Assuming all the assumptions of CLRM are fulfilled, obtain

a. b_1 and b_2.

b. standard errors of these estimators.

c. r^2.

d. Establish 95% confidence intervals for B_1 and B_2.

e. On the basis of the confidence intervals established in (*d*), can you accept the hypothesis that $B_2 = 0$?

3.10. Based on data for the United States for the period 1965 to 2006 (found in Table 3-4 on the textbook's Web site), the following regression results were obtained:

$$\mathrm{GNP}_t = -995.5183 + 8.7503 M_{1t} \quad r^2 = 0.9488$$
$$\mathrm{se} = (\quad) \qquad (0.3214)$$
$$t = (-3.8258) \qquad (\quad)$$

where GNP is the gross national product (\$, in billions) and M_1 is the money supply (\$, in billions).

Note: M_1 includes currency, demand deposits, traveler's checks, and other checkable deposits.

a. Fill in the blank parentheses.

b. The monetarists maintain that money supply has a significant positive impact on GNP. How would you test this hypothesis?

c. What is the meaning of the negative intercept?

d. Suppose M_1 for 2007 is \$750 billion. What is the mean forecast value of GNP for that year?

3.11. *Political business cycle:* Do economic events affect presidential elections? To test this so-called political business cycle theory, Gary Smith[20] obtained the following regression results based on the U.S. presidential elections for the four yearly periods from 1928 to 1980 (i.e., the data are for years 1928, 1932, etc.):

$$\hat{Y}_t = 53.10 - 1.70 X_t$$
$$t = (34.10)\,(-2.67) \qquad r^2 = 0.37$$

where Y is the percentage of the vote received by the incumbent and X is the unemployment rate change—unemployment rate in an election year minus the unemployment rate in the preceding year.

a. A priori, what is the expected sign of X?

b. Do the results support the political business cycle theory? Support your contention with appropriate calculations.

c. Do the results of the 1984 and 1988 presidential elections support the preceding theory?

d. How would you compute the standard errors of b_1 and b_2?

3.12. To study the relationship between capacity utilization in manufacturing and inflation in the United States, we obtained the data shown in Table 3-5 (found on the textbook's Web site). In this table, Y = inflation rate as measured by the

[20]Gary Smith, *Statistical Reasoning*, Allyn & Bacon, Boston, Mass., 1985, p. 488. Change in notation was made to conform with our format. The original data were obtained by Ray C. Fair, "The Effect of Economic Events on Votes for President," *The Review of Economics and Statistics*, May 1978, pp. 159–173.

3.15. Refer to the S.A.T. data given in Table 2-15 on the textbook's Web site. Suppose you want to predict the male math scores on the basis of the female math scores by running the following regression:

$$Y_t = B_1 + B_2X_t + u_t$$

where Y and X denote the male and female math scores, respectively.

a. Estimate the preceding regression, obtaining the usual summary statistics.
b. Test the hypothesis that there is no relationship between Y and X whatsoever.
c. Suppose the female math score in 2008 is expected to be 490. What is the predicted (average) male math score?
d. Establish a 95% confidence interval for the predicted value in part (*c*).

3.16. Repeat the exercise in Problem 3.15 but let Y and X denote the male and the female critical reading scores, respectively. Assume a female critical reading score for 2008 of 505.

3.17. Consider the following regression results:[22]

$$\hat{Y}_t = -0.17 + 5.26X_t \qquad \bar{R}^2 = 0.10, \text{ Durbin-Watson} = 2.01$$
$$t = (-1.73)(2.71)$$

where Y = the real return on the stock price index from January of the current year to January of the following year
X = the total dividends in the preceding year divided by the stock price index for July of the preceding year
t = time

Note: On Durbin-Watson statistic, see Chapter 10.
The time period covered by the study was 1926 to 1982.
Note: $\bar{R}^2$ stands for the adjusted coefficient of determination. The Durbin-Watson value is a measure of autocorrelation. Both measures are explained in subsequent chapters.

a. How would you interpret the preceding regression?
b. If the previous results are acceptable to you, does that mean the best investment strategy is to invest in the stock market when the dividend/price ratio is high?
c. If you want to know the answer to part (*b*), read Shiller's analysis.

3.18. Refer to Example 2.1 on years of schooling and average hourly earnings. The data for this example are given in Table 2-5 and the regression results are presented in Eq. (2.21). For this regression

a. Obtain the standard errors of the intercept and slope coefficients and r^2.
b. Test the hypothesis that schooling has no effect on average hourly earnings. Which test did you use and why?
c. If you reject the null hypothesis in (*b*), would you also reject the hypothesis that the slope coefficient in Eq. (2.21) is not different from 1? Show the necessary calculations.

[22]See Robert J. Shiller, *Market Volatility*, MIT Press, Cambridge, Mass., 1989, pp. 32–36.

that if family income is zero, the mean math score will be about 432.4138. Very often such an interpretation has no economic meaning. For example, we have no data where an annual family income is zero. *As we will see throughout the book, often the intercept has no particular economic meaning.* In general you have to use common sense in interpreting the intercept term, for very often the sample range of the X values (family income in our example) may not include zero as one of the observed values. *Perhaps it is best to interpret the intercept term as the mean or average effect on Y of all the variables omitted from the regression model.*

Reffer this
Q3-18

2.10 SOME ILLUSTRATIVE EXAMPLES

Now that we have discussed the OLS method and learned how to estimate a PRF, let us provide some concrete applications of regression analysis.

