practical

PROBLEMS

4.7. You are given the following data:

Y	X_2	X_3
1	1	2
3	2	1
8	3	−3

Based on these data, estimate the following regressions (*Note:* Do not worry about estimating the standard errors):

a. $Y_i = A_1 + A_2X_{2i} + u_i$
b. $Y_i = C_1 + C_3X_{3i} + u_i$
c. $Y_i = B_1 + B_2X_{2i} + B_3X_{3i} + u_i$
d. Is $A_2 = B_2$? Why or why not?
e. Is $C_3 = B_3$? Why or why not?

What general conclusion can you draw from this exercise?

4.8. You are given the following data based on 15 observations:

$$\bar{Y} = 367.693; \quad \bar{X}_2 = 402.760; \quad \bar{X}_3 = 8.0; \quad \sum y_i^2 = 66{,}042.269$$
$$\sum x_{2i}^2 = 84{,}855.096; \quad \sum x_{3i}^2 = 280.0; \quad \sum y_i x_{2i} = 74{,}778.346$$
$$\sum y_i x_{3i} = 4{,}250.9; \quad \sum x_{2i} x_{3i} = 4{,}796.0$$

where lowercase letters, as usual, denote deviations from sample mean values.

a. Estimate the three multiple regression coefficients.
b. Estimate their standard errors.
c. Obtain R^2 and $\bar{R}^2$.
d. Estimate 95% confidence intervals for B_2 and B_3.
e. Test the statistical significance of each estimated regression coefficient using $\alpha = 5\%$ (two-tail).
f. Test at $\alpha = 5\%$ that all partial slope coefficients are equal to zero. Show the ANOVA table.

4.9. A three-variable regression gave the following results:

Source of variation	Sum of squares (SS)	d.f.	Mean sum of squares (MSS)
Due to regression (ESS)	65,965	—	—
Due to residual (RSS)	—	—	—
Total (TSS)	66,042	14	

a. What is the sample size?
b. What is the value of the RSS?
c. What are the d.f. of the ESS and RSS?
d. What is R^2? And $\bar{R}^2$?
e. Test the hypothesis that X_2 and X_3 have zero influence on Y. Which test do you use and why?
f. From the preceding information, can you determine the individual contribution of X_2 and X_3 toward Y?

4.10. Recast the ANOVA table given in problem 4.9 in terms of R^2.

4.13. In the illustrative Example 4.2 given in the text, test the hypothesis that X_2 and X_3 together have no influence on Y. Which test will you use? What are the assumptions underlying that test?

4.14. Table 4-7 (found on the textbook's Web site) gives data on child mortality (CM), female literacy rate (FLR), per capita GNP (PGNP), and total fertility rate (TFR) for a group of 64 countries.

- **a.** A priori, what is the expected relationship between CM and each of the other variables?
- **b.** Regress CM on FLR and obtain the usual regression results.
- **c.** Regress CM on FLR and PGNP and obtain the usual results.
- **d.** Regress CM on FLR, PGNP, and TFR and obtain the usual results. Also show the ANOVA table.
- **e.** Given the various regression results, which model would you choose and why?
- **f.** If the regression model in (*d*) is the correct model, but you estimate (*a*) or (*b*) or (*c*), what are the consequences?
- **g.** Suppose you have regressed CM on FLR as in (*b*). How would you decide if it is worth adding the variables PGNP and TFR to the model? Which test would you use? Show the necessary calculations.

4.15. Use formula (4.54) to answer the following question:

Value of R^2	n	k	$\bar{R}^2$
0.83	50	6	—
0.55	18	9	—
0.33	16	12	—
0.12	1,200	32	—

What conclusion do you draw about the relationship between R^2 and $\bar{R}^2$?

4.16. For Example 4.3, compute the F value. If that F value is significant, what does that mean?

4.17. For Example 4.2, set up the ANOVA table and test that $R^2 = 0$. Use $\alpha = 1\%$.

4.18. Refer to the data given in Table 2-12 (found on the textbook's Web site) to answer the following questions:

- **a.** Develop a multiple regression model to explain the average starting pay of MBA graduates, obtaining the usual regression output.
- **b.** If you include both GPA and GMAT scores in the model, a priori, what problem(s) may you encounter and why?
- **c.** If the coefficient of the tuition variable is positive and statistically significant, does that mean it pays to go to the most expensive business school? What might the tuition variable be a proxy for?
- **d.** Suppose you regress GMAT score on GPA and find a statistically significant positive relationship between the two. What can you say about the problem of multicollinearity?
- **e.** Set up the ANOVA table for the multiple regression in part (*a*) and test the hypothesis that all partial slope coefficients are zero.
- **f.** Do the ANOVA exercise in part (*e*), using the R^2 value.