Statistical Inference Project2

Owusu

March 27, 2016

Project Overview

Now in this second portion of the 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.

1. Load Data

```
# load neccesary libraries
library(ggplot2)
library(datasets)
library(gridExtra)

# The Effect of Vitamin C on Tooth Growth in Guinea Pigs
data(ToothGrowth)
TG <- ToothGrowth
TG$dose <- as.factor(TG$dose) # convert to factor
```

2. Basic Summary

```
str(TG)
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(TG)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```r
head(TG)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```r
table(TG$supp, TG$dose)
```

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

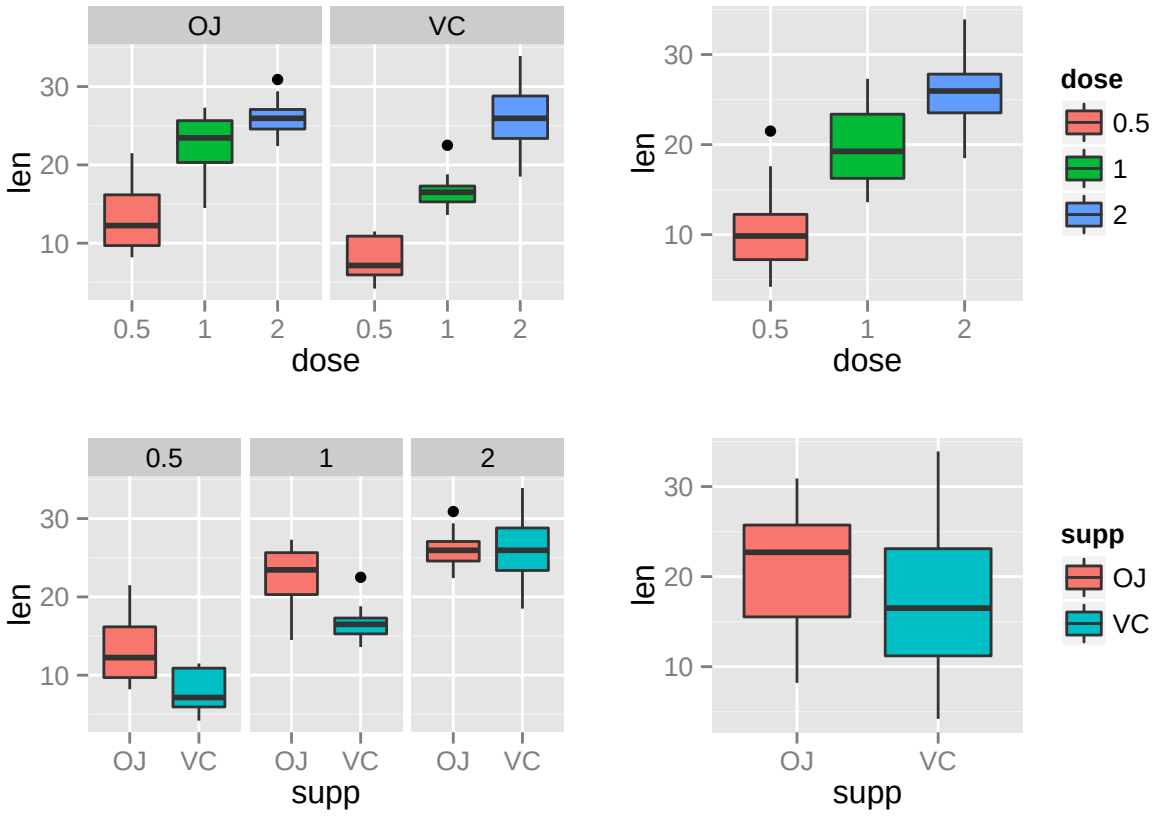
```r
p1 <- ggplot(data=TG, aes(x=dose,y=len,fill=dose)) +
  geom_boxplot() +
  theme(legend.position="none") +
  facet_grid(.~supp)

p2 <- ggplot(data=TG, aes(x=supp,y=len,fill=supp)) +
  geom_boxplot() +
  theme(legend.position="none") +
  facet_grid(.~dose)

p3 <- ggplot(data=TG, aes(x=supp,y=len,fill=supp)) +
  geom_boxplot()

p4 <- ggplot(data=TG, aes(x=dose,y=len,fill=dose)) +
  geom_boxplot()

grid.arrange(p1, p4, p2, p3, ncol = 2, nrow=2)
```



3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

```
# checking for the group differences due to the different supplement type
# assuming there are unequal variances between the two groups
t.test(len ~ supp, data = TG)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The p-value is 0.06, and the confidence interval contains zero. This indicates that we can not reject the null hypothesis that *the different supplement types have no effect on tooth length.*

```
# first create three sub-groups as per dose level pairs
TG.doses_0.5_1.0 <- subset (TG, dose %in% c(0.5, 1.0))
TG.doses_0.5_2.0 <- subset (TG, dose %in% c(0.5, 2.0))
TG.doses_1.0_2.0 <- subset (TG, dose %in% c(1.0, 2.0))
```

```r
# Checking for the group differences due to the different dose levels (0.5, 1.0)
# assuming there are unequal variances between the two groups
t.test(len ~ dose, data = TG.doses_0.5_1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
# Checking for the group differences due to the different dose levels (0.5, 2.0)
# assuming there are unequal variances between the two groups
t.test(len ~ dose, data = TG.doses_0.5_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```r
# Checking for the group differences due to the different dose levels (1.0, 2.0)
# assuming there are unequal variances between the two groups
t.test(len ~ dose, data = TG.doses_1.0_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

For all three dose level pairs, the p-value is less than 0.05, and the confidence interval does not contain zero. The mean tooth length increases on raising the dose level. This indicates that we can reject the null hypothesis, and establish that *increasing the dose level leads to an increase in tooth length.*

4. Conclusion and Assumptions

Conclusion

1. Supplement type has no effect on tooth growth.
2. Inreasing the dose level leads to increased tooth growth.

Assumptions

1. Members of the sample population, i.e. the 60 guinea pigs, are representative of the entire population of guinea pigs.
2. The experiment was done randomly by assigning the guinea pigs with different dose level categories and supplement type to control.
3. For the t-tests, the variances are assumed to be different for the two groups being compared. This assumption is less stronger than the case in which the variances are assumed to be equal.