



Essex County Math League

May 22, 2024

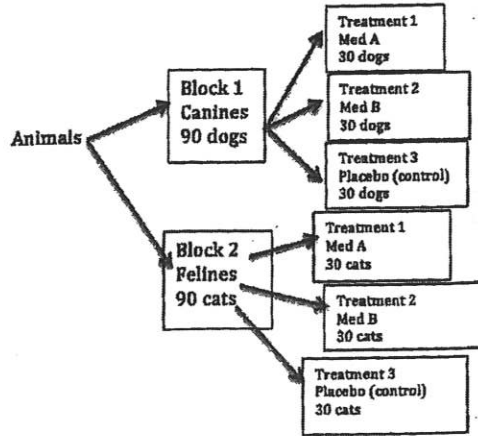
STATISTICS

DIRECTIONS: You may write on this test. Mark the answer sheet with dark, careful marks using a #2 pencil. Your score will be determined by the number of correct answers. Incorrect answers will **NOT** lower your score. You may 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.

Before the contest starts, fill in question 21 or 22 with the name of your school/town. Fill in the SEAT ID number as stated on the table.

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO

1. An experiment is conducted to determine if medication helps decrease flea rates on animals. You have available to use two medications (A and B) as well as a placebo. There are 180 test subjects in total-- 90 canines and 90 felines. A randomized block design is proposed, as shown in the following diagram:



At the end of the experiment, it is reported that the outcome is statistically significant. What is the best interpretation of the phrase “statistically significant” in context that is most appropriate for this experimental design?

- A) We cannot interpret the meaning of the phrase “statistically significant” since it is not a completely randomized design.
- B) The placebo had no effect on the rates of fleas in the dog block and the cat block.
- C) The rates of fleas between the 6 treatment groups are so different that this difference could not be attributed to random chance.
- D) The rates of fleas between the 3 treatment groups in the dog block and the rates of fleas between the 3 treatment groups in the cat block were so different within their respective blocks that this difference could not be attributed to random chance.
- E) The rates of fleas for the 3 treatment groups in the dog block were so different than the rates of fleas for the 3 treatments in the cat block that this difference could not be attributed to random chance.

2. A man possesses five coins, one of which is double-headed, one is double-tailed, and three are normal (fair). He shuts his eyes, picks a coin at random, and tosses it. He opens his eyes and the coin is showing heads face-up. What is the probability that it is the double-headed coin?

- A) 3/10
- B) 2/5
- C) 3/5
- D) 1/10
- E) 2/3

3. Consider the given population:

Person	Value
A	8
B	16
C	22
D	45

Which of the following gives the mean and standard deviation of the sampling distribution of sample means for samples of size $n = 3$ from this population?

- A) 22.5, 37
- B) 22.75, 7.952
- C) 22.75, 15.903
- D) 22.75, 5.301
- E) 22.5, 13.773

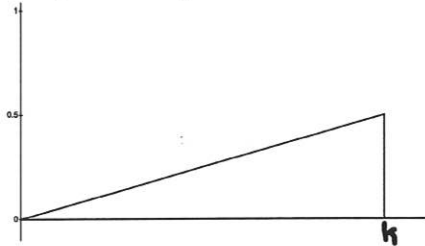
4. An apartment complex installed new carpeting in all of its apartment units. Five years later, they would like to assess the wear of those carpets. They do not have time to survey every apartment unit, so they would like to select a sample using a stratified sampling method. Which of the following criteria would be the most useful in determining the strata for this purpose?

- A) Which floor the apartment is on
- B) Whether the apartment has north-facing or south-facing windows
- C) Whether or not the apartment's residents own a car
- D) The genders of the apartment's residents
- E) Whether or not pets (i.e. cats/dogs) live in the apartment

5. The hiring director for a new tech startup found that (82.5, 89.3) is a 95% confidence interval for the mean coding proficiency score for all new software developers hired in the past year. Which of the following is the best interpretation of the confidence level in that statement?

- A) There is a 95% probability that the interval 82.5 to 89.3 contains μ .
- B) There is a 95% probability that the interval 82.5 to 89.3 contains the sample mean.
- C) The interval was constructed using the method that produces intervals that capture the true mean in 95% of all possible samples.
- D) If we take many samples, 95% of them will contain the interval (82.5, 89.3).
- E) About 95% of all coding proficiency scores are between 82.5 and 89.3.

6. Let X be a continuous random variable such that $f(x) = \frac{1}{8}x$ is a legitimate probability density function, where $0 \leq x \leq k$ as shown. What is the value of k ?



