

# Stat 021 Homework 1 Solution

Suzanne Thornton

Due: Wed, Sept 11 by 5pm

**Q 1)** (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 truly being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (5 points)

**Solution:**

For full credit, the solution must identify the following elements:

- the population under study
- the method of collecting a sample from the population
- potential biases against getting a truly random sample
- the different treatment (possibly more than one) and/or control groups
- specific random assignment process for treatment(s) and/or control groups.

**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. Create a new data set called *group1.sleep* that only contains data for individuals from group 1.
2. Perform a t-test on “*group1.sleep*” to test if 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  $H_0 : \mu \geq 0.5$  vs.  $H_1 : \mu < 0.5$  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:**

**Part 1)**

```
data(sleep)

sleep_tibble <- as_tibble(sleep)
sleep_tibble <- sleep_tibble %>% mutate(group1_sleep = replace(extra, (sleep_tibble$group==2), NA))
##Here I still have a data set of size 20 but I have indicated that data for variable group1_
head(sleep_tibble)

## # A tibble: 6 x 4
##   extra group ID      group1_sleep
##   <dbl> <fct> <fct>          <dbl>
## 1    0.7 1      1              0.7
```

```
## 2 -1.6 1 2 -1.6
## 3 -0.2 1 3 -0.2
## 4 -1.2 1 4 -1.2
## 5 -0.1 1 5 -0.1
## 6 3.4 1 6 3.4
```

```
tail(sleep_tibble)
```

```
## # A tibble: 6 x 4
##   extra group ID   group1_sleep
##   <dbl> <fct> <fct>         <dbl>
## 1 -0.1 2 5 NA
## 2 4.4 2 6 NA
## 3 5.5 2 7 NA
## 4 1.6 2 8 NA
## 5 4.6 2 9 NA
## 6 3.4 2 10 NA
```

Part 2)

```
t.test(sleep_tibble$group1_sleep, alternative=c('greater'), my=0.5, conf.level=0.90, na.exclude=TRUE)
```

```
##
## One Sample t-test
##
## data: sleep_tibble$group1_sleep
## t = 1.3257, df = 9, p-value = 0.1088
## alternative hypothesis: true mean is greater than 0
## 90 percent confidence interval:
## -0.03242711 Inf
## sample estimates:
## mean of x
## 0.75
```

At a  $\alpha = 0.1$  confidence level, we fail to reject the null hypothesis with a p-value of 0.2176 and conclude there is not enough evidence to support the claim that group 1 gets more than a half hour of extra sleep.

**Part 3)** From the above output we have a one-sided 90% CI of  $[-0.032, \infty)$  which contains the hypothesized value 0.5, so at a significance level of 0.10, we fail to reject the null hypothesis and conclude there is not enough evidence to suggest that group 1 gets more than a half hour of extra sleep.

In terms of the problem, this CI means that, if we were to run this experiment many times, selecting a new random sample each time, we can expect that the true average sleep time for the entire population of “group 1”-people is no less than missing 0.032 hours of sleep.

Part 4)

```
sleep_tibble %>% mutate(extra1_cat = ifelse(extra>=0, 'A', 'B')) %>% count(extra1_cat)
```

```
## # A tibble: 2 x 2
##   extra1_cat    n
##   <chr>      <int>
## 1 A         15
## 2 B          5
```

