A BRIEF INTRODUCTION TO MINITAB

Minitab Inc. (http://www.minitab.com/) is a leading global provider of software and services for quality improvement and statistics education. Their mission is to provide the tools and resources professionals need to analyze complex problems, improve their processes, and train their students.

Minitab is best known for its flagship product, the Minitab® Statistical Software. The package was originally created in 1972 to help professors teach statistics, but has since evolved into the premier software organizations use when analyzing business data to improve the quality of their goods and services. It has driven virtually every major Six Sigma improvement initiative around the world, and is the package students use to learn statistics in more than 4,000 colleges and universities.

Basic principles and simple tools for data analysis

The Minitab software is very easy to use. Minitab is a spreadsheet program. Data get entered into columns and rows of a spreadsheet, and calculation and graphing operations are executed through simple pull-down menus. As of July 2008, the latest version of Minitab is Version 15. Minitab divides its worksheet (spreadsheet) into columns (labeled C1, C2, C3, ...) and rows (labeled 1, 2, 3, …). Take, for example, the data in Table 1.5-1 which lists the weights (in 1000 pounds), the fuel efficiencies (in gallons per 100 travelled miles), and the names of n = 10 cars. Weights are entered into column 1, fuel efficiencies are entered into column 2, and labels are entered into column 3. Each row of the spreadsheet represents a different car. The data can be entered through the keyboard, or it can be entered by clicking (and opening up) the Minitab file that has been prepared for this particular data set (see section 2 of this manual). Click on the file Section1.5Table1.5-1Cars. The Minitab program will open and one of its windows will show a worksheet with three columns of data. Informative labels are attached to the columns; here column C1 is labeled as X=Weight, column C2 as Y=GPM, and column C3 as Car.

The Worksheet containing the data is one of the windows that you see when calling up Minitab. Another window that you see is the Session window. The Session window collects the output that is generated during a Minitab analysis. Some operations will generate Graph windows. All windows can be saved to files.

The command line on the top of the Minitab window (with its tabs: File, Data, Edit, Calc, Stat, Graph, Editor, Tools, Windows, and Help) contains pull-down menus for carrying out operations. For example, click on the tab File. The commands within this folder allow you to save the worksheet, enter previous worksheets (such as the worksheet files we have prepared for the data sets in this text), print the worksheet, and save the current project (which consists of the worksheet and the output that has been generated by the current Minitab session).

The commands under the tab Data allow you to sort, rank, and copy the information in specified columns to other columns of the worksheet. The commands under the tab Calc allow you to create new variables (using calculator), generate random variables, and carry out probability calculations. The commands under the tab Stat carry out the various statistical analyses, and commands under the tab Graph provide many of the displays that we discuss in our book. The instruction Enable under the Editor tab enables a record of all instructions that are carried out during a Minitab session. The Help tab is important for getting information on how Minitab works; it gives detailed descriptions of the procedures and explains how to carry them out. We suggest that you start with the simplest versions of the commands, before learning how to tweak each procedure to get the maximum benefit. It should take you almost no time to get familiar with the basic features of the software, and you will become very proficient in a matter of days.

Click on the prepared file Section1.5Table1.5-1Cars. You will see the worksheet with the data, and a session window. Go to the command line on the top of the Minitab window and click on Editor and then on Enable commands. You will see the greater symbol (">") on the command line of the session window. Minitab uses ">" as its prompt. Enter the line "print C1 C2 C3" (without the quotes; you can also use either lower or upper case letters) and hit return. This will print out the three columns. There are two ways of executing tasks. You can execute commands from the session window by entering certain text instructions, or from the command line by clicking on tasks that are included under the various tabs. Enabling commands (under the Editor tab) translates the instructions from the pull down menus into text instructions in the session window.

Click on the Stat tab, and then click on Basic Statistics and on Display Descriptive Statistics. We indicate this path by writing "Stat > Basic Statistics > Display Descriptive Statistics". A dialog box will open. Enter C1 and C2 for the variables (you can do this by clicking on the columns in the area on the left). Running this command (by clicking "OK") provides the summary statistics on these two columns. You can change the desired output by clicking on the tab Statistics; for example, you can calculate the trimmed mean if you wish. Also, you can stratify the analysis by adding a categorical variable into the "by variable" box. (In this data set, no such variable is available). Note that the text variable (label) does not show up as a variable in the dialog box; this makes sense as you would not want to calculate numerical statistics for text data.

