

ON BEHALF OF ALL THE
SOS TUTORING STAFF,

WELCOME!

Before you begin this mock exam, please read the following instructions:

1. This is a simulated exam. The questions are based on our instructor's review of past exams, experience working with students, and assessment of those topics students have the most troubles with.
2. Review and attempt the SOS Mock Exam at your own pace.
3. You are free to finish as little or as much as you would like. We recommend you attempt each question to the best of your ability.
4. On the day of the mock exam review, the instructor will work through the solutions for the group.

**To make your stay with us as enjoyable as possible,
we please ask that you abide by the following policies:**

1. Turn cell phones off or on vibrate.
2. Ensure your area is clean at the end of the day.

How did we do?

Please remember to fill out the evaluation form found in the back of your package before leaving the exam and drop it off (face down) in the designated area!

Copyright Reminder

The entire SOS Team works very hard to prepare up-to-date, comprehensive materials for your Crash Courses and Mock Exams, and we ask that you respect that effort by not copying, reproducing, or transmitting any part of our Mock Exams or Crash Course Packages.

Multiple-Choice Questions

1. Consider a multiple regression analysis attempting to predict the price of a car from several variables. One variable of interest is the season during which the car is being sold. How many dummy variables would we need to include in this model in order to account for the season?
 - a. 1
 - b. 4
 - c. 2
 - d. 3

2. SSE represents that part of the total variation in the y data values that is _____.
 - a. explained by our linear regression model
 - b. incorporated in the regression model via the use of dummy variables
 - c. not explained by our linear regression model
 - d. none of the above

3. For an F test with the following values: $p\text{-value} = 0.0087$, $R^2 = 0.77$, $s = 0.123$. The null hypothesis should be _____ at the 1% level of significance.
 - a. rejected
 - b. accepted
 - c. not rejected
 - d. cannot be determined

4. Which of the following is *NOT* an assumption for the multiple regression model?
 - a. Normality of error terms
 - b. Constant variation of error terms
 - c. Independence of error terms
 - d. Positive autocorrelation of error terms

5. The range of feasible values for the multiple coefficient of determination is from:
 - a. -1 to 1
 - b. -0.5 to 0.5
 - c. 0 to 1
 - d. 0 to 2

6. If we reject the null hypothesis in the F -test then we can conclude that _____.
- none of our independent variables makes a significant contribution in predicting y in our model and thus our overall model is not a good one
 - at least one of our independent variables makes a significant contribution in predicting y in our model and thus our overall model is a good one
 - all of our independent variables make a significant contribution in predicting y in our model
 - none of the above

7. A simple linear regression gave the following results:

$$\hat{y} = 123.45 + 6.88x, n = 15, MSE = 23.45, SS_{XX} = 9.81$$

What is the value of the test statistic required to test the significance of the slope?

- 16.04
 - 1.55
 - 4.45
 - 1.34
8. In a sample of 40 observations, the sample correlation found to be -0.85. The least squares regression line was computed to be $\hat{y} = 11 - 2x$. This means that the proportion of the variation in y explained by a linear relationship with x is _____.
- 0.7225
 - 0.85
 - 0.9220
 - 0.7225
9. If the simple coefficient of determination $R^2 = 0.64$ and the regression model estimates $b_0 = 87.6$ and $b_1 = -3.69$, then the correlation coefficient is _____.
- 0.64
 - 0.80
 - 0.80
 - 0.64
10. Which of the following is *NOT* an assumption for the simple regression model?
- Normality of error terms
 - Constant variance of error terms
 - Mean of zero of error terms
 - Variance of one of error terms

11. A simple regression model developed for $n = 10$ resulted in a sum of squares of error, $SSE = 164$. The standard error of the estimate is _____.

- a. 16.4
- b. 4.53
- c. 20.5
- d. 4.05

12. A simple linear regression gave the following results:

$$\hat{y} = 123.45 + 6.88x, n = 15, r^2 = 0.6744, s^2 = 0.2934$$

For each unit change in the independent variable x , the estimated change in the mean value of the dependent variable y is equal to:

- a. 0.6744
- b. 6.88
- c. 0.2934
- d. 123.45

13. The _____ measures the strength and direction of the linear relationship between the dependent and the independent variable.

