Problem Set for Chapter 3: Simple Regression Analysis
ECO382 Econometrics Queens College K.Matsuda

Excel Assignments

You are required to hand in these Excel Assignments by the due Matsuda specifies. Legibility is very important.

[1]Use the following data. Produce the Excel output of regression analysis. Compare this least-squares line in Excel output with your hand-calculated least-squares line. And comment.

Table 3.2	Example Data					
	x	1	2	3	4	5	6
	y	3	2	8	8	11	13

[2]For each of the data sets discussed in this section, use a computer to read the data, construct a scatterplot of y versus x, and produce the regression output relating y to x.

Example 3.2 Estimating Residential Real Estate Values

The Tarrant County Appraisal District must appraise properties for all of the county. The appraisal district uses data such as square footage of the individual houses as well as location, depreciation, and physical condition of an entire neighborhood to derive individual appraisal values on each house. This avoids labor-intensive reinspection each year.

Regression can used to establish the weight assigned to various factors used in assessing values. For example, Table 3.5 shows the value and size in square feet for a sample of 100 Tarrant county homes (these data are from 1990). A scatterplot of value (y) versus size (x) is shown in Figure 3.5.

Using a statistical package, the regression equation relating value to size can be determined as

$$VALUE = -50{,}038 + 72.8SIZE$$

If size were the only factor thought to be of importance in determining value, this equation could be used by the appraisal district. But obviously, other factors need to be considered. Developing an equation that includes more than one important factor (explanatory variable) is discussed in Chapter 4.

Example 3.3 Pricing Communication Nodes

In recent years, the growth of data communications networks has been amazing. The convenience and capabilities afforded by such networks are appealing to business with locations scattered throughout the United States and the world. Using networks allows centralization of a main computer with access through personal computers at remote locations.

The cost of adding a new communications node at a location not currently included in the network was of concern for a major Fort Worth manufacturing company. To try to predict the price of a new communications nodes, data were obtained on a sample of existing nodes. The installation cost and the number of ports available for access in each existing node were readily available information. These data are shown in Table 3.6 and a scatterplot of cost (y) versus number of ports (x) is shown in Figure 3.6.

Again, using a statistical package, the equation relating the price of the new communications node to the number of access ports to be included at the node is

$$\text{COST} = 16{,}594 + 650\text{NUMPORTS}$$

Where NUMPORTS represents the number of ports. This equation could be used to help predict the cost of installing new communications nodes based on the number of access ports to be included.

Example 3.4 Forecasting Housing Starts

Forecasts of various economic measures are important to the U.S. government and to various industries throughout the United States. The construction industry is concerned with the number of housing starts in a given year. Accurate forecasts can help with plans for expansion or cutbacks within the industry.

Table 3.7 shows data on the number of housing starts for the years 1963 to 1996. Also shown are data on home mortgage rates for new home purchases (U.S. average) for the same years. These data were obtained *from Business Statistics of the united States, 1997 Edition.* A scatterplot of housing starts (y) versus mortgage rates (x) is shown in Figure 3.7. Note that the relationship appears to be considerably “weaker” than in the other scatterplots presented in this section. Intuitively, we might expect the relationship between housing starts and mortgage rates to be a strong one. But from the data, this does not appear to be the case. Perhaps there are other variables that might be more strongly related to housing starts that could be used to provide accurate forecasts for future years. From viewing the scatterplot, mortgage rates alone do not appear to be particularly to be helpful.

[3]Estimating Residential Real Estate Values (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

1. What is the sample regression equation relating SIZE to VALUE?
2. Is there sufficient evidence to conclude that a linear relationship exists between VALUE and SIZE?
3. Find a 95% confidence interval estimate of β_1.
4. Interpret what this sample regression equation means.

[4]Pricing Communications Nodes (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

1. What is the sample regression equation relating NUMPORTS to COST?
2. Is there sufficient evidence to conclude that a linear relationship exists between COST and NUMPORTS?
3. Find a 95% confidence interval estimate of β_1.
4. A claim is made that each new access port adds at least $1000 to the installation of a communications node. Test this claim.
5. Interpret what this sample regression equation means.

[5] Forecasting Housing Starts (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

1. What is the sample regression equation relating RATES to STARTS?
2. Is there sufficient evidence to conclude that a linear relationship exists between STARTS and RATES?
3. Find a 95% confidence interval estimate of β_1.
4. Interpret what this sample regression equation means.

[6]Estimating Residential Real Estate Values (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

1. What percentage of the variation in VALUE is explained by the regression?
2. Use the F test and a 5% level of significance to test the hypotheses

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

[7]Pricing Communications Nodes (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

1. What percentage of the variation in COST is explained by the regression?
2. Use the F test and a 5% level of significance to test the hypotheses

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

[8] Forecasting Housing Starts (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

1. What percentage of the variation in STARTS is explained by the regression?
2. Use the F test and a 5% level of significance to test the hypotheses

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

[9]Estimating Residential Real Estate Values (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

For an individual house with 3,000 square feet, find a prediction of its real estate value.

[10]Pricing Communications Nodes (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

For an individual communication node with 40 access ports, find a prediction of cost.

[11] Forecasting Housing Starts (Continued)

Use the Excel outputs of regression analysis to answer the following questions.

When the home mortgage rate is 6.5%, find the prediction of the number of housing starts.

[12]ABX Company Sales

The ABX Company sells winter sports merchandise including skis, ice skates, sleds, and so on. Quarterly sales (in thousands of dollars) for the ABX Company are shown in Table 3.11. The time period represented starts in the first quarter of 1990 and ends in the fourth quarter of 1999.

a. Use Excel and produce time-series plot of the sales figures.
b. What does this time-series plot suggest?
c. Use Excel and get the regression output of the linear trend model $y_i = \beta_0 + \beta_1 t + e_i$.
d. Test whether the linear trend component is useful in explaining the variation in sales. Write down the hypotheses to be tested, decision rule using a 5% level of significance, and decision.
e. Use the linear trend equation estimated by Excel and forecast the sales for 2000.1, 2000.2, 2000.3, and 2000.4.