Part 5)

```
library('gridExtra')
```

```
##
```

```
## Attaching package: 'gridExtra'

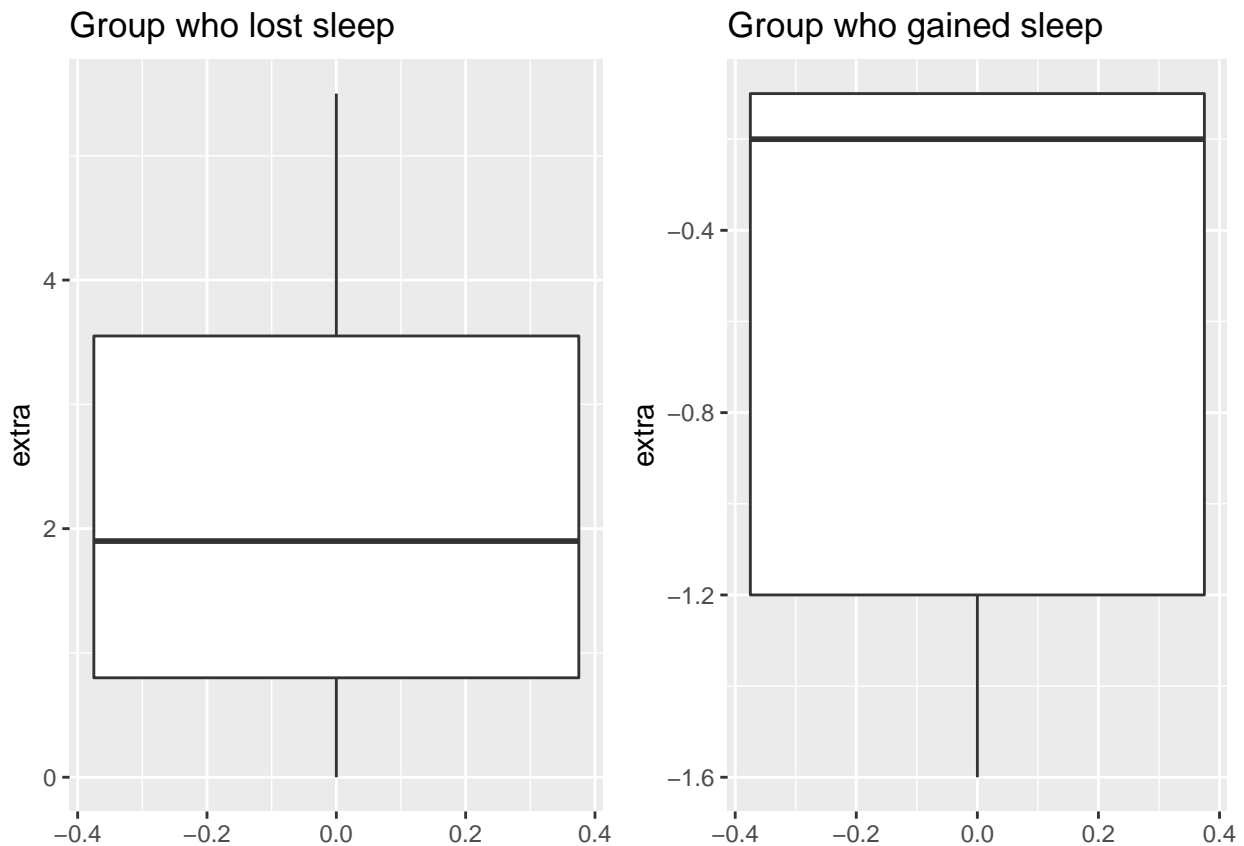
## The following object is masked from 'package:dplyr':
##
##      combine

sleep_tibble2 <- sleep %>% mutate(extra1_cat = ifelse(extra>=0, 'lost_sleep', 'gained_sleep'))

extra_lost <- sleep_tibble2 %>% filter(extra1_cat=='lost_sleep')
extra_gained <- sleep_tibble2 %>% filter(extra1_cat=='gained_sleep')

hist1 <- ggplot(data=extra_lost, aes(y=extra)) + geom_boxplot() + labs(title='Group who lost sleep')
hist2 <- ggplot(data=extra_gained, aes(y=extra)) + geom_boxplot() + labs(title='Group who gained sleep')

grid.arrange(grobs=list(hist1, hist2), ncol=2)
```



**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  $Var[Z] = 1$ . To receive full credit, **show all of your steps** (you may hand write 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:

$$E[Z] = E\left[\frac{Z - \mu}{\sigma}\right] = E\left[\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right] = \frac{E[X]}{\sigma} - \frac{\mu}{\sigma} = 0$$

In the above, the first equals sign is just the definition of  $Z$  in terms of  $X$ , the next equals sign is just an algebraic simplification of the term in parentheses, the next equals sign uses the linearity of expectation, and the penultimate equals sign uses the fact that we are given  $E[X] = \mu$ .

$$\text{Var}[Z] = \text{Var}\left[\frac{Z - \mu}{\sigma}\right] = \text{Var}\left[\frac{X}{\sigma} - \frac{\mu}{\sigma}\right] = \left(\frac{\text{Var}[Z]}{\sigma^2}\right) = \frac{\sigma^2}{\sigma^2} = 1$$

Note the following property of the variance function: For any constants  $a, b$  and any random variable  $X$ ,

$$\text{Var}[aX + b] = a^2 \text{Var}[X].$$

Again, in the above, the first equals sign is just plugging in the definition of  $Z$ , the second equals sign is an algebraic simplification, the third applies the property mentioned above, and the penultimate equals sign uses the fact that we are given  $\text{Var}[X] = \sigma^2$ .