Minitab® Study Card

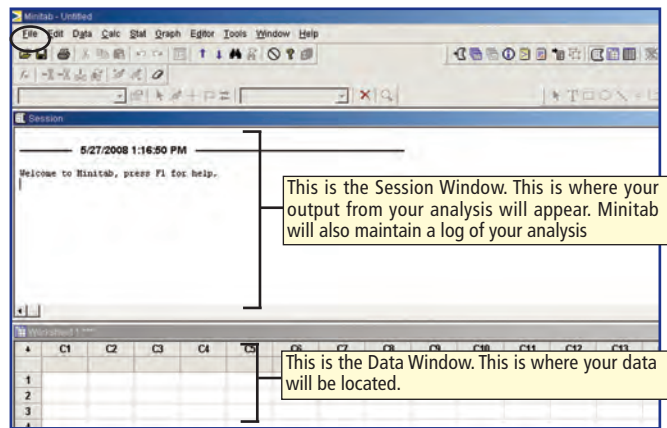JENNIFER LEWIS PRIESTLEY, PH.D.

PEARSON
Education

Introduction to Minitab

The interface for Minitab is very user-friendly, with a spreadsheet orientation. When you first launch Minitab, you will see the following window:

This is the Session Window. This is where your output from your analysis will appear. Minitab will also maintain a log of your analysis

This is the Data Window. This is where your data will be located.

To open an existing dataset select **File**, as circled in the screen to the left, and then **Open Worksheet**. This process will open the following window:
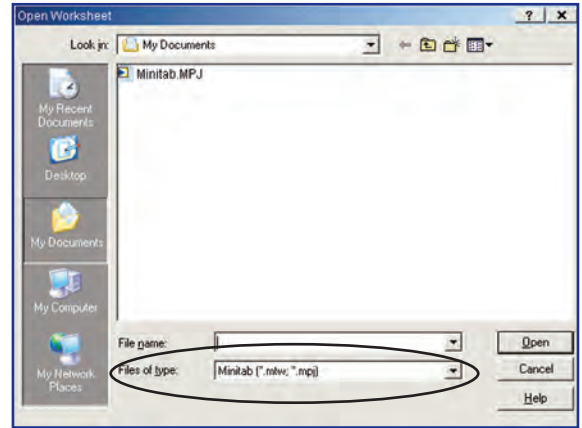
Note that there is an option to **Open Project**. A project in Minitab is a file that already includes analysis from a previous work session. Projects in Minitab have a .mpj extension. Since you will be starting a new work session, you will open a worksheet with data.

Browse to where your file is saved, and ensure that your file type is set to the required format—the **Files of type** box is circled above. Select **Open**. Note that if you open an Excel workbook with more than one sheet, that every sheet will open separately in Minitab. Just close (click on the X in the upper right corner) the sheets that you do not want to analyze.

Descriptive Statistics, Histogram, and Boxplot

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to find the complete description of a quantitative variable, including the mean, median and standard deviation, histogram and boxplot:

1. Select **Stat>Basic Statistics>Display Descriptive Statistics**. You should see the screen shown to the right:

   Notice that only the quantitative variables are listed in the column on the left.

2. Select the variables to be analyzed by clicking on them (multiple variables can be selected by holding down the **Ctrl** button) and clicking on the **Select** button (this button is grayed out until a variable has been highlighted from the list).

3. Click on the **Graphs** button. Select from the following options:
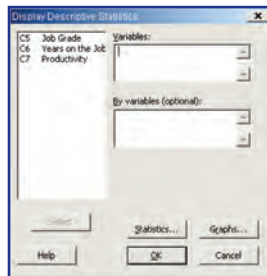
4. Select **OK**. The descriptive statistics can be found in the Session Window and the individual graphs (histogram and boxplot) can be found in separate windows (click on **Windows** at the top of the screen to see a listing of each window that was generated).

5. If you require statistics that are not included in the default output, additional statistics can be obtained by clicking on the **Statistics** button after the variable(s) has been selected.

