Analysis of the Tooth Growth data in R

Dung Nguyen

August 18, 2017

Overview

In this analysis, I explored the ToothGrowth data from the datasets package in R. According to R Documentation, this data contains measures of the length (len) of odotoblasts (cells that are responsible for tooth growth) in 60 guinea pigs. There are 2 groups of pigs that were either received vitamin C (supp) via orange juice (coded as OJ) or ascorbic acid (coded as VC). Each pig received one of three dose levels of vitamin C (0.5, 1, 2 mg/day). A quick look at the data in R confirms the information given by the Document.

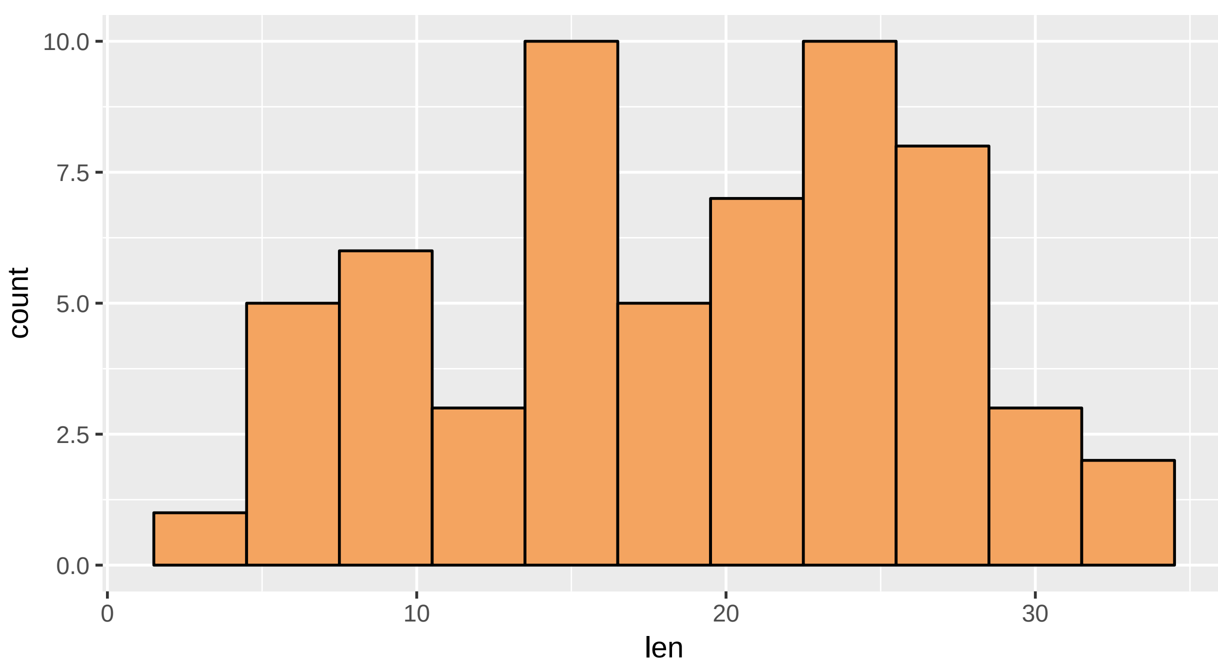
```
library(datasets)
data("ToothGrowth")
str(ToothGrowth)
```

```
## '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 ...
```

