January Regional **Statistics Test**

NOTA stands for None of the Above answers are correct.

1. Which of the following correlation coefficients would yield a coefficient of determination of 0.64?
   A. 0.4096 B. – 0.8 C. – 0.64 D. – 0.4096 E. NOTA

2. The six sides of Die A have the numbers 2,4,6,8,10,12. The six sides of Die B have the numbers 1,3,4,5,7,9. If Die A and Die B are both rolled one time, what is the probability that the sum of the resulting numbers is an even number?
   A. 1/6 B. ½ C. 1/3 D. ¼ E. NOTA

3. Selecting every 5th person that enters a classroom is known as _____ sampling.
   A. Cluster B. Systematic. C. Multistage D. Volunteer E. NOTA

4. The following data set is in order from smallest to largest: { a, b, c, d, e, f, g, h, i } Find the five number summary of the data set.
   A. { a, c, e, g, i } B. { a, 0 .5(b+ c), e, 0.5(g + h) , i }
   C. { a, b, e, h, i } D. { a, 0 .5(b+ c), 0.5(d + e), 0.5(g + h) , i } E. NOTA

5. Using the data set in question #4, which of the following is an **outlier boundary**:
   A. 0.5 (g + h) – 1.5 {0.5 (g + h) - 0.5 (b + c) } B. 0.5 (g + h) + 1.5 {0.5 (b + c) - 0.5 (g + h) }
   C. 0.5 (b + c) – 1.5 {0.5 (g + h) - 0.5 (b + c) } D. 0.5 (b + c) + 1.5 {0.5 (g + h) - 0.5 (b + c) } E. NOTA

6. You and seven of your friends (none of whom are related) are in a room. What is the probability that at least two of those people in the room were born on the same **day** of the week?
   A. 0.0 B. 0.20 C. 0.50 D. 1.0 E. NOTA

7. Which of the following statements is **false**?
   A. Correlation and regression describe only linear relationships
   B. The correlation coefficient and the least squares regression line are very resistant
   C. Correlation does not imply causation.
   D. A correlation coefficient of –0.9 is just as strong as a correlation coefficient of +0.9.
   E. NOTA

8. The distribution of shoe sizes of adult women is N(7, 2). A shoe store wanting to purchase 100 shoes that would fit the largest percentage of adult women should buy shoes from size ___ to size ___.
   A. 2, 7 B. 5, 9 C. 8, 15 D. 4, 10 E. NOTA

9. Consider the possible values of the standard deviation, correlation coefficient, coefficient of determination. If you add the smallest possible value to the largest possible value, which sum is the greatest?
   A. Standard deviation B. Correlation coefficient C. Coefficient of determination D. The sums of each are equal E. NOTA

10. If you multiply every number in a data set by a positive number greater than 1, which of the following will **not** change?
    A. Range B. IQR C. Median D. Standard Deviation E. NOTA

11. The test scores for the first nine weeks exam in a Statistics class are normally distributed. The teacher adds 5 extra credit points to every test score in the class. What will happen to the z-score of a randomly selected student in the class?
    A. It will go up by 5 points B. It will go up by 0.5 points C. It will go up by $(5/\sigma)$ points D. It will remain unchanged E. NOTA

12. Which statement does **not** contain an error in the proper usage of the Statistics term **correlation**?
    A. There is a high correlation between the new and old Social Studies book.
    B. There is a high correlation between race of a worker and his/her income.

