Aashwin Vats

Homework 1

1. **Exercises 3, 5, and 11 from page 23.**

- **Ex 3:** *A study found that individuals who lived in houses with more than two bathrooms tended to have higher blood pressure than individuals who lived in houses with two or fewer bathrooms. (a) Can cause and effect be inferred from this? (b) What confounding variables may be responsible for the difference?*

  a) No, the cause & effect can't be inferred from this as this is an observational study and it is impossible to draw causal conclusion from statistical analysis alone.
  b) The disparity in blood pressure levels can be a result of dietary habits, fitness level, older age, drug intake, genetics, thyroid disorders and other confounding variables & not just because of the number of bathrooms alone. Thus, the effect of number of bathrooms can't be separated from the effect of blood pressure levels.

- ***Ex 5:*** *In 1930 an experiment was conducted on 20,000 school children in England. Teachers were responsible for randomly assigning their students to a treatment group—to receive 34 pint of milk each day—or to a control group—to receive no milk supplement. Weights and heights were measured before and after the four-month experiment. The study found that children receiving milk gained more weight during the study period. On further investigation, it was also found that the controls were heavier and taller than the treatment group before the milk treatment began (more so than could be attributed to chance). What is the likely explanation and the implication concerning the validity of the experiment?*

  When the collection of available units from distinct groups was examined, it was found out that controls were heavier & taller than the treatment group before the milk treatment began (more so than could be attributed to chance).
  This may mean that the random assignment of students made by teachers to different groups wasn't random after all. There is a high chance that there was a bias in the division of groups & the teachers assigned students in the groups as per their judgment of the right person for the said group. This would imply that the results inferred from the study aren't reliable and hence the experiment's results are not valid.

- ***Ex 11:*** *A number of volunteers were randomly assigned to one of two groups, one of which received daily doses of vitamin C and one of which received daily placebos (without any active ingredient). It was found that the rate of colds* *was lower in the vitamin C group than in the placebo group. It became evident, however, that many of the subjects in the vitamin C group correctly guessed that they were receiving vitamin C rather than placebo, because of the taste. Can it still be said that the difference in treatments is what caused the difference in cold rates?*

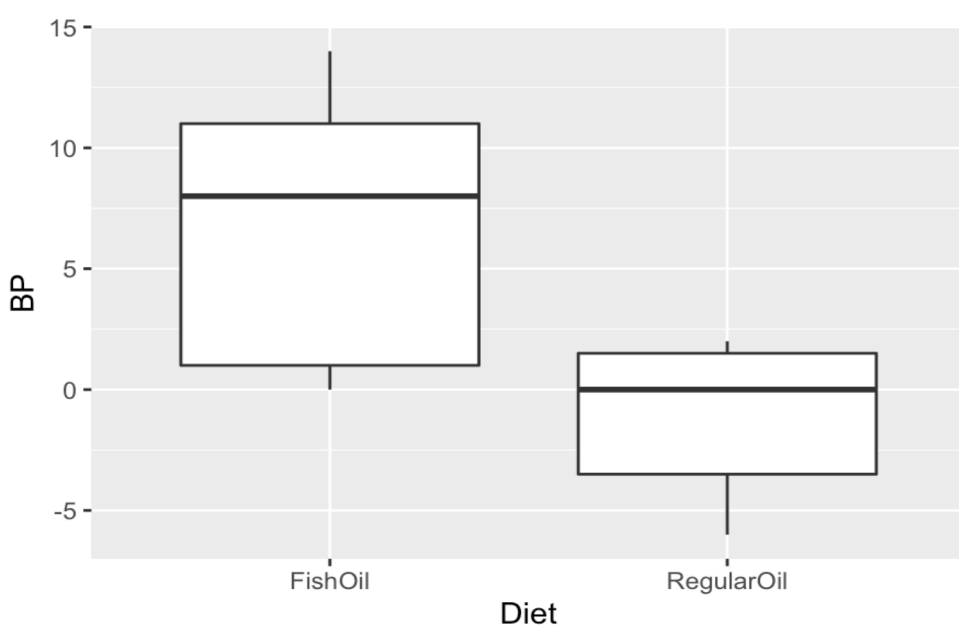
The research group that conducted this study needs to recognize how the two groups differed. They need to acknowledge that in their study, one group received daily doses of vitamin C and knew that they were receiving vitamin C and the other group received daily placebos. Research group should make sure that both groups are on the same page with respect to the treatment received and knowledge of the drug they are taking (to have a fair study). In addition to that, there may also be certain confounding variables affecting the results of this random experimental study. Hence, we don't have a convincing argument to say that the difference in treatments is what caused the difference in cold rates.

