Essex County Math League

Wednesday, May 25, 2022

Statistics

Essex County Math League
May 25, 2022
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 county 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. The following are the summary statistics for a set of grades for a class of students on a 100-point midterm exam: $\bar{x} = 77.07$ and $s_x = 5.09$. Assume the grades are approximately normally distributed. Put the following in order from least to greatest:
A: an exam grade at the 87th percentile of the class
B: an exam grade with a z-score of 0.87
C: an exam grade of 87 points

A) A < C < B
B) B < C < A
C) C < A < B
D) A < B < C
E) B < A < C

2. A researcher conducts a study on 9 subjects to determine if there is a linear relationship between a person's height (in inches) and pulse rate (in beats per minute). The 9 subjects had an average height of 69.1 in. and standard deviation of about 5.23 in. Their average pulse rate during the study was 79.2 bpm with a standard deviation of about 28.97 bpm. The equation of the least-squares regression line is: $\widehat{pulse\ rate} = 275.05 - 2.8335(height)$. Which of the following could be the approximate percent of variation in pulse rate that is explained by this linear model with height?

A) 5.5%
B) 8%
C) 26%
D) 51%
E) 79%

3. An industrial psychologist administered a personality inventory test for passive aggressive traits to 120 employees. Individuals were given a score from 1 to 5, where 1 was extremely passive and 5 extremely aggressive. The following frequency table shows the results of the test:

Score, *x*	1	2	3	4	5
Frequency	14	33	36	26	11

What is the average and standard deviation of the scores of these employees?

A) Average of 3 with a standard deviation of about 1.58.
B) Average of 3 with a standard deviation of about 1.41.
C) Average of 2.89 with a standard deviation of about 1.31.
D) Average of 2.89 with a standard deviation of about 1.15.
E) Average of 24 with a standard deviation of about 11.16.

4. The weights of a certain type of melon in a supermarket are normally distributed with a mean of 22 pounds and a standard deviation of 2.5 pounds. If you select 5 of these melons at random, what is the probability that exactly 1 of them will weigh more than 27 pounds?

A) 0.0228
B) 0.2
C) 0.9772
D) 0.995
E) 0.104

5. The blood cholesterol levels for all men aged 20 to 34 follows a Normal distribution with mean 188 mg/dL and standard deviation σ_1 mg/dL. For 14-year-old boys, blood cholesterol levels follow a Normal distribution with mean 170 mg/dL and standard deviation σ_2 mg/dL. Assume the values of σ_1 and $\sigma_2$are approximately equal. A sample of 25 members from each population is taken. In the sampling distribution of the difference between sample means, there is about a 0.1444 probability of obtaining a difference higher than 30 mg/dL. What is the approximate population standard deviation of the blood cholesterol levels of each population?

A) 20 mg/dL
B) 30 mg/dL
C) 40 mg/dL
D) 50 mg/dL
E) Cannot be determined.

6. Attendees at a concert range in age from 18 years old to 72 years old. The Q_1 of the ages (in years) of the concert attendees is half as big as the Q_3 of the ages of the concert attendees. Which of the following is the smallest possible value of Q_1 such that a person who is 72-years old is *not* considered an outlier at this concert?

A) 20
B) 21
C) 22
D) 23
E) 24

7. Consider the probability density function $f(x) = \frac{1}{5}$ over the interval $0 \leq x \leq 5$. Find $P(x < 1.5) \cup P(x > 3.5)$.

A) 1/5
B) 2/5
C) 3/5
D) 4/5
E) 1

8. The heights of full-grown maple trees are normally distributed with a mean of 87.5 feet and standard deviation of 5.75 feet. An arboriculturist is concerned that maple trees in a certain region are not growing tall enough. He takes a random sample of 12 trees and performs a test of significance to determine if the trees in this region are below average height. Which of the following is the *largest* sample mean for which he would reject the null hypothesis at the .05 level of significance?

A) 83.25 ft
B) 83.75 ft
C) 84.25 ft
D) 84.75 ft
E) 85.25 ft

9. Suppose the eye color of a particular population is distributed as follows: 40% brown eyes, 30% blue eyes, and 30% green eyes. A random sample of 220 people is taken from this population, 76 of whom have brown eyes. If the resulting test statistic for the chi-square test for goodness of fit is $\chi^2 = \frac{26}{3}$, which of the following is the number of blue-eyed and green-eyed people in the sample, respectively?

