Important Instructions for this Test: Please pay close attention to and carefully follow all rounding instructions. Round any 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."

Please Use the Following Information to Help You Answer Questions 1 through 20 (inclusive):
The infamous Dr. Rampal is at it again! Now, he is desperately trying to curry his influence with the new Mu Alpha Theta (MAΘ) president, Señor Nieve, by convincing him that his secret sleep audio program can help MAΘ Calculus student competitors become better at Calculus! For those of you unfamiliar with the evil Dr. Rampal's claim, here it is verbatim: "All you have to do is listen to my program while you sleep at night, and you will get better at Calculus in two months. Does that make sense?" Naturally, since Señor Nieve is a faithful disciple of Stats G-Sus, he requires statistically significant evidence that Dr. Rampal's Calculus program works before he allows it to be marketed and sold to MAΘ Calculus competition coaches so that they can apply it to their student competitors in hopes of becoming more competitive at Calculus. If this happens, Señor Nieve and Dr. Rampal both stand to make a boat-load of money; of which they will donate half to the Florida MAΘ organization. What great guys they are!

Here are the data from an SRS of 16 independent MAΘ Calculus competitors who were all given the same MAΘ-style Calculus pretest (Pretest Score) and then required to listen to Dr. Rampal's secret sleep audio program for two months after which they were all given the same parallel-form MAΘ-style Calculus posttest (Posttest Score). You may reasonably assume that all requisite assumptions and conditions for any inference procedures you are about to actually perform upon this data set are met.

Student Competitor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Pretest Score	77	73	69	68	66	62	62	61	61	55	55	54	51	50	48	48
Posttest Score	89	84	83	80	79	77	76	76	75	72	71	70	69	67	67	65
Difference (Posttest – Pretest)	12	11	14	12	13	15	14	15	14	17	16	16	18	17	19	17

1. What is the sample mean difference of the set of differences in score on these two Calculus tests (Posttest Score – Pretest Score) for this sample of 16 MAΘ Calculus student competitors?

A: 60 B: 75 C: 15 D: –15 E: NOTA

2. What is the sample variance of the set of differences in score on these two Calculus tests (Posttest Score – Pretest Score) for this sample of 16 MAΘ Calculus student competitors as an exact value?

A: 5 B: $\frac{16}{3}$ C: $\sqrt{5}$ D: $\frac{4\sqrt{3}}{3}$ E: NOTA

3. What is the sample correlation coefficient between the set of Pretest Scores and the paired set of Posttest Scores for this sample of 16 MAΘ Calculus students rounded to the thousandths place?

A: 0.993 B: 0.986 C: 0.758 D: 0.831 E: NOTA

4. Which of the following graphical displays of data is useful for determining whether or not a linear relationship exists between the set of Pretest Scores (treated as the explanatory variable) and the paired set of Posttest Scores (treated as the response variable) based upon this sample of 16 MAΘ Calculus students?

A: A two-way table of the Posttest Scores versus the Pretest Scores.
B: A Scatterplot of the Posttest Scores versus the Pretest Scores.
C: A residual plot of the residuals calculated from the differences between the observed Posttest Scores and the predicted Posttest scores using the least-squares linear regression model equation calculated from the data
D: Either B or C.
E: NOTA

Here are some additional instructions to help you answer the next eight questions:
Perform a full least-squares linear regression analysis of the data in the table on the previous page treating the Pretest Scores as the explanatory variable and the Posttest Scores as the response variable. Recall that you may assume all requisite assumptions and conditions are met.

5. Which of the following is the correct equation of the least-squares linear regression model built from this data with the coefficients rounded to four decimal places?

