Basic Inferential Data Analysis

Emilie Smith

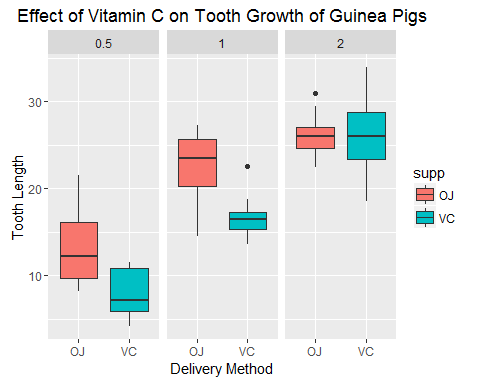
November 16, 2016

### Overview

##### In this project I analyzed the ToothGrowth data (<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html>) that recorded the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice OJ or ascorbic acid (a form of vitamin C and coded as VC).

#### 1)Load the ToothGrowth data and perform some basic exploratory data analyses

library(datasets)  
data(ToothGrowth)



##### From the graph above we can see that an increase in the dose of vitamin C increases the tooth length. At the lowest dose OJ seems to give an increase in tooth lenght compared to VC, but at the highest dose there seems to be almost no difference between the two supplements.

#### 2)Provide a basic summary of the 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 ...

##### There are 60 observations of 3 variables: len (which represents the length of the odontoblasts), supp (which describdes the supplement used (OJ or VC)), and dose (which is the dose of vitamin C administered: 0.5, 1 and 2 mg/day)

#### 3)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)

##### Let's first compare the tooth growth by supplement type by calculating the confidence interval and p-value accross all doses. The full results are located in the appendix.

t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)

##### The 95% confidence interval contains 0, and the p-value is higher than 0.01, so we cannot reject the null hypothesis that there is no difference between the two supplements. In other words, there is not a significant difference between the two supplements when all doses are taken into account.

##### Now let's take a look at the different doses. The full results are located in the appendix.

subdata1 <- ToothGrowth[ToothGrowth$dose!=0.5, ]#Excludes 0.5 mg dose  
subdata2 <- ToothGrowth[ToothGrowth$dose!=1, ]#Excludes 1 mg dose  
subdata3 <- ToothGrowth[ToothGrowth$dose!=2, ]#Excludes 2 mg dose  
  
t.test(len ~ dose, paired = F, var.equal = F, data = subdata1)#compare dose 1mg and 2mg

##### In this case the 95% confidence interval does not contain 0, and the p-value is small(1.906e-05). We can reject the null hypothesis that there is no difference between the 1mg and 2mg dose, meaning there is a significant difference between the 2mg and 1 mg dose.

t.test(len ~ dose, paired = F, var.equal = F, data = subdata2)#compare dose 0.5mg and 2mg

##### The difference between the 0.5mg and 2mg dose appears to be even more pronounced than the difference between 1mg and 2mg dose, as seen by the difference between the mean of 2mg dose (26.100) and the mean of the 0.5mg dose (10.605). The 95% confidence interval does not contain 0, and the p-value is small (less than 0.01).

t.test(len ~ dose, paired = F, var.equal = F, data = subdata3)#compare dose 0.5mg and 1mg

##### Once again, the 95% confidence interval does not contain 0, and the p-value (1.268e-07) is small, which means that there is a significant difference between the 0.5mg and 1mg.

##### Now let's look at each dose and see if there's an effect between the different supplements. The full results are located in the appendix section.

dose0.5<- ToothGrowth[ToothGrowth$dose==0.5, ]#Includes 0.5 mg dose  
dose1<- ToothGrowth[ToothGrowth$dose==1, ]#Includes 1 mg dose  
dose2<- ToothGrowth[ToothGrowth$dose==2, ]#Includes 2 mg dose  
  
t.test(len ~ supp, paired = F, var.equal = F, data = dose0.5)  
t.test(len ~ supp, paired = F, var.equal = F, data = dose1)  
t.test(len ~ supp, paired = F, var.equal = F, data = dose2)

##### It looks like the smaller the dose, the more effective OJ is compared to VC on tooth length:

##### - The 0.5mg dose has a 95% confidence interval of [1.719057 8.780943] with a p-value less than 0.01, which indicates that there is a significant difference between OJ and VC. OJ has a mean tooth length of 13.23 and VC a mean of 7.98, pointing that OJ is more effective than VC.

