Chapter 12: Simple Linear Regression

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

SCENARIO 12-3

The director of cooperative education at a state college wants to examine the effect of cooperative education job experience on marketability in the work place. She takes a random sample of 4 students. For these 4, she finds out how many times each had a cooperative education job and how many job offers they received upon graduation. These data are presented in the table below.

Student	CoopJobs	JobOffer
1	1	4
2	2	6
3	1	3
4	0	1

1) Referring to Scenario 12-3, set up a scatter plot. 1) _____

SCENARIO 12-10

The management of a chain electronic store would like to develop a model for predicting the weekly sales (in thousands of dollars) for individual stores based on the number of customers who made purchases. A random sample of 12 stores yields the following results:

Customers	Sales (Thousands of Dollars)
907	11.20
926	11.05
713	8.21
741	9.21
780	9.42
898	10.08
510	6.73
529	7.02
460	6.12
872	9.52
650	7.53
603	7.25

2) Referring to Scenario 12-10, generate the scatter plot. 2) _____

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

3) The *Y-intercept* (b_0) represents the 3) _____

A) predicted value of *Y*.
B) variation around the sample regression line.
C) change in estimated *Y* per unit change in *X*.
D) predicted value of *Y* when $X = 0$.

4) The slope (b_1) represents 4) _____

A) the predicted value of Y.
B) variation around the line of regression.
C) predicted value of Y when $X = 0$.
D) the estimated average change in Y per unit change in X.

SCENARIO 12-1

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 charges to a company is the company's sales revenue (X) —measured in millions of dollars. Data for 21 companies who use the bank's services were used to fit the model:

$Y_i = \beta_0 + \beta_1 X_i + E_i$

The results of the simple linear regression are provided below.

$\hat{Y} = -2{,}700 + 20X$, $S_{YX} = 65$, two-tail p-value = 0.034 (for testing β_1)

5) Referring to Scenario 12-1, interpret the estimate of β_0, the Y-intercept of the line. 5) _____

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

SCENARIO 12-6

The following Excel tables are obtained when "Score received on an exam (measured in percentage points)" (Y) is regressed on "percentage attendance" (X) for 22 students in a Statistics for Business and Economics course.

Regression Statistics	
Multiple R	0.142620229
R Square	0.02034053
Standard Error	20.25979924
Observations	22

	Coefficients	*Standard Error*	*T Stat*	*P-value*
Intercept	39.39027309	37.24347659	1.057642216	0.302826622
Attendance	0.340583573	0.52852452	0.644404489	0.526635689

6) Referring to Scenario 12-6, which of the following statements is true? 6) _____

A) If the score received increases by 39.39%, the estimated mean attendance will go up by 1%.
B) If attendance increases by 1%, the estimated mean score received will increase by 39.39 percentage points.
C) If attendance increases by 1%, the estimated mean score received will increase by 0.341 percentage points.
D) If attendance increases by 0.341%, the estimated mean score received will increase by 1 percentage point.

7) What do we mean when we say that a simple linear regression model is "statistically" useful? 7) _____

A) The model is an excellent predictor of *Y*.
B) The model is "practically" useful for predicting *Y*.
C) The model is a better predictor of *Y* than the sample mean, $\overline{Y}$.
D) All the statistics computed from the sample make sense.

8) The least squares method minimizes which of the following? 8) _____

I. *SSR*
II. *SSE*
III. *SST*

A) II only B) III only C) All of these D) I only

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

SCENARIO 12-4

The managers of a brokerage firm are interested in finding out if the number of new clients a broker brings into the firm affects the sales generated by the broker. They sample 12 brokers and determine the number of new clients they have enrolled in the last year and their sales amounts in thousands of dollars. These data are presented in the table that follows.

