Míry polohy a variability

Dominik Heger

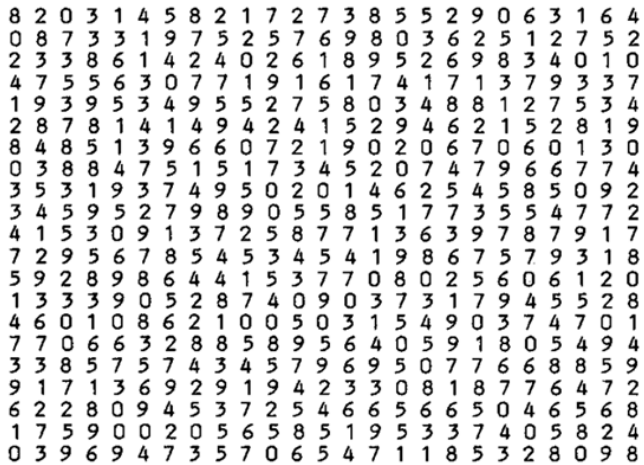
Masaryk University

hegerd@chemi.muni.cz

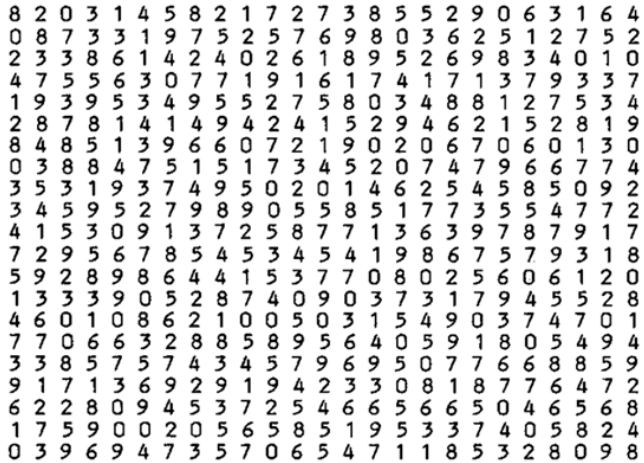
STDT 03 Míry polohy a variability

K čemu sumarizovat data? Není nejlepší mít všechna data?

K čemu sumarizovat data? Není nejlepší mít všechna data?



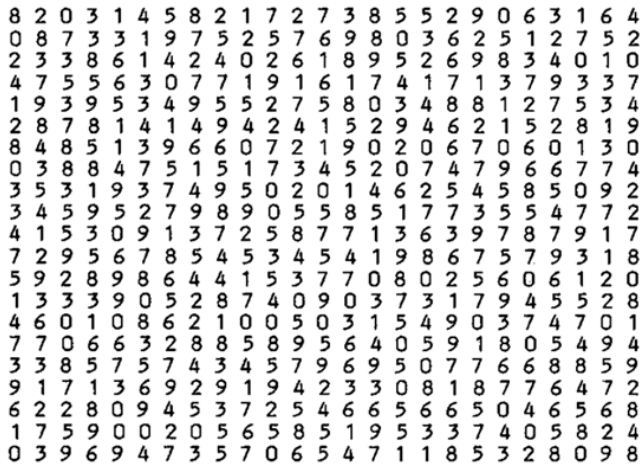
K čemu sumarizovat data? Není nejlepší mít všechna data?



Grafický popis

Numerický popis

K čemu sumarizovat data? Není nejlepší mít všechna data?

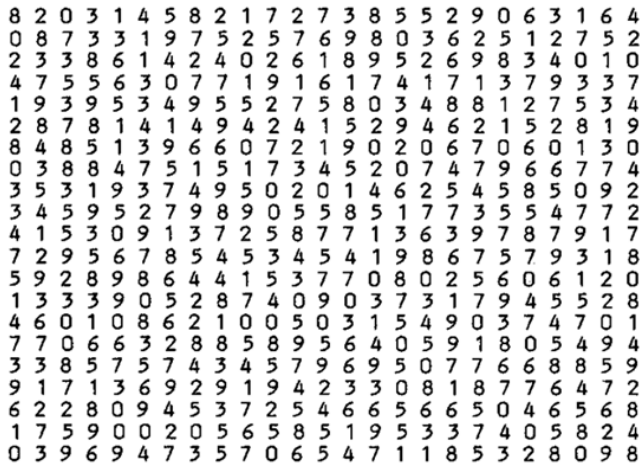


Grafický popis

Numerický popis

1. - **jedné proměnné** (stupeň vzdělání, příjem, oblíbená barva)

K čemu sumarizovat data? Není nejlepší mít všechna data?

