Statistical Inference Course Project Part Two: Data Analysis

John Higgins

December 3, 2017

Overview

We will explore the effects of vitamin supplementation on guinea pig tooth growth. Supplementation is provided either by ascorbic acid (VC) or by orange juice (OJ). Three dosage levels were applied: 0.5 mg/day, 1 mg/day, 2 mg/day. Subseqeunt tooth length is given in unspecified units.

Examining The Data

We load in the data and acquaint ourselves with its basic details.

```
library(datasets)
data("ToothGrowth")
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 ...
```

```
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
```

```
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
```

Let us look at aggregate summary data to determine how supplement and dose might affect tooth length.

```
aggregate(len ~ supp + dose, data = ToothGrowth, mean)
```

```
##   supp dose   len
## 1   OJ  0.5 13.23
## 2   VC  0.5  7.98
## 3   OJ  1.0 22.70
```

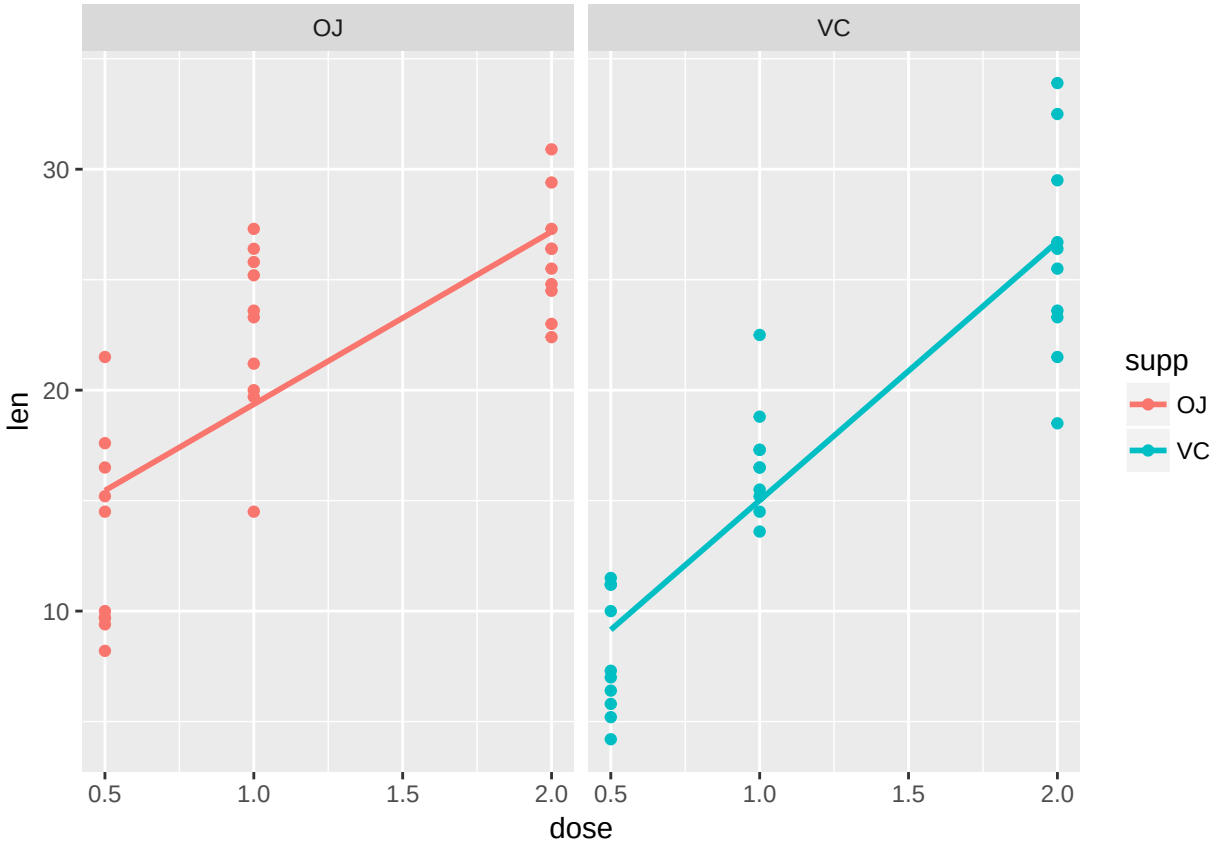
```
## 4   VC  1.0 16.77
## 5   OJ  2.0 26.06
## 6   VC  2.0 26.14
```

