Exam

Name____________________________________

Answer the question True or False.

1) 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. 1) ________

A) True B) False

2) A low value of the correlation coefficient r implies that x and y are unrelated. 2) ________

A) True B) False

3) A high value of the correlation coefficient r implies that a causal relationship exists between x and y. 3) ________

A) True B) False

4) 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. 4) ________

A) True B) False

5) The coefficient of correlation is a useful measure of the linear relationship between two variables. 5) ________

A) True B) False

Solve the problem.

6) 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 6) ________

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

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

7) The data for $n = 25$ points were subjected to a simple linear regression with the results: $\hat{\beta}_1 = 0.83$ and $s_{\hat{\beta}_1} = 0.14$. 7) ________________

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.

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

9) 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: 9) ________

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) We are 43% confident that the design pressure will be a useful predictor of number of man–hours required to build a steam drum.
B) About 43% of the sample variation in number of man–hours can be explained by the simple linear model.
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.

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 10) ________

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

The results of the simple linear regression are provided below.

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

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

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

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. 11) ________

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.

12) 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: 12) ________

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

where y = appraised value of the house (in thousands of dollars) and x = number of rooms. Using data collected for a sample of $n = 91$ houses in East Meadow, the following results were obtained:

$\hat{y} = 91.80 + 19.72x$

What are the properties of the least squares line, $\hat{y} = 74.80 + 19.72x$?

A) It is normal, mean 0, constant variance, and independent.
B) All 91 of the sample y–values fall on the line.
C) Average error of prediction is 0, and SSE is minimum.
D) It will always be a statistically useful predictor of y.

13) 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. 13) ________

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

Give a practical interpretation of $r = .81$.

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

14) 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: 14) ______________

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.

15) 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: 15) ________

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

16) 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. 16) _______

$\hat{\beta}_0 = -92040$ $\hat{\beta}_1 = 228$ $s = 3213$ $r^2 = .66$ $r = .81$ 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 increase in SALARY for a 600–point increase in GMAT 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 SALARY of a top business school graduate will fall between $37,915 and $51,984.
D) 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.

17) Suppose you fit a least squares line to 22 data points and the calculated value of SSE is .678. 17) _______________

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?

18) Construct a 90% confidence interval for β_1 when $\hat{\beta}_1 = 49$, $s = 4$, $SS_{XX} = 55$, and $n = 15$. 18) _______________

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. 19) ________

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) −17% B) 70% C) 0% D) 8%

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. 20) ________

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.

Construct the indicated prediction interval for an individual y.

21) The paired data below consists of test scores and hours of preparation for 5 randomly selected students. The equation of the regression line is $\hat{y} = 44.845 + 3.524x$ and the standard error of estimate is $s_e = 5.40$. Find the 99% prediction interval for the test score of a person who spent 7 hours preparing for the test. 21) ________

x Hours of preparation	5	2	9	6	10
y Test score	64	48	72	73	80

A) $62 < y < 78$ B) $32 < y < 107$ C) $58 < y < 82$ D) $35 < y < 104$

Solve the problem.

22) 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: 22) ________

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

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. $H_a: \beta_1 \neq 0$

23) 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. 23) ______________

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 $\bar{y} = 7.4571$. Calculate a 90% confidence interval for $E(y)$ when $x = 5$ months. Assume $s = 1.577$ and the prediction equation is $\hat{y} = -.25 + 1.315x$.

Use the given data to find the equation of the regression line. Round the final values to three significant digits, if necessary.

24) 24) ________

x	0	3	4	5	12
y	8	2	6	9	12

A) $\hat{y} = 4.88 + 0.525x$ B) $\hat{y} = 4.98 + 0.425x$
C) $\hat{y} = 4.98 + 0.725x$ D) $\hat{y} = 4.88 + 0.625x$

1) B
2) B
3) B
4) A
5) A
6) C
7) a. $t = 5.929 > t_{.05} = 1.714$, reject H_0; we concluded that x and y are positively linearly related.
   b. (0.590, 1.070); We can be 90% confident that the true slope is between 0.590 and 1.070.
8) $\text{SSE} = \text{SS}_{\text{yy}} - \hat{\beta}1\text{SS}_{\text{xy}} = 100 - .8(60) = 52$; $s^2 = \dfrac{\text{SSE}}{n-2} = \dfrac{52}{30-2} \approx 1.857$
9) B
10) A
11) $\hat{\beta}1 = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{236.5}{21{,}752} = .01087$

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

   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 $\alpha = .05$.
12) C
13) A
14) $r = \dfrac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} = \dfrac{25}{\sqrt{(10.5)(112)}} = .729$
15) B
16) D
17) 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. $2s = 2(.1841) = .3682$
18) $\hat{\beta}1 \pm t_{.05}\, s_{\hat{\beta}1} = 49 \pm 1.771\left(\dfrac{4}{\sqrt{55}}\right) = 49 \pm .96$
19) B
20) $r^2 = \dfrac{SS_{yy} - SSE}{SS_{yy}} = \dfrac{176.5171 - 12.435}{176.5171} = .9296$

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

22) A

23) 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:

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

24) A