Arkansas Council of Teachers of Mathematics **Name _________________________**

2022 Statistics Regional Competition

Work the multiple-choice questions first, choosing the single best response from the choices available. Indicate your answer here and on your answer sheet. Then attempt the tie-breaker questions at the end starting with tie breaker #1, then #2, and then #3. Turn in your answer sheet, your tie-breaker pages, and your scratch work when you are finished.

1) All of the following decrease the margin of error for a confidence interval except:
   A. Increased variability
   B. Decreased confidence level
   C. Increased sample size
   D. All of these decrease the margin of error
   E. None of these decrease the margin of error

2) The statement that "$P(A|B) = P(B|A)$ whenever A and B are the same event" is:
   A. Always True
   B. Never True
   C. Not enough information; we would need to know if A and B are disjoint events
   D. Not enough information; we would need to know if A and B are mutually exclusive
   E. Not enough information; we would need to know if the events are equally likely

3) The p-value in hypothesis testing represents which of the following:
   A. The probability of failing to reject the null hypothesis, given the observed results
   B. The probability that the null hypothesis is true, given the observed results
   C. The probability that the observed results are statistically significant, given that the null hypothesis is true
   D. The probability that the observed results are statistically significant, given that the null hypothesis is false
   E. The probability of observing results as extreme or more extreme than currently observed, given that the null hypothesis is true

4) Suppose that we have a sample space, $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and events, $A = \{2, 4, 6, 9\}$, and $B = \{1, 3, 7, 8\}$. Which of the following represents $\overline{A \cap B}$, the complement of the intersection of A and B ?
   A. $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$
   B. $\{1, 3, 5, 7, 8\}$
   C. $\{2, 4, 5, 6, 9\}$
   D. $\{5\}$
   E. $\emptyset$

5) The heights of 490 male high school students were recorded in inches. Summary statistics were then calculated for the dataset. However, after the summary statistics were calculated, it was realized that all of the original height measurements errored by one inch. To correct this error, one inch was added to the heights of all 490 original values. Following this one inch correction, which of the following summary statistics would change when recalculated with data from the 'new' sample?

   A. The standard deviation
   B. The variance
   C. The range
   D. The mean
   E. None of the above statistics would change

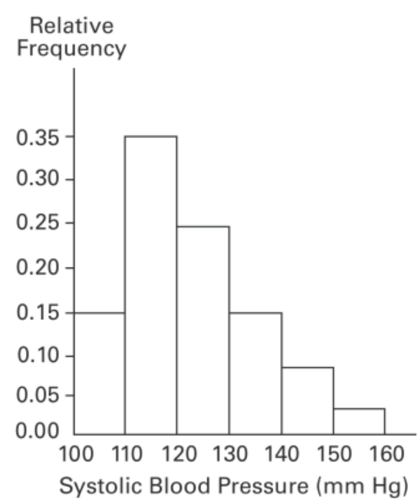
6) A psychologist focusing on popular culture and media believes that the average number of hours per week (hrs/week) spent using social media is greater for women than for men. The psychologist examines two independent simple random samples of 100 individuals each. If the average number of hrs/week spent using social media for the sample of women is 1 hour greater than that for the sample of men, what conclusion can be made from a hypothesis test where:

$$\begin{cases} H_0: \mu_W = \mu_M \\ H_1: \mu_W > \mu_M \end{cases}$$

   A. A hypothesis test cannot ever be performed in scenarios like this, even with the addition of more information
   B. The observed difference in average number of hrs/week spent using social media is not significant
   C. The observed difference in average number of hrs/week spent using social media is significant
   D. A conclusion is not possible without knowing the average number of hrs/week spent using social media in each sample
   E. A conclusion is not possible without knowing the standard deviations for each sample

For Questions 7–10, refer to the histogram, which shows relative frequency of systolic blood pressure for people aged 25 – 40 years.

Systolic Blood Pressure for People Aged 25 – 40 Years



7) How many individuals were included in this sample?

   A. 100
   B. 160
   C. 1
   D. 6
   E. It cannot be determined from the graph

For Questions 7–10, refer to the histogram shown on the previous page.

