

1. You performed a right tailed hypothesis test and your sample gave you a test statistic in the left tail. Which of the following is considered the appropriate next step in this hypothesis test?
  - A. Fail to reject the null.
  - B. Change the direction of the hypotheses.
  - C. Change the level of significance of your test.
  - D. Change the sample and do the test again.
  - E. All of the above.
  
2. You bought a 25 pack of blank CD-R discs. The rumor is that the manufacturer has some quality control problem; some batches of their CD-R discs are real good, while others are defective. Overall there is a 50-50 chance that a given disc burns and plays flawlessly. You would like to know the chance that all of your 25 discs will work great. Your E370 team member said, "Well it's simple--just use =BINOM.DIST(25,25,0.5,0)." You treat your team member to a cold brewski for such a brilliant idea and as you stumble home it hits you: Problem! What was it you remembered when you thought of a problem with the BINOM.DIST solution?
  - A. The result of one trial is not independent of the results of the other trials.
  - B. The number of trials should have been at least 30, but it's only 25.
  - C. The outcome of each trial is not one of two mutually exclusive categories.
  - D. The random variable is not discrete, as a binomial variable should be.
  
3. A simple linear regression generated a correlation coefficient of 0.01. This tells us that
  - A. we shall reject the null at less than a 5% significance level.
  - B. the two variables are barely related to each other.
  - C. SSE is almost zero.
  - D. SSR is almost zero.
  
4. In order to reduce a standard error by half, what must be done to the sample size,  $n$ ?
  - A. multiply  $n$  by 2.
  - B. multiply  $n$  by  $\frac{1}{2}$ .
  - C. multiply  $n$  by 4.
  - D. multiply  $n$  by  $\frac{1}{4}$ .

**Classic Catering, Inc.** is updating their pricing structure for Holiday Parties. Past experience has shown them that guests at holiday parties consume the following quantities of food items on average. Cost per item is also designated. Use this information to answer the next **THREE** questions.

| Item            | Cost | Mean | Variance |
|-----------------|------|------|----------|
| Cocktail Shrimp | .75  | 5.5  | 7        |
| Cookies         | .10  | 7.2  | 3        |
| Cups of Punch   | .05  | 2.3  | 1.5      |

5. What is the expected cost for the average guest?
- A. \$0.90      B. \$4.96      C. \$15.00      D. \$20.86
6. Assuming that the consumption of each item is independent of every other item, what is the standard deviation of the expected cost per guest?
- A. \$1.99      B. \$2.37      C. \$3.97      D. \$5.63
7. Cocktail Shrimp appears to be independent of the other items. However, when a guest eats more cookies the guest also drinks more punch. How would this relationship affect the standard deviation you calculated above?
- A. It would make the standard deviation larger.  
 B. It would make the standard deviation smaller.  
 C. It would have no effect.  
 D. It is impossible to tell if there would be an effect.
8. Ceteris paribus, which is wider, a 95% confidence interval calculated with a known  $\sigma$  or a 99% confidence interval calculated with an unknown  $\sigma$ ?
- A. The 99% interval with the unknown  $\sigma$  is wider.  
 B. The 95% interval with the known  $\sigma$  is wider.  
 C. The intervals are the same width.  
 D. The size of the standard deviation is needed to decide.

A sample of former E370 students' starting salaries was drawn with  $n=25$  and a confidence interval was calculated. Such salaries are approximately normally distributed. The mean of the sample was \$49,500 and its standard deviation was \$2,427. The margin of error for the confidence interval was \$1000. Use this information to answer the following **THREE** questions. Some useful numbers are these:

|                    |      |
|--------------------|------|
| =NORM.S.INV(0.95)  | 1.64 |
| =NORM.S.INV(0.975) | 1.96 |
| =NORM.S.INV(0.98)  | 2.05 |
| =NORM.S.INV(0.99)  | 2.33 |
| =T.INV(0.975,24)   | 2.06 |
| =T.INV(0.99,24)    | 2.49 |

9. What is the distribution of the random variable described above?
- A.  $\bar{X} \sim N(\$49,500, \$2,427)$
- B.  $\bar{X} \sim N(\$49,500, \sqrt{\frac{\$2,427}{25}})$
- C.  $\bar{X} \sim N(\$49,500, \frac{\$2,427}{\sqrt{25}})$
- D. Insufficient information.
10. What is the level of confidence of the interval described above?
- A. 99%      B. 95%      C. 98%      D. None of the above.
11. You would like a confidence interval that is \$1000 wide. How large a sample would be required in order to be 98% confident of your estimate?
- A. 147      B. 128      C. 37      D. 32

Human intelligence (IQ level) is normally distributed with mean of 100. A researcher measures the IQ of 9 people and finds that their average IQ is 102 with a standard deviation of 20. Use this information to answer the next **TWO** questions.