- a. Coefficient of correlation
- b. Coefficient of determination
- c. Y intercept
- d. Distance value

14. Assume that you are working with a sample $n = 50$ distributed over 8 categories. A chi-square test for goodness-of-fit will be used. How many degrees of freedom will be used in determining the chi-square test statistic?

- a. 50
- b. 8
- c. 49
- d. 7

15. If $H_0: p_1 = p_2 = p_3 = p_4 = p_5 = 0.20$. You take a sample of size 1000 and compute your test statistic to be $\chi^2 = 18.52$. At $\alpha = 0.10$, what conclusion would you reach?
- Reject H_0 . Conclude that at least one of p_1, p_2, p_3, p_4 and p_5 is not equal to 0.20
 - Reject H_0 . Conclude that p_1, p_2, p_3, p_4 and p_5 all exceed 0.20
 - Reject H_0 . Conclude that p_1, p_2, p_3, p_4 and p_5 are all less than 0.20
 - Do not reject H_0 . Conclude that $p_1 = p_2 = p_3 = p_4 = p_5 = 0.20$
16. The degrees of freedom for a Test of Independence of 10X10 table is _____.
- 99
 - 100
 - 98
 - 81
17. What would be the best test to determine if beer preference is really independent of the color of hair?
- Simple regression
 - Chi square Contingency test
 - Chi square Goodness of fit
 - Multiple Regression
18. If the population mean is \$87.60, what is the expected value of the sampling distribution?
- \$87.60
 - Less than \$87.60
 - More than \$87.60
 - Cannot be determined
19. The sampling distribution of the sample means must contain _____ possible samples of a given size.
- all
 - most
 - some
 - none of the above

20. According to the central limit theorem, if samples of size 100 are drawn from a population with a mean of 85 and standard deviation of 25, the mean of all sample means would equal ____.
- 0.85
 - 85
 - 8.5
 - 2.5
21. According to the central limit theorem, if samples of size 100 are drawn from a population with a mean of 85 and standard deviation of 25, the variance of all sample means would equal ____.
- 2.5
 - 6.25
 - 62.5
 - 225
22. As the n increases, the sampling distribution of the sample mean approaches a _____.
- Normal distribution
 - Binomial distribution
 - Discrete distribution
 - Poisson distribution
23. NAYA measures the amount of water in bottles before shipping them to the customers. If the water amounts in bottles have a population mean and standard deviation of 500 ml and 20 ml, respectively then, based on a sample size of 25 bottles, the probability that the sum will be between 12,425 ml and 12,625 ml is _____.
- 0.1583
 - 0.3322
 - 0.8417
 - 0.6678
24. Assuming that the speed of cars travelling on the 520 is normally distributed with a mean of 105 km/h and a standard deviation of 4 km/h, what is the probability that a car is going less than 113 km/h given that is going above the posted speed limit of 100 km/h?
- 0.8944
 - 0.9772
 - 0.9745
 - 0.8716

25. Andrew has determined that the grades of the assignments are normally distributed with a mean of 80% and a standard deviation of 4%. What is the minimum number of students he needs enrolled in the two classes if he wants a maximum error of 1% with a 98% confidence level?
- a. 87
 - b. 8687
 - c. 62
 - d. 133

Long-Answer Questions

1. The scores of all applicants taking a GMAT test required by a graduate business school follow a normal distribution with mean 620 and standard deviation 100.
 - a. What is the probability that a randomly selected applicant will score less than 650 in this test?

- b. What is the probability that two randomly (and independently) selected applicants will both score less than 650 in this test?

- c. What is the probability that out of two randomly (and independently) selected applicants, at least one will score less than 650 in this test?

- d. A random sample of 16 scores is taken. What is the probability that the mean score of this sample of 16 applicants is less than 650?

