Stat 021 Homework 1

Terell Dale

Due: Fri, Sept 13

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:

Purose of Experiment: Test the effect of caffine in coffee on the heart rate of Swarthmore students by measuring heart rate vs time for 2 hours after drinking the caffine.

Population: The population under study is the Swarthmore College academic year 2019-2020 student population.

Sampling Methodology: The sample would be a group of volunteer respondants to a mass email to all swarthmore students.

Potential Sampling Bias: This sampling method would create a volunteer bias in the sample. The act of some of the population choosing to respond and some choosing not to respond creates an inherent trait distinction in the sample. Also, students abroad would be unable to participate in the study.

Control Groups: A control group for this experiment would be a group given decaf coffee.

Random Assignment Methodology: To randomly assign each participant to a group, we could assign each one a number and use a random number generator to randomly choose half of the numbers for placement.

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.)

Solution:

```
group1.sleep <- sleep %>% filter(group == 1)
t.test(group1.sleep$extra, alternative="less", mu=.5, conf.level=0.9)
```

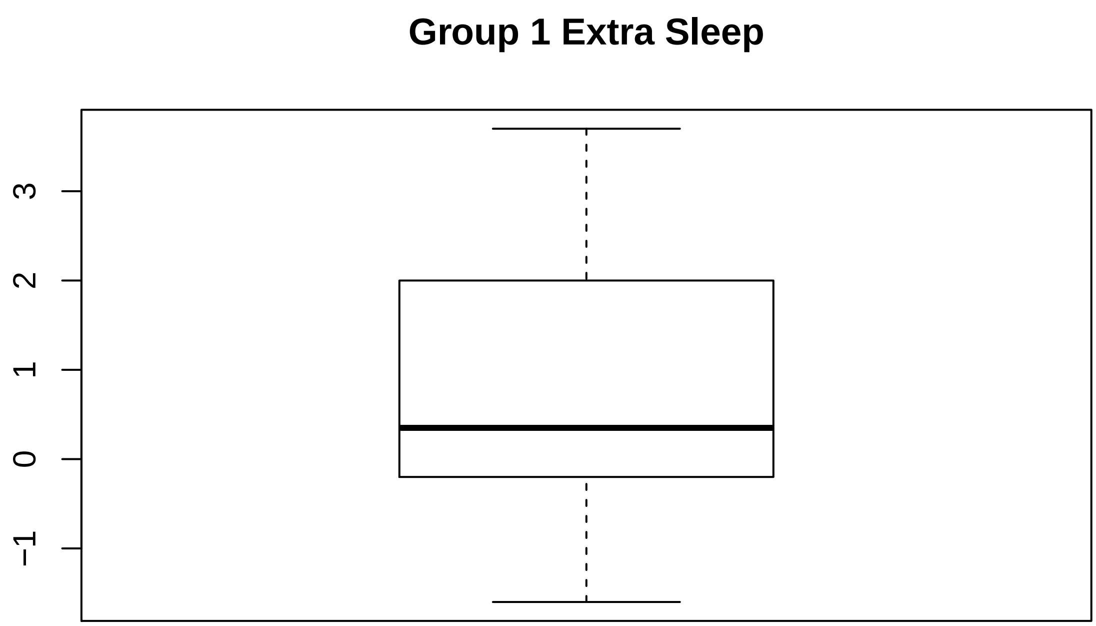
```
##
##  One Sample t-test
##
## data:  group1.sleep$extra
## t = 0.4419, df = 9, p-value = 0.6655
## alternative hypothesis: true mean is less than 0.5
## 90 percent confidence interval:
##      -Inf 1.532427
## sample estimates:
## mean of x
##      0.75
```

The 90% confidence interval is $(-\infty, 1.532427)$. There is a 90% chance that this confidence interval the true population mean.

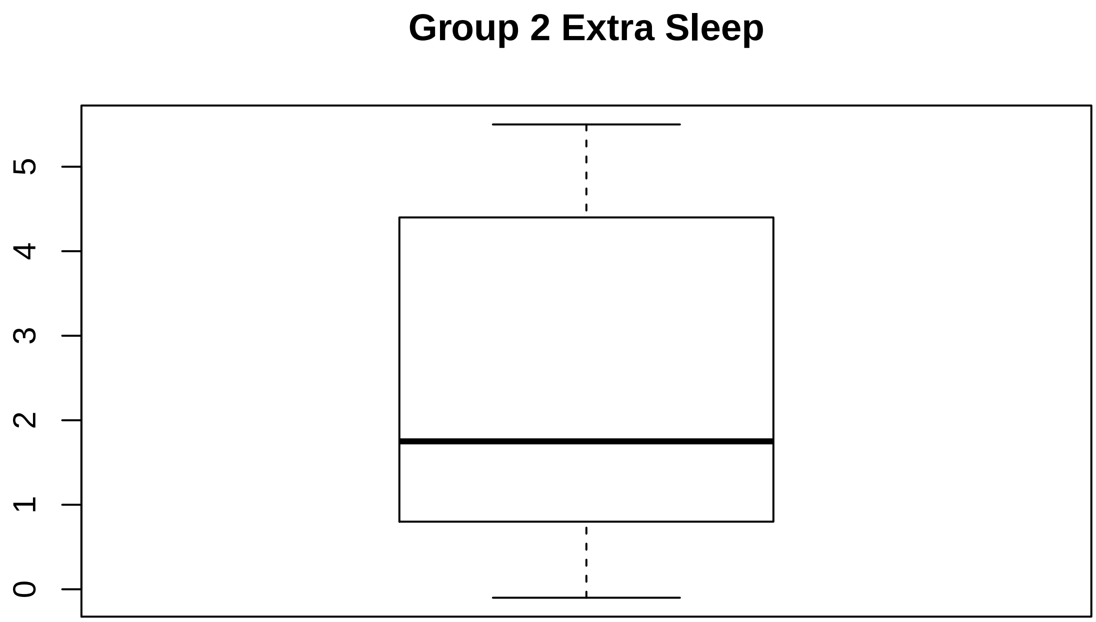
```
sleep %>%
  mutate(extra1.cat = ifelse(extra > 0, "Positive", "Negative")) %>%
    select(extra1.cat) %>%
      table()
```

```
## .
## Negative Positive
##        6       14
```

```r
sleep %>% filter(group == 1) %>% select(extra) %>% boxplot(main = "Group 1 Extra Sleep")
```




```r
sleep %>% filter(group == 2) %>% select(extra) %>% boxplot(main = "Group 2 Extra Sleep")
```



Q 4) Suppose $X \sim N(\mu, \sigma^2)$. Show that the random variable defined as $Z = \frac{X-\mu}{\sigma}$ 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.

Solution:

$$E(Z)$$
$$E(\frac{X-\mu}{\sigma})$$
$$\sigma E(X-\mu)$$
$$\sigma(E(X)) - \sigma(E(\mu))$$
$$\sigma\mu - \sigma\mu$$
$$E(Z) = 0$$

$$Var(Z)$$
$$Var(\frac{X-\mu}{\sigma})$$
$$\frac{1}{\sigma^2}Var(X-\mu)$$
$$\frac{1}{\sigma^2}Var(X)$$

$$\frac{1}{\sigma^2}\sigma^2$$

$$Var(Z) = 1$$