Broker	Clients	Sales
1	27	52
2	11	37
3	42	64
4	33	55
5	15	29
6	15	34
7	25	58
8	36	59
9	28	44
10	30	48
11	17	31
12	22	38

9) Referring to Scenario 12-4, the least squares estimate of the slope is _____. 9) _____

10) Referring to Scenario 12-4, the least squares estimate of the *Y*-intercept is _____. 10) _____

SCENARIO 12-5

The managing partner of an advertising agency believes that his company's sales are related to the industry sales. He uses Microsoft Excel to analyze the last 4 years of quarterly data (i.e., $n = 16$) with the following results:

Regression Statistics	
Multiple R	0.802
R Square	0.643
Adjusted R Square	0.618
Standard Error SYX	0.9224
Observations	16

ANOVA

	df	SS	MS	F	Sig.F
Regression	1	21.497	21.497	25.27	0.000
Error	14	11.912	0.851		
Total	15	33.409			

Predictor	Coef	StdError	t Stat	*p*-value
Intercept	3.962	1.440	2.75	0.016
Industry	0.040451	0.008048	5.03	0.000

Durbin-Watson Statistic 1.59

11) Referring to Scenario 12-5, the value of the quantity that the least squares regression line minimizes is _____. 11) _____

12) Referring to Scenario 12-5, the estimates of the *Y*-intercept and slope are _____ and _____, respectively. 12) _____

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

SCENARIO 12-9

It is believed that, the average numbers of hours spent studying per day (HOURS) during undergraduate education should have a positive linear relationship with the starting salary (SALARY, measured in thousands of dollars per month) after graduation. Given below is the Excel output for predicting starting salary (Y) using number of hours spent studying per day (X) for a sample of 51 students. NOTE: Only partial output is shown.

Regression Statistics	
Multiple R	0.8857
R Square	0.7845
Adjusted R Square	0.7801
Standard Error	1.3704
Observations	51

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	335.0472	335.0473	178.3859	
Residual			1.8782		
Total	50	427.0798			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1.8940	0.4018	-4.7134	0.0000	-2.7015	-1.0865
Hours	0.9795	0.0733	13.3561	0.0000	0.8321	1.1269

Note: 2.051E - 05 = 2.051*10^{-05} and 5.944E - 18 = 5.944*10^{-18}.

13) Referring to Scenario 12-9, the estimated change in mean salary (in thousands of dollars) as a result of spending an extra hour per day studying is 13) _____

A) 0.9795 B) 0.7845 C) 335.0473 D) -1.8940

SCENARIO 13-2

A candy bar manufacturer is interested in trying to estimate how sales are influenced by the price of their product. To do this, the company randomly chooses 6 small cities and offers the candy bar at different prices. Using candy bar sales as the dependent variable, the company will conduct a simple linear regression on the data below:

City	Price ($)	Sales
River Falls	1.30	100
Hudson	1.60	90
Ellsworth	1.80	90
Prescott	2.00	40
Rock Elm	2.40	38
Stillwater	2.90	32

14) Referring to Scenario 13-2, what is the percentage of the total variation in candy bar sales explained by the regression model? 14) _____

A) 100% B) 78.39% C) 48.19% D) 88.54%

15) True or False: The Regression Sum of Squares (*SSR*) can never be greater than the Total Sum of Squares (*SST*). 15) _____

A) True B) False

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

SCENARIO 13-4

The managers of a brokerage firm are interested in finding out if the number of new clients a broker brings into the firm affects the sales generated by the broker. They sample 12 brokers and determine the number of new clients they have enrolled in the last year and their sales amounts in thousands of dollars. These data are presented in the table that follows.

Broker	Clients	Sales
1	27	52
2	11	37
3	42	64
4	33	55
5	15	29
6	15	34
7	25	58
8	36	59
9	28	44
10	30	48
11	17	31
12	22	38

16) Referring to Scenario 13-4, _____ % of the total variation in sales generated can be explained by the number of new clients brought in. 16) _____

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

17) The coefficient of determination (r^2) tells you 17) _____

A) the proportion of total variation that is explained.
B) whether *r* has any significance.
C) that you should not partition the total variation.
D) that the coefficient of correlation (*r*) is larger than 1.

