#### -Directions-

This exam includes 25 multiple-choice questions and three open-response questions that might be used as tie breakers. For questions 1 through 25 (the multiple-choice questions), mark your answer choice in the appropriate location on the sheet provided. After completing questions 1 through 25, answer each tie breaker question in sequential order (i.e., complete Question #1 first, then Question #2, and then Question #3 last). Be sure that your name is printed on each of the tie breaker questions. When time is called, you will be asked to turn in your multiple-choice question answer sheet and your written responses to the tie breaker questions.

- 1. The graph to the right represents the observed data distribution for some random variable, X. Referring to the graph, what is the sample size?
  - a. 70
  - b. 249
  - c. 7
  - d. Cannot Be Determined
- 2. Refer to the graph in Question 1. Which of the following is a possible value for the median of the data distribution?
  - a. 13
  - b. 27
  - c. 35
  - d. Cannot Be Determined
- 3. Assume for two events, A and B: P(A) = 0.64 and P(B|A) = 0.64.
  Do not assume anything further about the events A and B. Are the events A and B disjoint? Please select the best answer of those provided below.
  - a. Yes, because P(B|A) = P(A)
  - b. No, because  $P(B \cup A) \neq 0$
  - c. No, because  $P(B \cap A) \neq 0$
  - d. Cannot Be Determined



NAME:

For Questions 4-5, refer to the setting and statistical output provided below.

NAME:

Spring break is almost here! For many families this might be a wonderful week-long staycation (i.e., relaxing at home) but for other families this break offers the perfect time to take a vacation! According to one travel site, the most common types of spring break vacations include beach trips or ski trips.

A study was conducted to determine if the type of spring break vacation impacted total trip cost. Total trip cost was recorded for 47 families taking spring break vacations that were beach trips (labeled 'Beach') and 29 families taking spring break vacations that were ski trips (labeled 'Ski'). Assume that the two samples are independent simple random samples selected from normally distributed populations.

Provided below are sample summary statistics as well as the results of a two-sample independent t-test of means testing the claim that the mean total trip cost for beach trips differs from that of ski trips (i.e.,  $\mu_1 \neq \mu_2$ ), where  $\mu_1$  represents the population mean total trip cost for beach trips and  $\mu_2$  represents the population mean total trip cost for ski trips.

Sample	n	Mean	Std. dev.	
Beach	47	1971.1472	635.57887	
Ski	29	3594.9522	781.05898	

#### Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
μ <sub>1</sub> - μ <sub>2</sub>	-1623.805	172.13721	50.430036	-9.4332015	< 0.0001

- 4. Referring to the setting provided above, which of the following is the alternative hypothesis?
  - a. The mean total trip cost for beach trips does not significantly differ from than the mean total trip cost for ski trips.
  - b. The mean total trip cost for beach trips significantly differs from the mean total trip cost for ski trips.
  - c. The mean total trip cost for beach trips is significantly less than the mean total trip cost for ski trips.
  - d. The mean total trip cost for beach trips is not significantly less than the mean total trip cost for ski trips.

- 5. Refer to the statistical output. Using a 0.05 significance level, which of the following is the most appropriate conclusion for the hypothesis test given the results?
  - a. Reject the null hypothesis; there is sufficient evidence to suggest that the mean total trip cost for beach trips significantly differs from the mean total trip cost for ski trips.
  - b. Reject the null hypothesis; there is not sufficient evidence to suggest that the mean total trip cost for beach trips significantly differs from the mean total trip cost for ski trips.
  - c. Reject the null hypothesis; there is sufficient evidence to suggest that the mean total trip cost for beach trips is significantly less than the mean total trip cost for ski trips.
  - d. Accept the null hypothesis; there is sufficient evidence to suggest that the mean total trip cost for beach trips significantly differs from the mean total trip cost for ski trips.
- 6. What happens to a confidence interval estimate for a population mean when one of the original observations in the sample data is replaced by an extreme outlier? Please select the best answer of those provided below.
  - a. The confidence interval will remain the same width because confidence interval estimates are robust to outliers.
  - b. The confidence interval will narrow because the sample size will have increased.
  - c. The confidence interval will widen because the sample standard deviation will have increased.
  - d. Cannot Be Determined.
- 7. Assume the following Positive Predictive Value (PPV) for the Rapid Influenza Diagnostic Test (RIDT):

P(Influenza | Positive Test Result) = 0.903

Describe the above probability in words. Please select the best answer of those provided below.

- a. 90.3% of people with influenza receive a positive test result.
- b. 90.3% of people with a positive test result have influenza.
- c. 90.3% of people tested have influenza and receive a positive test result.
- d. 90.3% of people tested have influenza.

