

Statistical inference – ToothGrowth data

Marija Zoldin

05.07.2015

Overview

The goal of this task was to load ToothGrowth data and apply learned knowledge from statistical inference class on Coursera. Tooth growth among six independent group will be examined.

Exploratory analysis and summary statistics

Experiment was made on 60 guinea pigs. They were divided into 6 groups and each group was threated with different supplement dose(0.5,1 and 2mg) and delivery methods(OJ, AC). Data set contains 60 observations (10 guinea pigs per each group) and three variable(Len, supp and dose). It was loaded via R with following code:

```
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 ...
```

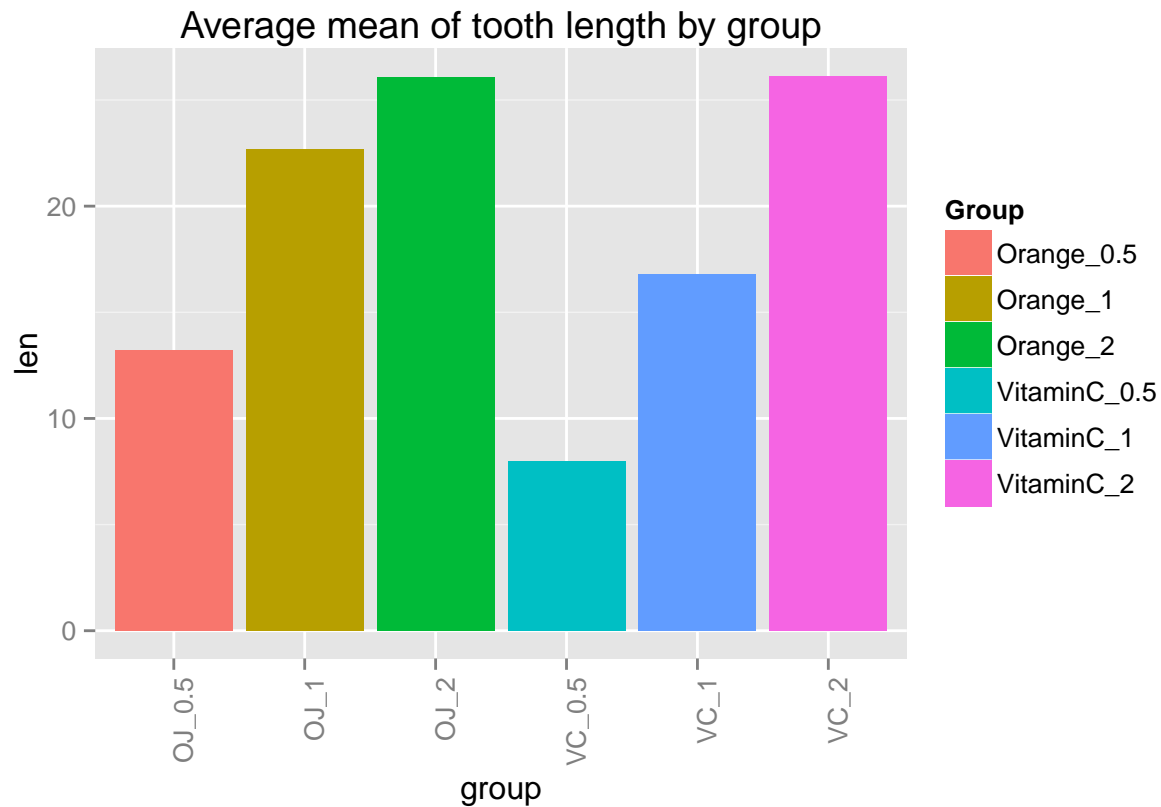
Data set contains informations about 6 different groups per supplement dose and delivery methods. Six independent groups needs to be compared by column len which represents Toothgrowth length.

New variable group was created – it represents each group of 10 pigs with different treatment:

```
ToothGrowth$group<-"OJ_0.5";
ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose==1,$group<-"OJ_1";
ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose==2,$group<-"OJ_2";
ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose==0.5,$group<-"VC_0.5";
ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose==1,$group<-"VC_1";
ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose==2,$group<-"VC_2";
```

Mean value of the each group was plotted:

```
library(ggplot2)
t1<-aggregate(len~group, FUN=mean,data=ToothGrowth)
ggplot(t1,aes(group,len,fill=group))+geom_bar(stat="identity",position="dodge")+
  theme(axis.text.x =element_text(size = 10,angle = 90,hjust = 1,vjust = 1))+
  scale_fill_discrete(name="Group",breaks=c("OJ_0.5", "OJ_1", "OJ_2","VC_0.5","VC_1","VC_2"),
  labels=c("Orange_0.5", "Orange_1", "Orange_2","VitaminC_0.5","VitaminC_1","VitaminC_2"))+ggtitle("Average length of tooth growth")
```



Summary statistics of length variable per each group was calculated:

```

tapply(ToothGrowth$len, ToothGrowth$group, summary)

```

```

## $OJ_0.5
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.20   9.70   12.25   13.23  16.18   21.50
##
## $OJ_1
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  14.50  20.30   23.45   22.70  25.65   27.30
##
## $OJ_2
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  22.40  24.58   25.95   26.06  27.08   30.90
##
## $VC_0.5
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.20   5.95   7.15   7.98  10.90   11.50
##
## $VC_1
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  13.60  15.27   16.50   16.77  17.30   22.50
##
## $VC_2
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  18.50  23.38   25.95   26.14  28.80   33.90

