

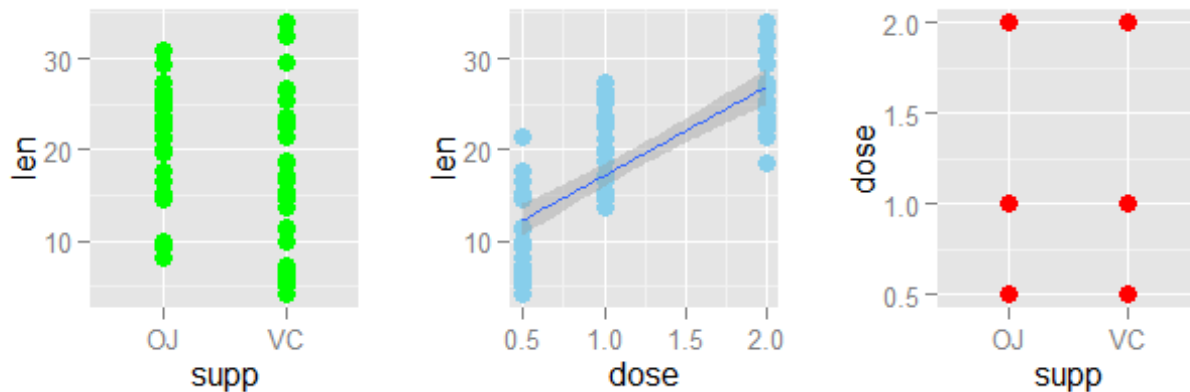
Analysis of ToothGrowth data

J. Thatcher

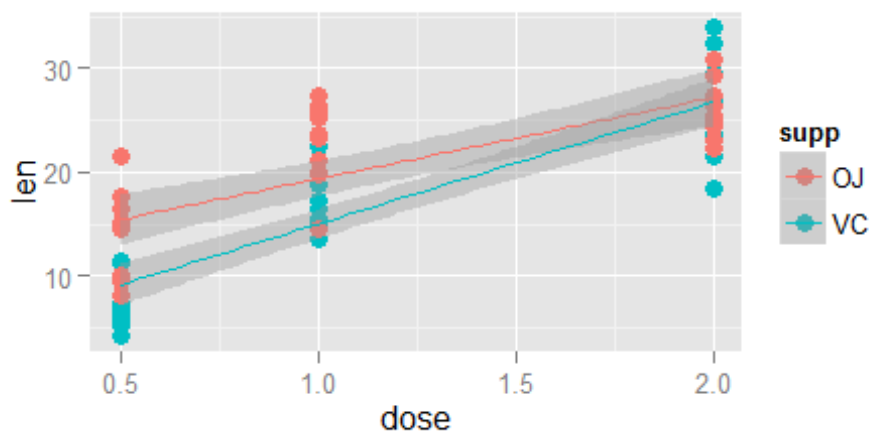
Load the ToothGrowth data

```
## Warning: package 'ggplot2' was built under R version 3.1.3
## Warning: package 'xtable' was built under R version 3.1.3
```

There are three variables in the toothgrowth dataset. They are length (len, levels = continuous), supplement (supp, levels = OJ and VC), and dose (dose, levels = 0.5, 1.0, and 2.0). We can quickly look to see any correlation between the variables by pairwise plots of each variable.



There appears to be positive correlation between the dose and length variables. Which means that length of the tooth can be predicted by the dose. Furthermore, the supp variable also seems to have a minor correlation with len. This becomes more obvious when we separate the dose ~ len plot by the factor supp in the following plot:



Hypotheses

The following hypotheses are relevant to this dataset:

For the supp variable

H0: the effect of supp OJ = the effect of supp VC on tooth length

H1: the effect of supp OJ \neq the effect of supp VC on tooth length

For all pairwise comparisons of the dose variable

H0: the effect of dose A = the effect of dose B on tooth length

H1: the effect of dose A \neq the effect of dose B on tooth length

For all pairwise comparisons of the supp variable within a single level of the dose variable

H0: the effect of supp OJ = the effect of supp VC on tooth length

H1: the effect of supp OJ \neq the effect of supp VC on tooth length

The effect of supp

For the first set of hypotheses:

H0: the effect of supp OJ = the effect of supp VC on tooth length

H1: the effect of supp OJ \neq the effect of supp VC on tooth length

First we will ask, does the mean of the supp OJ fall within the mean of the supp VC?

We can quickly calculate the 95% confidence interval with the `t.test` function, as well as the p-value for two-tailed comparison.

```
##  
## Two Sample t-test  
##  
## data: TGdata[1:30, 1] and TGdata[31:60, 1]  
## t = -1.9153, df = 58, p-value = 0.06039  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -7.5670064 0.1670064  
## sample estimates:  
## mean of x mean of y  
## 16.96333 20.66333
```

We can interpret this results table in such a way that the 95% confidence interval for the difference in means does include zero and the p-value is slightly larger than 0.05. This suggests that we cannot reject H0 and rule-out the possibility that the difference in means was a result of chance.

