Doing It with MINITAB: A Supplement to Applied Statistics for Engineers and Scientists [1] [2]

J. D. Petruccelli B. Nandram M. Chen

[1]Development of these materials was partially supported by the National Science Foundation's Division of Undergraduate Education under Grant DUE 9254087.

Contents

Introduction

This document contains detailed instructions of how to use MINITAB to do the kinds of analyses documented in the text **Applied Statistics for Engineers and Scientists**. Each of the first fifteen chapters is keyed to one chapter of the text, and references material in that chapter. For example, Chapter 1 begins by describing where to find the electric usage data from Chapter 1 of the book and how to use those data and MINITAB to generate Figures 1.1 and 1.3 in the text. Throughout, the instructions assume you are using release 12 of MINITAB.

We have written a collection of MINITAB macros to facilitate the use of MINITAB in statistical applications and in labs. The uses of these macros are described where appropriate in this supplement. Lists of these macros are found in Appendix B. The macros are included on the accompanying computer disk.

To run a macro in MINITAB, select *Edit:Command Line Editor*. A window will appear, and the user types the appropriate commands to run the macro. These commands are submitted by clicking on the *Submit Commands* button. Alternatively, these commands can be submitted directly from the MINITAB command line. As an example, consider the macro LAB4_1, which is used in lab 4-1 of the text. This macro requires two inputs, N: the number of the student's birth month modulo 9, and NSIMS: the number of simulations to be run. A student born in October has birth month number 10, which equals 1 modulo 9, so N=1. The lab asks the student to use 10,000 simulations, so the command to be submitted to run this macro is:

%*pathname* lab4_1 1 10000

where *pathname* is the pathname to the folder or directory containing the macro.[1] In the rest of this document, we will omit the pathname when giving instructions for running macros.

The data sets used in the text are also found on the accompanying CD. These data sets are provided in two formats:

- Portable MINITAB worksheet file format. The MINITAB export files are identified by the suffix .MTP.
- ASCII file format. ASCII files are text files. The data files in this format have a header containing the variable names, and are space-delimited. The ASCII files are identified by the suffix .dat. Missing values in these files are denoted by periods.

[1] A pathname is unnecessary if the macro is stored in one of the default folders or directories MINITAB searches for macros.

Chapter 1

Doing It with MINITAB: Chapter 1

1.1 Data Sets

Name	Description
ELECE	Electric usage data from Professor P.'s household.
WASHER5	Washer thickness data, Example 1.2
WASHER7	Washer thickness data, Example 1.3
EG1_5	Bearing play data, Example 1.5
EG1_5A	Bearing play data, Example 1.5

1.2 Stationarity of Processes

Figures 1.1 and 1.3 of the Text

Figures similar to Figures 1.1 and 1.3 can be produced with MINITAB. To get a plot similar to Figure 1.1, choose *Graph:Histogram* and then selecting KWH from the resulting dialog window.

A plot similar to Figure 1.3 can be produced by selecting *Graph:Marginal Plot...* and then choosing KWH as the Y variable and INDEX as the X variable in the dialog window.

Figures 1.4, 1.5 and 1.6 of the Text

A plot similar to Figure 1.4 can be produced by choosing *Graph:Time Series Plot*, and plots similar to Figures 1.5 and 1.6 can be produced by choosing *Stat:Time Series:Moving Average...* .

1.3 Identifying Possible Causes of Variation

Creating a New Ishikawa Diagram

While you can draw effective Ishikawa diagrams by hand, presentation-quality diagrams are easily drawn using MINITAB as follows:

1. You can specify up to six main branches. For each main branch, have one column in alpha format in your worksheet. The column entries will be the causes to be put on the corresponding main branch of the diagram.

2. Choose *Stat:Quality Tools:Cause-and-Effect...* . In the resulting dialog box, give the main branch names and which worksheet column corresponds to it.

As with any other MINITAB graph, this Ishikawa diagram can be edited, saved or printed using MINITAB's graph manager (choose *Window:Manage Graphs...*).

Chapter 2

Doing It with MINITAB: Chapter 2

2.1 Data Sets

Name	Description
TECHSAL	Salaries of technical support workers, Example 4.1
PLANETS	Planet diameters
BEARINGS	Weights and diameters of 100 ball bearings,
ULTRACAL	Ultrasound calibration data, Example 2.6

2.2 Histograms and Bar Charts

Frequency histograms and bar charts are obtained by choosing *Graph:Histogram*. The dialog box will let you select the variable to be plotted and will guide you through optional choices for the display.

2.3 Boxplots

You can easily generate boxplots in MINITAB by choosing *Graph:Boxplot...* and responding to the resulting dialog box. For example, the side-by-side boxplots shown in Figure 2.13 of the text compare the salaries of men and women in the TECHSAL data set. Similar plots can be produced in MINITAB by selecting SALARY as the Y variable and GENDER as the X variable in the dialog box.

You can add information to the boxplots. For instance, the annotation button in the dialog box will allow you to label the median or outliers with their values.

2.4 Numerical Summaries

The command *Stat:Basic Statistics:Display Descriptive Statistics...* will produce numerical summaries such as the mean, median and standard deviation. These basic statistics can also be saved to the worksheet for further use. By choosing the *Graphs* option, graphical displays such as histograms and boxplots can also be generated.

2.5 Resistant Measures

In addition to the median as a resistant measure of location, choosing *Stat:Basic Statistics:Display Descriptive Statistics...* will produce a 5% trimmed mean.

Doing Lab 2-1 with MINITAB

The instructions below are numbered to correspond to the step numbers in the Experimental Procedure section of Lab 2-1.

1. Access MINITAB and read in the MINITAB worksheet crime (located in the file crime.MTP).

2.,3. Select *Stat: Basic Statistics: Display Descriptive Statistics...* from the menu bar at the top of the window. From the dialog box select AUTO as the Y variable. Select the *Graphs...* button in the dialog box, and then select *Graphical summary.* A Distribution Window will appear with a frequency histogram, a boxplot, and a table of summary measures for AUTO. The IQR is not explicitly computed, but can be obtained as $Q_3 - Q_1$. Summary measures will also appear in the Session window. These measures will include the trimmed mean, labeled TRMEAN.[1]

5.,6. Data values can be changed directly in the worksheet. However, the graphs and summary measures are not automatically updated, so they must be recomputed after each change.

[1] Actually, MINITAB computes the 5% trimmed mean. The 5% trimmed mean is obtained by first computing the number representing 5% of the data: here, $0.05 \times 50 = 2.5$. This number is then rounded: here 2.5 is rounded to 3. The corresponding number of observations is removed from the top and bottom of the data. Thus, for these data the 5% trimmed mean is the same as the 3-times trimmed mean.

Chapter 3

Doing It with MINITAB: Chapter 3

3.1 Data Set

Name	Description
WATCHES	Watch assembly times, Example 3.9

3.2 Randomly Assigning Treatments to Experimental Units

To see how to use MINITAB to randomly assign treatments to experimental units, consider again the example of watch assemblers and assembly methods from Example 3.9.

1. Begin with a worksheet consisting only of column C1 containing the numbers 1 through 15 (one for each assembler).
2. Next choose *Calc:Random Data:Sample From Columns...* . In the resulting dialog box, tell MINITAB to sample 15 (that is, all) rows from C1 and to store the samples in C2.
3. Assign the first 5 of the randomly-ordered assembler numbers in C2 to assembly method 1, the next 5 to assembly method 2 and the last 5 to assembly method 3.

3.3 Doing Lab 3-2 with MINITAB

MINITAB can only graph scatterplots of two numeric variables, so in order to produce a scatterplot of PRESS versus STUDENT, you must first assign a number to each student. Call this variable STUDNO. Assume HAND, the variable denoting the hand used, is in C2. Create and label the scatterplot by choosing *Graph:Plot...* . In the resulting dialog box, assign PRESS as the Y variable and STUDNO as the X variable. To label which hand was used, click on *Annotation:Data Labels.* In the resulting dialog box, check *Show data labels*, then select the *Use labels from column* button, and enter C2 in the response area.

