

# Effect of Vitamin C on tooth growth of guinea pigs.

**1. Exploratory data analysis.** The data for the exercise is the R dataset ToothGrowth on the effect of Vitamin C on tooth growth of guinea pigs. As it is possible to read from the documentation (command ?ToothGrowth from R), the data set contains 60 observations about 10 guinea pigs. The tooth length has been measured for each of 10 guinea pigs at each of three dose levels of Vitamin C with each of two delivery methods. Each observation has 3 variables described in the table below:

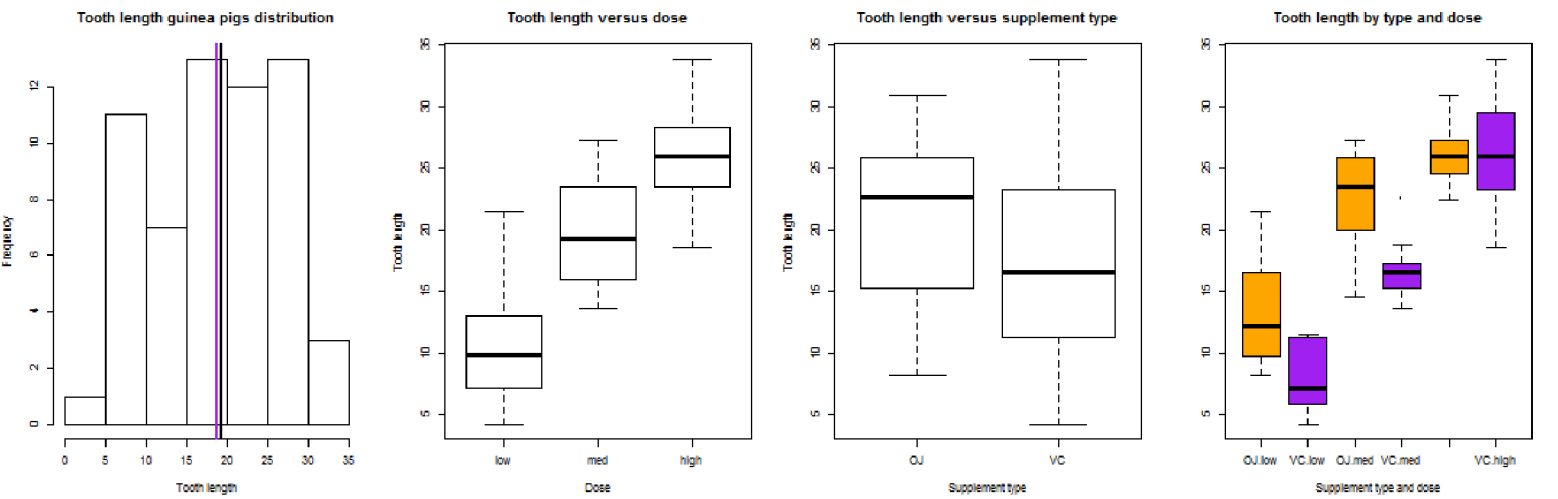
##	Name	Description	Type	values
##	:	:	:	:
##	len	Tooth length of guinea pigs	numeric	
##	supp	Supplement type	factor	VC(ascorbic acid), OJ(orange juice)
##	dose	Dose level in milligrams	numeric	0.5, 1, and 2 mg.

The variable dose is numeric but it can assume only 3 values so we transform it in a factor variable for convenience.

```
ToothGrowth$dose = factor(ToothGrowth$dose, labels = c("low", "med", "high"))
```

Below there are four plots. The first plot is the histogram of the distribution of the tooth length (the mean is the purple line, the median the black one). The second plot is a side-by-side boxplot that shows the relation between tooth length and dose and it suggests a positive association: as the dose of Vitamin C increases, the tooth length does the same. In addition, it is possible to see from the third plot that the tooth growth is greater if Vitamin C is given by orange juice than by ascorbic acid. In the last box plot, it is possible to compare the tooth length distribution by dose and supplement type: the difference between the two delivery methods is more evident at low and medium dosage, if the dosage is high the type of supplement seems to be not important.

```
par(mfrow = c(1, 4))
hist(ToothGrowth$len, xlab = "Tooth length", main = "Tooth length guinea pigs distribution")
abline(v = mean(ToothGrowth$len), col = "purple", lwd = 2)
abline(v = median(ToothGrowth$len), col = "black", lwd = 2)
boxplot(len ~ dose, ToothGrowth, main = "Tooth length versus dose", xlab = "Dose", ylab = "Tooth length")
boxplot(len ~ supp, ToothGrowth, main = "Tooth length versus supplement type", xlab = "Supplement type", ylab = "Tooth length")
boxplot(len ~ supp * dose, ToothGrowth, col = c("orange", "purple"), main = "Tooth length by type and dose", xlab = "Supplement type and dose")
```



**2. Data summaries.** The summaries below show the mean of tooth length grouped by dose, supplement and both variables. The values of the means quantify the previous considerations about variables associations.

```
aggregate(list(Mean = ToothGrowth$len), list(Dose = ToothGrowth$dose), FUN = mean)
```

```
##      Dose  Mean
## 1   low 10.61
## 2   med 19.73
## 3   high 26.10
```

```
aggregate(list(Mean = ToothGrowth$len), list(Supplement = ToothGrowth$supp), FUN = mean)
```

```
##      Supplement  Mean
## 1             OJ 20.66
## 2             VC 16.96
```

```
aggregate(list(Mean = ToothGrowth$len), list(Dose = ToothGrowth$dose, Supplement = ToothGrowth$supp), FUN = mean)
```

```
##      Dose Supplement      Mean
## 1 low             OJ 13.23
## 2 med             OJ 22.70
## 3 high            OJ 26.06
## 4 low             VC  7.98
## 5 med             VC 16.77
## 6 high            VC 26.14
```

**3. Comparisons and t-tests.** A t-test is conducted to verify if there is a difference between the tooth growth average between the two supplements for the same dosage. The data are not paired and we assume different variance.

```
VC <- subset(ToothGrowth, supp == "VC" & dose == "low")
OJ <- subset(ToothGrowth, supp == "OJ" & dose == "low")
t.test(OJ$len, VC$len, alternative = "greater")
```

```
##
##  welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 3.17, df = 14.97, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.346      Inf
## sample estimates:
## mean of x mean of y
##    13.23      7.98
```

```
VC <- subset(ToothGrowth, supp == "VC" & dose == "med")
OJ <- subset(ToothGrowth, supp == "OJ" & dose == "med")
t.test(OJ$len, VC$len, alternative = "greater")
```

```
##
##  welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 4.033, df = 15.36, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356      Inf
## sample estimates:
## mean of x mean of y
##    22.70    16.77
```

```
VC <- subset(ToothGrowth, supp == "VC" & dose == "high")
OJ <- subset(ToothGrowth, supp == "OJ" & dose == "high")
t.test(OJ$len, VC$len)
```

```
##
##  welch Two Sample t-test
##
## data:  OJ$len and VC$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:
## -3.798  3.638
## sample estimates:
## mean of x mean of y
##    26.06    26.14
```

In the first two test, the p-value is lower than 0.05, so we reject the null hypothesis and it is possible to conclude that the mean of the orange juice group is greater than the mean of the acid ascorbic group. In the third test, the p-value is greater than 0.05 so we cannot reject the null hypothesis and it is not possible to conclude that there is a difference between the two means.

**4. Conclusions.** The result of the t-tests allows the conclusion that the tooth growth is greater if Vitamin C is given by orange juice than by ascorbic acid but only if the dosage is low or medium.