

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

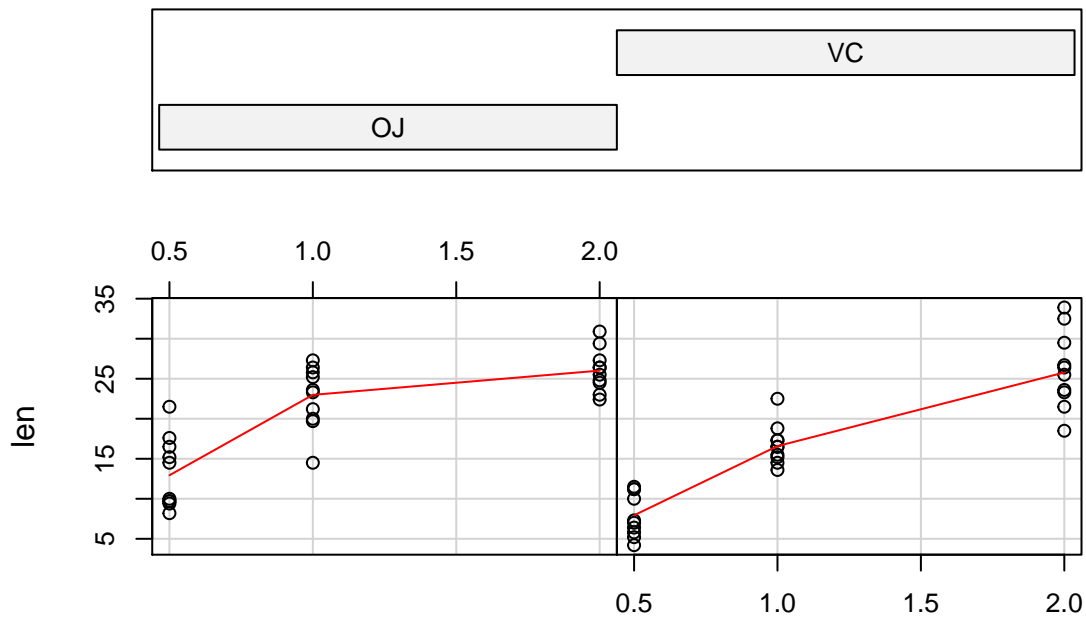
The ToothGrowth dataset is included in R and represents the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid). It is a data frame with 60 observations on the following 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)

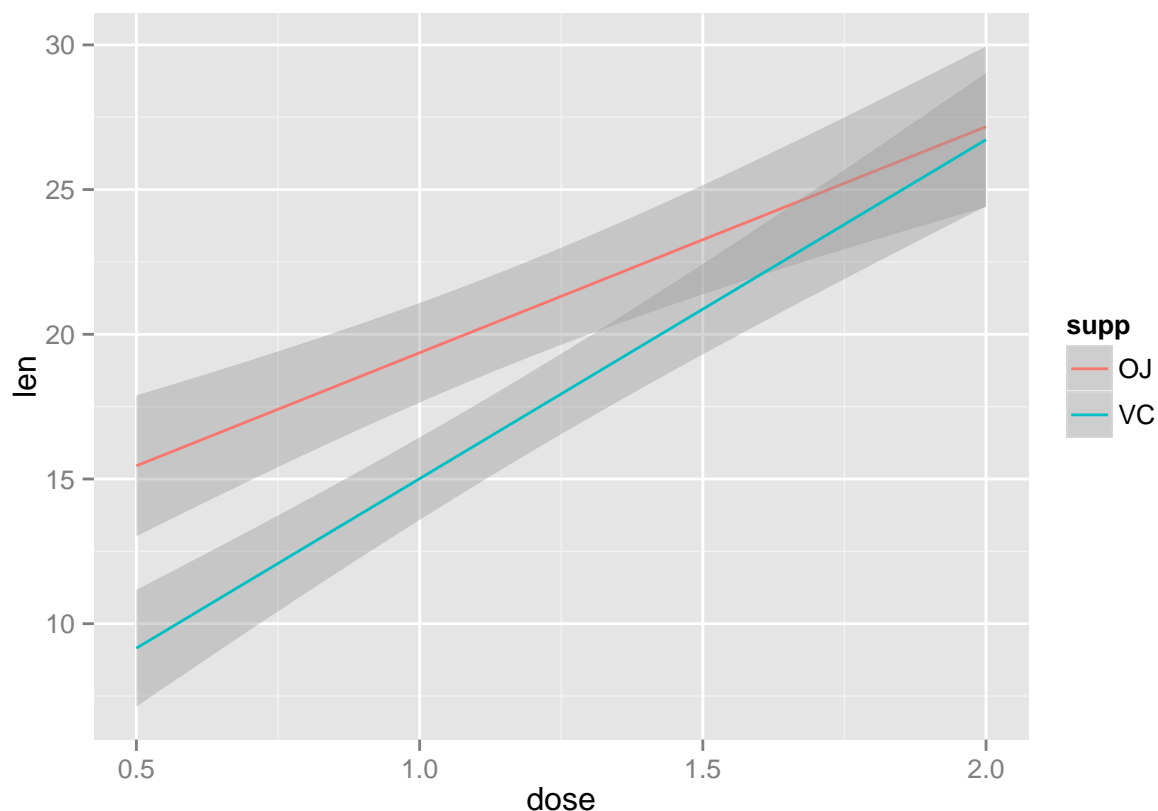
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```

Given : supp



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 measurement 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).

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

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

```
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 ($\text{var} = 68.33$) over the orange juice observations ($\text{var} = 43.63$)

Confidence Intervals and Hypothesis Tests

The

Conclusion

The