

Statistical Inference Project Report Part 2

Wien Wong

Sunday, January 25, 2015

Brief Introduction

This is the project report part 2 for the Online Coursera course – Statistical Inference. Basic tasks of this project includes fundamental inferential data analysis for the “ToothGrowth” data.

Load the ToothGrowth data and perform some basic exploratory data analyses.

```
TG <- data.frame(ToothGrowth)
head(TG,3)
```

```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

Three variables ‘len’, ‘supp’ and ‘dose’ can be seen. Detailed exam the dataset reveals 60 observations. Teeth length are classified into sample groups, each with 10 guinea pigs. Three dose levels of 0.5mg, 1mg and 2mg with either OJ (orange juice) or VC (vitamin C) delivery method are used for either group.

The ‘dose’ is of type numeric, convert it to factor for ggplot and t-test analyse.

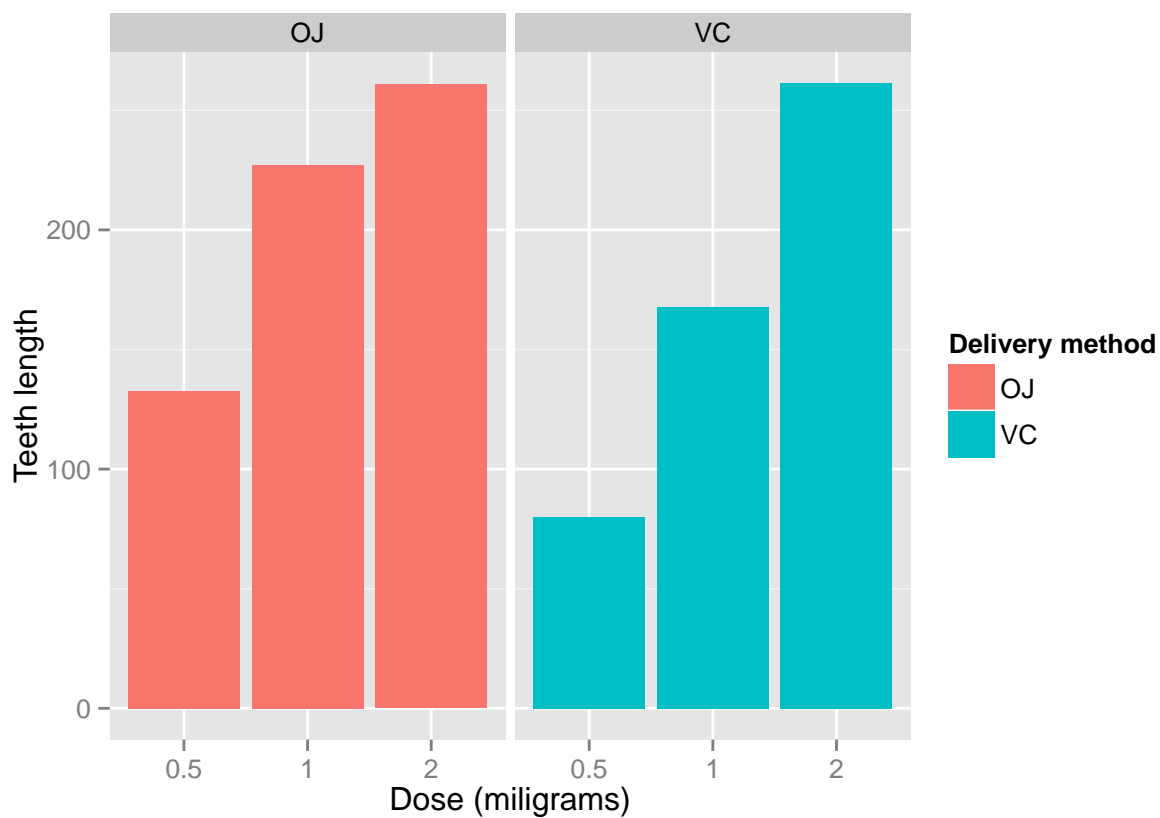
```
TG$dose <- as.factor(TG$dose)
```

To visualize the data, a plot is necessary.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

```
ggplot(data = TG, aes(x = dose, y = len, fill = supp)) +
  geom_bar(stat = "identity",) +
  facet_grid(. ~ supp) +
  xlab("Dose (milligrams)") +
  ylab("Teeth length") +
  guides(fill = guide_legend(title = "Delivery method"))
```



For both delivery methods, there is a positive correlation between the teeth length and the dose levels.

Provide a basic summary of the data.

Summary of the data set.