- A) 1/8
- B) 0.5
- C) 4
- D) 2
- E) 1.75

7. Below is the computer output for a least-squares regression of y = protein (in grams) and x = calories for 16 different sandwiches produced by a major sandwich franchise.

Predictor	Coef	SE Coef	T	P
Constant	-4.308	8.968	-0.48	0.638
Calories	0.06515	0.02006	3.25	0.006

$s = 5.72328$ $R\text{-Sq} = 43.0\%$ $R\text{-Sq}(\text{adj}) = 38.9\%$

Suppose that for one sandwich in the sample with 440 calories, the least-squares regression line overestimated its protein by about 5.6 grams. Which is closest to the actual number of grams of protein in the sandwich?

- A) 18.8 grams
- B) 29.9 grams
- C) 24.4 grams
- D) 12.2 grams
- E) 17.7 grams

8. Assume that the diameters of a population of oranges grown in a certain grove follow a normal distribution and vary with a standard deviation of 1.1 in. Which of the following represents the approximate diameter of an orange that is at the 92nd percentile of orange diameters in this grove?

- A) $(\mu + 1.405)$ in.
- B) $(\mu + 1.55)$ in.
- C) $(\mu + 0.92)$ in.
- D) $(\mu + 1.012)$ in.
- E) $(\mu + 2.02)$ in.

9. Consider testing the following hypotheses:

$$H_0: \mu = 300$$

$$H_a: \mu > 300$$

Which of the following statements are true about the power of the test?

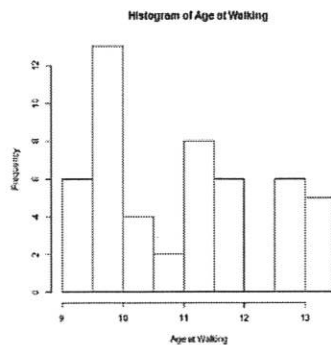
I. For any value of μ exceeding 300, the power of a test based on a sample of size 100 is higher than the power of a test based on a sample of size 75, assuming the same significance level.

II. For any value of μ exceeding 300, the power of a test using a significance level of .05 is higher than the power of a test using a significance level of .01, assuming the same sample size.

III. For any value of μ exceeding 300, the power of a test is greater if the true mean is 350 than if the true mean is 325, assuming the same sample size and significance level.

- A) I only
- B) I and III only
- C) II only
- D) I, II, and III
- E) None of these are true about the power of a test.

10. Below is a histogram displaying the age at walking (in months) in a sample of children:



Which of the following information could be concluded about the third quartile of the distribution based on the graph?

- A) It is equal to 12 months.
- B) It is at least 11 but less than 11.5 months.
- C) It is at least 11.5 but less than 12 months.
- D) It is greater than 12 months.
- E) None of these.

11. A test of significance results in a z-test statistic of 1.92. Which of the following could be the corresponding p -value of the test?

- I. 0.0274
- II. 0.0549
- III. 0.9726
- IV. 0.9451

- A) I only.
- B) I and II only.
- C) I, II, and III only.
- D) II and IV only.
- E) I, II, III, and IV.

12. Of 300 college football games, x games are won by the home team. Of 200 college baseball games, y games are won by the home team. Assuming that all games result in either a win or a loss, which of the following must be true in order to produce a segmented bar graph that shows there is no association between home-team advantage and sport within the given data?

- A) $x = y$
- B) $x = 150$ and $y = 100$
- C) $\frac{x+y}{500} > .50$
- D) $x + y \neq 250$
- E) $\frac{x}{300} = \frac{y}{200}$

13. The faces of a four-sided die are numbered 1 through 4, respectively. For a certain game, the die is tossed and the number that lands facedown is recorded. The table below summarizes the points a player earns for the number that lands facedown.

Number landing facedown	1	2	3	4
Points	0	1	0	2

Consider a game where you throw two independent tosses of the die. Let the random variable S represent the sum of the points earned from the first toss plus twice the points earned from the second toss. What is the probability the value of S is at least 4 points?

- A) 0.1875
- B) 0.3125
- C) 0.25
- D) 0.375
- E) 0.4375

14. A random sample of 20 Americans is selected. Let the random variable X be the number of people in the sample who live in an urban area. The standard deviation of X is approximately 1.45. For repeated samples of 200 Americans, which of the following could be the mean number of people who live in an urban area?

- A) 100
- B) 88
- C) 36
- D) 24
- E) 16

15. It is standard to use α to denote the probability of a Type I error and β to denote the probability of a Type II error. Let k represent the probability a particular null hypothesis is true. When performing a hypothesis test, which of the following would denote the probability of making a decision to reject the null hypothesis?

