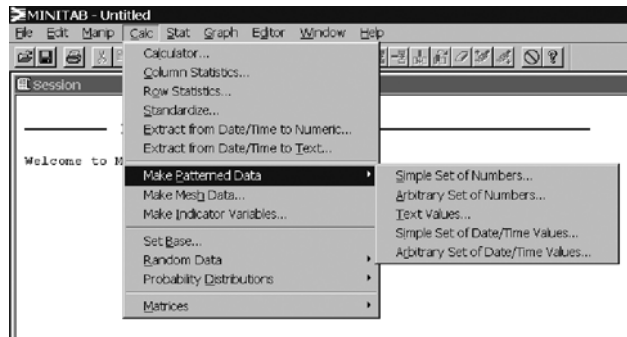
Figure A.3   Patterned Data menu.
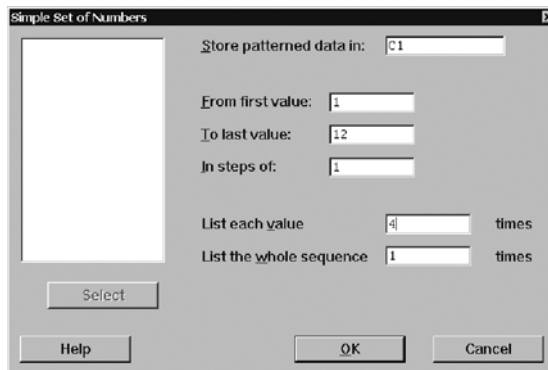
Figure A.4   Dialog box for simple set of patterned numbers.

A.3 CALCULATING DERIVED CHARACTERISTICS

To calculate new characteristics from the existing data, we use the calculator:

Calc → Calculator

In the label height measurement example, the minimum and maximum heights (as the bottle was rotated) were recorded. In Figure A.5, we use the calculator to record the difference (max–min) for each feeler gage measurement in column C9.

To use the calculator, we enter a formula involving the existing characteristics in the expression window. The expression can be typed in directly or identified using the *select* button. The expression may include any of the functions given in the list. Many standard statistical functions such as average and standard deviation are available. Note that (most of) these functions act simultaneously on all rows across different columns. Column summaries (for example, the mean or standard deviation of a characteristic) are discussed in Appendix B.
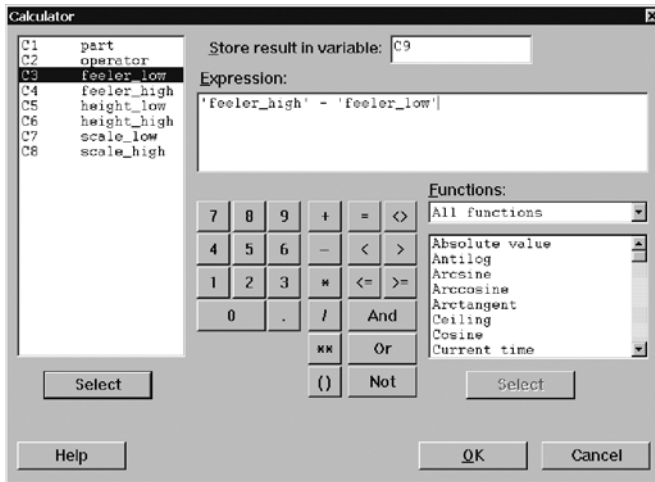
Figure A.5   Calculator dialog box.

A.4 SELECTING A DATA SUBSET

In many circumstances, it is useful to stratify the data in some way. Many numerical and graphical summaries (see appendices B and C) allow stratification without subsetting the data. We sometimes stratify data into separate worksheets using (see Figure A.6):

Manip → Subset Worksheet [Data → Subset Worksheet, in release 14.11]

or

Manip → Split Worksheet [Data → Split Worksheet, in release 14.11]
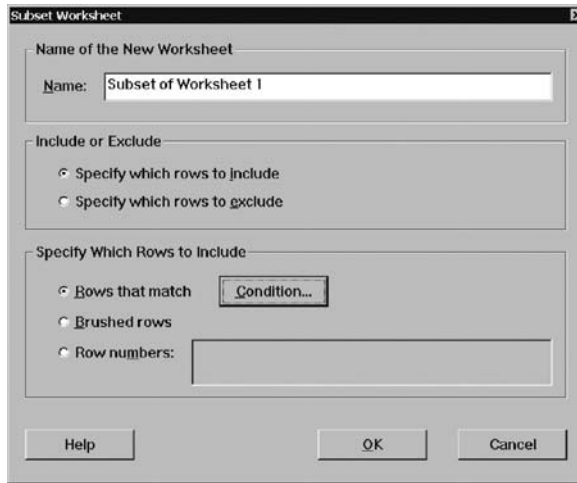
Figure A.6   Subset Worksheet dialog box.

When selecting the subset, we select the data observations to either include or exclude. There are a variety of ways of specifying the subset of observations. We can give the row numbers, select the rows beforehand using brushing, or give a condition that must be satisfied.

A.5   STACKING COLUMNS

In some circumstances, it is convenient to have the data stored in more than one way in order to conduct the desired analysis. Consider the brake rotor balance example discussed in Chapter 13. In the verification experiment there were eight runs at each of the eight treatments and, within each run, eight rotors were produced. The data are stored in the file *brake rotor balance verification* and the worksheet is shown in Figure A.7. The balance weights required for the eight rotors within each treatment are stored in separate columns labeled *r1* through *r8*.

To start the analysis, we plot the weight by treatment as shown in Figure 13.4. To produce this plot, however, we need all 64 balance weights stored in a single column. We stack the columns using:

Manip → Stack → Stack Columns

Figure A.7   Brake rotor balance verification experiment plan and data.
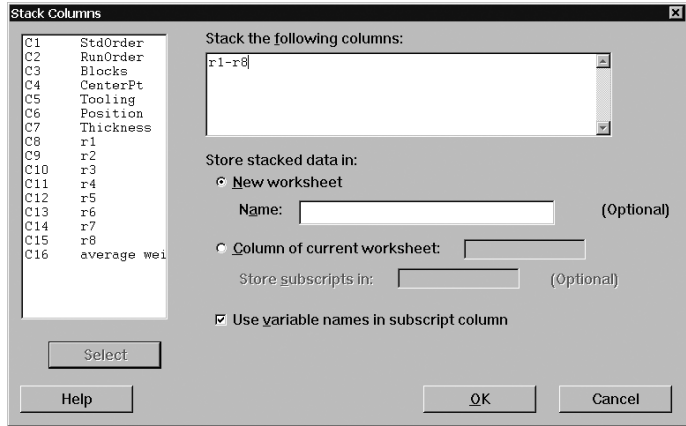
Figure A.8   MINITAB Stack Columns dialog box.

Figure A.9   Worksheet resulting from stacking columns.

The corresponding MINITAB dialog box with the required information filled in is given in Figure A.8.

Stacking the columns produces the first two columns shown in Figure A.9, where the second column has been given a descriptive label. To produce Figure 13.4, we need an additional column that gives the treatment corresponding to each weight. Figure A.9 shows the results of using the command to make patterned data as described in Section A.2.

The data were originally stored as in Figure A.7 because, for subsequent analysis, we calculate main and interaction effects based on the average balance weight.

A.6 MINITAB MACROS

Macros are useful for automating a series of commands repeated many times. Macros are also needed if you want do calculations row by row (that is, observation by observation) where the action depends on the previous value(s). For example, a macro can be used to simulate the effect of a proposed feedback control scheme on existing process data.

MINITAB macros are similar to the macros available in other software such as Excel. See the MINITAB help for more information on writing macros.

The sample macro illustrates testing a proposed feedback control scheme for the piston diameter example discussed in Chapter 18. The data are given in the file *V6 piston diameter 270*. The feedback scheme adjusts the process center back to the target of 6.7 if the observed diameter is less than 2.7 or greater than 10.7. Note that in the data file there are 200 diameter readings.

```
MACRO
macrotest  diameter  adjusted

mconstant cumad I diameter2 nadj
mcolumn diameter adjusted

let  I=1
let cumad=0  #cumulative adjustment needed based on rule
let  nadj=0

while  I<=200
let diameter2=diameter(I)+cumad  #current value
let adjusted(I)=diameter2
#determine if further adjustment needed
if diameter2<2.7
let cumad=cumad+(6.7-diameter2)
let nadj=nadj+1
endif
if  diameter2>10.7
let cumad=cumad-(diameter2-6.7)
let nadj=nadj+1
endif

let  I=I+1
endwhile

ENDMACRO
```

We save the macro in a separate text file and call the macro using the command line editor. The command line editor is available using:

Edit → Command Line Editor

To run the macro from the command line we type:

%macro name (including path if necessary) v1 v2 ...

where v1 and v2 are the characteristics from the current worksheet the macro acts on or characteristics the macros creates. For example, we may issue the command as shown in Figure A.10.
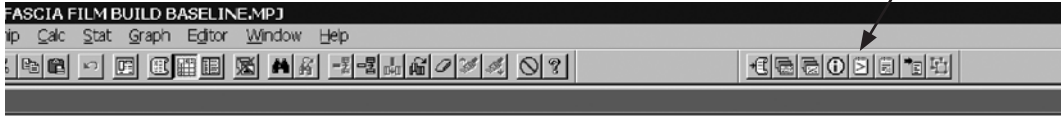

Figure A.10 Running a MINITAB macro.

The result of executing the macro is shown in the data file snapshot in Figure A.11. At observation 39, the observed diameter was 11.7. In the example, this was the first time an adjustment was required. At that time, the cumulative adjustment was set to –5. We can see the effect of the adjustment on subsequent diameters. Later in the data file (not shown in Figure A.11) further adjustments are necessary.

Figure A.11 MINITAB worksheet showing effect of running the macro.

Figure A.12   MINITAB show history dialog button.

   To write MINITAB macros, we can use the Show History dialog button that gives the command line interpretation for any MINITAB commands run using the pull-down menus. The History dialog button is the small yellow button with a prompt (>) at the far right at the top of the MINITAB window, as shown in Figure A.12.
   Another example macro to simulate the effect of selective fitting is given as part of the solution to Exercise 17.2.

Appendix B

Numerical Summaries

We use a manifold sand scrap investigation to illustrate various numerical data summaries in MINITAB. In the investigation, each manifold was classified as scrap or not and linked (as well as possible) to a number of sand and pour characteristics. We have a total of 17 characteristics measured on 970 castings. The data are given in the file *manifold sand scrap comparison*.

B.1 NUMERICAL SUMMARIES FOR CONTINUOUS CHARACTERISTICS

Numerical summaries appropriate for continuous characteristics are available using:

   Stats → Basic Statistics → Display Descriptive Statistics

Selecting the input characteristics pour time and temperature gives:

Descriptive Statistics: pourtime, temperature

Variable	N	Mean	Median	TrMean	StDev	SE Mean
pourtime	970	4.6990	4.7000	4.6794	0.5132	0.0165
temperature	970	92.446	92.450	92.442	2.751	0.088

Variable	Minimum	Maximum	Q1	Q3
pourtime	3.3000	7.5000	4.4000	5.0000
temperature	86.100	98.500	90.500	94.100

To define the summary measures, *Mean* is the sample average and *StDev* is the sample standard deviation. If we rank the values, the smallest value is given by *Minimum*, the largest value by *Maximum*, and the middle value by *Median*. *Q1* and *Q3* define the first and
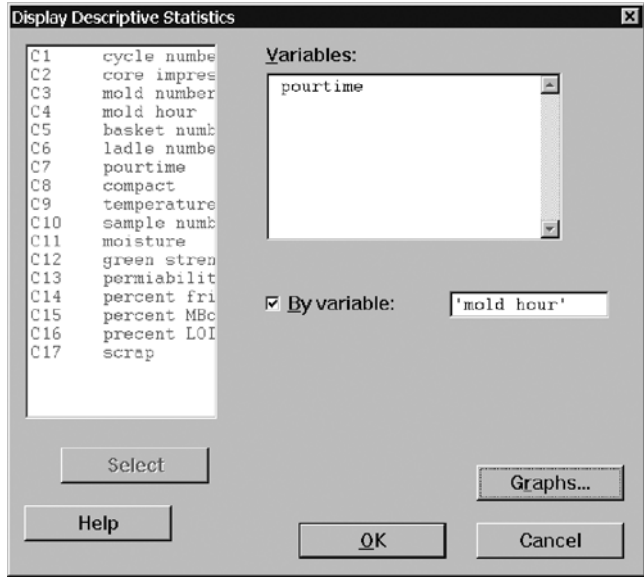
Figure B.1   Display Descriptive Statistics dialog box.

third quartiles—that is, the observation halfway in the ordered list between the minimum and the median, and halfway between the median and the maximum, respectively. The median defines the second quartile.

We can also produce numerical summaries for any characteristic stratified by a discrete characteristic. The dialog box in Figure B.1 shows how to stratify the summary for pour time by the mold hour.