# ARKANSAS COUNCIL OF TEACHERS OF MATHEMATICS 2020 STATISTICS REGIONAL EXAM

NAME:

- 8. Using a representative sample of high school mathletes, a 95% confidence interval (CI) was constructed for the population percentage of high school mathletes that would be excited to participate in a  $\pi$  day  $\pi$ K run (i.e., an ~ 3.14 kilometer run on March 14<sup>th</sup>). Given that the constructed 95% CI was (85%, 97%), which of the following is false?
  - a. The null hypothesis would be rejected at the 5% significance level when testing the following:  $H_0: p = 0.60$  vs.  $H_1: p \neq 0.60$
  - b. The sample percentage of high school mathletes is 91%
  - c. The standard error of the sample proportion (i.e., the standard deviation of the sample proportion) is 0.06
  - d. Using the same data, a 99% CI for the population percentage of high school mathletes that would be excited to participate in a  $\pi$  day  $\pi$ K run would be wider than the current 95% CI.

For Questions 9-10, suppose that X is a discrete random variable (RV) with a probability distribution function, m(x), defined as follows:

$$m(x) = \begin{cases} 0.235, & x = 1\\ C, & x = 3\\ 0.375, & x = 5 \end{cases}$$

- 9. What is the value of *C*? Round to 3 decimal places.
  - a. 0.390
  - b. 0.610
  - c. 0.140
  - d. 0.627
- 10. Three balls are drawn (with replacement) from an urn containing 9 white balls and X black balls. Let B be the event that exactly two of the three balls drawn from the urn are black. Find P(B | X = 1). Round to 3 decimal places.
  - a. 0.042
  - b. 0.127
  - c. 0.009
  - d. 0.027

- 11. The length of study abroad is recorded (in days) for several high school and college students. The distribution of the observed lengths is positively (i.e., right) skewed with a mean of 84 days and a standard deviation of 16 days. Which of the following is a possible value for the sample median length of study abroad?
  - a. 37
  - b. 84
  - c. 112
  - d. Cannot Be Determined
- 12. As in Question 11 above, the distribution of the observed study abroad lengths is positively (i.e., right) skewed with a mean of 84 days and a standard deviation of 16 days. If the set of all observed lengths are converted to a z-score scale, which of the following describes the shape, center, and spread of the new distribution of these z-scores?
  - a. Normally distributed with mean of 0 and a standard deviation of 1
  - b. Normally distributed with mean of 84 and a standard deviation of 16
  - c. Positively skewed with a mean of 0 and a standard deviation of 1
  - d. Cannot Be Determined

For Questions 13-16, refer to the table, which summarizes the types of beds used and the bedroom sharing status for a sample of 112 children.

	Type of Bed				
Shared Rom	Single	Loft	Bunk		
No	31	15	16		
Yes	14	12	24		

- 13. What is the probability of one randomly selected child sharing a room (i.e., 'Yes') given that the child utilizes a bunk bed? Round to 3 decimal places.
  - a. 0.589
  - b. 0.214
  - c. 0.480
  - d. 0.600

- 14. Assume simple random sampling for the data summarized in the table above. The goal of the study was to determine if there was a statistically significant association between the type of bed used and the bedroom sharing status of children. Which of the following statistical procedures would be most appropriate to test the claim that the type of bed used is associated with the bedroom sharing status of children? Assume that all necessary requirements hold.
  - a. Two-tailed one-sample z-test of proportions
  - b. Two-tailed two-sample z-test of proportions
  - c. Chi-square independence test
  - d. One-tailed two-sample z-test of proportions
- 15. Refer to Question 14. What is the appropriate p-value for testing the claim that that the type of bed used is associated with the bedroom sharing status of children? Round to 3 decimal places.
  - a. 0.028
  - b. 0.008
  - c. 0.004
  - d. 0.210
- 16. Refer to Questions 14-15. Using a 0.01 significance level, what would be the most appropriate conclusion for the hypothesis test given the results? What type of error *might* you be making? Provide both answers, selecting the best answer choice of those provided below.
  - a. Reject H<sub>0</sub>, Possible Type I Error
  - b. Reject H<sub>0</sub>, Possible Type II Error
  - c. Fail to Reject H<sub>0</sub>, Possible Type I Error
  - d. Fail to Reject H<sub>0</sub>, Possible Type II Error
- 17. The US Department of Education reports that 91% of all preK-12 students attend public schools. Among 68 preK-12 students, what is the expected number that do *not* attend public schools? Round to the nearest whole number.
  - a. 62
  - b. 6
  - c. 68
  - d. 2

18. Tired of the school bus?

Assume that waiting times for a school bus in the morning are uniformly distributed between 7 and 18 minutes with an approximate mean of 12.5 minutes and standard deviation of 10.1 minutes. What is the probability that a randomly selected individual waiting for a school bus waits for less than 10 minutes? Round to 3 decimal places.