6. To generate side-by-side boxplots, select a categorical variable in step 2 above to be placed into the **By variables** box. Continue the rest of the steps as outlined.
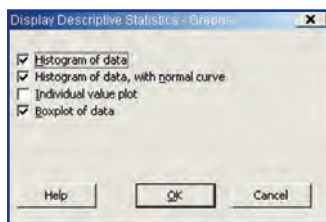
Displaying Categorical Variables (Graphs)

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to generate graphical displays of categorical variables:

1. Select **Graph>Bar Chart** (or **Pie Chart**). Click on the **Simple** option. Click **OK**.
2. From the list of variables, select the categorical variable(s) to be displayed. Click the **Select** button. Select the **Labels** button to create titles. Select **OK>OK**.
3. Typical results are shown to the right:
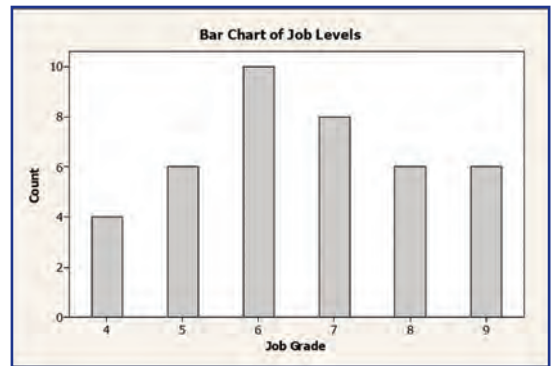
Scatterplot

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to generate a scatterplot for two quantitative variables:

1. Select **Graph>Scatterplot**. Select the first option—**Simple**—then **OK**. This will bring up the following screen:
2. Place your cursor inside the **Y variables** cell as shown to the right. Click on a quantitative variable from the list and click on the **Select** button. Next place your cursor inside the **X variables** (row 1) cell to the right. Click on a quantitative variable from the list and click on the **Select** button. Then click **OK**.
3. Once the graphic is generated, you can return to the datasheet by selecting **Window** and clicking on the datasheet name.
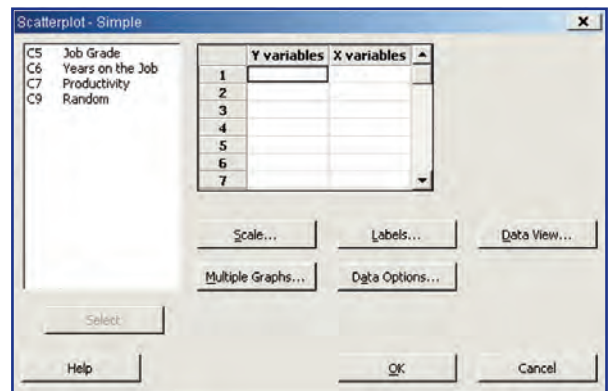
Regression and Residual Plot

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to generate a regression equation and associated plot of residuals:

1. Select **Stat>Regression>Regression**.
2. The cursor should start in the **Response** box. Click on a quantitative variable in the list that you want to assign as the dependent (response) variable and click on the **Select** button. Place your cursor inside the **Predictors** box. Click on a variable (or variables) in the list and click the **Select** button.
3. Click the **Graphs** button. Several options are presented for residuals plots. Identify if you want the Regular (unstandardized) or the Standardized residuals. Select the plots that you need.
4. A subset of typical results is shown to the right:

These results would indicate that the regression equation can explain 24.1% of the variance in the dependent variable (20.0% if the Adjusted R^2 is used). In addition, each of the predictors has a p-value less than .05, indicating that they are significant.

```
The regression equation is
Productivity = 77.8 + 1.69 Job Grade - 0.520 Years on the Job

Predictor          Coef   SE Coef     T      P
Constant         77.750     5.138  15.13  0.000
Job Grade         1.6885    0.6730   2.51  0.017
Years on the Job -0.5205    0.2449  -2.12  0.040

S = 6.49843   R-Sq = 24.1%   R-Sq(adj) = 20.0%
```

Correlation

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to generate the correlation coefficient and p-value for two quantitative variables:

1. Select **Stat>Basic Statistics>Correlation**.
2. From the variable list, select two variables (only quantitative variables are shown). Note that variables can be selected in one step by using the **Ctrl** button. Click **Select>OK**.

Generating Random Numbers

FROM AN OPEN MINITAB WORKSHEET
Use the following procedure to generate random numbers:

1. Select **Calc>Random Data**. This process will bring up a number of distribution options. The most common options include Uniform, Normal, and Binomial. For example, if you select Uniform, the screen shown to the right will be generated:

2. For the number of rows of data to generate, enter the number of rows in your dataset. Click in the **Store in column(s)** box and enter a new name like Random. Select **OK**. You should now have a new column titled Random in your dataset that contains a string of random numbers between 0 and 1.
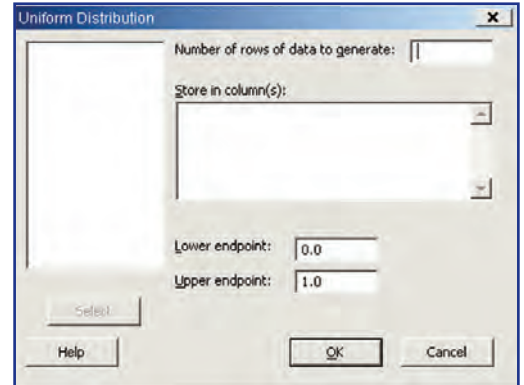
Sampling

FROM AN OPEN MINITAB WORKSHEET
Use the following procedure to create a sample from a dataset:

1. Note that this process will change your existing datasheet—save your data right now (**File>Save Current Worksheet As**).

