The Effect of Vitamin C on Tooth Growth in Guinea Pigs.

Basic Inferential Data Analysis. Statistical Inference, Coursera Project.

Malgorzata MS

January 31, 2017

Basic Inferential Data Analysis Instructions

In this project, 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
2. Provide a basic summary of the data.
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)
4. State your conclusions and the assumptions needed for your conclusions.

About `ToothGrowth` dataset

The results of the study of the effect of vitamin C on tooth growth in guinea pigs. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Load and explore data

```
data("ToothGrowth")
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Summary by supply method and dose. Data frames "sub_OJ" and "sub_VC" subset the observations for orange juice and ascorbic acid supply method, respectively.

```r
# Subset the data by supply method
sub_OJ <- ToothGrowth[which(ToothGrowth$supp == as.character("OJ")),]
sub_VC <- ToothGrowth[which(ToothGrowth$supp == as.character("VC")),]

# Summary by dose for orange juice supply method
tapply(sub_OJ$len,factor(sub_OJ$dose), summary)
```

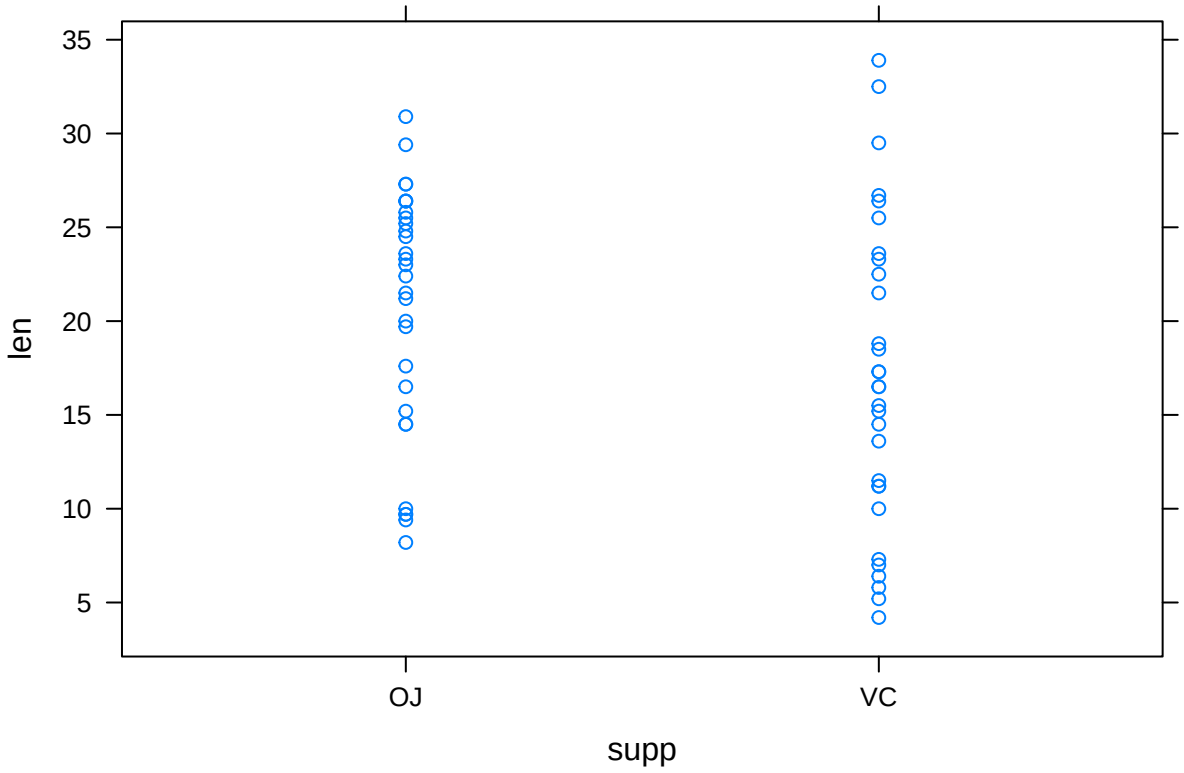
```
## $`0.5`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
##
## $`2`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.58   25.95   26.06   27.08   30.90
```