8) What percentage of peopled aged 25 – 40 years have systolic blood pressure between 110 and 140 inclusive?
   A. 75%
   B. 0.75%
   C. 95%
   D. 0.95%
   E. 9.5%

9) Does this histogram appear to show a sample that follows a Normal distribution?
   A. No, the frequencies start off low, increase to a maximum, and then decrease to low again, and are roughly symmetric.
   B. Yes, the frequencies start off low, increase to a maximum, and then decrease to low again, and are roughly symmetric.
   C. Yes, the distribution appears to be skewed to the right.
   D. No, the distribution appears to be skewed to the right.
   E. It cannot be determined from the graph.

10) Is it unlikely to find someone who has a systolic blood pressure strictly less than 110? (Consider an event to be unlikely if it occurs less than 5% of the time)
   A. Yes, the appropriate probability, 0.15, is less than 0.05
   B. Yes, the appropriate probability, 0.85, is less than 0.05
   C. No, the appropriate probability, 0.15, is greater than 0.05
   D. No, the appropriate probability, 0.85, is greater than 0.05
   E. It cannot be determined from the graph

For Questions 11–14, refer to the table to the right.

	Positive Test Result	**Negative Test Result**
Subject has the disease	111	4
Subject does not have the disease	11	172

11) What is the probability of randomly selecting a patient who has the disease?
   A. 0.386
   B. 0.5
   C. 0.614
   D. 0.910
   E. 0.965

For Questions 11–14, refer to the table on the previous page.

12) What is the probability of randomly selecting a patient who has a positive test result, given that the patient has the disease?
   A. 0.386
   B. 0.5
   C. 0.614
   D. 0.910
   E. 0.965

13) What is the probability of randomly selecting a patient who has the disease, given that the patient has a positive test result?
   A. 0.386
   B. 0.5
   C. 0.614
   D. 0.910
   E. 0.965

14) Based on the previous answers, can we conclude that 'subject has the disease' and 'positive test result' are independent?
   A. Yes, $P(disease) = P(disease|positive)$
   B. No, $P(disease) \neq P(positive|disease)$
   C. Yes, $P(disease) \neq P(disease|positive)$
   D. Yes, $P(disease) = P(positive|disease)$
   E. No, $P(disease) \neq P(disease|positive)$

15) A 99% t-based confidence interval for the mean price for a gallon of milk (dollars) is calculated using a simple random sample of gallon milk prices for 50 groceries. Given that the 99% confidence interval is \$3.52 < μ < \$4.18, what is the sample mean price for a gallon of milk (dollars)?
   Please select the best answer of those provided below.
   A. \$0.33
   B. \$4.00
   C. \$3.85
   D. Not enough information; we would need to know the variation in the sample of gallon gasoline prices
   E. Not enough information; we would need to know the variation in the population of gallon gasoline prices

16) A quiz consists of 10 True/False questions. Assume that the questions are independent. In addition, assume that (T) and (F) are equally likely outcomes when guessing on any one of the questions. What is the probability of guessing on each of the 10 quiz questions and getting three or more of the True/False questions wrong? Round to 3 decimal places.

A. 0.828
B. 0.945
C. 0.117
D. 0.172
E. 0.055

x	$P(x)$
1	0.14
2	0.16
3	0.12
4	0.14
5	0.13
6	0.31

17) Refer to the discrete probability distribution provided in the table to the right. Find the probability that x is equal to 1 or 6.

A. 0.45
B. 0.14
C. 0.17
D. 0.31
E. 0.33

18) Refer to the discrete probability distribution provided in the table on the previous problem. Find the expected value.

A. 0.00
B. 3.50
C. 3.89
D. 4.11
E. 6.00

19) What percentage of measurements in a dataset fall above the median?

A. 0%
B. 0.5%
C. 50%
D. 100%
E. Cannot Be Determined

20) Which of the following exam scores is better relative to other students enrolled in the course?

- A psychology exam grade of 96; the mean grade for the psychology exam is 94 with a standard deviation of 2
- An economics exam grade of 66; the mean grade for the economics exam is 78 with a standard deviation of 7
- A chemistry exam grade of 71; the mean grade for the chemistry exam is 72 with a standard deviation of 5
- A mathematics exam grade of 81; the mean grade for the mathematics exam is 75 with a standard deviation of 4

A. The psychology exam score is relatively better
B. The economics exam score is relatively better
C. The chemistry exam score is relatively better
D. The mathematics exam score is relatively better
E. All of the exam scores are relatively equivalent

21) The statement "If there is sufficient evidence to reject a null hypothesis at the 5% significance level, then there is sufficient evidence to reject it at the 10% significance level" is: Please select the best answer of those provided below.