C. There is a high correlation (r = -.93) between the travel speed and the driving time to work.
D. There is a high correlation (r = 1.13) between the height and income of American workers.
E. NOTA

13. Find the slope the least squares regression line with the following values: $S_x = 0.45$, $\bar{x} = 4$, $S_y = 0.9$, $\bar{y} = 3$, $r = 0.8$.
A. 1.6 B. 1.28 C. 0.4 D. 0.2 E. NOTA

14. Find the equation of the least squares regression line with the following values: $S_x = 0.45$, $\bar{x} = 4$, $S_y = 0.9$, $\bar{y} = 3$, $r = 0.8$
A. $\hat{y} = -3.4 + 1.6x$ B. $\hat{y} = -2.12 + 1.28x$ C. $\hat{y} = 13 + 0.4x$ D. $\hat{y} = 2.2 + 0.2x$ E. NOTA

15. A statistician at the Dept. of Motor Vehicles has noticed that the number of drivers needing license plates is rapidly approaching the number of 6 digit license plate tags currently available. The statistician suggests that the state abandon the use of 6 numbers in favor of a plate with 6 letters. The statistician claims that using 6 letters instead of 6 numbers will provide _____ **additional** unique license plates.
A. 308,915,776 B. 307,915,776 C. 6^{26} D. 26,000,000 E. NOTA

16. Which of the following is true about experiments and observational studies?
A. An experiment must have a control group B. All experiments can be made into a blind or double-blind experiments C. Your study is not considered an experiment if the subjects select their own treatment D. In an observational study, cause and effect may only be attributed to the treatment if a control group is present. E. NOTA

17. Fifty-two percent of voters surveyed planned to vote for John Dough for mayor. There was a margin of error of 10%. Which of the following is true?
A. We are 90% confident that John Dough will be elected mayor. B. There is a 10% chance that John Dough will not be elected mayor if the election were held today. C. Ten percent of those responding to the survey might change their minds before the election. D. The election is too close to call since we have a certain level of confidence that 42% to 62% of the voters favor John Dough. E. NOTA

18. Which of the following statements (concerning probability) is false?
A. The addition rule finds the probability that at least one of two things happens.
B. The multiplication rule finds the probability that two things both happen.
C. Multiplying the unconditional probabilities of two events requires them to be independent.
D. Two events are mutually exclusive if the occurrence of one event does not change the chances for the other.
E. NOTA

19. Which of the following are false?
A. If a control group is comparable to the treatment group (apart from the treatment) then a difference in the responses of the two groups is likely to be due to the effect of the treatment.
B. If a treatment group is different from the control group with respect to other factors, the effects of these other factors are likely to be confounded with the effect of the treatment.
C. To make sure that the treatment group is like the control group, investigators put subjects into control or treatment groups at random in a randomized controlled experiment.
D. In a blind experiment, the subjects do not know whether they are in the treatment group or the control group and neither does the person who evaluates the responses.
E. NOTA

20. Find P(A and B) if P(A) is 0.3 and the P(B) is 0.4, and events A and B **are** independent.
A. 0.012 B. 0.12 C. 0.1 D. 0.7 E. NOTA

21. Find P(A and B) if P(A) is 0.3 and the P(B) is 0.4, and events A and B **are** mutually exclusive.
A. 0.012 B. 0.12 C. 0.1 D. 0.7 E. NOTA

22. When finding a p-value, you are finding....
    A. The probability of getting the exact value that you have gotten
    B. The probability of getting a value that far away from the mean or farther
    C. The probability of getting a value that far away from the mean or farther if the null hypothesis is true
    D. The probability of getting a value that far away from the mean or farther if the alternative hypothesis is true.
    E. NOTA

23. A die has 9 sides. Three of those sides have two pips (the little dots on the side of the die). Four of the sides have 1 pip. One side has 8 pips and the last side has nine pips. How many times would you expect to roll the die (on average) in order to get a side with two pips.
    A. 2 B. 3 C. 8 D. 9 E. NOTA

24. Find the mean of the following:

Value	X	2X	7X
Prob	0.4	0.1	0.5

    A. 2X B. 3.1X C. 4.1X D. 9.5X E. NOTA

25. The _____ of x-bar describes how the _____x-bar varies in all possible samples from the population.
    A. Standard deviation, statistic B. Standard deviation, parameter C. Sampling distribution, parameter D. Sampling Distribution, statistic E. NOTA

