Important Instructions for this Test: Please pay close attention to and carefully follow all rounding instructions. Round any intermediate steps as indicated or as necessary to make the final answer as accurate as possible. Good luck, have fun, and as always: "NOTA" stands for "None of These Answers is correct."

Use the following set of bivariate data from explanatory variable X and response variable Y to answer the first fifteen questions. NOTE: Perform all requested calculations even though the relationship between X and Y is not actually linear. You will appreciate why as you go along! HINT: Always graph your data!

X	1	2	3	4	5	6	7	8	9	10
Y	25	17	11	7	5	5	7	11	17	25

1. What is the sample mean of the data set from variable Y rounded to the tenths place?

A: 4.5 B: 5.0 C: 5.5 D: 13.0 E: NOTA

2. What is the sample standard deviation of the data set from variable X rounded the thousandths place?

A: 7.659 B: 7.266 C: 3.028 D: 2.872 E: NOTA

3. According to the 1.5(IQR) Rule and the sample of data from variable Y above, a data value is considered a high outlier in data set Y if it is greater than _____. The missing value in the blank also corresponds to being how many standard deviations above the mean (to the nearest hundredth)?

A: 1.96 B: 2.48 C: 2.61 D: 2.70 E: NOTA

4. Suppose we randomly select an ordered pair (x, y) from the table above. What is the probability that both the x-coordinate and the y-coordinate of the ordered pair are a prime?

A: $\frac{2}{5}$ B: $\frac{1}{2}$ C: $\frac{4}{5}$ D: $\frac{9}{10}$ E: NOTA

5. Which of the following graphical representations best conveys the relationship between X and Y?

A: A pair of side-by-side bar graphs.
B: A pair of side-by-side histograms.
C: A back-to-back stem plot.
D: A scatterplot of Y versus X.
E: NOTA

6. What is the sample Pearson coefficient of linear correlation between X and Y rounded to four decimal places? Remember, perform this calculation even though the relationship between X and Y is not actually linear!

A: 0.0000 B: 1.0000 C: 0.5000 D: It is undefined. E: NOTA

7. Use the data in the table above to build a least squares linear regression model to predict Y from X. What is the predicted value of Y for X = 5 rounded to the nearest tenth? Again, do this even though the actual relationship between X and Y is not linear!

A: 5.0 B: 5.5 C: 13.0 D: It is undefined. E: NOTA

8. According to the linear regression model you just constructed in the previous question, which of the following represents the standard error of the sample slope rounded to four decimal places?

A: 0.8944 B: 3.0277 C: 7.6594 D: 8.1240 E: NOTA

9. Let R^2 represent the sum of the squared residuals of the linear regression model for predicting Y from X you used in the previous questions. Let V^2 represent the sample variance of the values of the dependent variable Y from the table above. Compute $R^2 - 9V^2$ as an exact value.

A: 0 B: 59 C: 528 D: -528 E: NOTA

10. Since the bivariate relationship between X and Y displayed in the data above is non-linear, which of the following transformations of the data absolutely guarantees a linear relationship between the transformed data sets?

A: ln(Y) vs. ln(X)
B: Y vs. ln(X)
C: ln(Y) vs. X
D: Either the relationship between X and Y is already linear and so there is no need to transform the data or there is no possible way to transform either X and / or Y so that the resulting transformed data shows a linear pattern.
E: NOTA

11. It is often possible for a set of bivariate data to show a statistically significant linear association both before and after a transformation such as one of the ones suggested in the previous question in choices A, B, and C. Perform the appropriate statistical test on the original bivariate data in the table and for the three transformations suggested in the previous question. How many of the relationships in the following set {X vs. Y; ln(Y) vs. ln(X); Y vs. ln(X); ln(Y) vs. X} show statistically significant evidence of a linear association at the 5% level?

A: 0 B: 1 C: 2 D: 3 E: NOTA

12. Construct a 95% confidence interval for the population mean of the response variable μ_Y based on the sample of data from variable Y in the table above. Round only the final limits to the hundredths place and you may assume all inference assumptions and conditions are met.

A: (8.25, 17.75) B: (3.33, 7.67) C: (7.52, 18.48) D: (3.62, 7.38) E: NOTA

13. Let's assume (at least for the moment) that data set X and data set Y were each independently drawn from two independent populations and we wish to perform the appropriate statistical test of the following hypotheses about the two populations' means based on these two independent samples: $H_0: \mu_X = \mu_Y \ vs. \ H_A: \mu_X \neq \mu_Y$. Compute the sum of the absolute value of the test statistic, the p-value, and the "exact" degrees of freedom provided by your calculator rounding only the final answer to the nearest integer. What is the sum of the digits of the integer you obtain? You may assume all inference assumptions and conditions are met.

A: 3 B: 4 C: 5 D: 6 E: NOTA

14. Which of the following is the best description of the most likely relationship in the population between the variables Y and X from which this set of bivariate data was drawn based upon your analysis of these data thus far?

