Statistical Inference - Tooth Growth Data Analysis

David Junor

August 20, 2016

Overview

This is the second part of the two-part Coursera Statistical Inference course project. In this section we examine the tooth growth data found in R. It is the purpose of this document to provide a summary of the data and use confidence intervals and/or hypotheiss testing to compare tooth growth by delivery method and dose.

Exploratory Data Analysis

Load and Describe Data

```
#Load the data and describe
library(datasets)
data("ToothGrowth")
head(ToothGrowth, 6)
```

```
##    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
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

```
sum(!complete.cases(ToothGrowth))
```

```
## [1] 0
```

The data has no missing values with 60 observations of three variables:
len: length of teeth
supp: delivery method (OJ: orange juice; VC: ascorbic acid)
dose: amount of vitamin C (0.5, 1, and 2 mg)

```
#Summarize the statistical properties of the data
tgTable <- ToothGrowth %>% group_by(supp,dose) %>% summarize(n=length(len),mean=mean(len),sd=sd(len),se=
kable(tgTable)
```

supp	dose	n	mean	sd	se
OJ	0.5	10	13.23	4.459708	1.4102837
OJ	1.0	10	22.70	3.910953	1.2367520
OJ	2.0	10	26.06	2.655058	0.8396031
VC	0.5	10	7.98	2.746634	0.8685620
VC	1.0	10	16.77	2.515309	0.7954104
VC	2.0	10	26.14	4.797731	1.5171757

```
#Graph the SE of the data
ggplot(tgTable, aes(x = dose, y = mean, color = supp)) +
geom_errorbar(aes(ymin = mean - se, ymax = mean + se), width = .1) +
        geom_line(size=1.25) +
        geom_point(size=2, shape=21, fill="black") +
        xlab("Dose (mg)") + ylab("Tooth Length") +
        theme(legend.justification = c(1,0), legend.position = c(1,0))
```

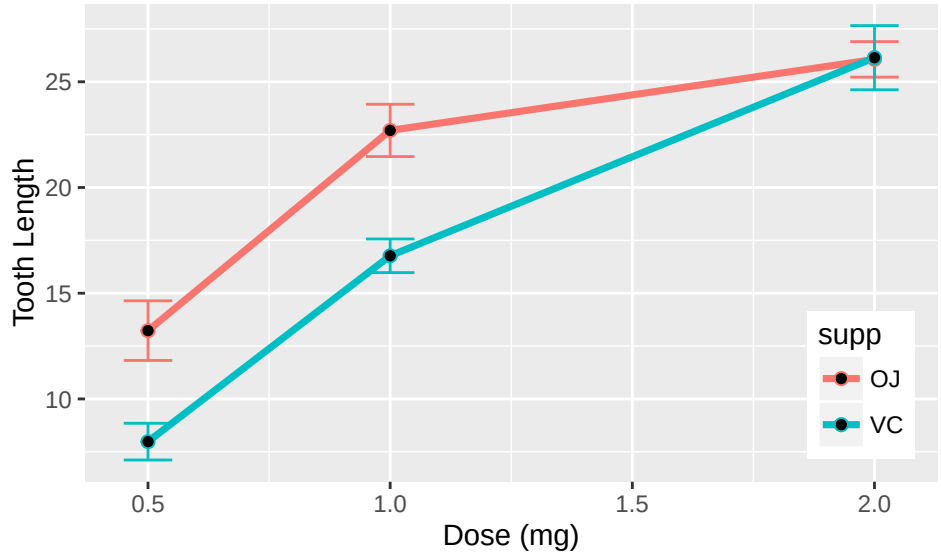

Figure 1. Graph of the standard error of the mean of guinea pig tooth length.

Summary

The summary table and graph appear to indicate that delivery method and dose have an effect on tooth growth. The mean of tooth growth increases as dose increases with the VC and OJ delivery methods. The VC delivery method and dose of 2mg provided the largest increase in growth but had the largest variance. Overall the OJ and VC delivery methods across doses appears to increase teeth length.

Data Analysis

Prepare Data

```r
#Supplement
vc <- ToothGrowth %>% filter(supp == "VC")
oj <- ToothGrowth %>% filter(supp == "OJ")
#Dose
dose5 <- ToothGrowth %>% filter(dose == 0.5)
dose1 <- ToothGrowth %>% filter(dose == 1.0)
dose2 <- ToothGrowth %>% filter(dose == 2.0)
```

Testing by Dosage

Null Hypothesis: Dose does not have an effect on teeth growth in guinea pigs.

```r
#95% CI - Dose 0.5 vs Dose 1.0
t.test(dose5$len, dose1$len, paired = FALSE, var.equal = FALSE)$conf
```

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

```r
t.test(dose5$len, dose1$len, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 1.268301e-07
```

```r
#95% CI - Dose 0.5 vs Dose 2.0
t.test(dose5$len, dose2$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(dose5$len, dose2$len, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 4.397525e-14
```

```r
#95% CI - Dose 1.0 vs Dose 2.0
t.test(dose1$len, dose2$len, paired = FALSE, var.equal = FALSE)$conf
```

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

```r
t.test(dose1$len, dose2$len, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 1.90643e-05
```

Summary

All dosage intervals have p-values less than alpha at the 95% CI. The null hypothesis is rejected. It appears that as the dose of Vitamin C increases teeth length in guinea pigs increases.

Testing by Delivery Method

Null Hypothesis: Delivery method does not have an effect on teeth growth in guinnea pigs.

```
t.test(oj$len, vc$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj$len and vc$len
## 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 of x mean of y
##  20.66333  16.96333
```

Summary

The p-value for the test is greater than alpha at 95% and the CI includes 0. We fail to reject the null hypothesis that delivery method has no effect on teeth growth in guinea pigs.

Conclusion

It appears that the teeth growth in guinea pigs increases with increasing dosages of vitamin C. The delivery method of vitamin C to guinea pigs did not appear to have an effect on teeth growth.