CHAPTER 18

Visualizing Regression Models

CONCEPTS

- Transformations, Linear Model, Elasticity Model, Exponential Growth Model, Declining Returns Model, Power Model, Beta Weight, Polynomial Model, Interaction Variable

OBJECTIVES

- Know the types of commonly-used variable transformations and their purposes

- Learn the effects of variable transformations on the fitted regression and its statistics of fit

- Recognize and interpret models that utilize log, reciprocal, standardized, or power transformations of X and/or Y

- Understand how to use polynomial models

- Be able to use and interpret an interaction variable

Overview of Concepts

The familiar **linear model** has the form $Y_i = \beta_0 + \beta_1 X_i + u_i$. β_0 and β_1 can be estimated using the method of Ordinary Least Squares (OLS) even if we perform **transformations** of the data. Common transformations of the variables are the logarithmic transformation [$\ln(X_i)$ or $\ln(Y_i)$], raising to a power [X_i^c or Y_i^c], inversion [$1/X_i$ or $1/Y_i$], or standardization by subtracting the mean and dividing by the standard deviation.

OLS can also be used for some alternative model forms. One example is the family of **polynomial models**. Although they can capture nonlinear patterns and improve the R^2, polynomial models are usually difficult to interpret. In this module, we limit the choices to the quadratic model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X^2_i$ and cubic model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X^2_i + \beta_3 X^3_i$ because they are familiar and easily graphed. In both cases, the sign of the slope coefficients affect the specific shape of the polynomial. Of course, the linear model is a special case of the polynomial model family.

In the **elasticity model**, $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i)$, a 1% change in X_i leads to a β_1% change in Y. For instance, if the fitted model is $\ln(Y_i) = 2.055 – 0.643 \ln(X_i)$ then a 1% change in X_i leads to a –0.643% change in Y. When $\beta_1 > 1$, the function increases at an increasing rate; when $0 < \beta_1 < 1$, it increases at a decreasing rate; and when $\beta_1 < 0$, it is asymptotic to $Y = 0$ from above. In the **exponential growth model**, $\ln(Y_i) = \beta_0 + \beta_1 X_i$, a 1-unit change in X_i causes Y_i to change by $100\beta_1$% . For example, if $\ln(Y_i) = 200 + .075 X_i$, then a 1-unit increase in X would cause Y to increase by 7.5%. When $\beta_1 > 0$, the function increases at an increasing rate; and when $\beta_1 < 0$, it decreases at an decreasing rate. In the **declining returns model,** $Y_i = \beta_0 + \beta_1 \ln(X_i)$, a 1% change in X_i causes Y_i to change by $\beta_1/100$. For example, if $Y_i = 200 + 57.5 \ln(X_i)$ then a 1% increase in X would increase Y by 0.575 units. This function either increases at a decreasing rate ($\beta_1 > 0$) or decreases at a decreasing rate ($\beta_1 < 0$). These models are useful when there is a theoretical reason to expect a nonlinear relationship (such as diminishing returns to scale in the theory of production) or when the data demands a nonlinear form.

Standardized variables are denoted $\text{Std}(Y) = (y_i - \bar{y}) / s_y$ and $\text{Std}(X) = (x_i - \bar{x}) / s_x$. The model $\text{Std}(Y_i) = \beta_0 + \beta_1 \text{Std}(X_i)$ yields the same R^2 as the linear model, since standardizing is a linear transformation of X and Y. Each transformed variable is expressed in standard deviations, and the coefficient β_1 is called a **beta weight**. This model is of interest in fields such as finance, human resources, and psychology. It eliminates units of measurement, and avoids issues of scaling in variables with dissimilar magnitudes. In the model, a 1-standard deviation change in X results in a β_1 standard deviation change in Y. The estimated intercept β_0 is always zero and hence, can be omitted from the model.

A **power model** has the form $Y_i = \beta_0 + \beta_1 X_i^c + u_i$ where c is a constant specified *a priori*. Special cases are c = 2 (quadratic model without the X term), c = 1 (the linear model), c = –1 (the *reciprocal model*), and c = 0.5 (a type of diminishing growth model). Alternatively, if Y_i is raised to the power c, the power model has the form $Y_i^c = \beta_0 + \beta_1 X_i + u_i$. In both cases if c is not known beforehand, a search procedure can be used.

An **interaction variable** has the form X_iZ_i (i.e., X_i is multiplied by Z_i), where Z_i is a variable believed to interact with X_i. The model becomes $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_iZ_i$. One case of interest is when Z_i is a variable that has only two values, but this is not a necessary restriction. Z_i can be multi-valued (e.g., 1, 2, 3) or even continuous. The effect of Z_i is to alter the slope of the regression line (in effect, a different fitted function for each value of Z_i).

Illustration of Concepts

For a sample of 52 architectural firms that specialize in health care, an industry analysis shows the number of architects employed by each firm (X) and the firm's total fees for the year (Y) in millions of dollars. The **linear model** (Figure 1) yields $R^2 = 0.78$. The quadratic model (Figure 2) improves the fit slightly to $R^2 = 0.80$. The cubic model (Figure 3) yields $R^2 = 0.81$ (but note that the coefficient of the X^3 term is nearly zero). Adding an additional variable to the model always improves R^2, even if it adds nothing to the model's explanatory power. The **declining returns model** (Figure 4) requires a Ln **transformation** of the independent variable. It yields $R^2 = 0.67$. In this model, a 1% increase in the number of architects yields an extra 0.083711 (= 8.3711/100) in fees or \$83,711 since the units are in thousands. In contrast, in the linear model each additional architect yielded an extra 0.1755 in fees (i.e., \$175,500). There is no straightforward interpretation of either **polynomial model**. However, all three nonlinear models suggest that total fees increase at a declining rate relative to the number of architects. This is theoretically appealing since it is unlikely that the 200th architect will have the same impact on total fees as the 5th architect.

