Statistical Inference part 2

Asfa Lohani

Overview

This project aims to analyze the ToothGrowth data in the R datasets package. The data looks at the length of 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). We will load the data and perform some basic exploratory analysis. We will then provide a basic summary of the data. Furthermore, we will Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. Lastly, we will state our conclusions and the necessary assumptions we made.

Basic Exploratory Data Analysis

Let's load the data and call the str function to get some more information.

```
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 ...
```

The three variables are len (tooth length), supp(supplement of orange juice or acorbic acic) and dose (the dose levels .5,1,or2mg). Lets look at a quick plot showing the relationship between the supplement and tooth length. (see Appendix: Plot 1 for figure)

```
library(ggplot2)
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```

Let's also look at the relationship between dose and tooth length. (See Appendix: Plot 2 for figure)

```
library(ggplot2)
ggplot(aes(x=as.factor(dose), y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
```

From the first graph, OJ seems to have the greater impact on tooth length as can be seen with the higher mean. From the second graph it appears that an increased dose results in increased tooth length.

Basic Summary of data

A quick summary of the data gives us some more information regarding the mean and quantiles of tooth length.

```
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
```

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

Let's first compare tooth growth (len) by supplement (supp). Our null and alternative hypothesis are: Ho: There is no difference in tooth growth depending on the supplement used. Ha: There is a difference in tooth growth depending on the supplement used.

Let's perform a t test (See Appendix: ttest len by supp for output) :

```
t.test(len~supp,data=ToothGrowth)
```

The p-value is greater than .05 and The 95% confidence interval contains zero. This indicates that we cannot reject the null hypothesis and that there is no difference in tooth growth depending on the supplement used.

Lets take this a little further and test len by supp depending on the the dose levels. We will subset the dataset for each dose level and compare len by supp.(See Appendix: ttest len by supp at each dose level for output)

```
dose.5 <- ToothGrowth[ToothGrowth$dose==0.5,]
  t.test(dose.5$len ~ dose.5$supp, data=ToothGrowth)

dose1 <- ToothGrowth[ToothGrowth$dose==1,]
  t.test(dose1$len ~ dose1$supp, data=ToothGrowth)

dose2 <- ToothGrowth[ToothGrowth$dose==2,]
  t.test(dose2$len ~ dose2$supp, data=ToothGrowth)
```

From the results we can see that for the .5 and 1 dose levels we can reject the null and conclude that there is a difference in tooth growth depending on which supplement is used as the p value is less than .05 and the confidence interval does not contain 0. For the 2 dose level, we accept the null and conclude that at this dose level there is no difference in tooth length depending on the supplement used as the p value is greater than .05 and the confidence interval contains 0.

Let's also compare len by dose. You can only compare two levels so we will subset the data set to only include two levels in the dose column and then perform the t test. Our null hypothesis will be Ho: There is no difference in tooth growth depending on the dose amount.(See Appendix: ttest len by dose for output)

```
len1<- subset(ToothGrowth, dose==0.5 | dose == 1 )
t.test(len1$len ~ len1$dose, data = ToothGrowth)

len2<- subset(ToothGrowth, dose==0.5 | dose == 2 )
t.test(len2$len ~ len2$dose, data = ToothGrowth)

