-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 break 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. All of the following increase the width of a confidence interval except:
   a. Increased confidence level
   b. Increased variability
   c. Increased sample size
   d. Decreased sample size

2. The statement that "$P(A|B) = P(B|A)$ whenever A and B are independent events" is: Please select the best answer of those provided below.
   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 the events are equally likely

3. The p-value in hypothesis testing represents which of the following: Please select the best answer of those provided below.
   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 of observing results as extreme or more extreme than currently observed, given that the null hypothesis is true

4. Assume that the difference between the observed, paired sample values is defined in the same manner and that the specified significance level is the same for both hypothesis tests. Using the same data, the statement that "a paired/dependent two sample t-test is equivalent to a one sample t-test on the paired differences, resulting in the same test statistic, same p-value, and same conclusion" is:
Please select the best answer of those provided below.

    a. Always True
    b. Never True
    c. Sometimes True
    d. Not Enough Information

5. Note for this question that the odds in favor of an event A are defined as follows: $\frac{P(A)}{1-P(A)}$
For fraternal twins, the odds in favor of having children that are twins are $\frac{1}{16}$. Based upon this information, what is the probability of a fraternal twin not having children that are twins?

    a. $\frac{1}{16}$
    b. $\frac{15}{16}$
    c. $\frac{1}{17}$
    d. $\frac{16}{17}$

6. Refer to the back-to-back stemplot at the right.
Which of the following are true statements regarding the data summarized?

	0	348
	1	01256
843	2	29
65210	3	2557
92	4	
7552	5	6
	6	1458
6	7	09
8541	8	
90	9	

I. The distributions have the same mean
II. The distributions have the same range
III. The distributions have the same variance
IV. The distributions have the same coefficient of variation

    a. II only
    b. II and III
    c. I, III, and IV
    d. II, III, and IV

For Questions 7–10, refer to the table, which relates to the possible epilepsy-depression link.

	Depressive Disorder	
Diagnosed Epilepsy	Present (Yes)	Absent (No)
Epilepsy	37	51
No Epilepsy	24	78

7. What is the probability of one randomly selected individual presenting with a depressive disorder given the individual has diagnosed epilepsy? Round to 3 decimal places.

   a. 0.421
   b. 0.420
   c. 0.195
   d. 0.607

8. Assume simple random sampling for the data summarized in the table above.
   Let p_E represent the proportion of individuals with diagnosed epilepsy ('Epilepsy') that present with a depressive disorder. Let p_{NE} represent the proportion of individuals without diagnosed epilepsy ('No Epilepsy') that present with a depressive disorder.
   What is the 95% confidence interval to estimate $p_E - p_{NE}$, the difference between the population proportions of individuals presenting with a depressive disorder among those with diagnosed epilepsy and among those without diagnosed epilepsy? Round to 3 decimal places.

   a. (0.040, 0.386)
   b. (0.258, 0.577)
   c. (-0.005, 0.142)
   d. (0.053, 0.317)

9. A researcher believes that the proportion of individuals with diagnosed epilepsy that present with a depressive disorder, p_E, is higher than the proportion of individuals without diagnosed epilepsy that present with a depressive disorder, p_{NE}. Testing this claim, what would the resulting p-value be? Round to 3 decimal places.

   a. 0.006
   b. 0.069
   c. 0.003
   d. 0.035

10. Refer to Question 9. Using a 0.10 significance level, which of the following is the most appropriate conclusion given the results?

    a. Reject the null hypothesis; there is sufficient evidence to support the researcher's claim.
    b. Fail to reject the null hypothesis; there is sufficient evidence to support the researcher's claim.
    c. Accept the null hypothesis; there is not sufficient evidence to support the researcher's claim.
    d. Accept the null hypothesis; there is sufficient evidence to support the researcher's claim.

11. A sociologist 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. Examining two independent simple random samples of 100 individuals each, the researcher calculates sample standard deviations of 2.3 hrs/week and 2.5 hrs/week for women and men respectively. 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:

$$\begin{cases} H_0: \mu_W - \mu_M = 0 \\ H_1: \mu_W - \mu_M > 0 \end{cases}$$

    a. The observed difference in average number of hrs/week spent using social media is not significant
    b. The observed difference in average number of hrs/week spent using social media is significant
    c. A conclusion is not possible without knowing the average number of hrs/week spent using social media in each sample
    d. A conclusion is not possible without knowing the population sizes

12. A 99% t-based confidence interval for the mean price for a gallon of gasoline (dollars) is calculated using a simple random sample of gallon gasoline prices for 50 gas stations. Given that the 99% confidence interval is $\$3.32 < \mu < \$3.98$, what is the sample mean price for a gallon of gasoline (dollars)?
Please select the best answer of those provided below.

   a. $0.33
   b. $3.65
   c. Not Enough Information; we would need to know the variation in the sample of gallon gasoline prices
   d. Not Enough Information; we would need to know the variation in the population of gallon gasoline prices

13. A quiz consists of 9 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 9 quiz questions and getting more than one of the True/False questions wrong? Round to 3 decimal places.

   a. 0.998
   b. 0.018
   c. 0.020
   d. 0.980

14. Five students take AP Calculus AB one year and AP Calculus BC the next year. Their overall course grades (%) are listed below for both courses. Which of the following statistical procedures would be most appropriate to test the claim that student overall course grades are the same in both courses? Assume that any necessary normality requirements hold.