A: $\widehat{Posttest\ Score} = 0.7583 + 29.5017(Pretest\ Score)$
B: $\widehat{Posttest\ Score} = 29.5017 + 0.7583(Pretest\ Score)$
C: $\widehat{Posttest\ Score} = 29.5017 + 0.8310(Pretest\ Score)$
D: $\widehat{Posttet\ Score} = 0.7583 + 0.8310(Pretest\ Score)$
E: NOTA

6. Compute the set of 16 residuals for this least-squares regression analysis using the correct linear regression model equation from the previous question. HINT: This is fairly easily done if you know how to make a residual plot or use the lists in your calculator! What is the ratio of positive residuals to negative residuals for this regression analysis?

A: 8:8 B: 1:1 C: 9:7 D: 7:9 E: NOTA

7. One of the linear regression inference conditions that is required is known as "homoskedasticity" whereby the variance (and equivalently the standard deviation) of the response variable remains constant across all given values of the explanatory variable. According to the least-squared regression analysis you just performed upon this data, what is the estimated value of the variance of the response variable (*Posttest Score*) for any given value of the explanatory variable (*Pretest Score*)? Round only your FINAL answer to four decimal places. HINT: Make sure to take into account the correct degrees of freedom in your calculation of this value!

A: 0.9331 B: 0.9862 C: 0.8310 D: 0.6905 E: NOTA

8. Speaking of degrees of freedom, what is the sum of the test statistic, p-value, and the correct degrees of freedom for a linear regression t-test of $H_0: \beta = 0 \; vs. H_A: \beta > 0$ where β represents the true population regression line slope for the linear relationship between all Pretest Scores (the explanatory variable) and all Posttest Scores (the response variable) based upon this data? Round only your FINAL answer to four decimal places.

A: 45.6648 B: 44.6648 C: 46.6648 D: 43.6648 E: NOTA

9. Which of the following is a 98% confidence interval for β, the true population regression line slope for the linear relationship between all Pretest Scores (the explanatory variable) and all Posttest Scores (the response variable) based upon this data? Round only the FINAL confidence interval limits to four decimal places.

A: $0.7069 < \beta < 0.8097$
B: $0.7041 < \beta < 0.8125$
C: $0.6955 < \beta < 0.8212$
D: $0.6870 < \beta < 0.8296$
E: NOTA

10. Based upon the correct results from the previous two questions, which of the following claims about β is NOT REJECTED (where β is defined as it is in the previous two questions)?

A: $\beta = 0$ B: $\beta < 0$ C: $\beta = 0.65$ D: $\beta = 0.75$ E: NOTA

11. Recall the properties of the sampling distribution of the sample least-squares regression line slope, b:

- The mean of the sampling distribution of b is $\mu_b = \beta$.
- The standard deviation of the sampling distribution of b is given by $\sigma_b = \frac{\sigma_e}{\sigma_x \sqrt{n}}$.
- The shape of the sampling distribution of b is approximately Normal.

In the formula for σ_b above, σ_e represents the constant standard deviation about the population regression line for the response variable, Y, across all given values of the explanatory variable, X; and σ_x represents the population standard deviation of the explanatory variable, X. As usual, n is the sample size. Suppose that for the scenario we have discussed thus far throughout this test, we have a population regression line slope of $\beta = 0.76$, a standard deviation about the population regression line for the response variable of $\sigma_e = 0.81$, a population standard deviation for the explanatory variable of $\sigma_x = 9$, and a random sample of size $n = 16$ from our population of interest. Compute the following probability rounding only your FINAL answer to four decimal places: $P(0.75 < b < 0.80)$.

A: 0.6993 B: 0.6339 C: 0.9336 D: 0.6393 E: NOTA

12. The fact that the mean of the sampling distribution of b is $\mu_b = \beta$ (or equivalently that the expected value of the sample slope is equal to the population slope: $E(b) = \beta$) is what it means for the sample regression line slope to be a(an) ______ estimator of the population regression line slope.