len3<- subset(ToothGrowth, dose==1 | dose == 2 )
t.test(len3$len ~ len3$dose, data = ToothGrowth)
```

From the results we can see that for all three tests the p value is very small (less than .05) and the confidence interval does not contain 0. Therefore we reject the null hypothesis and conclude that tooth growth is affected by the dose amount.

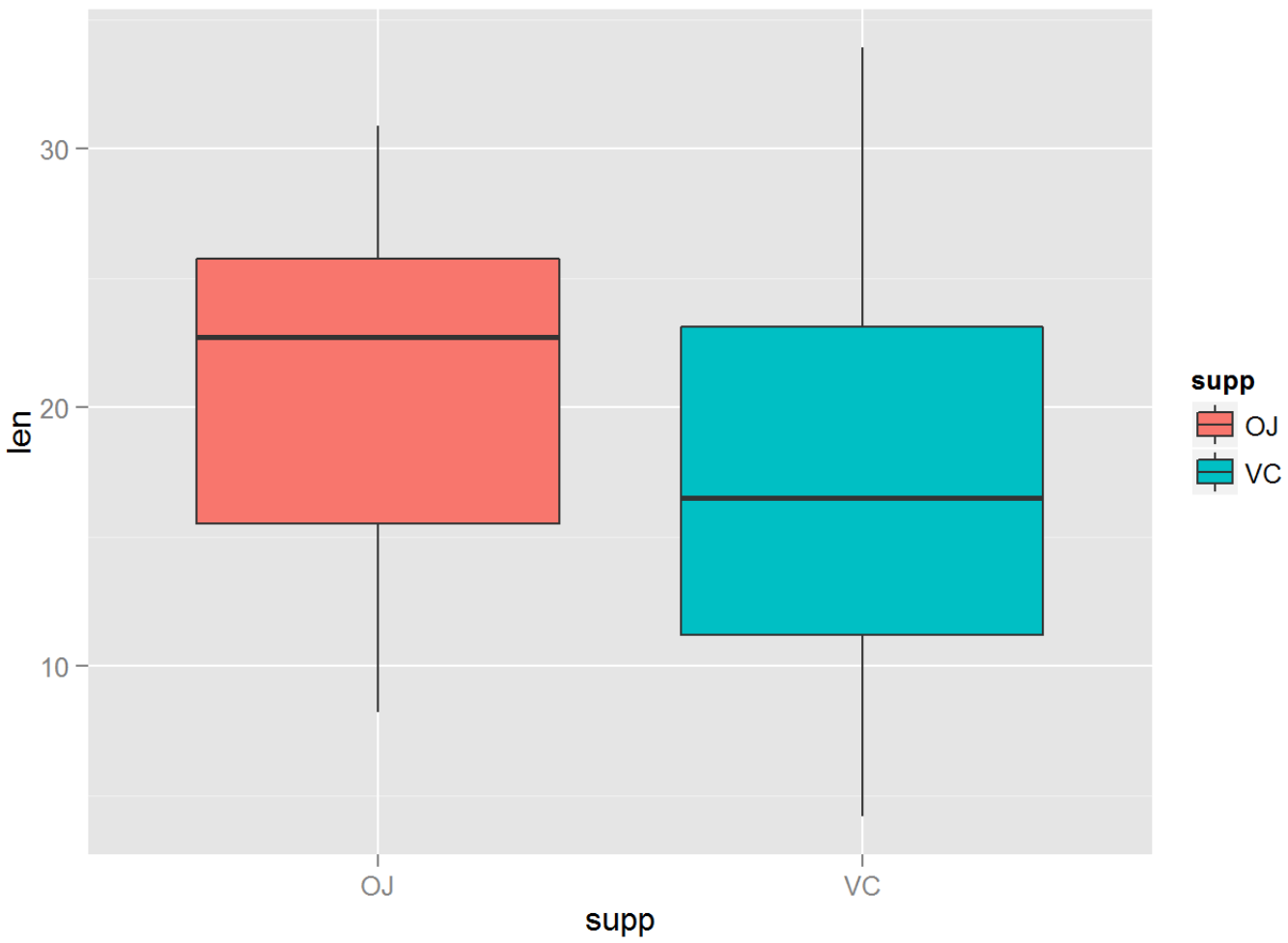
Conclusions and Assumptions

Conclusions: 1. Tooth growth is not affected by the type of supplement used. 2. At the specific dose level of .5 and 1 tooth growth is affected by the the type of supplement used. OJ has greater affect. 3.Tooth growth is affected by the dose level. An increase in dose increases tooth growth.

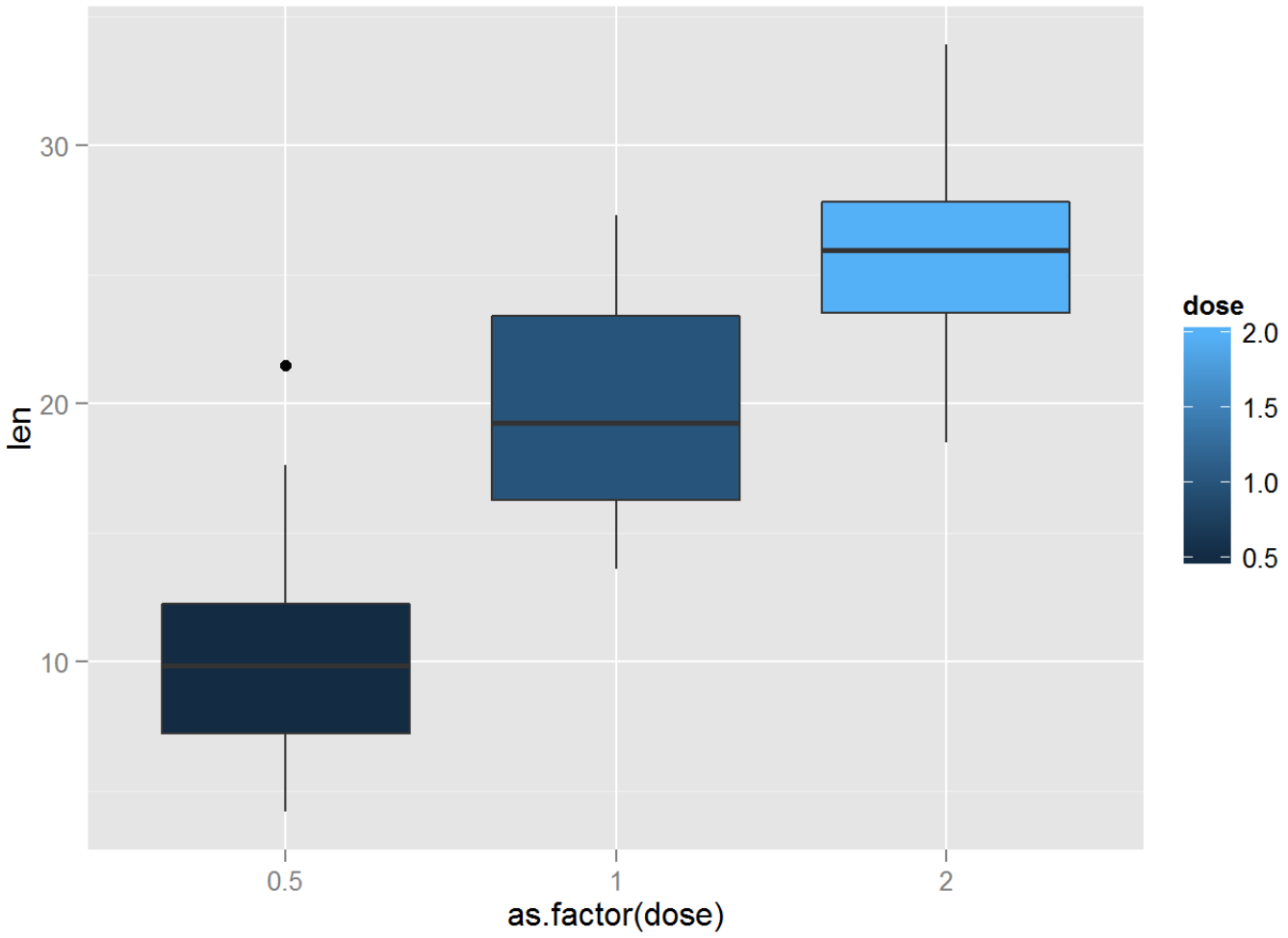
Assumptions: 1. The data are iid normal 2. Equal Variance

Appendix

PLOT 1



Plot2



ttest len by supp output

```
##
##  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
```

ttest len by supp at each dose level output

```
##
##  Welch Two Sample t-test
##
## data:  dose.5$len by dose.5$supp
## 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 in group OJ mean in group VC
##            13.23             7.98
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len by dose1$supp
## 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 in group OJ mean in group VC
##            22.70             16.77
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2$len by dose2$supp
## t = -0.0461, 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 in group OJ mean in group VC
##            26.06            26.14
```

ttest len by dose output

```
##
##  Welch Two Sample t-test
##
## data:  len1$len by len1$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
```

```
##
##  Welch Two Sample t-test
##
## data:  len2$len by len2$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
```

```
##
##  Welch Two Sample t-test
##
## data:  len3$len by len3$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
```