

Introduction

- Descriptive statistics (in the broad sense of the term) is a branch of statistics aiming at summarizing, describing and presenting a series of values or a dataset.
- Descriptive statistics is often the first step and an important part in any statistical analysis.
- It allows to check the quality of the data and it helps to “understand” the data by having a clear overview of it.
- If well presented, descriptive statistics is already a good starting point for further analyses.

Types of Descriptive Summary

There exists many measures to summarize a dataset. They are divided into two types:

- location measures and
- dispersion measures

Working with Toy Dataset

As a first step load the data set to R:

```
{r} dat <- iris # load the iris dataset and renamed it dat
{r} head(dat) # first 6 observations
```

Structure of a dataset

```
“{r} str(dat) # structure of dataset
```

```
## Basic summary statistics
```

min, max, mean, median, range, IQR, quantiles

```
```{r}
print("Minimum")
min(dat$Sepal.Length)
median(dat$Sepal.Length)
quantile(dat$Sepal.Length, c(0.25,0.5,0.75)) # three quartile
```

## Standard deviation and variance

The standard deviation and the variance is computed with the `sd()` and `var()` functions:

```
{r} sd(dat$Sepal.Length) var(dat$Sepal.Length) sqrt(var(dat$Sepal.Length))
```

**Tip:** to compute the standard deviation (or variance) of multiple variables at the same time, use `lapply()` with the appropriate statistics as second argument:

```
{r} lapply(dat[, 1:4], sd)
```

## Five point Summary

```
{r} summary(dat)
```

**Group-wise summary**

```
“{r} by(dat, dat$Species, summary)
```

```
Coefficient of variation
```

The coefficient of variation can be found by computing manually (remember that the coefficient of variation is the standard deviation divided by the mean)

```
sd(dat$Sepal.Length) / mean(dat$Sepal.Length)
```

## Mode

```
{r} tab <- table(dat$Sepal.Length) # number of occurrences for each unique value
sort(tab, decreasing = TRUE) # sort highest to lowest
```

## Takeaway

- In R programming, basic descriptive statistic functions are simple and exactly same as in statistical definitions.