

Pravděpodobnostní metody ve strojírenství

7. Úvod do matematické statistiky

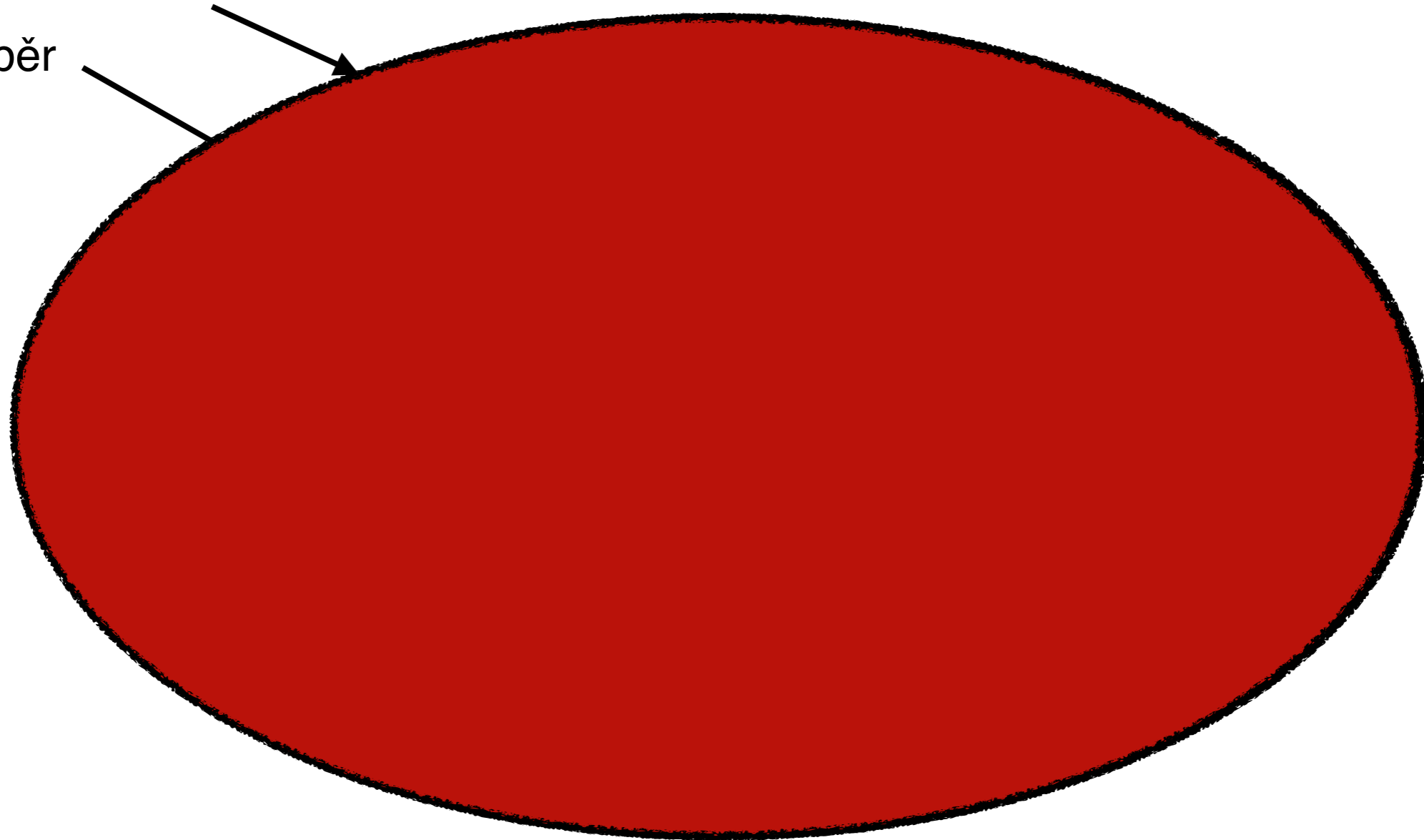


7. Úvod do matematické statistiky

- Klíčové pojmy:**
- Základní soubor, výběr
 - Statistická indukce, podmínky použití
 - Reprezentativnost výběru, náhodný výběr
 - Uspořádaný výběr, pořadová statistika, pořadí
 - Výběrové charakteristiky - střední hodnota, rozptyl, směrodatná odchylka, šikmost, špičatost
 - Krabicový diagram, kvartily, medián
 - Empirická distribuční funkce
 - Četnosti (absolutní, relativní, prosté kumulativní)
 - Histogram

Statistická indukce

- Základní soubor - nositel sledovaného znaku (veličiny)
- Výběr

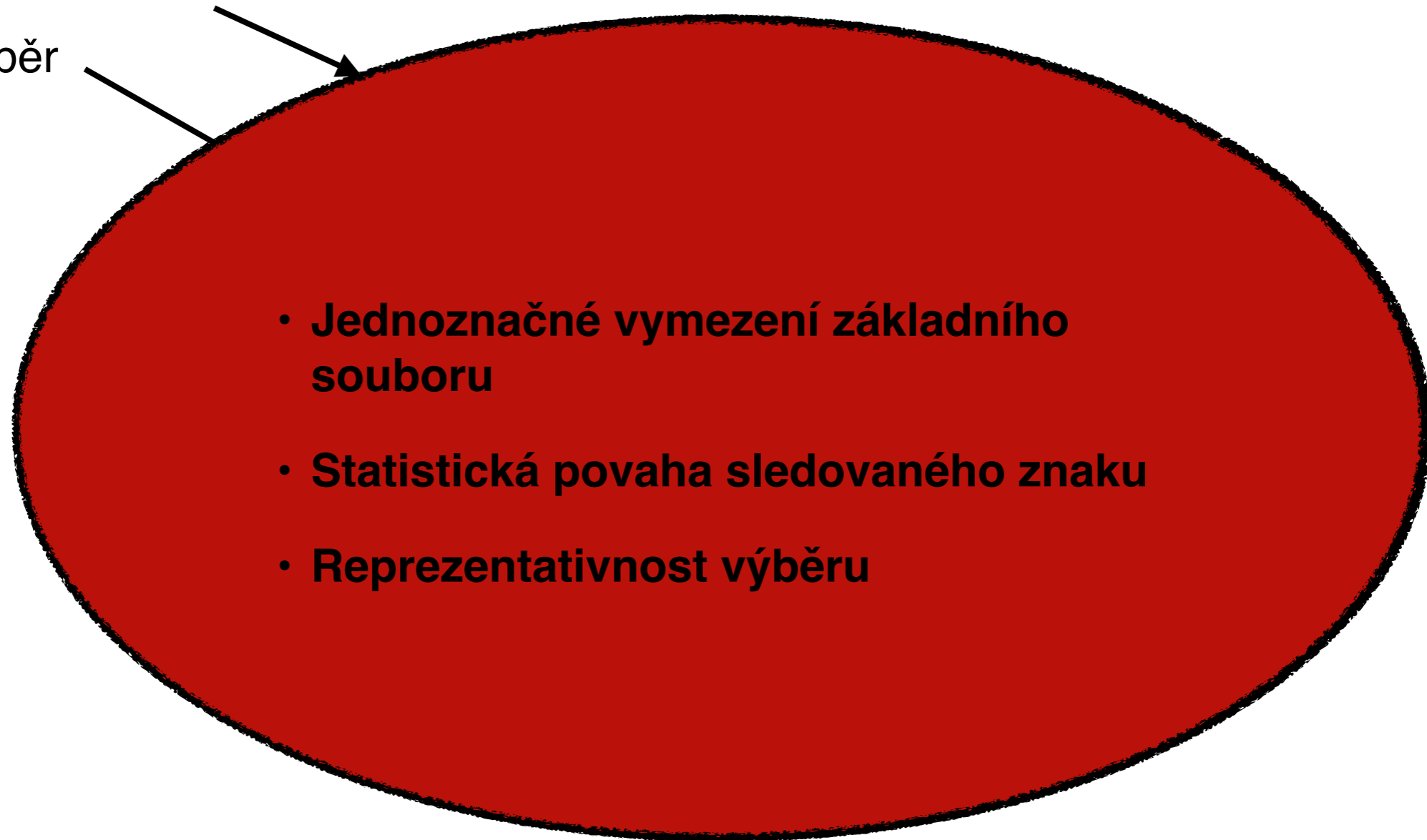


- Pozorování výběru
(měření sledovaného znaku) \Rightarrow Zjištění vlastností výběru
- Zobecnění na celý základní soubor



Statistická indukce

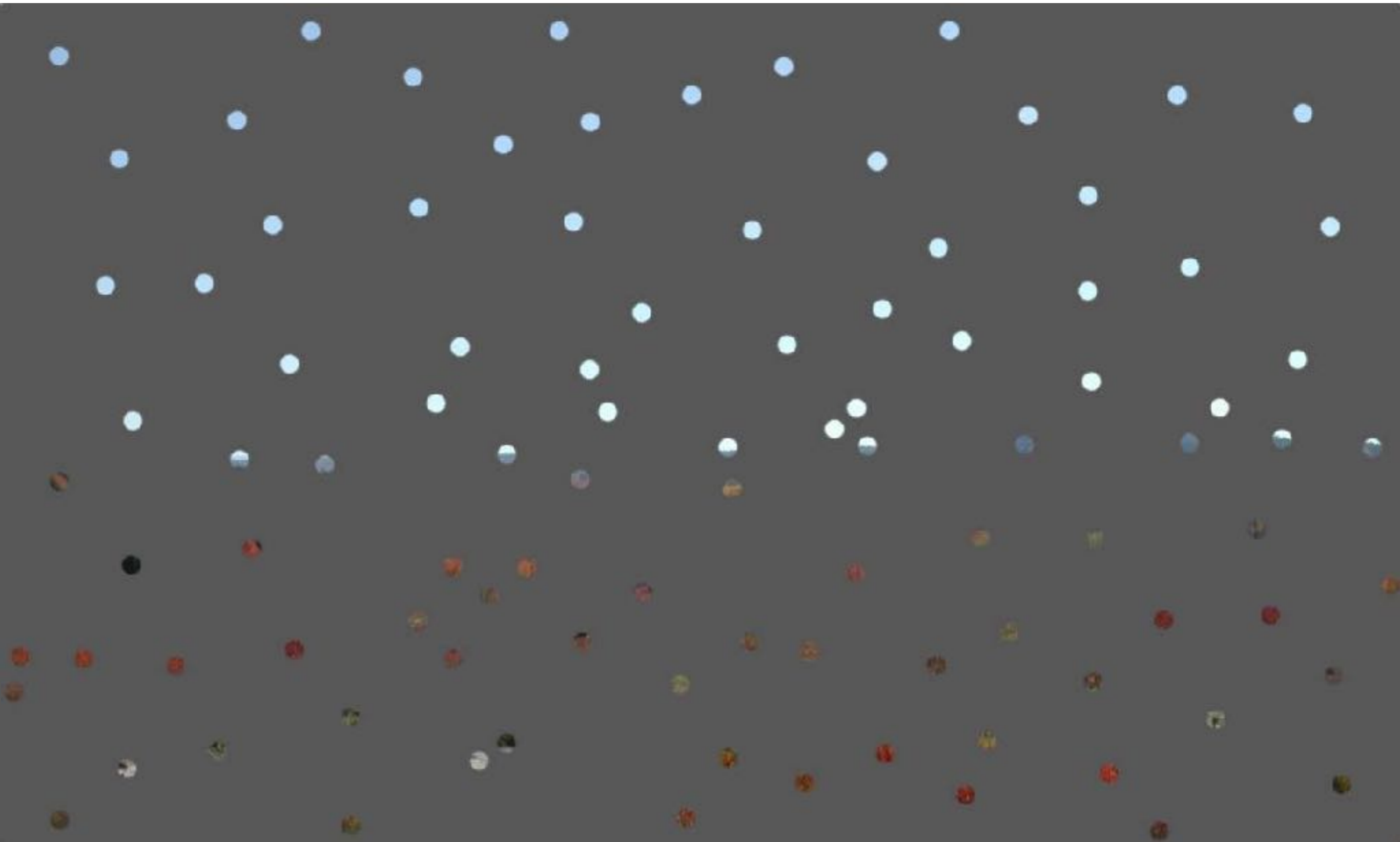
- Základní soubor - nositel sledovaného znaku (veličiny)
- Výběr



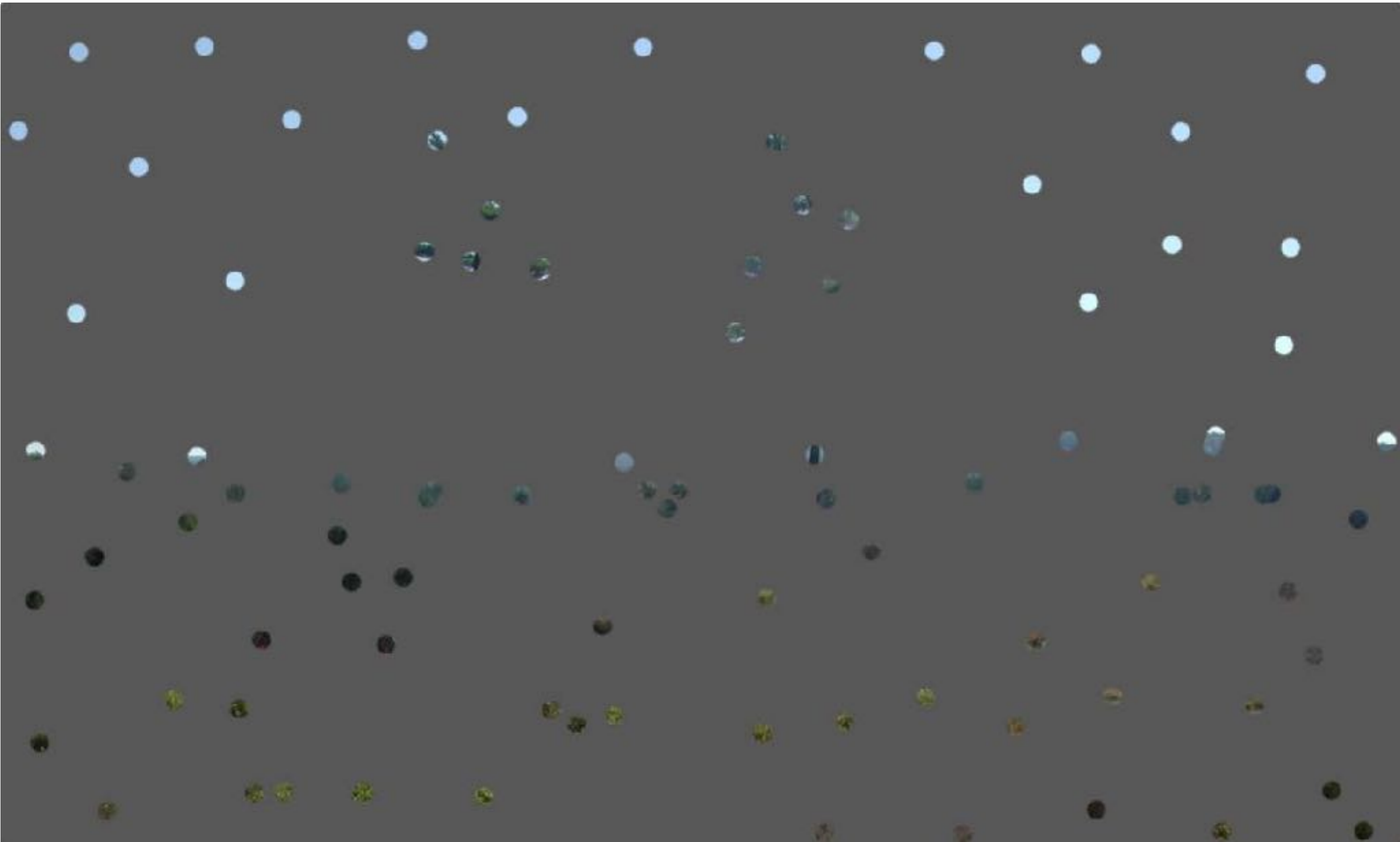
- Pozorování výběru
(měření sledovaného znaku) => Zjištění vlastností výběru
- Zobecnění na celý základní soubor



Statistická indukce



Statistická indukce



Statistická indukce



Základní pojmy popisné statistiky

Co je třeba znát a porozumět tomu:

- **Základní soubor (populace, universum):** množina objektů, na nichž provádíme statistické zkoumání; musí být přesně specifikována
- **Výběr (ze základního souboru):** n -tice nezávislých, stejně rozdělených náhodných veličin X_1, X_2, \dots, X_n odpovídající nezávislým pozorováním vybraných objektů základního souboru na nichž pozorujeme nějakou veličinu X reprezentující určitou měřitelnou (a přesně danou) vlastnost všech objektů základního souboru. Počet n objektů zahrnutých do výběru nazýváme **rozsahem výběru**.
- **Reprezentativnost výběru** je vlastnost výběru, zaručující rovnoměrné zastoupení charakteristických vlastností objektů základního souboru.
- **Náhodný výběr** vznikne tehdy, když každý objekt základního souboru má stejnou pravděpodobnost být zahrnut do výběru.
- **Realizace výběru:** je množina naměřených (napozorovaných) číselných hodnot x_1, x_2, \dots, x_n veličin z výběru X_1, X_2, \dots, X_n .



Statistická indukce

Tedy vždy musíme vycházet z :

- náhodného výběru X_1, X_2, \dots, X_n ← náhodné veličiny
- a jeho pozorování x_1, x_2, \dots, x_n . ← reálná čísla

Pravděpodobnostní charakteristiky	Výběrové charakteristiky
Střední hodnota $E(X) = \int_{-\infty}^{\infty} x f(x) dx$	Výběrový průměr $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$
Momenty $\mu_k(X) = E(X - E(X))^k$	Výběrové momenty $m_k(X) = \sum_{i=1}^n (x_i - \bar{X})^k$
Rozptyl $var(X) = E(X - E(X))^2$	Výběrový rozptyl $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$
Kvantily $\tilde{x}_{100\alpha}$	Výběrové kvantily $X_{([np]+1)}$



Statistické charakteristiky

Statistické charakteristiky: jsou spočteny na základě pozorování x_1, x_2, \dots, x_n výběru X_1, X_2, \dots, X_n .