SCENARIO 13-11

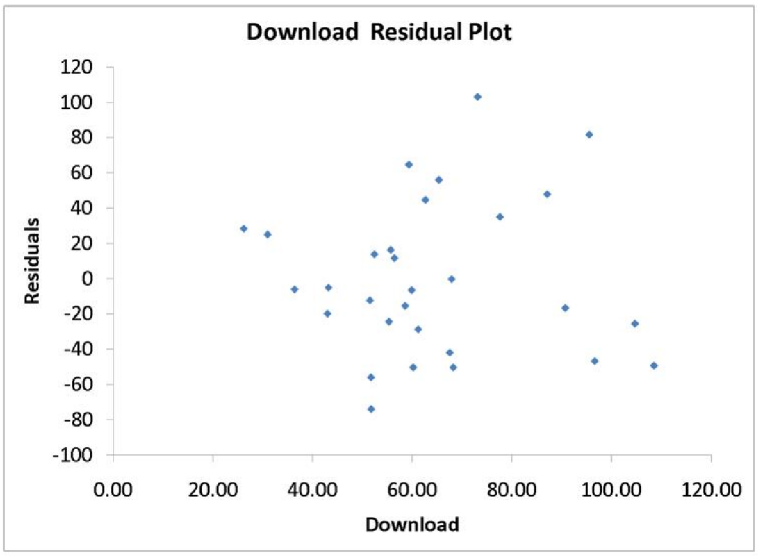
A computer software developer would like to use the number of downloads (in thousands) for the trial version of his new shareware to predict the amount of revenue (in thousands of dollars) he can make on the full version of the new shareware. Following is the output from a simple linear regression along with the residual plot and normal probability plot obtained from a data set of 30 different sharewares that he has developed:

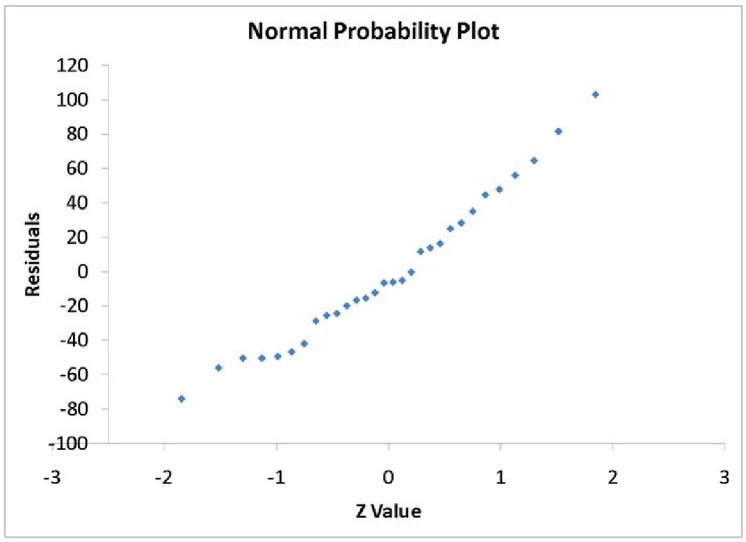
Regression Statistics	
Multiple R	0.8691
R Square	0.7554
Adjusted R Square	0.7467
Standard Error	44.4765
Observations	30.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	171062.9193	171062.9193	86.4759	0.0000
Residual	28	55388.4309	1978.1582		
Total	29	226451.3503			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-95.0614	26.9183	-3.5315	0.0015	-150.2009	-39.9218
Download	3.7297	0.4011	9.2992	0.0000	2.9082	4.5513





18) Referring to Scenario 13-11, which of the following is the correct interpretation for the coefficient of determination? 18) _____

A) 75.54% of the variation in the number of downloads can be explained by the variation in revenue.
B) 74.67% of the variation in the number of downloads can be explained by the variation in revenue.
C) 74.67% of the variation in revenue can be explained by the variation in the number of downloads.
D) 75.54% of the variation in revenue can be explained by the variation in the number of downloads.

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

SCENARIO 13-13

In this era of tough economic conditions, voters increasingly ask the question: "Is the educational achievement level of students dependent on the amount of money the state in which they reside spends on education?" The partial computer output below is the result of using spending per student ($) as the independent variable and composite score which is the sum of the math, science and reading scores as the dependent variable on 35 states that participated in a study. The table includes only partial results.

