ToothGrowth.Rmd

Beau Norgeot

May 21, 2015

we're going to analyze the ToothGrowth data in the R datasets package. This data set contains results examing the Effect of Vitamin C on Tooth Growth in Guinea Pigs. 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).

Parts 1 & 2: Exploratory Analysis and Basic Summary

```
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
```

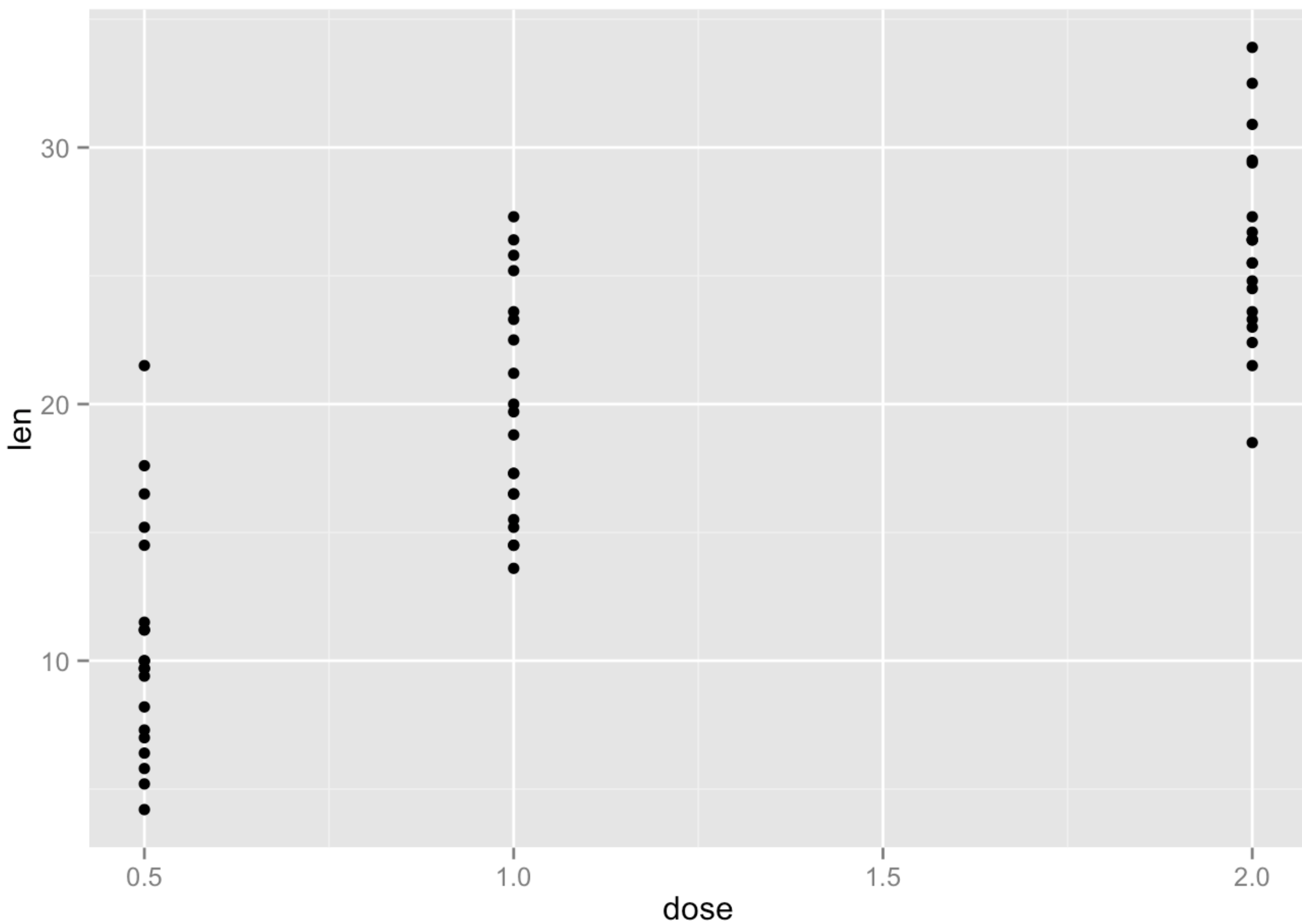
```
summary(ToothGrowth)
```

```
##       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  
```

From the above Exploratory Analysis, we can see that the ToothGrowth data set is composed of 60 observations of 3 variables. I can't determine the units that the length of the teeth are measured in, but tooth length is obviously the 'len' variable with a mean of 18.81. Some asshole decided to label the Delivery Method variable 'supp' just to confuse everyone, and further decided that 'VC' would be a good factor name for pure ascorbic acid (ascorbic acid IS Vitamin C, while OJ CONTAINS Vitamin C). This is where statistians get their bad rep from.. Dose is pretty straight forward, except we can see that 'dose' is a numeric type. Since dose levels are descrete and take only 3 values, which we probably care about, dose should probably be converted to a factor.