A) 72 and 72
B) 66 and 78
C) 58 and 86
D) 44 and 100
E) None of these.

10. Suppose a bottling machine that is supposed to fill bottles with an average of 12 ounces of cola per bottle is not functioning correctly and is actually underfilling the bottles. A test of significance is performed on a random sample of bottles in which the null hypothesis represents the machine functioning correctly (i.e. filling the bottles to an average of 12 ounces per bottle), and the alternative hypothesis represents the machine not functioning correctly. If the resulting p-value of this significance test is 0.243, which of the following is most likely to occur as a result?

A) There is no error and it is assumed the machine is functioning correctly.
B) There is a Type I error and it is assumed the machine is functioning correctly.
C) There is a Type II error and it is assumed the machine is functioning correctly.
D) There is a Type I error and it is assumed the machine is not functioning correctly.
E) There is no error and it is assumed the machine is not functioning correctly.

Use the following information to answer #11-12:
The following is a 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-Sq = 43.0%   R-Sq(adj) = 38.9%
```

11. Suppose a sandwich in the sample with 350 calories actually contains 15 grams of protein. Find the value of the residual for this sandwich.

A) -3.33075 grams
B) -3.4945 grams
C) 18.33075 grams
D) 18.4945 grams
E) -7.8025 grams

12. Which of the following gives a 95% confidence interval estimate for the true slope of the least-squares regression line that can be used to predict protein from calories of the sandwich?

A) $0.06515 \pm 3.25(0.02006)$
B) $0.06515 \pm 2.145(0.02006)$
C) $0.06515 \pm 1.96(0.02006)$
D) $-4.308 \pm 1.96(8.968)$
E) $-4.308 \pm 0.48(8.968$

13. You are playing a game in which the random variable *X* represents the points earned in the first round of the game and the random variable *Y* represents the points earned in the second round of the game. Let $E(X) = 13$ points, $SD(X) = 3$ points, $E(Y) = 15$ points, and $SD(Y) =$ 7 points. Suppose a player's total score for the game is their points earned in the first round added to twice their points earned in the second round. What are the expected value and standard deviation of a player's total score T for the game?

A) $E(T) = 43, SD(T) = 14.32$
B) $E(T) = 43, SD(T) = 17$
C) $E(T) = 41, SD(T) = 13$
D) $E(T) = 43, SD(T) = 10.34$
E) $E(T) = 28, SD(T) = 7.62$

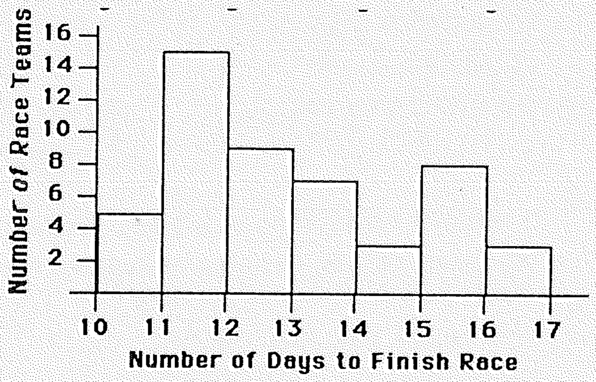
14. A manager of a gym would like to estimate the mean amount of time members spend using the facilities per week. They would like their estimate to be within 1 hour of the true mean with 90% confidence. What is the minimum number of members they should include in the sample? Assume a reasonable estimate for σ is 4 hours.

A) 7
B) 13
C) 24
D) 44
E) 62

15. The following histogram shows the distribution of race times (in days) for the top 50 finishers in the 1994 Iditarod Dog Sled Race.

Which of the following could be the median race time?

A) 11.480 days
B) 12.412 days
C) 13.572 days
D) 14.460 days
E) 16.668 days



16. A manufacturer of calculators receives computer chips from two suppliers: 40% of the chips come from Supplier A and 60% come from Supplier B. It is found that there is a 1 out of 250 chance a chip from Supplier A is defective, and a 1 out of 400 chance a chip from Supplier B is defective. A chip chosen at random is found to be defective. What is the probability it came from Supplier A?