A: Variable Y is a linear transformation of variable X.
B: Variable Y is a logarithmic transformation of variable X.
C: There is no association of any kind between variable Y and variable X.
D: Variable Y is some form of a non-linear transformation of variable X other than a logarithmic one.
E: NOTA

15. Which of the following statements is / are true?

I. If two variables are uncorrelated, then they are independent.
II. If two variables are independent, then they are uncorrelated.

A: Only Statement I is true.
B: Only Statement II is true.
C: Neither Statement I nor Statement II is true because they contradict each other.
D: Both Statement I and Statement II are true because they are essentially equivalent statements.
E: NOTA

16. Which of the following is the minimum sample size required to estimate an unknown population proportion with 95% confidence and a margin of error of 3.1% when no prior estimate for the proportion is known?

A: 999 B: 1000 C: 1067 D: 1068 E: NOTA

17. Which of the following is the minimum sample size required to estimate the unknown mean of a normally distributed population using a 95% confidence interval with a margin of error of at most 2 under the assumption that the population variance is $\sigma^2 = 16$?

A: 246 B: 245 C: 16 D: 15 E: NOTA

18. Suppose the mean alcohol content of beer across all individual brands and all individual varieties is normally distributed with a mean of $\mu_A = 5\%$ and a standard deviation of $\sigma_A = 0.75\%$. What is the probability that the mean alcohol content of an SRS of 25 independent individual beers is between 4.8% and 5.4%? Round the final answer to the thousandths place.

A: 0.905 B: 0.308 C: 0.095 D: 0.692 E: NOTA

19. Suppose that a statistical hypothesis test is conducted at the 5% level of significance with a power level of 0.95 and the resulting p-value is 0.0783. After the completion of the test, what is the probability that this particular test resulted in both a Type-I Error and a Type-II Error at the same time?

A: 0 B: 0.0783 C: 0.10 D: 0.1783 E: NOTA

20. Whenever the design of a study systematically favors a particular result, we say that the study's results are…

A: skewed. B: biased. C: confounded. D: controlled. E: NOTA

Use the following information to answer the final ten questions:
A casino offers a cash prize for flipping a fair coin n times in either one of two games: Game H or Game T. To win Game H, players must get exactly 50% heads in the n coin flips (and lose otherwise). To win Game T, players must get strictly between 44% and 56% tails in the n coin flips. Thus, players lose Game T if they get exactly 44% tails or exactly 56% tails in the n flips or any other proportion outside of 44% to 56% tails in the n flips. The game played is chosen randomly by an initial flip of the same fair coin with heads indicating a play of Game H and tails indicating a play of Game T. Then, players have the option to decide whether to toss the coin n = 20 subsequent times or n = 50 subsequent times in the randomly assigned game. Otherwise, they have the option to flip the coin again and if it turns up heads, they must use n = 20 subsequent flips and if it turns up tails, they must use n = 50 subsequent flips for the randomly assigned game by the initial flip of the coin.

21. How many coin flips should a player choose in each game in order to have the best chance to win the prize?

A: 20 flips for Game H and 50 flips for Game T.
B: 50 flips for Game H and 20 flips for Game T.
C: 20 flips for either game.
D: 50 flips for either game.
E: NOTA

22. The payoff odds for each game described above are 2:1. This means that for every dollar bet, the player wins double the bet and so the net profit on a $1 bet is $2 and the player loses the $1 when he loses. Suppose some player has absolutely no idea whether it is better to choose n = 20 flips or n = 50 flips for either game. So, he chooses the option to flip the coin a second time to determine whether to use n = 20 subsequent flips or n = 50 subsequent flips in the randomly assigned game based upon the first flip of the coin. What is the expected value of the player's net winnings by using this strategy if he bets $100 per play? Round all steps to six decimal places (if necessary) and express the final answer in dollars and cents rounded to the nearest whole cent.

A: -$2.11 B: -$4.45 C: -$4.23 D: -$0.02 E: NOTA

23. Now, suppose instead that some player absolutely knows which value of n is better to choose (20 or 50) for the subsequent flips of the fair coin after the initial flip determines the game to play. What is the expected value of the net winnings for the player on a $100 bet if the payoff odds are still 2:1? Again, round all steps to six decimal places (if necessary) and express the final answer in dollars and cents rounded to the nearest whole cent.

A: $2.11 B: $4.45 C: $4.23 D: $0.02 E: NOTA

24. The casino offering these games is hoping that well over half the people who play the games have absolutely no clue as to which value of n to choose for the number of flips of the coin in their randomly assigned game (and so they automatically choose the option of the second coin flip for determining n). Now, suppose that in fact exactly 75% of the people who play the games have absolutely no clue which value of n is better for which game and suppose that exactly 25% of those who play the games know exactly which value of n is optimal for their randomly assigned game. If these percentages hold true in the population of all independent players who play these games, then the Casino is expected to do which of the following in the long run under this scenario?

