Statistical_Inference_Project_Part_2

freestander

Sunday, July 26, 2015

Problem Description

We will analyze the ToothGrowth data in the R datasets package.

The project is divided into 4 sections:

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

ToothGrowth Data and Exploratory Analysis

First we load the ToothGrowth dataset into the workspace.

```
# load the dataset library
library(datasets)
# load the ToothGrowth dataset
data(ToothGrowth)
# attach the ToothGrowth dataset
attach(ToothGrowth)
```

Check the format of the dataset by listing the first few records.

```
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
```

Check the structure of the dataset. We can see that supplement only has two levels “OJ” and “VC”.

```
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 ...
```

Summary of the Data

Show the summary of the dataset.

```
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
```

Check the sample size partitioned by supplement and dose.

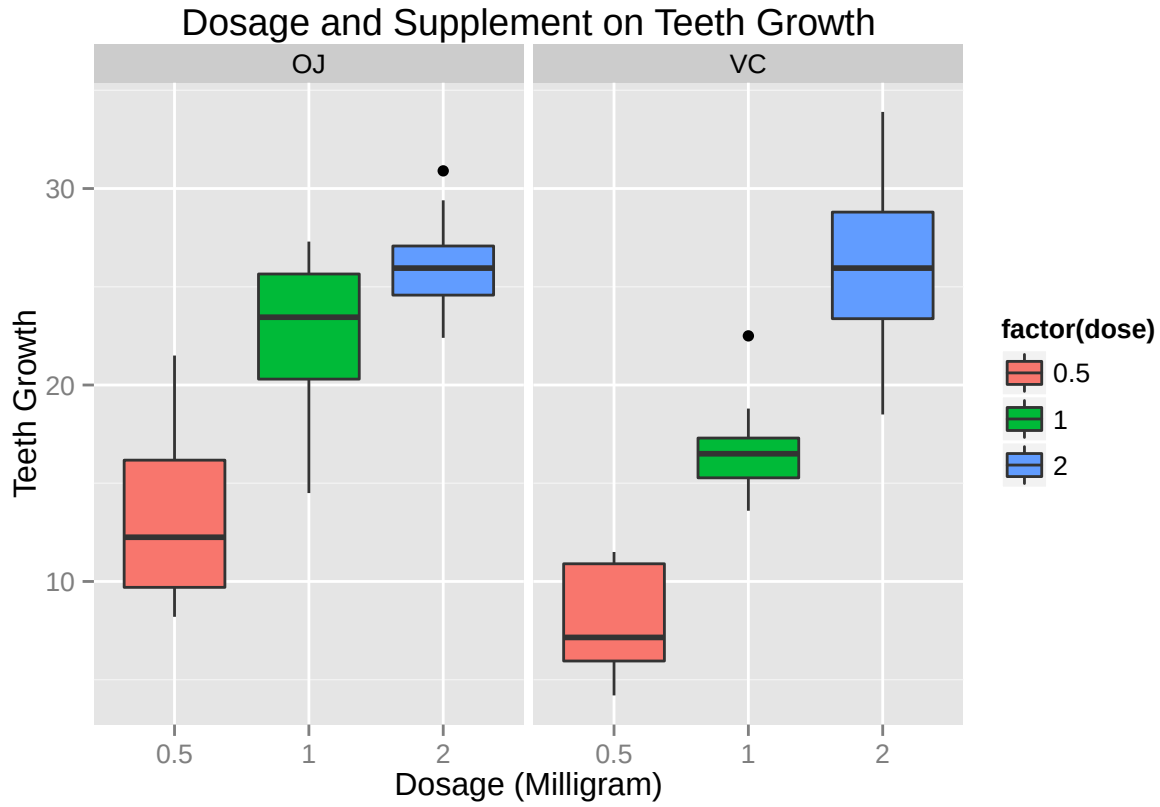
```
# convert dose to a factor variable
dose <- as.factor(dose)
# tabulate the sample size by supplement and dose
table(supp, dose)
```

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

We can see that each supplment and dose value combination has 10 samples. The ToothGrowth dataset contains teech growth at three dose sizes (0.5, 1, 2) with two supplements ("OJ" and "VC").

