Statistical Inference: Peer Assessment, Part 2

S C NG

20 June, 2015

Overview

In the second portion of the assignment, we're going to analyze the ToothGrowth data in the R datasets package. We will use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose, and make the conclusion with assumptions needed.

Setup

First of all, the following default settings and libraries are loaded

```
#preset default options for Rmd, codes not shown in report
require(knitr)
```

```
## Loading required package: knitr
```

```
opts_chunk$set(cache=TRUE, echo=TRUE)

#load required libraries for data analysis
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(datasets)
```

Exploratory of Data

According to the help page, the ToothGrowth dataset recorded 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). Below R code shows help page and the data structure of ToothGrowth dataset

```
?ToothGrowth
str(ToothGrowth)
```

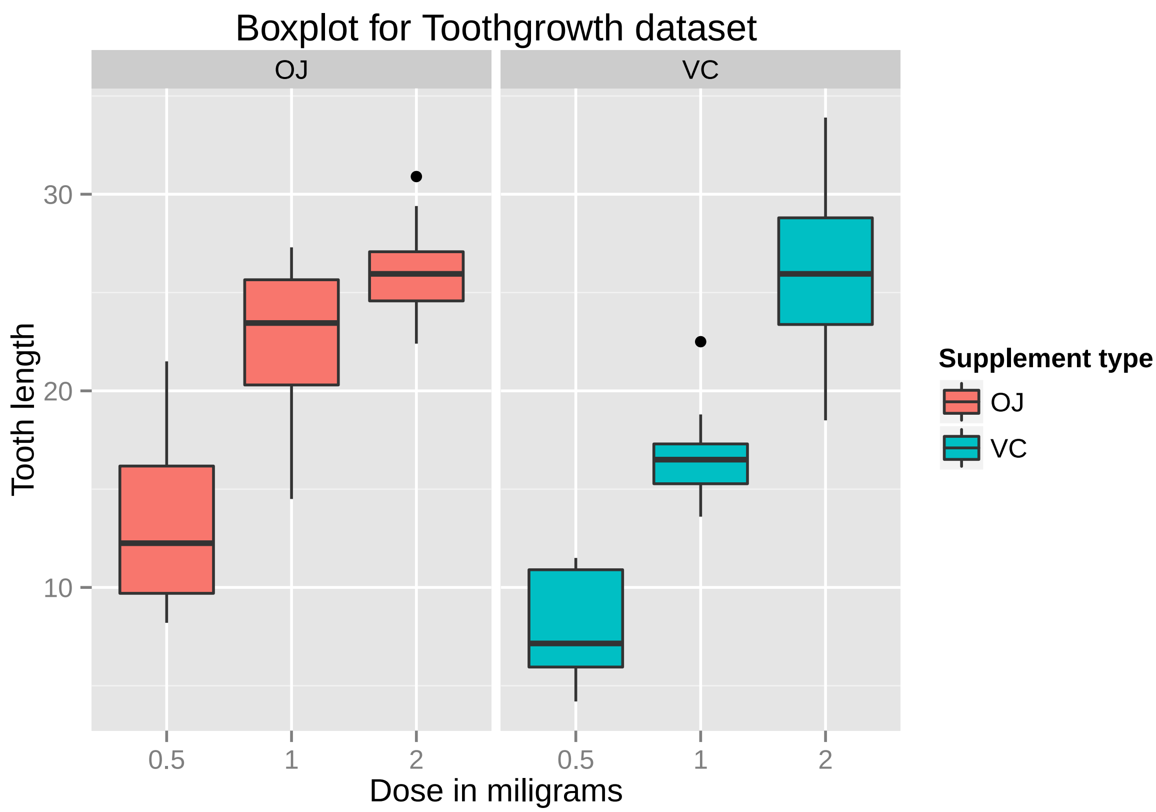
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

The dataset consists of 60 observations on 3 variables:

1. len (numeric) = Tooth length
2. supp (factor) = Supplement type (VC or OJ).
3. dose (numeric) = Dose in milligrams.