Chapter 4

Doing It with MINITAB: Chapter 4

4.1 Data Sets

Name	Description
TECHSAL	Salaries of technical support workers, Example 4.1
GASKET	Gasket thicknesses, Example 4.23

4.2 Computing Probabilities

Probability distributions can be calculated by choosing *Calc:Probability Distributions...* , and the desired distribution. By choosing the *Cumulative probability* button in the resulting dialog box, you will calculate $P(Y \leq b)$. This quantity can be calculated for a column of values b, by choosing *Input column*, or for just one value, by choosing *Input constant*.

If the distribution is discrete, choosing the *Probability* button, instead of the *Cumulative probability* button, will give the value of the probability mass function, $p_Y(y) = P(Y = y)$, for either a column or single y value. If the distribution is continuous, choosing the *Probability density* button, instead of the *Cumulative probability* button, will give the value of the probability density function, $p_Y(y)$, for either a column or single y value.

4.3 Fitting a Normal Distribution

Here is a sequence of steps a data analyst might use in analyzing the gasket data in Example 4.23.

1. First, produce a time series plot of thickness versus production order (which, if done at the outset, would have saved the quality personnel in Example 4.23 a great deal of trouble). To do so, choose *Graph:Time Series Plot...* . Select THICK as the Y variable; the data are already in time order. The plot should reveal the outliers as the first two values.

2. To look at the distribution of thickness, choose *Stat: Basic Statistics: Display Descriptive Statistics...* . From the dialog box select THICK as the Y variable. Select the *Graphs...* button in the dialog box, and then select *Graphical summary.* A Distribution Window will appear with a frequency histogram with a normal density superimposed, a boxplot, and a table of summary measures. What are the salient features of the data as displayed on the boxplot and histogram? How well do you think the normal curve fits the data?

3. To produce a normal quantile plot, choose *Graph:Probability Plot...* . The default probability distribution is the normal, though you may choose other distributions. The normal quantile plot

will appear with the values of the data quantiles (or actually percentiles) on the vertical axis and those of the normal distribution quantiles on the horizontal axis: just the opposite of the graphs in the text. To assess normality, it really doesn't matter which quantities are plotted on which axes. The plot will also have a reference line and a dotted line on each side. These lines are guides: If the data are normal, approximately 95% of all points should lie within them. If much more than 5% of the points lie outside them, the normality assumption is suspect.

4. Look at the normal curve fit to the histogram and at the normal quantile plot. How do they look? Those two outliers are really causing problems, aren't they? Remove the most extreme one (the first one) by clicking on its cell in the worksheet and selecting *Edit:Delete Cells.*[1] Then re-compute the histogram and quantile plot.

   Do you like this fit any better? Perhaps you should remove the other outlier now. Proceed as in the last paragraph. With the two outliers removed, the normal density fits the histogram well, and the normal quantile plot is nearly a straight line.

4.4 Transformations

A selection of transformations is available in MINITAB by choosing *Calc:Calculator...* .

Doing Lab 4-1 with MINITAB

To do lab 4-1, merely run the macro LAB4_1. To do this, select *Edit:Command Line Editor*, and in the resulting window, type:

```
%lab4_1 N 10000
```

where N is your birth month modulo 9.[2] The required density histogram and the plot of the cumulative proportion of values $Y = 1$ versus trial will be automatically produced, and the percentage of 0s and 1s will be displayed.

Doing Lab 4-2 with MINITAB

The macro LAB4_2 will produce the necessary histogram. Required input consists of the number of simulations, N (choose 10000), and n (choose 5):

```
%lab4_2 N 5 10000
```

Output from the macro LAB4_2 consists of a density histogram for 10,000 trials just like the one you produced for the 10 trials you conducted by hand. The relative frequency of each of 0—5 successes for the 10000 measurements will be output as well.

Doing Lab 4-3 with MINITAB

The Central Limit Theorem for Rolls of a Die

1. You are going to call the macro MAKEDATA. This macro will generate random data from the discrete uniform distribution having an equal probability of producing any of the integers 1,2,3,4,5 or 6, just like a fair die. The macro will simulate the trial of rolling a fair die 50 times. It replicates this trial 250 times, producing a total of 250 times 50 or 12,500 simulated die rolls. Call the macro with the following command:

[1] If you prefer to retain the original worksheet, you may choose *Manip: Subset Worksheet* to create a subset worksheet excluding the outlier.

[2] See the introduction to this document for more about macros.

```
%makedata c1-c54
```

MAKEDATA simulates the 50 rolls of the fair die, putting the result of the i^{th} roll in variable ci. Thus each row in c1- c50 represents one replication of the trial. There are 250 such rows corresponding to the 250 replications of the trial. The macro also computes the means of the first 2, 10, 30 and 50 rolls from each trial, putting them in columns c51-c54. You should name these columns MEAN2, MEAN10, MEAN30 and MEAN50, respectively.[3] The data will be put in the current worksheet.

2. You may produce a density histogram by choosing *Graph:Histogram...*, clicking on the *Options...* button, and choosing *Density* as the type of histogram. To obtain density histograms of C1, MEAN2, MEAN10, and MEAN50 all plotted on the same scale, you can input midpoints for the bars in the Options window under *Midpoint/cutpoint positions.* We have found that the values 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 work reasonably well.
3. The construction of normal quantile plots was described earlier in this chapter.
4. Production of summary statistics was described in Chapter 2.
5. To compute the standardized means of C1, MEAN2, MEAN10, and MEAN50, choose *Calc:Calculator...*. The expression giving the standardized mean for MEAN2 is (MEAN2-3.5)/(1.71/sqrt(2)). That giving the standardized mean for MEAN50 is (MEAN50-3.5)/(1.71/sqrt(50)).

An Example Where the Central Limit Theorem Fails

The macro MAKECAU will generate 250 data sets each of 50 observations from a Cauchy distribution model. Call it just as you called MAKEDATA:

```
%makecau c1-c54
```

[3] Actually, you can specify any 54 columns for the macro: the first 50 columns named will contain the simulated die rolls and the last 4 columns named will contain the means.

Chapter 5

Doing It with MINITAB: Chapter 5

5.1 Data Sets

Name	Description
SOL	One hundred measurements of the speed of light.
BEARINGS	Weights and diameters of 100 ball bearings,

5.2 Estimation Using MINITAB

Before any inference procedure for measurement data, you should investigate the data for stationarity, outliers and non-normality. The basic measures and graphs described in Chapters 1 and 2 and the normal quantile plot described in Chapter 4 are the best ways to do this. The command *Stat:Basic Statistics:Display Descriptive Statistics...* followed by the *Graphs...* option and *Graphical Summary* will also produce one sample t confidence intervals (equation (5.8) of the text), for any confidence level you select.
Confidence intervals can also be produced for

- One sample z (known variance: equation (5.6) of the text), by choosing *Basic Statistics:1-Sample Z...* .
- One sample t (equation (5.8) of the text), by choosing *Basic Statistics:1-Sample t...* .
- Two sample t for independent populations, in either the pooled variance or unequal variance case, by choosing *Basic Statistics:2-Sample t...* .
- Paired t, by choosing *Basic Statistics:Paired t...* .
- A large sample interval for a single population proportion (equation (5.17) of the text), by choosing *Basic Statistics:1-proportion...* .
- A large sample interval for the difference between two population proportions (equation (5.25) of the text), by choosing *Basic Statistics:2-proportion...* .

5.3 MINITAB Macros

Estimation

- The macro **BIEXACT** will compute exact one sample confidence intervals for population proportions. For $Y \sim b(n, p)$ and an observed number of successes $Y = y$, a level L confidence interval is obtained by submitting the command:

```
%biexact n y L
```

- The macro **CEPRED** will compute prediction intervals for future observations from a C+E model. If the data are in column c1, a level L prediction interval is obtained by submitting the command:

```
%cepred c1 L
```

Doing Lab 5-1 with MINITAB

1. Generate the 5 data sets of size 20 by choosing *Calc:Random Data:Normal...* . From the dialog box, choose 20 rows of data and specify 5 columns in which to store the data (c1-c5, for instance). Specify the desired mean and standard deviation. Each of the 5 columns is one data set. Use the descriptive measures and graphs in MINITAB to analyze them as indicated in the lab instructions.