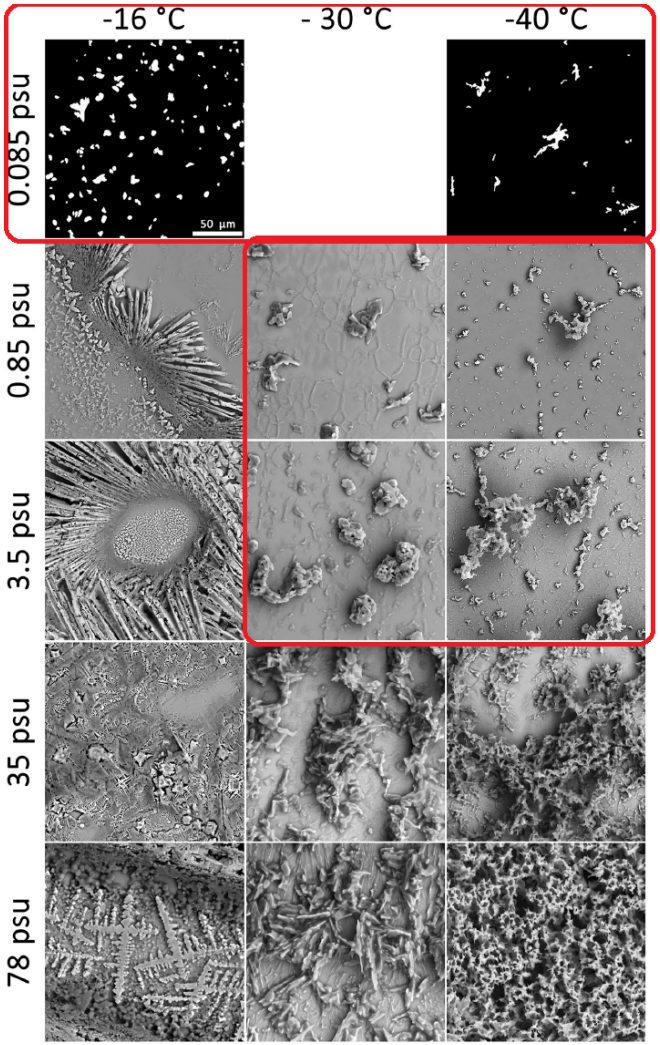


Grafický popis

Numerický popis

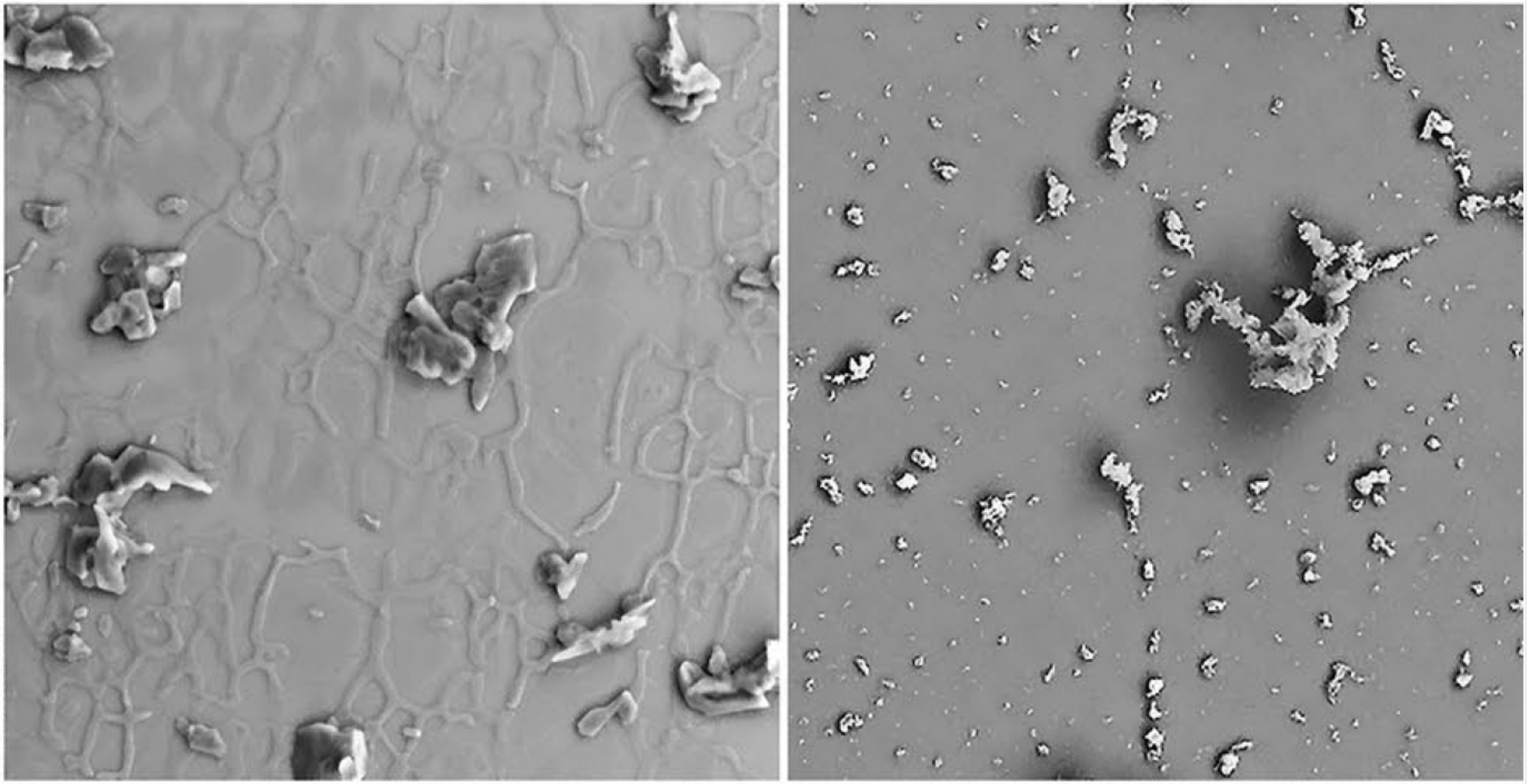
1. - **jedné proměnné** (stupeň vzdělání, příjem, oblíbená barva)
2. - **vzath mezi dvěmi proměnnými** (Jak stupeň vzdělání ovlivní příjem?)

Relationship between the Particle Sizes, Sea Salt Concentration, and Sublimation Temperature



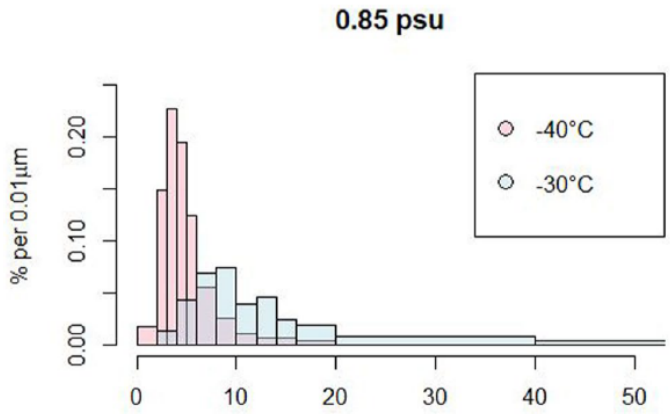
