Statiscal Inference Final Project

Maria Cristina Alameda Salas

25/4/2023

Part 2: Basic Inferential Data Analysis

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

Loading the ToothGrowth data

```
# load "tidyverse" plotting package
library(ggplot2)



# Load data
data(ToothGrowth)
```

Performing some basic exploratory data analyses

In the following code, we can see that 3 columns compose our dataset: len, supp & dose.

```
# Show first rows
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
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

Provide a basic summary of the data.

This data contains 3 columns. The first column (*len*) talks about the length of tooth of guinea pigs. It is a continious value in range 4.20 - 33.90 with a mean of 18.81. The columns *supp* is about the supplement took by the animal. This column is discrete, with just two possible values (OJ & VC).

The last column (*dose*) shows data about dosage of supplement in mg. It goes from 0.5 mg to 2 mg.

```
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 plot the previous information by *supp* group.

```
qplot(factor(supp), len, data = ToothGrowth, facets=~dose,
      main="Tooth growth of guinea pigs by supplement type and dosage (mg)",
      xlab="Supplement type", ylab="Tooth length") +
      geom_boxplot(aes(fill = supp)) +
      scale_fill_manual(values=c("#f9b4ab", "#679186"))
```