A) 0.004
B) 0.0065
C) 0.0016
D) 0.484
E) 0.516

17. A soft drink company is offering a promotion in which there is a winning code contained in the bottle cap of 1 out of every 16 bottles. What is the probability you will need to open at least six bottles in order to find your first winning code?

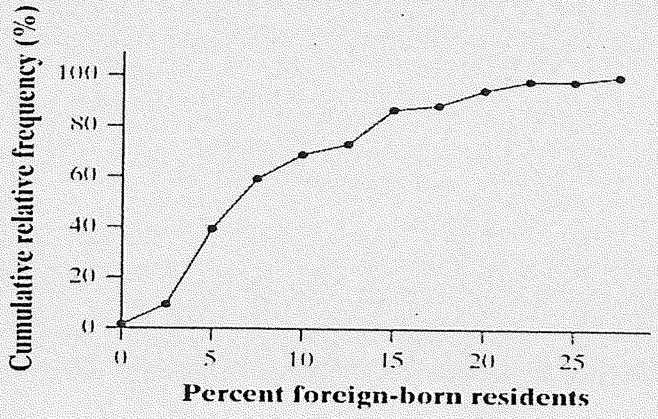
A) 0.724
B) 0.3798
C) 0.274
D) 0.824
E) 0.945

18. A two-proportion z-test is conducted using two independent random samples to test $H_0: p_1 - p_2 = 0$ vs. $H_a: p_1 - p_2 \neq 0$. The resulting p-value is 0.082. For which of the following level(s) of confidence will the two-proportion z-interval constructed using the same sample data contain the value "0"?

I. 90%
II. 95%
III. 99%

A) I only
B) I, II, and III
C) III only
D) II and III only
E) None of these.

19. Use the given cumulative relative frequency graph shown. Which of the following could be the approximate interquartile range of the percent of foreign-born residents in the 50 states?



A) 50%
B) 2%
C) 45%
D) 8%
E) 92%

20. Researchers want to test a drug manufacturer's claim that fewer than 10% of patients who take its new drug will experience side effects. They decide to carry out a test of $H_0: p = 0.10$ vs. $H_a: p < .10$ where p is the true proportion of patients similar to those in the study who would experience side effects when taking the new drug. They use a random sample of 300 patients and a level of significance of .05. Suppose the true proportion of patients who experience side effects when taking the new drug is actually 8%. In this situation, which of the following would increase the power of the test?

I. Use $\alpha = .01$ instead of $\alpha = .05$
II. If the true proportion is $p = .06$ instead of $p = .08$
III. Use $n = 200$ instead of $n = 300$

A) II only
B) I, II, and III
C) III only
D) I and III only
E) None of these.

Tie Breaker

A one-proportion z-interval is constructed to estimate the true proportion of people in a given community who are in favor of building a new shopping mall based on a random sample of *n* residents of the community. The resulting confidence interval is (0.225, 0.575). A new random sample that includes 4 times as many residents as the original sample results in the same sample proportion. If the level of confidence remains unchanged, what is the margin of error for the confidence interval constructed based on the new sample? Round to 4 decimal places.

ANSWER KEY (w/ WORK)

ECML STATISTICS 2022

1. The following are the summary statistics for a set of grades for a class of students on a 100-point midterm exam: $\bar{x} = 77.07$ and $s_x = 5.09$. Assume the grades are approximately normally distributed. Put the following in order from least to greatest:

A: an exam grade at the 87th percentile of the class $z = \text{invNorm}(.87) = 1.13$

B: an exam grade with a z-score of 0.87

C: an exam grade of 87 points

$$z = \frac{87 - 77.07}{5.09} = 1.95$$

$B < A < C$

A) A < C < B
B) B < C < A
C) C < A < B
D) A < B < C
[✓ E)] B < A < C

2. A researcher conducts a study on 9 subjects to determine if there is a linear relationship between a person's height (in inches) and pulse rate (in beats per minute). The 9 subjects had an average height of 69.1 in. and standard deviation of about 5.23 in. Their average pulse rate during the study was 79.2 bpm with a standard deviation of about 28.97 bpm. The equation of the least-squares regression line is: $\widehat{pulse\ rate} = 275.05 - 2.8335(height)$. Which of the

