Ch. 11 Simple Linear Regression

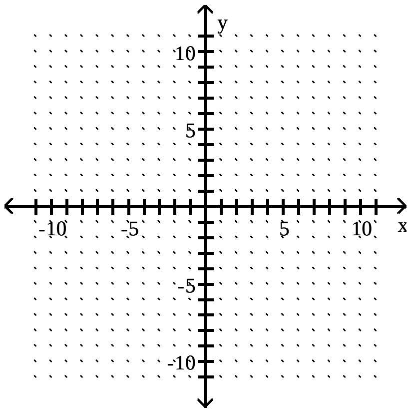
11.1 Probabilistic Models

1 Graph Line

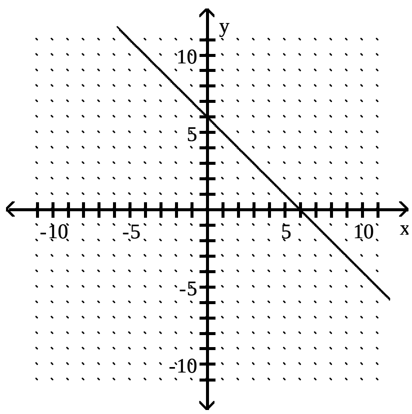
MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Graph the line that passes through the given points.

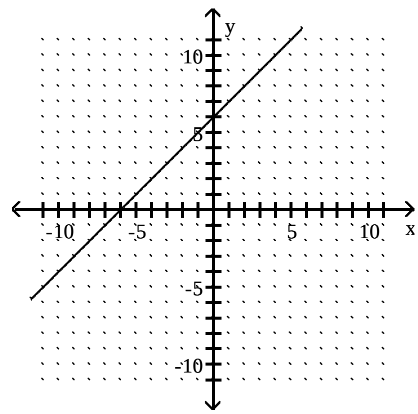
1) (0, 6) and (6, 0)



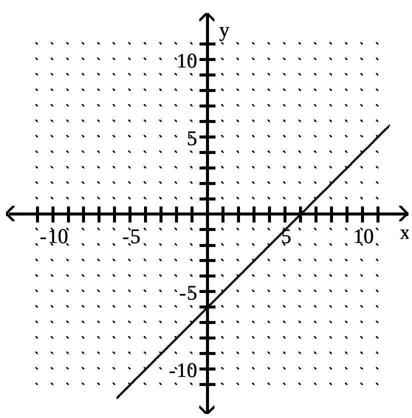
A)



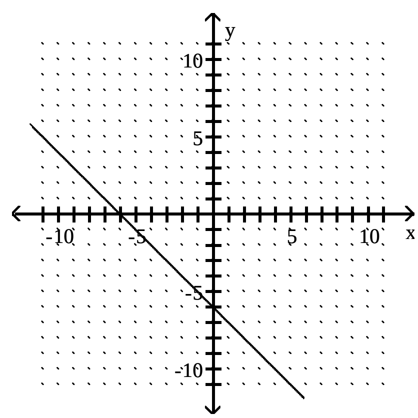
B)



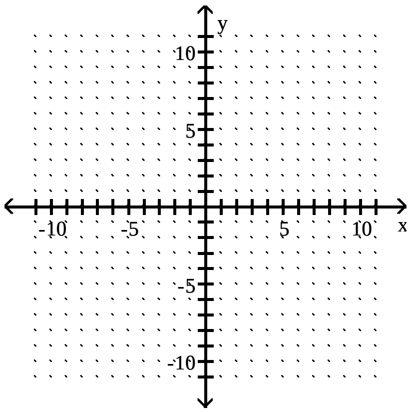
C)



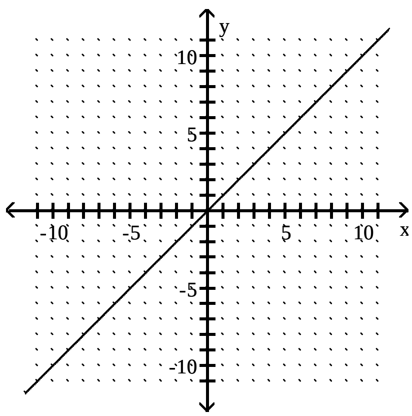
D)



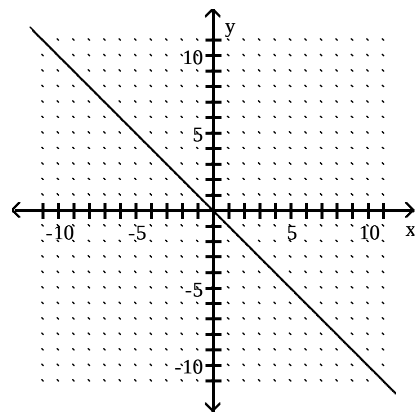
2) $(-8, -8)$ and $(4, 4)$



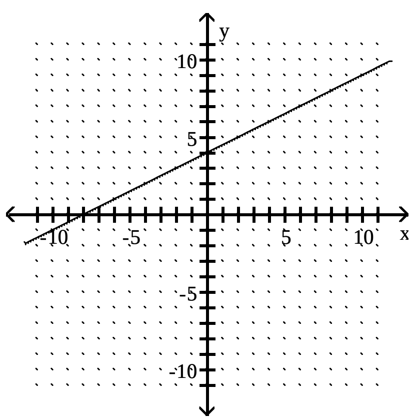
A)



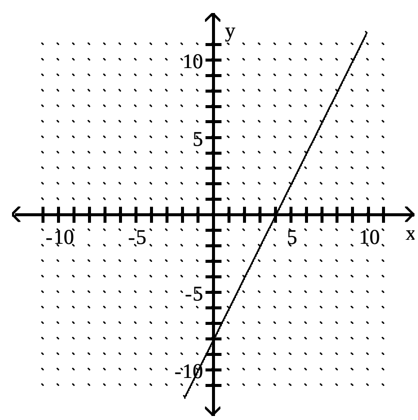
B)



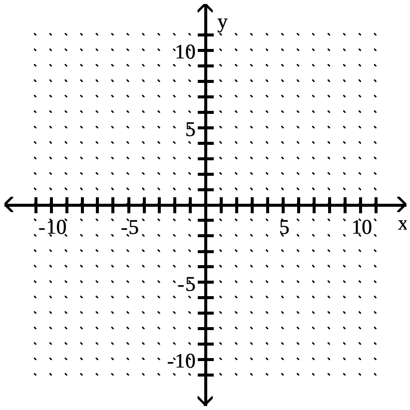
C)



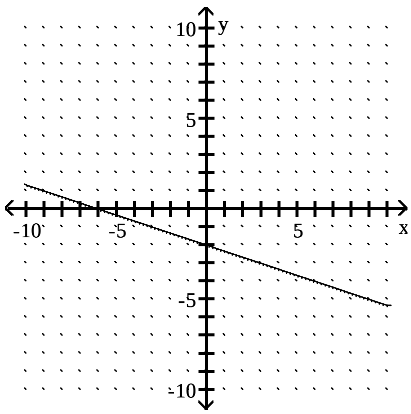
D)



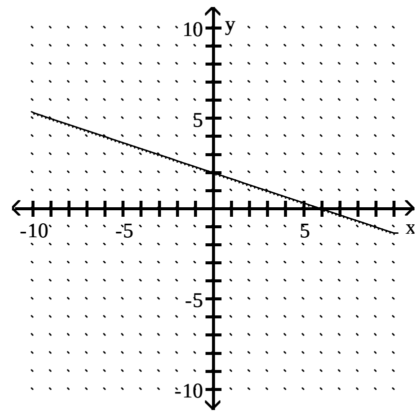
3) (–6, 0) and (–3, –1)



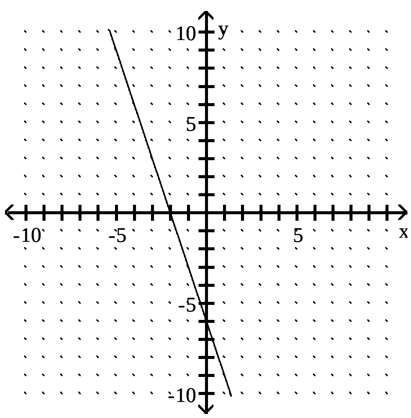
A)



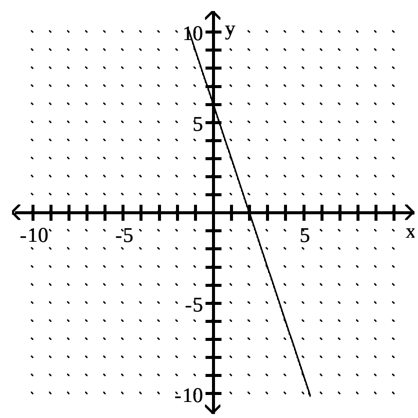
B)



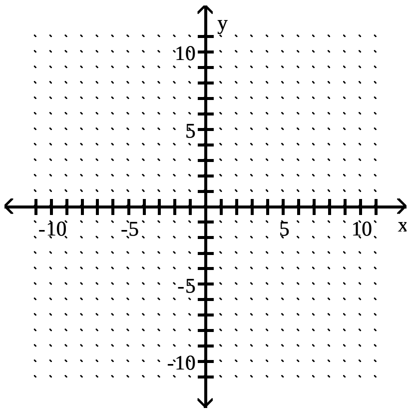
C)



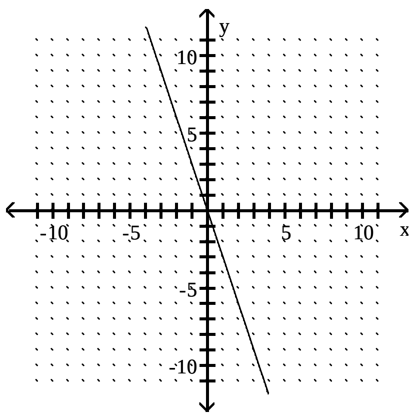
D)



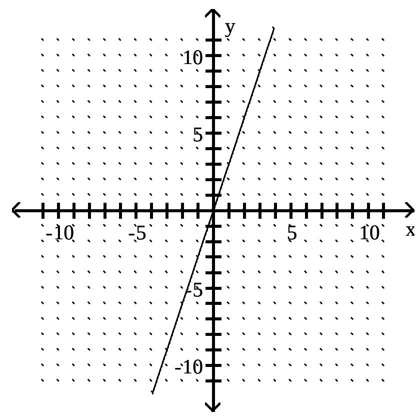
4) (2, –6) and (–1, 3)



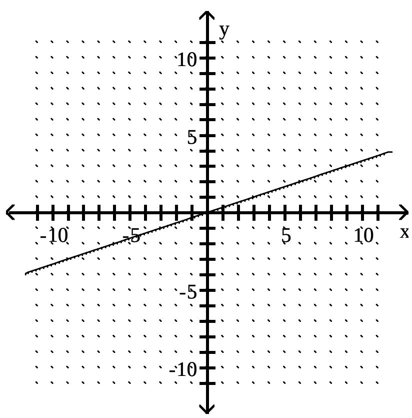
A)



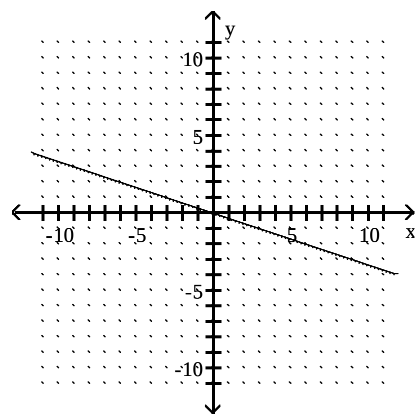
B)



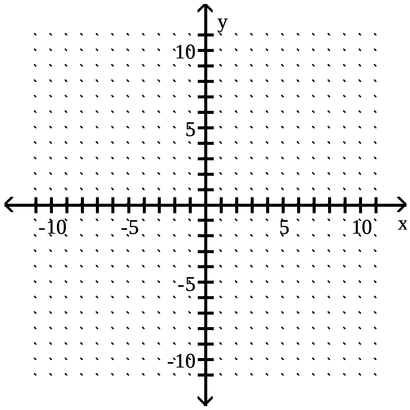
C)



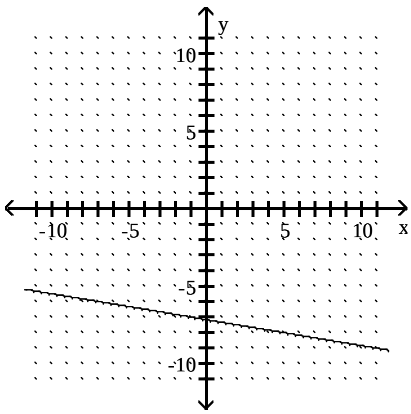
D)