ZÁVACKÁ, K., V. NEDĚLA, M. OLBERT, E. TIHLAŘÍKOVÁ, Ľ. VETRÁKOVÁ, X. YANG AND D. HEGER Temperature and Concentration Affect Particle Size Upon Sublimation of Saline Ice: Implications for Sea Salt Aerosol Production in Polar Regions. Geophysical Research Letters, 2022, 49(8),10.1029/2021gl097098.

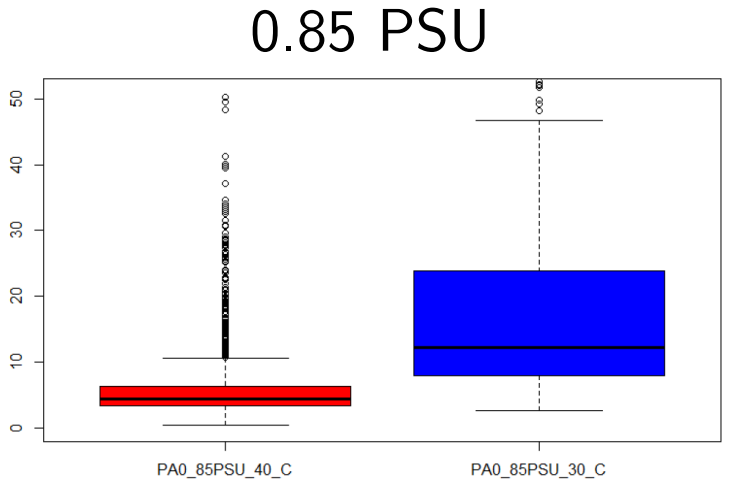
At -30 °C, Brine is Still Fluid to Become Mostly Solid at -40 °C (0.85 PSU) ESEM pictures

