Analysis of Toothgrowth Data

Nitin Aggarwal

September 26, 2015

Data

Let us load the `ToothGrowth` data as follows.

```
data("ToothGrowth")
```

A description on the help server on R about the `ToothGrowth` dataset shows following description:

The response is the 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).

Assumtion Although it is mentioned that there are just 10 guinea pigs because the indicators are not given about which one is which, we are going to take the observation as coming from independent sample, i.e. we will assume that there are 60 guinea pigs instead of just 10.

Let us look at the dimensions and the head of the data to see what we are dealing with here.

```
dim(ToothGrowth)
```

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

```
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
```

Exploration

Let us have a look at the summary of this dataset using the R function `summary`.

```
summary(ToothGrowth)
```

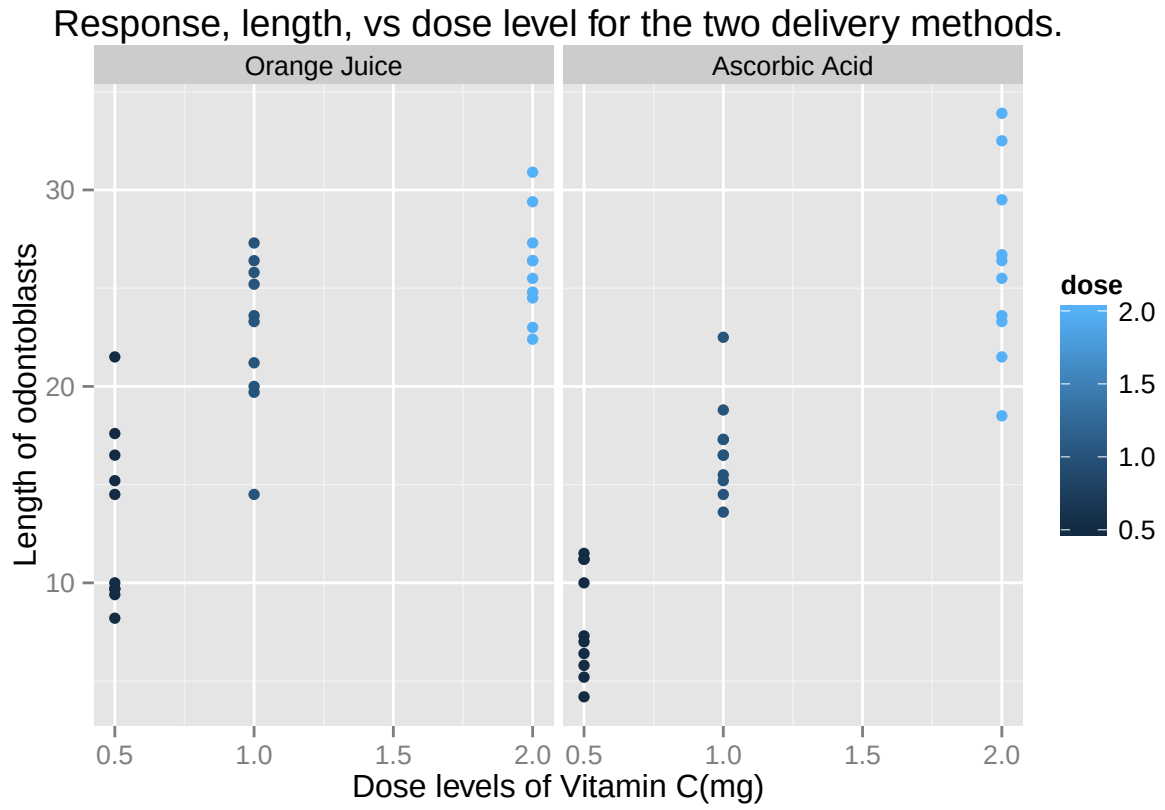
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

So, the column `supp` must be a factor class columns with the two classes OJ, VC representing the two delivery methods **orange juice** or **ascorbic acid**. Let us make some modifications in leu of this.

```
levels(ToothGrowth$supp)<-c("Orange Juice", "Ascorbic Acid")
```

Let us draw a facet plot distinguishing the responses of the dosages given to the guinea pigs using these two methods as follows.

```
library(ggplot2)
ggplot() +
  geom_point(data = ToothGrowth,
             aes(x = dose, y = len, colour = dose, type = supp)) +
  xlab("Dose levels of Vitamin C(mg)") +
  ylab("Length of odontoblasts")  +
  ggtitle("Response, length, vs dose level for the two delivery methods.") +
  facet_grid(.~supp)
```



So, in this 60x3 table, we have three columns one is the response of the test in column named `len` representing the length of odontoblasts(teeth) in each of 10 guinea pigs. Now, this is response to two features first is the dosage of Vitamin C mentioned in the column called `dose` taking values (0.5, 1, and 2mg) and the column `supp` to give the delivery method as discussed earlier.