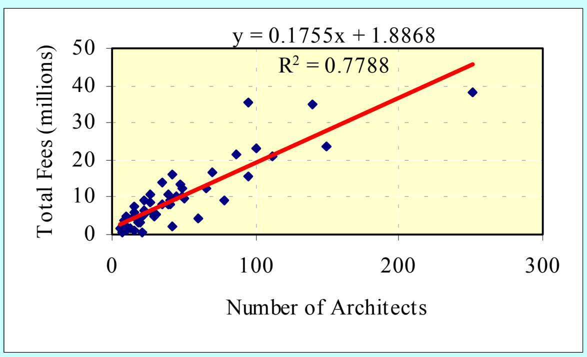


Figure 1: Linear Model

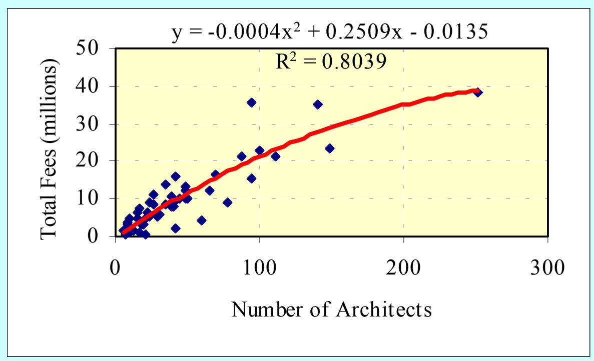


Figure 2: Quadratic Model

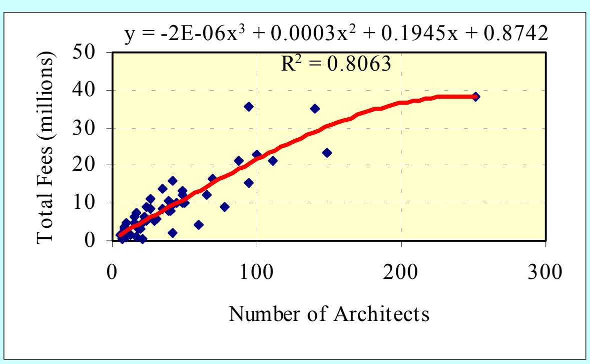


Figure 3: Cubic Model

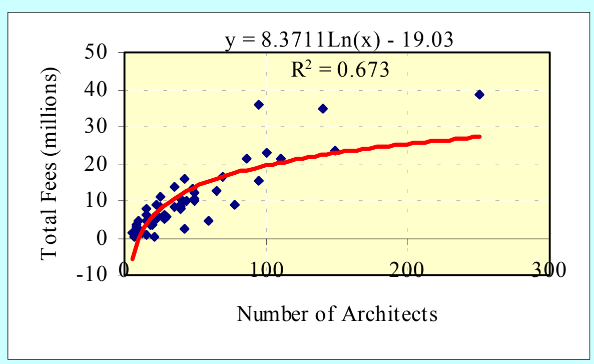


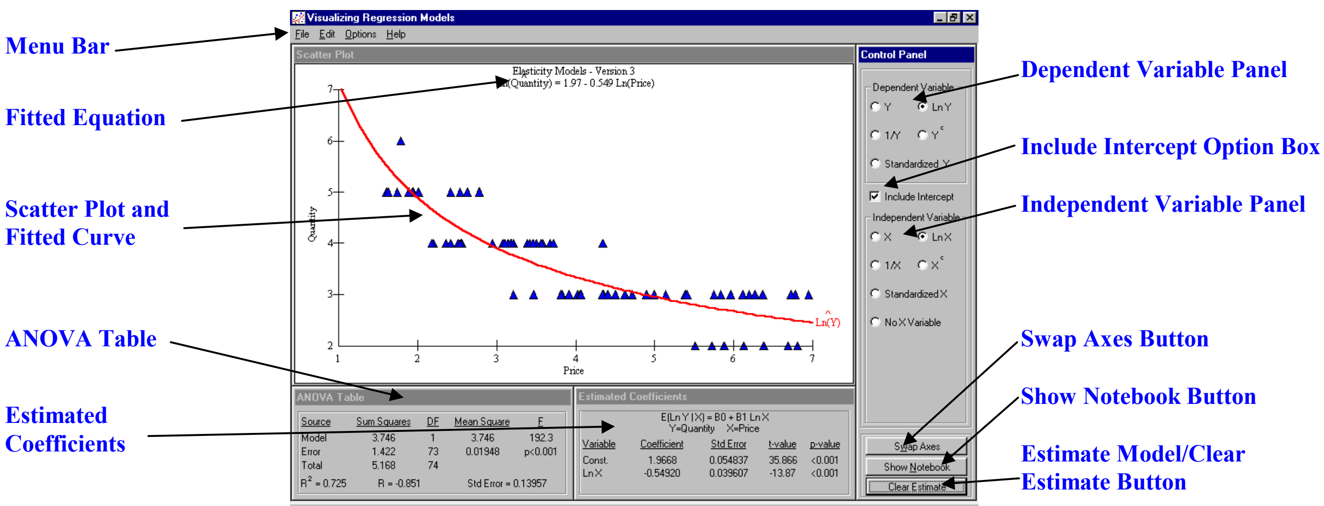
Figure 4: Declining Growth Model

Other models are possible. An **elasticity model** tells us the percentage change in total fees due to a 1% change in the number of architects. A *standardized model* gives the same fit as a linear model, but its **beta weight** tells us the standard deviation change in total fees due to a 1-standard deviation change in the number of architects. An **exponential growth model** does not make sense, given the shape of the data. The **power model** requires *a priori* specification of an exponent, for which we lack any real basis. An **interaction variable** might be appropriate, perhaps by using the number of engineers employed by the firm (Z) since engineers and architects are likely to be complementary in the design of health care facilities.

