Vitamin C effect on Guinea Pig Tooth Growth

Jacques wagstaff

18 November 2016

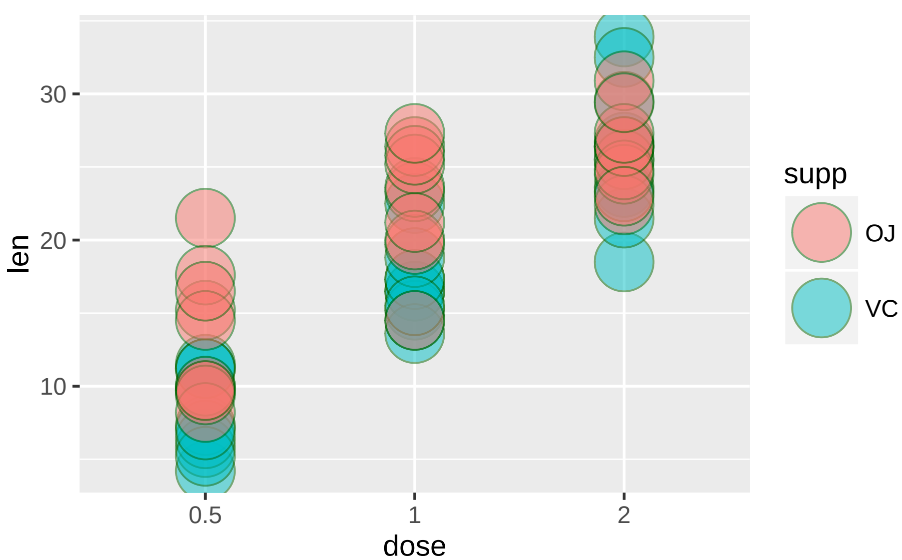
Overview

For the second part of the project, we analyze the ToothGrowth data in the R datasets package. The data looks at the effect of vitamin C on tooth growth in 60 guinea pigs. The data frame has 60 observations of 3 variables; the length of odontoblasts (cells responsible for tooth growth) (len), the Supplement type (supp) taking either ascorbic acid (VC) or orange juice (OJ), and the dose in milligrams/day taking one of (0.5, 1, and 2 mg/day). A basic summary of the data is given below:

Loading the data and Basic Summary

First we load and plot the data by groups, and obtain a summary to get an overview

```
data(ToothGrowth); dat<-ToothGrowth ;library(ggplot2)
dat$dose<-as.factor(dat$dose) # turn dose into factor variable
g <- ggplot(dat, aes(x = dose, y = len, group = supp))
g <- g + geom_point(size =10, pch = 21, colour="darkgreen" ,alpha = .5, aes(fill = supp))
g
```




```
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
```

The aim of the project is to identify if the delivery Method and/or Dosage affect tooth grow in Guinea Pigs. We can already see some apparent effect of method and dose from the plot above. We now quantify these effects.

Hypothesis testing

In this section we perform Hypothesis testing to compare tooth growth by supp and dose. There are 6 groups, labeled by one of method (OJ, VC) and one of dose (0.5,1,2). The Guinea Pigs are different in all groups, therefore we perform independent group analysis. We perform 2-group intervals for all possible group combinations, where the data is not paired (since the guinea pigs are different).

Comparing method VS vs OJ

First we split the data into groups by dose, so that we can compare method VS vs OJ.

```
dat05<-subset(dat, dose %in% c(0.5))
dat1<-subset(dat, dose %in% c(1))
dat2<-subset(dat, dose %in% c(2))
```

The Null hypothesis is that the difference in the means between the two groups is zero. We perform an independent group t-test to test this hypothesis. The results for the p-values and 95% confidence intervals are presented in the table below:

```
##   dose  supp pValues Interval1 Interval2
## 1  0.5 OJ-VC 0.00636      1.72      8.78
## 2  1.0 OJ-VC 0.00104      2.80      9.06
## 3  2.0 OJ-VC 0.96400     -3.80      3.64
```

The R-code for this table is given in the appendix.

We claim any p-value below $\alpha = 0.05$ as significant. We observe small p-values for small dose levels (0.5 and 1). This supports the alternative hypothesis: true difference in means is not equal to 0. Hence, for these dose values the type of delivery method has a significant effect on the tooth growth length. More specifically, the method by orange juice "OJ" is significantly more effective than the method by "VC" (ascorbic acid). Whereas for large dose (2), where the p-value is large and the interval contains 0, the delivery method has no effect i.e. both delivery methods are equally effective.

Comparing dose groups

Next we subset by OJ method, and compare dose groups

```
datOJ05_1<-subset(dat, supp %in% c("OJ") & dose %in% c("0.5","1"))
datOJ05_2<-subset(dat, supp %in% c("OJ") & dose %in% c("0.5","2"))
datOJ1_2<-subset(dat, supp %in% c("OJ") & dose %in% c("1","2"))
```

Finally we subset by VC method, and compare dose groups

```
datVC05_1<-subset(dat, supp %in% c("VC") & dose %in% c("0.5","1"))
datVC05_2<-subset(dat, supp %in% c("VC") & dose %in% c("0.5","2"))
datVC1_2<-subset(dat, supp %in% c("VC") & dose %in% c("1","2"))
```

