A company wants to see if their new energy gummies actually improves 800 meter times in Men's Division I collegiate track runners. The company decides to give 10 coaches from 10 different schools 8 bags, which contain 3 gummies (which are red) and 3 placebo gummies (which are blue) each. Both the coaches and athletes are told the gummies have the same capability. Each team's coach is instructed to randomly assign 3 of his or her six runners into the "red" category and the other 3 into the"blue" category. Before the first meet, each team's coach will administer one gummy to each of the ten runners depending on their previous group assignment. This administering of gummies will continue for every meet of the season. At the end of the season, the average time improvement between the "blue" and "red" groups at each school from the beginning to the end of the year will be compared.

One potential bias that may arise in this experiment is the placebo effect. The runners with the placebo gummies may run faster because they believe they have something to help them. Also, this experiment relies on the honorability of the coaches. If a coach believes the gummies will help them win a large meet, they may save some for then, thereby increasing dosage at a larger event.

Stat 021 Homework 1

Your Name

Due: Wed, Sept 11

Instructions: A hard copy of your homework must be handed in to me at the end of class on the due date or I must have recieved via email a **pdf** version of your homework by **noon** on the due date. If you are having trouble getting your *.Rmd* file to compile, you need to get help with this **before** the due date.

You are allowed to hand in **only one** late homework assignment throughout the semester. If you need to hand in this particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will recieve a grade of 0 on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework.

Q 1) Download and install R and R Studio following the instructions in class. Install the package *swirl()* using the command "install.packages("swirl")". Once the package is installed, call the package to your R session using the command "library("swirl"). Follow the instructions that pop up in your console. Select the course option"1: R Programming: The basics of programming in R" and complete the following lessons (about 20 minutes to complete each)

- 1: Basic Building Blocks
- 2: Workspace and Files
- 8: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the tutorial is covering something that you already know how to do, use the *skip()* command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (0 points)

Q 2) Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truely being random and define different treatment/control groups and describle how you will randomly assign treatments to your observational units. (5 points)

Solution: Simply start typing your solution here. If you use **double asterisks** to surround a phrase or word to make it bold. Use *single asterisks* to make italics.

To start a new paragraph, make sure you leave enough line breaks between your text. To include mathematical expressions in a R Markdown document, use the same format as you would for a LaTex document and surround the equation with dollar signs like this: $\sum_{i=1}^n (y_i^2 + \bar{y})^2 = 5$ for inline expressions and with double dollar signs for expressions centered on their own line such as

$$\sum_{i=1}^n (y_i^2 + \bar{y})^2 = 5.$$

Q 3) Access the data set called *sleep* in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (10 points)

1. Define a new variable called *group1.sleep* that includes only the values of the variable *extra* for individuals from group 1.

2. Perform a t-test on “group1.sleep" to test of the extra hours slept by group 1 is smaller than or equal to 0.5 hours at an $\alpha = 0.1$ level. I.e. Test the hypothesis $H_0 : \mu = 0.5$ vs. $H_1 : \mu \neq 0.5$ at an $\alpha = 0.1$ significance level.

3. Report and interpret the 90% CI for the average extra hours of sleep for group 1.

4. Form a new categorical variable called *extra1.cat* that categorizes the variable *extra* into two groups, the first where extra hours slept is at least zero hours and the second where extra hours slept is less than zero hours. Print a table that counts the total number of observations in each group. (You may want to use the *ifelse()* function.)

5. Produce two boxplots for the variable *extra*, one corresponding to each group. Make sure each plot has a title. (You may find the function *grid.arrange()* in the package *gridExtra* useful for displaying two plots at once.)

Do me a favor and write your solutions to the different parts of Q 3 all in the same space (not between each bulletted list item). You can include a chunk of R code like this:

```
data(sleep)
sleep %>% head()
```

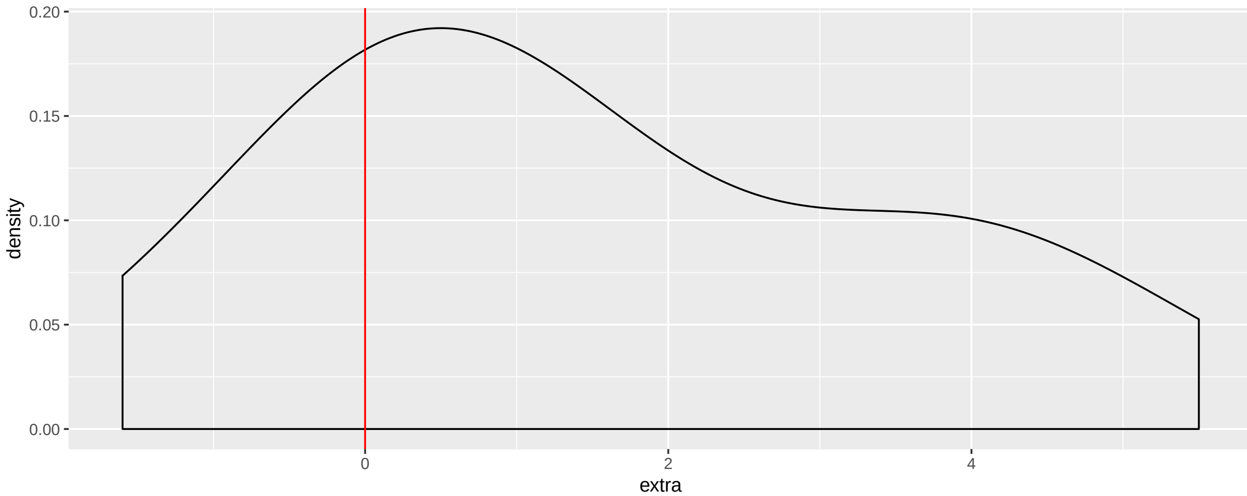
```
##   extra group ID
## 1   0.7     1  1
## 2  -1.6     1  2
## 3  -0.2     1  3
## 4  -1.2     1  4
## 5  -0.1     1  5
## 6   3.4     1  6
```