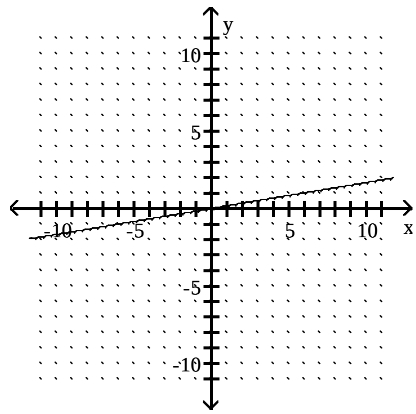
5) (–7, –6) and (–1, –7)



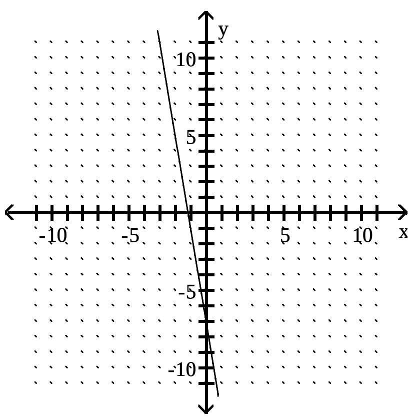
A)



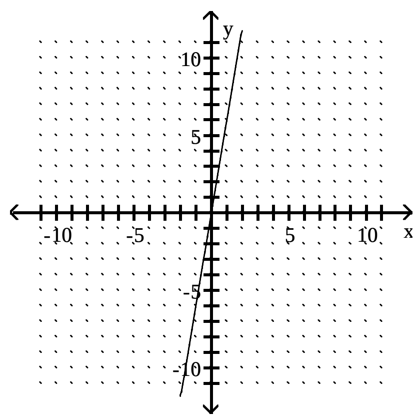
B)



C)



D)



SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

6) Plot the line $y = 3x$. Then give the slope and y–intercept of the line.

7) Plot the line $y = 4 - 2x$. Then give the slope and y–intercept of the line.

8) Plot the line $y = 1.5 + .5x$. Then give the slope and y–intercept of the line.

9) The equation for a (deterministic) straight line is $y = \beta_0 + \beta_1 x$. If the line passes through the points (1, 6) and (8, 9), find the values of β_0 and β_1, respectively.

2 Understand Probabilistic Models

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

10) A county real estate appraiser wants to develop a statistical model to predict the appraised value of houses in a section of the county called East Meadow. One of the many variables thought to be an important predictor of appraised value is the number of rooms in the house. Consequently, the appraiser decided to fit the linear regression model:

$$E(y) = \beta_0 + \beta_1 x,$$

where y = appraised value of the house (in thousands of dollars) and x = number of rooms. Using data collected for a sample of $n = 74$ houses in East Meadow, the following result was obtained:

$$\hat{y} = 74.80 + 19.72x$$

Which of the following statements concerning the deterministic model, $E(y) = \beta_0 + \beta_1 x$ is true?

A) In theory, a plot of the mean appraised value $E(y)$ against the number of rooms x for the entire population of houses in east Meadow would result in a straight line.
B) In theory, if the appraised values y and number of rooms x for the entire population of houses in East Meadow were obtained and the (x, y) data points plotted, the points would fall in a straight line.
C) A plot of the predicted appraised values $\hat{y}$ against the number of rooms x for the sample of houses in East Meadow would not result in a straight line.
D) All of the above statements are true.

11) Is there a relationship between the raises administrators at County University receive and their performance on the job?

A faculty group wants to determine whether job rating (x) is a useful linear predictor of raise (y). Consequently, the group considered the linear regression model

$$E(y) = \beta_0 + \beta_1 x.$$

The faculty group obtained the following prediction equation:

$$\hat{y} = 14{,}000 - 2{,}000x$$

Which of the following statements about the model $E(y) = \beta_0 + \beta_1 x$ is correct?

A) The model hypothesizes a line of means; as rating (x) increases, the mean raise $E(y)$ moves up or down along a straight line.
B) The model hypothesizes that the raises for the administrators fall in a perfect straight line.
C) The model hypothesizes that knowing an administrator's rating (x) will determine exactly the administrator's raise (y).
D) The model hypothesizes that, on average, administrators make more money than professors.

Answer the question True or False.

12) The probabilistic model allows the $E(y)$ values to fall around the regression line while the actual values of y must fall on the line.

A) True B) False

11.2 Fitting the Model: The Least Squares Approach

1 Perform Calculations and Find Least Squares Line

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

1) Consider the data set shown below. Find the estimate of the slope of the least squares regression line.

y	0	3	2	3	8	10	11
x	-2	0	2	4	6	8	10

A) 1.5
B) 0.94643
C) 1.49045
D) 0.9003

2) Consider the data set shown below. Find the estimate of the y-intercept of the least squares regression line.

y	0	3	2	3	8	10	11
x	-2	0	2	4	6	8	10

A) 1.5
B) 0.94643
C) 1.49045
D) 0.9003

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

3) In a comprehensive road test for new car models, one variable measured is the time it takes the car to accelerate from 0 to 60 miles per hour. To model acceleration time, a regression analysis is conducted on a random sample of 129 new cars.

TIME60: y = Elapsed time (in seconds) from 0 mph to 60 mph
MAX: x = Maximum speed attained (miles per hour)

The simple linear model $E(y) = \beta_0 + \beta_1 x$ was fit to the data. Computer printouts for the analysis are given below:

NWEIGHTED LEAST SQUARES LINEAR REGRESSION OF TIME60

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	18.7171	0.63708	29.38	0.0000
MAX	-0.08365	0.00491	-17.05	0.0000

R-SQUARED	0.6960	RESID. MEAN SQUARE (MSE)	1.28695
ADJUSTED R-SQUARED	0.6937	STANDARD DEVIATION	1.13444

SOURCE	DF	SS	MS	F	P
REGRESSION	1	374.285	374.285	290.83	0.0000
RESIDUAL	127	163.443	1.28695		
TOTAL	128	537.728			

CASES INCLUDED 129   MISSING CASES 0

Find and interpret the estimate $\hat{\beta}_1$ in the printout above.

4) Is the number of games won by a major league baseball team in a season related to the team's batting average? Data from 14 teams were collected and the summary statistics yield:

$$\sum y = 1{,}134, \ \sum x = 3.642, \ \sum y^2 = 93{,}110, \ \sum x^2 = .948622, \text{ and } \sum xy = 295.54$$

Find the least squares prediction equation for predicting the number of games won, y, using a straight-line relationship with the team's batting average, x.

5) To investigate the relationship between yield of potatoes, y, and level of fertilizer application, x, a researcher divides a field into eight plots of equal size and applies differing amounts of fertilizer to each. The yield of potatoes (in pounds) and the fertilizer application (in pounds) are recorded for each plot. The data are as follows:

x	1	1.5	2	2.5	3	3.5	4	4.5
y	25	31	27	28	36	35	32	34

Summary statistics yield $SS_{xx} = 10.5$, $SS_{yy} = 112$, $SS_{xy} = 25$, $\bar{x} = 2.75$, and $\bar{y} = 31$. Find the least squares prediction equation.

6) In a study of feeding behavior, zoologists recorded the number of grunts of a warthog feeding by a lake in the 15 minute period following the addition of food. The data showing the number of grunts and and the age of the warthog (in days) are listed below:

Number of Grunts	Age (days)
83	118
61	134
32	148
37	153
56	160
33	167
55	176
10	182
13	188

a. Write the equation of a straight-line model relating number of grunts (y) to age (x).
b. Give the least squares prediction equation.
c. Give a practical interpretation of the value of $\hat{\beta}_0$, if possible.
d. Give a practical interpretation of the value of $\hat{\beta}_1$, if possible.

7) a. Complete the table.

	x_i	y_i	x_i^2	x_iy_i
	2	3		
	5	2		
	3	4		
	8	0		
Totals	$\Sigma x_i =$	$\Sigma y_i =$	$\Sigma x_i^2 =$	$\Sigma x_iy_i =$

b. Find SS_{xy}, SS_{xx}, β_1, $\bar{x}$, $\bar{y}$, and $\hat{\beta}_0$.
c. Write the equation of the least squares line.

2 Plot Least Squares Line with Scattergram

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

8) Consider the following pairs of measurements:

x	1	3	4	6	7
y	3	6	8	12	13

a. Construct a scattergram for the data.
b. What does the scattergram suggest about the relationship between x and y?
c. Find the least squares estimates of β_0 and β_1.
d. Plot the least squares line on your scattergram. Does the line appear to fit the data well?

3 Interpret Least Squares Line

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

9) A county real estate appraiser wants to develop a statistical model to predict the appraised value of houses in a section of the county called East Meadow. One of the many variables thought to be an important predictor of appraised value is the number of rooms in the house. Consequently, the appraiser decided to fit the simple linear regression model:

$$E(y) = \beta_0 + \beta_1 x,$$

where y = appraised value of the house (in thousands of dollars) and x = number of rooms. Using data collected for a sample of n = 80 houses in East Meadow, the following results were obtained:

$$\hat{y} = 80.80 + 19.72x$$

What are the properties of the least squares line, $\hat{y} = 80.80 + 19.72x$?

A) Average error of prediction is 0, and SSE is minimum.
B) It is normal, mean 0, constant variance, and independent.
C) All 80 of the sample y-values fall on the line.
D) It will always be a statistically useful predictor of y.

10) A county real estate appraiser wants to develop a statistical model to predict the appraised value of houses in a section of the county called East Meadow. One of the many variables thought to be an important predictor of appraised value is the number of rooms in the house. Consequently, the appraiser decided to fit the simple linear regression model:

$$E(y) = \beta_0 + \beta_1 x,$$

where y = appraised value of the house (in thousands of dollars) and x = number of rooms. Using data collected for a sample of $n = 74$ houses in East Meadow, the following results were obtained:

$\hat{y} = 74.80 + 21.66x$

Give a practical interpretation of the estimate of the slope of the least squares line.

A) For each additional room in the house, we estimate the appraised value to increase $21,660.
B) For each additional room in the house, we estimate the appraised value to increase $74,800.
C) For each additional dollar of appraised value, we estimate the number of rooms in the house to increase by 21.66.
D) For a house with 0 rooms, we estimate the appraised value to be $74,800.

11) A county real estate appraiser wants to develop a statistical model to predict the appraised value of houses in a section of the county called East Meadow. One of the many variables thought to be an important predictor of appraised value is the number of rooms in the house. Consequently, the appraiser decided to fit the simple linear regression model:

$$E(y) = \beta_0 + \beta_1 \mathrm{x},$$

where y = appraised value of the house (in thousands of dollars) and x = number of rooms. Using data collected for a sample of $n = 74$ houses in East Meadow, the following results were obtained:

$\hat{y} = 74.80 + 19.72x$

Give a practical interpretation of the estimate of the y-intercept of the least squares line.

A) There is no practical interpretation, since a house with 0 rooms is nonsensical.
B) For each additional room in the house, we estimate the appraised value to increase $74,800.
C) For each additional room in the house, we estimate the appraised value to increase $19,720.
D) We estimate the base appraised value for any house to be $74,800.

12) Is there a relationship between the raises administrators at State University receive and their performance on the job?

A faculty group wants to determine whether job rating (x) is a useful linear predictor of raise (y). Consequently, the group considered the straight–line regression model

$$E(y) = \beta_0 + \beta_1 x.$$

Using the method of least squares, the faculty group obtained the following prediction equation:

$$\hat{y} = 14{,}000 - 2{,}000x$$

Interpret the estimated slope of the line.

A) For a 1–point increase in an administrator's rating, we estimate the administrator's raise to decrease $2,000.
B) For a 1–point increase in an administrator's rating, we estimate the administrator's raise to increase $2,000.
C) For an administrator with a rating of 1.0, we estimate his/her raise to be $2,000.
D) For a $1 increase in an administrator's raise, we estimate the administrator's rating to decrease 2,000 points.

13) Is there a relationship between the raises administrators at State University receive and their performance on the job?

A faculty group wants to determine whether job rating (x) is a useful linear predictor of raise (y). Consequently, the group considered the straight–line regression model

$$E(y) = \beta_0 + \beta_1 x.$$

Using the method of least squares, the faculty group obtained the following prediction equation:

$$\hat{y} = 14{,}000 - 2{,}000x$$

Interpret the estimated y–intercept of the line.

A) For an administrator who receives a rating of zero, we estimate his or her raise to be $14,000.
B) The base administrator raise at State University is $14,000.
C) For a 1–point increase in an administrator's rating, we estimate the administrator's raise to increase $14,000.
D) There is no practical interpretation, since rating of 0 is nonsensical and outside the range of the sample data.