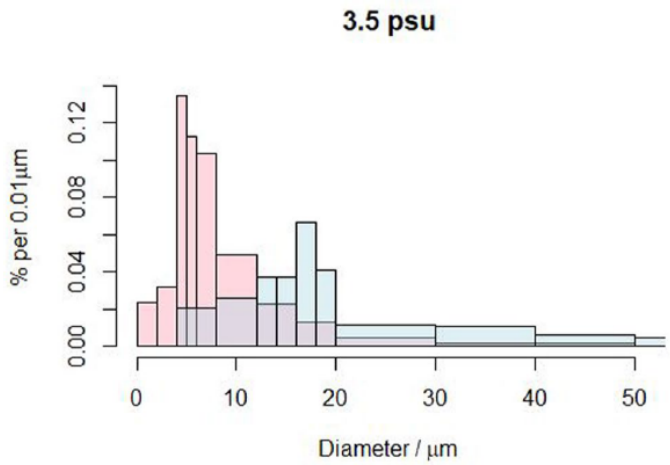


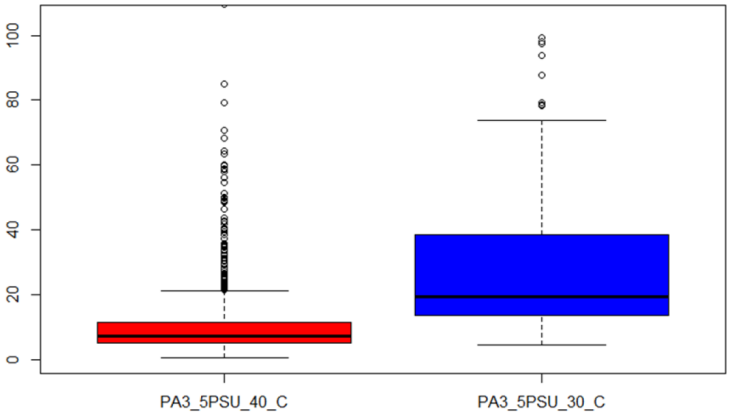
ZÁVACKÁ, K., V. NEDĚLA, M. OLBERT, E. TIHLAŘÍKOVÁ, Ľ. VETRÁKOVÁ, X. YANG AND D. HEGER Temperature and Concentration Affect Particle Size Upon Sublimation of Saline Ice: Implications for Sea Salt Aerosol Production in Polar Regions. Geophysical Research Letters, 2022, 49(8),10.1029/2021gl097098.

SSAs: the Size Decreases with the Temperature and Concentration: Histogram and Box-Plot