- *Výběrová střední hodnota (výběrový průměr):* $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$
- *Výběrové momenty:* $m_k(X) = \sum_{i=1}^n (x_i - \bar{X})^k$
- *Výběrový rozptyl:* $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$ a *výběrová směrodatná odchylka* s

Za předpokladu, že náhodný výběr je nezávislý a je z normálního rozdělení, t.j. X_1, X_2, \dots, X_n jsou i.i.d. a $X_k \sim N(\mu, \sigma^2)$, $k = 1, 2, \dots, n$, lze určit rozdělení pravděpodobnosti některých charakteristik:

- Pokud je μ a σ^2 známé, má výběrový průměr \bar{X}_n rozdělení $N(\mu, \sigma^2/n)$
- Pokud μ a σ^2 neznáme, má veličina $T = (X - \bar{X})/s$ tzv. Studentovo neboli t -rozdělení $t(n-1)$
- Veličina $S^2 = (n-1) \cdot s^2 / \sigma^2$ má $\chi^2(n-1)$ rozdělení (o $n-1$ stupních volnosti)



Statistické charakteristiky

Statistické charakteristiky: jsou spočteny na základě pozorování x_1, x_2, \dots, x_n výběru X_1, X_2, \dots, X_n .

Další důležité výběrové charakteristiky:

- Výběrová šikmost (skewness): $Skew(X) = \frac{m_3(X)}{m_2^{3/2}(X)}$
- Výběrová špičatost (kurtosis): $Kurt(X) = \frac{m_4(X)}{m_2^2(X)} - 3$

pro $X \sim N(\mu, \sigma^2)$ je $E(Skew(X)) = 0$ $var(Skew(X)) = \frac{6(n-2)}{(n+1)(n+3)}$

$$E(Kurt(X)) = -\frac{6}{n+1} \quad var(Kurt(X)) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

Máme-li dostatečný počet pozorování (řádově stovky), mají statistiky

$$T_3 = \frac{S_{kew}^{norm}}{\sqrt{Var(S_{kew}^{norm})}} \quad T_4 = \frac{K_{urt}^{norm} - E(K_{urt}^{norm})}{\sqrt{Var(K_{urt}^{norm})}}$$

přibližně standardní normální rozdělení pravděpodobnosti.



Základní pojmy popisné statistiky

Co je třeba znát a porozumět tomu:

- **Výběrový průměr** nahrazuje neznámou střední hodnotu veličiny X :
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$
- **Výběrový rozptyl** je charakteristika odpovídající rozptylu náhodné veličiny X :
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Výběrová směrodatná odchylka s je druhou odmocninou z výběrového rozptylu
- **k -tý výběrový centrální moment** veličiny X se počítá podle vztahu:
$$m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$
- **Výběrový index šikmosti** je výběrovou variantou indexu šikmosti pozorované veličiny X :
$$Skew(X) = \frac{m_3(X)}{m_2^{3/2}(X)}$$
- **Výběrový index špičatosti** je výběrovou variantou indexu špičatosti pozorované veličiny X :
$$Kurt(X) = \frac{m_4(X)}{m_2^2(X)} - 3$$
- **Výběrový modus** je nejčastější hodnota, která se vyskytuje v realizaci výběru. Tato hodnota nemusí existovat.



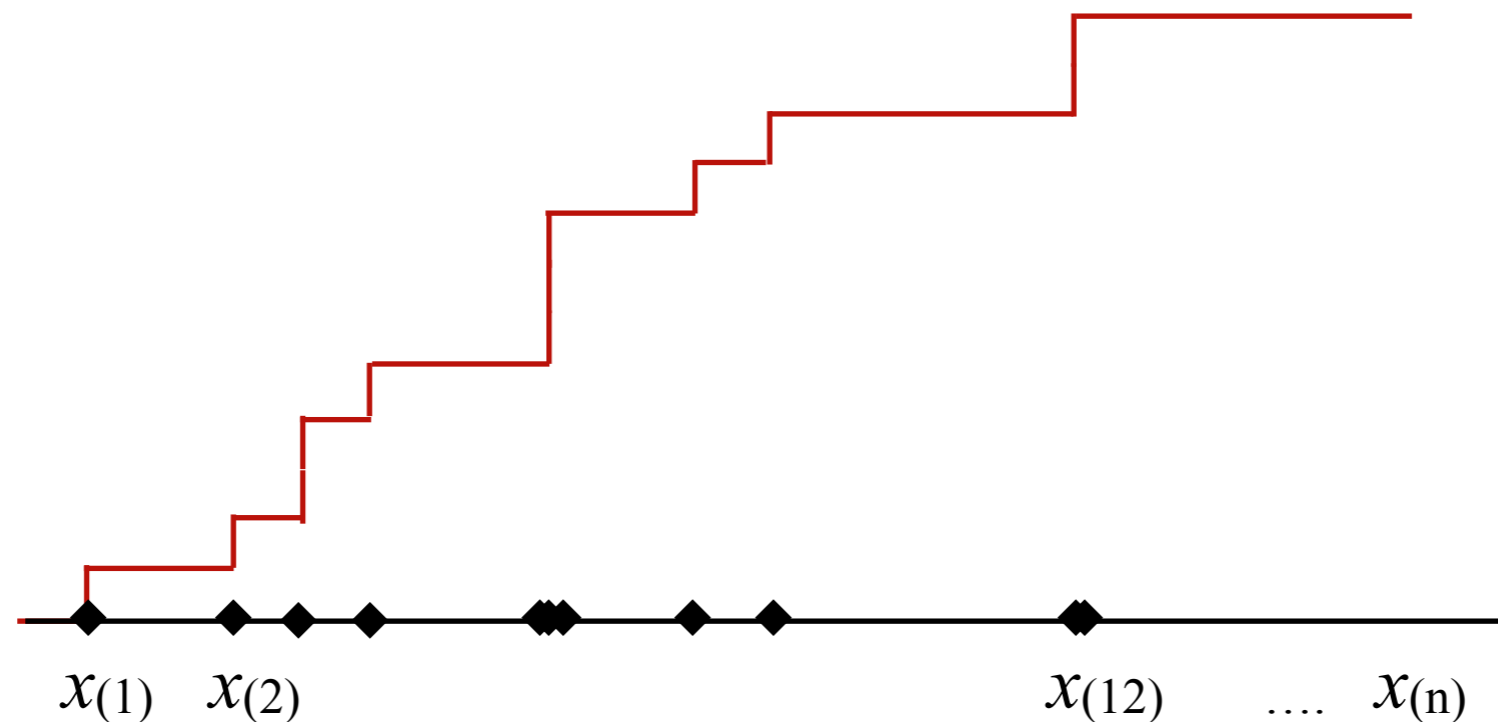
Statistické charakteristiky

Statistické charakteristiky: jsou spočteny na základě pozorování x_1, x_2, \dots, x_n výběru X_1, X_2, \dots, X_n . Uspořádáme-li veličiny ve výběru podle velikosti naměřených hodnot pozorování x_1, x_2, \dots, x_n , dostaneme tzv. **uspořádaný výběr** $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

Empirická distribuční funkce:

vycházíme z uspořádaného výběru: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. Potom $F_n(x_{(i)}) = \frac{i}{n}$

a tedy $F_n(x) = \frac{\max\{k : X_{(k)} \leq x\}}{n}, \quad x \in \mathbf{R}$



Základní pojmy popisné statistiky

Co je třeba znát a porozumět tomu:

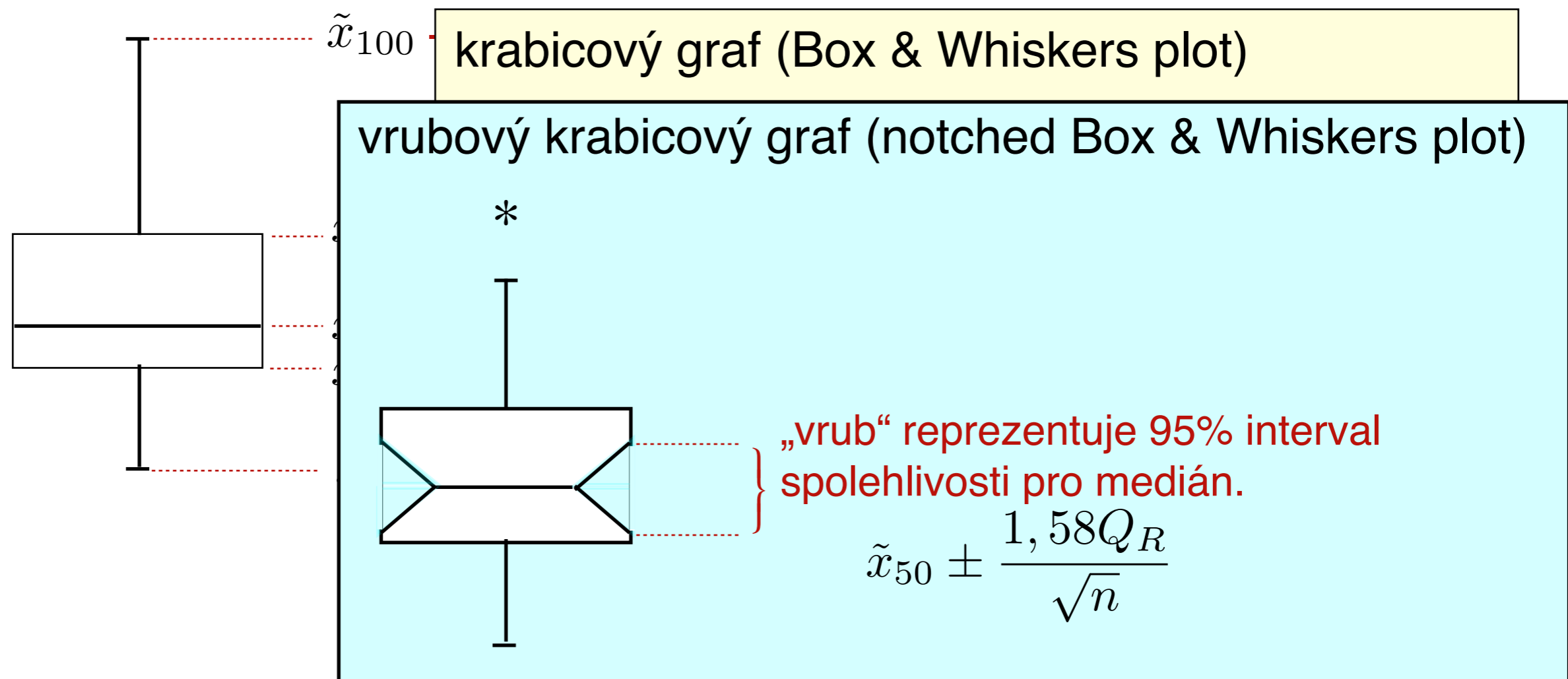
- **Uspořádaný výběr:** $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ vznikne z původního výběru X_1, X_2, \dots, X_n upořádáním podle velikosti pozorovaných hodnot x_1, x_2, \dots, x_n .
- **Pořadová statistika:** $X_{(k)}$ je náhodná veličina X_m , která je k -tá v pořadí podle velikosti pozorovaných hodnot x_1, x_2, \dots, x_n .
- **Pořadí** m -tého pozorování veličiny X_m ve výběru: pokud $X_m = X_{(k)} \Rightarrow R_m = k$.
- $X_{(1)}$ se nazývá **(výběrové) minimum**, $X_{(n)}$ je **(výběrové) maximum**
- **medián** je prostřední hodnota ve výběru: je-li n liché, je roven $X_{([n/2]+1)}$
pro n sudé je roven $(X_{(n/2)} + X_{(n/2+1)})/2$
- **dolní kvartil:** $X_{([n/4]+1)}$ resp. $(X_{(n/4)} + X_{(n/4+1)})/2$
- **horní kvartil:** $X_{([3n/4]+1)}$ resp. $(X_{(3n/4-1)} + X_{(3n/4)})/2$



Statistické charakteristiky

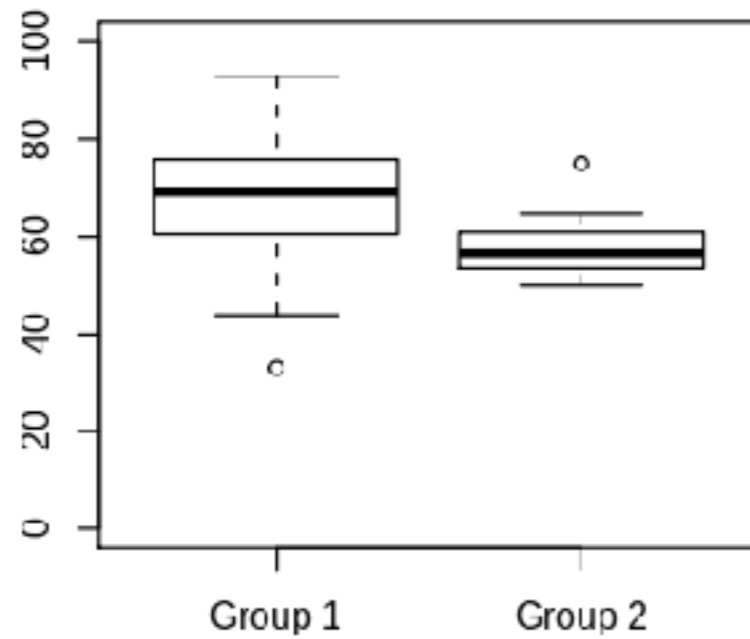
Statistické charakteristiky: jsou spočteny na základě pozorování x_1, x_2, \dots, x_n výběru X_1, X_2, \dots, X_n . Uspořádáme-li veličiny ve výběru podle velikosti naměřených hodnot pozorování x_1, x_2, \dots, x_n , dostaneme tzv. **uspořádaný výběr** $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

krabicový graf (Box & Whiskers plot)

