DATA → INFORMACE

Statistická analýza je založena na **zhušťování informace** – tj. jak s co nejmenšího množství vhodně zvolených údajů vytěžit maximum relevantních informací.

1. **prvotní zápis** – údaje v té podobě, jak jsou naměřeny
2. **tříděný soubor** – jednotlivá měřená data jsou tříděna do tříd
3. **statistické charakteristiky** – speciální veličiny, které **podávají koncentrovanou formou informaci o podstatných statistických vlastnostech** studovaného souboru

ZHUŠŤOVÁNÍ INFORMACE

STATISTICKÉ CHARAKTERISTIKY

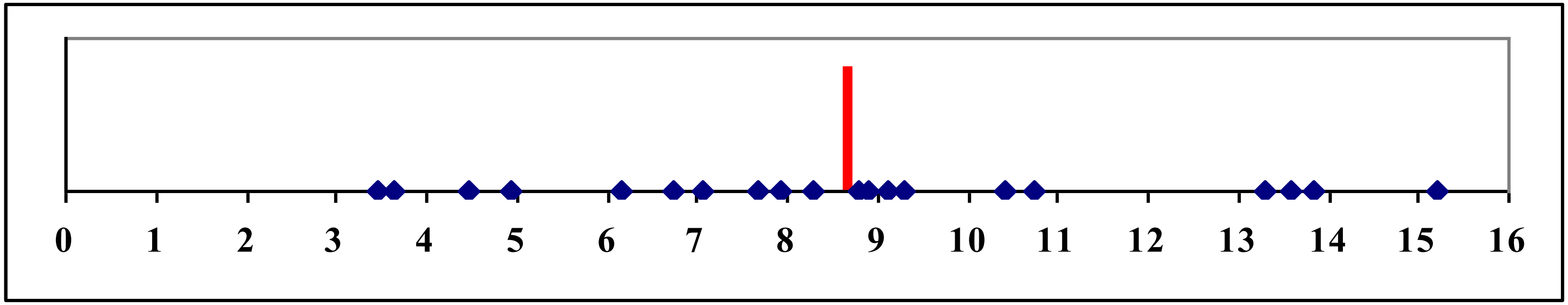
Statistické charakteristiky

- **polohy**
  - **momentové**
  - **kvantilové**
- **variability**
  - **momentové**
  - **kvantilové**
- **tvaru**
  - **momentové**
  - **kvantilové**

STATISTICKÉ CHARAKTERISTIKY

Typy charakteristik:

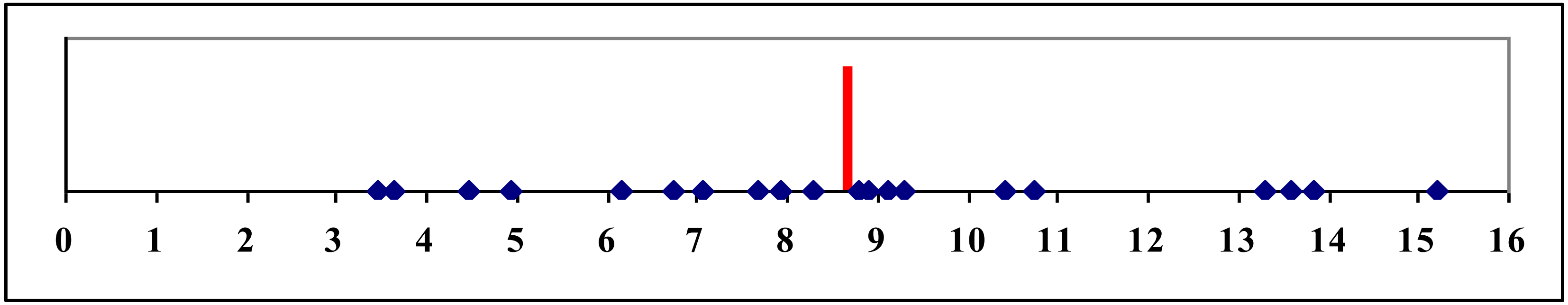
1. polohy – **reprezentace souboru na číselné ose**

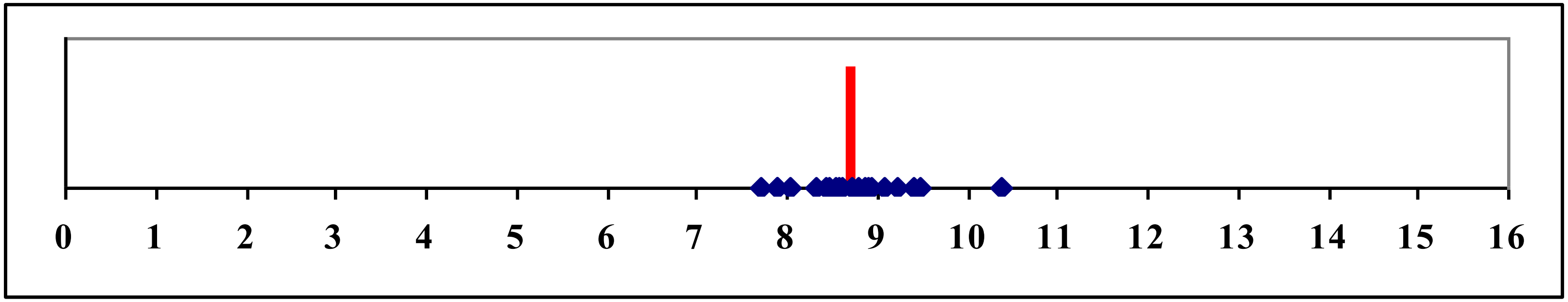


STATISTICKÉ CHARAKTERISTIKY

Typy charakteristik:

2. variability – rozptýlení hodnot po číselné ose navzájem a vůči charakteristice polohy

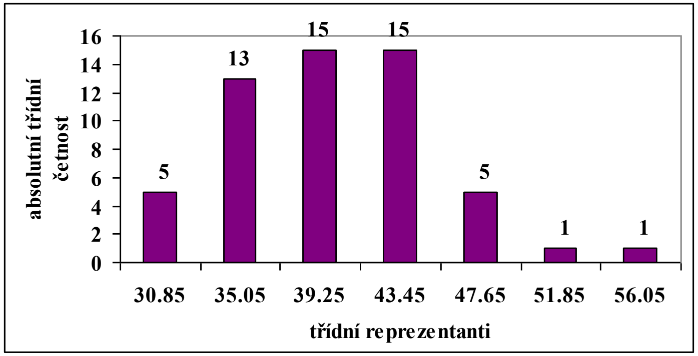


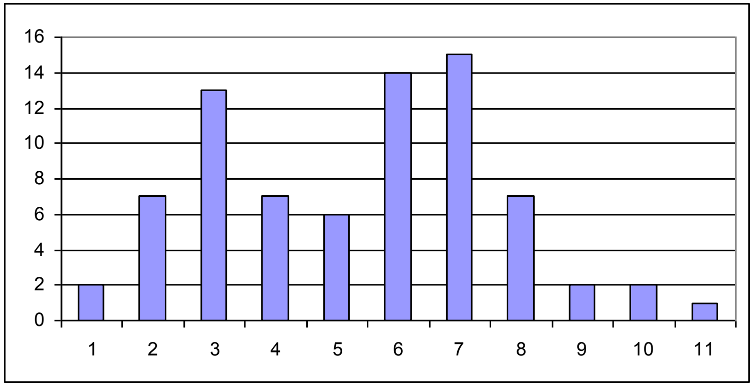


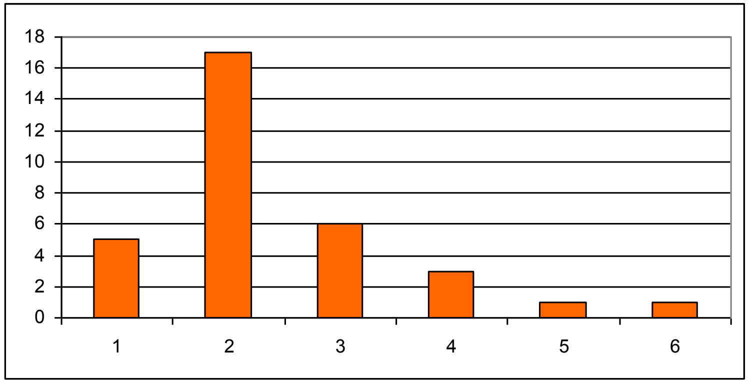
STATISTICKÉ CHARAKTERISTIKY

Typy charakteristik:

3. tvaru – rozložení četností hodnot







CHARAKTERISTIKY POLOHY

- **ARITMETICKÝ PRŮMĚR – hodnota reprezentující všechny hodnoty souboru s nejmenší chybou**
- **MEDIÁN – 50% kvantil, prostřední hodnota vzestupně uspořádaného souboru**
- **MODUS – nejčastěji se vyskytující hodnota v souboru**

ARITMETICKÝ PRŮMĚR

- základní statistická **MOMENTOVÁ** charakteristika polohy
- je to hodnota, která reprezentuje **VŠECHNY** hodnoty souboru s nejmenší chybou
- fyzikálně je možné jej považovat za těžiště souboru

$$\bar{x}_1 = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{m} n_i \cdot \bar{x}_i}{N}$$

MEDIÁN

- základní statistická **KVANTILOVÁ** charakteristika polohy
- je to hodnota, která reprezentuje **PROSTŘEDNÍ PRVEK VZESTUPNĚ USPOŘÁDANÉHO SOUBORU**

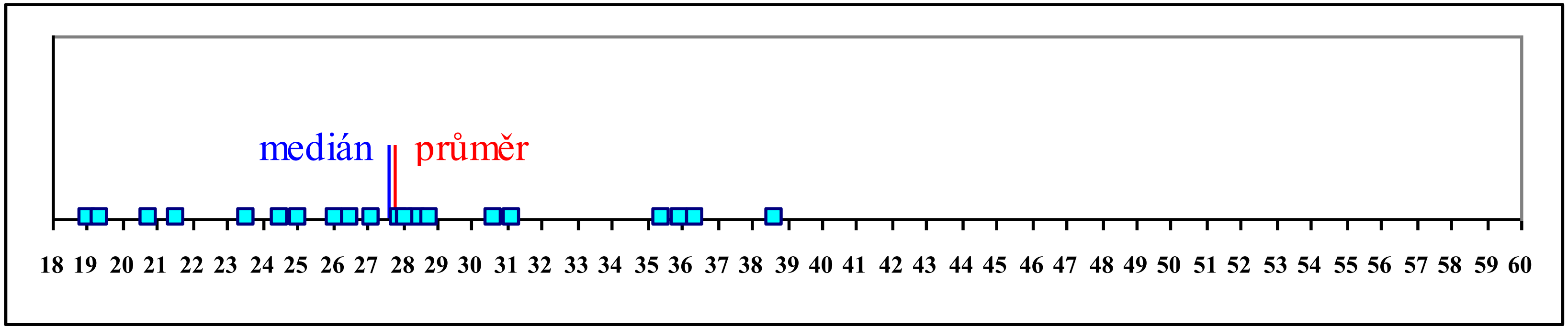
$$\tilde{x} = \begin{cases} x_{\left(\frac{N+1}{2}\right)} & \text{pro N liché} \\ \frac{1}{2} \cdot \left(x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)} \right) & \text{pro N sudé} \end{cases}$$

MODUS

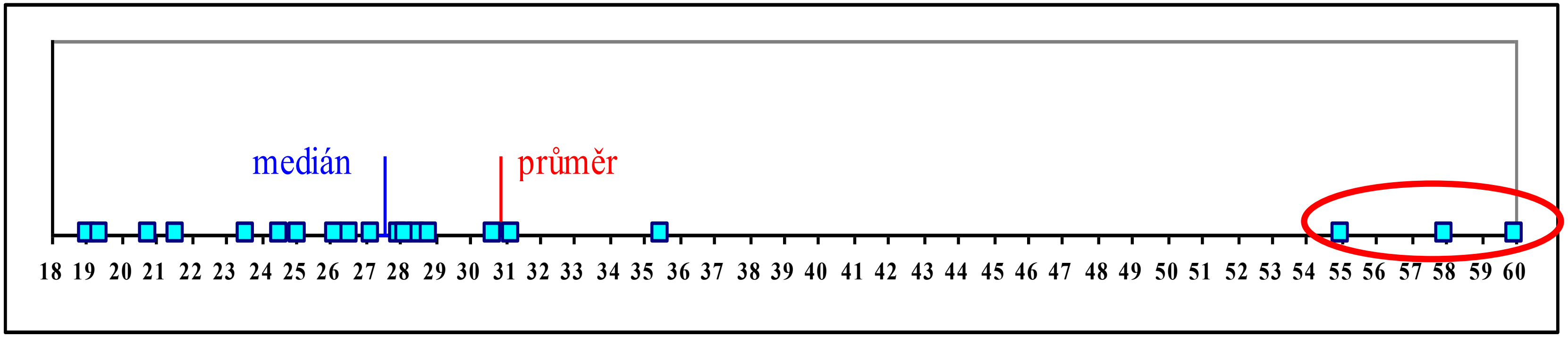
- **nejčastěji se vyskytující hodnota souboru**
- **existují soubory:**
  - **amodální** – bez modu (všechny prvky souboru mají stejnou četnost)
  - **unimodální** – jeden modus
  - **polymodální** – dva a více modů
- **nemá příliš velkou vypovídací schopnost**

POUŽITÍ PRŮMĚRU A MEDIÁNU

Soubor bez extrémních hodnot:



Soubor s extrémními hodnotami:



CHARAKTERISTIKY VARIABILITY

- informují o tom, jak jsou jednotlivé hodnoty souboru **rozptýleny**, tj. jak se jednotlivé hodnoty znaku **liší vzhledem k sobě navzájem** nebo **vzhledem ke střední hodnotě**
- existují dva typy:
  - **absolutní** - mají rozměr studované veličiny
  - **relativní (poměrné)** - bez rozměru nebo v procentech.