2. Select **Calc>Random Data>Sample from Columns**.

3. In the **Number of Rows to Sample** box, enter the number of observations required for the sample.

4. Click inside the **From columns** box. This will bring up the list of variables on the left. Click on the first variable, hold the Shift key down and click on the last variable. Click **Select**. Place your cursor in the next box and do the same thing. The dataset now has a random sample of observations. Save this file with a new name—such as Sample.

Displaying Categorical Data (Frequency Tables)

FROM AN OPEN MINITAB WORKSHEET
Use the following procedure to generate a frequency table:

1. Select **Stat>Tables>Tally Individual Variables**.

2. From the list of variables, click on a categorical variable and click **Select**.

3. Under **Display**, select all of the elements required. Click **OK**.

Confidence Intervals and Hypothesis Tests for Proportion

FROM AN OPEN MINITAB WORKSHEET
Use the following procedure to generate a confidence interval for a proportion:

1. Select **Stat>Basic Statistics>1 Proportion**.

2. Click inside the **Samples in columns** box. This will bring up the variable list on the left. Click on the categorical variable of interest (must have two values) and click the **Select** button.

3. Click on the **Options** box and enter the confidence level required (the default is 95%). Click **OK**.

4. To execute a hypothesis test, select the perform hypothesis test option and enter the proportion value to be tested against. (Note that the category that occurs first alphabetically will be tested.)

5. Click **OK**.

6. This process will execute a two-tailed test. If a one tailed test is required, click on the **Options** button and change the **Alternative-test** option from Not Equal to Greater Than or Less Than.
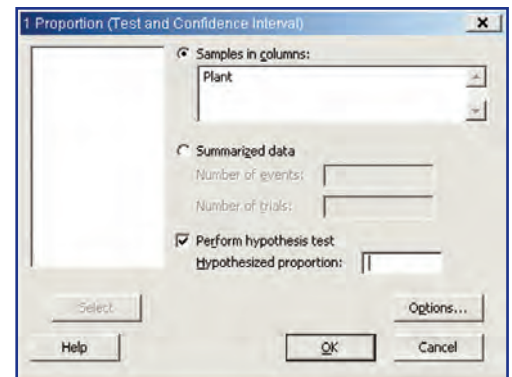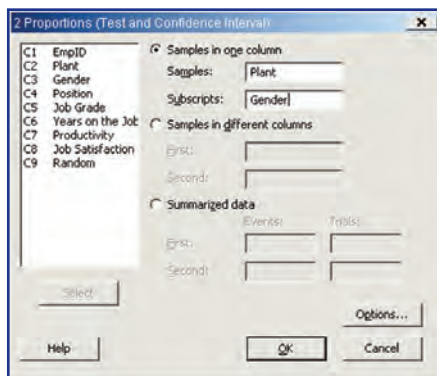
Minitab Study Card

Inference for the Difference between Proportions

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to generate a confidence interval for the difference between two proportions:

1. Select **Stat>Basic Statistics>2 Proportions**.

2. Click inside the **Samples** box. Click on the primary categorical variable (must have two values) and click **Select**. Click inside the **Subscripts** box. Click on the secondary categorical variable (must have 2 values) and click **Select**.

3. Select the **Options** button. Enter the required confidence level, the test difference, the alternative hypothesis (not equal, greater than or less than), and if you are using a pooled estimate. Click **OK>OK**. Typical results are shown below:

```
Test and CI for Two Proportions: Plant, Gender
Event = Norcross
Gender   X    N   Sample p
Female   7    20   0.350000
Male    10    20   0.500000
Difference = p (Female) - p (Male)
Estimate for difference:  -0.15
95% CI for difference:  (-0.452845, 0.152845)
Test for difference = 0 (vs not = 0):  Z = -0.96  P-Value = 0.337
Fisher's exact test: P-Value = 0.523
```

> These results would indicate that the difference in the proportions is not statistically different from 0.