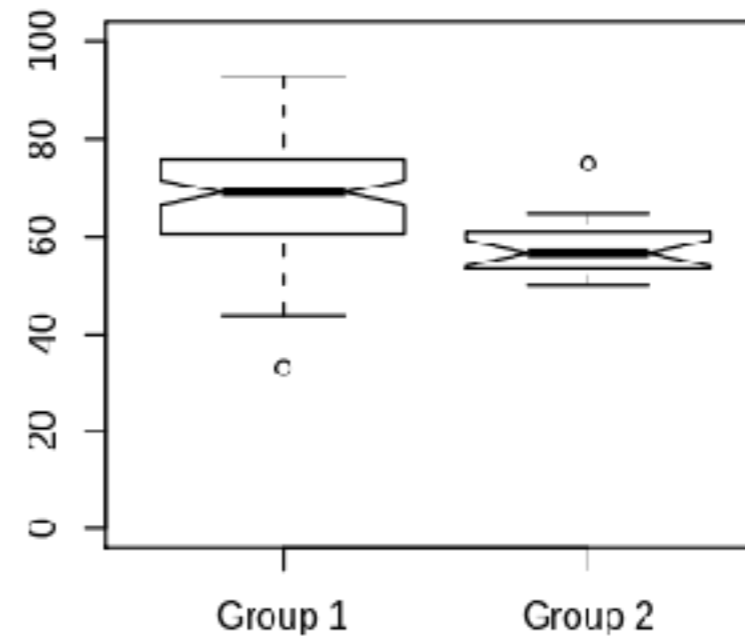


Statistické charakteristiky

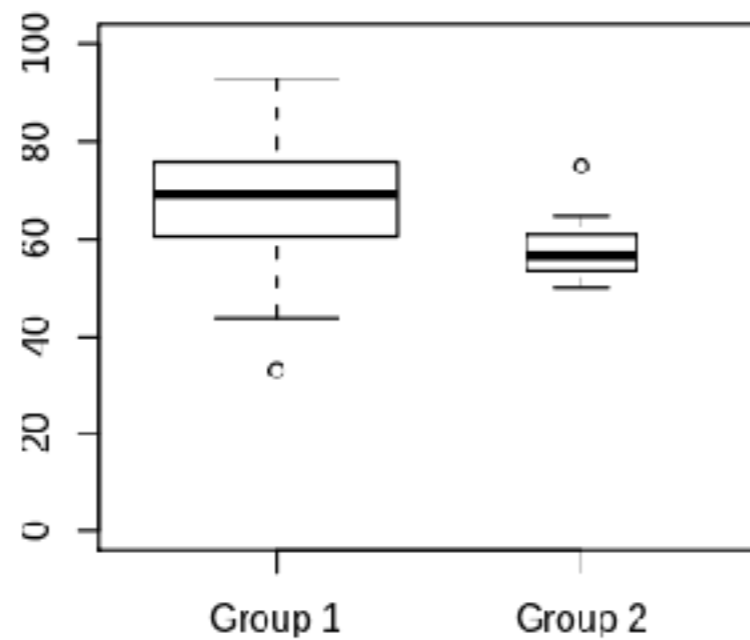
Traditional Box Plot



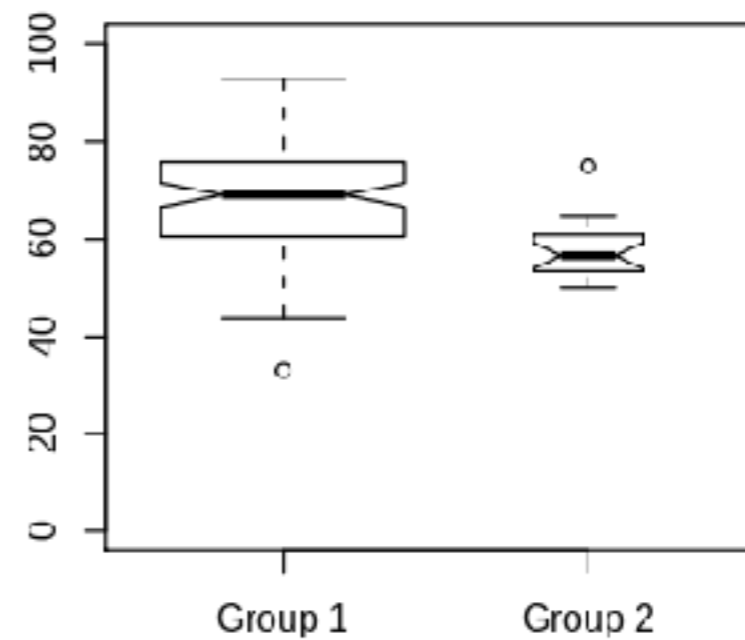
Notched Box Plot



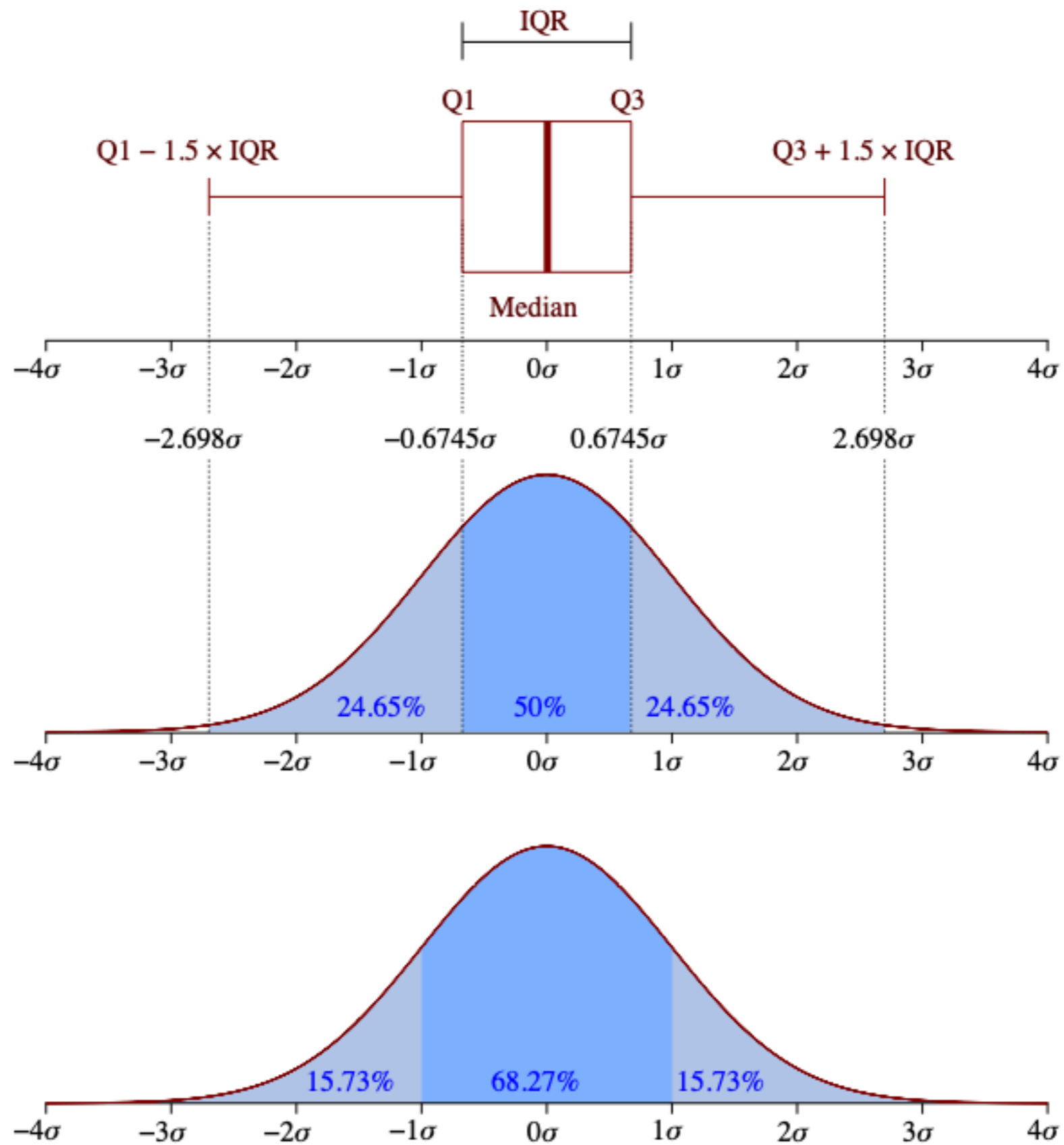
Variable Width Box Plot



Variable Width Notched Box Plot



Statistické charakteristiky



Základní pojmy popisné statistiky

Co je třeba znát a porozumět tomu:

- **Krabicový diagram** je grafické zobrazení rozdělení pozorovaných hodnot pomocí pěti (Tukey's) charakteristik: minima, dolního kvartilu, mediánu, horního kvartilu a maxima.
- **Empirická distribuční funkce** je grafické zobrazení realizace výběru formou grafu po částech konstantní funkce. Vychází z uspořádaného výběru: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. Potom

$$F_n(x_{(i)}) = \frac{i}{n} \quad \text{a tedy} \quad F_n(x) = \frac{\max\{k : X_{(k)} \leq x\}}{n}, \quad x \in \mathbf{R}$$



STATISTICA

Version 10

STATISTICA 10

Home Server Edit View Insert Format Tables Analysis Tools

Extract Clean Analyze Base Identify Predict

Advanced Models
Multiple Regression
Power Analysis

Copyright © StatSoft, Inc. 1984 - 2011. All Rights Reserved.
This program is protected by US and international copyright.



StatSoft



File Home Edit View Insert Format Statistics Data Mining Graphs Workbook

Clipboard: Paste, Cut, Copy, Delete, Screen Catcher

View: Controls, Grid, Status Bar

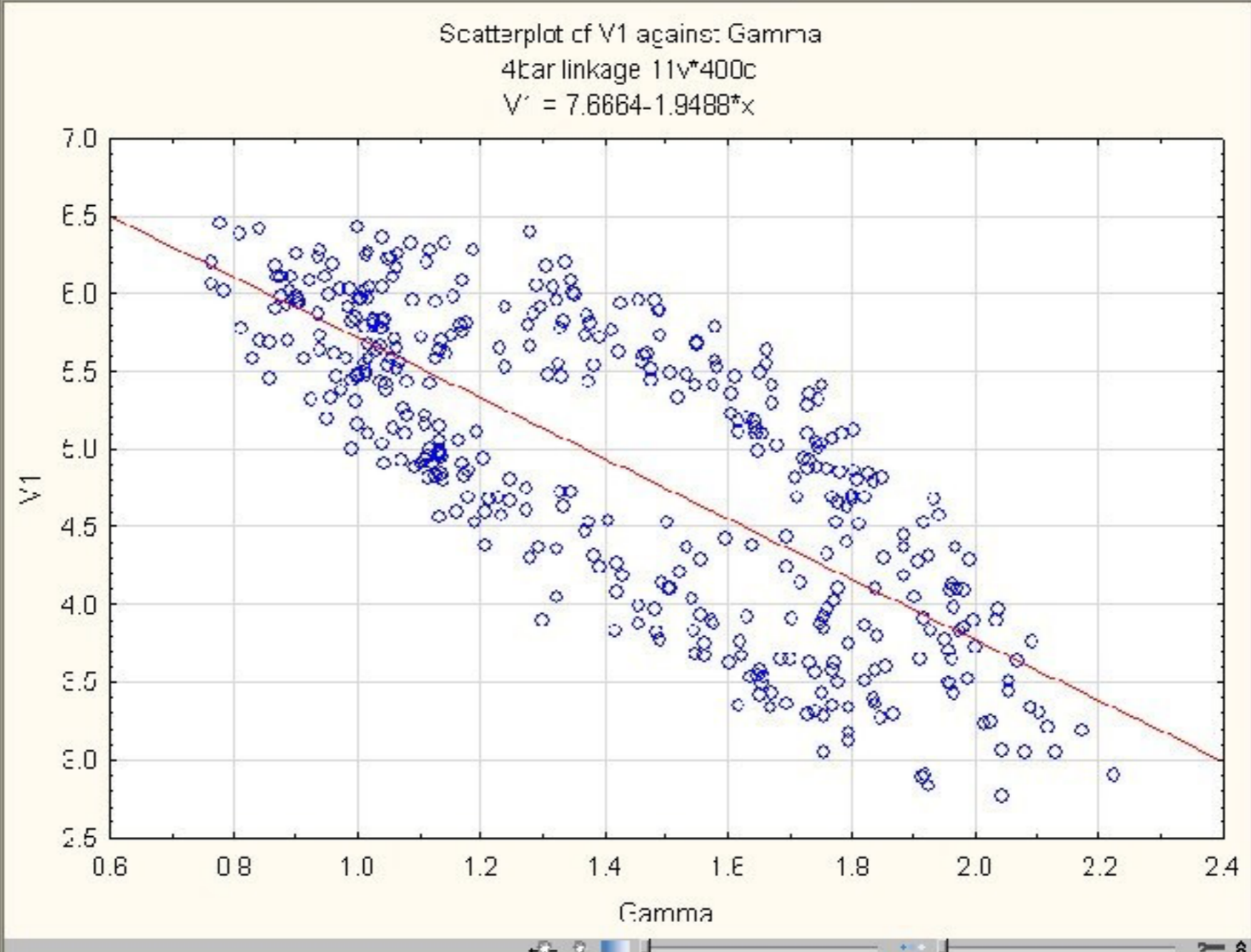
Customize Graph: Normal Graph [modi...]

Options: Selection, Graph

Tools: MultiDoc Style, Plots, Lock Graph, Graph Data, Header/Footer, Events, Zoom subgraph, Merge Graph

Links: OLE, Object

- Workbook1*
- 2D Scatterplots (4bar linkage)
- Scatterplot of V1 against



Scatterplot of V1 against Gamma

2D Scatterplots





Minitab®

Minitab - Minitab Demonstration.mpx

File Edit Data Calc Stat Graph View Help Assistant

Navigator

- Bubble Plot of Hours vs Spe...
- General Linear Model: Hour...
- Pareto Chart of Vehicle Color
- Chart of Mean(Correction ...
- Scatterplot of Hours vs Speed
- Summary Report for Thicken...
- Process Capability Report fo...
- Boxplot of Thickness, Data
- Process Capability Sixpack R...
- One-way ANOVA: Correctio...
- Regression Analysis: Hours ...
- Regression Analysis: Hours ...
- Regression Analysis: Hours ...
- Regression Analysis: Hours ...

Regression Analysis: Hours ...

BELTWEAR

Regression Analysis: Hours versus Speed

Model Summary

S	R-sq	R-sq(adj)
126,186	17,93%	16,36%

Fitted Line Plot
Hours = 1510 - 0,2028 Speed

+	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
	Belt Type	Hours	Speed	Humidity													
12	HS-91	689	3660	77													
13	HS-91	975	3120	50													
14	NS-33	1015	2550	59													
15	HS-91	827	3510	48													
16	NS-33	663	3010	58													

H < > + BeltWear Measurements ProcessImprovementData.MTW Prkg Violation

BeltWear 90%

DDS 2004 v 2.2.7 - [OnLine]

Tree Graph Report Route Online Tools Help

Start Stop A3716 Manual Auto

Data Collection Units Refresh

OnLine

- Unit1
 - Device 1
 - Pump1 [M]
 - CH1 [10]
 - Vel [R, C]
 - Acc [R, C]
 - Acc-Time [R, C]
 - Vel-Spec [R, C]
 - Acc-Spec [R, C]
 - CH2 [10]
 - CH3 [10]
 - CH4 [10]
 - Pump2 [M]
 - Pump3 [M]
 - Device 2
 - Fan1 [M]

OnLine/Unit1/Device 1/Pump1/CH1/Vel

OnLine/Unit1/Device 1/Pump1/CH1/Acc

OnLine/Unit1/Device 1/Pump1/CH1/Vel-Spec 18.8.2015 9:34:03.894

ZOOM TOT 0,1035mm/s (RMS); TOT 0,1037mm/s (RMS)

[Filtered]

OnLine/Unit1/Device 1/Pump1/CH1/Acc-Spec 18.8.2015 9:34:03.364

ZOOM TOT 0,0097g (RMS); TOT 0,0244g (RMS)

#	Date/Time	Value[g]	Alarm
11	18.8.2015 9:31:57.659	0,0858	●
12	18.8.2015 9:31:58.719	2,006	●
13	18.8.2015 9:31:59.780	0,0882	●
14	18.8.2015 9:32:25.770	0,0921	●
15	18.8.2015 9:32:26.831	0,2317	●
16	18.8.2015 9:32:33.395	0,1308	●
17	18.8.2015 9:32:50.434	0,0229	●

Preview

The Wolfram Solution for Statistics

Pull in your data or ours, do standardized or custom analysis and visualization, then generate and deploy interactive reports—all in one system, with one integrated workflow.

The Wolfram statistics solution includes powerful optimization and statistics functionality alongside unique capabilities like free-form linguistic input, instant interactivity and the reliability of any-precision numerics and symbolic analysis.

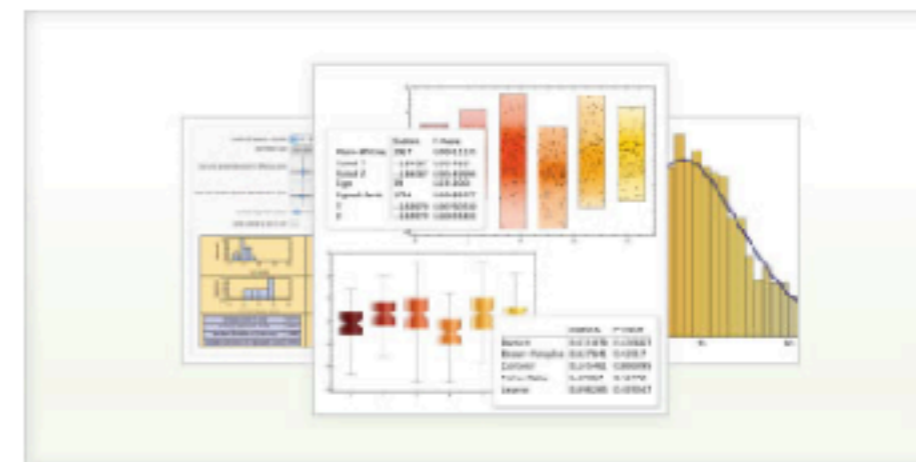
The Wolfram Edge

How Wolfram Compares

Key Capabilities

Wolfram technologies include thousands of **built-in functions** and **curated data** on many topics that let you:

- **Model and analyze system reliability**, including the life expectancy of a component or the reliability of engineering materials, using built-in reliability analysis functionality
- **Build and analyze statistical models** for futures trading and other applications in finance
- **Determine insights** by analyzing current and historical built in financial, socioeconomic, geographic and demographic data

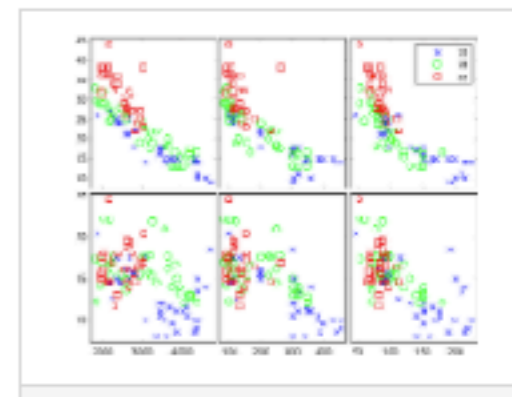
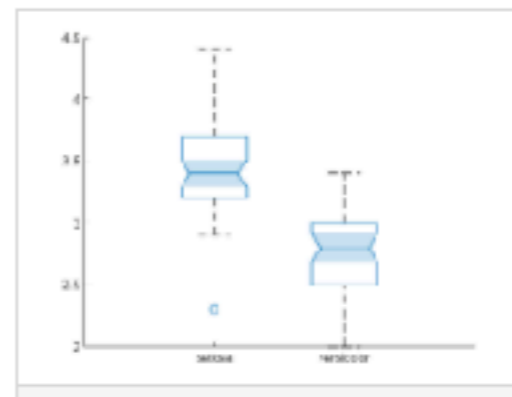
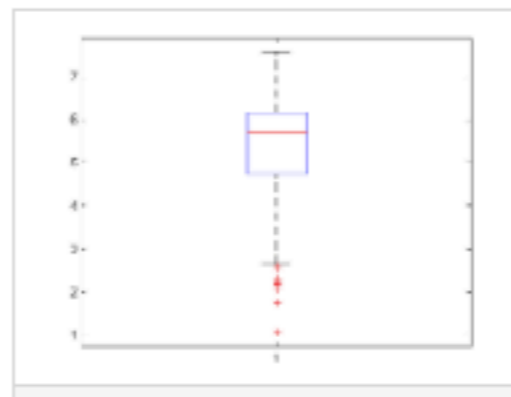
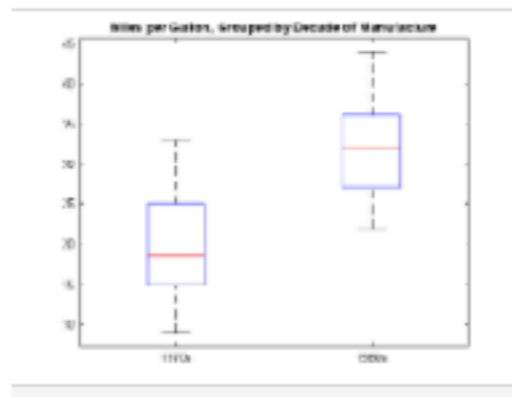
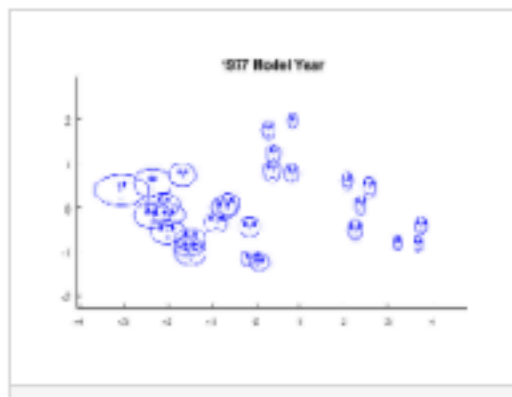


