Statistical Inference Course Project - part 2

Arkadiusz Oliwa

1 March 2018

Dataset

The Effect of Vitamin C on Tooth Growth in Guinea Pigs
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).

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

Let see the first part of dataset 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
```

Display the structure of dataset ToothGrowth

```
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 ...
```

Exploratory Data Analysis

Display descriptive statistics of dataset

```
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
```

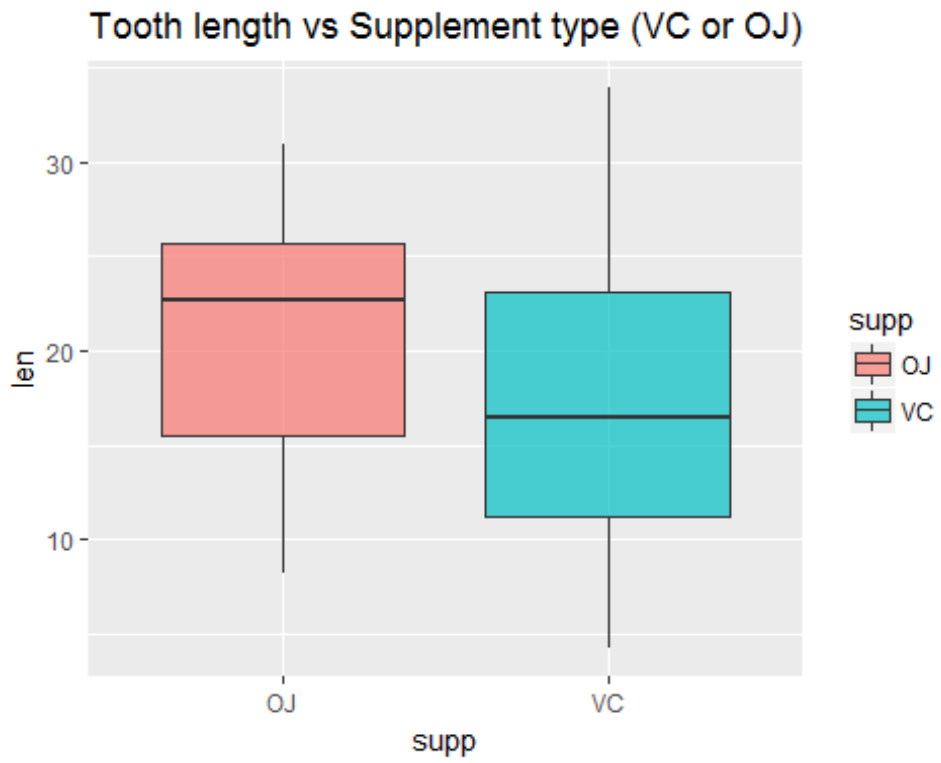
Group data by supplement type and display mean, median and standard deviation

```
ToothGrowth %>%
  group_by(supp) %>%
  summarise(lenMean = mean(len), lenMedian = median(len), lenSD = sd(len))
```

```
## # A tibble: 2 x 4
##   supp  lenMean lenMedian lenSD
##   <fct>   <dbl>     <dbl> <dbl>
## 1 OJ       20.7      22.7  6.61
## 2 VC       17.0      16.5  8.27
```

Plot Tooth length vs Supplement type (VC or OJ)

