Analysis of ToothGrowth data

Stefano Masneri

16 April 2016

Loading the ToothGrowth dataset

In this project we will analyze the dataset “ToothGrowth”. The first thing to do is then some exploratory data analysis.

```
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 ...
```

The data contains a set of 60 observations about the length of the tooth of 10 guinea pigs using different amounts of vitamin C and two differents delivery methods. For more info check https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html)

Explore the dataset

It is clear that there are just two types of supplments (column “supp”), that is orange juice (“OJ”) and vitamin C (“VC”), while there are 3 different types of dosage:

```
unique(ToothGrowth$dose)
```

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

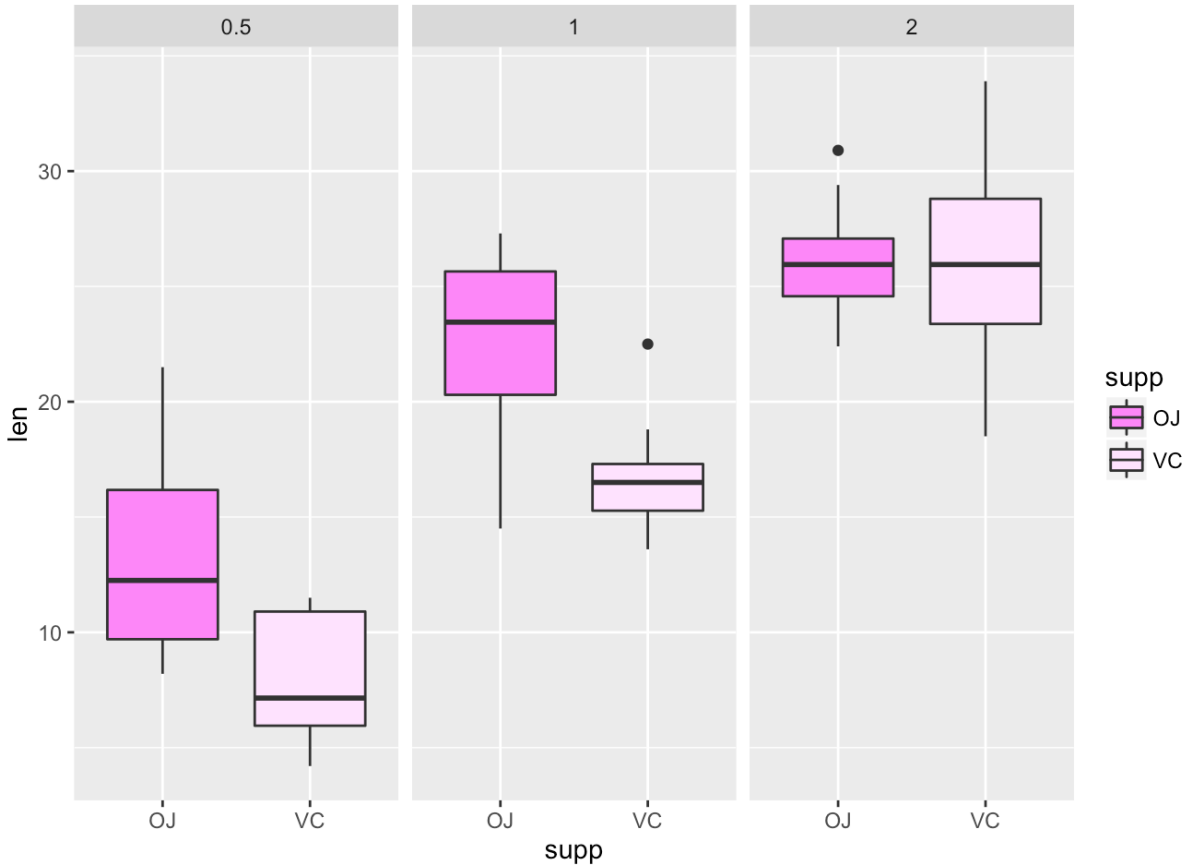
We want to get an idea of how the dosage and the supplement influence the length of the tooth.

```
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
```

This can be also be done with graphical inspection of the data:

```
ggplot(aes(x = supp, y = len, fill = supp), data = ToothGrowth) +
  geom_boxplot() +
  scale_fill_manual(values = c('orchid1', 'thistle1')) +
  facet_wrap(~ dose)
```



It seems clear that the tooth length is highly influenced by the dosage. Regarding the supplement used, it seems like oranje juice is more effective then vitamin C for lower dosage, while for a dosage of 2 there seems to be little difference between the type of supplement used. The variability of the length varies a lot across supplement and dosage, so not much can be said about that.

Confidence Intervals

We will now run t-tests to verify that indeed the amount of dosage influences the tooth length. Since we have three different amounts of dosage, we will consider all the pairs of possible dosage.

```
dose_0.5_1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose_0.5_2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose_1_2   <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

And then we just run the t.test three times. We specify that the varoables are not paired and the variances are different. Without additional information, this is the most general case. So our null

hypotheses is that the dosage does not influence the length of the tooth or, equivalently, that the mean length of the tooth with dosage A is the same as with dosage B. The alternative hypothesis is that the mean is different.

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose_0.5_1)$con
f.int
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose_0.5_2)$con
f.int
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose_1_2)$conf.
int
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Since none of the confidence intervals calculated contains zero, we can reject the null hypothesis for each test we ran. So for this dataset the higher the dosage, the bigger the tooth growth.

We also want to check whether the type of supplement influences the tooth length. Specifically, our null hypothesis is that the mean length is the same when both for “OJ” and “VC”, while the alternative hypothesis is that it is bigger when using “OJ” as supplement

```
t.test(len ~ supp, alternative = "greater", paired = FALSE, var.equal = FALS
E, data = ToothGrowth)$conf.int
```

```
## [1] 0.4682687       Inf
## attr(,"conf.level")
## [1] 0.95
```

The confidence interval does not contain the zero, so we can reject the null hypothesis in this case as well.

Finally, we want to check the effects of the supplement only when using a dosage of 2, since in that case the boxplot was showing very similar data for both OJ and VC. We will run a two sided test in this case, so the alternative hypothesis is that the two means are different.

```
t.test(len ~ supp, alternative = "two.sided", paired = FALSE, var.equal = FA
LSE, data = dose_1_2)$conf.int
```

```
## [1] -0.3166175  6.1666175
## attr(,"conf.level")
## [1] 0.95
```

As expected, the interval does contain zero, hence we fail to reject the null hypothesis.

Conclusions

We performed data analysis on the ToothGrowth dataset. By visual inspection and by performing statistical tests we were able to determine that the dosage and the supplement used during the experiment does affect the tooth growth. We were also able to determine that, when higher dosages are used, the supplement is not that important to determine the growth. When running the tests, we assumed that the sample population used was independent, that each group had different variances and the population was not paired, that is different guinea pigs were used for each combination of dose and supplement.