Statistical Inference - Course Project Part 2

Author : Davy Meesemaecker

Overview

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data.
2. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
3. State your conclusions and the assumptions needed for your conclusions.

1 : Exploratory analysis

Let's start with the required libraries

```
library(ggplot2)
library(datasets)
library(dplyr)
```

Now we can load the data with

```
data("ToothGrowth")
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
```

Let's have a closer look at 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 ...
```

```
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
```

So the data contains 60 observations on 3 variables. Thanks to help("ToothGrowth"), we also learn that the response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal

received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

By reading, we should then have 3 levels for dose and 2 levels for supp. Let's check if it's fair and also check if we have Nas :

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
sum(is.na(ToothGrowth))
```

```
## [1] 0
```

Now we'd like to compare our data but we need to fix a problem. Our dose column class in numeric and we might convert it to a factor for our future plots :

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

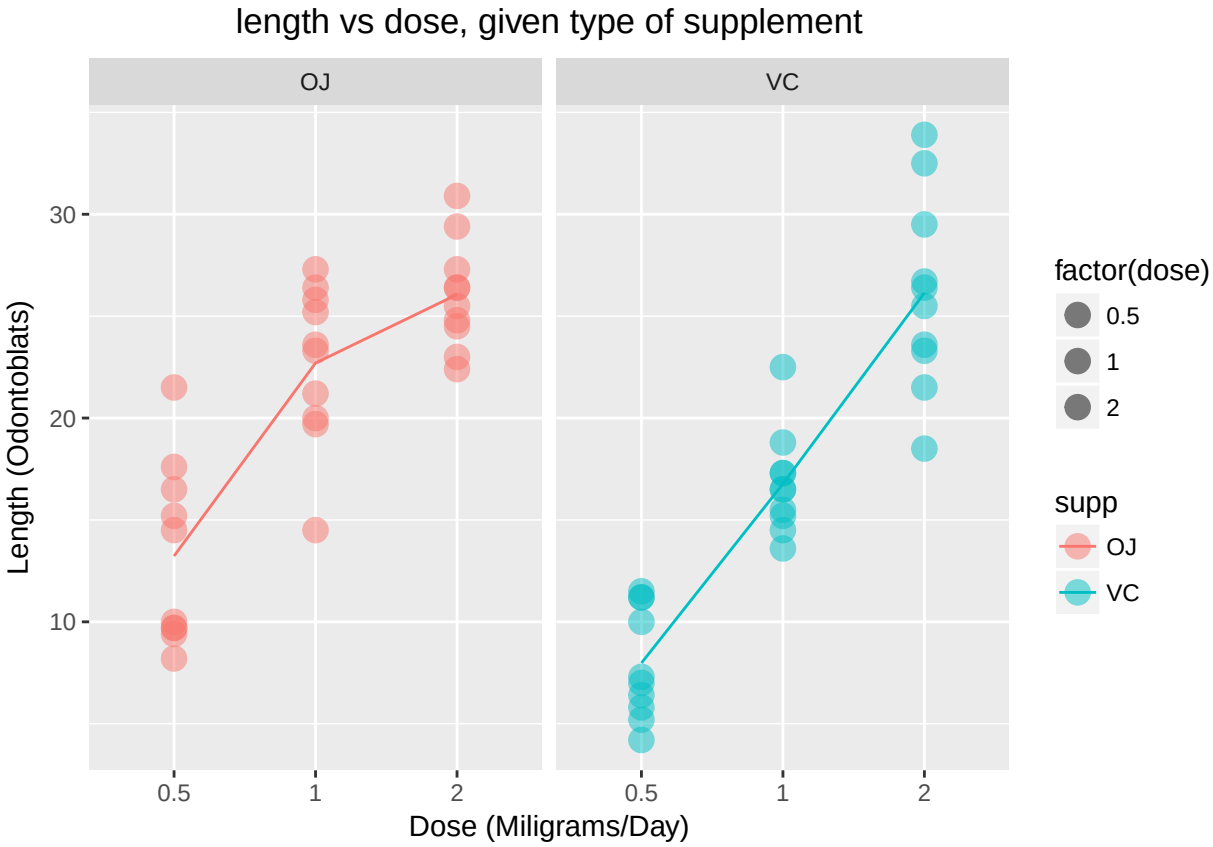
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

Our dose class in now factor as expected, we can start ploting. We'll compare tooth length versus dose, given type of supplement.

```
mean_ToothGrowth <- ToothGrowth %>%
        group_by(dose, supp) %>%
        summarize(len = mean(len))
mean_ToothGrowth
```

```
## # A tibble: 6 x 3
## # Groups:   dose [?]
##     dose   supp   len
##   <fctr> <fctr> <dbl>
## 1    0.5     OJ 13.23
## 2    0.5     VC  7.98
## 3      1     OJ 22.70
## 4      1     VC 16.77
## 5      2     OJ 26.06
## 6      2     VC 26.14
```

```
g <- ggplot(ToothGrowth, aes(x=dose, y=len))
g + geom_point(data = ToothGrowth, aes(fill = factor(dose), colour = supp), size = 4, alpha = 0.5) +
        geom_line(data = mean_ToothGrowth, aes(group = supp, colour = supp)) +
        facet_grid(~supp) +
        labs(x = "Dose (Miligrams/Day)", y = "Length (Odontoblats)", title = "length vs dose, given typ
        theme(plot.title = element_text(hjust = 0.5))
```



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

When watching the previous plot, it seems like there's no difference between effects of supplement type on the tooth growth. We'll use t-test to try to confirm our hypothesis. Knowing from the help file 60 pigs received

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

So, our p-value is 0.0606 which is greater than 0.05, meaning we can't reject the null hypothesis that there's no difference between effects of supplement type on tooth growth. How about dosage ? We can try with the lowest dosage, 0.5mg, 1mg and 2mg :

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  len by 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 in group OJ mean in group VC
##            13.23             7.98
```

P-value is here 0.006, so it's lower than 0.05, we then reject the null hypothesis and conclude that supplement type has an effect on tooth growth at a dosage of 0.5mg.

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by 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 in group OJ mean in group VC
##            22.70            16.77
```

P-value is here 0.001, so as above, we reject the null hypothesis and confirm that supplement type has an effect on tooth growth with a dosage of 1mg

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, 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 in group OJ mean in group VC
##            26.06            26.14
```

Here the p-value is 0.9639, a value much bigger than 0.05. We can't reject our null hypothesis here, meaning there's no difference between effects of supplement type on tooth growth with a 2mg dosage.

3 : Conclusion

Based on the hypothesis tests we run earlier, we can conclude that there is no difference between using orange juice or ascorbic acid as supplement on the growth of tooth. However, when we look at specific dosage levels, there appears to be differences. For dosage levels of 0.5mg and 1mg, orange juice appears to be more effective on the growth of tooth when there's no difference with a 2mg level.