**Version: K Statistician: Guy Barrows**

**Directions:**

You may have a pencil, a calculator, and your R output document produced from the **Lakesurvey.csv** data on your desk. All other materials should be fully stored out of sight and your computer should be turned off.

Do not write anything on your document of results except to add labels – e.g., “Output #1” or “Figure 1” – for referring to when answering the questions below. When you are finished with the exam, you should staple this sheet, your handwritten answers, and your document of R results together, in that order.

You should answer all questions below with as much information as necessary to fully answer the question. All answers should be completed by using and referring to specific R output. Some questions require further calculations for which you are allowed to use your calculator. You are not allowed to make any further calculations in R. Your answers should be legibly handwritten on the sheets of paper provided, clearly labeled with the question number, and, *when marked by an asterisk*, written with complete sentences.

# **11 Steps for any Significance Test**

1. **[1]** state the rejection criterion (),

2. **[2]** state the null and alternative hypotheses to be tested – define the parameter,

3. **[1]** determine which hypothesis test to use – thoroughly explain why,

4. **[1]** collect the data (address type of study and randomization),

5. **[2]** check all necessary assumptions – explain how you tested the validity,

6. **[1]** calculate the appropriate statistic(s),

7. **[2]** calculate the appropriate test statistic,

8. **[2]** calculate the p‑value,

9. **[1]** state rejection decision,

10. **[2]\*** summarize your findings in terms of the problem, and

11. **[2]\* If reject H0,** compute a **100(1-)%** *confidence region* for the parameter.

**Questions:**

1. **[3pts]** Identify what type of variable each of the following is: age, sumDays, and fishing.

2. **[5pts]\*** Perform a thorough EDA for sumDays.

3. **[2pts]\*** Perform a thorough EDA for educat.

4. **[5pts]\*** Perform a thorough EDA for the relationship between sumDays and age.

5a. **[2pts]\*** Interpret the slope of the linear regression that you performed.

5b. **[2pts]** Predict the sumDays for a respondent with a median age.

5c. **[2pts]** What proportion of total variability in sumDays is explained by knowing the age of the respondent?

6. **[15pts]** Test at the 5% level that the distribution of individuals into the time spent fishing categories differs between those whose property is and is not their primary residence

7. **[15 or 17 pts]** Test at the 5% level that the mean value of the house differs between those whose property is and is not their primary residence.

8. **[8 pts]\*** Describe the importance of statistics (as a field of study or a collection of methods). Among other things make sure you describe the two major goals of statistics, identify at least three major concepts or ideas of statistics, and identify how some of the methods that you have learned this semester illustrate or are related to why you think statistics is important.