

Analysis of ToothGrowth data in the R datasets package

Carlos Hernández

25/9/2020

Contents

ToothGrowth Dataset	1
Basic exploratory data analysis	1
Hypothesis Testing	2
Assumptions	2
Hypothesis 1: Variation of tooth length when using OJ or VC	2
Hypothesis 2: Variation of tooth length when using different doses	3
Conclusion	4

ToothGrowth Dataset

ToothGrowth data set contains the result from an experiment studying the effect of vitamin C on 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).

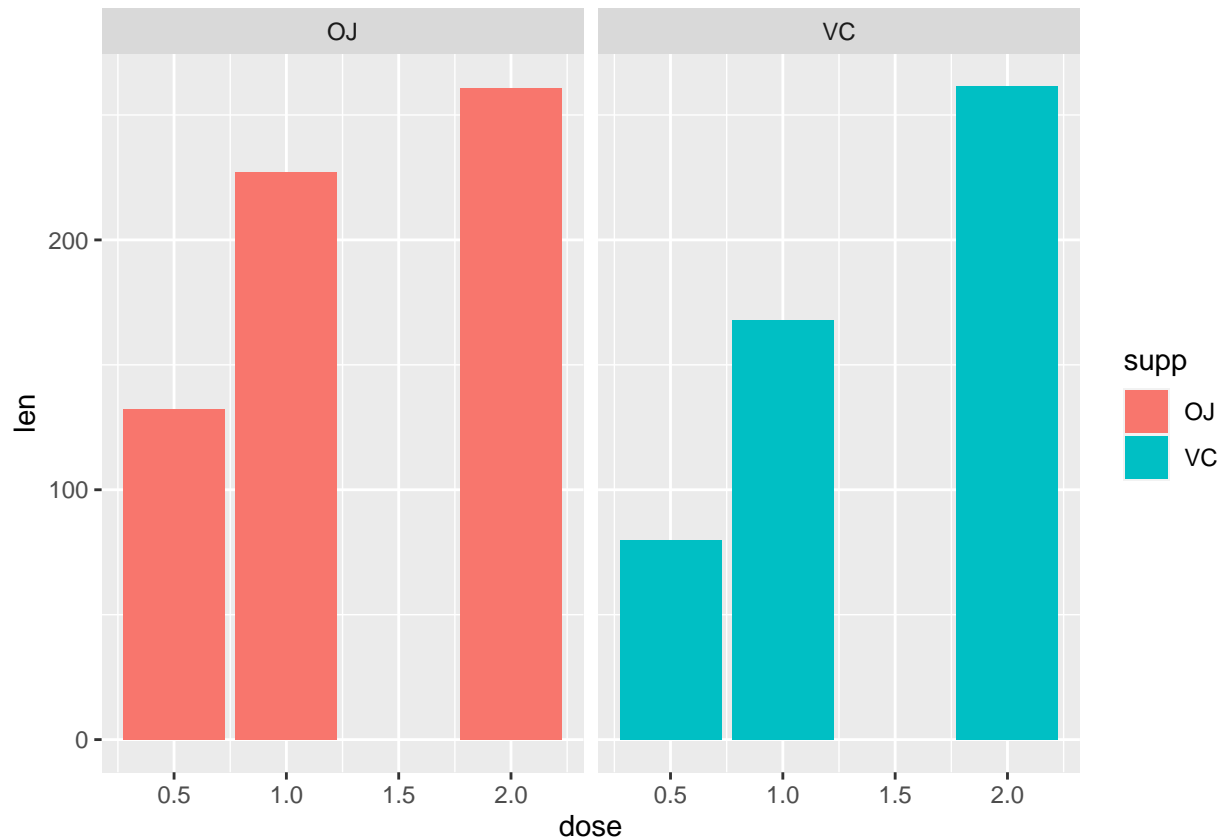
```
library(kableExtra)
head(ToothGrowth,5) %>%
  kbl() %>%
  kable_material(c("striped", "hover"))
```

len	supp	dose
4.2	VC	0.5
11.5	VC	0.5
7.3	VC	0.5
5.8	VC	0.5
6.4	VC	0.5

1. len: Tooth length
2. supp: Supplement type (VC or OJ).
3. dose: numeric Dose in milligrams/day

