Statical Inference Project Work - Question 2

Hardeep

25 April 2015

Overview

Here we're going to analyze the ToothGrowth data in the R datasets package and report on this analysis.

The ToothGrowth data is 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).

A data frame with 60 observations on 3 variables.

Analysis

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

Let us look load and look at the dataset first

```
library(datasets)
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 ...
```

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Data has 60 observations, 30 for each supplement (Orange Juice & Vitamin C). Further it has 10 observations for each of 6 dose/supplement combinations.

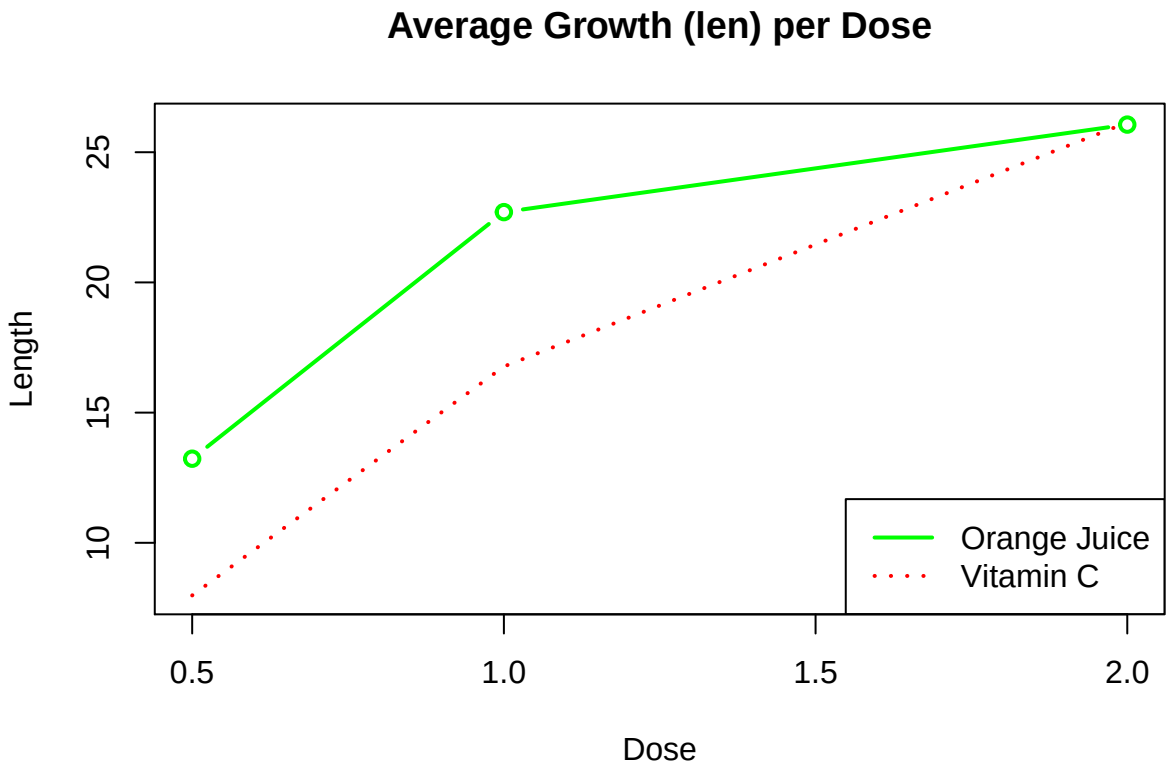
Let us see this visually

```
TGOJ <- ToothGrowth[ToothGrowth$supp == "OJ", ]
TGVC <- ToothGrowth[ToothGrowth$supp == "VC", ]

TGOJAgg <- aggregate(TGOJ$len ~ TGOJ$dose, , mean)
names(TGOJAgg) <- c("len","dose")
TGVCAgg <- aggregate(TGVC$len ~ TGVC$dose, , mean)
names(TGVCAgg) <- c("len","dose")

ylm <- c(min(min(TGOJAgg$dose),min(TGVCAgg$dose)),max(max(TGOJAgg$dose),max(TGVCAgg$dose)))
xlm <- c(min(min(TGOJAgg$len),min(TGVCAgg$len)),max(max(TGOJAgg$len),max(TGVCAgg$len)))
```

```r
plot(TGOJAgg,  xlim = xlm, ylim = ylm,
    type = "b", col = "green", lwd = 2, lty = 1,
    xlab = "Dose", ylab = "Length", main = "Average Growth (len) per Dose")
lines(TGVCAgg, col = "red", lwd = 2, lty = 3)
legend("bottomright", c("Orange Juice", "Vitamin C"),
  lwd = 2, lty = c(1, 3), col = c("green", "red"))
```



It shows us that tooth lengths are higher for OJ in most cases and also it increases with the dosage of supplements.

