S&DS 220: Homework 10

Due Friday April 19

Braeden Cullen

Instructions

1. Complete the questions below. Upload your knitted PDF solutions to Gradescope by the due date.
2. Your solutions should be a combination of writing and R code. When writing, use complete sentences.
3. Previous homework assignments already had code chunks created for you. Now it is up to you to insert R code chunks within each problem as needed.
4. You should aim for clear and concise communication (in both words and R code).

Problem set questions

Question 1: (8.31) Fish oil and blood pressure

Consider the `ex0112` data in the `Sleuth3` package. The researchers randomly assigned 7 people to receive fish oil and 7 people to receive regular oil, and measured the decrease in their diastolic blood pressure. Is there sufficient evidence to conclude that the mean decrease in blood pressure of patients taking fish oil is different from the mean decrease in blood pressure of those taking regular oil?

```
library(Sleuth3)

glimpse(ex0112)
```

```
## Rows: 14
## Columns: 2
## $ BP   <int> 8, 12, 10, 14, 2, 0, 0, -6, 0, 1, 2, -3, -4, 2
## $ Diet <fct> FishOil, FishOil, FishOil, FishOil, FishOil, FishOil, FishOil, Re~
```

```
t.test(data = ex0112, BP ~ Diet)
```

```
##
##  Welch Two Sample t-test
##
## data:  BP by Diet
## t = 3.0621, df = 9.2643, p-value = 0.01308
## alternative hypothesis: true difference in means between group FishOil and group RegularOil is not e
## 95 percent confidence interval:
##   2.039893 13.388678
## sample estimates:
##    mean in group FishOil mean in group RegularOil
##                 6.571429                -1.142857
```

By examining the p-value we can determine if the mean decrease in blood pressure of patients taking fish oil is different from the mean decrease in blood pressure of those taking regular oil. The p-value is 0.013, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is sufficient evidence to suggest that the mean decrease in blood pressure of patients taking fish oil is different from the mean decrease in blood pressure of those taking regular oil.

Question 2: (8.42) Drug trials

In a drug trial, patients are given a drug to test if it will cure their illness.

(a) What is the null hypothesis for a drug trial?

The null hypothesis for a drug trial is the following: the drug has no effect on the illness.

(b) What would a type I error be for a drug trial?

A type I error would be to determine that the drug has an effect when in reality it doesn't

(c) What would a type II error be for a drug trial?

A type II error is to conclude that the drug has no effect when in reality it does.

(d) What are the implications for each type of error for ill patients?

Implications of a type I error for ill patients would be that they think they are given a drug that cures their illness when in reality it does not actually cure their illness. Implications of a type II error for ill patients would be that they are given a drug that is effective but they think that it is not effective.

(e) What are the implications for each type of error for drug manufacturers?

Implications of a type I error for drug manufacturers would be that they are able to sell a drug that does not actually cure the illness, therefore reducing the reputability of that manufacturer. Implications of a type II error for drug manufacturers would be that they are not able to sell a drug that could potentially cure the illness, therefore making them lose out on potential profits.

Question 3: (8.43) Determining sample size

Suppose you wish to test the side effects of a new vaccine by forming two groups of subjects, one that gets the vaccine and another that gets a placebo. You would like to detect a $1°F$ change in body temperature at the $\alpha = 0.05$ level of significance with a power of 99%. Data on human body temperature (such as `fosdata::normtemp`) suggests that the standard deviation of body temperature is around $0.73°F$. How many subjects will you need in each group?

```
standard_deviation <- sd(fosdata::normtemp$temp)

power.t.test(delta = 1, sd = 0.73, sig.level = 0.05, power = 0.99)
```

```
##
##      Two-sample t test power calculation
##
##               n = 20.59948
##           delta = 1
##              sd = 0.73
##       sig.level = 0.05
##           power = 0.99
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Based on this result, we will need atleast 21 subjects in each group in order to detect a $1°F$ change in body temperature at the $\alpha = 0.05$ level of significance with a power of 99%.

Question 4: (8.47) A study on treating wrist fractures I: Statistical power

This exercise uses the data set `fosdata::wrist`, which comes from a study by Raittio et al.

The authors performed a power analysis in order to determine sample size. They wanted a power of 80% at the $\alpha = 0.05$ level, and they assumed that 30% of their patients would drop out. Previous studies had shown that the standard deviation of the `prwe12m` variable would be about 14. They concluded that they needed 40 participants in each group. What difference in means did they use in their power computation? (You can check your answer by looking up their paper.)

```
power.t.test(n = 40*0.7, sd = 14, sig.level = 0.05, power = 0.80, type = "two.sample")
```

```
##
##      Two-sample t test power calculation
##
##               n = 28
##           delta = 10.67366
##              sd = 14
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

We can conclude that they utilized a difference in means of 10.67 in their power computation.

Question 5: (8.48) A study on treating wrist fractures II: t-test

This exercise uses the data set `fosdata::wrist`, which comes from a study by Raittio et al.

The authors used the `fosdata::wrist` data set to find a 95% confidence interval for the difference in the mean of `prwe12m` between the functional cast position and the volar-flexion and ulnar deviation cast position. They **assumed** equal variances in the two populations to find the 95% confidence interval.

(a) Find the 95% confidence interval of the mean difference assuming equal variances.

‘*Solution.*

```
t.test(data = wrist, prwe12m ~ cast_position, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  prwe12m by cast_position
## t = -1.1739, df = 84, p-value = 0.2438
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -13.091655   3.372615
## sample estimates:
## mean in group 1 mean in group 2
##        15.51282        20.37234
```

(b) Compare your answer to the result in the published paper.

This answer is the same as the result in the published paper.

(c) Was assuming equal variance justified?

We are justified in assuming equal variance. If we examine the plot drawn, we can see that while cast position does have an impact on the variance of the data, the variance is not significantly different between the two groups. This is further shown by the t.test performed, which shows that while the confidence interval is narrowed when assuming equal variance, the confidence interval is still relatively wide, which suggests that the assumption of equal variance is justified.

```
t.test(data = wrist, prwe12m ~ cast_position, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  prwe12m by cast_position
## t = -1.2106, df = 82.237, p-value = 0.2295
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -12.844871   3.125831
## sample estimates:
## mean in group 1 mean in group 2
##        15.51282        20.37234
```

```r
# visualization of differences in variance, not assuming equal variance
ggplot(data = wrist, aes(x = cast_position, y = prwe12m)) +
  geom_boxplot() +
  geom_jitter(width = 0.2, height = 0, alpha = 0.5) +
  labs(title = "Boxplot of PRWE12M by Cast Position",
       x = "Cast Position",
       y = "PRWE12M") +
  facet_wrap(~cast_position, scales = "free_y") +
  theme_minimal()
```

```
## Warning: Continuous x aesthetic
## i did you forget 'aes(group = ...)'?

## Warning: Removed 19 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning: Removed 19 rows containing missing values or values outside the scale range
## ('geom_point()').
```