Fit, test and describe properties of data

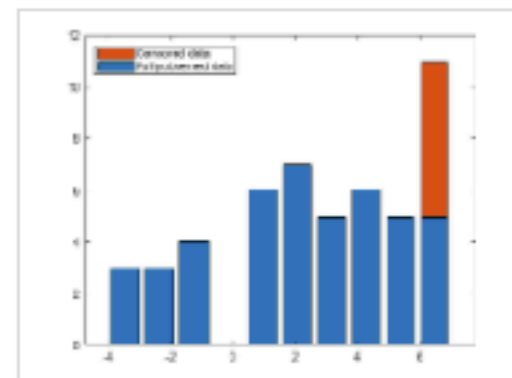
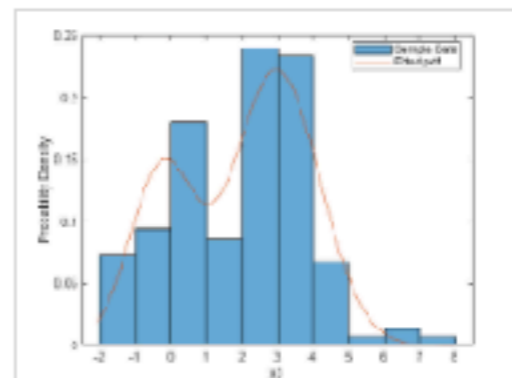
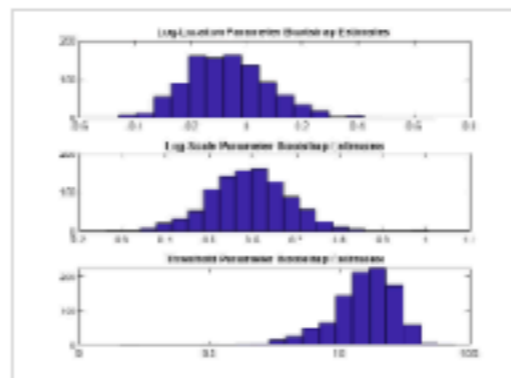
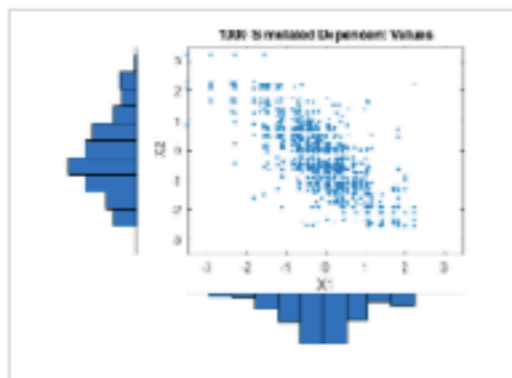
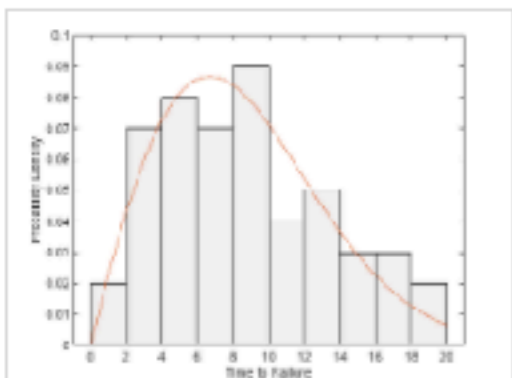


Statistics and Machine Learning Toolbox – Examples

Descriptive Statistics and Visualization



Probability Distributions



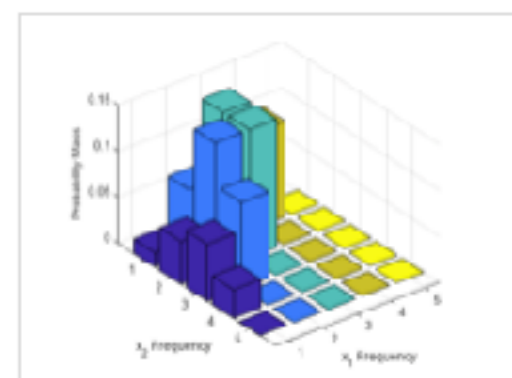
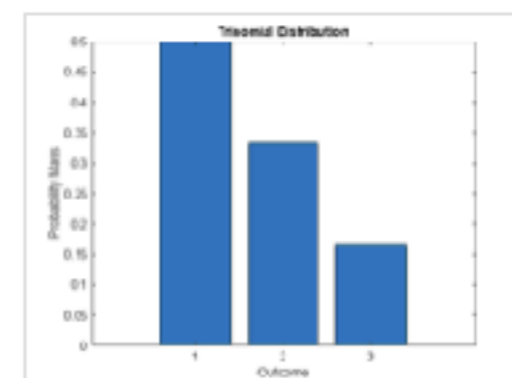
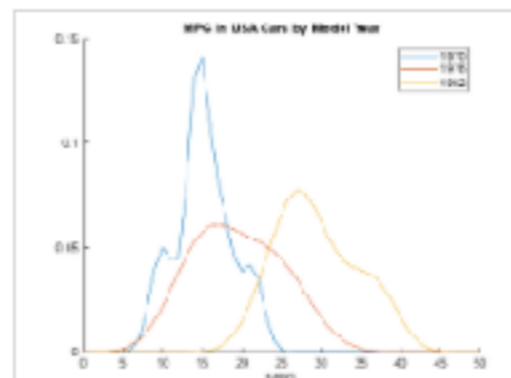
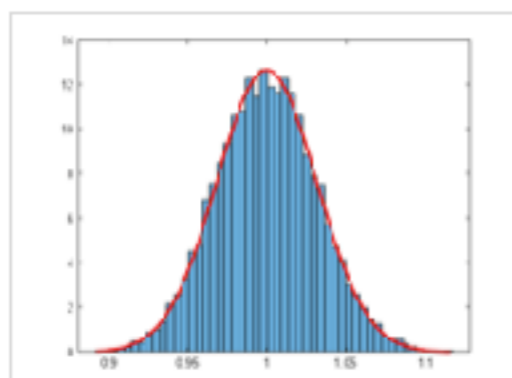
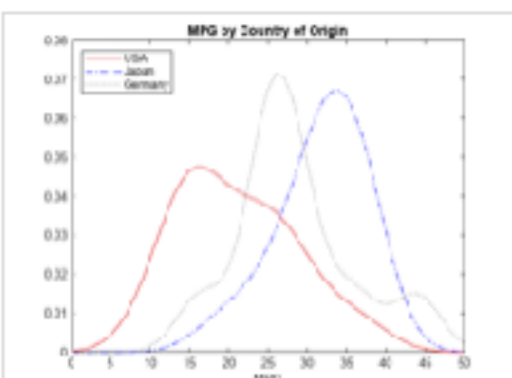
Curve Fitting and Distribution Fitting

Simulating Dependent Random Variables Using Copulas

Fitting a Univariate Distribution Using Cumulative Probabilities

Fit Custom Distributions

Avoid Numerical Issues When Fitting Custom Distributions



Fit Probability Distribution Objects to Grouped Data

Maximum Likelihood Estimation

Compare Multiple Distribution Fits

Multinomial Probability Distribution Objects

Multinomial Probability Distribution Functions



Statistics for Data Science with Python

Python statistics library that is open source

There are numerous open-source Python libraries and Python statistics packages for data manipulation, data visualization, statistics, mathematics, machine learning, and natural language processing.

1. Python's Statistics

It is a built-in Python module for descriptive statistics. If your datasets are not too large or if you cannot rely on importing other libraries, you can utilize it.

2. NumPy

It is a third-party numerical computing package that is optimized for working with single- and multi-dimensional arrays. Its primary type is an array known as ndarray. This package offers a large number of statistical analysis routines.

3. SciPy

It is a NumPy-based third-party library for scientific computing. It provides more capabilities than NumPy, such as `scipy.stats` for statistical analysis.

4. Pandas

It is a NumPy-based third-party library for numerical computing. It excels at labelled one-dimensional (1D) data handling with Series objects and two-dimensional (2D) data handling with DataFrame objects.

5. Matplotlib

It is a third-party data visualization package. It is useful in conjunction with NumPy, SciPy, and Pandas.





[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Get Involved: Contributing](#)

[Developer Pages](#)

[R Blog](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 4.4.0 \(Puppy Cup\) prerelease versions](#) will appear starting Sunday 2024-03-24. Final release is scheduled for Wednesday 2024-04-24.
- [R version 4.3.3 \(Angel Food Cake\)](#) has been released on 2024-02-29.
- [Registration for useR! 2024](#) has opened with early bird deadline March 31 2024.
- [R version 4.2.3 \(Shortstop Beagle\)](#) has been released on 2023-03-15.
- You can support the R Foundation with a renewable subscription as a [supporting member](#).



Welcome

Preface

I Part 1

1 Project 1: Weighted Dice

2 The Very Basics

3 Packages and Help Pages

II Part 2

4 Project 2: Playing Cards

5 R Objects

6 R Notation

7 Modifying Values

8 Environments

III Part 3

Project 3: Slot Machine

9 Programs

10 S3

11 Loops

12 Speed

Appendix

[A Installing R and RStudio](#)

A.1 How to Download and Install R

A Installing R and RStudio

To get started with R, you need to acquire your own copy. This appendix will show you how to download R as well as RStudio, a software application that makes R easier to use. You'll go from downloading R to opening your first R session.

Both R and RStudio are free and easy to download.

A.1 How to Download and Install R

R is maintained by an international team of developers who make the language available through the web page of [The Comprehensive R Archive Network](#). The top of the web page provides three links for downloading R. Follow the link that describes your operating system: Windows, Mac, or Linux.

A.1.1 Windows

To install R on Windows, click the "Download R for Windows" link. Then click the "base" link. Next, click the first link at the top of the new page. This link should say something like "Download R 3.0.3 for Windows," except the 3.0.3 will be replaced by the most current version of R. The link downloads an installer program, which installs the most up-to-date version of R for Windows. Run this program and step through the installation wizard that appears. The wizard will install R into your program files folders and place a shortcut in your Start menu. Note that you'll need to have all of the appropriate administration privileges to install new software on your machine.

A.1.2 Mac

To install R on a Mac, click the "Download R for Mac" link. Next, click on the `R-3.0.3` package link (or the package link for the most current release of R). An installer will download to guide you through the installation process, which is very easy. The installer lets you customize your installation, but the defaults



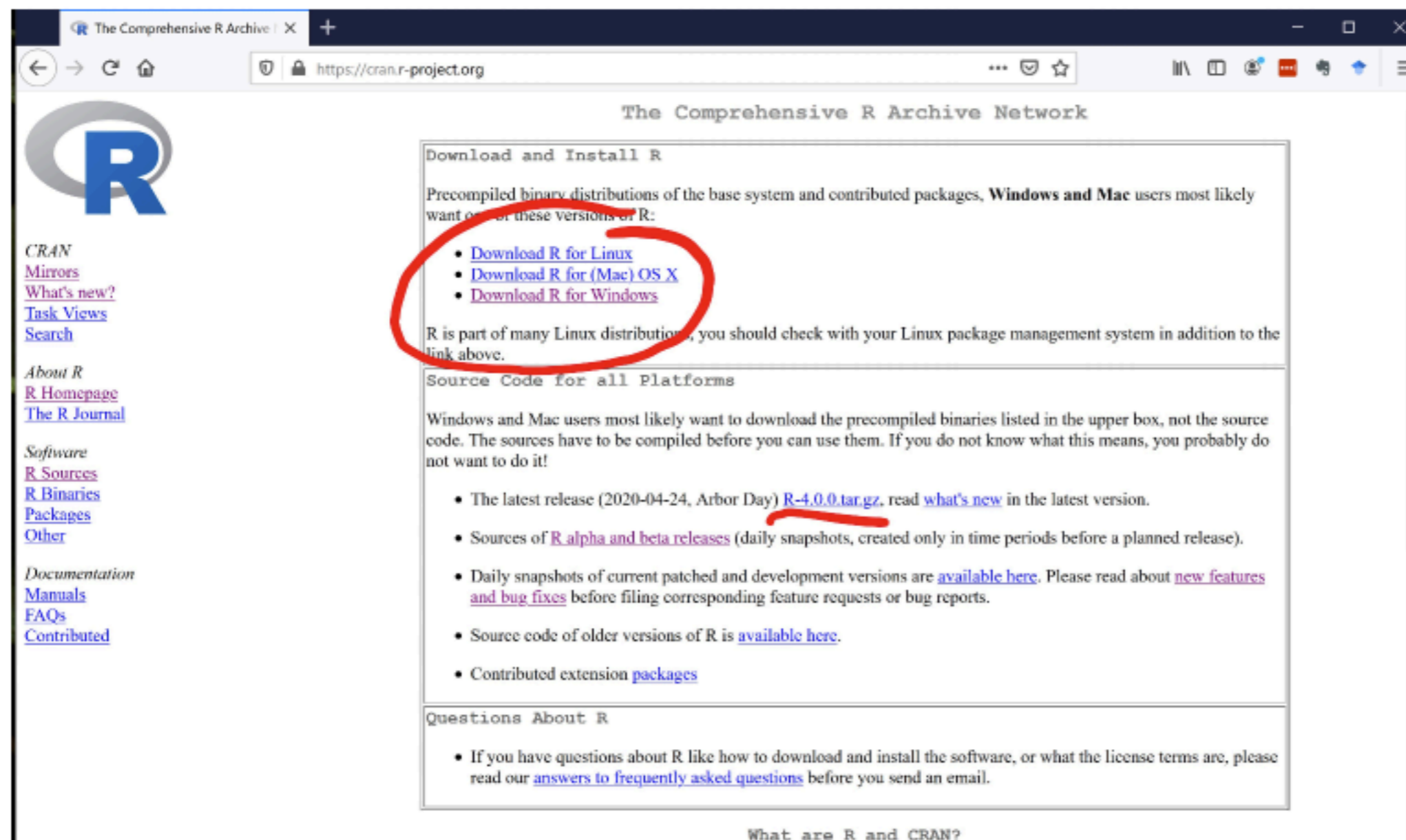
Install R and RStudio

We need to install two things onto your computer. The first is **R**, this is the programming language we will use. After that we need to install **RStudio**, this is a front end program that lets you write R code, view plots, and do many other useful things.

If you have a previous installation of R and RStudio, that you have not used much, please remove these first from your computer, and download the current versions as per the instructions below.

Install R

1. Download the R installer from <https://cran.r-project.org/>.
 - a. Click on the link for your operating system. Make sure the installer is for R version 4.0.0.



The screenshot shows the CRAN website with the following content:

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-04-24, Arbor Day) [R-4.0.0.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?



```

16 ecdf(x)
17 plot(ecdf(x), verticals = FALSE, do.points="FALSE")
18 xfit<-seq(min(x),max(x),length=40)
19 zfit<-pnorm(xfit,mean=mean(x),sd=sd(x))
20 lines(xfit, zfit, col="red", lwd=2)
21 #
22 # Histogram
23 hist(x)
24 hx <- hist(x)
25 print(hx)
26 h<-hist(x, main="Belicci automat na kávu", xlab="hmotnost beličků v dekagramech",
27       col="yellow", freq=FALSE, breaks = 8)
28 yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
29 lines(xfit, yfit, col="red", lwd=2)
30 lines(density(x))
31 #
32 #Testy normality
33 qqnorm(x, main='Normal')

```

Environment History Connections Tutorial

R - Global Environment

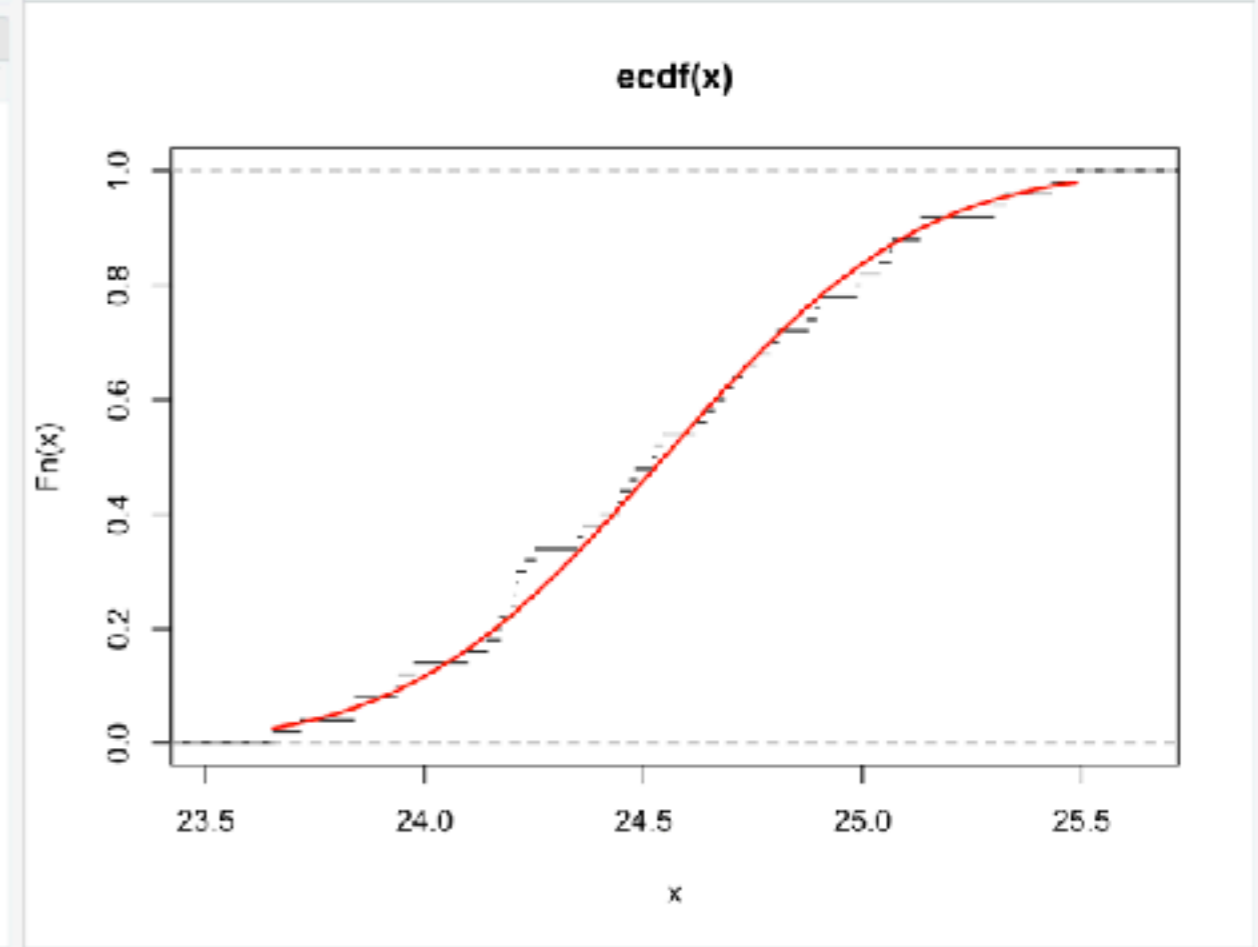
Data	
h	List of 6
hx	List of 6
x	num [1:50, 1] 24.5 24.2 24.5 24.4 23.9 ...

