Statistical Inference Project 2

Eric VACHON

May 2015

Overview

In this work we decsribe and analyse the "ToothGrowth data" and we are going to compare tooth growth by supp and dose. We are going to use confidence intervals and/or hypothesis tests to compare them before concluing.

Description

The data "ToothGrowth" is the Effect of Vitamin C on Tooth Growth in Guinea Pigs.
The response is 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 : OJ or ascorbic acid : VC).

Load the data set and Basic summary

```
data(ToothGrowth)
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
```

```
str(ToothGrowth)
```

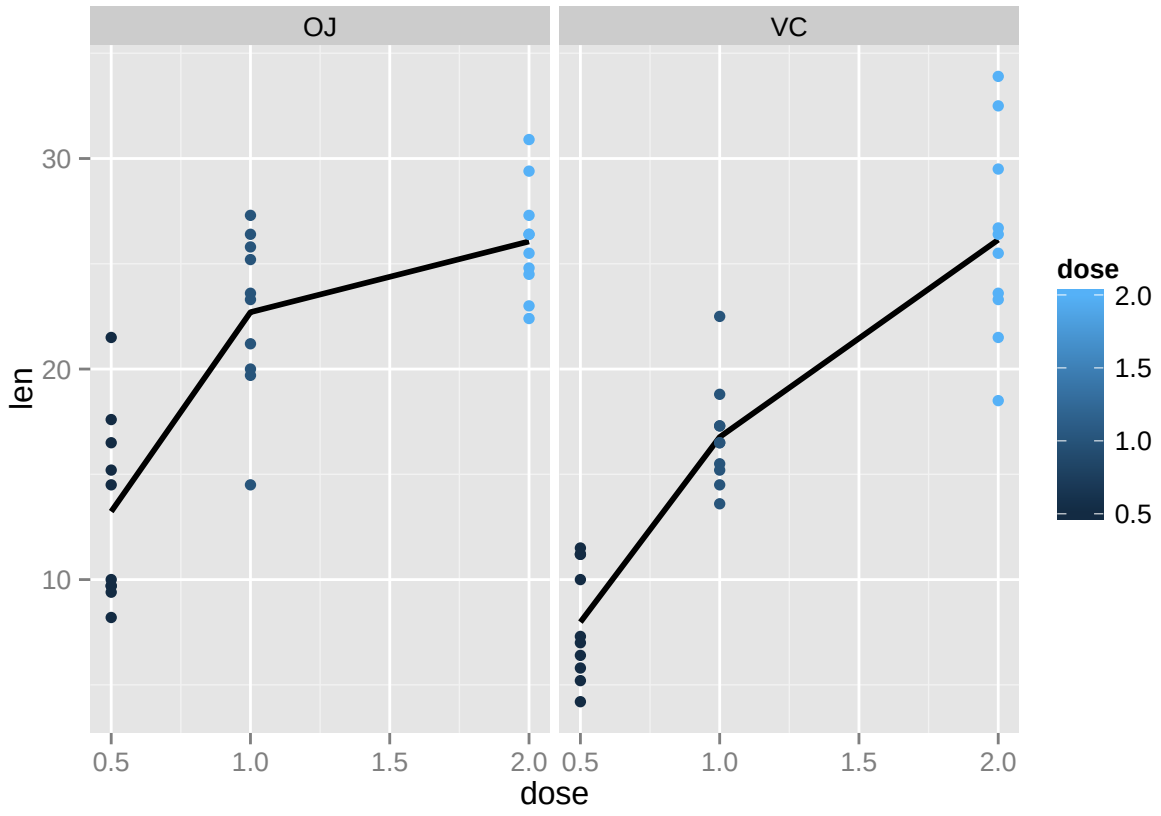
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
head(ToothGrowth)
```

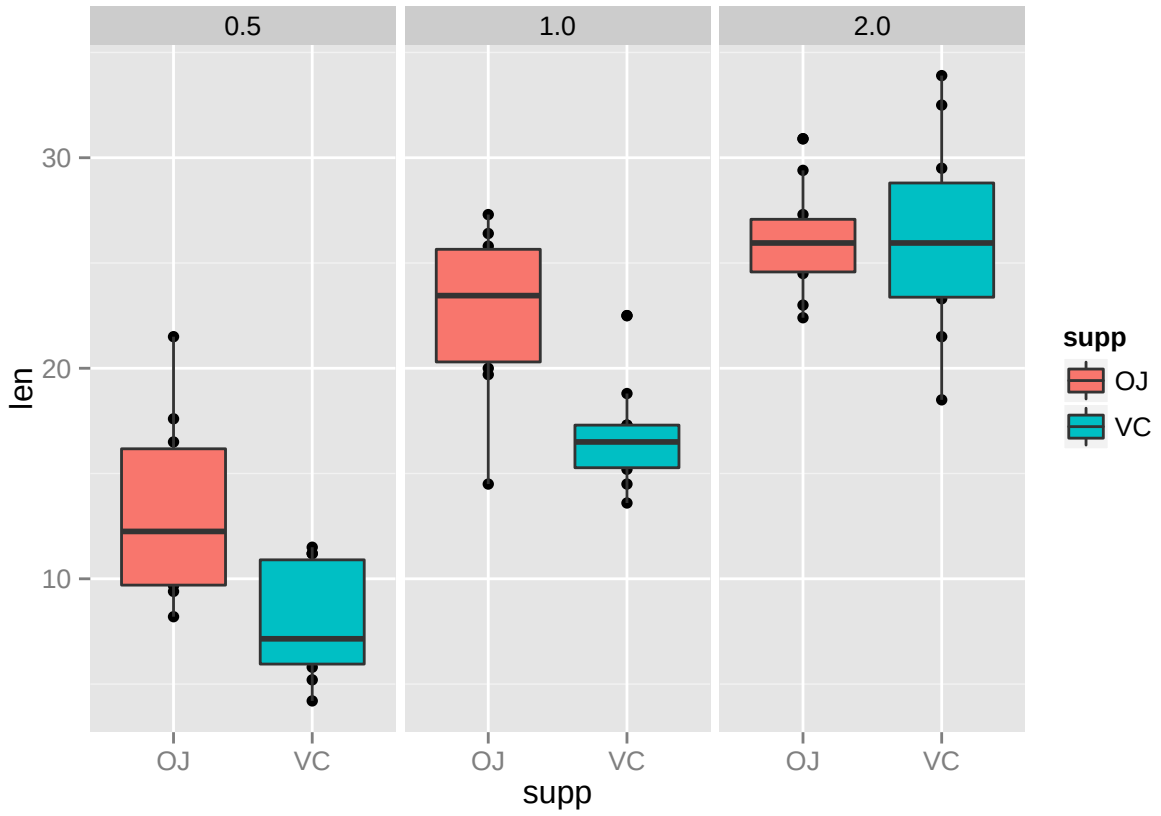
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

This data frame with 60 observations on 3 variables : len, supp and dose. ## Ploting the raw data