The results are:

**Descriptive Statistics: pourtime by mold hour**

Variable	mold hour	N	Mean	Median	TrMean	StDev
pourtime	3	166	4.8470	4.7000	4.8320	0.6400
	4	172	4.7488	4.7000	4.7130	0.5893
	5	176	4.5409	4.5000	4.5196	0.5540
	6	180	4.6289	4.6000	4.6228	0.4833
	7	142	4.7423	4.7000	4.7359	0.3190
	8	134	4.7075	4.7000	4.7100	0.2568

Variable	mold hour	SE Mean	Minimum	Maximum	Q1	Q3
pourtime	3	0.0497	3.7000	6.2000	4.3000	5.3000
	4	0.0449	3.8000	7.5000	4.3000	5.1000
	5	0.0418	3.3000	7.3000	4.1250	4.8750
	6	0.0360	3.5000	6.5000	4.3000	4.9000
	7	0.0268	4.1000	5.5000	4.5000	5.0000
	8	0.0222	4.1000	5.2000	4.5000	4.9000

Sometimes we need to save numerical summaries in the worksheet for further analysis. We request storage of some of the descriptive summaries using:

Stats → Basic Statistics → Store Descriptive Statistics

From the dialog box (see Figure B.2) selecting *Statistics*, we check off the data summaries we want to save. This is useful in the analysis of multivari investigations that involve a part-to-part family or other family that is expected to have a haphazard effect.
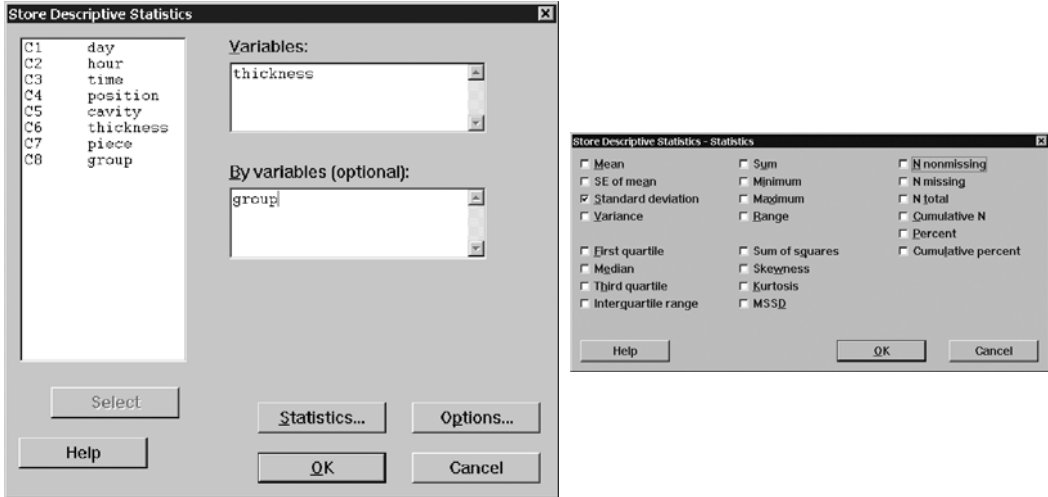
Figure B.2   Store Descriptive Statistics dialog box.

Note that numerical summaries are sometimes given using exponential notation. Exponential notation is convenient for very small or very large numbers. The integer given after the *E* tells us how many positions (to the right, if the integer is positive; to the left, if the integer is negative) to shift the decimal point; for example, 4.49E–03 = 0.00449.

B.2 NUMERICAL SUMMARIES FOR DISCRETE CHARACTERISTICS

For discrete characteristics, count summaries are available through the MINITAB menu selection:

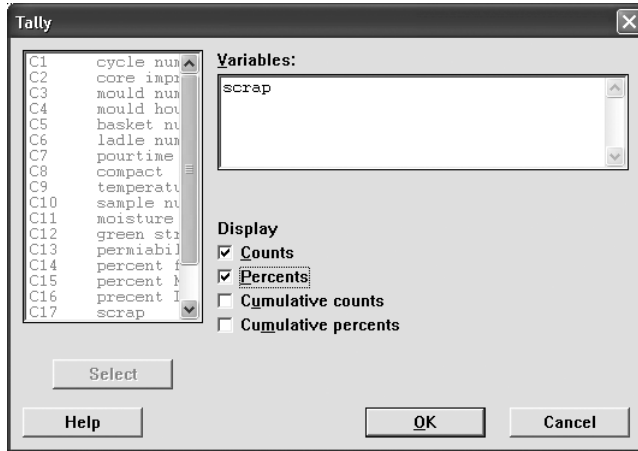*Stat → Tables → Tally*

Figure B.3 shows the *Tally* dialog box.

Figure B.3 Tally dialog box.

Using the tally summary for the characteristic scrap, we get

Tally for Discrete Variables: scrap

scrap	Count	Percent
0	853	87.94
1	117	12.06
N=	970	

We see that there are only two possible values for scrap (where 0 represents a passed manifold and 1 a manifold that was scrapped). The proportion of scrap is roughly 12%.

In such situations, we may also be interested in determining whether a relationship exists between two discrete characteristics. We use the MINITAB menu selection:

Stat → Tables → Cross Tabulation

For example, as illustrated by Figure B.4, we may be interested in determining if there is a relationship between mold hour and whether the casting is scrapped. The MINITAB results given as follows suggest the scrap rate is lower in hours 7 and 8 though the sample size in each hour is small.
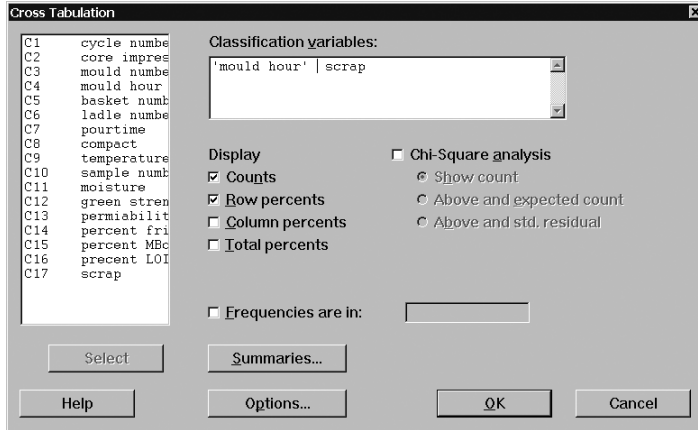
Figure B.4   Cross Tabulation dialog box.

Tabulated Statistics: mold hour, scrap

Rows: mold hour   Columns: scrap

	0	1	All
3	141	25	166
	84.94	15.06	100.00
4	148	24	172
	86.05	13.95	100.00
5	151	25	176
	85.80	14.20	100.00
6	156	24	180
	86.67	13.33	100.00
7	133	9	142
	93.66	6.34	100.00
8	124	10	134
	92.54	7.46	100.00
All	853	117	970
	87.94	12.06	100.00

Cell Contents --
            Count
   **% of Row**

Appendix C

Graphical Summaries

In all investigations, graphical summaries play an essential role in the analysis. Often we use only a graphical summary to draw conclusions.

C.1 HISTOGRAM

A histogram summarizes the distribution of a continuous characteristic in the data set. Histograms are available through the MINITAB menu selection:

Graph → Histogram

To make comparing histograms easier, we recommend using the *option* button and choosing *percent* rather than the default *frequency* display. In the manifold sand scrap example introduced in Appendix B, we enter the characteristic *pour time* as the graph variable to get the histogram in Figure C.1. The data are given in the file *manifold sand scrap comparison*.
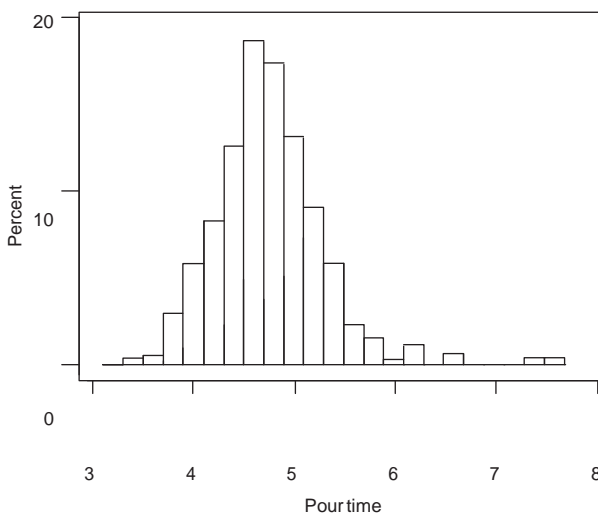
Figure C.1   Histogram of pour times.

The histogram shows the distribution of pour times. In Figure C.1, we see that the average pour time is roughly 4.75 seconds and that there are a few large pour times above 7 seconds. Note that the time order of the data is lost in a histogram.

If there is a large number of observations, it may be difficult to identify outliers (unusually small or large values) in a histogram. This occurs, for instance, in the histogram of alignment pull given in Figure 1.2. Box plots, covered in Section C.3, are a good alternative when the number of observations is large or when we wish to compare distributions.
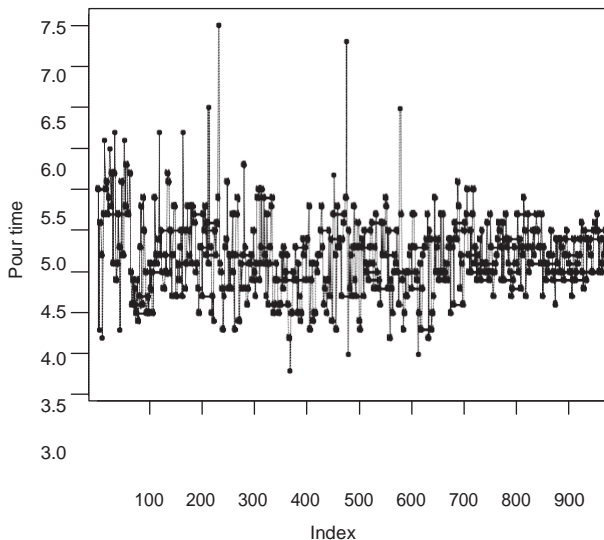
C.2 RUN CHART

We use run charts (also called time series plots) to look for patterns over time. A run chart is available through:

Graph → Time Series Plot

The run chart of pour time in Figure C.2 shows the time order. Figure C.2 suggests the variation in pour times is greater at the beginning of the investigation. At the end, say after casting 700, the variation is noticeably smaller.

The run chart assumes the observations are equally spaced in time. If this is not the case, we use a scatter plot (see Section C.4) where the horizontal axis is defined in terms of time. If the data are collected in subgroups over time, say five units measured each hour, we can also use a multivari plot (see Section C.5).
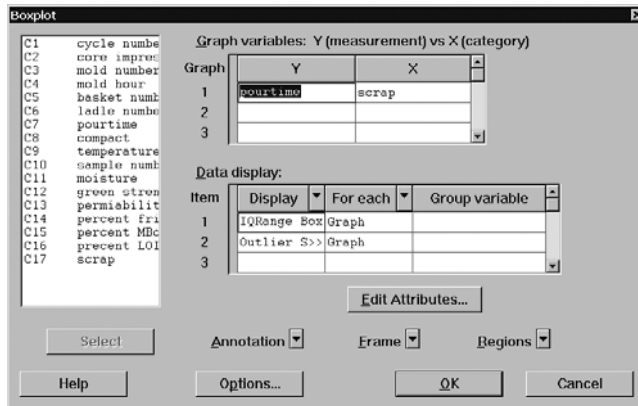

**Figure C.2**   Run chart of pour times.

C.3 BOX PLOT

A box plot is a simplified histogram useful for comparing data subsets and identifying outliers. Box plots are available using (see Figure C.3):

Graph → Boxplot

Figure C.3   Boxplot dialog box.

In Figure C.4 we stratify pour time by whether or not the part is scrapped. The horizontal line in the middle of each box gives the *median* value. The upper and lower edges of the rectangle show the first and third quartiles, *Q1* and *Q3*, defined in the descriptive statistics summary in Appendix B. The so-called whiskers are the lines coming out of the central rectangle. The ends of the whiskers identify the range of the data with the exception of unusual values plotted as separate stars. The plotting of unusual values with separate symbols is useful for finding outliers.
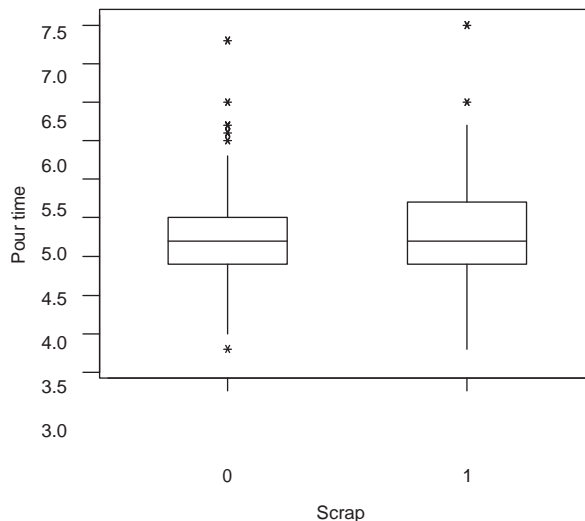

Figure C.4   Box plot of pour time by scrap.

If the number of observations summarized in the box plot is small, we prefer displaying the individual observations rather than the summary statistics. In MINITAB we change the display items to *Individual Symbol*, rather than the default of *IQRange Box* and *Outlier Symbol*. In MINITAB release 14.11, we can plot individual observations using:

Graph → Individual Value Plot

For example, Figure C.5 shows measurement error stratified by part number for the camshaft diameter relative bias investigation discussed in Chapter 7. The data come from the file *camshaft journal diameter measurement2*. In Figure C.5, we see the that measurement error is negative in all cases and that there is little difference in the distribution of the errors among the parts.
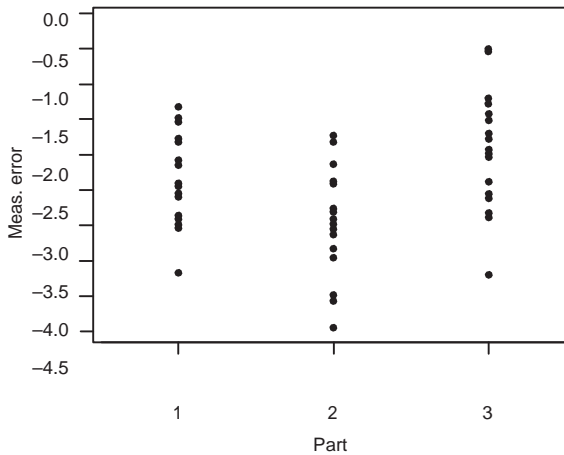
Figure C.5   Box plot (showing individual observations) of measurement error by part number.