26. If the power of a test is 0.8, find the probability of a Type II error.
    A. 0.2 B. 0.4 C. 0.64 D. 0.8

27. If you wish to compare a sample proportion to a know proportion of the population, you would use a _____. If you wish to compare two sample proportion each other, you would use a _____. If you wish to compare a three of more proportions, you would use a _____.
    A. One Proportion Z-test, Two Proportion Z-test, Multiple Proportion T-Test
    B. One Proportion Z-test, Two Proportion Z-test, Chi Square Goodness of Fit,
    C. One Proportion Z-test, Two Proportion Z-test, Chi Square Test for Independence
    D. One Proportion Z-test, Chi Square Test for Independence, Chi Square Goodness of Fit,
    E. NOTA

28. Exactly 9% of the 2004 taxpayers filed extensions on their tax returns. The mean of the sampling distribution of proportion of 2004 taxpayers who filed extensions on their tax returns is....
    A. An unknown parameter B. An unknown statistic C. Approximately 9% D. Exactly 9% E. NOTA

29. Which of the following probabilities is higher?
    A. Guessing the 4 numbers that will be chosen in the lottery. (Numbers from 1 to 40 may be used. The order doesn't matter and no number may be chosen twice.)
    B. Guessing the exact order that 10 horses place in a race with no prior knowledge of the track record of the horses. (There are only 10 horses in the race).
    C. Guessing exactly 8 out of 10 questions correct on a multiple choice test with answer choices A – D, and no prior knowledge of the content of the test.
    D. Guessing a three number locker "combination" on a lock with numbers from 0 through 40 when all possible sequences of numbers are possible "combinations".
    E. NOTA

30. If you draw independent observations at random from any population with a finite mean, μ, you will notice that as the number of observations drawn increases, the mean of the observed values eventually approaches the mean, μ, of the population and stays there. This is known as the...
    A. Central Limit Theorem B. Law of Large Numbers C. Rule of Thumb for Means D. Empirical Rule

January Regional **Statistics Test**

NOTA stands for None of the Above answers are correct.

Solutions

1. Which of the following correlation coefficients would yield a coefficient of determination of 0.64?
   A. 0.4096 B. – 0.8 C. – 0.64 D. – 0.4096 E. NOTA

   B. The correlation is the square root of the coefficient of determination, and one of the square roots of 0.64 is –0.8 which is choice B.

2. The six sides of Die A have the numbers 2,4,6,8,10,12. The six sides of Die B have the numbers 1,3,4,5,7,9. If Die A and Die B are both rolled one time, what is the probability that the sum of the resulting numbers is an even number?
   A. 1/6 B. ½ C. 1/3 D. ¼ E. NOTA

   A. There are two ways to have an even sum: two even numbers or two odd numbers. Die A has only even numbers, so the number from Die B that must pair up with any of these must also be even. There is only one even number of out 6 on Die B, so the answer is 1/6.
   Another Solution: P(A) x P(B)
   1/1 x 1/6 = 1/6

3. Selecting every 5th person that enters a classroom is known as _____ sampling.
   A. Cluster B. Systematic. C. Multistage D. Volunteer E. NOTA

   B. Systematic sampling is taking every nth person.

4. The following data set is in order from smallest to largest: { a, b, c, d, e, f, g, h, i } Find the five number summary of the data set.
   A. { a, c, e, g, i } B. { a, 0 .5(b+ c), e, 0.5(g + h) , i }
   C. { a, b, e, h, i } D. { a, 0 .5(b+ c), 0.5(d + e), 0.5(g + h) , i } E. NOTA

   B. The min and max are obviously a and i. In a nine number data set, the median is the 5th number (which is e). The average of the 2nd and 3rd number will be Q_1. The average of the 7th and 8th number will be Q_3.

