**Biost 517: Applied Biostatistics I**
**Biost 515: Biostatistics I**
Autumn 2019

**Homework #1**
Due: Monday, October 7, 2019 by 9:00 AM

**Written problems:** To be submitted as a pdf or MS-Word compatible file via the canvas course website.

*On this (as all homeworks) R code and unedited R output is **TOTALLY** unacceptable. Instead, prepare a table of statistics gleaned from the R output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*

This homework uses a subset of information that was collected for a study to examine magnetic resonance imaging (MRI) changes in the brain in a sample of generally healthy elderly subjects in four U.S. communities and the relationship with aging, cardiovascular disease, cerebrovascular disease, and mortality. The data can be found on the Canvas course web page by clicking on the "Files" link and then accessing the "Datasets" folder. The text file "mri.txt" contains the data for this study, and the data is in free-field format. RStudio can be used to read in the datafile using the file import tools. Alternatively, the following command can be used to read in the data:

```
read.table("mri.txt",header=TRUE)
```

Documentation of study and the variables in the dataset can be found in the file "mri.pdf"

For this homework, we will focus on the 7 variables measuring patient age, sex (*male*), weight, height, brain atrophy (*atrophy*), length of follow-up (*obstime*), and vital status at the end of follow-up (*death*).

**Questions:**

1. Provide a brief description of the population of interest for this study. Provide a brief description of the study sample, including how many subjects are included in this study.

2. For the 7 variables of interest, provide a full description of the type of variable each is: Quantitative or qualitative? Discrete or continuous? Nominal, ordinal, interval, or ratios? Censored or uncensored?

3. Where relevant, provide descriptive statistics for each of the variables of interest in the entire sample, as well as within groups defined by sex. The descriptive statistics should provide information on the number of valid observations, the number of missing observations, the mean, the standard deviation, the minimum, 25th percentile, median,

50th percentile, 75th percentile, and the maximum, where such statistics are of scientific interest.

4. Note that the participants were followed for varying times. Because of this, the proportion of observed deaths are not immediately interpretable. What is the minimum follow-up, **in years**, among the patients still alive at the time of data extraction?