

Exercise 1 (Univariate descriptives)

In this exercise we consider the same dataset as last week. It contains a survey of school children and it's stored in CSV format. You can download it from:

<https://polybox.ethz.ch/index.php/s/bA45tIopHB7dBRX>

- (a) Read in the data. Use the URL or download the data and read it from the local directory. The direct access link to the file is

<https://polybox.ethz.ch/index.php/s/bA45tIopHB7dBRX/download>.

- (b) To gain an overview over the data calculate some characteristic measures of the distribution:

- Determine the mean and the median of `Arm.span` (**R-Hint:** `mean()`, `median()`).
- Determine the 10% quantile of `Arm.span` (**R-Hint:** `quantile()`).
- Calculate the range, variance, standard deviation and interquartile range of `Arm.span` (**R-Hint:** `range()`, `var()`, `sd()`, `IQR()`).

- (c) Now we would like to visualize the distribution of the variable `Arm.span`:

- Visualize the data as a boxplot and add the notches (**R-Hint:** `boxplot(..., notch=TRUE)`).
- Visualize the difference between students who take part in a sports team (`Sports.team`) and those who don't (**R-Hint:** `boxplot(Arm.span ~ ...)`).
- Visualize the four variables `Arm.span`, `Height`, `Age`, `Hand.span` in one boxplot. Does this visualization make sense?
(**R-Hint:** `boxplot(dat[,c("Arm.span", "Height", "Age", "Hand.span")])`)

- (d) Now, we want to compare two variables to each other:

- Determine the contingency table between `Eye.color` and `Hair.color` (**R-Hint:** `table()` or `xtabs(...)`).
- Display the frequencies of the contingency table as mosaic plot (**R-Hint:** `mosaicplot()`). What do you observe?
- Visualize the relation between `Arm.span` and `Height` in a scatterplot (**R-Hint:** `plot()`). What do you observe?

Exercise 2 (Descriptives)

The dataset of this exercise is from a study on guinea pigs. The study investigates the effects of Vitamin C consumption on the length of the teeth growth. Therefore, the guinea pigs were fed by orange juice (OJ) or ascorbic acid (VC) using different doses of Vitamin C (0.5, 1.0, 2.0). The data contains the following variables:

len mean of teeth length
supp supplement type (OJ or VC)
dose vitamin C dose in mg

In order to access the data, you can use the following code:

```
# The data is contained in the R package datasets
# with data() the data is loaded into the workspace
data("ToothGrowth")

# then we can rename the dataset and store it in dat
# (easier for coding purposes)
dat <- ToothGrowth

# Consider the data
head(dat)

##      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
```

- (a) How many guinea pigs have been included into this study?
- (b) Investigate the three variables of the dataset by appropriate plots (**R-Hint:** `hist()`, `barplot()`).
- (c) Does the distribution of the tooth length depend on the supplement type? Illustrate your answer with an appropriate plot (**R-Hint:** `boxplot()`).
- (d) Does the distribution of the tooth length depend on the Vitamin C dose? Illustrate your answer with an appropriate plot (**R-Hint:** `boxplot()`).
- (e) Find out if the influence of the Vitamin C dose is different for the two delivery types. (**R-Hint:** Take subsets of the data using e.g. `dat_oj <- subset(dat, supp=="OJ")`, `dat_vc <- subset(dat, supp=="VC")` and do boxplots for the two subsets.)