Data

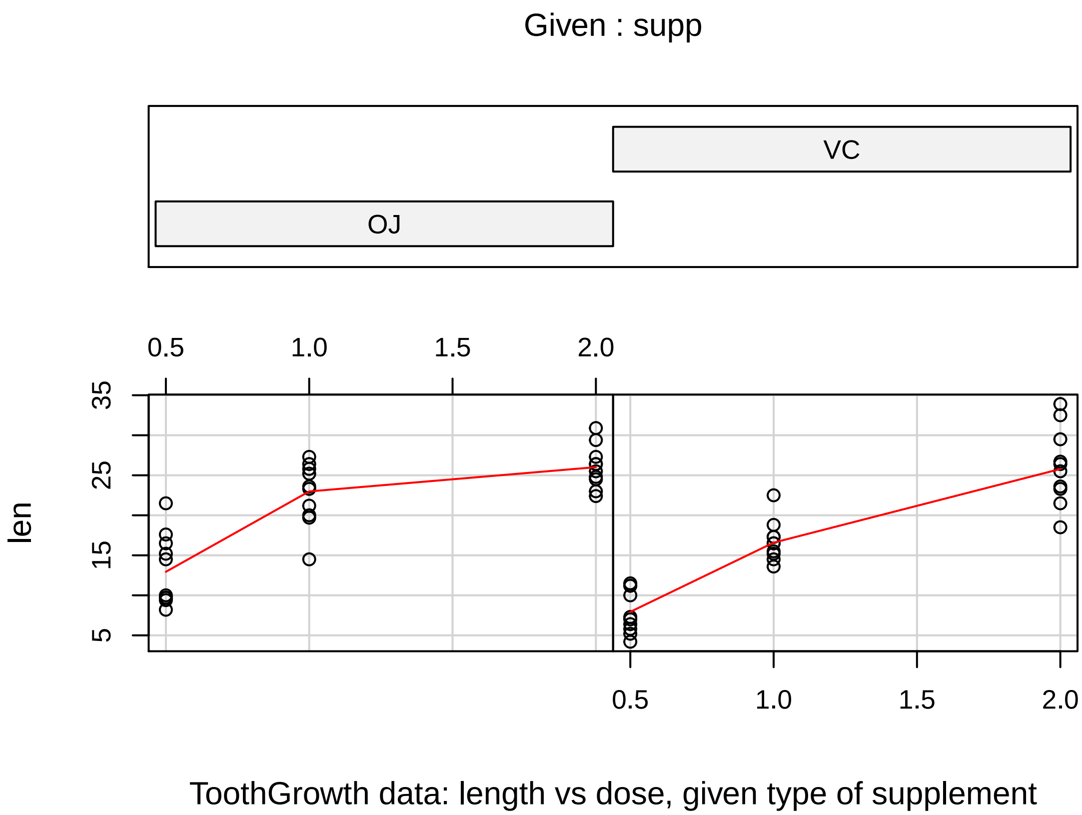
Quick plots of the distributions of responses in 2 different groups of OJ and VC and in 3 different dose levels of 0.5, 1, and 2 mg/day suggests that the distributions of the responses are not normal.

```
library(ggplot2)
qplot(len, data = ToothGrowth, fill=I("sandybrown"), col = I("black"), binwidth = 3)
```



The coplot suggests that the length of odotoblasts seems, in general, shorter among pigs that consumed vitamin C via ascorbic acid than that of those via orange juice. In addition, higher doses may lead to longer odotoblasts in guinea pigs.

```
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



I want to test 3 hypotheses:

Hypothesis 1: The length of odotoblasts in pigs that were given vitamin C via orange juice is longer than that in pigs given via ascorbic acid.

Hypothesis 2: Pigs on 0.5mg of vitamin C per day have shorter odotoblasts than those on 1mg/day.

Hypothesis 3: Pigs on 1mg of vitamin C per day have shorter odotoblasts length than those on 2mg/day.

Results

Hypothesis 1: OJ vs VC

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

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.0302
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  0.4708204       Inf
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The mean in group OJ is significantly larger than the mean in group VC with a p-value under the null hypothesis of no difference of 0.03, smaller than 0.05. Therefore, I reject the null hypothesis and conclude that pigs got vitamin C via orange juice have significantly longer odotoblasts than those did via ascorbic acid at 5% level of significance.

Hypothesis 2: 0.5mg vs 1mg

```
dose1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
t.test(len ~ dose, paired = FALSE, alternative = "less", var.equal = TRUE, data = dose1)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 6.331e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753344
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

The mean in group 0.5mg/day is smaller than the mean in group 1mg/day, p-value = 6.331e-08 < 0.05. As a result, I reject the null hypothesis of no mean difference and conclude that pigs with 0.5mg/day of vitamin C have significantly shorter odotoblasts in pigs than those with 1mg/day of vitamin C at 5% level of significance.