Values	
xfit	num [1:40] 23.7 23.7 23.7 23.8 23.8 ...
y	num [1:50] 24.5 24.2 24.5 24.4 23.9 ...
yfit	num [1:40] 0.0257 0.0325 0.0407 0.0505 0.0621 ...
zfit	num [1:40] 0.0257 0.0325 0.0407 0.0505 0.0621 ...

```

R 4.3.1 - ~/Nextcloud/Vyuka/ZS/R/
> #
> #Grafy
> plot(x)
> boxplot(x)
> boxplot(x, main="Hmotnost beličků dávkovaných automatem", ylab="hmotnost v dag")
> ecdf(x)
Empirical CDF
Call: ecdf(x)
 x[1:50] = 23.654, 23.717, 23.837, ..., 25.43, 25.489
> plot(ecdf(x), verticals = FALSE, do.points="FALSE")
> xfit<-seq(min(x),max(x),length=40)
> zfit<-norm(xfit,mean=mean(x),sd=sd(x))
Error in norm(xfit, mean = mean(x), sd = sd(x)) :
  unused arguments (mean = mean(x), sd = sd(x))
> zfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
> lines(xfit, zfit, col="red", lwd=2)
> yfit<-pnorm(xfit,mean=mean(x),sd=sd(x))
> ecdf(x)
Empirical CDF
Call: ecdf(x)
 x[1:50] = 23.654, 23.717, 23.837, ..., 25.43, 25.489
> plot(ecdf(x), verticals = FALSE, do.points="FALSE")
> xfit<-seq(min(x),max(x),length=40)
> zfit<-pnorm(xfit,mean=mean(x),sd=sd(x))
> lines(xfit, zfit, col="red", lwd=2)
>

```



Statistické charakteristiky

Příklad: Balící automat na kávu - popisná statistika v programu R:

```
R version 4.3.1 (2023-06-16) -- "Beagle Scouts"  
Copyright (C) 2023 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin20 (64-bit)
```

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.



Statistické charakteristiky

Příklad: Balící automat na kávu - popisná statistika v programu R:

```
# Nastavení pracovního adresáře:  
> setwd(„/Users/dohnal/Nextcloud/Vyuka/ZS/R“)  
#  
# Volání potřebných knihoven:  
> library(moments)  
> library(nortest)  
#  
# Načtení dat  
> x <- data.matrix(read.table("davkovac.txt"))  
> x
```

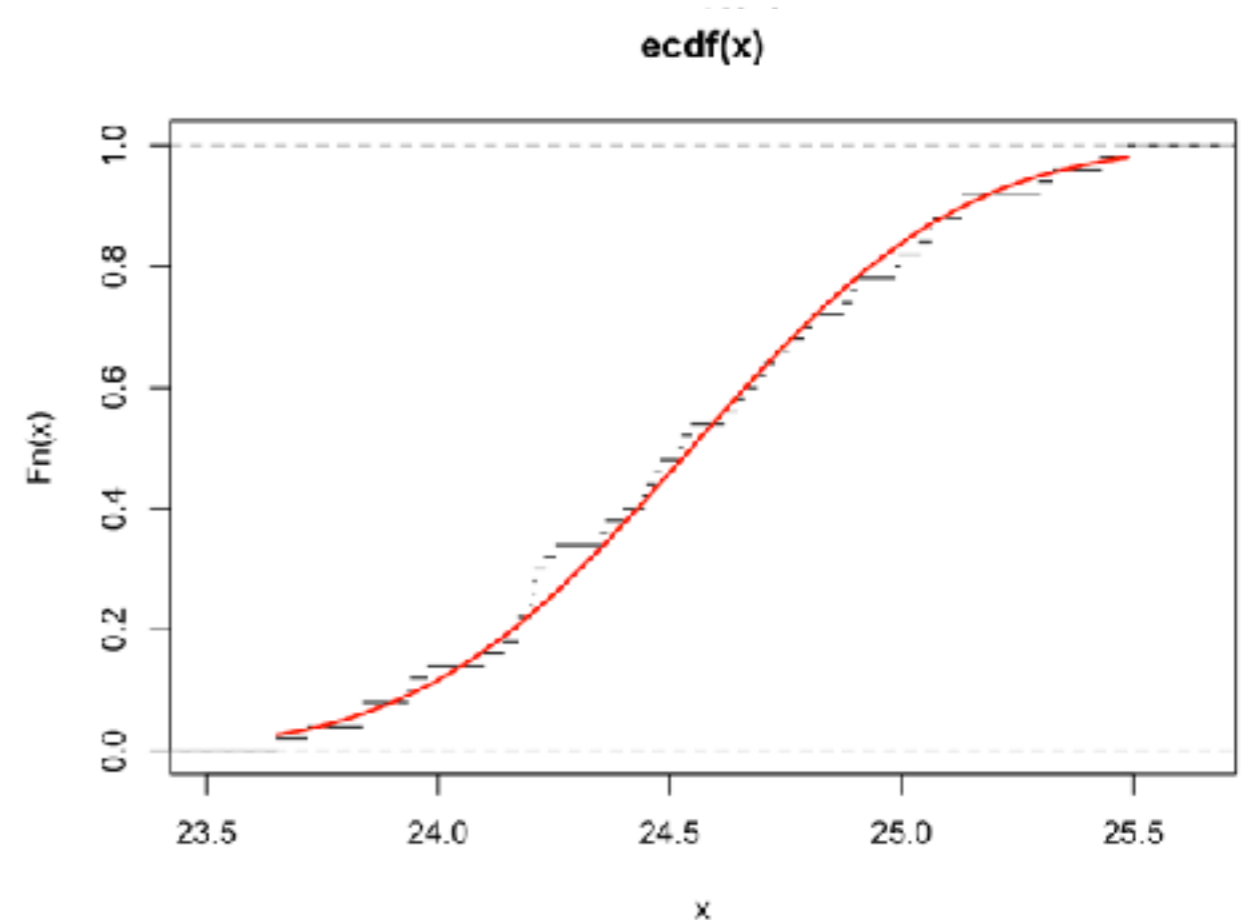
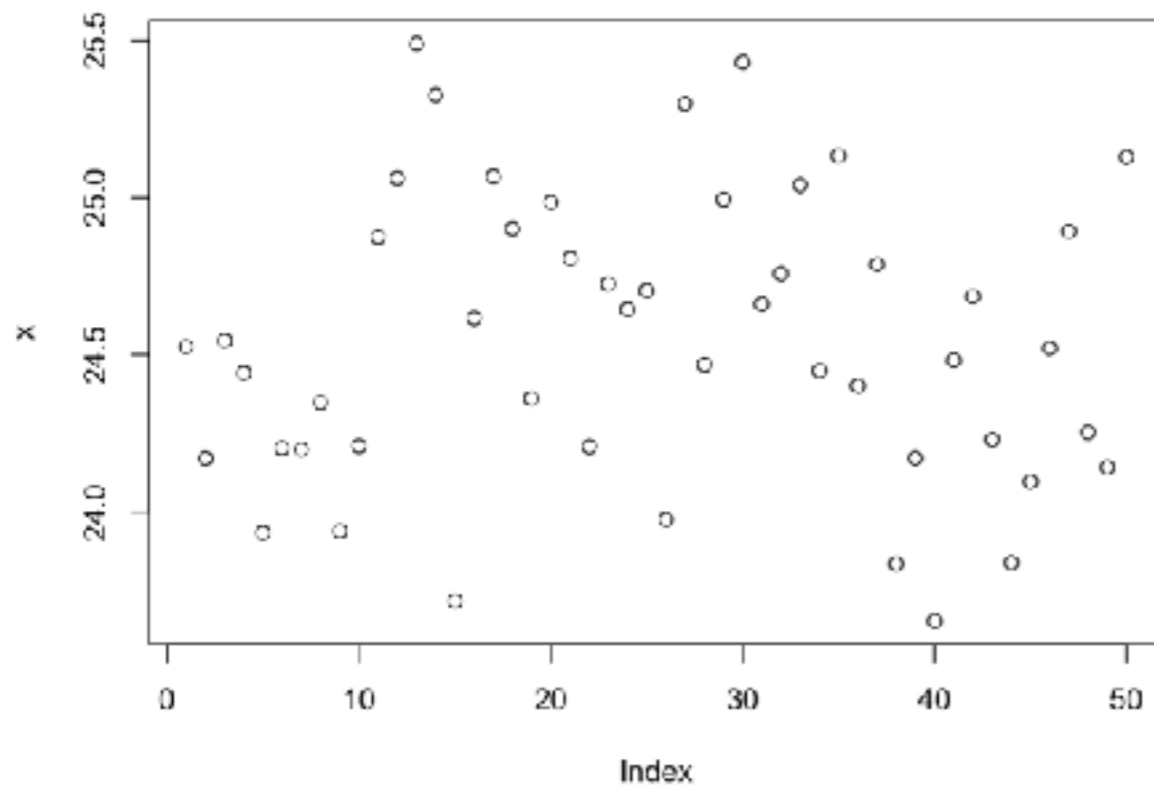
```
> x  
v1  
[1,] 24.52586  
[2,] 24.17119  
[3,] 24.54486  
[4,] 24.44240  
[5,] 23.93455  
[6,] 24.20389  
[7,] 24.19974  
[8,] 24.34851  
[9,] 23.94024  
[10,] 24.21022  
[11,] 24.87474  
[12,] 25.06155  
[13,] 25.48924  
[14,] 25.32572  
[15,] 23.71721  
[16,] 24.61622  
[17,] 25.06676  
[18,] 24.90055  
[19,] 24.36213  
[20,] 24.98580  
[21,] 24.80591  
[22,] 24.20853  
[23,] 24.72623  
[24,] 24.64437  
[25,] 24.70405  
[26,] 23.97645  
[27,] 25.29837  
[28,] 24.46910  
[29,] 24.99453  
[30,] 25.42994  
[31,] 24.66147  
[32,] 24.75773  
[33,] 25.03970  
[34,] 24.44901  
[35,] 25.13285  
[36,] 24.40205  
[37,] 24.78721  
[38,] 23.83656  
[39,] 24.17186  
[40,] 23.65390  
[41,] 24.48244  
[42,] 24.68550  
[43,] 24.22988  
[44,] 23.83956  
[45,] 24.09777  
[46,] 24.52098  
[47,] 24.89240  
[48,] 24.25332  
[49,] 24.14259  
[50,] 25.12906  
> |
```

```
24.52586  
24.17119  
24.54486  
24.44240  
23.93455  
24.20389  
24.19974  
24.34851  
23.94024  
24.21022  
24.87474  
25.06155  
25.48924  
25.32572  
23.71721  
24.61622  
25.06676  
24.90055  
24.36213  
24.98580  
24.80591  
24.20853  
24.72623  
24.64437  
24.70405  
23.97645  
25.29837  
24.46910  
24.99453  
25.42994  
24.66147  
24.75773  
25.03970  
24.44901  
25.13285  
24.40205  
24.78721  
23.83656  
24.17186  
23.65390  
24.48244  
24.68550  
24.22988  
23.83956  
24.09777  
24.52098  
24.89240  
24.25332  
24.14259  
25.12906
```

Statistické charakteristiky

Příklad: Balící automat na kávu - popisná statistika v programu R:

> plot(x)



- > ecdf(x)
- > plot(ecdf(x), verticals = FALSE, do.points="FALSE")
- > xfit<-seq(min(x),max(x),length=40)
- > zfit<-pnorm(xfit,mean=mean(x),sd=sd(x))
- > lines(xfit, zfit, col="red", lwd=2)



Statistické charakteristiky

Příklad: Balící automat na kávu - popisná statistika v programu R:

```
> summary(x)
```

```
      V1
```

```
Min.   :23.65
```

```
1st Qu.:24.21
```

```
Median :24.52
```

```
Mean   :24.55
```

```
3rd Qu.:24.89
```

```
Max.   :25.49
```

```
> skewness(x)
```

```
      V1
```

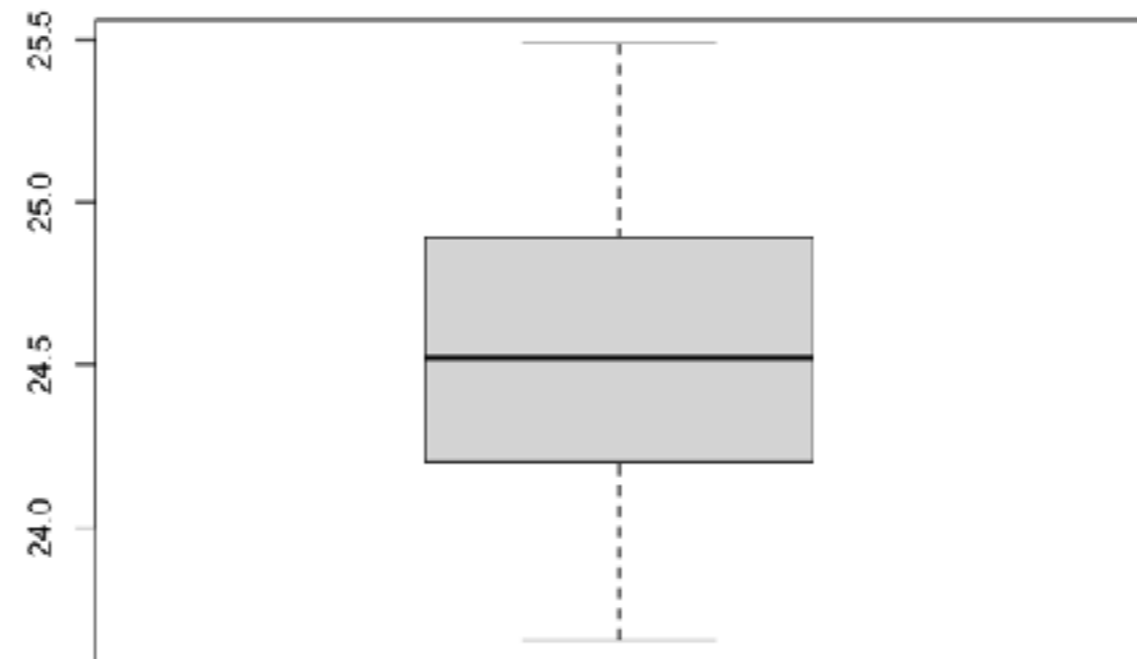
```
0.09303432
```

```
> kurtosis(x)
```

```
      V1
```

```
2.27457
```

```
> boxplot(x)
```



Statistické charakteristiky

Příklad: Balící automat na kávu - popisná statistika v programu R:

```
> summary(x)
```

```
      V1
```

```
Min.   :23.65
```

```
1st Qu.:24.21
```

```
Median :24.52
```

```
Mean   :24.55
```

```
3rd Qu.:24.89
```

```
Max.   :25.49
```

```
> skewness(x)
```

```
      V1
```

```
0.09303432
```

```
> kurtosis(x)
```

```
      V1
```

```
2.27457
```

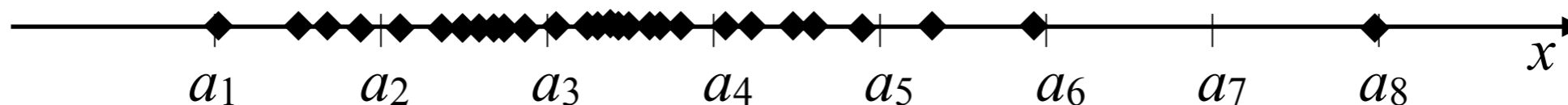
```
> boxplot(x, main="Hmotnost balíčků  
dávkových automatem",  
ylab="hmotnost v dag")(x)
```



