

DIRECTIONS:

- Some exercises include special instructions which modify or clarify textbook instructions.
 - Use a 5% significance level for tests unless noted otherwise.
 - Data sets for all exercises are available on the MINITAB Data Sets link.
-

Textbook Exercises: Exercises have been *grouped together* around the same word problem, for convenience. (The word problems are labeled A, B, C, etc.)

A. (Commercial Bank Data)

- 11.3 (p. 622)
- 11.4

Follow these specific steps:

1. First, download the Table 11.2 data and make a Matrix Plot to scan for patterns between the three variables:
Graph > Matrix Plot > Simple > OK > (Select variables Assets, Deposits, Number) > OK
2. To get a closer view, also make two Fitted Line Plots in MINITAB, putting the response variable Assets on the y -axis each time.
3. Finally, get the multiple regression output (using both predictor variables.)

Answer the following questions:

- (a) Which predictor variable seems from the graphs to be most valuable for predicting Assets?
- (b) Scan the graphs and the data table. Which state appears to be an *outlier* in the graphs? Can you think of a (business) reason why?
- (c) Write down the estimated (sample) regression line from the MINITAB output.
- (d) Let $x_1 = \text{Deposits}$ and $x_2 = \text{Number of Banks}$. Interpret $\hat{\beta}_1$ from multiple regression.
- (e) Interpret $\hat{\beta}_2$ from multiple regression.
- (f) If two states have the same Deposits but the first state has more banks than the second state, which state would you expect to have more Assets?
- (g) Suppose that a new commercial bank is planning to open in Iowa in 2012 but that the impact of that opening on Deposits in Iowa is unclear or unknown.
What's the best estimate of the impact of the bank's opening on Assets in Iowa? (Use 10% significance for this question.)

(continued)

(h) How can the *negative* slope for (Number of Banks) in (e) be reconciled with the *positive* slope in (g)?

- i. Find the Total Effect of Number on Assets.
- ii. Find the Direct Effect of Number on Assets.
- iii. Find the Indirect Effect of Number on Assets.
- iv. Explain the *sign* (positive or negative) of the Indirect Effect after first getting a correlation matrix to supplement the matrix plot:

Stat > Basic Statistics > Correlation > (Select the 3 variables)
> (Uncheck box for Display p-values)

B. (Dow Jones Company Stocks)

(First read the one-paragraph description of Case 11.1 in the middle of page 620 and briefly scan Table 11.1 on page 621.)

- 11.8 Use MINITAB output for multiple regression to answer this question.
- 11.13 Find s^2 , s , and MSE from the MINITAB output.
- 11.14 The easiest way to find s_y is by displaying descriptive statistics for $y = \text{Profits}$ in MINITAB.

C. (Home Prices in Ames, Iowa)

- Briefly read the introduction to Example 1 on page 83 in the Notes.
- Then turn to page 91. Review calculations on that page for Direct and Indirect Effects for both Area and Age upon Sales Price.

(1) Page 91 shows the Indirect Effect of Age upon Sales Price (through Area) as

$$(-7.154 \text{ square feet}) \times (\$74.27/\text{square foot}) = \boxed{-\$531}$$

The Logic: (Age \uparrow by one year) \implies (Area \downarrow by 7.154 feet)
and each square foot is worth \$74.27 in Sales Price.

(2) Page 91 shows a different way to calculate the Indirect Effect of Area upon Sales Price:

$$\begin{aligned} \text{Indirect Effect} &= \text{Total Effect} - \text{Direct Effect} \\ &= 92.82 - 74.27 \\ &= \boxed{\$18.55} \end{aligned}$$

Questions:

- (a) Calculate the Indirect Effect of Area similar to Method (1) above.
- (b) Also explain the logic of the calculation.

Hint: First make a Fitted-Line Plot to predict Age from Area using the data set Table 2.13.

D. (Online Stock Brokerages)

- 11.25 (p. 635)

Make a Matrix Plot for the three variables and answer the following questions in addition to (a) and (b) from the text:

- (c) From the Matrix Plot, how does Market Share seem to be related to Assets? to Number of Accounts?
- (d) What does the F test indicate?
- (e) Which of the predictor variables are significant at the 5% level?
- (f) The multiple regression model implies that *increased* market share is associated with *fewer* accounts. What number from the regression output describes this association? Does this mean that a brokerage should close some customer accounts if its goal is to increase market share? Explain.
- (g) Choose the best conservative model for predicting Market Share. (Use this model for all subsequent questions. See the definition of “best conservative model” on page 102 in the Notes.)
What is the prediction equation for this chosen model?
- (h) Predict the market share for an online brokerage which has 500,000 accounts and \$20 billion in assets.
- (i) Predict with 95% certainty the market share for an online brokerage which has 500,000 accounts and \$20 billion in assets.