The data suggest that tooth growth is dose-dependent, and for lower doses, orange juice is superior to ascorbic acid. This effect decreases as dose increases. Let us visualize the data to see if it strengthens this hypothesis.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
ggplot(ToothGrowth,
       aes(x = dose, y = len)) +
  geom_point(aes(col = supp)) +
  geom_smooth(aes(col = supp),
              method = "lm",
              se = FALSE) +
  facet_wrap(~supp, nrow = 1)
```



Indeed, we see that both regression lines terminate at a similar value for dose 2 mg/day, but the line for ascorbic acid is steeper. Let us test these hypotheses.

Hypothesis Testing

Our tentative hypotheses are these: For a given supplement, greater dose implies greater tooth growth. Between supplements, OJ produces greater growth than VC at lower doses, but this difference vanishes as the dose increases.

To check these hypotheses, we will run T-tests. We will compare a supplement to itself at increasing dosage, and we will compare one supplement to the other at identical doses. Our null hypothesis is that the means of tooth length will be identical across supplement and dosage.

Dose Dependency

OJ, 0.5 mg/day vs. 1 mg/day

Alternative hypothesis: the mean tooth length for 1 mg/day is greater than that for 0.5 mg/day.

```
oj.halfmg <- with(ToothGrowth, len[supp == "OJ" & dose == 0.5])
oj.onemg <- with(ToothGrowth, len[supp == "OJ" & dose == 1.0])
t.test(oj.onemg, oj.halfmg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.onemg and oj.halfmg
## t = 5.0486, df = 17.698, p-value = 4.392e-05
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.214316      Inf
## sample estimates:
## mean of x mean of y
##     22.70     13.23
```

For a p-value of .05, the null hypothesis is rejected.

OJ, 1 mg/day vs. 2 mg/day

Alternative hypothesis: the mean tooth length for 2 mg/day is greater than that for 1 mg/day.

```
oj.twomg <- with(ToothGrowth, len[supp == "OJ" & dose == 2.0])
t.test(oj.twomg, oj.onemg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.twomg and oj.onemg
## t = 2.2478, df = 15.842, p-value = 0.0196
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.7486236       Inf
## sample estimates:
## mean of x mean of y
##     26.06     22.70
```

For p = .05, the null hypothesis is rejected.

VC, 0.5 mg/day vs. 1 mg/day

Alternative hypothesis: the mean tooth length for 1 mg/day is greater than that for 0.5 mg/day.

```
vc.halfmg <- with(ToothGrowth, len[supp == "VC" & dose == 0.5])
vc.onemg <- with(ToothGrowth, len[supp == "VC" & dose == 1.0])
t.test(vc.onemg, vc.halfmg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  vc.onemg and vc.halfmg
## t = 7.4634, df = 17.862, p-value = 3.406e-07
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.746867      Inf
## sample estimates:
## mean of x mean of y
##     16.77      7.98
```

For p = .05, the null hypothesis is rejected.

VC, 1 mg/day vs. 2 mg/day

Alternative hypothesis: the mean tooth length for 2 mg/day is greater than that for 1 mg/day.

```
vc.twomg <- with(ToothGrowth, len[supp == "VC" & dose == 2.0])
t.test(vc.twomg, vc.onemg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  vc.twomg and vc.onemg
## t = 5.4698, df = 13.6, p-value = 4.578e-05
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.346525      Inf
## sample estimates:
## mean of x mean of y
##     26.14     16.77
```

For p = .05, the null hypothesis is rejected.

Conclusion

Supplement dose correlates strongly with tooth growth. An increase in supplementation level, regardless of the supplement, corresponds to greater tooth growth, p < .05.

Supplement Choice

OJ vs. VC, 0.5 mg/day

Alternative hypothesis: OJ elicits greater tooth growth than VC.

```
t.test(oj.halfmg, vc.halfmg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.halfmg and vc.halfmg
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604     Inf
```

```
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

For p = .05, the null hypothesis is rejected.

OJ vs. VC, 1 mg/day

Alternative hypothesis: OJ elicits greater tooth growth than VC.

```
t.test(oj.onemg, vc.onemg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.onemg and vc.onemg
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

For p = .05, the null hypothesis is rejected.

OJ vs. VC, 2 mg/day

Alternative hypothesis: OJ elicits greater tooth growth than VC.

```
t.test(oj.twomg, vc.twomg, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.twomg and vc.twomg
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

For p = .05, we fail to reject the null hypothesis.

Let us see if there is any detectable difference in mean between the two supplements, with a two-sided T-test.

```
t.test(oj.twomg, vc.twomg, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.twomg and vc.twomg
## 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
```

Again, for p = .05 or for any reasonable p-value, we fail to reject the null hypothesis.

Conclusion

For the two smaller dosages, OJ elicits greater tooth growth than VC ($p < .05$). For a dose of 2 mg/day, there is no difference between supplements.

Summary

Tooth growth increases as supplement growth increases. For a dose of 0.5 mg/day or 1 mg/day, orange juice results in greater tooth growth than ascorbic acid, however at a dose of 2 mg/day the tooth growth between supplements is identical.