Peer Assessment for Coursera Statistical Inference Class

Edoardo "Dado" Marcora

2015-04-26 - Part 2

Here, we analyse the effect of vitamin C on tooth growth in guinea pigs.

First, we load the ToothGrowth dataset in R, rename it to tg, and inspect it.

```
library(datasets)
data("ToothGrowth")
tg = ToothGrowth
head(tg)
```

```
##    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
```

```
summary(tg)
```

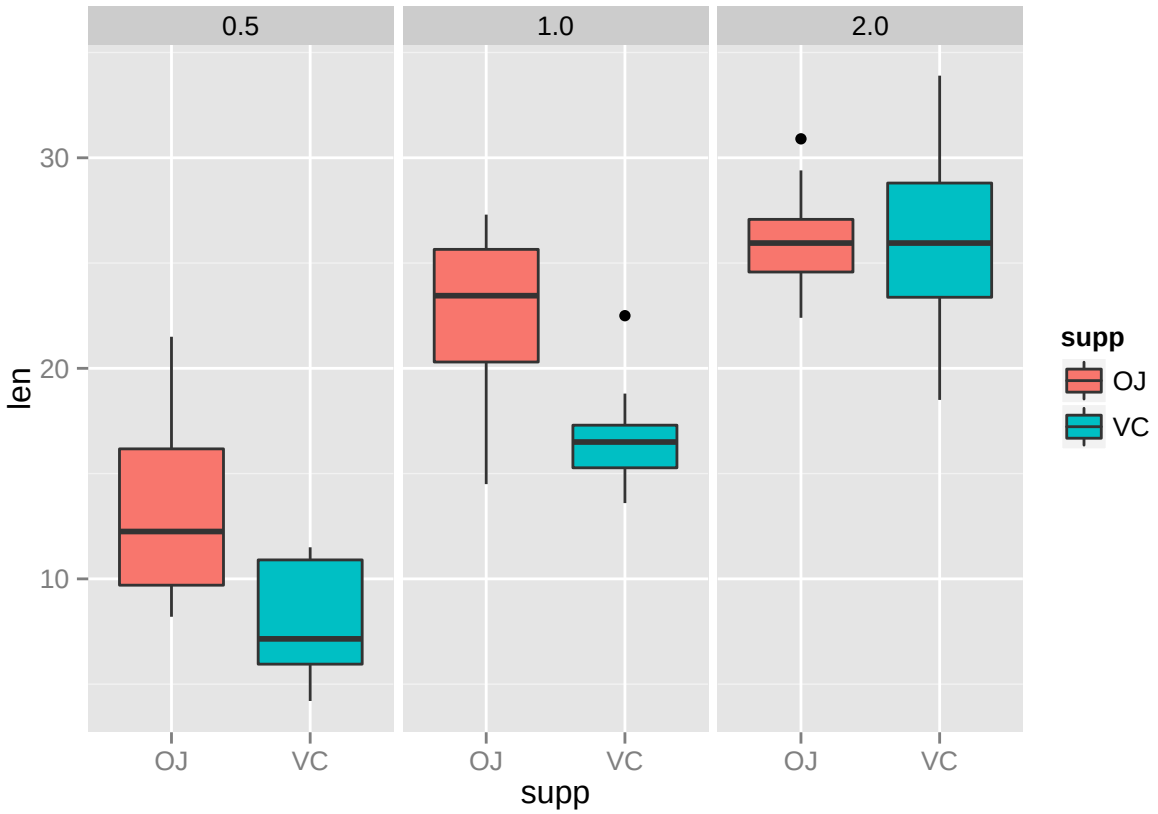
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

This dataset consists of a dataframe with 60 observations on 3 variables. The response is the tooth length (len, in mm) in each of 10 guinea pigs at each of three dose levels of vitamin C (dose, 0.5, 1, and 2 mg) with each of two delivery methods (supp, OJ = orange juice and VC = ascorbic acid, a form of vitamin C).

First, we perform some exploratory data analysis by plotting the tooth length according to delivery method and dose of vitamin C.

```
library(ggplot2)

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



It appears that, at the two lower doses of vitamin C, tooth length is shorter when delivered as ascorbic acid. At the highest dose of vitamin C, both delivery methods (ascorbic acid or orange juice) are equivalent. It also appears that there is a dose-dependent effect (the highest the dose the longest the tooth) irrespective of delivery method, but also that this effect seem to plateu at the highest dose with orange juice. But let's inspect these conclusions using t-tests and confidence intervals.

First, we check whether the groups that we are going to compare meet the assumption of equal variance because it does not look like so in the previous plot.

```
library(dplyr)
library(magrittr)

tg %>% group_by(dose, supp) %>%
  summarise(n = n(), len.mean = mean(len), len.var = var(len))
```

```
## Source: local data frame [6 x 5]
## Groups: dose
##
##   dose supp  n len.mean   len.var
## 1  0.5   OJ 10    13.23 19.889000
## 2  0.5   VC 10     7.98  7.544000
## 3  1.0   OJ 10    22.70 15.295556
## 4  1.0   VC 10    16.77  6.326778
## 5  2.0   OJ 10    26.06  7.049333
## 6  2.0   VC 10    26.14 23.018222
```

It is clear, from the summary table, that for most group comparisons the assumption of equal variance is not a valid one. Therefore, we will proceed accordingly by NOT assuming equal variance in all of the tests that follow.

First, we group the observations according to delivery method and dose of vitamin C.

```
oj.low = filter(tg, supp == 'OJ' & dose == 0.5)
vc.low = filter(tg, supp == 'VC' & dose == 0.5)
oj.mid = filter(tg, supp == 'OJ' & dose == 1)
vc.mid = filter(tg, supp == 'VC' & dose == 1)
oj.high = filter(tg, supp == 'OJ' & dose == 2)
vc.high = filter(tg, supp == 'VC' & dose == 2)
```

Last, we perform group comparisons using t-tests.

```
t.test(oj.low$len, vc.low$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.low$len and vc.low$len
## 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
```

```
t.test(oj.mid$len, vc.mid$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.mid$len and vc.mid$len
## 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
```

```
t.test(oj.high$len, vc.high$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.high$len and vc.high$len
## 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
```

Indeed, at the two lower doses of vitamin C, tooth length is significantly shorter (p-val < 0.05) when delivered as ascorbic acid. At the highest dose of vitamin C, both delivery methods (ascorbic acid or orange juice) are equivalent (i.e., we can't reject the null hypothesis with alpha $= 0.05$).

```
t.test(oj.low$len, oj.mid$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.low$len and oj.mid$len
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

```
t.test(oj.mid$len, oj.high$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj.mid$len and oj.high$len
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean of x mean of y
##     22.70     26.06
```

```
t.test(vc.low$len, vc.mid$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc.low$len and vc.mid$len
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean of x mean of y
##      7.98     16.77
```

```
t.test(vc.mid$len, vc.high$len, paired = FALSE, var.equal = FALSE)
```

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

```
## data:  vc.mid$len and vc.high$len
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean of x mean of y
##     16.77     26.14
```

Moreover, as evidenced in the t-tests above, there is a significant dose-dependent effect (the highest the dose the longest the tooth) irrespective of delivery method, but also that this effect seem to plateu, i.e., is no longer significant at the highest dose with orange juice.