ıple of the use of these formulas, consider again the data in Table 3.2. The :e computations necessary for finding b_0 and b_1 are shown in Figure 3.5. b_1, can now be computed using the formula in Equation (3.4):

$$b_1 = \frac{196 - \frac{1}{6}(21)(45)}{91 - \frac{1}{6}(21)^2} = \frac{38.5}{17.5} = 2.2$$

:pt, b_0, is computed as in Equation (3.3):

$$b_0 = 7.5 - 2.2(3.5) = -0.2$$

$$\bar{x} = \frac{21}{6} = 3.5 \text{ and } \bar{y} = \frac{45}{6} = 7.5$$

ıquares regression line for these data is

$$\hat{y} = -0.2 + 2.2x$$

is no longer any guesswork associated with computing the best-fitting ı criterion has been stated that defines "best." Using the criterion of mini- of squared errors, the regression line we computed provides the best de- f the relationship between the variables x and y. Any other values used for :esult in larger sum of squared errors. For example, Figure 3.6(a) shows ıtation of the sum of squared errors for the original "guessed" line · $2.5x$, and Figure 3.6(b) shows the same computation for the least- ıe $\hat{y} = -0.2 + 2.2x$. As expected, the sum of squared errors for the line rger than that for the least-squares line.

x_i	y_i	$x_i y_i$	x_i^2
1	3	3	1
2	2	4	4
3	8	24	9
4	8	32	16
5	11	55	25
6	13	78	36
21	45	196	91

FIGURE 3.6

(a) Computation of Sum of Squared Errors for: $\hat{y} = -1 + 2.5x$

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
1	3	1.5	1.5	2.25
2	2	4.0	−2.0	4.00
3	8	6.5	1.5	2.25
4	8	9.0	−1.0	1.00
5	11	11.5	−0.5	0.25
6	13	14.0	−1.0	1.00

$$\sum_{i=1}^{6}(y_i - \hat{y}_i)^2 = 10.75$$

(b) Computation of Sum of Squared Errors for: $\hat{y} = -0.2 + 2.2x$

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
1	3	2.0	1.0	1.00
2	2	4.2	−2.2	4.84
3	8	6.4	1.6	2.56
4	8	8.6	−0.6	0.36
5	11	10.8	0.2	0.04
6	13	13.0	0.0	0.00

$$\sum_{i=1}^{6}(y_i - \hat{y}_i)^2 \doteq 8.80$$

EXERCISES

Exercises 1 and 2 should be done by hand.

1. **Flexible Budgeting.** A budget is an expression of management's expectations and goals concerning future revenues and costs. To increase their effectiveness, many budgets are flexible, including allowances for the effect of variation in uncontrolled variables. For example, the costs and revenues of many production plants are greatly affected by the number of units produced by the plant during the budget period, and this may be beyond a plant manager's control. Standard cost-accounting procedures can be used to adjust the direct-cost parts of the budget for the level of production, but it is often more difficult to handle overhead. In many cases, statistical methods are used to estimate the relationship between overhead (y) and the level of production (x) using historical data. As a simple example, consider the historical data for a certain plant:

Production (in 10,000) units:	5	6	7	8	9	10	11
Overhead costs (in \$1000):	12	11.5	14	15	15.4	15.3	17.5

   a. Construct a scatterplot of y versus x.
   b. Find the least-squares regression line relating overhead costs to production.
   c. Graph the regression line on the scatterplot.

2. **Central Company.** The Central Company manufactures a certain specialty item once a month in a batch production run. The number of items produced in each run varies from month to month as demand fluctuates. The company is interested in the relationship between the size of the production run (x) and the number of hours of labor (y) required for the run. The company has collected the following data for the ten most recent runs:

Number of items:	40	30	70	90	50	60	70	40	80	70
Labor (hours):	83	60	138	180	97	118	140	75	159	144

a. Construct a scatterplot of y versus x.

b. Find the least-squares line relating hours of labor to number of items produced.

c. Graph the regression line on the scatterplot.

3.2 EXAMPLES OF REGRESSION AS A DESCRIPTIVE TECHNIQUE

EXAMPLE 3.2 **Pricing Communications Nodes** In recent years the growth of data communications networks has been amazing. The convenience and capabilities afforded by such networks are appealing to businesses with locations scattered throughout the United States and the world. Using networks allows centralization of an information system 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 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 3.3 and a scatterplot of cost (y = COST) versus number of ports (x = NUMPORTS) is shown in Figure 3.7. (See the file COMNODE3 on the CD.)

