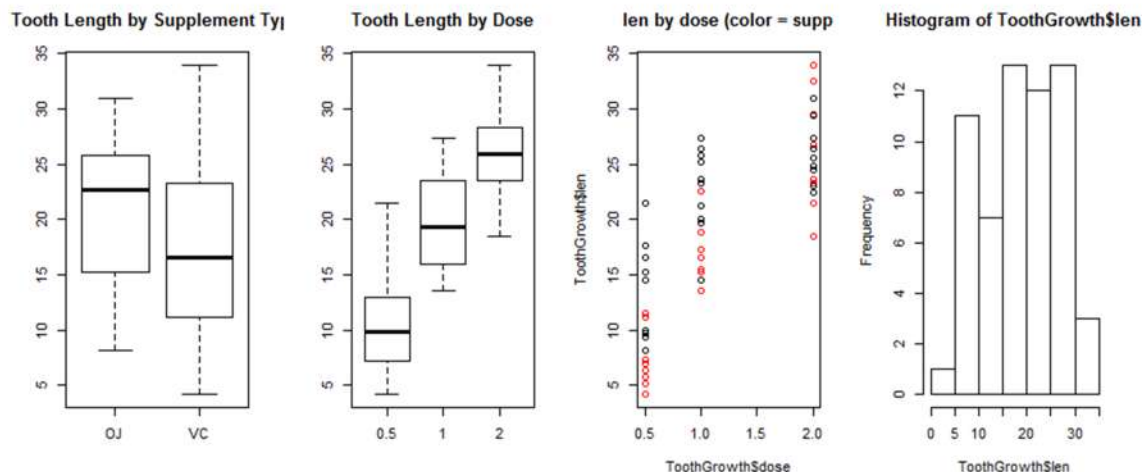


Tooth Growth Inferential Data Analysis

We are examining the ToothGrowth data set from the datasets package in R. From the R help file: "The response is 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)."

Basic summary of data

The data consist of 60 observations of 3 variables: Tooth Length (numeric len), Supplement type (factor supp), and Dose in mg (numeric dose). The mean tooth length is 18.813333 with a standard deviation of 7.6493152. We now explore the data and see if there is a relationship between tooth length, dose, and supplement type.



From the 2 boxplots above, we can see two things: 1) the mean tooth length increases with dose size and 2) the mean tooth length is higher for orange juice than it is for ascorbic acid. Additionally, from the scatter plot, we see that the positive relationship between tooth length and dose seems to hold regardless of supplement type.

Analysis

This suggests a more formal analysis is warranted. We begin by establishing that although the tooth length variable is not iid normal, it is roughly symmetrical and mound shaped from the histogram in the far right plot.

This suggests a student's t-test is appropriate to determine if: 1) higher doses are associated with higher tooth length measurements 2) orange juice is associated with higher tooth length measurements than ascorbic acid 3) combinations of dose and supplement type are related to tooth length measurements. The description of the data suggests that there were 6 groups with 10 guinea pigs each chosen at random implying they are independent and not paired.

Are higher doses associated with higher tooth length measurements?

Variance of tooth length for doses of 0.5, 1.0, 2.0

```
## [1] 20.24787 19.49608 14.24421
```

Before beginning our analysis, we must establish if the variance of the groups are equal. It appears from the above calculation the variance of the group given a dose of 2mg is not equal to the other two groups. Therefore, we will assume unequal variance in our t-test to be conservative.

```
mean(len|dose=0.5) - mean(len|dose=1.0)
```

```
p-value = 1.268300710^{-7}
```

```
## [1] -11.983781 -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
mean(len|dose=1.0) - mean(len|dose=2.0)
```

```
p-value = 1.906429510^{-5}
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Both t-tests result in p-values well below 0.01 suggesting we would reject the null hypothesis that the mean tooth length was equal for dose pairs {0.5,1.0} and {1,2} at any reasonable p-value. The confidence intervals are both negative indicating lower doses were associated with lower mean tooth lengths. We can also infer that mean tooth lengths for doses of 0.5mg are less than those at doses of 2.0mg with 95% confidence since the 95% confidence intervals do not overlap.

Is orange juice associated with higher tooth length measurements than ascorbic acid?

Variance of tooth length by supplement OJ and VC

```
## [1] 43.63344 68.32723
```

Before beginning our analysis, we must establish if the variance of the groups are equal. It appears from the above calculation the variance of the groups are not equal.