- a. 0.273
- b. 0.402
- c. 0.598
- d. 0.727
- 19. In a particular area, 78.5% of individuals are either public transit users or female or both, 13.5% of individuals are public transit users and female, and 57.7% of individuals are female. Based upon this information, what is the probability of one randomly selected individual in the area being a public transit user given they are *not* female? Round to 3 decimal places.
  - a. 0.319
  - b. 0.766
  - c. 0.492
  - d. Cannot Be Determined
- 20. Assume that the number of pages in required books for a high school literature course are normally distributed, with a mean of 255 and a variance of 144. The instructor for the literature course announces that the number of pages in the next required book correspond to a z-score of -1.1. How many pages are in this particular book? Round to the nearest whole number.
  - a. 268
  - b. 97
  - c. 255
  - d. 242

# ARKANSAS COUNCIL OF TEACHERS OF MATHEMATICS 2020 STATISTICS REGIONAL EXAM

21. Below is a visual representation of three labeled distributions. Assume all three of these distributions are approximately normal.



Referring to the visual provided above, what can we say about the coefficient of variation (CV) for the three distributions,  $CV_1$ ,  $CV_2$ , and  $CV_3$  for distributions 1, 2, and 3 respectively.

Please select the best answer of those provided below.

- a.  $CV_1 = CV_2 = CV_3$
- b.  $CV_1 < CV_2 < CV_3$
- c.  $CV_1 > CV_2 > CV_3$
- d. Cannot Be Determined
- 22. In the scatterplot provided, two data points are labeled as being potential outliers or influential points. Which of the two points (if included when fitting a simple linear regression model) has the most significant effect on the simple linear regression line? Also, describe the expected effect.

Please select the best answer of those provided below.



- a. 1, Including the point would increase the intercept term and increase the slope term.
- b. 2, Including the point would increase the intercept term and increase the slope term.
- c. 1, Including the point would increase the intercept term and decrease the slope term.
- d. 2, Including the point would increase the intercept term and decrease the slope term.

- 23. Traditionally during prom, a corsage would be purchased for a female date and a boutonniere would be purchased for a male date. Assume that the cost of a boutonniere is normally distributed with a mean of \$17.50 and a standard deviation of \$8.75. If, during a certain prom, 500 boutonnieres are sold, approximately how many cost more than \$30.00? Round to the nearest whole number.
  - a. 0
  - b. 38
  - **c**. 1
  - d. 462
- 24. The deluxe edition of Lizzo's album 'Cuz I Love You' was nominated for Album of the Year at the most recent Grammy Awards and consists of 14 distinct tracks. A Lizzo fan plans to make a short sample 'album' for their mother consisting of 3 randomly selected tracks from the album (in any order). What is the probability that the short sample 'album' of 3 tracks includes the hit single 'Truth Hurts' (one of the fourteen tracks on the album)? Round to 3 decimal places.
  - a. 0.185
  - b. 0.071
  - c. 0.333
  - d. 0.214
- 25. The third quartile of a distribution is 42, the median is 27, the range is 57, and the minimum value is -2. What is the maximum value for the distribution? Please select the best answer of those provided below.
  - a. 59.0
  - b. 57.0
  - c. 55.0
  - d. Cannot Be Determined

### -Tie Breaker Question 1-

A cohort study of several first generation college students was conducted to examine their experiences during their undergraduate career, with particular attention being paid to retention and academic progress.

Correct descriptions of variables in a study is a benefit to researchers by providing them with information that can guide their data analysis and presentation choices. Describe the following data, which was recorded during the study, as Categorical or Quantitative.

<u>Note</u>: If you describe the data as Categorical you must also indicate whether it is nominal, binary/dichotomous, or ordinal. If you describe the data as Quantitative you must also indicate if it is discrete or continuous.

- State of Residence
- Number of Children
- Time Spent Working per Week (hours)
- Pell Grant Received (Yes/No)
- Distance from Hometown (miles)

### -Tie Breaker Question 2-

The table below represents data from a survey of product sale prices (US Dollars, \$) offered through two of the largest online retailers, Amazon and Walmart.

The goal of the study was to determine if average product prices were significantly higher on Amazon compared to Walmart.

Assume that all necessary requirements hold.

Product	Nespresso	Tracfone 60	TI-30XIIS	Peanut M&Ms,	Everest Fanny
	Machine	Minute Card	Calculator	42 OZ	Waist Pack
Amazon	272.05	29.95	13.95	10.54	9.99
Walmart	215.99	19.99	11.99	7.98	8.00

Do the data indicate that average product prices are significantly higher on Amazon compared to Walmart? 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.

### -Tie Breaker Question 3-

Coronavirus is scary! Wash your hands! Take care of yourself and others!

While the virus is spreading across the world in a strange manner, possibly due to global travel (especially air travel) and airport hubs, it is still hypothesized that distance to the source of the outbreak (i.e., distance in kilometers between a given country and the epicenter of the outbreak in Wuhan) is related to the total number of novel coronavirus cases a given country has recorded.

Data (current as of February 23, 2020) resulted in the following simple linear regression equation  $\hat{y} = 12244.12 - 1.86x$ , where x represents the distance between a country and the outbreak epicenter (in kilometers) and y represents the number of novel coronavirus cases in a country.

One currently affected country is approximately 6085 kilometers from the outbreak epicenter and has a residual of - 480. What was the *observed* number of novel coronavirus cases in the country? Round to the nearest whole number. You must provide reasoning for your answer.