Národní informační středisko pro podporu jakosti

Konzultační středisko statistických metod při NIS-PJ

Základy zpracování, vyhodnocování a prezentování dat

Ing. Josef Křepela

Odborná skupina statistických metod ČSJ

Ing. Jiří Tonar

Ústav strojírenské technologie ČVUT Praha

21. dubna 2005.

Matematicko-statistické metody se uplatňují:

- při **zpracování a analýze časově uspořádaných údajů**, kde závěr zhodnocení má zpětnovazební charakter;
- při **ověřování účinnosti navržených opatření** před tím, než jsou tato opatření začleněna do dokumentace;
- při **řešení problémů, jejichž výstup je ovlivňován celou řadou faktorů** a zjištění výstupu vyžaduje určitý pokus;
- při **analýze výrobního procesu**, odhalování zvláštních příčin variability a jejich odstraňování, při postupném dosahování stabilizovaného a statisticky zvládnutého procesu který pracuje na požadované úrovni jakosti;

- ve vstupní kontrole při **ověřování jakosti dávek výrobků a surovin**;
- v oblasti činností technologů, konstruktérů, projektantů, pracovníků zkušeben a laboratoří při řešení problémů jako jsou:
  - **objektivní postupy pro ověřování určitých tvrzení** za specifikovaných podmínek jako např. ověřování deklarovaných hodnot, očekávané účinnosti navržených opatření v technologii a pod.,
  - **navrhování experimentů**, které je nutno realizovat např. při zavádění nových výrob, při zlepšování nebo zhospodárňování výrobního procesu a při zjišťování stupně interakce mezi některými faktory.

Správné uplatňování statistických metod vyžaduje jistý způsob specifického myšlení:

„ Statistické myšlení “

H. G. Wells (1866 - 1946):

„**Statistické myšlení bude jednoho dne pro zdatného občana právě tak nezbytné, jako je schopnost číst a psát**“.

(Citace z článku: S. S. Wilks : Undergraduated Statistical Education, Jour. Am. Stat. Ass., vol. 46, 1951)

Statistické myšlení:

1) *v oblasti logiky:*

indukce **namísto** **dedukce**;

2) *v oblasti filosofie:*

náhoda **namísto** **nutnosti**

3) *v oblasti typu zákonitostí:*

indeterminismus **namísto** **determinismu**;

4) *výchozí kvantitativní charakteristika:*

pravděpodobnost.

J. Janko: Základy STATISTICKÉ INDUKCE

„Statistickou indukcí rozumíme **zobecnění statistických výsledků získaných zpracováním určitého souboru**, za předpokladu, že je možno **vztáhnout hodnoty zjištěných charakteristik na rozsáhlejší soubor než je ten, ze kterého byly skutečně odvozeny** “.

„Tohoto postupu užíváme stále v praktické statistice, ačkoliv často nebývají předpoklady zcela splněny. "

Předpoklady realizace statistického myšlení:

∇ **respektování náhody jako objektivní součásti reálu,**

∇ **znalost statistické indukce,**

∇ **znalost základů teorie pravděpodobnosti a matematické statistiky,**

∇ **schopnost zohlednit pravděpodobnostní prvky:**

- při technické formulaci problému,
- při specifikaci podmínek pokusu,
- při technickém řešení problému,
- při interpretaci výsledků,
- při zohledňování rizik spojených se závěry.

Statistického myšlení v praxi vyžaduje:

- jasné vymezení problému, který má být řešen;
- stanovení rozhodující veličiny – jakostní vlastnosti a způsobu jejího zjišťování ;
- zabezpečení stálých podmínek při jejím zjišťování;
- uvědomění si, že výsledky měření vykazují jistou (často jen částečně odstranitelnou) variabilitu;

  {rozdíl mezi náhodnými a zvláštními – vymezitelnými - příčinami proměnlivosti} ;
- vytváření podskupin homogenních výsledků, zahrnujících pouze náhodnou proměnlivost ;
- respektovat náhodné odebírání jednotek do náhodných výběrů, tak aby každá jednotka v souboru měla stejnou pravděpodobnost, že může být vybrána do výběru;

- studium ne jen celkové variability, ale i variability uvnitř podskupin a variability mezi podskupinami (v čase) ;
- provádění dostatečného počtu pozorování ;
- vážení rizik chybných závěrů, činěných na základě neúplné informace z náhodných výběrů ;
- prezentování dat přehledně, ve zhuštěné formě číselně, nebo graficky ;
- charakterizování dat číselně, udáním polohy na číselné ose a míry proměnlivosti – variability ;
- uvědomění si nejen variability studované náhodné veličiny, ale i z ní odvozené variability vypočítaných statistik – výběrových charakteristik .

Matematická statistika (MS)

- studuje hromadné jevy,
- činí o nich závěry pomocí indukce.

MS se soustřeďuje především na:

- sběr údajů, jejich popis a analýzu,
- rozšíření platnosti závěrů z malého počtu vzorků na soubor, z něhož vzorky pocházejí,
- zpracování a vyhodnocování informací o realitě, která není známá.

Věrohodnost závěrů analýzy vyžaduje, aby:

- výrobní dávky byly vyrobeny za **stejných podmínek,**
- **podmínky pokusu** byly specifikovány předem a byly dodržovány během celého pokusu,
- vzorky byly odebrány **náhodně a byly reprezentativní** pro soubor, z něhož jsou odebrány.

Znaky jakosti - náhodné veličiny

a) **KVANTITATIVNÍ** - měřitelné
spojité náhodné veličiny

b) **KVALITATIVNÍ** - neměřitelné, nebo neměřené
diskrétní náhodné veličiny

POKUS je proces zjišťování hodnoty znaku jakosti - hodnoty náhodné veličiny - (měření, vážení, přiřazování), který probíhá a opakuje se za **stále stejných podmínek.**

Výsledek pokusu (zjištěná hodnota znaku jakosti; zjištěná hodnota náhodné veličiny) je vyjádřen reálným číslem.

NÁHODNÝ POKUS je takový pokus, který může dávat různé výsledky i při dodržení stejných podmínek.

NÁHODNÝ JEV je tvrzení o výsledku náhodného pokusu, o kterém lze po jeho uskutečnění jednoznačně rozhodnout, zda je či není pravdivé.

Např. Průměr vyrobené osičky je 10,37 mm; je menší než 10,40 mm; je v mezích 13,3 mm až 10,4 mm.

Doba bezporuchového provozu televizoru Tesla M63 je 1500 h.

Počet neshodných jednotek v náhodném výběru z dodávky je menší než 4.

Výsledky náhodného pokusu (realizace náhodné veličiny) tedy ani realizace náhodného jevu nelze s jistotou předpovědět. Některé nastávají častěji, některé méně často, některé jen zřídka.

TEORIE PRAVDĚPODOBNOSTI

se zabývá studiem zákonitostí výsledků, jejich popisem a vytvořením pravidel pro určení počtu výskytu možných náhodných jevů.

MATEMATICKÁ STATISTIKA využívá jejích výsledků.

Matematická statistika pracuje s výsledky *opakovaných náhodných pokusů - údaji získanými na základě:*

NÁHODNÉHO VÝBĚRU - *konečného souboru náhodně vybraných jednotek - reprezentujícího hypotetický základní soubor, ze kterého je náhodný výběr odebrán*.

(Při opakování náhodného výběru obvykle dostáváme poněkud jiné výsledky, vlivem náhody při výběru jednotek, na kterých je zjišťován znak jakosti.)

ZÁKLADNÍ SOUBOR je konečný, ale častěji nekonečný soubor, všech možných výsledků náhodného pokusu - realizací náhodné veličiny (hodnot znaku jakosti) - které všechny **nemůžeme zjistit**. Základní soubor odpovídá realitě, která v praxi není známa a kterou **pouze odhadujeme**, nebo o ní **činíme závěry**, na základě informace získané z náhodného výběru.

Statistický soubor se obvykle rozumí konečný soubor hodnot náhodné veličiny, které, při jejich znalosti, poskytují přesně, vyčerpávajícím způsobem, informaci o statistických vlastnostech souboru.

OBLAST TEORIE

Základní soubor -

všech možných hodnot náhodné veličiny, popsaný jejich **rozdělením pravděpodobnosti** (*typ* a *parametry*)

Náhodný výběr -

n nezávisle na sobě vybraných hodnot náhodné veličiny popsaný jejich **rozdělením četností** (*výběrové charakteristiky*)

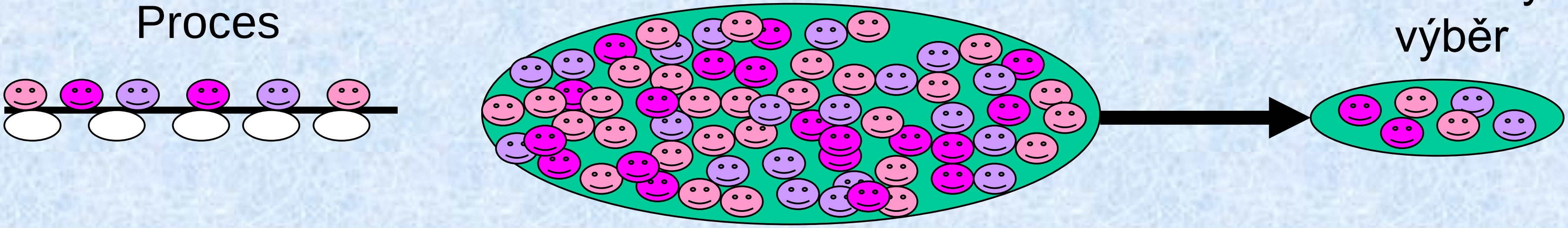
OBLAST REALITY

- ***Základní soubor*:** Soubor dat, který zahrnuje všechna měření prvků určité hypotetické skupiny.
- ***Náhodný výběr*:** Soubor dat, který se skládá z měření pouze části prvků náhodně vybraných ze základního souboru.