```
mean(len|supp=OJ) - mean(len|supp=VC)
```

```
p-value = 0.0606345
```

```
## [1] -0.1710156 7.5710156
## attr(,"conf.level")
## [1] 0.95
```

The t-test results in a p-value of 0.06063, so we fail to reject the null hypothesis at a 5% significance level that the mean tooth length was equal for guinea pigs fed orange juice and ascorbic acid. We reach the same conclusion by examining the the 95% confidence interval [-0.1710156,7.5710156] which includes 0.

Are combinations of dose and supplement type related to tooth length measurements?

One way to examine this question is to see if there is a difference in tooth length by dose, holding the supplement type constant and separately if there is a difference in tooth length by supplement, holding the dose constant. (h/t Keith Williams on the discussion forum)

Variance of tooth length by dose (0.5, 1.0, 2.0 respectively) for orange juice supplement

```
## [1] 19.889000 15.295556 7.049333
```

Variance of tooth length by dose (0.5, 1.0, 2.0 respectively) for ascorbic acid supplement

```
## [1] 7.544000 6.326778 23.018222
```

Before beginning our analysis, we must establish if the variance of the groups are equal. It appears from the above calculation the variance of the groups are not equal.

What is the difference in tooth length by dose, holding the supplement type constant?

```
mean(len|supp=OJ&dose=0.5) - mean(len|supp=OJ&dose=1.0)
```

```
p-value = 8.784919110^{-5}
```

```
## [1] -13.415634 -5.524366
```

```
## attr(,"conf.level")
```

```
## [1] 0.95
```

```
mean(len|supp=OJ&dose=1.0) - mean(len|supp=OJ&dose=2.0)
```

```
p-value = 0.0391951
```

```
## [1] -6.5314425 -0.1885575
```

```
## attr(,"conf.level")
```

```
## [1] 0.95
```

```
mean(len|supp=VC&dose=0.5) - mean(len|supp=VC&dose=1.0)
```

```
p-value = 6.811017710^{-7}
```

```
## [1] -11.265712 -6.314288
```

```
## attr(,"conf.level")
```

```
## [1] 0.95
```

```
mean(len|supp=VC&dose=1.0) - mean(len|supp=VC&dose=2.0)
```

```
p-value = 9.155603110^{-5}
```

```
## [1] -13.054267 -5.685733
```

```
## attr(,"conf.level")
```

```
## [1] 0.95
```

All the t-tests result in p-values below 0.05 meaning we would reject the null hypothesis that the mean tooth length was equal each dose pair at a 5% significance level. The confidence intervals are all negative indicating lower doses were associated with lower mean tooth lengths for both orange juice and ascorbic acid. We can also infer that mean tooth lengths for doses of 0.5mg are less than those at doses of 2.0mg with 95% confidence since the 95% confidence intervals do not overlap. Of note is that the p-value for the t-test comparing doses of 1mg versus 2mg when administered through orange juice was much higher than the other 3 combinations or the earlier t-test with pooled supplements. This indicates the strength of the relationship was not as strong.

What is the difference in tooth length by supplement, holding the dose constant?

