Tooth Gowth exploratory analysis

Matteo

Sunday, February 22, 2015

Exporatory Analysis of Guinea Pig Teeth Growth with Vitamin C

The Tooth Growth dataset reports the results of an experiment correlating Vitamin C to teeth growth in 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).

1. Load the ToothGrowth data and perform some basic exploratory data analyses

We calculated the average size of the teeth with respecto the dose and the delivery method:

```
library(datasets)
data(ToothGrowth)
tg <- ToothGrowth
```

2. Provide a basic summary of the data.

The data has a quantitative variable (the length of the tooth) and two categorical categories (dosage and delivery). An interesting way to look at this data is the mean of the combinations of the variables and its standard deviation:

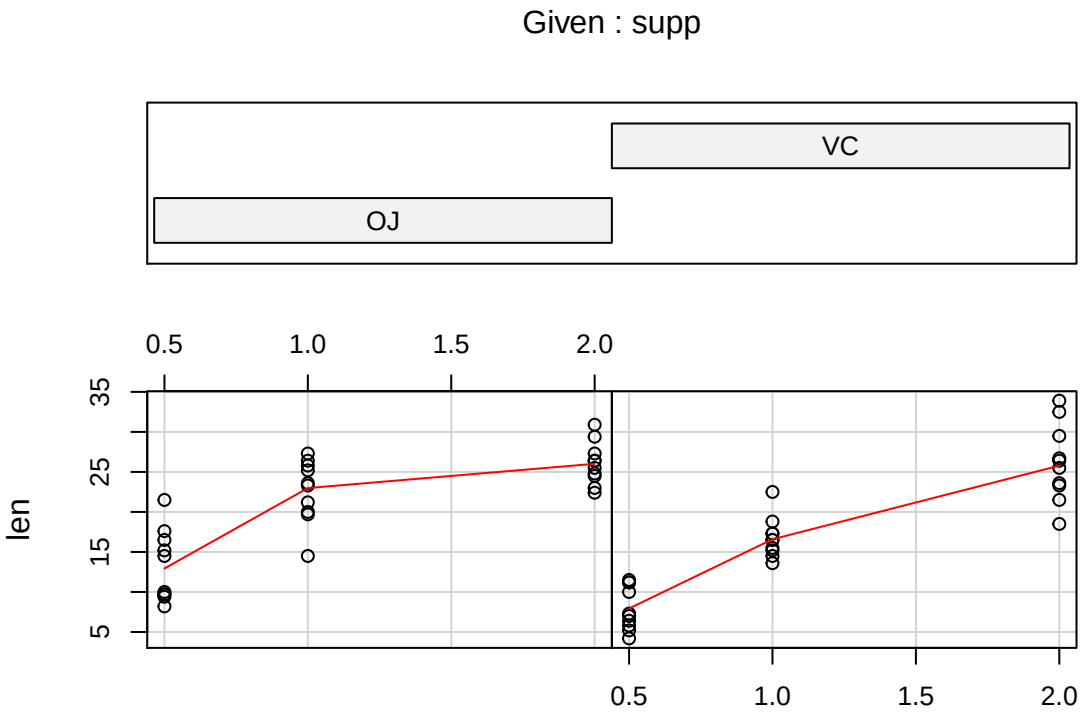
```
aggtg <- aggregate(len ~ supp + dose, data = tg, mean)
aggtgSd <- aggregate(len ~ supp + dose, data = tg, sd)
aggtg ## mean of the tooth length by variable
```

```
##   supp dose   len
## 1   OJ  0.5 13.23
## 2   VC  0.5  7.98
## 3   OJ  1.0 22.70
## 4   VC  1.0 16.77
## 5   OJ  2.0 26.06
## 6   VC  2.0 26.14
```

```
aggtgSd ## standard deviation of tooth length by variable
```

```
##   supp dose      len
## 1   OJ  0.5 4.459709
## 2   VC  0.5 2.746634
## 3   OJ  1.0 3.910953
## 4   VC  1.0 2.515309
## 5   OJ  2.0 2.655058
## 6   VC  2.0 4.797731
```

We can take a quick look at the data with a graph:



ToothGrowth data: length vs dose, given type of supplement

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

There is some difference in the means of the ltooth length with the way the vitamin C is delivered, with the orange juice having an higher mean, and lower standard deviation:

```
aggDel<-aggregate(len ~ supp, data = tg, mean)
aggDelSd<-aggregate(len ~ supp, data = tg, sd)
aggDel ## mean of the tooth length by method of delivery
```

```
##   supp      len
## 1   OJ 20.66333
## 2   VC 16.96333
```

```
aggDelSd ## standard deviation of the tooth length by method of delivery
```

```
##   supp      len
## 1   OJ 6.605561
## 2   VC 8.266029
```

However looking at the 95% confidence interval, the superimposition suggests that it is not a statistically significant effect:

```
conDel1 <- aggDel[1, 2] + c(-1,1) * qt(.975, 9) * aggDelSd[1,2]/sqrt(10)
conDel2 <- aggDel[2, 2] + c(-1,1) * qt(.975, 9) * aggDelSd[2,2]/sqrt(10)
conDel1  ## confidence interval of Vitamin C delivered as orange juice
```

```
## [1] 15.93800 25.38867
```

```
conDel2  ## confidence interval of vitamin C delivered as ascorbic acid
```

```
## [1] 11.05017 22.87649
```

The mean of tooth size follows a clear direct trend:

```
aggDose<-aggregate(len ~ dose, data = tg, mean)
aggDoseSd <-aggregate(len ~ dose, data = tg, sd)
aggDose  ## mean of the tooth length by vitamin C dosage
```

```
##   dose    len
## 1  0.5 10.605
## 2  1.0 19.735
## 3  2.0 26.100
```

```
aggDoseSd ## standard deviation of the tooth by vitamin C dosage
```

```
##   dose      len
## 1  0.5 4.499763
## 2  1.0 4.415436
## 3  2.0 3.774150
```

And we can see that the 95% confidence intervals have superimposed areas, suggesting a significant effect.

```
conDose1 <- aggDose[1,2] + c(-1,1) * qt(.975, 9) * aggDoseSd[1,2]/sqrt(10)
conDose2 <- aggDose[2,2] + c(-1,1) * qt(.975, 9) * aggDoseSd[2,2]/sqrt(10)
conDose3 <- aggDose[3,2] + c(-1,1) * qt(.975, 9) * aggDoseSd[3,2]/sqrt(10)
conDose1  ## confidence interval of 0.5 mg of Vitamin C
```

```
## [1]  7.386063 13.823937
```

```
conDose2  ## confidence interval of 1.0 mg of Vitamin C
```

```
## [1] 16.57639 22.89361
```

```
conDose3  ## confidence interval of 2.0 mg of Vitamin C
```

```
## [1] 23.40014 28.79986
```

4. State your conclusions and the assumptions needed for your assumptions.

Administration of vitamin C has an effect on teeth growth on guinea pigs. The delivery of the substance seems to have a positive effect, with orange juice being the most effective way, with respect to the use of ascorbic acid. The delivery effect doesn't seem to be statistically significant. The dosage seems to have a proportinal effect, espcially when comparing 0.5 mg and 2.0 mg, which is possibly statistically significant, since the two intervals do not superimpose.