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 was computed to be

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

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. ■

TABLE 3.3 Cost and Number of Ports for Communications Nodes Example*

COST	NUMPORTS
52,388	68
51,761	52
50,221	44
36,095	32
27,500	16
57,088	56
54,475	56
33,969	28
31,309	24
23,444	24
24,269	12
53,479	52
33,543	20
33,056	24

*These data have been modified as requested by the company to provide confidentiality.

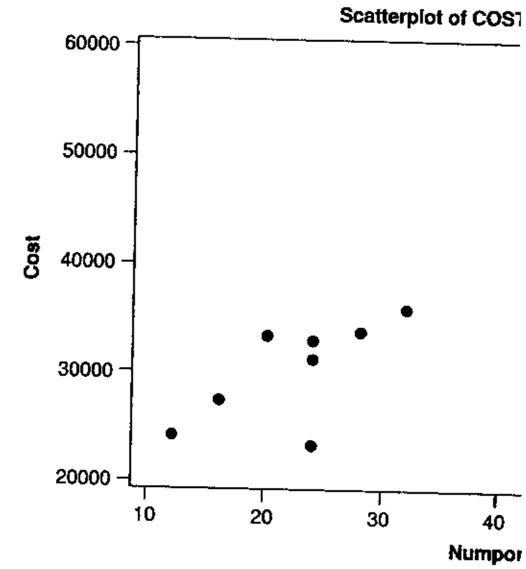

FIGURE 3.7 Scatterplot of Cost Versus Number of Ports for the Communications Nodes Example.

EXAMPLE 3.3 **Estimating Residential Real Estate Values** The must appraise properties for the entire county. The square footage of the individual houses as well as l condition of an entire neighborhood to derive indivic This avoids labor-intensive inspections each year.

Regression can be used to establish the weigh in assessing values. For example, Table 3.4 show for a sample of 100 Tarrant County homes (these of value (y = VALUE) versus size (x = SIZE) is REALEST3 on the CD.)

Using a statistical package, the regression equ determined as

$$\text{VALUE} = -50{,}035 + 7$$

If size were the only factor thought to be of im this equation could be used by the appraisal distric need to be considered. Developing an equation that tant factor (explanatory variable) is discussed in Ch

EXAMPLE 3.4 **Forecasting Housing Starts** Forecasts of va important to the U.S. government and to various i States. The construction industry is concerned with a given year. Accurate forecasts can help with plans the industry.

Answer:

(a) Using the standardized test statistic:

Decision rule: Reject H_0 if $F > 4.75$
Do not reject H_0 if $F \leq 4.75$
Test Statistic: $F = 94.41$
Decision: Reject H_0
Conclusion: There is evidence to conclude that COST and NUMPORTS are linearly related.

(b) Using the p value:

Decision rule: Reject H_0 if p value < 0.05
Do not reject H_0 if p value ≥ 0.05
Test Statistic: p value $= 0.000$
Decision: Reject H_0

EXERCISES

Exercises 8 and 9 should be done by hand.

8. Flexible Budgeting (continued) Refer to Exercise 1.

a. Compute the coefficient of determination (R^2) for the regression of overhead costs on production.

b. What percentage of the variation in overhead costs has been explained by the regression?

c. Use the F test to test the hypotheses H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ at the 5% level of significance. Be sure to state the decision rule, the test statistic value, and your decision.

d. From the result in part c, are production and overhead costs linearly related?

9. Central Company (continued) Refer to Exercise 2.

a. Compute the coefficient of determination (R^2) for the regression of the number of labor hours on number of items produced.

b. What percentage of the variation in hours of labor has been explained by the regression?

c. Use the F test to test the hypotheses H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ at the 5% level of significance. Be sure to state the decision rule, the test statistic value, and your decision.

d. From the result in part c, are hours of labor and number of items produced linearly related?

10. Dividends (continued) Use the regression results in Figure 3.20 to help answer the questions.

a. What percentage of the variation in dividend yield has been explained by the regression?

b. Use the F test to test the hypotheses H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ at the 5% level of significance. Be sure to state the decision rule, the test statistic value, and your decision.