Go to the Graph tab next. You can get dot plots of the data in the two columns (use "Graph > Dotplot"). A scatter plot of the fuel efficiency against the weight of the car can be obtained from "Graph > Scatterplot." Select the simplest version to get started. All you need is to enter the variables into the dialog box. If you wish, you can add labels and titles. With time and practice you will find that there are many other useful options. For example you may want to add to this graph the least squares line. For that you need to go to the window "With Regressions". You may want to overlay two scatter plots on the same graph. For that you have to go to "With Groups." Selecting the "Multiple Graphs" window in the following dialog box will give you many graphing options.

The correlation coefficient is calculated from the "Stat > Basic Statistics > Correlation" dialog window. You can achieve the same from the session window by typing in "corr c1 c2" and hitting the return key.

Time sequence plots can be obtained through "Graph > Time Series Plot". For example click on the file Section1.2Exercise1.2-1Thermostat. It contains the sales from 52 consecutive weeks. You can add informative labels. You can also change the labels in the graph by double clicking on the labels and changing them. You can change the scales of the axes by pointing your mouse to the desired axis (either x or y), clicking the right button of your mouse, and going to the "Edit X Scale" tab for further instructions.

Consider the file Section1.3Exercise1.3-8Thickness. It contains the thickness measurements of n = 150 ears of paint cans. Calculate summary statistics (by "Stat > Basic Statistics > Display Descriptive Statistics"), construct a dotplot ("Graph > Dotplot"), a histogram ("Graph > Histogram"), a steam-and-leaf display ("Graph > "Stem-and-Leaf"), and a box plot ("Graph > Boxplot"). Look at the available options for constructing histograms. Once you have created the histogram, you can point your mouse to the x-axis, right click, and edit scale. The options in the "binning" tab will allow you to change the number of bins as well as the cut- and mid-points of the histogram.

Consider the data on the lead concentrations in the file Section1.4Lead1976&1977. The first column contains the lead concentration for 1976, while the second column contains the data for 1977. The numbers of observations happen to be the same in the two groups. Stratification is important in this analysis as we want to compare the two distributions Hence dot diagrams, box plots and histograms should be graphed for each of the two years on the same sheet and on the same scale. You can construct comparative plots by executing "Graph > Dotplot," entering C1 and C2 into the dialog box, and clicking "Multiple Y's".

Sometimes observations are missing. Consider the file Chapter8Project2Wine. The price for the 1954 and 1956 vintage is missing. Minitab uses the symbol "*". Commands will skip over the rows that contain the missing value. For example the summary statistics for column C5 (price) calculates the statistics from the 27 available rows in that column. Summary statistics on the other columns (such as rain in C3) use all available rows for that column (that is, not just those that have information on all columns). Command such as scatter plot of price (in C5) on rain (in C3) use the 27 available pairs.

Determining probabilities and percentiles of various distributions

The "Calc > Probability Distributions" tab can calculate cumulative probabilities and percentiles for all distributions discussed in this text. For example, selecting Normal will open up a dialog box. For cumulative probabilities, we click "Cumulative probability," enter the mean and standard deviation, and a constant (which specifies the argument of the c.d.f.). For example with mean 3, standard deviation 2, and constant 1, we obtain the cumulative probability $P(X \leq 1) = 0.158655$. For percentiles, we need to click "Inverse

cumulative probability," enter the mean and standard deviation, and a constant (which is now the specified proportion). For example with mean 3, standard deviation 2, and constant 0.80, we obtain the 80th percentile as 4.68324. Clicking on the "probability density" will give us the value of the density at a specified constant. For example, $f(x = 1.0) = 0.120985$ for the normal distribution with mean 2 and standard deviation 3.