2. Provide a basic summary of the data.

Let us look at the data summary

```r
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
```

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)

In order to test our initial judgement from exploratory analysis we will use the `t.test` function here to perform a Student's t-Test on the data with the defaults of 95% confidence level and equal variances.

NULL Hypothesis

- Mean growth rate in both supplements is equal
- Mean growth rate stays same with increase in dosage

We will reject the null hypothesis in favor of the alternative hypothesis when we find a p-value of less than 0.05.

Let's partition the data for conducting the testing

```
TGVC05 <- TGVC[TGVC$dose==0.5,]
TGVC10 <- TGVC[TGVC$dose==1.0,]
TGVC20 <- TGVC[TGVC$dose==2.0,]
TGOJ05 <- TGOJ[TGOJ$dose==0.5,]
TGOJ10 <- TGOJ[TGOJ$dose==1.0,]
TGOJ20 <- TGOJ[TGOJ$dose==2.0,]
```

CASE I : Effects of Supplement Type

- Mean growth rate in both supplements is equal

```
# Perform t test on vc vs oj at 0.5mg dose
T05Supps <- t.test(len ~ supp, data=rbind(TGVC05,TGOJ05), var.equal=FALSE)

# Perform t test on vc vs oj at 1.0mg dose
T10Supps <- t.test(len ~ supp, data=rbind(TGVC10,TGOJ10), var.equal=FALSE)

# Perform t test on vc vs oj at 2.0mg dose
T20Supps <- t.test(len ~ supp, data=rbind(TGVC20,TGOJ20), var.equal=FALSE)
```

Summary of the results:

```
##   supplements dosage     p-value -conf.int +conf.int reject H0
## 1   VC and OJ  0.5mg 0.006358607  1.719057  8.780943      TRUE
## 2   VC and OJ  1.0mg 0.001038376  2.802148  9.057852      TRUE
## 3   VC and OJ  2.0mg 0.963851589 -3.798070  3.638070     FALSE
```

CASE II : Effects of Dosage

- Mean growth rate stays same with increase in dosage

```
# Perform t test on 0.5mg vs 1.0mg, within each supplement
TVCDose05and10 <- t.test(len ~ dose, data=rbind(TGVC05,TGVC10), var.equal=TRUE)
TOJDose05and10 <- t.test(len ~ dose, data=rbind(TGOJ05,TGOJ10), var.equal=TRUE)
```

```r
# Perform t test on 1.0mg vs 2.0mg, within each supplement
TVCDose10and20 <- t.test(len ~ dose, data=rbind(TGVC10,TGVC20), var.equal=TRUE)
TOJDose10and20 <- t.test(len ~ dose, data=rbind(TGOJ10,TGOJ20), var.equal=TRUE)

# Perform t test on 0.5mg vs 2.0mg, within each supplement
TVCDose05and20 <- t.test(len ~ dose, data=rbind(TGVC05,TGVC20), var.equal=TRUE)
TOJDose05and20 <- t.test(len ~ dose, data=rbind(TGOJ05,TGOJ20), var.equal=TRUE)
```

Summary of the results:

```
##            dosage supplement      p-value -conf.int  +conf.int reject H0
## 1 0.5mg and 1.0mg         VC 6.492265e-07 -11.26435  -6.315654      TRUE
## 2 0.5mg and 1.0mg         OJ 6.492265e-07 -11.26435  -6.315654      TRUE
## 3 1.0mg and 2.0mg         VC 3.397578e-05 -12.96896  -5.771040      TRUE
## 4 1.0mg and 2.0mg         OJ 3.397578e-05 -12.96896  -5.771040      TRUE
## 5 0.5mg and 2.0mg         VC 4.957286e-09 -21.83284 -14.487157      TRUE
## 6 0.5mg and 2.0mg         OJ 4.957286e-09 -21.83284 -14.487157      TRUE
```

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

- Mean growth rate in both supplements is equal

We see there is a correlation between the supplement type (Orange Juice vs Vitamin C) and tooth growth that disappears at a higher dose. For 0.5mg and 1.0mg doses, we see that the p-values are below our threshold and the confidence interval does not include zero, indicating that the increase in tooth growth that's seen with OJ vs VC is statistically significant. Hence we reject the null hypothesis.

However, at 2.0mg this discrepancy vanishes and our null hypothesis at this level is accepted.

- Mean growth rate stays same with increase in dosage

There is a definite dependence on the dosage level on tooth growth. p-values for all tests were less than the threshold and confidence intervals do not include zero, indicating that the increase in tooth growth when supplement dose is increased is statistically significant and hence we reject the null hypothesis.