ST 314 Practice Final Exam

The following questions are meant to give you sense of what the midterm exam might look like. This should not be your only source of study material! See Canvas for suggestions of other resources you can use to prepare for the exam.

Use for the following 4 questions. For each scenario indicate the matching test by filling in the corresponding letter in the \_\_\_\_\_\_\_\_. Each test can only be used once.

|  |  |  |  |
| --- | --- | --- | --- |
| A. Two-sample t test | B. Single-factor ANOVA | C. One Sample t test | D. Simple linear Regression. |

1. Matching test letter: C  Jenny would like to know whether she has budgeted enough money per month for groceries while she is in college. She takes a representative random sample of 50 college students and asks them how much they spend a month on groceries per month. Jenny has budgeted $200. Which test can she use to test whether she has budgeted enough for groceries per month?
2. Matching test letter: D Research has found that there is a relationship between neck circumference and body fat percentage. A nationwide fitness franchise would like to be able to predict body fat percentage from neck circumference. Which type of analysis is appropriate for this scenario?
3. Matching test letter: B The lumen output is measure on 6 light bulbs each for 4 different brands of comparable bulbs. A contractor is interested in finding whether there is significant difference in the average lumen output for the different bulb brands. What type of procedure would you recommend?
4. Matching test letter: A  A student would like test whether the average rent per room is higher for dwellings that are closer to campus versus further away. They take a random sample of 40 dwellings within 1 mile of their University campus, and another random sample of 40 dwellings more than 1 mile from campus. For each sample they record the cost of rent per bedroom. Which type of analysis is appropriate for this scenario?
5. If the p-value for a hypothesis test is very small, there is convincing evidence to support the alternative hypothesis.
   * + 1. True
       2. False
6. Suppose you have two single factor ANOVA experiments with the same degrees of freedom. The resulting *F* statistics are:  
   Experiment 1 *F = 5.68*Experiment 2 *F = 20.15*Which statement is true in regards to comparing Experiment 1 and Experiment 2?
7. Procedure 1 has a smaller test statistic and therefore will result in stronger evidence in favor of the alternative hypothesis.
8. Procedure 2 has a larger test statistic and therefore will result in strong evidence in favor of the alternative hypothesis.
9. We should reject the null for both tests.
10. Impossible to know with this information.

Use the following for the next two questions. In an experiment to study automobile engine operating efficiency for five different brands of gasoline, mpg was measured over a controlled distance and speed for eight cars in each group. Let represent the average mpg for the *i*th brand.

1. Identify the null and alternative hypotheses for the single factor ANVOA F test used to test if there is a difference in average mpg between the five brands.
   * 1. vs. the average for a least one brand differs from the other four brands.
     2. vs. the average for a least one brand differs from the other four brands.
     3. vs.
     4. vs.
2. Assume mpg is normally distributed for all five gasoline brands. Identify the null distribution for the *F* test statistic in this hypothesis testing scenario.
   * 1. *t* distribution with 39 degrees of freedom
     2. Standard normal distribution
     3. *F* distribution with 4 and 35 degrees of freedom
     4. Left skewed distribution

**Use the following to answer the next 3 questions.** A lumber manufacturing company would like to compare the average weight capacity of two types of boards they currently manufacture. The sampled boards were obtained using a random mechanism.

Chart, box and whisker chart

Description automatically generatedTable

Description automatically generated

1. Describe the side-by-side box plot. Include a comparison of the two group, the center, shape, and spread of each.   
     
   Board B looks to have an overall higher weight, on average, in comparison to board A. The variability of B also looks to be greater for board B. Board A weights look to be somewhat symmetric, yet board B weights have a longer negative tail.

1. State the null and alternative hypotheses to test whether there is a difference between the two boards.   
     
   ­­
2. Using the R output provided, write a 4-part conclusion based on the software output for the hypothesis test.

There is moderately suggestive evidence the weights of board A and board b differ on average.  
The null hypothesis is rejected based on a significance level of 0.05 (t = -2.0875, df = 36, p-value = 0.0439). The 95% confidence interval estimates the weights of board A will be 0.76 lbs to 51.9 lbs less than board B, on average.

The best guess for the difference in average weight between the boards is 26 lbs, with board A weighing in less than board B.

**Use the following to answer the next 3 questions.** According to the National Institute on Alcohol Abuse and Alcoholism (NIAAA), and the National Institutes of Health (NIH), 41% of college students nationwide engage in “binge drinking” behavior, having 5 or more drinks in one occasion during the past two weeks. A college president wonders if the proportion of students enrolled at her college that binge drink is lower than the national proportion. In a commissioned study, 462 students are selected randomly from a list of all students enrolled at the college. Of these, 162 admitted to having engaged in binge drinking.

1. The college president is more interested in testing her suspicion that the proportion of students at her college that binge drink is lower than the national proportion of 0.41. Write the null and alternative hypotheses for this test.
2. Write an expression for the test statistic needed to perform this hypothesis test. You do not need to evaluate the expression.
3. Which of the following conclusions is reasonable, based on a p-value of 0.0047?
   * 1. There is little evidence to support a conclusion that the proportion of students at this particular college that binge drink is lower than the national proportion of 0.41.
     2. There is no evidence that the proportion of binge drinking students at this college is lower than the national proportion of 0.41.
     3. There is strong evidence that the proportion of students at this college that binge drink is lower than the national proportion of 0.41.
     4. Chart, scatter chart

        Description automatically generatedWe can’t reach any reasonable conclusion because the assumptions necessary for a significance test for a proportion are not met in this case.

**Use the following information for the next 2 questions.** Airlines have increasingly outsourced the maintenance of their planes to other companies. Flight delays are often due to maintenance problems. Critics are concerned that the maintenance may be done less carefully, such that outsourcing creates safety hazards and delays. The following is a simple linear regression analysis on data from 2005 and 2006 on the percent of outsourcing for 12 airlines and their respective percent of delayed flights. Does the data support the concerns of the critics?

Table

Description automatically generated

1. Assuming the conditions are met, what is the least squares regression equation for estimated percent of flights delayed given the percent of maintenance outsourced by the airline.
   * 1. 
     2. 
     3. 
     4. 
2. From the regression analysis, which statement is true about the relationship between the percent of outsourced maintenance and percent of flight delays?
   * 1. The critics were right. Outsourcing causes delays.
     2. The LSRL estimates that, on average, for every 1% increase in outsources maintenance, flight delays increase by 27.48%.
     3. There is a very strong correlation between percent of outsources maintenance from airlines and percent of flight delays.
     4. The LSRL estimates that for every 1% increase in outsources maintenance the average percent of flight delays is expected to decrease by 0.1636.
3. Which of the following statements is FALSE about the correlation coefficient, *r*?
   * 1. The correlation coefficient is a unitless number and must always lie between 0 and 1, inclusive.
     2. The correlation coefficient can only describe the relationship between two quantitative variables.
     3. If r =1, then there is a perfect positive association between x and y.
     4. The correlation coefficient is a unitless number and must always lie between –1 and 1, inclusive.