r^2 $\{$ following could be the approximate percent of variation in pulse rate that is explained by this linear model with height?

$$\text{slope} = r \cdot \left(\frac{s_y}{s_x}\right)$$

$$-2.8335 = r\left(\frac{28.97}{5.23}\right)$$

$$r^2 = (-.511)^2 = .26\%$$

A) 5.5%
B) 8%
[✓ C)] 26%
D) 51%
E) 79%

3. An industrial psychologist administered a personality inventory test for passive aggressive traits to 120 employees. Individuals were given a score from 1 to 5, where 1 was extremely passive and 5 extremely aggressive. The following frequency table shows the results of the test:

Score, x	1	2	3	4	5
Frequency	14	33	36	26	11

$n = 120$

What is the average and standard deviation of the scores of these employees?

A) Average of 3 with a standard deviation of about 1.58.
B) Average of 3 with a standard deviation of about 1.41.
C) Average of 2.89 with a standard deviation of about 1.31.
[✓ D)] Average of 2.89 with a standard deviation of about 1.15.
E) Average of 24 with a standard deviation of about 11.16.

$$E(x) = 1\left(\frac{14}{120}\right) + 2\left(\frac{33}{120}\right) + 3\left(\frac{36}{120}\right) + 4\left(\frac{26}{120}\right) + 5\left(\frac{11}{120}\right) = 2.89$$

$$SD(X) = \sqrt{(1-2.89)^2\left(\frac{14}{120}\right) + (2-2.89)^2\left(\frac{33}{120}\right) + (3-2.89)^2\left(\frac{36}{120}\right) + (4-2.89)^2\left(\frac{26}{120}\right) + (5-2.89)^2\left(\frac{11}{120}\right)} =$$

$$\sqrt{1.313..} \approx 1.1459...$$

$\mu = 22 \quad \sigma = 2.5 \quad P(x > 27) = P\left(z > \frac{27-22}{2.5}\right) = P(z > 2) = .0228$

4. The weights of a certain type of melon in a supermarket are normally distributed with a mean of 22 pounds and a standard deviation of 2.5 pounds. If you select 5 of these melons at random, what is the probability that exactly 1 of them will weigh more than 27 pounds?

A) 0.0228
B) 0.2
C) 0.9772
D) 0.995
[✓ E) 0.104]

Binomial $n = 5 \quad p = .0228 \quad X = 1$

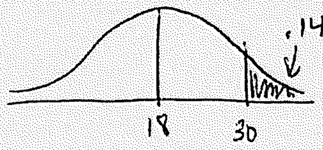
$\binom{5}{1}(.0228)(.9772)^4$

5. The blood cholesterol levels for all men aged 20 to 34 follows a Normal distribution with mean 188 mg/dL and standard deviation σ_1 mg/dL. For 14-year-old boys, blood cholesterol levels follow a Normal distribution with mean 170 mg/dL and standard deviation σ_2 mg/dL. Assume the values of σ_1 and σ_2 are approximately equal. A sample of 25 members from each population is taken. In the sampling distribution of the difference between sample means, there is about a 0.1444 probability of obtaining a difference higher than 30 mg/dL. What is the approximate population standard deviation of the blood cholesterol levels of each population?

$_1 = \sigma_2$

A) 20 mg/dL
B) 30 mg/dL
[✓ C) 40 mg/dL]
D) 50 mg/dL
E) Cannot be determined.

$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 188 - 170 = 18$



$z^* = \text{invNorm}(.8556) = 1.06$

$$z = \frac{30-18}{\sqrt{\frac{\sigma_1^2}{25} + \frac{\sigma_2^2}{25}}} = 1.06 \qquad \frac{12}{\sqrt{\frac{2\sigma^2}{25}}} = 1.06$$

$$\sqrt{\frac{2\sigma^2}{25}} = \frac{12}{1.06} = 11.32$$

$\frac{2\sigma^2}{25} = 128.16$

$2\sigma^2 = 3204$

$\sigma^2 = 1602 \quad \sigma = 40$

6. Attendees at a concert range in age from 18 years old to 72 years old. The Q₁ of the ages (in years) of the concert attendees is half as big as the Q₃ of the ages of the concert attendees. Which of the following is the smallest possible value of Q₁ such that a person who is 72-years old is *not* considered an outlier at this concert?

$Q_1 = \frac{1}{2} Q_3$

$2Q_1 = Q_3$

A) 20
[✓ B) 21]
C) 22
D) 23
E) 24

$72 < Q_3 + 1.5(Q_3 - Q_1)$