11. Sales/Advertising (continued) Use the regression results in Figure 3.22 to help answer the questions.

a. What percentage of the variation in sales has been explained by the regression?

b. Use the F test to test the hypotheses H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ at the 5% level of significance. Be sure to state the decision rule, the test statistic value, and your decision.

3.5 PREDICTION OR FORECASTING WITH A SIMPLI LINEAR REGRESSION EQUATION

One of the possible goals for fitting a regression the regression equation to predict or forecast val Given that a value of x has been observed, what sponse value, y? To discuss how to best predict using predictions based on a random sample, two are considered.

3.5.1 ESTIMATING THE CONDITIONAL MEAN OF y GIVEN x

In Example 3.6, suppose that the network administ ble nodes with 40 communications ports. The qu will the cost be, on average, for nodes with 40 port

The average cost of all nodes with 40 commur Thus. an estimate of the conditional mean cost $\mu_{y|x=40}$, is required (see Figure 3.27). If a relation the best estimate of this point on the population reg

$$\hat{y}_m = b_0 + b_1 x_m$$

where b_0 and b_1 are the least-squares estimates of ports for which an estimate is desired, and $\hat{y}_m$ is the cost. In this case, $\hat{y}_m$ represents the estimate of the p sponding to (or conditional on) $x = x_m$. Thus, it mean. The variance of this estimate can be shown t

$$\sigma_m^2 = \sigma_e^2\left(\frac{1}{n} + \frac{(x_m -}{(n - 1}\right.$$

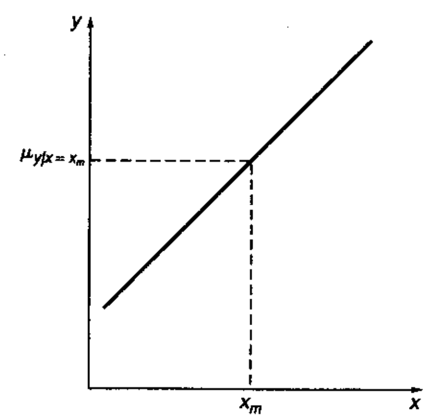

FIGURE 3.27
Estimating a Conditional Mean $\mu_{y|x=x_m}$.

ed Values for New Observations

Fit	SE Fit	95% CI	95% PI
42600	1178	(40035, 45166)	(32872, 52329)

of Predictors for New Observations

NUMPORTS
40.0

Output Statistics

cted lue	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
305	2414	55545	66065	50047	71563	-8417
402	1559	47006	53799	40423	60382	1359
201	1262	42452	47950	35423	54979	5020
399	1186	34814	39984	27666	47132	-1304
996	1780	23119	30874	16843	37150	503.6461
003	1751	49189	56818	42873	63133	4085
003	1751	49189	56818	42873	63133	1472
798	1278	32015	37582	25010	44587	-829.3840
198	1414	29116	35280	22321	42075	-888.7073
198	1414	29116	35280	22321	42075	-8754
396	1991	20057	28735	14057	34734	-126.6772
402	1559	47006	53799	40423	60382	3077
597	1585	26143	33051	19598	39596	3946
198	1414	29116	35280	22321	42075	858.2927
500	1178	40035	45166	32872	52329	.

siduals 0
uared Residuals 222594146
Residual SS (PRESS) 345066019

tion Output for Example 3.10.

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

wer: The point prediction is again $42,600. A 95% prediction interval he individual node is ($32,872, $52,329). Note that the prediction inters considerably wider than the confidence interval, reflecting the addial uncertainty of predicting for an individual as opposed to estimating an age. ■

EXERCISES

Exercises 12 and 13 should be done by hand.

12. Flexible Budgeting (continued), Refer to Exercise 1.

a. Find a point estimate of the overhead costs, on average, for production runs of 80,000 units.

b. Find a 95% confidence interval estimate of overhead costs, on average, for production runs of 80,000 units.

c. Find a point prediction of the overhead costs for a single production run of 80,000 units.

d. Find a 95% prediction interval for overhead costs for a single production run of 80,000 units.

e. State why the prediction interval is wider than the confidence interval.

13. Central Company (continued), Refer to Exercise 2.

a. Find a point estimate for the number of hours of labor required, on average, when 60 units are produced.

b. Find a 95% confidence interval estimate of hours of labor required, on average, when 60 units are produced.

c. Find a point prediction of the number of hours of labor required for one run producing 60 units.

d. Find a 95% prediction interval for the number of hours of labor required for one run producing 60 units.

