assignment_P2

Marcelladane

18 August 2017

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

- Load the ToothGrowth data and perform some basic exploratory data analyses -Provide a basic summary of the data. -Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
- State your conclusions and the assumptions needed for your conclusions.

Start by loading the necessary packages, here, just one for plot.

```
packages <- c("datasets","ggplot2", "gridExtra", "GGally")
sapply(packages, require, character.only = TRUE, quietly = TRUE)
```

Now you can load the data

```
data(ToothGrowth)
toothGrowth <- ToothGrowth
toothGrowth$dose <- as.factor(toothGrowth$dose)
```

This dataset shows the effect of vitamin C in tooth growth of guinea pigs

Now lets take a look in how the data is organized in the table

```
summary(toothGrowth)
```

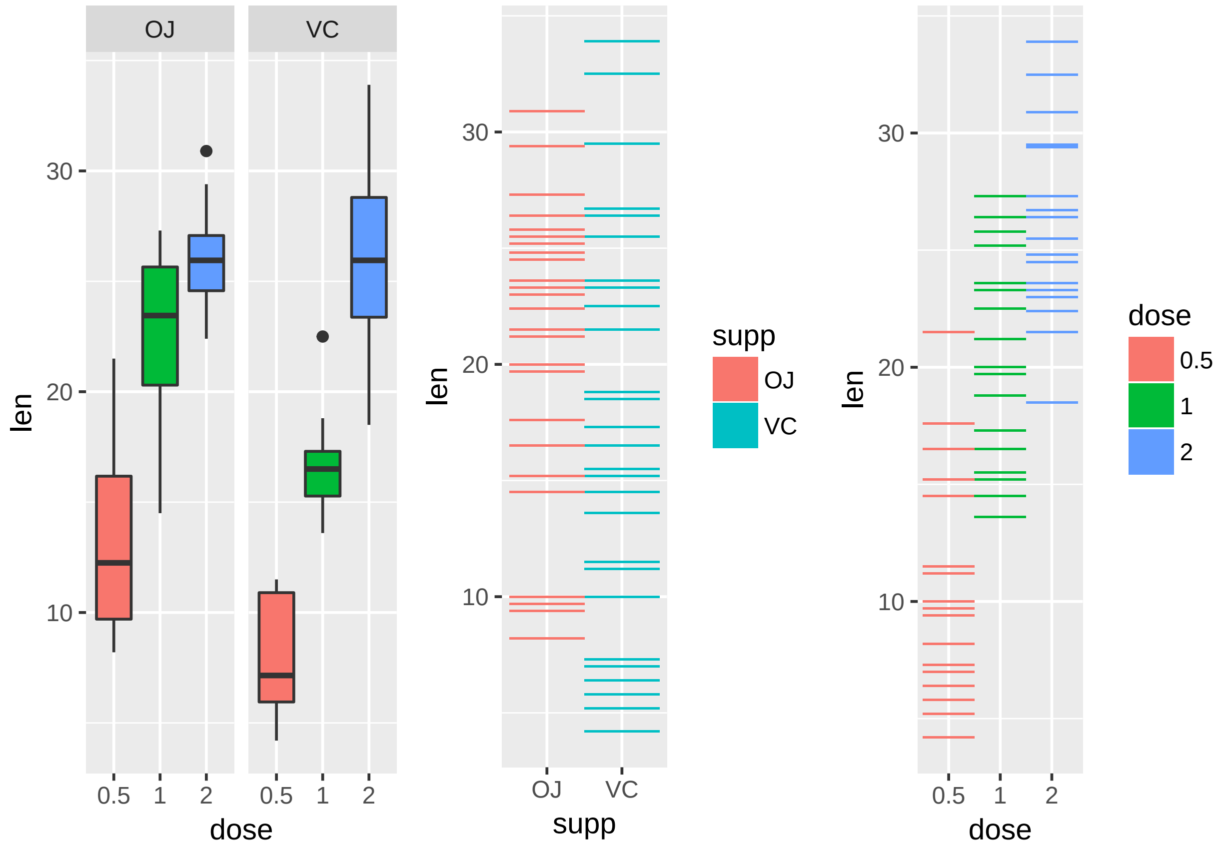
```
##       len         supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
head(toothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

We have basically 3 columns, one with the lenght of growth, one with the supplementation type and one with the dose.

One way to analyze the effects of vitamin C in the growth is basically plotting the effects of supplement and dose versus the lenght.



In the first plot you can observe that we have a dose response, for both treatments, OJ and VC. The other 2 plots show how disperse is the data. as you can see, the VC treatment has quite high effect in some but really low in ther treatments. If counting just the dose, in tern, increase the dose increase effect, independ of treatment.

Therefore we can conclude that dose has a strong effect and treatment might have as well. Not by itself, but by the union treatment + dose. To check this assumption, we can perform an ANOVA (Analysis of variance) analysis.

