# *Essex County Math League*

## May 24, 2023

# 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 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.

1. In Denali National Park, Alaska, the late-winter wolf population in the park can be predicted from the fall caribou population, such that about 83.7% of the variation in the wolf population can be explained by the linear model with the caribou population.
In a random sample of recent years, the average fall caribou population was 25.2 hundred with a standard deviation of 5.8 hundred caribou. The average late-winter wolf population was 62.6 with a standard deviation of 10.2 wolves. Which of the following could be the slope of the least-squares regression line used to predict the late-winter wolf population from the fall caribou population?

A) 0.520
B) 0.837
C) 1.472
D) 1.609
E) 1.759

2. A machine that fills bottles is supposed to do so according to a normal model with $\mu = 237 \ mL$. A quality control manager suspects the machine is under-filling the bottles and wants to take a random sample of bottles to test the hypotheses $H_0: \mu = 237$ vs. $H_a: \mu < 237$. Suppose the reality is that $\mu = 231 \ mL$. For which of the following would the manager's test have the greatest power?

A) $\alpha = .05$, n = 25
B) $\alpha = .01$, n = 40
C) $\alpha = .10$, n = 50
D) $\alpha = .10$, n = 25
E) The power will always be the same as long as $\mu = 231 \ mL$.

3. A student who attends a very large high school would like to estimate the true proportion of students at their high school who will be attending a summer camp program. Which of the following is the minimum sample size the student should use in order to be 95% confident and have a margin of error within 5% of the true proportion?

A) 95
B) 271
C) 385
D) 664
E) 1221

4.      The following frequency table gives the number of days in each daily high temperature range for the month of March:

| Temperature °F | Number of Days |
|:---:|:---:|
| 31-36 | 9 |
| 37-42 | 7 |
| 43-48 | 5 |
| 49-54 | 5 |
| 55-60 | 3 |
| 61-66 | 2 |

Which of the following is a possible value of the median daily high temperature in the month of March?

A) 35°F
B) 42°F
C) 45°F
D) 51°F
E) 56°F

5.      An apartment complex has four buildings, each of which has four floors, with four apartments on each floor. You would like to take a random sample of 8 apartments and survey the residents in those apartments. Which of the following describes a <u>cluster</u> sampling method for selecting the 8 apartments to be surveyed?

A) Assign each apartment a unique number from 1 to 64. Use a random number generator to select 8 unique numbers and survey those apartments selected.

B) Assign each floor a unique number from 1 to 16. Use a random number generator to select 2 unique numbers and survey all four apartments on both floors that were selected.

C) Assign each building a unique number from 1 to 4. Use a random number generator to select 2 buildings. Then randomly select four apartments from each of the 2 selected buildings to be surveyed.

D) Assign each floor a unique number from 1 to 16. Use a random number generator to select 4 floors. Then randomly select two apartments from each of the 4 selected floors to be surveyed.

E) Assign each apartment a unique number from 1 to 64. Use a random number generator to select a number between 1 and 8, and call it $k$. Survey every $k$th apartment until 8 apartments have been surveyed.

6.  A survey collected data on the annual salaries of a random sample of registered nurses. Let $x$ = years of experience and $y$ = annual salary in thousands of dollars. A least-squares regression analysis of $y$ on $x$ resulted in the following equation: $\hat{y} = 42.122 + 1.398x$. Suppose one nurse in the data set had 20 years of experience. The value of the residual for that observation was $2,563. What was this nurse's actual salary?

A) $72,645
B) $70,082
C) $67,519
D) $44,685
E) $42,122

7.  In an experiment used to test the effects of advertising, all subjects watched a 60-minute television program that included an ad for a specific product. Some subjects saw a 30-second commercial and others saw a 90-second version. The same commercial was shown either 1, 3, or 5 times during the program. After viewing, all subjects answered questions about their recall of the ad, their attitude toward the product, and their intention to purchase it. What is the number of factors and number of treatments in this experiment?

A) 2 factors; 3 treatments
B) 1 factor; 5 treatments
C) 5 factors; 6 treatments
D) 3 factors; 9 treatments
E) 2 factors; 6 treatments

8.  A two-sample t-test for the mean is conducted testing the hypotheses $H_0: \mu_1 = \mu_2$ vs. $H_a: \mu_1 > \mu_2$ in which the resulting *p-value* is 0.016. A two-sided confidence interval for the true population parameter $\mu_1 - \mu_2$ is constructed using the same data. For which level(s) of confidence will the resulting interval be entirely positive?