Proces

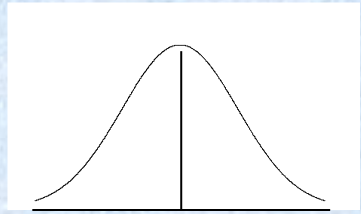
Základní soubor

Náhodný výběr



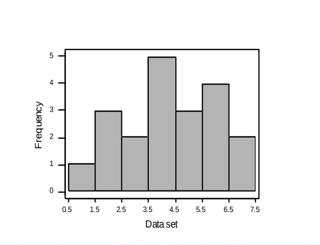
Střední hodnota základního souboru - μ

Směrodatná odchylka základního souboru - σ



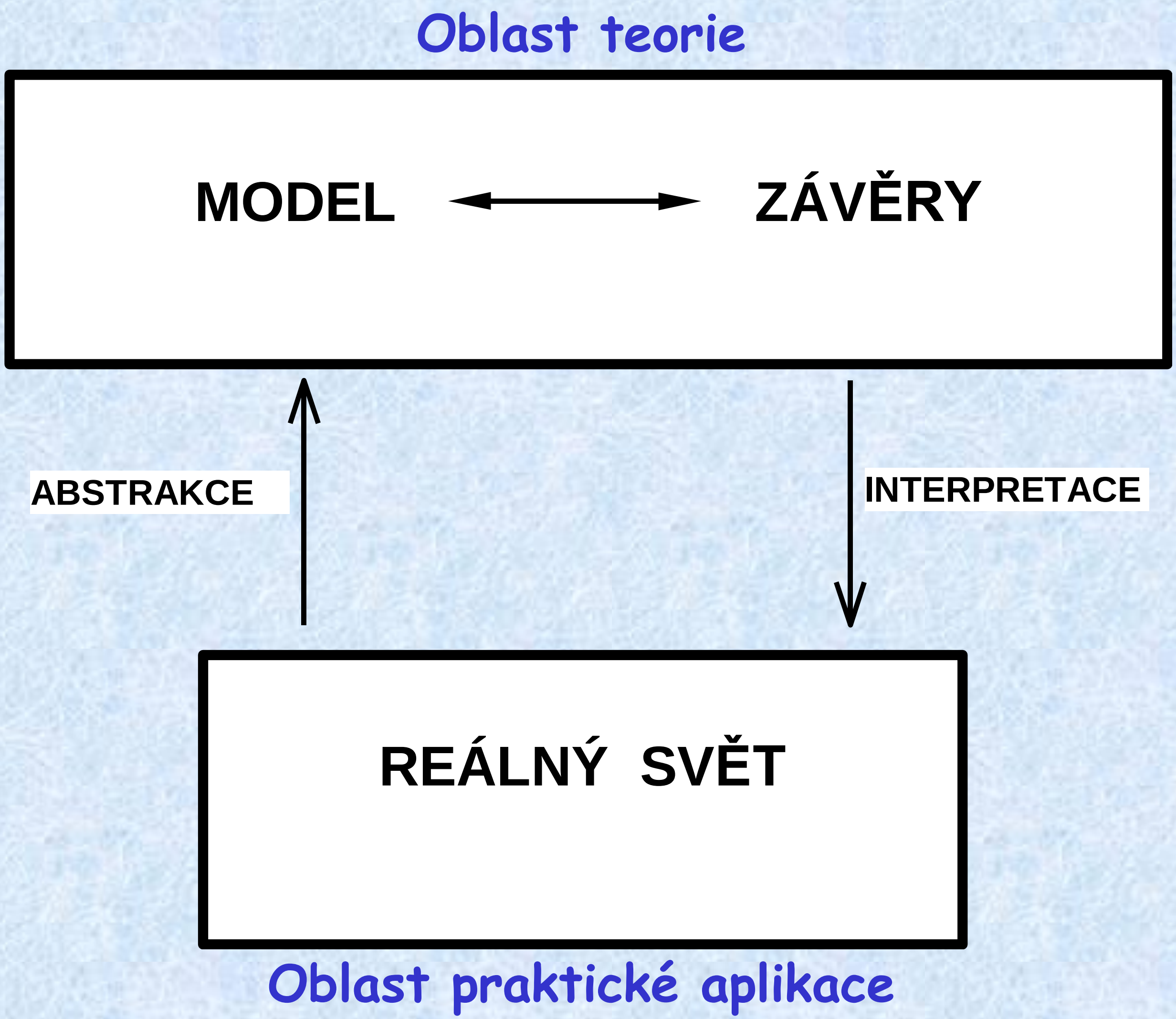
Výběrový průměr - $\overline{X}$

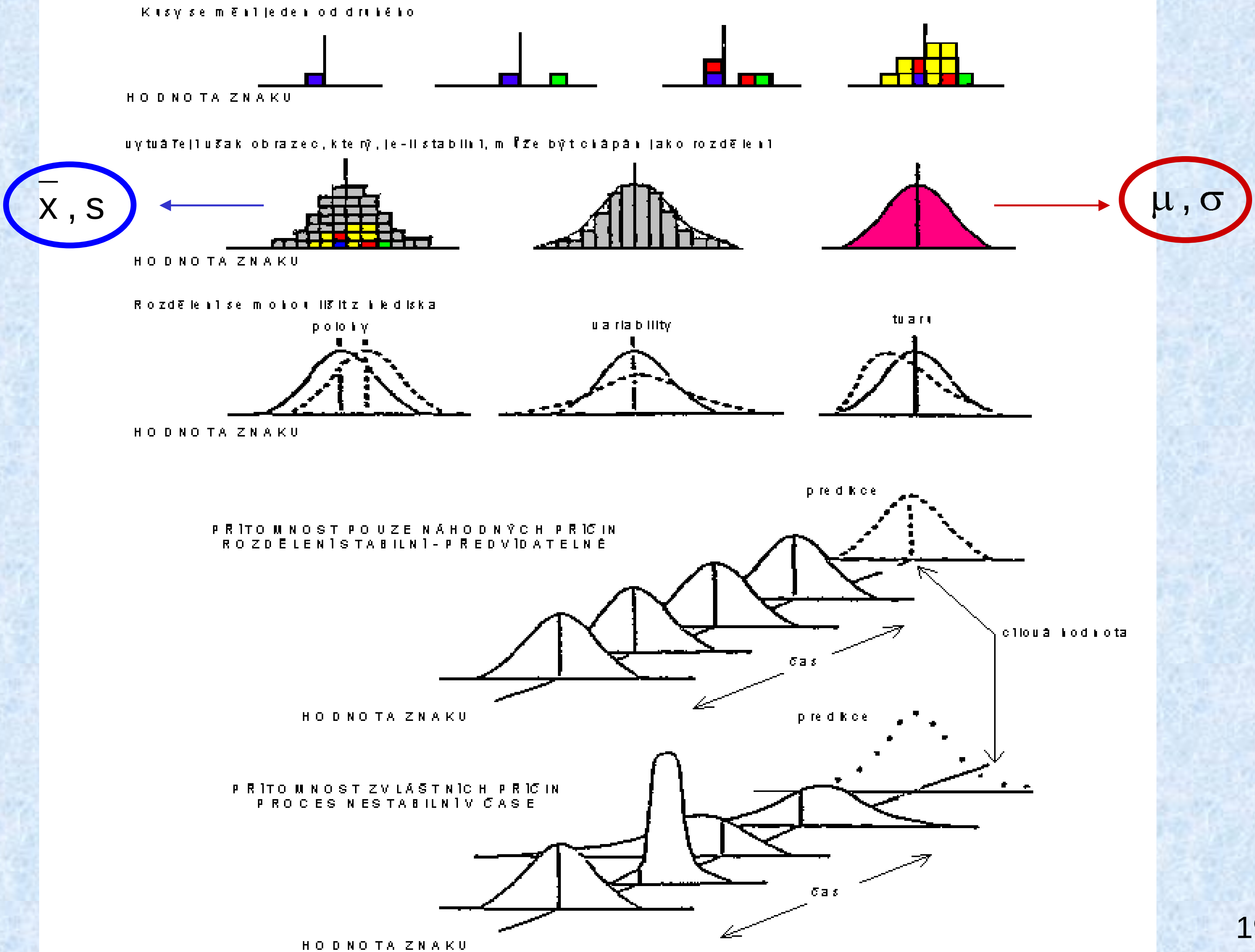
Výběrová směrodatná odchylka - s



- Výsledků náhodného výběru se používá k ohodnocení vlastností základního souboru. Náhodný výběr poskytuje informaci o neznámé realitě (vlastnostech základního souboru, které neznáme).

Teorie pravděpodobnosti a matematická statistika




HODNOTA ZNAKU
$\bar{x}$, s
μ , σ
HODNOTA ZNAKU
polohy
variability
tvaru
HODNOTA ZNAKU
predikce
PŘÍTOMNOST POUZE NÁHODNÝCH PŘÍČIN
ROZDĚLENÍ STABILNÍ - PŘEDVÍDATELNÉ
čas
HODNOTA ZNAKU
predikce
PŘÍTOMNOST ZVLÁŠTNÍCH PŘÍČIN
PROCES NESTABILNÍ V ČASE
čas
HODNOTA ZNAKU

Náhodné příčiny variability

Široká škála **neidentifikovatelných příčin variability** (proměnlivosti), z nichž každá přispívá jen nepatrně k celkové variabilitě, ale jejich součet je měřitelný. Jsou vlastním rysem procesu, jsou neodstranitelné a za jejich působení je proces stabilní a predikovatelný.

Zvláštní (vymezitelné) příčiny variability

Reálná změna ve výrobním procesu, zjistitelná z průběhu sledovaného znaku jakosti (např. pomocí regulačního diagramu). Nutno takovou příčinu variability, která není vlastní sledovanému procesu IDENTIFIKOVAT **=>** ODSTRANIT **=>** ZABEZPEČIT proti opakování.

Jedná se

a) o předpověditelné systematické vlivy způsobené fyzikálními zákonitostmi nebo plynoucí z praktických zkušeností, nebo

b) o vlivy poruch.

POZNÁMKA:

Náhodný výběr a logické podskupiny

NÁHODNÝ VÝBĚR - konečný soubor náhodně vybraných jednotek reprezentující základní soubor, ze kterého je náhodný výběr odebrán.

Např.: Náhodný výběr rozsahu n (může být větší) prověřovaný při statistické přejímce přejímané dávky.

LOGICKÁ PODSKUPINA - jednotky shromážděné za dokonale identických podmínek, nemusí být vybrány náhodně, ale záměrně.

Např.: Posledních n (obvykle 3 až 10) vyrobených jednotek při statistické regulaci.)

Poskytují informaci o **variabilitě uvnitř logických podskupin**, která je způsobena pouze náhodnými příčinami a tvoří východisko pro stanovení regulačních mezí k ověření krátkodobé stability a informaci o **variabilitě mezi logickými podskupinami**, která je obvykle způsobena zvláštními příčinami a tvoří východisko pro ověření dlouhodobé stability.

ZPRACOVÁNÍ DAT

Výpočet výběrových charakteristik

A. výpočet výběrových charakteristik přímo z napozorovaných hodnot

- rozsah výběru: **n**
- napozorované hodnoty: $\mathbf{x_1 , x_2 , ... , x_n}$

Charakteristiky polohy :

Výběrový průměr $\bar{x}$:

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

tj. $\bar{x} = (x_1 + x_2 + x_3 + \ \ldots \ + x_n) / n$

Výběrový medián Me :

– hodnoty uspořádané podle velikosti :
$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \ldots\ldots\ldots \leq x_{(n)}$

a) pro n liché, prostřední hodnota ;

b) pro n sudé, průměr dvou prostředních hodnot .

V případě a): $x_{(1)} \leq x_{(2)} \leq \mathbf{x_{(3)}} \leq x_{(4)} \leq x_{(5)}$ je medián $x_{(3)}$.

V případě b): $x_{(1)} \leq \mathbf{x_{(2)}} \leq \mathbf{x_{(3)}} \leq x_{(4)}$ je medián $(x_{(2)} + x_{(3)}) / 2$.

Výběrový modus Mo :

nejčetnější hodnota .