2. It is easiest to put the 5 $\overline{y}$ values into one column and the 5 s^2 values into another to form the histograms.

3. To create the 500 data sets, you will do what you did in part 1, but turn things sideways: put the normal random numbers into 20 columns (say c1-c20) of 500 rows each. Each row will be one data set. You will generate a new column consisting of the means of each row and another consisting of the variances of each row as follows:

   a. Choose *Calc:Row Statistics...* . Select the mean, and c1-c20 as the input variables. Store the result in a different column, say, c21.

   b. Next, select the standard deviation, c1-c20 as the input variables, and yet another column, say, c22, to store the standard deviation.

   c. Choose *Calc:Calculator...* to square column c22, giving the variance.

   d. Now analyze the columns of means and variances as asked in the lab instructions.

Doing Lab 5-2 with MINITAB

1. The macro LAB5_2 will create 100 samples from the C+E model and calculate level L confidence intervals for μ. It is called as follows:

```
lab5_2 muin sigma nobs nsets L contam
```

   Where muin and sigma are the values of the parameters μ and σ (choose what you like for this lab), nobs is the number of observations in data set (choose 20), nsets is the number of data sets (choose 100), L is the confidence level (choose .95), and contam is the contamination level (choose 0).

   The Session window will display the mean width of the 100 intervals. A graph will display the true value of μ and the computed confidence intervals. The intervals that contain the true parameter value are displayed in green and the intervals that do not contain the true parameter value are displayed in red.

3. Run the macro using the same parameters as previously, but first with a 0.1 proportion of contamination and then with a 0.5 proportion of contamination.

Chapter 6

Doing It with MINITAB: Chapter 6

Hypothesis tests can can be conducted in the following settings:

- One sample z (known variance: test statistic (6.2) in the text), by choosing *Basic Statistics:1-Sample Z...* .
- One sample t (test statistic (6.3) in the text), by choosing *Basic Statistics:1-Sample t...* .
- Two sample t for independent populations, in either the pooled variance (test statistic (6.11) in the text) or unequal variance (test statistic (6.12)) case, by choosing *Basic Statistics:2-Sample t...* .
- Paired t, by choosing *Basic Statistics:Paired t...* .
- A large sample test for a single population proportion (test statistic (6.5) in the text), by choosing *Basic Statistics:1-proportion...* .
- A large sample test for the difference between two population proportions (test statistic (6.13) or (6.14) in the text), by choosing *Basic Statistics:2-proportion...* .

In each setting, you can choose the value of the parameter (or difference of parameters in the two sample setting) under the null hypothesis. You can also specify a two-sided or one-sided alternative.

For the exact test for the binomial parameter p, the appropriate tail areas are easily computed using *Calc:Probability Distributions:Binomial...* .

Doing Lab 6-1 with MINITAB

The instructions below are keyed to the instructions in the text.

The Meaning of Statistical Significance and p-values

1. Use the macro LAB6_1N to generate 1 set of 10 observations from a $N(25,1)$ distribution. The macro will also compute the t statistic and the p-value for testing $H_0 : \mu = 25$ versus $H_a : \mu \neq 25$ for this data set. The macro is called as follows:

   ```
   %lab6_1n muin sigma nobs nsets tstat pval
   ```

   where you choose muin=25, sigma=1, nobs=10, nsets=1, and you give two columns for tstat and pval, say c1 and c2. The t statistic will be output to the column you selected for tstat and its p-value will be output to the column you selected for pval.

3. Now use the same macro to generate 1000 sets of 10 observations each from a $N(25,1)$ distribution, and to compute the t statistic and p-values for each. Assuming you want to put the t statistics in c3 and the associated p-values in c4, the command will be

```
%lab6_1n 25 1 10 1000 c3 c4
```

4. You can use MINITAB to obtain the proportion of the 1000 test statistics that provide as much evidence against the null and in favor of the alternative hypothesis as does t^*, and the proportion of p-values as small as or smaller than the p-value associated with t^*. One way to do this is to sort the columns of t statistics and p-values, by choosing *Manip:Sort...* .

How Nonnormality Affects the Results

1.,3. Use macro LAB6_1E to generate the initial set of 10 observations from the exponential distribution with mean $\mu = 25$, and put the t statistic and the p-value for testing $H_0 : \mu = 25$ versus $H_a : \mu \neq 25$, in c1 and c2, respectively:

```
%lab6_1e 10 1 1 c1 c2
```

This command also creates a histogram based on 1000 observations from this distribution.

5. Now use the macro to generate 1000 sets of 10 observations each from an exponential distribution with mean 25, and to compute the t statistic and p-values for each. If you submit the command

```
%lab6_1e 10 1000 0 c3 c4
```

the 1000 sets of t statistics and p-values will be output to c3 and c4, respectively.

6. Use MINITAB to obtain the proportion of the 1000 test statistics that provide as much evidence against the null and in favor of the alternative hypothesis as does t^*. Obtain the proportion of p-values as small as or smaller than the p-value associated with t^*.

Chapter 7

Doing It with MINITAB: Chapter 7

7.1 Data Sets

Name	Description
TWEAR	Tool wear data
TWEAR8	Tool wear data for VELOCITY=800
FUEL	Fuel consumption versus equivalence ratio
DRAFTLOT	1970 draft lottery data
TRAPDATA	Bacterial trap data
DONNER	Donner party data
DERBY	Kentucky Derby data

7.2 The Median Trace

To use the macro MTRACE to compute a median trace, begin by reading the desired X and Y variables into the current worksheet, and then call the macro with the names of the X and Y variables. For example, to do a median trace like Figure 7.9 for the draft lottery data, submit the command

```
%mtrace 'bdate' 'number'
```

You will be prompted in the Session Window for the number of slices (enter 12), and numbers of observations in each slice (enter 31 29 31 30 31 30 31 31 30 31 30 31).

7.3 The Tool Wear Data

To generate a plot like Figure 7.1, choose *Graph:Plot...* . From the resulting dialog window, select WEAR as the Y and TIME as the X variable. A scatterplot window will appear.

To generate a plot like Figure 7.5, choose *Graph:Plot...* . From the resulting dialog window, select WEAR as the Y and TIME as the X variable, and under ***Data display:*** set ***Display*** to ***Symbol***, ***For each*** to ***Group*** and ***Group variables*** to *velocity*.

You can obtain a scatterplot like that in Figure 7.6 from the data set TWEAR8.

7.4 Correlation

To standardize variables in MINITAB, select *Calc:Standardize...* . Try this now for the two variables WEAR and TIME in the data set TWEAR8. Plot the standardized variables against each other.

To find the correlation of the tool wear data for VELOCITY=800, access TWEAR8 and choose *Stat:Basic Statistics:Correlation...* . From the resulting dialog window select TIME and WEAR and ORDER as the Y variables. The Pearson correlation between each pair will be displayed in the Session Window.

The macro CORR will compute a confidence interval for the Pearson correlation. To compute a level L interval with the variables in columns c1 and c2, submit

```
%corr c1 c2 L
```

7.5 Regression

Least Squares Fit

It is very easy to compute the least squares estimators using MINITAB: just choose *Stat:Regression:Regression...*, and select the X and Y variable from the dialog window. Output is sent to the Session Window and includes the fitted model, and the estimated standard error, the t statistic and the p-value for both slope and intercept. There are several buttons that allow you to customize the analysis:

- You may control what is output by selecting the *Results...* button.
- The *Graphs...* button allows you to specify the residual plots produced. Among the selections, are the type of residual: regular, which is the default, and Studentized, here called *Deleted*. You can also specify the type of plot: histogram, normal quantile, versus fitted, versus order and versus any other variable. After the regression has been run and the residuals (or Studentized residuals) obtained, much the same set of graphs is available by choosing *Stat:Regression:Residual Plots...* .
- Among the options available from the *Options...* button, is the creation of a prediction interval for a new observation at a specified value of the regressor. A confidence interval for the mean response at the chosen regressor value is also produced when this option is selected.
- the *Storage...* button stores different quantities, such as residuals, Studentized residuals and parameter estimates, in worksheet columns.

Choosing *Stat:Regression:Fitted Line Plot...* will graph the original data along with the fitted line, and, if desired, confidence and prediction bands.