A. Always True
B. Never True
C. Sometimes True; the hypothesis test would have to be a two-tailed test
D. Sometimes True; the p-value for the statistical test needs provided for a conclusion
E. Not enough information; this would depend on the type of statistical test used

22) Assuming weights of male athletes are normally distributed with a mean of 180 lbs and a standard deviation of 10 lbs, what is the probability that a randomly selected male athlete weighs less than 170 lbs? Round to 3 decimal places.

Also, is the probability on the previous page the same as the probability that a randomly selected sample of size n (where $n > 1$) has a mean weight less than 170 lbs?

A. 0.159; yes, these two probabilities would be the same
B. 0.159; no, these two probabilities would not be the same
C. 0.851; yes, these two probabilities would be the same
D. 0.851; no these two probabilities would not be the same
E. It cannot be determined with the information provided

For Questions 23–25, refer to the relevant results from a regression analysis provided below.

A simple random sample of 5k race times for 32 competitive female runners aged 15-24 years old resulted in a mean 5k race time of 16.79 minutes. The simple linear regression equation that fit the sample data was obtained and found to be $\hat{y} = 21.506 - 0.276x$ where x represents the age of the runner in years and y represents the 5k race time in minutes. When testing the claim that there is a linear correlation between age and 5k race times of competitive female runners, an observed test statistic of $t = -7.87$ resulted in an approximate p-value of 0.0001.

23) The proportion of variation in 5k race times that can be explained by the variation in the age of competitive female runners was approximately 0.64. What is the value of the sample linear correlation coefficient?

A. 0.64
B. 0.8
C. -0.8
D. 0.32
E. -0.64

24) Using all results provided, is it reasonable to predict the 5k race time (minutes) of a competitive female runner 78 years of age?

A. Yes; linear correlation between age and 5k race times is statistically significant
B. No; the age provided is beyond the scope of our available sample data
C. Yes; both the sample linear regression equation and an age in years is provided
D. No; linear correlation between age and 5k race times is not statistically significant
E. Yes; the age provided is beyond the scope of our available sample data

25) The statement "Since the p-value for the simple linear correlation is so small, a line is the best possible fit for this data and no other quadratic or cubic function could possibly fit better," is:

A. Always True
B. Never True
C. Sometimes True; the linear model will only be better than quadratic models
D. Not enough information: we would need to see a scatterplot of the data
E. Cannot be determined

TIEBREAKER #1

Name: ___________________

School: __________________

Assume the following probabilities for two events, A and B:

$$P(A) = 0.30, P(B) = 0.80, P(A \cup B) = 0.86$$

Are the events, A and B, independent in this situation?

You must provide reasoning for your answer.

TIEBREAKER #2

Name: __________________

School: _________________

The table below relates to a study where the length of hands and feet are measured for a group of college freshmen. The goal of the study was to determine if there was a statistically significant association between foot and hand length.

	Right foot longer	Left foot longer	Both feet same	Row sample size
Right hand longer	50.00%	13.64%	36.36%	22
Left hand longer	8.00%	36.00%	56.00%	25
Both hands same	22.64%	24.53%	52.83%	53

The table above provides row percentages and sample sizes. For example, 50.00% of students with 'Right hand longer' have their 'Right foot longer' and there are 22 total 'Right hand longer' students.

1) Based upon the row percentages provided, fill in the contingency table below with the approximate cell counts, rounding to the nearest whole number.

	Right foot longer	Left foot longer	Both feet same
Right hand longer			
Left hand longer			
Both hands same			

2) Do the data indicate an association between foot and hand length? Conduct an appropriate hypothesis test to answer this question using a 0.05 significance level. Provide the hypotheses, test statistic(s), p-value(s), and a formal conclusion.

TIEBREAKER #3

Name: __________________

School: _________________

For $k, n \in \mathbb{N}$ such that $0 \leq k \leq n$, is the following combinatorial statement true?

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$$

You must provide reasoning for your answer.

Multiple Choice Key

1. A	11. A	21. A
2. A	12. E	22. B
3. E	13. D	23. C
4. A	14. E	24. B
5. D	15. C	25. D
6. E	16. B	
7. E	17. A	
8. A	18. C	
9. D	19. C	
10. C	20. D	

-Tie Breaker Question 1-

Assume the following probabilities for two events, A and B:

$$P(A) = 0.30, P(B) = 0.80, P(A \cup B) = 0.86$$

Are the events, A and B, independent in this situation?