Uvažujme $x_{(1)} \leq \mathbf{x_{(2)}} = \mathbf{x_{(3)}} = \mathbf{x_{(4)}} \leq x_{(5)} \leq x_{(6)} \leq x_{(7)}$;

modus je $\mathbf{x_{(2)}}$ ($= x_{(3)} = x_{(4)}$) .

Charakteristiky variability :

Výběrový rozptyl s^2 :

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Výběrová směrodatná odchylka s :

$$s = \sqrt{s^2} \qquad \text{tj.} \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Po úpravě :

$$s = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]} = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]}$$

Poznámka:

Rozptyl statistického (základního) souboru s^2 :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Nejedná se o výběrový rozptyl vypočítaný z výběru několika náhodně vybraných jednotek z procesu nebo základního souboru,

ale o rozptyl vypočítaný ze všech prvků konečného statistického souboru.

Výběrové rozpětí R :

označíme $\mathbf{x_{min}}$ nejmenší $x_{(1)}$ hodnotu ve výběru

$\mathbf{x_{max}}$ největší $x_{(n)}$ hodnotu ve výběru rozsahu n

potom

$$\mathbf{R = x_{max} - x_{min}}$$

Schéma pro výpočet výběrových charakteristik :

i	x_i	x_i^2
1	x_1	x_1^2
2	x_2	x_2^2
3	x_3	x_3^2
atd.	atd.	atd.
n	x_n	x_n^2
součet	$\boxed{\sum_{i=1}^{n} x_i}$	$\boxed{\sum_{i=1}^{n} x_i^2}$

Příklad:

Uspořádané hodnoty:

(1)	1,1
(2)	1,3
(3)	1,3
(4)	1,4
(5)	1,6
(6)	1,7

i	x_i	x_i^2
1	1,7	2,89
2	1,4	1,96
3	1,6	2,56
4	1,1	1,21
5	1,3	1,69
6	1,3	1,69
Součet	**8,4**	**12,00**

Me = **1,35**

$\bar{x}$ = 8,4 / 6 = **1,4**

(Mo = **1,3**)

R = 1,7 - 1,1 = **0,6**

$s^2 = (1/5)\,(12{,}0 - (1/6)\,8{,}4^2) =$ **0,048**

$s = \sqrt{0{,}048} =$ **0,219**

B. výpočet výběrových charakteristik z hodnot seskupených do tříd

– rozsah výběru: **n**
– napozorované hodnoty: $\mathbf{x_1 , x_2 , ... , x_n}$
– počet tříd: **k**
– šíře třídy: **h**

Označíme pro j-tou třídu :

– $\mathbf{n_j}$ třídní četnost (absolutní)

– $\mathbf{f_j} = n_j / n$ relativní třídní četnost

– $\mathbf{N_j} = \sum_{i=1}^{j} n_i$ kumulovaná třídní četnost (absolutní)

– $\mathbf{F_j} = N_j / n$ kumulovaná relativní třídní četnost

– $\mathbf{z_j}$ = třídní znak (obvykle střed j-té třídy)

– $\mathbf{z_j + h/2}$ = horní mez j-té třídy

Schéma pro výpočet výběrových charakteristik :

i	z_j	n_j	$z_j n_j$	$z_j^2 n_j$
1	z_1	n_1	$z_1 n_1$	$z_1^2 n_1$
2	z_2	n_2	$z_2 n_2$	$z_2^2 n_2$
3	z_3	n_3	$z_3 n_3$	$z_3^2 n_3$
atd.	atd.	atd.	atd.	atd.
k	z_k	n_k	$z_k n_k$	$z_k^2 n_k$
Součet		$\sum_{j=1}^{k} n_j$	$\sum_{j=1}^{k} z_j n_j$	$\sum_{j=1}^{k} z_j^2 n_j$

$$n = \sum_{j=1}^{k} n_j$$

$$\bar{x} \cong \bar{z} = \frac{1}{n} \sum_{j=1}^{k} z_j n_j$$

$$s_x^2 \cong s_z^2 = \frac{1}{n-1} \left[\sum_{j=1}^{k} z_j^2 n_j - \frac{1}{n} \left(\sum_{j=1}^{k} z_j n_j \right)^2 \right]$$

Příklad:

Délka keramických tělísek

Výběr **n = 44**

7,77	7,69	7,73	7,75	7,32	7,69	7,77	7,49	7,69
7,60	7,64	7,84	7,68	7,72	7,87	7,72	7,73	7,59
7,80	7,99	7,88	7,67	7,77	7,68	7,99	7,71	7,84
7,65	7,77	7,70	7,69	7,75	7,74	7,72	7,88	7,61
7,80	7,84	7,81	7,70	7,91	7,79	8,03	7,73	

Seskupíme do tříd šíře **h = 0,1** , zvolíme třídní intervaly

j	**třídní interval**	**z_j**	**n_j**	**f_j**	**N_j**	**F_j**
1	7,30 - 7,40	7,35	1	0,022727	1	0,022727
2	7,40 - 7,50	7,45	1	0,022727	2	0,045455
3	7,50 - 7,60	7,55	2	0,045455	4	0,090909
4	7,60 - 7,70	7,65	12	0,272727	16	0,363636
5	7,70 - 7,80	7,75	17	0,386364	33	0,750000
6	7,80 - 7,90	7,85	7	0,159091	40	0,909091
7	7,90 - 8,00	7,95	3	0,068182	43	0,977273
8	8,10 - 8,20	8,05	1	0,022727	44	1,000000
Součet			44	1,000000		

Výpočet výběrových charakteristik $\bar{x}$ a s :

j	z_j	n_j	$z_j n_j$	$z_j^2 n_j$
1	7,35	1	7,35	54,0225
2	7,45	1	7,45	55,5025
3	7,55	2	15,10	114,0050
4	7,65	12	91,80	702,2700
5	7,75	17	131,75	1021,0625
6	7,85	7	54,95	431,3575
7	7,95	3	23,85	189,6075
8	8,05	1	8,05	64,8025
Součet		44	340,30	2632,6300

$$n = \sum_{j=1}^{k} n_j = 44$$

$$\bar{x} \cong \bar{z} = \frac{1}{n} \sum_{j=1}^{k} z_j n_j =$$

= 340,30 / 44 = **7,734**

$$s_x^2 \cong s_z^2 = \frac{1}{n-1} \left[\sum_{j=1}^{k} z_j^2 n_j - \frac{1}{n} \left(\sum_{j=1}^{k} z_j n_j \right)^2 \right] =$$

= (1/43)(2632,63 - 340,30^2 / 44) = **0,01672**

$$s_x \cong s_z = \sqrt{s_z^2} = \sqrt{0{,}01672} =$$ **0,1293**

ZPRACOVÁNÍ DAT

Histogram

Histogram – grafické znázornění dat seskupených do tříd

napozorované hodnoty $x_1, x_2, \ldots , x_n$
náhodný výběr rozsahu **n**

Konstrukce histogramu:

počet tříd ***k*** stejné šíře ***h*** ;

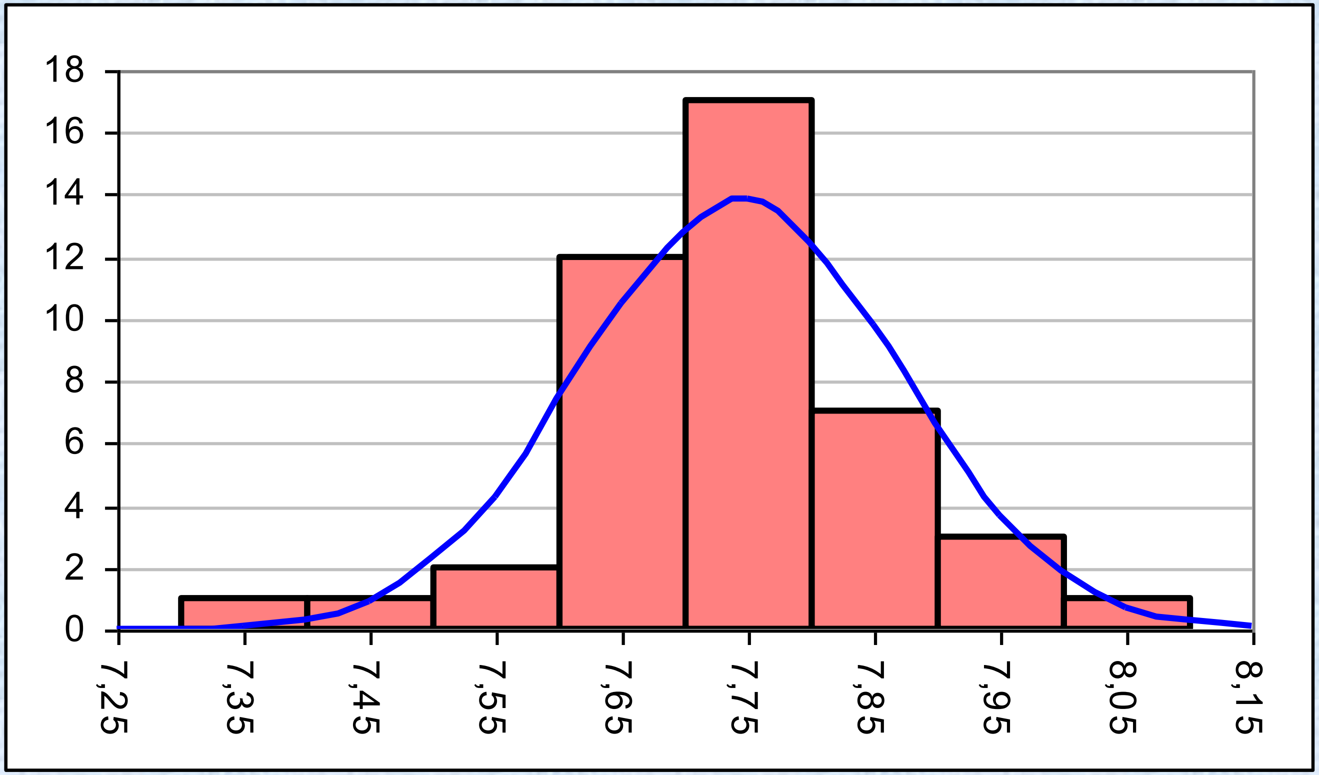
zjistí se absolutní třídní četnosti n_j ,případně relativní třídní četnosti f_j ;

na osu **x** se vynesou **hranice** třídních intervalů, případně třídní znaky z_j ;

na osu **y** se vynáší **třídní četnosti** n_j (absolutní) nebo f_j (relativní); nad třídními intervaly se sestrojí obdélníky.