```
anova.analysis <- aov(len ~ supp * dose, data=toothGrowth)
summary(anova.analysis)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## supp         1  205.4   205.4  15.572 0.000231 ***
## dose         2 2426.4  1213.2  92.000  < 2e-16 ***
## supp:dose    2  108.3    54.2   4.107 0.021860 *
## Residuals   54  712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results showed actually a significant effect for treatment, also for dose (what we already noticed before) and finally a interaction between dose and treatment.

But, ANOVA dont tell you where the difference is. Just tells you that it has a difference, it could be in any of the doses, moreover, could be in any interaction of dose and treatment.In order to analyze this, we should perform a post-hoc test (choose Tukey because its not too conservative neither too permisive).

```
TukeyHSD(anova.analysis)
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
##
## Fit: aov(formula = len ~ supp * dose, data = toothGrowth)
##
## $supp
##       diff       lwr       upr     p adj
## VC-OJ -3.7 -5.579828 -1.820172 0.0002312
##
## $dose
##         diff       lwr       upr   p adj
## 1-0.5  9.130  6.362488 11.897512 0.0e+00
## 2-0.5 15.495 12.727488 18.262512 0.0e+00
## 2-1    6.365  3.597488  9.132512 2.7e-06
##
## $`supp:dose`
##                diff        lwr        upr     p adj
## VC:0.5-OJ:0.5 -5.25 -10.048124 -0.4518762 0.0242521
## OJ:1-OJ:0.5    9.47   4.671876 14.2681238 0.0000046
## VC:1-OJ:0.5    3.54  -1.258124  8.3381238 0.2640208
## OJ:2-OJ:0.5   12.83   8.031876 17.6281238 0.0000000
## VC:2-OJ:0.5   12.91   8.111876 17.7081238 0.0000000
## OJ:1-VC:0.5   14.72   9.921876 19.5181238 0.0000000
## VC:1-VC:0.5    8.79   3.991876 13.5881238 0.0000210
## OJ:2-VC:0.5   18.08  13.281876 22.8781238 0.0000000
## VC:2-VC:0.5   18.16  13.361876 22.9581238 0.0000000
## VC:1-OJ:1     -5.93 -10.728124 -1.1318762 0.0073930
## OJ:2-OJ:1      3.36  -1.438124  8.1581238 0.3187361
## VC:2-OJ:1      3.44  -1.358124  8.2381238 0.2936430
## OJ:2-VC:1      9.29   4.491876 14.0881238 0.0000069
## VC:2-VC:1      9.37   4.571876 14.1681238 0.0000058
## VC:2-OJ:2      0.08  -4.718124  4.8781238 1.0000000
```

Considering a p<0.5 as significant.

The results show a dose effect for the treatment VC. That means, as bigger is the dose, bigger the effect.

Finally we can check the confidence interval, as I final check of our conclusions.

```r
confint(anova.analysis)
```

```
##                   2.5 %    97.5 %
## (Intercept)  10.9276907 15.532309
## suppVC       -8.5059571 -1.994043
## dose1         6.2140429 12.725957
## dose2         9.5740429 16.085957
## suppVC:dose1 -5.2846186  3.924619
## suppVC:dose2  0.7253814  9.934619
```

Conclusions

As mentioned above, both treatments show some level of effect.

But treatment OJ have a plato of response at dose 1.

In other hand, treatment VC have the highest effect and present an increase effect as the doses increase.

Full code at: https://github.com/marcelladane/statistical_inference