

course6Part2

Rohit Jain

9/13/2018

Part2: Basic Inferential Data Analysis Instructions:

Analysis of the ToothGrowth data in the R datasets package

1. Load the ToothGrowth data and perform some basic exploratory data analyses And 2. Provide a basic summary of the data.

Summary Of Data:

The data 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). It is a data frame with 60 observations on 3 variables.

- [,1] len numeric Tooth length
- [,2] supp factor Supplement type (VC or OJ).
- [,3] dose numeric Dose in milligrams/day

```
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

Exploratory Data Analyses:

- Figure1 in the Appendix distribution depict the number of guinea pigs for specific range of tooth length and as per the distribution the maximum number guinea pigs are in the (15, 20) and (25,30) range of tooth length.
- Figure2 in the Appendix is the distribution for the toothlength on the basis of delivery methods that is orange juice or ascorbic acid (a form of vitamin C and coded as VC). We have plotted a mean tooth length with a black and the inference from the distribution is “Number of guinea pigs having tooth length greater than the mean tooth length is for the OJ delivery method.”
- Figure3 in the Appendix is distribution for the tooth length on the basis of dose levels of vitamin C (0.5, 1, and 2 mg/day). We have plotted a mean tooth length with a black line and the inference from the distribution is “Number guinea pigs having tooth length greater than the mean tooth length is in the order $2 > 1 > 0.5$.”

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