A: resistant B: unbiased C: robust D: maximum likelihood E: NOTA

Use the following additional information to help you answer the next eight questions:
Let us now turn our attention to the matched-pairs analysis of the set of differences in *Posttest Scores* minus the corresponding paired *Pretest Scores*. Let us define μ_G as the population mean gain in score on the parallel form Calculus tests for all MAΘ Calculus student competitors who would use Dr. Rampal's secret sleep audio program. Señor Nieve considers a population mean gain in score, μ_G, of over 12 points to be of practical value since it represents a student getting at least three more questions correct on any given MAΘ Calculus test. Again, perform the following analyses at the 1% level of significance.

13. Perform a matched-pairs t-test of $H_0: \mu_G = 12$ vs. $H_A: \mu_G > 12$ at the 1% (0.01) level of significance using the set of gains in score: $Difference = Posttest\ Score - Pretest\ Score$. Compute the sum of the test statistic, p-value, and the degrees of freedom of the test rounded to three decimal places. Then, add the level of significance to this rounded sum if the null hypothesis is rejected or subtract the level of significance from this rounded sum if the null hypothesis is not rejected. Make sure to treat the level of significance as a decimal, 0.01, and NOT as an integer percentage for this calculation.

A: 21.196 B: 20.196 C: 20.206 D: 21.206 E: NOTA

14. Let us now follow up the matched-pairs hypothesis test from the previous question with a 98% confidence interval for μ_G; which is a natural thing to do, and we will, but in the next question. For now, please select the reason why confidence intervals, generally speaking, are better than hypothesis tests when performing statistical inference when all the requisite conditions for either procedure are met.

A: A confidence interval is more likely to result in correctly rejecting the null hypothesis when it is false.
B: A hypothesis test will result in either a Type-1 error and / or Type-2 error far more often than a confidence interval will.
C: A confidence interval gives us a plausible range of values for what the parameter of interest IS equal to and well as what values it is plausibly NOT equal to.
D: A hypothesis test is less likely to support the alternative hypothesis when it is true.
E: NOTA

15. Which of the following is the correct 98% confidence interval for μ_G (as it is defined above) with the confidence interval limits rounded to the tenths place?

A: $13.3 < \mu_G < 16.7$
B: $13.8 < \mu_G < 16.2$
C: $13.7 < \mu_G < 16.3$
D: $13.5 < \mu_G < 16.5$
E: NOTA

16. According to the data in the table on page 1 of this test, which of the following is the best unbiased point estimate for the population proportion of all MAΘ Calculus student competitors who would gain over 12 points in their score if they used Dr. Rampal's secret sleep audio program for two months?

A: 0.8125 B: 0.75 C: 0.9375 D: 0.875 E: NOTA

17. The previous question hints at the possibility of constructing a one-proportion Z-interval for p, the true population proportion of all MAΘ Calculus student competitors who would gain over 12 points in their score if they used Dr. Rampal's secret sleep audio program for two months. However, constructing such a confidence interval is not permissible using this data because one (or more) of the following requisite conditions for constructing just such an interval is (are) violated. Which one (or more) is it?

A: The sample of data is based upon an SRS from the population of interest.
B: The data observations are independent of each other.
C: The sample size of the data is sufficiently large so that the normal approximation to the binomial distribution can be used to construct the interval. This is often referred to as the "large counts condition" or the "sufficient sample size condition" by most textbooks.
D: At least two of the above conditions are violated.
E: NOTA

18. Based upon the correct results from questions 13 and 15, is there convincing evidence that the population mean gain in score on an MAΘ Calculus test for all MAΘ Calculus student competitors who would use Dr. Rampal's secret sleep audio program for two months is greater than 12; and hence, should Señor Nieve market and sell it to MAΘ Calculus competition coaches so that they can apply it to their student competitors in hopes of becoming more competitive at Calculus?
NOTE: There are two "yes or no" questions posed here and the answer choices below are in the form of an ordered pair of (the answer to the first question, the answer to the second question).