Again, the Null hypothesis is that the difference in the means between pairs of groups is zero. The results for the p-values are presented in the table below: The R-code for this table is given in the appendix.

```
##   supp  dose  pValues Interval1 Interval2
## 1   OJ 0.5-1 8.78e-05    -13.40    -5.520
## 2   OJ 0.5-2 1.32e-06    -16.30    -9.320
## 3   OJ   1-2 3.92e-02     -6.53    -0.189
## 4   VC 0.5-1 6.81e-07    -11.30    -6.310
## 5   VC 0.5-2 4.68e-08    -21.90   -14.400
## 6   VC   1-2 9.16e-05    -13.10    -5.690
```

We observe small p-values $p < 0.05$ for all combinations, and all negative intervals. Hence an increase in dose results in a significant increase in the length of odontoblasts, for both delivery methods. The p-values are generally smaller for the "VC" method, this tells us that the difference in effectiveness between the "VC" and "OJ" methods seems to get smaller as the dose increases. This can also be seen in the data plot above.

Assumptions

For the statistical testing we carried out we assumed the following:

1. The underlying data is Normally distributed
2. We assumed unequal variances in the paired t-tests
3. We tested many hypotheses, and assumed that no correcting for this is required. This is a fair assumption because most of our p-values are very small anyway.

Conclusions

Here we answer the question: Do delivery Method and/or Dosage affect tooth grow in Guinea Pigs?

1. An increase in dose results in a significant increase in the length of odontoblasts, for both delivery methods.
2. For dose values of 0.5 and 1 mg/day the type of delivery method has a significant effect on the tooth growth length. More specifically, the method by orange juice "OJ" is significantly more effective than the method by "VC" (ascorbic acid). Whereas for large dose (2 mg/day), the delivery method has no effect i.e. both delivery methods are equally effective.
3. Hence the effectiveness of the "VC" method increases more rapidly as dose increase than the "OJ" method.

Appendix

Here is the R-code for table 1:

```r
pvalsumm<-data.frame(dose=c(0.5,1,2),supp=c("OJ-VC","OJ-VC","OJ-VC"),
                     pValues=signif(c(
                             t.test(len~supp, paired=F, var.equal=F, data=dat05)$p.value,
                             t.test(len~supp, paired=F, var.equal=F, data=dat1)$p.value,
```

```r
                                t.test(len~supp, paired=F, var.equal=F, data=dat2)$p.value), 3),
                        Interval1=signif(c(
                                t.test(len~supp, paired=F, var.equal=F, data=dat05)$conf.int[1],
                                t.test(len~supp, paired=F, var.equal=F, data=dat1)$conf.int[1],
                                t.test(len~supp, paired=F, var.equal=F, data=dat2)$conf.int[1]),3),
                        Interval2=signif(c(
                                t.test(len~supp, paired=F, var.equal=F, data=dat05)$conf[2],
                                t.test(len~supp, paired=F, var.equal=F, data=dat1)$conf.int[2],
                                t.test(len~supp, paired=F, var.equal=F, data=dat2)$conf.int[2]),3)

                        )
pvalsumm
```

```
##   dose  supp pValues Interval1 Interval2
## 1  0.5 OJ-VC 0.00636      1.72      8.78
## 2  1.0 OJ-VC 0.00104      2.80      9.06
## 3  2.0 OJ-VC 0.96400     -3.80      3.64
```

Here is the R-code for table 2:

```r
pvalsumm2<-data.frame(supp=c("OJ","OJ","OJ","VC","VC","VC"),
                         dose=c("0.5-1","0.5-2","1-2","0.5-1","0.5-2","1-2"),
                        pValues=signif(c(
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ05_1)$p.value,
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ05_2)$p.value,
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ1_2)$p.value,
                                t.test(len~dose, paired=F, var.equal=F, data=datVC05_1)$p.value,
                                t.test(len~dose, paired=F, var.equal=F, data=datVC05_2)$p.value,
                                t.test(len~dose, paired=F, var.equal=F, data=datVC1_2)$p.value),3),
                        Interval1=signif(c(
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ05_1)$conf.int[1],
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ05_2)$conf.int[1],
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ1_2)$conf.int[1],
                                t.test(len~dose, paired=F, var.equal=F, data=datVC05_1)$conf.int[1],
                                t.test(len~dose, paired=F, var.equal=F, data=datVC05_2)$conf.int[1],
                                t.test(len~dose, paired=F, var.equal=F, data=datVC1_2)$conf.int[1]),3),
                        Interval2=signif(c(
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ05_1)$conf.int[2],
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ05_2)$conf.int[2],
                                t.test(len~dose, paired=F, var.equal=F, data=datOJ1_2)$conf.int[2],
                                t.test(len~dose, paired=F, var.equal=F, data=datVC05_1)$conf.int[2],
                                t.test(len~dose, paired=F, var.equal=F, data=datVC05_2)$conf.int[2],
                                t.test(len~dose, paired=F, var.equal=F, data=datVC1_2)$conf.int[2]),3))
pvalsumm2
```

```
##   supp  dose  pValues Interval1 Interval2
## 1   OJ 0.5-1 8.78e-05    -13.40    -5.520
## 2   OJ 0.5-2 1.32e-06    -16.30    -9.320
## 3   OJ   1-2 3.92e-02     -6.53    -0.189
## 4   VC 0.5-1 6.81e-07    -11.30    -6.310
## 5   VC 0.5-2 4.68e-08    -21.90   -14.400
## 6   VC   1-2 9.16e-05    -13.10    -5.690
```