Student	1	2	3	4	5
AP Cal AB	80.0%	72.6%	99.0%	91.3%	68.9%
AP Cal BC	85.5%	71.0%	93.2%	93.0%	74.8%

   a. Two-tailed two-sample paired/dependent t-test of means
   b. Two-tailed two-sample independent t-test of means
   c. Two-tailed two-sample independent z-test of means
   d. One-tailed two-sample z-test of proportions

15. Referring to the setting and data provided in Question 14 above, what is the test statistic for testing the claim that student overall course grades are the same in both courses? Round to 3 decimal places.

    a. -0.516

    b. -0.157

    c. 4.306

    d. Not Enough Information; we would need to know the variation in the population

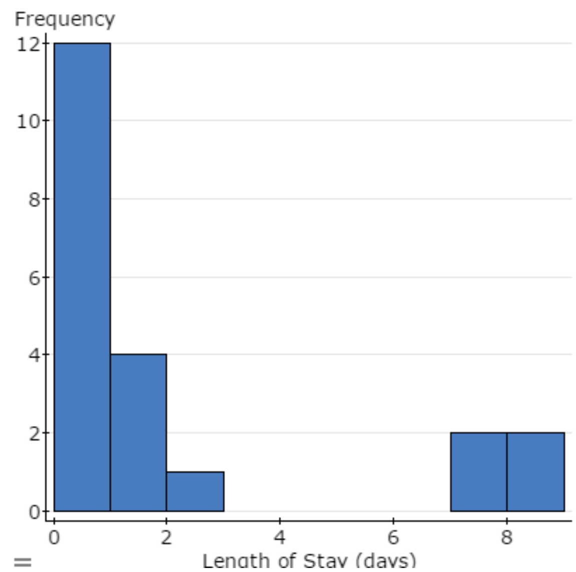
16. The histogram to the right represents the hospital length of stay (in days) for patients at a nearby medical facility. How many patients are included in the histogram?

    a. 5

    b. 21

    c. 17

    d. 9



17. Using the histogram to the right that represents the hospital lengths of stay (in days) for patients at a nearby medical facility, determine the relationship between the mean and the median.

    a. Mean = Median

    b. Mean ≈ Median

    c. Mean < Median

    d. Mean > Median

18. Refer to the discrete probability distribution provided in the table below.

X=x	0	1	2	3	4
P(X=x)	0.040	0.110	0.450	0.230	?

Find the probability that x is equal to 0 or 4. Round to 3 decimal places.

    a. 0.040

    b. 0.210

    c. 0.007

    d. 1.000

19. Green sea turtles have normally distributed weights, measured in kilograms, with a mean of 134.5 and a variance of 49.0. A particular green sea turtle's weight has a z-score of -2.4. What is the weight of this green sea turtle? Round to the nearest whole number.

    a. 17 kg
    b. 151 kg
    c. 118 kg
    d. 252 kg

20. What percentage of measurements in a dataset fall above the median?

    a. 49%
    b. 50%
    c. 51%
    d. Cannot Be Determined

21. Which of the following exam scores is better relative to other students enrolled in the course?
    - A psychology exam grade of 85; the mean grade for the psychology exam is 92 with a standard deviation of 3.5
    - An economics exam grade of 67; the mean grade for the economics exam is 79 with a standard deviation of 8
    - A chemistry exam grade of 62; the mean grade for the chemistry exam is 62 with a standard deviation of 5

    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. All of the exam scores are relatively equivalent

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

    a. Always True
    b. Never True
    c. Sometimes True; the p-value for the statistical test needs to be provided for a conclusion
    d. Not Enough Information; this would depend on the type of statistical test used

23. Assuming weights of female athletes are normally distributed with a mean of 140 lbs and a standard deviation of 15 lbs, what is the probability that a randomly selected female athlete weighs more than 170 lbs? Round to 3 decimal places.

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

    a. 0.023; yes, these two probabilities would be the same
    b. 0.023; no, these two probabilities would not be the same
    c. 0.977; yes, these two probabilities would be the same
    d. 0.977; no these two probabilities would not be the same

For Questions 24–25, refer to the relevant results from a regression analysis provided below.