```
library(ggplot2)
ggplot(ToothGrowth, aes(x = dose, y = len, colour = dose)) + stat_summary(aes(group = 1),
       geom = "line", fun.y = mean, size = 1, col = "black") + facet_grid(. ~ supp) + geom_point()
```




```
ggplot(ToothGrowth, aes(x = supp, y = len)) + geom_point() + facet_wrap(~ dose) +
       geom_boxplot(aes(fill = supp))
```



Hypothesis tests

Now we are goinf to compare tooth growth by supp and dose.

Hypothesis tests :
- H0 is : there is no differnce on growth by giving a specific dose with VC or OJ.
- Ha is : there is difference on growth by giving a specific dose with VC or OJ.

So we have 3 hypothesis because we have 3 doses :
Hypothesis 1 (0.5g) : H0 : mu(OJ/0.5) == mu(VC/0.5) Vs Ha : mu(OJ/0.5) <> mu(VC/0.5)
Hypothesis 2 (1g) : H0 : mu(OJ/1.0) == mu(VC/1.0) Vs Ha : mu(OJ/1.0) <> mu(VC/1.0)
Hypothesis 3 (2g) : H0 : mu(OJ/2.0) == mu(VC/2.0) Vs Ha : mu(OJ/2.0) <> mu(VC/2.0)

Confidence intervals

There we are going to do a t test with **var.equal = FALSE** (cf. Appendix)

- Hypothesis 1

```
p1<- t.test(ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5,]$len,
            ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5,]$len,var.equal = FALSE)
```

- Hypothesis 2

```
p2 <- t.test(ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1,]$len,
             ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1,]$len,var.equal = FALSE)
```

- Hypothesis 3

```
p3 <- t.test(ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2,]$len,
             ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2,]$len,var.equal = FALSE)
```

Conclusion

In first we can notice that the vitamine C **has effect** on Tooth Growth.

Second we are going to analyse the delivery methods by dose :

- Hypothesis 1 t test value = 3.1697328
  P value = 0.0063586
  conf int = 1.7190573, 8.7809427
  ⇒ p value les than 0.05 so **we can reject H0** : the is a **significant difference** between the two delivery methods with the same dose of 0.5 mg.
- Hypothesis 2 t test value = 4.0327696
  P value = 0.0010384
  conf int = 2.8021482, 9.0578518
  ⇒ p value les than 0.05 so **we can reject H0** : the is a **significant difference** between the two delivery methods with the same dose of 1 mg.
- Hypothesis 3 t test value = -0.0461361
  P value = 0.9638516
  conf int = -3.7980705, 3.6380705
  ⇒ p value les than 0.05 so we can **not** reject H0 : this is **no** significant difference between the two delivery methods with the same dose of 2 mg.

In conclusion we ca said that the orange juice has more effect on growth with dose 0.5 and 1g dose but with 2g dose there is no significant difference.

Appendix

We done three t.test but before we had test if the variances are egals or not :

- Hypothesis 1

```
var.test(ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5,]$len,
         ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5,]$len)$p.value
```

```
## [1] 0.1648902
```

```
p1
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5,  and ToothGrowth[ToothGrowth$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 of x mean of y
##     13.23      7.98
```

p value $> 5\%$ so we are going to do a t test with **var.equal = FALSE**

- Hypothesis 2

```
var.test(ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1,]$len,
         ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1,]$len)$p.value
```

```
## [1] 0.2046214
```

```
p2
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1,  and ToothGrowth[ToothGrowth$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 of x mean of y
##     22.70     16.77
```

p value $> 5\%$ so we are going to do a t test with **var.equal = FALSE**

- Hypothesis 3

```r
var.test(ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2,]$len,
         ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2,]$len)$p.value
```

```
## [1] 0.09274336
```

```r
p3
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2,  and ToothGrowth[ToothGrowth$supp ==
## 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:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

p value $> 5\%$ so we are going to do a t test with **var.equal = FALSE**

End of the document