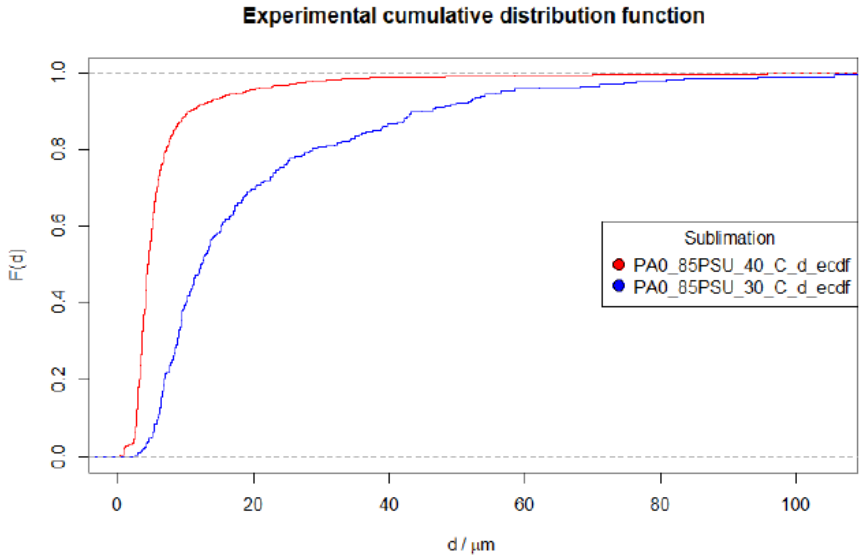


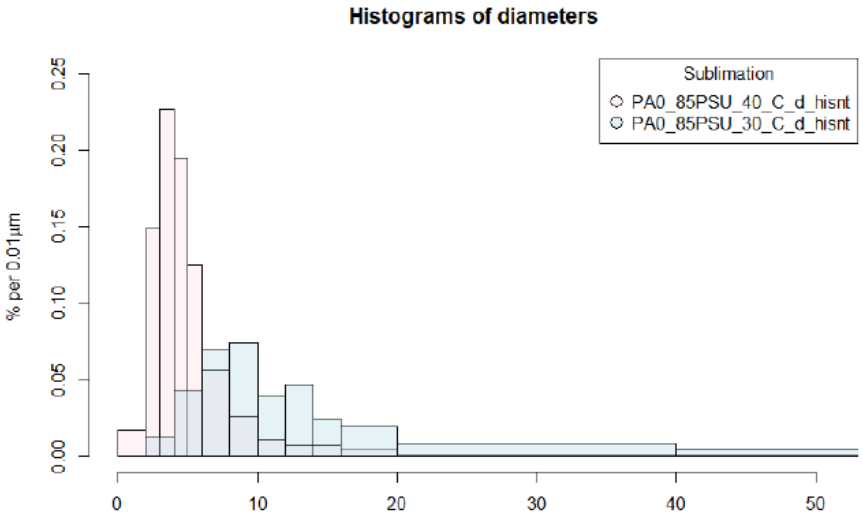




ZÁVACKÁ, K., V. NEDĚLA, M. OLBERT, E. TIHLAŘÍKOVÁ, Ľ. VETRÁKOVÁ, X. YANG AND D. HEGER Temperature and Concentration Affect Particle Size Upon Sublimation of Saline Ice: Implications for Sea Salt Aerosol Production in Polar Regions. Geophysical Research Letters, 2022, 49(8) 10.1029/2021gl097098