The macro TQPLOT will produce a plot of the studentized residuals versus quantiles of the appropriate t distribution. Before calling TQPLOT, you must save the Studentized residuals to the worksheet. The arguments to TQPLOT are studres and df, where studres is the studentized residual produced from the regression and df is the degrees of freedom for error (i.e., associated with MSE). For instance, suppose that TWEAR8 is the current worksheet, that WEAR has been regressed on TIME, and that the Studentized residuals are in the variable STUDRES. Since there are 7 degrees of freedom associated with error, the following will produce the plot for the studentized residuals from the regression of WEAR on TIME:

```
%tqplot 'studres' 7
```

7.6 Categorical Data

In MINITAB you can graphically analyze data for a single categorical variable using a bar chart. A bar chart is obtained by selecting *Graph:Chart...*, and choosing the variable defining the categories as the X variable. In the DONNER data, for example, a bar chart showing the numbers of males and females is obtained by specifying GENDER as the X variable. Mosaic plots for displaying the association between two categorical variables are not available in MINITAB.

A one-way table can be created by choosing *Tables:Tally...* . Two-way tables can be generated by selecting *Tables:Cross Tabulation...* . Input includes the names of the two classification variables (to compare GENDER and FATE for the DONNER data, these are GENDER and FATE). You can customize the table by choosing what to display (counts, row, column and/or total percents). You can also ask for a χ^2 test.

Doing Lab 7-1 with MINITAB

The instructions below are keyed to the instructions in the text.

1. Access the macro LAB7_1. This macro will generate a bivariate data set, and will display a plot of the response versus regressor variable.

2. In the Session Window, you will be asked to give an intercept and slope for a line you think best fits the data. Take a few minutes to formulate an educated guess before answering. Following instructions in the window, enter your guesses.

3. The program will then plot your line superimposed on a plot of the data, and also plot the residuals from your "fit" versus the regressor. How did you do? From the plot you should see how your fitted line can be improved. The SSE for the line you fit will appear in the Session Window. Mark the SSE down.

4. Using the feedback from the data plots, try to improve your fit. Make another guess, submitting numbers as you did before. In terms of SSE and of the data plots how did you do?

   Keep track of your best fit and keep trying until you think you've done as well as you can.

5. When you are satisfied you have done as well as you can, answer 'n' to the prompt asking if you want to try again. This will get you the least squares fits of slope and intercept and the minimum SSE. How does your best fit compare?

Chapter 8

Doing It with MINITAB: Chapter 8

8.1 Data Sets

Name	Description
TREES	Volumes, heights and diameters of 31 black cherry trees, Example 8.1

8.2 The Graphical Exploration of Multi-Variable Data

Scatterplot Arrays and Brushing

To create a scatterplot array for the tree data in MINITAB choose *Graph:Matrix Plot...* . In the dialog box, select the variables to be plotted.

To create a brush on the scatterplot array, or any other MINITAB graph for that matter, first choose *Editor:Brush*. This will convert the cursor into a brushing tool, indicated by a small hand with a pointing index finger. Position the brushing tool where you want one corner of the brush to appear and click and hold the left mouse button while moving the cursor toward where you want the diagonally opposite vertex of the rectangular brush to be. See Section 8.5 of the text for more on this. The brush may be moved by placing the cursor on one of the sides of the rectangle (away from a vertex), clicking the left mouse button and dragging.

3-D Plots

To create a 3-D plot in MINITAB, choose *Graph:3D Plot...* . Do this now for the tree data. From the resulting window choose V as the Z variable, H as the Y variable and D as the X variable. A graph window will appear. These plots cannot be rotated using the cursor, but by choosing the *3D Effects* button and then *View* from the 3D Graph dialog box, you can change the orientation and other features of the display. Generating a sequence of 3-D plots, each with a different orientation, is the next best thing to being able to rotate the plot.

Fitting Models (8.22) and (8.23)

To fit model (8.23), follow the steps for simple linear regression given in Chapter 7, with the difference that two regressor variables (D and H) are selected. In order to fit model (8.22), you must first create the product term DH. Use *Calc:Calculator...* to do this. The regression is then conducted with the regressors D, H and their product.

Centering Predictors

To avoid the computational and statistical difficulties associated with multicollinearity, we might want to center both D and H by subtracting the mean of the tree diameters from each tree's diameter and the mean of the tree heights from each tree's height. This has already been done in this data set with the variables CD and CH being the centered variables.

To center the predictor D, for instance, choose *Calc:Calculator...* and submit the expression D-mean(D), with the result stored in a column of your choosing.

Residuals, Studentized residuals and their plots are generated just as for the simple regression case described in Chapter 7. The macro TQPLOT will generate a plot of Studentized residuals versus quantiles of the appropriate t distribution in the multiple regression setting as well.

Confidence and prediction intervals are generated just as for the simple regression case described in Chapter 7. As an example, suppose we want to use model (8.22) and the tree data to obtain level 0.95 intervals for the mean volume and to predict the volume of a new tree having diameter 10 inches and height 70 feet. To get the desired results, follow these steps:

1. Create the variable $DH = D \cdot H$ using the calculator (*Calc:Calculator...*).
2. Call up the regression dialog box by choosing *Stat:Regression:Regression...* .
3. In the dialog box, enter V as the response and D, H and DH as the "predictors".
4. Click on the *Options* button. In the resulting dialog box,
   a. Enter 10 70 700 in *Prediction intervals for new observations.*
   b. Check that the confidence level is set to its default value, 95.
   c. Check the boxes for confidence limits and prediction limits.

The results will appear in the Session Window.

Backward Elimination

MINITAB offers a particularly easy way to remove one variable at a time from a fitted regression model. At each step, the variable with the smallest individual t ratio (equivalently, largest p-value) will be removed. Here are the steps involved:

1. Select *Stat:Regression:Stepwise...* .
2. In the resulting dialog box,
   a. Enter the response where indicated.
   b. Enter all the regressors you want to begin with in the *Predictors:* space.
   c. Enter all the regressors you want to begin with in the *Enter:* space.
   d. Click on the *Options* button. In the resulting dialog box, enter 10000 as *F to enter* and 9999 as *F to remove.*

The results will be displayed in the Session Window.

Doing Lab 8-1 with MINITAB

The instructions below are keyed to instructions in the text.

Experimental Procedure

Data Generation

To generate the data sets as in 1.-3., call the MINITAB macro LAB8_1. The arguments to this macro are the columns in which you wish to store the response and the two regressors, in that order. For instance, if you wanted to store the response in c8 and the two regressors in c3 and c4, you would submit the command

```
%lab8_1 c8 c3 c4
```

You will be prompted for the number of observations, the parameters of the model, and the desired correlation between the predictor variables. All quantities except the first and last remain the same for all three data sets. To refresh your memory, they are:

- Number of observations: 20
- $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $\sigma^2 = 1$

Analysis

1. The Pearson correlation between the regressors may be calculated in MINITAB using *Stat:Basic Statistics:Correlation...* .

2.-3. All required output can be obtained using *Stat:Regression...* . In the resulting dialog box, click on the *Options:* button, and in the dialog box that appears, check the *Display variance inflation factors* box.

Doing Lab 8-2 with MINITAB

Experimental Procedure

The instructions below are keyed to instructions in the text.

Data Generation

To generate the data set, call the MINITAB macro LAB8_1, as documented above. Choose the correlation to be 0.5.

Analysis

Look at the Data.

1.-2. Use MINITAB to plot X_1 versus X_2 and to regress Y on X_1 and X_2.

3. Generate Studentized residuals. How to do this in MINITAB is described earlier in this chapter.

4. Generate a t quantile plot of the Studentized residuals by running the MINITAB macro TQPLOT as described earlier in this chapter. Now look at the plot. Are any major problems evident?

Create an Outlier and See What Happens. To change a data value in MINITAB, click on the cell in the data window containing the value, type in the new value and hit the return key. The new value will now replace the old one. To get updated output, you will have to rerun the procedures you used to get the original output.

