Coursera.org

Statistical Inference Assignment December 2014

Martin Sawtell

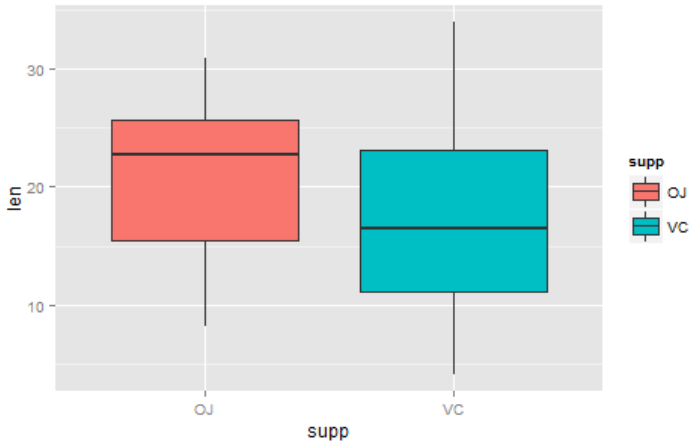
Part 2: Tooth Growth

1. Load the data and do some basic EDA:

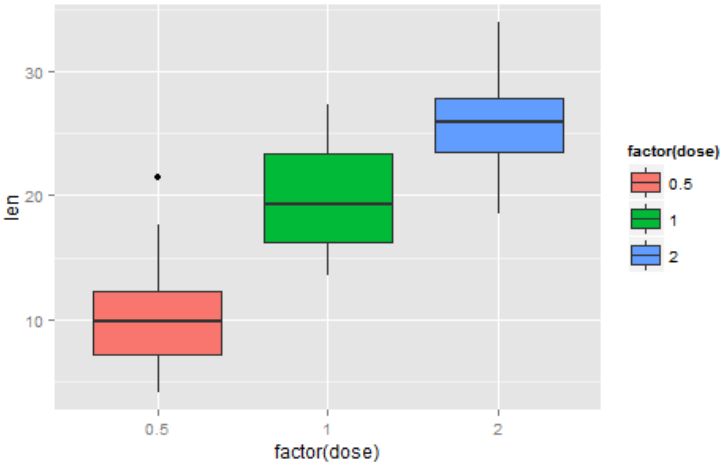
Rstudio output:

```
> data(ToothGrowth)
```

```
> library(ggplot2)
> ggplot(aes(x = supp, y = len), data = Tooth-
Growth) + geom_boxplot(aes(fill = supp))
```




```
> ggplot(aes(x = supp, y = len), data = Tooth-
Growth) + geom_boxplot(aes(fill = supp))
```



Question 1 ANSWER:
Based on a simple EDA it seems the higher the dose the longer the tooth length, and that OJ is more effective than VC.

2. Provide a basic summary of the data:

Rstudio output:

```
> 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
```

```
> str(ToothGrowth)
$ 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 ...
```

```
> 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
```

```
> table(ToothGrowth$supp,
ToothGrowth$dose)
```

Rstudio output:

```
   0.5  1  2
OJ  10 10 10
VC  10 10 10
```

Question 2 ANSWER:
We can see that there are 60 data measurements spread evenly through 6 categories, measuring OJ (Orange Juice) and VC (Ascorbic Acid) against 0.5mg, 1mg and 2mg doses. Note that the groups are not paired (the experiment did not repeat on the same guinea pigs).

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)

Compare by supplement type via grouping the data by dose first.

```
> dose_0.5 <- subset(ToothGrowth, dose == 0.5)
> dose_1.0 <- subset(ToothGrowth, dose == 1.0)
> dose_2.0 <- subset(ToothGrowth, dose == 2.0)
```

Null hypothesis: There is no significant increase in tooth length between the two supplement types.

```
> t.test(len ~ supp, data = dose_0.5)
```

```
        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, data = dose_1.0)
```

```
        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, data = dose_2.0)
```

```
        Welch Two Sample t-test

data:  len by supp
t = -0.0461, 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
```

The confidence interval of [1.72, 8,78] for mean(OJ)-mean(VC) at a dose of 0.5 mg results in rejecting the null hypothesis and stating that there is a significant difference in tooth length between the two supplement types at this dose.

The confidence interval of [2.80, 9.06] for mean(OJ)-mean(VC) at a dose of 1.0 mg results in rejecting the null hypothesis and stating that there is a significant difference in tooth length between the two supplement types at this dose.

The confidence interval of [-3.80, 3.64] for mean(OJ)-mean(VC) at a dose of 2.0 mg results in NOT rejecting the null hypothesis which claims there is a significant difference in tooth length between the two supplement types at this dose.

Coursera.org

Statistical Inference Assignment December 2014

Martin Sawtell

Compare by dose via grouping the data by supplement first.

Null hypothesis: There is no significant increase in tooth length between the three dosages.

```
> dose.1.2 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
> dose.1.3 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
> dose.2.3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

```
> t.test(len ~ dose, paired = F, var.equal = F, data = dose.1.2)

        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
```

```
> t.test(len ~ dose, paired = F, var.equal = F, data = dose.1.3)

        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 = dose.2.3)

        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
```

The confidence interval of [-11.98, -6.276] for mean(dose0.5)-mean(dose1.0) results in rejecting the null hypothesis and stating that there is a significant difference in tooth length between dosages of 0.5 and 1.0 mg (1.0 mg results in longer teeth).

The confidence interval of [-18.16, -12.83] for mean(dose0.5)-mean(dose2.0) results in rejecting the null hypothesis and stating that there is a significant difference in tooth length between dosages of 0.5 and 2.0 mg (2.0 mg results in longer teeth).

The confidence interval of [-8.996, -3.734] for mean(dose1.0)-mean(dose2.0) results in rejecting the null hypothesis and stating that there is a significant difference in tooth length between dosages of 1 and 2.0 mg (2.0 mg results in longer teeth).

Question 4 ANSWER:

Assumptions:

- the populations are independent as the subjects did not have the tests repeated on them (a guinea pig cannot grow a tooth twice!)
- the variances are not the same
- other standard rules were followed; ie accurate measurements by trained staff, iid sampling of guinea pigs from a larger population, etc

Conclusions:

The bigger factor in influencing tooth growth was the dosage, but on the whole orange juice was more effective than ascorbic acid in increasing tooth length. If one had to choose between {orange juice or ascorbic} OR {max dosage} one would choose max dosage. For maximum results orange juice and max dosage is recommended.