Orientation to Basic Features

This module estimates a model based on the functional form and variable transformations that you specify. You can examine the shape of the fitted function that is displayed on a scatter plot, or study its estimated coefficients, ANOVA statistics, and examine confidence intervals for E(y|x) and prediction intervals for y|x. Data is brought into the module from the notebook.

1. **Opening Screen**
   Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. Read the questions and then click the Concepts tab to see the concepts that you will learn. Click on the Scenarios tab. Click on Growth Models, select the Elasticity Models scenario, and read it. Enter a number in the version window at the bottom of the notebook page. The version number allows you to duplicate this same scenario later. The default (0) gives a randomly chosen scenario between 1 and 99. Version 3 is shown below. When you have selected a version, press OK. A scatter plot appears with X and Y control panels on the right. Click the Estimate Model button and you will see a screen similar to the one shown here with the fitted model displayed in red on the scatter plot. After the model has been fitted, the Estimate Model button becomes a Clear Estimates button.



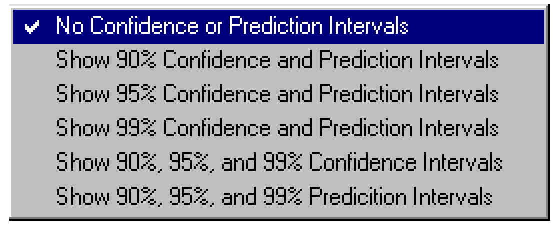
2. **ANOVA Table and Estimated Coefficients**
   Examine the ANOVA table and estimated coefficients at the bottom of the screen. Except for rounding, the coefficients are the same as the fitted equation shown at the top of the scatter plot. Each scenario will generate a different fitted equation. Click the Swap Axes button to interchange the independent and dependent variables.

3. **Control Panel for X and Y**
   The Ln X and Ln Y options are selected in the Control Panel. Select X and Y, then press the Estimate Model button. Note the effects on the fitted equation and displayed curve. Repeat, selecting the 1/X and 1/Y options. Restore the original display by selecting the options Ln X and Ln Y. *To prevent illegal transformations, when the axes are swapped, or data is obtained from the notebook (except scenarios), the model is always returned to options X and Y.*

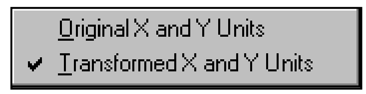
4. **Options**

   a. On the menu bar, click Options. Select Show Data. You will see a list of the X and Y data values that are displayed on the scatter plot.

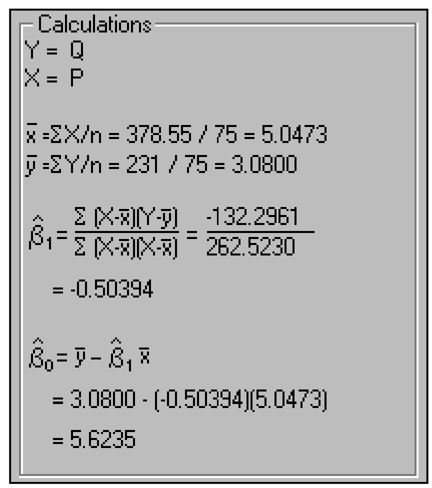
   b. On the menu bar, click Options. Select Show Confidence and Prediction Intervals and then Show 90% Confidence and Prediction Intervals. The confidence bands follow the shape of the estimated regression function with additional bend. Note that the prediction intervals (for individual Y values) are wider than the confidence intervals (for the *conditional mean* of Y). Select Show 90%, 95%, and 99% Confidence Intervals to see a visual comparison of three confidence levels. Note that the 95% intervals are wider than the 90% intervals. Finally, select No Confidence or Prediction Intervals to restore the display to its default appearance.



   c. On the menu bar, click Options. Select Axis Scale and choose Transformed X and Y Units. The fitted regression function immediately changes its appearance. In this example, it becomes a straight line, because *the elasticity function is linear in terms of Ln X and Ln Y*. Observe that the axis scales now are quite different, and are labeled Ln X and Ln Y instead of X and Y. Return to Original X and Y Units.



   d. On the menu bar, click Options. Select Show Calculations (note that this option is unavailable unless you have already clicked the Estimate Model button). You will see a dialog box that shows the transformations and formulas used in the OLS calculations for the slope and intercept. This can be useful for understanding how the fitted equation is derived. This option is *not* available if the model has more than one independent variable since the estimation equations are much more complex and not overly instructive.



5. **Copying a Display**

   Click on any graph or the ANOVA Table. Press the Copy button on the toolbar or select Copy from the Edit menu on the menu bar. The copied display can be pasted into another application.

6. **Help**

   Click Help on the menu bar at the top of the screen. Search for Help lets you search an index, Contents shows a table of contents for this module, Using Help gives instructions on Help, and About gives licensing and copyright information.

7. **Exit**

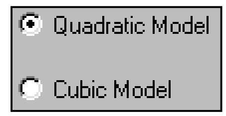
   Close the module by selecting Exit in the File menu (or click ☒ in the upper right-hand corner of the window). You will be returned to the *Visual Statistics* main menu.

Orientation to Additional Features

1. **Interaction Variable**
Press the Show Notebook button. Click on the Scenarios tab. Click on Interaction Models, select the Interaction Variable with Three Values scenario, read it, enter 43 in the Version label box, and press the OK button. The new option Interaction Model appears in the Independent Variable panel and has been selected. This option can also be obtained from the menu bar: click Options, select Interaction Model and Allow. Press the Estimate Model button. Three separate regression lines are fitted, each color-coded (red, green, and blue in this example) as are the data points that match the coding of the interaction variable. In this scenario, the interaction variable is number of car cylinders (1 = 4 cylinders, 2 = 6 cylinders, 3 = 8 cylinders).