Chapter 9

Doing It with MINITAB: Chapter 9

9.1 Data Sets

Name	Description
PROSTATE	Efficacy of different treatments on benign prostate hyperplasia, Example 9.1
WATCHES	Watch assembly times, Example 9.3

9.2 Interval Plots: An Alternative to Mean Diamonds

MINITAB does not produce mean diamonds, but it has a similar plot called an *interval plot.* The default interval plot graphs the group means versus groups just as in the mean diamond plots. Over each group mean is plotted a vertical interval with endpoints $\overline{y} \pm s/\sqrt{n}$, where $\overline{y}$ is the group mean, s is the group standard deviation and n is the number of data in the group.[1]

Interval plots can be customized in a number of ways. One is to have the interval endpoints at $\overline{y} \pm k \cdot s/\sqrt{n}$, where k is a user-specified constant. Another is to display a level L confidence interval for the mean, where the user selects L.

Interval plots may be obtained by choosing *Stat:ANOVA:Interval Plots...* . Different options may be selected from the dialog box.

9.3 Model Fitting in MINITAB

To get means and standard deviations, as well as other basic summaries, of the response for each factor level, choose *Stat:Basic Statistics:Display Descriptive Statistics...* . Under *Variables:* put the response, then check the *By variable:* box, and enter the name of the factor.

The one-way model can be fit by choosing *Stat:ANOVA:One-way...*, but for users who have the full version of MINITAB, we suggest using *Stat:ANOVA:General Linear Model...* .[2] There are several reasons to prefer the General Linear Model routine:

1. The General Linear Model routine can be used to analyze both the one-way and RCB models of Chapter 9 and the factorial models of Chapter 10.

[1] Note that $s/\sqrt{n}$ is the standard error of the mean. Contrast this with the values $\overline{y} \pm s$ which are the upper and lower vertices of the mean diamonds.

[2] The general linear model is a general framework that encompasses a large number of models, including the regression models of Chapters 7 and 8, the one-way and RCB models of Chapter 9 and the factorial models of Chapter 10. Unfortunately, the General Linear Model routine is not available in the student version of MINITAB.

2. The General Linear Model routine gives the correct analysis for more complicated designs. For example, it does the Type III analysis, and can do the TYPE I analysis as discussed in Chapter 10.

3. The General Linear Model routine can produce a greater range of output, such as the studentized residuals that we favor in the analysis of model fit.

Model Fitting Using the General Linear Model Routine

To fit the one-way model, select *Stat:ANOVA:General Linear Model...* . In the dialog box, input the name of the response where indicated, and the name of the factor under *Model:* Many choices for analysis and output are available. Among those you will want to consider are:

- Under *Comparisons...*, the Tukey and Bonferroni comparisons.
- Under *Storage...*, Studentized residuals (called *Deleted t residuals*).
- Under *Graphs...*, various plots of regular and Studentized residuals. (As usual, we recommend running the macro TQPLOT to get a proper t quantile plot of the Studentized residuals, using the Studentized residuals saved to the worksheet, and the degrees of freedom for error taken from the ANOVA table.)

For the RCB model, there are two factors, corresponding to the blocks and treatments. In the General Linear Model dialog box, put both factors under *Model:*.

To produce interaction plots, select *Stat:ANOVA:Interactions Plot...* . The macro TUKEY will also produce the interaction plots in addition to conducting Tukey's test for additivity. It is called as

```
%tukey y treat block
```

where y is the column of responses, treat is the column denoting the treatment and block is the column denoting the block.

Model Fitting Using the Oneway Routine

For those who do not have the General Linear Model routine, one-way models may be fit by choosing *Stat:ANOVA:One-way...* . The *Comparisons...* button brings up a dialog box that allows you to select multiple comparisons such as Tukey and Fisher (i.e., LSD). The *Graphs...* button brings up a dialog box that allows you to select plots of the data and of the ordinary residuals. If the data are in separate columns, you should choose *Stat:ANOVA:One-way(Unstacked)...* .

Model Fitting Using the Twoway Routine

For those who do not have the General Linear Model routine, the RCB model may be fit by choosing *Stat:ANOVA:Two-way...* . This routine has no provision for doing multiple comparisons.

To produce interaction plots, select *Stat:ANOVA:Interactions Plot...* . The macro TUKEY will produce the interaction plots and Tukey's test for additivity. It is called as

```
%tukey y treat block
```

where y is the column of responses, treat is the column denoting the treatment and block is the column denoting the block.

9.4 Doing Lab 9-2 with MINITAB

The following sections correspond to items 1-3 of the lab description in the text.

1. The MINITAB macro LAB9_2A will generate data sets from the one-way model with five populations having means 5, 2, 2, 2, and 2 and common variance 1. The data sets all have equal sample sizes of five from each population.

   Use this macro now to generate three data sets each with five observations per population, by submitting the command

   ```
   %lab9_2a c1-c4
   ```

   The response variables will be put in columns c1-c3 and the variable denoting the population will be in c4 (You may, of course, select any set of four columns).

2. Use MINITAB to compute individual (LSD) and multiple (TUKEY) comparisons for all three data sets. Take the confidence level to be 0.95. For the individual comparisons count the number of the three data sets in which there is at least one mistaken conclusion (i.e. an interval which does not contain the true mean difference). Record the result. Now do the same for the Tukey multiple comparisons.

3. The MINITAB macro LAB9_2B does exactly what you did in generating the three sets of data from the one-way model, computing individual and Tukey multiple comparison confidence intervals and checking to see for each type of comparison how many of the data sets have at least one mistaken conclusion. The only difference is that the macro will do all this for any number of data sets (not just three), and will do it all much faster than you can. You need only input the number of data sets you want generated. The output is the number of those data sets which contain at least one mistaken conclusion. Run this macro now for 1000 data sets by submitting the command

   ```
   %lab9_2b 1000
   ```

   What results do you observe?

Chapter 10

Doing It with MINITAB: Chapter 10

10.1 Data Sets

Name	Description
FSHAKER2	Balanced pulse oximetry data, Example 10.1
FSHAKER4	Unalanced pulse oximetry data, Example 10.6
PEANUTS4	Peanut data, Example 10.5

10.2 Model Fitting in MINITAB

To get means and standard deviations, as well as other basic summaries, of the response for each factor level, choose *Stat:Tables:Cross Tabulation...* . Under *Classification variables:* put the factors, then click on the *Summaries...* button, and enter the response as an *Associated variable.* Check the desired summaries, which should include at least the mean, standard deviation and number of non-missing observations.

To fit the additive model ((10.14) in the text), select *Stat:ANOVA:General Linear Model...* . In the dialog box, input the name of the response where indicated, and the names of the factors under *Model:*. You may customize the output as discussed in chapter 9. To produce interaction plots, select *Stat:ANOVA:Interactions Plot...* .

To fit the general model ((10.16) in the text), proceed as with the additive model, but after selecting the factors, include the interaction term as a*b, where a is the name of factor 1 and b is the name of factor 2. For example, for the pulse oximetry data, what appears under model is

```
shivtype intensiy shivtype*intensiy
```

This can be done using the shorthand

```
shivtype|intensiy
```

Note, however, that this shortand, if applied to three factors will produce the model with all main effects and two factor interactions as well as the three factor interaction.

As described in Chapter 9, options in the generalized linear model dialog box allow you to produce residual and Studentized residual plots, multiple comparisons and to store the Studentized residuals for plotting using the TQPLOT macro.

Chapter 11

Doing It with MINITAB: Chapter 11

11.1 Data Sets

Name	Description
TISSUEPH	Actual and estimated pH of rabbit tissue, Example 11.2
FUEL	Fuel consumption versus equivalence ratio, Example 11.4
PROSTATE	Efficacy of different treatments on benign prostate hyperplasia, Example 11.5
WATCHES	Watch assembly times, Example 11.6

11.2 The Sign Test and Estimation of the Median

To perform the sign test or to obtain the associated point estimate and confidence interval, choose *Stat:Nonparametrics:1-Sample Sign...* . When performing either procedure for paired data, such as those given in Example 11.2, you must take the differences of each pair first. You can do so using the calculator or by choosing *Stat:Nonparametrics:Pairwise Differences...* .

