**Memorandum**

To: NTRHD Intern

From: Dr. Brad Cannell, Director, NTRHD

Re: Descriptive Analysis II

**Overview:**

In this week’s lab, you will practice describing relationships between variables. Specifically, you will practice basic descriptive techniques appropriate for analysis of a continuous predictor variable and a continuous outcome variable, a categorical predictor variable and a continuous outcome variable, or a categorical predictor variable and a categorical outcome variable. Practice by completing the tasks below using the WHAS 500 data set.

**Task 1.** Please use the following links to download the WHAS 500 data to your computer and view the codebook:

* [Click here to view/download the WHAS 500 data codebook](https://www.dropbox.com/s/hx37w38ruytjwi9/WHAS%20500%20Description.pdf?dl=0). Look over the codebook to learn a little bit about the data you will be working with.
* [Click here to view/download the WHAS 500 data](https://www.dropbox.com/s/c14gqhjbny0mlra/whas500.txt?dl=1).

**Task 2.** Import **whas500.txt** into your R global environment as **whas500**.

**Task 3.** Please view the structure of **whas500** using the str() function or the glimpse() function.

|  |
| --- |
| **Question:**  1. The **whas500** data frame contains \_\_\_\_ rows and \_\_\_\_ columns. |

**Task 4.** Use the cor.test() function to explore the relationship between **age** and **initial heart rate.** Examine the correlation coefficient and p-value.

**Questions:**

2**.** Is there a positive, negative, or no correlation between **age** and **initial heart rate**?

3. The p-value returned by the cor.test() function is \_\_\_\_\_\_\_\_ . Please copy and paste the entire value.

**Task 5.** Explore the relationship between **age** and **initial heart rate** by creating a scatter plot. Make sure to add an OLS regression line. View this scatter plot again by gender.

**Question:**

4**.** Based on the scatter plot, is there an apparent relationship between **age** and **initial heart rate**?

**Task 6.** Calculate mean **bmi** within levels of gender.

**Question:**

5**.** Is the mean **bmi** for females lower than the mean **bmi** for males?

**Task 7.** Produce side-by-side histograms of **sysbp** comparing gender.

**Question:**

6. Is the systolic blood pressure normally distributed in both genders?

**Task 8.** Produce a boxplot of **bmi** comparing gender.

**Question:**

7. Based on the boxplots, what can we conclude about the relationship between **bmi** and gender in our sample?

**Task 9.** Create a two-way frequency table for **cohort year** and **history of cardiovascular disease**.

**Questions:**

8. What proportion of people from the cohort year 1999 reported a history of cardiovascular disease?

9. Which cohort year had the highest percentage of people reporting cardiovascular disease?

10. Of those reporting a history of cardiovascular disease, what percentage was in the cohort year 1997?