Tooth Growth Analysis

Y. Zheng

Friday, October 14th, 2016

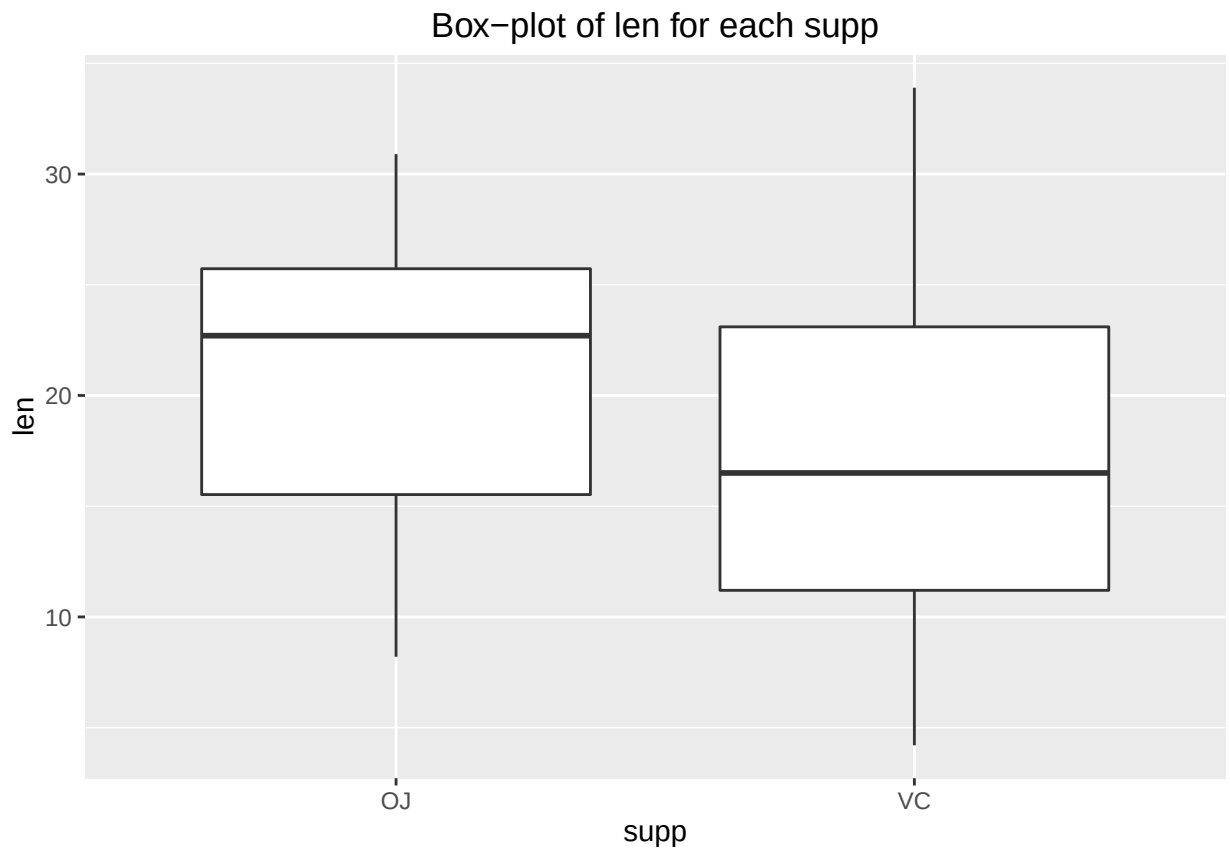
Loading Data

```
# Extracting Tooth Growth data and ggplot2 library
library(datasets)
library(ggplot2)
data(ToothGrowth)
```

