Statistical Inference Assignment 1 - Question 2

Lachlan Glascott

April 2020

Question 2

Overview

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package and compare tooth growth by supp and dose.

Firstly the data is loaded.

```
data("ToothGrowth")
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

The data has 60 rows and three columns. Each row indicates an individual guinea pig, while the columns len (tooth length) and dose (dose level) are numerical, and supp (supplement) which is a categoerical variable.

The mean and standard deviation of tooth length by supplement and dose are shown below.

```
ToothGrowth %>%
  dplyr::group_by(supp, dose) %>%
  dplyr::summarise(mean_len = mean(len),
                   sd_len = sd(len))
```

```
## # A tibble: 6 x 4
## # Groups:   supp [2]
##   supp   dose mean_len sd_len
##   <fct> <dbl>    <dbl>  <dbl>
## 1 OJ      0.5    13.2    4.46
## 2 OJ      1      22.7    3.91
## 3 OJ      2      26.1    2.66
## 4 VC      0.5     7.98   2.75
## 5 VC      1      16.8    2.52
## 6 VC      2      26.1    4.80
```

The data can also be displayed in a boxplot (refer to Appendix).

The summary data suggests that guinea pigs which recieved the OJ supplement have a higher average tooth length than those which recieved VC. We can test this under the null hypothesis that there is no difference between mean tooth length for guinea pigs which recieved OJ and VC.

```
VC <- ToothGrowth %>% dplyr::filter(supp == "VC")
OJ <- ToothGrowth %>% dplyr::filter(supp == "OJ")

t.test(OJ$len, VC$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Since the 95% confidence interval includes 0, we fail to reject the null hypothesis that that there is no difference between mean tooth length for guinea pigs which recieved OJ and VC. However, since the p value is below 0.1 we would reject the null hypothesis at a 90% confidence level.

The summary data also shows that average tooth length also tends to increase as the dose increases. We can also test this under three seperate hypothesis tests: 1. The difference in means of groups which recieved dosage of 2 and dosage of 0.5 is equal to zero 2. The difference in means of groups which recieved dosage of 2 and dosage of 1 is equal to zero 3. The difference in means of groups which recieved dosage of 1 and dosage of 0.5 is equal to zero

```
dose05 <- ToothGrowth %>% dplyr::filter(dose == 0.5)
dose1 <- ToothGrowth %>% dplyr::filter(dose == 1)
dose2 <- ToothGrowth %>% dplyr::filter(dose == 2)
t.test(dose2$len, dose05$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose05$len
## t = 11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83383 18.15617
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

```
t.test(dose2$len, dose1$len)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  dose2$len and dose1$len
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
t.test(dose1$len, dose05$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose05$len
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

For all three tests we reject the null hypothesis that the difference in the means of groups which recieved different dose levels is zero at the 95% confidence level.

Note that using multiple hypotheis tests is subject to multiple comparisons issues such as false positives and false discovery. False positives can be controlled using the Bonferroni correction.

```
p1 <- t.test(dose2$len, dose05$len)$p.value
p2 <- t.test(dose2$len, dose1$len)$p.value
p3 <- t.test(dose1$len, dose05$len)$p.value

p_values <- c(p1, p2, p3)

alpha <- 0.05
hypotheses <- 3

bonferroni <- alpha/hypotheses

p_values<bonferroni
```

```
## [1] TRUE TRUE TRUE
```

Controlling for FWER we still reject the null hypothesis for each of the hypotheses.

There is insufficient evidence to conclude with 95% confidence that OJ or VC have a different impact on average tooth length. However, there is evidence that average tooth length increases as the dose level increases.

Appendix

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose)))+
               geom_boxplot()+
               facet_grid(.~supp)
```