Let's take a quick look at Length vs Dose

```
library(dplyr)
library(ggplot2)
TG <- tbl_df(ToothGrowth)
LvD <- ggplot(TG,aes(x=dose, y=len))
LvD + geom_point()
```



Looks like dosing amount has a postive correlation with tooth length.

Let's get some actual values to compare

```
LvDdata <- TG %>%
  group_by(dose) %>%
  summarise(n=n(),meanGROWTH = mean(len), SD = sd(len))
```

Yup, we can see that growth increases with dose. Tooth growth doubles when dose doubles from .5mg to 1mg, but only increases by 60% when dose increases from 1mg to 2mg.

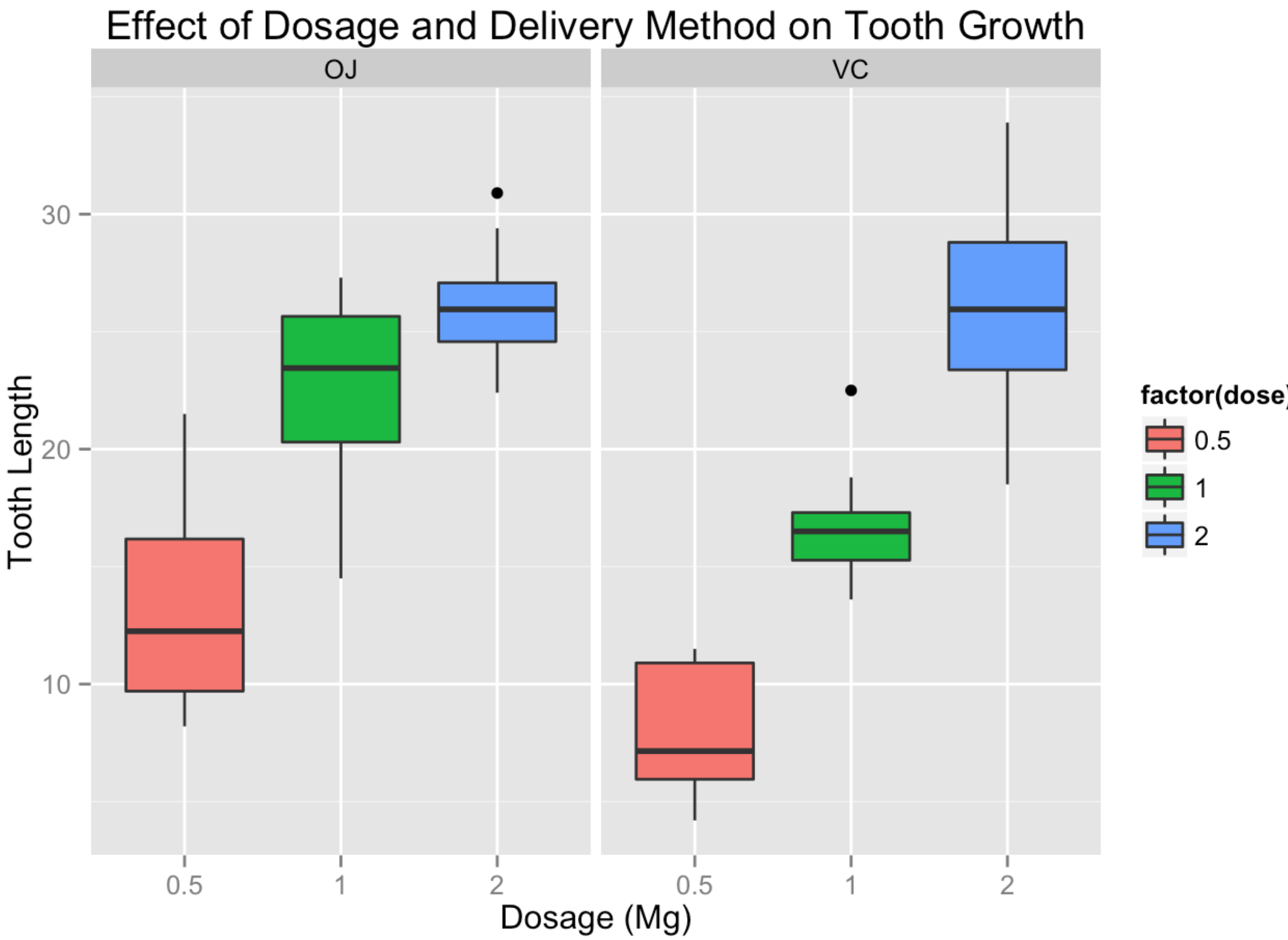
Let’s see if delivery method impacts growth

```
LvDMdata <- TG %>%
  group_by(supp) %>%
  summarise(n=n(), meanGROWTH = mean(len),SD = sd(len))
```

Delivery method does indeed impact growth. Growth when OJ was used to delivery the Vitamin C was quite a bit higher. Perhaps the sugar provided extra energy for growth?