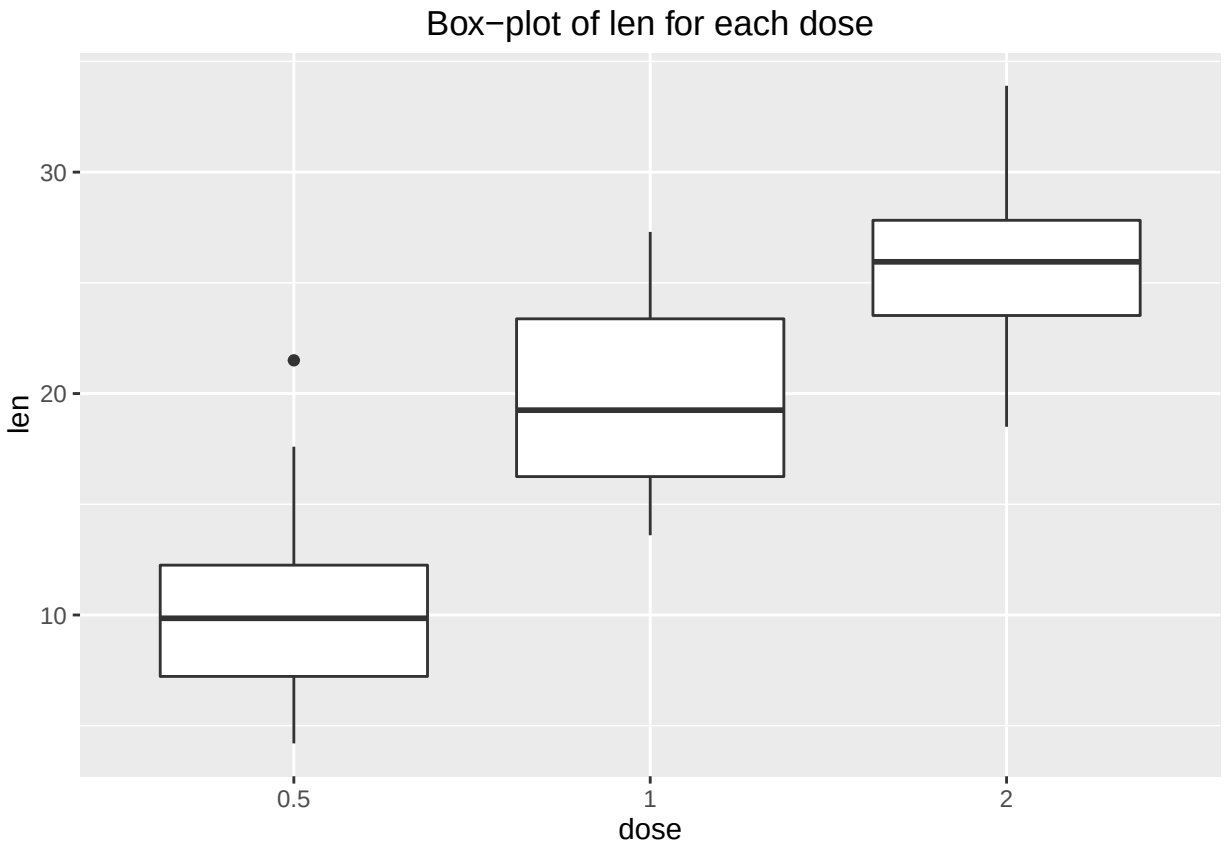
Summary of the data

We investigate the relationships between len and each supp as well as len and each dose via exploratory data analysis.

```
# Look at how len varies for each supp
qplot(supp, len, data = ToothGrowth, geom = "boxplot") +
    labs(title = "Box-plot of len for each supp")
```



```r
# Look at how len varies for each dose
ToothGrowth$dose = factor(ToothGrowth$dose)
qplot(dose, len, data = ToothGrowth, geom = "boxplot") +
    labs(title = "Box-plot of len for each dose")
```



From exploratory data analysis, we expect that OJ is the more effective supplement for tooth growth. We also expect higher dosage to correspond to increased tooth growth (len).

Hypothesis Testing

We will divide our test into 2 portions: one for supplement and one for dosage.

Supplement Analysis

Definitions:

- μ_{OJ} = population mean of len using OJ supplement
- μ_{VC} = population mean of len using VC supplement

From our exploratory plots, we make the following hypothesis:
$H_0 : \mu_{OJ} - \mu_{VC} \leq 0$
$H_1 : \mu_{OJ} > \mu_{VC}$

```r
t.test(len ~ supp, paired = FALSE, var.equal = FALSE,
       data = ToothGrowth, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

p-value = 0.03032
95% confidence interval = [0.4682687, ∞]

The 95% confidence interval is entirely positive.
The p-value perspective tells us that the ¶$(Observation|H_0) = 0.03032 < 0.05$.
We have sufficient evidence to reject H_0 in preference for $H_1 : \mu_{OJ} > \mu_{VC}$.

Dosage Analysis

Definitions:

```
* ${\mu}_{0.5}$ = population mean of len with 0.5 dose
```

- μ_1 = population mean of len with 1 dose
- μ_2 = population mean of len with 2 dose

Similarly, we expect 1 dose has greater len than 0.5 dose:
$H_0 : \mu_{0.5} - \mu_1 \geq 0$
$H_1 : \mu_{0.5} - \mu_1 < 0$

```r
t.test(len ~ dose, paired = FALSE, var.equal = FALSE,
       data = ToothGrowth[ToothGrowth$dose!=2,], alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

p-value = 6.342e-08

The probability of us observing the data given H_0 as prior is vanishingly small.

We have sufficient evidence to reject H_0 in preference for $H_1 : \mu_{0.5} - \mu_1 < 0$.

From the exploratory plot, we expect 2 dose has greater len than 1 dose: $H_0 : \mu_1 - \mu_2 \geq 0$ $H_1 : \mu_1 - \mu_2 < 0$

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE,
       data = ToothGrowth[ToothGrowth$dose!=0.5,], alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

p-value = 9.532e-06

The probability of us observing the data given H_0 as prior is vanishingly small.
We have sufficient evidence to reject H_0 in preference for $H_1 : \mu_1 - \mu_2 < 0$.

By the *transitive property of inequality*, we conclude that $\mu_2 > \mu_1 > \mu_{0.5}$.

Conclusions and Assumptions

Assumptions

The above analysis makes the following assumptions:

- len is normally distributed
- Each sample is an independent and random observation
- Variables between factors are uncorrelated - this justifies the use of unpaired t-test

We use weighted variances and did not assume equal variance.
This applies to equal variances without loss of generality (WLOG).

Conclusions

- **Dosage effectiveness in terms of tooth growth is in this order: 0.5 dose < 1 dose < 2 dose**
- **Supplement OG is more effective than supplement VC in assisting tooth growth**