```r
# Summary by dose for ascorbic acid supply method
tapply(sub_VC$len,factor(sub_OJ$dose), summary)
```

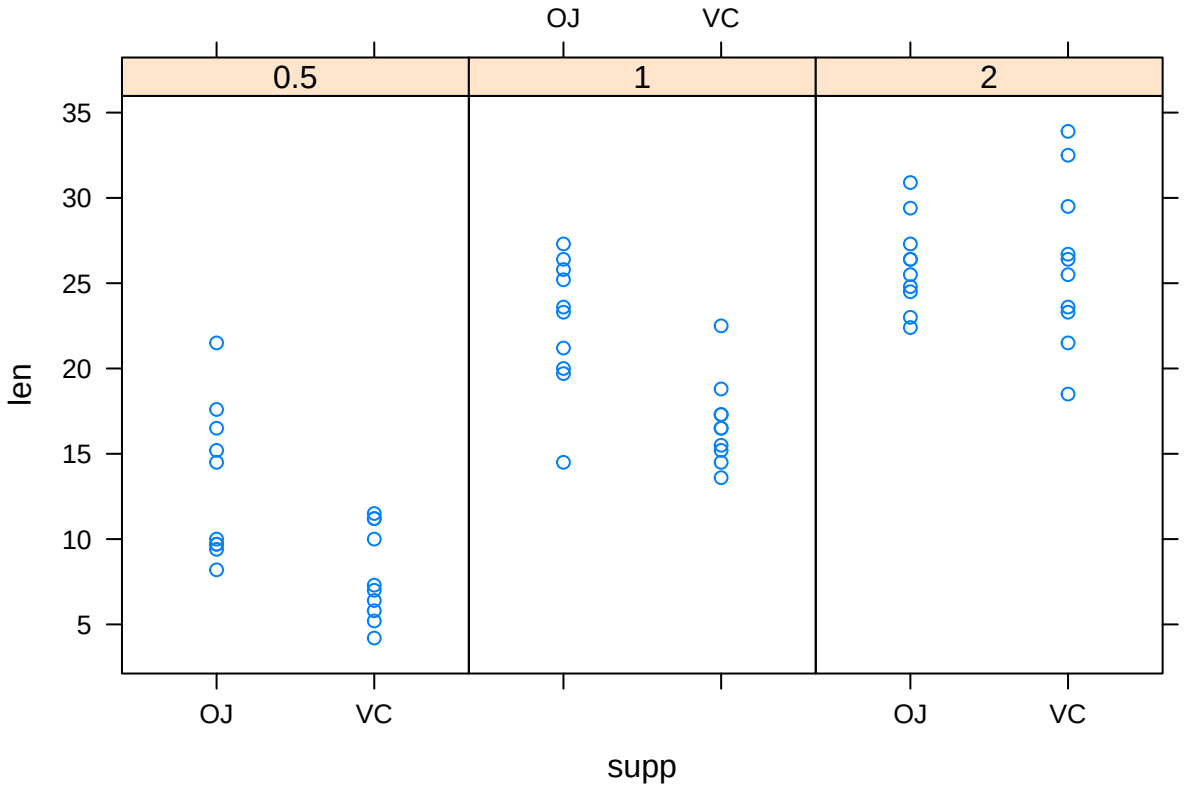
```
## $`0.5`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
##
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.27   16.50   16.77   17.30   22.50
##
## $`2`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

Plot data, (1) by supply method and (2) by supply method + dose.

```r
library(lattice)
with(ToothGrowth, xyplot(len~supp))
```




```
with(ToothGrowth, xyplot(len~supp | factor(dose)))
```



Compare tooth growth by supp and dose

- H0 - no difference (the two compared data subsets are equal)
- Ha - the two compared data subsets are different

1. Comparison between dose 0.5

```
t.test(sub_OJ[which(sub_OJ$dose == 0.5),1], sub_VC[which(sub_VC$dose == 0.5),1])
```

```
##
##  Welch Two Sample t-test
##
## data:  sub_OJ[which(sub_OJ$dose == 0.5), 1] and sub_VC[which(sub_VC$dose == 0.5), 1]
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The p-value is equal to 0.006359, therefore with significance level alfa=0.05 we reject the H0. Result: Higher response in the length of tooth cells for orange juice supply method.

2. Comparison between dose 1.0

```
t.test(sub_OJ[which(sub_OJ$dose == 1.0),1], sub_VC[which(sub_VC$dose == 1.0),1])
```

```
##
##  Welch Two Sample t-test
##
## data:  sub_OJ[which(sub_OJ$dose == 1), 1] and sub_VC[which(sub_VC$dose == 1), 1]
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The p-value is equal to 0.001038, therefore with significance level alfa=0.05 we reject the H0. Result: Higher response in the length of tooth cells for orange juice supply method.

3. Comparison between dose 2.0

```
t.test(sub_OJ[which(sub_OJ$dose == 2.0),1], sub_VC[which(sub_VC$dose == 2.0),1])
```

```
##
##  Welch Two Sample t-test
##
## data:  sub_OJ[which(sub_OJ$dose == 2), 1] and sub_VC[which(sub_VC$dose == 2), 1]
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

The p-value is equal to 0.9639, therefore with significance level alfa=0.05 we fail to reject the H0. Result: No significant difference in response in the length of tooth cells by supply method.

Assumptions:

The variables in analyzed data are iid normal distributed. The variance is assumed unequal.

Conclusions

The difference in tooth growth is observed only for lower doses of vit C: 0.5 and 1.0 mg/day. for lower doses supplying orange juice cause better response in compare to supplying ascorbic acid. For higher dose, equal to 2.0 mg/day there is no difference between supply methods.

Plot of averages vs dose for different supply methods.

```
plot(c(0.5,1,2), tapply(sub_OJ$len, sub_OJ$dose, mean), pch=19, cex=1.5, col="blue",
     ylim=c(min(ToothGrowth$len), max(ToothGrowth$len)),
     main = "Averages of Tooth Growth length", xlab="Dose", ylab="Average length")
legend(0.5, 35,c("Orange juice", "Ascorbic acid"), title="Supply method",
       lwd=c(2,2), col=c("blue", "red"))
lines(c(0.5,1,2), tapply(sub_OJ$len, sub_OJ$dose, mean), lty=1, lwd=2, col="blue")
points(c(0.5,1,2), tapply(sub_VC$len, sub_VC$dose, mean), pch=19, cex=1.5, col="red")
lines(c(0.5,1,2), tapply(sub_VC$len, sub_VC$dose, mean), lty=1, lwd=2, col="red")
```