**Statistician: Frank Wroblewski Version: K**

**Directions**

The final exam will consist of several questions related to the following topics – univariate EDA (quantitative & categorical), bivariate EDA (quantitative & categorical), linear regression, one-sample t-test, two-sample t-test, and chi-square. You will be asked to answer questions from results that you prepared prior to the exam using R.

The data that you will examine is introduced below, including actual questions that will be on the exam. You should use R to create output that can be used to answer each question. Your R input and output should be printed and brought to the exam to be used to answer the exam questions. The document that you bring to the exam must meet the following strict criteria:

* The document must be produced by you and you alone!! You may not ask anyone but me (including the tutors) for specific help on using R. If you have questions, I will be available in my office or via e-mail.
* The document can only contain R commands related to expressions, assignments, functions, or objects; R output; or R created graphics. You may not type or write ***any other*** material on the document (including labeling figures, tables, output, or sections). You may not type any “notes” (i.e., non-R-related expressions, assignments, functions or objects) as “R code.” You cannot use R comments. The document should contain no code that results in errors.
* The document must be produced by compiling your script in RStudio (as described in a FAQ on the class webpage).

Failure to follow all of these criteria will result in a 0 for the final exam!!

**Data Set – LakeSurvey**

Researchers at the University of Wisconsin surveyed property owners on lakes in Vilas County, WI. These data are in **LakeSurvey.csv**, with information about these data in **LakeSurvey\_meta.txt**. **Make sure to follow the code for “cleaning” these data given in the metadata**. Prepare results for each of the following items …

1. Univariate EDA for the age, sumDays, educat, and fishing variables.

2. Bivariate EDA for all pairs of the age, sumDays, and value variables. [*May use one graph and one table.]*

3. Bivariate EDA between the prmryRes and fishing variables.

4. Linear regression results (equation results and r2) for predicting sumDays from age.

5. Results for testing the following research hypotheses (use 5% level for each)

a. The mean sumDays is greater than 31 days.

b. The distribution of individuals into the time spent fishing categories differs between those whose property is and is not their primary residence.

c. The mean age differs between those whose property is and is not their primary residence.

d. The mean value of the house differs between those whose property is and is not their primary residence.

e. The mean year for obtaining the property for those whose property **is their primary residence** was before 1990.

1. Describe the importance of statistics (as a field of study or a collection of methods). Make sure you **DESCRIBE** the two major goals of statistics, at least three major concepts or ideas of statistics, and how methods that you learned this semester illustrate or are related to why you think statistics is important. [*This question is not specific to these data. An answer with <400 thoughtful words is probably inadequate.*]