STATISTIX 8 Tutorial

This tutorial shows you some of the most important and frequently used statistical procedures that can be accomplished using the Student Version of STATISTIX 8. The best way to learn how to use STATISTIX is to use the software, so this tutorial will employ a learn-by-doing approach. First, you will see how to load STATISTIX 8 onto your computer from the CD that came with your textbook. Then, the tutorial guides you step-by-step through several of the statistical examples in the textbook. Keep in mind that STATISTIX 8 has an excellent HELP menu on the main toolbar. This tutorial is no substitute for using the HELP menu provided in the software.

1.   LOADING THE SOFTWARE FOR THE FIRST TIME

The first time you use STATISTIX on your computer, you will need to load the software onto your computer's hard drive — usually this drive is called the "C:" drive. After the software is loaded for the first time, it is permanently available thereafter. The first step in loading the software is to close all your Windows applications that may be running (e.g., close Word, Excel, Internet Explorer, and so on). Next, insert into your computer's CD-ROM drive the STATISTIX 8 CD, which is the blue CD that came with your textbook. After a few seconds, instructions for installation will appear on your screen. Follow the simple installation instructions. When finished, remove the CD and begin using STATISTIX. To start STATISTIX in the future, you go to the START menu. Then, under the PROGRAMS menu, click on STATISTIX and then again on STATISTIX 8. [Alternatively, a shorter procedure can be followed if you place a "shortcut" to STATISTIX 8 on your Windows desktop. See the Help menu in Windows to learn how to add shortcuts.]

2.   SIMPLE LINEAR REGRESSION

To illustrate how to perform simple linear regression with STATISTIX 8, we will replicate the results presented in Chapter 4 for the advertising and sales relation facing travel agents. The data are shown in Table 4.2 on page 125 and the regression results are shown in Table 4.3 on page 129.

The first step is to enter the data on advertising expenditure (A) and sales (S) shown in Table 4.2 (page 125). After starting the STATISTIX program, you will see the opening screen, which is a blank spreadsheet. This is where you will enter the data to be analyzed. There are several ways to enter data in STATISTIX; you can enter data by hand, or open preexisting STATISTIX data sets, or import data from sources such as EXCEL spreadsheets. The HELP menu in STATISTIX gives details on how to import data from external sources such as EXCEL or Lotus 1-2-3, etc. In this example, we will enter the data rather than import it from a previously created STATISTIX dataset or an EXCEL spreadsheet.

Begin entering the data in Table 4.2 by moving through the toolbar menus: DATA ➤ INSERT ➤ VARIABLES. In the "New Variable Names" box, type the names of the variables with either a comma or a blank separating the variable names. In this case, type "S,A" or "S A" and then click on OK. Notice that you now have two columns in the data matrix, and the columns are labeled "S" and "A" in the order that you named them.

Now enter Firm A's sales of $15,000 by typing "15000" (do not type commas or dollar signs). Press the TAB key to move to Firm A's advertising level and type $2,000 as "2000." Press ENTER to move to the next observation, which is Firm B in Table 4.2. Pressing ENTER causes a second row with two columns to appear so that you can enter values of S and A for Firm B. Continue entering data until the data for all 7 firms have been entered. You will have a data matrix that looks like the following:

		S	A
	1	15000	2000
	2	30000	2000
	3	30000	5000
	4	25000	3000
	5	55000	9000
	6	45000	8000
▶	7	60000	7000
✳			

For two reasons, it is now a good idea to save the data you have entered. First, a dataset that is saved can be retrieved in the future making it unnecessary to retype the data for later use. Second, the saved dataset is in a special format used by STATISTIX (such saved files are denoted by the suffix, ".sx"). A previously created and saved STATISTIX dataset can be reopened simply by clicking on its name in Windows Explorer. This is a very handy feature. To save the data, hold the CTRL key down and simultaneously press S. If this is a new dataset, you will be asked to provide a name for the dataset. Choose an appropriate name and click the ENTER button. Now we are ready to "run the regression."

To estimate a linear regression model — i.e., to estimate the values of the parameters of a linear equation — follow these steps:

1. Make sure you have a dataset open and showing on the screen.

2. Move through the menu of the toolbar as follows: STATISTICS ▶ LINEAR MODELS ▶ LINEAR REGRESSION.

3. In the "Linear Regression" box, choose *S* to be the dependent variable by highlighting *S* and pressing the ▶ button to move *S* into the "Dependent Variable" box. Note: Only one variable can be moved to this box.

4. Next move A into the box named "Independent Variables."

5. Be sure the box named "Fit Constant" has a check mark in it. Note: "Fit Constant" tells the computer to estimate the intercept (or constant) parameter.