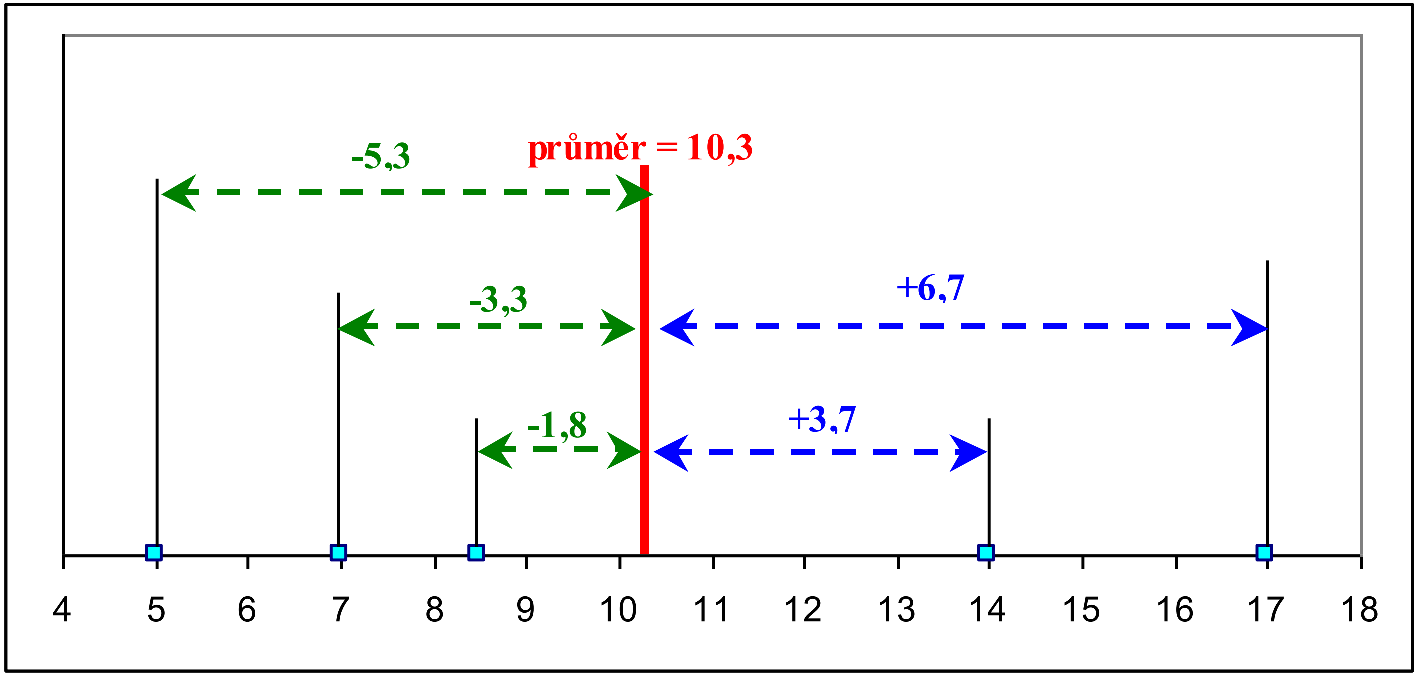
  Jsou vhodné pro porovnání variability různých souborů

CHARAKTERISTIKY VARIABILITY

- **variační rozpětí** – rozdíl maximální a minimální hodnoty
- **rozptyl** – základní momentová míra variability, průměr odchylek od průměru
- **směrodatná odchylka** – odmocnina z rozptylu, využívaná hlavně pro popis souborů
- **variační koeficient** – relativní míra variability užívaná ke srovnání variability různých souborů
- **kvantilové odchylky** – kvantilová míra variability počítaná obvykle z kvartilů nebo decilů
- **interkvartilové rozpětí** – rozdíl horního a dolního kvartilu

ROZPTYL

Rozptyl je základní mírou variability. Je to **aritmetický průměr čtverců odchylek od průměru** a je tedy konstruován k vyjádření variability hodnot kolem průměru, ale vyjadřuje i vzájemnou odlišnost hodnot znaku.



ROZPTYL

pro ZS:

$$\sigma^2 = \operatorname{var} X = \frac{\sum_{j=1}^{N}\left(x_j - \mu\right)^2}{N}$$

pro VS:

$$S^2 = \operatorname{var} X = \frac{\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2}{n-1}$$

pro tříděný soubor:

$$\mathrm{S}^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{n}_\mathrm{i}\left(\bar{\mathrm{x}}_\mathrm{i} - \bar{\mathrm{x}}\right)^2}{\mathrm{N}}$$

SMĚRODATNÁ ODCHYLKA

je odmocnina z rozptylu. **Rozměr** směrodatné odchylky **je stejný jako rozměr veličiny**, což je její hlavní výhodou oproti rozptylu pro účely popisné statistiky.

VARIAČNÍ KOEFICIENT

je **relativní mírou variability** a používá se **k vzájemnému porovnávání variability** různých souborů.

$$S\% = \frac{S}{\bar{x}} \cdot 100$$

KVANTILOVÉ MÍRY VARIABILITY

Kvantilové odchylky jsou horší mírou variability než momentové charakteristiky. **Používají se tam, kde nelze použít momentové charakteristiky** (silně nenormální rozdělení, výskyt extrémních hodnot, apod.)

Kvartilová odchylka:

$$Q = \frac{(\tilde{x}_{75} - \tilde{x}) + (\tilde{x} - \tilde{x}_{25})}{2} = \frac{\tilde{x}_{75} - \tilde{x}_{25}}{2}$$

Interkvartilové rozpětí:

$$R_F = \tilde{x}_{75} - \tilde{x}_{25}$$

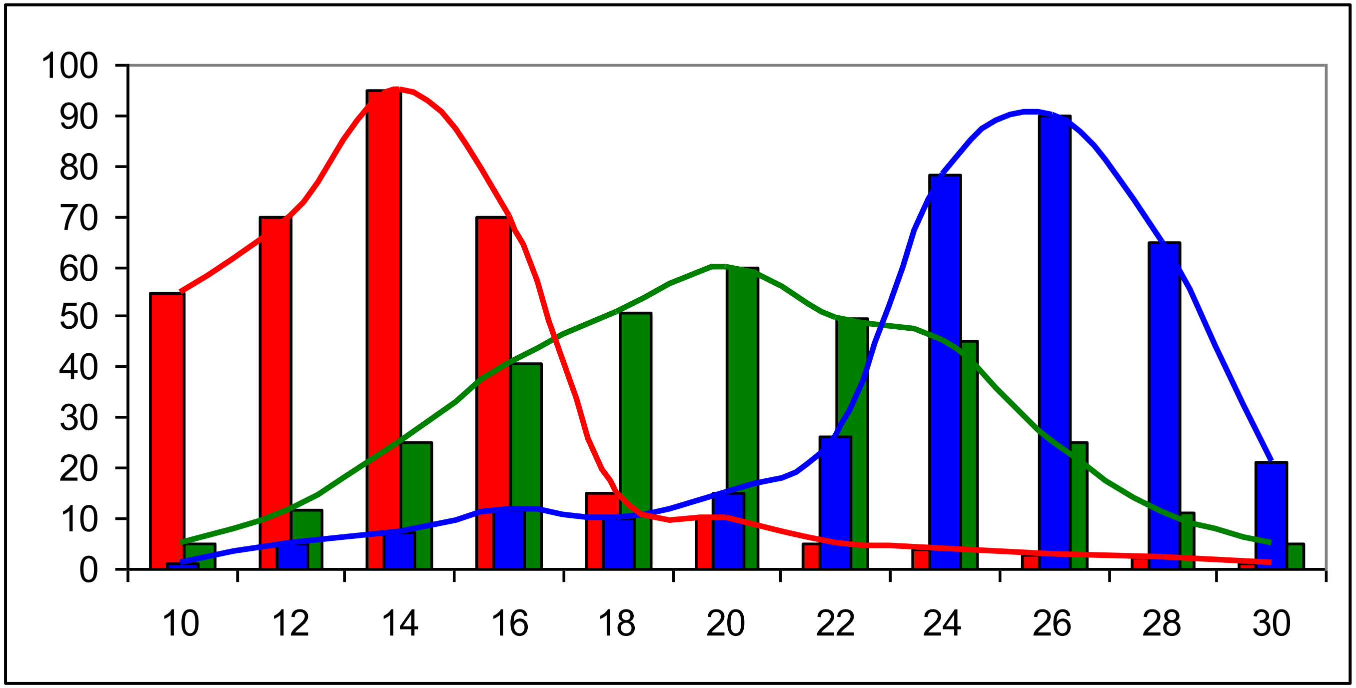
CHARAKTERISTIKY TVARU

měří odchylku v rozložení četností hodnot oproti danému referenčnímu rozdělení četností (obvykle normálnímu): Skládá se ze dvou složek:

- **nesouměrnosti (šikmosti, asymetrie)**
- **špičatosti (zahrocenosti, excesu)**

NESOUMĚRNOST

se projevuje tím, že v souboru je **více hodnot menších než větších ve srovnání se střední hodnotou** (**levostranná nesouměrnost**) nebo **více hodnot větších než menších ve srovnání se střední hodnotou (pravostranná nesouměrnost).**