5. Using the data set in question #4, which of the following is an **outlier boundary**:
   A. 0.5 (g + h) – 1.5 {0.5 (g + h) - 0.5 (b + c) } B. 0.5 (g + h) + 1.5 {0.5 (b + c) - 0.5 (g + h) }
   C. 0.5 (b + c) – 1.5 {0.5 (g + h) - 0.5 (b + c) } D. 0.5 (b + c) + 1.5 {0.5 (g + h) - 0.5 (b + c) } E. NOTA
   C. The formula for the lower boundary is :
   Quartile 1 - 1.5 (Quartile 3 – Quartile 1) Q_1 = 0.5 (b + c) Q_3= 0.5 (g + h)
   Substituting... 0.5 (b + c) – 1.5 {0.5 (g + h) - 0.5 (b + c) }

6. You and seven of your friends (none of whom are related) are in a room. What is the probability that at least two of those people in the room were born on the same **day** of the week?
   A. 0.0 B. 0.20 C. 0.50 D. 1.0 E. NOTA

   D. "You and seven of your friends" means there are 8 people in a room. There are only 7 days on which to be born. Obviously at least 2 people must be born on the same day.

7. Which of the following statements is **false**?
   A. Correlation and regression describe only linear relationships
   B. The correlation coefficient and the least squares regression line are very resistant
   C. Correlation does not imply causation.
   D. A correlation coefficient of –0.9 is just as strong as a correlation coefficient of +0.9.
   E. NOTA

   B. **The correlation coefficient and the least squares regression line are NOT very resistant.**

8. The distribution of shoe sizes of adult women is N(7, 2). A shoe store wanting to purchase 100 shoes that would fit the largest percentage of adult women should buy shoes from size ___ to size ___.
A. 2, 7 B. 5, 9 C. 8, 15 D. 4, 10 E. NOTA

D. Answer A includes about 49% of the population. Answer B includes 68% of the population. Answer C includes about 31% of the population. Answer D includes 86% of the population.

9. Consider the possible values of the standard deviation, correlation coefficient, and coefficient of determination. If you add the smallest possible value to the largest possible value, which sum is the greatest?
A. Standard deviation B. Correlation coefficient C. Coefficient of determination D. The sums of each are equal E. NOTA

A. The smallest standard deviation is 0 and there is no maximum standard deviation. However, the
Smallest correlation is –1 and largest is 1, and that sum is 0. The smallest coefficient of determination is 0 and the largest is 1, so that sum is 1. This means that as long as standard deviations can be greater than 1 (which they can), the sum of the smallest and largest will exceed the sums of the other two choices.

10. If you multiply every number in a data set by a positive number greater than 1, which of the following will **not** change?
A. Range B. IQR C. Median D. Standard Deviation E. NOTA

E. Each of the choices listed will change when a data set is multiplied by a positive number greater than one.

11. The test scores for the first nine weeks exam in a Statistics class are normally distributed. The teacher adds 5 extra credit points to every test score in the class. What will happen to the z-score of a randomly selected student in the class?
A. It will go up by 5 points B. It will go up by 0.5 points C. It will go up by $(5/\sigma)$ points D. It will remain unchanged E. NOTA

D. A z-score remains unchanged as it is a standardized value.

12. Which statement does **not** contain an error in the proper usage of the Statistics term **correlation**?
A. There is a high correlation between the new and old Social Studies book.
B. There is a high correlation between race of a worker and his/her income.
C. There is a high correlation (r = -.93) between the travel speed and the driving time to work.
D. There is a high correlation (r = 1.13) between the height and income of American workers.
E. NOTA

C. Travel speed and driving time are negatively related, and – 0.93 is probably a realistic correlation.
We would not expect a perfect correlation since traffic lights will add to the travel time.

