Analysis of ToothGrowth Data

Rajesh Sinha

22 March 2015

Contents

1 Loading and Summary of Data

```
library(datasets)
data(ToothGrowth)
df <- data.frame(ToothGrowth)
df$supp <- factor(df$supp)
df$dose <- factor(df$dose)
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
```

2 Exploratory Analysis

We can see that the dataset contains length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

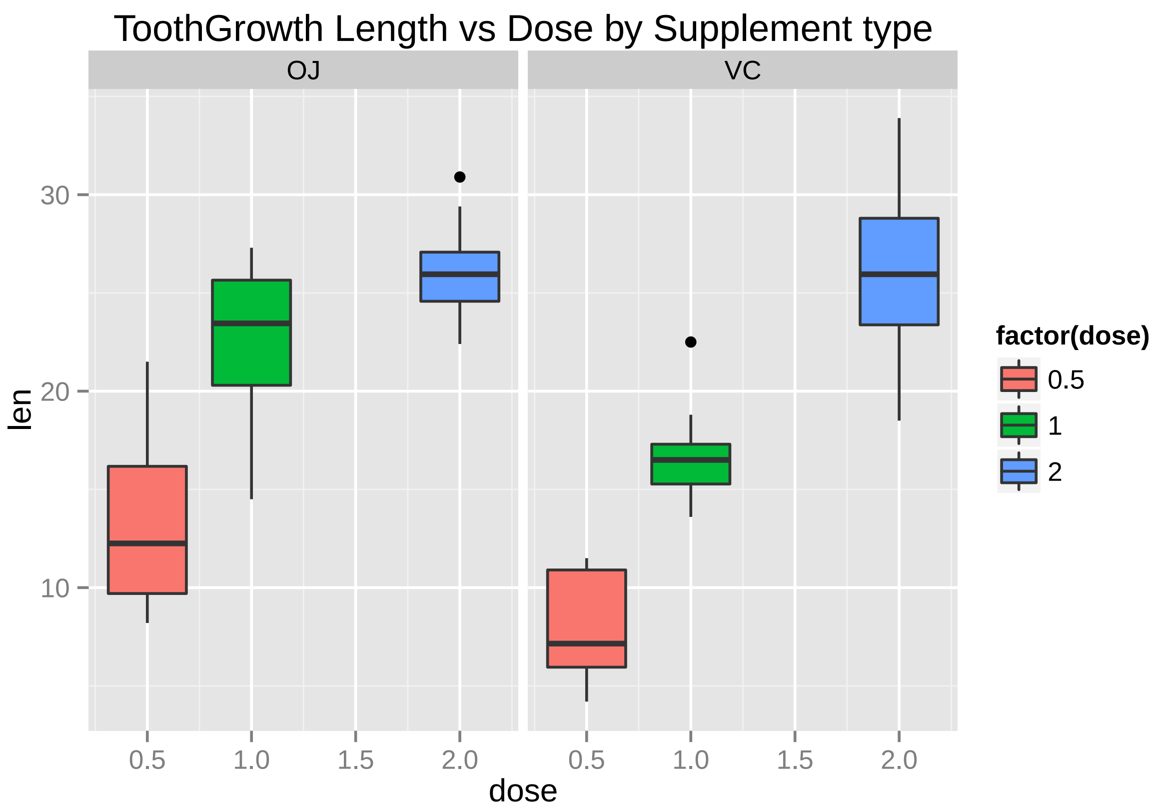
Summarizing the data by supplement type (VC or OJ) and dose we get following averages and standard deviation

```
df.s <- df %>% group_by (supp) %>% summarize(avg=mean(len), sd=sd(len), count=n()); # Grouped by supple
df.sd <- df %>% group_by (supp, dose) %>% summarize(avg=mean(len), sd=sd(len), count=n());
df.sd # Grouped by supplement and dose
```

```
## Source: local data frame [6 x 5]
## Groups: supp
##
##   supp dose   avg       sd count
## 1   OJ  0.5 13.23 4.459709    10
## 2   OJ    1 22.70 3.910953    10
## 3   OJ    2 26.06 2.655058    10
## 4   VC  0.5  7.98 2.746634    10
## 5   VC    1 16.77 2.515309    10
## 6   VC    2 26.14 4.797731    10
```