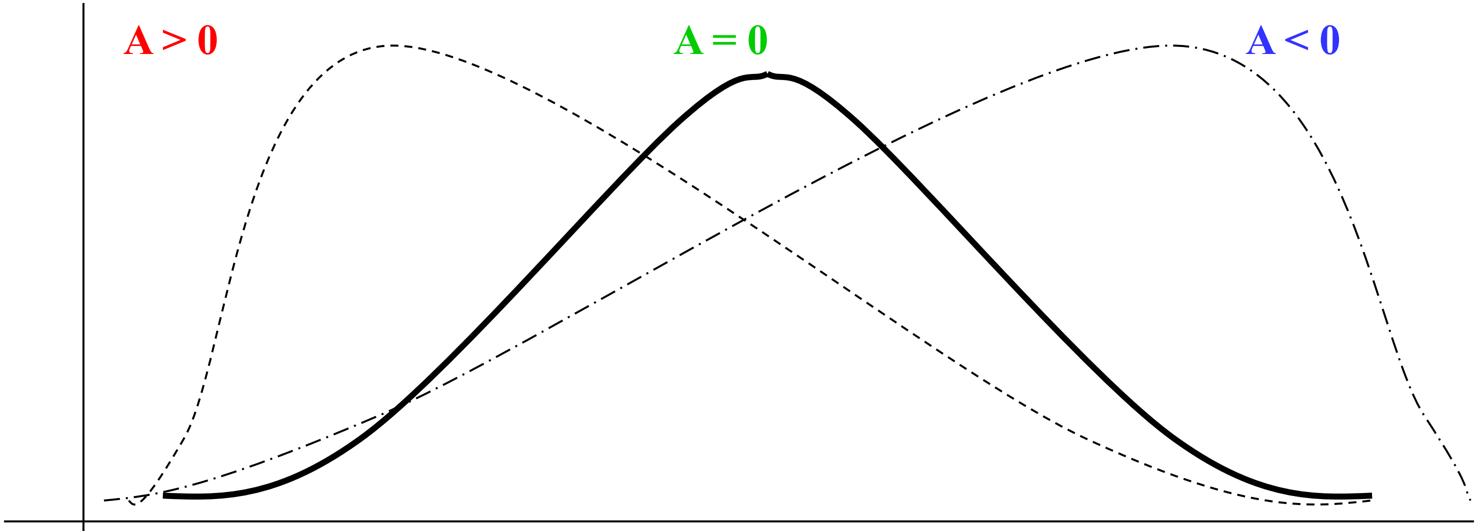


NESOUMĚRNOST

měříme **koeficientem nesouměrnosti**

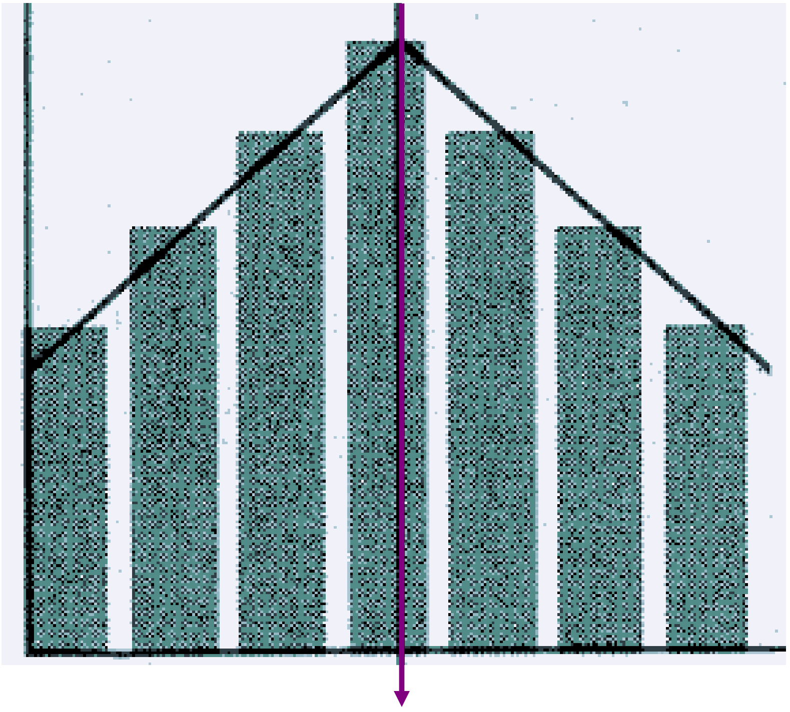
$$A = \frac{\sum_{j=1}^{N} \left(x_j - \bar{x} \right)^3}{n \cdot S^3}$$

$$A = \frac{\sum_{i=1}^{m} n_i \left(\bar{x}_i - \bar{x} \right)^3}{n \cdot S^3}$$



NESOUMĚRNOST

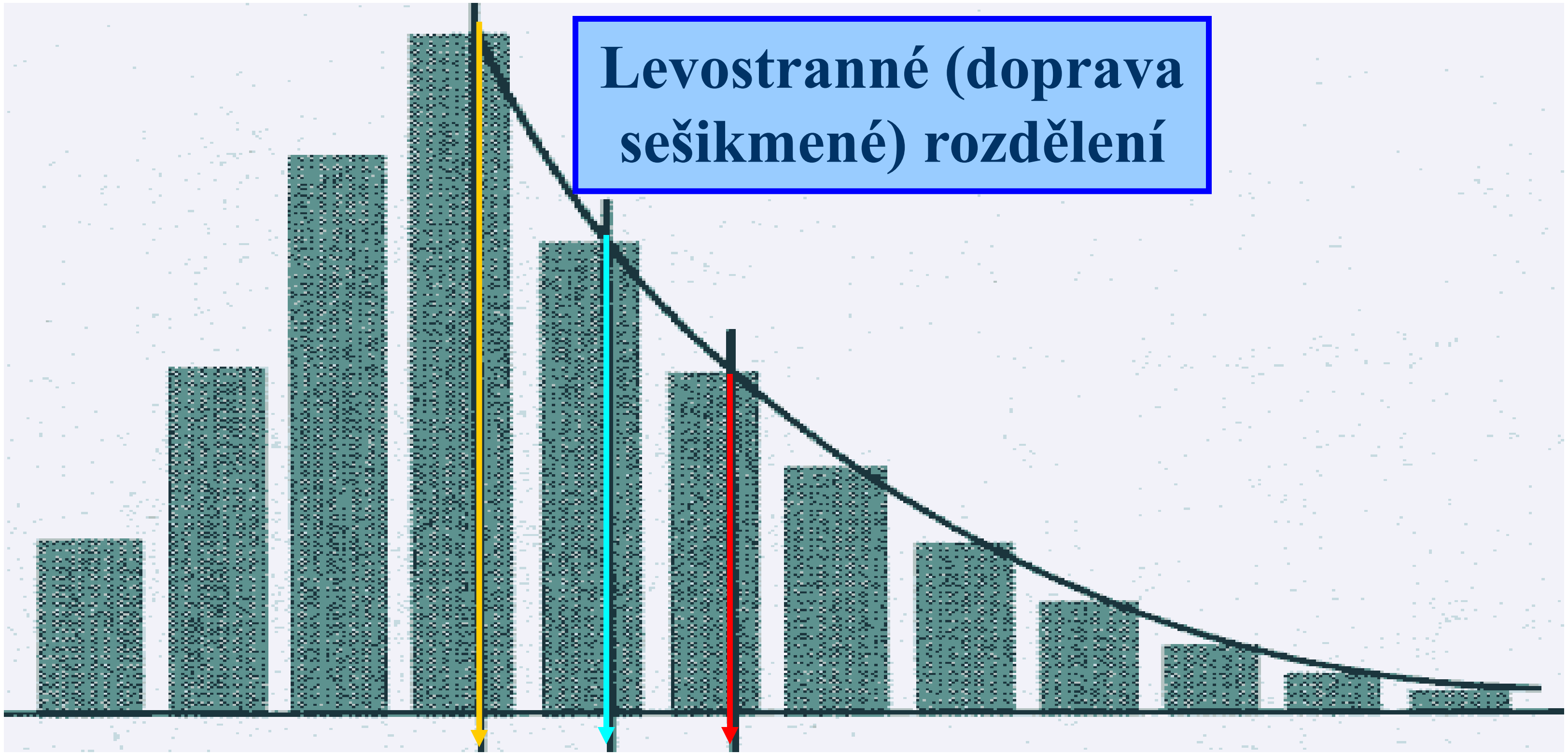
Souměrné rozdělení:



A = 0

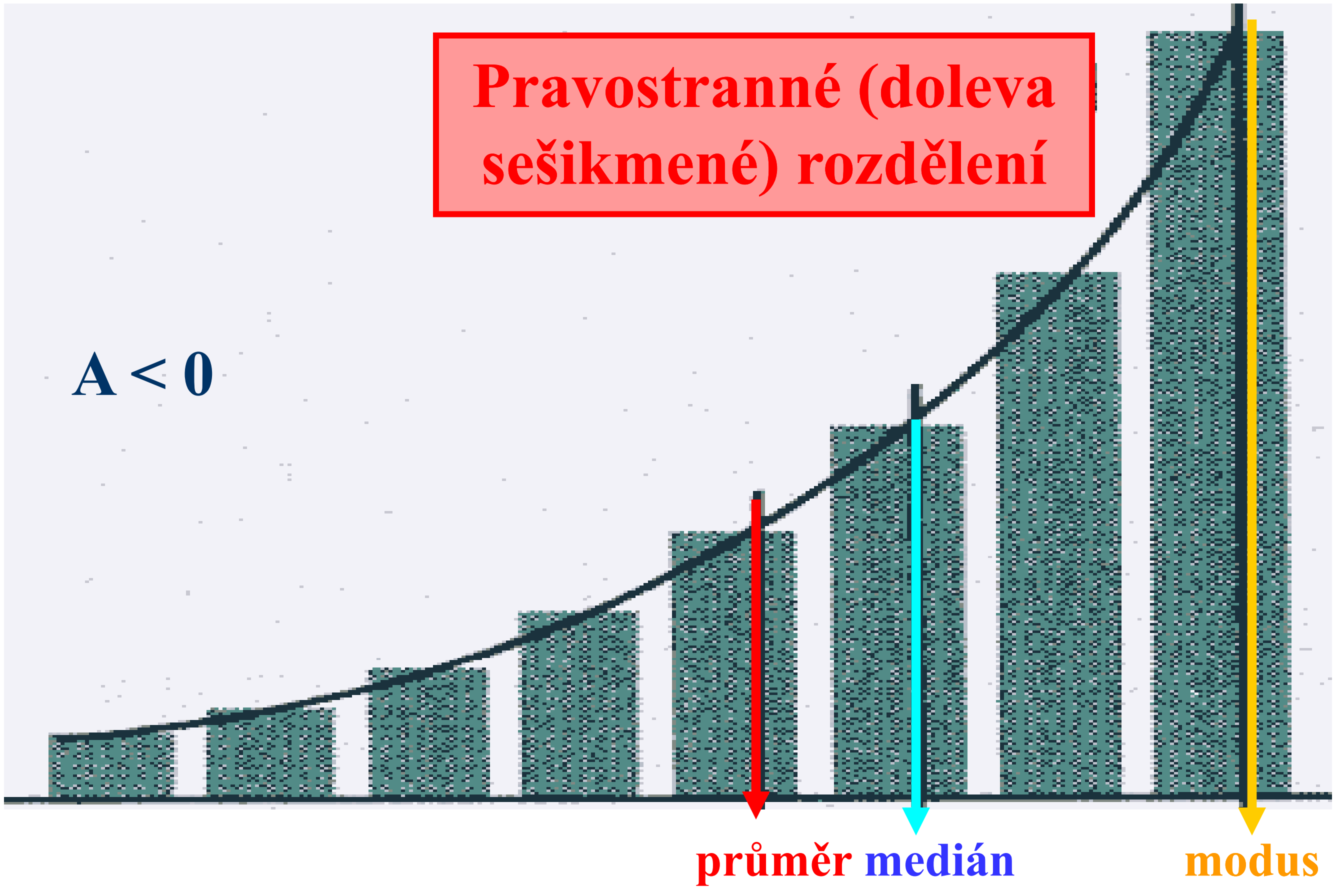
Průměr = medián = modus

NESOUMĚRNOST



modus medián průměr

NESOUMĚRNOST

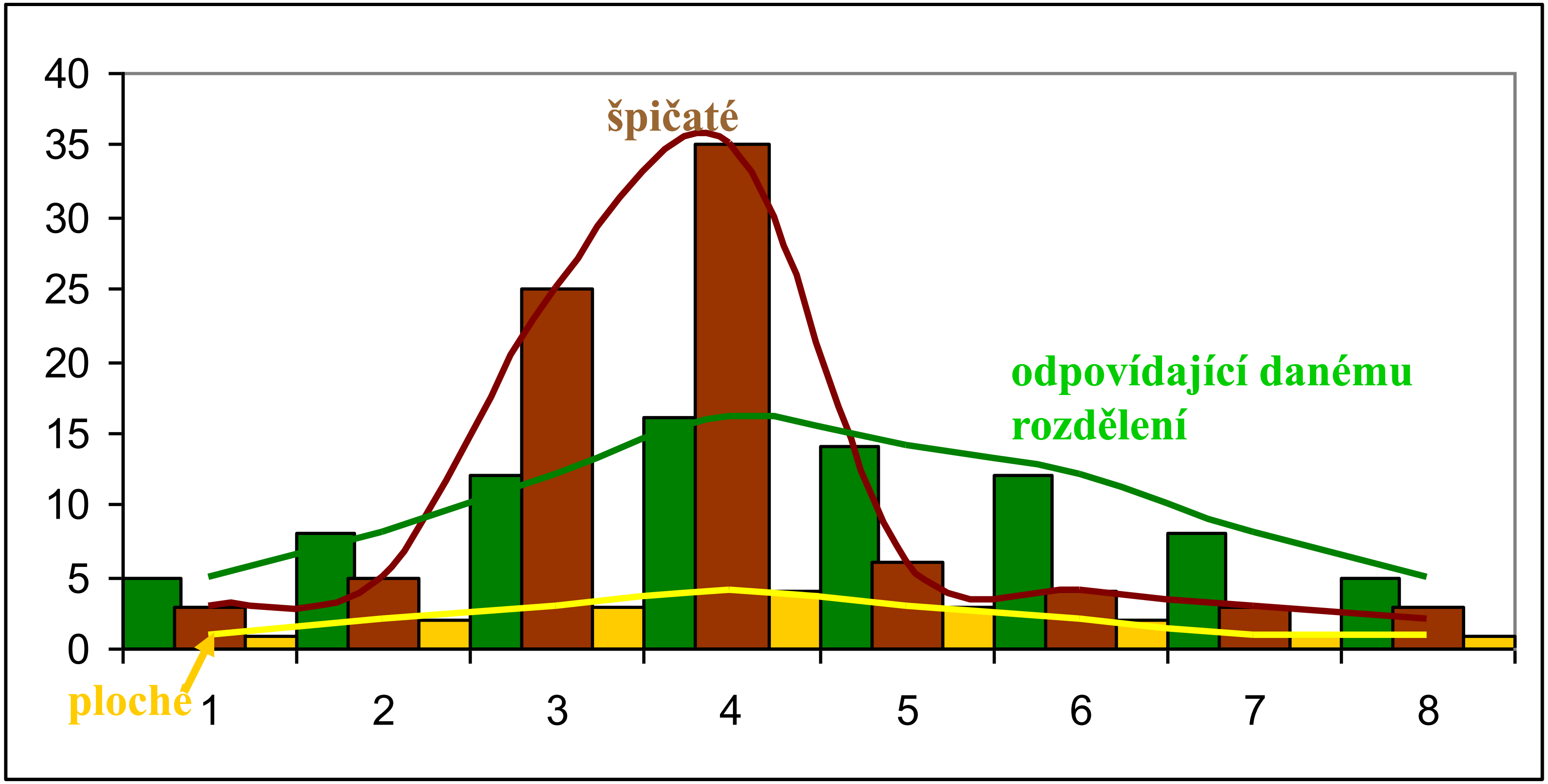


ŠPIČATOST

je mírou **koncentrace dat** kolem určité hodnoty nebo skupiny hodnot ve srovnání s určitým definovaným rozdělením veličiny (např. normálním). Rozlišujeme rozdělení:

- **ploché** – **koncentrace dat** kolem určité hodnoty **je NIŽŠÍ** než odpovídá definovanému rozdělení (tedy četnosti kolem této hodnoty jsou nižší)
- **špičaté** - **koncentrace dat** kolem určité hodnoty je **VYŠŠÍ** než odpovídá definovanému rozdělení(tedy četnosti kolem této hodnoty jsou vyšší)
- odpovídající danému definovanému rozdělení (např. normální)

ŠPIČATOST



špičaté
odpovídající danému rozdělení
ploché
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8

ŠPIČATOST

Mírou špičatosti je **koeficient špičatosti**:

$$E = \frac{\sum_{j=1}^{N}\left(x_j - \mu\right)^4}{N \cdot \sigma^4}\left[-3\right]$$

$$E = \frac{\sum_{i=1}^{m} n_i\left(\bar{x}_i - \bar{x}\right)^4}{n \cdot S^4}\left[-3\right]$$

Pro normální rozdělení platí:

E = 0 (3) normálně zahrocené

E < 0 (3) ploché

E > 0 (3) špičaté

BODOVÉ ODHADY ZÁKLADNÍCH PARAMETRŮ

Odhad střední hodnoty:

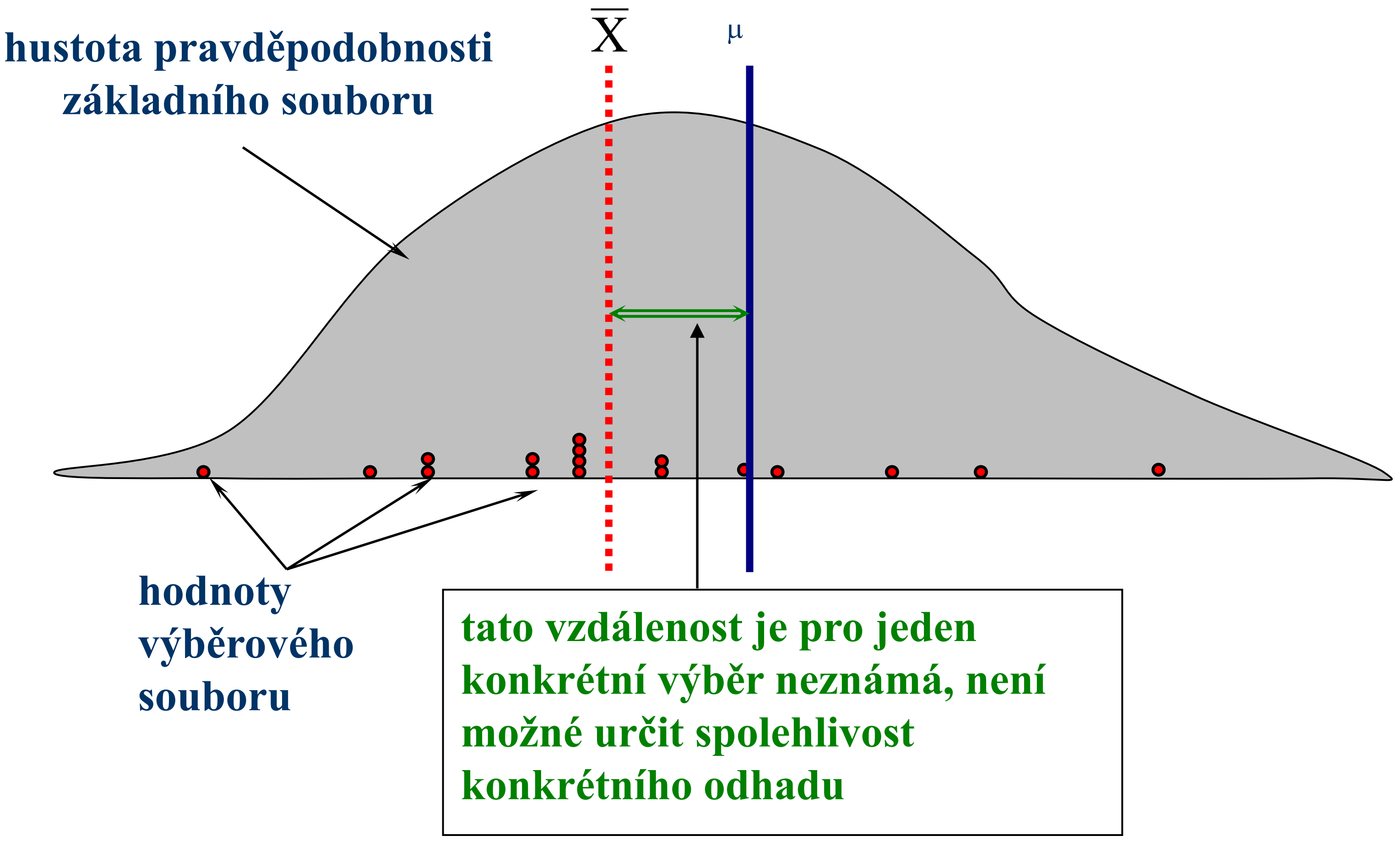
$$E\left(\overline{X}\right) = \mu$$

Odhad rozptylu:

$$S^2 \cdot \boxed{\frac{n}{n-1}} = \sigma^2$$

korekce vychýlení

BODOVÉ ODHADY ZÁKLADNÍCH PARAMETRŮ

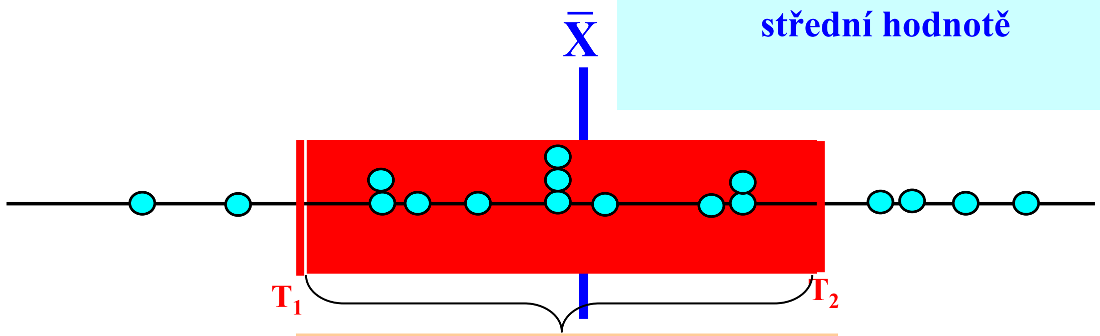


INTERVALOVÉ ODHADY PARAMETRŮ ZS

Interval spolehlivosti pro parametr τ při **hladině významnosti $\alpha \in (0,1)$** je určen statistikami T_1 a T_2:.

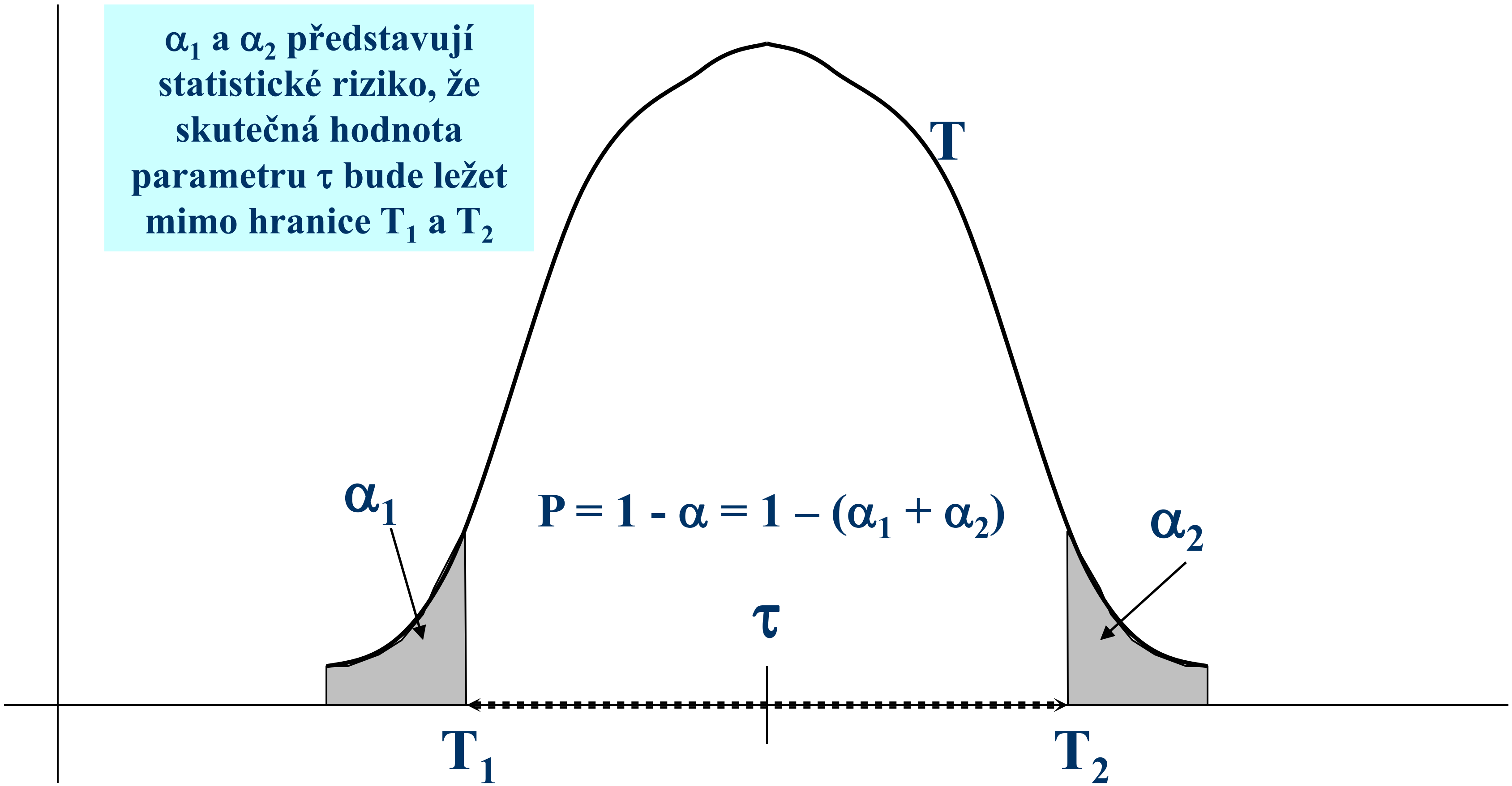
$$P\left(T_1 \le \tau \le T_2\right) = 1 - \alpha$$

toto je bodový odhad neznámé střední hodnoty μ vypočítaný z prvků výběru – nevíme nic o jeho vztahu ke skutečné střední hodnotě



toto je intervalový odhad neznámé střední hodnoty - předpokládáme, že s <u>pravděpodobností P =1-α</u> leží μ kdekoli v tomto úseku číselné osy

