Homework

1. Which of the following is not a statistic used in comparing 2 binomial proportions?

|  |  |
| --- | --- |
|  | Risk Difference |
|  | Relative Risk |
|  | Odds Ratio |
|  | Welch-Satterthwaite degrees of freedom  2. Suppose a 95% confidence interval for the proportion of US adults who are obese ranges from 25% to 29%.  Which one of the following statements is true?   |  |  | | --- | --- | |  | It is reasonable to conclude that more than 29% of US adults are currently obese. | |  | It is reasonable to conclude that fewer than 25% of US adults are currently obese. | |  | The hypothesis that 27% of US adults are currently obese can be rejected at the alpha=0.05 level. | |  | We can be 95% confident that the interval in question covers the true proportion of US adults who are obese. |   3. A 2-sample t-test was performed, and the resulting t-statistic was 2.6 with 62.6 degrees of freedom.  Based on this information alone, we can conclude:   |  |  | | --- | --- | |  | The t-test assumed equal variances. | |  | The t-test assumed unequal variances. | |  | A pooled standard deviation was used in the calculation of the t-statistic. | |  | The null hypothesis was not rejected.  4. In order to test whether or not the variance of a continuous random variable is the same in 2 groups, the best approach is:   |  |  | | --- | --- | |  | An F-test | |  | A 2-sample T-test | |  | A paired T-test | |  | Wilcoxon rank sum test |   5. A researcher wants to design a study comparing reaction times in 2 groups of mice.  Group 1 receives a standard diet the day before the reaction time testing, and Group 2 receives a diet that includes caffeine.  The researcher can afford to include n=6 mice per group (12 mice total).  In these mice, reaction times are known to be right skewed.  Before any data are collected, what type of statistical hypothesis test would you propose to use to compare the reaction times between the 2 groups?   |  |  | | --- | --- | |  | 2-sample t-test | |  | Paired t-test | |  | Wilcoxon rank sum test | |  | Wilcoxon signed rank test |   6. Suppose you construct a 95% confidence interval around a difference in means between 2 independent groups.  Also, suppose you also conduct a 2-sample t-test to compare the group means and obtain a p-value. Which of the following statements bests describes the relationship the confidence interval and the p-value?   |  |  | | --- | --- | |  | If the 95% confidence interval for the difference in means overlaps 0, then the p-value must be less than 0.05. | |  | If the 95% confidence interval for the difference in means overlaps 0, then the p-value must be greater than 0.05. | |  | If the 95% confidence interval for the difference in means does not overlap 0, then the p-value must be greater than 0.05. | |  | There is no inherent relationship between this type of confidence interval and the p-value. |   7. An investigator collects data on a group of 1000 surgeries to investigate whether the surgery type is associated with patients’ outcomes.  There is equal representation in each of 4 unique types of surgery, and patient outcome is categorized into 3 groups (discharged home, discharged to nursing home, and in-hospital death).  The investigator proposed to use a chi-square test to assess this association.  How many degrees of freedom would be required for such a chi-square test?   |  |  | | --- | --- | |  | 1 | |  | 3 | |  | 4 | |  | 6 | |  | 12 |   8. Suppose you conduct a 1-sample t-test, and you observe values for the sample mean and sample standard deviation with a sample size of n=50 that leads to the rejection of H0: µ = 0.  You calculate a p-value of 0.043.  What would happen to the p-value (for the same type of hypothesis test) if you observed the same sample mean and sample standard deviation, but in a sample that includes data from n=75 subjects?   |  |  | | --- | --- | |  | The p-value would increase. | |  | The p-value would decrease. | |  | The p-value would not change. | |  | The p-value may either increase or decrease. |   9. The table below includes sample statistics for the level of a particular toxic agent for a group of nonsmokers (1) and a group of current cigarette smokers (2).  Sample means and standard deviations are provided.  The study team is interested in knowing whether the mean levels are different between the 2 groups.  Define µ1 and µ2 as the true mean levels in nonsmokers and smokers, respectively, and σ21 and σ22 as the true variances in nonsmokers and smokers, respectively.   |  |  |  |  | | --- | --- | --- | --- | | Group | N | Sample  Mean | Sample  Standard Deviation | | 1)      Nonsmokers | 50 | 26 | 16 | | 2)      Smokers | 25 | 30 | 40 |   Which of the following represents a set of hypotheses for a hypothesis test that reflects the research question?   |  |  | | --- | --- | |  | H0: µ1 ≤ µ2; HA: µ1 > µ2 | |  | H0: σ21 = σ22; HA: σ21 ≠ σ22 | |  | H0: µ1 = µ2; HA: µ1 ≠ µ2 | |  | H0: µ1 < µ2; HA: µ1 ≥ µ2 |   10. A scientist collects behavioral outcomes on male and female gerbils.  The table below summarizes the data.  Using these data, estimate the odds ratio quantifying the association between sex and behavioral outcome.   |  |  |  |  | | --- | --- | --- | --- | |  | Behavioral Outcome | |  | |  | Positive Response | Negative Response | Total | | Males | 6 | 14 | 20 | | Females | 11 | 9 | 20 | | Total | 17 | 23 | 40 | |  | 0.35 | | | |  | 0.55 | | | |  | 0.85 | | | |  | 1.27 | | |   11. Researchers performed a study in which they compared subjects who had an event (n=150) to subjects who didn't have an event (n=250).  The age distributions are displayed in the boxplots below.  You can assume the sample standard deviations are similar to each other.  The researchers would like to compare the 2 groups to determine whether the average ages are different.  What type of statistical hypothesis test would you propose to use in this situation? (Assume that the ages are relatively normally distributed within each group.)   |  |  | | --- | --- | |  | 1-sample t-test | |  | Paired t-test | |  | 2 sample t-test with the assumption of equal variances | |  | 2 sample t-test with the assumption of unequal variances | |  | Exact test for the difference in binomial proportions  12. Which of the following statements are true about confidence intervals?   |  |  | | --- | --- | |  | A 99% confidence interval for a given mean will tend to be narrower than the corresponding 95% confidence interval. | |  | A 95% confidence interval is always preferred over a 99% confidence interval. | |  | The 95% confidence interval for a mean does not depend on the standard error of the sample mean. | |  | A 95% confidence interval for a given mean based on a sample size of n=50 will tend to be wider than the corresponding 95% confidence interval based on a sample size of n=100. |   13. A medical educator would like to test whether students’ knowledge of the cardiovascular system improves during a virtual class.  Each student takes a knowledge-based test before the class begins and then takes the identical test at the end of the class.  There are 600 students enrolled in the class.  Before any data are collected, what type of statistical hypothesis test would you propose to use to compare the test scores before and after the virtual class?   |  |  | | --- | --- | |  | 2-sample t-test | |  | Paired t-test | |  | Wilcoxon rank sum test | |  | Wilcoxon signed rank test |   14. The Wilcoxon rank sum test and the Wilcoxon signed rank test   |  |  | | --- | --- | |  | Cannot be used to analyze data that represent counts (e.g. number of clinic visits) | |  | Should only be used for small sample sizes | |  | Do not assume underlying normality of the data | |  | Are generally more powerful than their analogous t-tests  15. The table below includes sample statistics for the level of a particular toxic agent for a group of nonsmokers (1) and a group of current cigarette smokers (2).  Sample means and standard deviations are provided.   |  |  |  |  | | --- | --- | --- | --- | | Group | N | Sample  Mean | Sample  Standard Deviation | | 1)      Nonsmokers | 50 | 26 | 16 | | 2)      Smokers | 25 | 30 | 40 |   The standard error of the mean level of the toxic agent in smokers, rounded to the nearest hundredths, is:   |  |  | | --- | --- | |  | 2.26 | |  | 6.00 | |  | 6.50 | |  | 8.00 | |  | 16.00 | |  | Not able to be determined from the information provided |   16. Suppose you collect some data and find that a 95% confidence interval for a mean extends from 26.4 to 30.2.  Which of the following is true?   |  |  | | --- | --- | |  | There's a 95% probability that the sample mean lies between 26.4 and 30.2. | |  | There's a 95% chance that a randomly selected observation from the sample would lie between 26.4 and 30.2. | |  | If the experiment were repeated many times, there is a 95% probability that the sample mean would lie between 26.4 and 30.2. | |  | If the experiment were repeated many times, about 95% of such confidence intervals would include the true population mean. |   17. What does it mean for a result to be statistically significant?   |  |  | | --- | --- | |  | The result proves that the study design was unbiased. | |  | The data from the study were within the margin of error. | |  | The study produced a result that is likely biased. | |  | The result is too extreme to be consistent with random chance alone.  