Below boxplot summarize the Tooth length values per each three dose levels, 1 plot per each Supplement type

```
g2 <- ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp))
g2 <- g2 + geom_boxplot()
g2 <- g2 + facet_grid(. ~ supp)
g2 <- g2 + xlab("Dose in miligrams") + ylab("Tooth length")
g2 <- g2 + guides(fill = guide_legend(title ="Supplement type"))
 g2 <- g2 + ggtitle("Boxplot for Toothgrowth dataset")
g2
```



Data Analysis

The following section tried to use confidence intervals and hypothesis tests to compare the tooth growth by supplement type and dose

Tooth Length by Supplement Type

We will start by comparing the tooth length by supplement type through the following test in R

```r
t1 <- t.test(len~supp, paired = FALSE, data=ToothGrowth)
t1
```

```
##
##  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 95% confidence interval contains zero, and the p-value is slightly larger than 0.05. It was failed to reject the null hypothesis that there would be no difference on tooth length if we deliver Vitamin C through orange juice or ascorbic acid

Tooth Length by Dose

Then we will check the tooth length by dose. we will start doing this by separating the dataset in 3 groups by dose

```r
ToothGrowth.05 <- ToothGrowth[ToothGrowth$dose == 0.5, "len"]
ToothGrowth.10 <- ToothGrowth[ToothGrowth$dose == 1.0, "len"]
ToothGrowth.20 <- ToothGrowth[ToothGrowth$dose == 2.0, "len"]
```

Next we will perform the t-test between different doses

```r
t.05.10 <- t.test(ToothGrowth.05, ToothGrowth.10, paired = FALSE)
t.05.10
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth.05 and ToothGrowth.10
## 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 of x mean of y
##    10.605    19.735
```

```r
t.10.20 <- t.test(ToothGrowth.10, ToothGrowth.20, paired = FALSE)
t.10.20
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  ToothGrowth.10 and ToothGrowth.20
## 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 of x mean of y
##    19.735    26.100
```

```
t.05.20 <- t.test(ToothGrowth.05, ToothGrowth.20, paired = FALSE)
t.05.20
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth.05 and ToothGrowth.20
## 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 of x mean of y
##    10.605    26.100
```

All the 95% confidence intervals do not contain zero, and the p-values are much lower than 0.05. The difference betwen the means of teeth length for different doses are significant, and we can reject the null hypothesis that there would be no difference on tooth length if we use different dose of Vitamin C

Tooth Length by Supplement Type Under The Same Dose

Finally, we will compare the effect on different different delivery types under the same dose level. First of all we will prepare separate datasets

```
ToothGrowth.05.VC <- ToothGrowth[ToothGrowth$dose == 0.5 & ToothGrowth$supp == "VC", "len"]
ToothGrowth.05.OJ <- ToothGrowth[ToothGrowth$dose == 0.5 & ToothGrowth$supp == "OJ", "len"]
ToothGrowth.10.VC <- ToothGrowth[ToothGrowth$dose == 1.0 & ToothGrowth$supp == "VC", "len"]
ToothGrowth.10.OJ <- ToothGrowth[ToothGrowth$dose == 1.0 & ToothGrowth$supp == "OJ", "len"]
ToothGrowth.20.VC <- ToothGrowth[ToothGrowth$dose == 2.0 & ToothGrowth$supp == "VC", "len"]
ToothGrowth.20.OJ <- ToothGrowth[ToothGrowth$dose == 2.0 & ToothGrowth$supp == "OJ", "len"]
```

Then we will conduct t-tests to compare the effect of different delivery under the same dose

```
t.05 <- t.test(ToothGrowth.05.VC, ToothGrowth.05.OJ, paired = FALSE)
t.05
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth.05.VC and ToothGrowth.05.OJ
## 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:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

```r
t.10 <- t.test(ToothGrowth.10.VC, ToothGrowth.10.OJ, paired = FALSE)
t.10
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth.10.VC and ToothGrowth.10.OJ
## 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:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

```r
t.20 <- t.test(ToothGrowth.20.VC, ToothGrowth.20.OJ, paired = FALSE)
t.20
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth.20.VC and ToothGrowth.20.OJ
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

With doses 0.5 or 1.0 mg of Vitamin C, there are significant difference on the effect of tooth growth if we choose a different delivery method, since the 95% confidence intervals do not contain zero, and the p-values are lower than 0.05, and we can reject the null hypothesis. However, with dose 2.0 mg of Vitamin C, the difference on the effect of tooth growth is not significant if we change the delivery method, since the 95% confidence interval contains zero, and the p-value is higher than 0.05, and it is failed to reject the null hypothesis.

Conclusions

1. Changing delivery methods (orange juice or ascorbic acid) may not have significant effect on the tooth growth
2. Changing Vitamin C dose have significant effect on the tooth growth
3. With lower Vitamin C dose level (0.5 or 1.0 mg), the change of delivery method have significant effect on the tooth growth. However, the effect of delivery change may not significant when the dose level is highe (2.0 mg)

Assumptions

For the above tests, we assume that the 60 guinea pigs in the sample are randomly sampled from the population that are normally distributed. The samples are all independent.