```
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
```

Summary of the delivery method of OJ.

```
summary(TG[TG$supp == "OJ", ])
```

```
##      len      supp  dose
##  Min.   : 8.20   OJ:30  0.5:10
##  1st Qu.:15.53   VC: 0  1  :10
##  Median :22.70           2  :10
```

```
## Mean :20.66
## 3rd Qu.:25.73
## Max. :30.90
```

Summary of the delivery method of VC.

```
summary(TG[TG$supp == "VC", ])
```

```
##      len      supp      dose
## Min.   : 4.20    OJ: 0    0.5:10
## 1st Qu.:11.20    VC:30    1 :10
## Median :16.50          2 :10
## Mean   :16.96
## 3rd Qu.:23.10
## Max.   :33.90
```

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

Set the p-value as 0.05 and the null hypothesis as the difference in length means is equal to 10.

Under dose of 0.5mg, compare which delivery method contributes more for teeth growth:

```
t.test(len ~ supp, TG[TG$dose == 0.5, ], mu = 10, paired = FALSE, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: len by supp
## t = -2.8679, df = 18, p-value = 0.01023
## alternative hypothesis: true difference in means is not equal to 10
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean in group OJ mean in group VC
##           13.23           7.98
```

The p-value of 0.01023 is less than 0.05, thus the null hypothesis should be rejected. One other line shows evidence of the alternative hypothesis that the true difference in means is not equal to 10. Notice that it also gives us a 95% confidence interval of the true difference in means of length from 1.770262 to 8.729738. Based on the sample mean of each method, delivery method of OJ imposes greater teeth growth than delivery method of VC.

Under dose of 1mg, compare which supplement type contributes more for teeth growth:

```
t.test(len ~ supp, TG[TG$dose == 1, ], mu = 10, paired = FALSE, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: len by supp
## t = -2.7679, df = 18, p-value = 0.01268
## alternative hypothesis: true difference in means is not equal to 10
```

```
## 95 percent confidence interval:
## 2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
## 22.70 16.77
```

Similar analysis applies for this comparison. Still, delivery method of OJ imposes greater teeth growth than delivery method of VC.

Under dose of 2mg, compare which supplement type contributes more for teeth growth:

```
t.test(len ~ supp, TG[TG$dose == 2, ], mu = 10, paired = FALSE, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: len by supp
## t = -5.8131, df = 18, p-value = 1.656e-05
## alternative hypothesis: true difference in means is not equal to 10
## 95 percent confidence interval:
## -3.722999 3.562999
## sample estimates:
## mean in group OJ mean in group VC
## 26.06 26.14
```

The null hypothesis should be rejected. In dose of 2mg case, there's no significant difference between the two supplement types. This is also reflected from the plot.

Compare dose of 0.5mg with dose of 1mg:

```
t.test(len ~ dose, TG[TG$dose != 2, ], mu = 10, paired = FALSE, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: len by dose
## t = -13.5705, df = 38, p-value = 3.776e-16
## alternative hypothesis: true difference in means is not equal to 10
## 95 percent confidence interval:
## -11.983748 -6.276252
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

The null hypothesis should be rejected. Dose of 1mg causes more teeth growth than dose of 0.5mg, no matter which method applied.

Compare dose of 1mg with dose of 2mg:

```
t.test(len ~ dose, TG[TG$dose != 0.5, ], mu = 10, paired = FALSE, var.equal = TRUE)
```

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

```
##
## data: len by dose
## t = -12.5996, df = 38, p-value = 3.844e-15
## alternative hypothesis: true difference in means is not equal to 10
## 95 percent confidence interval:
## -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The null hypothesis should be rejected. Dose of 2mg causes more teeth growth than dose of 1mg, no matter which method applied.

Given results of the previous tests, there is no need to compare dose of 0.5mg with dose of 2mg.

State your conclusions and the assumptions needed for your conclusions.

For different delivery methods, the OJ delivery method has more influence on teeth growth than the VC method under low dose (0.5mg and 1mg) cases, however, in high dose (2mg) case there's no significant difference between these two methods.

For different doses, this data set shows an overall tendency that the larger amount of the dose, the larger the teeth growth, no matter which method applied.

Assumptions could be: 1) There is no other factor that would contribute on the teeth length. 2) The delivery methods and the dose amount are mutual independent. 3) All the guinea pigs are randomly independent selected. 4) Guinea pigs do not have preference to any dose amount or delivery methods.