Course 6 - Statistical Inference Week 4

Stephan Leijtens

Part 2 of Peer Assignment

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
library(ggplot2)

head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
str(ToothGrowth)
```

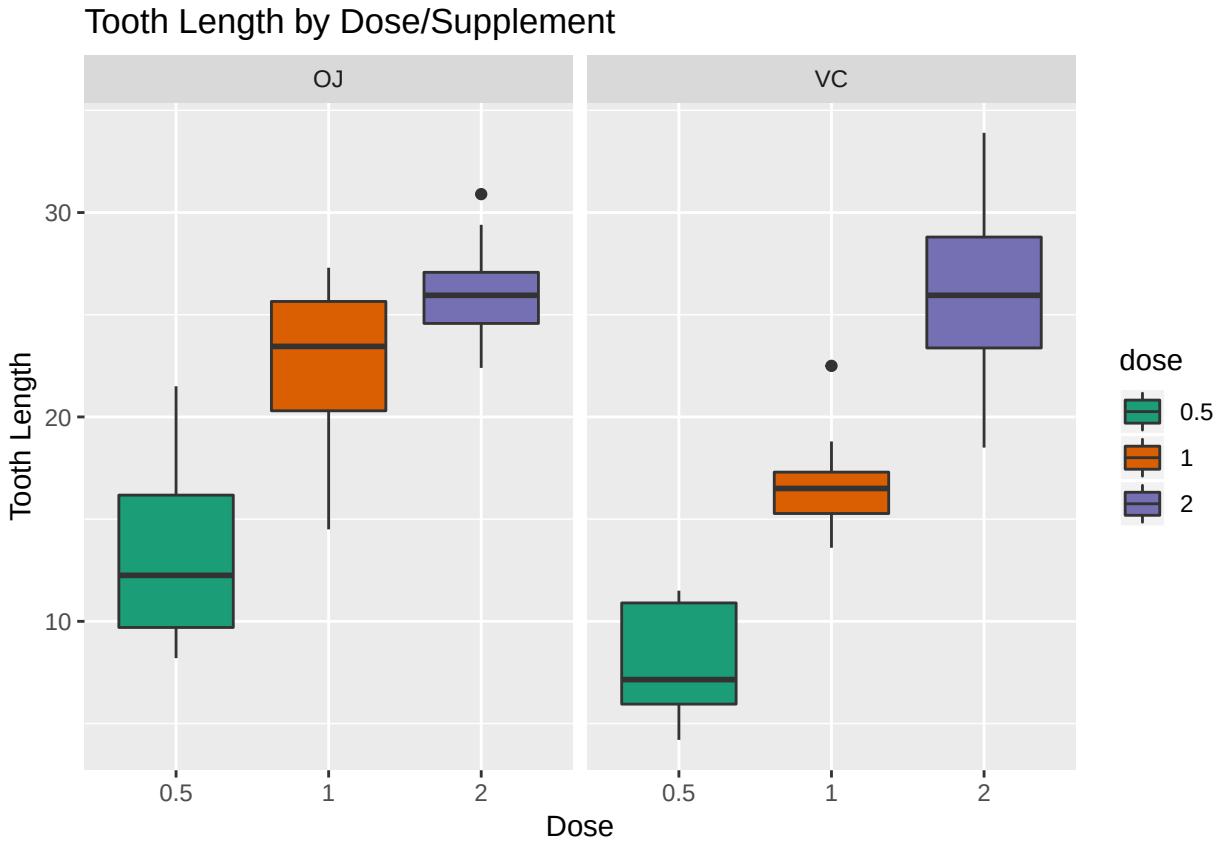
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

2. Provide a basic summary of the data.

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ggplot(aes(x=dose, y=len), data = ToothGrowth) +
        geom_boxplot(aes(fill=dose)) +
        ggtitle("Tooth Length by Dose/Supplement") +
        xlab("Dose") +
        ylab("Tooth Length") +
        facet_grid(~supp) + scale_fill_brewer(palette = "Dark2")
```



This gives us a visual indication that higher doses improve tooth growth, but it is still unclear if one type of supplement is more effective than the other.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

Our first hypothesis that we want to check is whether or not tooth growth is correlated with taking supplements. More specifically: - H0: There is no significant difference in tooth growth between delivery methods (OJ and VC) - H1: There is a significant difference in tooth growth between delivery methods (OJ and VC)

```
t.test(len~supp,paired=FALSE,var.equal=FALSE,data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

We see a p-value of 0.06 which is larger than the 0.05 we need to be able to reject the null hypothesis.

The second hypothesis is the following: - H0: Tooth growth is not correlated with dose amount - H1: Tooth growth is correlated with dose amount

```r
dose1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
dose2 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
dose3 <- ToothGrowth$len[ToothGrowth$dose == 2.0]

t.test(dose1, dose2, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1 and dose2
## 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:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

The p-value is very small which allows us to reject the null hypothesis. Continuing:

```r
t.test(dose1, dose3, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1 and dose3
## 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:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

Also here we reject the null hypothesis since we have a very small p-value. And finally:

```r
t.test(dose2, dose3, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2 and dose3
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

And in the final case we reject the null hypothesis as well.

4. State your conclusions and the assumptions needed for your conclusions.

Following our hypothesis tests it is safe to say that tooth length increases when dosage amount increases, regardless of administration method. We could not draw any conclusions if either OJ or VC was better. The

assumptions made are that this sample is a proper representation of the population and that the distribution of the means is normal.