6. Now click OK to get the regression output to pop onto your computer screen.

7. You can print the output by clicking on the printer icon on the top tool bar or you can save it: FILE ▶ SAVE AS.

Your regression output should look like:

```
STUDENT EDITION OF STATISTIX 8.0        PRACTICE DATA, 10/15/2004, 10:22:07 PM

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF S

PREDICTOR
VARIABLES     COEFFICIENT     STD ERROR      STUDENT'S T        P
---------     -----------     ---------      -----------      ------
CONSTANT        11573.0        7150.83          1.62          0.1665
A               4.97191        1.23154          4.04          0.0100

R-SQUARED              0.7652      RESID. MEAN SQUARE (MSE)   7.713E+07
ADJUSTED R-SQUARED     0.7183         STANDARD DEVIATION        8782.64

SOURCE          DF       SS            MS            F        P
---------       ---    ----------    ----------    -----    ------
REGRESSION       1     1.257E+09     1.257E+09     16.30    0.0100
RESIDUAL         5     3.857E+08     7.713E+07
TOTAL            6     1.643E+09

CASES INCLUDED 7   MISSING CASES 0
```

This printout is also shown in Panel B of Figure 4.3 on page 129 of your textbook.

3.   READING STATISTIX 8 OUTPUT

The printout of regression results from STATISTIX looks a little bit different than the computer printouts shown in the textbook. To see the difference, the printout from Illustration 4.3 on page 145 is compared to the corresponding STATISTIX 8 output you would get if you did the regression analysis presented in Illustration 4.3. The diagrams below show the relationship between the computer printouts in your textbook and the computer printouts from STATISTIX 8.

Textbook regression output from page 145:

```
STATISTIX 7.0                    BODILY INJURY CLAIMS..., 09/07/2002, 12:00:05 PM

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF (P)  A

PREDICTOR
VARIABLES      COEFFICIENT      STD ERROR     STUDENT'S T       P        VIF
---------      -----------      ---------     -----------     ------     -----
CONSTANT        -74.1352  F      34.6122  G     -2.14  H      0.0470  I
N                11.3203         0.95258        11.88         0.0000     1.1
C                0.01155         0.00403         2.87         0.0107     1.1

R-SQUARED           0.9116  C    RESID. MEAN SQUARE (MSE)    176.815
ADJUSTED R-SQUARED  0.9012       STANDARD DEVIATION           13.2972

SOURCE        DF        SS            MS           F          P
----------    ---    ----------    ----------    ------    ------
REGRESSION     2       30997.9       15499.0     87.66  D   0.0000  E
RESIDUAL      17       3005.86        176.815
TOTAL         19       34003.8
                                                    B
CASES INCLUDED (20)    MISSING CASES 0
```

Notice that the STATISTIX regression output provides more statistical information than is used in the textbook. When you take a course in statistics, you will learn how to use the rest of the information in the STATISTIX printout. For the purposes of learning managerial economics, this extra information is not essential.

4.   CREATING A SCATTER PLOT OF DATA

You may wish to create a scatter plot of data points such as the one shown in Figure 4.2 on page 126 of the text. To create a scatter plot of the sales (S) and advertising (A) data using STATISTIX, proceed according to the following steps:

1.   Open the STATISTIX data file (on the CD that came with your text) named Travel Agency Sales and Advertising.sx.

2.   Move through the menu of the toolbar as follows: STATISTICS ▸ SUMMARY STATISTICS ▸ SCATTER PLOT.

3.   For the "X variable" select A. For the "Y variable" select S.

4.   You could click on OK now and get a scatter plot. However, you may wish to impose some restrictions on how the vertical and horizontal axes are rendered in the plot. In this example, the Figure 4.2 can be closely approximated by setting the following options before clicking on the OK button:

Set X-axis settings: Low = 0   High = 10000   Step = 1000

Set Y-axis settings: Low = 0   High = 70000   Step = 10000

Click the empty box "Display Regression Line" to get the regression line to print over the data.