Note that the deliminators for r code are **not** apostraphes but are the tick marks found in the upper left hand corner of your keyboard. You will never need to print out an entire data set for me in your homework, just the fist few rows using the *head()* function is fine.

Another note, the “echo=TRUE” and “warning=FALSE” options in your R code chunk are settings that will make the incorporation of your code into your document a lot neater. These options tell R to print the output of the code to your document and to not print any warning signs that may come up in the console, respectively.

To include a plot, I recomend the following options for your R code chunk:

```
myPlot <- ggplot(sleep, aes(x=extra)) + geom_density() +
                                        geom_vline(xintercept=0, col="red")
myPlot
```



In the code above, *aes()* is short for aesthetic which doesn't make a whole lot of sense to me, regardless, it is the function that enables you to define your x (and y) variable(s).

Q 4) Suppose $X \sim N(\mu, \sigma^2)$. Show that the random variable defined as $Z = \frac{Z-\mu}{\sigma}$ has $E[Z] = 0$ and $Var[Z] = 1$. Show all of your steps (you may handwrite your answer to this question). (5 points)

Hint: Recall/look up some common properties of normal random variables and the rules of the expectation and variance operations.

title: "Eason_Major_HW1" author: "Major Eason" date: "9/12/2019" output: pdf_document: default html_document: default —

```
#create three vectors
extra <- c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, 0.8, 0.0, 2.0, 1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.
group <- c("1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "2", "2", "2", "2", "2", "2", "2", "2", "2
ID <- c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "1", "2", "3", "4", "5", "6", "7", "8", "9",

#Create data frame
sleep <- data.frame(extra, group, ID)
sleep
```

```
##    extra group ID
## 1    0.7     1  1
## 2   -1.6     1  2
## 3   -0.2     1  3
## 4   -1.2     1  4
## 5   -0.1     1  5
## 6    3.4     1  6
## 7    3.7     1  7
## 8    0.8     1  8
## 9    0.0     1  9
## 10   2.0     1 10
## 11   1.9     2  1
## 12   0.8     2  2
## 13   1.1     2  3
## 14   0.1     2  4
## 15  -0.1     2  5
## 16   4.4     2  6
```

```
## 17   5.5     2  7
## 18   1.6     2  8
## 19   4.6     2  9
## 20   3.4     2 10
```

t.test(sleep$extra, mu=0.5, alternative="less", conf.level=0.9)

"We are 90% confident that the mean number of extra hours of sleep for the population is 2.139098 hours"

```
extra1.cat <- ifelse(extra < 0,1,0)
print.table(extra1.cat)
```

```
##  [1] 0 1 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
```

```
#install.packages("gridExtra")
#install.packages("dplyr")
```

```
#filter(extra1.cat)
sleep$extra
```

```
##  [1]  0.7 -1.6 -0.2 -1.2 -0.1  3.4  3.7  0.8  0.0  2.0  1.9  0.8  1.1  0.1
## [15] -0.1  4.4  5.5  1.6  4.6  3.4
```

