Analyze the ToothGrowth data

Lenny Fenster

Friday, September 19, 2014

Instructions

Analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

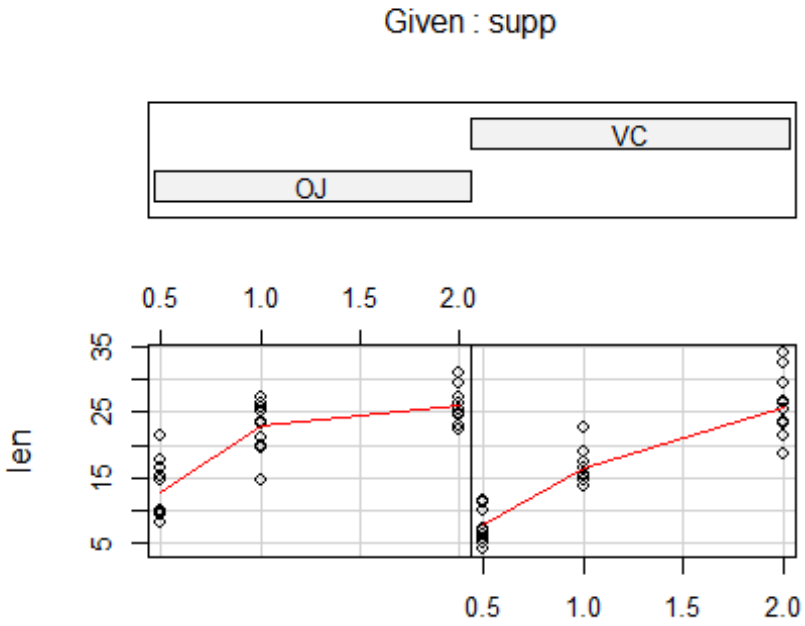
Exploratory Data Analyses

The ToothGrowth dataset is included in R and represents 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). It is a data frame with 60 observations on the following 3 variables:

- [,1] len numeric Tooth length
- [,2] supp factor Supplement type (VC or OJ).
- [,3] dose numeric Dose in milligrams.

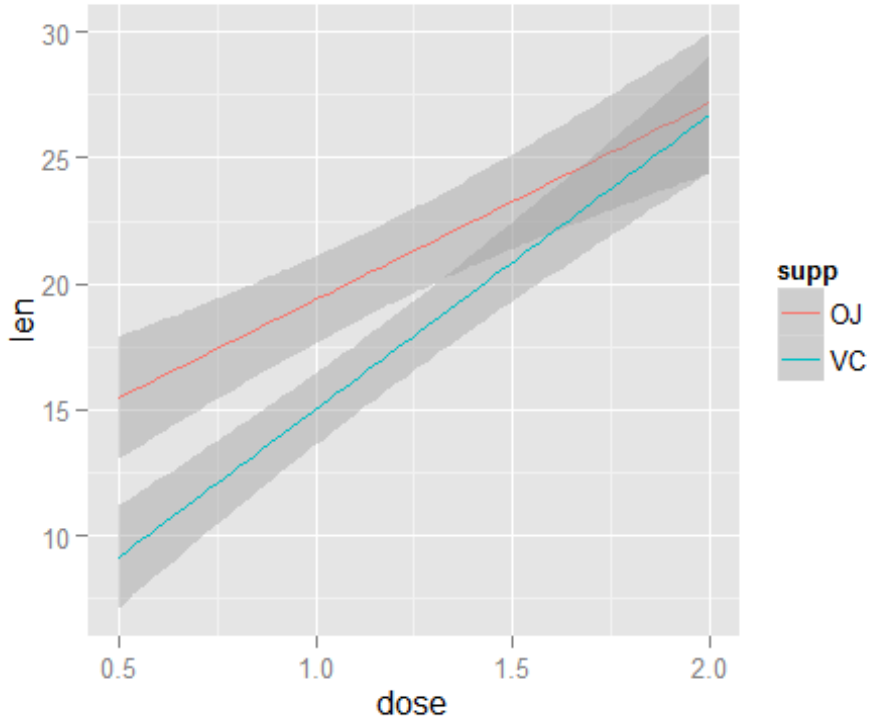
```
library(datasets)
library(ggplot2)

require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



ToothGrowth data: length vs dose, given type of supplement

```
g<-ggplot(ToothGrowth, aes(x=dose, y=len, color=supp))
g<-g+geom_smooth(method="lm")
print(g)
```



Initial exploratory analysis indicates shows the effects of Orange Juice and Vitamin C on 10 guinea pigs with three different doses of each. The initial analysis seems to indicate that a low dose of orange juice results in an increase in odontoblast; however, as the dosages increase, the vitamin C becomes more effective.

Basic Summary of the Data

The dataset includes three variables: the tooth length (numeric), the supplement taken (factor), and the dosage for the supplement (numeric) for the ten guinea pig subjects. The tooth lengths vary from 4.2 to 33.9 with a mean of 18.8 and a median of 19.2. The dosages are fairly consistent from 0.5 to 2.0 by 0.5 increases.