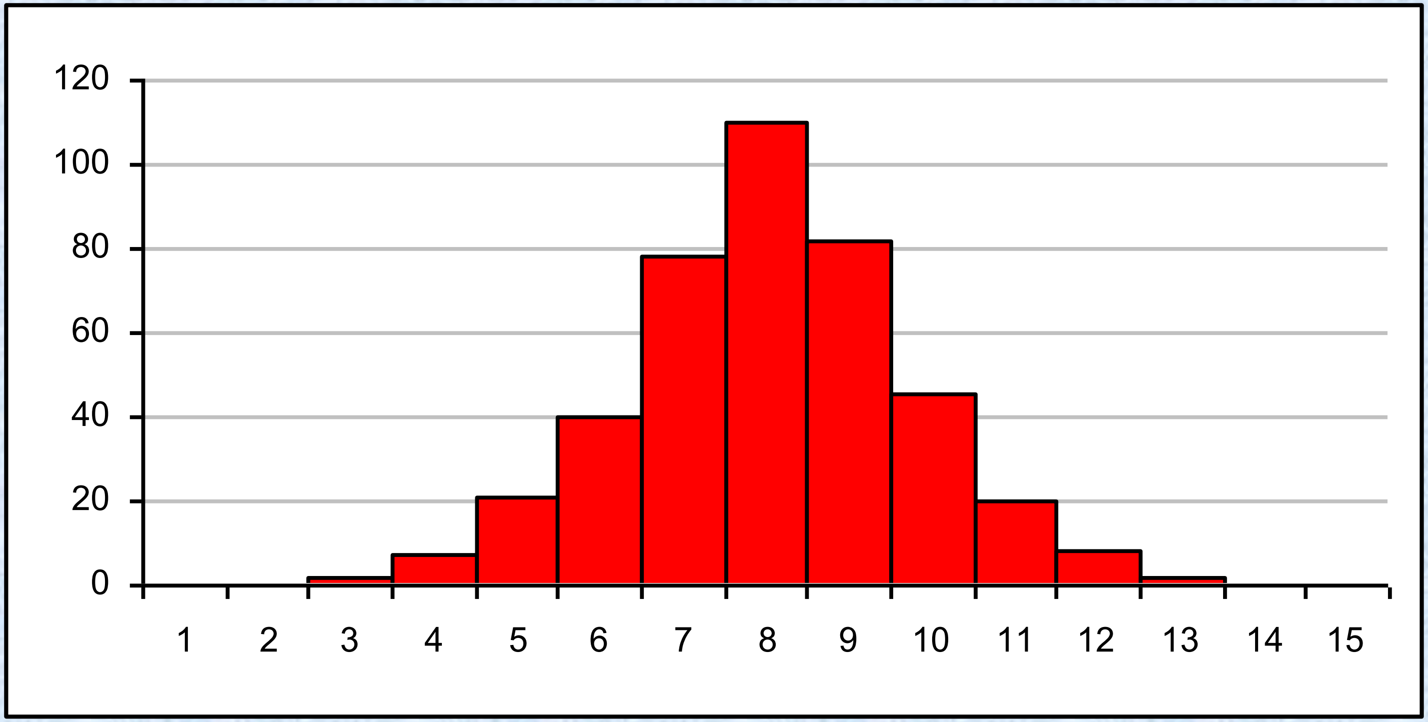
Příklad :

j	třídní interval	z_j	n_j	f_j
1	7,30 - 7,40	7,35	1	0,022727
2	7,40 - 7,50	7,45	1	0,022727
3	7,50 - 7,60	7,55	2	0,045455
4	7,60 - 7,70	7,65	12	0,272727
5	7,70 - 7,80	7,75	17	0,386364
6	7,80 - 7,90	7,85	7	0,159091
7	7,90 - 8,00	7,95	3	0,068182
8	8,10 - 8,20	8,05	1	0,022727
Součet			44	1,000000

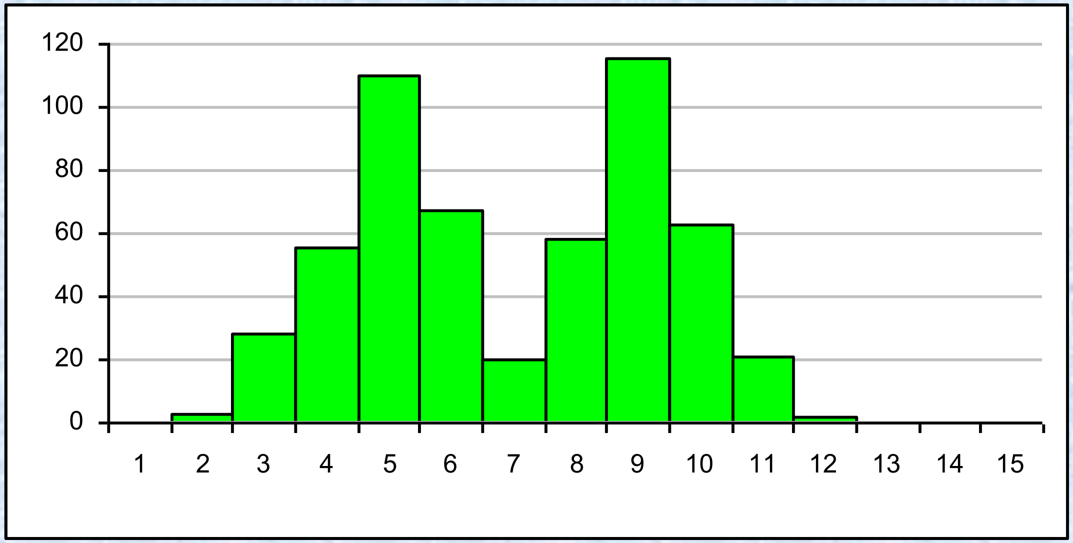


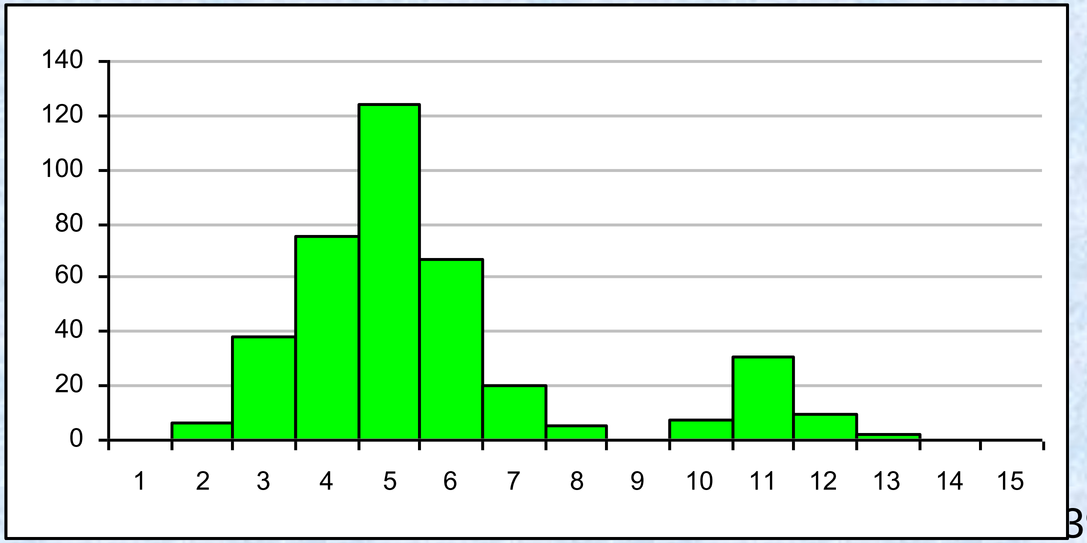
Ukázky některých základních typů histogramů