14) A large national bank charges local companies for using its services. A bank official reported the results of a regression analysis designed to predict the bank's charges (y), measured in dollars per month, for services rendered to local companies. One independent variable used to predict the service charge to a company is the company's sales revenue (x), measured in $ million. Data for 21 companies who use the bank's services were used to fit the model

$$E(y) = \beta_0 + \beta_1 x.$$

The results of the simple linear regression are provided below.

$\hat{y} = 2{,}700 + 20x$

Interpret the estimate of β_0, the y–intercept of the line.

A) There is no practical interpretation since a sales revenue of $0 is a nonsensical value.
B) All companies will be charged at least $2,700 by the bank.
C) About 95% of the observed service charges fall within $2,700 of the least squares line.
D) For every $1 million increase in sales revenue, we expect a service charge to increase $2,700.

15) An academic advisor wants to predict the typical starting salary of a graduate at a top business school using the GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT was created from a set of 25 data points.

Which of the following is not an assumption required for the simple linear regression analysis to be valid?

A) SALARY is independent of GMAT.
B) The errors of predicting SALARY are normally distributed.
C) The errors of predicting SALARY have a mean of 0.
D) The errors of predicting SALARY have a variance that is constant for any given value of GMAT.

16) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight–line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	–2298.36	158.531	–14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R–Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R–Squared	0.8922	Standard Deviation	1117.56

Interpret the estimated slope of the regression line.

A) For every 1–carat increase in the size of a diamond, we estimate that the price of the diamond will increase by $11,598.90.
B) For every 1–carat increase in the size of a diamond, we estimate that the price of the diamond will decrease by $2298.36.
C) For every $1 decrease in the price of the diamond, we estimate that the size of the diamond will increase by 11,598.9 carats.
D) For every 2298.36–carat decrease in the size of a diamond, we estimate that the price of the diamond will increase by $11,598.90.

17) What is the relationship between diamond price and carat size? 307 diamonds were sampled (ranging in size from 0.18 to 1.1 carats) and a straight–line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	–2298.36	158.531	–14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R–Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R–Squared	0.8922	Standard Deviation	1117.56

Interpret the estimated y–intercept of the regression line.

A) When a diamond is 0 carats in size, we estimate the price of the diamond to be $11,598.90.
B) When a diamond is 0 carats in size, we estimate the price of the diamond to be $2298.36.
C) When a diamond is 11598.9 carats in size, we estimate the price of the diamond to be $2298.36.
D) No practical interpretation of the y–intercept exists since a diamond of 0 carats cannot exist and falls outside the range of the carat sizes sampled.

18) A study of the top 75 MBA programs attempted to predict the average starting salary (in $1000's) of graduates of the program based on the amount of tuition (in $1000's) charged by the program. The results of a simple linear regression analysis are shown below:

Least Squares Linear Regression of Salary

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	18.1849	10.3336	1.76	0.0826
Size	1.47494	0.14017	10.52	0.0000

R–Squared	0.6027	Resid. Mean Square (MSE)	532.986
Adjusted R–Squared	0.5972	Standard Deviation	23.0865

Interpret the estimated slope of the regression line.

A) For every $1000 increase in the tuition charged by the MBA program, we estimate that the average starting salary will decrease by $1474.94.
B) For every $1000 increase in the tuition charged by the MBA program, we estimate that the average starting salary will increase by $1474.94.
C) For every $1474.94 increase in the tuition charged by the MBA program, we estimate that the average starting salary will increase by $18,184.90.
D) For every $1000 increase in the average starting salary, we estimate that the tuition charged by the MBA program will increase by $1474.94.

Answer the question True or False.

19) The Method of Least Squares specifies that the regression line has an average error of 0 and has an *SSE* that is minimized.

A) True
B) False

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

20) A company keeps extensive records on its new salespeople on the premise that sales should increase with experience. A random sample of seven new salespeople produced the data on experience and sales shown in the table.

Months on Job	Monthly Sales y ($ thousands)
2	2.4
4	7.0
8	11.3
12	15.0
1	.8
5	3.7
9	12.0

Summary statistics yield $SS_{xx} = 94.8571$, $SS_{xy} = 124.7571$, $SS_{yy} = 176.5171$, $\bar{x} = 5.8571$, and $\bar{y} = 7.4571$. State the assumptions necessary for predicting the monthly sales based on the linear relationship with the months on the job.

11.3 Model Assumptions

1 Understand Model Assumptions

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

1) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight-line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	Coefficient	Std Error	T	P
Constant	−2298.36	158.531	−14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R−Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R−Squared	0.8922	Standard Deviation	1117.56

Which of the following assumptions is not stated correctly?

A) The probability distribution of ε is normal.
B) The mean of the probability distribution of ε is 0.
C) The variance of the probability distribution of ε is constant for all settings of the independent variable.
D) The values of ε associated with any two observations are dependent on one another.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

2) State the four basic assumptions about the general form of the probability distribution of the random error ε.

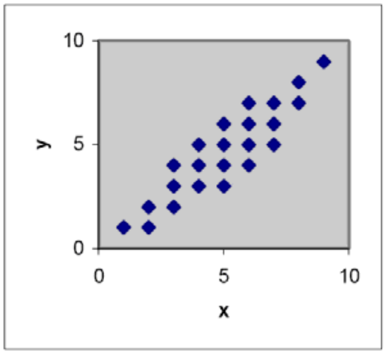
2 Interpret Scattergram

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

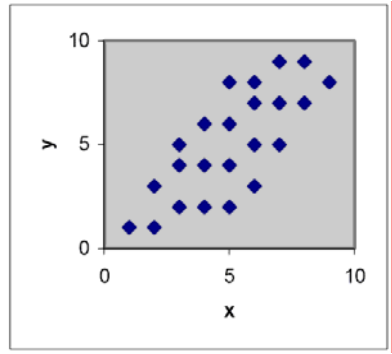
Solve the problem.

3) If a least squares line were determined for the data set in each scattergram, which would have the smallest variance?

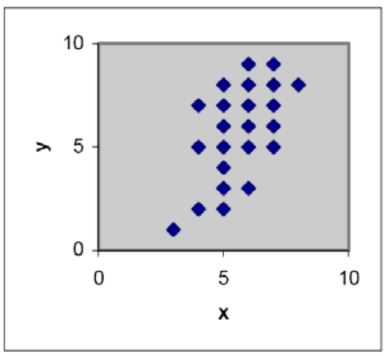
A)



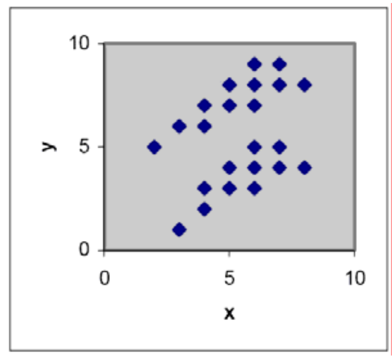
B)



C)



D)



3 Find and Interpret s^2 and s

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

4) Calculate SSE and s^2 for $n = 30$, $SS_{yy} = 100$, $SS_{xy} = 60$, and $\hat{\beta}_1 = .8$.

5) Calculate SSE and s^2 for $n = 25$, $\sum y^2 = 950$, $\sum y = 65$, $SS_{xy} = 3000$, and $\hat{\beta}_1 = .2$.

6) Suppose you fit a least squares line to 22 data points and the calculated value of SSE is .678.

a. Find s^2, the estimator of σ^2.

b. Find s, the estimator of σ.

c. What is the largest deviation you might expect between any one of the 22 points and the least squares line?

7) Suppose you fit a least squares line to 25 data points and the calculated value of SSE is 0.42.

a. Find s^2, the estimator of σ^2.

b. What is the largest deviation you might expect between any one of the 25 points and the least squares line?

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

8) Consider the data set shown below. Find the standard deviation of the least squares regression line.

y	0	3	2	3	8	10	11
x	−2	0	2	4	6	8	10

A) 1.5 B) 0.94643 C) 1.49045 D) 0.9003

9) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight−line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	Coefficient	Std Error	T	P
Constant	−2298.36	158.531	−14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R−Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R−Squared	0.8922	Standard Deviation	1117.56

Interpret the standard deviation of the regression model.

A) We can explain 89.25% of the variation in the sampled diamond prices around their mean using the size of the diamond in a linear model.
B) We expect most of the sampled diamond prices to fall within $1117.56 of their least squares predicted values.
C) We expect most of the sampled diamond prices to fall within $2235.12 of their least squares predicted values.
D) For every 1−carat increase in the size of a diamond, we estimate that the price of the diamond will increase by $1117.56.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

10) In a study of feeding behavior, zoologists recorded the number of grunts of a warthog feeding by a lake in the 15 minute period following the addition of food. The data showing the number of grunts and the age of the warthog (in days) are listed below:

Number of Grunts	Age (days)
115	143
93	159
64	173
69	178
88	185
65	192
87	201
42	207
47	213

a. Find *SSE*, s^2, and s.
b. Interpret the value of s.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

11) A county real estate appraiser wants to develop a statistical model to predict the appraised value of houses in a section of the county called East Meadow. One of the many variables thought to be an important predictor of appraised value is the number of rooms in the house. Consequently, the appraiser decided to fit the simple linear regression model:

$$E(y) = \beta_0 + \beta_1 x,$$

where y = appraised value of the house (in thousands of dollars) and x = number of rooms. Using data collected for a sample of $n = 74$ houses in East Meadow, the following results were obtained:

$\hat{y} = 74.80 + 19.72x$

Give a practical interpretation of the estimate of σ, the standard deviation of the random error term in the model.

A) We expect to predict the appraised value of an East Meadow house to within about $58,000 of its true value.
B) We expect to predict the appraised value of an East Meadow house to within about $29,000 of its true value.
C) We expect 95% of the observed appraised values to lie on the least squares line.
D) About 29% of the total variation in the sample of y–values can be explained by the linear relationship between appraised value and number of rooms.

12) The dean of the Business School at a small Florida college wishes to determine whether the grade–point average (GPA) of a graduating student can be used to predict the graduate's starting salary. More specifically, the dean wants to know whether higher GPAs lead to higher starting salaries. Records for 23 of last year's Business School graduates are selected at random, and data on GPA (x) and starting salary (y, in $thousands) for each graduate were used to fit the model

$$E(y) = \beta_0 + \beta_1 x$$

The results of the simple linear regression are provided below.

$\hat{y} = 4.25 + 2.75x$, $SS_{xy} = 5.15$, $SS_{xx} = 1.87$
$SS_{yy} = 15.17$, $SSE = 1.0075$

Compute an estimate of σ, the standard deviation of the random error term.

A) 0.219 B) 1.0075 C) .689 D) .048

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

13) Is the number of games won by a major league baseball team in a season related to the team's batting average? Data from 14 teams were collected and the summary statistics yield:

$\sum y = 1{,}134$, $\sum x = 3.642$, $\sum y^2 = 93{,}110$, $\sum x^2 = .948622$, and $\sum xy = 295.54$

Assume $\hat{\beta}1 = 455.27$. Estimate and interpret the estimate of σ.

14) A breeder of Thoroughbred horses wishes to model the relationship between the gestation period and the length of life of a horse. The breeder believes that the two variables may follow a linear trend. The information in the table was supplied to the breeder from various thoroughbred stables across the state.

Horse	Gestation period x (days)	Life Length y (years)	Horse	Gestation period x (days)	Life Length y (years)
1	416	24	5	356	22
2	279	25.5	6	403	23.5
3	298	20	7	265	21
4	307	21.5			

Summary statistics yield $SS_{xx} = 21{,}752$, $SS_{xy} = 236.5$, $SS_{yy} = 22$, $\bar{x} = 332$, and $\bar{y} = 22.5$. Calculate SSE, s^2, and s.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

15) Locate the values of SSE, s^2, and s on the printout below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.859	.737	.689	11.826

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4512.024	1	4512.024	32.265	.001
	Residual	1678.115	12	139.843		
	Total	6190.139	13			

A) $SSE = 1678.115$; $s^2 = 139.843$; $s = 11.826$

B) $SSE = 4512.024$; $s^2 = 4512.024$; $s = 32.265$

C) $SSE = 6190.139$; $s^2 = 4512.024$; $s = 32.265$

D) $SSE = 4512.024$; $s^2 = 139.843$; $s = 11.826$

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

16) Consider the following pairs of measurements:

x	5	8	3	4	9
y	6.2	3.4	7.5	8.1	3.2

a. Construct a scattergram for the data.

b. Use the method of least squares to model the relationship between x and y.

c. Calculate SSE, s^2, and s.

d. What percentage of the observed y-values fall within $2s$ of the values of $\hat{y}$ predicted by the least squares model?

