

Statistical Inference: Inference project

(c) Marcos Gestal. March 2015.

Overview

This second part of the project will analyze the ToothGrowth data in the R datasets package and provide some conclusions about the growth by supp and dose.

Analysis

Load data and basic exploratory data analyses

ToothBrowth data represents the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
data(ToothGrowth)  # loading data
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 ...
```

See appendix 1 for a graphical view of the data with supplement and dose influencees.

Provide a basic summary of the data

```
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
```

```
aggregate(data = ToothGrowth, len ~ supp * dose , mean)
```

```
##  supp dose  len
## 1   OJ  0.5 13.23
## 2   VC  0.5  7.98
## 3   OJ  1.0 22.70
## 4   VC  1.0 16.77
## 5   OJ  2.0 26.06
## 6   VC  2.0 26.14
```

```
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##      OJ VC
## 0.5 10 10
## 1   10 10
## 2   10 10
```

The two supplement methods (with 3 different doses) are applied to the 10 pigs, so the next tests will use a paired approximation when it will be required.

Appendix 2 shows a boxplot of the data where the mean and variance per dose and per supplement can be graphically checked. As the boxplot shows, the length increases as the dose also increases from 0,5 to 2mg. The influence of the delivery methods (Orange Juice; Ascorbic Acid) is not so clear in this graph but it seems that Orange Juice has a better performance, so some tests will be required.

Statistical Inference

Influence of the Supplement type over the length increase

Firstly, we will compare the influence of the supplement type without taking into account the dose.

```
suppOJ <- ToothGrowth[which(ToothGrowth$supp=="OJ"), ]
suppVC <- ToothGrowth[which(ToothGrowth$supp=="VC"), ]
```

Appendix 3 shows the Confidence Interval obtained with the Wald method and with a one sample test for the median of the difference of length by the supplement used. However, the number of samples is small so T Test will be preferred. As we said earlier, supplements are applied to the same population of guinea pigs so a paired test with grouped data should be used. In this test, the Null Hypothesis will be that "There are NO differences between the effect of delivery method (orange juice or ascorbic acid)".

```
t.test(suppVC$len, suppOJ$len, paired = T)
```

```
##
## Paired t-test
##
## data:  suppVC$len and suppOJ$len
## t = -3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.991341 -1.408659
## sample estimates:
## mean of the differences
##                -3.7
```

Since the p-value is 0.00255 (p-value < 0.05) null hypothesis can be rejected, so we accept that *there are differences due to delivery method*. Furthermore, the estimated mean is -3.7, so another time the best supplement seems to be the Orange Juice.

Influence of the Supplement type over the length increase (by dose)

Once we stated that there is a significant difference between the supplements, we will test the influence of the supplement type depending on the dose value

```
dose05_supp0J <- supp0J[(supp0J$dose==0.5), ]
dose05_suppVC <- suppVC[(suppVC$dose==0.5), ]

t.test(dose05_suppVC$len, dose05_supp0J$len, paired = T)

##
## Paired t-test
##
## data: dose05_suppVC$len and dose05_supp0J$len
## t = -2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -9.236542 -1.263458
## sample estimates:
## mean of the differences
## -5.25
```

```
dose1_supp0J <- supp0J[(supp0J$dose==1), ]
dose1_suppVC <- suppVC[(suppVC$dose==1), ]

t.test(dose1_suppVC$len, dose1_supp0J$len, paired = T)

##
## Paired t-test
##
## data: dose1_suppVC$len and dose1_supp0J$len
## t = -3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -9.908089 -1.951911
## sample estimates:
## mean of the differences
## -5.93
```

```
dose2_supp0J <- supp0J[(supp0J$dose==2), ]
dose2_suppVC <- suppVC[(suppVC$dose==2), ]

t.test(dose2_suppVC$len, dose2_supp0J$len, paired = T)

##
## Paired t-test
##
## data: dose2_suppVC$len and dose2_supp0J$len
## t = 0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -4.168976 4.328976
```