Example 2.1. Years of Schooling and Average Hourly Earnings

Based on a sample of 528 observations, Table 2-5 gives data on average hourly wage $Y(\$)$ and years of schooling (X).

Suppose we want to find out how Y behaves in relation to X. From human capital theories of labor economics, we would expect average wage to increase with years of schooling. That is, we expect a positive relationship between the two variables; it would be bad news if such were not the case.

The regression results based on the data in Table 2-5 are as follows:

$$\hat{Y}_i = -0.0144 + 0.7241X_i \tag{2.21}$$

TABLE 2-5 AVERAGE HOURLY WAGE BY EDUCATION

Years of schooling	Average hourly wage ($)	Number of people
6	4.4567	3
7	5.7700	5
8	5.9787	15
9	7.3317	12
10	7.3182	17
11	6.5844	27
12	7.8182	218
13	7.8351	37
14	11.0223	56
15	10.6738	13
16	10.8361	70
17	13.6150	24
18	13.5310	31

Source: Arthur S. Goldberger, *Introductory Econometrics*, Harvard University Press, Cambridge, Mass., 1998, Table 1.1, p. 5 (adapted).

As these results show, there is a positive association between education and earnings, which accords with prior expectations. For every additional year of schooling, the mean wage rate goes up by about 72 cents per hour.[9] The negative intercept in the present instance has no particular economic meaning.

Example 2.2. Okun's Law

Based on the U.S. data for 1947 to 1960, the late Arthur Okun of the Brookings Institution and a former chairman of the President's Council of Economic Advisers obtained the following regression, known as Okun's law:

$$Y_t = -0.4(X_t - 2.5) \tag{2.22}$$

where Y_t = change in the unemployment rate, percentage points
X_t = percent growth rate in real output, as measured by real GDP
2.5 = the long-term, or trend, rate of growth of output historically observed in the United States

In this regression the intercept is zero and the slope coefficient is −0.4. Okun's law says that for every percentage point of growth in real GDP above 2.5 percent, the unemployment rate declines by 0.4 percentage points.

Okun's law has been used to predict the required growth in real GDP to reduce the unemployment rate by a given percentage point. Thus, a growth rate of 5 percent in real GDP will reduce the unemployment rate by 1 percentage point, or a growth rate of 7.5 percent is required to reduce the unemployment rate by 2 percentage points. In Problem 2.17, which gives comparatively more recent data, you are asked to find out if Okun's law still holds.

This example shows how sometimes a simple (i.e., two-variable) regression model can be used for policy purposes.

Example 2.3. Stock Prices and Interest Rates

Stock prices and interest rates are key economic indicators. Investors in stock markets, individual or institutional, watch very carefully the movements in the interest rates. Since interest rates represent the cost of borrowing money, they have a vast effect on investment and hence on the profitability of a company. Macroeconomic theory would suggest an inverse relationship between stock prices and interest rates.

As a measure of stock prices, let us use the S&P 500 composite index (\$1941–1943 = 10), and as a measure of interest rates, let us use the three-month

[9]Since the data in Table 2-5 refer to the mean wage for the various categories, the slope coefficient here should strictly be interpreted as the average increase in the mean hourly earnings.

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.

4.11. To explain what determines the price of air conditioners, B. T. Ratchford[24] obtained the following regression results based on a sample of 19 air conditioners:

$$\hat{Y}_i = -68.236 + 0.023X_{2i} + 19.729X_{3i} + 7.653X_{4i} \quad R^2 = 0.84$$
$$\text{se} = \qquad (0.005) \qquad (8.992) \qquad (3.082)$$

where Y = the price, in dollars
X_2 = the BTU rating of air conditioner
X_3 = the energy efficiency ratio
X_4 = the number of settings
se = standard errors

a. Interpret the regression results.
b. Do the results make economic sense?
c. At $\alpha = 5\%$, test the hypothesis that the BTU rating has no effect on the price of an air conditioner versus that it has a positive effect.
d. Would you accept the null hypothesis that the three explanatory variables explain a substantial variation in the prices of air conditioners? Show clearly all your calculations.

4.12. Based on the U.S. data for 1965-IQ to 1983-IVQ ($n = 76$), James Doti and Esmael Adibi[25] obtained the following regression to explain personal consumption expenditure (PCE) in the United States.

$$\hat{Y}_t = -10.96 + 0.93X_{2t} - 2.09X_{3t}$$
$$t = (-3.33)\,(249.06) \quad (-3.09) \qquad R^2 = 0.9996$$
$$F = 83{,}753.7$$

where Y = the PCE ($, in billions)
X_2 = the disposable (i.e., after-tax) income ($, in billions)
X_3 = the prime rate (%) charged by banks

a. What is the marginal propensity to consume (MPC)—the amount of additional consumption expenditure from an additional dollar's personal disposable income?
b. Is the MPC statistically different from 1? Show the appropriate testing procedure.
c. What is the rationale for the inclusion of the prime rate variable in the model? A priori, would you expect a negative sign for this variable?
d. Is b_3 significantly different from zero?
e. Test the hypothesis that $R^2 = 0$.
f. Compute the standard error of each coefficient.

[24] B. T. Ratchford, "The Value of Information for Selected Appliances," *Journal of Marketing Research*, vol. 17, 1980, pp. 14–25. Notations were adapted.
[25] James Doti and Esmael Adibi, *Econometric Analysis: An Applications Approach*, Prentice-Hall, Englewood Cliffs, N.J., 1988, p. 188. Notations were adapted.

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.