14. Dividends (continued), Consider the dividend-yield problem in Exercise 6 and the associated computer results in Figure 3.20. An analyst wants an estimate of dividend yield for all firms with earnings per share of $3. Does the equation developed provide a more accurate estimate than simply using the sample mean dividend yield for all 42 firms examined? State why or why not.

15. Sales/Advertising (continued), Use the results in Figure 3.33 to help solve these problems. These results were obtained requesting a prediction with $x = 200$, 250, 300 and 350, respectively (representing $20,000, $25,000, $30,000 and $35,000).

a. Find an estimate of average sales for all sales districts with advertising expenditures of $25,000. Find a point estimate and a 95% confidence interval estimate.

b. Predict sales for individual districts having advertising expenditures of $20,000, $25,000, $30,000, and $35,000. Find point predictions as well as 95% prediction intervals.

Predicted Value	SE Fit	95% Confidence Int.	95% Prediction Int.
3457	211	(3013, 3900)	(2131, 4783)
4335	156	(4007, 4663)	(3043, 5627)
5214	133	(4934, 5493)	(3933, 6494)
6092	157	(5763, 6421)	(4800, 7384)

FIGURE 3.33
Prediction Output for Exercise 15.

3.6 FITTING A LINEAR TREND TO TIME-SERIES DATA

Data gathered on individuals at the same point in time are called *cross-sectional data. Time-series data* are data gathered on a single individual (person, firm, and so on) over a sequence of time periods, which may be days, weeks, months, quarters, years, or virtually any other measure of time. In a given problem, however, it is assumed that the data are gathered over only one interval of time (daily and weekly data are not combined, for example).

When dealing with time-series data, the primary goal often is to be able to produce forecasts of the dependent variable for future time periods. Two separate approaches to this problem can be identified. On the one hand, a researcher may

nt. In 1990, the city
d a study examin-
ne aspect helpful in
nitoring the quality
our into the Trinity
ing water for Fort
n the river is filtered
from entering the
ful in maintaining
city monitored to
e river from storm

ory test
termined by com-
r sample

hydrocarbon SHEEN: hydrocarbon (oil) sheen on surface of water

sewage BACTERIA: filamentous sewage bacteria

These are monthly data from January 1986 through December 1989. In all cases, lower numbers are better.

These data are in a file named WATER3 on the CD.

Your job is to use time-series plots and linear trend regression to examine the performance of the city's water department in improving the quality of storm drain water entering the Trinity River. Which of the variables show a significant decrease? Are there areas where the city might concentrate its efforts to achieve future improvements? Use a 5% level of significance in any tests.

IN INTERPRETING REGRESSION RESULTS

RSUS CAUSALITY

n mistake made when using regression analysis is to assume that a strong ^{2}) of a regression of y on x automatically means that "x causes y." This is sarily true. Some alternative explanations for the good fit include:

everse is true; y causes x. Linear regression computations pay no attention direction of causality. If x and y are highly correlated, a high R^2 value ts even if the causal order of the variables is reversed.

e may be a third variable related to both x and y. It may be that neither x s y nor y causes x. Both variables may be related to some third common . As an example, consider the price and gasoline mileage of automobiles. e two variables are inversely related. As mileage rises, price goes down verage). But it is not the rise in mileage that "causes" the price to drop. A variable, size of car, may be influencing both of the other two variables. ize increases, price increases and mileage drops. There are a variety of esting examples in this category. For example, the mortality rate in coun- is inversely related to the number of televisions. As the number of televi- increases, mortality rate decreases. I don't think this is a causal ionship.

nfer that x causes y requires that additional conditions be satisfied. A high egression of y on x might be considered supporting evidence for causality, own, this is not enough to ensure that x causes y.
that the absence of causality is not necessarily a drawback in regression An equation showing a relationship between x and y can be important and en if it is recognized that x does not cause y.