```
t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose == 0.5,])
```

```
##
## 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
```

The above t-test for the lowest dose level that is “0.5” shows that the tooth growth is significantly large for OJ than the VC delivery method.

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

The above t-test for the dose level “1” shows that the tooth growth is significantly large for OJ then the VC delivery method.

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

The above t-test for the dose largest level that is “2” shows that there is no influence of the delivery method OJ or VC on tooth growth.

4. State your conclusions and the assumptions needed for your conclusions

Conclusion:

- Tooth growth is faster in OJ than VC delivery method for the 0.5 and 1 dose level.

- Delivery method doesn't have any impact on the tooth growth for the 2 dose level.

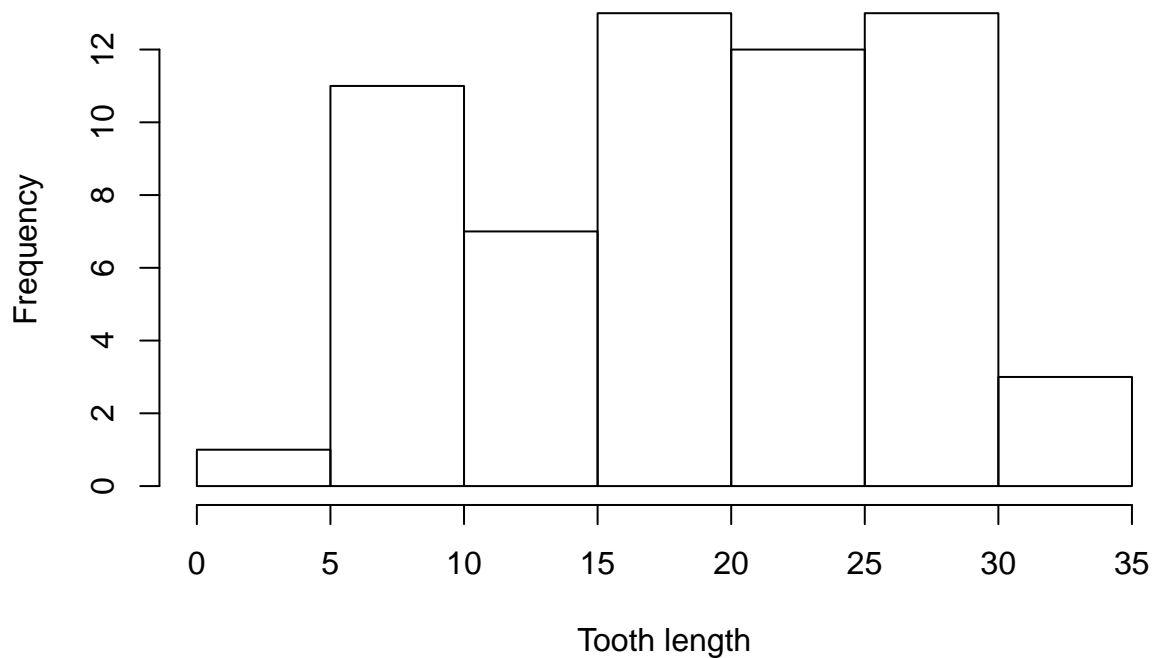
Assumptions:

- The sample data is not paired.
- The variances between the sample data are not equal.

Appendix:

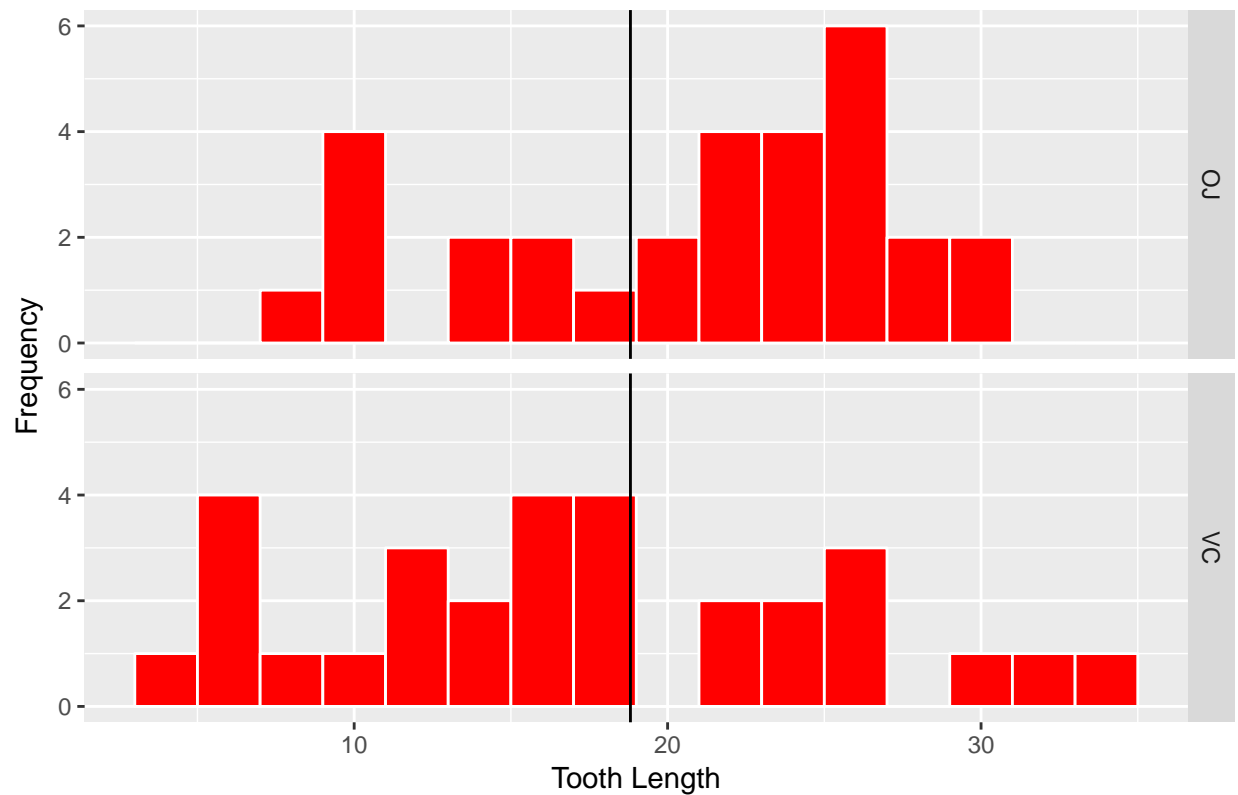
```
hist(ToothGrowth$len,
     main = "Figure1. Frequency distribution of Toothlength",
     xlab = "Tooth length",
     ylab = "Frequency")
```

Figure1. Frequency distribution of Toothlength



```
library(ggplot2)
g <- ggplot(data = ToothGrowth, aes(x=len)) +
  facet_grid(supp ~ .) +
  geom_histogram(binwidth=2, colour="white", fill = "red") +
  geom_vline(xintercept = mean(ToothGrowth$len)) +
  labs(title = "Figure2. Frequency distribution of Toothlength on the basis of delivery method") +
  xlab("Tooth Length") +
  ylab("Frequency")
print(g)
```

Figure2. Frequency distribution of Toothlength on the basis of delivery method



```
g1 <- ggplot(data = ToothGrowth, aes(x=len)) +
  facet_grid(dose ~ .) +
  geom_histogram(binwidth=2, colour="white", fill = "blue")+
  geom_vline(xintercept = mean(ToothGrowth$len))+
  labs(title = "Figure3. Frequency distribution of Toothlength on the basis of dose level") +
  xlab("Tooth Length") +
  ylab("Frequency")
print(g1)
```

Figure3. Frequency distribution of Toothlength on the basis of dose level