$72 < Q_3 + 1.5(Q_3 - \frac{1}{2}Q_3)$

$72 < Q_3 + 1.5\ Q_3 - 0.75Q_3$

$72 < 1.75 \cdot Q_3$

$41.143 < Q_3$

$72 < Q_3 + 1.5(Q_3 - Q_1)$

$72 < 2Q_1 + 1.5(2Q_1 - Q_1)$

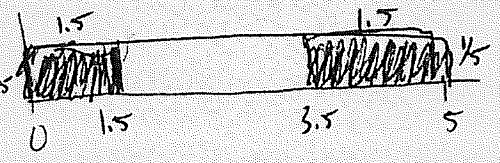
$72 < 2Q + 3Q_1 - 1.5Q_1$

$\frac{72}{3.5} < \frac{3.5Q_1}{3.5}$

$\boxed{20.57 < Q_1}$

7. Consider the probability density function $f(x) = \frac{1}{5}$ over the interval $0 \leq x \leq 5$. Find $P(x < 1.5) \cup P(x > 3.5)$.

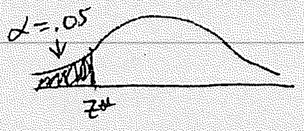
A) 1/5
B) 2/5
[✓ C) 3/5]
D) 4/5
E) 1



$\frac{1}{5}(1.5) + \frac{1}{5}(1.5) = .6$ or $3/5$

8. The heights of full-grown maple trees are normally distributed with a mean of 87.5 feet and standard deviation of 5.75 feet. An arboriculturist is concerned that maple trees in a certain region are not growing tall enough. He takes a random sample of 12 trees and performs a test of significance to determine if the trees in this region are below average height. Which of the following is the *largest* sample mean for which he would reject the null hypothesis at the .05 level of significance?

Ho: $\mu = 87.5$
Ha: $\mu < 87.5$



A) 83.25 ft
B) 83.75 ft
C) 84.25 ft
[✓ D) 84.75 ft]
E) 85.25 ft

Reject Ho if
$Z < -1.645$

$Z^* = \text{invNorm}(.05) = -1.645$

$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < -1.645$ $\quad \frac{\bar{X}-87.5}{\frac{5.75}{\sqrt{12}}} < -1.645$

$\bar{X} - 87.5 < -2.73$

$\bar{X} < 84.764$

9. Suppose the eye color of a particular population is distributed as follows: 40% brown eyes, 30% blue eyes, and 30% green eyes. A random sample of 220 people is taken from this population, 76 of whom have brown eyes. If the resulting test statistic for the chi-square test for goodness of fit is $\chi^2 = \frac{26}{3}$, which of the following is the number of blue-eyed and green-eyed people in the sample, respectively?

	Br	Bl	GR	Total
obs	76	x	y	220
exp	88	66	66	220

~~A) 72 and 72~~
~~B) 66 and 78~~
[✓ C) 58 and 86]
D) 44 and 100
E) None of these.

$$\chi^2 = \Sigma \frac{(\text{obs}-\text{exp})^2}{\text{exp}} = \frac{(76-88)^2}{88} + \frac{(x-66)^2}{66} + \frac{(y-66)^2}{66} = \frac{26}{3}$$

$$\frac{18}{11} + \frac{(x-66)^2}{66} + \frac{(y-66)^2}{66} = \frac{26}{3}$$

$$-\frac{18}{11} \qquad -\frac{18}{11}$$

$$66 \cdot \left(\frac{(x-66)^2}{66} + \frac{(y-66)^2}{66} = \frac{232}{33}\right)$$

$$(x-66)^2 + (y-66)^2 = 464$$

10. Suppose a bottling machine that is supposed to fill bottles with an average of 12 ounces of cola per bottle is not functioning correctly and is actually underfilling the bottles. A test of significance is performed on a random sample of bottles in which the null hypothesis represents the machine functioning correctly (i.e. filling the bottles to an average of 12 ounces per bottle), and the alternative hypothesis represents the machine not functioning correctly. If the resulting p-value of this significance test is 0.243, which of the following is most likely to occur as a result?

p-value > α Fail to Reject Ho

A) There is no error and it is assumed the machine is functioning correctly.
B) There is a Type I error and it is assumed the machine is functioning correctly.
[✓ C)] There is a Type II error and it is assumed the machine is functioning correctly.
D) There is a Type I error and it is assumed the machine is not functioning correctly.
E) There is no error and it is assumed the machine is not functioning correctly.

Use the following information to answer #11-12:

The following is a 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-Sq = 43.0%    R-Sq(adj) = 38.9%
```

