Essex County Math League
May 22, 2019
Statistics

Directions: You may write on this test. Be sure that your name, subject, and school (including town name) are on the answer sheet. Mark the answer sheet with dark, careful marks using a #2 pencil. Your score will be determined by your number of correct answers, incorrect answers will NOT lower your score. You MAY only use a calculator on this test that is approved for use on the SAT's. The answer to the tie-breaker should be placed on the answer sheet in the place indicated by the proctors. The tie-breaker will be scored only in the case of a tie between the top scorers, and will not count as part of the team score. The fifth choice for each question is, NG, which means, "not given" and is a valid answer that indicates that the correct answer is not among the answers given.

1. Assume the standard deviation σ of a normally distributed population is known to be 12. A one-sample z-test for the mean is conducted to test the hypotheses $H_0: \mu = 80$ vs. $H_a: \mu > 80$ using a sample of size $n = 100$. Which of the following is the minimum sample mean that would result in a decision to reject the null hypothesis at a level of significance of $\alpha = 0.05$?

A) 99.741
B) 78.021
C) 81.975
D) 82.353
E) 83.092

2. A high school guidance counselor was interested in finding out how well student grade point averages (GPA) predict ACT scores. The data of a sample of 15 seniors was reviewed to obtain their GPA and ACT scores. Part of the regression output is given below.

variable	**coefficient**	**s.e. of coeff**	**t-ratio**	**prob**
Constant	-0.427035	3.382	-0.126	0.9011
GPA	7.30697	1.087	6.80	<0.0001

Which of the following is the 95% confidence interval estimate for the true slope of the least-squares regression line relating GPA and ACT scores?

A) (-0.085, 14.699)
B) (4.959, 9.655)
C) (5.401, 9.212)
D) (5.176, 9.437)
E) (6.220, 8.394)

3. A fair six-sided die has faces numbered 1,2,3,4,5, and 6. Consider the following two games:
Game A: The die is rolled thrice and the total score is the sum of the faces of all three rolls.
Game B: The die is rolled once and the total score is triple the face of the single roll.
Which of the following is true about the mean and standard deviation of the expected score of Game A compared to Game B?

A) Game A and Game B have the same mean and the same standard deviation.
B) Game A and Game B have the same mean; Game A has a larger standard deviation.
C) Game A and Game B have the same mean; Game B has a larger standard deviation.
D) Game A and Game B have the same standard deviation; Game A has a larger mean.
E) Game A and Game B have the same standard deviation; Game B has a larger mean.

4. A researcher conducted a two-tailed hypothesis test on a set of data and obtained a *p-value* of 0.38. If the researcher had conducted a one-tailed test on the same set of data, which of the following is true about the possible *p*-value(s) that the researcher could have obtained?

A) The only possible *p*-value is 0.38.
B) The only possible *p*-value is 0.19.
C) The only possible *p*-value is 0.62.
D) The possible *p*-values are 0.38 and 0.62.
E) The possible *p*-values are 0.19 and 0.81.

5. We would like to conduct a test of significance to determine if New Jersey has a statistically significant higher graduation rate than New York. A random sample of 1,200 New Jersey public high school students found that 1,076 of them graduated high school within 4 years and a random sample of 1,800 New York public high school students found that 1,426 of them graduated high school within 4 years. Which of the following gives the correct equation for the test statistic when testing the hypotheses $H_0: p_{NJ} = p_{NY}$ versus $H_0: p_{NJ} > p_{NY}$?

A) $z = \dfrac{(0.897-0.792)-0}{\sqrt{(0.834)(0.166)\left(\frac{1}{1200}+\frac{1}{1800}\right)}}$

B) $z = \dfrac{0.897-0.792}{\sqrt{\frac{(0.834)(0.166)}{3000}}}$

C) $z = \dfrac{(0.897-0.792)-0}{\sqrt{\frac{(0.897)(0.103)}{1200}-\frac{(0.792)(0.208)}{1800}}}$

D) $z = \dfrac{(1076-1426)-0}{\sqrt{\frac{0.897^2}{1200}+\frac{0.792^2}{1800}}}$

E) $z = \dfrac{0.897-0.792}{\frac{0.834^2}{\sqrt{3000}}}$

6. Suppose the total first-year cost of owning a dog averages $1,270 with a standard deviation of $381 and the total first-year cost of owning a cat averages $1,070 with a standard deviation of $321. What is the expected value for the first-year cost of a family that owns one dog and two cats? With what standard deviation?

A) $3,410; $55.35
B) $3,410; $592.66
C) $3,410; $746.54
D) $3,610; $627.19
E) Not enough information is given to determine the standard deviation.