11.3 The Wilcoxon Signed Rank Test

To perform the sign test or to obtain the associated point estimate and confidence interval, choose *Stat:Nonparametrics:1-Sample Wilcoxon...* . The Walsh averages may be computed and saved to the worksheet by choosing *Stat:Nonparametrics:Pairwise Averages...* .

11.4 The Wilcoxon Rank Sum Test

The Mann-Whitney test is a statistical test of the difference in central location in two independent populations that is equivalent to the Wilcoxon rank sum test. Either the test or a confidence interval for the difference in location can be obtained by choosing *Stat:Nonparametrics:Mann-Whitney...* .

11.5 Spearman Correlation

To compute the Spearman correlation, first compute the ranks of the variables to be correlated by choosing *Manip:Rank...* . The Spearman correlation is the Pearson correlation of the ranks, which is computed by choosing *Stat:Basic Statistics:Correlation...* .

11.6 The Kruskal-Wallis Test

To perform the Kruskal-Wallis test, choose *Stat:Nonparametrics:Kruskal-Wallis...* .

11.7 Friedman's Test

To perform the Friedman test, choose *Stat:Nonparametrics:Friedman...* .

11.8 The Two Sample Pitman Test

The macro TWORAND will approximate the p-value closely using a randomiztion test. To get a good approximation of the p-value, choose a large number of randomizations when prompted: 10,000 should be do-able on most computers.

11.9 Fisher's Exact Test

Fisher's exact test is not available in MINITAB.

11.10 The One Sample Pitman Test

The macro ONERAND will approximate the p-value closely using a randomiztion test. ONERAND is called as follows:

```
%onerand y theta0 nrands
```

where y is the column containing the data, theta0 is the hypothesized median value under H_0, and nrands is the number of randomizations. To get a good approximation of the p-value, choose a large number of randomizations when prompted: 10,000 should be do-able on most computers.

11.11 The Generalized Kruskal-Wallis Test

The macro GKWRAND will conduct a randomization test version of the generalized Kruskal-Wallis test. GKWRAND is called as follows:

```
%gkwrand y class nrands
```

where y is the column of responses, class is the column containing the classification variable and nrands is the number of randomizations. To get a good approximation of the p-value, choose a large number of randomizations when prompted: 10,000 should be do-able on most computers.

By using the ranks of the data as the response variable, you will obtain a randomization test version of the Kruskal-Wallis test.

11.12 The Generalized Friedman Test

The macro GFRAND will conduct a randomization test version of the generalized Friedman test. GFRAND is called as follows:

```
%gfrand y treat block nrands
```

where y is the column of responses, treat is the column indicating the treatment, block is the column indicating the block and nrands is the number of randomizations. To get a good approximation of the p-value, choose a large number of randomizations when prompted: 10,000 should be do-able on most computers.

By using the ranks of the data as the response variable, you will obtain a randomization test version of Friedman's test.

11.13 Bootstrap Inference

Before any bootstrap inference procedure for measurement data, you should investigate the data for outliers. MINITAB graphics are the easiest way to do this.

The C+E Model

The macro CEBOOT1 computes one sample (equation (11.16) of the text) bootstrap BC_a confidence intervals for the mean of a C+E model. CEBOOT1 uses the sample mean as the estimator. It is called as follows:

```
%ceboot1 y level nboots
```

where y is the column of responses, level is the desired confidence level (between 0 and 1) and nboots is the number of bootstrap samples. CEOOT1 also computes the bootstrap prediction interval given by equation (11.17) of the text.

The macro CEBOOT2 computes two sample bootstrap (equation (11.20)) BC_a confidence intervals for the difference in the means of two independent populations, each defined by the C+E model. CEBOOT2 uses the difference of the sample means as the estimator. CEBOOT2 is called as follows:

```
%ceboot2 y1 y2 level nboots
```

where y1 and y1 are the columns of responses, level is the desired confidence level (between 0 and 1) and nboots is the number of bootstrap samples.

The macro VBOOT computes two sample bootstrap (equation (11.21) of the text) BC_a confidence intervals for the quotient of variances of two independent populations, each defined by the C+E model. VBOOT uses the quotient of the sample variances as the estimator. VBOOT is called as follows:

```
%vboot y1 y2 level nboots
```

where y1 and y1 are the columns of responses, level is the desired confidence level (between 0 and 1) and nboots is the number of bootstrap samples.

The Binomial Model

The macro BIBOOT2 will compute two sample bootstrap BC_a confidence intervals for population proportions (equation (11.22) of the text). It is called as follows:

```
%biboot2 n1 n2 y1 y2 level nboots
```

where y1 is the observed number of successes in n1 trials from population 1, y2 is the observed number of successes in n2 trials from population 2, level is the desired confidence level (between 0 and 1) and nboots is the number of bootstrap samples.

Chapter 12

Doing It with MINITAB: Chapter 12

12.1 Data Sets

Name	Description
SF	Surface finish data, Example 12.1
SF31	Unreplicated surface finish data, Example 12.2
SF32	Surface finish data with center points, Example 12.2
WASH	Washing test scores, Example 12.3
PLUGS	Sparkplug removal times, Example 12.4
PLANES	Paper airplane flight times, Example 12.5

12.2 Generating 2^k Designs

In order to analyze 2^k designs using MINITAB's DOE (Design Of Experiment) capability, the design must either be

- Created by choosing *Stat:DOE:Create Factorial Design...* . This method will create the design and additional information needed by MINITAB for analysis and store it all in the current worksheet. You must then add the responses in the appropriate worksheet cells.
- An existing design, such as one found in the list of data sets at the beginning of this chapter, modified by MINITAB in the current worksheet to make it suitable for analysis by MINITAB. This is done by putting the design in the current worksheet and choosing *Stat:DOE:Define Custom Factorial Design...* .

Creating A New Design

You can create a new design by choosing *Stat:DOE:Create Factorial Design...* . Follow these steps:

1. Make sure you choose *2-level factorial (default generators)* by clicking its button. Select the number of factors.

2. Click the *Designs...* button. In the resulting dialog box, highlight the *Full factorial design.*

3. Specify the options you want to change: the number of center points, the number of replicates of corner points, and the number of blocks.

4. Click *OK* to return to the Create Factorial Design dialog box.

5. Now you can choose to customize the design. Here are some possibilities:

   - Click on the *Factors...* button to name and specify the levels of the factors.
   - Click on the *Options...* button to change default settings which randomize the order of the runs and store the design in the worksheet.
   - Click on the *Results...* button to customize the output.

If you do not tell MINITAB otherwise in the *Options...* window, the design will be put into the worksheet ready for you to enter the responses.

Preparing An Existing Design for Analysis

Suppose you have a pre-existing design in your worksheet, such as the surface finish data in the data set SF32 from Example 12.2 of the text. Then you will need to prepare it for analysis for MINITAB. Here is how to do it for the mold data:

1. Since (1) the data have center points, (2) MINITAB needs a separate column to identify the center points and (3) the data have no such column, you must create a column which contains the value 0 for each center point and 1 otherwise.

2. Choose *Stat:DOE:Define Custom Factorial Design...* .

3. In the resulting dialog box, select the factors and make sure the *2-level factorial* button is selected (It should be: it's the default).

4. Since the design has center points, click on the *Design...* button and identify the column you created to specify the center points.

When you are done, the additional information MINITAB needs to analyze the design will be written to the worksheet, and you will be ready to proceed with the analysis.

12.3 Analysis of Unreplicated 2^k Experiments

Once the design and responses are in the current worksheet, the design is analyzed by choosing *Stat:DOE:Analyze Factorial Design...* . The dialog box asks you to input the response(s), and offers a number of option buttons, among which are:

- *Graphs...* Options available here include selection of residual plots to display and the type of residual to plot (deleted residuals are Studentized residuals). You can also choose a normal quantile plot of the estimated effects.

- *Terms...* This window allows you to choose which effects to include in the model.

- *Results...* This window allows you to customize what is output.

- *Storage...* Allows the storage of various quantities produced by the analysis, such as residuals and fitted values, in the worksheet.