INTERVALOVÉ ODHADY PARAMETRŮ ZS



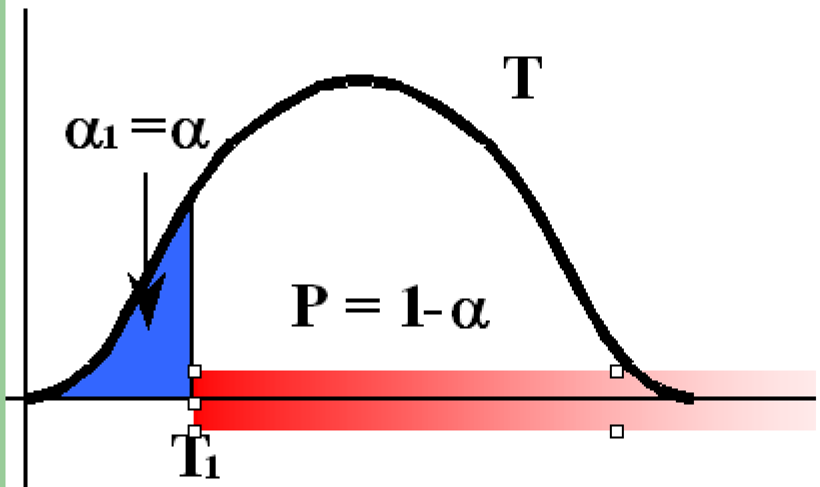
JEDNOSTRANNÉ INTERVALOVÉ ODHADY

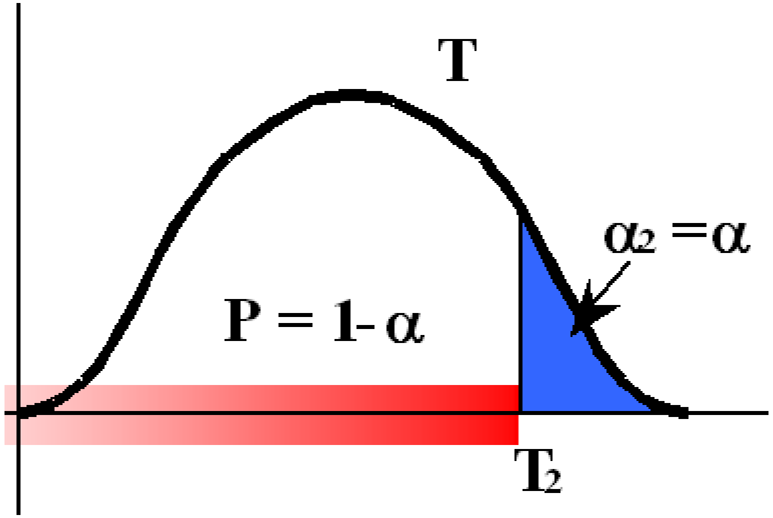
levostranný odhad

pravostranný odhad

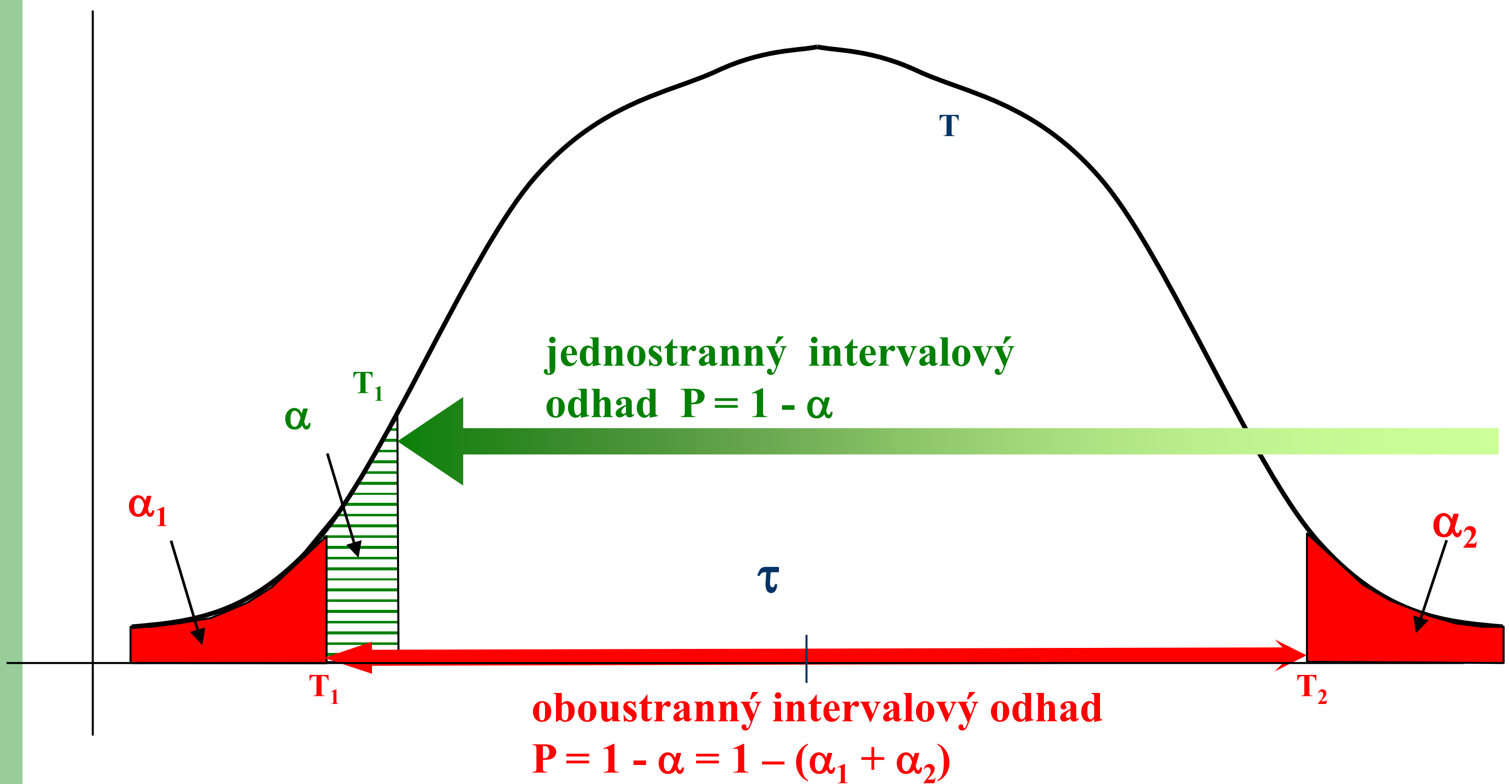
$$P(\tau > T_1) = 1 - \alpha$$

$$P(\tau < T_2) = 1 - \alpha$$

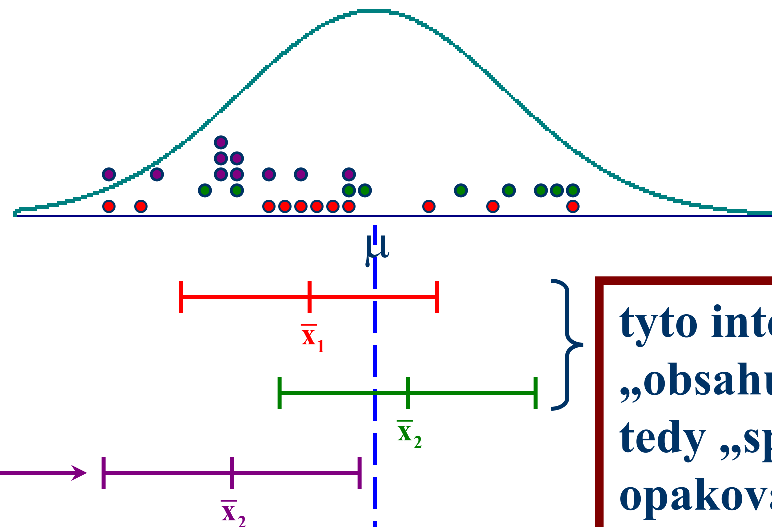




POROVNÁNÍ JEDNOSTRANNÉHO A ODOUSTRANNÉHO ODHADU



HLADINA VÝZNAMNOSTI α V INTERVALOVÝCH ODHADECH



tyto intervaly spolehlivosti „obsahují“ střední hodnotu (jsou tedy „správné“), těch (při opakovaných výběrech) bude nejméně (1- α).100 %

tento interval spolehlivosti „neobsahuje“ střední hodnotu (je tedy „chybný“), těchto intervalů se objeví nejvýše (100α) %

INTERVAL SPOLEHLIVOSTI STŘEDNÍ HODNOTY μ

❶ **je známa směrodatná odchylka σ základního souboru nebo je používán velký výběr (nad 30 prvků)**

$$\underbrace{\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{dolní hranice}} \le \mu \le \underbrace{\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{horní hranice}}$$

v případě velkého výběru lze použít místo σ výběrovou směrodatnou odchylku S

$z_{\alpha/2}$ je kvantil normovaného normálního rozdělení pro hladinu významnosti $\alpha/2$

INTERVAL SPOLEHLIVOSTI STŘEDNÍ HODNOTY μ

❷ **není známa směrodatná odchylka σ základního souboru a je používán malý výběr (do 30 prvků)**

Platí, že veličina $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ **má t-rozdělení s k =(n – 1) stupni volnosti**

$$\Downarrow$$

$$\bar{x} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}$$

$t_{\alpha/2,n-1}$ je kvantil Studentova t-rozdělení pro hladinu významnosti $\alpha/2$ a (n-1) stupňů volnosti

INTERVAL SPOLEHLIVOSTI STŘEDNÍ HODNOTY μ

❸ **velikost základního souboru je známa (N) a výběrový soubor je relativně velký (n > 5 % N)**

Používá se **korekce na konečný základní soubor**:

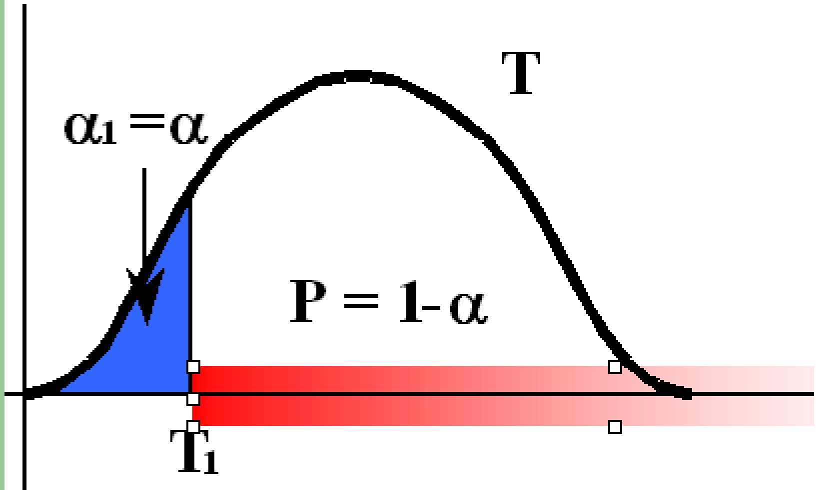
$$\bar{x} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \cdot \boxed{\sqrt{1 - \frac{n}{N}}} \leq \mu \leq \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$$

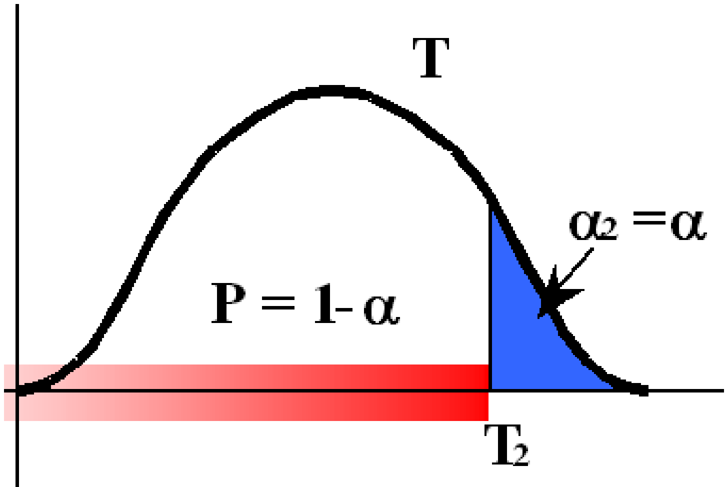
Účelem korekce je zmenšit standardní chybu $\bar{\mathbf{x}}$