a) Symetrický histogram zvonovitého tvaru

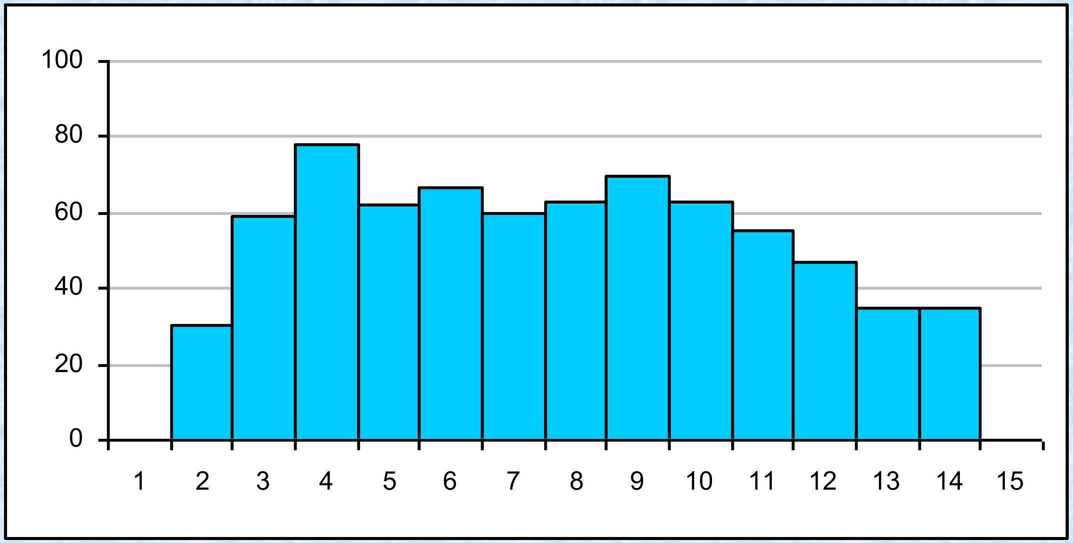


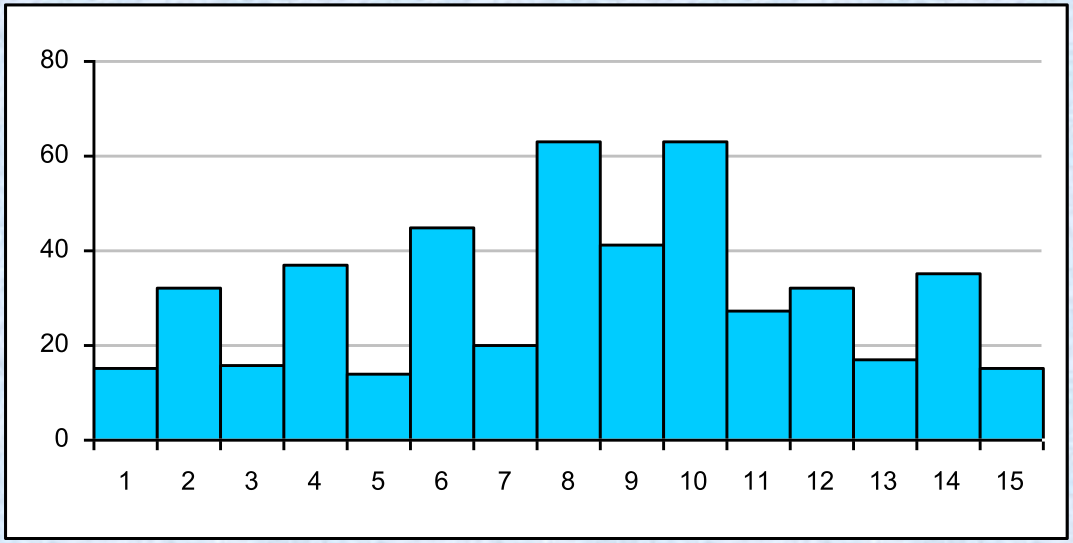
b) Dvojvrcholové histogramy



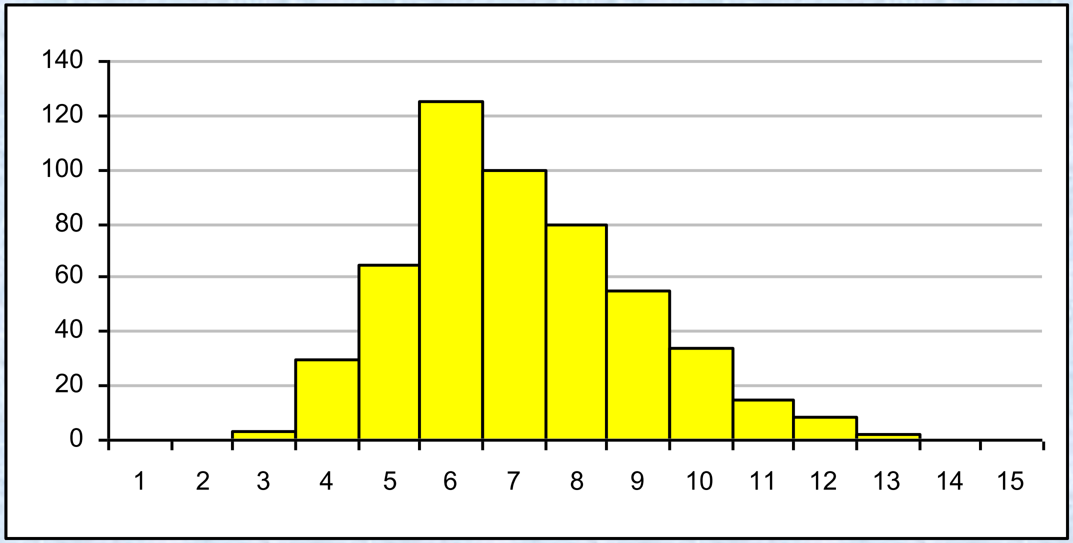


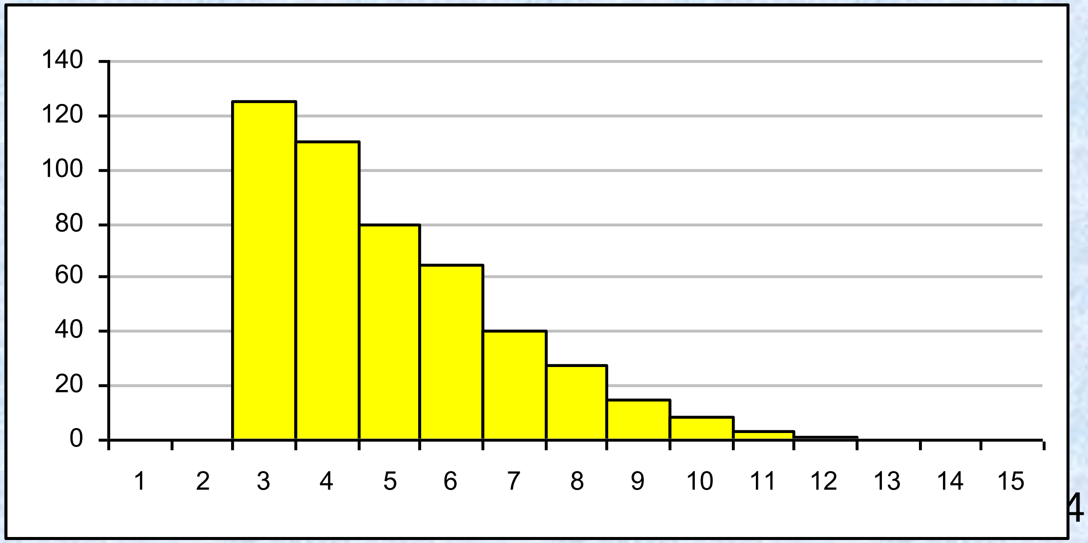
c) Histogramy plochého a hřebenovitého tvaru



