Statistical Inference

Peer Assessment Part 2

Sergei Titov

Overview

This part of Statistical Inference project consider with the ToothGrowth data in the R datasets package short analysis. The goal of this report is:

- Perform some basic exploratory data analysis
- Provide a basic summary of the data.
- Compare tooth growth by supp and dose using confidence intervals and hypothesis tests.
- State conclusions.

Load and describe the data

Loading the data from datasets package.

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

Quick summary on the data.

```
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 provide the Effect of Vitamin C on Tooth Growth in Guinea Pigs. Data frame has 60 observations on 3 variables.

[,1] len. Tooth length of guinea pigs
[,2] supp. Supplement type: VC or OJ (ascorbic acid or orange juice).
[,3] dose. Dose in milligrams.

```
head (ToothGrowth, 3)
```

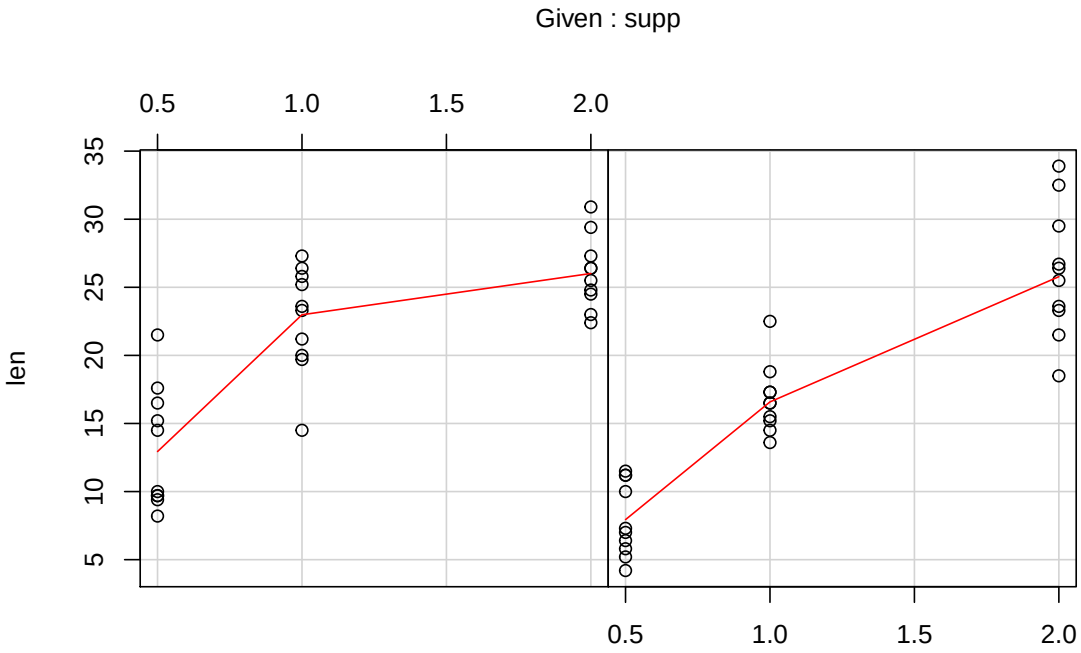
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

```
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
```

Plot the data.

```
coplot(len ~ dose | supp, data = ToothGrowth,
       panel = panel.smooth, show.given = F,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



ToothGrowth data: length vs dose, given type of supplement

Statistical analysis

To analyze tooth growth, we provide two-sample (orange juice vs. ascorbic acid) T-tests at each dosage level. The T-test at 0.5 mg yields the following:

```
tt.5 <- t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ], var.equal = T)
tt.5
```

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

```
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

P value equals 0.53% and the 95% confidence interval doesn't contain zero.
We reject the null hypothesis and take alternative - true difference in means is not equal to 0.

The T-test at 1 mg yields the following:

```
tt1 <- t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1,  ], var.equal = T)
tt1
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
##            22.70             16.77
```

P value equals 0.08% and the 95% confidence interval doesn't contain zero.
We reject the null hypothesis and take alternative - true difference in means is not equal to 0.

The T-test at 2 mg yields the following:

```
tt2 <- t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2,  ], var.equal = T)
tt2
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.722999  3.562999
## sample estimates:
## mean in group OJ mean in group VC
##            26.06             26.14
```

P value equals 96.37% and the 95% confidence interval contains zero.
We fail to reject the null hypothesis - true difference in means is equal to 0.

The overall T-Test, not including dose variable provide the folowing:

```r
t.test (len ~ supp, ToothGrowth, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

P value equals 6.04% and the 95% confidence interval contains zero.
We fail to reject the null hypothesis - true difference in means is equal to 0.

Conclusions

For this experiment, we've assumed there's a common variance in the guinea pigs population (var.equal = TRUE) and the test is not paired (default paired = FALSE).

From the exploratory data analysis, we see that increased vitamin C dosages (in either orange juice or ascorbic acid form) is an effective of tooth growth.

From the T-test analysis, we see that for dosages of 0.5 mg and 1 mg, orange juice is more effective at promoting tooth growth than just ascorbic acid. From the p-value for the 2 mg, we cannot conclude that orange juice promotes tooth growth more effectively than just ascorbic acid.

The overall view on supplement type (not considering doze value) give us a conclusion, that supplement type has no effect on tooth growth