7. Assume that the heights of men are normally distributed with a mean of 69.0 inches and a standard deviation of 2.8 inches. Assume that the heights of women are normally distributed with a mean of 63.6 inches and a standard deviation of 2.5 inches. At a conference of 2000 attendees, 1180 are women and 820 are men. A conference attendee is selected at random. Assuming the attendees match these population parameters, what is the probability he or she is more than 67 inches tall?

A) 0.2419
B) 0.3639
C) 0.4855
D) 0.6361
E) 0.5145

8. Independent samples from two populations are taken, and a *t*-statistic is used to test the null hypothesis $H_0: \mu_1 = \mu_2$ against the alternative $H_0: \mu_1 < \mu_2$. The resulting *p-value* is 0.016. Using the same samples, a two-sided confidence interval will now be constructed for $\mu_1 - \mu_2$. Of the following, which is the smallest confidence level for which the interval will contain zero?

A) 90% confidence
B) 92% confidence
C) 94% confidence
D) 96% confidence
E) 98% confidence

9. Here is a portion of the computer output from a least squares regression analysis using a random sample of 25 students where x = foot length (cm) and y = height (cm). Assume conditions about the slope of the population regression line are met.

Predictor	**Coeff**	**SE Coef**	**T**	**P**
Constant	91.9766	10.2204	8.999	0.000
Foot Length	3.0867	0.4117	7.498	0.000

Which of the following is the best interpretation of the value 0.4117 in the computer output?

A) For each increase of 1 cm. in foot length, the average height increases by about 0.4117 cm.
B) In repeated samples of size 25, the slope of the sample regression line for predicting height from foot length will typically vary from the population slope by about 0.4117.
C) About 41.17% of the variation in height is accounted for by the linear model with foot length.
D) When using this model to predict height, about 41.17% of the predictions will be accurate.
E) There is a moderate, positive, linear relationship between foot length and height.

10. A buyer for a t-shirt shop wants to compare the proportion of t-shirts of each size that are sold in the store to the proportion that were ordered for stock. The buyer counts a random selection of t-shirts that were sold in a week: 25 small, 41 medium, 91 large, and 68 extra-large shirts were sold. If the buyer ordered 300 small, 600 medium, 1200 large, and 900 extra-large shirts, which of the following gives the correct test statistic, and *p-value* for the appropriate chi-square test of significance the buyer would perform?

A) $\chi^2 = 19.24, p - value = 2.4 \cdot 10^{-4}$
B) $\chi^2 = 0.37, p - value = 0.946$
C) $\chi^2 = 0.65, p - value = 0.885$
D) $\chi^2 = 5.22, p - value = 0.1564$
E) $\chi^2 = 45.42, p - value = 7.5 \cdot 10^{-10}$

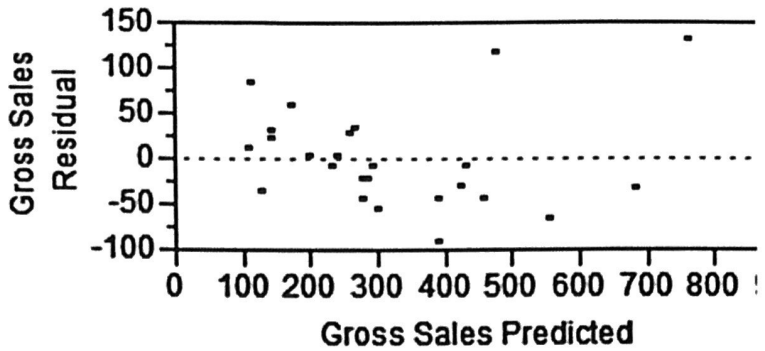
11. Suppose the probability of a member of a population having a certain disease is ψ. If a person has the disease, the probability of a diagnostic test yielding a positive result is k; whereas the probability the diagnostic test yields a positive result is h when the person does not have the disease. If a randomly selected diagnostic test yields a positive result, which of the following is equivalent to the probability the person has the disease?

A) ψ
B) $\frac{k \cdot \psi}{k \cdot \psi + h \cdot \psi}$
C) $\frac{h+k}{k \cdot \psi}$
D) $\frac{k \cdot \psi}{k \cdot \psi + h \cdot (1-\psi)}$
E) $\frac{k}{\psi}$

12. A multiple-choice test contains 32 questions, each with 6 answer choices. A student just randomly guesses on each question, without leaving any questions blank. If this process were to be repeated many times, what is the average number of questions that would be answered correctly, and with what standard deviation?

A) An average of 5 questions correct, with a standard deviation of about 0.2.
B) An average of 6.4 questions correct, with a standard deviation of about 0.17.
C) An average of 6.4 questions correct, with a standard deviation of about 0.2.
D) An average of 5.3 questions correct, with a standard deviation of about 1.06.
E) An average of 5.3 questions correct, with a standard deviation of about 2.1.

