Part 2: Basic Inferential Data Analysis Instructions

Andrea Eoli - Statistical Inference Course Project [2022-11-07]

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

Q1-2: Exploratory analysis and summary

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

```
library(dplyr)
library(ggplot2)
data("ToothGrowth")
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
```

```
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
```

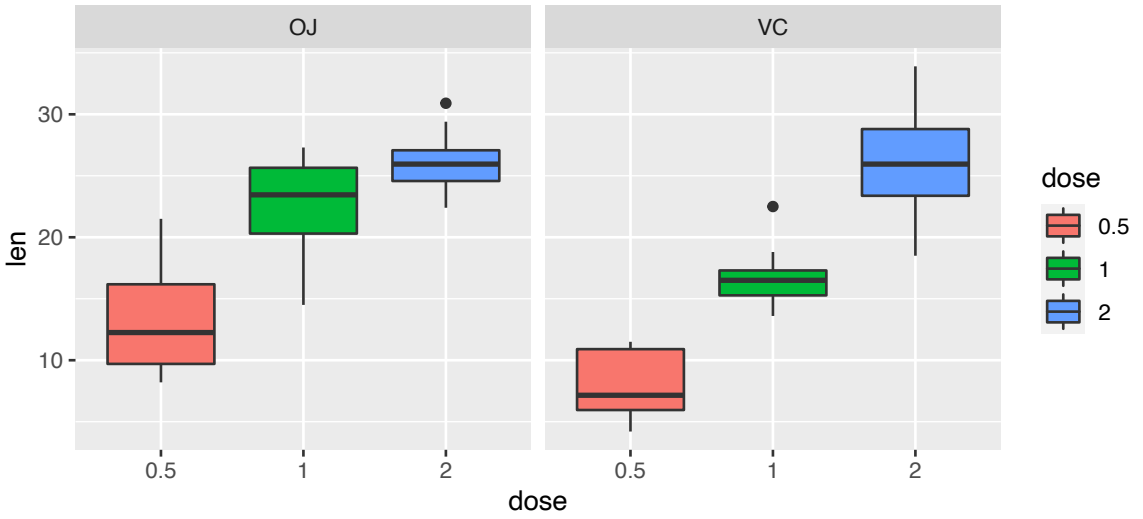
```
unique(ToothGrowth$dose) # Check if dose is really numeric -> only 3 values, convert to factor
```

```
## [1] 0.5 1.0 2.0
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

Let's visualise the data as a boxplot of the tooth growth by dosage and type of supplement.

```
ggplot(aes(x=dose, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=dose)) +
  facet_wrap("supp")
```



Q3: Compare tooth growth by supp and dose

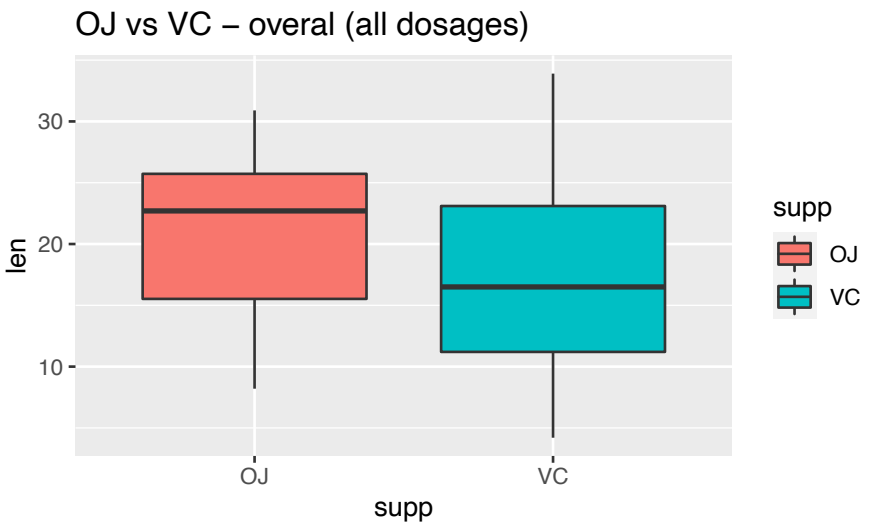
Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering) We assume unequal variances between the two groups and test, for each group, growth length between low and high doses.

