

ToothGrowth Data Analysis

Tim Chopoorian

December 27, 2015

Overview

The following report explores the ToothGrowth data available in R. The data are measures of the length of odontoblasts (teeth) in each of 60 guinea pigs that were given one of three dose levels of Vitamin C, 0.5, 1, and 2 mgs, via one of two delivery methods, orange juice (OJ) and ascorbic acid (VC). We're interested in discovering which combinations of delivery method and dosage, if any, affect tooth growth.

Load Data

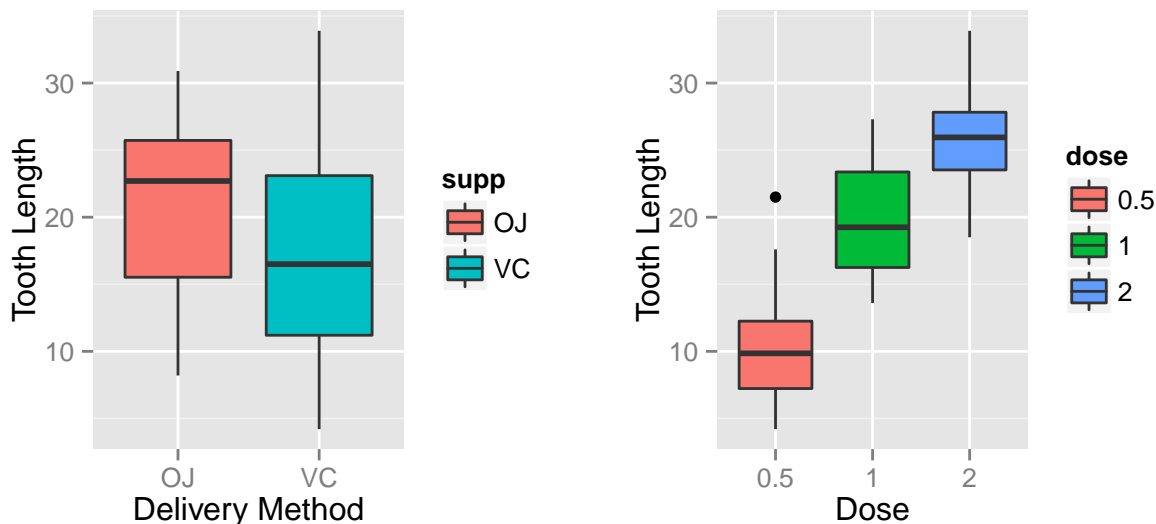
First, we'll load the data, require some libraries, and take a quick look at the structure of our data.

```
data("ToothGrowth")
library(ggplot2); library(broom); library(pander); library(pastecs); library(grid); library(lattice)
panderOptions('table.alignment.default', 'left'); panderOptions('table.alignment.rownames', 'left')
panderOptions('table.style', 'grid'); panderOptions('table.split.cells', 50); panderOptions('table.continues',
panderOptions('table.split.table', 100)
dat <- ToothGrowth
str(dat)
```

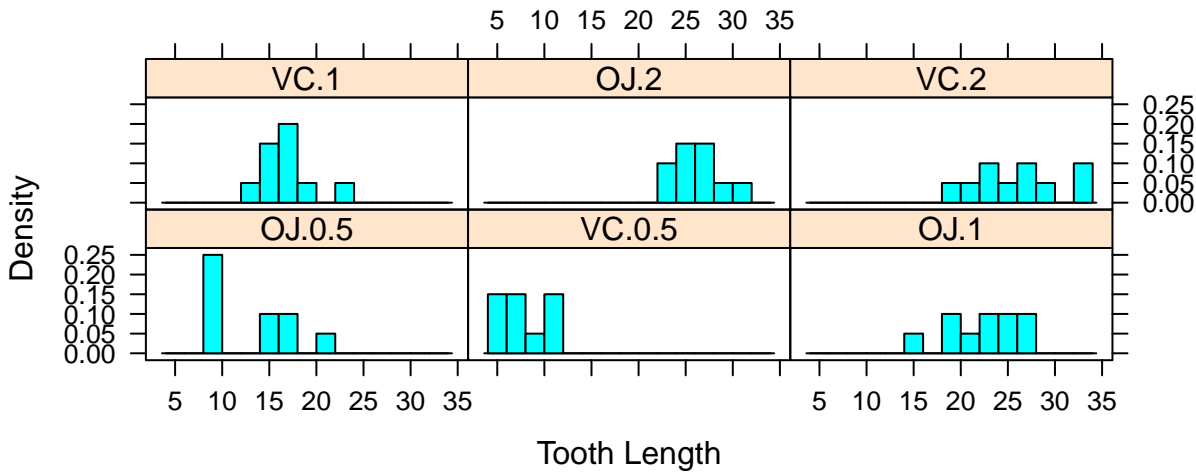
```
## 'data.frame':    60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
dat$dose <- as.factor(dat$dose) # convert dose levels to factor
```

