Statistical Inference - Course Project 2

Mark C

Saturday, November 17, 2014

Part 2 : Basic Inferential Data Analysis

Project Info

In the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.Perform some exploratory data analysis and perform hypothesis testing on the data.

Data Info

The response is 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).

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

```
#loading the library required
library(ggplot2)

# loading the dataset
library(datasets)
data(ToothGrowth)

# Check how many rows in the dataset
nrow(ToothGrowth)
```

```
## [1] 60
```

```
# Preview part of the data
head(ToothGrowth)
```

```
##    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
```

```
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 ...
```

```r
# Perform some pre-processing of the data
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

2. Provide a basic summary of the data.

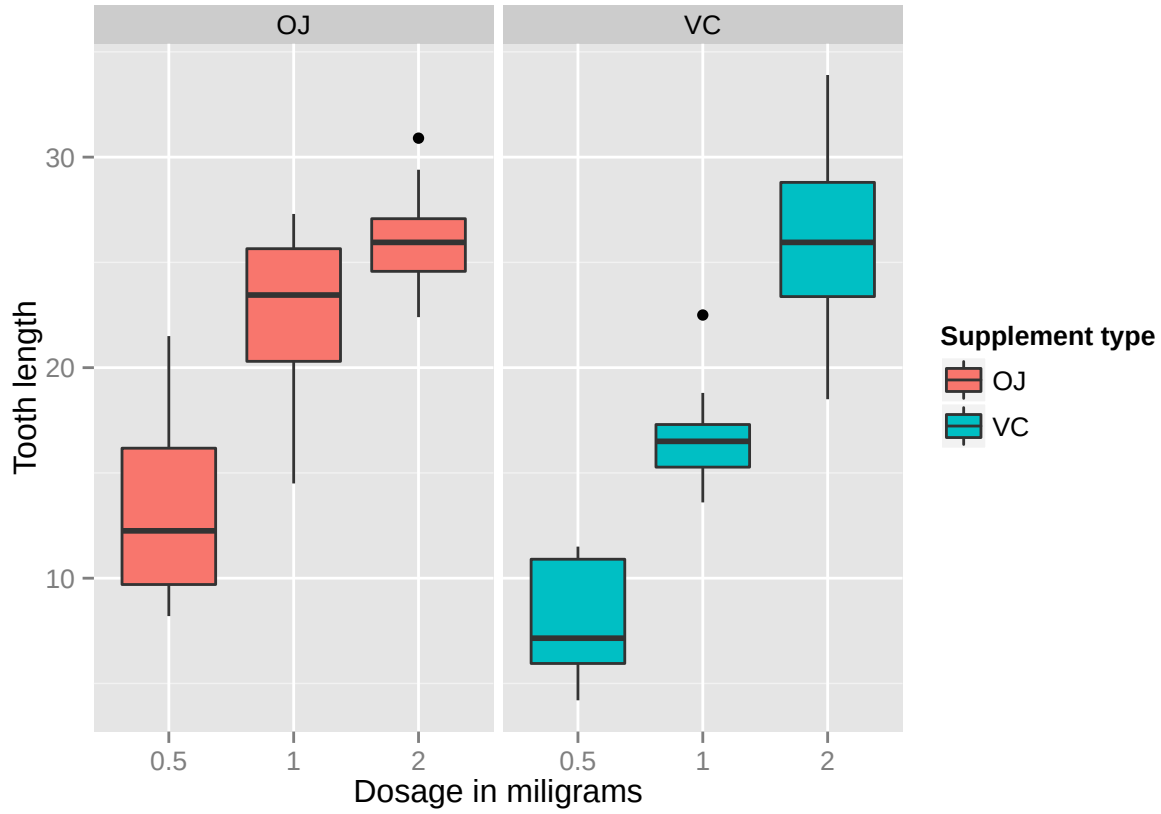
```r
#generate the summary statistics of the data
summary(ToothGrowth)
```

```
##       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
# split the data  between different dose levels and delivery methods
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

```r
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_boxplot() +
    facet_grid(. ~ supp) +
    xlab("Dosage in miligrams") +
    ylab("Tooth length") +
    guides(fill=guide_legend(title="Supplement type"))
```



Fromt the graph above , it shows that there is a positive correlation between the tooth length and the dosage levels for both delivery method . The length of the tooth increases as the dosage increases. However, it can be seen that the OJ delivery method yields a greater length increate than the VC.

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

Statistical inference

Now , we conduct the hypothesis testing to compare whether the same dosage of each supplement yields the same tooth growth length. We perform the null hypothesis test as follows:

$H_0 : \mu_{OJ|0.5} = \mu_{VC|0.5}$
$H_0 : \mu_{OJ|1.0} = \mu_{VC|1.0}$
$H_0 : \mu_{OJ|2.0} = \mu_{VC|2.0}$

```
# We split the data up by dosages
d0.5 <- subset(ToothGrowth, dose == 0.5)
d1.0 <- subset(ToothGrowth, dose == 1.0)
d2.0 <- subset(ToothGrowth, dose == 2.0)

# conduct a t-test between supplements
test0.5 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d0.5)
test0.5$p.value; test0.5$conf[1]
```

```
## [1] 0.006358607
```

```
## [1] 1.719057
```

```
test1.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d1.0)
test1.0$p.value; test1.0$conf[1]
```

```
## [1] 0.001038376
```

```
## [1] 2.802148
```

```
test2.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d2.0)
test2.0$p.value; test2.0$conf[1]
```

```
## [1] 0.9638516
```

```
## [1] -3.79807
```

Observations

Dosages 0.5 and 1 have p-values of 0.0063586 and 0.0010384 respectively and is below 0.01. This shows there is a strong presumption against the null hypothesis, indicating that the difference in mean values between the supplements is significant.

Dosage 1.0 has a confidence interval between 1.7190573 - 8.7809427 and dosage 2.0 has a confidence interval between 2.8021482 - 9.0578518.

Dosage 2 has a very high p-value of 0.9638516 and a confidence interval below zero (-3.7980705 - 3.6380705). The p-value this high indicates that there is no presumption against the null hypothesis. This means that there is no significance between the supplemen at the dosage at this amount.

Conclusion

From the analysis, it can be seen that the tooth length growth is much more effective at lower dosages (0.5 & 1) for both supplements and the effects on tooth growth wears off at dosage 2.0.

Assumptions Applied

1)Samples obtained are unpaired, with unequal variances. 2)Test subjects are similar (size, age, diet). 3)No other factors other than the supplements are the factors affecting tooth growth in this case.