The same type of operations ("cumulative probability" for probabilities, "inverse cumulative probability" for percentiles, and "probability density" for the value of the density function) apply to all other continuous distributions. The only changes are in the parameters of the distribution. For example, for the Gamma distribution with parameters $\alpha = 2$ (shape) and $\beta = 5$ (scale), we obtain $P(X \leq 10) = 0.593994$; note the mean of this distribution is (2)(5) = 10. The 90th percentile of this distribution is 19.4486.

The same instructions are carried out for discrete distributions such as the binomial distribution with parameters n = 20 and p = 0.1. "Cumulative probability" provides the cumulative probabilities up to the selected constant c, $P(X \leq c)$. For example, $P(X \leq 1.4) = P(X \leq 1) = 0.391747$. "Probability density" provides the probabilities, $P(X = c)$. For example, $P(X = 1) = 0.270170$ and $P(X = 1.4) = 0$, since this particular binomial is a discrete distribution with support on the integers from 0 to 20. "Inverse cumulative probability" provides the percentiles; the 80th percentile equals 3 [Minitab lists the cumulative probabilities $P(X \leq 3) = 0.86704$ and $P(X \leq 2) = 0.676927$]

It's easy to draw the p.d.f. of the binomial distribution with n = 20 and p = 0.1. First enter the integers 0, 1, 2, …, 20 into the 21 rows of column C1 [you can do this manually, or you can use the command "Calc > Make Patterned Data > Simple Set of Numbers".] Then use the "Calc > Probability Distributions" tab and go to the binomial distribution. Click "probability" and add C1 into the input column field. Enter C2 (optional storage) to store the probabilities $P(X = x), x = 0,1,...,20$. Then go to "Graph > Barchart," click "values from a table," and enter columns C2 and C1.

Generating random variables

"Calc > Random Data > Normal" can be used to generate a fixed number of realizations from a normal distribution with specified mean and standard deviation. The data can be stored in any column(s). Use this command to generate 1,000 realizations from a normal distribution with mean 10 and standard deviation 3. Calculate the summary statistics, and plot the histogram to convince yourself that this function is doing the right thing.

You can use the "Calc > Random Data" command to generate realizations from many other distributions. For illustration, generate realizations (say n = 100) from such distributions as the geometric with probability 0.2 (a discrete distribution), the continuous uniform between 0 and 1, the exponential with mean 6, the gamma with $\alpha = 2$ (shape) and $\beta = 5$ (scale), and the chi-square with 10 degrees of freedom. If you want to learn more about these distributions, click the help feature in the dialog box.

There is no carry-over among consecutive random numbers. You can check this by computing the autocorrelations of the generated data sequence at lags 1, 2, … Use "Stat > Time Series > Autocorrelations" to obtain the autocorrelations. They should be within the 2 sigma limits that are indicated on the graph. You can also lag the series C1 by one period (using "Stat > Time Series > Lag"), store the lagged series in another column, say C2, construct a scatter plot of C1 against C2, and calculate the correlation coefficient (corr C1 C2). You should see no patterns in the scatter plot, and the correlation (which is the lag 1 autocorrelation) should be small.

Constructing probability plots

Probability plots are easy to obtain. Take the generated normal random variables (in column C1), for example, and check whether or not these observations are from a normal distribution. Use "Graph > Probability Plot > Single," enter the column that contains the data, and check Normal under the Distribution tab. You can check many other distributions such as the Weibull, gamma, etc. The added line in the Minitab probability plot helps you judge whether the data can be modeled with the selected distribution. Violations from the linearity indicate that the selected distribution is not a good fit.

Confidence intervals and testing hypotheses

The "Stat > Basic Statistics" folder includes procedures for the calculation of confidence intervals and the testing of hypotheses. "1-Sample Z" is used for the inference about a population mean from a single random sample assuming that the standard deviation is known; it uses the normal distribution. "1-Sample t" is used for the inference about a population mean from a single random sample with estimated standard deviation; it uses the t-distribution. "2-Sample t" is used for the comparison of two means when the samples are independent. "Paired t" is used in the paired or blocked situation. "1 Proportion" is used for the inference about a single proportion, while "2 Proportions" is used for the comparison of two proportions. "1 Variance" is used for the inference about a single variance, while "2 Variances" covers the comparison of two variances.