Exploratory Data Analysis



Tooth Growth by Delivery & Dose



From our preliminary exploration, the boxplots seem to indicate that tooth length increases with dosage, while the effect of delivery method is not so clear. The boxplots also show that most of the data is roughly symmetrical with even distribution around the mean (except the OJ plot), suggesting normality. The shape of the density curves for each combination of delivery method and dosage tell us that our data is roughly Gaussian. The histograms also support the hypothesis that tooth length increases with dosage level. (See Appendix 1 for code to produce above plots.)

Basic Summary of Data

% Data Summary by Delivery Method Across All Dosages

	VC	OJ
median	16.5	22.7
mean	16.96	20.66
SE.mean	1.509	1.206
CI.mean.0.95	3.087	2.467
var	68.33	43.63
std.dev	8.266	6.606
coef.var	0.4873	0.3197

% Data Summary For Each Pair of Delivery Method and Dosage

	vc05	vc1	vc2	oj05	oj1	oj2
median	7.15	16.5	25.95	12.25	23.45	25.95
mean	7.98	16.77	26.14	13.23	22.7	26.06
SE.mean	0.8686	0.7954	1.517	1.41	1.237	0.8396
CI.mean.0.95	1.965	1.799	3.432	3.19	2.798	1.899
var	7.544	6.327	23.02	19.89	15.3	7.049
std.dev	2.747	2.515	4.798	4.46	3.911	2.655
coef.var	0.3442	0.15	0.1835	0.3371	0.1723	0.1019

From our basic data summary, we're further assured of the normal distribution of our data by the close approximation of medians and means. (See Appendix 2 for code to produce basic data summary tables.)

Comparison of Tooth Growth by Delivery Method and Dosage

The following tables show the results of two-sample t-tests performed on the data for each combination of delivery method and dosage. Column names indicate the data pairs. For example, vc05_1 indicates that delivery method was held constant to compare dosages of .05 and 1 (See Appendix 3 for a complete key to column names). In each case, the null hypothesis is that there is no difference in the means of tooth length while the alternative hypothesis is that the true difference in means is not zero. (See Appendix 4 for code to produce tables.)

% T-tests of All Dosage Pairs with Delivery Method Held Constant

	vc05_1	vc05_2	vc1_2	oj05_1	oj05_2	oj1_2
estimate	-8.79	-18.2	-9.37	-9.47	-12.8	-3.36
estimate1	7.98	7.98	16.8	13.2	13.2	22.7
estimate2	16.8	26.1	26.1	22.7	26.1	26.1
statistic	-7.46	-10.4	-5.47	-5.05	-7.82	-2.25
p.value	6.81e-07	4.68e-08	9.16e-05	8.78e-05	1.32e-06	0.0392
parameter	17.9	14.3	13.6	17.7	14.7	15.8
conf.low	-11.3	-21.9	-13.1	-13.4	-16.3	-6.53
conf.high	-6.31	-14.4	-5.69	-5.52	-9.32	-0.189

% T-tests of Delivery Method Pairs with Dosage Held Constant

	vc05_oj05	vc1_oj1	vc2_oj2
estimate	-5.25	-5.93	0.08
estimate1	7.98	16.77	26.14
estimate2	13.23	22.7	26.06
statistic	-3.17	-4.033	0.04614
p.value	0.006359	0.001038	0.9639
parameter	14.97	15.36	14.04
conf.low	-8.781	-9.058	-3.638
conf.high	-1.719	-2.802	3.798

Conclusions

For the t-tests of dosage pairs with delivery method held constant, in each case the 95% confidence intervals indicate a difference in the negative range when testing smaller dosage minus larger dosage, which tells us that there is a statistically significant increase in tooth growth with an increase in Vitamin C dosage. Also, the p-values are all below 0.05, well below that in almost all cases, which tells us that the probability of getting the results we did **and** the null hypothesis being true (no difference in tooth growth with increased dosage) is less than 5%.

The t-tests of delivery methods with dosages held constant provided mixed results. At dosages of 0.5 mg and 1.0 mg, the results were statistically significant with ascorbic acid (VC) resulting in greater tooth growth based, as with the above tests, on the confidence intervals and p-values. At dosages of 2.0 mg, the p-value was 0.96 and the confidence interval contained zero, which indicates a very high likelihood that the null hypothesis is correct – there is no real difference in tooth growth between delivery methods at the highest dosage level.