11. Suppose a sandwich in the sample with 350 calories actually contains 15 grams of protein. Find the value of the residual for this sandwich.

A) -3.33075 grams
[✓ B)] -3.4945 grams
C) 18.33075 grams
D) 18.4945 grams
E) -7.8025 grams

$\hat{y} = 0.06515(350) - 4.308 = 18.4945$

$y - \hat{y}$ [illegible] $= 15 - 18.4945$

12. Which of the following gives a 95% confidence interval estimate for the true slope of the least-squares regression line predict[illegible] protein from calories of the sandwich?

→ that can be use to

A) $0.06515 \pm 3.25(0.02006)$
[✓ B)] $0.06515 \pm 2.145(0.02006)$
C) $0.06515 \pm 1.96(0.02006)$
D) $-4.308 \pm 1.96(8.968)$
E) $-4.308 \pm 0.48(8.968$

$t^* = \text{invT}(.975, df =$ [illegible] $14) = 2.145$

13. You are playing a game in which the random variable X represents the points earned in the first round of the game and the random variable Y represents the points earned in the second round of the game. Let $E(X) = 13$ points, $SD(X) = 3$ points, $E(Y) = 15$ points, and $SD(Y) = 7$ points. Suppose a player's total score for the game is their points earned in the first round added to twice their points earned in the second round. What are the expected value and standard deviation of a player's total score T for the game?

[✓ A)] $E(T) = 43, SD(T) = 14.32$
B) $E(T) = 43, SD(T) = 17$
C) $E(T) = 41, SD(T) = 13$
D) $E(T) = 43, SD(T) = 10.34$
E) $E(T) = 28, SD(T) = 7.62$

$E(x) + 2(E(y)) = 13 + 2(15) = 43$

$SD(x+2y) = \sqrt{3^2 + 2^2 \cdot 7^2} = \sqrt{9+4.49}$

$= \sqrt{205} = 14.32$

14. A manager of a gym would like to estimate the mean amount of time members spend using the facilities per week. They would like their estimate to be within 1 hour of the true mean with 90% confidence. What is the minimum number of members they should include in the sample? Assume a reasonable estimate for σ is 4 hours.

A) 7
B) 13
C) 24
[✓ D) 44]
E) 62

$z^* = \pm 1.645$

$1 \geq 1.645\left(\frac{4}{\sqrt{n}}\right)$

$1 \geq \frac{6.58}{\sqrt{n}}$

$\sqrt{n} \geq 6.58$

$n \geq (6.58)^2$

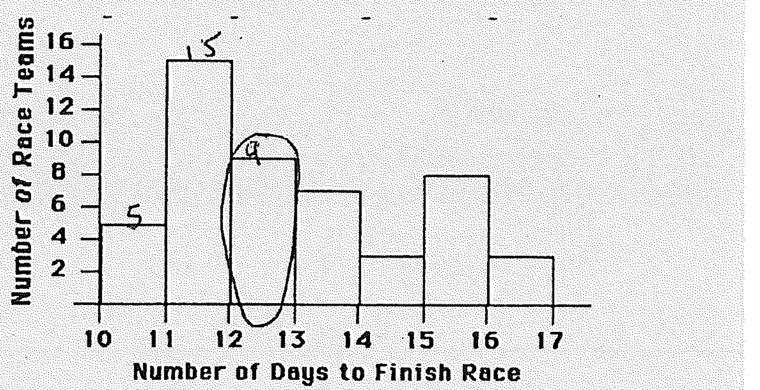
$n \geq 43.2964$

15. The following histogram shows the distribution of race times (in days) for the top 50 finishers in the 1994 Iditarod Dog Sled Race.

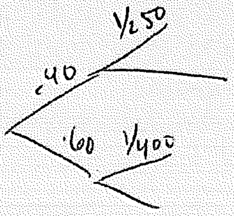
25th ; 26th

Which of the following could be the median race time?

A) 11.480 days
[✓ B) 12.412 days]
C) 13.572 days
D) 14.460 days
E) 16.668 days



16. A manufacturer of calculators receives computer chips from two suppliers: 40% of the chips come from Supplier A and 60% come from Supplier B. It is found that there is a 1 out of 250 chance a chip from Supplier A is defective, and a 1 out of 400 chance a chip from Supplier B is defective. A chip [illegible] chosen at random [illegible] is found to be defective. What is the probability it came from Supplier A?