We can see a side-by-side box plot of the same for better visualization

```
ggplot(ToothGrowth, aes(x = dose, y = len)) + ggtitle("ToothGrowth Length vs Dose by Supplement type") +
    geom_boxplot(aes(fill = factor(dose))) + facet_grid(. ~ supp)
```



From the above, we can start making following possible tests

- Comparing the average tooth growth **by supplement irrespective of dose levels** (given that there are 10 observations each in both supplements at the 3 dosage levels)

- Comparison of average tooth growth **by supplement type and at same dose levels.**
- Comparison of average tooth growth **by dose levels irrespective of supplement types.**
- Comparison of average tooth growth **by dose levels for specific supplement type.**
- Doing the same as above for individual supplement types.

3 Assumptions

- It is not clear whether we have paired data or not - whether the tooth growth for different doses with the same supplement OR across different supplements are the same guinea pigs. We assume that this is not the case
- We are applying t-tests as the number of observations is small (<= 30)
- We have assumed independence between guinea pigs within the same supplement group and across groups.
- We are assuming unequal variances across groups for simplification purposes.

4 Hypothesis Tests

4.1 Hypothesis Tests for Supplements

We compare whether the difference of means of two supplements are indeed different with null hypothesis that they are same. $H_o : \mu_{oj} = \mu_{vc}$ and $H_a : \mu_{oj} != \mu_{vc}$

```
t.supp <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = df)
```

Since p-value 0.0606345 is above the 5% significance level and the 95% confidence interval (-0.1710156,7.5710156) has zero within it, we fail to reject null hypothesis.

4.2 Tests for Supplements at same dose levels

Next we examine whether there is a material difference between the supplements at same dosage levels. Essentiall our null hypothesis are

$H_o : \mu_{oj0.5} = \mu_{vc0.5}$ and $H_a : \mu_{oj0.5} != \mu_{vc0.5}$

$H_o : \mu_{oj1.0} = \mu_{vc1.0}$ and $H_a : \mu_{oj1.0} != \mu_{vc1.0}$

$H_o : \mu_{oj2.0} = \mu_{vc2.0}$ and $H_a : \mu_{oj2.0} != \mu_{vc2.0}$

The results are shown below using t-test

```
d0.5 <- filter(df, dose == 0.5)
d1.0 <- filter(df, dose == 1)
d2.0 <- filter(df, dose == 2)
t <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d0.5)
t0.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d1.0)
```

```r
t1.0result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "1.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d2.0)
t2.0result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])
dose.result1 <- rbind(t0.5result, t1.0result, t2.0result)
dose.result1
```

```
##            p.value  conf.low conf.high Ho.Rejected. diff.means
## 0.5 mg 0.006358607  1.719057  8.780943           No       5.25
## 1.0 mg 0.001038376  2.802148  9.057852           No       5.93
## 2.0 mg 0.963851589 -3.798070  3.638070          Yes      -0.08
```

4.3 Tests for Dose levels irrespective of Supplements

Now we test for statistical significance for dose levels (3 combinations) i.e. comparing mean growth of tooth when given 0.5 mg of any supplement versus 1 mg, 1 mg versus 2 mg and 0.5 mg versus 2 mg.

```r
d0.51.0 <- filter(df, dose != 2)
d1.02.0 <- filter(df, dose != 0.5)
d0.52.0 <- filter(df, dose != 1)
t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d0.51.0)
t0.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 vs 1.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d1.02.0)
t1.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "1.0 vs 2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d0.52.0)
t2.0result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 vs 2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])
dose.result2 <- rbind(t0.5result, t1.5result, t2.0result)
dose.result2
```

```
##                    p.value   conf.low  conf.high Ho.Rejected. diff.means
## 0.5 vs 1.0 mg 1.268301e-07 -11.983781  -6.276219           No     -9.130
## 1.0 vs 2.0 mg 1.906430e-05  -8.996481  -3.733519           No     -6.365
## 0.5 vs 2.0 mg 4.397525e-14 -18.156167 -12.833833           No    -15.495
```