Frekvenční analýza - analýza četností

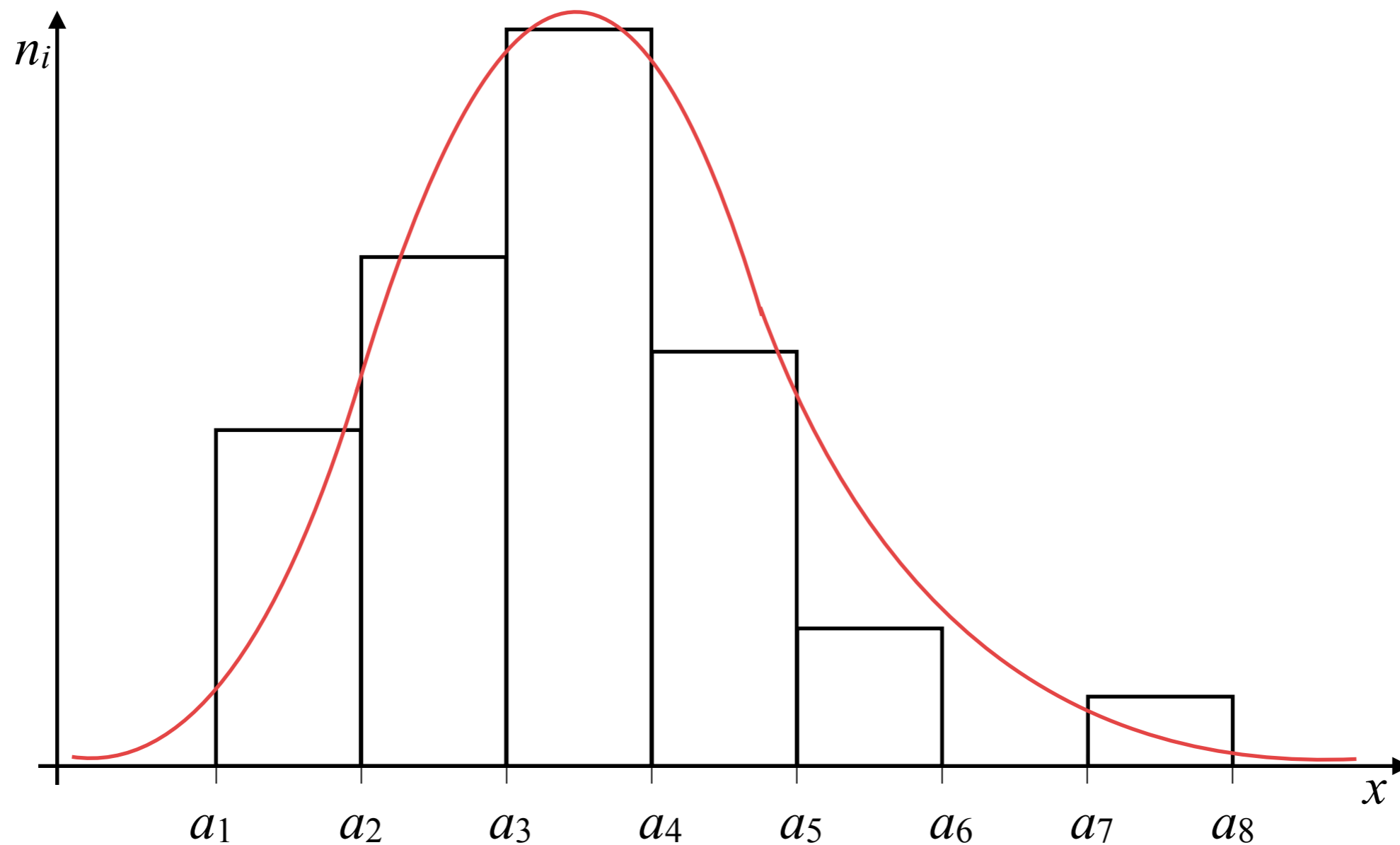
Máme pozorování x_1, x_2, \dots, x_n náhodného výběru X_1, X_2, \dots, X_n .

pořadí třídy	třídní intervaly	(prosté) absolutní četnosti	(prosté) relativní četnosti	kumulativní četnosti	kumulativní relativní četnosti
1	$a_2 - a_1$	n_1	$f_1 = n_1/n$	$c_1 = n_1$	$d_1 = c_1/n$
2	$a_3 - a_2$	n_2	$f_2 = n_2/n$	$c_2 = n_1 + n_2$	$d_2 = c_2/n$
⋮	⋮	⋮	⋮	$c_j = \sum_{i=1}^j n_i$	⋮
k	$a_k - a_{k-1}$	n_k	$f_k = n_k/n$	$c_k = n$	$d_k = c_k/n = 1$



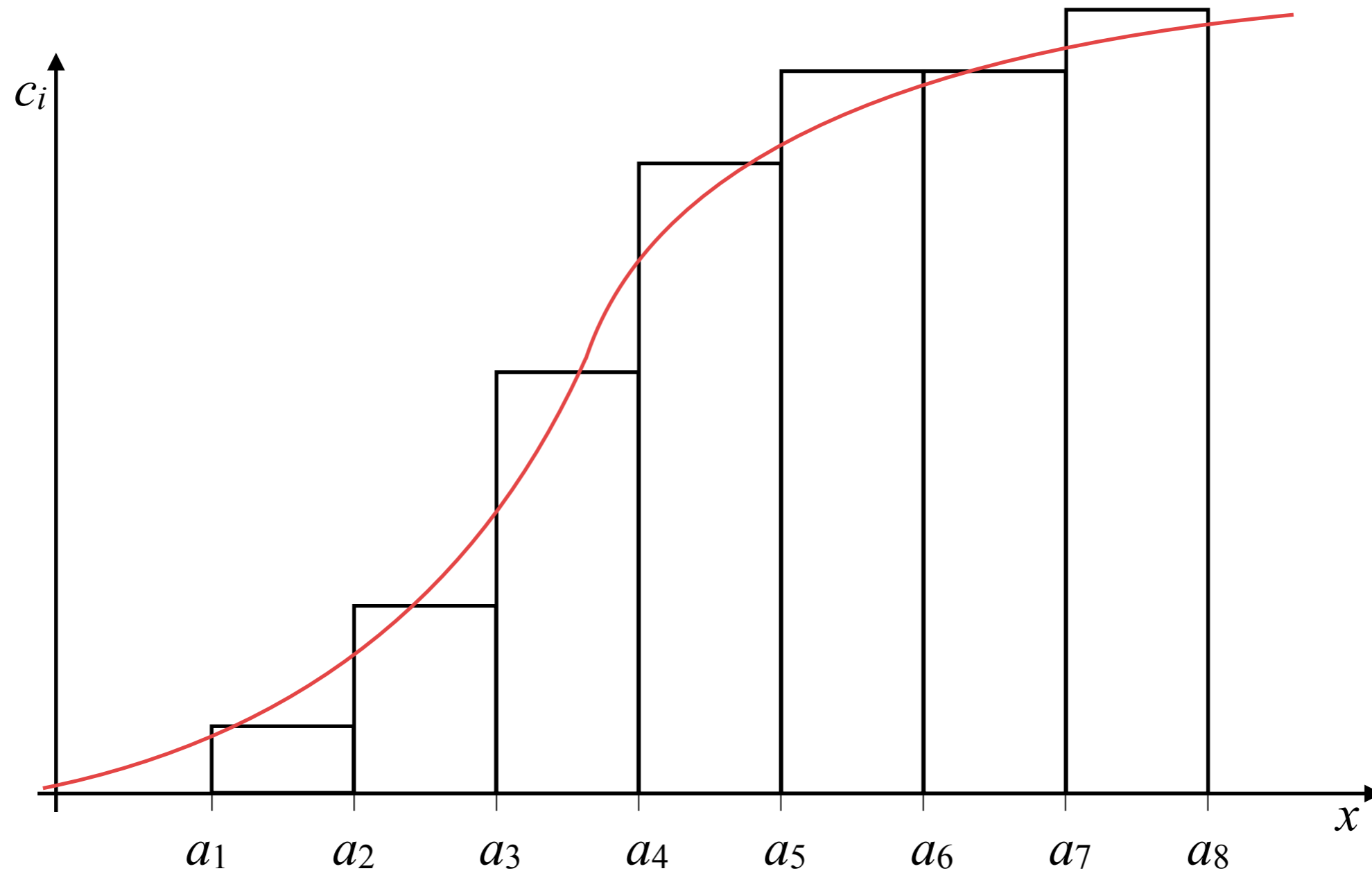
Frekvenční analýza - histogram četností

Máme pozorování x_1, x_2, \dots, x_n náhodného výběru X_1, X_2, \dots, X_n .



Frekvenční analýza - histogram četností

Máme pozorování x_1, x_2, \dots, x_n náhodného výběru X_1, X_2, \dots, X_n .



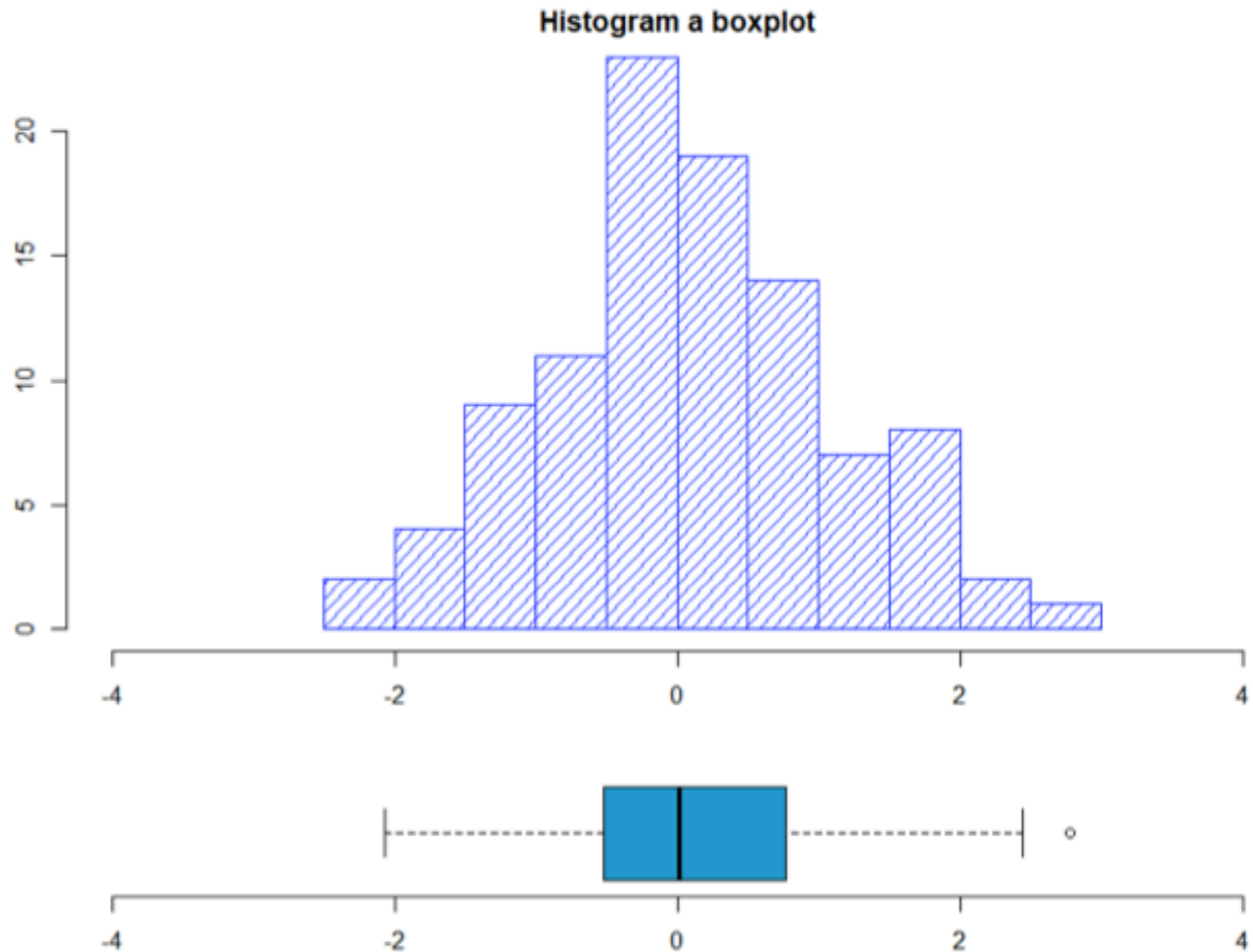
Frekvenční analýza - analýza četností

Co je třeba znát a porozumět tomu:

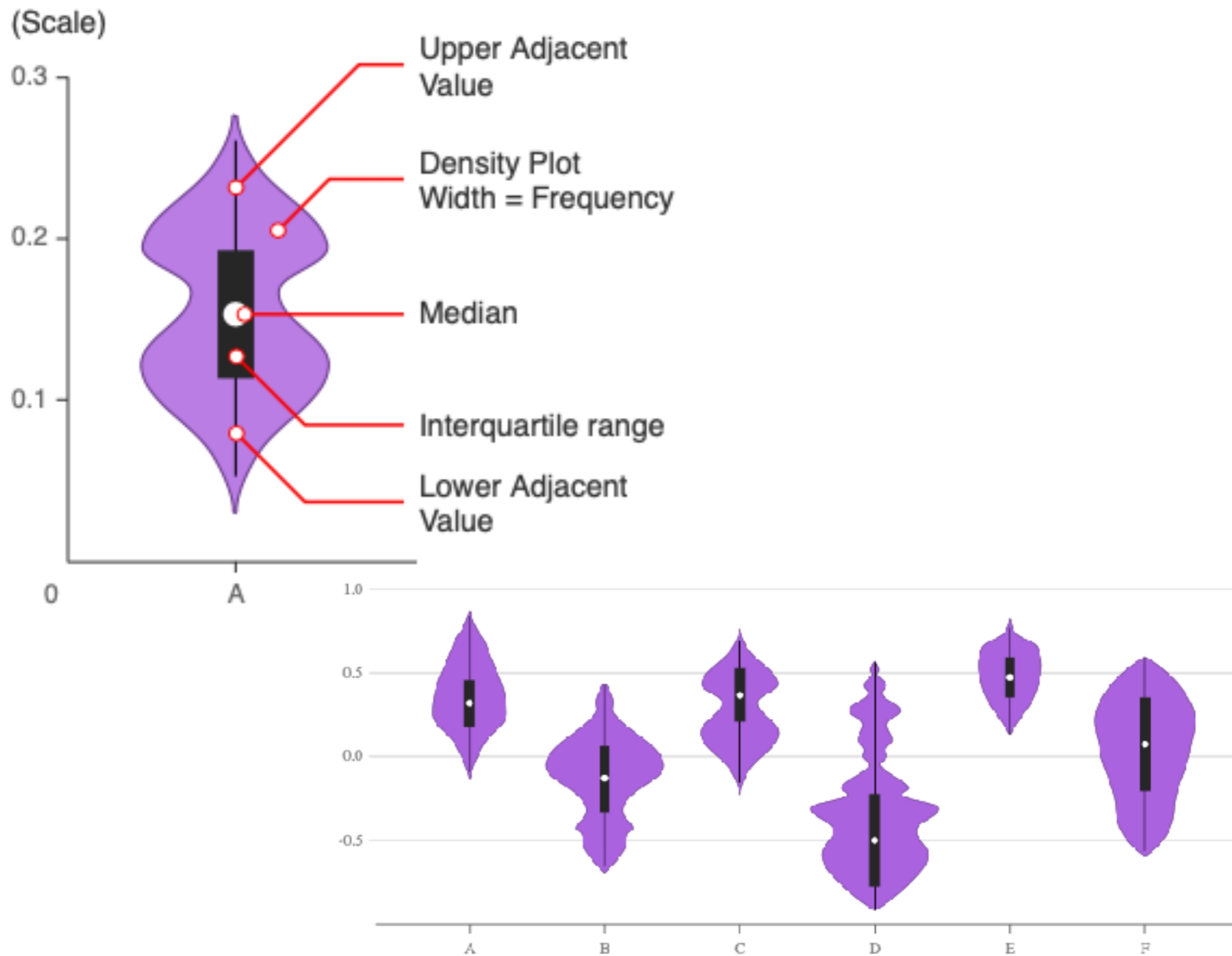
- **Třídní intervaly** rozdělují maximální rozsah pozorovaných hodnot náhodné veličiny (od minima do maxima) na k stejných dílů.
- **(prostá absolutní) četnost i -té třídy** je počet pozorování náhodné veličiny X v i -té třídě, $i = 1, \dots, k$.
- **(prostá) relativní četnost i -té třídy** je poměr počtu pozorování náhodné veličiny X v i -té třídě ku rozsahu výběru n , $i = 1, \dots, k$.
- **kumulativní (absolutní) četnost i -té třídy** je počet pozorování náhodné veličiny X od minima až do i -té třídy včetně, $i = 1, \dots, k$.
- **kumulativní relativní četnost i -té třídy** je součet relativních četností pozorování náhodné veličiny X až do i -té třídy včetně, $i = 1, \dots, k$.
- **Histogram četností** je grafické zobrazení četností ve formě sloupkového grafu. Relativní četnosti lze zobrazovat i ve formě kruhového (koláčového) grafu. Existuje celá řada variant.