11.4 Assessing the Utility of the Model: Making Inferences about the Slope β1

1 Construct Confidence Interval for β1

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

1) Consider the data set shown below. Find the 95% confidence interval for the slope of the regression line.

y	0	3	2	3	8	10	11
x	−2	0	2	4	6	8	10

A) 0.94643 ± 0.27603 B) 0.94643 ± 0.28377 C) 0.94643 ± 0.33306 D) 0.94643 ± 0.36203

2) A large national bank charges local companies for using their services. A bank official reported the results of a regression analysis designed to predict the bank's charges (y), measured in dollars per month, for services rendered to local companies. One independent variable used to predict service charge to a company is the company's sales revenue (x), measured in $ million. Data for 21 companies who use the bank's services were used to fit the model

$$E(y) = \beta_0 + \beta_1 x.$$

Suppose a 95% confidence interval for β_1 is (15, 25). Interpret the interval.

A) We are 95% confident that service charge (y) will increase between $15 and $25 for every $1 million increase in sales revenue (x).
B) We are 95% confident that the mean service charge will fall between $15 and $25 per month.
C) We are 95% confident that sales revenue (x) will increase between $15 and $25 million for every $1 increase in service charge (y).
D) We are 95% confident that service charge (y) will decrease between $15 and $25 for every $1 million increase in sales revenue (x).

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

3) Construct a 90% confidence interval for β_1 when $\hat{\beta}1 = 49$, $s = 4$, $SS_{xx} = 55$, and $n = 15$.

4) Construct a 95% confidence interval for β_1 when $\hat{\beta}1 = 49$, $s = 4$, $SS_{xx} = 55$, and $n = 15$.

2 Perform Hypothesis Test for Linearity

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

5) The data for $n = 37$ points were subjected to a simple linear regression with the results: $\hat{\beta}1 = 0.95$ and $s_{\hat{\beta}1} = 0.14$.

a. Test whether the two variables, x and y, are positively linearly related. Use $\alpha = .05$.
b. Construct and interpret a 90% confidence interval for β_1.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

6) In a comprehensive road test on new car models, one variable measured is the time it takes a car to accelerate from 0 to 60 miles per hour. To model acceleration time, a regression analysis is conducted on a random sample of 129 new cars.

TIME60: y = Elapsed time (in seconds) from 0 mph to 60 mph
MAX: x = Maximum speed attained (miles per hour)

The simple linear model $E(y) = \beta_0 + \beta_1 x$ was fit to the data. Computer printouts for the analysis are given below:

NWEIGHTED LEAST SQUARES LINEAR REGRESSION OF TIME60

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	18.7171	0.63708	29.38	0.0000
MAX	−0.08365	0.00491	−17.05	0.0000

R–SQUARED	0.6960	RESID. MEAN SQUARE (MSE)	1.28695
ADJUSTED R–SQUARED	0.6937	STANDARD DEVIATION	1.13444

SOURCE	DF	SS	MS	F	P
REGRESSION	1	374.285	374.285	290.83	0.0000
RESIDUAL	127	163.443	1.28695		
TOTAL	128	537.728			

CASES INCLUDED 129 MISSING CASES 0

Fill in the blank: "At α =.05, there is ________________ between maximum speed and acceleration time."
A) sufficient evidence of a negative linear relationship
B) insufficient evidence of a negative linear relationship
C) sufficient evidence of a positive linear relationship
D) insufficient evidence of a linear relationship

7) A manufacturer of boiler drums wants to use regression to predict the number of man-hours needed to erect drums in the future. The manufacturer collected a random sample of 35 boilers and measured the following two variables:

MANHRS: y = Number of man-hours required to erect the drum
PRESSURE: x_1 = Boiler design pressure (pounds per square inch, i.e., psi)

The simple linear model $E(y) = \beta_0 + \beta_1 x$ was fit to the data. A printout for the analysis appears below:

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF MANHRS

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	1.88059	0.58380	3.22	0.0028
PRESSURE	0.00321	0.00163	2.17	0.0300

R-SQUARED	0.4342	RESID. MEAN SQUARE (MSE)	4.25460
ADJUSTED R-SQUARED	0.4176	STANDARD DEVIATION	2.06267

SOURCE	DF	SS	MS	F	P
REGRESSION	1	111.008	111.008	5.19	0.0300
RESIDUAL	34	144.656	4.25160		
TOTAL	35	255.665			

Fill in the blank. At α =.01, there is ____________ between man-hours and pressure.

A) insufficient evidence of a positive linear relationship
B) sufficient evidence of a positive linear relationship
C) sufficient evidence of a negative linear relationship
D) sufficient evidence of a linear relationship

8) A county real estate appraiser wants to develop a statistical model to predict the appraised value of houses in a section of the county called East Meadow. One of the many variables thought to be an important predictor of appraised value is the number of rooms in the house. Consequently, the appraiser decided to fit the simple linear regression model:

$$E(y) = \beta_0 + \beta_1 x,$$

where y = appraised value of the house (in thousands of dollars) and x = number of rooms.

What set of hypotheses would you test to determine whether appraised value is positively linearly related to number of rooms?

A) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 > 0$
B) H_0: $\beta_1 < 0$ vs. H_a: $\beta_1 > 0$
C) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 < 0$
D) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$

9) A large national bank charges local companies for using their services. A bank official reported the results of a regression analysis designed to predict the bank's charges (y), measured in dollars per month, for services rendered to local companies. One independent variable used to predict service charge to a company is the company's sales revenue (x), measured in \$ million. Data for 21 companies who use the bank's services were used to fit the model

$$E(y) = \beta_0 + \beta_1 x.$$

The results of the simple linear regression are provided below.

$\hat{y} = 2{,}700 + 20x$, $s = 65$, 2-tailed p-value = .064 (for testing β_1)

Interpret the p-value for testing whether β_1 exceeds 0.

A) There is sufficient evidence (at $\alpha = .05$) to conclude that service charge (y) is positively linearly related to sales revenue (x) .
B) For every \$1 million increase in sales revenue (x), we expect a service charge (y) to increase \$.064.
C) There is insufficient evidence (at $\alpha = .05$) to conclude that service charge (y) is positively linearly related to sales revenue (x).
D) Sales revenue (x) is a poor predictor of service charge (y).

10) The dean of the Business School at a small Florida college wishes to determine whether the grade-point average (GPA) of a graduating student can be used to predict the graduate's starting salary. More specifically, the dean wants to know whether higher GPAs lead to higher starting salaries. Records for 23 of last year's Business School graduates are selected at random, and data on GPA (x) and starting salary (y, in \$thousands) for each graduate were used to fit the model

$$E(y) = \beta_0 + \beta_1 x.$$

The value of the test statistic for testing β_1 is 17.169. Select the proper conclusion.

A) There is sufficient evidence (at $\alpha = .05$) to conclude that GPA is positively linearly related to starting salary.
B) There is insufficient evidence (at $\alpha = .05$) to conclude that GPA is positively linearly related to starting salary.
C) There is insufficient evidence (at $\alpha = .10$) to conclude that GPA is a useful linear predictor of starting salary.
D) At any reasonable α, there is no relationship between GPA and starting salary.

11) An academic advisor wants to predict the typical starting salary of a graduate at a top business school using the GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT using 25 data points is shown below.

$\hat{\beta}_0 = -92040$ $\hat{\beta}1 = 228$ $s = 3213$ df = 23 $t = 6.67$

Set up the null and alternative hypotheses for testing whether a linear relationship exists between SALARY and GMAT.

A) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$
B) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 > 0$
C) H_0: $\hat{\beta}1 = 228$ vs. H_a: $\hat{\beta}_1 > 228$
D) H_0: $\beta_1 > 0$ vs. H_a: $\beta_1 < 0$

12) Consider the following model $y = \beta_0 + \beta_1 x + \in$, where y is the daily rate of return of a stock, and x is the daily rate of return of the stock market as a whole, measured by the daily rate of return of Standard & Poor's (S&P) 500 Composite Index. Using a random sample of $n = 12$ days from 2007, the least squares lines shown in the table below were obtained for four firms. The estimated standard error of $\hat{\beta}_1$ is shown to the right of each least squares prediction equation.

Firm	Estimated Market Model	Estimated Standard Error of β_1
Company A	$y = .0010 + 1.40x$	.03
Company B	$y = .0005 - 1.21x$	.06
Company C	$y = .0010 + 1.62x$	1.34
Company D	$y = .0013 + .76x$	.15

Calculate the test statistic for determining whether the market model is useful for predicting daily rate of return of Company A's stock.

A) 46.7 B) 161.6 C) 1.40 ± .067 D) 1.40

13) Consider the following model $y = \beta_0 + \beta_1 x + \in$, where y is the daily rate of return of a stock, and x is the daily rate of return of the stock market as a whole, measured by the daily rate of return of Standard & Poor's (S&P) 500 Composite Index. Using a random sample of n = 12 days from 1980, the least squares lines shown in the table below were obtained for four firms. The estimated standard error of $\hat{\beta}_1$ is shown to the right of each least squares prediction equation.

Firm	Estimated Market Model	Estimated Standard Error of β_1
Company A	$y = .0010 + 1.40x$	.03
Company B	$y = .0005 - 1.21x$	.06
Company C	$y = .0010 + 1.62x$	1.34
Company D	$y = .0013 + .76x$	.15

For which of the three stocks, Companies B, C, or D, is there evidence (at $\alpha = .05$) of a positive linear relationship between y and x?

A) Company D only
B) Company C only
C) Companies B and C only
D) Companies B and D only

14) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight–line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	–2298.36	158.531	–14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R–Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R–Squared	0.8922	Standard Deviation	1117.56

Which of the following conclusions is correct when testing to determine if the size of the diamond is a useful positive linear predictor of the price of a diamond?

A) There is insufficient evidence to indicate that the size of the diamond is a useful positive linear predictor of the price of a diamond when testing at $\alpha = 0.05$.
B) There is sufficient evidence to indicate that the size of the diamond is a useful positive linear predictor of the price of a diamond when testing at $\alpha = 0.05$.
C) There is insufficient evidence to indicate that the price of the diamond is a useful positive linear predictor of the size of a diamond when testing at $\alpha = 0.05$.
D) The sample size is too small to make any conclusions regarding the regression line.

15) A study of the top 75 MBA programs attempted to predict the average starting salary (in $1000's) of graduates of the program based on the amount of tuition (in $1000's) charged by the program. The results of a simple linear regression analysis are shown below:

Least Squares Linear Regression of Salary

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	18.1849	10.3336	1.76	0.0826
Size	1.47494	0.14017	10.52	0.0000

R–Squared	0.6027	Resid. Mean Square (MSE)	532.986
Adjusted R–Squared	0.5972	Standard Deviation	23.0865

Fill in the blank. At $\alpha = 0.05$, there is ________________ between the amount of tuition charged by an MBA program and the average starting salary of graduates of the program.

A) …sufficient evidence of a negative linear relationship…
B) …insufficient evidence of a positive linear relationship…
C) …sufficient evidence of a positive linear relationship…

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

16) Operations managers often use work sampling to estimate how much time workers spend on each operation. Work sampling—which involves observing workers at random points in time—was applied to the staff of the catalog sales department of a clothing manufacturer. The department applied regression to data collected for 40 randomly selected working days.

The simple linear model $E(y) = \beta_0 + \beta_1 x$ was fit to the data. The printouts for the analysis are given below:

TIME: y = Time spent (in hours) taking telephone orders during the day

ORDERS: x = Number of telephone orders received during the day

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF TIME

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	10.1639	1.77844	5.72	0.0000
ORDERS	0.05836	0.00586	9.96	0.0000

R–SQUARED	0.7229	RESID. MEAN SQUARE (MSE)	11.6175
ADJUSTED R–SQUARED	0.7156	STANDARD DEVIATION	3.40844

SOURCE	DF	SS	MS	F	P
REGRESSION	1	1151.55	1151.55	99.12	0.0000
RESIDUAL	38	441.464	11.6175		
TOTAL	39	1593.01			

CASES INCLUDED 40 MISSING CASES 0

Conduct a test of hypothesis to determine if time spent (in hours) taking telephone orders during the day and the number of telephone orders received during the day are positively linearly related. Use $\alpha = .01$.

17) Is the number of games won by a major league baseball team in a season related to the team's batting average? Data from 14 teams were collected and the summary statistics yield:

$$\sum y = 1{,}134,\ \sum x = 3.642,\ \sum y^2 = 93{,}110,\ \sum x^2 = .948622, \text{ and } \sum xy = 295.54$$

Assume $\hat{\beta}1 = 455.27$ and $\hat{\sigma} = 9.18$. Conduct a test of hypothesis to determine if a positive linear relationship exists between team batting average and number of wins. Use $\alpha = .05$.