5.   Now click OK.

The scatter plot resulting from the above steps looks like:

Scatter Plot of S vs A
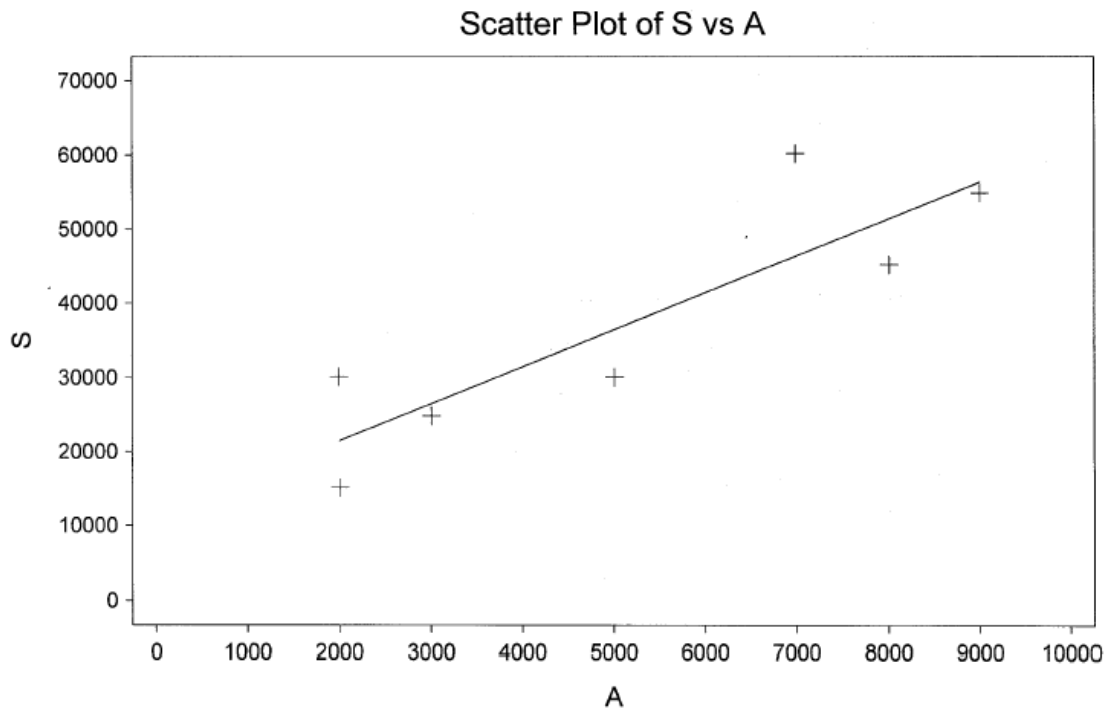
5. QUADRATIC REGRESSION MODELS

Quadratic regression, as discussed in the textbook beginning on page 143, is accomplished in STATISTIX using the Linear Regression procedure. There is NO "Quadratic Regression" procedure in STATISTIX (or any other regression package for that matter).

To illustrate how to perform quadratic regression with STATISTIX 8, we will explain the steps employed in Figure 4.5 on page 146 of your textbook. Proceed according to the following steps:

1. Following the procedure explained in Section 2 of this tutorial, enter the data for Y and X shown below:

Y	X
83	3
107	3
61	4
76	5
68	6
30	8
57	10
40	12
81	14
68	15
102	17
110	18

2. Notice that the data set does not include the variable X^2. This variable must be created. Following the explanation in the textbook, let's call this variable Z, and define Z to be equal

to X^2. The new variable Z is created in STATISTIX by holding down the CTRL key while pressing T (i.e., CTRL-T). This brings up the "Transformations" dialog box.

3.   In the Transformations dialog box, click your cursor into the "Transformation Expression" box. Then type in the new variable and its definition: $Z = X*X$ or $Z = X\hat{\ }2$; either one will work.

4.   Click GO and you will see the "Transformation Complete" notice. Now the dataset has a new variable Z, which is X^2 in this example.

5.   Click on CLOSE to leave the Transformations dialog box and return to the data.

6.   Move through the menu on the toolbar as follows: STATISTICS ▸ LINEAR MODELS ▸ LINEAR REGRESSION.

7.   Run a linear regression just as explained in Section 2 of this tutorial. Treat Y as the dependent variable and use BOTH X and Z as the Independent Variables in the regression dialog box. (Also be sure a check mark shows in the box "Fit Constant." You will get the following regression output:

```
STATISTIX 8.0                                    FIGURE 5, 10/07/2004, 12:12:31
PM

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

PREDICTOR
VARIABLES      COEFFICIENT    STD ERROR      STUDENT'S T        P        VIF
---------      -----------    ---------      -----------     ------     -----
CONSTANT          140.082       16.8043          8.34        0.0000
X                -19.5072       4.05454         -4.81        0.0010      29.1
Z                 1.00581       0.19647          5.12        0.0006      29.1

R-SQUARED             0.7542      RESID. MEAN SQUARE (MSE)     189.107
ADJUSTED R-SQUARED    0.6995        STANDARD DEVIATION          13.7516

SOURCE        DF       SS          MS          F        P
----------    ---    ----------   ----------   -----    ------
REGRESSION     2      5220.95      2610.47      13.80    0.0018
RESIDUAL       9      1701.97      189.107
TOTAL         11      6922.92

CASES INCLUDED 12    MISSING CASES 0
```

6.   LOG-LINEAR REGRESSION MODELS

Log-linear regression using STATISTIX is illustrated here using the example in the textbook in Figure 4.6 on page 149 of the textbook. Proceed according to the following steps:

1.   Open the STATISTIX data file (on the CD that came with your text) named Log Linear Regression.sx. These are the data for Y and X shown in Panel A of Figure 4.6.