3.7.2 FORECASTING OUTSIDE THE RANGE OF THE EXPLANATORY VARIABLE

When using an estimated regression equation to construct estimates of $\mu_{y|x}$ or to predict individual values of the dependent variable, some caution must be used if forecasts are outside the range of the x variable. Consider the communications nodes example. The explanatory variable was NUMPORTS, the number of access ports. The sample values ranged from 12 to 68. The estimated regression model can be expected to be reliable over this range of the x variable. If, however, a node is to be installed with 100 ports, there is some question as to how reliable the model will be. The relationship that holds over the range from 12 to 68 may differ from the relationship outside this range. Estimates of $\mu_{y|x}$ or predictions outside the range of the x variable require some caution for this reason.

There are often occasions where forecasts outside the range of the x variable must be made. One common example is when time-series data are used and forecasts for future time periods are desired. It may be that the values of the explanatory variables in future time periods are outside the range observed in the past, as, for example, when the linear trend model is used. In such cases, it must be recognized that the quality of the forecasts depends on whether the estimated relationship still holds for values of the explanatory variables that are outside the observed range.

EXERCISES

17. **Sales/Advertising (continued).** Use the results in Figure 3.22 to help answer the following questions.
   a. Find a point estimate of average sales for all sales districts with advertising expenditures of \$60,000. Are there any cautions that should be exercised regarding this estimate?
   b. A district sales manager examines the model developed. The manager points out that \$0 advertising expenditure results in sales of −\$5700, which is impossible. She suggests that this means the model is of no use. Do you agree or disagree with her assessment? Explain why.

ADDITIONAL EXERCISES

18. Indicate whether the following statements are true or false:
   a. If the hypothesis $H_0: \beta_1 = 0$ is rejected, then it can be safely concluded that x causes y.
   b. Suppose a regression of y on x is run and the t statistic for testing $H_0: \beta_1 = 0$ versus $H_a: \beta_1 \neq 0$ has a p value of 0.0295 associated with it. Using a 5% level of significance, the null hypothesis should be rejected.
   c. If the correlation between y and x is 0.9, then the R^2 value for a regression of y on x is 90%.
   d. As long as the R^2 value is high for an estimated regression equation, it is safe to use the equation to predict for any value of x.
   e. If the R^2 value for a regression of y on x is 75%, then the R^2 value for a regression of x on y is also 75%.

19. Suppose a regression analysis provides the following results:

$$b_0 = 1, \quad b_1 = 2, \quad s_{b_0} = 0.05,$$
$$s_{b_1} = 0.25, \; SST = 117.2873, \; SSE = 30.0$$

and $n = 24$. Use this information to solve the following problems.
   a. Test the hypotheses

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

using a 5% level of significance. State the decision rule, the test statistic, and your decision. Use a t test.

b. Perform the same test as in part a using an F test. Use a 10% level of significance.

c. Compute the R^2 for the regression.

20. Suppose a regression analysis provides the following results:

$$b_0 = 4.0, \quad b_1 = 10.0, \quad s_{b_0} = 1.0,$$
$$s_{b_1} = 4.0, \quad SST = 67.36, \quad SSE = 50.0$$

and $n = 20$. Use this information to solve the following problems:

a. Test the hypotheses

$$H_0: \beta_0 = 0$$
$$H_a: \beta_0 \neq 0$$

using a 5% level of significance. State the decision rule, the test statistic, and your decision.

b. Test the hypotheses

$$H_0: \beta_1 \leq 0$$
$$H_a: \beta_1 > 0$$

using a 5% level of significance. State the decision rule, the test statistic, and your decision. What conclusion can be drawn from the test result?

c. Compute the R^2 for the regression.

21. Fill in the missing blanks on the following ANOVA table:

ANOVA Source	DF	SS	MS	F
Regression	1		1000	
Error (Residual)		800		
Total	81			

22. Fill in the missing blanks on the following ANOVA table:

ANOVA Source	DF	SS	MS	F
Regression			100	10
Error (Residual)	40			
Total				

23. **Salary/Education.** Data on beginning salary ($y = SALARY$) and years of education ($x = EDUC$) for 93 employees of Harris Bank Chicago in 1977 are provided in a data file named SALED3 on the CD. These data were obtained from an article by Daniel W. Schafer, "Measurement-Error Diagnostics and the Sex Discrimination Problem," *Journal of Business and Economic Statistics*, 5: 529–537, 1987. (Copyright 1987 by the American Statistical Association. Used with permission. All rights reserved.)