I. C = 90%        II. C = 95%        III. C = 98%        IV. C = 99%

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

9.     The following computer output shows the least-squares regression analysis from data on a random sample of 116 Savannah sparrows at Kent Island. The weight (in grams) and wing length (in mm) were obtained for birds from nests that were reduced, controlled or enlarged.

```
The regression equation is  Weight = 1.37 + 0.467 WingLength

Predictor      Coef  SE Coef      T     P
Constant     1.3655   0.9573   1.43  0.156
WingLength   0.4674  0.03472  13.46  0.000

S = 1.39959   R-Sq = 61.4%   R-Sq(adj) = 61.1%
```

Which of the following gives the 90% confidence interval estimate for the true slope of the least-squares regression line that could be used to predict a Savannah sparrow's weight from its wing length?

A) $0.4674 \pm 1.43(0.03472)$
B) $1.3655 \pm 1.645(0.93573)$
C) $1.3655 \pm 13.46(0.93573)$
D) $0.4674 \pm 1.658(0.03472)$
E) $0.4674 \pm 1.645(139959)$


10.    About 80% of lightbulbs made by a particular manufacturer are fluorescent bulbs that have an average lifetime of 800 hours with a standard deviation of 50 hours. The remaining lightbulbs made by this manufacturer are halogen bulbs that have an average lifetime of 900 hours with a standard deviation of 66 hours. If you choose 10 lightbulbs made by this manufacturer at random, what is the probably that exactly 5 of them will last longer than 750 hours?

A) .2101
B) .0046
C) .0097
D) .0192
E) .3394

11.     A particular candy comes in eight different flavors per package, where each flavor is supposed to represent an equal proportion of the package, as claimed by the manufacturer. To test this claim, you purchase a package containing 60 of these candies and find the distribution of the eight flavors in the package is as follows:

3, 4, 6, 6, 7, 9, 12, 13.

Which of the following is the correct test statistic, p-value, and degrees of freedom (if applicable) for the most appropriate test of significance to test the manufacturer's claim?

A) $t = 5.916, \ p = .00059, df = 7$
B) $t = 0.75, \ p = .2339, df = 12$
C) $z = -0.276, \ p = .7825, df = NA$
D) $\chi^2 = 12, \ p = .1006, \ df = 7$
E) $\chi^2 = 7.5, \ p = .3787, \ df = 7$

12.     Let $X$ represent a continuous random variable than can take on any value between 0 and 500. Which of the following gives the equation of the probability density curve for $X$ where $0 \leq X \leq 500$?

A) $f(x) = 0.002$
B) $f(x) = 0.004$
C) $f(x) = 0.025$
D) $f(x) = 0.01$
E) $f(x) = 1$

13.     A game involves two players and a bag of 5 marbles in which 3 marbles are red and 2 marbles are white. Two marbles are randomly chosen without replacement. If both marbles are the same color, player A wins $3 from player B. Otherwise, player B wins $2 from player A. If this game is played many times, what is player A's expected value?

A) $0
B) -$0.33
C) -$0.67
D) $0.15
E) $1.35

14.     A teacher gives a multiple-choice quiz with 20 questions, each with 4 answer choices. Each question is worth 1 point, or 5% of the quiz score, such that 20 correct answers = 100%, 19 correct answers = 95%, etc. Suppose that every student who takes the quiz randomly guesses on each question with leaving no questions blank. What is the expected mean and standard deviation of the students' quiz scores?

A) mean of 50% with a standard deviation of about 25%
B) mean of 50% with a standard deviation of about 5%
C) mean of 40% with a standard deviation of about 1.93%
D) mean of 25% with a standard deviation of about 2.24%
E) mean of 25% with a standard deviation of about 9.7%

15.     A one proportion z-test is to be conducted to test $H_0: p = 0.37$ vs. $H_a: p < 0.37$. If $n = 200$, which of the following is the largest possible value of the sample proportion $\hat{p}$ that would result in a decision to Reject $H_0$ when using a 1% level of significance?

A) 0.2219
B) 0.2594
C) 0.2638
D) 0.2905
E) 0.3117

16.     The following table summarizes the number of public schools in two California counties.

| School | County Los Angeles | San Diego | Total |
|---|---|---|---|
| Elementary | 1,395 | 498 | 1,893 |
| Middle | 422 | 165 | 587 |
| High | 570 | 191 | 761 |
| Total | 2,387 | 854 | 3,241 |

Suppose that a certain number of schools that were recorded as Los Angeles Elementary schools were done so in error, and should have actually been recorded as Middle schools. If about 18.4% of all public schools in Los Angeles County are Middle schools, how many schools were incorrectly recorded as Elementary schools in the table above?