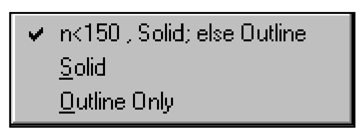
2. **Other Options**

a. On the menu bar, click Options. Select Allow Polynomial Models. Two new options appear on the bottom of the Independent Variable panel. Select Quadratic Model and select Y in the Dependent Variable panel. Press the Estimate Model button. Note the shape of the new fitted function. Select Cubic Model and then press the Estimate Model button. Observe that the new function has an inflection point, where the line's slope changes from becoming flatter to becoming steeper, at about 3.2 on the Weight axis. Also note that the R^2 increased slightly. This always occurs when additional variables are added to the model.



b. On the menu bar, click Options. Select Show Y Intercept. The display shows X = 0. The Y intercept (104.18) can now be seen. If X = 0 is showing, the option has no effect.

c. Press the Show Notebook button. Select the Interaction Variable with Four Values scenario, read it, and press the OK button. Press the Estimate Model button. On the menu bar, click Options and select Graph Display. The menu offers four choices. The Color Printer option is selected. Select the Black and White Printer. The colored lines and symbols change to lines and symbols that they are easier to differentiate on a non-color printer. If you click Symbols you will see another submenu showing three choices. The default is to display solid symbols unless the sample size exceeds 150, and then to use outline symbols (to reveal more detail as the scatter plot becomes more crowded). However, you can override this choice and choose Solid or Outline Only symbols all the time. If you click Type of Plot, choose Time Plot to connect the points on the scatter plot in their temporal order. *For cross-sectional data, which doesn't have a temporal order, this can result in a graph with a very odd appearance.*



3. **Databases**
Press the Show Notebook button. Select the Databases tab, select a type of database, and click on a particular database. Read about the database and press OK to create a model. Scroll down the list of variables and select a dependent variable. Press the dependent variable ⇨ button. To remove the variable, press the ⇦ button. Select an independent variable and press the independent variable ⇨ button. The Sample Size button or Time Period button (depending on the database) enables you to change the sample size. You can select an interaction variable now or later if needed. Press OK to create your scatter plot.

Basic Learning Exercises   Name ________________

Exponential Growth Model

1. Press the Show Notebook button. Click on the Scenarios tab, select Growth Models and select Exponential Growth Model. Read the scenario. a) What are the variables? Enter 90 in the Version label box, and press OK. b) Describe the scatter plot. b) Press the Estimate Model button. Describe the fitted line. c) Write the estimated equation. d) What does Ln stand for? e) Why is there a ^ above the dependent variable? f) What are the axis labels? **Hint:** This model is sometimes called a Log-Linear model.

2. Even though the estimated model is nonlinear the model was estimated with Ordinary Least Squares (OLS). Select Options on the Menu bar and select Show Calculations. Notice that Y is the Ln of GDP. Also, notice that the estimated slope is the same equation you learned in simple regression. The transformation is what caused the fitted line to be curved. Press the Close button. Select Options on the Menu bar and select Axis Scale and then Transformed X and Y Units. Describe the estimated line and axes labels.

3. Select Options on the Menu bar and select Show Confidence and Prediction Intervals and then Show 95% Confidence and Prediction Intervals. a) What does the confidence interval show? b) What does the prediction interval show? c) Why are they both closest to the estimated line at Period 12.5? d) Select Options on the Menu bar and select Axis Scale and then Original X and Y Units. Describe the confidence and prediction intervals. When are they closest to the estimated nonlinear fitted line?

4. Remove the confidence and prediction intervals using Options on the Menu bar. The table shows the fitted Y value that corresponds to each X value. The ratio Y_2/Y_1 tells us that Y grew 20.7% from period 5 to 10. Raising the ratio to the 1/5 power (5 is the number of periods) gives the growth per period (3.83% from period 5 to 10). Which parameter estimates this growth rate?

X	Y	Y_2/Y_1	$(Y_2/Y_1)^{1/5}$	Growth
5	5.27	NA	NA	NA
10	6.36	1.207	1.0383	3.83%
15	7.68	1.208	1.0384	3.84%
20	9.26	1.206	1.0381	3.81%

5. Press the Show Notebook button. Click on the Examples tab, select Macroeconomic Examples, and select Consumer Price Index. Read the example. a) What years will be displayed on the graph? b) Press the OK button. Select Ln Y option button in the Dependent Variable panel to estimate an exponential growth model. Press the Estimate Model button. What has inflation (growth in prices) averaged over the time period? c) In the late 1990's was inflation above or below this historic average?

Declining Returns Model

6. Press the Show Notebook button. Click on the Scenarios tab, select Growth Models and select Declining Returns Model. Read the scenario. a) What are the dependent and independent variables? b) Enter 40 in the Version label box, and press the OK button. Describe the scatter plot. c) Press the Estimate Model button. Describe the fitted line. d) Write the estimated equation. e) Why is this model also called a Linear-Log model?

7. Select Options on the Menu bar and select Axis Scale and then Transformed X and Y Units. Describe the estimated line and axes labels?

8. The table shows the fitted Y value (in thousands) that corresponds to each X value. The difference in Y's, ΔY, is about 64,000 when X is doubled. This constant change due to a constant *percentage* increase in X (100% in this case) is characteristic of the declining returns model. *The estimated model tells how much Y increases if X is increased by 1%.* In this case, a 1% increase in Employment (X) causes Output (Y) to increase by 917.24 units. How was 917.24 obtained? Why didn't doubling X increase Y by 91,724 units?

X	Y(000's)	ΔY(000's)
400	614	NA
800	677	63
1600	741	64
3200	805	64

9. Press the Show Notebook button. Click on the Examples tab, select Financials Examples, and Baseball Salaries. Read the example. a) What are the variables and in what units are they measured? b) Press the OK button. Select Ln X option button in the Independent Variable panel to estimate a declining returns model. Press the Estimate Model button. What does this model suggest about baseball salaries and years in professional baseball?

