**Practice Final: First Part (Non-Analysis Question) (45 pts) Solution**

Note: Every final is different… this does not mean the questions on your final will look the same. The final exam is aimed at testing concepts and those can be tested from varying types and numbers of questions. With that said, this could have been a final exam in our class.

1. (2pts each) True / False. Simply highlight your answer.

* 1. T / F For a two independent sample t-test, If we decrease alpha the power increases.
  2. T / F The 2 Sample T-Test is resistant to outliers.
  3. T / F A Rank-Sum test is a good test to use with censored data.
  4. T / F For a fixed , the quantile will never be greater than the quantile , no matter the degrees of freedom.
  5. T / F For a two-sided, two sample t-tools pooled confidence interval for the difference in means (µ1-µ2), if both ends of the confidence interval are positive, that is evidence that µ1 > µ2
  6. T / F The MSE in an ANOVA is the pooled estimate of the variance for the assumed normal distributions of the variable for each group in the study.
  7. T / F We wish to test the equality of 3 group means, if the normality assumption is met, the ANOVA is robust to the standard deviation assumption as long as the sample sizes of the groups are the same.
  8. T / F In general, in an observational study the result cannot be generalized to the population the sample is taken from.

1. (5 pts) Define the Pvalue:

The probabily of observing a result by chance as extreme or more extreme than what was observed in the study under the assumption that that null hypothesis is true.

1. (7 pts) Describe the strategy of the 6-step hypothesis test.

The strategy is to first make an assumption about a parameter of interest (example: the mean).

Next, we gather evidence (data) to see if it is consistent with the assumption or not.

Next, we assess the probability of observing by random chance the result we saw our data if the assumption made in the first step is true. This probability is known as the “pvalue”.

Finally, we make a decision based on the pvalue. If what we observed is “rare” enough (less than alpha) then we will conclude that there was enough evidence to suggest the assumption is not correct.

1. For question 4, consider the output from the Barbie Bungee Below ☺

|  |  |
| --- | --- |
| **proc** **glm** data = Barbie;  model distance = bands / solution;  **run**; | **proc** **glm** data = Barbie;  class bands;  model distance = bands / solution;  **run**; |
| Table  Description automatically generated  Diagram  Description automatically generated | Graphical user interface, table  Description automatically generated  Chart  Description automatically generated |

1. (5 pts) We would like to conduct a lack of fit test with respect to the simple linear regression fit of the distance bungeed to the number of rubber bands. Please fill in the ANOVA table below and **provide a conclusion**.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F | P-value |
| Model (Extra) | 5 | 0.714283 | 0.1428566 | 0.2307684 | .9427 |
| Error(Reduced ) | 14 | 8.666667 | 0.6190476 |  |  |
| Total(Full Model  Separate Means) | 19 | 9.38095 |  |  |  |

Critical Value given to be 2.958

Conclusion:

There is not sufficient evidence to suggest the regression model has a lack of fit with respect to the equal means model since the test statistic (.2308) is considerably less than the critical value (2.958)… or the student could have said the pvalue was bigger than alpha (pvalue = .9427

1. (6 pts) Show the associated 6 step hypothesis test for the slope of the bands variable from the parameter estimate table in the simple linear regression fit above (the first column).

Critical Value: t = A picture containing text

Description automatically generated

t statistic = 40.68

pvalue < .0001

Reject Ho

There is overwhelming evidence to suggest that there is a positive linear relationship between the number of bands and the distance bungeed by the action figure. (pvalue < .0001)

1. (3pts) Find (show your work) and interpret a 95% confidence interval for the “bands” variable’s slope from b.

95% CI: 3.12 +/- 2.093 \* .0767 -> (2.96, 3.28)

We are 95% confident that for each one band increase the mean distance bungeed will increase between 2.96 and 3.28 inches.

1. (3 pts) With respect to the last problem, let’s say that Ver 3 and Ver 4 used video content while Ver 1 and Ver 2 did not. You would thus like to compare the average of the click rates of Version 3 and 4 with the Original but you are assuming the standard deviations to all be the same so you would like to include Ver 1 and Ver 2 in the estimation of the standard deviation. What are the contrast weights that would be used to test the hypothesis:

Assume the order of the groups is alphabetical: “Original Version”, “Ver 1”, “Ver 2”, “Ver 3”, “Ver 4”

Assume the data is contained in a dataset called ***WebsiteTest***, the groups are in a variable called ***WebsiteType*** and the click rates are in a variable called ***ClickRate***. Simply finish the code below to perform the desired contrast. Place your additional code in place of the green placeholders.

**proc** **glm** data = **WebsiteTest**;

class **WebsiteType**;

model **ClickRate = WebsiteType**;

lsmeans WebsiteType / pdiff;

estimate **“Original versus the Average of Ver2, Ver3 and Ver4” WebsiteType 3 0 -1 -1 -1 divisor = 3**

**run**;