Name:\_\_Rahul Avadhoot\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last 4 Digits of Student ID Number:\_\_1787\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Introduction to R for Data Analysis in the Health Sciences**

**BIOST 509**

**In-Class Exercise 1**

**Due 5:00 PM (ideally by 3:30 PM, the end of class) on 9/27/2017**

Instructions

Enter your numerical and/or written answers to the questions using this file or create your own doc or pdf file with your answers. Submit your answers in a Word (.doc or .docx) or pdf file to the course canvas page by the end of the lab session today. Your work will be graded as 1 (complete) if at least 70% of the questions are answered correctly (and/or an obvious “good faith” effort), or 0 (incomplete) otherwise. For this in-class exercise, also include the R commands that you used to obtain your answers, when applicable.

Before you begin:

* Install RStudio from <http://www.rstudio.com/ide/download/> - use the standard installation.

Go to the course canvas page: and find the file folders for class lecture notes, datasets, in-class exercises, and homework assignments.: <https://canvas.uw.edu/courses/1115070/>

* Open In-Class Exercise 1

1. Ensure you can read in the “mammals.csv” datasets used in the lecture, and that you obtain the same data summaries seen in the slides. Datasets are available from canvas (mammals.csv in the datasets file folder). If you use the datasets from canvas, you need to download them to your computer before importing them to R. Datasets are also available on the web: <http://faculty.washington.edu/tathornt/Biost509/DataSets/mammals.csv>
2. Read in the “crab.txt” dataset using the import dataset button within the environment window of RStudio. Note that this file is a space delimited file, and not a comma separated file (csv). The data are available from canvas (crab.txt in the datasets files folder) or from web: <http://faculty.washington.edu/tathornt/Biost509/DataSets/crab.txt>.

This dataset describes 173 female horseshoe crabs from a study of nesting horseshoe crabs (J. Brockman, Ethology, 1996). It is a commonly used dataset for illustrating categorical data analysis; see also Agresti (2002), Sect 4.3.

Each female horeshoe crab in the study had a male crab attached to her chest. The study investigated whether the female crab had other males, called satellites, residing near her. Variables include: the female crab's color, spine condition, weight, width, and number of satellites.

* 1. What is the mean number of satellites among all 173 female crabs?

2.919

Command: summary(crab)

* 1. What is the mean number of satellites among female crabs whose width is <26cm?

1.899

Command: is.thin<-crab$width<26

summary(crab[is.thin,])

* 1. What is the mean number of satellites among female crabs whose width is >= 26 cm?

3.777

Command: is.thick<-crab$width>=26

summary(crab[is.thick,])

* 1. Tabulate the crabs color, for
     1. All the crabs

Command: table(crab[1])

dark dark medium light medium

22 44 12

medium

95

* + 1. Those with width < 26cm

Command: table(crab[is.thin,1])

dark dark medium light medium

15 23 3

medium

38

* + 1. Those with width >=26cm?

Command: table(crab[is.thick,1])

dark dark medium light medium

7 21 9

medium

57

1. The othercrab data set is also available in the data files folder on canvas, and from web: <http://faculty.washington.edu/tathornt/Biost509/DataSets/othercrab.csv>. The othercrab data set contains the same crab data but it is a comma-separated file, and has a .csv extension. Read in the othercrab dataset. Verify that you obtain the same answers to question #2.
2. Which biostatistics or statistics courses have you taken previously at the UW? (If you have only taken courses elsewhere briefly describe the level and/or content of the courses)

Data Science Methods for Clean Energy Research (Cheme 599)

Introduction to statistical methods, hypothesis testing and basic statistical analysis

1. Which biostatistics or statistics courses are you taking concurrently with this course?

None

1. Have you had any prior experience with any of the following statistical packages: R, SPSS, SAS, Stata, or Other? If yes, please list them.

Self taught R

Proficient in Python

1. What are your major goals for this course?

To learn valuable skills in R to use in my research and MS thesis

1. Are there any special R topics that you would like to see covered in this course? Please list them.

Regression analysis and other statistical methods