Statistical Inference Project Part 2

James Hancock

July 30, 2016

Process/Execution Details

1. Load the ToothGrowth data

Load Library

```
library(ggplot2)
library(datasets)
library(graphics)
}
```

Get information on the dataset

```
?ToothGrowth
data(ToothGrowth)}
```

2. 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 ...

```
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

```
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

```
unique(ToothGrowth$dose)}
```

0.5 1.0 2.0

```
unique(ToothGrowth$supp)}
```

VC OJ

Levels: OJ VC

```
by(ToothGrowth$len, INDICES = list(ToothGrowth$supp, ToothGrowth$dose), summary)}
```

: OJ

: 0.5

Min. 1st Qu. Median Mean 3rd Qu. Max.

8.20 9.70 12.25 13.23 16.18 21.50

: VC

: 0.5

Min. 1st Qu. Median Mean 3rd Qu. Max.

4.20 5.95 7.15 7.98 10.90 11.50

: OJ

: 1

Min. 1st Qu. Median Mean 3rd Qu. Max.

14.50 20.30 23.45 22.70 25.65 27.30

: VC

: 1

Min. 1st Qu. Median Mean 3rd Qu. Max.

13.60 15.27 16.50 16.77 17.30 22.50

: OJ

: 2

Min. 1st Qu. Median Mean 3rd Qu. Max.

22.40 24.58 25.95 26.06 27.08 30.90

: VC

: 2

Min. 1st Qu. Median Mean 3rd Qu. Max.

18.50 23.38 25.95 26.14 28.80 33.90

(supp). Is there a correlation between supp and does? 2.Analyze the Tooth lenght relative to Supplement alone. 3.Analyze the Tooth lenght relative to Dosage alone. 4.Determine if within a Dosage - supplements have different effects on tooth growth.

1. Analyze the Tooth Lenght (len) relative to Dosage(dose) & Supplement (supp).

Calculate len mean for every dose and supp

```
avg <- aggregate(len~.,data=ToothGrowth,mean)}
```

Plot the Tooth Lenght (len) relative to Dosage(dose) & Supplement (supp)

```
g <- ggplot(data = ToothGrowth,aes(x=dose,y=len))
g <- g + geom_point(aes(group=supp,colour=supp,size=2,alpha=0.6))
g <- g + geom_line(data=avg,aes(group=supp,colour=supp))
g <- g + labs(title="Fig 1: Tooth Lenght (len) relative to Dosage(dose) & Supplement (supp)")
print(g)}
```

2. Analyze the Tooth lenght relative Supplement alone.

Analyze the Tooth Lenght (len) relationship to Supplement(supp) independent of dose

```
g <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp))
g <- g + labs(title="Fig 2: Tooth Lenght (len) relationship to Supplement(supp)")
print(g)}
```

3. Analyze the Tooth lenght relative Dosage alone.

Analyze the Tooth Lenght (len) relationship to Dose (does) independent of Supplement

```
g <- ggplot(aes(x = factor(dose), y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = factor(dose)))
g <- g + labs(title="Fig 3: Tooth Lenght (len) relationship to Dosage(dose)")
print(g)}
```

4. Determine if within a Dosage - supplements have different effects on tooth growth.

Determine if within a Dosage - supplements have different effects on tooth growth

```
g <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose)
g <- g + labs(title="Fig 4: Tooth Lenght (len) by Supplement(supp) & Dosage(dose)")
print(g)
}
```

4. Confidence Intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Test 1. Check for group differences due to different supplement type assuming unequal variances between the two groups

Two Sample Test lenght relative to supplement

```
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

Test 2. Dosage as a Factor - Analyzing the data for correlation between the dose level and change in tooth growth

```
dose1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
```

```
dose3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = F, var.equal = F, data = dose1)}
```

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 = dose2)}
```

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 = dose3)}
```

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

Test 3. Supplement as a Factor within Dose Levels - Analyzing the data for correlation between dose level and change in tooth growth within each dose level:

```
Tooth.dose4 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose5 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose6 <- subset(ToothGrowth, dose == 2.0)
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose4)}
```

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 = Tooth.dose5)}
```

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 = Tooth.dose6)}
```

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

5. Conclusions and Assumptions

To make conclusions with the data in this dataset, we must assume that the poplulations are independent, that the variances between populations are different, a random population was used, the population was comprised of similar guinea pigs, measurement error was accounted for with significant digits, and double blind research methods were used.

If all the preceding assumptions are true, A higher dose level consistently led to longer teeth. Initially it appeared that the delivery method had no significant impact on tooth length, but when controlling for dose level we discovered that there was a significant difference at 0.5mg and 1.0mg, but not at 2.0mg. Based on this evidence, it appears that orange juice is a better delivery method with a larger impact on tooth length for a given dose of Vitamin C, but above a maximum dose level there is no further improvement.