```

Assumptions

How the groups are independent (unpaired – each 10 pigs have different treatment) and only 10 pigs per group are present t-test was applied (mostly applied when number of sample data set is very small). Unequal variance parameter was used because variances differ in groups. Variance calculation is presented below:

```
tapply(ToothGrowth$len, ToothGrowth$group, var)
```

```
##      OJ_0.5      OJ_1      OJ_2      VC_0.5      VC_1      VC_2
## 19.889000 15.295556  7.049333  7.544000  6.326778 23.018222
```

Confidence intervals and t-test

Confidence intervals per groups(all possible groups were examined) were calculated based on t.test formula. Here is an example for CI calculation for group 1(supp orange juice with 0.5 mg dose) and group 2 (supp orange juice with 1 mg dose):

```
#take the subset
group1_2<-ToothGrowth[ToothGrowth$group %in% c("OJ_0.5","OJ_1"),]
#calculate CI
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group1_2)$conf

## [1] -13.415634 -5.524366
## attr("conf.level")
## [1] 0.95
```

Below are calculations for all groups:

```
group1_2<-ToothGrowth[ToothGrowth$group %in% c("OJ_0.5","OJ_1"),]
#calculate CI
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group1_2)$conf

## [1] -13.415634 -5.524366
## attr("conf.level")
## [1] 0.95
```

```
group1_3<-ToothGrowth[ToothGrowth$group %in% c("OJ_0.5","OJ_2"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group1_3)$conf

## [1] -16.335241 -9.324759
## attr("conf.level")
## [1] 0.95
```

```
group1_4<-ToothGrowth[ToothGrowth$group %in% c("OJ_0.5","VC_0.5"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group1_4)$conf

## [1] 1.719057 8.780943
## attr("conf.level")
## [1] 0.95
```

```
group1_5<-ToothGrowth[ToothGrowth$group %in% c("OJ_0.5","VC_1"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group1_5)$conf
```

```
## [1] -7.00810903 -0.07189097  
## attr("conf.level")  
## [1] 0.95
```

```
group1_6<-ToothGrowth[ToothGrowth$group %in% c("OJ_0.5","VC_2"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group1_6)$conf
```

```
## [1] -17.263522 -8.556478  
## attr("conf.level")  
## [1] 0.95
```

```
group2_3<-ToothGrowth[ToothGrowth$group %in% c("OJ_1","OJ_2"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group2_3)$conf
```

```
## [1] -6.5314425 -0.1885575  
## attr("conf.level")  
## [1] 0.95
```

```
group2_4<-ToothGrowth[ToothGrowth$group %in% c("OJ_1","VC_0.5"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group2_4)$conf
```

```
## [1] 11.51851 17.92149  
## attr("conf.level")  
## [1] 0.95
```

```
group2_5<-ToothGrowth[ToothGrowth$group %in% c("OJ_1","VC_1"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group2_5)$conf
```

```
## [1] 2.802148 9.057852  
## attr("conf.level")  
## [1] 0.95
```

```
group2_6<-ToothGrowth[ToothGrowth$group %in% c("OJ_1","VC_2"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group2_6)$conf
```

```
## [1] -7.5643336 0.6843336  
## attr("conf.level")  
## [1] 0.95
```

```
group3_4<-ToothGrowth[ToothGrowth$group %in% c("OJ_2","VC_0.5"),]  
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group3_4)$conf
```

```
## [1] 15.54182 20.61818  
## attr("conf.level")  
## [1] 0.95
```

```
group3_5<-ToothGrowth[ToothGrowth$group %in% c("OJ_2","VC_1"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group3_5)$conf
```

```
## [1] 6.859667 11.720333
## attr("conf.level")
## [1] 0.95
```

```
group3_6<-ToothGrowth[ToothGrowth$group %in% c("OJ_2","VC_2"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group3_6)$conf
```

```
## [1] -3.79807 3.63807
## attr("conf.level")
## [1] 0.95
```

```
group4_5<-ToothGrowth[ToothGrowth$group %in% c("VC_0.5","VC_1"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group4_5)$conf
```

```
## [1] -11.265712 -6.314288
## attr("conf.level")
## [1] 0.95
```

```
group4_6<-ToothGrowth[ToothGrowth$group %in% c("VC_0.5","VC_2"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group4_6)$conf
```

```
## [1] -21.90151 -14.41849
## attr("conf.level")
## [1] 0.95
```

```
group5_6<-ToothGrowth[ToothGrowth$group %in% c("VC_1","VC_2"),]
t.test(len ~ group, paired = FALSE, var.equal = FALSE, data = group5_6)$conf
```

```
## [1] -13.054267 -5.685733
## attr("conf.level")
## [1] 0.95
```

Conclusions

From the results above we can conclude that there is a statistically significant mean difference among all treatments – exceptions are groups OJ(dose 1) and Vitamin C (dose 2) and OJ and Vitamin C (both with dose 2). In confidence intervals that are calculated for those groups 0 is included in interval so we can assume that there is no difference in average tooth growth. Negative confidence intervals means that group 1 having slower tooth growth than the group2. Positive confidence interval tells us that group one have higher average tooth growth and if 0 is not included means that this positive value is statistically significant.