Statistical Inference, Course Project - Part2

Ivan Maksimov

18.01.2015

1 Introdaction

In this report we should analyze the ToothGrowth data in the R datasets package.

2 Data analysis

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

Load the dataset

```
library(datasets)
data(ToothGrowth)
```

Look at the dataset variables

```
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
```

Number of rows of dataset

```
nrow(ToothGrowth)
```

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

Convert variable dose from numeric to factor

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

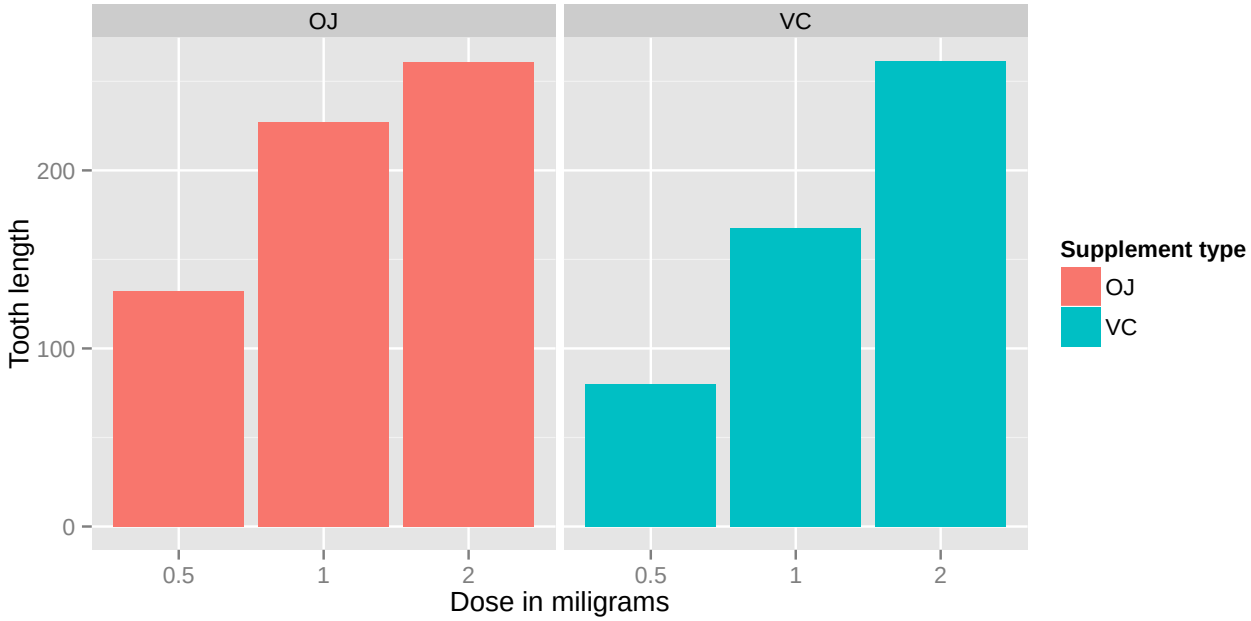
2.2 Provide a basic summary of the data

There are 3 variables in the data set. len is the continuous variable stands for the length of tooth with the range from 4.2 to 33.9. supp is the delivery method with 2 values (VC or OJ). And dose is the dose of Vitamin C with 3 values(0.5, 1, and 2).

```
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
```

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



As can be seen above, there is a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods.

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

P-value

Split the data based on dosage

```
d0.5 <- subset (ToothGrowth, dose == 0.5)
d1.0 <- subset (ToothGrowth, dose == 1.0)
d2.0 <- subset (ToothGrowth, dose == 2.0)
```

T-test between supplements

```r
test0.5 <- t.test (len ~ supp, paired = FALSE, var.equal = FALSE, data = d0.5)
test0.5$p.value
```

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

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

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

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

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

The results of the T-test shows that dosage of 0.5mg and 1mg have relatively low p-value (0.006358 and 0.001038) as compared to the 2.0mg dosage which have a p-value of 0.963851.

For p-values that are more than 0.05, we can conclude that there is little differences in the effect on the length of tooth for the same Vitamin C dosage for the 2 different delivery methods.

For p-values that are less than 0.05, there is a noticeable difference in the effect on the length of tooth for the same Vitamin C dosage for the 2 different delivery methods. In this particular study, the effect of the 2 different delivery is more visible for the 1.0mg dosage (smallest p-value).

2.4 State your conclusions and the assumptions needed for your conclusions

For different delivery methods, there are significant difference in length of tooth when the dose is small (0.5 or 1), but no significant difference when the dose is large.

For different doses, there are always significant difference in length of tooth with any delivery methods. Here is the assumptions:

1. There is no other factor that would pose confounding effect on the length of tooth.
2. All the guinea pigs are randomly selected, and there is no significant difference in any attribute that might influence the length of tooth (such as the size of guinea pigs).
3. Different delivery methods would not affect the effect of dose size, that's to say the delivery methods are independent of dose size, and vice versa.
4. Guinea pigs do not have perference to any dose size or delivery methods, or guinea pigs' reaction to dose size and dose size will not affect the length of tooth.