Frekvenční analýza - histogram četností

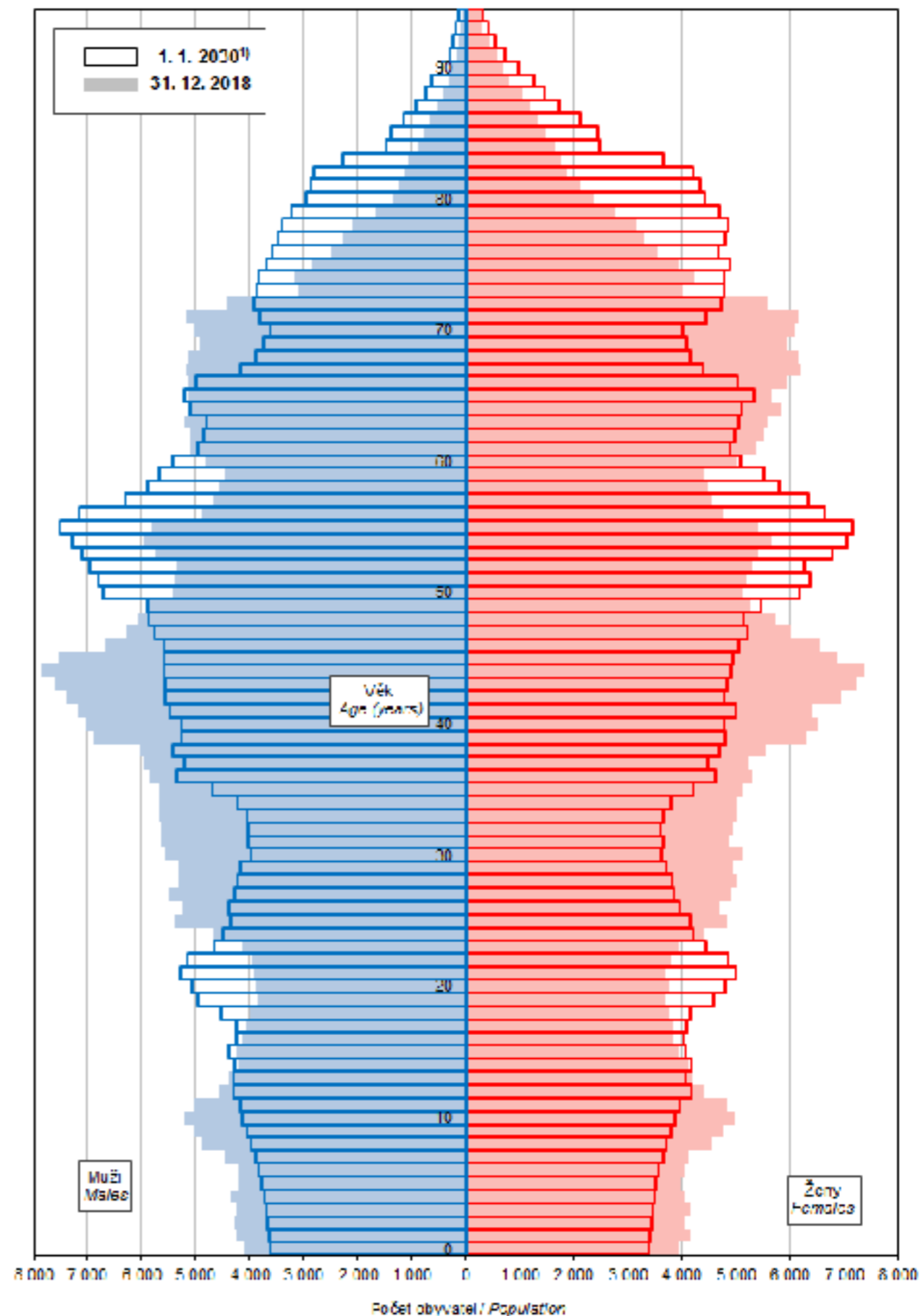


Frekvenční analýza - huoslový graf



Frekvenční analýza - dvourozměrný histogram

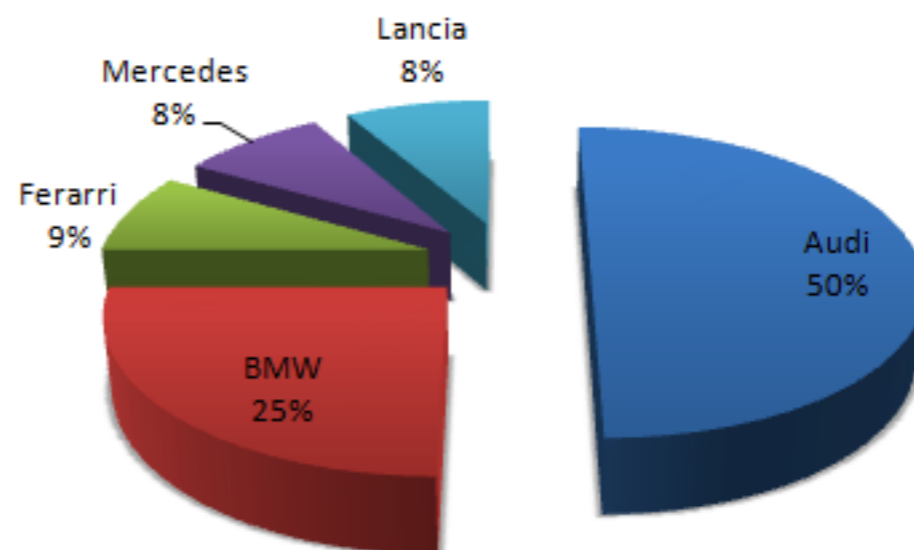
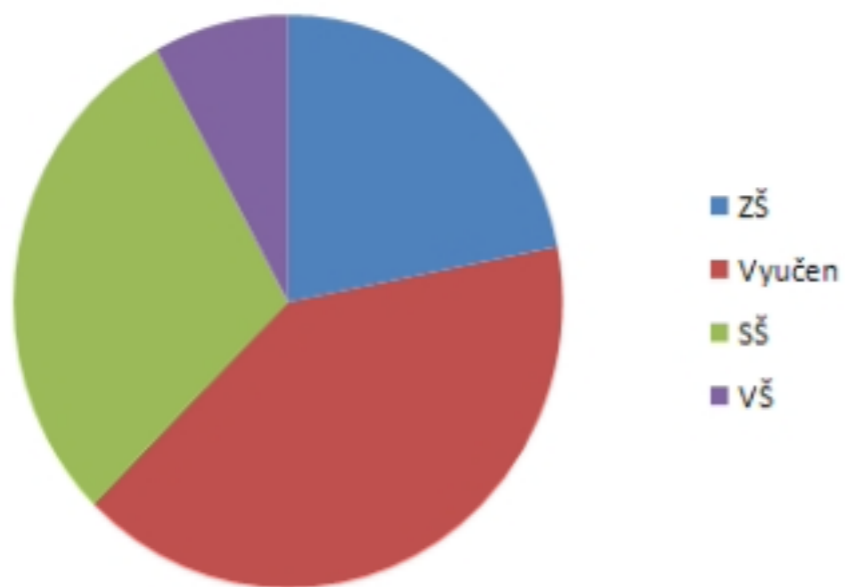
Věkové složení obyvatelstva Ústeckého kraje k 31. 12. 2018 a k 1. 1. 2030
Age distribution of the population in the Ústecký Region as at 31 December 2018 and 1 January 2030



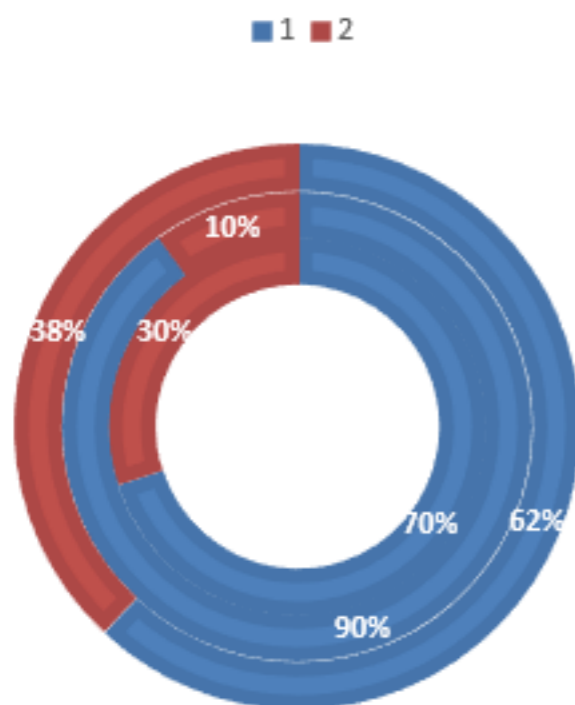
¹⁾ Zdroj: Projekce obyvatelstva v krajích ČR do roku 2070
²⁾ Source: CZSO publication "Projekce obyvatelstva v krajích ČR do roku 2070" (Czech only)



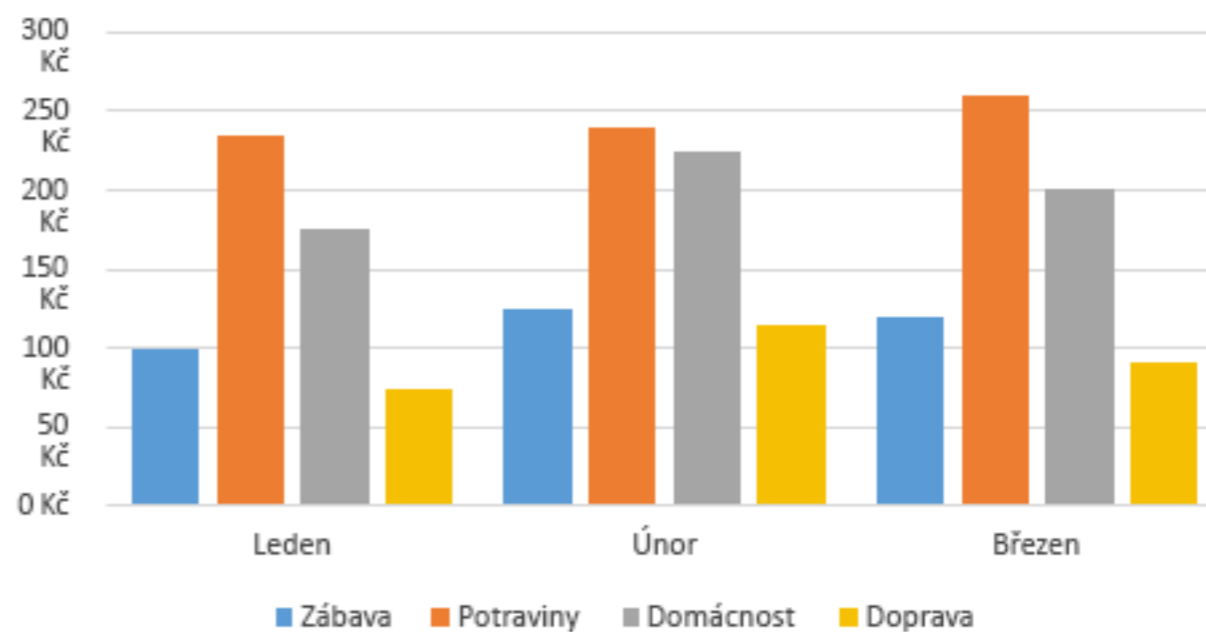
Frekvenční analýza - kruhový graf



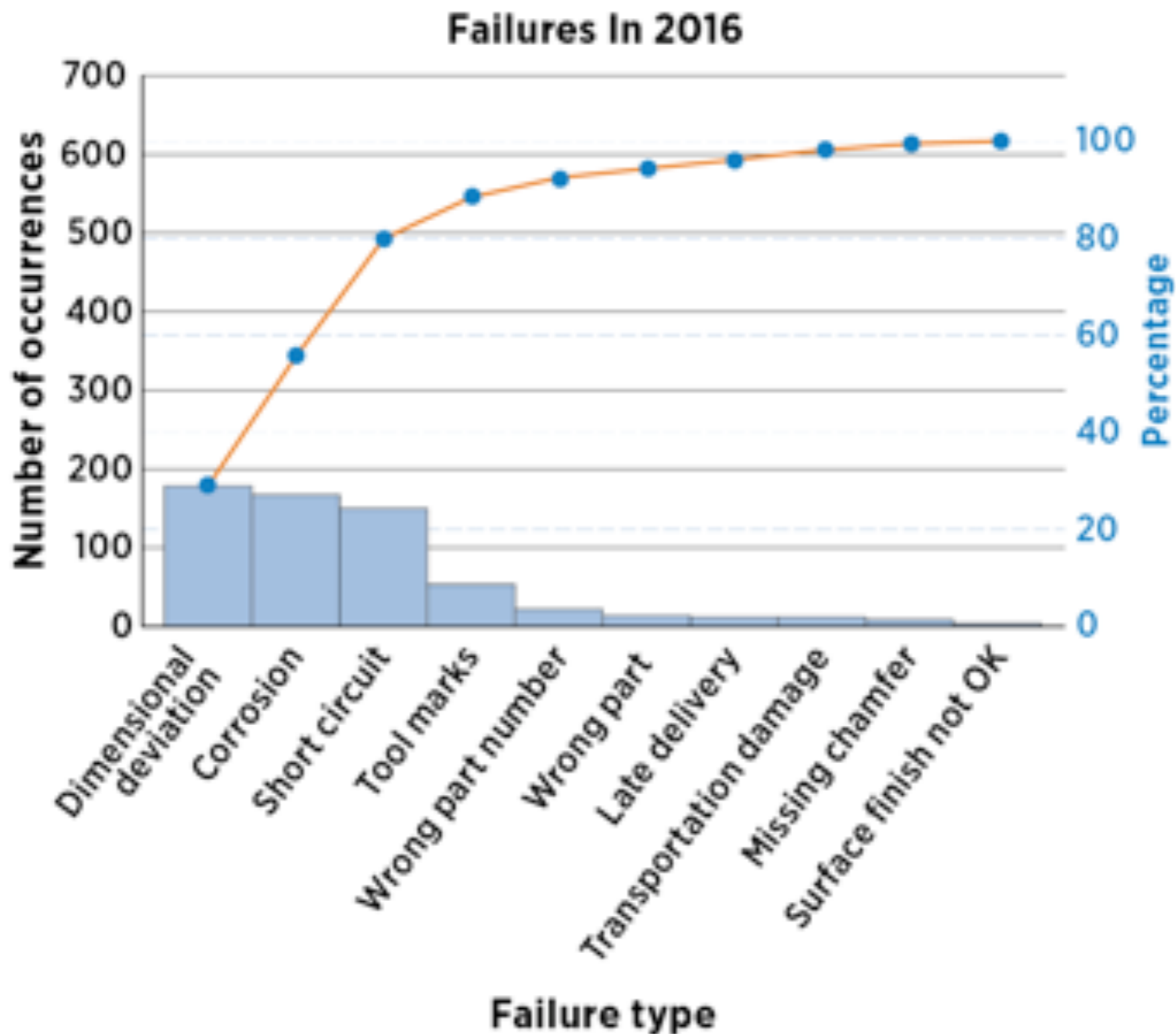
PRSTENCOVÝ GRAF – 3 HODNOTY



KONTINGENČNÍ GRAF



Frekvenční analýza - Paretův graf



Frekvenční analýza

Příklad: Balící automat na kávu

```
> hist(x)  
> hx <- hist(x)  
> print(hx)
```

\$breaks

```
[1] 23.6 23.8 24.0 24.2 24.4 24.6  
    24.8 25.0 25.2 25.4 25.6
```

\$counts

```
[1] 2 5 5 7 8 8 6 5 2 2
```

\$density

```
[1] 0.2 0.5 0.5 0.7 0.8 0.8 0.6 0.5 0.2 0.2
```

\$mids

```
[1] 23.7 23.9 24.1 24.3 24.5 24.7 24.9 25.1 25.3 25.5
```

\$xname

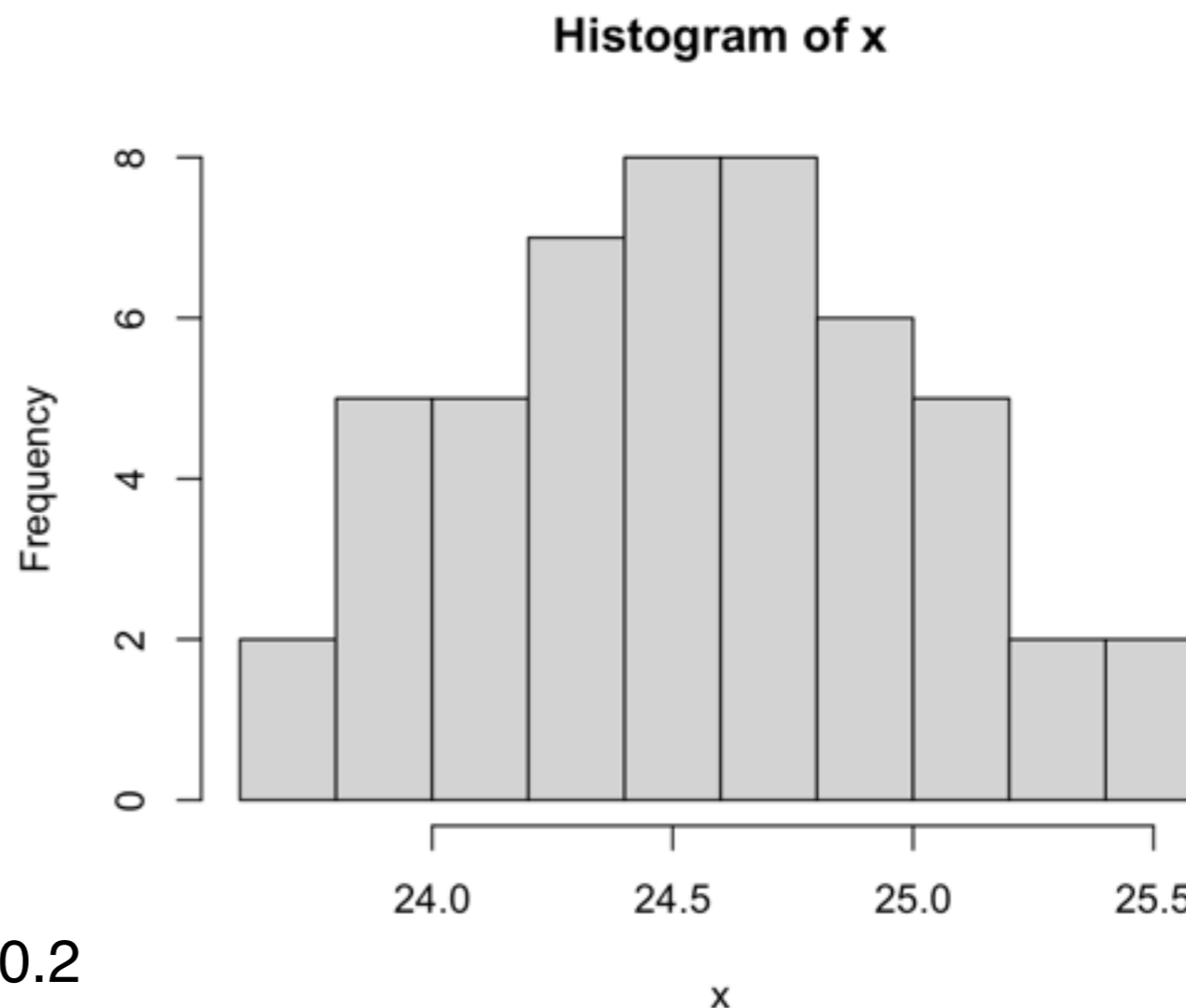
```
[1] "x"
```

\$equidist

```
[1] TRUE
```

attr("class")

