

An Exploration of Tooth Growth Data

barb dornseif - saoirsegirl

February, 2015

1 Loading and Summarization of Data Set: ToothGrowth

Let's load our data and look at a summary and the structure of its contents.

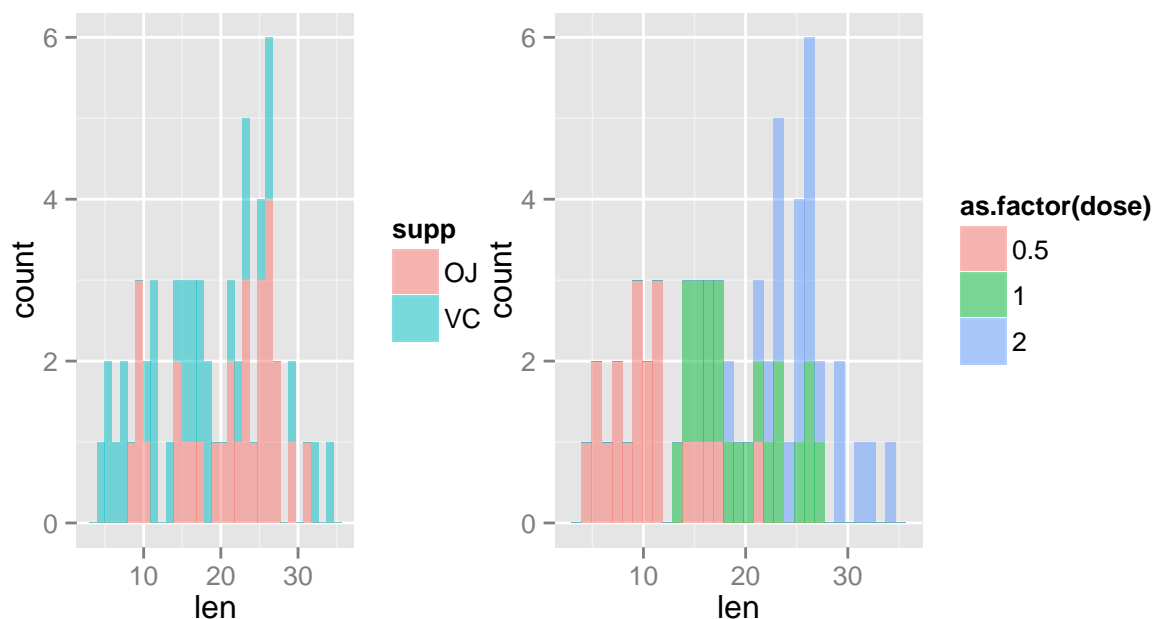
```
data(ToothGrowth); str(ToothGrowth); summary(ToothGrowth)
```

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

```
##      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
```

We can see there are 60 observations of 3 columns, that the first and third columns are numerics - and the second column is a factor with two levels. The summary further tells us the quantile breakdown of each of the numeric columns. If we lookup the short documentation on the dataset [link](#) we see the 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).”

So we know that we have a growth in the length of teeth for six(6) groupings of 10 guinea pigs where each group is given one of three(3) dosage amounts via one or the other of two(2) delivery methods (column = “supp”). Let's take a look at how length groups by dose and delivery separately.



We can see by the clustering by color, that there are distinct distributions of lengths by dose (when used as a factor variable) and less so by delivery. And although we see there is some overlap between each grouping, it is pretty safe to say that the greater the dose, the greater the efficacy. Our first hypothesis is that the overlaps may suggest that we are not sure how or if the delivery method impacts this simple observation. So we will test each dosage against the two delivery methods using a standard t-Test to determine if the variability is intrinsic to the dosage level, or is impacted by the delivery method.

Because we see that the spread of each density is different as well, we will need to assume that the variance for each group varies, even though the number of observations in each sample is the same; this will impact the results generated by the t-Test.