INTERVAL SPOLEHLIVOSTI STŘEDNÍ HODNOTY μ

❹ jednostranné intervaly

Jednostranné intervaly se počítají podle stejných vztahů jako oboustranné, pouze **hladina významnosti je α místo α/2** (veškeré statistické riziko „chybného“ intervalu je na jedné straně)

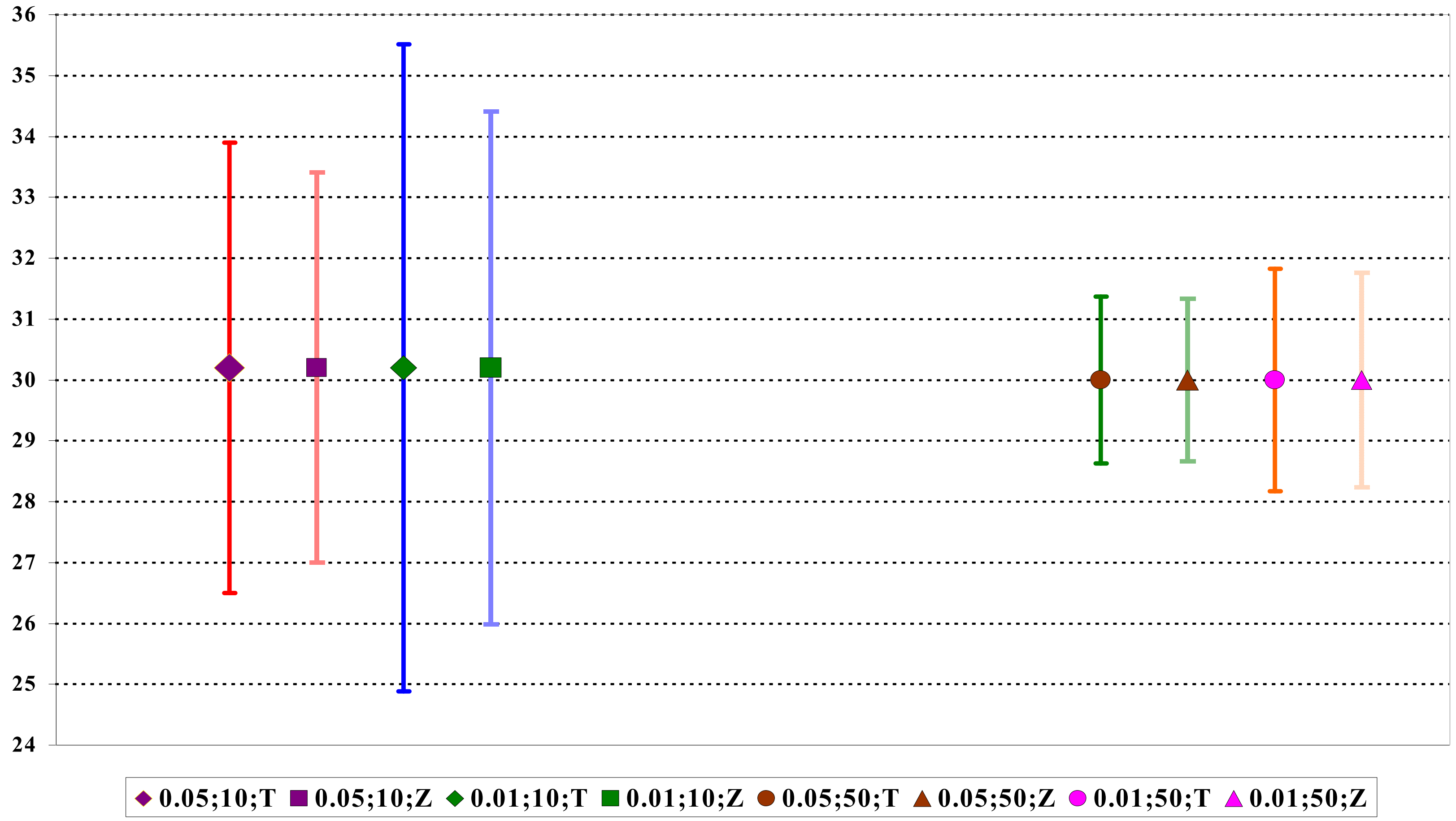




FAKTORY OVLIVŇUJÍCÍ VELIKOST INTERVALU SPOLEHLIVOSTI (IS)

- velikost výběru (čím větší výběr, tím užší IS)
- hladina významosti α(čím vyšší hodnota α, tím užší interval – nižší hladina významnosti (např. 0,01 místo 0,05) znamená požadavek vyšší spolehlivosti určení IS - pokud určíme α =0,01, požadujeme spolehlivost IS P=99%, pokud určíme α =0,05, požadujeme spolehlivost IS P=95%, IS musí být širší pro P=99% než pro P=95%, protože musíme zaručit vyšší spolehlivost)
- variabilita (čím vyšší hodnota směrodatné odchylky, tím širší IS)
- použitý vzorec (pokud používáme t-rozdělení, je IS širší než při použití N(0,1), rozdíl je markantnější u malých výběrů)

FAKTORY OVLIVŇUJÍCÍ VELIKOST INTERVALU SPOLEHLIVOSTI



INTERVAL SPOLEHLIVOSTI SMĚRODATNÉ ODCHYLKY σ

❶ **pro malé výběry**

Výpočet intervalu spolehlivosti směrodatné odchylky využívá χ^2-rozdělení a je nesouměrný – nesouměrnost je vyšší u odhadů vycházejících z malých výběrů.

$$\sqrt{\frac{n \cdot S^2}{\chi^2_{\frac{\alpha}{2}}}} \leq \sigma \leq \sqrt{\frac{n \cdot S^2}{\chi^2_{1-\frac{\alpha}{2}}}}$$

INTERVAL SPOLEHLIVOSTI SMĚRODATNÉ ODCHYLKY σ

❷ **pro velké výběry (nad 30 prvků)**

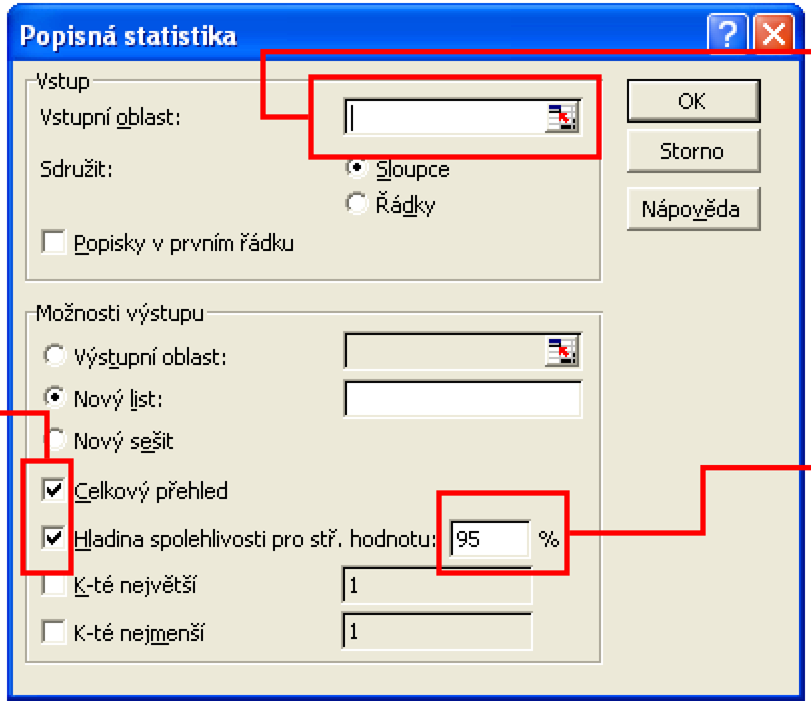
Výpočet intervalu spolehlivosti směrodatné odchylky pro velké výběry využívá normovaného normálního rozdělení a je souměrný.

$$\sigma = S \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{2n}}$$

INTERVALY SPOLEHLIVOSTI – PROVEDENÍ V EXCELU

❶ interval spolehlivosti střední hodnoty

a) pomocí doplňku Analýza dat



rozsah dat výběru

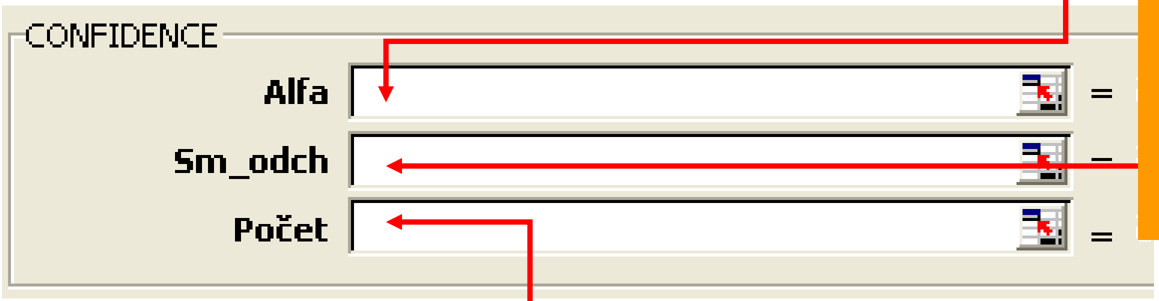
hodnota 100.(1-α)%

musí být zatrženo !!

INTERVALY SPOLEHLIVOSTI – PROVEDENÍ V EXCELU

❷ pomocí funkce CONFIDENCE

hodnota α



směrodatná odchylka (např. vypočítaná pomocí modulu „Popisná statistika“

velikost výběru

Způsob ❶ počítá interval spolehlivosti podle vzorce $t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}$

Způsob ❷ počítá interval spolehlivosti podle vzorce $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

VÍCEROZMĚRNÝ STATISTICKÝ SOUBOR

Vícerozměrný statistický soubor je množina C **souběžných realizací určitého počtu veličin $X_1, X_2, \ldots, X_m$.**

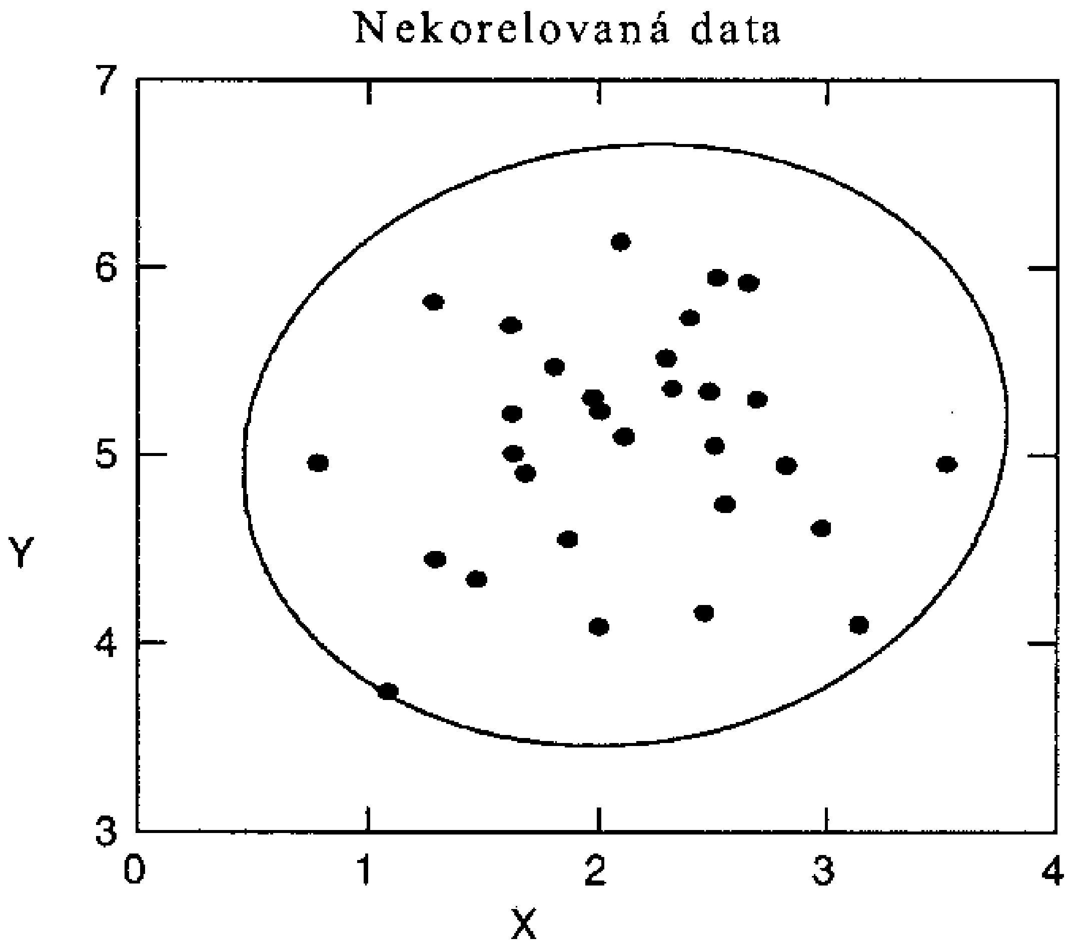
Množina C vznikne získáním hodnot znaků $X_1, X_2, \ldots, X_m$ na prvcích množiny n. C je potom množina uspořádaných *m-tic* hodnot $[x_1, x_2, \ldots, x_m]$ znaků $X_1, X_2, \ldots, X_m$.

