Statistical Interference Course Project. Part 2

Daria Kravets

Task requirments

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.

Analysis

First of all, we need to load the data, and we can look at few rows of dataset:

```
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
```

We can see that there are 3 columns:

`len` numeric Tooth length

`supp` factor Supplement type (VC or OJ)

`dose` numeric Dose in milligrams/day

Basic summary

```
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
```

Let's look at some characteristics of Length column:

```
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

```
sd(ToothGrowth$len)
```

```
## [1] 7.649315
```

```
var(ToothGrowth$len)
```

```
## [1] 58.51202
```

Correlation. Step 1

Let's check if there is any correlation between tooth length and dose:

```
cor.test(ToothGrowth$len, ToothGrowth$dose)
```

```
##
##  Pearson's product-moment correlation
##
## data:  ToothGrowth$len and ToothGrowth$dose
## t = 10.25, df = 58, p-value = 1.233e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6892521 0.8777169
## sample estimates:
##       cor
## 0.8026913
```

As we can see, it is, and it's big. What about supplement type? We want to find if it has impact on length too.

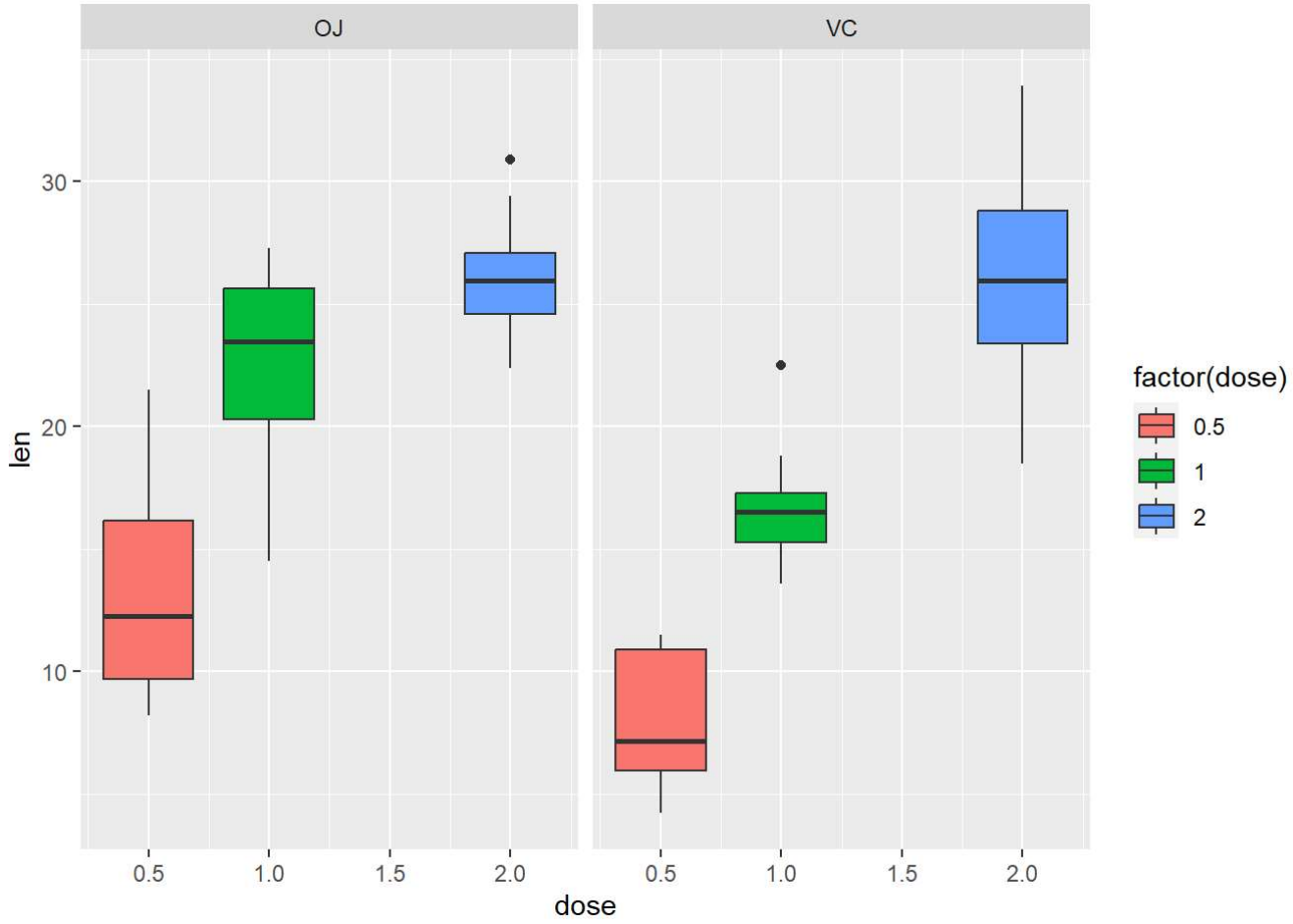
Plots

Let's take a look at the box plots:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
ggplot(ToothGrowth, aes(x = dose, y = len, fill = factor(dose))) +
  geom_boxplot(position = position_dodge()) +
  facet_wrap(~supp)
```



We can see there is a difference between how these two types of supplement affecting tooth growth.

95% confidence intervals

```
mean(ToothGrowth$len) + c(-1, 1) * qnorm(0.975) * sd(ToothGrowth$len) / sqrt(length(ToothGrow
th$len))
```

```
## [1] 16.87783 20.74884
```

T-tests

```
t.test(len~supp, 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 between group OJ and group VC is not equa
l to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

T-test shows that we cannot reject null hypothesis that the difference in group means is zero (in our case that supplement type doesn't impact tooth length).

Let’s divide our dataset into three different sets by dosage:

```
d1 <- subset(ToothGrowth, dose == 0.5)
d2 <- subset(ToothGrowth, dose == 1)
d3 <- subset(ToothGrowth, dose == 2)
```

And then we will repeat our T-test for all three groups:

```
t.test(len~supp, data = d1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and group VC is not equa
l to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len~supp, data = d2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and group VC is not equa
l to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len~supp, data = d3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means between group OJ and group VC is not equa
l to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

And here are our final results: for dosage 0.5 mg/day and 1 mg/day we can reject null hypothesis, and for dosage 2 mg/day we cannot reject null hypothesis.

Conclusions

- Tooth growth is not completely affected by supplement type.
- If pig is receiving vitamin C in 0.5 mg/day or 1 mg/day, there will be difference of having it as orange juice or ascorbic acid. As shown on plots, orange juice will have better effect.
- If dose is 2 mg/day, we can state that there is no big difference in received supplement type.