18. A scientist comes to you for assistance with data analysis.  He wants to compare a variable called "response" between 2 independent groups, each of which contains n=250 subjects.  It's safe to assume that the one sample standard deviation is quite larger than the other.  The data are displayed in the box plots below.    The most appropriate statistical test to compare the mean response between groups is:   |  |  | | --- | --- | |  | Paired t-test | |  | 2-sample t-test assuming equal variance | |  | 2-sample t-test assuming unequal variance | |  | Wilcoxon signed rank test  19. According to the CDC, 16.7% of U.S. adults binge drinks on weekly basis.  Suppose a researcher at MUSC would like to conduct a study of MUSC students to determine whether they binge drink at a rate similar to that of U.S. adults.  Which of the following hypothesis tests would be most appropriate to use?   |  |  | | --- | --- | |  | 1-sample t-test | |  | 1-sided t-test | |  | 1-sample test for a binomial proportion | |  | 1-sided test for a binomial proportion |   20. A p-value represents:   |  |  | | --- | --- | |  | the probability that the null hypothesis is true given the observed data. | |  | the probability that the result is clinically significant assuming the null hypothesis is true. | |  | the probability of observing a result as or more extreme than what was seen in our data assuming the null hypothesis is true. | |  | the probability of observing a result as or more extreme than what was seen in our data assuming the alternative hypothesis is true. | |  | the ratio of the probability that the null hypothesis is true versus the alternative hypothesis is true.  21. In a study of n=40 people with cancer and n=160 people without cancer, 30% were found to have been exposed to a known carcinogen.  Assuming there is no association between exposure and disease, what is the expected number of people in this study to both have cancer and have been exposed to the carcinogen? In other words, if A is the set of people with cancer, and B is the set of people who were exposed to the carcinogen, what is the expected number of people in the set represented by the intersection of A and B?  Answer:  22. After conducting a Wilcoxon rank sum test to compare LDL cholesterol values between 2 independent groups of people, a researcher realized that there was one subject whose LDL value was recorded as 165, but that was a mistake.  It should have been recorded as 195. Then he corrected the data point and conducted the Wilcoxon rank sum test again.  Which of the following statements is true?   |  |  | | --- | --- | |  | If the sample median LDL was 165, the 2 p-values would be the same. | |  | If all other subjects in the study had LDL values that were less than 165, the 2 p-values would be the same. | |  | The p-value associated with the 2nd Wilcoxon test would be identical to the p-value associated with a 2-sample t-test. | |  | The p-value associated with the 2nd Wilcoxon test is meaningless, because the researcher should have used a Wilcoxon signed rank test. |   23. Which of the following statements is true about the 2-sample t-test?   |  |  | | --- | --- | |  | The 2-sample t-test is excellent for comparing measurements obtained on individuals at 2 points in time. | |  | The 2-sample t-test requires that the variable being compared between groups is normally distributed. | |  | When determining whether or not a 2-sample t-test is appropriate for a given scenario, the sample size is irrelevant. | |  | In conducting a 2-sample t-test, a pooled sample standard deviation may be needed. |   24. Which of the following is a good reason for selecting a non-parametric statistical test in favor of a parametric test?   |  |  | | --- | --- | |  | Non-parametric tests tend to be less statistically powerful than parametric tests. | |  | Non-parametric tests do not assume that the data being analyzed come from a normal distribution. | |  | Non-parametric tests are typically not valid for small sample sizes, whereas parametric tests may be. | |  | Non-parametric tests tend to be more sensitive to outliers than parametric tests. |   25. Imagine a researcher sends out an e-mail survey to a 20% randomly selected sample from 3,000 employees of XYZ University about their use of fitness centers.  Assume that no one refused to participate in the survey.  Among the employees who respond to the survey, 70% say they use fitness centers weekly. A 95% confidence interval for weekly use of fitness centers in the study population is:   |  |  | | --- | --- | |  | (0.13, 0.15) | |  | (0.19, 0.21) | |  | (0.28, 0.32) | |  | (0.66, 0.74) | |  | (0.68, 0.72) | | | | | | | |