n-tý OBJEKT

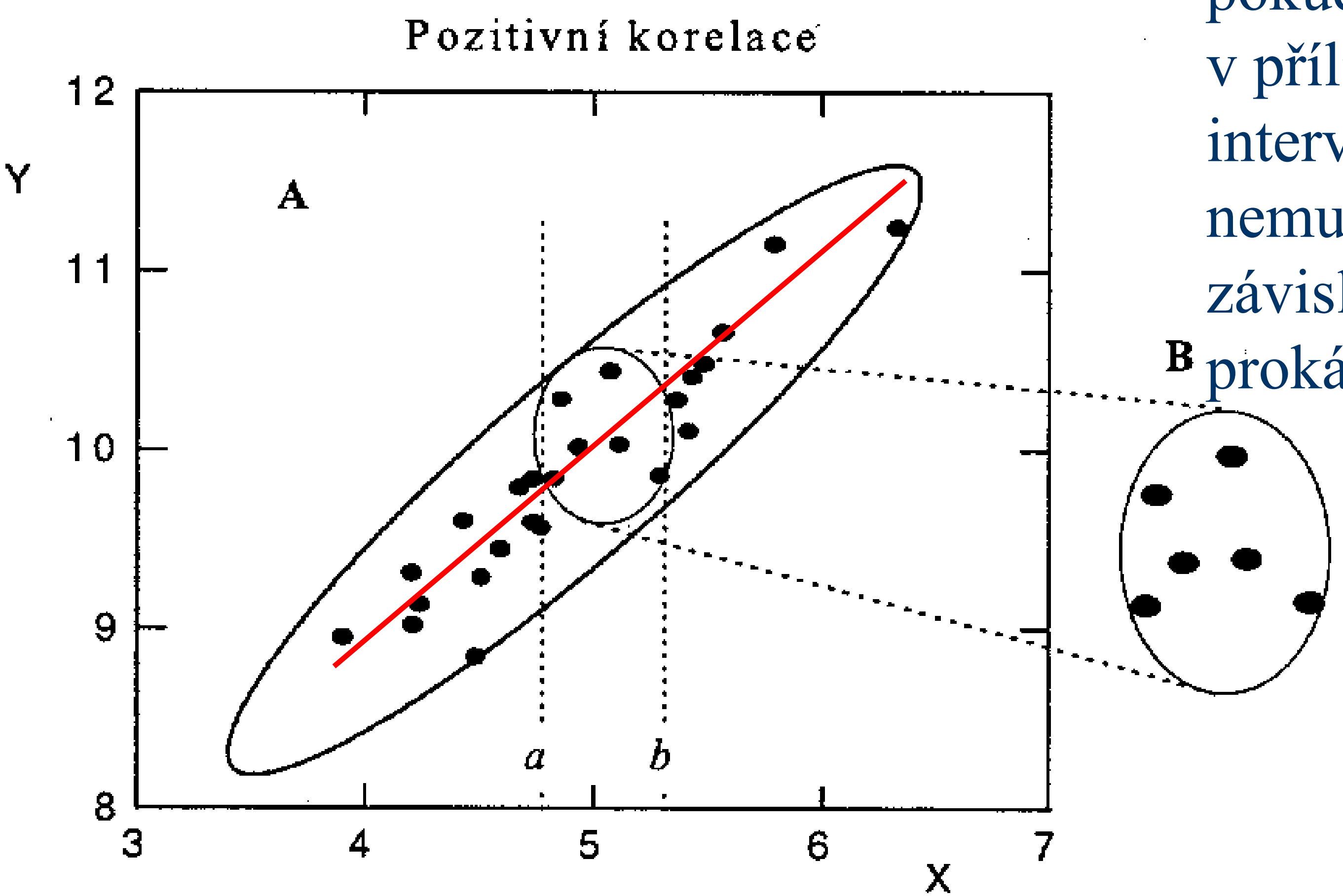
$$\mathbf{C} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_j^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,i} & \cdots & x_{1,m} \\ \vdots & & \vdots & & \vdots \\ x_{j,1} & \cdots & x_{j,i} & \cdots & x_{j,m} \\ \vdots & & \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,i} & \cdots & x_{n,m} \end{bmatrix}$$

m-tá VELIČINA

STATISTICKÁ ZÁVISLOST

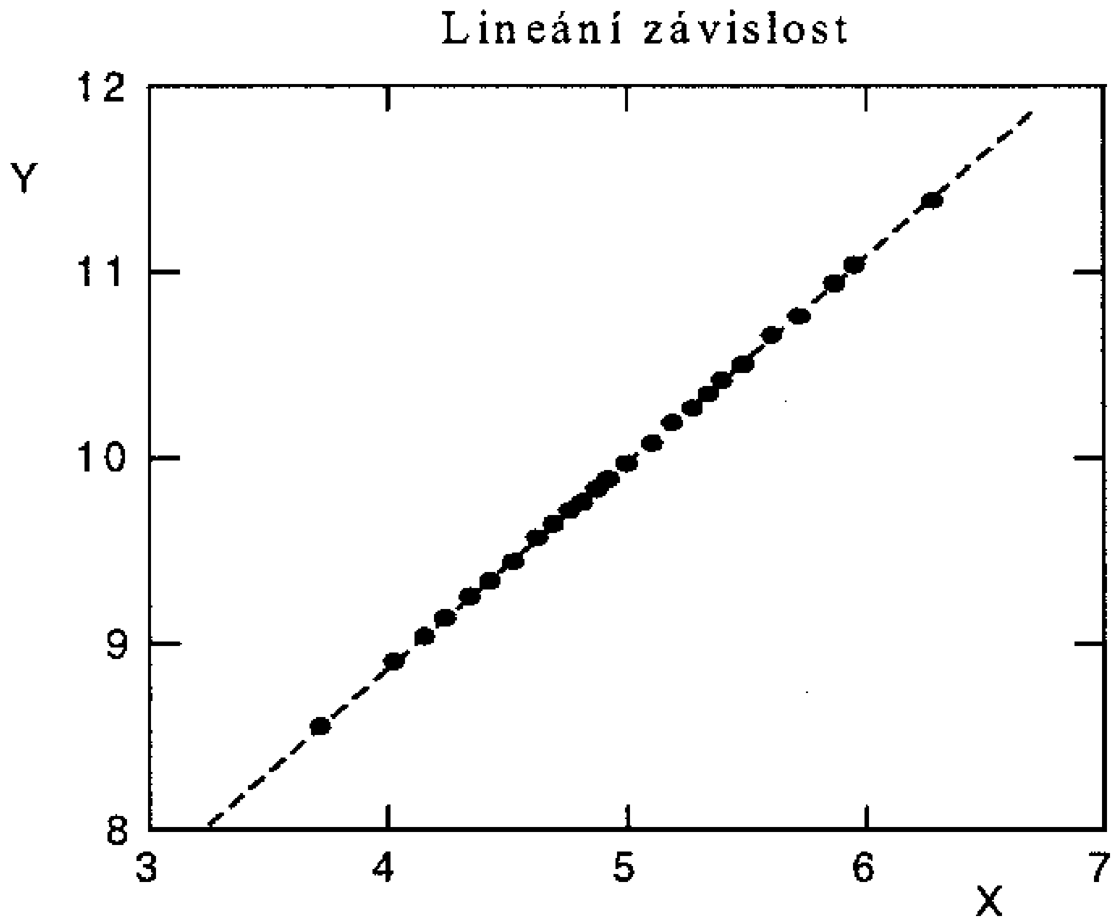


STATISTICKÁ ZÁVISLOST



pokud měříme v příliš malém intervalu, nemusí se závislost prokázat!!

STATISTICKÁ ZÁVISLOST



jedna proměnná je násobkem druhé – v tom případě je možné jednu proměnnou z analýzy vyloučit bez ztráty informace

STATISTICKÁ ZÁVISLOST

- **korelace** – popisuje vliv změny úrovně jednoho znaku na změnu úrovně jiných znaků a platí pro **kvantitativní (měřené) znaky**;

- **kontingence** – popisuje závislost **kvalitativních** (slovních, popisných) znaků, které mají více než dvě alternativy, tzv. **množných znaků** (např. druh dřeviny, národnost, apod.);

- **asociace** - popisuje závislost **kvalitativních** (slovních, popisných) znaků, které mají pouze dvě alternativy, tzv. **alternativních znaků** (např. pohlaví, odpovědi typu ano/ne, …).

KORELACE

typy podle počtu korelovaných znaků

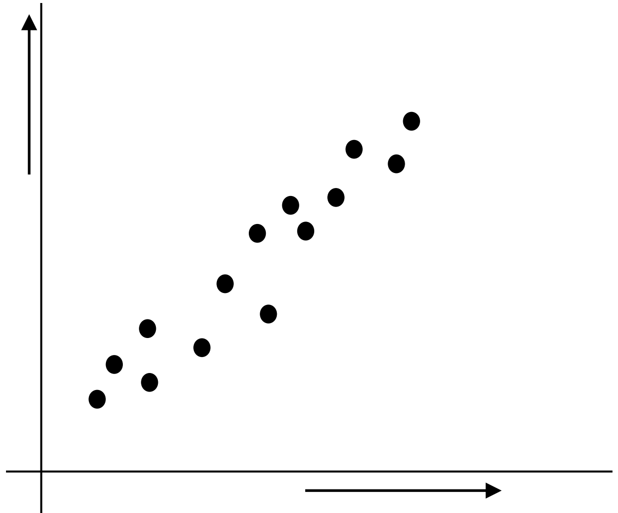
- ***jednoduchá*** – popisuje vztah dvou znaků,
- ***mnohonásobná*** – popisuje vztahy více než dvou znaků,
- ***parciální*** – popisuje závislost dvou znaků ve vícerozměrném statistickém souboru při vyloučení vlivu ostatních znaků na tuto závislost·

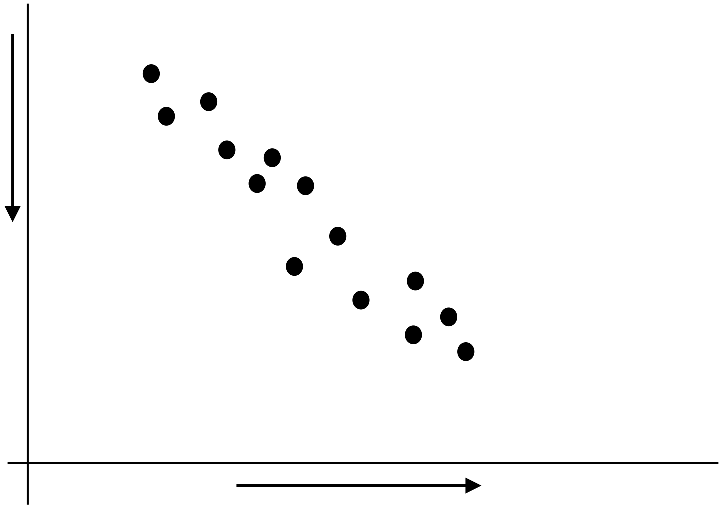
KORELACE

typy podle smyslu změny hodnot

- ***kladná*** – se zvyšováním hodnot jednoho znaku se zvyšují i hodnoty druhého znaku

- ***záporná*** - se zvyšováním hodnot jednoho znaku se zmenšují hodnoty druhého znaku





KORELACE

typy podle tvaru závislosti

- ***přímková (lineární)*** – grafickým obrazem závislosti je přímka (lineární trend)
- ***křivková (nelineární)*** – grafickým obrazem závislosti je křivka (nelineární trend)

KORELAČNÍ POČET

korelační analýza

- zjišťuje ***existenci závislosti*** a její druhy,
- měří ***těsnost závislosti***,
- ověřuje ***hypotézy o statistické významnosti*** *závislosti*;

regresní analýza

- zabývá se ***vytvořením vhodného matematického modelu*** závislosti,
- stanoví ***parametry*** tohoto *modelu*,
- ověřuje ***hypotézy o vhodnosti a důležitých vlastnostech modelu***.