The scatterplot of salary verses education is shown in Figure 3.39. The regression results are shown in Figure 3.40. Use the results to answer the following questions:

a. Is there a linear relationship between salary and education? State the hypotheses to be tested, the decision rule, the test statistic, and your decision. Use a 10% level of significance.

b. What percentage of the variation in salary has been explained by the regression?

c. For an individual with 12 years of education, find a point prediction of beginning salary.

d. For all individuals with 12 years of education, find a point estimate of the conditional mean beginning salary.

e. What other factors, in addition to education, might be useful in helping to estimate beginning salary?

24. **Cost Estimation.** The file COSTEST3 on the CD contains data on production runs at a manufacturing plant. There are two columns of data:

y = COST is the total cost of the production run.

x = NUMBER is the number of items produced during that run.

Run the regression using COST as the dependent variable and NUMBER as the independent variable and use the result to help answer the following questions:

a. What is the estimated regression equation relating y to x?

b. What percentage of the variation in y has been explained by the regression?

c. Are y and x linearly related? Conduct a hypothesis test to answer this question and use a 5% level of significance. State the hypotheses to be tested, the decision rule, the test statistic, and your decision. What conclusion can be drawn from the result of the test?

d. Estimate the fixed cost involved in the production process. Find a point estimate and a 95% confidence interval estimate.

e. Estimate the variable cost involved in the production process. Find a point estimate and a 95% confidence interval estimate.

25. **Income/Consumption.** The following data are annual disposable income and total annual consumption for 12 families selected at random from a large metropolitan area. Regard annual disposable income as the explanatory variable and total

annual con
From the r
tions that f
INCONS3

Annua
In

FIGURE 3.39
Scatterplot for Salary and Education Exercise.

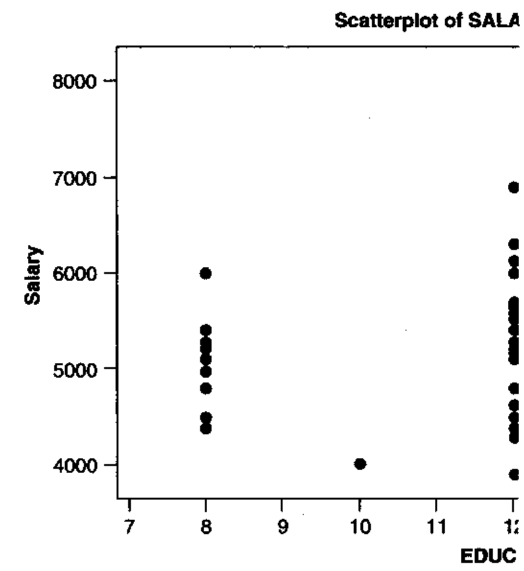

FIGURE 3.40
Regression Results for Salary and Education Exercise.

Variable	Coefficient	Std Dev	T Stat
Intercept	3818.6	377.4	10.12
EDUC	128.1	29.7	4.31

Standard Error = 650.112   R-Sq = 17.0%   R-S

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Regression	1	7862534	7862534
Error	91	38460756	422646
Total	92	46323290	

c. What percentage of the variation in y has been explained by the regression?

d. Construct a 95% confidence interval estimate of β_1.

(*Source*: Data are from D. A. Bessler and R. A. Babubla, "Forecasting Wheat Exports: Do Exchange Rates Really Matter?" *Journal of Business and Economic Statistics*, 5, 1987, pp. 397–406. Copyright 1987 by the American Statistical Association. Used with permission. All rights reserved.)

30. Major League Baseball Salaries. The owners of Major League Baseball (MLB) teams are concerned with rising salaries (as are owners of all professional sports teams). Table 3.9 provides the average salary (AVESAL) of the 30 MLB teams for the 2002 season. Also provided is the number of wins (WINS) for each team during the 2002 season. Is there evidence that teams with higher total payrolls tend to be more successful? Justify your answer. These data are available in a file named BBALL3 on the CD.