$P(A|def) = \frac{P(A \cap def)}{P(def)}$

$= \frac{.40(1/250)}{.40(\frac{1}{250}) + .60(\frac{1}{400})} = \frac{.0016}{.0031}$

$.0016 + .0015$ $= .516$

A) 0.004
B) 0.0065
C) 0.0016
D) 0.484
[✓ E) 0.516]

17. A soft drink company is offering a promotion in which there is a winning code contained in the bottle cap of 1 out of every 16 bottles. What is the probability you will need to open at least six bottles in order to find your first winning code?

$p = \frac{1}{16}$ geometric $P(x \geq 6) = 1 - P(x \leq 5)$

1 - geometcdf $(\frac{1}{16}, 5)$

$1 - .276$

$.724$

[✓ A) 0.724]
B) 0.3798
C) 0.274
D) 0.824
E) 0.945

18. A two-proportion z-test is conducted using two independent random samples to test $H_0: p_1 - p_2 = 0$ vs. $H_a: p_1 - p_2 \neq 0$. The resulting p-value is 0.082. For which of the following level(s) of confidence will the two-proportion z-interval constructed using the same sample data contain the value "0"?

Two-tailed

Fail to Reject Ho

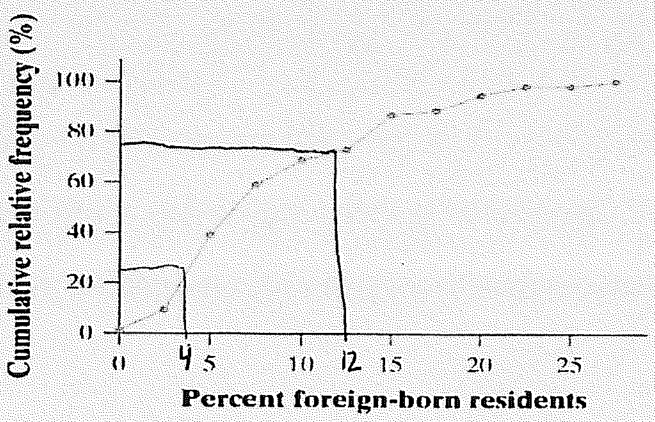
Pvalue $> \alpha$

I. 90% $\alpha = .10$
II. 95% $\alpha = .05$ ✓
III. 99% $\alpha = .01$ ✓

$0.082 > .01$ ✓
$0.082 > .05$ ✓
$0.082 < .10$ ✗

A) I only
B) I, II, and III
C) III only
[✓ D) II and III only]
E) None of these.

19. Use the given cumulative relative frequency graph shown. Which of the following could be the approximate interquartile range of the percent of foreign-born residents in the 50 states?



A) 50%
B) 2%
~~C) 45%~~
[✓ D) 8%]
~~E) 92%~~

20. Researchers want to test a drug manufacturer's claim that fewer than 10% of patients who take its new drug will experience side effects. They decide to carry out a test of $H_0: p = 0.10$ vs. $H_a: p < .10$ where p is the true proportion of patients similar to those in the study who would experience side effects when taking the new drug. They use a random sample of 300 patients and a level of significance of .05. Suppose the true proportion of patients who experience side effects when taking the new drug is actually 8%. In this situation, which of the following would increase the power of the test?

~~I. Use $\alpha = .01$ instead of $\alpha = .05$~~
II. If the true proportion is $p = .06$ instead of $p = .08$
~~III. Use $n = 200$ instead of $n = 300$~~

[✓ A) II only]
~~B) I, II, and III~~
C) III only
D) I and III only
E) None of these.

Tie Breaker

A one-proportion z-interval is constructed to estimate the true proportion of people in a given community who are in favor of building a new shopping mall based on a random sample of n residents of the community. The resulting confidence interval is (0.225, 0.575). A new random sample that includes 4 times as many residents as the original sample results in the same sample proportion. If the level of confidence remains unchanged, what is the margin of error for the confidence interval constructed based on the new sample? Round to 4 decimal places.

$\hat{p} = 0.4$

$E = .175$

$$.175 = z^* \sqrt{\frac{(.4)(.6)}{n}}$$

$$E_{New} = z^* \sqrt{\frac{(.4)(.6)}{4n}}$$

$$E_{New} = \underbrace{\left(z^* \sqrt{\frac{(.4)(.6)}{n}}\right)}\left(\frac{1}{2}\right)$$

$$E_{New} = (.175)\left(\frac{1}{2}\right) = \boxed{.0875}$$