2. *The research question is if including fish oil in diet has any effect on diastolic blood pressure, and how large the difference is. Using these data, do the following.*
   a. *Produce side-by-side boxplots for the two groups. Turn in your plot and your R code.*

R Code:

```
qplot(Diet, BP, data=ex0112, geom="boxplot")
```

Plot:



b. *I did a permutation test by calculating the difference in sample means for 10,000 random shuffles of the data into two equal-sized groups. Below is a histogram. Label the observed value of the difference on this histogram. Comment on the plausibility of the null hypothesis of no difference between the groups.*

Calculating the observed value of the difference in sample mean:

R Code:

```
with(ex0112,(mean(BP[Diet=="FishOil"]))) -
with(ex0112,(mean(BP[Diet=="RegularOil"])))
```

Values:

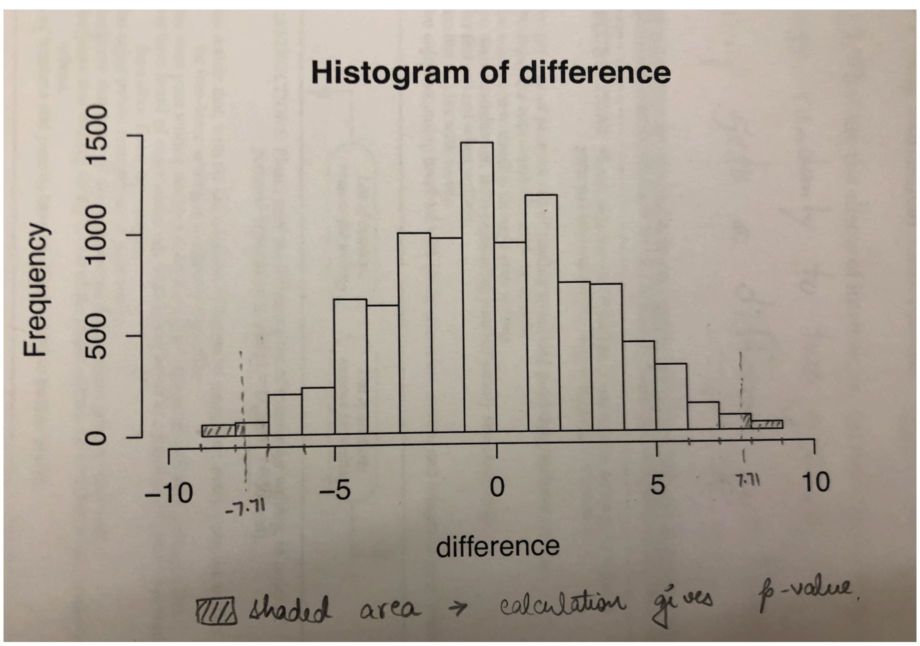
6.571429 and -1.142857 respectively

Observed difference in sample mean/output of above R code:

= 6.571429 – (-1.142857)

= 7.714286

The observed test statistic (7.71) is very unusual if the H_0 is true because it's in the far tail of the distribution. Hence, H_0 is <u>not plausible</u> for the given data.



In the above histogram, the observed value of the difference (7.71) has been labeled and the calculation of the shaded area gives the p-values.

c. *Perform a two-sided t-test using R's t.test function. Turn in your R code, but not the output. Instead, write a brief "Statistical Conclusion" similar to those given for the two case studies in Chapter 1.*

Performing a two sided t-test in R

R Code:

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

p-value: 0.009861

Statistical conclusion:

There is substantial evidence that the sample mean difference in blood pressure of subjects including fish oil in diet vs. those including regular oil is non zero (p-value = 0.009861, two-sided t-test) & hence receiving fish oil in the diet resulted in reduction of diastolic blood pressure of the subjects in the study. It is estimated that the mean reduction in BP is 7.71mm & a 95% confidence interval was observed for the difference from -1.14 to 6.57mm.