2.   Create two new variables: the natural logarithm of Y (call it LNY) and the natural logarithm of X (call it LNX). These are the data shown in Panel B of Figure 4.6. Hit CTRL-T to enter the Transformation dialog box. In the Transformations dialog box, click your cursor into the "Transformation Expression" box. Then type: LNY = ln(Y). Click on GO. Repeat for LNX: LNX = ln(X). Notice that you must use "ln( )". You will get base 10 logarithms if you use "log( )". This is a subtle, but important, distinction. You must use base e logarithms (known as "natural" logarithms), and this is accomplished in STATISTIX by using the ln( ) transformation. Check your LNY and LNX values to make sure they match the values shown in Panel B on page 149 of your text.

3.   Move through the menu on the toolbar as follows: STATISTICS ▸ LINEAR MODELS ▸ LINEAR REGRESSION. Be careful here: Do NOT go to LOGISTIC REGRESSION.

Log-linear regression is a LINEAR regression procedure.

4. Run a linear regression as explained in Section 2 of this tutorial. Treat LNY as the dependent variable and LNX as the Independent Variables in the regression dialog box. (Also be sure a check mark shows in the box "Fit Constant.") You will get the following regression output:

```
STATISTIX 8.0                    LOG LINEAR REGRESSION, 09/07/2004, 2:38:23 PM

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF LNY

PREDICTOR
VARIABLES     COEFFICIENT    STD ERROR     STUDENT'S T        P
---------     -----------    ---------     -----------     ------
CONSTANT        11.0642       0.47895         23.10         0.0000
LNX            -0.96221       0.11519         -8.35         0.0000

R-SQUARED            0.8746     RESID. MEAN SQUARE (MSE)    0.07091
ADJUSTED R-SQUARED   0.8621     STANDARD DEVIATION          0.26628

SOURCE        DF       SS           MS          F        P
----------    ---   ----------   ----------   -----   ------
REGRESSION     1     4.94732      4.94732     69.77   0.0000
RESIDUAL      10     0.70906      0.07091
TOTAL         11     5.65638

CASES INCLUDED 12   MISSING CASES 0
```

7. TWO-STAGE LEAST-SQUARES (2SLS)

Two-stage least-Squares regression (2SLS) using STATISTIX is illustrated here using the world copper market example in the Chapter 7 of the textbook. In order to estimate the world DEMAND for copper, proceed according to the following steps:

1. Open the STATISTIX data file (on the CD that came with your text) named World Copper Market.sx. These are the data shown in Table A on page 306 of your textbook.

2. Move through the menu on the toolbar as follows: STATISTICS ▸ LINEAR MODELS ▸ TWO STAGE LEAST SQUARES.

3. In the 2SLS dialog box, first enter the dependent variable, which is QC in this example.

4. Next enter the RIGHT HAND VARIABLES. The right hand variables are the independent or explanatory variables in the DEMAND equation. In this case, the explanatory variables for copper demand are PC, M_, and PA. (NOTE: STATISTIX reserves "M" for its own use so the income variable was named "M_" to avoid a conflict with the software.)

5. Finally, enter the list of exogenous variables in the entire system of demand and supply equations. In other words, list ALL of the exogenous variables in BOTH demand and supply, even though you are only estimating demand with this one regression. (NOTE: When you estimate supply, you must run the 2SLS regression again changing ONLY the list of RIGHT HAND VARIABLES.) In the dialog box for EXOGENOUS VARIABLES, enter X, M_, PA, T. (Also be sure a check mark shows in the box "Fit Constant.") Now click OK.

6. You will get the following regression output:

TWO STAGE LEAST SQUARES REGRESSION OF QC
4 EXOGENOUS VARIABLES: X, M_, PA, T

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	-6837.83	1264.46	-5.41	0.0000
PC	-66.4950	31.5338	-2.11	0.0472
M_	13997.7	1306.34	10.72	0.0000
PA	107.662	44.5098	2.42	0.0247

R-SQUARED	0.9421	RESID. MEAN SQUARE (MSE)	184327
ADJUSTED R-SQUARED	0.9339	STANDARD DEVIATION	429.333

CASES INCLUDED 25   MISSING CASES 0