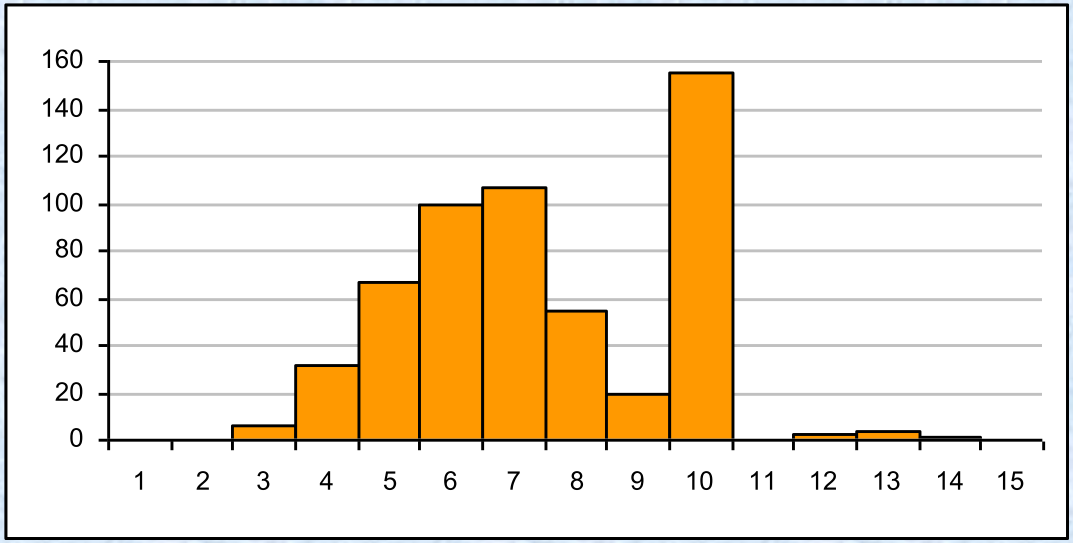


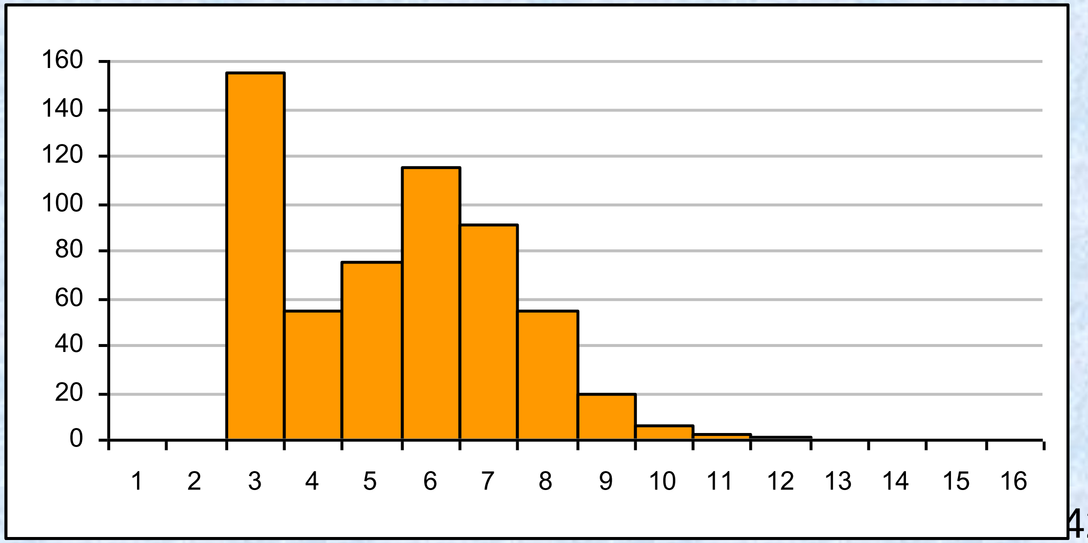
d) Histogramy asymetrického tvaru

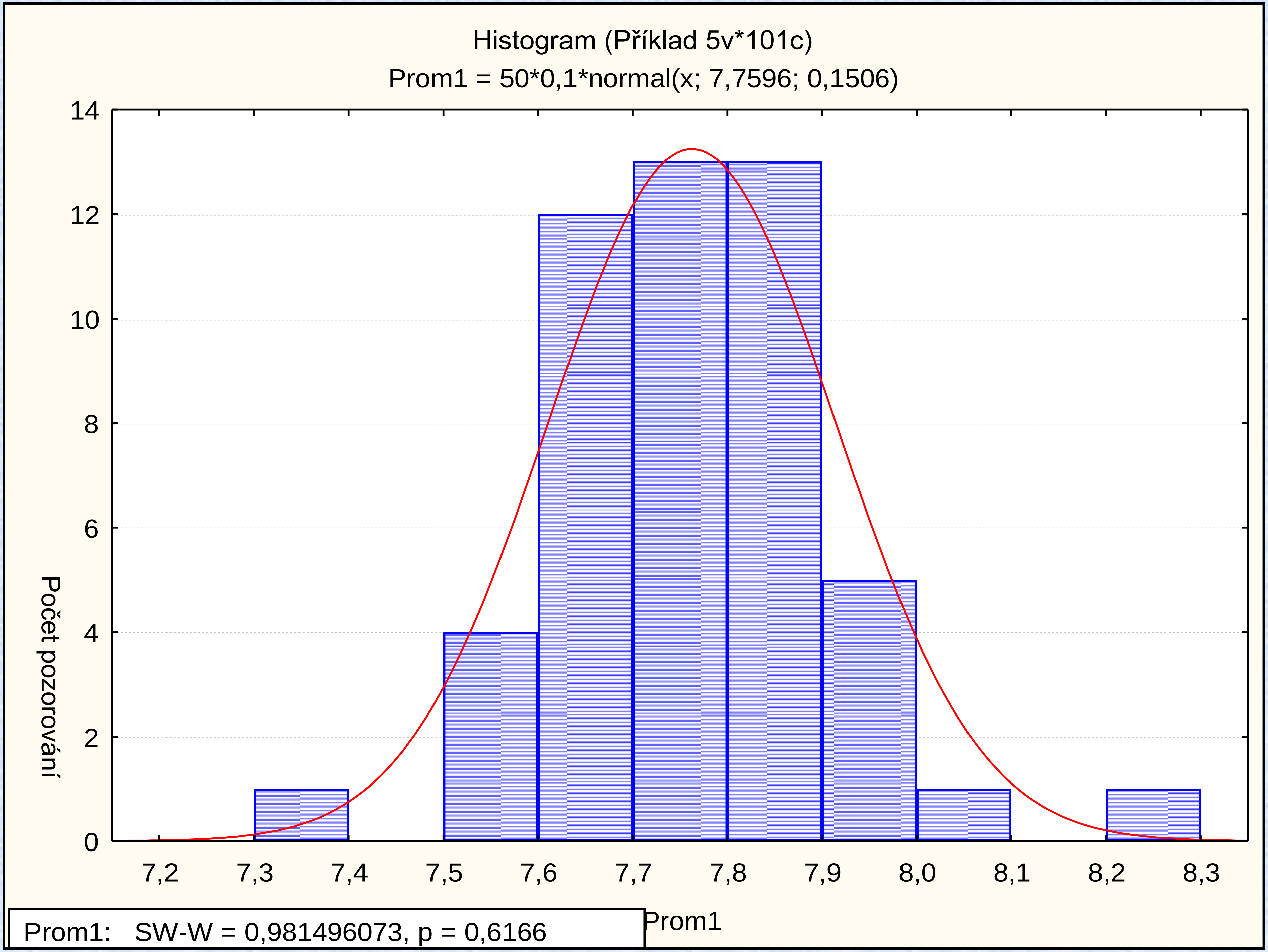




e) Dvojvrcholové histogramy s výraznou četností v krajní třídě



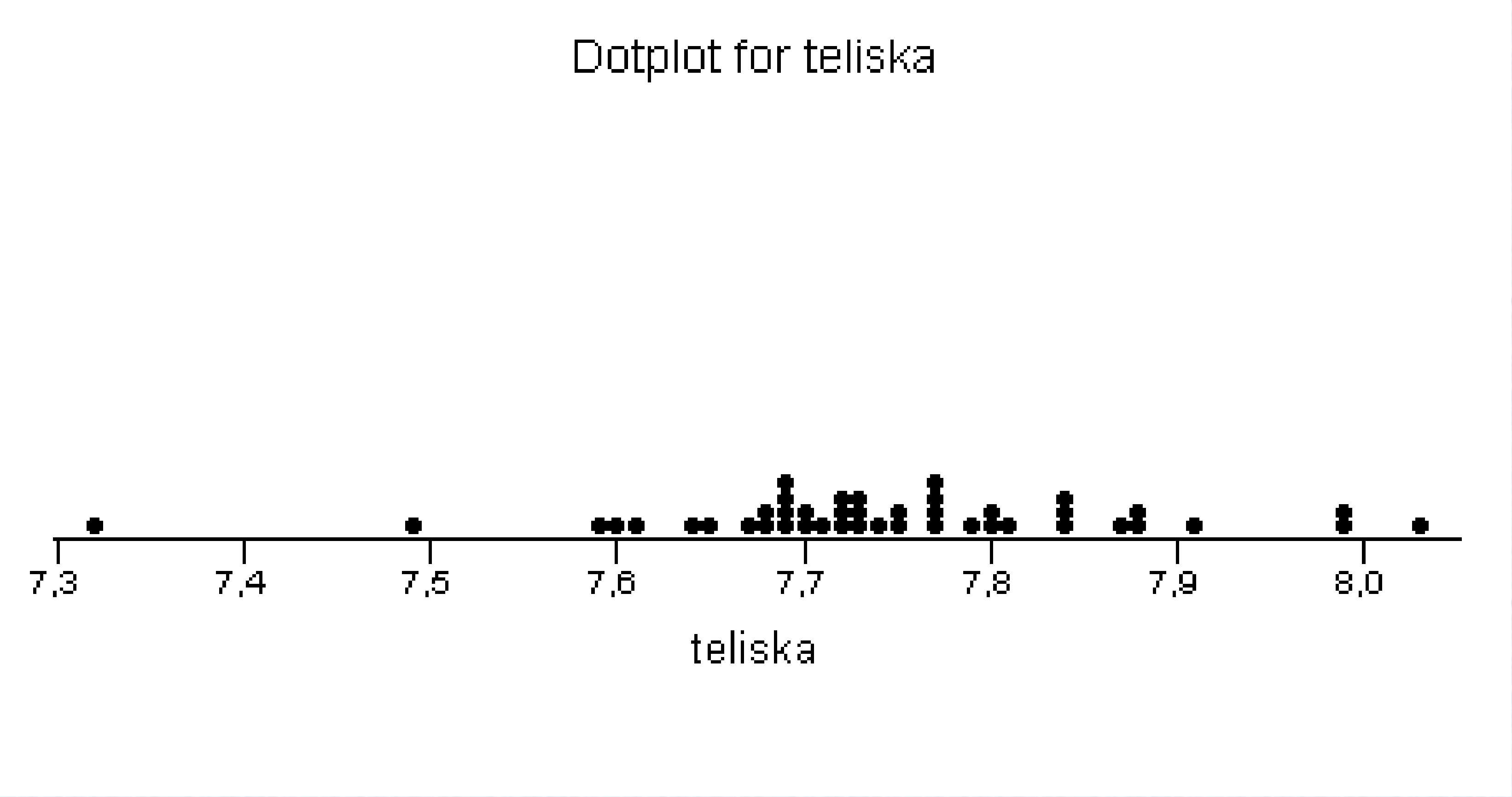



Histogram (Příklad 5v*101c)
Prom1 = 50*0,1*normal(x; 7,7596; 0,1506)
14
12
10
8
6
4
2
0
Počet pozorování
7,2
7,3
7,4
7,5
7,6
7,7
7,8
7,9
8,0
8,1
8,2
8,3
Prom1
Prom1: SW-W = 0,981496073, p = 0,6166

ZPRACOVÁNÍ DAT

Bodový diagram

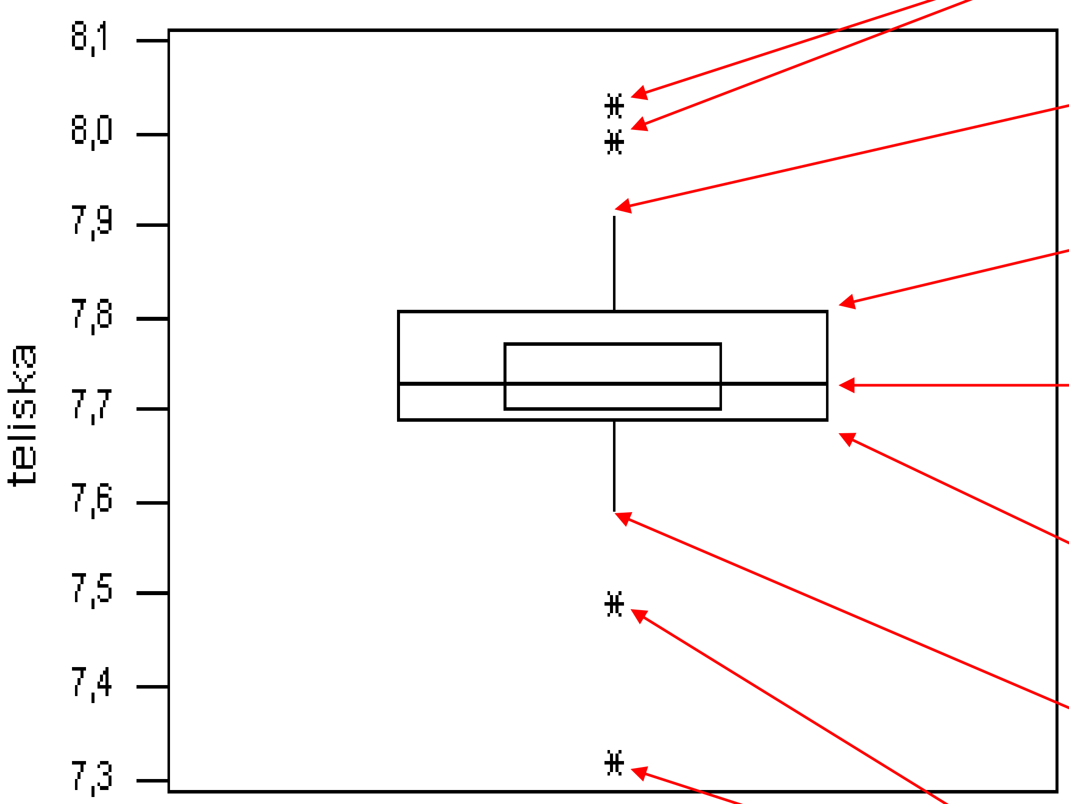
(Dot Plot)


Dotplot for teliska
7,3
7,4
7,5
7,6
7,7
7,8
7,9
8,0
teliska

ZPRACOVÁNÍ DAT

Krabicový diagram

(Box Plot)



Nejvíce a nejméně vzdálené hodnoty ležící nad horní mezí

Max. hodnota ležící uvnitř meze,
Horní mez = Q3 + 1.5 (Q3 – Q1)

Q3, třetí kvartil (75 %)
hodnota menší nebo rovná 75 %

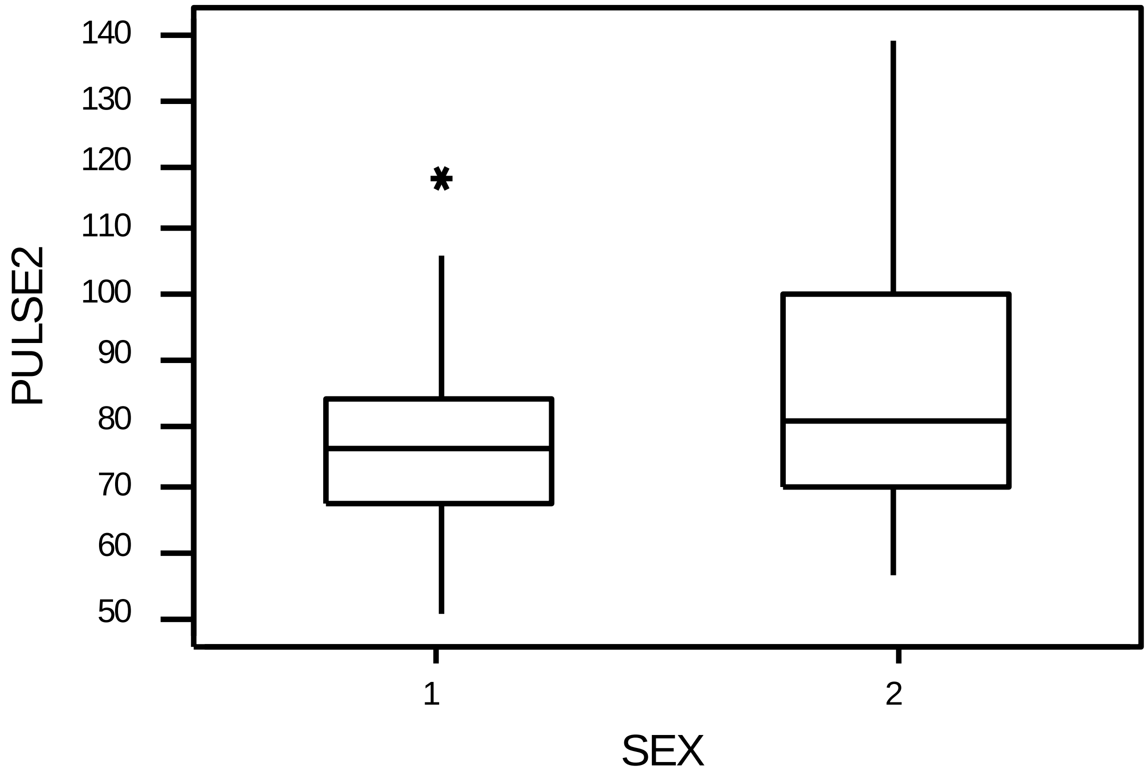
Medián, prostřední hodnota
Hodnota odpovídající druhému kvartilu Q2 (50 %)