Hypothesis 3: 1mg vs 2mg

```
dose2 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = FALSE, alternative = "less", var.equal = TRUE, data = dose2)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 9.054e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -4.175196
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The mean in group 1mg/day is smaller than the mean in group 2mg/day, p-value = 9.054e-06 < 0.05. As a result, I reject the null hypothesis of no mean difference and conclude that pigs on 1mg/day of vitamin C have significantly shorter odotoblasts in pigs than those on 2mg/day of vitamin C at 5% level of significance.

Discussion:

Assumptions

I assumed that 3 different dose levels and 2 different methods of delivering vitamin C were randomized among pigs. Therefore, conclusions about mean difference is reliable. It is also reasonable that the variances of responses should be the same in those groups if subjects were randomized.

One thing to keep in mind is false positives due to multiple testing. Here I performed 3 tests at 5% level of significance. If the Bonferroni correction is used to adjust the significance level from 5% to 1.67% for each test, the null hypotheses of no difference between 0.5mg and 1mg dose level and no difference between 1mg and 2mg dose level would still be rejected since their p-values are close to 0 (6.331e-08 and 9.054e-06 respectively). However, the null hypothesis 1 will be failed to reject.

However, since the responses are not normally distributed, the inference made by t-test may be inaccurate. I decided to use a nonparametric procedure to compare the above results. Permutation tests are shown below:

Hypothesis 1

```
y <- ToothGrowth$len
group <- as.character(ToothGrowth$supp)
testStat <- function(w, g) mean(w[g == "OJ"]) - mean(w[g == "VC"])
observedStat <- testStat(y, group)
permutations <- sapply(1 : 10000, function(i) testStat(y, sample(group)))
g1 = ggplot(data.frame(permutations = permutations), aes(permutations))
g1 = g1 + geom_histogram(fill = "lightblue", color = "black", binwidth = 1)
g1 = g1 + geom_vline(xintercept = observedStat, size = 1)
mean(permutations > observedStat)
```

```
## [1] 0.0316
```

The p-value calculated using permutation test is 0.03, which is less than 5% level of significance. Since this p-value is a binomial variable, quickly calculating the 95% confidence interval of the p-value using Wald confidence interval in my head results from 0.02 to 0.04, still less than 0.05. Therefore, this confirms the result above

Hypothesis 2 and 3