C.4 SCATTER PLOT

A scatter plot is a generalization of a run chart where the horizontal axis is defined by any characteristic, not just time. Scatter plots are available through:

Graph → Plot [Graph → ScatterPlot in release 14.11*]*

Scatter plots are useful for examining the relationship between two characteristics. In the manifold sand scrap example we examine how pour time changes with ladle number (see Figure C.6).

The resulting scatter plot, given in Figure C.7, suggests there is no association between pour time and ladle number.
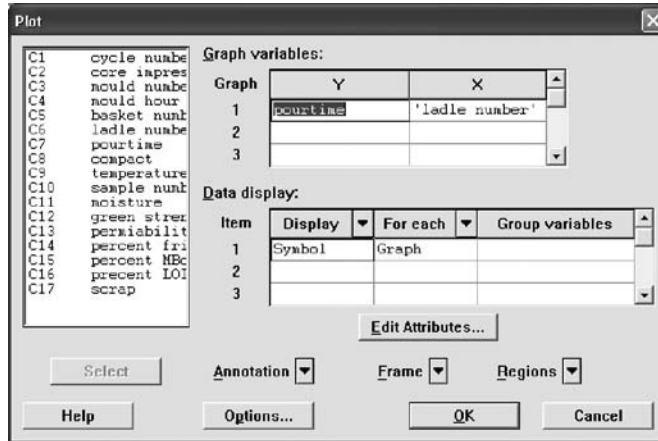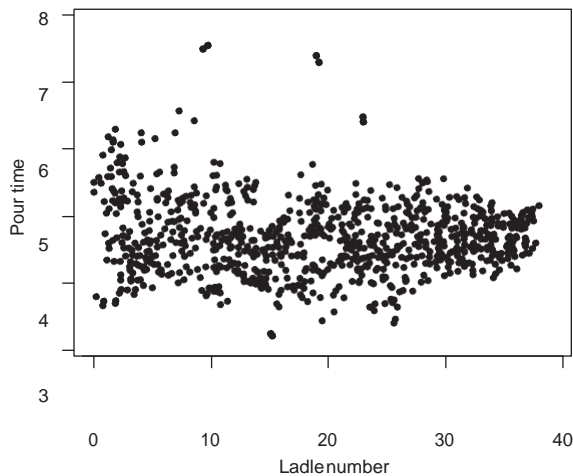
Figure C.6 (Scatter) Plot dialog window.

**Figure C.7** Scatter plot of pour time versus ladle number.

Adding Jitter

In the crossbar dimension robustness investigation from Chapter 19, eight treatments were used, and each run consisted of five consecutive parts for each treatment. The data are given in the file *crossbar dimension robustness*. The output "burn" could take only four possible values. Plotting burn by treatment, as shown in the left panel of Figure C.8, is helpful but difficult to interpret, since we cannot see the output for all five parts in each run. Many observations are plotted on the same position, called overplotting. We add jitter (small
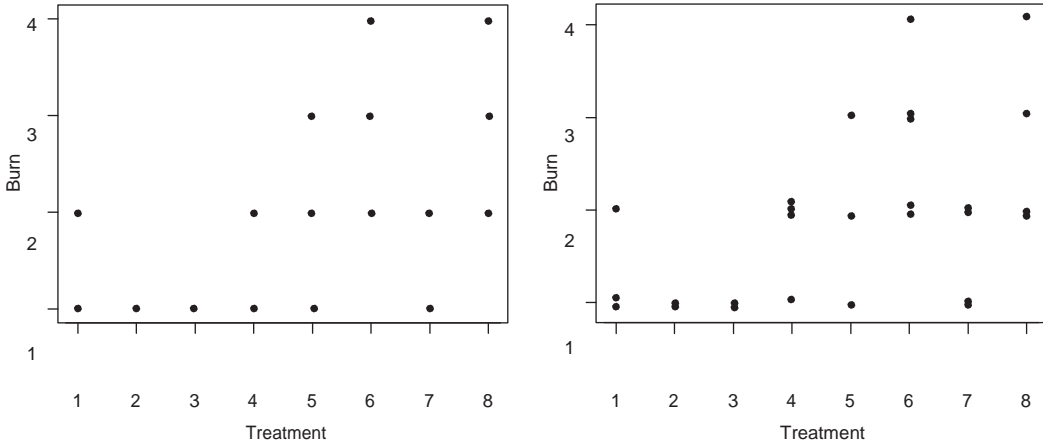
Figure C.8   Scatter plot of burn by treatment—regular on left, with jitter in vertical direction on right.
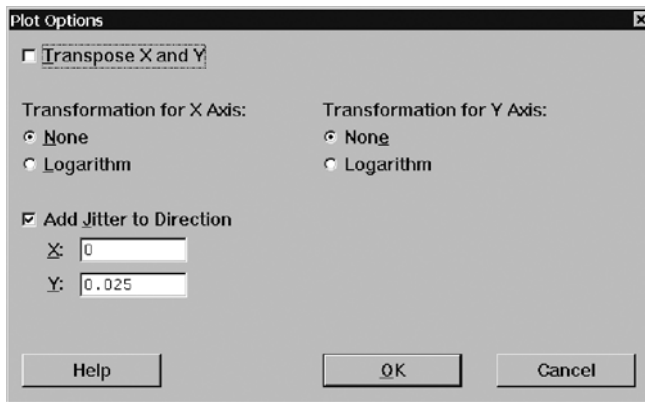
Figure C.9   Plot Options dialog box showing selecting of jitter.

random perturbations) to each observation to reduce the effect of overplotting as given in Figure C.9. Here we illustrate adding jitter in the vertical, or *Y*, direction. The scatter plot with added jitter is shown in the right panel of Figure C.8. The dialog box shown in Figure C.9 is obtained by selecting options in Figure C.6.

Labeling Points

The plotted points can be labeled using the values of any discrete characteristic. Consider the cylinder head scrap example discussed in Chapter 12. The data are given in the file *cylinder head scrap multivari*. The team wanted to see if the relationship between side shift (the output) and time depended on pattern.

They used the dialog box, as shown in Figure C.10, to make the scatter plot given in Figure C.11. Note the change to the data display in the Plot dialog box. There are four different plotting symbols, one for each pattern. Generally, pattern 1 has the largest side shift and pattern 2 the smallest, but there is no dependency on time.
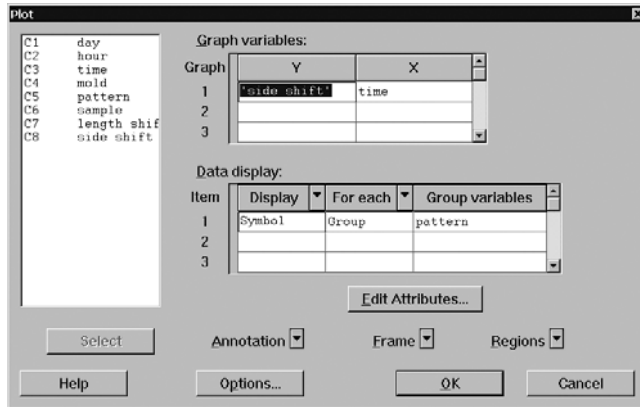
Figure C.10   (Scatter) Plot dialog box showing a group variable.
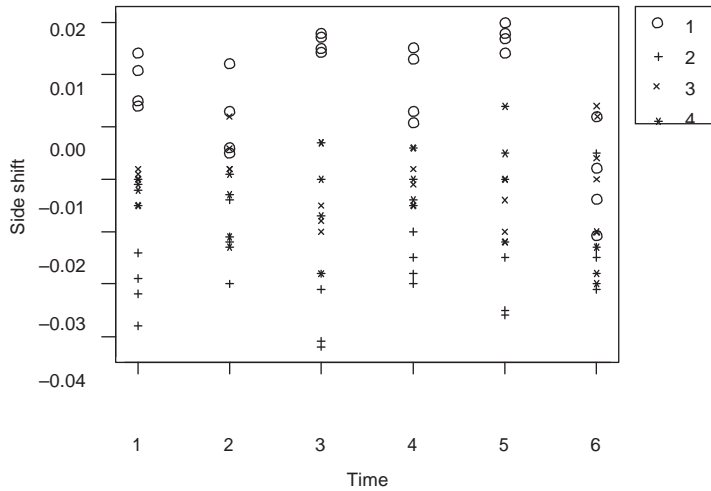
Figure C.11   Scatter plot of side shift by time with labels for different patterns.

Draftsman and Matrix Plots

To create many scatter plots simultaneously, MINITAB provides a way to produce multiple graphs with defined output and input characteristics. A draftsman plot is available through the MINITAB menu selection:

*Graph → Plot → Draftsman Plot [*option *each Y versus each X* under *Graph □ Matrix Plot* in release 14.11]*

The draftsman plot automatically produces all scatter plots that involve the output characteristic (*Y* variable) and the list of input characteristics (X variables). Figure C.12 shows the dialog box.
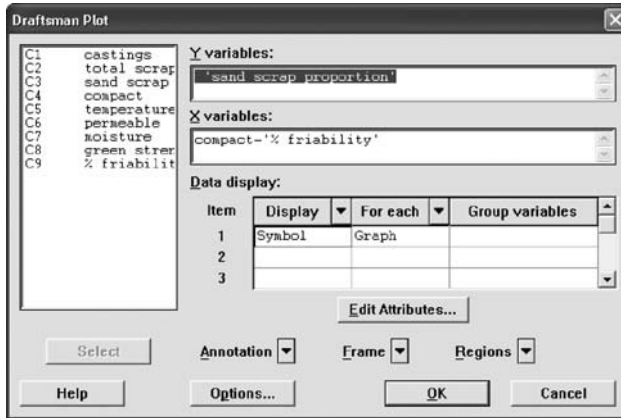
Figure C.12   Draftsman Plot dialog box.

In Figure C.13, we see no clear relationship between the sand scrap proportion and various sand characteristics. Upon closer inspection, however, a quadratic relationship between temperature and sand scrap proportion becomes clear. This example illustrates the point that if too many characteristics are chosen, draftsman plots (and matrix scatter plots, as will be discussed) will be hard to interpret, since each individual plot will be very small. In that case we look at the scatter plots one at a time.

A matrix plot is another way of producing multiple scatter plots simultaneously. Unlike with a draftsman plot, we now make no distinction between inputs and outputs. The window leaks problem discussed in Chapter 12 provides an example, where we also label the
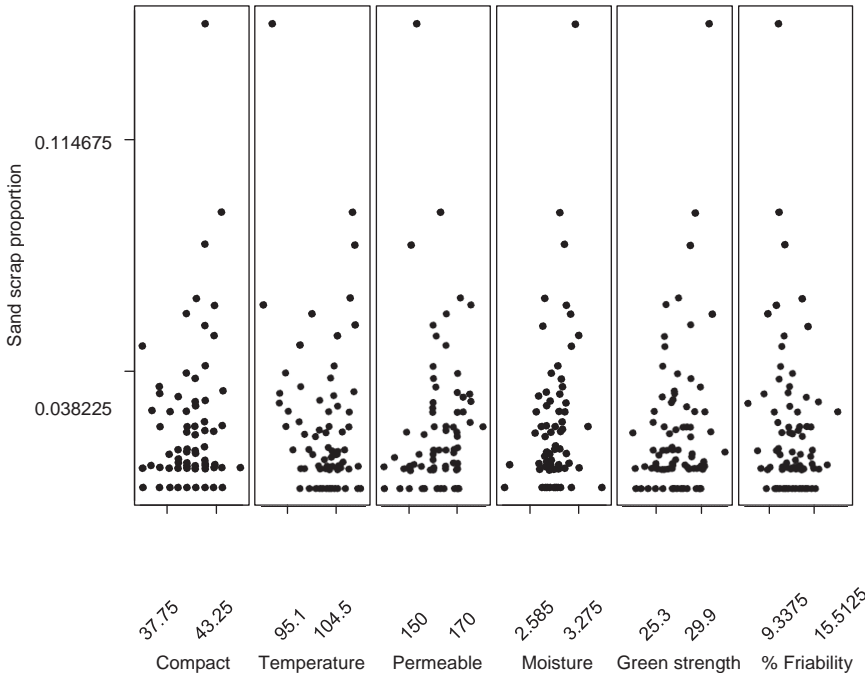
Figure C.13   Draftsman plot.

plotted points to show leakers and nonleakers. The data are given in the file *window leaks comparison*. A matrix scatter plot is available using:

Graph → Plot → Matrix Plot

Selecting all the inputs and using *class* as a group variable, as shown in Figure C.14, we obtain the matrix scatter plot in Figure C.15.
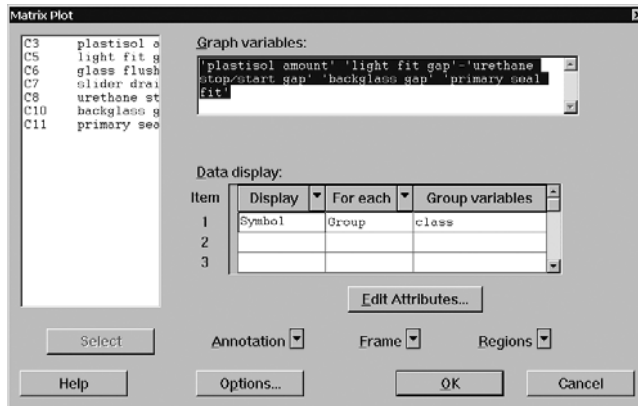
Figure C.14   Matrix (scatter) Plot dialog window.
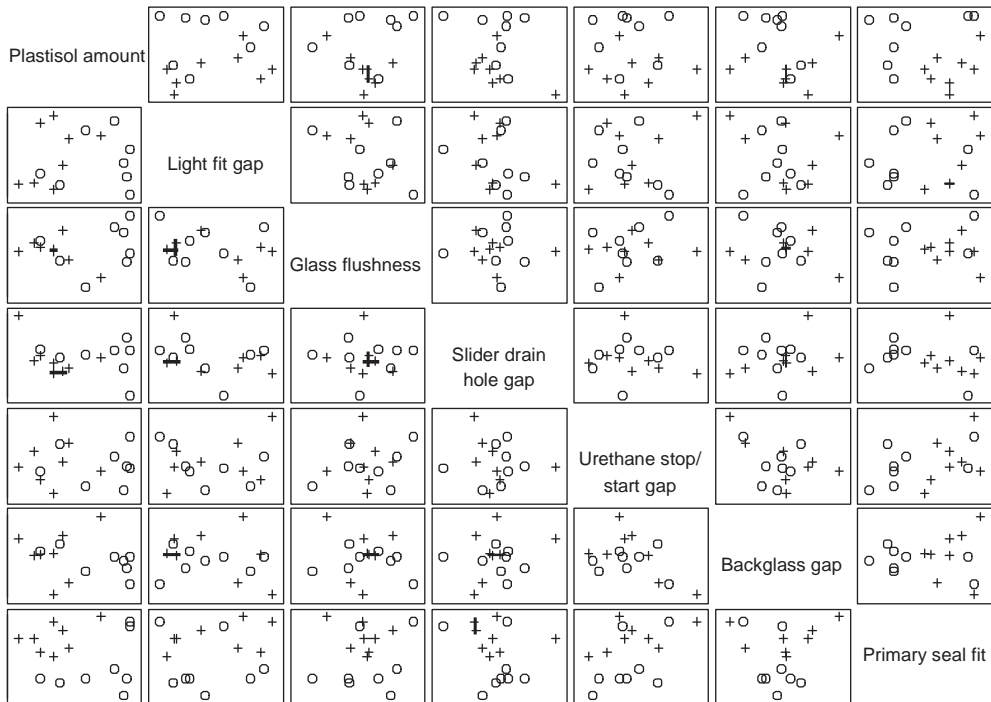
Figure C.15   Matrix scatter plot with labeled points.

Leaking windows are shown with circles, and nonleakers are shown with plus signs. There is a separation between leakers and nonleakers if we look simultaneously at the two inputs: plastisol amount and primary seal fit. The plot in the lower left corner of Figure C.15 is reproduced in a larger format in the right panel of Figure 12.2.

C.5 MULTIVARI CHART

A multivari chart displays the variation in the output due to families labeled by a discrete or categorical input. Multivari charts are available using:

Stat → Quality Tools → Multi-Vari Chart

The background and multivari investigation for the cylinder head scrap problem are described in Chapter 11. The side and end shift data for the 96 measured parts are stored in the file *cylinder head scrap multivari*.

To create a multivari chart, specify an output characteristic (called *response* in MINITAB) and up to four inputs (factor 1, factor 2, and so on). The order of the inputs affects the graphical display. Choose either time or the characteristic with the most different levels as the last factor, because this factor defines the labels on the horizontal axis.

Figure C.16 shows the Multi-Vari Chart dialog box. We recommend always using the Options button to check *Display individual data points*.
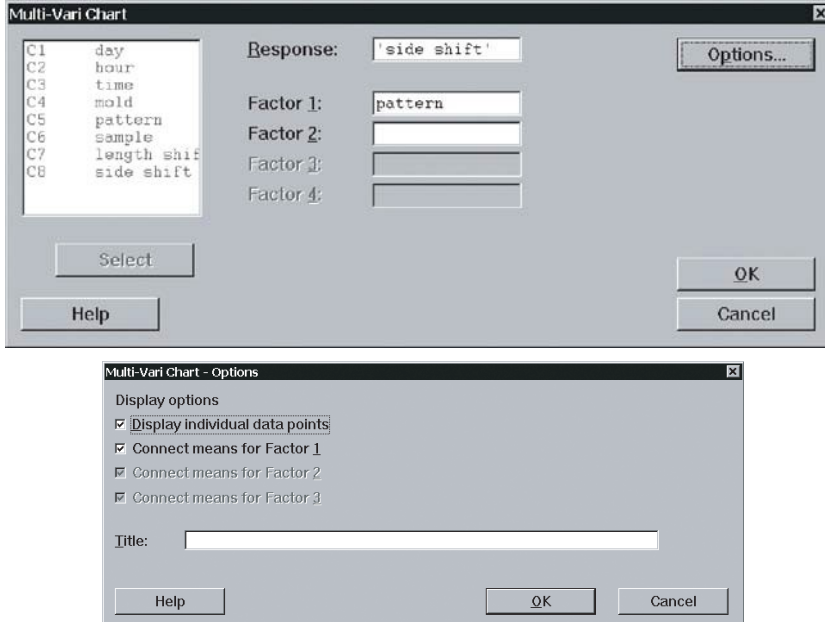
Figure C.16   Multi-Vari Chart dialog box with Options window.
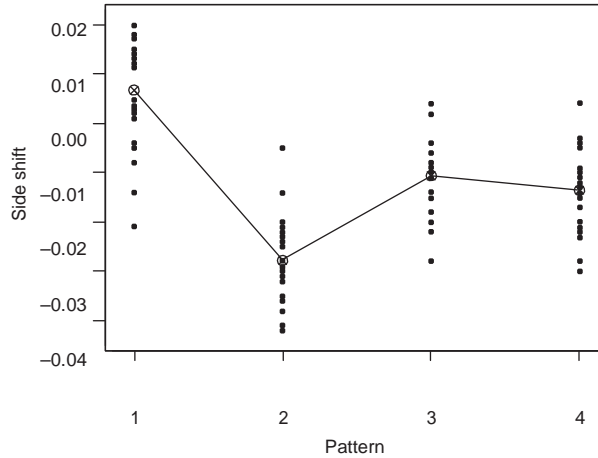
Figure C.17   Multivari chart for cylinder head scrap example.

Using the selections made in Figure C.16 results in the multivari chart given in Figure C.17.

Figure C.17 shows large systematic differences among the patterns (cavities). The variation within each pattern is about 60% of the overall variation.

The dialog box in Figure C.18 creates a multivari chart involving two inputs—time and pattern. From the resulting chart in Figure C.19, we conclude that the dominant cause acts only in the pattern-to-pattern family and does not involve the time-to-time family.

A multivari chart with three or four inputs can be difficult to interpret. Using a variety of different charts and orders for the inputs may help.
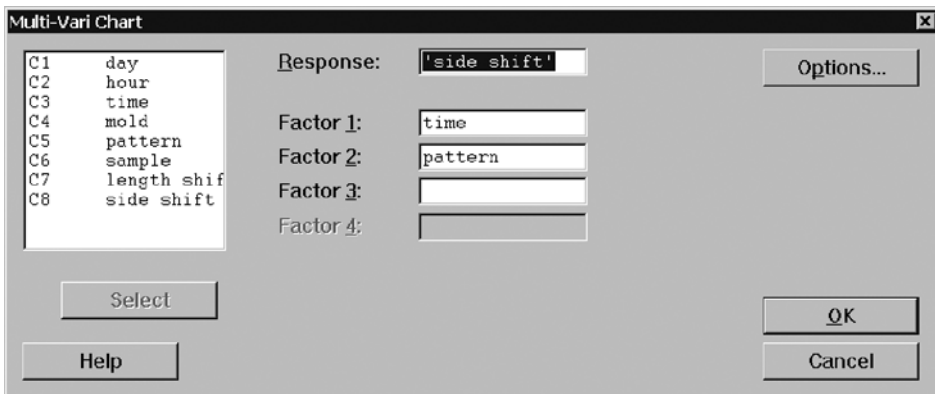
Figure C.18   Multi-Vari Chart dialog box showing two inputs.
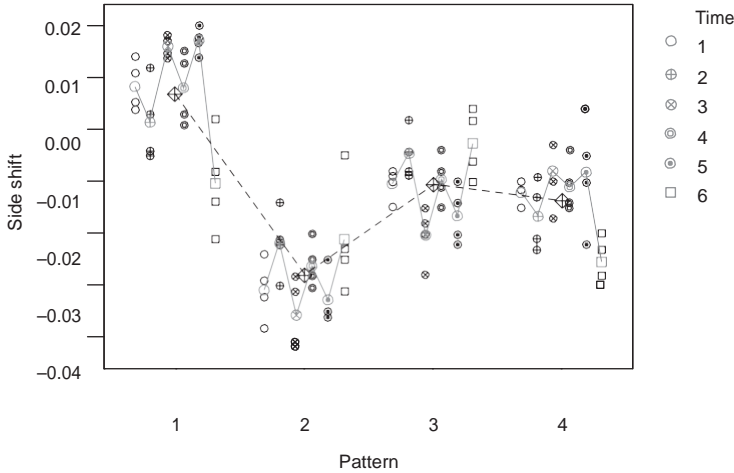
Figure C.19   Multivari chart of side shift versus pattern and time.

Exploring Haphazard Families

Consider the casting thickness example introduced in the supplement to Chapter 11, where the casting-to-casting family has haphazard effect. We want to determine if changing one of the other inputs results in changes in variation within different levels of the haphazard family.

To conduct the analysis, we first define a new output based on the standard deviation of original output within each level of group, where group = (time – 1)*24 + (cavity – 1)*4 + position. The variation within each group is due to causes acting in the casting-to-casting family. We use the MINITAB calculator, described in Appendix A, to define group. Next, we define a new output characteristic using:

Stat → Basic Statistics → Store Descriptive Statistics

The dialog box is shown in Figure C.20. This creates two new columns in the MINITAB worksheet. The first column (labeled "ByVar1") gives the group number, while the second column gives the within-group standard deviation. To ease interpretation of the subsequent analysis, we relabel the second new column "group stdev." Part of the resulting MINITAB worksheet is shown in Figure C.21.
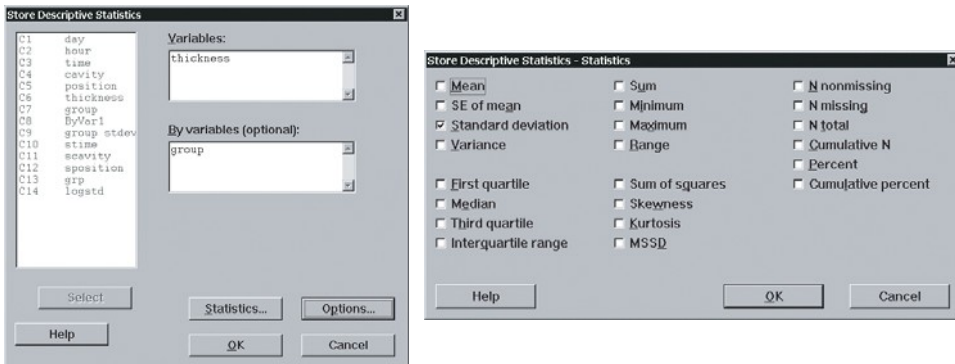
Figure C.20   MINITAB dialog box to store group standard deviation.

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9
	day	hour	time	cavity	position	thickness	group	ByVar1	group stdev
1	1	1	1	1	1	-3	1	1	1.0000
2	1	1	1	1	1	-2	1	2	1.0000
3	1	1	1	1	1	-1	1	3	3.5119
4	1	1	1	1	2	-3	2	4	1.5275
5	1	1	1	1	2	-1	2	5	2.0000
6	1	1	1	1	2	-2	2	6	4.5826
7	1	1	1	1	3	-2	3	7	4.0415
8	1	1	1	1	3	-6	3	8	8.6217
9	1	1	1	1	3	-9	3	9	1.5275
10	1	1	1	1	4	9	4	10	2.0000
11	1	1	1	1	4	8	4	11	6.0828
12	1	1	1	1	4	6	4	12	2.0817
13	1	1	1	2	1	-5	5	13	1.1547
14	1	1	1	2	1	-7	5	14	0.0000
15	1	1	1	2	1	-3	5	15	6.0277
16	1	1	1	2	2	-4	6	16	5.5076
17	1	1	1	2	2	-10	6	17	1.5275
18	1	1	1	2	2	-1	6	18	2.6458
19	1	1	1	2	3	-6	7	19	5.1316
20	1	1	1	2	3	-9	7	20	1.7321

Figure C.21   MINITAB worksheet showing group standard deviation.

To create the multivari charts for the group standard deviation, we also need to define new input columns that give the values of the original inputs for each level of the characteristic group. This is best accomplished using:

Manip → Code → Use Conversion Table

We define a new input column for cavity as shown in Figure C.22. We label the new input column "scavity," since it tells us the value of the cavity that corresponds to the group standard deviation column. We similarly define "stime" and "sposition."