Next we draw a boxplot to visualize the effects of dose and supplement on teeth growth.

```
# load ggplot library
library(ggplot2)
# draw the plot
p <- ggplot(ToothGrowth, aes(x=factor(dose), y=len,fill=factor(dose)))
p <- p + geom_boxplot(notch=F) + facet_grid(.~supp)
p <- p + scale_x_discrete("Dosage (Milligram)")
p <- p + scale_y_continuous("Teeth Growth")
p <- p + ggtitle("Dosage and Supplement on Teeth Growth")
p
```



From the boxplot we can roughly tell:

1. For both supplements ("OJ" and "VC"), the teech growth increases as the dose increases.
2. When dose = 0.5 and 1, "OJ" is more effective than "VC" on teech growth. However, when dose = 2, there is no obvious difference between "OJ" and "VC" on teeth growth.

Next we are going to verify the visualization results using T tests assuming non-paired and unequal variance between the tested groups.

Confidence Interval and Hypothesis Test

The following T test shows the difference of teech gowth between dose 0.5 and 1. Both the confidence interval (<0) and small p-value (<0.05) suggest more teech growth of dose = 1 over dose = 0.5.

```
# T Test by dose between 0.5 and 1.
t.test(len ~ dose, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose %in% c(0.5, 1), ])
```

```
##
##  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
```

The following T test shows the difference of teech gowth between dose 0.5 and 2. Both the confidence interval (<0) and small p-value (<0.05) suggest more teech growth of dose = 2 over dose = 0.5.

```
# T Test by dose between 0.5 and 2.
t.test(len ~ dose, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose %in% c(0.5, 2), ])
```

```
##
##  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
```

The following T test shows the difference of teech gowth between dose 1 and 2. Both the confidence interval (<0) and small p-value (<0.05) suggest more teech growth of dose = 2 over dose = 1.

```
# T Test by dose between 1 and 2.
t.test(len ~ dose, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose %in% c(1, 2), ])
```

```
##
##  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
```

The following T test shows the difference of teech gowth between two supplements. The confidence interval contains 0 and p-value (>0.05) suggest there is no significant difference between both supplements.

```
# T Test by supplement type
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
##  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 following T test shows the difference of teech gowth between two supplements when dose = 0.5. Both the confidence interval (<0) and small p-value (<0.05) suggest more teech growth of supplement "OJ" over "vc".

```
# T Test by supplement type when when dose=0.5.
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose == 0.5, ])
```

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

The following T test shows the difference of teech gowth between two supplements when dose = 1. Both the confidence interval (<0) and small p-value (<0.05) suggest more teech growth of supplement "OJ" over "vc".

```
# T Test by supplement type when dose=1.
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose == 1, ])
```

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

The following T test shows the difference of teech gowth between two supplements when dose = 2. The confidence interval contains 0 and p-value (>0.05) suggest there is no significant difference between both supplements.

```
# T Test by supplement type when dose=2.
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth[ToothGrowth$dose == 2, ])
```

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

```
##
## data:  len by supp
## t = -0.046136, 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
```

Conclusions and Assumptions

1. The p-value and confidence intervals suggest the dose size has significant effect on teech growth, which aligns with our visualization result.
2. The p-value and confidence intervals suggest "OJ" over "vc" has more significant impact on teech effect when dose = 0.5 and 1, but fail to prove significant difference when dose = 2, which also aligns with our visualization result.

As a result, we can conclude that there is significant difference in teech growth when dose size increases. However, we cannot conclude there is significant difference in teech growth caused by the two supplements.

Finally, all the T tests assume non-paired and unequal variance between the tested groups.