Q1, první kvartil (25 %)
hodnota menší nebo rovná 25 %

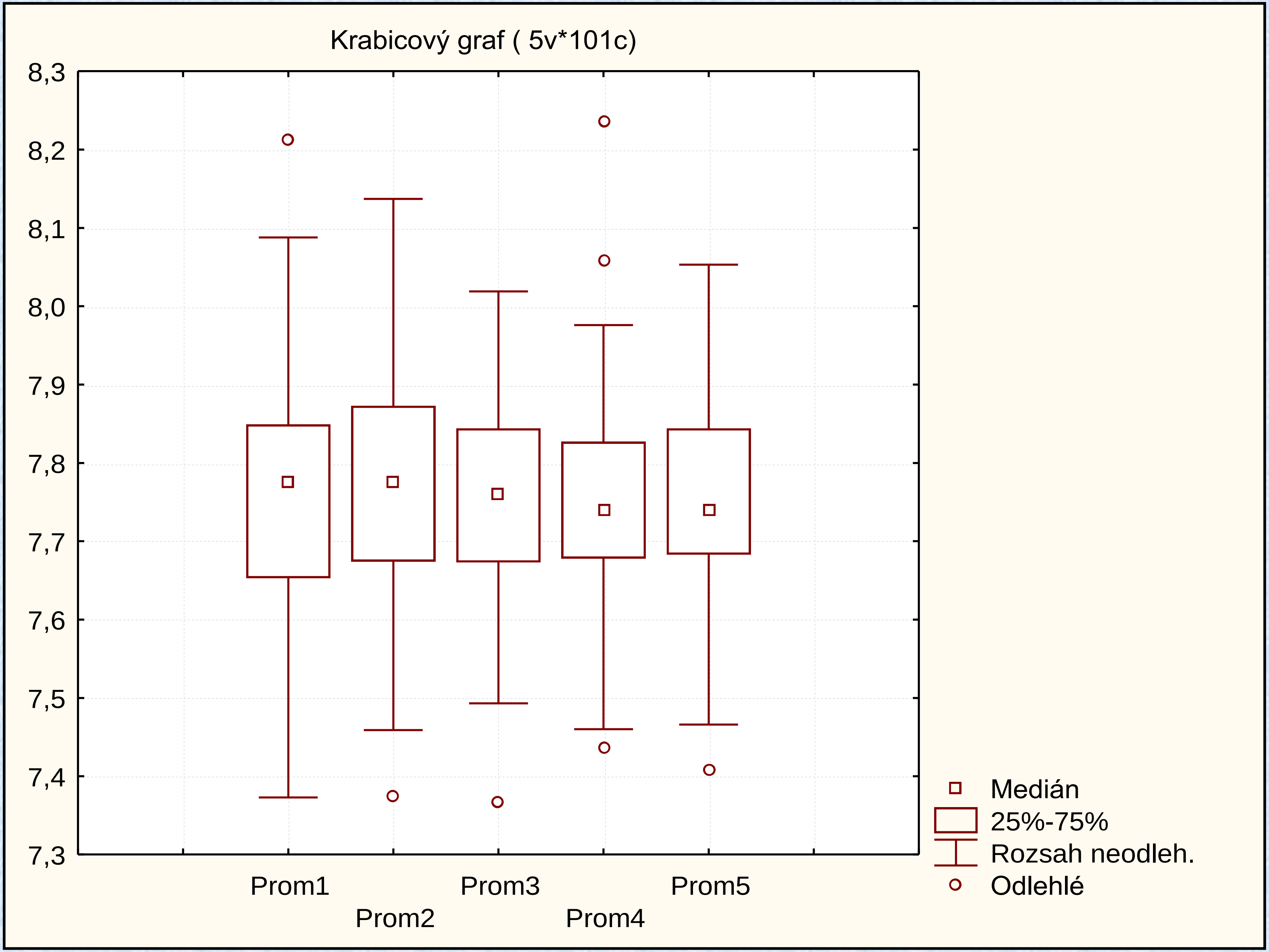
Min. hodnota ležící uvnitř meze,
Dolní mez = Q1 - 1.5 (Q3 – Q1)

Nejvíce a nejméně vzdálené hodnoty ležící pod dolní mezí

Konfidenční interval pro střední hodnotu (malý vnitřní obdélník)



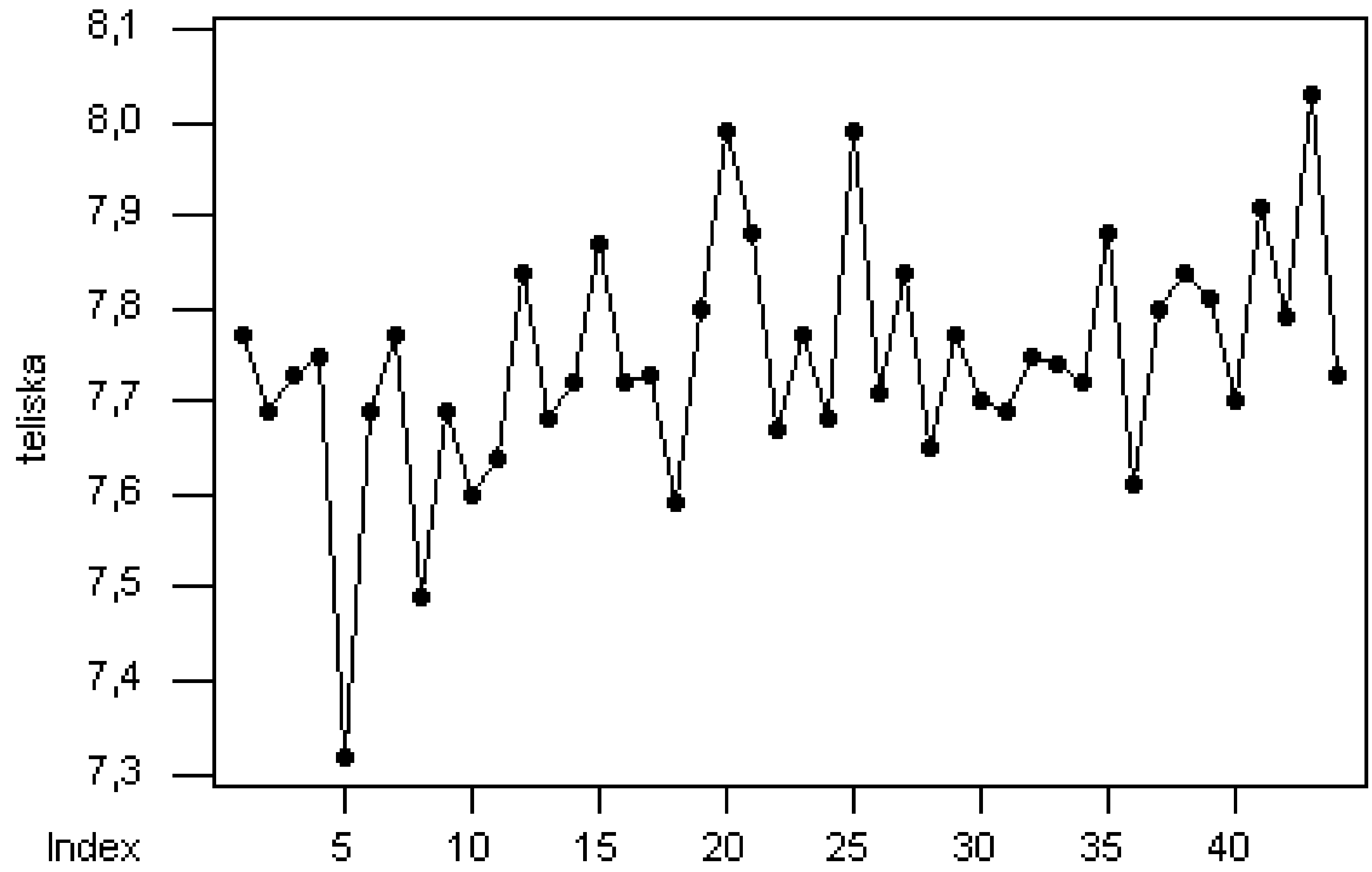
Porovnání srdečního tepu po zátěži podle pohlaví

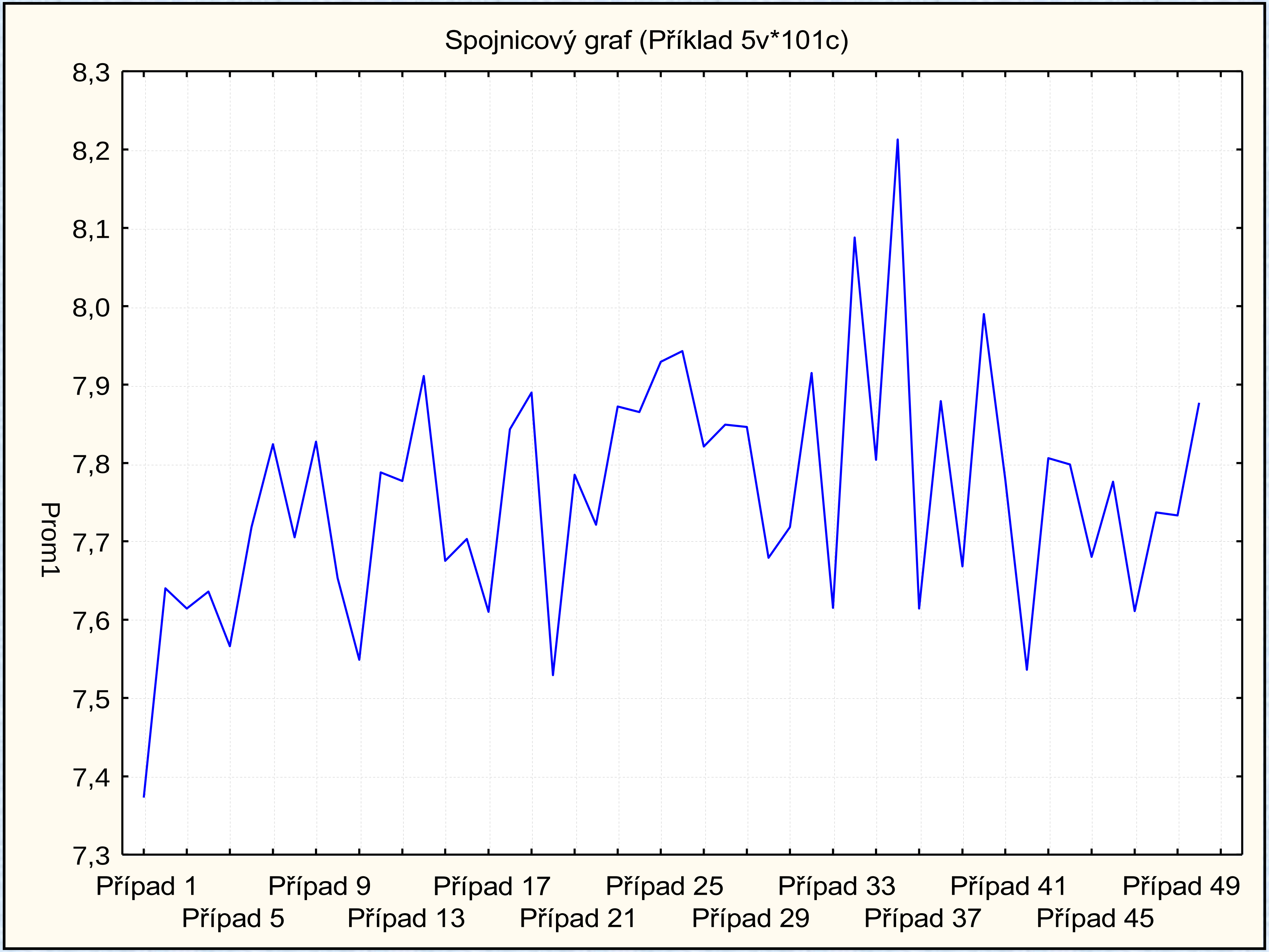

Krabicový graf (5v*101c)
8,3
8,2
8,1
8,0
7,9
7,8
7,7
7,6
7,5
7,4
7,3
Prom1
Prom2
Prom3
Prom4
Prom5
Medián
25%-75%
Rozsah neodleh.
Odlehlé

