Basic Inferential Data Analysis

The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Eddy Delta

01/09/2018

Basic summary of the data

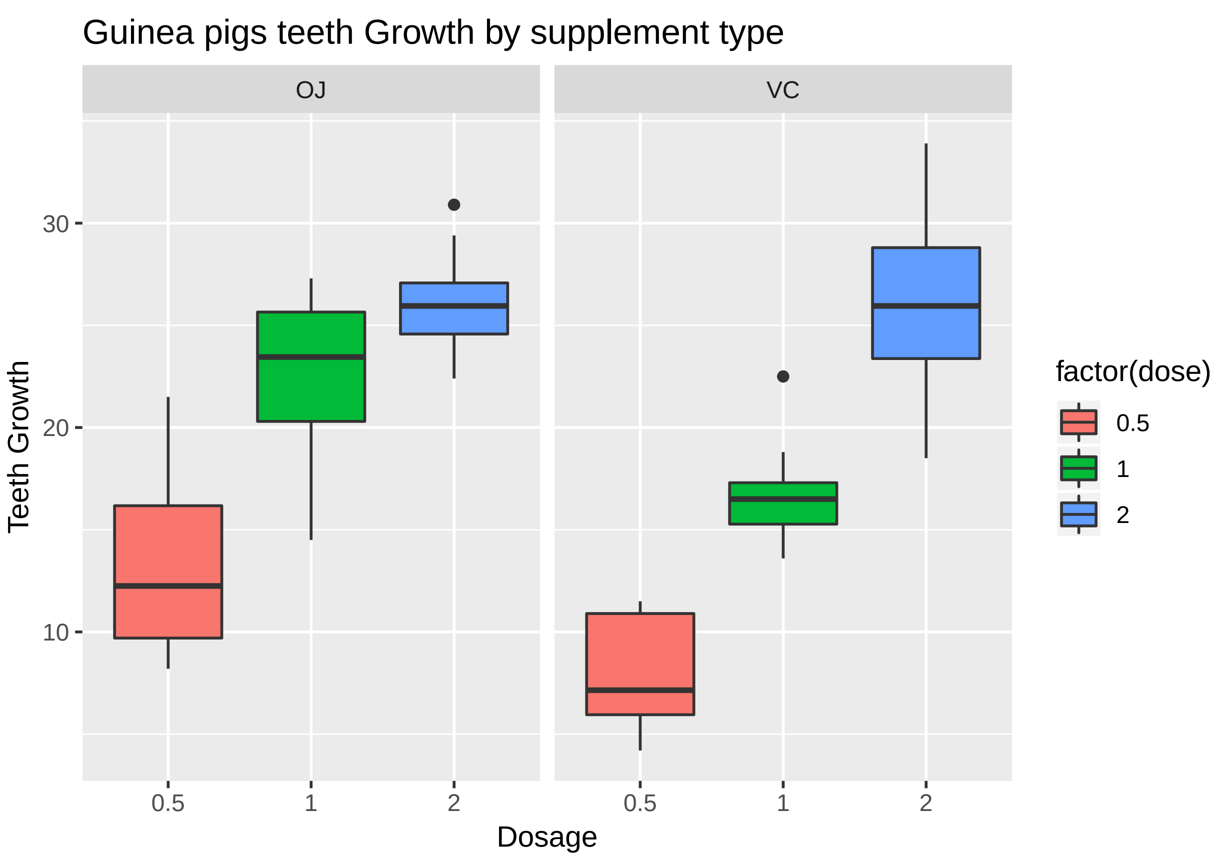
The ToothGrowth data in the R datasets package contains the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Overview

```
library(ggplot2)
library(datasets)
data(ToothGrowth)
```

Data distribution

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose))) +
  geom_boxplot() +
  facet_grid(.~supp) +
  xlab("Dosage") +
  ylab("Teeth Growth") +
  ggtitle("Guinea pigs teeth Growth by supplement type")
```



Quick look at the data contents

```r
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 ...
```

Some meaning full values about the data content

```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
```

The data shows a possible correlation between the teeth growth and the dosage on both supplement type.

Analysis

Dosage influence

To analyse the correlation between the teeth growth and the dosage, we'll consider the differents pairs of dosage. With $H_0 : corr(dosage_x, dosage_y) = 0$ and $H_A : corr(dosage_x, dosage_y) \neq 0$

0.5 vs 1.0

```r
t.test(len~dose, data=subset(ToothGrowth, dose %in% c(0.5, 1.0)), alternative = 'two.sided', paired=FALSE)
```

```
## 
##  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
```

0.5 vs 2.0

```r
t.test(len~dose, data=subset(ToothGrowth, dose %in% c(0.5, 2.0)), alternative = 'two.sided', paired=FALSE)
```

```
## 
##  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
```

1.0 vs 2.0

```r
t.test(len~dose, data=subset(ToothGrowth, dose %in% c(1.0, 2.0)), alternative = 'two.sided', paired=FALSE)
```

```
##
##  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
```

0 is out of the 95% confidence interval in every cases so H_0 is rejected, which makes us conclude with 95% of confidence that there is a correlation between teeth growth and Vitamin C dose.

Delivery methods influence

Let's analyse the correlation between the teeth growth and the delivery methods. With $H_0 : corr(delivery_{OJ}, delivery_{VC}) = 0$ and $H_A : corr(delivery_{OJ}, delivery_{VC}) \neq 0$

```r
t.test(len ~ supp, paired = F, var.equal = F, 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 is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The H_0 can't be rejected with the 95% of confidense because the 95% confidence interval contains 0. We then can't conclude with 95% of confidence that there is a correlation between the teeth growth and the delivery method. However the min value is pretty close to 0, so it's a fair to think that under a little smaller confidence interval it can fall out of the confidence interval.

```r
t.test(len ~ supp, paired=F, var.equal=F, data=ToothGrowth, alt='two.sided', conf.level=.939)$conf.int
```

```
## [1] 0.005406982 7.394593018
## attr(,"conf.level")
## [1] 0.939
```

H_0 can be rejected under 93.9% confidence because at this confidence level the confidence interval doesn't contains 0. Which means that with 93.9% of confidence it's possible to consider a correlation between teeth growth and delivery method.

Conclusion

We can assume assume with 95% of confidence the the teeth growth is positively correlated to the teeth growth. However there is no enough data to conclude any correlation between the teeth growth and the delivery method with 95% confidence. It's interesting to take note about this last point that under 93.9% confidence it's possible to fairly consider a correlation between teeth growth and delivery method.