12. What percent of groups of 9 people should the researcher expect to have IQ levels more than 102 on average?

- A. =1-NORM.S.DIST(0.1,1)
- B. =1-NORM.S.DIST(0.3,1)
- C. =1-T.DIST(0.1, 8, 1)
- D. =1-T.DIST(0.3, 8, 1)

13. Of the two commands below, which will underestimate the likelihood that the mean IQ of 9 people will be more than 125?

|                        |                       |
|------------------------|-----------------------|
| =1-NORM.S.DIST(3.75,1) | =1-T.DIST(3.75, 8, 1) |
|------------------------|-----------------------|

- A. The TDIST command.
  - B. The NORMSDIST command.
  - C. The estimates will be identical.
  - D. Excel is required to answer this question.
14. For a continuous random variable X, which of the following is true?
- A. The probability of any particular value occurring at random is 0.
  - B. The values it can assume are limited by the number of times the experiment can be performed.
  - C. The expected value is determined by adding up the number of values that the variable can assume and dividing by the number of values.
  - D. The number of values it can assume is finite or countably infinite.

A fruit company interviews 100 people and finds that 50% of Americans like mangoes. Answer the following **SIX** questions based on this information. **Note: the questions continue onto the next page.**

15. Assuming that the survey is accurate, in a random sample of 20 American adults, what is the likelihood that 7 or more of them like mangoes?

- A. =1-BINOM.DIST(7,20,0.5,1)
- B. =1-BINOM.DIST(6,20,0.5,1)
- C. =BINOM.DIST(7,20,0.5,1)
- D. =BINOM.DIST(6,20,0.5,1)

16. Is it appropriate to approximate the probability calculated in the previous question using the Normal Distribution?
- A. No, the Normal is a continuous distribution.
  - B. Yes, the Normal is a discrete distribution.
  - C. No, the distribution above is not sufficiently symmetric to use the Normal.
  - D. Yes, the parameters of the distribution above give a sufficiently large number of expected successes and expected failures to use the Normal.
17. Assuming that the poll is accurate, in a random sample of 15 Americans, what is the likelihood that more than 8 but less than 12 of them will report that they like mangoes?
- A. =BINOM.DIST(11,15,0.5,1)-BINOM.DIST(7,15,0.5,1)
  - B. =BINOM.DIST(12,15,0.5,1)-BINOM.DIST(7,15,0.5,1)
  - C. =BINOM.DIST(12,15,0.5,1)-BINOM.DIST(8,15,0.5,1)
  - D. =BINOM.DIST(11,15,0.5,1)-BINOM.DIST(8,15,0.5,1)
18. What will be the standard deviation of the above binomial distribution (n=15)?
- A. 1.53
  - B. 1.84
  - C. 1.94
  - D. 1.20
19. A 99% confidence interval constructed from the results of this survey is [0.471, 0.628]. It is correctly interpreted as which of the following?
- A. We are confident that between 47.1% and 62.8% of Americans truly like mangoes 99% of the time.
  - B. We are between 47.1% and 62.8% confident that a randomly selected American will like mangoes 99% of the time.
  - C. We are 99% confident that the true proportion of Americans who like mangoes is between 47.1% and 62.8%.
  - D. We are 99% confident that a sample of Americans will have a proportion which likes mangoes between 47.1% and 62.8% of the time.
20. Using the 99% interval estimate mentioned above, [0.471, 0.628], what conclusion would we reach if we test  $H_0: \pi = 0.56$  versus  $H_1: \pi \neq 0.56$  at  $\alpha = 0.01$ ?
- A. Reject the alternative and conclude the null.
  - B. Fail to reject the null.
  - C. Reject the null and conclude the alternative.
  - D. Conclude the alternative.
  - E. We cannot tell what our decision will be from the information given.