With this preliminary work, we can now create the desired multivari charts for *group stdev*.
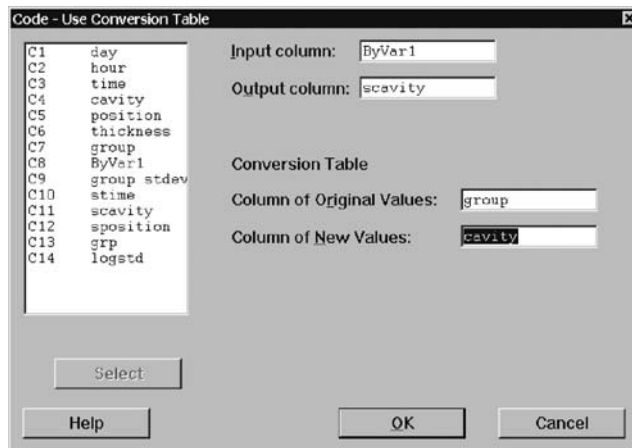
Figure C.22   Defining a new input column for cavity.

C.6 EXPONENTIAL SMOOTHING

When determining feasibility and deciding how to implement a feedback controller, we can fit an EWMA (exponential smoother) to an equally spaced time series of output values. See Chapter 18 and its supplement for details.

As shown in Figure C.23, to fit an EWMA we use:

Stat → Time Series → Single Exp Smoothing

To illustrate, we use the fascia film build example explored in Chapter 18. The data are given in the file *fascia film build baseline*. A plot of the results of the EWMA smoothing is given in Figure C.24.
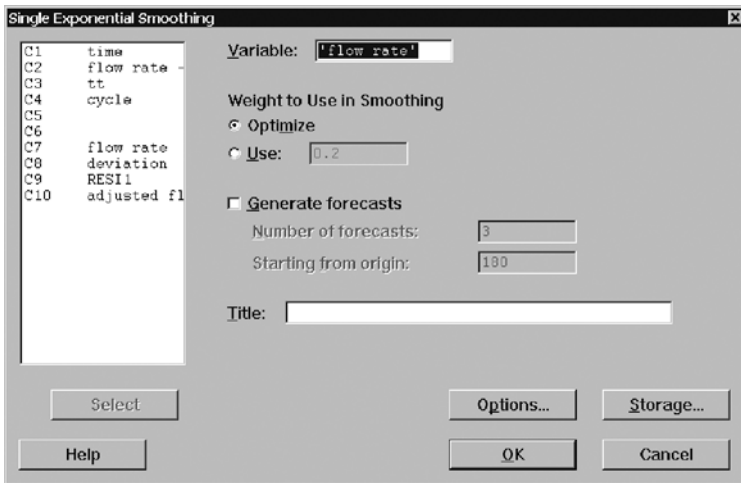
Figure C.23   MINITAB Single Exponential Smoothing dialog box.

We see that the time-to-time variation is captured well by the exponential smoothing. The smoothing constant (alpha) is 0.17, and the standard deviation of the smoothed series is given by the square root of the mean squared deviation (MSD). In the example, we have $\sqrt{3.83} = 1.96$.
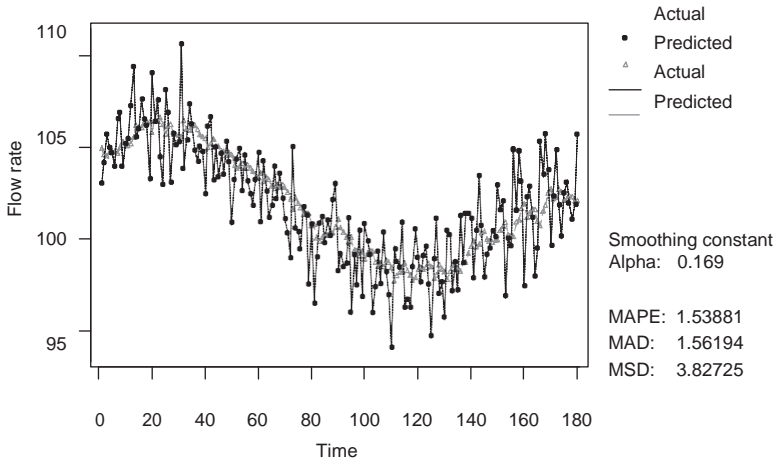
Figure C.24   EWMA smoothing of flow rate.

We can also use MINITAB to determine how the adjusted series will look (assuming no adjustment errors) by calculating the output minus the target value and fitting the single exponential smoother to the translated output. Saving the fitted values (using the Storage button from Figure C.23), we can determine the adjusted series by subtracting the fitted values from the original film build values.

Appendix D

Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is the main formal numerical analysis tool associated with the search for a dominant cause of variation. While in most investigations we draw conclusions from graphical displays, we can use ANOVA to:

- Assess a measurement system (Chapter 7).

- Compare contributions of two families to output variation (Chapter 10).

- Supplement the analysis of multivari investigations (Chapter 11).

- Set a new goal for a reformulated problem (Chapter 14).

ANOVA is applicable when the output is continuous and we have one or more discrete inputs. See Box et al. (1978) and Neter et al. (1996). Here we describe a few forms of ANOVA that we may need when applying the Statistical Engineering algorithm.

ANOVA partitions the overall variation (as quantified by the total sum of squares) into components attributable to various inputs or families. We recommend using ANOVA as a supplement to graphical displays if the results are unclear.

D.1 ONE-WAY ANOVA

A one-way ANOVA is appropriate for a continuous output and a single discrete input. We use the MINITAB menu selection:

Stat → ANOVA → One-way

Consider the cylinder head scrap example discussed in Chapter 11. The data are given in the file *cylinder head scrap multivari*. Figure D.1 shows a plot of the side shift (the output) versus pattern (an input). With ANOVA, we quantify how much of the variation in the side shift can be explained by differences among the four mold patterns.
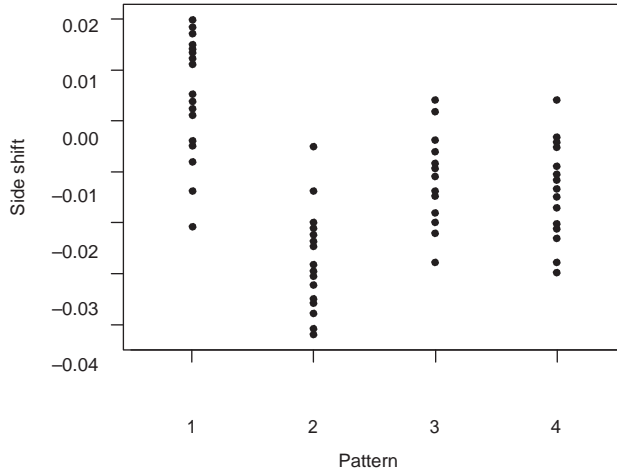
Figure D.1   Plot of side shift versus pattern.

Using the dialog box shown in Figure D.2, we get the ANOVA results:

One-way ANOVA: side shift versus pattern

Analysis of Variance for side shift

Source	DF	SS	MS	F	P
pattern	3	0.0146454	0.0048818	60.56	0.000
Error	92	0.0074168	0.0000806		
Total	95	0.0220622			

Figure D.2   MINITAB One-way ANOVA dialog box.

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	--------+---------+---------+---------
1	24	0.006833	0.011001	(--*--)
2	24	–2.8E–02	0.008371	(--*--)
3	24	–1.1E–02	0.007812	(--*--)
4	24	–1.4E–02	0.008387	(--*--)

-------+---------+---------+---------

Pooled StDev = 0.008979   –0.024   –0.012   –0.000

Pooled StDev provides an estimate of the variation within each pattern pooled across all patterns. That is, Pooled StDev provides an estimate of remaining output variation if we could eliminate all pattern-to-pattern differences in the average output. It estimates the variation attributable to all families other than the pattern-to-pattern family.

We can also use a one-way ANOVA model to estimate the variation due to the measurement system. In Chapter 7, we described a measurement investigation for the camshaft journal diameter example. The data are given in the file *camshaft journal diameter measurement*. A one-way ANOVA by part is:

One-way ANOVA: diameter versus part

Analysis of Variance for diameter

Source	DF	SS	MS	F	P
part	2	5008.014	2504.007	4383.00	0.000
Error	51	29.136	0.571		
Total	53	5037.151			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev	-------+---------+---------+---------
1	18	–10.901	0.638	*)
2	18	1.204	0.796	(*
3	18	12.685	0.820	*)

-------+---------+---------+---------

Pooled StDev =   0.756   -7.0   0.0   7.0

The measurement variation is estimated as 0.756, as given by Pooled StDev. This matches the result obtained in Chapter 7.

In interpreting the ANOVA results, we ignore the *p*-value and confidence intervals (CIs), which are formal statistical procedures. See the supplement to Chapter 10 for our reasons.

D.2 ANOVA FOR TWO OR MORE INPUTS

To fit an ANOVA model with two or more inputs we use the MINITAB menu selection:

   Stat → ANOVA → Balanced ANOVA

To illustrate, we consider the casting thickness example previously discussed in the supple-
ment to Chapter 11. The data are given in the file *casting thickness multivari*.

   The MINITAB Balanced ANOVA dialog box is shown in Figure D.3. The notation
shown in Figure D.3, using the symbol "|" between inputs, requests a model with all possi-
ble interaction terms. In the casting thickness example, this includes up to the three-way
interaction involving all of time, cavity, and position.
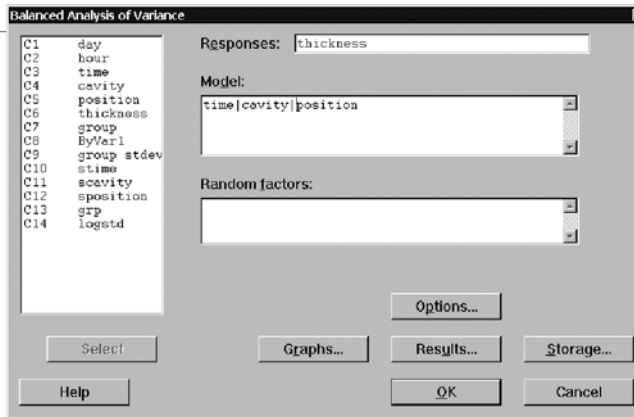
Figure D.3   MINITAB Balanced ANOVA dialog box.

The MINITAB results are:

ANOVA: thickness versus time, cavity, position

Factor	Type	Levels	Values						
time	fixed	12	1	2	3	4	5	6	7
			8	9	10	11	12		
cavity	fixed	6	1	2	3	4	5	6	
position	fixed	4	1	2	3	4			

Analysis of Variance for thickness

Source	DF	SS	MS	F	P
time	11	9008.82	818.98	31.81	0.000
cavity	5	16994.99	3399.00	132.00	0.000
position	3	16697.24	5565.75	216.15	0.000
time*cavity	55	1544.02	28.07	1.09	0.311
time*position	33	7580.19	229.70	8.92	0.000
cavity*position	15	5363.43	357.56	13.89	0.000
time*cavity*position	165	1634.06	9.90	0.38	1.000
Error	576	14832.00	25.75		
Total	863	73654.75			

To compare the relative sizes of the families, we look at the sum of squares (SS) column. We see large, roughly equal-sized effects due to position, cavity, and error. The error sum of squares includes the effect of the casting-to-casting family and all interactions between the casting-to-casting family and the other families. The calculated sum of squares tells us approximately how much we could expect to reduce the total sum of squares if we could eliminate all variation due to the given family. It is complicated to translate these sums of squares into estimates for the standard deviation attributable to each family, but there is a rough correspondence. See the supplement to Chapter 11 for a complete discussion of the casting thickness example.

If some data are lost, the assumption of balance (equal number of observations at each level of each input) is violated and the balanced ANOVA analysis is no longer appropriate. For unbalanced data, we use a general linear model:

Stat → ANOVA → General Linear Model

Appendix E

Regression Models and Analysis

With regression, we model the relationship between an output characteristic and one or more inputs. We find regression models useful for the following tasks:

- Investigate variation transmission (chapters 10 and 11).
- Investigate the relationship between the output and inputs (Chapter 12).
- Set the goal for a reformulated problem (Chapter 14).
- Find a prediction equation for a feedforward controller (Chapter 17).
- Calibrate an adjuster (Chapter 18).

There are many good references. See Box et al. (1978), Ryan (1989), and Montgomery et al. (2001).

E.1 REGRESSION WITH A SINGLE INPUT

A regression model relating the output to a single input is:

$$output = b_0 + b_1 input + residual$$

The term $b_0 + b_1 input$ captures the effect of the input, and *residual* describes the variation in the output due to all other inputs. The regression analysis consists of estimating (also called *fitting*) the unknown constants b_0 and b_1 and the residual variation.

In MINITAB, regression analysis is available using:

Stat → Regression → Regression
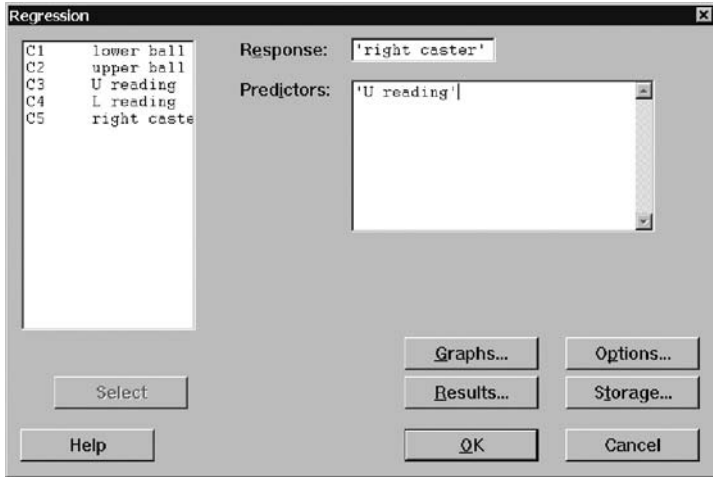
Figure E.1 Regression dialog box.