It is important to note that our analysis assumes that the 60 guinea pigs were chosen randomly in such a way that they represent the actual population of guinea pigs, and that the choice of dosage and delivery method for each guinea pig was done randomly so as not to introduce any unknown confounders. Also, our t-tests assume unequal variance between each pair of samples that we tested.

Appendices

Appendix 1: Code to produce exploratory data analysis plots:

```
g1 <- ggplot(data=dat, aes(x=supp, y=len)) + geom_boxplot(aes(fill=supp)) +  
  xlab("Delivery Method") + ylab("Tooth Length")  
g2 <- ggplot(data=dat, aes(x=dose, y=len)) + geom_boxplot(aes(fill=dose)) +  
  xlab("Dose") + ylab("Tooth Length")  
# Define grid layout to locate plots and print each graph  
pushViewport(viewport(layout = grid.layout(1, 2)))  
print(g1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))  
print(g2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))  
  
histogram(~len|interaction(supp,dose),data=dat, xlab="Tooth Length",  
  type="density", main="Tooth Growth by Delivery & Dose", breaks=seq(from=4,to=34,by=2) )
```

Appendix 2: Code to produce tables for Basic Data Summary

```
vcLength <- dat[dat$supp=="VC",]$len  
ojLength <- dat[dat$supp=="OJ",]$len  
df1 <- data.frame("VC"= vcLength, "OJ"= ojLength)  
pander(stat.desc(df1, basic=FALSE))  
  
vc05 <- dat[dat$supp=="VC" & dat$dose=="0.5", ]$len  
vc1 <- dat[dat$supp=="VC" & dat$dose=="1", ]$len  
vc2 <- dat[dat$supp=="VC" & dat$dose=="2", ]$len  
oj05 <- dat[dat$supp=="OJ" & dat$dose=="0.5", ]$len  
oj1 <- dat[dat$supp=="OJ" & dat$dose=="1", ]$len  
oj2 <- dat[dat$supp=="OJ" & dat$dose=="2", ]$len  
  
df2 <- data.frame(vc05, vc1, vc2, oj05, oj1, oj2)  
pander(stat.desc(df2, basic=FALSE))
```

Appendix 3: Key to column names in t-test tables:

vc05_1	Delivery method VC held constant, comparing dosages of 0.5 vs. 1 mg
vc05_2	Delivery method VC held constant, comparing dosages of 0.5 vs. 2 mg
vc1_2	Delivery method VC held constant, comparing dosages of 1 vs. 2 mg
oj05_1	Delivery method OJ held constant, comparing dosages of 0.5 vs. 1 mg
oj1_2	Delivery method OJ held constant, comparing dosages of 1 vs. 2 mg
oj05_2	Delivery method OJ held constant, comparing dosages of 0.5 vs. 2 mg
vc05_oj05	Dosage of 0.5 mg held constant, comparing delivery methods VC vs. OJ
vc1_oj1	Dosage of 1 mg held constant, comparing delivery methods VC vs. OJ
vc2_oj2	Dosage of 2 mg held constant, comparing delivery methods VC vs. OJ

Appendix 4: Code to produce t-test tables:

```
vc05 <- dat[dat$supp=="VC" & dat$dose=="0.5", ]$len  
vc1 <- dat[dat$supp=="VC" & dat$dose=="1", ]$len  
vc2 <- dat[dat$supp=="VC" & dat$dose=="2", ]$len  
oj05 <- dat[dat$supp=="OJ" & dat$dose=="0.5", ]$len  
oj1 <- dat[dat$supp=="OJ" & dat$dose=="1", ]$len
```

```
oj2 <- dat[dat$supp=="OJ" & dat$dose=="2", ]$len
```

```
combos <- data.frame("vc05_1" = t( tidy(t.test(vc05, vc1))))  
combos <- cbind(combos, "vc05_2" = t( tidy(t.test(vc05, vc2))))  
combos <- cbind(combos, "vc1_2" = t( tidy(t.test(vc1, vc2))))  
combos <- cbind(combos, "oj05_1" = t( tidy(t.test(oj05, oj1))))  
combos <- cbind(combos, "oj05_2" = t( tidy(t.test(oj05, oj2))))  
combos <- cbind(combos, "oj1_2" = t( tidy(t.test(oj1, oj2))))  
pander(combos)
```

```
combos2 <- data.frame("vc05_oj05" = t( tidy(t.test(vc05, oj05))))  
combos2 <- cbind(combos2, "vc1_oj1" = t( tidy(t.test(vc1, oj1))))  
combos2 <- cbind(combos2, "vc2_oj2" = t( tidy(t.test(vc2, oj2))))  
pander(combos2)
```

Appendix 4: Complete RMD file

The complete R Markdown file used to compile this pdf report can be found here...

<https://github.com/tchopoorian/Statistical-Inference>