4.4 Tests for Dose levels for OJ supplement

Now we test for statistical significance for dose levels (3 combinations) i.e. comparing mean growth of tooth when given 0.5 mg of supplement versus 1 mg, 1 mg versus 2 mg and 0.5 mg versus 2 mg. Supplement is set to OJ only

```r
d0.51.0 <- filter(df, dose != 2 & supp == "OJ")
d1.02.0 <- filter(df, dose != 0.5 & supp == "OJ")
d0.52.0 <- filter(df, dose != 1 & supp == "OJ")
t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d0.51.0)
t0.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 vs 1.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d1.02.0)
t1.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "1.0 vs 2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d0.52.0)
t2.0result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 vs 2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])
dose.result3 <- rbind(t0.5result, t1.5result, t2.0result)
dose.result3
```

```
##                    p.value   conf.low  conf.high Ho.Rejected. diff.means
## 0.5 vs 1.0 mg 8.784919e-05 -13.415634 -5.5243656           No      -9.47
## 1.0 vs 2.0 mg 3.919514e-02  -6.531443 -0.1885575           No      -3.36
## 0.5 vs 2.0 mg 1.323784e-06 -16.335241 -9.3247594           No     -12.83
```

4.5 Tests for Dose levels for VC supplement

Now we test for statistical significance for dose levels (3 combinations) i.e. comparing mean growth of tooth when given 0.5 mg of supplement versus 1 mg, 1 mg versus 2 mg and 0.5 mg versus 2 mg. Supplement is set to VC only

```r
d0.51.0 <- filter(df, dose != 2 & supp == "VC")
d1.02.0 <- filter(df, dose != 0.5 & supp == "VC")
d0.52.0 <- filter(df, dose != 1 & supp == "VC")
t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d0.51.0)
t0.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 vs 1.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d1.02.0)
t1.5result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "1.0 vs 2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])

t <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d0.52.0)
t2.0result <- data.frame(`p-value` = t$p.value, `conf-low` = t$conf[1], `conf-high` = t$conf[2],
    `Ho Rejected?` = ifelse(t$p.value >= 0.05, "Yes", "No"), row.names = "0.5 vs 2.0 mg",
    `diff-means` = t$estimate[1] - t$estimate[2])
dose.result4 <- rbind(t0.5result, t1.5result, t2.0result)
dose.result4
```

```
##                    p.value  conf.low  conf.high Ho.Rejected. diff.means
```

```
## 0.5 vs 1.0 mg 6.811018e-07 -11.26571  -6.314288           No      -8.79
## 1.0 vs 2.0 mg 9.155603e-05 -13.05427  -5.685733           No      -9.37
## 0.5 vs 2.0 mg 4.681577e-08 -21.90151 -14.418488           No     -18.16
```

5 Conclusions

- We failed to find complelling evidence for inferencing that average tooth growth will be higher when using one of the two supplement over the other (irrespective of dosage) at 5% significance level or within 95% confidence interval.
- However, At 5% significane level, we found evidence that average tooth growth by 2 mg dosage of VC is higher than that of OJ supplement. This was proved using 95% confidence interval. The sample difference was 0.08 unit length.
- At the same significance level, we failed to find any such statistically significant relationship for other dose level (1.0 or 0.5 mg) between VC and OJ supplements
- The average tooth growth in the sample data shows higher growth with higher dosages - but we have not found statistically significant result at 5% significance level to conclusively accept this for wider population for following three conditions - increasing dose irrespective of supplement type, increasing dose for OJ supplement, increasing dose for VC supplement.