For example, take the 1976 lead concentrations (with n = 64) in file Section1.4Lead1976&1977 and obtain a 95 percent confidence interval for the mean lead concentration in 1976. Furthermore, test the research hypothesis that the mean lead concentration is different than 7 ppm. The following result is obtained with the command "Stat > Basic Statistics > 1-Sample t". You need to enter the column that contains the data, the hypothesized mean, and you need to specify a 2-sided test alternative (under the Options tab).

```
Test of mu = 7 vs not = 7
Variable   N   Mean  StDev  SE Mean      95% CI        T     P
Lead1976  64  7.291  2.025    0.253  (6.785, 7.797)  1.15  0.255
```

Next, assume that you want to test whether the mean concentration for 1976 is smaller than the mean concentration for 1977. The result is obtained with the command "Stat > Basic Statistics > 2-Sample t". You need to specify the columns that contain the 1976 and 1977 data (two different columns) and, under Options, you need to specify the confidence coefficient (say 95 percent) and the fact that the alternative hypothesis specifies that the mean in the first column (1976) is smaller than the mean in the second column (1977). The Welch approximation (see book Section 4.6-1) is used if you can not assume that the variances are equal. The pooled standard deviation is used if the variances are assumed equal. You can also create graphs that support your test conclusions.

```
Two-sample T for Lead1976 vs Lead1977

           N   Mean   StDev   SE Mean
Lead1976   64   7.29   2.03     0.25
Lead1977   64   9.42   2.08     0.26

Difference = mu (Lead1976) - mu (Lead1977)
Estimate for difference:  -2.131
95% upper bound for difference:  -1.530
T-Test of difference=0 (vs <):T-Value=-5.87 P-Value=0.000 DF=125
```

Finally, let us illustrate how you can use "Stat > Basic Statistics > 2 Variances" to test whether the variances of the observations in 1976 and 1977 are the same. The F-test explained in Section 4.6-4 is calculated by Minitab. (In addition, Minitab calculates the Levine test which is another useful procedure). The large p value (0.832 being much larger than 0.05) indicates that we cannot reject the null hypothesis that the two variances are equal.
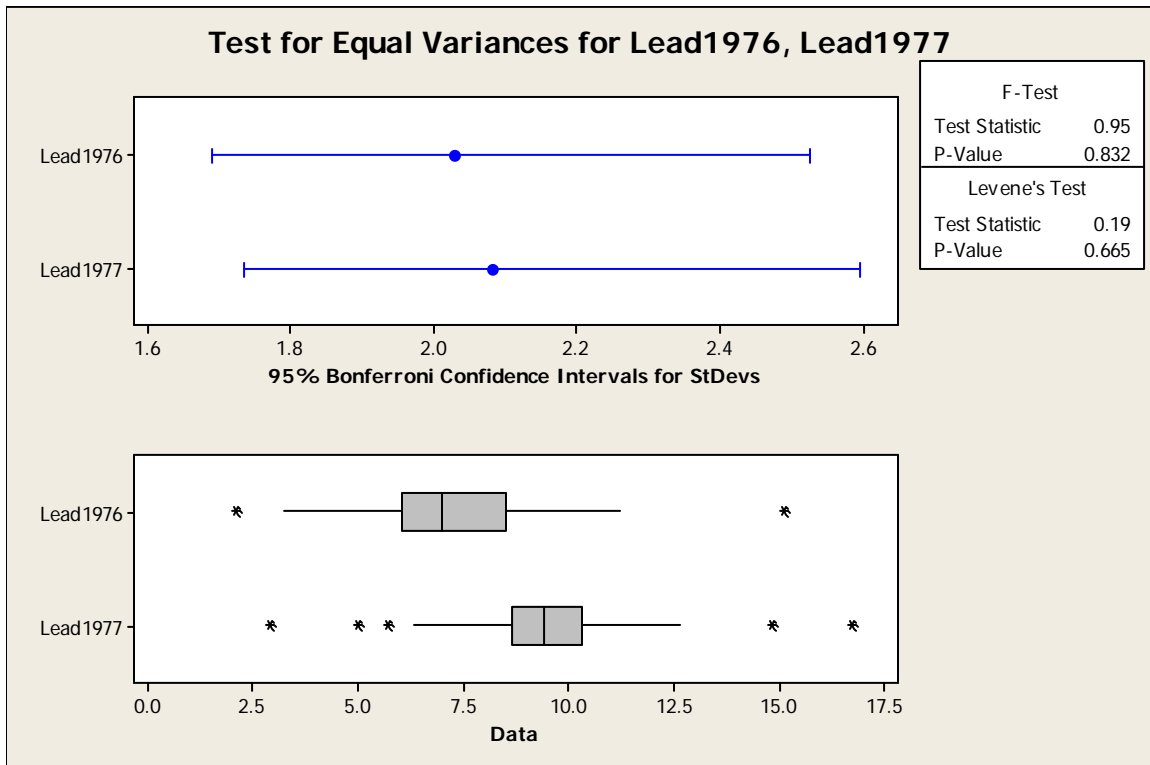
Power calculations