Elasticity Model

10. Press the Show Notebook button. Click on the Scenarios tab, select Growth Models and select Elasticity Model. Read the scenario. a) What are the dependent and independent variables and their units of measurement? b) Enter 20 in the Version label box, and press the OK button. Describe the scatter plot. c) Press the Estimate Model button. Describe the fitted line. d) Write the estimated equation. **Hint:** *This model is also called a Log-Log or multiplicative model.*

11. Select Options on the Menu bar and select Axis Scale and then Transformed X and Y Units. Describe the estimated line and axis labels.

12. The table shows the fitted Y value that corresponds to each X value. The ratio Y_2/Y_1 tells us that Y grew by –35.8% (= 1 – 0.642 changed to percent) when price increased from \$1 to \$2 or increased by 100%. This constant *percentage* change due to a constant *percentage* increase in X (100% in this case) is characteristic of the elasticity model. The estimated model tells the percentage Y increases if X is increased by *1%.* In this case, a 1% increase in Price (X) results in Quantity (Y) *decreasing* by 0.635%. –0.635 is called an elasticity. Which parameter estimates this elasticity? Why didn't doubling X *decrease* Y by 63.5%?

X	Y	Y_2/Y_1	Growth
1	8.1	NA	NA
2	5.2	0.642	–31.4%
4	3.4	0.654	–33.3%
8	2.2	0.647	–31.2%

13. Press the Show Notebook button. Click on the Examples tab, select Health Examples, and Birth Weights. Read the example. a) What are the independent and dependent variables and in what units are they measured? b) Press the OK button. Select Ln X option button in the Independent Variable panel and Ln Y in the Dependent Variable panel to estimate an elasticity model. Press the Estimate Model button. What does this model suggest about gestation period and birth weight? c) What would be the shape of the fitted line if the slope coefficient were –2.57?

14. Press the Show Notebook button. Click on Next Page in the lower right corner and select Utility Costs. Read the example. a) What are the independent and dependent variables and in what units are they measured? b) Press the OK button. Select Ln X and Ln Y to estimate an elasticity model. Press the Estimate Model button. What does this model suggest about gas used and energy cost? c) What would be the shape of the fitted line if the slope coefficient were –0.272?

Intermediate Learning Exercises Name ________________

The Polynomial Model

15. Press the Show Notebook button. Click on the Scenarios tab, select Models Using Exponents, and select Quadratic Model. Read the scenario. a) What are the dependent and independent variables and their units of measurement? b) Enter 5 in the Version label box, and press OK. Describe the scatter plot. c) Notice that two new options have been added to the Independent Variable panel box. These options can be selected by selecting Options on the Menu bar and selecting Allow Polynomial Models. Which model is selected? d) Press the Estimate Model button. Describe the fitted line. e) Write the estimated equation.

16. In a quadratic model selecting Options on the Menu bar and selecting Axis Scale and then Transformed X and Y Units will have no effect since there are two independent variables. The table shows the fitted Y value that corresponds to each X value. The ΔX and ΔY columns show the change in X and Y. The ΔY/ΔX column shows the slope of the fitted curve. Although this table shows how to calculate the slope using the graph, unlike the Ln transformation, there is no straight-forward interpretation of the quadratic model. An alternative approach to calculating the slope is to take the function's derivative: 0.438 + 2 (–0.505) (Revenue). a) Calculate the slope using this formula (use the average value of X for Revenue). b) Why aren't the slopes the same as above? c) Why does this parabola have a maximum?

X	ΔX	Y	ΔY	ΔY/ΔX
2	NA	4.88	NA	NA
3	1	5.07	0.19	0.19
4	1	5.15	0.08	0.08
5	1	5.14	–0.01	–0.01

17. Press the Show Notebook button. Click on the Examples tab, select Time Series Examples, and Gasoline Prices. Read the example. a) What are the variables and in what units are they measured? b) Press the OK button. Select Quadratic Model in the Independent Variable panel. Press the Estimate Model button. What does this model suggest about gas prices during this the late 1990's? c) Why does this parabola have a minimum?

18. Select the Cubic Model. Press the Estimate Model button. Describe the fitted line.

The Standardized Model

19. Press the Show Notebook button. Click on the Scenarios tab, select Linear Models, and select Standardized Variables. Read the scenario. a) Define the dependent and independent variables and their units of measurement. b) Enter 20 in the Version label box, and press OK. Describe the scatter plot. c) Press the Estimate Model button. Describe the fitted line. d) Write the estimated equation.

20. In a standardized model selecting Options on the Menu bar and selecting Axis Scale and then Transformed X and Y Units will have no effect since the model is already linear. a) What is the standardization transformation? b) Why is this considered a linear transformation? c) Why is the estimated intercept equal to 0? **Hint:** Unlike the Ln transformation, it is rare to use this transformation on just the independent or dependent variable.

21. a) What is the coefficient of the independent variable called in this model? b) What is the interpretation of the estimated slope? c) Select Y and X in the panels. Press the Estimate Model button. Did the fitted line change? **Hint:** Because this transformation removes all scaling, the size of the beta weight measures the variable's impact on the dependent variable. This is useful in a multiple regression model where you have many independent variables.

22. Press the Show Notebook button. Click on the Examples tab, select Examples with People, and Statistics Course Grades. Read the example. a) What are the independent and dependent variables and in what units are they measured? b) Press the OK button. Select Standardized X in the Independent Variable panel and Standardized Y in the Dependent Variable panel to estimate a standardized model. Press the Estimate Model button. What does this model suggest about the final grade and grade on exam 1?

