Inferencial Statistics Project, Part 2

Irina White

18/04/2021

Part 2: Basic Inferential Data Analysis

The goal is to analyze the ToothGrowth data in the R datasets package. *Overview:* The ToothGrowth data set measures the growth of teeth [their length] in guinea pigs based on vitamin C given with orange juice or ascorbic acid. The Vitamin C dosage varies throughout the study.

The analysis is set to test two hypotheses: - The effect if the supplement on the tooth grow - The effect of the dosage on the tooth grow

```r
head(ToothGrowth,3)
```

```
##     len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

Basic summary of the data.

```r
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
```

Overview: 60 cases in total, 30 with orange juice (OJ) and 30 with ascorbic acid (VC). The min dosage is 0.5 and max of 2. The observed growth of the teeth ranges between 4.2 and 33.9mm.

```r
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
ggplot(data=ToothGrowth, aes(dose, len, group=dose))+
  geom_boxplot(aes(fill=dose))+
  facet_grid(.~supp)+
  ggtitle('Tooth Growth by Dosage with different supplements', )+
  ylab('Tooth Length')+
  xlab('Dose, mg')+
  theme(plot.title=element_text(size=23), axis.text = element_text(size=16),
        strip.text.x = element_text(size=20))
```

Tooth Growth by Dosage with different supplements
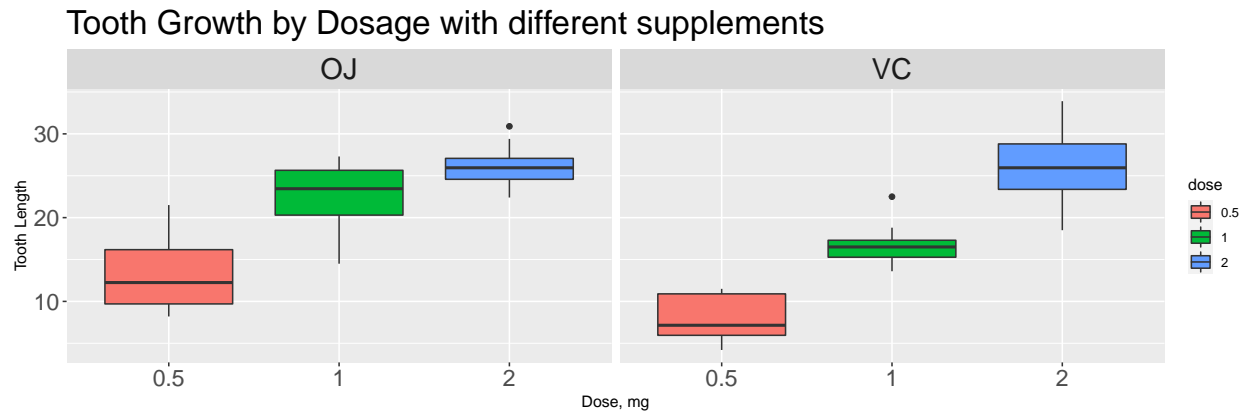
As it can be seen from the graph on average the higher dosage results in a better tooth grow. However, the effect of the supplements on the tooth growth is not as pronounced. Also there are not many outliers, except one in VC for 1mg dose and one in OJ for 2mg dose.

Basic exploratory data analyses

X0.5	X1	X2	OJ	VC
Min. : 4.200	Min. :13.60	Min. :18.50	Min. : 8.20	Min. : 4.20
1st Qu.: 7.225	1st Qu.:16.25	1st Qu.:23.52	1st Qu.:15.53	1st Qu.:11.20
Median : 9.850	Median :19.25	Median :25.95	Median :22.70	Median :16.50
Mean :10.605	Mean :19.73	Mean :26.10	Mean :20.66	Mean :16.96
3rd Qu.:12.250	3rd Qu.:23.38	3rd Qu.:27.82	3rd Qu.:25.73	3rd Qu.:23.10
Max. :21.500	Max. :27.30	Max. :33.90	Max. :30.90	Max. :33.90

Basic data already shows significant difference in means and medians between the dosage groups, and some difference between the mean and median of the OJ vs VC group. Further investigation is required to confirm the significance of these differences.

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

Supplements First is the hypothesis test on a two independent samples of supplements

```
t.test(data_supp$OJ,data_supp$VC, var.equal = FALSE)[3:4]
```

```
## $p.value
## [1] 0.06063451
##
## $conf.int
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

The p value is 0.06, which is higher than alpha (0.05). The Confidence intervals include 0. Therefore, there is no any statistical difference for any supplement.

Dose First is the hypothesis test on set of two independent samples of supplements, total 3 pairs

```
s1<-t.test(data_dose$X0.5,data_dose$X1, var.equal = FALSE)[3:4]
s2<-t.test(data_dose$X0.5,data_dose$X2, var.equal = FALSE)[3:4]
s3<-t.test(data_dose$X1,data_dose$X2, var.equal = FALSE)[3:4]
p_value<-list(s1$p.value,s2$p.value,s3$p.value)
conf_int<-list(round(s1$conf.int,2),round(s2$conf.int, 2),round(s3$conf.int,2))
p_value<-as.data.frame(p_value, col.names = c('mg0.5vs1', 'mg0.5vs2', 'mg1vs2'))
conf_int<-as.data.frame(conf_int, col.names = c('mg0.5vs1', 'mg0.5vs2', 'mg1vs2'), row.names = c('lower
t1<-kable(list(p_value, conf_int), caption='P values and confidence intervals for dosage')
kable_styling(t1, latex_options = 'hold_position')
```

Table 1: P values and confidence intervals for dosage

mg0.5vs1	mg0.5vs2	mg1vs2		mg0.5vs1	mg0.5vs2	mg1vs2
1e-07	0	1.91e-05	lower limit	-11.98	-18.16	-9.00
			upper limit	-6.28	-12.83	-3.73

All p-values are extremely small, meanwhile t-test confidence intervals do not include zero and all show significant change as the dosage change. Especially in the pair 0.5mg vs 2mg.

Conclusions and the assumptions.

Basic inferential statistics has been applied during this analysis to investigate the effect that various dosages of vitamin C (0.5mg, 1mg, 2mg) per day would have on the guinea pigs tooth growth rate. Similarly, the same approach has been used to analyze if there is any affect of supplements (Orange juice and Ascorbic Acid) that contain the Vitamin C dosage on the tooth growth. Due to the small number of the sample the two sample two-sided t-test (95% confidence level) has been used for the analysis. **Key assumption:** the sample contain the independent population.

Conclusion: * There is no significant difference between supplement usage. * There is a significant change influenced by the various dosage values. The higher the dosage - the longer the tooth length.