The Lenth Procedure

On the normal quantile plot of effects, MINITAB labels those effects determined to be significant at the 0.90 confidence level by the Lenth procedure. However, MINITAB does not make MOE or SMOE available. In order to get MOE and SMOE, first analyze the model and save the estimated effects to a column in the worksheet using the*Storage...* button. Then run the macro LENTH:

```
%lenth effects level
```

where effects is the column containing the effects and level is the confidence level desired.

12.4 Analysis of Replicated 2^k Experiments

Analysis of replicated 2^k experiments is conducted in exactly the same way as for unreplicated 2^k experiments. The kind of output obtained is different, however.

The analysis done by MINITAB is the same as that conducted in the text, but the output is packaged differently. Basically, MINITAB does a regression, and produces regression output. Instead of comparing the estimated effects to MOE and SMOE, you will look at their p-values from the individual t tests.

12.5 Interaction Plots

Interaction plots can be generated by selecting *Stat:DOE:Factorial Plots...* . In addition to the plots for two factors illustrated in the text, MINITAB can produce main effect plots, which compare responses for levels of main effects, and three dimensional cube plots, which compare three factors.

Chapter 13

Doing It with MINITAB: Chapter 13

13.1 Data Sets

Name	Description
SF32	Surface finish data with center points, Example 13.1
MOLD	EVA ring data, Example 13.7
HANGER	Picture hanger data, Example 13.9
HANGERR	Reduced picture hanger data, Example 13.9

13.2 Obtaining a Design

Creating A New design

You can create new designs of a given resolution by choosing *Stat:DOE:Create Factorial Design...* . To see if a design of the desired resolution is possible, click on the *Display Available Designs...* button. Once you have chosen the design, click on *OK* to return to the Create Factorial Design dialog box. Now follow these steps:

1. Make sure you choose *2-level factorial (default generators)* by clicking its button. Select the number of factors.
2. Click the *Designs...* button. In the resulting dialog box, highlight the design you want.
3. Specify which, if any, options you want to change: the number of center points, the number of replicates of corner points, and the number of blocks.
4. Click *OK* to return to the Create Factorial Design dialog box.
5. Now you can choose to customize the design. Here are some possibilities:
    - Click on the *Factors...* button to name and specify the levels of the factors.
    - Click on the *Options...* button to create a foldover design (choose *Fold on all factors* to get the kind of foldover design discussed in the text.) You can also change default settings which randomize the order of the runs and store the design in the worksheet.

- Click on the *Results...* button to customize the output.

If you do not tell MINITAB otherwise in the *Options...* window, the design will be put into the worksheet ready for you to enter the responses.

Preparing An Existing Design for Analysis

Suppose you have a pre-existing design in your worksheet, such as that in the data set MOLD from Example 13.7 of the text. Then you will need to prepare it for analysis for MINITAB. Here is how to do it for the mold data:

1. Choose *Stat:DOE:Define Custom Factorial Design...* .
2. In the resulting dialog box, select the factors and make sure the *2-level factorial* button is selected (It should be: it's the default).
3. Since the design has blocks, click on the *Design...* button and identify the blocks column.

When you are done, the additional information MINITAB needs to analyze the design will be written to the worksheet, and you will be ready to proceed with the analysis.

13.3 Analyzing the Design

Analysis of a 2^{k-p} design proceeds exactly as for a 2^k design and is described in Chapter 12. In addition to the output described there, the aliasing structure of the design is also written to the Session Window.

13.4 Robust Parameter Design

MINITAB has tools to help analyze robust parameter designs. In order to make use of these tools, the data must be in the same form as shown in Table 13.13. That is, each row contains one set of inner array factor levels and all the response values corresponding to those inner array factor levels.

Once you have the data in this form, you can access MINITAB's robust parameter design tools by choosing *Stat:DOE:Analyze Inner/Outer Array Design...* . In the resulting dialog box:

1. Under *Responses data are in:*, enter the columns containing the responses.
2. Under *Inner array data are in:*, enter the columns constituting the inner array.
3. Click on the *Graphs...* button. If you must analyze the signal-noise ratio,[1] we recommend selecting all options: signal-noise ratio, means and standard deviations, and also using the log of the standard deviations.
4. Click on the *Options...* button and choose the signal-noise ratio desired.
5. Click on the *Store...* button and choose the quantities to store. We recommend the means and standard deviations, and if you must, the signal-noise ratio.

When you submit the commands by clicking on *OK* in the main dialog box, two things will happen: (1) A set of main effect and interaction graphs will be generated for each response (signal-noise ratio, mean or standard deviation) chosen. (2) The desired response will be put in the worksheet. The graphs are useful in suggesting which effects are significant, and why. The computed values can then be input as responses to *Stat:DOE:Analyze Factorial Design...* for a detailed analysis.

[1] Recall from the text that we do not recommend using the signal-noise ratio, but rather urge users to analyze the mean and the log of the standard deviation separately.

Chapter 14

Doing It with MINITAB: Chapter 14

14.1 Data Sets

Name	Description
CAM1	Cam data 2^2 design, Example 14.1
CAM2	Cam data CCD design, Example 14.1

14.2 Creating a Central Composite Design

Creating A New design

You can create a new CCD by choosing *Stat:DOE:Create RS Design...* . Once you have chosen the design, click on *OK* to return to the Create Factorial Design dialog box. Now follow these steps:

1. In the resulting dialog box, select Central Composite and the number of factors.

2. Click the *Designs...* button. In the resulting dialog box, highlight the design you want.

3. Specify which options, if any, you want to change: the number of center points and the *Value of Alpha*, which will give you a rotatable design (default), a face-centered design, or one of your own choosing.

4. Click *OK* to return to the Create RS Design dialog box.

5. Now you can choose to customize the design. Here are some possibilities:

    - Click on the *Factors...* button to name and specify the levels of the factors.
    - Click on the *Options...* button to change default settings which randomize the order of the runs and store the design in the worksheet.
    - Click on the *Results...* button to customize the output.

If you do not tell MINITAB otherwise in the *Options...* window, the design will be put into the worksheet ready for you to enter the responses.

Preparing An Existing Design for Analysis

Suppose you have a pre-existing design in your worksheet, such as that in the data set CAM2 from Example 14.1 of the text. Then you will need to prepare it for analysis for MINITAB. Here is how to do it for the mold data:

1. Choose *Stat:DOE:Define Custom RS Design...* .
2. In the resulting dialog box, select the factors.
3. If the design has blocks, or if it has columns designating the run order or standard order as being different than the order of the data, click on the *Design...* button.

When you are done, the additional information MINITAB needs to analyze the design will be written to the worksheet, and you will be ready to proceed with the analysis.

14.3 Analyzing the Design

Analysis of a response surface design begins with *Stat:DOE:Analyze RS Design...*, which proceeds exactly as for a 2^k design and is described in Chapter 12. MINITAB just fits a quadratic regression to the data.

MINITAB does not perform an eigen-analysis of the fitted model. To obtain the stationary point and the eigenvectors and eigenvalues that give valuable information about the shape of the fitted surface, run the macro EIGEN. To do so, first put the estimated first-order coefficients $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$, into a column of the worksheet. Put the estimated second-order coefficients that constitute the matrix $\hat{B}$ into k columns, corresponding to the k columns of the matrix:

$$\begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \hat{\beta}_{13}/2 & \cdots & \hat{\beta}_{1k}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} & \hat{\beta}_{23}/2 & \cdots & \hat{\beta}_{2k}/2 \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\beta}_{1k}/2 & \hat{\beta}_{2k}/2 & \hat{\beta}_{3k}/2 & \cdots & \hat{\beta}_{kk} \end{bmatrix}.$$

Then run the macro:

```
%eigen betac bc
```

where betac is the column containing the estimated first-order coefficients and bc is a list of columns containing the matrix $\hat{B}$.

MINITAB does produce contour and 3-D surface plots of the fitted surface. Choosing *Stat:DOE:RS Plots...* will generate these plots.