Let’s visualize Delivery Method vs Dose for tooth growth

```
plot <- ggplot(TG,aes(x=factor(dose),y=len,fill=factor(dose)))
plot + geom_boxplot(notch=F) + facet_grid(.~supp) +
     scale_x_discrete("Dosage (Mg)") +
     scale_y_continuous("Tooth Length") +
     ggtitle("Effect of Dosage and Delivery Method on Tooth Growth")
```



Part 3: Compare tooth growth by Delivery Method and Dose Volume

Unfortunately, this data set doesn't include any controls, so we can't even ask the biggest question which would be 'does Vitamin C effect tooth growth'? However, given the data, we can ask some other questions.

1. Does the volume of dose effect tooth growth?
2. Does Delivery Method effect tooth growth?
3. Does the effectiveness of a given Delivery Method vary w/dosage?

T-Tests are a relatively common method for answering questions of this kind. Our Null Hypothesis will always assume equality, a p-value of less than .05 will be considered sufficient to reject the Null Hypothesis and assume that the element in question has SOME effect on tooth growth.

Some Assumptions

- All of our sample data is normally distributed
- The sample data is not paired, ie all measurements are taken on different pigs. No pig was measured twice under different conditions.

Dosage Effects

Our Null Hypothesis is that dosage has no effect, our Alternative Hypothesis will be that dosage effects tooth growth. We need to make 3 comparisons: .5mg to 1mg, .5mg to 2mg, 1mg to 2mg

```
low <- TG %>% filter(dose == 0.5)
medium <- TG %>% filter(dose == 1.0)
high <- TG %>% filter(dose == 2.0)

lowVSmedium <- t.test(low$len,medium$len); lowVSmedium; lowVSmedium$p.value
```

```
##
##  Welch Two Sample t-test
##
## data:  low$len and medium$len
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
## [1] 1.268301e-07
```

```
lowVShigh <- t.test(low$len,high$len); lowVShigh; lowVShigh$p.value
```

```
##
##  Welch Two Sample t-test
##
## data:  low$len and high$len
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```
## [1] 4.397525e-14
```

```
mediumVShigh <- t.test(medium$len,high$len); mediumVShigh; mediumVShigh$p.value
```

```
##
##  Welch Two Sample t-test
##
## data:  medium$len and high$len
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

```
## [1] 1.90643e-05
```

Dosage Conclusions

The p.values for all trials is extremely small, we can reject the Null Hypotheis and infer that dosage has an effect on tooth growth

Delivery Method

Our Null Hypothesis is that the two methods have the same effect, our Alternative Hypothesis will be that method of delivery matters.

```
OJ <- TG %>% filter(supp == "OJ", !is.na(len))
VC <- TG %>% filter(supp == "VC", !is.na(len))
DeliveryResult <- t.test(OJ$len,VC$len); DeliveryResult; DeliveryResult$p.value
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

```
## [1] 0.06063451
```

Delivery Method Initial Conclusions

Not really conclusive. A p.value of .061 isn't enough to reject the Null, however it is far lower than random chance and my initial investigation certainly made me think that the method matter (favoring OJ). Let's go on to question 3 and see if that clears anything up.

Does the effectiveness of a given Delivery Method vary w/dosage

Our Null Hypothesis is that the delivery methods will have the same effect regardless of dosage, our Alternative Hypothesis will be that one method is more effective than the other at some dosage levels. We'll compare delivery methods with equal doses

```
OJlow <- TG %>% filter(supp == "OJ", !is.na(len), dose == 0.5)
VClow <- TG %>% filter(supp == "VC", !is.na(len), dose == 0.5)
lowDoses <- t.test(OJlow$len,VClow$len); lowDoses
```

```
##
##  Welch Two Sample t-test
##
## data:  OJlow$len and VClow$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```
OJmed <- TG %>% filter(supp == "OJ", !is.na(len), dose == 1.0)
VCmed <- TG %>% filter(supp == "VC", !is.na(len), dose == 1.0)
medDoses <- t.test(OJmed$len,VCmed$len); medDoses
```

```
##
##  Welch Two Sample t-test
##
## data:  OJmed$len and VCmed$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
OJhigh <- TG %>% filter(supp == "OJ", !is.na(len), dose == 2.0)
VChigh <- TG %>% filter(supp == "VC", !is.na(len), dose == 2.0)
highDoses <- t.test(OJhigh$len,VChigh$len); highDoses
```

```
##
##  Welch Two Sample t-test
##
## data:  OJhigh$len and VChigh$len
## t = -0.0461, 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
```

```
values <- c(lowDoses$p.value, medDoses$p.value, highDoses$p.value); values
```

```
## [1] 0.006358607 0.001038376 0.963851589
```

Conclusions about Dosage effect on Delivery

The p.values for low/med/high dosage comparing OJ vs VC were [0.006358607 0.001038376 0.963851589]. From this we can infer that dosage plays a significant role the effectives of the delivery method when the dosage is below 2mg, and that tooth growth is delivery method agnostic at 2mg.

Wrap Up

Increasing Vitamin C dosage changes the effect of tooth growth for all dosage levels. Delivery method effects tooth growth when the dosage level is below 2mg.