```
a <- t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose %in% c(0.5,2) & ToothGrowth$supp == "OJ",])
b <- t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose %in% c(0.5,2) & ToothGrowth$supp == "VC",])
```

For both supplements the p-value is much smaller than 0.05 (1.3237839×10^{-6} for OJ, 4.6815774×10^{-8} for VC), so we reject the null hypotheses: an higher dose has a significantly higher mean, for both supplements.

Let's now test if one supplement is better than the other.

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=supp)) + ggtitle("OJ vs VC - overal (all dosages)")
```

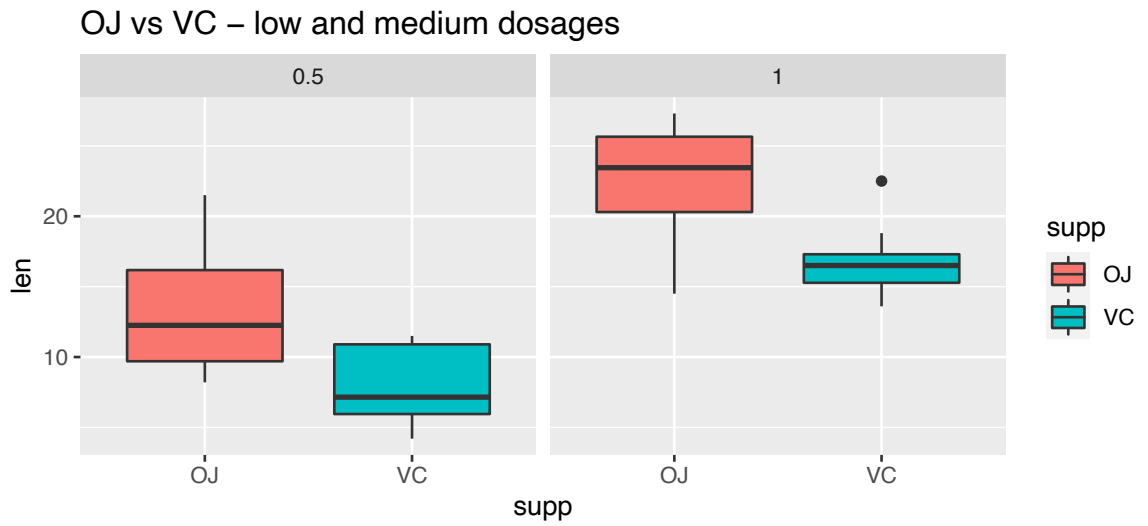


```r
t <- t.test(len ~ supp, data = ToothGrowth)
```

The p-value is 0.0606345, so we fail to reject the null hypothesis, meaning that there is no significant different between the mean of the two supplements.

However, looking at the first boxplot, it seems that OJ has a greater effect on tooth length at a low (0.5) and medium (1) dosage. Let's test it.

```r
ggplot(aes(x=supp, y=len), data=ToothGrowth[ToothGrowth$dose %in% c(0.5,1),]) +
  geom_boxplot(aes(fill=supp)) + facet_wrap("dose") +
  ggtitle("OJ vs VC - low and medium dosages")
```




```r
# "OJ vs VC - low dosage"
a <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 0.5,])
# "OJ vs VC - medium dosage"
b <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 1,])
```

For both tests, the p-value is smaller than 0.05 (0.0063586 for low dose, 0.0010384 for medium dose) so we reject the null hypotheses and we can confirm that at the same dose the supplement OJ has a greater effect on tooth growth.

Q4: Conclusions

1. Overall, there is no difference in supplement type on tooth growth
2. However, when looking at individual dosages, OJ is better than VJ in stimulating tooth growth at low (0.5) and medium (1) doses.
3. For each supplement type, the high dose is better than the low dose in stimulating tooth growth.

Assumptions
For each t test performed above, variances are assumed to be different.