13. Find the slope the least squares regression line with the following values: $S_x = 0.45$, $\bar{x} = 4$, $S_y = 0.9$, $\bar{y} = 3$, $r = 0.8$.
A. 1.6 B. 1.28 C. 0.4 D. 0.2 E. NOTA

A. The formula for slope is (r) (S_y/S_x).
Substituting we get: (0.8) (0.9/0.45) = 1.6

14. Find the equation of the least squares regression line with the following values: $S_x = 0.45$, $\bar{x} = 4$, $S_y = 0.9$, $\bar{y} = 3$, $r = 0.8$.
A. $\hat{y} = -3.4 + 1.6x$ B. $\hat{y} = -2.12 + 1.28x$ C. $\hat{y} = 13 + 0.4x$ D. $\hat{y} = 2.2 + 0.2x$ E. NOTA

A. The formula is y-hat = a + bx. **In the previous problem we established that b = 1.6. The formula for a = y-bar – b (x-bar). Substituting a = 3 – 1.6(4) A = -3.4**

If a = - 3.4 and b= 1.6, then y-hat = - 3.4 + 1.6 x

15. A statistician at the Dept. of Motor Vehicles has noticed that the number of drivers needing license plates is rapidly approaching the number of 6 digit license plate tags currently available. The statistician suggests that the state abandon the use of 6 numbers in favor of a plate with 6 letters. The statistician claims that using 6 letters instead of 6 numbers will provide _____ **additional** unique license plates.
A. 308,915,776 B. 307,915,776 C. 6^{26} D. 26,000,000 E. NOTA

B. The current license plates available are 10^6 plates. The plates available with letters will be 26^6. The additional plates available will be $26^6 - 10^6$, which is answer B. Answer A assumes they forgot to subtract the current amount of license tags available.

16. Which of the following is true about experiments and observational studies?
A. An experiment must have a control group B. All experiments can be made into a blind or double-blind experiments C. Your study is not considered an experiment if the subjects select their own treatment D. In an observational study, cause and effect may only be attributed to the treatment if a control group is present. E. NOTA

C. If subjects select their own treatment, you are merely observing, and hence it is not an experiment. An experiment does not have to have a control group, so A is false There are experiments in which it is not possible to be blind or double blind (Example: One group swims twice a week and the other runs twice a week to reduce cholesterol. Both groups know they are running or swimming.)

17. Fifty-two percent of voters surveyed planned to vote for John Dough for mayor. There was a margin of error of 10%. Which of the following is true?
A. We are 90% confident that John Dough will be elected mayor. B. There is a 10% chance that John Dough will not be elected mayor if the election were held today. C. Ten percent of those responding to the survey might change their minds before the election. D. The election is too close to call since we have a certain level of confidence that 42% to 62% of the voters favor John Dough. E. NOTA

D. If you subtract 10% from 52% you get 42%. Adding 10%, you get 62%. The election is too close to call since you are saying that the candidate will either win or lose.

18. Which of the following statements (concerning probability) is false?
A. The addition rule finds the probability that at least one of two things happens.
B. The multiplication rule finds the probability that two things both happen.
C. Multiplying the unconditional probabilities of two events requires them to be independent.
D. Two events are mutually exclusive if the occurrence of one event does not change the chances for the other.
E. NOTA

D. The reason D is false is that this is the definition of independence, not mutually exclusive.

19. Which of the following are false?
A. If a control group is comparable to the treatment group (apart from the treatment) then a difference in the responses of the two groups is likely to be due to the effect of the treatment.
B. If a treatment group is different from the control group with respect to other factors, the effects of these other factors are likely to be confounded with the effect of the treatment.
C. To make sure that the treatment group is like the control group, investigators put subjects into control or treatment groups at random in a randomized controlled experiment.

D. In a blind experiment, the subjects do not know whether they are in the treatment group or the control group and neither does the person who evaluates the responses.
E. NOTA

D. The reason D is false is that D describes a double blind situation.

20. Find P(A and B) if P(A) is 0.3 and the P(B) is 0.4, and events A and B **are** independent.
A. 0.012 B. 0.12 C. 0.1 D. 0.7 E. NOTA

B. For independent events, multiply the probabilities. 0.3 X 0.4 = 0.12

21. Find P(A and B) if P(A) is 0.3 and the P(B) is 0.4, and events A and B **are** mutually exclusive.
A. 0.012 B. 0.12 C. 0.1 D. 0.7 E. NOTA