Regression Statistics				
Multiple R	0.3122			
R Square	0.0975			
Adjusted R Square	0.0701			
Standard Error	26.9122			
Observations	35			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	1	2581.5759		
Residual			724.2674	
Total	34	26482.4000		
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	595.540251	22.115176		
Spending per Student ($)	0.007996	0.004235		

19) Referring to Scenario 13-13, what percentage of the variation in composite score can be explained by the variation in spending per student? 19) _____

SCENARIO 13-3

The director of cooperative education at a state college wants to examine the effect of cooperative education job experience on marketability in the work place. She takes a random sample of 4 students. For these 4, she finds out how many times each had a cooperative education job and how many job offers they received upon graduation. These data are presented in the table below.

Student	CoopJobs	JobOffer
1	1	4
2	2	6
3	1	3
4	0	1

20) Referring to Scenario 13-3, the coefficient of determination is _____. 20) _____

SCENARIO 13-5

The managing partner of an advertising agency believes that his company's sales are related to the industry sales. He uses Microsoft Excel to analyze the last 4 years of quarterly data (i.e., $n = 16$) with the following results:

Regression Statistics	
Multiple R	0.802
R Square	0.643
Adjusted R Square	0.618
Standard Error SYX	0.9224
Observations	16

ANOVA

	df	SS	MS	F	Sig.F
Regression	1	21.497	21.497	25.27	0.000
Error	14	11.912	0.851		
Total	15	33.409			

Predictor	Coef	StdError	t Stat	*p*-value
Intercept	3.962	1.440	2.75	0.016
Industry	0.040451	0.008048	5.03	0.000

Durbin-Watson Statistic 1.59

21) Referring to Scenario 13-5, the coefficient of determination is _____. 21) _____

SCENARIO 13-10

The management of a chain electronic store would like to develop a model for predicting the weekly sales (in thousands of dollars) for individual stores based on the number of customers who made purchases. A random sample of 12 stores yields the following results:

Customers	Sales (Thousands of Dollars)
907	11.20
926	11.05
713	8.21
741	9.21
780	9.42
898	10.08
510	6.73
529	7.02
460	6.12
872	9.52
650	7.53
603	7.25

22) Referring to Scenario 13-10, what is the value of the coefficient of determination? 22) _____

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

SCENARIO 13-1

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 charges to a company is the company's sales revenue (X) —measured in millions of dollars. Data for 21 companies who use the bank's services were used to fit the model:

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

The results of the simple linear regression are provided below.

$\hat{Y} = -2{,}700 + 20X$, $S_{YX} = 65$, two-tail p-value = 0.034 (for testing β_1)

23) Referring to Scenario 13-1, interpret the estimate of σ, the standard deviation of the random error term (standard error of the estimate) in the model. 23) _____

A) About 95% of the observed service charges fall within $65 of the least squares line.
B) For every $1 million increase in sales revenue, we expect a service charge to increase $65.
C) About 95% of the observed service charges equal their corresponding predicted values.
D) About 95% of the observed service charges fall within $130 of the least squares line.

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

SCENARIO 13-3

The director of cooperative education at a state college wants to examine the effect of cooperative education job experience on marketability in the work place. She takes a random sample of 4 students. For these 4, she finds out how many times each had a cooperative education job and how many job offers they received upon graduation. These data are presented in the table below.