18) A breeder of Thoroughbred horses wishes to model the relationship between the gestation period and the length of life of a horse. The breeder believes that the two variables may follow a linear trend. The information in the table was supplied to the breeder from various thoroughbred stables across the state.

Horse	Gestation period x (days)	Life Length y (years)	Horse	Gestation period x (days)	Life Length y (years)
1	416	24	5	356	22
2	279	25.5	6	403	23.5
3	298	20	7	265	21
4	307	21.5			

Summary statistics yield $SS_{xx} = 21{,}752$, $SS_{xy} = 236.5$, $SS_{yy} = 22$, $\overline{x} = 332$, and $\overline{y} = 22.5$. Test to determine if a linear relationship exists between the gestation period and the length of life of a horse. Use $\alpha = .05$ and use $s = 1.97$ as an estimate of σ.

19) Consider the following pairs of observations:

x	2	3	5	5	6
y	1.3	1.6	2.1	2.2	2.7

a. Construct a scattergram for the data. Does the scattergram suggest that y is positively linearly related to x?
b. Find the slope of the least squares line for the data and test whether the data provide sufficient evidence that y is positively linearly related to x. Use $\alpha = .05$.

20) A realtor collected the following data for a random sample of ten homes that recently sold in her area.

House	Asking Price	Days on Market
A	\$114,500	29
B	\$149,900	16
C	\$154,700	59
D	\$159,900	42
E	\$160,000	72
F	\$165,900	45
G	\$169,700	12
H	\$171,900	39
I	\$175,000	81
J	\$289,900	121

a. Construct a scattergram for the data.
b. Find the least squares line for the data and plot the line on your scattergram.
c. Test whether the number of days on the market, y, is positively linearly related to the asking price, x. Use $\alpha = .05$.

11.5 The Coefficients of Correlation and Determination

1 Interpret Correlation Coefficient

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

1) The coefficient of correlation is a useful measure of the linear relationship between two variables.
A) True B) False

2) A high value of the correlation coefficient r implies that a causal relationship exists between x and y.
A) True
B) False

3) A low value of the correlation coefficient r implies that x and y are unrelated.
A) True
B) False

Solve the problem.

4) An academic advisor wants to predict the typical starting salary of a graduate at a top business school using the GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT using 25 data points is shown below.

$$\hat{\beta}_0 = -92040 \quad \hat{\beta}_1 = 228 \quad s = 3213 \quad r^2 = .66 \quad r = .81 \quad \text{df} = 23 \quad t = 6.67$$

Give a practical interpretation of $r = .81$.
A) There appears to be a positive correlation between SALARY and GMAT.
B) We estimate SALARY to increase 81% for every 1-point increase in GMAT.
C) 81% of the sample variation in SALARY can be explained by using GMAT in a straight-line model.
D) We can predict SALARY correctly 81% of the time using GMAT in a straight-line model.

5) A study of the top 75 MBA programs attempted to predict the average starting salary (in \$1000's) of graduates of the program based on the amount of tuition (in \$1000's) charged by the program. The results of a simple linear regression analysis are shown below:

Least Squares Linear Regression of Salary

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	18.1849	10.3336	1.76	0.0826
Size	1.47494	0.14017	10.52	0.0000

R-Squared	0.6027	Resid. Mean Square (MSE)	532.986
Adjusted R-Squared	0.5972	Standard Deviation	23.0865

In addition, we are told that the coefficient of correlation was calculated to be r = 0.7763. Interpret this result.
A) There is a fairly strong negative linear relationship between the amount of tuition charged and the average starting salary variables.
B) There is a fairly strong positive linear relationship between the amount of tuition charged and the average starting salary variables.
C) There is a very weak positive linear relationship between the amount of tuition charged and the average starting salary variables.
D) There is almost no linear relationship between the amount of tuition charged and the average starting salary variables.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

6) In team-teaching, two or more teachers lead a class. A researcher tested the use of team-teaching in mathematics education. Two of the variables measured on each teacher in a sample of 169 mathematics teachers were years of teaching experience x and reported success rate y (measured as a percentage) of team-teaching mathematics classes.

The correlation coefficient for the sample data was reported as $r = -0.34$. Interpret this result.

2 Calculate Correlation Coefficient

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

7) Consider the data set shown below. Find the coefficient of correlation for between the variables x and y.

y	0	3	2	3	8	10	11
x	−2	0	2	4	6	8	10

A) 0.9003 B) 0.8804 C) 0.9489 D) 0.9383

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

8) To investigate the relationship between yield of potatoes, y, and level of fertilizer application, x, a researcher divides a field into eight plots of equal size and applies differing amounts of fertilizer to each. The yield of potatoes (in pounds) and the fertilizer application (in pounds) are recorded for each plot. The data are as follows:

x	1	1.5	2	2.5	3	3.5	4	4.5
y	25	31	27	28	36	35	32	34

Summary statistics yield $SS_{xx} = 10.5$, $SS_{yy} = 112$, $SS_{xy} = 25$, and $SSE = 52.476$. Calculate the coefficient of correlation.

9) In a study of feeding behavior, zoologists recorded the number of grunts of a warthog feeding by a lake in the 15 minute period following the addition of food. The data showing the number of grunts and and the age of the warthog (in days) are listed below:

Number of Grunts	Age (days)
88	123
66	139
37	153
42	158
61	165
38	172
60	181
15	187
18	193

Find and interpret the value of r.

10) Consider the following pairs of observations:

x	2	3	5	5	6
y	1.3	1.6	2.1	2.2	2.7

Find and interpret the value of the coefficient of correlation.

3 Perform Correlation Test

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

11) A study of the top 75 MBA programs attempted to predict the average starting salary (in \$1000's) of graduates of the program based on the amount of tuition (in \$1000's) charged by the program. We are told that the coefficient of correlation was calculated to be $r = 0.7763$. Use this information to calculate the test statistic that would be used to determine if a positive linear relationship exists between the two variables.

A) $t = 10.52$ B) $t = 1.475$ C) $t = 1.760$ D) $t = 0.6027$

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

12) In team-teaching, two or more teachers lead a class. An researcher tested the use of team-teaching in mathematics education. Two of the variables measured on each sample of 177 mathematics teachers were years of teaching experience x and reported success rate y (measured as a percentage) of team-teaching mathematics classes.

a. The researcher hypothesized that mathematics teachers with more years of experience will report higher perceived success rates in team-taught classes.
State this hypothesis in terms of the parameter of a linear model relating x to y.
b. The correlation coefficient for the sample data was reported as $r = -0.3$.
Interpret this result.
c. Does the value of r support the hypothesis? Test using $\alpha = .05$.

4 Calculate Coefficient of Determination

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

13) Consider the data set shown below. Find the coefficient of determination for the simple linear regression model.

y	0	3	2	3	8	10	11
x	-2	0	2	4	6	8	10

A) 0.9003 B) 0.8804 C) 0.9489 D) 0.9383

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

14) In a study of feeding behavior, zoologists recorded the number of grunts of a warthog feeding by a lake in the 15 minute period following the addition of food. The data showing the number of grunts and and the age of the warthog (in days) are listed below:

Number of Grunts	Age (days)
93	128
71	144
42	158
47	163
66	170
43	177
65	186
20	192
23	198

Find and interpret the value of r^2.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

15) The dean of the Business School at a small Florida college wishes to determine whether the grade–point average (GPA) of a graduating student can be used to predict the graduate's starting salary. More specifically, the dean wants to know whether higher GPAs lead to higher starting salaries. Records for 23 of last year's Business School graduates are selected at random, and data on GPA (x) and starting salary (y, in $thousands) for each graduate were used to fit the model

$$E(y) = \beta_0 + \beta_1 x$$

The results of the simple linear regression are provided below.

$\hat{y} = 4.25 + 2.75x$, $SS_{xy} = 5.15$, $SS_{xx} = 1.87$
$SS_{yy} = 15.17$, $SSE = 1.0075$

Calculate the value of r^2, the coefficient of determination.

A) 0.934 B) 0.661 C) 0.872 D) 0.339

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

16) A company keeps extensive records on its new salespeople on the premise that sales should increase with experience. A random sample of seven new salespeople produced the data on experience and sales shown in the table.

Months on Job	Monthly Sales y ($ thousands)
2	2.4
4	7.0
8	11.3
12	15.0
1	.8
5	3.7
9	12.0

Summary statistics yield $SS_{xx} = 94.8571$, $SS_{xy} = 124.7571$, $SS_{yy} = 176.5171$, $\overline{x} = 5.8571$, and $\overline{y} = 7.4571$. Using $SSE = 12.435$, find and interpret the coefficient of determination.

17) To investigate the relationship between yield of potatoes, y, and level of fertilizer application, x, a researcher divides a field into eight plots of equal size and applies differing amounts of fertilizer to each. The yield of potatoes (in pounds) and the fertilizer application (in pounds) are recorded for each plot. The data are as follows:

x	1	1.5	2	2.5	3	3.5	4	4.5
y	25	31	27	28	36	35	32	34

Summary statistics yield $SS_{xx} = 10.5$, $SS_{yy} = 112$, $SS_{xy} = 25$, and $SSE = 52.476$. Calculate the coefficient of determination.

18) Consider the following pairs of observations:

x	2	3	5	5	6
y	1.3	1.6	2.1	2.2	2.7

Find and interpret the value of the coefficient of determination.

5 Interpret Coefficient of Determination

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

19) In a comprehensive road test on new car models, one variable measured is the time it takes the car to accelerate from 0 to 60 miles per hour. To model acceleration time, a regression analysis is conducted on a random sample of 129 new cars.

TIME60: y = Elapsed time (in seconds) from 0 mph to 60 mph
MAX x = Maximum speed attained (miles per hour)

The simple linear model $E(y) = \beta_0 + \beta_1 x$ was fit to the data. Computer printouts for the analysis are given below:

NWEIGHTED LEAST SQUARES LINEAR REGRESSION OF TIME60

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	18.7171	0.63708	29.38	0.0000
MAX	−0.08365	0.00491	−17.05	0.0000

R-SQUARED	0.6960	RESID. MEAN SQUARE (MSE)	1.28695
ADJUSTED R-SQUARED	0.6937	STANDARD DEVIATION	1.13444

SOURCE	DF	SS	MS	F	P
REGRESSION	1	374.285	374.285	290.83	0.0000
RESIDUAL	127	163.443	1.28695		
TOTAL	128	537.728			

CASES INCLUDED 129 MISSING CASES 0

Approximately what percentage of the sample variation in acceleration time can be explained by the simple linear model?

A) 70% B) 0% C) −17% D) 8%

20) A manufacturer of boiler drums wants to use regression to predict the number of man–hours needed to erect drums in the future. The manufacturer collected a random sample of 35 boilers and measured the following two variables:

MANHRS: y = Number of man–hours required to erect the drum
PRESSURE: x = Boiler design pressure (pounds per square inch, i.e., psi)

The simple linear model $E(y) = \beta_1 + \beta_1 x$ was fit to the data. A printout for the analysis appears below:

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF MANHRS

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	1.88059	0.58380	3.22	0.0028
PRESSURE	0.00321	0.00163	2.17	0.0300

R–SQUARED	0.4342	RESID. MEAN SQUARE (MSE)	4.25460
ADJUSTED R–SQUARED	0.4176	STANDARD DEVIATION	2.06267

SOURCE	DF	SS	MS	F	P
REGRESSION	1	111.008	111.008	5.19	0.0300
RESIDUAL	34	144.656	4.25160		
TOTAL	35	255.665			

Give a practical interpretation of the coefficient of determination, r^2.

A) About 43% of the sample variation in number of man–hours can be explained by the simple linear model.
B) We are 43% confident that the design pressure will be a useful predictor of number of man–hours required to build a steam drum.
C) Approximately 95% of the actual man–hours required to build a drum will fall within 43 hours of their predicted values.
D) About 2.06% of the sample variation in number of man–hours can be explained by the simple linear model.

21) An academic advisor wants to predict the typical starting salary of a graduate at a top business school using the GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT using 25 data points is shown below.

$\hat{\beta}_0 = -92040$ $\hat{\beta}_1 = 228$ $s = 3213$ $r^2 = .66$ $r = .81$ df = 23 $t = 6.67$

Give a practical interpretation of $r^2 = .66$.

A) 66% of the sample variation in SALARY can be explained by using GMAT in a straight–line model.
B) We expect to predict SALARY to within $2\left[\sqrt{.66}\right]$ of its true value using GMAT in a straight–line model.
C) We estimate SALARY to increase $.66 for every 1–point increase in GMAT.
D) We can predict SALARY correctly 66% of the time using GMAT in a straight–line model.

22) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight–line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	–2298.36	158.531	–14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R–Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R–Squared	0.8922	Standard Deviation	1117.56

Interpret the coefficient of determination for the regression model.

A) We can explain 89.25% of the variation in the sampled diamond prices around their mean using the size of the diamond in a linear model.

B) There is sufficient evidence to indicate that the size of the diamond is a useful predictor of the price of a diamond when testing at alpha = 0.05.

C) We expect most of the sampled diamond prices to fall within $2235.12 of their least squares predicted values.

D) For every 1–carat increase in the size of a diamond, we estimate that the price of the diamond will increase by $1117.56.

11.6 Using the Model for Estimation and Prediction

1 Calculate and Compare Confidence Intervals for Mean of y

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Solve the problem.

1) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight-line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	-2298.36	158.531	-14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R-Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R-Squared	0.8922	Standard Deviation	1117.56

The model was then used to create 95% confidence and prediction intervals for y and for E(Y) when the carat size of the diamond was 1 carat. The results are shown here:

95% confidence interval for E(Y): ($9091.60, $9509.40)
95% prediction interval for Y: ($7091.50, $11,510.00)

Which of the following interpretations is correct if you want to use the model to estimate E(Y) for all 1-carat diamonds?

A) We are 95% confident that the price of a 1-carat diamond will fall between $7091.50 and $11,510.00.
B) We are 95% confident that the average price of all 1-carat diamonds will fall between $7091.50 and $11,510.00.
C) We are 95% confident that the price of a 1-carat diamond will fall between $9091.60 and $9509.40.
D) We are 95% confident that the average price of all 1-carat diamonds will fall between $9091.60 and $9509.40.

2) A study of the top 75 MBA programs attempted to predict the average starting salary (in \$1000's) of graduates of the program based on the amount of tuition (in \$1000's) charged by the program. The results of a simple linear regression analysis are shown below:

Least Squares Linear Regression of Salary

Predictor Variables	**Coefficient**	**Std Error**	**T**	**P**
Constant	18.1849	10.3336	1.76	0.0826
Size	1.47494	0.14017	10.52	0.0000

R–Squared	0.6027	Resid. Mean Square (MSE)	532.986
Adjusted R–Squared	0.5972	Standard Deviation	23.0865

The model was then used to create 95% confidence and prediction intervals for y and for E(Y) when the tuition charged by the MBA program was \$75,000. The results are shown here:

95% confidence interval for E(Y): (\$123,390, \$134,220)
95% prediction interval for Y: (\$82,476, \$175,130)

Which of the following interpretations is correct if you want to use the model to estimate E(Y) for all MBA programs?

A) We are 95% confident that the average starting salary for graduates of a single MBA program that charges \$75,000 in tuition will fall between \$123,390 and \$134,220.
B) We are 95% confident that the average starting salary for graduates of a single MBA program that charges \$75,000 in tuition will fall between \$82,476 and \$175,130.
C) We are 95% confident that the average of all starting salaries for graduates of all MBA programs that charge \$75,000 in tuition will fall between \$123,390 and \$134,220.
D) We are 95% confident that the average of all starting salaries for graduates of all MBA programs that charge \$75,000 in tuition will fall between \$82,476 and \$175,130.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

3) A company keeps extensive records on its new salespeople on the premise that sales should increase with experience. A random sample of seven new salespeople produced the data on experience and sales shown in the table.

Months on Job	Monthly Sales y (\$ thousands)
2	2.4
4	7.0
8	11.3
12	15.0
1	.8
5	3.7
9	12.0

Summary statistics yield $SS_{xx} = 94.8571$, $SS_{xy} = 124.7571$, $SS_{yy} = 176.5171$, $\overline{x} = 5.8571$, and $\overline{y} = 7.4571$. Calculate a 90% confidence interval for $E(y)$ when $x = 5$ months. Assume $s = 1.577$ and the prediction equation is $\hat{y} = -.25 + 1.315x$.

4) A realtor collected the following data for a random sample of ten homes that recently sold in her area.

House	Asking Price	Days on Market
A	\$114,500	29
B	\$149,900	16
C	\$154,700	59
D	\$159,900	42
E	\$160,000	72
F	\$165,900	45
G	\$169,700	12
H	\$171,900	39
I	\$175,000	81
J	\$289,900	121

a. Find a 90% confidence interval for the mean number of days on the market for all houses listed at \$150,000.
b. Suppose a house has just been listed at \$150,000. Find a 90% prediction interval for the number of days the house will be on the market before it sells.

2 Find and Interpret Predication Interval for y

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer the question True or False.

5) Probabilistic models are commonly used to estimate both the mean value of y and a new individual value of y for a particular value of x.
A) True B) False

6) The least squares model provides very good estimates of y for values of x far outside the range of x values contained in the sample.
A) True B) False

Solve the problem.

7) An academic advisor wants to predict the typical starting salary of a graduate at a top business school using the GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT using 25 data points is shown below.

$\hat{\beta}_0 = -92040$ $\hat{\beta}_1 = 228$ $s = 3213$ $r^2 = .66$ $r = .81$ df $= 23$ $t = 6.67$

A 95% prediction interval for SALARY when GMAT = 600 is approximately (\$37,915, \$51,984). Interpret this interval.
A) We are 95% confident that the SALARY of a top business school graduate with a GMAT of 600 will fall between \$37,915 and \$51,984.
B) We are 95% confident that the mean SALARY of all top business school graduates with GMATs of 600 will fall between \$37,915 and \$51,984.
C) We are 95% confident that the increase in SALARY for a 600-point increase in GMAT will fall between \$37,915 and \$51,984.
D) We are 95% confident that the SALARY of a top business school graduate will fall between \$37,915 and \$51,984.

8) The dean of the Business School at a small Florida college wishes to determine whether the grade–point average (GPA) of a graduating student can be used to predict the graduate's starting salary. More specifically, the dean wants to know whether higher GPAs lead to higher starting salaries. Records for 23 of last year's Business School graduates are selected at random, and data on GPA (x) and starting salary (y, in $thousands) for each graduate were used to fit the model

$$E(y) = \beta_0 + \beta_1 x.$$

The results of the simple linear regression are provided below.

$\hat{y} = 4.25 + 2.75x$, $SSxy = 5.15$, $SSxx = 1.87$
$SSyy = 15.17$, $SSE = 1.0075$
Range of the x–values: 2.23 – 3.85
Range of the y–values: 9.3 – 15.6

Suppose a 95% prediction interval for y when $x = 3.00$ is (16, 21). Interpret the interval.

A) We are 95% confident that the starting salary of a Business School graduate with a GPA of 3.00 will fall between $16,000 and $21,000.
B) We are 95% confident that the mean starting salary of all Business School graduates with GPAs of 3.00 will fall between $16,000 and $21,000.
C) We are 95% confident that the starting salary of a Business School graduate will increase between $16,000 and $21,000 for every 3–point increase in GPA.
D) We are 95% confident that the starting salary of a Business School graduate will fall between $16,000 and $21,000.

9) What is the relationship between diamond price and carat size? 307 diamonds were sampled and a straight–line relationship was hypothesized between y = diamond price (in dollars) and x = size of the diamond (in carats). The simple linear regression for the analysis is shown below:

Least Squares Linear Regression of PRICE

Predictor Variables	Coefficient	Std Error	T	P
Constant	–2298.36	158.531	–14.50	0.0000
Size	11598.9	230.111	50.41	0.0000

R–Squared	0.8925	Resid. Mean Square (MSE)	1248950
Adjusted R–Squared	0.8922	Standard Deviation	1117.56

The model was then used to create 95% confidence and prediction intervals for y and for E(Y) when the carat size of the diamond was 1 carat. The results are shown here:

95% confidence interval for E(Y): ($9091.60, $9509.40)
95% prediction interval for Y: ($7091.50, $11,510.00)

Which of the following interpretations is correct if you want to use the model to determine the price of a single 1–carat diamond?

A) We are 95% confident that the price of a 1–carat diamond will fall between $7091.50 and $11,510.00.
B) We are 95% confident that the average price of all 1–carat diamonds will fall between $7091.50 and $11,510.00.
C) We are 95% confident that the price of a 1–carat diamond will fall between $9091.60 and $9509.40.
D) We are 95% confident that the average price of all 1–carat diamonds will fall between $9091.60 and $9509.40.

10) A study of the top 75 MBA programs attempted to predict the average starting salary (in \$1000's) of graduates of the program based on the amount of tuition (in \$1000's) charged by the program. The results of a simple linear regression analysis are shown below:

Least Squares Linear Regression of Salary

Predictor Variables	Coefficient	Std Error	T	P
Constant	18.1849	10.3336	1.76	0.0826
Size	1.47494	0.14017	10.52	0.0000

R–Squared 0.6027 Resid. Mean Square (MSE) 532.986
Adjusted R–Squared 0.5972 Standard Deviation 23.0865

The model was then used to create 95% confidence and prediction intervals for y and for E(Y) when the tuition charged by the MBA program was \$75,000. The results are shown here:

95% confidence interval for E(Y): (\$123,390, \$134,220)
95% prediction interval for Y: (\$82,476, \$175,130)

Which of the following interpretations is correct if you want to use the model to predict Y for a single MBA programs?

A) We are 95% confident that the average starting salary for graduates of a single MBA program that charges \$75,000 in tuition will fall between \$123,390 and \$134,220.
B) We are 95% confident that the average starting salary for graduates of a single MBA program that charges \$75,000 in tuition will fall between \$82,476 and \$175,130.
C) We are 95% confident that the average of all starting salaries for graduates of all MBA programs that charge \$75,000 in tuition will fall between \$123,390 and \$134,220.
D) We are 95% confident that the average of all starting salaries for graduates of all MBA programs that charge \$75,000 in tuition will fall between \$82,476 and \$175,130.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

11) A breeder of Thoroughbred horses wishes to model the relationship between the gestation period and the length of life of a horse. The breeder believes that the two variables may follow a linear trend. The information in the table was supplied to the breeder from various thoroughbred stables across the state.

Horse	Gestation period x (days)	Life Length y (years)	Horse	Gestation period x (days)	Life Length y (years)
1	416	24	5	356	22
2	279	25.5	6	403	23.5
3	298	20	7	265	21
4	307	21.5			

Summary statistics yield $SS_{xx} = 21,752$, $SS_{xy} = 236.5$, $SS_{yy} = 22$, $\overline{x} = 332$, and $\overline{y} = 22.5$. Find a 95% prediction interval for the length of life of a horse that had a gestation period of 300 days. Use $s = 2$ as an estimate of σ and use $\hat{y} = 18.89 + .01087x$.

12) Consider the following pairs of observations:

x	2	0	3	3	5
y	1	3	4	6	7

a. Construct a scattergram for the data.
b. Find the least squares line, and plot it on your scattergram.
c. Find a 99% confidence interval for the mean value of y when $x = 1$.
d. Find a 99% prediction interval for a new value of y when $x = 1$.

11.7 A Complete Example

1 Apply Simple Linear Regression

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

(Situation P) Below are the results of a survey of America's best graduate and professional schools. The top 25 business schools, as determined by reputation, student selectivity, placement success, and graduation rate, are listed in the table. For each school, three variables were measured: (1) GMAT score for the typical incoming student; (2) student acceptance rate (percentage accepted of all students who applied); and (3) starting salary of the typical graduating student.

	School	GMAT	Acc. Rate	Salary
1.	Harvard	644	15.0%	$ 63,000
2.	Stanford	665	10.2	60,000
3.	Penn	644	19.4	55,000
4.	Northwestern	640	22.6	54,000
5.	MIT	650	21.3	57,000
6.	Chicago	632	30.0	55,269
7.	Duke	630	18.2	53,300
8.	Dartmouth	649	13.4	52,000
9.	Virginia	630	23.0	55,269
10.	Michigan	620	32.4	53.300
11.	Columbia	635	37.1	52,000
12.	Cornell	648	14.9	50,700
13.	CMU	630	31.2	52,050
14.	UNC	625	15.4	50,800
15.	Cal-Berkeley	634	24.7	50,000
16.	UCLA	640	20.7	51,494
17.	Texas	612	28.1	43,985
18.	Indiana	600	29.0	44,119
19.	NYU	610	35.0	53,161
20.	Purdue	595	26.8	43,500
21.	USC	610	31.9	49,080
22.	Pittsburgh	605	33.0	43,500
23.	Georgetown	617	31.7	45,156
24.	Maryland	593	28.1	42,925
25.	Rochester	605	35.9	44,499

The academic advisor wants to predict the typical starting salary of a graduate at a top business school using GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT using the 25 data points in the table are shown below.