Advanced Learning Exercises Name ________________

Power and Reciprocal Models

23. Press the Show Notebook button. Click on the Scenarios tab, select Models Using Exponents, and select Power of X Transformation. Read the scenario. a) Define the dependent and independent variables and their units of measurement. b) To what power is X raised? c) Enter 25 in the Version label box, and press OK. Describe the scatter plot. d) Press the Estimate Model button. Notice that the independent variable option selected is $X^{0.2}$. Describe the fitted line. e) Write the estimated equation.

24. The downside of this type of model is that you must know the exponent *before* you estimate the model. One way around this problem is to re-estimate the model with different values of the exponent and select the one that minimizes the Error Sum of Squares (ESS). The key to doing this search efficiently is to set up a systematic approach. Fill in the ESS in the table below (the first one is entered). After each entry, write down *why* the next exponent was selected. To change the power to which X is raised, select the X option and then re-select the X^c option. This will bring up a dialog box where you can enter your exponent. Press OK and the Estimate Model button. Write down the estimate of your final model. *This same process could be used if Y is raised to a power.*

Exponent	ESS	Why Select the Next Exponent to Estimate
0.2	1.558	
0.3		
0.1		
0.01		
0.05		
0.03		
0.02		
Final Model –		

25. Press the Show Notebook button. Select the Reciprocal of X Model. Read the scenario. a) Define the dependent and independent variables and their units of measurement. b) Press OK. Describe the scatter plot. c) Press the Estimate Model button. Describe the fitted line. d) Write the estimated equation. e) What is the elasticity of this demand function? **Hint:** This is a special case of the power model where the exponent is –1.

Interaction Variables

26. Press the Show Notebook button. Click on the Scenarios tab, select Interaction Models, and select Interaction Variable with Two Values. Read the scenario. a) Define the dependent and independent variables and their units of measurement. b) Define the interaction variable and its unit of measurement. c) Enter 35 in the Version label box, and press OK. Describe the scatter plot. d) Press the Estimate Model button. Write the estimated equation. e) Why are there two lines? Describe the fitted lines.

27. a) What is the p-value of the interaction term (i.e., SqFt People)? b) If we tested the coefficient at $\alpha = 0.05$, what does it tell us about the value of the condominium and the number of people living in it? c) Is the size of the condominium important in part (b)?

28. Press the Show Notebook button. Select Interaction Variable with Four Values. Read the scenario. a) What are the dependent and independent variables and their units of measurement? b) What is the interaction variable and its unit of measurement? c) Press OK. Press the Estimate Model button. Write the estimated equation. d) Why are there four lines? Describe the fitted lines.

29. a) What is the p-value of the interaction term (i.e., Exp Ed)? b) If we tested the coefficient at $\alpha = 0.01$, what does it tell us about Income and Education? c) Is Experience important in part (b)?

Individual Learning Projects

Write a report on one of the three topics listed below. Use the cut-and-paste facilities of the module to place the appropriate graphs in your report.

1. Use one of the Databases in this module to select variables that can be estimated with a simple linear model ($Y = \beta_0 + \beta_1 X + u$). Copy the graph with the estimated model for your paper. Define your variables, explain the model, and interpret the estimated coefficients. Repeat each step using a standardized model, an exponential growth model, a declining returns model, and an elasticity model. Your data should be appropriate for the model used.
2. Use data from an Example or a Database in this module, to create a nonlinear model (not a polynomial). Estimate the model and display a 95% confidence and prediction interval on the graph. Copy the graph for your paper. Transform the axes to display the transformed variable(s) on the axes. Copy the graph for your paper. Define your variables, explain the model and interpret the estimated coefficients. Explain in *detail* why graph 2 shows a linear line while graph 1 shows a nonlinear line. Review how a 95% confidence and prediction interval are created. Explain how the intervals were created for graph 2 and then for graph 1. Explain why the two graphs display the intervals differently.
3. Use data from an Example or a Database in this module to create a power model in X. Make sure that neither X nor Y is less than or equal to zero. Define your variables. Estimate the model using the search procedure outlined in Exercise 24. Create a table with your exponents and Error Sum of Squares. Copy at least 2 graphs for your paper. Use the table and graphs to explain how you searched for the best model. Explain your final estimated model. Repeat the process using the same data and a power model in Y. Compare and contrast the two models.

Team Learning Projects

Select one of the three projects listed below. In each case, produce a team project that is suitable for an oral presentation. Use presentation software or large poster boards to display your results. Graphs should be large enough for your audience to see. Each team member should be responsible for producing some of the graphs. Ask your instructor if a written report is also expected.

1. This project is for a team of two or three. Using Examples and Databases from this module, you are to illustrate and interpret 11 different models: simple linear model, standardized model, elasticity model ($\beta_1 > 1$, $0 < \beta_1 < 1$, and $\beta_1 < 0$), exponential growth model ($\beta_1 > 0$, $\beta_1 < 0$), decreasing returns model ($\beta_1 > 0$, $\beta_1 < 0$), and quadratic model ($\beta_2 > 0$, $\beta_2 < 0$). Although you may be able to use the same variables more than once, most models will use different data. For each model define your variables, explain the model, and interpret the coefficients (if they have an interpretation). Each model should have a graphical display. Models that have similar explanations should be presented together. The purpose of this project is to demonstrate the variety of different models that can be estimated with OLS, to understand the similarities and differences in these models, and to understand the importance of the slope coefficient in differentiating these models.
2. This project is for a team of three. The purpose of this project is to investigate and explain different types of interaction variables. Data will come from any cross sectional database in this module. The team will select the dependent and independent variable, and sample size. Each variable should be defined. One team member will use an interaction variable with two or three values, another member will use an interaction variable with 5 to 10 values, and the last member will use a continuous interaction variable (more than 12 different values). Each team member will estimate their model and create a graph of this estimation. The model will be explained and the interaction term interpreted and tested for significance. The team should compare and contrast the three models. The project should be repeated using a time series database from this module. Compare and contrast the process using cross-sectional data and time series data.
3. This project is for a team of three or four. Use data from an Example or a Database in this module, to create a power model in X. Make sure that neither X nor Y is less than or equal to zero. Define your variables. Estimate the model using the search procedure outlined in Exercise 24. Create a table with your exponents and Error Sum of Squares. Show at least 2 graphs for your presentation. Use the table and graphs to explain how you searched for the best model. Explain your final estimated model. Repeat the process using the same data and a power model in Y. Use the final exponents for X and Y to *begin* a search for the best model using *both* an exponent for X and Y. In searching for the best exponents in this new model you must search simultaneously over both exponents. Your table of ESS should have exponents for Y across the top and exponents for X down the side. The table should be filled in completely (i.e., for any value of the exponent of Y that is used you need to estimate the model with every value of the exponent of X that you used). You should use at least 8 different exponents for both X and Y. At least 3 graphs from this search should be included in your presentation. The focus of this project is how to estimate a model using a search procedure when both X and Y are raised to a power.