Mama Jane's Pizzerias make only 14 inch pizzas. The pizza size is actually a normally distributed random variable with a mean of 14 inches and a standard deviation of 0.1 inches. Use this information to answer the next ***FIVE*** questions.

21. What is the largest diameter of the 60% of Mama Jane's pizzas symmetric around the mean?
- A. =NORM.INV(0.8,14,0.1)  
 B. =NORM.INV(0.6,14,0.1)  
 C. = NORM.INV(0.2,14,0.1)  
 D. = NORM.INV(0.4,14,0.1)
22. What proportion of pizzas has a diameter between 14 inches and 14.5 inches?
- A. =0.5 - NORM.DIST(14.5,14,0.1,1)  
 B. =NORM.DIST(14,14,0.1,1) - NORM.DIST(14.5,14,0.1,1)  
 C. =NORM.DIST(14.5,14,0.1,1)+ NORM.DIST(14,14,0.1,1)  
 D. =NORM.DIST(14.5,14,0.1,1) - NORM.DIST(14,14,0.1,1)
23. An unacceptably small pizza is a pizza smaller than 13.75 inches in diameter. What is the Z-score of the smallest pizza that is acceptable?
- A. -2.5    B. -1.25    C. -0.25    D. 1.25    E. 2.5
24. Which is less likely, the probability that a random sample of 16 pizzas will have a mean diameter greater than 14.1 inches or the probability that a random sample of 36 pizzas will have a mean diameter greater than 14.1 inches?
- A. The probability that a random sample of 16 pizzas will have a mean diameter greater than 14.1 inches is smaller.  
 B. The probability that a random sample of 36 pizzas will have a mean diameter greater than 14.1 inches is smaller.  
 C. The two probabilities are equal.  
 D. There is insufficient information to answer this question.
25. Mama Jane is tracking the type of drink her customers order with their pizzas. A sample of 79 customers selected at random times on random days revealed that 31 of the customers ordered the house beer with their pizzas. Calculate a 91% confidence interval for the proportion of Mama Jane's customers prefer house beer with their pizza. One of the following numbers will be useful.
- =T.INV(0.955,78) = 1.717                      =NORM.S.INV(.955) = 1.695  
 =T.INV(0.91,78) = 2.037                      =NORM.S.INV(0.91) = 1.341
- A. [0.281, 0.504]  
 B. [0.298, 0.487]  
 C. [0.299, 0.486]  
 D. [0.319, 0.466]

At Cheapie-Cheap Grocery Store, a random sample of 100 apples was taken. Of those apples, 6 turned out to be rotten. The quality control manager there claims that no more than 5% of apples are rotten. You wish to test this claim. Use this information to answer the next six questions.

26. The correct set of hypotheses for the test described above is which of the following?
- |    |                      |                      |
|----|----------------------|----------------------|
| A. | $H_0: \pi = 0.05$    | $H_1: \pi \neq 0.05$ |
| B. | $H_0: p \leq 0.06$   | $H_1: p > 0.06$      |
| C. | $H_0: \pi \leq 0.05$ | $H_1: \pi > 0.05$    |
| D. | $H_0: p = 0.06$      | $H_1: p \neq 0.06$   |
27. The p-value to use to test the null hypothesis is
- |    |   |
|----|---|
| A. | $=2*(1-T.DIST(((0.06 - 0.05)/0.022),99,1))$   |
| B. | $=1-T.DIST(((0.06 - 0.05)/0.022),99,1)$       |
| C. | $=2*(1 - NORM.S.DIST((0.06 - 0.05)/0.022),1)$ |
| D. | $=(1 - NORM.S.DIST((0.06 - 0.05)/0.022),1)$   |
28. If the p-value from the above test was 0.068, the null hypothesis would be
- |    |   |
|----|---|
| A. | rejected at $\alpha=0.05$ but not rejected at $\alpha=0.10$ .     |
| B. | not rejected at $\alpha=0.05$ but rejected at $\alpha=0.10$ .     |
| C. | not rejected at $\alpha=0.05$ and not rejected at $\alpha=0.10$ . |
| D. | rejected at $\alpha=0.05$ and rejected at $\alpha=0.10$ .         |
29. On the basis of a p-value of 0.004, the quality manager at Cheapie-Cheap Grocery Store rejected the null hypothesis when the population proportion of rotten apples was actually 0.05. We would say that the quality manager has
- |    |                                 |
|----|---------------------------------|
| A. | committed a Type I error.       |
| B. | committed a Type II error.      |
| C. | made a correct decision.        |
| D. | decided not to make a decision. |
30. The area under the Standard Normal from -2 to -3 is always \_\_\_\_\_ the area under the Student's t distribution from -2 to -3.
- |    |   |
|----|---|
| A. | the same as                                   |
| B. | larger than                                   |
| C. | smaller than                                  |
| D. | The answer depends on the degrees of freedom. |