Student	CoopJobs	JobOffer
1	1	4
2	2	6
3	1	3
4	0	1

24) Referring to Scenario 13-3, the total sum of squares (*SST*) is _____. 24) _____

25) Referring to Scenario 13-3, the regression sum of squares (*SSR*) is _____. 25) _____

26) Referring to Scenario 13-3, the error or residual sum of squares (*SSE*) is _____. 26) _____

27) Referring to Scenario 13-3, the standard error of estimate is _____. 27) _____

SCENARIO 13-5

The managing partner of an advertising agency believes that his company's sales are related to the industry sales. He uses Microsoft Excel to analyze the last 4 years of quarterly data (i.e., $n = 16$) with the following results:

Regression Statistics	
Multiple R	0.802
R Square	0.643
Adjusted R Square	0.618
Standard Error SYX	0.9224
Observations	16

ANOVA

	df	SS	MS	F	Sig.F
Regression	1	21.497	21.497	25.27	0.000
Error	14	11.912	0.851		
Total	15	33.409			

Predictor	Coef	StdError	t Stat	*p*-value
Intercept	3.962	1.440	2.75	0.016
Industry	0.040451	0.008048	5.03	0.000

Durbin-Watson Statistic 1.59

28) Referring to Scenario 13-5, the standard error of the estimate is _____. 28) _____

29) Referring to Scenario 13-5, the standard error of the estimated slope coefficient is _____. 29) _____

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

30) The standard error of the estimate is a measure of 30) _____
A) the variation of the *X* variable.
B) total variation of the *Y* variable.
C) the variation around the sample regression line.
D) explained variation.

SCENARIO 13-9

It is believed that, the average numbers of hours spent studying per day (HOURS) during undergraduate education should have a positive linear relationship with the starting salary (SALARY, measured in thousands of dollars per month) after graduation. Given below is the Excel output for predicting starting salary (*Y*) using number of hours spent studying per day (*X*) for a sample of 51 students. NOTE: Only partial output is shown.

Regression Statistics	
Multiple R	0.8857
R Square	0.7845
Adjusted R Square	0.7801
Standard Error	1.3704
Observations	51

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	335.0472	335.0473	178.3859	
Residual			1.8782		
Total	50	427.0798			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1.8940	0.4018	-4.7134	0.0000	-2.7015	-1.0865
Hours	0.9795	0.0733	13.3561	0.0000	0.8321	1.1269

Note: 2.051*E* - 05 = $2.051*10^{-05}$ and 5.944*E* - 18 = $5.944*10^{-18}$.

31) Referring to Scenario 13-9, the error sum of squares (SSE) of the above regression is 31) _____
A) 1.878215 B) 92.0325465 C) 427.079804 D) 335.047257

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

SCENARIO 13-11

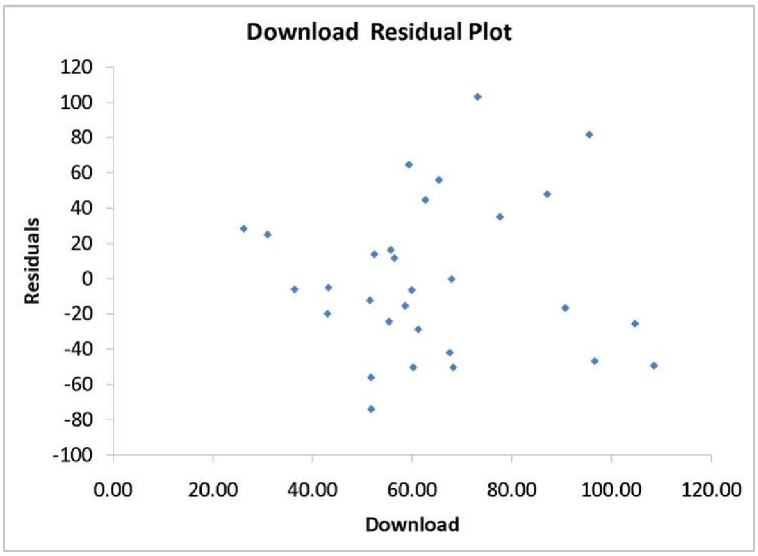
A computer software developer would like to use the number of downloads (in thousands) for the trial version of his new shareware to predict the amount of revenue (in thousands of dollars) he can make on the full version of the new shareware. Following is the output from a simple linear regression along with the residual plot and normal probability plot obtained from a data set of 30 different sharewares that he has developed:

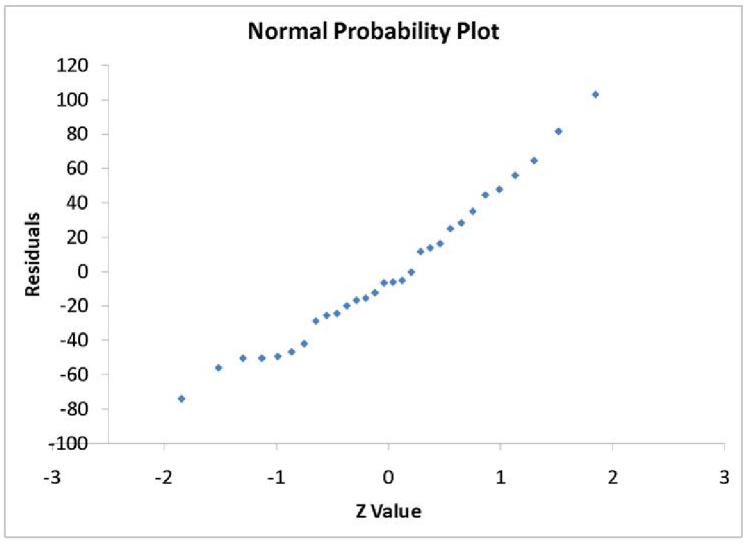
Regression Statistics	
Multiple R	0.8691
R Square	0.7554
Adjusted R Square	0.7467
Standard Error	44.4765
Observations	30.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	171062.9193	171062.9193	86.4759	0.0000
Residual	28	55388.4309	1978.1582		
Total	29	226451.3503			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-95.0614	26.9183	-3.5315	0.0015	-150.2009	-39.9218
Download	3.7297	0.4011	9.2992	0.0000	2.9082	4.5513





32) Referring to Scenario 13-11, what is the standard error of estimate? 32) _____

SCENARIO 13-12

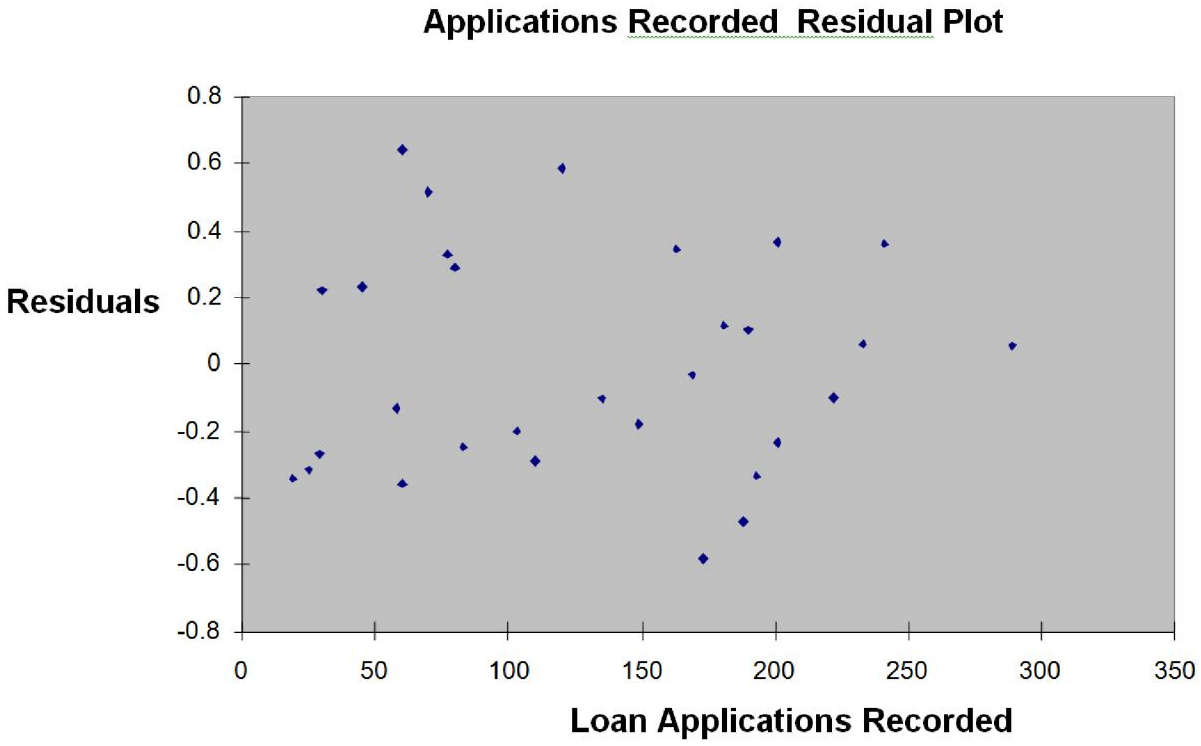
The manager of the purchasing department of a large saving and loan organization would like to develop a model to predict the amount of time (measured in hours) it takes to record a loan application. Data are collected from a sample of 30 days, and the number of applications recorded and completion time in hours is recorded. Below is the regression output:

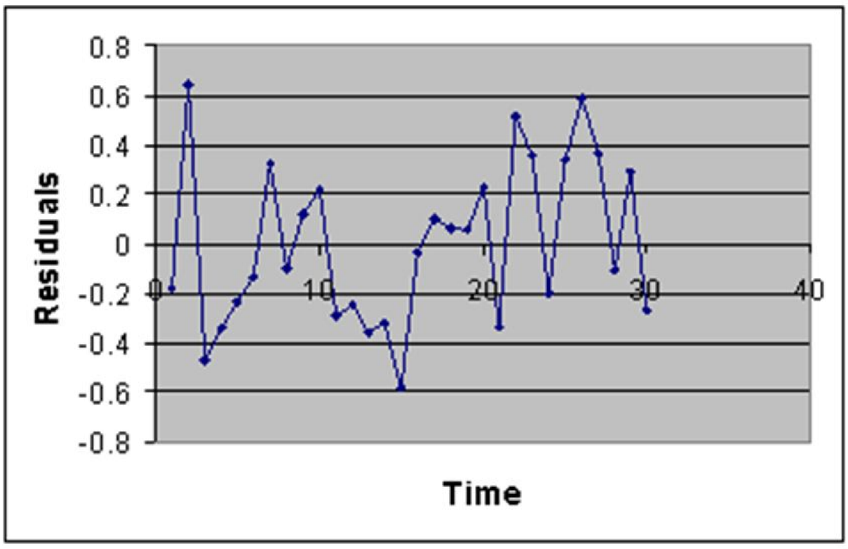
Regression Statistics	
Multiple R	0.9447
R Square	0.8924
Adjusted R Square	0.8886
Standard Error	0.3342
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	25.9438	25.9438	232.2200	4.3946E-15
Residual	28	3.1282	0.1117		
Total	29	29.072			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.4024	0.1236	3.2559	0.0030	0.1492	0.6555
Applications Recorded	0.0126	0.0008	15.2388	0.0000	0.0109	0.0143





33) Referring to Scenario 13-12, what percentage of the variation in the amount of time needed can be explained by the variation in the number of invoices processed? 33) _____

SCENARIO 13-13

In this era of tough economic conditions, voters increasingly ask the question: "Is the educational achievement level of students dependent on the amount of money the state in which they reside spends on education?" The partial computer output below is the result of using spending per student ($) as the independent variable and composite score which is the sum of the math, science and reading scores as the dependent variable on 35 states that participated in a study. The table includes only partial results.