$\beta_0 = -92040$ $\hat{\beta}_1 = 228$ $s = 3213$ $r^2 = .66$ $r = .81$ df = 23 $t = 6.67$

---

1) For the situation above, write the equation of the probabilistic model of interest.

A) Salary = $\beta_0 + \beta_1$ (GMAT)
B) Salary = $\beta_0 + \beta_1$ (GMAT) + $\in$
C) GMAT = $\beta_0 + \beta_1$ (SALARY)
D) GMAT = $\beta_0 + \beta_1$ (SALARY) + $\in$

2) For the situation above, write the equation of the least squares line.

A) SALARY = –92040 + 228 (GMAT)
B) SALARY = 228 + 92040 (GMAT)
C) GMAT = –92040 + 228 (SALARY)
D) GMAT = 228 – 92040 (SALARY)

3) For the situation above, give a practical interpretation of $\hat{\beta}_0 = -92040$.

A) We expect to predict SALARY to within 2(92040) = \$184,080 of its true value using GMAT in a straight-line model.
B) We estimate SALARY to decrease \$92,040 for every 1-point increase in GMAT.
C) We estimate the base SALARY of graduates of a top business school to be -\$92,040.
D) The value has no practical interpretation since a GMAT of 0 is nonsensical and outside the range of the sample data.

4) For the situation above, give a practical interpretation of $\hat{\beta}_1 = 228$.

A) We expect to predict SALARY to within 2(228) = \$456 of its true value using GMAT in a straight-line model.
B) We estimate SALARY to increase \$228 for every 1-point increase in GMAT.
C) We estimate GMAT to increase 228 points for every \$1 increase in SALARY.
D) The value has no practical interpretation since a GMAT of 228 is nonsensical and outside the range of the sample data.

5) For the situation above, give a practical interpretation of s = 3213.

A) We expect to predict SALARY to within 2(3213) = \$6,426 of its true value using GMAT in a straight-line model.
B) We estimate SALARY to increase \$3,213 for every 1-point increase in GMAT.
C) Our predicted value of SALARY will equal 2(3213) = \$6,426 for any value of GMAT.
D) We expect the predicted SALARY to deviate from actual SALARY by at least 2(3213) = \$6,426 using GMAT in a straight-line model.

6) For the situation above, give a practical interpretation of $r^2 = .66$.

A) We expect to predict SALARY to within $2(\sqrt{.66})$ of its true value using GMAT in a straight-line model.
B) We estimate SALARY to increase \$.66 for every 1-point increase in GMAT.
C) 66% of the sample variation in SALARY can be explained by using GMAT in a straight-line model.
D) We can predict SALARY correctly 66% of the time using GMAT in a straight-line model.

7) For the situation above, give a practical interpretation of $r = .81$.

A) We estimate SALARY to increase 81% for every 1-point increase in GMAT.
B) There appears to be a positive correlation between SALARY and GMAT.
C) 81% of the sample variation in SALARY can be explained by using GMAT in a straight-line model.
D) We can predict SALARY correctly 81% of the time using GMAT in a straight-line model.

8) Set up the null and alternative hypotheses for testing whether a positive linear relationship exists between SALARY and GMAT in the situation above.

A) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 > 0$
B) H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$
C) H_0: $\beta_1 = 228$ vs. H_a: $\beta_1 > 228$
D) H_0: $\beta_1 > 0$ vs. H_a: $\beta_1 < 0$

9) For the situation above, give a practical interpretation of $t = 6.67$.

A) There is evidence (at $\alpha = .05$) to indicate that $\beta_1 = 0$.
B) Only 6.67% of the sample variation in SALARY can be explained by using GMAT in a straight-line model.
C) We estimate SALARY to increase \$6.67 for every 1-point increase in GMAT.
D) There is evidence (at $\alpha = .05$) of at least a positive linear relationship between SALARY and GMAT.

10) A 95% prediction interval for SALARY when GMAT = 600 is ($37,915, $51,948). Interpret this interval for the situation above.
   A) We are 95% confident that the SALARY of a top business school graduate with a GMAT of 600 will fall between $37,915 and $51,984.
   B) We are 95% confident that the mean SALARY of all top business school graduates with GMATs of 600 will fall between $37,915 and $51,984.
   C) We are 95% confident that the increase in SALARY for a 600-point increase in GMAT will fall between $37,915 and $51,984.
   D) We are 95% confident that the SALARY of a top business school graduate will fall between $37,915 and $51,984.

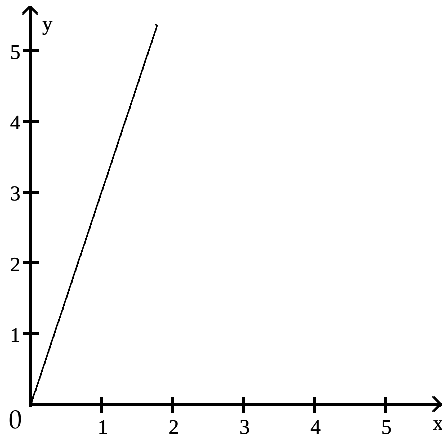
11) For the situation above, which of the following is *not* an assumption required for the simple linear regression analysis to be valid?
   A) The errors of predicting SALARY are normally distributed.
   B) The errors of predicting SALARY have a mean of 0.
   C) The errors of predicting SALARY have a variance that is constant for any given value of GMAT.
   D) SALARY is independent of GMAT.

Ch. 11 Simple Linear Regression
Answer Key

11.1 Probabilistic Models

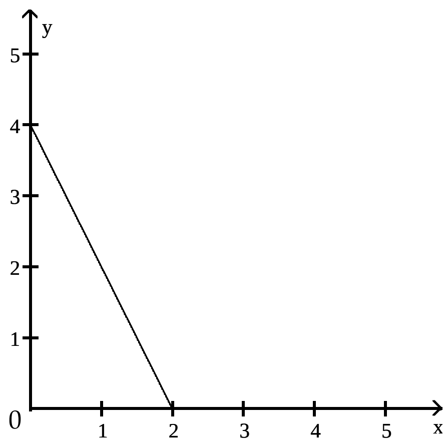
1 Graph Line

1) A
2) A
3) A
4) A
5) A
6)



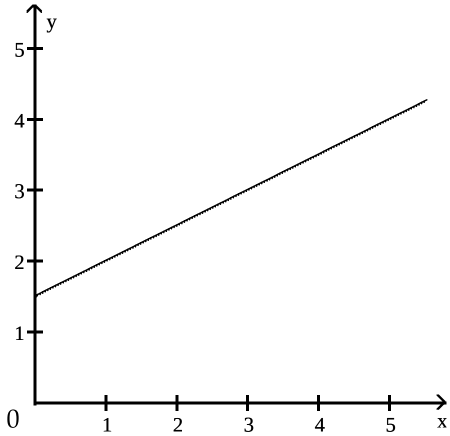
slope: 3; y-intercept: 0

7)



slope: −2; y-intercept: 4

8)



slope: .5; y–intercept: 1.5

9) $\beta_0 = \frac{39}{7}$, $\beta_1 = \frac{3}{7}$

2 Understand Probabilistic Models

10) A
11) A
12) B

11.2 Fitting the Model: The Least Squares Approach

1 Perform Calculations and Find Least Squares Line

1) B
2) A
3) $\hat{\beta}_1 = -.08365$. For every 1 mile per hour increase in the maximum speed attained, we estimate the elapsed 0–to–60 acceleration time to decrease by .08365 second.

4) $SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = .948622 - \frac{(3.642)^2}{14} = .00118171$

$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 295.54 - \frac{(3.642)(1{,}134)}{14} = .538$

$\bar{y} = \frac{\sum y}{n} = \frac{1{,}134}{14} = 81$

$\bar{x} = \frac{\sum x}{n} = \frac{3.642}{14} = .26014$

$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{.538}{.00118171} = 455.27$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 81 - 455.27(.26014) = -37.434$

The least squares equation is $\hat{y} = -37.434 + 455.27x$.

5) $\beta_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{25}{10.5} \approx 2.3810$

$\hat{\beta}_0 = \bar{y} - \beta_1 x = 31 - 2.3810(2.75) = 24.4523$

The least squares prediction equation is $\hat{y} = 24.4523 + 2.3810x$

6) a. $E(y) = \beta_0 + \beta_1 x$

b. $\hat{y} = \hat{\beta}0 + \hat{\beta}1x = 172.06 - .8195x$

c. We would expect approximately 172 grunts after feeding a warthog that was just born. However, since the value 0 is outside the range of the original data set, this estimate is highly unreliable.

d. For each additional day, we estimate the number of grunts will decrease by .8195.

7) a.

	x_i	y_i	x_i^2	x_iy_i
	2	3	4	6
	5	2	25	10
	3	4	9	12
	8	0	64	0
Totals	$\Sigma x_i = 18$	$\Sigma y_i = 9$	$\Sigma x_i^2 = 102$	$\Sigma x_iy_i = 28$

b. $SS_{xy} = 28 - \frac{(18)(9)}{4} = -12.5$; $SS_{xx} = 102 - \frac{18^2}{4} = 21$;

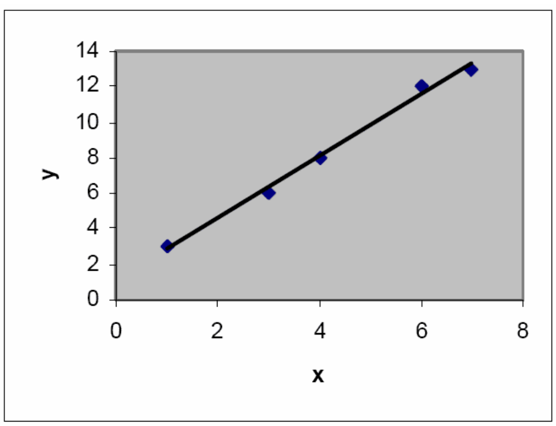
$\hat{\beta}_1 = \frac{-12.5}{21} \approx -.5952$; $\bar{x} = \frac{18}{4} = 4.5$; $\bar{y} = \frac{9}{4} = 2.25$;

$\hat{\beta}_0 = 2.25 + .5952(4.5) = 4.9284$

c. $\hat{y} = 4.9284 + .5952x$

2 Plot Least Squares Line with Scattergram

8) a.



b. y increases as x increases in an approximately linear pattern.

c. $\beta_1 \approx 1.7368$; $\beta_0 \approx 1.1053$

d. The line appears to fit the data well.

3 Interpret Least Squares Line

9) A
10) A
11) A
12) A
13) A
14) A
15) A

16) A
17) D
18) B
19) A
20) The assumptions necessary are:

1. The random errors are normally distributed.
2. The random errors are independent of one another.
3. The mean of the random errors is 0.
4. The variance of the random errors, σ^2, is constant for all levels of the independent variable x.

11.3 Model Assumptions

1 Understand Model Assumptions

1) D
2) Assumption 1: The mean of the probability distribution of ε is 0.

Assumption 2: The variance of the probability distribution of ε is constant for all settings of the independent variable, x.

Assumption 3: The probability distribution of ε is normal.

Assumption 4: The values of ε associated with any two observed values of y are independent.

2 Interpret Scattergram

3) A

3 Find and Interpret s^2 and s

4) $\text{SSE} = \text{SS}_{yy} - \hat{\beta}1\text{SS}_{xy} = 100 - .8(60) = 52;\ s^2 = \dfrac{\text{SSE}}{n-2} = \dfrac{52}{30-2} \approx 1.857$

5) $\text{SS}_{yy} = 950 - \dfrac{65^2}{25} = 781,\ \text{SSE} = \text{SS}_{yy} - \hat{\beta}1\text{SS}_{xy} = 781 - .2(3000) = 181;$

$s^2 = \dfrac{\text{SSE}}{n-2} = \dfrac{181}{25-2} \approx 7.87$

6) a. $s^2 = \dfrac{\text{SSE}}{n-2} = \dfrac{.678}{22-2} \approx .0339$

b. $s = \sqrt{s^2} \approx \sqrt{.0339} \approx .1841$
c. $2s = 2(.1841) = .3682$

7) a. $s^2 = 0.018$; b. 0.268
8) C
9) C
10) a. $SSE = 1650.36;\ s^2 = 235.77;\ s = 15.35$
b. We expect most of the observed y values to lie within 30.70 units of their respective least squares predicted values, $\hat{y}$.
11) A
12) A

13) $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 295.54 - \frac{(3.642)(1{,}134)}{14} = .538$$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 93{,}110 - \frac{(1{,}134)^2}{14} = 1{,}256$$

$SSE = 1{,}256 - 455.27(.538) = 1{,}011.06$

$$s^2 = \frac{SSE}{n-2} = \frac{1011.06}{14-2} = 84.26$$

$$s = \sqrt{s^2} = \sqrt{84.26} = 9.179$$

We expect most of the sample number of games won, y, to fall within $2s \Rightarrow 2(9.179) \Rightarrow 18.358$ of their least squares predicted values.

