**Statistician: James Weber Version: P**

**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 – Possum1**

Researcher’s in Australia collected a variety of morphological (body characteristics) data on the mountain brushtail possum (*Trichosurus caninus gilby*). The data are in **Possum1.csv** and information about the data are in the **Possum1\_meta.txt** file. **Make sure to follow the code for “cleaning” the data given in the metadata**. Possible questions asked from these data are …

1. Univariate EDA for total length, chest girth, foot length, and sex.

2. Bivariate EDA for each pair of total length, chest girth, and belly girth. [*May use one graph and one table.]*

3. Bivariate EDA for sex and location.

4. Linear regression results (equation results and r2) for predicting chest girth from total length.

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

a. The mean tail length of possums is greater than 36 cm.

b. The proportion of female possums differs between the two populations.

c. The mean chest girth is different for possums captured from Victoria than for possums captured from “other” populations.

d. The mean belly girth is less for male than female possums.

e. The mean age of **FEMALE** possums is greater than 3.

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.*]