To illustrate, we use the truck alignment example introduced in Chapter 1 and discussed in Chapter 12. The data are given in the file *truck pull input-output*. Figure E.1 is the MINITAB Regression dialog box, used to fit a regression model with output right caster and input U-reading.

The corresponding MINITAB regression results are:

Regression Analysis: right caster versus U reading

The regression equation is
right caster = 4.55 + 0.171 U reading

Predictor	Coef	SE Coef	T	P
Constant	4.55157	0.03544	128.44	0.000
U reading	0.17115	0.05309	3.22	0.003

S = 0.1926   R-Sq = 27.1%   R-Sq(adj) = 24.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.38566	0.38566	10.39	0.003
Residual Error	28	1.03893	0.03710		
Total	29	1.42459			

Unusual Observations

Obs	U readin	right ca	Fit	SE Fit	Residual	St Resid
22	2.09	5.1663	4.9089	0.1121	0.2574	1.64 X

X denotes an observation whose X value gives it large influence.
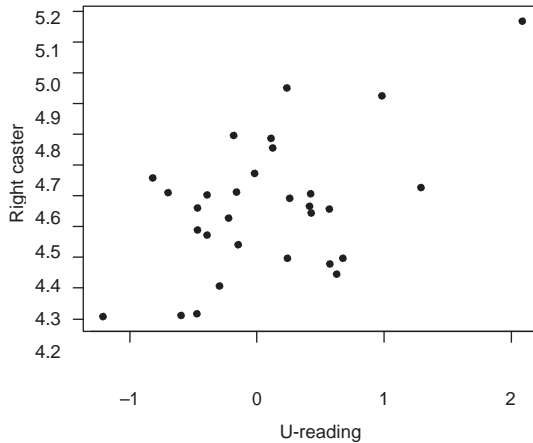
Figure E.2   Scatter plot of right caster versus U-reading.

The regression equation, right caster = 4.55 + 0.17*U-reading, provides a summary of the relationship between right caster and U-reading. In other words, the model suggest that a one-unit increase in U-reading would, on average, result in a 0.17-unit increase in right caster. However, the scatter plot in Figure E.2 suggests the regression model is not a very good summary. The value of $S = 0.193$ provides an estimate of the residual variation, the standard deviation in right caster if we could hold U-reading fixed. Holding the input U-reading fixed, we could reduce the variation in caster from around 0.22 (the baseline variation and, as it happens, also the caster variation in the regression data) to 0.19. We conclude that U-reading is not a dominant cause.

In the MINITAB regression results, the list of unusual observations flags outliers, observations that are either not close to the fitted regression line or that have a large influence on the estimated regression parameters. In this case, the one observation in the upper right-hand corner of the plot in Figure E.2 has been flagged because it has a large influence on the slope of the fitted line.

Fitted Line Plot

To superimpose the fitted regression equation on a scatter plot of the output versus the input, we use:

Stat → Regression → Fitted Line Plot

We illustrate using the crossbar dimension example covered in Chapter 12. The data from the investigation are given in the file *crossbar dimension input-output*. The left panel in Figure E.3 shows the resulting graph and numerical regression model summary when fitting a model for dimension as a linear function of barrel temperature. We see that the linear model fits quite well and (since the dimension variation in the regression data matches the baseline) that barrel temperature is a dominant cause. However, we also notice that the relationship between dimension and barrel temperature seems to be nonlinear.
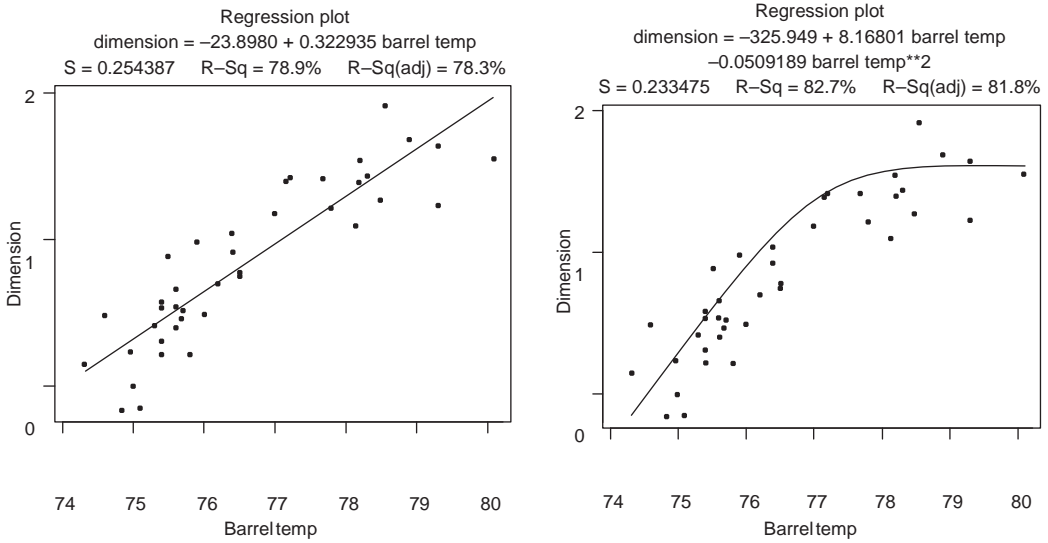
Figure E.3 Regression plots with fitted lines.
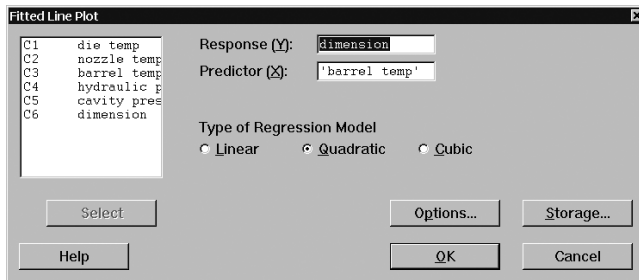
Figure E.4 Fitted Line Plot dialog box.

Using the fitted line plot dialog box as shown in Figure E.4, where we request a quadratic model, we get the plot given in the right panel of Figure E.3. In this way, we can explore the relationship between the input and output. Using the Options button we can also request a log transformation of the input, output, or both.

Best Subsets Regression

To quickly fit many possible models, we use automatic model building. See MINITAB help or Neter et al. (1996) for more details.

In investigations where many inputs are measured, the Best Subsets Regression dialog box, as shown in Figure E.5, can quickly summarize the results from fitting regression models with all possible single inputs.

Best subsets regression is available using:

Stat → Regression → Best Subsets

Figure E.5   Best Subsets Regression dialog box.

We use the best subsets regression options, as shown in the right panel of Figure E.5, to request a minimum and maximum of one free predictor (that is, input) and to display five models of each size to print. To illustrate, we use the brake rotor balance example discussed in one of the case studies. In a group comparison investigation, 26 foundry-related inputs were measured on 30 balance rejects and 30 well-balanced brake rotors. The data are given in the file *brake rotor balance comparison*. Previous analysis for this data was based on the binary output: balanced or not balanced. However, since the balance weight required for each rotor was also recorded, we can also analyze these data using regression models. Filling in the Best Subsets Regression dialog box as in Figure E.5 gives the following (edited) MINITAB results:

Best Subsets Regression: balance versus v1, v2, …

Response is balance

```
                                    o i     i i o o       h h h h h
                                    u n o o o n n u u h h h h o o o o o
                                    t n v f f n n t t e e e e l l l l l
                                    e e e f f e e e e i i i i e e e e e
                                    r r r s s r r r r g g g g
                                    a e e         h h h h s s s s s
                              v v v v v v v v v t t l t t d d d d t t t t i i i i i
R-Sq  R-Sq(adj)  C-p      S    1 2 3 4 5 6 7 8 h h l 1 2 i i i i 1 2 3 4 z z z z z

62.6    61.9     2.4   0.30960                       X
48.2    47.3    24.9   0.36435              X
44.1    43.1    31.3   0.37863                   X
34.7    33.6    45.9   0.40900                 X
 5.5     3.9    91.5   0.49213          X
```

The results summarize the five best regression models that involve only a single input. The models are ranked by largest R-Sq value (or equivalently, since all models have the same number of parameters, by the smallest value for *S*). Overall thickness variation is the single input that explains the most variation in the output. We would then look at the scatter plots that correspond to the best inputs. To determine the regression equation for any particular input, we need to fit the individual regression model.

Using the best subsets regression routine, we can avoid examining all scatter plots (and fitting all the corresponding regression models) involving the output and all possible single inputs. There are 26 such plots (models) in this case. We do not recommend the use of the best subsets regression routine unless the number of inputs is large. Looking at the individual plots is preferred because the regression summary can miss patterns like nonlinear relationships and the effect of outliers.

E.2 REGRESSION WITH MULTIPLE INPUTS

Fitting regression models with more than one input can be useful:

- As a screening tool to analyze the results of investigations to look at the relationship between inputs and the output (Chapter 12)

- For fitting a prediction equation for a feedforward controller (Chapter 17)

In cases where we have many input characteristics the regression model is extended to

$$output = b_0 + b_1 input_1 + ... + b_k input_k + residual$$

Regression analysis with multiple inputs is available using:

Stat → Regression → Regression

To illustrate, consider the truck pull feedforward investigation discussed in Chapter 17. The goal of the investigation was to find a prediction equation for caster that could be used to build a feedforward controller. The data are given in the file *truck pull feedforward*. We fit a regression model to describe the relationship between the output left caster and the inputs given by the truck frame geometry using the Regression dialog box, as shown in Figure E.6.
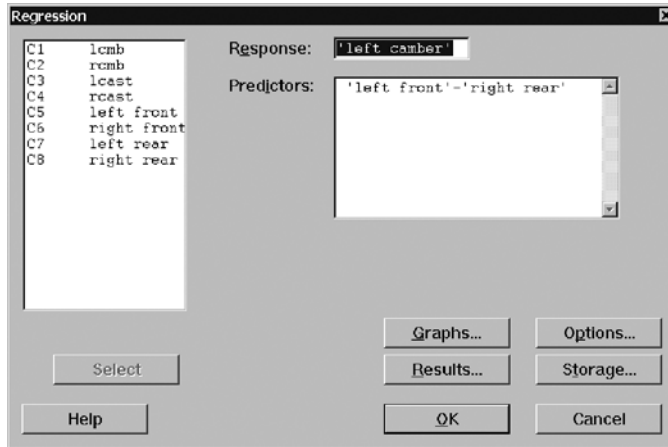
Figure E.6   Regression dialog box showing multiple inputs.

The MINITAB results are:

The regression equation is
left caster =   –18.6 + 1.24 left front + 0.677 right front + 0.140 left rear +
1.156 right rear

Predictor	Coef	SE Coef	T	P
Constant	–18.6097	0.6100	–30.51	0.000
left front	1.24243	0.04018	30.92	0.000
right front	0.67669	0.03585	18.88	0.000
left rear	0.13953	0.02763	5.05	0.000
right rear	0.15624	0.04162	3.75	0.000

S = 0.1760   R-Sq = 96.4%   R-Sq(adj) = 96.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	77.827	19.457	627.82	0.000
Residual Error	95	2.944	0.031		
Total	99	80.771			

In this example, the regression equation summarizes the relationship between the left caster and the four truck frame geometry inputs. The value of S in the regression results gives an estimate of the remaining variation in left caster (quantified in terms of standard deviation) if we hold *all* of the four truck frame inputs fixed or if we could perfectly compensate for their effects using a feedforward controller.

For regression models with many inputs, we need to use caution when trying to interpret the estimated regression parameters (given by the "Coef" column in the MINITAB regression results). We may interpret the parameter estimate corresponding to the input "left front" (1.24) as the expected change in the left caster for a unit change in left caster if we *hold all the other inputs fixed*. If the inputs used in the regression analysis do not vary independently, it may make no physical sense to think of changing one input while holding all the others fixed. In the truck pull example, where we want a prediction equation, we do not interpret the individual regression parameter estimates, so this is not a major concern.

Appendix F

Planning and Analysis of Designed Experiments

Designed experiments are used in the Statistical Engineering algorithm to:

- Verify a dominant cause (Chapter 13).

- Find adjusters that move the process center (Chapter 15).

- Find candidate settings that desensitize the process to variation in a dominant cause (Chapter 16).

- Find candidate settings that make the process robust (Chapter 19).

There is an extensive literature on designed experiments. See, for example, Box et al. (1978) and Montgomery (2001). In this appendix we show how to set up and analyze the types of experiments used in this book.