In the simplest case, use the command "Stat > Power and Sample Size > 1-Sample Z."
For example, assume that the standard deviation of a measurement is 2. You want to test
for an increase in the mean (one-sided alternative) and want to detect a one-unit increase
with power 0.8; that is, the probability of not detecting such a change is 0.20 if the mean
is indeed increased by one unit (type II error). The type I error probability is selected as
0.05. Then the required sample size is 25. The output is shown below.

```
1-Sample Z Test

Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05  Assumed standard deviation = 2

            Sample  Target
Difference    Size   Power  Actual Power
         1      25     0.8      0.803765
```

Chi-square test for independence

Use the command "Stat > Table > Chi-Square Test (Two way Table in Worksheet)." For
example, with the data from Table 4.7-2 stored in columns C1, C2, and C3 of the
worksheet, the output is as follows:

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

          C1      C2      C3  Total
    1     15      18       7     40
        8.89   17.78   13.33
        4.201  0.003   3.008

    2      5      22      23     50
       11.11   22.22   16.67
        3.361  0.002   2.407

Total     20      40      30     90

Chi-Sq = 12.982, DF = 2, P-Value = 0.002
```

Use the command "Stat > Table > Cross Tabulation and Chi-Square" if you work with
raw (not summarized) data. Minitab will create the table of counts and perform the chi-
square test.

Control charts

"Stat > Control Charts > Variables Charts for Subgroups" can be used for the x-bar and
R-charts, while "Stat > Control Charts > Attributes Charts" can be used for the p- and c-
charts discussed in Section 5.1.

Consider the information in the file <u>Section5.1Table5.1-1Chart</u>. Go to "Stat > Control Charts > Variables Charts for Subgroups" and then to "Xbar-R". Indicate that the observations for a subgroup are in one row of columns. Go to "Xbar-R chart Options;" under the "Estimate" tab, specify that you want to omit the subgroups when estimating parameters; here 11:12) and click "Rbar" so that $\overline{R}$ is used for the calculation of the control limits. This will result in Figures 5.1-1 and 5.1-2 of the text.

For the p-chart in Example 5.1-1 of the text, you enter the numbers of defectives "1 1 3 …. 0 2" into a column of the worksheet, say C1. Then go to "Stat > Control Charts > Attributes Charts," specify "P"(chart), and enter the column with the number of defectives (here C1) and the subgroup size (here n = 50). This will give you the chart in Figure 5.1-3.

Minitab includes numerous other useful control charts that have not been discussed in this text.

Analyzing data from designed experiments: ANOVA methods

"Stat > ANOVA" with its commands "One Way", "One Way (unstacked)", "Two Way", "Balanced ANOVA", "General Linear Model", and "Fully Nested ANOVA" are relevant.

Consider the data in file <u>Section6.1Table6.1-2Deflection</u>. Use the "Stat > ANOVA > One Way" (the stacked) command. The column of deflections (C1) is entered as the response, while the column of beam indicators (1, 2, 3 in C2) is entered as the factor. The Comparison tab allows us to obtain "Tukey" multiple comparisons, and under the Graph tab we can ask for useful graphs (such as an individual value plot or box plots of the data). Try it, and check that this will give you the output that we discussed in Sections 6.1 and 6.2.

Consider the data in file <u>Section6.3Table6.3-1Strength</u>. Use the "Stat > ANOVA > Two Way" command. Strength in column C1 is entered as the response, while the column of treatments (1, 2, 3 in C2) and batch (1, 2, …,5 in C3) are entered as the factors. The Graph tab allows for useful graphs (such as an individual value plot or box plots of the data). Try it, and check that this will give you the output that we discussed in Section 6.3.

Consider the data in file <u>Section7.1Table7.1-5Popcorn</u>. For our analysis we use the "Stat > ANOVA > Two Way" command. Yield in column C1 is entered as the response, while the popper (1 and 2 in C2) and brand (1, 2, 3 in C3) are entered as the factors. The output coincides with the one shown in Section 7.1 of the book. The commands "Stat > ANOVA > Main Effects Plot" and "Stat > ANOVA > Interactions Plot" display main effects and interactions graphically. The output of the interaction plot is shown in Figure 7.1-2 of the text.