A: (yes, yes) B: (yes, no) C: (no, yes) D: (no, no) E: NOTA

19. Suppose that the population mean gain in score for all MAΘ Calculus student competitors who would use Dr. Rampal's secret sleep audio program for two months is actually $\mu_G = 14$ and that the population standard deviation of the gain in score is $\sigma_G = 2.4$. What is the <u>power</u> of the test of the hypotheses H_0: $\mu_G = 12$ vs. H_A: $\mu_G > 12$ at the 1% level of significance for an SRS of $n = 16$ from the population of interest? NOTE: Since you are given $\sigma_G = 2.4$, you may use "z" instead of "t" for this calculation! Round only your FINAL answer to the <u>thousandths place</u> and you may assume all conditions are met.

A: 0.068 B: 0.843 C: 0.996 D: 0.816 E: NOTA

20. Speaking of power, the one-sided hypothesis test performed in question 13 used a significance level of 0.01. Suppose that the null hypothesis was truly false. If a significance level of 0.05 had been used instead of a significance level of 0.01, which one of the following would have happened?

A: The probability of a Type II error would increase and the power of the test would decrease.
B: Both the probability of a Type II error and the power of the test would both decrease.
C: Both the probability of a Type II error and the power of the test would both increase.
D: The probability of a Type II error would decrease and the power of the test would increase.
E: NOTA

The remainder of the questions on this test have ABSOLUTELY NOTHING to do with the previous 20 questions! THANK GOODNESS, EH? 😊

21. The Vera and Alsheikh polling firm is interested in surveying a representative random sample of 1000 registered voters in the United States to assess their opinion on the President's job performance. The firm has automated its sampling so that random phone numbers within the United States are called (including landlines, cell phones, and unregistered numbers). Each time a random number is called, the procedure below is followed.

• If there is no response or if a voicemail box is reached, another number is automatically called.
• If a person answers, a survey worker verifies that the person is at least 18 years of age.
• If the person is not at least 18 years of age, no response is recorded, and another number is called.
• If the person is at least 18 years of age, that person is surveyed.

Some people claim the procedure being used does not permit the results to be extended to all registered voters and/or that it may lead to biased results. All of the following are potentially legitimate concerns with the sampling methodology being employed by the polling firm; however; which of the following is clearly THE MOST legitimate concern about the sampling procedure being used by the polling firm?

A: Registered voters with children under the age of 18 years may be underrepresented in the sample.
B: Registered voters who have more than one telephone number may be overrepresented in the sample.
C: Registered voters who live in households consisting of more than one voter may be underrepresented.
D: People who are not registered to vote would get included in the survey which would almost certainly bias the poll's results.
E: NOTA

22. Bernie, Jessie, and Spencer love to get deranged! So, they buy a brand-new standard deck of cards, unwrap the box and then remove the two Jokers. This deck now has the 52 cards arranged such that all 13 cards of each suit are sequentially arranged in the order of Ace, 2, 3, …10, Jack, Queen, King with all the cards of each suit together and the suits are arranged in the order of Spades, Hearts, Clubs, and Diamonds. Bernie randomly selects one of the four sets of 13 cards all in one suit and then randomly lays them down on the table before her from left to right. What is the probability that she selects the set of 13 Clubs and that none of the cards are in their original position as they were in the original brand-new deck with the Jokers removed? Thus, she has achieved what is known as a "derangement" of the 13 cards. Round only your FINAL answer to the thousandths place.

A: 0.029 B: 0.386 C: 0.368 D: 0.092 E: NOTA

23. A two-sample t-test of $H_0: \mu_1 = \mu_2\ vs. H_A: \mu_1 \neq \mu_2$ resulted in a test statistic of $t = 4.24$ and the exact degrees of freedom of the test statistic were $df = 75.31633842$. The means of the two samples were $\bar{X}_1 = 25\ and\ \bar{X}_2 = 23$ and the sample standard deviations were $s_{x_1} = 2.3\ and\ s_{x_2} = 1.9$. Both samples were the same size, n. What is the sum of the digits of n? You may assume all conditions are met.