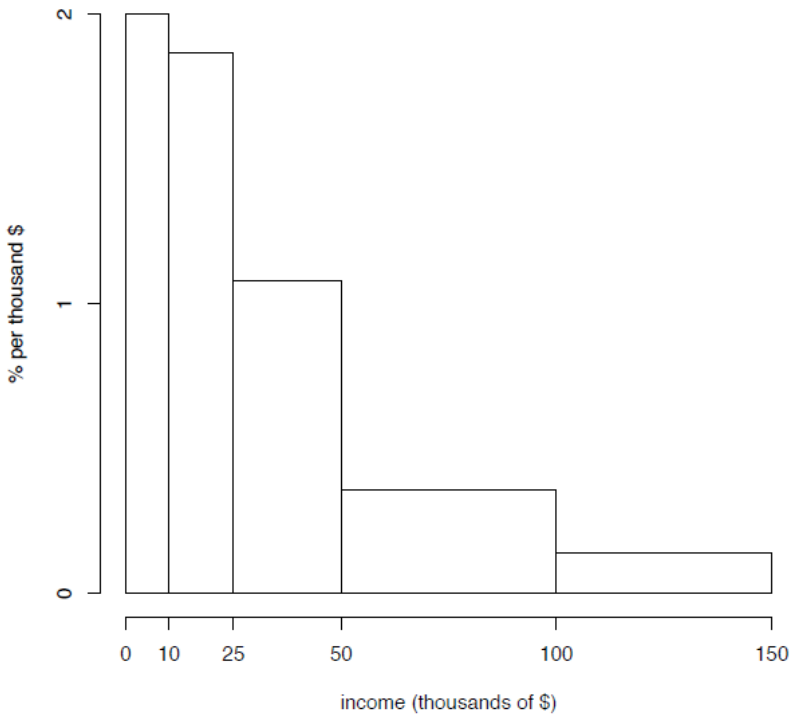
Empirická distribuční funkce





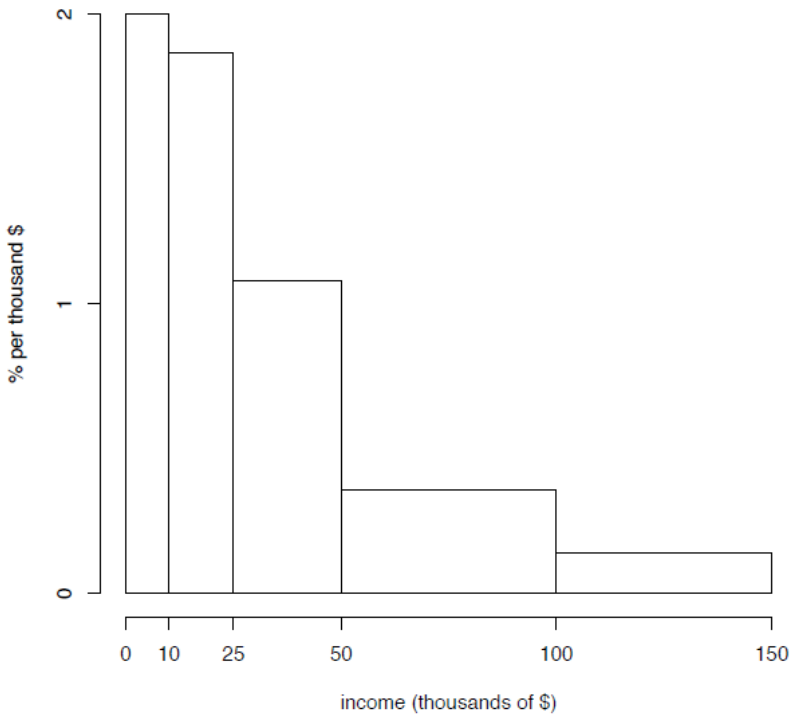
A right-skewed distribution = do prava zšikmělé rozložení

Income in USA 2010



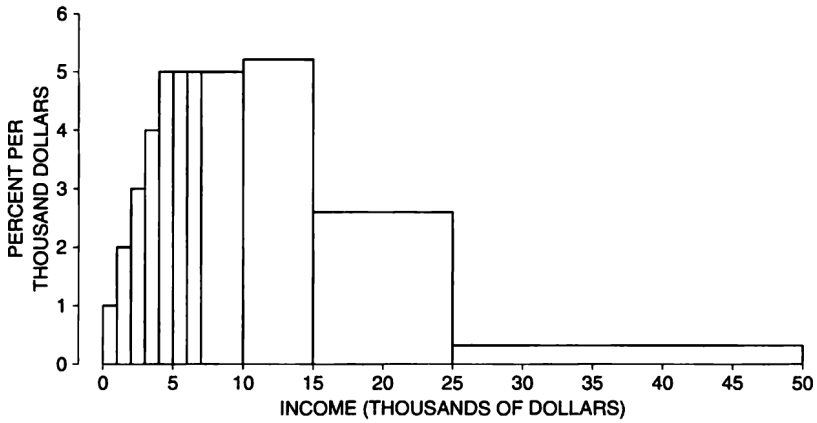
A right-skewed distribution = do prava zšikmělé rozložení

Income in USA 2010



Income in USA 1973

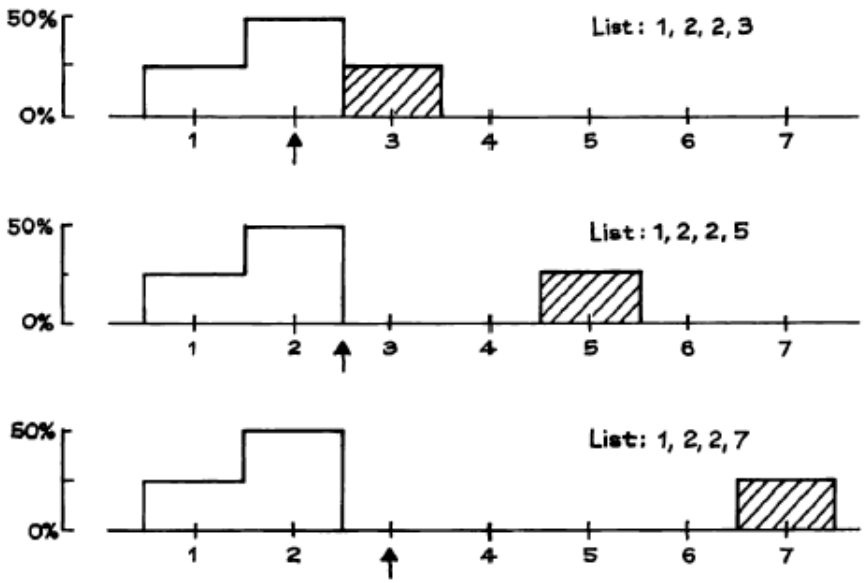
Figure 4. Distribution of families by income in the U.S. in 1973.

