

## Practice Final Exam

### Questions 1-4 are based on the following information:

An insurance company analyst is interested in analyzing the dollar value of damage in automobile accidents. She collects data from 115 accidents, and records the amount of damage as well as the age of the driver. The results of her regression analysis are listed below.

<i>Regression Statistics</i>	
Multiple R	0.187
R Square	0.035
Adjusted R Square	0.026
Standard Error	5652.090
Observations	115.000

#### ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	130433116.219	130433116.219	4.083	0.046
Residual	113	3609911959.868	31946123.539		
Total	114	3740345076.087			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	10725.802	1535.215	6.987	0.000
Age	69.964	34.625	2.021	0.046

- The regression equation is:
  - $\hat{y} = 10725.802 + 69.964x$
  - $\hat{y} = 114 + 113x$
  - $\hat{y} = 1535.215 + 34.625x$
  - $\hat{y} = 113 + 114x$
- How would you best explain the y-intercept in this situation?
  - For each additional 1-year increase in the age of the driver, we would expect damage to increase by \$10,726.
  - For each additional 1-year increase in the age of the driver, we would expect damage to increase by \$70.
  - It makes no sense to explain the intercept in this situation, since we cannot have a driver with an age of zero.
  - The average amount of damage was \$10,726.
- What would be the dollar value of an accident involving a 25-year-old driver?
  - \$11,836.56
  - \$10,795.47
  - \$13,372.58
  - \$12,474.90

## Practice Final Exam

4. Which of the following statements is the best explanation of the  $R^2$ ?
- A) 3.5% of the accident damage can be explained by the age of the driver.
  - B) 3.5% of the variation in accident damage can be explained by variation in the age of the driver.
  - C) 3.5% of the coefficients, t-stat, and p-value can be explained by the age of the driver.
  - D) 3.5% of the total error can be explained by the SSE.
5. The residual is defined as the difference between the actual value of:
- A) y and the estimated value of y.
  - B) x and the estimated value of x.
  - C) y and the estimated value of x.
  - D) x and the estimated value of y.

**Questions 6-10 are based on the following information:**

A sales manager is interested in determining the relationship between the amount spent on advertising and total sales. The manager collects data for the past 24 months and runs a regression of sales on advertising expenditures. The results are presented below but, unfortunately, some values identified by asterisks are missing.

<i>Regression Statistics</i>	
Multiple R	0.492
R Square	0.242
Adjusted R Square	0.208
Standard Error	40.975
Observations	24.000

<b>ANOVA</b>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	11809.406	11809.406	7.034	*
Residual	*	*	*		
Total	*	*			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	*	26.239	4.021	0.001
Advertising	2.015	*	2.652	0.015

6. What are the degrees of freedom for residuals?
- A) 21
  - B) 22
  - C) 23
  - D) 24
7. The value of mean square error (residual) is:
- A) 1678.9
  - B) 1,554.25
  - C) 1,493.63
  - D) 1,407.35

## Practice Final Exam

8. The total degrees of freedom is:
- A) 21
  - B) 22
  - C) 23
  - D) 24
9. A regression analysis between sales (in \$1000) and advertising (in \$100) resulted in the following least squares line:  $\hat{y} = 75 + 5x$ . This implies that if advertising is \$800, then the predicted amount of sales (in dollars) is:
- A) \$4075
  - B) \$115,000
  - C) \$164,000
  - D) \$179,000
10. In a regression problem, a coefficient of determination 0.90 indicates that:
- A) 90% of the y values are positive.
  - B) 90% of the variation in y can be explained by the regression line.
  - C) 90% of the x values are equal.
  - D) 90% of the variation in x can be explained by the regression line.
11. If the least squares regression line  $\hat{y} = -2.88 + 1.77x$  and the coefficient of determination is 0.81, the coefficient of correlation is:
- A) -0.88
  - B) +0.88
  - C) +0.90
  - D) -0.90
12. In a simple regression analysis (where Y is a dependent and X an independent variable), if the Y intercept is positive, then
- A) there is a positive correlation between X and Y
  - B) if X is increased, Y must also increase
  - C) if Y is increased, X must also increase
  - D) None of these alternatives is correct.
13. In regression analysis, the independent variable is
- A. used to predict other independent variables
  - B. used to predict the dependent variable
  - C. called the intervening variable
  - D. the variable that is being predicted