The increase in dosage of Vitamin C indicates the increase in the response in general. We will see this in numbers in the next section.

Analysis of the data

The plot in the previous section suggests following two hypothesis:

1. The length of odontoblasts(teeth) in guinea pigs increases with the increase in dosage of Vitamin C.
2. The delivery method of Vitamin C is more effective via orange juice versus that via ascorbic acid.

Let us study these two hypothesis separately as follows;

Hypothesis 1: Effect of dosage

First we test the hypothesis mentioned above: *Increase in dosage of Vitamin C increases the length of odontoblasts(teeth).*

If we denote the means of the length of teeth for the three different dose levels of Vitamin C, 0.5, 1, and 2mg as m1, m2, and m3 respectively. Then we can test three alternate hypothesis as follows:

1. H1_a: m1 $<$ m2,
2. H2_a: m2 $<$ m3, and
3. H3_a: m1 $<$ m3,

against the null hypothesis:

1. H1_0: m1 $=$ m2
2. H2_0: m2 $=$ m3, and
3. H3_0: m1 $=$ m3

We automatically get the third hypothesis if we select the first two. Let us do this testing for the level of testing alpha as 0.05.

Let us extract the vectors for the lengths for different doses and store them into the variables `x1`, `x2`, and `x3`, respectively.

```
x1 <- ToothGrowth$len[ToothGrowth$dose==0.5]
x2 <- ToothGrowth$len[ToothGrowth$dose==1]
x3 <- ToothGrowth$len[ToothGrowth$dose==2]
```

Now, the t-tests for the above hypotheses are as follows:

```
t.test(x1, x2, alternative = "less", mu = 0, paired = FALSE,
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x1 and x2
## 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 of x mean of y
##    10.605    19.735
```

```
t.test(x2, x3, alternative = "less", mu = 0, paired = FALSE,
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x2 and x3
## 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 of x mean of y
##    19.735    26.100
```

```
t.test(x1, x3, alternative = "less", mu = 0, paired = FALSE,
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x1 and x3
## t = -11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -13.27926
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

Verdict These p-values suggests that we can reject all of the null hypothesis and accept the alternate hypothesis, in short proving that increase in dose of Vitamin C increases the lengths of teeth.

Hypothesis 2: Effect of delivery methods

About the delivery method we can hypothesise only one thing and that is Orange Juice is a better medium than Ascorbic Acid for any dose level.

So, if we break the data into six parts as follows:

```
supp_levels <- levels(ToothGrowth$supp)
x1OJ <- ToothGrowth$len[ToothGrowth$supp == supp_levels[1] & ToothGrowth$dose == 0.5]
x2OJ <- ToothGrowth$len[ToothGrowth$supp == supp_levels[1] & ToothGrowth$dose == 1]
x3OJ <- ToothGrowth$len[ToothGrowth$supp == supp_levels[1] & ToothGrowth$dose == 2]
x1VC <- ToothGrowth$len[ToothGrowth$supp == supp_levels[2] & ToothGrowth$dose == 0.5]
x2VC <- ToothGrowth$len[ToothGrowth$supp == supp_levels[2] & ToothGrowth$dose == 1]
x3VC <- ToothGrowth$len[ToothGrowth$supp == supp_levels[2] & ToothGrowth$dose == 2]
```

Let us use a similar representation for the means also. For example mean for the response for dose level 0.5 and delivery method orange juice as mu1OJ, etc.

So, based on these we can divide the hypothesis for this section into three separate part, namely Orange Juice is more effective than Ascorbic Acid at dose level 0.5mg, 1mg, and 2mg respectively. For this the alternate hypothesis for the null hypothesis Hj_0: mujOJ = mujVC is Hj_a: mujOJ > mujVC for j = 1, 2, 3.

The t-tests for these hypothesis are as follows:

```
t.test(x1OJ, x1VC, alternative = "greater", mu = 0, paired = FALSE,
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x1OJ and x1VC
## 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
```

```
t.test(x2OJ, x2VC, alternative = "greater", mu = 0, paired = FALSE,
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x2OJ and x2VC
## 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
```

```
t.test(x3OJ, x3VC, alternative = "greater", mu = 0, paired = FALSE,
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x3OJ and x3VC
## 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
```

So, the null hypothesis is rejected at 1% level of significance, when the dose levels are 0.5mg and 1mg. But it is not rejected when dose level is 2mg.

Verdict Orange Juice is more effective than Ascorbic Acid at increasing the length of the teeth when the dose level is 0.5mg or 1mg. This is not the case when dose level is 2mg.

Conclusion

Based on the hyposthesis testing done above we found the following:

1. Increase in dose level of Vitamin C increases the length of teeth.
2. Orange Juice is more effective than Ascorbic Acid at increasing the length of the teeth when the dose level is 0.5mg or 1mg.

One assumption taken in this data is the independence of the observation inspite of the mention of relationship in the absense of the indicators.