Hypothesis Tests for the Mean of Paired Differences

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to determine if a statistical difference exists between two paired samples:

1. Select **Stat>Basic Statistics>Paired t**.

2. Under the **Samples in columns**, click inside the **First Sample** box. From the list of variables on the left, click on the first of the two paired variables (such as a prescore). Select **OK**. Click in the **Second Sample** box. From the list of variables on the left, click on the second of the two paired variables (such as a postscore). Click on the **Options** button to assign the confidence level, the difference to be tested, and the alternative test (not equal, less than, or greater than).

3. Click **OK>OK**. Typical results are shown below:

```
Paired T-Test and CI: PRE, POST

Paired T for PRE - POST
             N    Mean   StDev   SE Mean
PRE         47   26.32   31.44    4.59
POST        47   14.03   24.27    3.54
Difference  47   12.29   30.50    4.45

95% CI for mean difference: (3.33, 21.25)
T-Test of mean difference = 0 (vs not = 0): T-Value = 2.76  P-Value = 0.008
```

> These results would indicate that the difference in the paired samples is statistically significant.

Hypothesis Tests and Confidence Intervals for Means of Two Independent Samples

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to determine if a statistical difference exists between two independent sample means:

1. Select **Stat>Basic Statistics>2 Sample t**.

2. Under the **Samples in columns**, click inside the **Sample** box. From the list of variables on the left, click on the quantitative variable of interest. Click **Select**. Click inside the **Subscripts** box. From the list of variables on the left, click on the binary categorical variable of interest (this variable might be treatment A and treatment B). Click **Select**. Click on the **Options** button to assign the confidence level, the difference to be tested, and the alternative test (not equal, less than, or greater than).

3. Click **OK>OK**. Typical results are shown below:

```
Two-Sample T-Test and CI: PRE, Treatment
Two-sample T for PRE

Treatment   N   Mean   StDev   SE Mean
A          30   25.1    29.8    5.4
B          31   22.4    31.1    5.6
Difference = mu (A) - mu (B)
Estimate for difference:  2.65
95% CI for difference:  (-12.97, 18.27)
T-Test of difference = 0 (vs not =): T-Value = 0.34  P-Value = 0.736  DF = 58
```

> These results would indicate that the difference between the means of the two groups is not significant.

Hypothesis Test and Confidence Interval for One Sample Mean

FROM AN OPEN MINITAB WORKSHEET. Use the following procedure to determine if a statistical difference exists between a single sample mean and a specified value and to generate the confidence interval:

1. Select **Stat>Basic Statistics>1 Sample t**.

2. Under **Samples in columns**, click inside the **Samples** box. From the list of variables on the left, click on the quantitative variable of interest. Click **Select**. Check the **Perform Hypothesis Test** box and enter the hypothesized value to be tested against. Click on the **Options** button to assign the confidence level, the difference to be tested, and the alternative test (not equal, less than, or greater than).

3. Click **OK>OK**.

4. Typical results are shown below:

```
One-Sample T: Productivity

Test of mu = 82 vs not = 82


Variable        N    Mean  StDev  SE Mean        95% CI          T      P
Productivity   40   84.58   7.26     1.15  (82.26, 86.90)    2.25  0.030
```

Finding Inverse Normality (finding *z*-scores given an area) and the Area Under the Normal Curve

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to convert quantitative values into *z*-scores and their associated probability of occurrence given a normal distribution:

1. Select **Calc>Standardize**. This will generate the screen below:
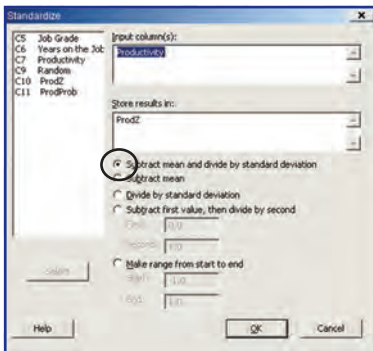
2. Click inside the **Input column** box and select a quantitative variable From the list on the left. Click **Select**. Click inside the **Store results in** box and type the name of a new variable where the *z*-scores can be stored.

3. Ensure that the first calculation option is selected, as circled above—this is the formula for a *z*-score calculation. Click **OK**.

4. You should see a column of *z*-scores on the right of your dataset.

5. To determine the associated (cumulative) probabilities with these *z*-scores, under the normal curve, select **Calc>Probability Distributions>Normal**. This will bring up the following screen:
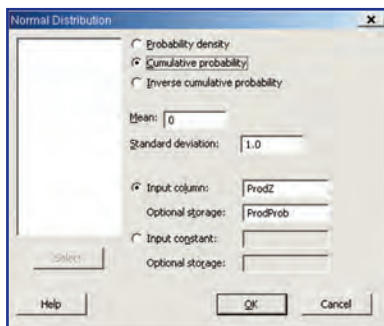
6. Click the second option—**Cumulative probability**. The mean should be set to 0 and the standard deviation should be set to 1.0. For the Input column, select the new column that was just created with the *z*-scores. For the **Optional storage**, type a new variable name—this is where the probability values or the area values will be stored. Click **OK**.