KORELAČNÍ KOEFICIENT

PRO JEDNODUCHOU KORELACI

párový - zvláštní případ vícenásobného korelačního koeficientu, kdy vyjadřuje míru lineární stochastické závislosti mezi náhodnými veličinami Xi a Xj**,**

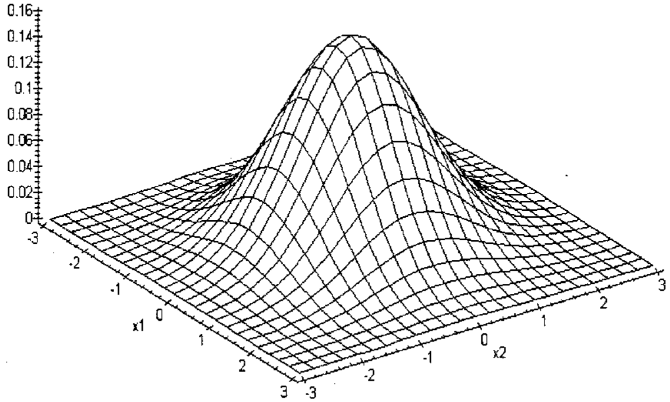
- **Pearsonův**
- **Spearmanův (korelace pořadí)**

PEARSONŮV KORELAČNÍ KOEFICIENT (r)

podmínkou je dodržení dvourozměného normálního rozdělení

= normovaná kovariance

$$r_{x_1 x_2} = r_{x_2 x_1} = \frac{\boxed{\mathrm{cov}_{x_1 x_2}}}{s_{x_1} \cdot s_{x_2}}$$



PEARSONŮV KORELAČNÍ KOEFICIENT (r)

KOVARIANCE:

- **míra intenzity vztahu** mezi složkami vícerozměrného souboru
- je mírou intenzity **lineární** závislosti
- je vždy **nezáporná**
- její **limitou je součin směrodatných odchylek**
- je **symetrickou funkcí** svých argumentů
- její **velikost je závislá na měřítku argumentů ⇒ nutnost normování**

$$\mathrm{cov}_{x_1 x_2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)\cdot\left(x_{2i} - \bar{x}_2\right)$$

PEARSONŮV KORELAČNÍ KOEFICIENT (r)

Základní vlastnosti Pearsonova korelačního koeficientu:

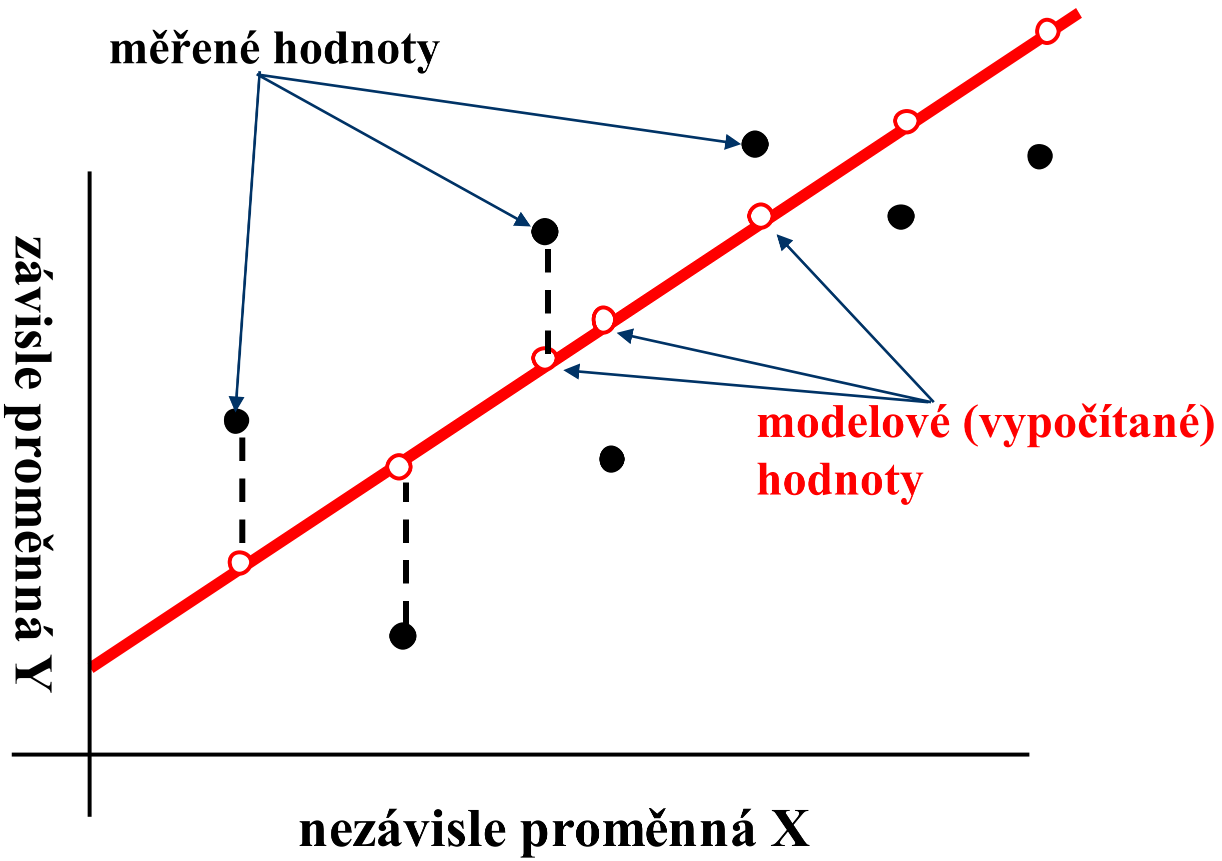
- je to **bezrozměrná** míra lineární korelace;
- nabývá hodnoty **0 – 1 pro kladnou korelaci, 0 – (-1) pro zápornou korelaci**;
- hodnota **0** znamená, že mezi posuzovanými veličinami **není žádný lineární vztah** (může být nelineární) nebo tento vztah zůstal na základě dat, které máme k dispozici, neprokázán;
- hodnota **1** nebo **(-1)** indikuje **funkční závislost**;
- hodnota korelačního koeficientu je stejná pro závislost x_1 na x_2 i pro opačnou závislost x_2 na x_1.

REGRESNÍ ANALÝZA

Základní úlohou **regresní analýzy** je nalezení **vhodného modelu studované závislosti.**

Snažíme se **nahradit každou měřenou** (experimentální, empirickou, zjištěnou) **hodnotu závisle proměnné** (vysvětlované proměnné) Y **hodnotou teoretickou** (modelovou, vyrovnanou, predikovanou), tj. hodnotou **ležící na spojité funkci** (modelu) **nezávisle proměnné** (vysvětlující proměnné) X ($\boldsymbol{X}$)

REGRESNÍ ANALÝZA



REGRESNÍ MODEL

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nk} \end{bmatrix}}_{\mathbf{X}} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

y závisle proměnná

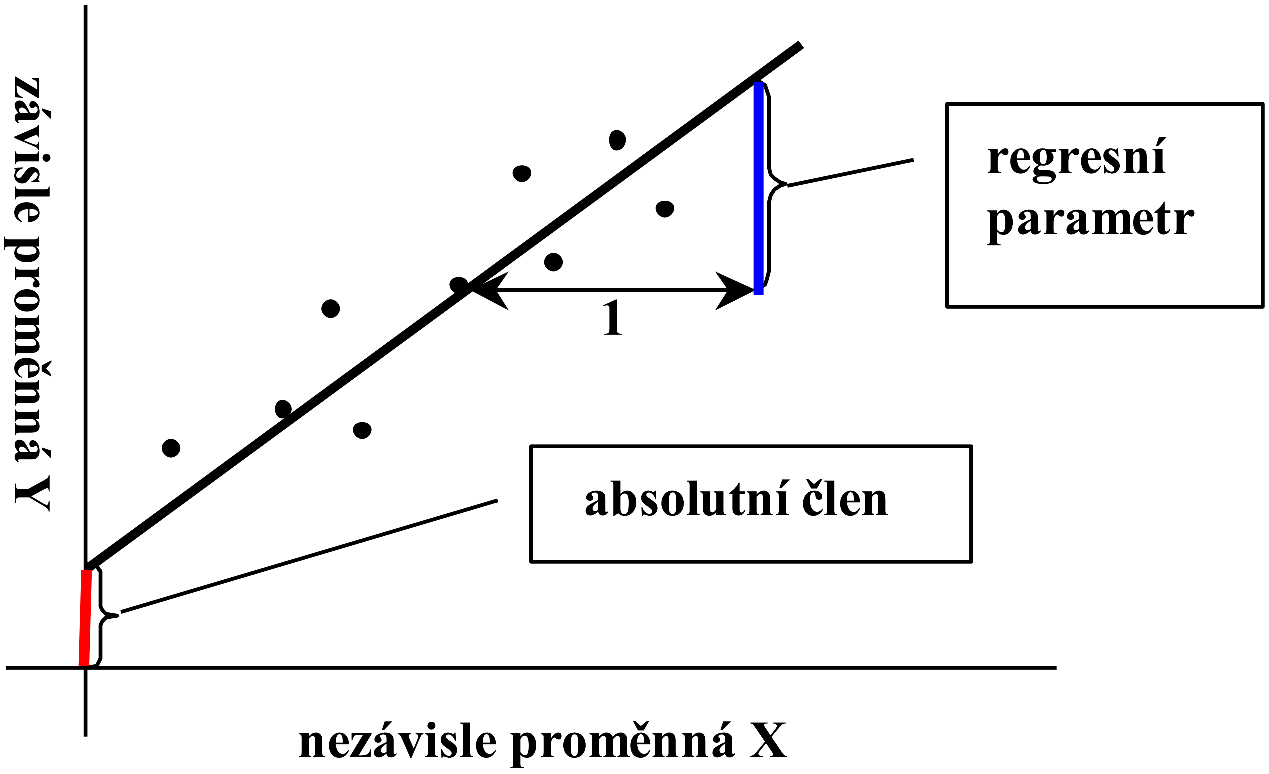
X nezávisle proměnná

β regresní parametry

ε náhodná chyba

$$\boxed{\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}}$$

REGRESNÍ MODEL



TEST VÝZNAMNOSTI REGRESNÍHO MODELU – co testujeme

$$Y = \mathbf{b_0} + \mathbf{b_1}x_1 + \mathbf{b_2}x_2 + \mathbf{b_3}x_3 + \ldots + \mathbf{b_m}x_m$$

Testujeme JEDNOTLIVÉ PARAMETRY (jestliže je daný parametr nevýznamný, příslušná proměnná x_j nijak nepřispívá ke zpřesnění odhadu závisle proměnné a je v modelu zbytečná).

Testujeme MODEL JAKO CELEK (zda příslušná kombinace nezávisle proměnných statisticky významně zpřesní odhad závisle proměnné oproti použití jejího průměru)

TEST VÝZNAMNOSTI REGRESNÍHO MODELU JAKO CELKU

1. Test významnosti korelačního koeficientu

2. Pomocí analýzy rozptylu

Zdroj variability	Součet čtverců odchylek	Počet stupňů volnosti	Průměrný čtverec odchylek (rozptyl)	Testové kritérium
regresní model	$S_{REG} = \sum_{i=1}^{n}(y'_i - \bar{y})^2$	$DF_{REG} = m - 1$	$M_{REG} = \frac{S_{REG}}{DF_{REG}}$	$F = \frac{M_{REG}}{M_R}$
reziduum (nevysvětleno regresním modelem)	$S_R = \sum_{i=1}^{n}(y_i - y'_i)^2$	$DF_R = n - m$	$M_R = \frac{S_R}{DF_R}$	
Celkový	$S_C = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$DF_C = n - 1$		

Testové kritérium F se porovná s kritickou hodnotou $F_{\alpha;m-1;n-m}$.

TEST VÝZNAMNOSTI REGRESNÍCH PARAMETRŮ

H$_0$: $\beta_j = 0$, *tj. j-tý regresní parametr je nevýznamný*

$$t = \frac{b_j - \beta_j}{s_b} \quad \text{pro } \beta_j = 0 \quad \Rightarrow \quad t = \frac{b_j}{s_b}$$

Pokud platí, že $|t| > t_{\alpha 2;n-m}$, potom je j-tý regresní parametr statisticky významný a příslušná proměnná musí zůstat v modelu.