```
summary(ToothGrowth)
```

```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

```
var(ToothGrowth$len)
```

```
## [1] 58.51
```

The *length* is the mesurement we are most interested in as it relates to dosages of supplements. There seems to be a fair amount of variability in the *length* (58.51).

```
summary(subset(ToothGrowth, supp=="OJ")$len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     8.2    15.5    22.7    20.7    25.7    30.9
```

```
var(subset(ToothGrowth, supp=="OJ")$len)
```

```
## [1] 43.63
```

```
summary(subset(ToothGrowth, supp=="VC")$len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     4.2    11.2    16.5    17.0    23.1    33.9
```

```
var(subset(ToothGrowth, supp=="VC")$len)
```

```
## [1] 68.33
```

Looking at the summaries for the orange juice and vitamin C observations independently, it appears that the bulk of the variability exists in the Vitamin C observations (var = 68.33) over the orange juice observations (var = 43.63)

Confidence Intervals and Hypothesis Tests

While the exploratory data analyis and basic summary of the data point to a difference in the variance between the the length of odontoblasts when guinea pigs are dosed with orange juice compared with doses of vitamin C, the means, medians, and maximum values for both groups are similar. Therefore, we will test the null hypothesis that the length of the odontoblasts when guinea pigs are dosed with orange juice is statistically equivalent to the odontoblasts for when they are dosed of vitamin C. In other words:

$H_0 : \mu_0 = \mu_\alpha$

where

μ_0 : mean for length with orange juice doses and

μ_α : mean for length with vitamin C doses

```
t.test(len ~ supp,paired=TRUE, data=ToothGrowth)
```

```
##
##  Paired t-test
##
## data:  len by supp
## t = 3.303, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.409 5.991
## sample estimates:
## mean of the differences
##                     3.7
```

```
t.test(len ~ supp,paired=TRUE, data=ToothGrowth, equal=FALSE)
```

```
##
##  Paired t-test
##
## data:  len by supp
## t = 3.303, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.409 5.991
## sample estimates:
## mean of the differences
##                     3.7
```

Additional comparison #1

While the variances for the total population appear to differ, the initial exploratory analysis does seem to indicate that the variance between the two population decreases as the dosage increases. In other words, it appears that as the dosage of orange juice and vitamin C increase, it appears that there may be no statistical difference in their effect on tooth

growth. Therefore, we will test the hypothesis that the length of the odontoblasts when guinea pigs are dosed with orange juice is statistically equivalent to the odontoblasts for when they are dosed of vitamin C when for doses of 2.0

```
t.test(subset(ToothGrowth, supp=="OJ" & dose==2.0)$len - subset(ToothGrowth,
supp=="VC" & dose==2.0)$len)

##
##  One Sample t-test
##
## data:  subset(ToothGrowth, supp == "OJ" & dose == 2)$len -
subset(ToothGrowth,     supp == "VC" & dose == 2)$len
## t = -0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.329  4.169
## sample estimates:
## mean of x
##     -0.08
```

Conclusion

We reject the initial null hypotheses that the length of the odontoblasts when guinea pigs are dosed with orange juice is statistically equivalent to the odontoblasts for when they are dosed with vitamin C. We reject it because the t-statistic, the difference in the average length between the two doses, is 3.3 standard deviations from the mean with 29 degrees of freedom (there were 30 pairs measured). Furthermore, the 95% confidence interval which ranges from 1.41 to 5.99, does not contain the value zero (the hypotheses for which we were testing). Because we noted when analyzing the basic summary data that the variances between the two groups differ, we also tested with unequal variances. The results were the same and our conclusion remains to reject the null hypotheses.

However, we will not reject the null hypothesis for the additional comparison where we examined the populations specifically for dosgaes of 2.0. The difference in the average length between the two samples is only -0.04 standard deviations from the mean with 9 degrees of freedom (there were 10 pairs measured in the subset for dosages of 2.0). Furthermore, the 95% confidence interval which ranges from -4.33 to 4.17 contains the value zero (the hypotheses for which we were testing). Thus, there is not enough evidence to reject this second null hypothese that the means of tooth growth are equal for dosages of 2.0 for orange juice and vitamin C.