You must provide reasoning for your answer.

Solution:

Yes. A and B are independent.

Reasoning (One Possible Answer):

Notice that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

So, $0.86 = 0.30 + 0.80 - P(A \cap B)$

Solving for $P(A \cap B)$, we find that $P(A \cap B) = 0.24$

If A and B are independent, $P(A \cap B) = P(A) * P(B)$

$P(A \cap B) = 0.24 = 0.30 * 0.80 = P(A) * P(B)$

Thus, A and B are independent

Rubric:

0 pts – Answer that A and B are NOT independent

1 pt – Answer that A and B are independent but with no reasoning or improper reasoning

2 pts – Answer that A and B are NOT independent but with an arithmetic error in proper reasoning

3 pts – Answer that A and B are independent with proper reasoning

*Note that 'proper reasoning' implies that the response includes a valid probabilistic definition of independent events such as $P(A \cap B) = P(A) * P(B)$ or $P(A|B) = P(A)$ or $P(B|A) = P(B)$ etc.

-Tie Breaker Question 2-

The table below relates to a study where the length of hands and feet are measured for a group of college freshmen. The goal of the study was to determine if there was a statistically significant association between foot and hand length.

	Right foot longer	Left foot longer	Both feet same	Row sample size
Right hand longer	50.00%	13.64%	36.36%	22
Left hand longer	8.00%	36.00%	56.00%	25
Both hands same	22.64%	24.53%	52.83%	53

The table above provides row percentages and sample sizes. For example, 50.00% of students with 'Right hand longer' have their 'Right foot longer' and there are 22 total 'Right hand longer' students.

Solution Provided in Purple:

3) Based upon the row percentages provided, fill in the contingency table below with the approximate cell counts, rounding to the nearest whole number.

	Right foot longer	Left foot longer	Both feet same
Right hand longer	**11**	**3**	**8**
Left hand longer	**2**	**9**	**14**
Both hands same	**12**	**13**	**28**

4) Do the data indicate an association between foot and hand length? Conduct an appropriate hypothesis test to answer this question using a 0.05 significance level. Provide the hypotheses, test statistic(s), p-value(s), and a formal conclusion

- Hypotheses $\begin{cases} H_0: No\ association\ between\ foot\ and\ hand\ length\ (independent) \\ H_1: There\ is\ an\ association\ between\ foot\ and\ hand\ length\ (dependent) \end{cases}$
- Test Statistic $\chi 2 = 11.942$, $df = 4$
- P-Value $p - value = 0.0178$
- Formal Conclusion at $\boldsymbol{\alpha} = \mathbf{0.05}$ (in terms of $\boldsymbol{H0}$) Reject the null hypothesis at the 5% significance level. There is sufficient evidence to support the claim that foot and hand length are associated/dependent.

Rubric: 0 pts to 5 pts Possible 1 point for each completely correct answer and 0 points for each incorrect answer of the following tie breaker components: (1) approximate cell counts, (2) hypotheses, (3) test statistic, (4) p-value, and (5) formal conclusion

-Tie Breaker Question 3-

For $k, n \in \mathbb{N}$ such that $0 \leq k \leq n$, is the following combinatorial statement true?

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$$

You must provide reasoning for your answer.

Solution:

Yes. The statement is true.

Reasoning (One Possible Arithmetic Based Proof):

Starting with the left hand side (LHS), we will work to get the right hand side (RHS)

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \frac{(n-1)!}{k!\,(n-1-k)!} + \frac{(n-1)!}{(k-1)!\,(n-k)!}$$

$$= (n-1)!\left[\frac{1}{k!\,(n-1-k)!} + \frac{1}{(k-1)!\,(n-k)!}\right]$$

$$= (n-1)!\left[\frac{n-k}{k!\,(n-k)!} + \frac{k}{k!\,(n-k)!}\right]$$

$$= (n-1)!\frac{n}{k!\,(n-k)!}$$

$$= \frac{n!}{k!\,(n-k)!}$$

$$= \binom{n}{k}$$

Thus, the statement is true.

Rubric:

0 pts – Answer the statement is NOT true with no reasoning

1 pt – Answer the statement is true with no reasoning

2 pts – Answer the statement is NOT true but with error in an attempt at proper reasoning

3 pts – Answer the statement is true with proper reasoning