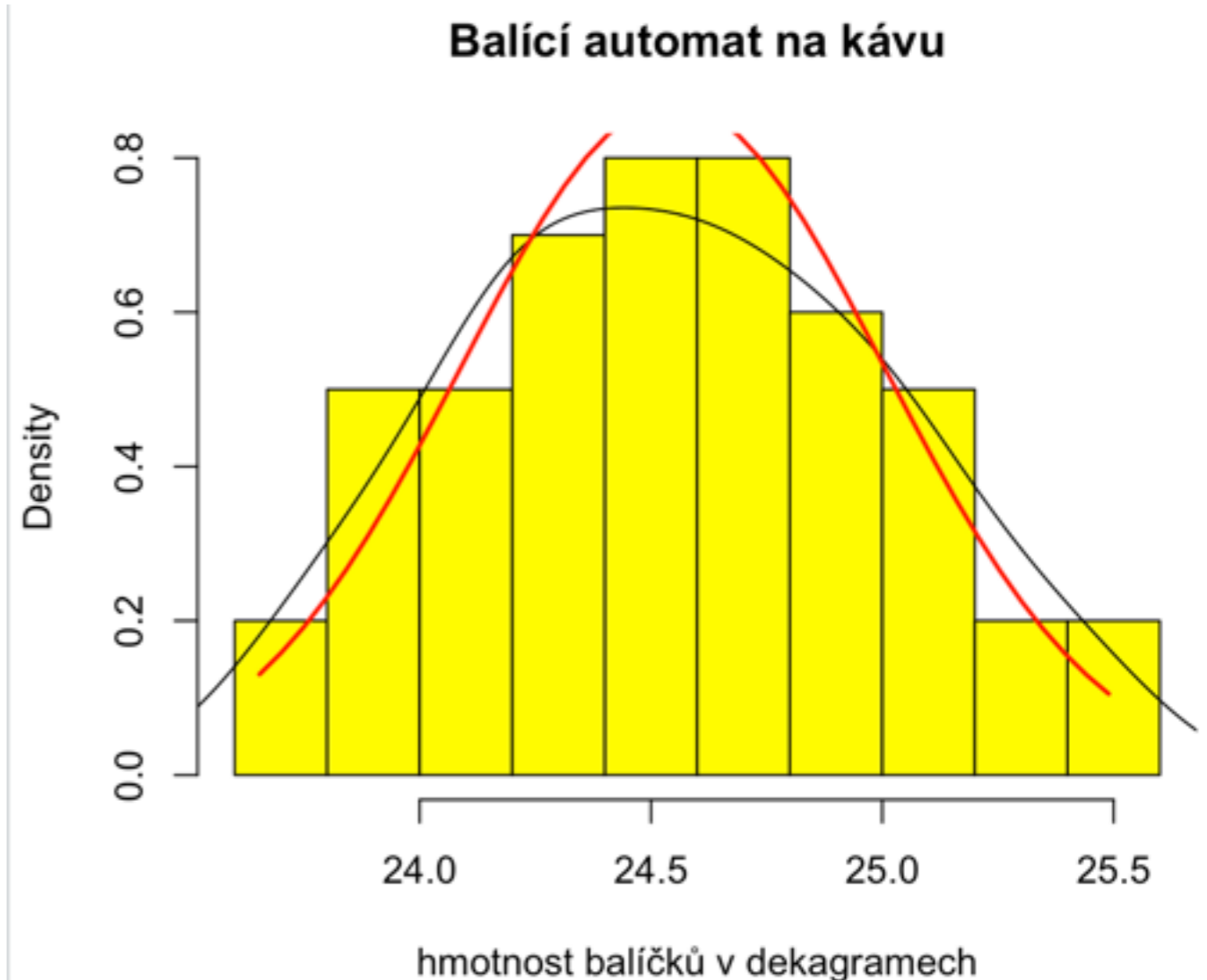
```
[1] "histogram"
```



Frekvenční analýza

Příklad: Balící automat na kávu

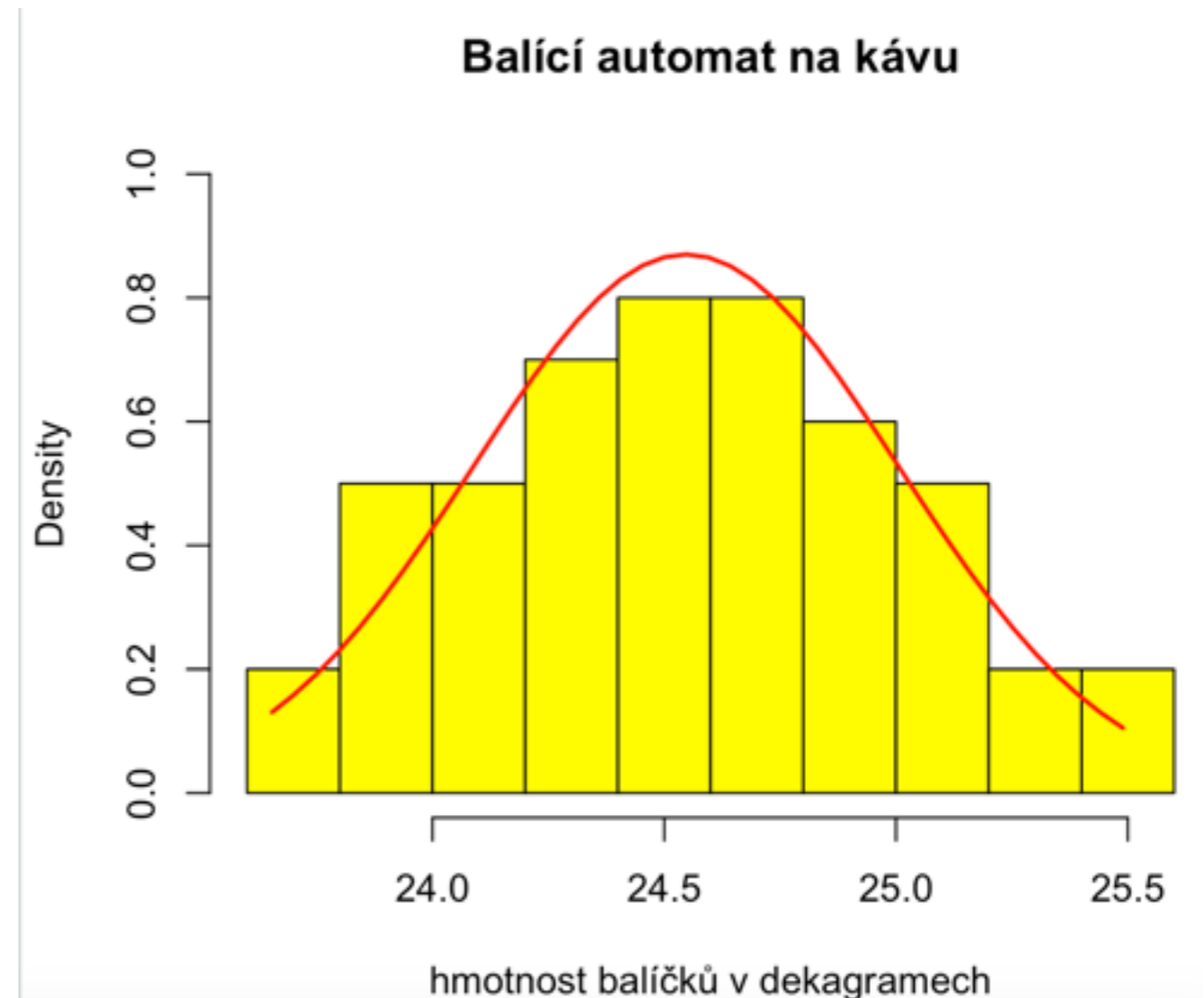
- > `h <- hist(x, main="Balící automat na kávu", xlab="hmotnost balíčků v dekagramech", col="yellow", freq=FALSE, breaks = 8)`
- > `lines(density(x))`
- > `xfit<-seq(min(x),max(x),length=40)`
- > `yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))`
- > `lines(xfit, yfit, col="red", lwd=2)`



Frekvenční analýza

Příklad: Balící automat na kávu

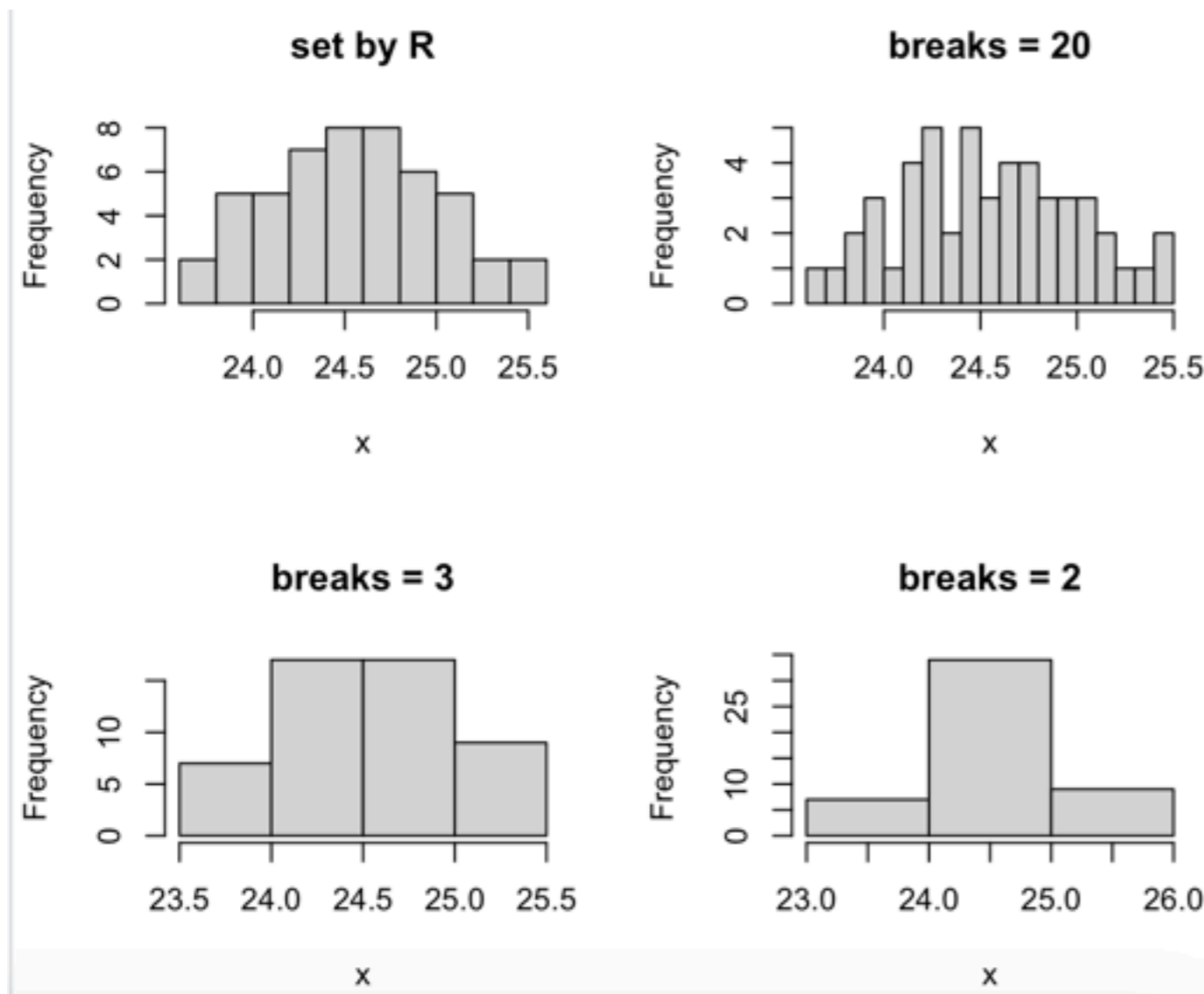
```
> h <- hist(x, ylim=c(0,1),  
  main="Balící automat na kávu",  
  xlab="hmotnost balíčků  
  v dekagramech", col="yellow",  
  freq=FALSE, breaks = 8)  
  
> xfit<-seq(min(x),max(x),length=40)  
> yfit<-dnorm(xfit,mean=mean(x),  
  sd=sd(x))  
> lines(xfit, yfit, col="red", lwd=2)
```



Frekvenční analýza

Příklad: Balící automat na kávu

```
> par(mfrow = c(2,2))  
> hist(x, main="set by R");  
hist(x, breaks = 20, main="breaks = 20");  
hist(x, breaks = 3, main="breaks = 3");  
hist(x, breaks = 2, main="breaks = 2")
```



Pravidla pro stanovení počtu tříd:

- Sturgesovo pravidlo:
 $1 + 3, 32 \log n$
- Yuleovo pravidlo:
 $k \approx 2, 5 \sqrt[4]{n}$
- obecně:
 $k \approx \sqrt{n}; k \leq 5 \log n$

$n=50 \Rightarrow$ Sturges: 6.64
Yule: 6.65

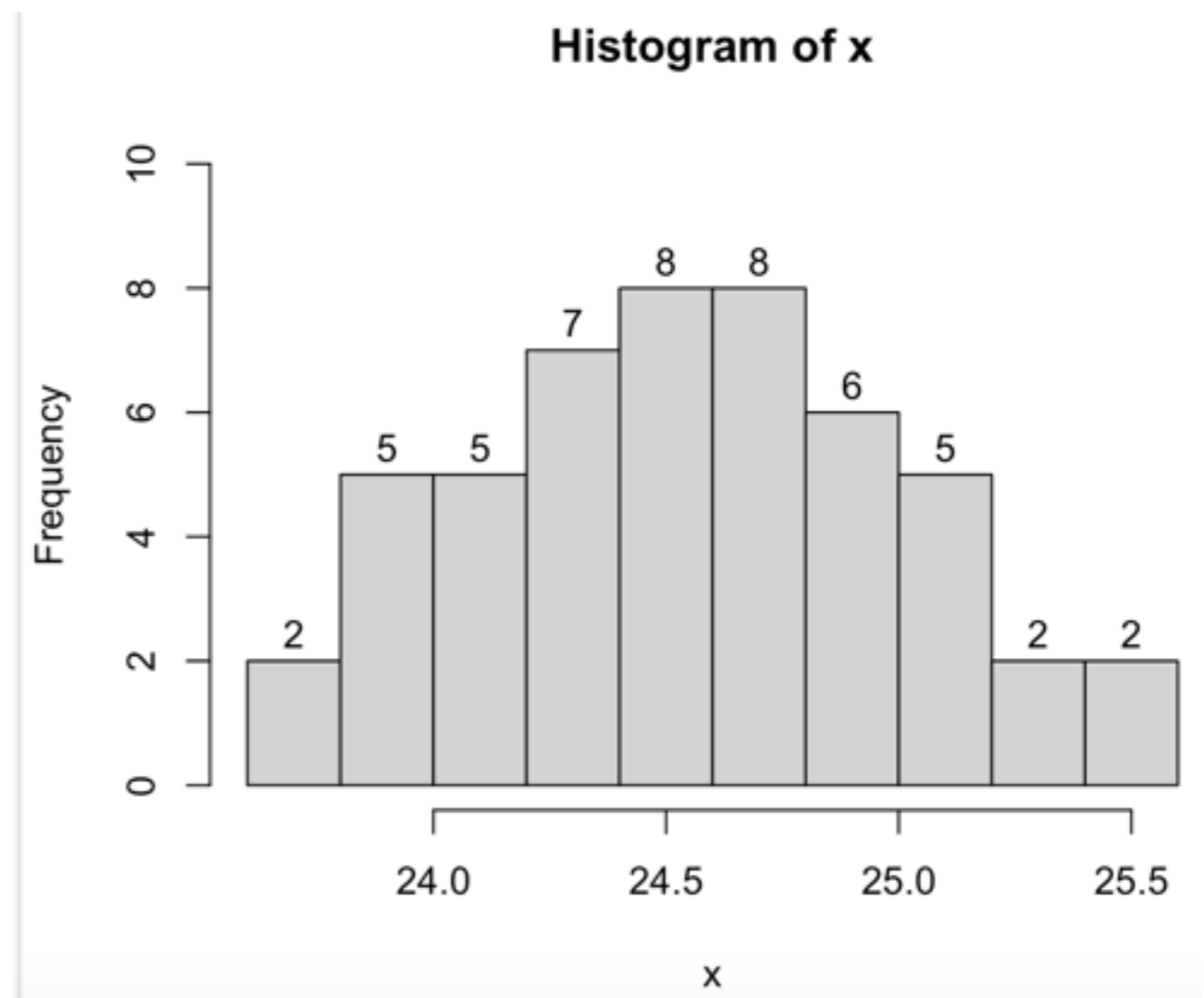


Frekvenční analýza

Příklad: Balící automat na kávu

```
> h <- hist(x,ylim=c(0,10))
```

```
> text(h$mids,h$counts,labels=h$counts, adj=c(0.5, -0.5))
```



Frekvenční analýza

Příklad: Balící automat na kávu

- > `h <- hist(x,ylim=c(0,10))`
- > `h$counts<-cumsum(h$counts)`
- > `plot(h)`
- > `xfit<-seq(min(x),max(x),length=40)`
- > `zfit<-50*pnorm(xfit,mean=mean(x),sd=sd(x))`
- > `lines(xfit, zfit, col="red", lwd=2)`

