SWI4 Part 2:

Pratik Patil

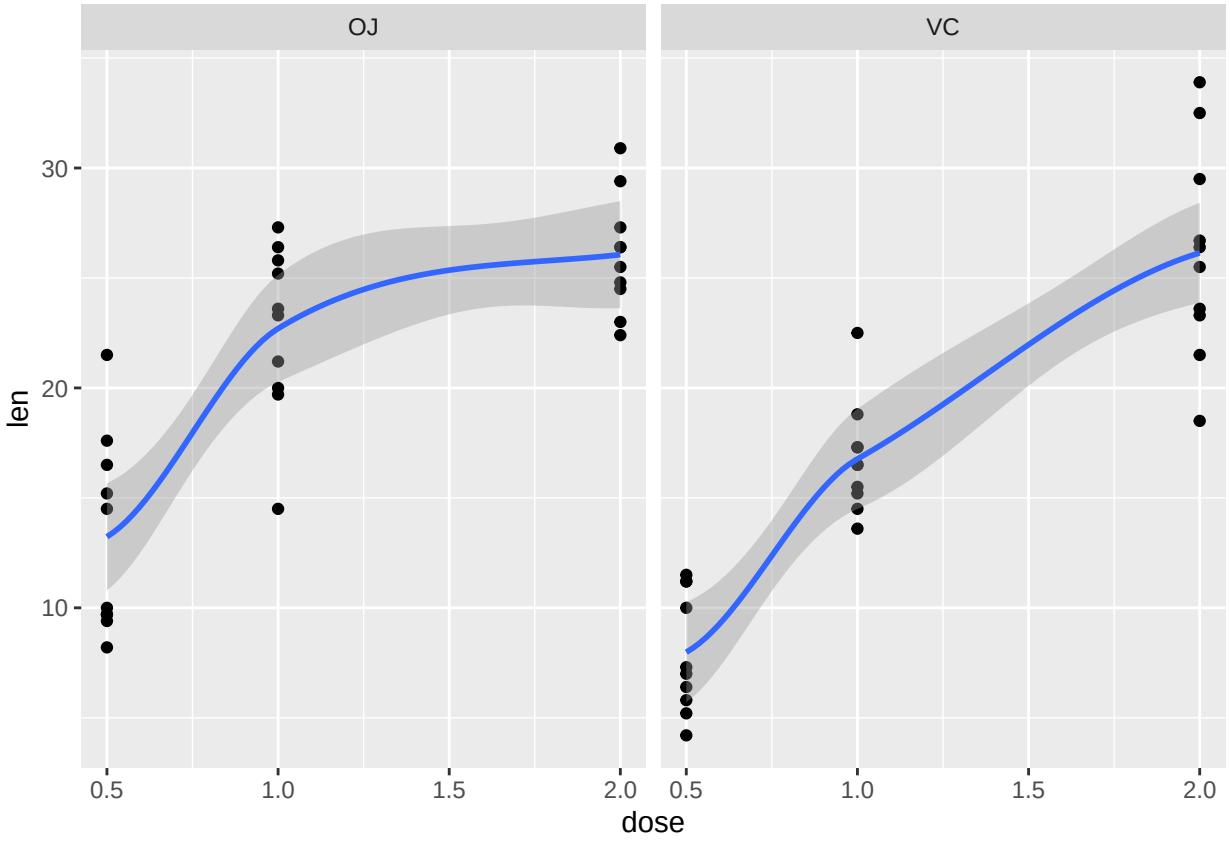
17/10/2020

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

Question 1:
Load the ToothGrowth data and perform some basic exploratory data analyses

Plot the length against the dose for each of the supplements. In order to gain a better view of growth rates, we also add a curve. It is seen 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")
```



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

```
##      supp dose sd       var      mu
## [1,] "OJ" 0.5  4.459709 19.889   13.23
## [2,] "OJ" 1    3.910953 15.29556 22.7
## [3,] "OJ" 2    2.655058 7.049333 26.06
## [4,] "VC" 0.5  2.746634 7.544    7.98
## [5,] "VC" 1    2.515309 6.326778 16.77
## [6,] "VC" 2    4.797731 23.01822 26.14
```

Question 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:

```
tt = 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)
  tt <- rbind(tt, list(id=id, p.value=t$p.value, ci.lo=t$conf.int[1], ci.hi=t$conf.int[2]))
}
tt
```

```
##      id                p.value      ci.lo     ci.hi
## [1,] "OJ 0.5 - VC 0.5" 0.005303661  1.770262  8.729738
## [2,] "OJ 1 - VC 1"     0.0007807262 2.840692  9.019308
## [3,] "OJ 2 - VC 2"     0.9637098    -3.722999 3.562999
```

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

Assumption:
1.We assume that variance in all groups should be expected to be equal. 2.The underlying assumption is that sampling of Guinea Pigs to assign them to a supplement and a dose was done properly.

Conclusions:
Observing the test results from the other questions we can **reject** the following hypotheses:

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

2. True difference in means between OJ 1 and VC 1 is equal to 0
3. True difference in means between OJ 2 and VC 2 is equal to 0