- e. If a sample of 25 applicants had been taken (instead of 16), would the probability of a sample mean being less than 650 be smaller than, larger than, or the same as the correct answer to part d. above?

- f. What is the probability that the mean score of a sample of 25 applicants will be between 600 and 650?

- g. There is a 5% probability that the mean score of a random sample of 25 applicants is higher than what number?

- h. There is a 85% probability that the mean score of a random sample of 25 applicants is less than what number?

2. The regional manager for Pizza Domino's Canada is investigating why some restaurants in the region are performing better than others. His consultants, being statisticians, found that three factors may be related to total sales: x_1 , the number of competitors in the region; x_2 , the population in the surrounding area (in millions); and x_3 , the amount spent on advertising (in thousands of dollars). A random sample of 30 restaurants in the region was selected and the following multiple regression equation was obtained:

$$Y = 14.0 - 1.0x_1 + 30.0x_2 + 0.2x_3$$

Where y =total sales last year in thousands of dollars. Additional information obtained included the estimated standard errors of the sample regression coefficients that are

$S_{b_1} = 0.70$, $S_{b_2} = 5.20$, and $S_{b_3} = 0.20$, as well as $SST = 5420$ and $MSE = 64.82$.

- a. What are the estimated sales for the Pizza Domino's that has 3 competitors, a regional population of 500,000, and an advertising expense of \$25,000.

- b. Compute the value of the standard error of the estimate.

- c. Compute the value of the coefficient of determination. Explain the meaning in the context of the problem.

- d. Is there sufficient evidence, at the 5% level of significance, to conclude that the model is significant?

- e. Is there sufficient evidence, at the 5% level of significance, to conclude that the number of competitors in the region is related to sales? What is the approximate p-value?

3. A consumer agency that proposes that lawyer rates are too high wanted to estimate the mean hourly rate for all lawyers in Montreal. A sample of 70 lawyers taken from the Montreal area showed that the mean hourly rate charged by them is \$203 with a standard deviation of \$55.
 - a. At the 10% significance level, can it be concluded that that the mean hourly rate of lawyers is higher than \$200? Compute the p-value.

- b. Construct a 90% confidence interval for the mean hourly rate for all lawyers in the Montreal area. Interpret your result.

- c. How many more lawyers would have to be sampled if we want a 99% confidence interval with the same margin of error as the margin of error of part b)?

- d. How large should you take a sample to ensure that the upper limit of the 95% confidence interval of the mean hourly rate does not exceed \$210?

4. The results of a sample of the participants who have filled out a questionnaire regarding their access to personal computers are summarized in the following table.

	North America	Europe	Asia
No Access	21	18	16
Home Access	12	17	21
Other Access	27	25	43

- a. Test, using a 10 percent level of significance, whether the type of access to personal computers depends significantly on where the person lives.

- b. The manager claims that fewer than 30% of all participants are from Europe.
Test the manager's claim.

- c. How large a sample would have to be taken in order to establish a 98% confidence interval for the true proportion of those with home computer access if the margin of error is 0.024?

- d. The manager said he believes that 40%, 20%, and 40% of participants have no access, home access, and other access respectively. Test his claim using $\alpha = 0.01$

- e. Construct the 99% confidence interval for the true proportion of Europeans.

5. The following table shows nine months of data on the price charged for an exotic fruit and its sales volume.

Price (\$)	Sales (hundreds of cases)
0.51	4.2
0.46	5.4
0.36	7.2
0.35	9.1
0.35	8.9
0.40	9.0
0.41	8.5
0.46	7.7
0.56	4.4

Calculations for the above data are provided:

$$\begin{array}{lll} \sum x = 3.86 & \sum y = 64.4 & \sum xy = 26.609 \\ \sum x^2 = 1.6996 & \sum y^2 = 492.56 & \end{array}$$

- a. Test at the 5% level whether price is a significant predictor of sales volume.

- b. Find a 95% prediction interval for the sales volume if price is 55 cents.

- c. On the average, how does the sales volume change with a 5-cent price increase?

- d. What proportion of the variation of sales volume is explained by the linear regression model?