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

The final seven questions on this test use the following set of ten INDEPENDENT random variables as they are defined below. A note on notation: μ_k denotes the mean of random variable X_k and σ_k^2 denotes the variance of random variable X_k, etc. Some of the means and/or variances are given to you since you are not expected to know how to calculate them, but you are expected to know how to calculate the rest.

X_1 is a discrete uniform distribution on the set of the 9 positive digits: 1, 2, 3, 4, 5, 6, 7, 8, and 9.
X_2 is a continuous uniform distribution on the closed interval $[0, 9]$ with a variance of $\sigma_2^2 = 6.75$.
X_3 is a binomial distribution with $n = 12$ trials and probability of success $p_3 = 0.75$.
X_4 is a geometric distribution with probability of success $p_4 = 0.25$.
X_5 is a standard Normal distribution.
X_6 is a standard Student's t-distribution with 12 degrees of freedom and a variance of $\sigma_6^2 = 1.2$.
X_7 is a chi-square distribution with 8 degrees of freedom, a mean of $\mu_7 = 8$, and a variance of $\sigma_7^2 = 16$.
X_8 is a gamma distribution with a mean of $\mu_8 = 2$ and a variance of $\sigma_8^2 = 4$.
X_9 is a beta distribution with a mean of $\mu_9 = 0.5$ and a variance of $\sigma_9^2 = 0.05$.
X_{10} is some unknown continuous distribution with a mean of $\mu_{10} = 10$ and a variance of $\sigma_{10}^2 = 100$.

24. What is $P(X_1 \text{ is a prime})$ as an exact value?

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

25. What is $P(3 \le X_2 < 6)$ as an exact value?

A: $\frac{2}{9}$ B: $\frac{1}{3}$ C: $\frac{4}{9}$ D: $\frac{1}{2}$ E: NOTA

26. What is $P(\mu_3 - 2\sigma_3 \le X_3 < \mu_3 + 2\sigma_3)$ rounded to three decimal places?

A: 0.914 B: 0.946 C: 0.954 D: 0.986 E: NOTA

27. What is $P\left[(X_1 = \mu_1) \cup \left(X_4 = \frac{\sigma_4^2}{\mu_4}\right)\right]$ as an exact value?

A: $\frac{17}{72}$ B: $\frac{146}{576}$ C: $\frac{77}{288}$ D: $\frac{1}{64}$ E: NOTA

28. Compute the following probability rounding only your FINAL answer the ten-thousandths place:
$P[(\mu_5 - 2\sigma_5 \le X_5 < \mu_5 + 2\sigma_5) \cap (\mu_6 - 2\sigma_6 \le X_6 < \mu_6 + 2\sigma_6) \cap (\mu_7 - 2\sigma_7 \le X_7 < \mu_7 + 2\sigma_7)]$.

A: 0.8693 B: 0.8963 C: 0.8369 D: 0.8639 E: NOTA

29. Let's define random variable Y as $Y = \sum_{k=1}^{10} X_k$ and let $\bar{Y}$ represent the mean of an SRS of $n = 100$ observations from random variable Y. Compute $P(41.1 \le \bar{Y} \le 44.4)$ rounding only your FINAL answer to the ten-thousandths place.

A: 0.8123 B: 0.8321 C: 0.8132 D: 0.8231 E: NOTA

30. Compute the following probability as an exact value: $P(X_k = \mu_k \text{ for all } k)$; or, written another way:
$P[(X_1 = \mu_1) \cap (X_2 = \mu_2) \cap (X_3 = \mu_3) \cap (X_4 = \mu_4) \cap (X_5 = \mu_5) \cap (X_6 = \mu_6) \cap (X_7 = \mu_7) \cap (X_8 = \mu_8) \cap (X_9 = \mu_9) \cap (X_{10} = \mu_{10})]$.

A: 0.25 B: 0.5 C: 0.75 D: 1 E: NOTA