A researcher is hired by a relief organization which wishes to understand and be able to predict the number of deaths related to an earthquake. The researcher suggests a model that explains number of deaths (in thousands) by the magnitude of the earthquake as measured by the Richter Scale, whether the earthquake resulted in a tsunami (1=tsunami, 0=no tsunami), the population of the area (in thousands), and a measure of the quality of the building techniques used in the area (1=low quality, 0 = high quality). A partial regression output from Excel is given below. Answer next few questions based on this information.

## SUMMARY OUTPUT

Regression Statistics

|                |      |
|----------------|------|
| Multiple R     |      |
| R Square       | 0.94 |
| Adj R Square   |      |
| Standard Error |      |
| Observations   | 111  |

## ANOVA

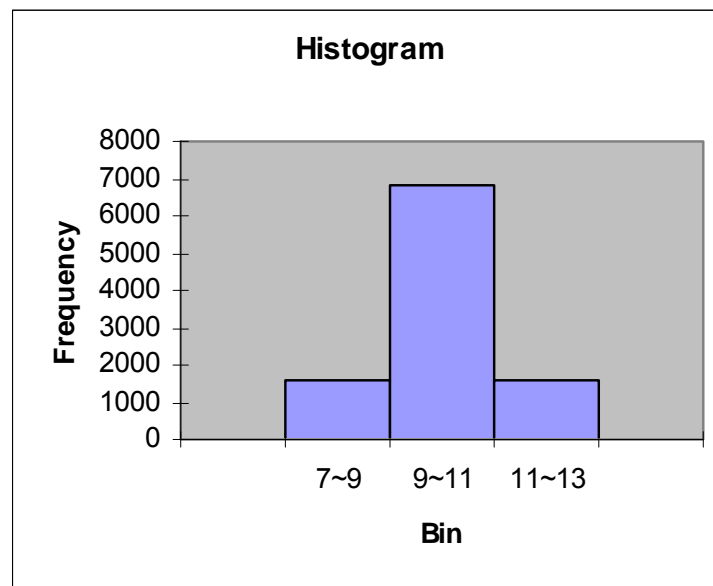
|            | <i>df</i> | <i>SS</i> |
|------------|-----------|-----------|
| Regression | 4         | 955461.13 |
| Residual   |           |           |
| Total      | 110       |           |

|            | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|------------|---------------------|-----------------------|---------------|----------------|------------------|------------------|
| Intercept  | -104.41             | 91.11                 | -1.15         | 0.2543         |                  | 76.17            |
| Tsunami    |                     | 7.74                  | 3.90          | 0.0174         | 14.85            | 45.53            |
| Magnitude  | 12.11               | 13.28                 | 0.91          | 0.3640         | -14.22           |                  |
| Population | 0.93                |                       | 35.02         |                | 0.88             | 0.98             |
| Housing    | 43.30               | 4.61                  | 9.40          | 0.0004         |                  | 52.44            |

31. What is the value of the total sum of squares from the regression?
- 1971909.14
  - 1016448.01
  - 955461.13
  - 60986.88
32. If the relief organization's purpose was simply to be able to predict deaths from earthquakes, which of the variables might they be justified in eliminating from the model?
- Tsunami and Magnitude
  - Magnitude and Population
  - Magnitude only
  - Tsunami only
  - Population only