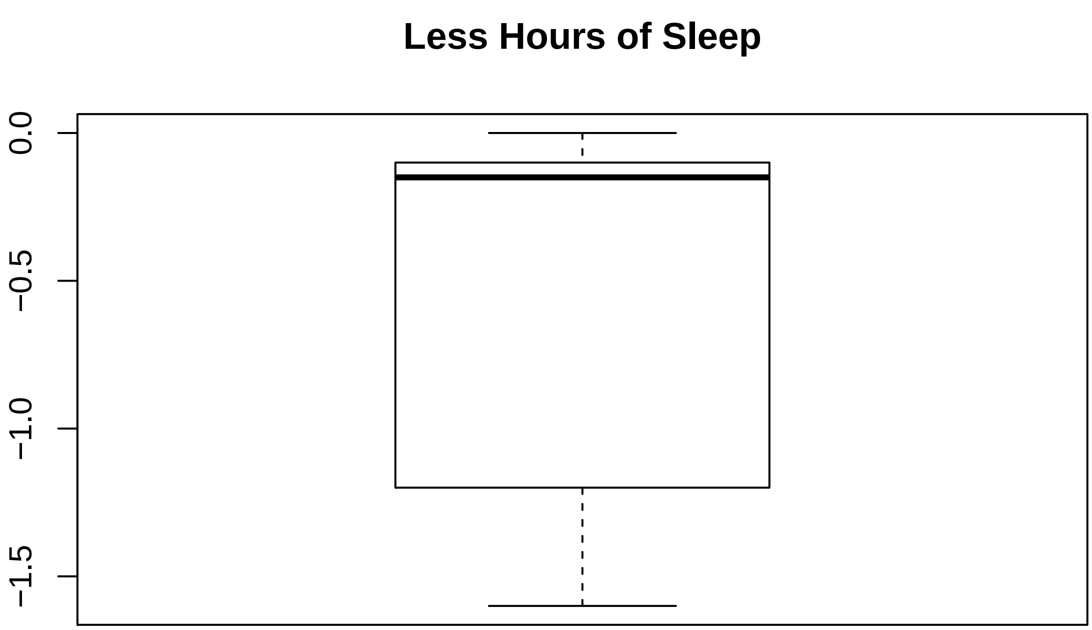
```
(sleep$extra<=0)
```

```
##  [1] FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
## [12] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
```

```
sleep$extra[(sleep$extra<=0)]
```

```
## [1] -1.6 -0.2 -1.2 -0.1  0.0 -0.1
```

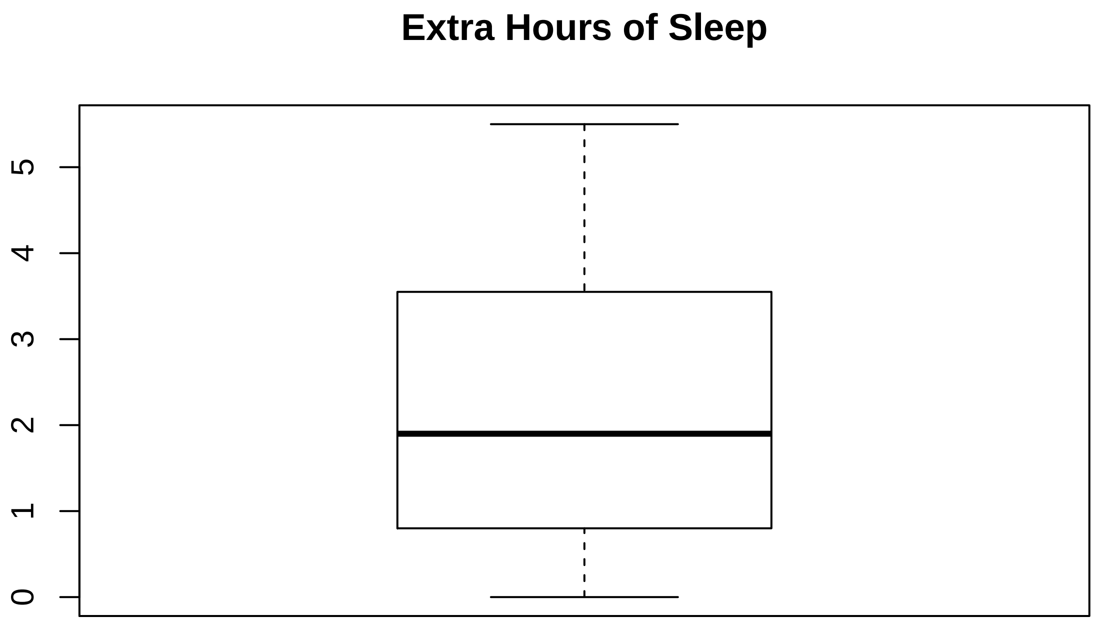
```
extra1.cat <- sleep$extra[(sleep$extra<=0)]
boxplot(extra1.cat, main="Less Hours of Sleep")
```



```r
sleep$extra[(sleep$extra>=0)]
```

```
##  [1] 0.7 3.4 3.7 0.8 0.0 2.0 1.9 0.8 1.1 0.1 4.4 5.5 1.6 4.6 3.4
```

```r
extra <- sleep$extra[(sleep$extra>=0)]
boxplot(extra, main="Extra Hours of Sleep")
```



Q4. $z = \frac{\bar{x} - \mu}{\sigma}$

sample mean ↓ Pop. mean ↙

$E(z) = E\left(\frac{\bar{x} - \mu}{\sigma}\right) = \frac{E(\bar{x} - \mu)}{\sigma} = \frac{E(\bar{x}) - E(\mu)}{\sigma} = 0$

$Var[z] = Var\left(\frac{\bar{x} - \mu}{\sigma}\right) = \frac{Var(\bar{x} - \mu)}{\sigma} = \frac{Var(\bar{x}) - Var(\mu)}{\sigma} =$

$\frac{1}{\sigma^2} var(\bar{x} - \mu) = \frac{var(\bar{x} - \mu)}{var} = 1$