

# STAT 408: Final Exam

Name:

Please turn in the exam to D2L. Please include the R Markdown code and a Word or PDF file with output. For any questions completed using SAS turn in the program summary that includes code and output and include the code in your R Markdown code (see examples in Q3 below for how to do this, but know the code will not be executed). You are welcome to turn in code and output for each question separately. Please verify that all of the code has compiled and the graphics look like you think they should on your Word or PDF file, you are welcome to upload image files directly if they look distorted in the Word or PDF file.

While the exam is open book, meaning you are free to use any resources from class, this is strictly an individual endeavor and **you should not discuss the problems with anyone outside the course instructor including group mates or class members.** The instructor will answer questions related to the data, expectations, and understanding of the exam, but will not fix broken code.

## 1. (22 points - Short Answer Questions)

For your responses, the target length should be a few sentences.

- a. (4 points) Describe the server and UI components of code to create a shiny app.
- b. (4 points) Discuss a scenario where it would be advantageous to create an R package. Then briefly detail the steps for creating an R package.
- c. (4 points) Define data scraping and discuss why a basic understanding HTML code is useful for data scraping.
- d. (4 points) Describe the functionality for SAS DATA steps and SAS PROC steps.
- e. (4 points) Compare and contrast writing code in R and SAS.
- f. (2 points) Do you prefer R or SAS, why? (There is no correct answer for this question.)

## 2. (14 points - Data Manipulation in R)

Note, the same questions are asked in SAS in question 3.

```
library(readr)
okcupid <- read_csv("http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/OKCupid_profiles_clean.csv")
```

- a. (2 points) How many profiles are contained in this dataset.
- b. (4 points) Compute the reported mean height for male and female profiles

c. (4 points) Compute the reported mean height for male and female profiles for those that self-reported a body type of athletic or fit.

d. (4 points) What is the most common age in the OkCupid profiles?

### 3. (14 points - Data Manipulation in SAS)

Note, the same questions are asked in R for question 2.

The code below will allow you to access the OKCUPID dataset in SAS.

```
LIBNAME STAT408 "/courses/d716b355ba27fe300";
```

```
DATA OKCUPID;  
    SET STAT408.OKCUPID;  
RUN;
```

SAS also has a procedure called PROC SQL that allows you to use SQL commands to create and modify data sets. The code below shows how the data set could be created with PROC SQL.

```
LIBNAME STAT408 "/courses/d716b355ba27fe300";
```

```
PROC SQL;  
    Create Table OKCUPID as  
    SELECT *  
    FROM STAT408.OKCUPID;  
RUN;
```

a. (2 points) How many profiles are contained in this dataset.

PROC SQL might make the remaining questions easier. Below is a solution using PROC SQL that you are welcome to use.

```
PROC SQL;  
    SELECT Count(*)  
    FROM OKCUPID;  
RUN;
```

b. (4 points) Compute the reported mean height for male and female profiles

c. (4 points) Compute the reported mean height for male and female profiles for those that self-reported a body type of athletic or fit.

d. (4 points) What is the most common age in the OkCupid profiles?

### 4. (20 points - Data Visualization)

Continuing with the OKCUPID dataset, we are now going to focus on data visualization. These graphs should be compelling and stand-alone with complete titles, labels, axes, and potentially annotations.

**a. (8 points)**

Create a set of boxplots in SAS that shows the relationship between body type and height graphics as a function of sex. You will likely need to consider using SGPanel <https://documentation.sas.com/?docsetId=grstatproc&docsetTarget=p07rpazajrf2wsn1nzd86x2nw6dv.htm&docsetVersion=9.4&locale=en> to do this.

**b. (8 points)**

Using R try to recreate (or make a better version) of the figure constructed with SAS.

**c. (4 points)**

Reflect on making graphics in R and SAS, which do you prefer? Why?