E. If the events are mutually exclusive, there is nothing in common so the union of these should yield nothing.

22. When finding a p-value, you are finding....
A. The probability of getting the exact value that you have gotten
B. The probability of getting a value that far away from the mean or farther
C. The probability of getting a value that far away from the mean or farther if the null hypothesis is true
D. The probability of getting a value that far away from the mean or farther if the alternative hypothesis is true.
E. NOTA

C. The important part of this answer is the "if the null hypothesis is true"

23. A die has 9 sides. Three of those sides have two pips (the little dots on the side of the die). Four of the sides have 1 pip. One side has 8 pips and the last side has nine pips. How many times would you expect to roll the die (on average) in order to get a side with two pips.
A. 2 B. 3 C. 8 D. 9 E. NOTA

B. We are looking for the expected value. Expected value = 1/p. The probability of success is 1/3. The reciprocal is 3.

24. Find the mean of the following:

Value	X	2X	7X
Prob	0.4	0.1	0.5

A. 2X B. 3.1X C. 4.1X D. 9.5X E. NOTA

C. To get the mean, you multiply vertically and add. Mean = 0.4X + 0.1(2X) + 0.5(7x). Mean = 4.1X

25. The _____ of x-bar describes how the _____x-bar varies in all possible samples from the population.
A. Standard deviation, statistic B. Standard deviation, parameter C. Sampling distribution, parameter D. Sampling Distribution, statistic E. NOTA

D. The _ Sampling Distribution of x-bar describes how the statistic, x-bar varies in all possible samples from the population.

26. If the power of a test is 0.8, find the probability of a Type II error.
A. 0.2 B. 0.4 C. 0.64 D. 0.8

A. Type II error = 1 - Power

27. If you wish to compare a sample proportion to a know proportion of the population, you would use a _____. If you wish to compare two sample proportion each other, you would use a _____. If you wish to compare a three of more proportions, you would use a _____.
    A. One Proportion Z-test, Two Proportion Z-test, Multiple Proportion T-Test
    B. One Proportion Z-test, Two Proportion Z-test, Chi Square Goodness of Fit,
    C. One Proportion Z-test, Two Proportion Z-test, Chi Square Test for Independence
    D. One Proportion Z-test, Chi Square Test for Independence, Chi Square Goodness of Fit,
    E. NOTA

B. This is the progression from proportions to Chi Square.

28. Exactly 9% of the 2004 taxpayers filed extensions on their tax returns. The mean of the sampling distribution of proportion of 2004 taxpayers who filed extensions on their tax returns is....
    A. An unknown parameter B. An unknown statistic C. Approximately 9% D. Exactly 9% E. NOTA

D. The mean of the sampling distribution is the EXACTLY equal to p. (This was on the free response
Of one of the first AP exams.)

29. Which of the following probabilities is higher?
    A. Guessing the 4 numbers that will be chosen in the lottery. (Numbers from 1 to 40 may be used. The order doesn't matter and no number may be chosen twice.)
    B. Guessing the exact order that 10 horses place in a race with no prior knowledge of the track record of the horses. (There are only 10 horses in the race).
    C. Guessing exactly 8 out of 10 questions correct on a multiple choice test with answer choices A – D, and no prior knowledge of the content of the test.
    D. Guessing a three number locker "combination" on a lock with numbers from 0 through 40 when all possible sequences of numbers are possible "combinations".
    E. NOTA

C. To solve for A, use $1/{}_{40}C_4$ and this is 0.0000109
To solve for B, use 1/ 10 ! and this is 0.000000276
To solve for C, use 1/ binomial pdf (10, 0.25, 8) and this is 0.000386
To solve for D, use $1/40^3{}_4$ and this is 0.0000156

30. If you draw independent observations at random from any population with a finite mean, μ, you will notice that as the number of observations drawn increases, the mean of the observed values eventually approaches the mean, μ, of the population and stays there. This is known as the...
    A. Central Limit Theorem B. Law of Large Numbers C. Rule of Thumb for Means D. Empirical Rule

B. This is the definition of LOLN.