33. The president of the relief organization is thinking of using the magnitude of the earthquake on the Richter Scale to determine the amount of aid a region hit by an earthquake will receive. However, the president will only suggest the idea to the board of directors if there is statistical evidence of the significance of Magnitude. Based on the above output, what should the president of the relief organization do?
- A. Reject the null that Magnitude is unimportant in determining the number of deaths from an earthquake and suggest the idea to the board.
  - B. Reject the null that Magnitude is unimportant in determining the number of deaths from an earthquake and do not mention the idea to the board.
  - C. Fail to reject the null that Magnitude is unimportant in determining the number of deaths from an earthquake and suggest the idea to the board.
  - D. Fail to Reject the null that Magnitude is unimportant in determining the number of deaths from an earthquake and do not mention the idea to the board.
34. The slope coefficient on Housing, 43.30, means
- A. Each additional low quality house increases deaths from earthquake by 43.3 on average.
  - B. Each additional high quality house increases deaths from earthquake by 43.3 on average.
  - C. Compared to high quality housing, low quality housing reduces deaths from earthquake by 43.3 thousand on average.
  - D. Compared to low quality housing, high quality housing reduces deaths from earthquake by 43.3 thousand on average.
35. The researcher argues that the slope coefficient on Tsunami will never be negative, and would like to know if it is greater than zero, a one tailed test. What is the p-value for this test?
- A. 0.0087      B. 0.0174      C. 0.0348      D. 0.05
36. The interpretation of  $R^2$  for this regression, 0.94, is
- A. the ratio of number of deaths due to earthquake to the variation in Magnitude, Tsunami, Population and Housing is 0.94.
  - B. the ratio of the variation in Magnitude, Tsunami, Population and Housing to number of deaths due to earthquake is 0.94.
  - C. the proportion of the variation of Magnitude, Tsunami, Population and Housing explained by variation in number of deaths due to earthquake is 0.94.
  - D. the proportion of variation of number of deaths due to earthquake to the variation in Magnitude, Tsunami, Population and Housing is 0.94.

37. Based on the output, the lower limit of the 95% confidence interval on Housing is
- A. 9.14      B. 18.28      C. 34.16      D. 43.3
38. Two samples, size  $n$  and  $2n$ , are drawn at random from the same population. If  $n$  is at least 30, which of the following statements are correct?
- A. Sample means from samples of size  $n$  will be approximately normally distributed, but sample means from samples of size  $2n$  will be exactly normally distributed.
- B. Samples of size  $2n$  will be exactly normally distributed, but samples of size  $n$  will be approximately normally distributed.
- C. Sample means from samples of size  $n$  will be centered at  $\mu$ , will be approximately normally distributed, and have a smaller standard deviation than will samples of size  $2n$ .
- D. Sample means from samples of size  $2n$  will be centered at  $\mu$ , will be approximately normally distributed, and will have a smaller standard deviation than samples of size  $n$ .
39. The following graph is a histogram of residuals from a regression. Based on the graph, which of the following regression assumptions have been violated in the regression?



- A. The assumption of multicollinearity.
- B. The assumption of homoscedasticity.
- C. The assumption that the distribution of  $\varepsilon$  is normal.
- D. The assumption that the distribution of  $\varepsilon$  is centered at 0.

40. Which of the following statements about multiple regression is TRUE?
- A. The total sum of squares in a regression model will never exceed the regression sum of squares.
  - B. The coefficient of multiple determination is calculated by taking the ratio of the regression sum of squares over the total sum of squares and subtracting that value from 1.
  - C. If we have taken into account all relevant explanatory factors, the residuals from a multiple regression should be random.
  - D. A multiple regression is called "multiple" because it has several data points.

| <b>Answer Key Sample Final C</b> |   |    |   |    |   |    |   |
|----------------------------------|---|----|---|----|---|----|---|
| 1                                | A | 11 | B | 21 | A | 31 | B |
| 2                                | A | 12 | D | 22 | D | 32 | C |
| 3                                | B | 13 | B | 23 | A | 33 | D |
| 4                                | C | 14 | A | 24 | B | 34 | D |
| 5                                | B | 15 | B | 25 | C | 35 | A |
| 6                                | A | 16 | D | 26 | C | 36 | D |
| 7                                | A | 17 | D | 27 | D | 37 | C |
| 8                                | A | 18 | C | 28 | B | 38 | D |
| 9                                | C | 19 | C | 29 | A | 39 | D |
| 10                               | B | 20 | B | 30 | C | 40 | C |