## Practice Final Exam

14. In a regression analysis if  $SSE = 200$  and  $SSR = 300$ , then the coefficient of determination is

- A. 0.6667
- B. 0.6000
- C. 0.4000
- D. 1.5000

15. Regression analysis was applied between sales (in €1000) and advertising (in €100) and the following regression function was obtained.

$$\hat{y} = 500 + 4x$$

Based on the above estimated regression line if advertising is €10,000, then the point estimate for sales (in euros) is

- A. €900
- B. €900,000
- C. €40,500
- D. €505,000

16. A regression analysis between demand (Y in 1000 units) and price (X in euros) resulted in the following equation

$$\hat{y} = 9 - 3x$$

The above equation implies that if the price is increased by €1, the demand is expected to

- A. increase by 6 units
- B. decrease by 3 units
- C. decrease by 6,000 units
- D. decrease by 3,000 units

17. Regression analysis was applied between sales (Y in €1,000) and advertising (X in €100), and the following estimated regression equation was obtained.

$$\hat{y} = 80 + 6.2x$$

Based on the above estimated regression line, if advertising is €10,000, then the point estimate for sales (in euros) is

- A. €62,080
- B. €142,000
- C. €700
- D. €700,000

### **Questions 18 and 19 are based on the following information:**

A large milling machine produces steel rods to certain specifications. The machine is considered to be running normally if the standard deviation of the diameter of the rods is at most 0.15 millimeters. The line supervisor needs to test the machine is for normal

## Practice Final Exam

functionality. The quality inspector takes a sample of 25 rods and finds that the sample standard deviation is 0.19.

18. What are the null and alternative hypotheses for the test?
- $H_0 : \sigma^2 > 0.15$  and  $H_1 : \sigma^2 \leq 0.15$
  - $H_0 : \sigma \geq 0.15$  and  $H_1 : \sigma < 0.15$
  - $H_0 : \sigma = 0.0225$  and  $H_1 : \sigma \neq 0.0225$
  - $H_0 : \sigma^2 \leq 0.0225$  and  $H_1 : \sigma^2 > 0.0225$
19. Which of the following statements is the most accurate?
- Reject null hypothesis at  $\alpha = 0.025$
  - Fail to reject null hypothesis at  $\alpha \leq 0.10$
  - Reject null hypothesis at  $\alpha = 0.05$
  - Reject null hypothesis at  $\alpha = 0.01$
20. Suppose you have the following null and alternative hypotheses:  $H_0 : \sigma^2 = 34.5$  and  $H_1 : \sigma^2 > 34.5$ . If you take a random sample of 15 observations from a normally distributed population and find that  $s^2 = 48.1$ , what is the most accurate statement that can be made about the p-value for this test?
- $0.01 < \text{p-value} < 0.025$
  - $0.025 < \text{p-value} < 0.05$
  - $\text{p-value} < 0.01$
  - $0.05 < \text{p-value} < 0.10$
21. A sample of 20 cans of tomato juice showed a standard deviation of 10.0 grams. A 95% confidence interval estimate of the variance for the population is
- 144.6 to 533.3
  - 139.0 to 495.3
  - 55.6 to 198.1
  - 57.8 to 213.3
22. The producer of bottling equipment claims that the variance of all their filled bottles is 0.027 or less. A sample of 30 bottles showed a standard deviation of 0.2. The p-value for the test is
- between 0.025 to 0.05
  - between 0.05 to 0.01
  - 0.05
  - 0.025