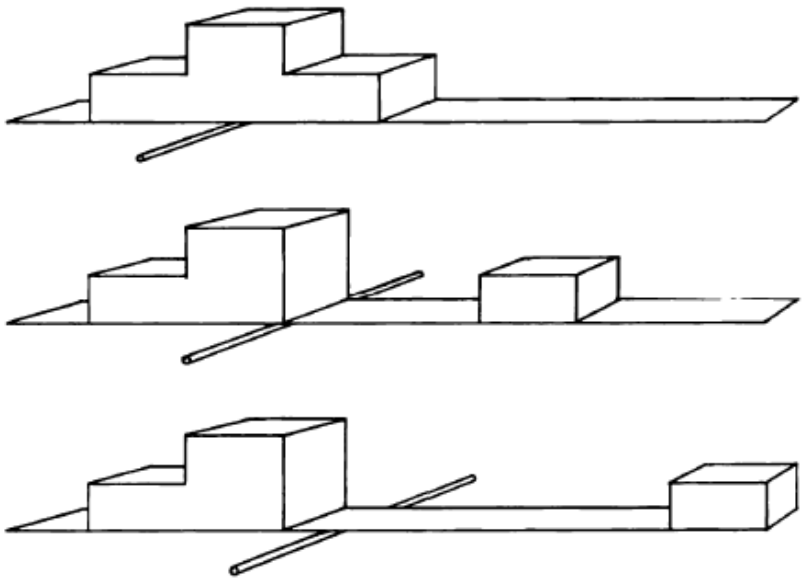


V histogramu

1. plocha pravoúhelníků představuje pravděpodobnosti
2. výška pravoúhelníků představuje hustotu - procenta celkové populace na jednotku horizontální osy.

Average


50%
0%
List: 1, 2, 2, 3
1
3
4
5
6
7
50%
0%
List: 1, 2, 2, 5
1
2
3
4
5
6
7
50%
0%
List: 1, 2, 2, 7
1
2
4
5
6
7



Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.
[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.
[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?
Data - **longitudinal** $\times$ **cross-sectional**

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.
[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Data - **longitudinal** $\times$ **cross-sectional**

The men in various categories are not the same.

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.
[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Data - **longitudinal** $\times$ **cross-sectional**

The men in various categories are not the same.

The older men were shorter, on average.

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.
[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?
Data - **longitudinal** $\times$ **cross-sectional**
The men in various categories are not the same.
The older men were shorter, on average.

When comparing averages first think:

1. How are the groups related to each other?
2. Take a look on the numerical averages.

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

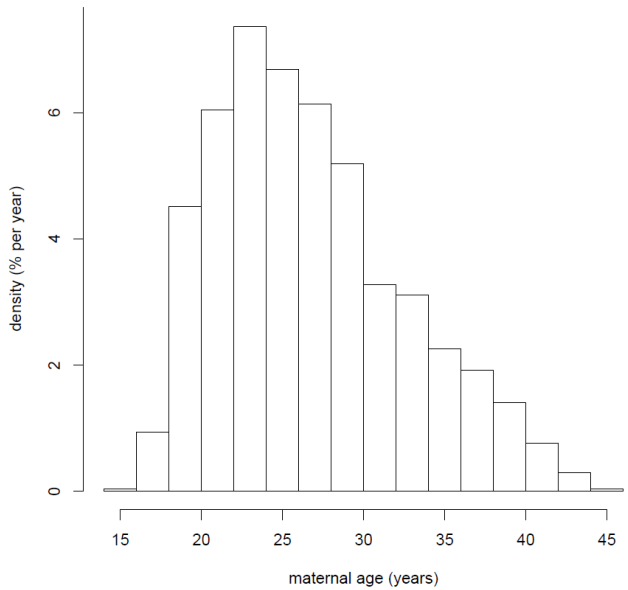
Range and interquartile range

How far from median?

Range and interquartile range

How far from median?