A) 9
B) 17
C) 31
D) 174
E) 309

17.     In a random sample of 250 records from Car Dealership A, 163 customers routinely use the dealership service department for maintenance. In a random sample of 280 records from Car Dealership B, 154 customers routinely use the dealership service department for maintenance. A confidence interval for the true difference between the proportion of customers at Car Dealership A and Car Dealership B who routinely use the dealership service department for maintenance is constructed, resulting in an interval of (0.025, 0.179). Which of the following is closest to the $z^*$ critical value that was used to construct the given interval?

A) 1.65
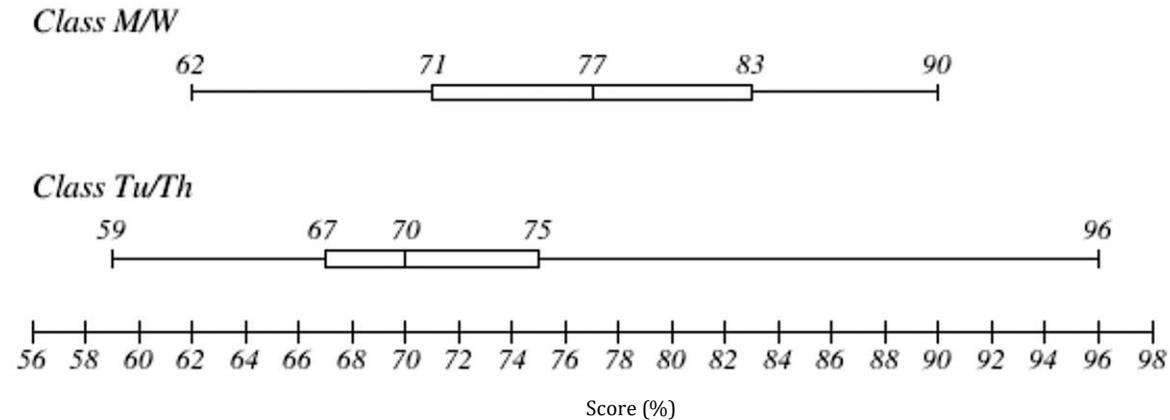B) 1.82
C) 1.99
D) 2.14
E) 2.58

18.     Water specimens are taken from water used for cooling as it is being discharged from a power plant into a river. It has been determined that as long as the mean temperature of the discharged water is at most 150°F, there will be no negative effects on the river's ecosystem, and the plant is considered "in compliance" with regulations. In order to test the plant's compliance, a random sample of 50 water specimens at randomly selected times are taken and the resulting data will be used to test the hypotheses $H_0: \mu = 150$°F versus $H_a: \mu > 150$°F. Which of the following best describes a Type II error in the context of this situation?

A) Being unable to make a conclusion about the plant's compliance.
B) Correctly concluding that the plant is not in compliance.
C) Concluding that the plant is not in compliance when it actually is.
D) Correctly concluding that the plant is in compliance.
E) Concluding that the plant is in compliance when it actually is not.

19.     A least-squares regression analysis is performed on variables $x$ and $y$, in which the resulting residual plot shows a distinct curved pattern. A least-squares regression analysis is performed on variables $x$ and $log(y)$ in which the resulting residual plot shows random scatter and no curved pattern. Which of the following equations should be used to best model the relationship between $x$ and $y$?

A) $\hat{y} = a + bx$
B) $\hat{y} = a + b \cdot log(x)$
C) $log(\hat{y}) = a + bx$
D) $log(\hat{y}) = a + b \cdot log(x)$
E) $\hat{y} = a + b^x$

20.    The following parallel boxplots give the distribution of scores on an exam given in a college course in which one section meets Monday/Wednesday and another section meets Tuesday/Thursday. A student takes a make-up exam and unfortunately ends up scoring the lowest grade in their respective section. Their score would be considered an outlier if they were in the Tu/Th section, but would not be considered an outlier if they were in the M/W section. Which of the following could be the student's score?

*Class M/W*

```
        62          71      77      83       90
        |------------[---|---------]---------|
```

*Class Tu/Th*

```
    59          67  70    75                    96
    |-----------[-|---]--------------------------|
```

```
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98
```

Score (%)

A) 41%
B) 47%
C) 50%
D) 54%
E) 58%

---

**TIE BREAKER**    A medical researcher is testing a new method to check for the presence of a certain gene, using 400 subjects, of whom the gene is actually present in 26 of the subjects. Suppose the new method correctly identifies the gene as being present 96.4% of the time in those who have it, and incorrectly identifies the gene as being present 2.2% of the time in those who don't have it. To the nearest whole person, how many of the subjects are expected to receive a correct result about the gene's presence when using the new method?

ANSWER KEY

1.  D
2.  C
3.  C
4.  B
5.  B
6.  A
7.  E
8.  B
9.  D
10. B
11. D
12. A
13. A
14. E
15. D
16. B
17. B
18. E
19. C
20. D
TB. 391