**Name:\_Sydney Wood\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**NOTE: Type your answers in the appropriate fields and make answer fields larger as needed. Please turn in a printed copy next Friday (October 12) before lab begins. Assignments will lose 5% of the total possible points for each day they are late.**

***For the following questions please use the HW01Data.csv file. Provide the syntax, commands, etc., that you used to complete each problem; report the answer to the problem in the space provided*.**

1. What are the dimensions of the data set?

Code/Syntax:

nrow(hw01data)

ncol(hw01data) OR dim(hw01data)

Answer:

rows- 233, columns- 5

2. What data type (e.g., numeric, character, etc.) is in each column?

Code/Syntax:

sapply(hw01data, class)

Answer:

biosex height\_in weight\_lbs age\_yr ed\_cmplt

"factor" "numeric" "numeric" "numeric" "factor"

3. How many participants were female? How many were male?

Code/Syntax:

table(hw01data$biosex)

Answer:

FEMALE MALE

119 114

4. What is the frequency (tally) of education completed (e.g., HS and MA) in the data set? That is, for each level of educational attainment, how many participants were in that level?

Code/Syntax:

table(hw01data$ed\_complt)

Answer:

College2YR College4YR HS MA PhD

51 51 39 55 37

5. In this data set, how many males are taller than 72 in? What are their row numbers in the data set?

Code/Syntax:

rtallmales <- which(hw01data$biosex == “MALE” & hw01data$height\_in > 72)

rtallmales

length(rtallmales)

Answer:

Row #s

[1] 1 72 152 203 216 222

# of Rows

[1] 6

6. (a) What is the average age of participants based on their level of educational attainment? That is, what is the mean age of participants who had completed HS, 2 years of college, etc.?

(b) What is the average age of participants based on their level of educational attainment and their biological sex? That is, what is the mean age of females who had completed HS, 2 years of college, etc., and the mean age of males who had completed HS, 2 years of college, etc. We recommend tabling your results.

Code/Syntax:

1. by(hw01data$age\_yr, hw01data$ed\_cmplt, mean) OR aggregate(age\_yr~ed\_cmplt, data = hw01data, mean)
2. aggregate(age\_yr~ed\_cmplt\*biosex, data = hw01data, mean) /can also use by() but I prefer aggregate()

Answer:

a) only showing aggregate() output

ed\_cmplt age\_yr

1 College2YR 20.01176

2 College4YR 23.14118

3 HS 24.80256

4 MA 25.10182

5 PhD 29.83784

b) only showing aggregate() output

ed\_cmplt biosex age\_yr

1 College2YR FEMALE 19.97895

2 College4YR FEMALE 23.08378

3 HS FEMALE 25.63333

4 MA FEMALE 25.06486

5 PhD FEMALE 30.01765

6 College2YR MALE 20.03125

7 College4YR MALE 23.29286

8 HS MALE 24.55333

9 MA MALE 25.17778

10 PhD MALE 29.68500

7. The formula for BMI is as follows:

(a) Create a new variable in the data frame called “bmi.” Calculate the BMI for the participants and place these values in the “bmi” column of the data frame.

(b) Calculate the average BMI separately for men and for women.

(c) The Centers for Disease Control ([CDC](https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/)) reports the following criteria for categorizing BMI:

|  |  |
| --- | --- |
| **Category** | **Body Mass Index** |
| Underweight | Less than or equal to 18.5 |
| Normal/Healthy Weight | Greater than 18.5 but less than 25 |
| Overweight | Greater than 25 but less than 30 |
| Obese | Greater than or equal to 30 |

Based on these criteria, how many men fall into each weight status category? How many women fall into each weight status category? Create an APA-style table to summarize your results.

Code/Syntax:

1. hw01data$bmi <- (hw01data$weight\_lbs/(hw01data$height\_in^2)) \* 703
2. aggregate(bmi~biosex, data = hw01data, mean)
3. hw01data$class[hw01data$bmi <= 18.5] <- "Underweight"

hw01data$class[hw01data$bmi > 18.5 & hw01data$bmi <= 25] <- "Normal/Healthy"

hw01data$class[hw01data$bmi > 25 & hw01data$bmi < 30] <- "Overweight"

hw01data$class[hw01data$bmi >= 30] <- "Obese"

table(list(hw01data$biosex, hw01data$class)

Answer:

a) has no output

b)

biosex bmi

1 FEMALE 22.71763

2 MALE 22.73494

c) modified output

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 1  *Frequency of Males and Female in each Body Mass Index Category* | | | | | |
|  | | Underweight | Normal/Healthy | Overweight | Obese |
|  | Female | 5 | 94 | 20 | 0 |
| Male | 4 | 94 | 15 | 1 |