13. A least-squares regression analysis was performed using data from the number of items sold by a retailer, x, and the gross sales, y. The equation of the resulting fitted regression line is $\hat{y} = 2.1386882 + 6.5953589x$. The residuals plot from that analysis is shown below:



Which of the following is closest to the actual value of the gross sales when 83 items were sold?

A) 105
B) 475
C) 550
D) 625
E) 750

14. A brand of cereal is offering a prize in each box. There are 3 different prizes, and each prize is equally likely to be in any box. You have already collected 2 of the prizes. What is the probability that you will have to open at least 5 more boxes of cereal in order to obtain the 3rd (different) prize?

A) $\frac{16}{243}$
B) $\frac{2}{243}$
C) $\frac{16}{81}$
D) $\frac{211}{243}$
E) $\frac{65}{81}$

15. A probability density curve is given by the function $f(x) = x$ for $0 \leq x \leq b$. What is the value of b?

A) 1
B) 4
C) $\frac{1}{2}$
D) $\sqrt{2}$
E) $\sqrt{\frac{1}{2}}$

16. The following data gives the weight lifted in the 2000 Olympics by athletes in various weight classes. All weights are in kg.

Class, x	56	62	69	77	85	94	105
Lifted, y	305	325	357.5	367.5	390	405	425

For which of the following models is the highest percentage of variation in the response variable attributed to the linear model with the explanatory variable?

A) $\widehat{lifted} = 179.9184 + 2.4007(class)$
B) $\log(\widehat{lifted}) = 2.3391 + 0.0029(\text{class})$
C) $\widehat{lifted} = -451.1893 + 434.6834(\log(class))$
D) $\sqrt{\widehat{lifted}} = 14.2335 + 0.0628(class)$
E) $\log(\widehat{lifted}) = 1.5828 + 0.5203(\log(class))$

17. In a carton of a dozen eggs, three are rotten. Which of the following could most effectively be used to simulate the number of rotten eggs that might be chosen when two eggs are randomly selected from that dozen?

A) Randomly generate digits 0, 1, or 2 to represent the number of rotten eggs you get.
B) Assign 0 = rotten and 1, 2, 3 = good. Generate two random numbers between 0 and 3 and count how many 0's you get.
C) Assign 0, 1, 2 = rotten and 3, 4, 5, …., 11 = good. Generate two random numbers between 0 and 11, ignoring repeats.
D) Assign even numbers as good and odd numbers as rotten. Generate two random numbers between 0 and 11, ignoring repeats.
E) All of the above.

18. A manufacturer is testing his machine to make sure it operates properly. "Operating properly" in this case means that the machine is producing fewer than 1% defective parts. The manufacturer tests the hypothesis $H_0: p = .01$ *vs.* $H_a: p < .01$. Unbeknownst to the manufacturer, the machine is actually producing 2% defective parts. He takes a random sample of 1,200 machine parts and finds that 3 are defective. What probably happens as a result?

A) He rejects H_0, making a Type I error.
B) He fails to reject H_0, making a Type II error.
C) He correctly rejects H_0.
D) He correctly fails to reject H_0.
E) Can not be determined.

19. Which of the following designs describes the use of a cluster sampling method?

A) Soybeans are planted on a 48-acre field. The field is divided into one-acre subplots. A sample is taken from each subplot to estimate the harvest.
B) After a hurricane, a disaster area is divided into 200 equal grids. Thirty of the grids are selected, and every occupied household in the grid is interviewed about the relief efforts.
C) From calls made with randomly generated telephone numbers, 1012 respondents are asked if they rent or own their residences.
D) Chosen at random, 500 rural and 500 urban persons age 65 or older are asked about their health and experience with prescription drugs.
E) Every tenth person entering a mall is asked to name his or her favorite store.

20. Suppose there are two cities, *A* and *B*, that have approximately the same population mean and population standard deviation of daily temperatures. The distribution of daily temperatures for both cities *A* and *B* are each approximately normal. A random sample of 12 temperatures is obtained from city *A*, whereas a random sample of 20 temperatures is obtained from city *B*, and the difference between sample mean temperatures is calculated. The sampling distribution of the difference between sample mean temperatures of cities *A* and *B* is approximately normal with a mean of 0°F and standard deviation of 1.86°F. What is the population standard deviation of the distribution of daily temperatures of each city?

A) 25.95°F
B) 3.63°F
C) 6.44°F
D) 13.95°F
E) 5.09°F

Tie Breaker:

The five-number-summary for a set of data is given below:

Min = *16* $Q_1 = x$ Med = *51* $Q_3 = 2x+10$ Max = *104*

What is the largest possible integer value of x such that the maximum value of 104 would qualify an outlier in this data set?

Solutions:

1. C
2. B
3. C
4. E
5. A
6. B
7. B
8. E
9. B
10. C
11. D
12. E
13. B
14. C
15. D
16. C
17. C
18. A
19. B
20. E

Tie Breaker: 22