The function "Stat > ANOVA > Balanced ANOVA" allows the analysis of balanced multi-factor experiments like the ones in Section 7.1 (2-factor factorial designs) and Section 7.3 (general factorial designs with more than two factors). We can declare factors as either fixed or random. If this approach is used, one must specify the model in terms of main effects (say C2 and C3 for popper and brand) and the interaction (C2*C3).

The commands "Stat > ANOVA > Two Way" and "Stat > ANOVA > Balanced ANOVA" do not work for unbalanced situations. Convince yourself of this fact by deleting the last row in the file Section7.1Table7.1-5Popcorn. For unbalanced designs we must use "Stat > ANOVA > General Linear Models". The interpretation of the sums of squares and their associated tests is more complicated than the one in the balanced situation. Now sums of squares need to be interpreted as partial (that is, partial to all other factors in the model). We explain this concept in Section 8.5 (chapter on regression).

Consider the data in file Section7.2Table7.2-1Rods. For our analysis we use the "Stat > ANOVA > Fully Nested ANOVA" command. Column C3 containing the measurements is entered as the response, while rods (1, 2, …, 5 in C1) and forging (1, 2, 3 in C2) are entered as the factors. The output of the analysis (ANOVA table and estimated variance components) is shown in Table 7.2-3 of the text.

Factorial and fractional factorial designs

Minitab constructs the designs (that is, tells us about the factor levels of the design runs and the order in which the runs should be carried out), and helps us analyze the resulting data once the experiment has been performed. Minitab's features are included under the "Stat > DOE > Factorial" tab. Minitab makes it easy for the user to construct 2-level factorial (Section 7.3) and fractional factorial designs (Section 7.4). The user enters the number of factors and is then presented a list of full and fractional designs with their respective run sizes. After deciding on the number of runs, the user can construct the design either through default generators that optimize the resolution of the design, or by writing out specific generators. In either situation, Minitab indicates the design columns, displays – if desired – a randomized arrangement of the runs, and lists the confounding patterns of the particular fraction that is being selected. If default generators are used, Minitab displays the adopted generators. The user can add center points, replicate the design, block the experiment by specifying blocking generators, and modify the design by considering fold-overs. These can be complete (full) fold-overs where the signs of all factors are changed, or fold-overs of individual factors.

Once the design has been carried out, Minitab facilitates an efficient analysis of the data. The analysis of 2-level factorial and fractional factorial designs includes estimates of the effects. Two different estimates are listed: estimates labeled "Coef" which are the estimates discussed in this book, and estimates called "Effects" (which are twice the size of the coefficients expressing the differences in the response averages at high and low levels). Standard errors of the estimated coefficients are calculated if replications are available. Normal probability plots and Pareto plots for assessing the importance of the

effects, and main effects and interaction plots for assessing the nature of the relationships are also available.

Consider the data file <u>Section7.3Table7.3-5Fabric</u>. This is a 2^4 full factorial in 16 runs. The runs are listed in standard order. We have already run the experiment and have collected the observations. Hence we first need to define the experiment by carrying out the command "Stat > DOE > Factorial > Define Custom Factorial Design". We communicate to the program the columns that contain the factors. As far as the subsequent calculations are concerned, we can ignore the information that gets added into columns C6 through C9. After this first "define" step we execute the "Stat > DOE > Factorial > Analyze Factorial Design" command, and indicate to Minitab the column that contains the response (here it is in C5). The output lists the effects and estimated coefficients (see the above