7. An example of the *z*-scores and area values is shown below:

ProdProb is the column of cumulative normal probabilities associated with productivity.

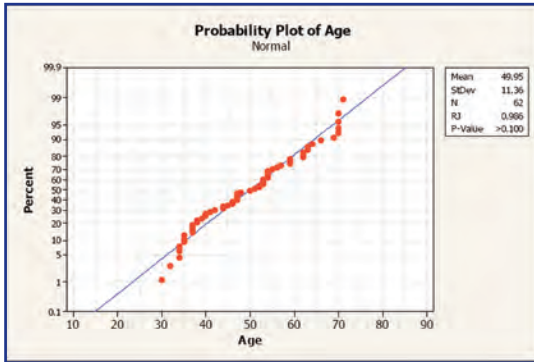**ProdZ** is the column of *z*-scores associated with productivity.

**Productivity** is the original quantitative variable.

Assessing Normality and Chi-Square Goodness of Fit

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to assess the normality (or not) of a quantitative variable of interest:

1. Select **Stat>Basic Statistics>Normality Test**.
2. Click inside the **Variable** box and select a quantitative variable from the list on the left. Click **Select**.
3. Select the normality test of interest (the Ryan Joiner—similar to the Shapiro Wilk test—is the preferred test for datasets with fewer than 2000 observations). Click **OK**.
4. Typical results are shown below:

The 45-degree line indicates perfect normality. As the plot of the actual values diverges from the line, the distribution of the data is less normal. The result of the Ryan Joiner test—indicated in the output as RJ—and the associated *p*-value would indicate that the distribution of the data is normal. In this test, the null hypothesis is normality.

5. To generate additional information on the normality of a variable, including percentiles, select **Stat>Basic Statistics> Graphical Summary**. From the list of variables on the left, select the variable of interest. Select **OK**. Typical results are shown below:
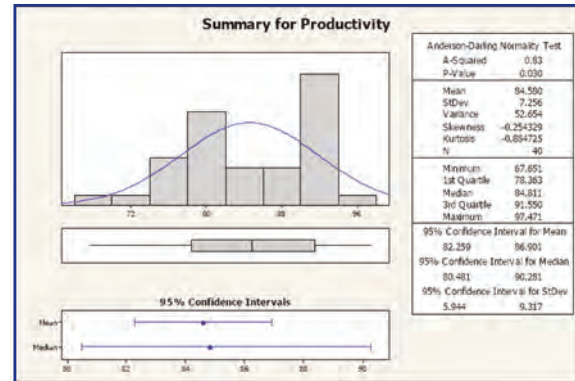
Note that in this output, the normality test is the Anderson Darling—the test is slightly different from the Ryan Jones test above, but the results (and corresponding p-value) are interpreted the same way.

One-Way ANOVA

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to execute a one-way ANOVA:

1. Select **Stat>ANOVA>One Way**.
2. Click inside the **Response** box. From the list on the left, click on the quantitative variable of interest. Click **Select**. Click inside the **Factor** box. From the list on the left, click on the categorical variable with three or more values of interest. Click **Select**. Click **OK**.
3. Typical results are shown below:

These results would indicate that the differences among the means are not significant.

4. To determine which groupings are significantly different, execute a Post Hoc test (such as Tukey's Post Hoc Test). This can be done by clicking on the **Comparisons** button from step 2 above.

Chi-Square Test of Independence (or Homogeneity)

FROM AN OPEN MINITAB WORKSHEET   Use the following procedure to execute a chi-squared test of independence (or homogeneity).

1. Select **Stat>Tables>Cross Tabulation and Chi Square**.
2. Click inside the **Rows** box and select a categorical variable from the list on the left. Click on the **Columns** box and select a categorical variable from the list on the left.
3. Click on the **Chi Square** button and check the first box—**Chi Square Analysis**. Click **OK**.
4. Typical results are shown below:

These results would indicate that the differences among the frequency counts are not significant.

5. Note that a chi-squared test of independence is used to determine if there is an association between a row and a column variable from a single population. The chi-squared test for homogeneity of proportions is used to determine if different populations have the same proportion of individuals with some characteristic. The execution in Minitab is the same, but the interpretation of the results is different.