F.1   SETTING UP THE EXPERIMENTAL DESIGN

In this section, we consider the design of factorial experiments with two or more levels for each input. In MINITAB, to create a factorial design we use the menu selection:

Stat → DOE → Factorial → Create Factorial Design

To illustrate, we use the brake rotor balance verification experiment discussed in Chapter 13 and in one of the case studies. In the experiment, two levels for each of the three suspects—tooling, position, and thickness variation—were chosen to capture their full range of variation. The team decided to make eight rotors for each of the eight treatments. That is, there were eight runs with eight repeats. No treatment was replicated. Figure F.1 shows how the experimental plan was entered into MINITAB. In MINITAB, the inputs that are changed in an experiment are called *factors*.
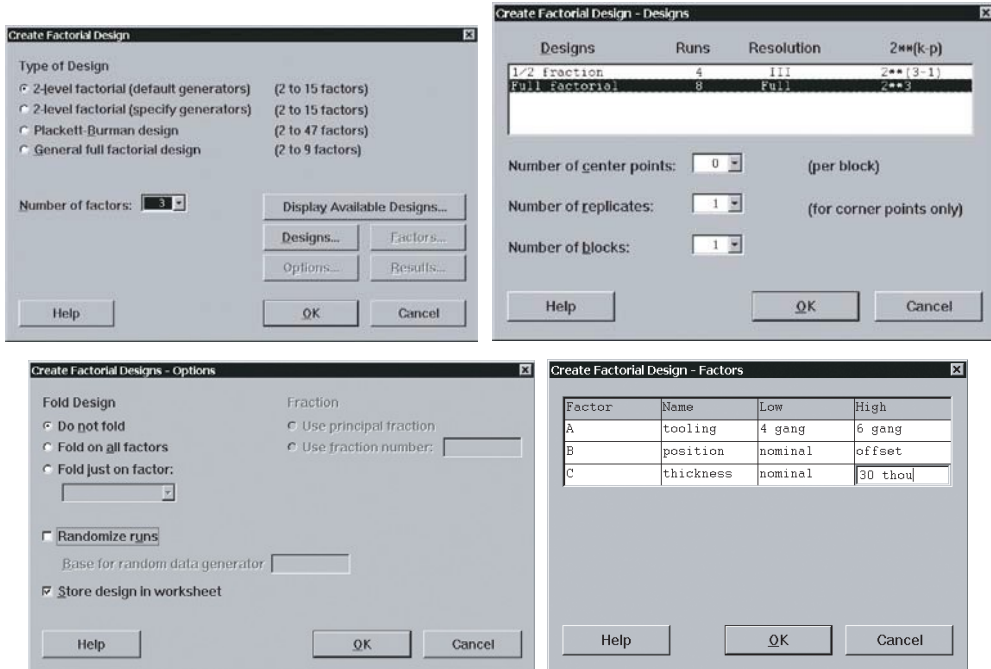
Figure F.1   Create Factorial Design dialog box.

For the choices shown in Figure F.1, MINITAB gives the summary:

Factorial Design

Full Factorial Design

Factors:	3	**Base Design:**	**3, 8**
Runs:	8	**Replicates:**	1
Blocks:	none	**Center pts (total):**	0

All terms are free from aliasing

The plan and the collected data are given in the file *brake rotor balance verification*. The worksheet with the data is shown in Figure F.2. Note that MINITAB has created four special columns in the worksheet: StdOrder, RunOrder, CenterPt, and Blocks. These columns are necessary to use any of the MINITAB subcommands under Stat ☐ DOE ☐ Factorial once the design has been created.

In Figure F.2, the columns r1, r2, ..., r8 give the balance weights for the eight rotors produced for each treatment. Much of the analysis in the brake rotor example uses the performance measure *average weight*. The average weight for each run is calculated as described in Appendix A. For experiments with a number of repeats for each run, such as the brake rotor balance verification experiment, we need to store the data in a different way

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8	C9	C10	C11	C12	C13	C14
	StdOrder	RunOrder	CenterPt	Blocks	Tooling	Position	Thickness	r1	r2	r3	r4	r5	r6	r7
1	1	8	1	1	4 gang	offset	30 thou	0.46	0.51	0.39	0.68	0.86	0.61	0.5
2	5	1	1	1	4 gang	offset	nominal	0.16	0.18	0.16	0.22	0.10	0.24	0.0
3	3	3	1	1	4 gang	nominal	30 thou	0.00	0.97	0.42	0.38	0.00	0.65	0.6
4	7	7	1	1	4 gang	nominal	nominal	0.11	0.09	0.09	0.09	0.04	0.08	0.0
5	2	2	1	1	6 gang	offset	30 thou	1.56	1.95	1.27	1.40	1.48	1.70	1.3
6	6	5	1	1	6 gang	offset	nominal	0.16	0.44	0.39	0.42	0.47	0.22	0.4
7	4	4	1	1	6 gang	nominal	30 thou	0.83	1.61	1.44	1.49	1.21	0.91	1.3
8	8	6	1	1	6 gang	nominal	nominal	0.04	0.03	0.04	0.03	0.03	0.05	0.0
9														
10														

Figure F.2   Experimental design setup in MINITAB worksheet.

to plot the output by treatment. Putting all the output values (that is, repeats) in a single col-
umn (use the stack columns command as described in Section A.5), we get the worksheet
in Figure A.9.

Custom Designs

In some cases, the experiment we wish to conduct is not a standard two-level factorial
design, or we already have the experimental data stored. In such cases, we can set up the
experiment in MINITAB using:

Stat → DOE → Factorial → Define Custom Factorial Design

For instance, in the oil pan scrap example discussed in Chapter 16, the experiment had four
inputs, with one at three levels and the other three at two levels each. We enter the design
as in Figure F.3. For two-level factorial designs, we need to also use the Low/High dialog
box to tell MINITAB what codes correspond to the high and low level of each input.

Entering the design as shown in Figure F.3 adds the columns StdOrder, RunOrder,
CenterPt, and Blocks to the existing data as shown in Figure F.4. The data are available in
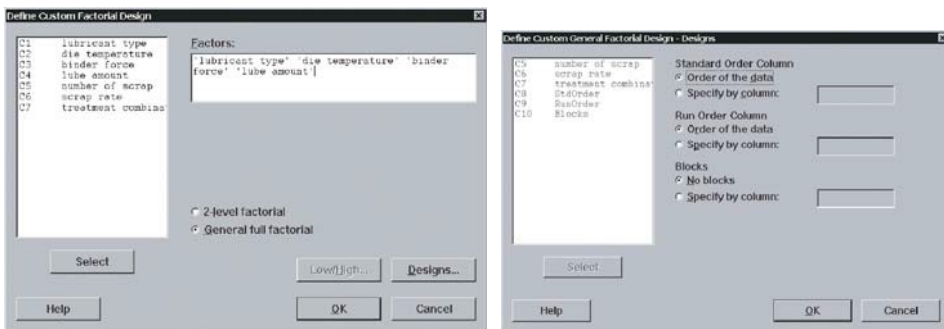the file *oil pan scrap desensitization.*

Figure F.3   Define Custom Factorial Design dialog box.

Figure F.4   Results from custom design setup.

For a proper interpretation of this experiment, recall that this is a desensitization experiment. Three of the inputs—namely, lubricant type, die temperature, and binder force—are candidates, while lube amount is the dominant cause. As shown in Figure F.4, we number the treatments based on combinations of the candidates.

Fractional Factorial Designs

Designs for fractional factorial experiments (where only a subset of all possible treatment combinations are used—see the supplement to Chapter 15) are set up in MINITAB with the same command as full factorial designs, namely,

$Stat \rightarrow DOE \rightarrow Factorial \rightarrow Create\ Factorial\ Design$

To plan a fractional factorial experiment, useful information about all possible 2^k factorial designs is given in MINITAB using the Display Available Designs button, as shown in Figure F.5.

To illustrate a fractional factorial experiment, we use the paint film build example discussed in the Chapter 19 exercises. The data are given in the file *paint film build robustness*. The goal was to explore the effect of five inputs on the variation in film build (as measured by the log standard deviation) over five consecutive panels. There were resources for 16 runs, a half fraction. We choose the design in MINITAB as shown in Figure F.6. We also used the Factors button to enter input names and levels.

Figure F.7 gives the resulting worksheet after adding appropriate labels, entering the experimental results, and calculating some performance measures. Performance measures were determined using the calculator function in MINITAB (see Appendix A).
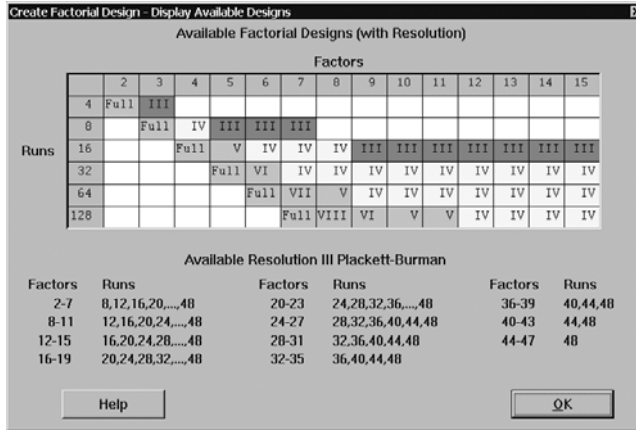
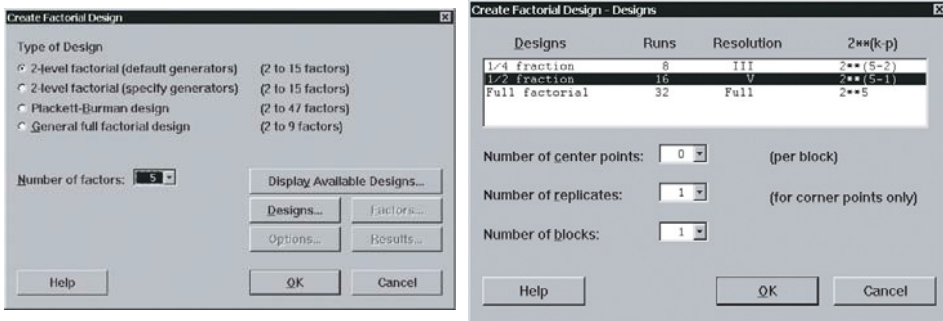**Figure F.5**   Two-level factorial experiments available in MINITAB.

Figure F.6   Design setup for the paint film build robustness experiment.

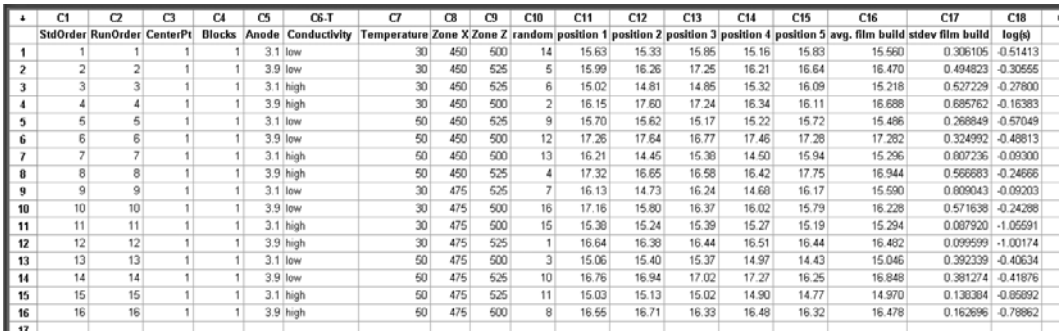|	C1	C2	C3	C4	C5	C6-T	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	StdOrder	RunOrder	CenterPt	Blocks	Anode	Conductivity	Temperature	Zone X	Zone Z	random	position 1	position 2	position 3	position 4	position 5	avg. film build	stdev film build	log(s)
1	1	1	1	1	3.1	low	30	450	500	14	15.63	15.33	15.85	15.16	15.83	15.560	0.306105	-0.51413
2	2	2	1	1	3.9	low	30	450	525	5	15.99	16.26	17.25	16.21	16.64	16.470	0.494823	-0.30555
3	3	3	1	1	3.1	high	30	450	525	6	15.02	14.81	14.86	15.32	16.09	15.218	0.527229	-0.27800
4	4	4	1	1	3.9	high	30	450	500	2	16.15	17.60	17.24	16.34	16.11	16.688	0.685762	-0.16383
5	5	5	1	1	3.1	low	50	450	525	9	15.70	15.62	15.17	15.22	15.72	15.486	0.266849	-0.57049
6	6	6	1	1	3.9	low	50	450	500	12	17.26	17.64	16.77	17.46	17.28	17.282	0.324992	-0.48813
7	7	7	1	1	3.1	high	50	450	500	13	16.21	14.45	15.38	14.50	15.94	15.296	0.807236	-0.09300
8	8	8	1	1	3.9	high	50	450	525	4	17.32	16.65	16.58	16.42	17.75	16.944	0.566683	-0.24666
9	9	9	1	1	3.1	low	30	475	525	7	16.13	14.73	16.24	14.68	16.17	15.590	0.809043	-0.09203
10	10	10	1	1	3.9	low	30	475	500	16	17.16	15.80	16.37	16.02	15.79	16.228	0.571638	-0.24288
11	11	11	1	1	3.1	high	30	475	500	15	15.38	15.24	15.39	15.27	15.19	15.294	0.087920	-1.05591
12	12	12	1	1	3.9	high	30	475	525	1	16.64	16.38	16.44	16.51	16.44	16.482	0.099599	-1.00174
13	13	13	1	1	3.1	low	50	475	500	3	15.06	15.40	15.37	14.97	14.43	15.046	0.392339	-0.40634
14	14	14	1	1	3.9	low	50	475	525	10	16.76	16.94	17.02	17.27	16.25	16.848	0.381274	-0.41876
15	15	15	1	1	3.1	high	50	475	525	11	15.03	15.13	15.02	14.90	14.77	14.970	0.138384	-0.85892
16	16	16	1	1	3.9	high	50	475	500	8	16.55	16.71	16.33	16.48	16.32	16.478	0.162696	-0.78862
17																		

Figure F.7   Fractional factorial design and output for paint film build experiment.

F2 ANALYSIS OF THE EXPERIMENTAL RESULTS

To start the analysis, we plot the output by treatment number (see Appendix C). For the paint film build robustness experiment discussed in Section F.1, we get Figure F.8. We look for outliers and differences in the average or variation across the treatments. In some cases, the output by treatment plot is sufficient to draw appropriate conclusions. Note that if the multiple observations for each treatment come from repeats in the experiment (rather than replicates), the variation within each treatment likely underestimates the long-term variation for that treatment.

To analyze the experimental results more fully we use:

Stat → DOE → Factorial → Analyze Factorial Design

The resulting dialog box is given in Figure F.9. We specify the output (response) to analyze and choose the form of the model using the Terms button.

