Coursera Statistics Inference Assignment

Keith Bailey

February 4, 2017

Part2

In this section of the report we will be looking at tooth growth associated to two supplements, OJ & VC, at differing dosing levels, 0.5, 1 & 2.

Lets look at the data:

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
table(ToothGrowth[2:3])
```

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

From this we can see there are 10 samples for each dose-supplement combination.

In order to investigate this I have first established the mean and standard deviation so we can establish the 95% confidence intervals of each combination of supplement and dose.

```
cToothGrowth<-ToothGrowth%>%
              group_by(supp, dose)%>%
              summarise(len.mean=mean(len),
                        count = n(),
                        len.sd=sd(len),
                        len.95=1.96*len.sd)

kable(cToothGrowth)
```

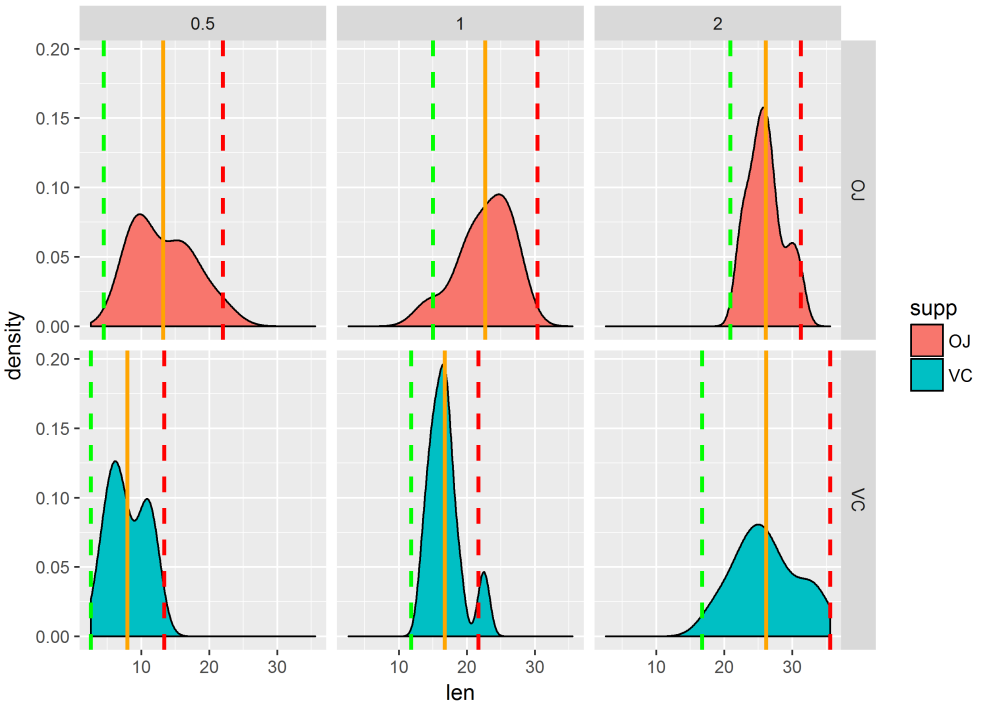
supp	dose	len.mean	count	len.sd	len.95
OJ	0.5	13.23	10	4.459708	8.741029
OJ	1.0	22.70	10	3.910953	7.665468
OJ	2.0	26.06	10	2.655058	5.203914
VC	0.5	7.98	10	2.746634	5.383403

supp	dose	len.mean	count	len.sd	len.95
VC	1.0	16.77	10	2.515309	4.930005
VC	2.0	26.14	10	4.797731	9.403553

We can see that for the lower doses there is clear separation between the growth achieved using different supplements. However, this would be much easier to see with a chart.

Below, we have a density plot for each combination, overlaid with the mean (orange), 2.5th percentile (green) and 97.5th percentile (red), to show the 95% confidence interval, so we can see what, overlap there are between these combinations.

```
ggplot(ToothGrowth, aes(x=len, fill=supp)) +
  geom_density() +
  geom_vline(data=cToothGrowth, aes(xintercept=len.mean),
             linetype="solid", size=1, colour="orange") +
  geom_vline(data=cToothGrowth, aes(xintercept=len.mean+len.95),
             linetype="dashed", size=1, colour="red") +
  geom_vline(data=cToothGrowth, aes(xintercept=len.mean-len.95),
             linetype="dashed", size=1, colour="green") +
  facet_grid(supp~dose)
```



Next, we would like to perform a t-test to conduct an unpaired test of each set of dose observations to assess the null hypothesis that there is no difference in the means between each set. From the chart above, we suspect that doses 0.5 and 1 will have low p-values such that we reject the null hypothesis, whereas does 2 looks likely that it will have a high p-value and so we will reject the alternative hypothesis. However it would be foolhardy to rely on such visual analysis alone.

Using the t-test function in R, we will asses each of these.

```
t.test(x=ToothGrowth$len[ToothGrowth$dose==0.5 & ToothGrowth$supp=="OJ"],
       y=ToothGrowth$len[ToothGrowth$dose==0.5 & ToothGrowth$supp=="VC"])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5 & ToothGrowth$supp ==  and ToothGrowth$len[ToothGrowth$d
ose == 0.5 & ToothGrowth$supp ==     "OJ"] and     "VC"]
## 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 of x mean of y
##     13.23      7.98
```

```
t.test(x=ToothGrowth$len[ToothGrowth$dose==1 & ToothGrowth$supp=="OJ"],
       y=ToothGrowth$len[ToothGrowth$dose==1 & ToothGrowth$supp=="VC"])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1 & ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowt
h$dose == 1 & ToothGrowth$supp == "VC"]
## 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 of x mean of y
##     22.70     16.77
```

```
t.test(x=ToothGrowth$len[ToothGrowth$dose==2 & ToothGrowth$supp=="OJ"],
       y=ToothGrowth$len[ToothGrowth$dose==2 & ToothGrowth$supp=="VC"])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2 & ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowt
h$dose == 2 & ToothGrowth$supp == "VC"]
## 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 of x mean of y
##     26.06     26.14
```

Conclusion

We can see from these that the when:

1. Dose == 0.5, we reject the NULL hypthesis as the p-value is very small, being less than 0.7%. This tells us that supplement OJ is a greater effect on tooth growth that is statistically significant with a growth difference of 1.7 to 8.7 greater than supplement VC with 95% confidence.

2. Dose == 1.0, again we reject the NULL hypthesis as the p-value is very small, being less than 0.1%. This tells us that supplement OJ is a greater effect on tooth growth that is statistically significant with a growth difference of at least 2.8 greater than supplement VC with 95% confidence.

3. Dose == 2.0, here we reject the ALTERNATIVE hypthesis as the p-value is very high, being almost 1. This tells us that there is no statistical difference between the two supplements at this dosage.