Self-Evaluation Quiz

1. In the model Ln(Y) = 4 + 5Ln(X) + u, the interpretation of the slope coefficient is that
   a. a 1% change in X results in a 5 unit change in Y.
   b. a 1% change in X results in a 5% change in Y.
   c. a 1% change in X results in a 0.05 unit change in Y.
   d. a 1-unit change in X results in a 5 unit change in Y.
   e. a 1-unit change in X results in a 500% change in Y.

2. Which model could show a 3% growth rate?
   a. $Std(Y) = \beta_0 + \beta_1\, Std(X)$.
   b. $Y = \beta_0 + \beta_1 X^2$.
   c. $\ln Y = \beta_0 + \beta_1 X$.
   d. $Y = \beta_0 + \beta_1 \ln X$.
   e. $\ln Y = \beta_0 + \beta_1 \ln X$.

3. Which of the following is a declining returns model?
   a. $Std(Y) = \beta_0 + \beta_1\, Std(X)$.
   b. $\ln Y = \beta_0 + \beta_1 \ln X$.
   c. $Y = \beta_0 + \beta_1 X^2$.
   d. $\ln Y = \beta_0 + \beta_1 X$.
   e. $Y = \beta_0 + \beta_1 \ln X$.

4. For any transformed bivariate model, a 95% prediction interval for y | x
   a. will be narrower than a 95% confidence interval for Y.
   b. will be widest near the mean of the transformed variable X.
   c. will contain the individual Y values about 95% of the time.
   d. will contain the true conditional mean of Y about 95% of the time.
   e. will be linear if the transformed equation is linear.

5. Which equation represents a nonlinear variable transformation?
   a. $\ln Y = \beta_0 + \beta_1 X$.
   b. $Y = \beta_0 + \beta_1 (1/X)$.
   c. $\ln Y = \beta_0 + \beta_1 \ln X$.
   d. $Y = \beta_0 + \beta_1 X^{1/2}$.
   e. All of the above.

6. Which do you expect from the polynomial model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$?
   a. It is generally easier to interpret than a linear or quadratic model.
   b. It is incapable of fitting data with an inflection point.
   c. It is likely to yield a better fit than the simple linear model.
   d. It is impossible to estimate using the OLS method.
   e. It cannot be graphed on a simple X-Y scatter plot.

7. The standardized variable model $Std(Y) = \beta_0 + \beta_1\ Std(X)$
   a. is a simple linear transformation of X and Y.
   b. is sometimes called a beta-weight model.
   c. improves the scaling of variables of dissimilar magnitudes.
   d. no longer uses the original units of X and Y (e.g., kilograms).
   e. All of the above.

8. Which model can be displayed as a curve that increases at a decreasing rate as X increases?
   a. $Y = 200 – 0.5\ (1/X)$.
   b. $Y = 200 + 5 \ln X$.
   c. $Y = 200 + 15\ X – 0.5\ X^2$.
   d. $Y = 200 + 5\ X^{1/2}$.
   e. All of the above.

9. Which is *not* a feature of the reciprocal model relating quantity (Q) and price (P)?
   a. It cannot be estimated using the OLS method because it is nonlinear.
   b. It has the form $Q = \beta_1\ (1/P) + u_i$.
   c. It represents constant elasticity.
   d. Its form is a hyperbola except for the disturbance.
   e. Other things being equal, a 1% increase in P will result in a 1% rise in P.

10. Which is a power of Y transformation that represents *declining* growth in Y?
    a. $Y^{0.5} = 12 + 10\ X$.
    b. $Y^2 = 12 + 10\ X$.
    c. $Y^{-0.5} = 12 + 10\ X$.
    d. $Y^{-2} = 12 + 10\ X$.
    e. None of the above.

11. Which model uses an interaction variable?
    a. $Y = \beta_0 + \beta_1\ X + \beta_2\ X^2$.
    b. $\ln Y = \beta_0 + \beta_1\ X + \beta_2\ Z$.
    c. $Y = \beta_0 + \beta_1\ X^{1/2}$.
    d. $\ln Y = \beta_0 + \beta_1\ X + \beta_2\ XZ$.
    e. Two of the above.

12. An interaction variable
    a. can have only two values (generally 0 or 1).
    b. may have several discrete values (such as 1, 2, 3).
    c. may be a continuous variable.
    d. can result in a changing slope of the fitted equation.
    e. could have any of these characteristics.

Glossary of Terms

ANOVA table Decomposition of variance in a regression, showing total sum of squares and its sources (regression, error) along with degrees of freedom and mean squares. *Total* degrees of freedom equals $n - 1$, *error* degrees of freedom equals $n - k - 1$, and the *regression* degrees of freedom equals k, where n is the sample size and k is the number of independent variables.

Beta weight Coefficient of the independent variable in a regression model of the form $Std(Y_i) = \beta_0 + \beta_1 \, Std(X_i)$ where $Std(Y) = (y_i - \overline{y}) / s_y$ and $Std(X) = (x_i - \overline{x}) / s_x$. β_0 is always zero. It is unusual not to transform both the independent and dependent variables.