```
mean(len|supp=OJ&dose=0.5) - mean(len|supp=VC&dose=0.5)
```

```
p-value = 0.0063586
```

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

```
mean(len|supp=OJ&dose=1.0) - mean(len|supp=VC&dose=1.0)
```

```
p-value = 0.0010384
```

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

```
mean(len|supp=OJ&dose=2.0) - mean(len|supp=VC&dose=2.0)
```

```
p-value = 0.9638516
```

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

The t-tests for doses of 0.5mg and 1.0mg had p-values below 0.01 while the t-test for doses of 2.0mg had a large p-value of 0.96. This means we reject the null hypothesis that the mean tooth length was equal for orange juice and ascorbic acid for doses of 0.5mg and 1.0mg, but not at 2.0mg. The confidence intervals for were positive for doses of 0.5mg and 1.0mg meaning orange juice was associated with higher mean tooth lengths.

```
##### Source code #####
```

```
#basic summary plot
```

```
data(ToothGrowth)
```

```
par(mfcol = c(1, 3))
```

```
boxplot(ToothGrowth$len ~ ToothGrowth$supp,main="Tooth Length by Supplement Type")
```

```
boxplot(ToothGrowth$len ~ ToothGrowth$dose,main="Tooth Length by Dose")
```

```
plot(ToothGrowth$dose,ToothGrowth$len, col=ToothGrowth$supp,main="len by dose (color =
supp)")
```

```
#plot 2
```

```
hist(ToothGrowth$len)
```

```

# calculating variance of tooth length for each dose
c(var(ToothGrowth[ToothGrowth$dose==0.5,1]),
  var(ToothGrowth[ToothGrowth$dose==1.0,1]),
  var(ToothGrowth[ToothGrowth$dose==2.0,1]))

#subsetting data to only compare two doses
TG.51 <- subset(ToothGrowth, dose %in% c(0.5,1.0))
TG12 <- subset(ToothGrowth, dose %in% c(1.0,2.0))

#t-tests
t1 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = TG.51)
t2 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = TG12)

# calculating variance of tooth length for each supplement
c(var(ToothGrowth[ToothGrowth$supp=="OJ",1]),
  var(ToothGrowth[ToothGrowth$supp=="VC",1]))

#t-test
t3 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)
# calculating variance of tooth length for each OJ supplement
c(var(ToothGrowth[(ToothGrowth$supp=="OJ" &
ToothGrowth$dose==0.5),1]),var(ToothGrowth[(ToothGrowth$supp=="OJ" &
ToothGrowth$dose==1.0),1]),var(ToothGrowth[(ToothGrowth$supp=="OJ" &
ToothGrowth$dose==2.0),1]))

# calculating variance of tooth length for each VC supplement
c(var(ToothGrowth[(ToothGrowth$supp=="VC" &
ToothGrowth$dose==0.5),1]),var(ToothGrowth[(ToothGrowth$supp=="VC" &
ToothGrowth$dose==1.0),1]),var(ToothGrowth[(ToothGrowth$supp=="VC" &
ToothGrowth$dose==2.0),1]))

#t-tests for supp = OJ
t4<- t.test(ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==0.5),1],
            ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==1.0),1],
            paired = FALSE, var.equal = FALSE)
t5 <- t.test(ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==1.0),1],
            ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==2.0),1],
            paired = FALSE, var.equal = FALSE)
#t-tests for supp = VC
t6 <- t.test(ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==0.5),1],
            ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==1.0),1],
            paired = FALSE, var.equal = FALSE)
t7 <- t.test(ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==1.0),1],
            ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==2.0),1],
            paired = FALSE, var.equal = FALSE)
#t-tests for dose = 0.5

```

```
t8 <- t.test(ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==0.5),1],
            ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==0.5),1],
            paired = FALSE, var.equal = FALSE)
#t-tests for dose = 1.0
t9 <- t.test(ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==1.0),1],
            ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==1.0),1],
            paired = FALSE, var.equal = FALSE)
#t-tests for dose = 2.0
t10 <- t.test(ToothGrowth[(ToothGrowth$supp=="OJ" &
                        ToothGrowth$dose==2.0),1],
            ToothGrowth[(ToothGrowth$supp=="VC" &
                        ToothGrowth$dose==2.0),1],
            paired = FALSE, var.equal = FALSE)
```