## Practice Final Exam

23. The chi-squared value for a one-tailed (upper tail) hypothesis test of the population standard deviation at 95% confidence and a sample size of 25 is
- 33.196
  - 36.415
  - 39.364
  - 37.652
24. We are interested in testing a claim that the standard deviation of a population is no more than 5.0. The null hypothesis for this test is
- $H_0: \sigma^2 > 25.0$
  - $H_0: \sigma^2 \leq 25.0$
  - $H_0: \sigma^2 \leq 5.0$
  - $H_0: \sigma^2 \geq 5.0$
25. The value of  $F_{0.025}$  with 15 numerator and 10 denominator degrees of freedom is
- 3.06
  - 2.54
  - 2.85
  - 3.52

**Questions 26 and 27 are based on the following information:**

The standard deviation of the ages of a sample of 16 executives from large companies was 8.2 years; while the standard deviation of the ages of a sample of 26 executives from small companies was 12.8 years.

26. What is the test statistic to test if there any difference in the standard deviations of the ages of all the large-company and small-company executives.
- 2.32
  - 2.25
  - 2.21
  - 3.19
27. We would like to test to see if there is any difference in the standard deviations of the ages of all the large-company and small-company executives. What is the critical value of F at 10% significance level?
- 2.12
  - 2.10
  - 2.21
  - 3.10

## Practice Final Exam

### Questions 28-30 are based on the following information:

Suppose two food preservatives are extensively tested and determined safe for use in meats. A processor wants to compare the preservatives for their effects on retarding spoilage. Suppose 16 cuts of fresh meat are treated with preservative A and another 12 cuts of meat are treated with preservative B. The number of hours until spoilage begins is recorded for each of the 28 cuts of meat. The results are summarized in the table below.

	Preservatives A	Preservatives B
Sample means	95.25	100.50
Sample standard deviation	13.45	10.55

28. The value of the test statistic for determining if there is a difference in the population variances for preservatives A and B is equal to:
- e. 0.784
  - f. 1.625
  - g. 1.275
  - h. 1.129
29. The numerator and denominator degrees of freedom associated with the test statistic for determining if there is a difference in the population variances for preservatives A and B are, respectively:
- i. 16 and 12
  - j. 12 and 16
  - k. 15 and 11
  - l. 11 and 15
30. The most accurate statement that can be made about the p-value for testing whether there is a difference in the population variances for preservatives A and B is:
- a.  $p\text{-value} > 0.05$
  - b.  $p\text{-value} = 0.00$
  - c.  $p\text{-value} < 0.01$
  - d.  $0.01 < p\text{-value} < 0.05$
31. An ANOVA procedure is used for data that was obtained from four sample groups each comprised of five observations. The degrees of freedom for the critical value of F are
- a. 3 and 20
  - b. 3 and 16
  - c. 4 and 17
  - d. 3 and 19
32. The analysis of variance is a procedure that allows statisticians to compare three or more population:
- a. Means.
  - b. Proportions.
  - c. Variances.
  - d. Standard deviations.

## Practice Final Exam

33. In analysis of variance, between-treatments variation is measured by the:
- SSE
  - SSTR
  - SST
  - All of the above

**Questions 34-38 are based on the following information:**

Consider the following ANOVA table:

Source of Variation	SS	df	MS	F
Treatments	SSTR=4	2	?	?
Error	SSE= ?	?	?	
Total	SST=34	14		

34. What are the degrees of freedom for the Errors?
- 3
  - 4
  - 13
  - 12
35. What is the value of SSE?
- 29
  - 20
  - 30
  - 33
36. What is the Mean-Square for Treatments?
- 2
  - 1
  - 4
  - 0
37. What is Mean-Square Errors?
- 3
  - 4
  - 3.5
  - 2.5
38. What is the value of the F-statistic?
- 0.5
  - 0.8
  - 1.2
  - 1.6