```
y <- dose1$len
group <- as.numeric(dose1$dose)
testStat <- function(w, g) mean(w[g == 0.5]) - mean(w[g == 1])
observedStat <- testStat(y, group)
permutations <- sapply(1 : 10000, function(i) testStat(y, sample(group)))
g2 = ggplot(data.frame(permutations = permutations), aes(permutations))
g2 = g2 + geom_histogram(fill = "lightblue", color = "black", binwidth = 1)
g2 = g2 + geom_vline(xintercept = observedStat, size = 1)
mean(permutations < observedStat)
```

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

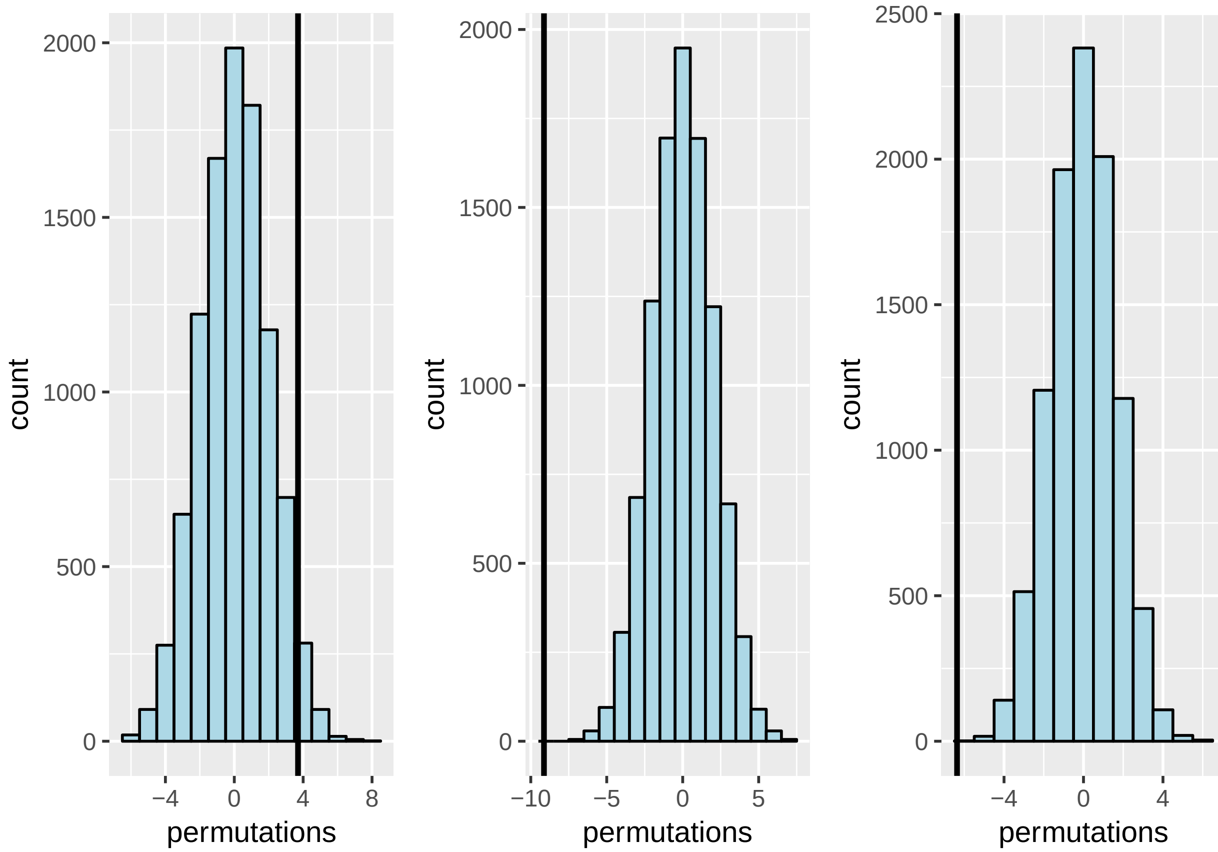
```
y <- dose2$len
group <- as.numeric(dose2$dose)
testStat <- function(w, g) mean(w[g == 1]) - mean(w[g == 2])
```

```r
observedStat <- testStat(y, group)
permutations <- sapply(1 : 10000, function(i) testStat(y, sample(group)))
g3 = ggplot(data.frame(permutations = permutations), aes(permutations))
g3 = g3 + geom_histogram(fill = "lightblue", color = "black", binwidth = 1)
g3 = g3 + geom_vline(xintercept = observedStat, size = 1)
mean(permutations < observedStat)
```

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

Permutation tests estimate p-values under the null hypotheses of no difference between 0.5 and 1 dose level and between 1 and 2 dose level to be both 0, thus confirming the results from performing t-tests.

```r
library(gridExtra)
grid.arrange(g1, g2, g3, ncol=3)
```



Conclusion

There is evidence supporting that higher doses of vitamin C results in greater lengths of odotoblasts, cells that are responsible for tooth growth, in guinea pigs. Vitamin C consumed via orange juice may be responsible for greater length of odotoblasts compared to that via ascorbic juice, although this conclusion of the different effect between vitamin C consumption via orange juice and ascorbic acid may need more investigations.