31. Computing Beta. In finance class you will discuss (or have discussed) the use of simple regression to estimate the relationship between the return on a stock and the market return. This relationship can be written as

$$y = \beta_0 + \beta_1 x + e$$

where y = return on the stock and x = the return on the market. The slope coefficient, β_1, is called the *beta coefficient* and is used to measure how responsive a stock's price is to movements in the market. The beta coefficient is used as a measure of a firm's systematic risk. In the file named BETA3 on the CD, the return on the stock of three companies is provided: Dell, Sabre, and Wal-Mart. Also provided is the return on the market (This is the value-weighted return computed by CRSP, the Center for Research on Security Prices). Five year's of monthly returns (January 1998 through December 2002) are used so there are a total of 60 observations for each company. Run the regression using the firm's return as the dependent variable and the market return as the independent variable for each of the three companies. Use the three regression results to answer the following questions:

a. What are the beta coefficients for each of the three companies?

b. Is there a relationship between the firm return and the market return for each of these three companies? Be sure to state the hypotheses to be tested, the decision rule, the test statistic, and your decision. Use a 5% level of significance.

c. The beta coefficient measures a security's responsiveness to movements in the market. For example, a beta of 2 would mean that a 1% increase (decrease) in the market return would result in, on average, a 2% increase (decrease) in the security's return. A beta of 1 would mean that movements in the market were matched, on average, by movements in the security's return. For each of the companies in the data file, test to see if the beta coefficient is equal to one or not. Be sure to state the decision rule, the test statistic, and your decision. Use a 5% level of significance.

d. Test to see if Dell's beta coefficient is greater than 1. Be sure to state the decision rule, the test statistic, and your decision. Use a 5% level of significance.

e. Test to see if Wal-Mart's beta coefficient is less than 1. Be sure to state the decision rule, the test statistic, and your decision. Use a 5% level of significance.

32. Major League Baseball Wins. What factor is most important in building a winning baseball team? Some might argue for a high batting average. Or it might be a team that hits for power as measured by the number of home runs. On the other hand, many believe that it is quality pitching as measured by the earned run average of the team's pitchers. The file MLB3 on the CD contains data on the following variables for the 30 major league baseball teams during the 2002 season:

WINS = number of games won
HR = number of home runs hit
BA = average batting average
ERA = earned run average

Using WINS as the dependent variable, use scatterplots and regression to investigate the relationship of the other three variables to WINS. Which of the three possible explanatory variables exhibits the strongest relationship to WINS? What might this suggest to managers of major league baseball teams?
(*Source*: Data courtesy of the *Fort Worth Star-Telegram.*)

TABLE 3.9 Data for Major League Baseball Salaries Exercise

Team	WINS	AVESAL	Team
Anaheim Angels	99	2160054	Atlanta Braves
Baltimore Orioles	67	1855318	Chicago Cubs
Boston Red Sox	93	3633457	Cincinnati Reds
Chicago White Sox	81	1791286	Colorado Rockies
Cleveland Indians	74	2106591	Florida Marlins
Detroit Tigers	55	1562847	Houston Astros
Kansas City Royals	62	1832594	Los Angeles Dodgers
Minnesota Twins	94	1430068	Milwaukee Brewers
New York Yankees	103	4902777	Montreal Expos
Oakland Athletics	103	1746264	New York Mets
Seattle Mariners	93	3337435	Philadelphia Phillies
Tampa Bay Devil Rays	55	1131474	Pittsburgh Pirates
Texas Rangers	72	3123803	St. Louis Cardinals
Toronto Blue Jays	78	1868356	San Diego Padres
Arizona Diamondbacks	98	3199608	San Francisco Giants

Source: Reprinted courtesy of the *Fort Worth Star-Telegram*.

33. Work Orders. During the construction phase of a nuclear plant, the number of corrective work orders open should gradually decline until reaching a steady state that would be present during the operational phase. The Nuclear Regulatory Commission has licensing requirements that the number of work orders open at licensing and for operational plants be less than 1000. (This was, of course, back in the days when nuclear plants were still being constructed in the United States.) This number is set to provide a goal indicating operational readiness. The number of work orders for a consecutive 120-working-day period during the construction phase of a nuclear plant are available in a file named WKORDER3 on the CD.

As a consultant to the plant, you have been asked to estimate how many days it will take to reach the operational level of 1000 work orders. In determining the number of days, state any assumptions you make and any caveats that might be in order.

34. Fanfare. Fanfare International, Inc., designs, distributes, and markets ceiling fans and lighting fixtures. The company's product line includes 120 basic models o fan light kits a marketed to ov electrical whol and new constr distributed by dent sales repr

In the summer to develop fore mine future sal tures, and so o contains mont additional vari through May 1 follows

SALES =

ADEX =

MTGRAT

HSSTAR

The data file c shown, plus c sales data hav confidentiality