A simple random sample of 5k race times for 32 competitive male 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 for a competitive male runner in minutes. When testing the claim that there is a linear correlation between age and 5k race times of competitive male runners, an observed test statistic of $(t = -7.87)$ resulted in an approximate p-value of 0.0001.

24. The proportion of variation in 5k race times that can be explained by the variation in the age of competitive male runners was approximately 0.663. What is the value of the sample linear correlation coefficient? Round to 3 decimal places.

    a. 0.663
    b. 0.814
    c. -0.814
    d. 0.440

25. Using all of the results provided, is it reasonable to predict the 5k race time (minutes) of a competitive male runner 73 years of age?

    a. Yes; linear correlation between age and 5k race times is statistically significant
    b. Yes; both the sample linear regression equation and an age in years is provided
    c. No; linear correlation between age and 5k race times is not statistically significant
    d. No; the age provided is beyond the scope of our available sample data

-Tie Breaker Question 1-

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

$P(A) = 0.50$, $P(B) = 0.70$, $P(A \cup B) = 0.85$

Are the events, A and B, independent in this situation?
You must provide reasoning for your answer.

-Tie Breaker Question 2-

The table below relates to a study where infants listened to three types of 'music' in utero and their advancement to crawling/walking was then observed and categorized as either early, on time, or late. The goal of the study was to determine if there was a statistically significant association between music in utero and time of advancement to crawling/walking in infants.

	Advancement to Crawling/Walking			Row Sample Sizes
Music In Utero	Early	On Time	Late	N
Mozart (Piano Sonata)	50.8%	30.2%	19.0%	63
Philip Glass (minimalist music)	40.0%	38.3%	21.7%	60
White Noise and Silence	17.9%	21.1%	61.0%	44

The table above provides row percentages and sample sizes. For example, 50.8% of 'Mozart' infants have 'Early' advancement to crawling/walking and there are 63 total 'Mozart' infants.

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

	Advancement to Crawling/Walking		
Music In Utero	Early	On Time	Late
Mozart (Piano Sonata)			
Philip Glass (minimalist music)			
White Noise and Silence			

b. Do the data indicate an association between music in utero and time of advancement to crawling/walking in infants? 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-

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

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

You must provide reasoning for your answer.

Multiple Choice Key

1. c
2. d
3. d
4. a
5. d
6. b
7. b
8. d
9. c
10. a
11. b
12. b
13. d
14. a
15. a
16. b
17. d
18. b
19. c
20. d
21. c
22. c
23. b
24. c
25. d

-Tie Breaker Question 1-

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

$P(A) = 0.50$, $P(B) = 0.70$, $P(A \cup B) = 0.85$

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, $0.85 = 0.50 + 0.70 - P(A \cap B)$

Solving for $P(A \cap B)$, we find that $P(A \cap B) = 0.35$

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

$P(A \cap B) = 0.35 = 0.50 * 0.70 = P(A) * P(B)$

Thus, A and B are independent

Rubric:

0 pts – Answer that A and B are NOT independent

1 pts – 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-

	Advancement to Crawling/Walking			Row Sample Sizes
Music In Utero	Early	On Time	Late	N
Mozart (Piano Sonata)	50.8%	30.2%	19.0%	63
Philip Glass (minimalist music)	40.0%	38.3%	21.7%	60
White Noise and Silence	17.9%	21.1%	61.0%	44

Solution Provided in Red:

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

	Advancement to Crawling/Walking		
Music In Utero	Early	On Time	Late
Mozart (Piano Sonata)	32	19	12
Philip Glass (minimalist music)	24	23	13
White Noise and Silence	8	9	27

b. Do the data indicate an association between music in utero and time of advancement to crawling/walking in infants? 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**
  $$\begin{cases} \mathrm{H}_0\text{: Music in Utero and Advancement are Independent (there is no association)} \\ \mathrm{H}_1\text{: Music in Utero and Advancement are Dependent (there is association)} \end{cases}$$
- **Test Statistic**
  $\chi^2 = 27.516,\ df = 4$
- **P-Value**
  $p - value < 0.0001 \approx 0$
- **Formal Conclusion at $\alpha = 0.05$ (in terms of H_0)**
  Reject the null hypothesis at the 5% significance level. There is sufficient evidence to support the claim that Music in Utero and advancement to crawling/walking 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 $k, m, n \in \mathbb{N}$ such that $0 \leq k \leq m \leq n$, is the following combinatorial statement true?

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

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)

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

Notice that an $m!$ can be canceled

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

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

Multiply both numerator and denominator by $(n-k)!$

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

Notice that this is the RHS

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

Thus, the statement is true.

Rubric:

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

1 pts – 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