A: Break even.
B: Suffer a loss.
C: Earn a profit.
D: It is impossible to know.
E: NOTA

25. Again, suppose that in fact exactly 75% of the people who play the games have absolutely no clue which value of n is better for which game and so they choose the option of letting another flip of the coin decide and again suppose that exactly 25% of those who play the games know exactly which value of n is optimal for their randomly assigned game. If these percentages hold true in the population of all independent players who play these games, then what is the probability that a randomly selected player of the games knows which value of n to choose for their randomly assigned game given that the player won? Round all steps to six decimal places (if necessary) and the final answer to four decimal places.

A: 0.2521 B: 0.3263 C: 0.5162 D: 0.2624 E: NOTA

26. A player is about to play one of the games after already having determined both the game and the number of flips of the fair coin for the assigned game. Let random variable X represent the number of heads in the n flips of the coin for the already determined game. What is the expected value of random variable X divided by the variance of random variable X? That is, what is $\frac{E(X)}{Var(X)}$?

A: 1 B: 2 C: 3 D: 4 E: NOTA

Use the following additional information to answer the remaining questions:
The casino expects a particular approximate breakdown of winners and losers for the games per every 100 players under the assumption that 75% of players do not know which value of n is better for which game and 25% do know. The table below displays their expected breakdown (on the left) and the actual observed breakdown (on the right) in an SRS of 100 actual independent players who recently played the games. Do these data provide evidence that the actual breakdown of winners and losers on the games in the population does not fit what the casino expects? We shall investigate and you may assume all inference assumptions and conditions are met.

KEY: PDK = Player Doesn't Know which value of n is better for which game.
PK = Player Knows which value of n is better for which game.

Expected Counts	PDK	PK	Total		Observed Counts	PDK	PK	Total
Win Either Game	24	9	33		Win Either Game	27	18	45
Lose Either Game	51	16	67		Lose Either Game	43	12	55
Total	75	25	100		Total	70	30	100

27. Perform the appropriate statistical test on the data in the table below in order to determine if the observed counts provide evidence that the population distribution of winners and losers on the two games (collectively) does not fit the distribution that the casino expects. What is the sum of the test statistic, p-value, and degrees of freedom of the test rounded to four decimal places? You may assume all inference conditions are met.

	Win and PDK	Lose and PDK	Win and PK	Lose and PK
Observed Counts	27	43	18	12
Expected Counts	24	51	9	16

A: 4.9445 B: 14.6387 C: 7.9445 D: 15.6387 E: NOTA

28. Which of the following is the correct conclusion that the casino can draw from the results of the appropriate statistical test in the previous question using a 1% level of significance?

A: The observed counts do provide statistically significant evidence that the population distribution of winners and losers on the two games (collectively) does not fit the distribution that the casino expects.
B: The observed counts do not provide statistically significant evidence that the population distribution of winners and losers on the two games (collectively) does not fit the distribution that the casino expects.
C: The observed counts do provide statistically significant evidence that the population distribution of winners and losers on the two games (collectively) does fit the distribution that the casino expects.
D: The observed counts do not provide statistically significant evidence that the population distribution of winners and losers on the two games (collectively) does fit the distribution that the casino expects.
E: NOTA

29. Are the outcomes of the games (winning vs. losing) and whether or not the player knows which value of n is better for which game independent of each other? Well, the table of the casino's expected counts essentially indicates that they believe they are, or at least that they are sufficiently close to being independent for their purposes. However, the table of observed counts seems to show otherwise. Perform the appropriate statistical test on the table of observed counts to determine if they provide evidence that the outcomes of the games (the player wins or the player loses) and whether or not the player knows which value of n is better for which game are independent of each other at the 5% level of significance. Which of the following is the correct conclusion? You may assume all inference assumptions and conditions are met.

A: The observed counts provide overwhelming evidence that the outcomes of the games and whether or not the player knows which value of n is better for which game are not independent of each other because the p-value of the test is much greater than 0.05.
B: The observed counts do not quite provide sufficient evidence that the outcomes of the games and whether or not the player knows which value of n is better for which game are not independent of each other because the p-value of the test is slightly greater than 0.05.
C: The observed counts provide almost no evidence that the outcomes of the games and whether or not the player knows which value of n is better for which game are not independent of each other because the p-value of the test is much greater than 0.05.
D: The observed counts provide statistically significant evidence that the outcomes of the games and whether or not the player knows which value of n is better for which game are not independent of each other because the p-value of the test is less than 0.05.
E: NOTA

30. Suppose the casino had deliberately selected a random sample of 75 players who do not know which value of n is better for which game and a random sample of 25 players who do know and combined these two samples in order to obtain their overall sample of 100 players. Then, instead of a simple random sample (or SRS), this type of a sample is known as a…

A: convenience sample.
B: stratified random sample.
C: cluster random sample.
D: systematic random sample.
E: NOTA