```
## sample estimates:
## mean of the differences
##                0.08
```

According with p_values the two supplement methods are different with a 0.5mg and 1mg doses (best performance for Orange Juice) and equal with 2mg doses

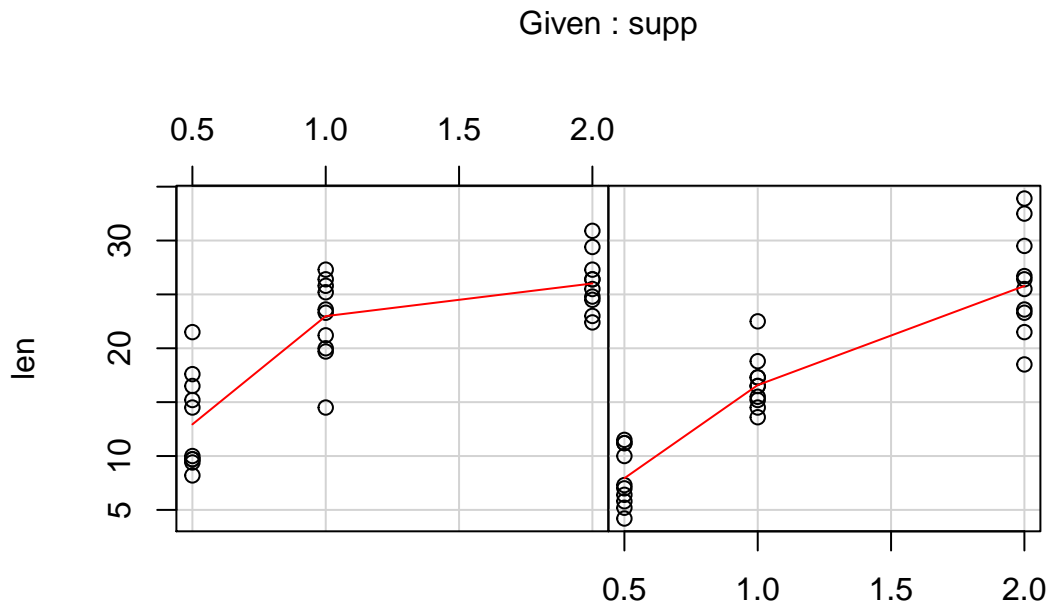
Conclusion

Confidence interval and hypothesis testing was performed (confidence interval for normal distribution was showed with educational and comparision purposes). The ToothGrowth dataset was 10 individuals. Each one of that pigs were used to test the different supplement types and the different doses for each one, so paired tests (dependent groups) were used (results are very similar without this assumption).

From these test, the following clonclusions can be extrated:

- The boxplot and scatterplot show how a increase of the dose implies an increase of the length for both supplements
- There are statistically significant differences between the type of supplement
- According Confidence Intervals, for 0.5mg and 1mg doses the best supplement is the Orange Juice. For 2mg doses, both methods achieve very similar results.

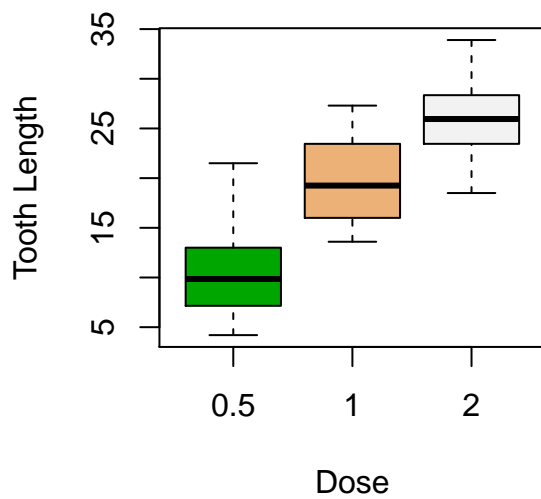
Appendix 1 : ToothGrowth data



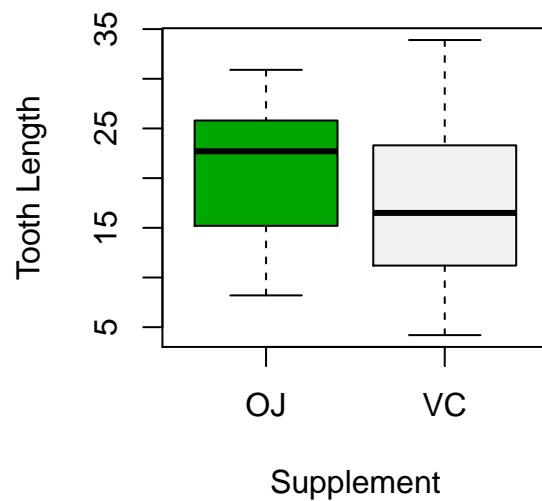
ToothGrowth data: length vs dose, given type of supplement (OJ & VC)

Appendix 2 : Boxplot

Boxplot of Tooth Growth Data



Boxplot of Tooth Growth Data



Appendix 3 : Walden approach to confidence interval for the mean of length

The variable `difLength` contains the differences in toothgrowth related with the supplement (a positive value means that VC/Ascorbic Acid offers a higher increase than Orange Juice)

```
difLength <- suppVC$len - suppOJ$len  
mean(difLength)
```

```
## [1] -3.7
```

```
mean(difLength)+c(-1,1)*qnorm(.975)*sd(difLength)/sqrt(length(difLength))
```

```
## [1] -5.895815 -1.504185
```

We can compare it with t test... and check how, as it is expected the average is equal, and the confidence interval is a little bit small in this case. Also, you can check how this test over `difLength` values is exactly equal to the paired test previously showed in the text.

```
t.test(difLength)
```

```
##  
## One Sample t-test  
##  
## data: difLength  
## t = -3.3026, df = 29, p-value = 0.00255  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -5.991341 -1.408659  
## sample estimates:  
## mean of x  
## -3.7
```

In both cases, the tests appoint that this difference between means should be negative, so Orange Juice should be considered as the best supplement