14.4 Doing Lab 14-1 with MINITAB

The macro QUADGEN will prompt you to input values for x_1 and x_2, and will output the value of the response, y. Use it to attempt OFAT optimization. QUADGEN is invoked by submitting the macro call

```
%quadgen
```

Later, you can use the macro SURFPLOT will produce a contour plot and a 3-D plot of the response surface. Use these plots to see how well the OFAT optimization did. SURFPLOT is invoked by submitting the macro call

```
%surfplot
```

Chapter 15

Doing It with MINITAB: Chapter 15

15.1 Data Sets

Name	Description
ALUM	Aluminum sheet thicknesses, Example 15.1
NUGGETS	Chicken nugget data, Example 15.2
DSTONE	Dressing stone data, Example 15.3
BOXES	Seal strength data: stratification version, Example 15.4
BOXMIX	Seal strength data: mixing version, Example 15.4
ELECE	Prof. P.'s electric data, Example 15.5
DICE	Data on defective dice, Example 15.6
TELEM	Telemetry data, Example 15.7
WAVE	Wave solder process data, Example 15.8

15.2 Checking Process Assumptions

If doing an X chart or an $\overline{X}$ chart, you will want to check the assumptions of normality and non-correlation over time. For the X chart, you can check the raw data, but for the $\overline{X}$ chart, you will want to check the subgroup means, $\overline{X}_i$. To obtain these means, choose *Stat:Basic Statistics:Store Descriptive Statistics...* . From the resulting dialog box, choose the variable containing the data to be analyzed and the variable designating the subgroups as the *by* variable. The subgroup means will be output to the current worksheet.

Once you have readied the data to be checked, choose *Graph:Probability Plot...* for a normal quantile plot and *Stat:Time Series:Autocorrelation...* for an autocorrelation plot.

15.3 Control Charts

All control charts can be created by choosing *Stat:Control Charts:*. The chart types that are covered in the text and available in MINITAB are:

$\overline{\mathbf{X}}$, S and R Charts

$\overline{X}$ and S charts can be selected together by specifying *Stat:Control Charts:Xbar-S...* . $\overline{X}$ and R charts can be selected together by specifying *Stat:Control Charts:Xbar-R...* . These charts can be selected individually as well, by selecting *Stat:Control Charts:Xbar...* ,*Stat:Control Charts:S...* or *Stat:Control Charts:R...* .

In each case, input the the dialog box depends on how the data are arranged:

- If the data are in a single column, choose the *Single column:* button. You must specify a subgroup size (valid if all subgroups are equal in size and if the data are in order), or the name of a column identifying the subgroup of each measurement.
- If the data are in rows, one per subgroup, then select *Subgroup across rows of:*.

Various options are available to customize the analysis and display. Of most interest to us is the option of selecting some or all of the tests for special causes mentioned in the text. This option is available by clicking on the *Tests...* button.

To produce the $\overline{X}$ and S charts for Example 15.1 of the text (Figure 15.3), read the data set ALUM into the worksheet and select *Stat:Control Charts:Xbar-S...* . In the resulting dialog box, click on the button in front of *Single column:*, and supply x as the variable and t as the subgroup size.

Individual Measurement and Moving Range Charts

You can produce individual measurement and moving range charts by choosing *Stat:Control Charts:I-MR...* . The charts can also be selected individually by selecting *Stat:Control Charts:I...* or *Stat:Control Charts:MR...* .

To produce the individual measurement and moving range charts for Example 15.5 of the text (Figure 15.12), read the data set ELECE into the worksheet. Create the fourth root of KWH using *Calc:Calculator...*, and call the resulting value KWH1. Select *Stat:Control Charts:I-MR...* . In the resulting dialog box, select KWH1 as the variable, and supply 2.11 and .21 as the historical mean and sigma.

p Charts

You can produce p charts by choosing *Stat:Control Charts:P...* . The dialog box asks you to supply the name of the variable to be charted and either a subgroup size (this can only be used if all subgroups are the same size) or the name of a variable containing the subgroup sizes. If you do not supply an estimate of p, it will be computed from the data.

To produce the chart for Example 15.6 of the text (Figure 15.13), read the data set DICE into the worksheet, choose x as the variable, check the *Subgroups in:* button and supply n as the subgroup variable, and input .05 as the historical p.

c Charts

You can produce c charts by choosing *Stat:Control Charts:C...* . The dialog box asks you to supply the name of the variable to be charted.

To produce the chart for Example 15.7 of the text (Figure 15.14), read the data set TELEM into the worksheet, choose d as the variable and supply the historical mean 1.

u Charts

You can produce u charts by choosing *Stat:Control Charts:U...* . The dialog box asks you to supply the name of the variable to be charted and either a subgroup size (this can only be used if all subgroups are the same size) or the name of a variable containing the subgroup sizes. If you do not supply an estimate of u, it will be computed from the data.

To produce the chart for Example 15.8 of the text (Figure 15.15), read the data set TELEM into the worksheet, choose x as the variable, check the *Subgroups in:* button and supply n as the subgroup variable. As the example explains, the control limits are to be computed using only the first 30 data values. To do this, click on the *Estimate...* button. From the resulting dialog box, type 31:50 into the space provided below the text that begins *Omit the following samples...* .

15.4 Process Capability

To compare the quantities

$$(\text{USL} - \bar{\bar{x}})/\bar{s} \text{ and } (\text{LSL} - \bar{\bar{x}})/\bar{s}$$

with the N(0,1) density, you must compute the area under the N(0,1) density above the former and below the latter. To do this, choose *Calc:Probability Distributions:Normal...* .

To obtain the estimated capability indices $\hat{C}_p$, $\hat{C}_{pk}$ and $\hat{C}_{pm}$ using MINITAB, proceed as follows (we will follow Example 15.1, which uses the ALUM data set):

1. Choose *Stat:Quality Tools:Capability Analysis (Normal)...* .

2. To understand what happens next, you should be aware that MINITAB seems to want to perform capability analysis only for data structured for control charts. Therefore, to get a capability analysis as described in the text, we need to make some nonintuitive choices from the dialog box. Specifically, the choices made for the ALUM data are:

   - Specify that the data are arranged as the single column x with subgroup size 1.
   - Input lower and upper specification limits as .149 and .151.
   - Specify .15005 (which is $\overline{x}$) as the historical mean, and .00033 (which is s) as the historical sigma.
   - Click on the *Options...* button and input .15 as the target.

Output will include basic descriptive statistics, $\hat{C}_p$, $\hat{C}_{pk}$ and $\hat{C}_{pm}$, and other process indices not discussed in the text.

Appendix A

MINITAB Macros

Where appropriate in the first fifteen chapters of this document, we have described MINITAB macros that we have written in order to enable students to do analyses not available or easily done in MINITAB. We have also written a set of MINITAB macros specifically in support of the labs at the ends of the chapters. These macros require the full version of MINITAB, since the student version of MINITAB does not support the macro capability needed to run these macros.

Application Macros

Macro	Description
BIBOOT2	Bootstrap BC_a confidence interval for difference of proportions $p_1 - p_2$
BIEXACT	Exact confidence interval for the binomial parameter p
CEOOT1	Bootstrap BC_a confidence interval for mean in C+E model
CEOOT2	Bootstrap BC_a confidence interval for difference of means $\mu_1 - \mu_2$
CEPRED	Level L prediction interval for new observation from C+E model
CORR	Confidence interval for Pearson correlation
EIGEN	Eigen analysis of response surface
GFRAND	Generalized Friedman randomization test
GKWRAND	Generalized Kruskal-Wallis randomization test
LENTH	Computes MOE and SMOE for 2^k or 2^{k-p} design using Lenth procedure
MTRACE	Median trace
ONERAND	One sample Pitman randomization test
TQPLOT	Plot of Studentized residuals versus t quantiles
TUKEY	Tukey's 1 df for additivity test and interaction plots for RCBD model
TWORAND	Two sample Pitman randomization test
VBOOT	Bootstrap BC_a confidence interval for ratio of variances

Lab Macros

Macro	Description
LAB4_1	Macro for lab 4-1
LAB4_2	Macro for lab 4-2
LAB5_2	Macro for lab 5-2
LAB6_1E	Macro for lab 6-1
LAB6_1N	Macro for lab 6-1
LAB7_1	Macro for lab 7-1
LAB8_1	Macro for labs 8-1 and 8-2
MAKECAU	Macro for lab 4-3
MAKEDATA	Macro for lab 4-3
QUADGEN	Macro for lab 14-1
SURFPLOT	Macro for lab 14-1