In this example, as shown in Figure F.9, we choose the performance measure "log(s)" as the output (upper left panel of Figure F.9) and select a model with all possible main effects and interactions (upper right panel of Figure F.9). Here we include interactions up to fifth order, since there are five inputs in the experiment.

From the Graphs dialog box (lower left panel in Figure F.9) we select a Pareto effects plot. In some cases, we may also want to have a numerical summary of some of the important effects. We can use the Results dialog box (lower right panel of Figure F.9) to request a display of the estimated average output for different levels of the input or inputs.
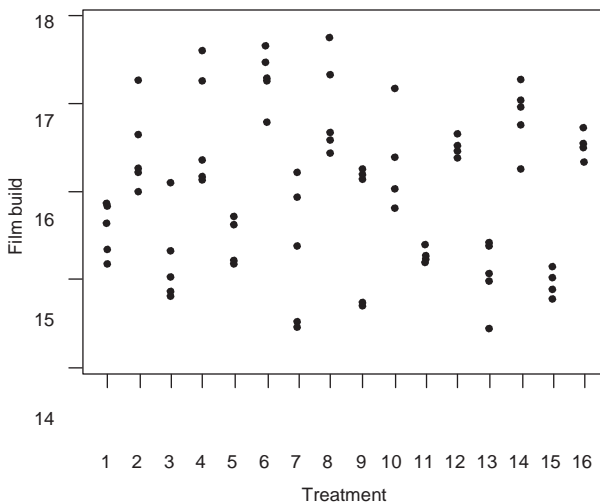
Figure F.8   Film build by treatment.
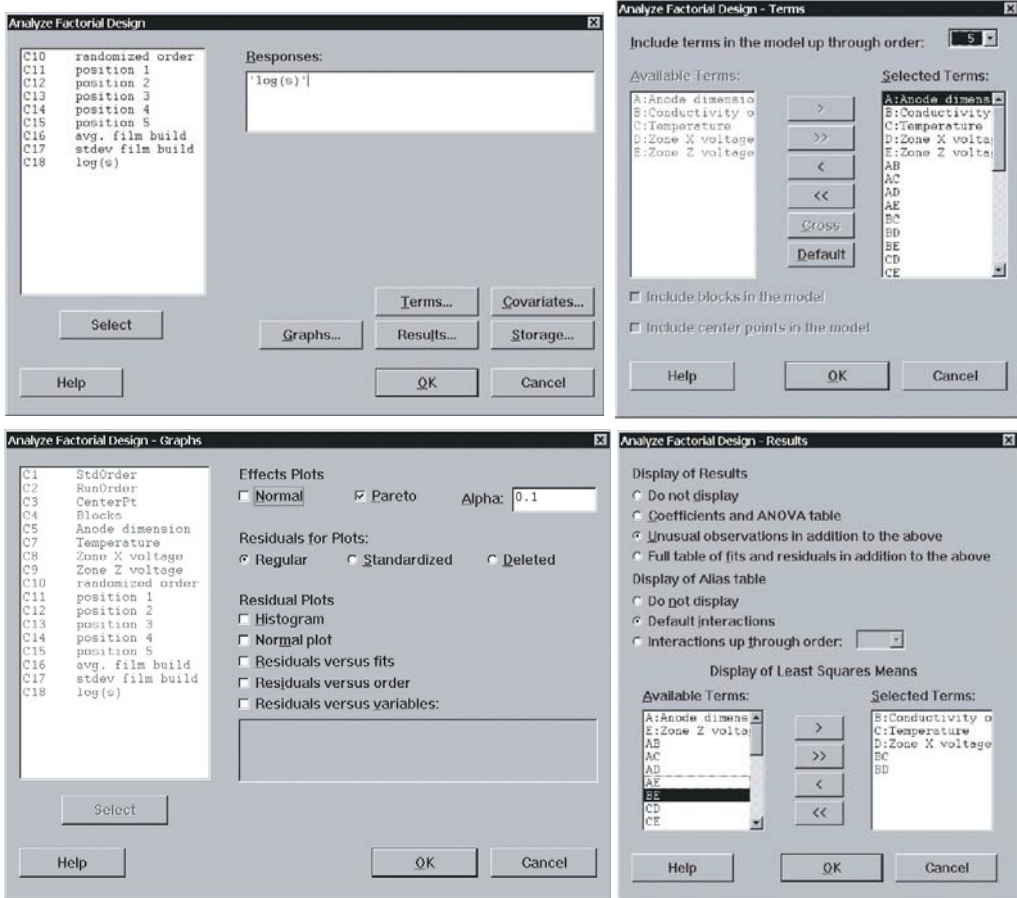
Figure F.9   MINITAB Analyze Factorial Design dialog box.

For the paint film build example, the resulting Pareto effects plot, which ranks the unsigned effects, is given in Figure F.10. In this plot any large effects (relative to the other effects) should be clearly evident. We see that the effects due to zone X voltage, conductivity, and the interaction between these two candidates are large.
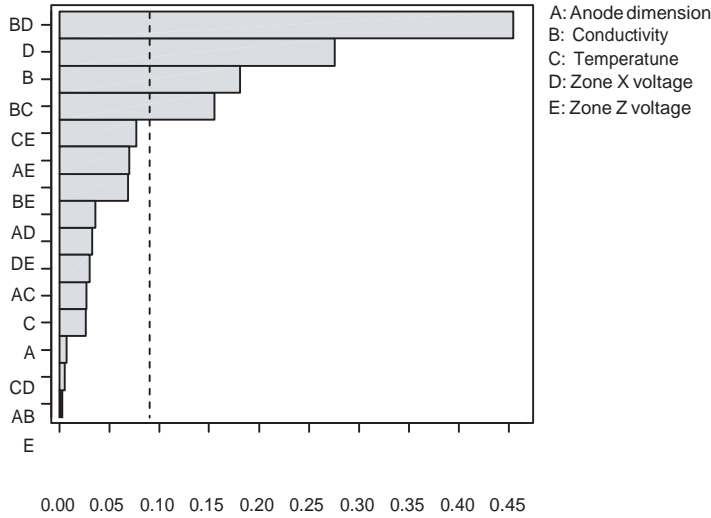
A: Anode dimension
B: Conductivity
C: Temperatune
D: Zone X voltage
E: Zone Z voltage

Figure F.10   Pareto plot of the effects for paint film build robustness experiment.

To make further sense of these experimental results, we consider factorial plots obtained using the MINITAB menu selection:

Stat → DOE → Factorial → Factorial Plots

We can request main effects, two-way interaction, and three-way interaction plots. Figure F.11 shows how we request main effects plots based on the performance measure log(*s*).

In the paint film build example, choosing main effects and interaction plots with all five candidates, we get the plots given in figures F.12 and F.13.
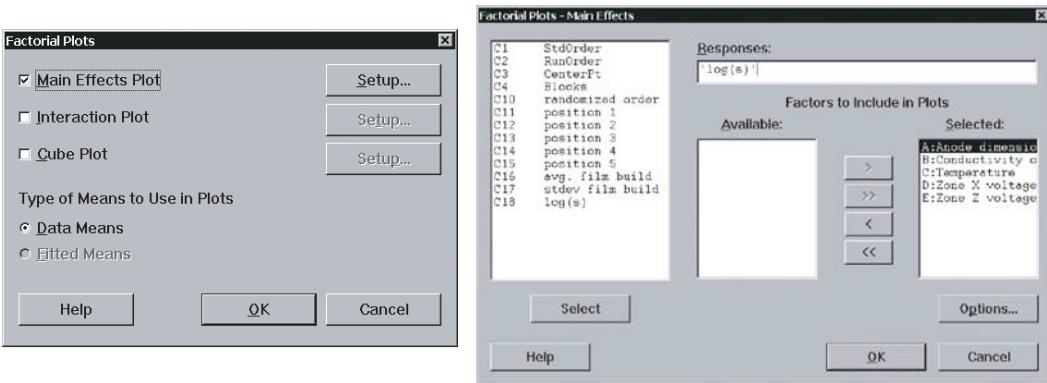

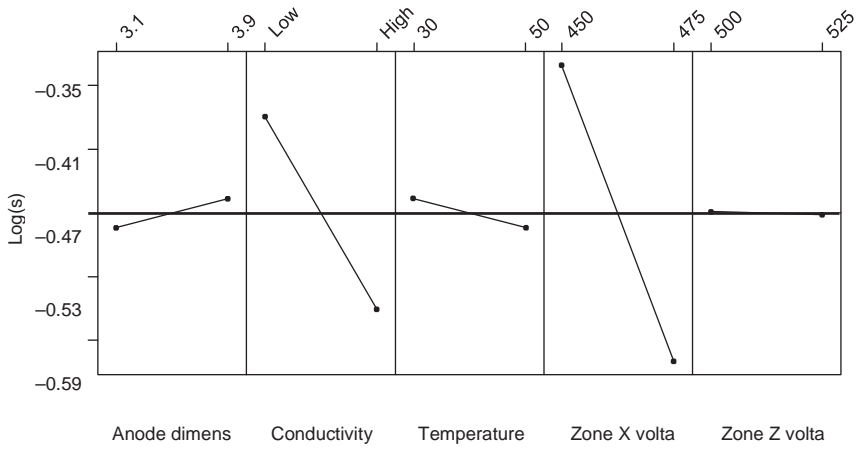**Figure F.11**   Factorial Plots dialog box.

Figure F.12   Main effects plot (in terms of log(s)) for film build robustness experiment.
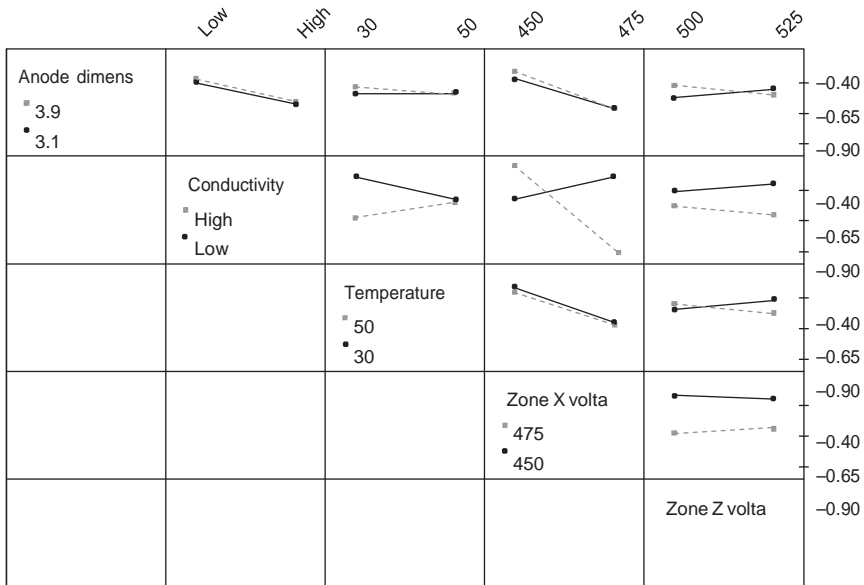

Figure F.13   Interactions plot (in terms of log(s)) for film build robustness experiment.

Note that in the paint film build robustness experiment, the design was resolution V; thus, the main effects and interactions can be estimated separately. If the design is resolution IV, some two-way interactions are aliased with other two-way interactions. In particular, for the standard half fraction resolution IV designs, half the individual interaction plots show the same information. Similarly, in resolution III design, the main effects are confounded with interactions. With resolution III designs, we need only consider the main effects plots, since the interaction plots will not provide any new information.
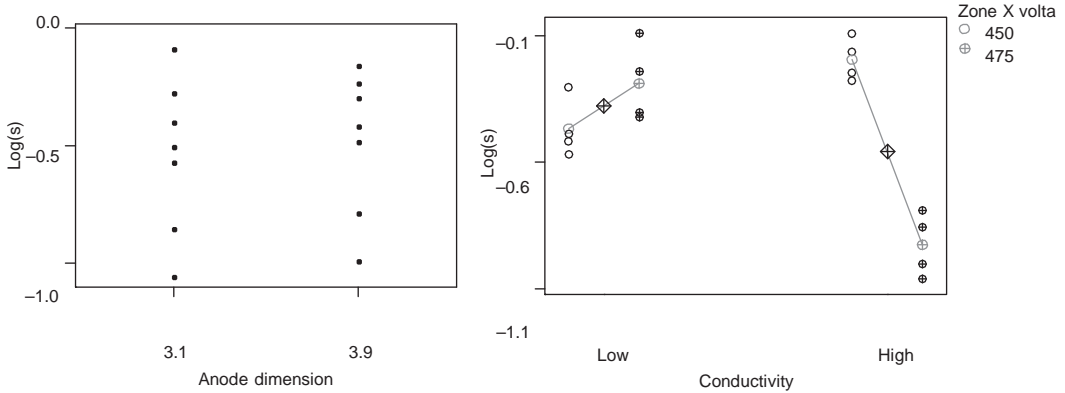
Figure F.14   Main effects and two-way interaction plots showing individual observations.

From figures F.12 and F.13, we conclude there are large effects due to conductivity, zone *X* voltage, and the interactions between conductivity and zone *X* voltage and between conductivity and temperature. Recall that smaller log(*s*) is better. Since there are large interactions, we draw conclusions based on the interaction plot. Figure F.13 suggests high zone *X* voltage, high conductivity, and low temperature are best. We are fortunate that high conductivity is best in both large interactions.

Note that the main effects and interaction effects plots produced by MINITAB show only averages. We can use box plots (showing individual output values) and multivari plots (see Appendix C) to create alternative displays of main effects and two-way interactions that show individual observations, as illustrated in Figure F.14.