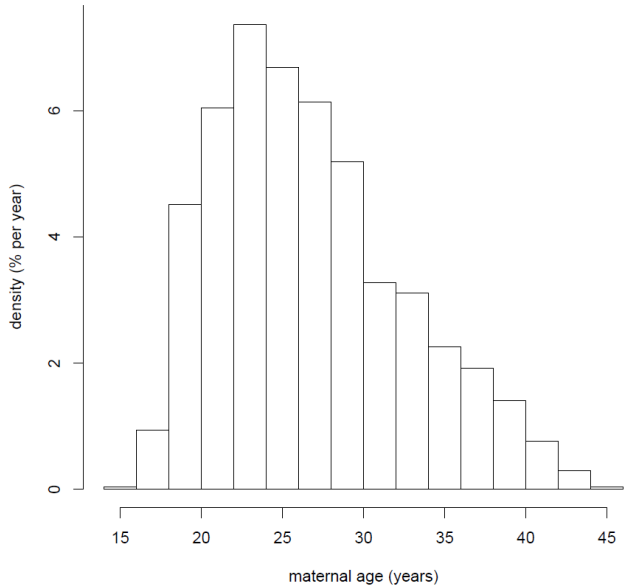
Maternal ages



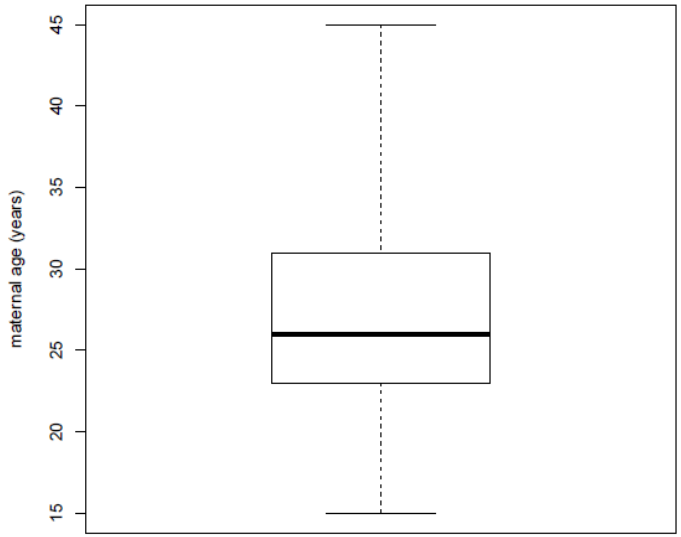
Range and interquartile range

How far from median?

Maternal ages



Box plot



Deviation from average

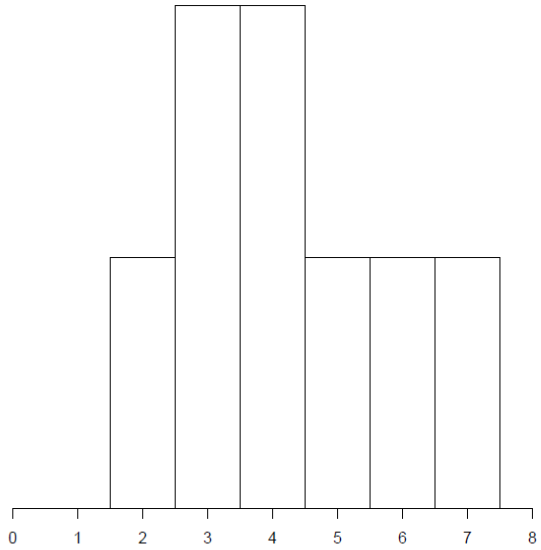
What is the typical (standard) deviation from average?

Deviation from average

What is the typical (standard) deviation from average?

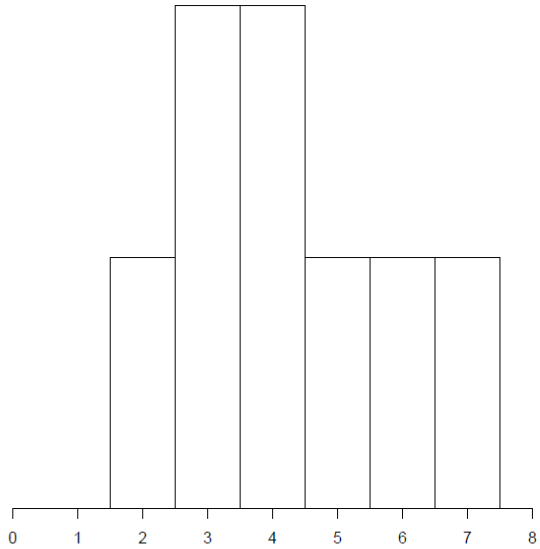
Deviation from average

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 meausres 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 $1/k^2$.

Infromation sources

https://courses.edx.org/courses/BerkeleyX/Stat_2.1x/
http://www.stat.berkeley.edu/ stark/SticiGui/Text/location.htm