- A) $k\alpha + (1 - k)\beta$
- B) $k\alpha + (1 - k)(1 - \beta)$
- C) $k(1 - \alpha) + (1 - k)(1 - \beta)$
- D) $k\alpha(1 - \beta)$
- E) $k\beta(1 - \alpha)$

16. A questionnaire about cellphone use asked participants to state their age and their response to the following question: "You need to contact a friend to arrange a time and place to meet. Would you call your friend or would you send a text?" The responses are summarized in the table below.

Age	Response	
	Would Call	Would Text
39 or under	212	404
40 or over	273	382

Which of the following could be the test statistic when performing an appropriate test of significance with the given data?

- I. $\chi^2 = 7.098$
- II. $\chi^2 = 8.288$
- III. $z = -2.664$
- IV. $t = 2.664$

- A) I only.
- B) II only.
- C) I and III only.
- D) I, II, and III only.
- E) I and IV only.

17. A police cadet took a sample of 10 days (out of the 90-day summer) and recorded the number of police officers on duty in Boston city parks, x , and the number of reported muggings on that day, y . From the 10 days, the mean and standard deviation of the number of police officers on duty was 10.3 and 5.6, respectively. The average number of reported muggings was 4.7 with a standard deviation of 2.9. Upon running a regression analysis, he found there was a very strong negative correlation of -0.969. Which of the following is the equation of the least squares regression line that could be used to predict the number of muggings from the number of police officers on duty that data, based on the sample data provided?

- A) $\hat{y} = 9.87 - 0.502x$
- B) $\hat{y} = 10.04 - 0.518x$
- C) $\hat{y} = 0.502 - 9.87x$
- D) $\hat{y} = 15 + 0.502x$
- E) $\hat{y} = 15 - 0.969x$

18. A left-tailed two-sample t-test is performed which results in a test statistic of $t = -2.06$ with $df = 16$. For which of the following levels of confidence would a two-sided confidence interval constructed using the same sample data contain the value of 0?

- I. 98%
- II. 95%
- III. 90%

- A) I and II only.
- B) I and III only.
- C) II and III only.
- D) I, II, and III.
- E) None of the above.

19. Two components work in series such that the failure of either will cause the system to fail. The time to failure for a new component is normally distributed with a mean of 125 days and a standard deviation of 16 days and 16 hours. One of the components has already been running for 116 days and 16 hours, while the other has been running for 104 days and 4 hours. Assuming component failures are independent, what is the probability the system survives for 10 more days?

- A) 17.95%
- B) 55.22%
- C) 51.98%
- D) 46.05%
- E) 24.74%

20. The GPAs of students at School A are approximately normal with a mean of 3.12 and a standard deviation of 0.40. The GPAs of students at School B are approximately normal with a mean of 3.06 and a standard deviation of 0.38. A random sample of 100 students is taken from each school, and the sample mean is obtained from each. Which of the following is equivalent to the probability that the sample mean from School B is higher than the sample mean from School A?

- A) $P\left(t > \frac{0.06}{\frac{0.78}{\sqrt{200}}}\right)$
- B) $P\left(z < \frac{-0.06}{\sqrt{\frac{0.4^2}{100} + \frac{0.38^2}{100}}}\right)$
- C) $P\left(t < \frac{(0.06-0)}{\sqrt{\frac{0.4^2}{100} + \frac{0.38^2}{100}}}\right)$
- D) $P\left(z > \frac{(0-0.02)}{\sqrt{\frac{3.12^2}{100} + \frac{3.06^2}{100}}}\right)$
- E) $P\left(z < \frac{(0-0.06)}{\sqrt{\frac{0.40^2}{100} + \frac{0.38^2}{100}}}\right)$

Tie-Breaker

A student in a college psychology lecture hall took note of the number of students who were wearing hoodies and were wearing glasses, as partially summarized in the table below:

	Wearing glasses	Not wearing glasses
Wearing a hoodie	$3x + 14$	$7y + 3$
Not wearing a hoodie	$8x - 7$	$10y + 11$

The probability that a randomly selected student is wearing a hoodie given they are wearing glasses is $\frac{2}{5}$. The probability that a randomly selected student is wearing glasses given they are wearing a hoodie is $\frac{19}{31}$. Find the number of students in the hall.

ANSWER KEY

1. D
 2. B
 3. D
 4. E
 5. C
 6. C
 7. A
 8. B
 9. D
 10. C
 11. C
 12. E
 13. B
 14. D
 15. B
 16. C
 17. A
 18. A
 19. B
 20. E
- TB. 320