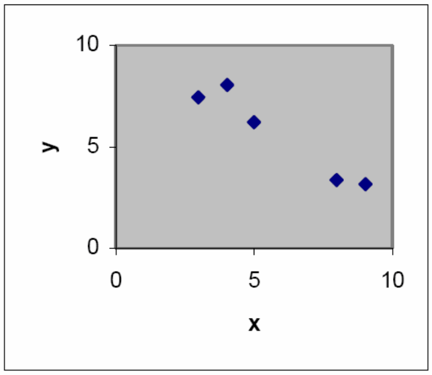
14) $\hat{\beta}1 = \frac{SS_{xy}}{SS_{xx}} = \frac{236.5}{21{,}752} = .01087$

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 22 - .01087(236.5) = 19.4286$

$s^2 = \frac{SSE}{n-2} = \frac{19.4286}{7-2} = 3.8857$ $\qquad s = \sqrt{s^2} = \sqrt{3.8857} = 1.971$

15) A

16) a.



b. $\hat{y} = 10.6187 = .8515x$

c. $SSE = 1.35677$, $s^2 = .4523$, and $s = .6725$

d. 100%

11.4 Assessing the Utility of the Model: Making Inferences about the Slope β1

1 Construct Confidence Interval for β1

1) D

2) A

3) $\hat{\beta}1 \pm t_{.05}\, s_{\hat{\beta}1} = 49 \pm 1.771\left(\frac{4}{\sqrt{55}}\right) = 49 \pm .96$

4) $\hat{\beta}1 \pm t_{.025}\, s_{\hat{\beta}1} = 49 \pm 2.160\left(\frac{4}{\sqrt{55}}\right) = 49 \pm 1.17$

2 Perform Hypothesis Test for Linearity

5) a. $t = 6.786 > t_{.05} = 1.69$, reject H_0; we concluded that x and y are positively linearly related.

b. (0.713, 1.187); We can be 90% confident that the true slope is between 0.713 and 1.187.

6) A

7) A

8) A

9) A

10) A

11) A

12) A

13) A

14) B

15) C

16) To determine if time spent taking telephone orders during the day is positively linearly related with the number of telephone orders received during the day, we test:

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

The test statistic is given on the printout as $t = 9.96$. The p-value for the desired test is $p = .0000/2 = .0000$ (divided in half because a one-tailed test is desired.)

Since $\alpha = .01 > p$-value ≈ 0, H_0 is rejected. There is sufficient evidence to indicate that the time spent taking telephone orders during the day is positively linearly related with the number of telephone orders received during the day.

17) We test: H_0: $\beta_1 = 0$
H_a: $\beta_1 > 0$

The test statistic is $t = \dfrac{\hat{\beta}_1 - 0}{s/\sqrt{SS_{xx}}}$.

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = .948622 - \frac{(3.642)^2}{14} = .00118$$

$$t = \frac{455.27 - 0}{9.18/\sqrt{.00118}} = 1.704$$

The rejection region requires $\alpha = .05$ in the upper tail of the t distribution with df $= n - 2 = 14 - 2 = 12$. From a t table, $t_{.05} = 1.782$. The rejection region is $t > 1.782$.

Since the observed value of the test statistic does not fall in the rejection region ($t = 1.704 \not> 1.782$), H_0 cannot be rejected. There is insufficient evidence to indicate that team wins is positively linearly related with team batting average.

18) $\hat{\beta}_1 = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{236.5}{21{,}752} = .01087$

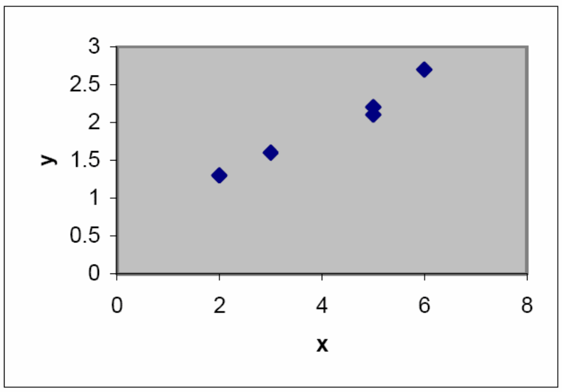
We test: H_0: $\beta_1 = 0$
H_a: $\beta_1 \neq 0$

The test statistic is $t = \dfrac{\hat{\beta}_1 - 0}{s/\sqrt{SS_{xx}}} = \dfrac{.01087 - 0}{1.97/\sqrt{21{,}752}} = .814$

The rejection region requires $\alpha/2 = .05/2 = .025$ in both tails of the t distribution with $n - 2 = 7 - 2 = 5$ df. From a t table, $t_{.025} = 2.571$. The rejection region is $t > 2.571$ or $t < -2.571$.

Since the observed value of the test statistic does not fall in the rejection region ($t = .814 \not> 2.571$), H_0 cannot be rejected. There is insufficient evidence to indicate that the gestation period and the length of life of a horse are linearly related at $\alpha = .05$.

19) a.



The scattergram does suggest that y is positively linearly related to x.

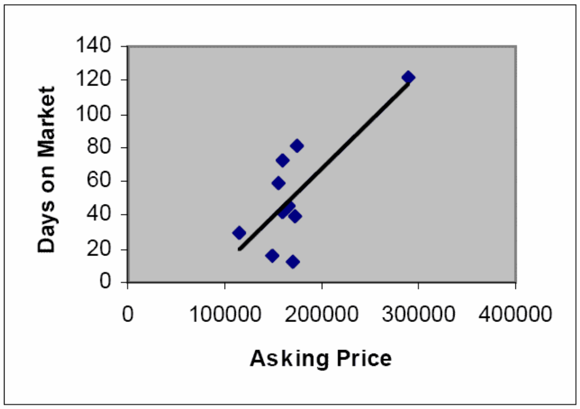
b. $SS_{xy} = 45.1 - \dfrac{(21)(9.9)}{5} = 3.52$; $SS_{xx} = 99 - \dfrac{(21)^2}{5} = 10.8$;

$\hat{\beta}_1 = \dfrac{3.52}{10.8} \approx .3259$; $s = .1165$

The test statistic is $t = \dfrac{.3259}{.1165/\sqrt{10.8}} \approx 9.19$.

Based on 3 degrees of freedom, the rejection region is $t > 2.353$. Since the test statistic falls in the rejection region, we reject the null hypothesis and conclude that y is positively linearly related to x.

20) a.



b. $SS_{xy} = 10{,}232{,}660$; $SS_{xx} \approx 1.833 \times 10^{10}$; $\hat{\beta}_1 = .0005583$; $\bar{x} = 171{,}140$;

$\bar{y} = 51.6$; $\hat{\beta}_0 = 51.6 - .0005583(171{,}140) = -43.94$

$\hat{y} = 43.94 + .0005583x$

c. The test statistic is $t = \dfrac{.0005583}{22.58/\sqrt{1.833 \times 10^{10}}} \approx 3.35$.

Based on 8 degrees of freedom, the rejection region is $t > 1.860$. Since the test statistic falls in the rejection region, we reject the null hypothesis and conclude that the number of days on the market, y, is positively linearly related to the asking price, x.

11.5 The Coefficients of Correlation and Determination

1 Interpret Correlation Coefficient

1) A
2) B
3) B
4) A
5) C
6) There is a weak negative correlation between years of teaching experience and success in team–teaching mathematics classes.

2 Calculate Correlation Coefficient

7) C

8) $r = \dfrac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} = \dfrac{25}{\sqrt{(10.5)(112)}} = .729$

9) $r = .792$; There is a positive linear correlation between age and number of grunts.

10) $SS_{xy} = 45.1 - \dfrac{(21)(9.9)}{5} = 3.52$; $SS_{xx} = 99 - \dfrac{(21)^2}{5} = 10.8$;

$SS_{yy} = 20.79 - \dfrac{(9.9)^2}{5} = 1.188$; $r = \dfrac{3.52}{\sqrt{10.8 \cdot 1.188}} \approx .9827$;

There is a strong linear relationship between x and y.

3 Perform Correlation Test

11) A

12) a. H_0: $\beta_1 > 0$.

b. There is a weak negative correlation between years of teaching experience and success in team–teaching mathematics classes.

c. No; $t = -4.16$

4 Calculate Coefficient of Determination

13) A

14) r^2 = .627; 62.7% of the variation in number of grunts can be explained by using age in a linear model.

15) A

16) $r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{176.5171 - 12.435}{176.5171} = .9296$

92.96% of the variation in the sample monthly sales values can be explained by using months on the job in a linear model.

17) $r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{112 - 52.4762}{112} = .53146$

18) $SS_{yy} = 1.188$; $SSE = .040741$; $r^2 = 1 - \frac{.040741}{1.188} \approx .9657$;

96.57% of the sample variation in y values can be attributed to the linear relationship between x and y.

5 Interpret Coefficient of Determination

19) A

20) A

21) A

22) A

11.6 Using the Model for Estimation and Prediction

1 Calculate and Compare Confidence Intervals for Mean of y

1) D

2) C

3) For $x = 5$, $\hat{y} = -.25 + 1.315(5) = 6.325$

The confidence interval is of the form:

$$\hat{y} \pm t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{SS_{xx}}}$$

Confidence coefficient $.90 = 1 - \alpha \Rightarrow \alpha = 1 - .90 = .10$. $\alpha/2 = .10/2 = .05$. From a t table, $t_{.05} = 2.015$ with $n - 2 = 7 - 2 = 5$ df. The confidence interval is:

$$6.325 \pm 2.015(1.577)\sqrt{\frac{1}{7} + \frac{(5 - 5.8571)^2}{94.8571}} \Rightarrow 6.325 \pm 1.233 \Rightarrow (5.092, 7.558)$$

4) a. The regression line is $\hat{y} = -43.94 + .0005583x$, so for $x = 150{,}000$ we have
$\hat{y} = -43.94 + .0005583(150{,}000) = 39.805$.
For df = 8, $t_{.05} = 1.860$.
$s = 22.58236$; $\bar{x} = 171{,}140$; $SS_{xx} = 1.8329 \times 10^{10}$
The interval is 39.805 ± 14.813.

b. The interval is 39.805 ± 44.539.

2 Find and Interpret Predication Interval for y

5) A

6) B

7) A

8) A

9) A

10) B

11) The prediction interval is of the form:

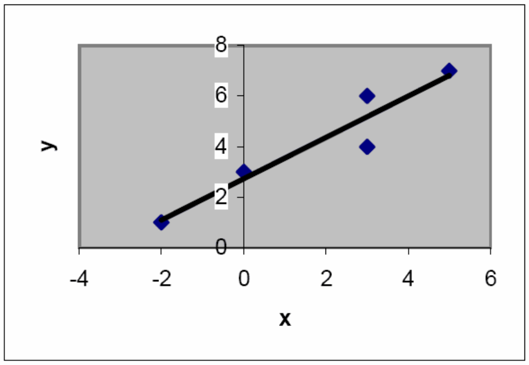
$$\hat{y} \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_{xx}}}$$

$\hat{y} = 18.89 + .01087(300) = 22.151$

Confidence coefficient $.95 = 1 - \alpha \Rightarrow \alpha = 1 - .95 = .05$. $\alpha/2 = .05/2 = .025$. From a t table, $t_{.025} = 2.571$ with $n - 2 = 7 - 2 = 5$ df. The 95% prediction interval is:

$$22.151 \pm 2.571(2)\sqrt{1 + \frac{1}{7} + \frac{(300 - 332)^2}{21{,}752}} \Rightarrow 22.151 \pm 5.609 \Rightarrow (16.542, 27.760)$$

12) a.



b. $\hat{y} = 2.7273 + .8182x$

c. For $x = 1$, $\hat{y} = 2.7273 + .8182(1) = 3.5455$.
For 3 degrees of freedom, $t_{.005} = 5.841$.
$s = .8528$; $\bar{x} = 1.8$; $SS_{xx} = 30.8$

The interval is $3.5455 \pm 5.841(.8528)\sqrt{\frac{1}{5} + \frac{(1 - 1.8)^2}{30.8}} \approx 3.5455 \pm 2.3405$.

d. The interval is $3.5455 \pm 5.841(.8528)\sqrt{\frac{1}{5} + \frac{(1 - 1.8)^2}{30.8}} \approx 3.5455 \pm 5.5037$.

11.7 A Complete Example

1 Apply Simple Linear Regression

1) B
2) A
3) D
4) B
5) A
6) C
7) B
8) A
9) D
10) A
11) D