discussion). Note that there is no error since the model is fully saturated, with 16 coefficients (average, main effects and interactions) estimated from 16 observations. Under the "Graph" tab one can request a normal probability plot of the estimated coefficients. This resulting graph is shown in Figure 7.4-2 of the text. One could specify a model with just main effects and 2-factor interactions. This is achieved by going to the tab "Terms" and selecting main effects and 2-factor interactions for the model. An estimate of the standard deviation of the error is obtained from pooling the higher-order interactions; in this case the ANOVA table lists an error mean square and tests for main effects and 2-factor interactions.

Next, let Minitab select a "good" fractional factorial design for, say, 8 factors in 16 runs. This amounts to a 2^{8-4} fractional factorial design. We run the "Stat > DOE > Factorial > Create Factorial Design" command. We enter the number of factors (there are 8) and learn, under available factorial designs, that a resolution IV design is possible. We click on "Design", and select from the offered choices the 1/16 resolution IV fraction. Under "Options", we click the tab for getting the runs in standard (that is, not randomized) order. The output lists the factor levels of the 16 runs. It also reveals the generating relationship (design generators: E = BCD, F = ACD, G = ABC, H = ABD) and the alias structure that is implied by this particular fraction. Once we have obtained the responses at these experimental conditions, we add them to the worksheet. [For illustration, generate 16 realizations of the standard normal distribution]. Then execute "Stat > DOE > Factorial > Analyze Factorial Design", and indicate to the program the column that contains the responses. Minitab then calculates the effects and estimated coefficients, and lists the alias relationships for the estimates.

Consider the data file <u>Section7.4Table7.4-3FractFact2</u>. It shows a 2^{7-4} fractional factorial design in 8 runs. Since we already have run the experiment and have collected the observations we need to define the experiment by carrying out the "Stat > DOE > Factorial > Define Custom Factorial Design" command. We basically tell Minitab where the factors are. Minitab recognizes the arrangement as the fully saturated factorial design. After this step, we execute the "Stat > DOE > Factorial > Analyze Factorial Design" command and indicate to Minitab the response column. The resulting output lists the effects and estimated coefficients (see above) and the alias patterns, and provides, if asked for, the normal probability graph of the estimated effects. The coefficients and the

alias patterns agree with the ones we have listed in Table 8.4-3 (under Results, specify interactions up to order 2 only).

Regression

The regression analysis (in Chapter 8 of this book) is executed through various programs listed under the "Stat > Regression" tab (folder). They are: Regression, Stepwise, Best Subsets, and Fitted Line Plot.

The "Stat > Regression > Regression" dialog box requires the specification of the response variable and the explanatory (regressor) variables. The output provides the coefficient estimates, their standard errors, t-ratios and probability values. It also displays the ANOVA table and the coefficient of determination. The output (residuals, fitted values, as well as measures of leverage and influence – which were not discussed in this book) can be stored in unused columns of the worksheet. Residual plots can be executed through the "Graph" dialog box, or by first storing the needed variables and then using standard plotting functions.

The "Stat > Regression > Fitted Line Plot" is used when there is only one regressor variable. The scatter plot of the data is shown, together with the superimposed fitted model (which can be either linear, quadratic, or cubic).

The "Stat > Regression > Best Subsets" feature fits all possible regressions and displays useful summary statistics that help select a parsimonious model.

The "Stat > Regression > Stepwise" feature includes the backward elimination method explained in the text as a special case.

A useful command for creating indicator variables: "Calc > Make Indicator Variables" creates m indicator variables from a single variable with m different values.