Statistical inference - basic inference

Harsh Reddy Gandavarapu

07/13/2017

We're going to analyze the ToothGrowth data in the R datasets package.

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

We plot the lengt vs the dose for each of the supplements. To gain a better view of groth rates, we also add a loess curve. We see that the growth rates seem to behave differently for both supplements.

```
library(ggplot2)
data(ToothGrowth)
qplot(dose, len, data=ToothGrowth, facets=.~supp, geom=c("point", "smooth"), method="loess")
```
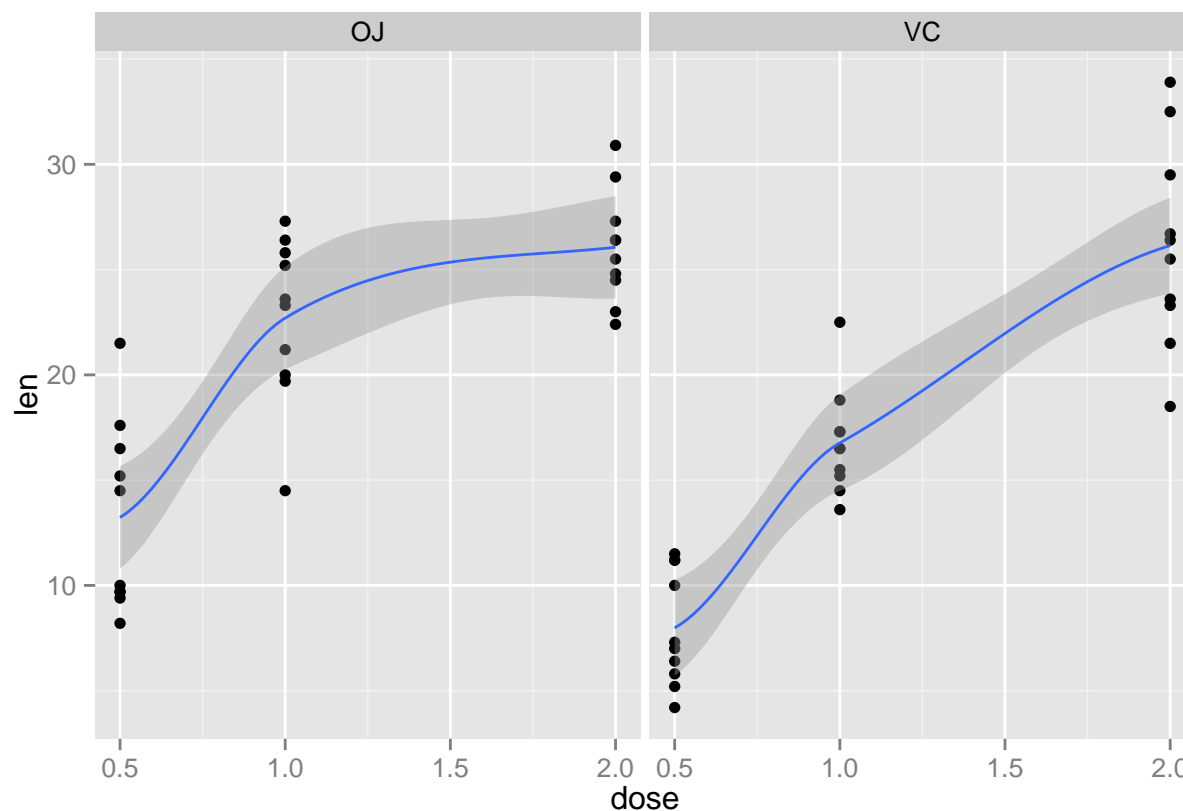

2. Provide a basic summary of the data.

This dataset contains three variables: supplement, dose and len. For each supplement, and each dose we calculate basic descriptive statistics: standard deviation, variance, and mean.

```
dose <- as.numeric(levels(as.factor(ToothGrowth$dose)))
supp <- levels(ToothGrowth$supp)

# Structured for further processing
data <- list()
```

```
x <- Map(function(s) {
  Map(function(d) {
    l <- ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == s]
    data <<- rbind(data, list(supp = s, dose = d, sd=sd(l), var=var(l), mu=mean(l)))
  }, dose)
}, supp)

data
```

```
##       supp dose sd     var    mu
## [1,] "OJ" 0.5  4.46   19.89 13.23
## [2,] "OJ" 1    3.911  15.3  22.7
## [3,] "OJ" 2    2.655  7.049 26.06
## [4,] "VC" 0.5  2.747  7.544 7.98
## [5,] "VC" 1    2.515  6.327 16.77
## [6,] "VC" 2    4.798  23.02 26.14
```

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)

We perform the student-t test for each dose level between the two supplements:

```
tests = list()
for (d in dose) {
  ojd <- ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "OJ"]
  vcd <- ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "VC"]
  t <- t.test(ojd, vcd, var.equal=T)
  id <- paste("OJ", d, "-", "VC", d)
  tests <- rbind(tests, list(id=id, p.value=t$p.value, ci.lo=t$conf.int[1], ci.hi=t$conf.int[2]))
}
tests
```

```
##       id                p.value    ci.lo  ci.hi
## [1,] "OJ 0.5 - VC 0.5" 0.005304   1.77   8.73
## [2,] "OJ 1 - VC 1"     0.0007807  2.841  9.019
## [3,] "OJ 2 - VC 2"     0.9637     -3.723 3.563
```

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

First, we assume that variance in all groups should be expected to be equal. The underlying assumption is that sampling of Guinea Pigs to assign them to a supplement and a dose was done properly.

Based on the test results from the previous question we need to **reject** the following hypotheses:

- True difference in means between OJ 0.5 and VC 0.5 is equal to 0
- True difference in means between OJ 1 and VC 1 is equal to 0
- True difference in means between OJ 2 and VC 2 is equal to 0