Basic exploratory data analysis

```
library(ggplot2)
ggplot(ToothGrowth, aes(x = dose, y = len, fill = supp)) +
  geom_col() +
  facet_grid(~supp, scales = "free")
```



At first glance, it seems that the dose given through orange juice is more effective, since greater growth is observed in the teeth when the dose is administered via orange juice and less when it is administered with ascorbic acid. We could also notice that when doses of 2mg are administered, it seems that the growth is the same regardless of which medium is administered.

But these are only initial guesses that we can verify or reject by performing a hypothesis test.

Hypothesis Testing

Assumptions

- The variables must be independent and identically distributed (i.i.d.).
- Variances of tooth growth are different when using different supplement and dosage.
- Tooth growth follows a normal distribution.

Hypothesis 1: Variation of tooth length when using OJ or VC

Our null hypothesis is that the length of the tooth does not vary when we use either of the two methods (VC or OJ).

Therefore, our alternative hypothesis would be that tooth length varies depending on the method through which the dose is delivered.

```
oj_len <- ToothGrowth[ToothGrowth$supp=="OJ",]$len
vc_len <- ToothGrowth[ToothGrowth$supp=="VC",]$len

t.test(oj_len,vc_len, paired = FALSE, var.equal = FALSE, alternative = "greater")

##
## Welch Two Sample t-test
##
## data:  oj_len and vc_len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687      Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

As we can see our p-value is greater than 0.05, therefore our null hypothesis is rejected and we accept that the length of the tooth varies according to the method used.

Furthermore, we can see that on average if we use OJ the tooth length is greater than using VC.

Hypothesis 2: Variation of tooth length when using different doses

Our null hypothesis is that tooth length does not vary between methods when we use different doses.

Therefore, our alternative hypothesis would be that the length of the teeth varies according to the method and dose delivered.

```
OJDoseHalf <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose==0.5,]$len
OJDoseOne <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose==1.0,]$len
OJDoseTwo <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose==2.0,]$len

VCDoseHalf <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose==0.5,]$len
VCDoseOne <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose==1.0,]$len
VCDoseTwo <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose==2.0,]$len
```

For dose equal to 0.5 mg:

```
t.test(OJDoseHalf, VCDoseHalf, paired = FALSE, var.equal = FALSE, alternative = "greater")

##
## Welch Two Sample t-test
##
## data:  OJDoseHalf and VCDoseHalf
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604      Inf
```

```
## sample estimates:
## mean of x mean of y
##      13.23      7.98
```

For dose equal to 1 mg:

```
t.test(OJDoseOne,VCDoseOne, paired = FALSE, var.equal = FALSE, alternative = "greater")
```

```
##
## Welch Two Sample t-test
##
## data: OJDoseOne and VCDoseOne
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##      22.70      16.77
```

For dose equal to 2 mg:

```
t.test(OJDoseTwo, VCDoseTwo, paired = FALSE, var.equal = FALSE, alternative = "greater")
```

```
##
## Welch Two Sample t-test
##
## data: OJDoseTwo and VCDoseTwo
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335      Inf
## sample estimates:
## mean of x mean of y
##      26.06      26.14
```

As we can see, for doses of 0.5 mg and 1 mg we obtained results similar to that of our hypothesis 1. In both cases the p-value is less than 0.5, therefore we can reject the null hypothesis and accept that the logitud of the teeth It varies according to the dose and greater lengths are obtained with doses of 1 mg being administered with OJ.

However, for doses of 2 mg, we obtain a p-value greater than 0.5, which we can interpret in that we must accept the null hypothesis. This means that regardless of the method used (VC or OJ) the length of the teeth obtained is the same for a dose of 2 mg.

Conclusion

As a conclusion we can say that after conducting this brief but interesting analysis, we have shown that for doses of 0.5 mg and 1 mg, orange juice results in greater tooth length. However for doses of 2mg, the length of teeth obtained will be the same regardless of whether OJ or VC is used.