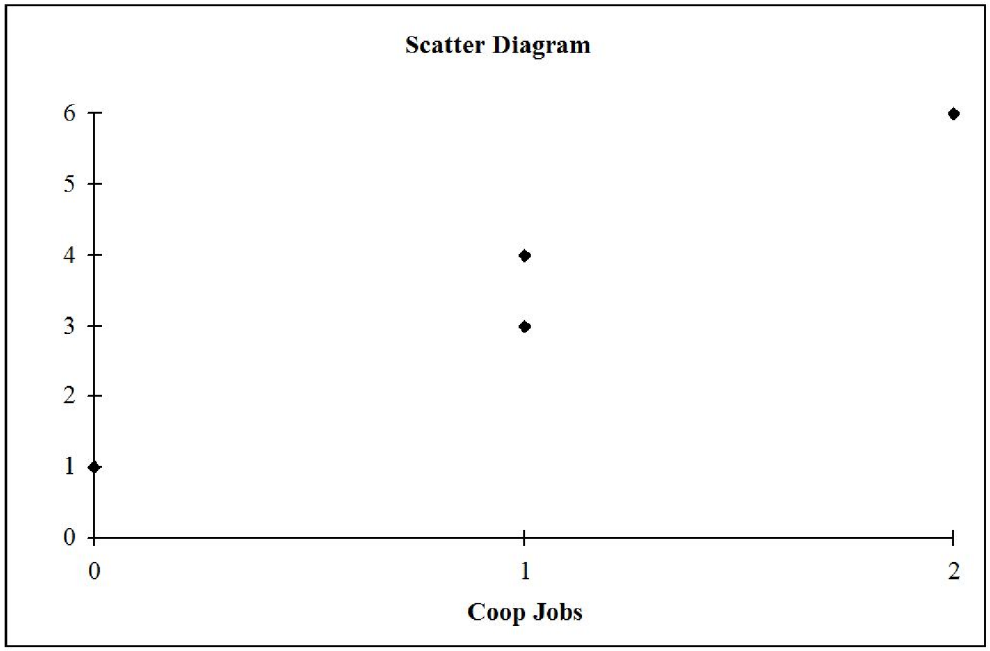
Regression Statistics				
Multiple R	0.3122			
R Square	0.0975			
Adjusted R Square	0.0701			
Standard Error	26.9122			
Observations	35			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	1	2581.5759		
Residual			724.2674	
Total	34	26482.4000		
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	595.540251	22.115176		
Spending per Student ($)	0.007996	0.004235		

34) Referring to Scenario 13-13, the error sum of squares (*SSE*) of the above regression is _____. 34) _____

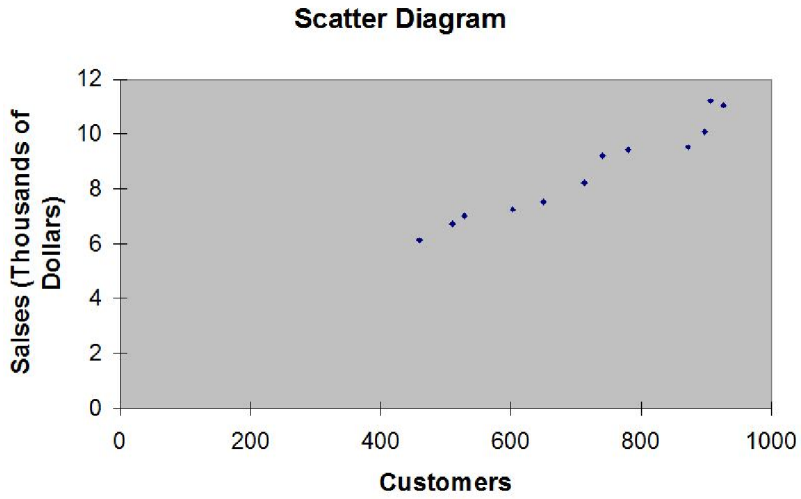
35) Referring to Scenario 13-13, the regression mean square (*MSR*) of the above regression is _____. 35) _____

36) Referring to Scenario 13-13, what is the standard deviation of the composite score around the regression line? 36) _____

1)



2)



3) D
4) D
5) D
6) C
7) C
8) A
9) 1.12
10) 17.7
11) 11.912
12) 3.962 and 0.040451
13) A
14) B
15) A
16) 78.5
17) A
18) D
19) 9.75%

20) 0.962
21) 0.643
22) 0.9453
23) D
24) 13.0
25) 12.5
26) 0.50
27) 0.50
28) 0.9224
29) 0.008
30) C
31) B
32) $44.4765 thousands
33) 89.24%
34) 23,900.8241
35) 2,581.5759
36) 26.9122