The effect of dose

For all pairwise comparisons of the dose variable

There are 3 total pairwise comparisons.

0.5 vs. 1.0

0.5 vs. 2.0

1.0 vs. 2.0

Results of these t-test can be quickly calculated, but we will consider that the possibility of getting a significant comparison increases with the number of tests performed on the data and therefore we need to apply a correction to alpha. We will use the Bonferroni correction, which is conservative and simple to implement.

##	Comparison of Dose	P-Value
## 1	0.5 vs. 1.0	1.266297e-07
## 2	0.5 vs. 2.0	2.837553e-14
## 3	1.0 vs. 2.0	1.810829e-05

In this case we find that the p values for all three comparisons are extremely small, and even with the conservative Bonferroni correction where $\alpha = 0.0166667$ we can still reject the null hypotheses and agree that the increase the variable len increases with increasing levels of the variable dose, regardless of the level of the variable supp.

supp vs len at each level of dose

finally, we will consider the effect of either supplement when we remove the effect of the dose. To do this, we will compare the two levels of supp, OJ and VC at each of the three levels of dose. Again, the Bonferroni correction will be used.

There are 3 total pairwise comparisons.

Dose 0.5: VC vs. OJ

Dose 1.0: VC vs. OJ

Dose 2.0: VC vs. OJ

possible comparisons and their results are as follows:

##	Comparison of Supp by Dose	P-Value
## 1	Dose 0.5: VC vs. OJ	0.00530
## 2	Dose 1.0: VC vs. OJ	0.00078
## 3	Dose 2.0: VC vs. OJ	0.96371

Here we find that the p-values for the first two levels of dose are less than the alpha value we selected ($0.05/3$ or 0.0166667). The third comparison, however, is not less than $\alpha = 0.0166667$. Therefore, we conclude that the response of len to supp is dependent on the level of dose and that the supp OJ is associated with higher values of len at the doses 0.5 and 1.0.

Assumptions

Our overall assumptions for these tests include:

1. groups of data were normally distributed

2. equal variances within each of the t-tests

Appendix: R-code

```
library(ggplot2)
library(xtable)
options(xtable.floating = FALSE)
source("multiplot.R")

TGdata <- ToothGrowth

p1 <- ggplot(TGdata, aes(supp,len)) + geom_point(size = 3, color="green")
p2 <- ggplot(TGdata, aes(dose,len)) + geom_point(size = 3, color="skyblue") +
  geom_smooth(method = "lm")
p3 <- ggplot(TGdata, aes(supp,dose)) + geom_point(size = 3,color="red")

multiplot(p1,p2,p3,cols=3)

p4 <-ggplot(TGdata, aes(x= dose,y= len, colour=supp)) + geom_point(size = 3)
+
  geom_smooth(method = "lm")
p4

t.test(TGdata[1:30,1], TGdata[31:60,1], var.equal = TRUE)

d1 <- c(TGdata[1:10,1], TGdata[31:40,1])
d2 <- c(TGdata[11:20,1], TGdata[41:50,1])
d3 <- c(TGdata[21:30,1], TGdata[51:60,1])
AvB <- t.test(d1, d2, var.equal = TRUE)
AvC <- t.test(d1, d3, var.equal = TRUE)
BvC <- t.test(d2, d3, var.equal = TRUE)

results <- as.data.frame(rbind(AvB$p.value, AvC$p.value, BvC$p.value))
results <- cbind(c("0.5 vs. 1.0 ", "0.5 vs. 2.0 ", "1.0 vs. 2.0 "),results)
colnames(results) <- c("| Comparison of Dose |", "| P-Value |")
results

VCd1 <- c(TGdata[1:10,1])
VCd2 <- c(TGdata[11:20,1])
VCd3 <- c(TGdata[21:30,1])

OJd1 <- c(TGdata[31:40,1])
OJd2 <- c(TGdata[41:50,1])
OJd3 <- c(TGdata[51:60,1])

A1vB1 <- t.test(VCd1, OJd1, var.equal = TRUE)
A2vB2 <- t.test(VCd2, OJd2, var.equal = TRUE)
A3vB3 <- t.test(VCd3, OJd3, var.equal = TRUE)
```

```
results2 <- as.data.frame(rbind(A1vB1$p.value, A2vB2$p.value, A3vB3$p.value))
results2 <- cbind(c("Dose 0.5: VC vs. OJ", "Dose 1.0: VC vs. OJ",
                    "Dose 2.0: VC vs. OJ"), round(results2, 5))
colnames(results2) <- c("| Comparison of Supp by Dose |", "| P-Value |")
results2
```