##### - The 1mg dose shows similar results as the 0.5mg dose, with a mean of 22.70 for OJ and a mean of 16.77 for VC.

##### - Finally the 2mg dose does not seem to differentiate between the two supplements, since the 95% confidence interval contains 0, the p-value is high (0.9639), and the difference between the 2 means is almost 0 (0.08)

#### 4)State your conclusions and the assumptions needed for your conclusions.

#### Conclusions

##### From the confidence interval calculations, I can conclude that the dose of Vitamin C correlates with tooth lenght: as the dose increases so does the tooth length.

##### As for the delivery system or supplement, OJ seems to have a better effect than VC only at a lower dose (0.5mg or 1mg).

#### Asumptions

##### 1)The tested groups are not paired (meaning the same guinea pig does not get a different treatment).

##### 2)The sample is representative of the population.

##### 3)Variances between the population are not equal.

#### Appendix

#### Graph code:

library(datasets)  
data(ToothGrowth)  
  
library(ggplot2)  
g = ggplot(ToothGrowth, aes(supp, len, fill = supp))  
g = g + geom\_boxplot()  
g = g + facet\_grid(~dose)  
g + ggtitle("Effect of Vitamin C on Tooth Growth of Guinea Pigs") + xlab("Delivery Method") + ylab("Tooth Length")

#### Full results:

t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)

##   
## Welch Two Sample t-test  
##   
## data: len by supp  
## t = 1.9153, df = 55.309, p-value = 0.06063  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1710156 7.5710156  
## sample estimates:  
## mean in group OJ mean in group VC   
## 20.66333 16.96333

t.test(len ~ dose, paired = F, var.equal = F, data = subdata1)#compare dose 1mg and 2mg

##   
## Welch Two Sample t-test  
##   
## data: len by dose  
## t = -4.9005, df = 37.101, p-value = 1.906e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -8.996481 -3.733519  
## sample estimates:  
## mean in group 1 mean in group 2   
## 19.735 26.100

t.test(len ~ dose, paired = F, var.equal = F, data = subdata2)#compare dose 0.5mg and 2mg

##   
## Welch Two Sample t-test  
##   
## data: len by dose  
## t = -11.799, df = 36.883, p-value = 4.398e-14  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -18.15617 -12.83383  
## sample estimates:  
## mean in group 0.5 mean in group 2   
## 10.605 26.100

t.test(len ~ dose, paired = F, var.equal = F, data = subdata3)#compare dose 0.5mg and 1mg

##   
## Welch Two Sample t-test  
##   
## data: len by dose  
## t = -6.4766, df = 37.986, p-value = 1.268e-07  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -11.983781 -6.276219  
## sample estimates:  
## mean in group 0.5 mean in group 1   
## 10.605 19.735

dose0.5<- ToothGrowth[ToothGrowth$dose==0.5, ]#Includes 0.5 mg dose  
dose1<- ToothGrowth[ToothGrowth$dose==1, ]#Includes 1 mg dose  
dose2<- ToothGrowth[ToothGrowth$dose==2, ]#Includes 2 mg dose  
  
t.test(len ~ supp, paired = F, var.equal = F, data = dose0.5)

##   
## Welch Two Sample t-test  
##   
## data: len by supp  
## t = 3.1697, df = 14.969, p-value = 0.006359  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.719057 8.780943  
## sample estimates:  
## mean in group OJ mean in group VC   
## 13.23 7.98

t.test(len ~ supp, paired = F, var.equal = F, data = dose1)

##   
## Welch Two Sample t-test  
##   
## data: len by supp  
## t = 4.0328, df = 15.358, p-value = 0.001038  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.802148 9.057852  
## sample estimates:  
## mean in group OJ mean in group VC   
## 22.70 16.77

t.test(len ~ supp, paired = F, var.equal = F, data = dose2)

##   
## Welch Two Sample t-test  
##   
## data: len by supp  
## t = -0.046136, df = 14.04, p-value = 0.9639  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.79807 3.63807  
## sample estimates:  
## mean in group OJ mean in group VC   
## 26.06 26.14