```
ggplot(data = ToothGrowth, aes(x = supp, y = len)) +
  geom_boxplot(aes(fill=supp), alpha=.7) +
  labs(title="Tooth length vs Supplement type (VC or OJ)")
```



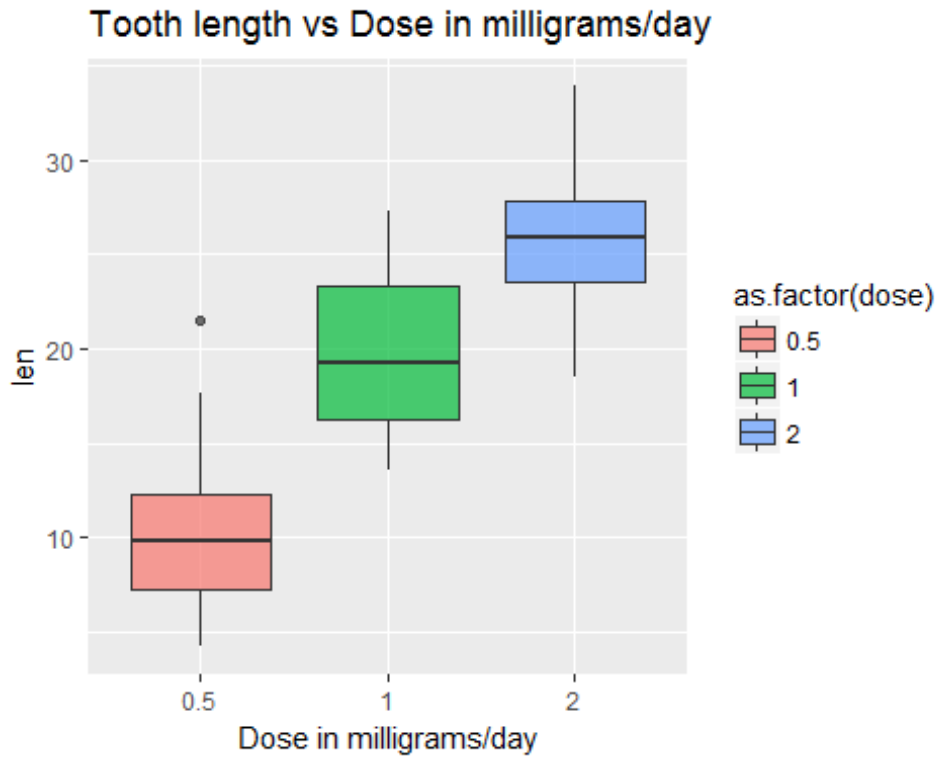
Group data by Dose in milligrams/day and display mean, median and standard deviation

```
ToothGrowth %>%
  group_by(as.factor(dose)) %>%
  summarise(lenMean = mean(len), lenMedian = median(len), lenSD = sd(len))
```

```
## # A tibble: 3 x 4
##   `as.factor(dose)` lenMean lenMedian lenSD
##   <fct>                <dbl>     <dbl> <dbl>
## 1 0.5                   10.6      9.85  4.50
## 2 1                     19.7     19.2   4.42
## 3 2                     26.1     26.0   3.77
```

Plot Tooth length vs Dose in milligrams/day

```
ggplot(data = ToothGrowth, aes(x = as.factor(dose), y = len)) +
  geom_boxplot(aes(fill=as.factor(dose)), alpha=.7) +
  labs(title="Tooth length vs Dose in milligrams/day", x="Dose in
milligrams/day")
```



Statistical Inference based on dataset

```
t.test(formula = len ~ supp, data = ToothGrowth, paired=FALSE, var.equal=FALSE)
```

```
##
##  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
```

As p-value is larger than the significance level of 0.05, null hypothesis can't be rejected and hence we infer that Orange Juice (OJ) and Ascorbic acid (VC) have the same effect on tooth growth.

```
t.test(formula = len ~ dose,
       data = ToothGrowth[which(ToothGrowth$dose == .5 | ToothGrowth$dose == 1
       ), ], paired=FALSE, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

As p-value is smaller than the significance level of 0.05, null hypothesis can be rejected and hence we infer that dose 0.5 and 1 milligrams/day do not have the same effect on tooth growth

```r
t.test(formula = len ~ dose,
       data = ToothGrowth[which(ToothGrowth$dose == 1 | ToothGrowth$dose == 2
       ), ], paired=FALSE, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

As p-value is smaller than the significance level of 0.05, null hypothesis can be rejected and hence we infer that dose 1 and 2 milligrams/day do not have the same effect on tooth growth