Course Project - Part II

Part 2: Basic Inferential Data Analysis Instructions

Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Task I

```
library(datasets)
library(dplyr)
library(ggplot2)
```

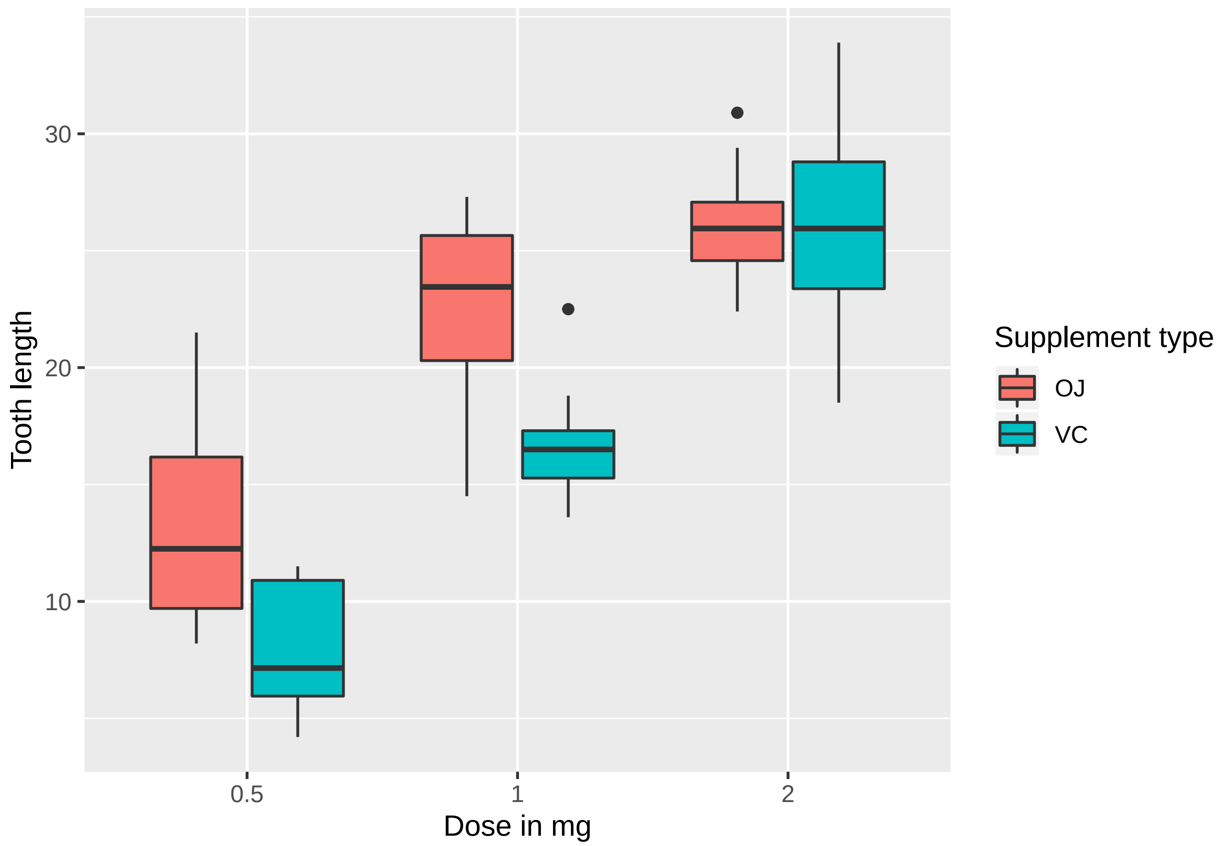
```
# Load ToothGrowth Data and explore it
data("ToothGrowth")
ToothGrowth %>%
  group_by(supp,dose) %>%
  summarise(avg_len = mean(len),
            Q25_len = quantile(len,0.25),
            Q75_len = quantile(len,0.75),
            sd_len = sd(len)) -> data
data
```

```
## # A tibble: 6 x 6
## # Groups:   supp [2]
##   supp   dose avg_len Q25_len Q75_len sd_len
##   <fct> <dbl>   <dbl>   <dbl>   <dbl>  <dbl>
## 1 OJ      0.5   13.2     9.7     16.2   4.46
## 2 OJ      1     22.7    20.3     25.6   3.91
## 3 OJ      2     26.1    24.6     27.1   2.66
## 4 VC      0.5    7.98    5.95    10.9   2.75
## 5 VC      1     16.8    15.3     17.3   2.52
## 6 VC      2     26.1    23.4     28.8   4.80
```

These dataset deals with the effect of vitamin C on tooth growth in guinea pigs. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Let's plot the data and explore it

```r
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_boxplot() +
  xlab("Dose in mg") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```



According to our plot, there seems to be:

1. There is a trend between dose and len.
2. There is a trend between dose and supp.

It's time to check both hypothesys:

Hyp I: Dose vs Len

H0- There is no effect of dosage on tooth growth

```r
# arrange our dateset
dose_05mg <- filter(ToothGrowth, dose == 0.5)
dose_1mg <- filter(ToothGrowth, dose == 1)
dose_2mg <- filter(ToothGrowth, dose == 2)
```

```r
# compare between dose of 0.5 and 1
t.test(dose_1mg$len, dose_05mg$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_1mg$len and dose_05mg$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```r
# compare between dose of 1 and 2
t.test(dose_2mg$len, dose_1mg$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_2mg$len and dose_1mg$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

As both p-values are less than 0.001, it is highly unlikely that there is no effect of dosage on tooth growth. Therefore, we reject our H0. The higher dosages the higher tooth growth.

Hyp II: Dose vs Sup

H0- There is no difference between supplements OJ and VC

```r
#Dosage: 0.5mg
dose_OJ_05mg <- filter(dose_05mg, supp == "OJ")
dose_VC_05mg <- filter(dose_05mg, supp == "VC")
t.test(dose_OJ_05mg$len, dose_VC_05mg$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_OJ_05mg$len and dose_VC_05mg$len
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604     Inf
```

```
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```r
#Dosage: 1mg
dose_OJ_1mg <- filter(dose_1mg, supp == "OJ")
dose_VC_1mg <- filter(dose_1mg, supp == "VC")
t.test(dose_OJ_1mg$len, dose_VC_1mg$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_OJ_1mg$len and dose_VC_1mg$len
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```r
#Dosage: 2mg
dose_OJ_2mg <- filter(dose_2mg, supp == "OJ")
dose_VC_2mg <- filter(dose_2mg, supp == "VC")
t.test(dose_OJ_2mg$len, dose_VC_2mg$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_OJ_2mg$len and dose_VC_2mg$len
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

According to the obtained p-values, for dosages of 0.5 and 1 mg, we must reject HO, so OJ is better than VC. On the other hand, for dosages of 2mg, our p-value is compatible with H0, therefore the are no differents between OJ and VC