(Hint: Use the commands
Options > (Enter one or more values into “Prediction intervals for new observations”)
as part of your MINITAB regression procedure.)
- (j) Use the Matrix Plot to explain why neither predictor in the *full* regression model is significant, even though each variable is significant in *simple* regression.

(continued)

E. (Architectural Firms) (First take a look at Table 1.3 on page 13 of the text.)

- 11.35 (p. 636)

Add these parts:

- (c) Make a Matrix Plot in MINITAB. What does the plot suggest about how 1998 Total Billings are related to each of the predictor variables?
- (d) Which of the variables are significant at the 5% level?
- (e) Choose a “best conservative” model for predicting and estimating total billings in 1998. What is the prediction equation? (Use this model for all subsequent analysis.)
- (f) Predict total billings for an architectural firm which employs 10 architects, 20 engineers, and 100 staff.
- (g) Predict with 90% certainty total billings for an architectural firm which employs 10 architects, 20 engineers, and 100 staff.
- (h) Estimate with 90% certainty mean total billings for all architectural firms which employ 10 architects, 20 engineers, and 100 staff.

F. (Computer Science Students)

(First read the one-paragraph description of Case 11.2 at the top of page 638.)

- 11.37 (p. 643)

Perform multiple regression in MINITAB with the three predictor variables HSM, HSS, HSE and perform the t test for HSS. What do you conclude?

- 11.38

- 11.43 (p. 646)

Add parts (b) and (c):

- (b) Choose the best conservative model. What is the regression equation of this model? (Hint: There are $2^3 - 1 = 7$ different models to consider.)
- (c) Interpret the slopes $\beta_1, \beta_2, \beta_3$ in the chosen regression model.

- 11.59 (p. 653)

- 11.60

G. (Vitamins in Bread)

- 11.109, 11.110 (p. 678) **Add this part:** Provide the most-accurate predictions of Vitamin C, A, E content after 4 days.

Hints:

1. See page 109 in the Notes – Example 4 (Engine oxygen) – to review how to create the variable $(\text{Days})^2$ in a MINITAB column.
2. See page 66 in the Notes for MINITAB steps to produce a Residuals Plot to accompany a Fitted-Line Plot.

(continued)

H. (Predicting Water Quality)

- Exercise 11.93 (p. 673)

Ignore the book's directions for this exercise. Instead, follow these steps:

1. Make a Matrix Plot.
2. Get the output for *multiple* regression.
3. Get output for *simple* regression applied to each predictor variable separately.

Answer the following questions, using 10% significance:

- (a) What does the F test from multiple regression indicate about the predictor variables?
- (b) What do the two t tests from multiple regression indicate about the predictor variables?
- (c) From the Matrix Plot, do the predictors Area and Forest appear to be positively related or negatively related? Find the correlation r between Area and Forest (use MINITAB.) If a watershed has greater area, does it tend to be more forested or less forested?
- (d) Write down all three sample regression equations from MINITAB output.
- (e) Split the Total Effect (overall slope) for Area on IBI into Direct and Indirect Effects. (Hint: See similar calculations for Topic 9 Part 1 Example 1 on page 91 in the Notes.) Give a reason for the *sign* (positive or negative) of the Indirect Effect.
- (f) Split the Total Effect for Forest on IBI into Direct and Indirect Effects. Give a reason for the *sign* (positive or negative) of the Indirect Effect.
- (g) Interpret the slope for Forest from multiple regression.
- (h) Suppose you wish to predict with 90% certainty the IBI for Stream A, whose watershed covers 50 square kilometers of area and is 30% forested. Provide the answer from the best conservative model.
- (i) Suppose you wish to predict with 90% certainty the IBI for Stream B, whose watershed covers 23.5 square km but whose percent forestation is unknown. Provide the best possible answer.
- (j) The watershed for Stream C is 90% forested, the watershed for Stream D is 57% forested, but the area for both watersheds has been lost in the database. By how much do we expect the IBI for Stream C to exceed the IBI for Stream D? (Provide the most-accurate possible answer.)