2 Using Means and Confidence Intervals to Compare Growth by Delivery and Dosage.

So let's start by isolating the dosage (2.0 mg) that appears to give the best growth results, and see if the delivery method shows any distinction in efficacy. The groups are sorted alphabetically, so OJ will take the position of having the larger mean with which to test VC against. Let's see if this is true for a dosage of 2.0 mg.

| | OJ = Orange Juice | VC = Asorbic Acid |
|------------------------------|-------------------|-------------------|
| Means of each Group | 26.06 | 26.14 |
| Confidence of the Difference | -3.80 | 3.64 |

Table 1: Means and Confidence Intervals for Dosage 2.0 mg

Using the t-test to derive the Means of two groups of equal size, and the Confidence Interval of the difference between those Means when both receive the same dosage of 2.0 mg. we see that there is little distinction between Mean tooth length when receiving orange juice (OJ) versus Asorbic Acid (VC) as the method of delivery. Further, the Confidence Interval clearly straddles zero, so we must conclude that at this dosage, method of delivery is not a relevant factor the growth outcome.

But given the observed histograms we cannot conclude that this holds true for all dosages. So let's try the 0.5 mg dosage, which shows the tightest clustering in the histogram.

| | OJ = Orange Juice | VC = Asorbic Acid |
|------------------------------|-------------------|-------------------|
| Means of each Group | 13.23 | 7.98 |
| Confidence of the Difference | 1.72 | 8.78 |

Table 2: Means and Confidence Intervals for Dosage 0.5 mg

Here we see the different relationship. Although the Mean itself is significantly lower than at the higher dosage, using OJ provides a significantly higher mean growth result (by 5.25) and the confidence interval is significantly above zero. Here, we can conclude that delivery method does impact efficacy and that orange juice is the preferred method.

Given such divergent outcomes, let's try our third dosage (1.0 mg) and see which camp it most resembles.

| | OJ = Orange Juice | VC = Asorbic Acid |
|------------------------------|-------------------|-------------------|
| Means of each Group | 22.70 | 16.77 |
| Confidence of the Difference | 2.80 | 9.06 |

Table 3: Means and Confidence Intervals for Dosage 1.0 mg

Interestingly, the 1.0 mg dosage not only appears to provide a better mean efficacy than the 0.5 mg dosage it also shows a greater distinction in efficacy between the two delivery methods with OJ providing higher

measured growth by 5.93 and with a narrower Confidence Interval that the actual difference in the population Means will consistently bear this out. However the amount of growth for OJ delivered Vitamin C does not exceed either delivery method of the 2.0 dosage. So we are left with a conondrum given we have no data on the doage 1.5 mg which lies between these two groups.

3 Conclusions

Not knowing the cost of each of the dosages by delivery method, the impacts of 1.5 mg dosing nor the potential negative impacts of overdosing on the test subjects, it would appear that using orange juice at the dosage of 1.0 mg is the most reliable method and dosage to acheive the greatest potential impact on tooth growth for the least amount of intervention/cost.

4 Appendix - Code Chunks Using `echo = FALSE`

Echo=TRUE is replaced with eval=TRUE

```
library(ggplot2)
ggplot(ToothGrowth, aes(len, fill = supp)) + geom_histogram(alpha = 0.5)

ggplot(ToothGrowth, aes(len, fill = as.factor(dose))) + geom_histogram(alpha = 0.5)
```

Manipulations of ToothGrowth data - repeated three times; once for each dosage grouping.

```
TGs2 <- subset(ToothGrowth, dose %in% 2)
TGs2 <- t.test(len~supp, paired = FALSE, var.equal=FALSE, data=TGs2) # grab the output of the function
TGs2c <- c(as.vector(TGs2$conf[1]), as.vector(TGs2$conf[2]))
TGs2m <- c(as.vector(TGs2$estimate[1]), as.vector(TGs2$estimate[2]))

library(xtable)
table1 <- rbind( TGs2m, TGs2c)
rownames(table1) <- c("Means of each Group", "Confidence of the Difference")
colnames(table1) <- c("OJ = Orange Juice", "VC = Asorbic Acid")
print(xtable(table1, format = "markdown", caption = "Means and Confidence Intervals for Dosage 2.0 mg"))
```