Statistical Inference Course Project Part 2

Author: Brynjólfur Gauti Jónsson

Synopsis

The purpose of this assignment is to analyze the ToothGrowth dataset. The dataset reports the growth of odontoblasts (cells responsible for tooth growth) in guinea pigs as a response to different dosages of vitamin c, administered as orange juice or as ascorbic acid (a vitamin compound).

Data Summary

First let us load the data set and take a quick look at it.

```
library(datasets)
data(ToothGrowth)
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 ...
```

```
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
```

We see that the *dose* variable is stored as a numeric, but we will change it into a factor variable. We will also make the names of the *supp* levels more descriptive.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
levels(ToothGrowth$supp) <- c('Orange Juice', 'Ascorbic Acid')
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##                 0.5  1  2
##   Orange Juice   10 10 10
##   Ascorbic Acid  10 10 10
```

So the sample is split into six different groups of size ten. Our aim is much clearer now that we have gotten a quick look at the data.

Analysis

Let's start the analysis by looking at a boxplot of the data.

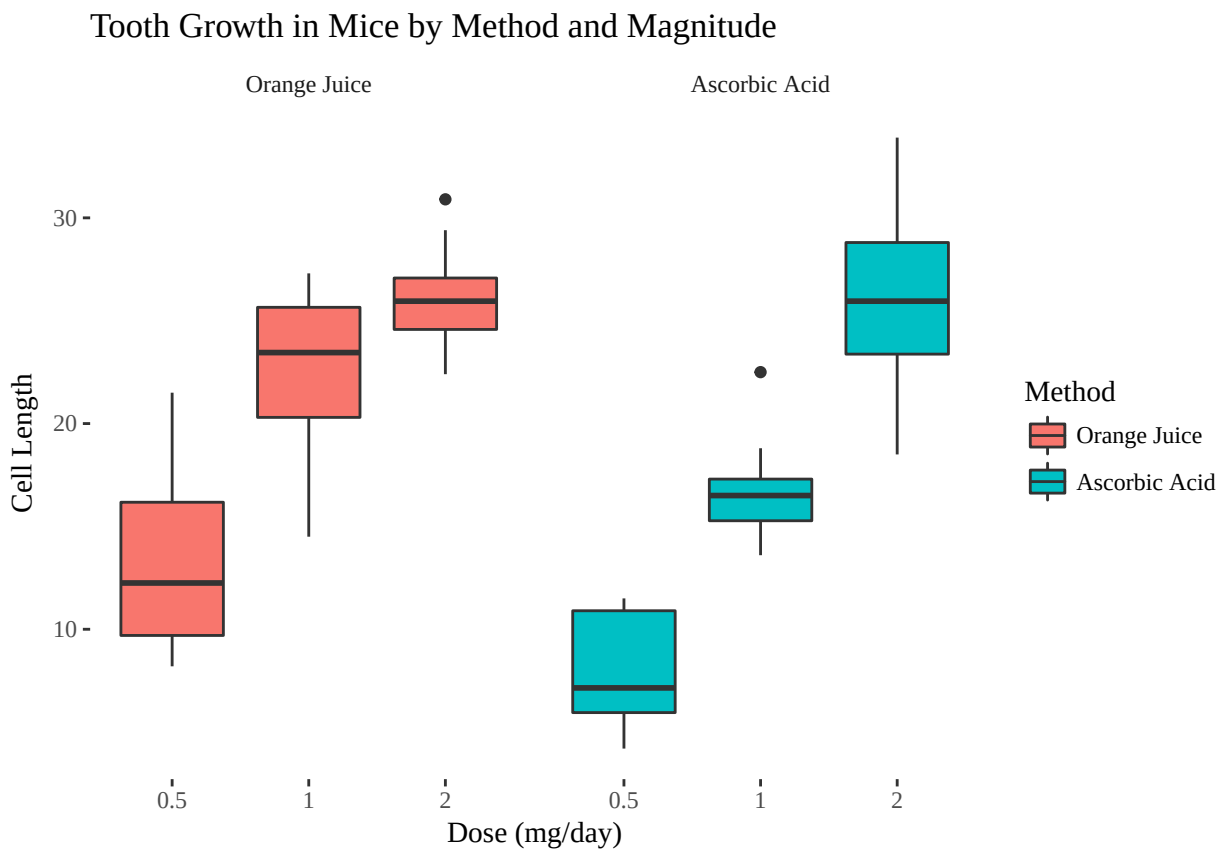
```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(ggthemes)
```

```
## Loading required package: ggthemes
```

```
g <- ggplot(ToothGrowth, aes(y=len, x=dose)) + theme_tufte()
g + geom_boxplot(aes(fill = supp)) + facet_grid(.~supp) + xlab('Dose (mg/day)') +
    ylab('Cell Length') + labs(fill = 'Method') +
    scale_color_manual(labels = c('Orange Juice', 'Ascorbic Acid')) +
    ggtitle('Tooth Growth in Mice by Method and Magnitude')
```



We see a clear difference in dose magnitude across both methods. It also looks like orange juice has a more pronounced effect when the dosage is small but the difference evens out in the largest dose.

Now we will perform t-tests to back our claims.

Testing differences in Orange Juice vs. Ascorbic Acid

We will print out three t-tests. One t-test comparing administration methods for each dosage.

```
for (i in c(0.5, 1, 2)) {
    print(paste('T Test of Difference Between Administration Methods for Doses of Size',
          i))
    print(t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose==i,]))
}
```

```
## [1] "T Test of Difference Between Administration Methods for Doses of Size 0.5"
##
##  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 is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
##  mean in group Orange Juice mean in group Ascorbic Acid
##                       13.23                        7.98
##
## [1] "T Test of Difference Between Administration Methods for Doses of Size 1"
##
##  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 is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
##  mean in group Orange Juice mean in group Ascorbic Acid
##                       22.70                       16.77
##
## [1] "T Test of Difference Between Administration Methods for Doses of Size 2"
##
##  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 is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
##  mean in group Orange Juice mean in group Ascorbic Acid
##                       26.06                       26.14
```

As we thought, the administration method seems to matter when the dose is small but had almost no effect when the dose is large. For doses of size 0.5mg and 1mg we can reject the null-hypotheses that method of administration has no effect.

Does the Size of the Dose Have an Effect?

```
for (i in c(1, 2, 3)) {
    sizes <- c(0.5, 1, 2)
    test <- t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose!=sizes[i],])
    sizes <- sizes[-i]
    print(paste('T Test for dosage sizes:', sizes[1], 'and', sizes[2]))
    print(test)
}
```

```
## [1] "T Test for dosage sizes: 1 and 2"
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 1 mean in group 2
##          19.735          26.100
##
## [1] "T Test for dosage sizes: 0.5 and 2"
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 2
##            10.605            26.100
##
## [1] "T Test for dosage sizes: 0.5 and 1"
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 1
##            10.605            19.735
```

We get p-values indicating significant results across the board. None of the confidence intervals contain zero. We say with 99% probability that a dose of 2mg will yield more growth than doses of 0.5mg and 1mg

Results

Given the assumptions that

- our sample is representative of the population from which it was drawn,
- there was no bias in the measurements and
- there are no hidden unmeasured variables affecting our data

We have found out that there is a significant increase in odontoblast growth rates with higher doses up to 2mg and for lower doses, administration of orange juice had a more pronounced effect than ascorbic acid