Confidence interval for mean of Y With probability $1 - \alpha$ the upper and lower values of the range of estimates for the *conditional mean* of Y. When plotted over the entire range of X they comprise a *confidence band*. Individual Y values often lie outside this confidence band.

Cubic model Regression model of the form $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$. It will have two extrema (over the displayed range of X, it may not be possible to see them) and an inflection point. See **Polynomial model**.

Declining returns model Regression model of the from $Y_i = \beta_0 + \beta_1 Ln(X_i) + u_i$. In this model, a 1% change in X results in a $\beta_1/100$ unit change in Y. It is also called a Linear-Log model.

Elasticity model Regression model of the form $Ln\, Y_i = \beta_0 + \beta_1 Ln(X_i) + u_i$. Economists use it to describe a relationship between price and quantity demanded or supplied. In this model, a 1% change in X_i leads to a β_1% change in Y_i. It is also called a Log-Log or multiplicative model.

Exponential growth model Regression model of the form $Ln\, Y_i = \beta_0 + \beta_1 X_i + u_i$. In this model, a 1-unit change in X results in a $100 \times \beta_1$% change in Y. It is also called a Log-Linear model. It is commonly used for financial or economic data.

F statistic In a regression ANOVA table, the ratio of the *regression* mean square to the *error* mean square. It is used to test the overall significance of the regression. See **p-value**.

Interaction variable An interaction variable is a transformation where a third variable Z_i enters the analysis through an interaction with X_i. It has the form $Z_i X_i$, i.e., Z_i is multiplied by X_i. If entered into a simple linear model it has the form $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i X_i + u_i$. Although it requires multivariate estimation techniques, it can be viewed on a scatter plot as line segments or as a series of different lines (one for each unique value of Z_i).

Multiple correlation coefficient Measure of overall fit in a regression. It is the square root of R^2. It may be interpreted as the correlation between Y_{actual} and Y_{fitted} over all n observations.

Ordinary Least Squares (OLS) Method of estimating a regression that guarantees the smallest possible sum of squared residuals. The residuals sum to 0 using the OLS method.

Parameter Numerical constant needed to define a particular model or distribution. A regression model's the parameters are the intercept and the coefficients of the k independent variables, whose true values are denoted $\beta_0, \beta_1, \beta_2, ..., \beta_k$. See **Estimated coefficient**.

Prediction interval for individual Y Upper and lower values of the range of estimates for an *individual value* of Y conditional upon the value of the independent variable(s). The prediction interval for Y is wider than the confidence interval for the mean of Y. Individual Y values generally will lie within the prediction interval unless the assumptions are violated.

P-value Probability (usually two-tailed) of committing Type I error if we reject the null hypothesis that a parameter is zero (for example, a regression coefficient). A small p-value (such as 0.01) would incline us to reject the hypothesis that the true parameter is zero.

Polynomial model Since a polynomial model only involves X and Y, it can be graphed in two dimensions. However, it requires multivariate methods of estimation because it has more than one independent variable. Adding polynomial terms will improve the R^2, but the resulting model may have no meaningful interpretation. Nonetheless, a the model does offer a useful test for non-linearity. If the quadratic and cubic coefficients are near zero (i.e., insignificant t-value) it essentially collapses to a linear model. See **Quadratic model** and **Cubic model**.

Power model Regression model of the form $Y_i = \beta_0 + \beta_1 X_i^c + u_i$ where c is specified *a priori*. Special cases are c = 2 (quadratic model without the X term), c = 1 (the linear model), c = –1 (the reciprocal model), and c = 0.5 (a declining growth model). Alternatively, it can be of the form $Y_i^c = \beta_0 + \beta_1 X_i + u_i$.

Quadratic model Regression model of the $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$. The quadratic model has a single peak or trough, However, over the displayed range of X, it may not be possible to see them. See **Polynomial model**.

Reciprocal model Regression model of the form $Y_i = \beta_0 + \beta_1(1/X_i) + u_i$ or of the form $1/Y_i = \beta_0 + \beta_1 X_i + u_i$. **See Power model**.

R-squared Ratio of the *regression* sum of squares to the *total* sum of squares. R^2 near 0 indicates the fit is poor while R^2 near 1 indicates the fit is good. Also called coefficient of determination.

Standard error Estimate of the standard deviation of the stochastic disturbances, using the square root of the sum of the squared residuals, divided by n – k – 1, where n is the sample size and k is the number of independent variables.

Transformed variable Common transformations include taking logarithms [$\ln(X_i)$ or $\ln(Y_i)$], raising to a power [X_i^c or Y_i^c], inversion [$1/X_i$ or $1/Y_i$], or standardizing by subtracting the mean and dividing by the standard deviation. As long it contains only two variables, the transformed model is still a simple regression model and may be estimated using the OLS method. However, on a graph the fitted regression equation may be nonlinear. Some variable transformations may be impossible if there are zero or negative values (e.g., 1/X).

Solutions to Self-Evaluation Quiz

1. b Do Exercises 10–14. Read the Overview of Concepts.
2. c Do Exercises 1–5. Read the Overview of Concepts.
3. e Do Exercises 6–9. Read the Overview of Concepts.
4. c Do Exercise 3.
5. e Do Exercises 1, 10, 23–25. Read both the Overview and the Illustration of Concepts.
6. c Do Exercises 15–18.
7. e Do Exercises 19–22. Read the Overview of Concepts.
8. e Do Exercises 6–9, 15–17, 23–25.
9. a Do Exercise 25.
10. b Do Exercises 6–9, 23, 24. Individual Learning Project 3.
11. d Do Exercises 26–29. Read the Overview of Concepts.
12. e Do Exercises 26–29. Read the Overview of Concepts.