

# Measures of Spread

Ľubica and Ján Krauskovi, Dominik Heger

Masaryk University

*hegerd@chemi.muni.cz*

STDT 04 Measures of Spread

# Measures of Spread

How can we quantify your distance from the median and/or mean?

# Measures of Spread

How can we quantify your distance from the median and/or mean?

How far from average am I?

# Measures of Spread

How can we quantify your distance from the median and/or mean?

How far from average am I?

How much am I deviating?

# Measures of Spread

How can we quantify your distance from the median and/or mean?

How far from average am I?

How much am I deviating?

The amount your score is off (from average) is the **deviation**.

# Range and interquartile range

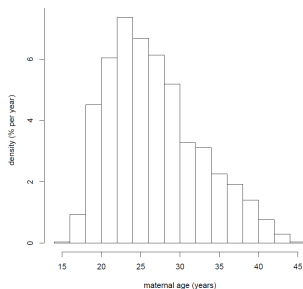
# Range and interquartile range

How far from median?

# Range and interquartile range

How far from median?

Maternal ages

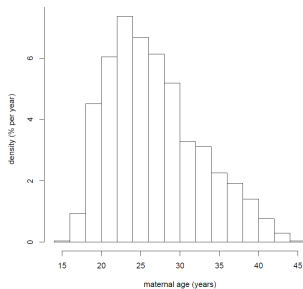




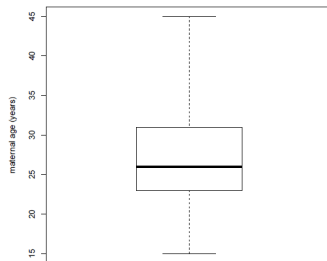
# Range and interquartile range

How far from median?

Maternal ages



Box plot



# Deviation from average

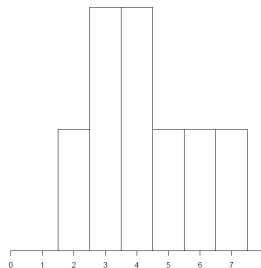
What is the typical (standard) deviation from average?

# Deviation from average

What is the typical (standard) deviation from average?

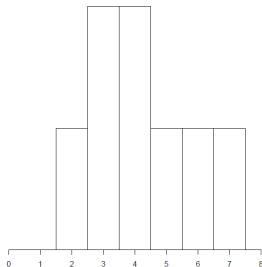
# Deviation from average

What is the typical (standard) deviation from average?



# Deviation from average

What is the typical (standard) deviation from average?



**SD = standard deviation** = směrodatná odchylka  
= (square) root mean square of deviations from average  
= (druhá) odmocnina průměru čtverců odchylek od průměru

**variance** = rozptyl  
= mean square of deviations from average  
= průměr čtverců odchylek od průměru

# Properties of SD

Why SD is so commonly used measure of spread?

SD for given distribution measures typical distance from average.

- 1 It is non negative
- 2 It has the same units as average and the list.
- 3 It measures the average distance from the data to their mean (rms of the deviations of the data from their mean)
- 4 Chebychev inequality

Pafnuty Lvovich Chebychev (1821 - 1894)

In any list, the proportion of entries that are  $k$  or more SDs away from the average is **at most**  $1/k^2$ .

Pro jakoukoli číselnou řadu platí: podíl členů, které jsou od průměru vzdáleny **alespoň**  $k$ -krát SD je **nejvíce**  $1/k^2$ .

[https://courses.edx.org/courses/BerkeleyX/Stat\\_2.1x/](https://courses.edx.org/courses/BerkeleyX/Stat_2.1x/)

<http://www.stat.berkeley.edu/~stark/SticiGui/Text/location.htm>