Statistical inference: Assignment part2

Gurunath Gandikota

September 17, 2018

Synopsis

The current study summarizes experiments carried out on guinea pigs to understand the impact of supplements on odontoblasts, the cells responsible for teeth growth.Three levels of dosages of vitamin C and orange juice are used as supplements and length of teeth are monitored. The results show that for lower levels of dosages, orange juice is more effective than vitamin C. However for higher dosages, the impact is more or less the same.

Load tooth growth data and carry out initial exploratory analysis

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
# Loading the data
Tooth_growth <- ToothGrowth
Tooth_growth$dose <- as.numeric(Tooth_growth$dose) # convert to factor
Tooth_growth$len <- as.numeric(Tooth_growth$len)

# Looking at the details of the data
str(Tooth_growth)
```

```
## '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 ...
```

```
unique(Tooth_growth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
summary(Tooth_growth)
```

```
##       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
```

It can be seen that the data set has 60 results. It has three columns of data: length (length of tooth), dose (dosage level) and supp (type of supplement used). The supplements used are

OJ: orange juice and VC: vitamin C. The unique dosage levels in the data set are: 0.5, 0.1 and 2. Summary of the data set shows that the length has a median value of 19.25 and dosage has a median value of 1. Tooth length values vary between 4.2 and 33.9 indicating that the suplements have a very large impact.

Now plotting tooth length vs dosage for the two supplements:

```
g <- ggplot(Tooth_growth, aes(x=factor(dose),y=len,fill=factor(dose)))
g + geom_boxplot(notch=F) + facet_grid(.~supp) +
  xlab("Dosages (Milligram)") + ylab("Tooth length") + ggtitle("Tooth Growth
as a function of dosage and supplement")
```
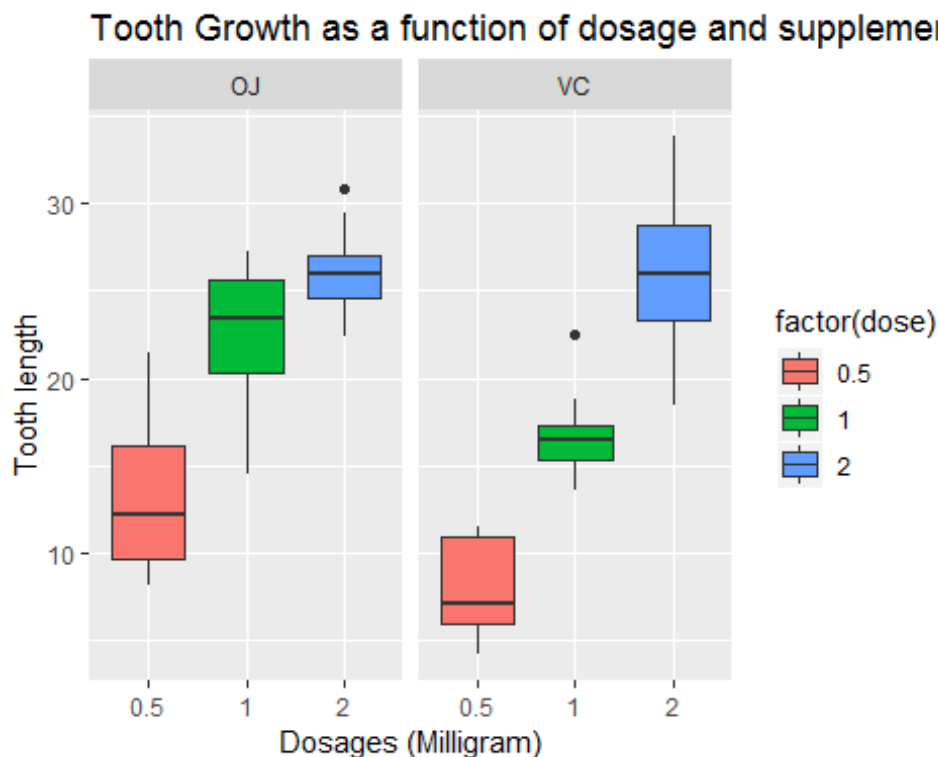

It can be seen from the box plot that the growth rate is better for orange juice for lower dosage levels. However, there seems to be a lot of variability in the data for orange juice at smaller dosage levels and for vitamin C at higher dosage levels.

Calculating confidence intervals using T tests and understand the p value

For the T-tests, our null hypothesis is that the difference in means of the tooth growth for both orange juice and vitamin C is 0.

```
conf_t <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
```

The two sample T test indicates that the 95% confidence interval, is -0.167 to 7.567. The T quantile, which is 1.9153 falls in the confidence interval, and since the p-value is 0.06, we fail to reject the null hypothesis.

Conclusion

```
Summary_TG <- aggregate(len~dose+supp,Tooth_growth,mean)
ggplot(Summary_TG, aes(x=dose,y=len,colour=supp)) +
  geom_line() + xlab("Dosage")+ ylab("Tooth length") + geom_point(aes(shape =
supp, size=dose))
```
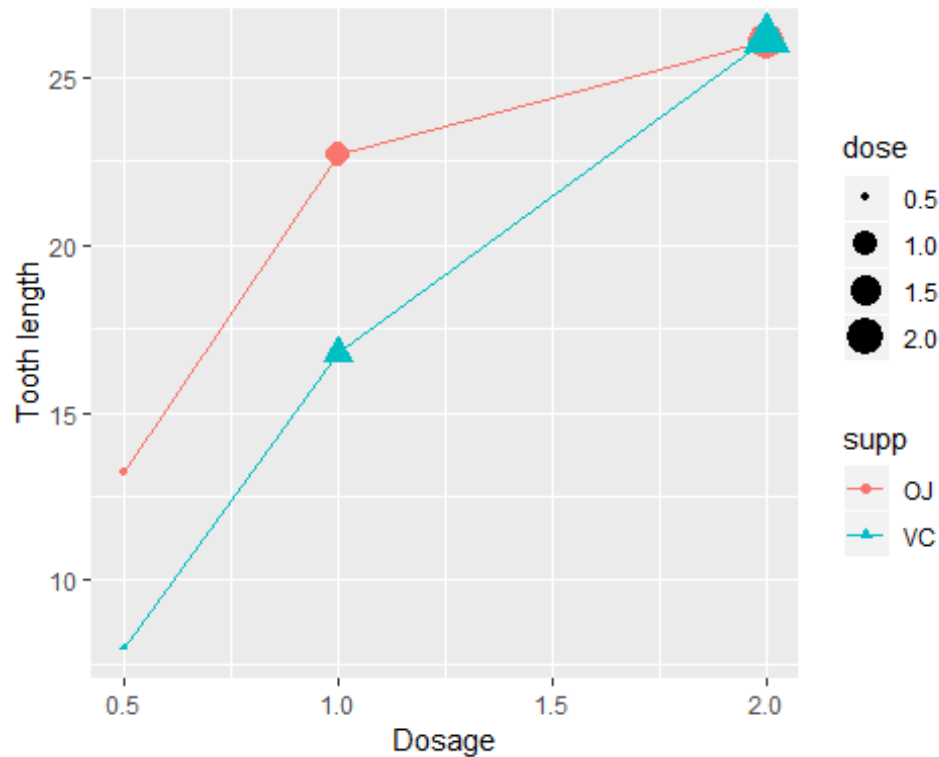

It can be seen from the summary plot that oranje juice is much more effective than vitamin C.