Stat 021 Homework 1

Harry Yeung

Due: Wed, Sept 11

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:

I will examine the effects of a new chemical in a recipe for a soda to see if the chemical makes that soda more popular over the original recipe. The population under study will consist of college students in Pennsylvania (to keep the study manageable). At each college, each student will be assigned a number, and 100 random numbers will be generated. Those 100 students will partake in the study. The experiment will be double-blind and the participants will try both sodas and say which one tastes better. Fifty kids will try the old soda first, and the other half will try the new soda. However, neither group will know which soda they are trying, nor will the person giving them the soda. Thus, the old soda acts somewhat as a control while the new soda acts as an experimental unit.

There could be a possible nonresponse bias, since if the selected participants don't drink soda or don't want to participate, then I would have to randomly select another number. In addition, there could be possible undercoverage, since the opinions of college students in Pennsylvania probably do not accurately represent the tastes of all American college students, let alone all Americans in general.

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.)

I am 90% confident that the true mean extra hours of sleep lies between 0 and 1.532427. If I were to repeat this experiment many times, then about 90% of the intervals produced would capture the true mean extra hours slept.

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

```
##   extra group ID
## 1   0.7     1  1
## 2  -1.6     1  2
```

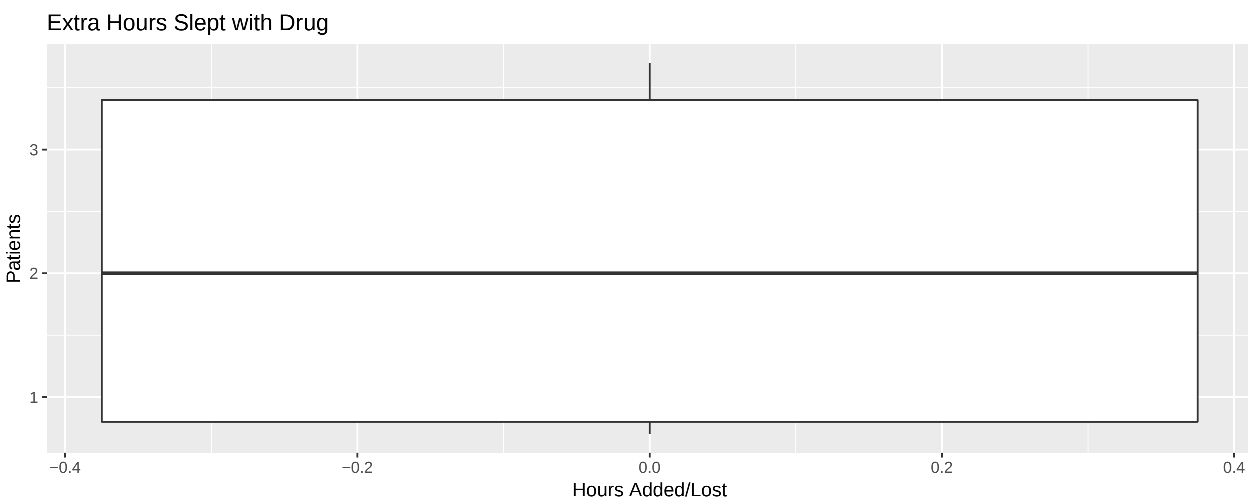
```
## 3  -0.2     1  3
## 4  -1.2     1  4
## 5  -0.1     1  5
## 6   3.4     1  6
```

```r
sleep_1 <- sleep %>% filter(group==1)
t.test(sleep_1$extra, alternative = c("less"), mu = .5, paired = FALSE, var.equal = FALSE, conf.level =
```

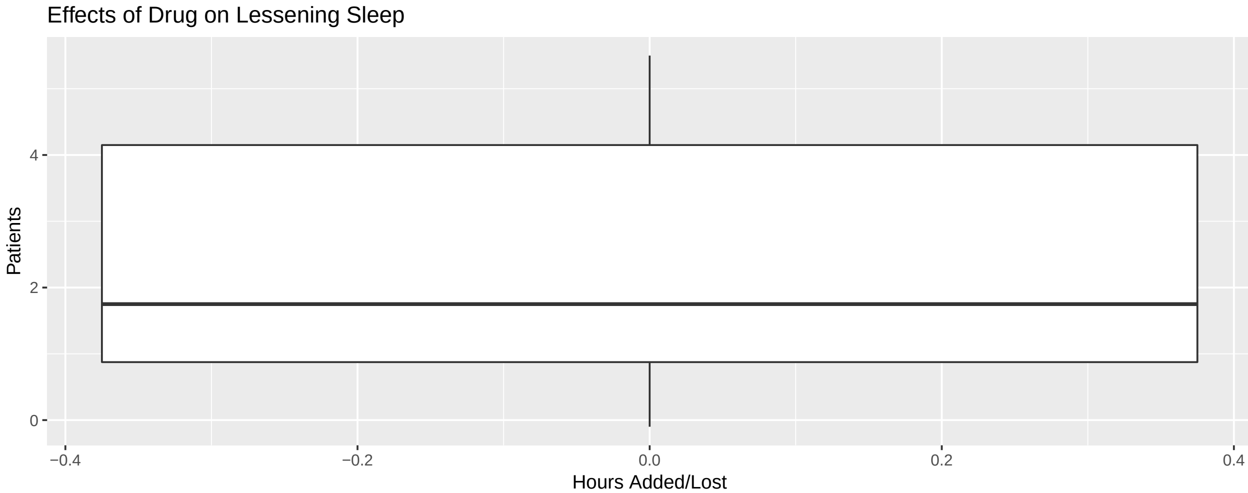
```
##
##  One Sample t-test
##
## data:  sleep_1$extra
## t = 0.4419, df = 9, p-value = 0.6655
## alternative hypothesis: true mean is less than 0.5
## 90 percent confidence interval:
##      -Inf 1.532427
## sample estimates:
## mean of x
##      0.75
```

```r
extra1.cat <- ifelse(sleep_1$extra>0, "pos","neg")
sleep_1.1 <- sleep_1 %>% mutate(extra1.cat = ifelse(sleep_1$extra>0, "pos","neg"))
sleep_2 <- sleep %>% filter(group==2)
  slept.extra <- sleep_1.1 %>% filter(extra1.cat == "pos")
```

```r
slept.extra %>% ggplot(aes(y=extra)) + geom_boxplot() + labs(title = "Extra Hours Slept with Drug", x =
```




```r
sleep_2 %>% ggplot(aes(y=extra)) + geom_boxplot() + labs(title = "Effects of Drug on Lessening Sleep", :
```


Effects of Drug on Lessening Sleep
Patients
4
2
0
−0.4
−0.2
0.0
0.2
0.4
Hours Added/Lost

4. $E[Z] = 0$

$Z = \frac{x-\mu}{\sigma}$, $E[Z] = E\left[\frac{x-\mu}{\sigma}\right]$

$N(\mu, \sigma^2)$ means X is normally distributed w/ mean $= \mu$

$\therefore$ $E[x] = \mu$

$$\frac{E[x] - E[\mu]}{E[\sigma]} = \frac{\mu - \mu}{E[\sigma]} = 0$$

$Var[Z] = 1$

$Var[Z] = Var\left[\frac{x-\mu}{\sigma}\right]$

$N(\mu, \sigma^2)$ means x is normally distributed w/ variance $= \sigma^2$

$$\frac{Var[x] - \cancel{var[\mu]}}{var[\sigma]} = \frac{\sigma^2}{\sigma^2} = 1$$