ZPRACOVÁNÍ DAT

Diagram časového průběhu

(Time Series Plot)

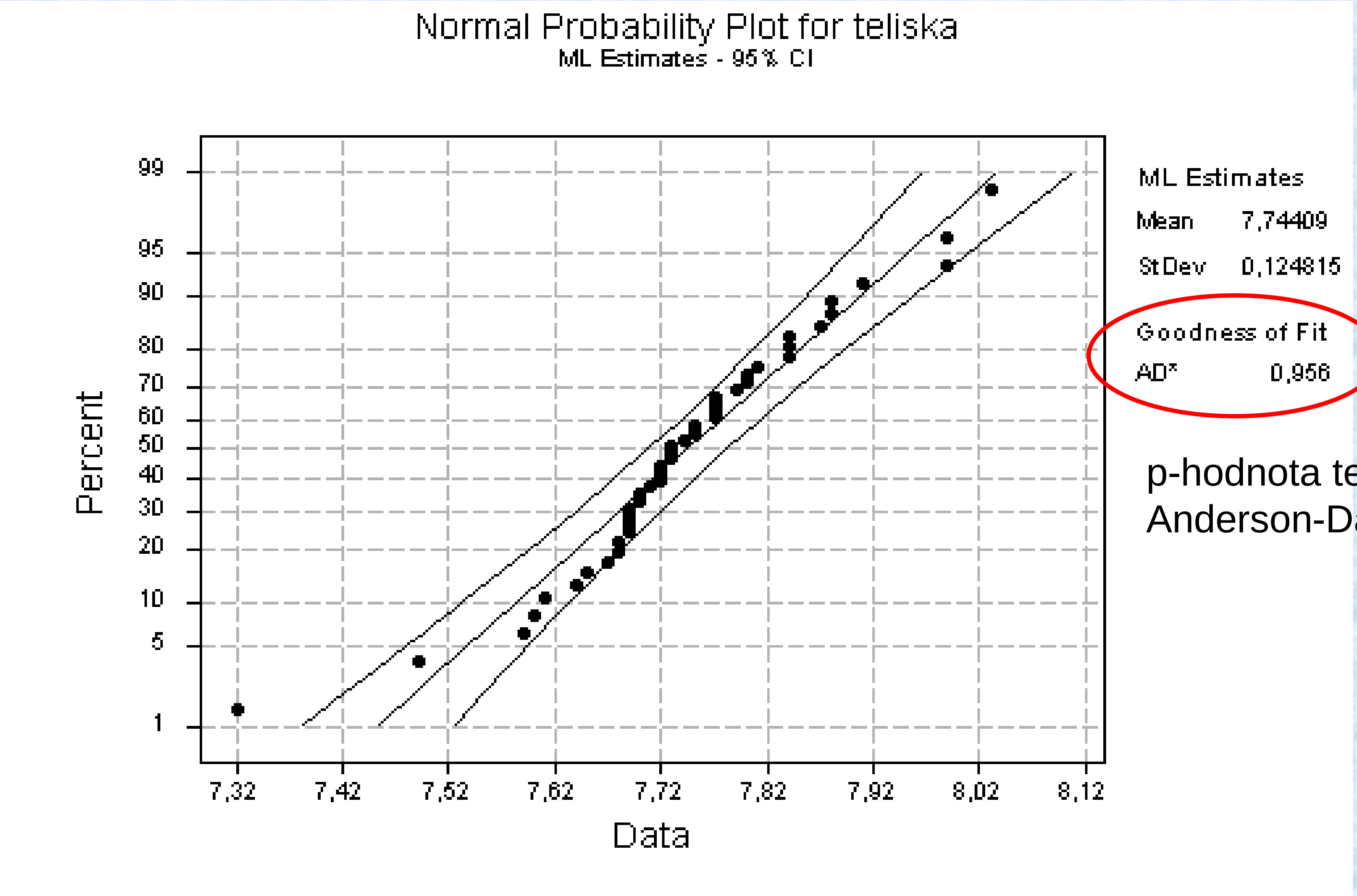

8,1
8,0
7,9
7,8
7,7
7,6
7,5
7,4
7,3
teliska
Index
5
10
15
20
25
30
35
40


Spojnicový graf (Příklad 5v*101c)
Prom1
8,3
8,2
8,1
8,0
7,9
7,8
7,7
7,6
7,5
7,4
7,3
Případ 1
Případ 5
Případ 9
Případ 13
Případ 17
Případ 21
Případ 25
Případ 29
Případ 33
Případ 37
Případ 41
Případ 45
Případ 49

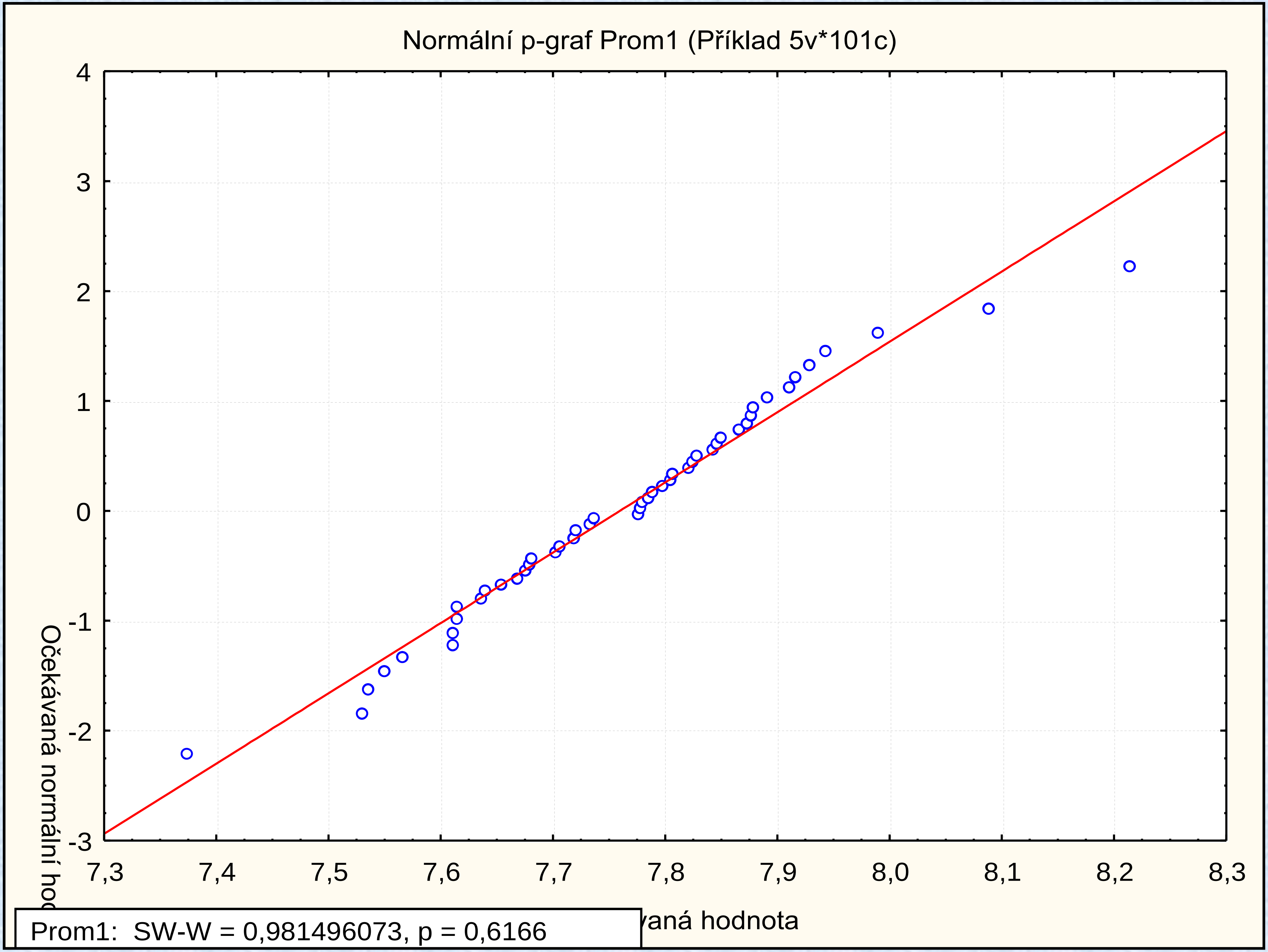
ZPRACOVÁNÍ DAT

Pravděpodobnostní diagram

(Probability Plot)



p-hodnota testu Anderson-Darling


Normální p-graf Prom1 (Příklad 5v*101c)
4
3
2
1
0
-1
-2
-3
7,3
7,4
7,5
7,6
7,7
7,8
7,9
8,0
8,1
8,2
8,3
Očekávaná normální hod
vaná hodnota
Prom1: SW-W = 0,981496073, p = 0,6166

ZPRACOVÁNÍ DAT

Korelační (bodový) diagram

(Scatter Plot)

Korelační (bodový) diagram zobrazuje *body jako dvojici souřadnic (x,y),* odpovídající dvěma proměnným:

- jedna proměnná určuje souřadnici na svislé ose (y)
- druhá proměnná určuje souřadnici na vodorovné ose (x)

Budeme uvažovat sílu desek z nichž každá má odlišný počet vrstev. Pro každou desku je počet vrstev zapsán do samostatného sloupce tabulky. K interpretaci napozorovaných údajů bude použit korelační diagram.

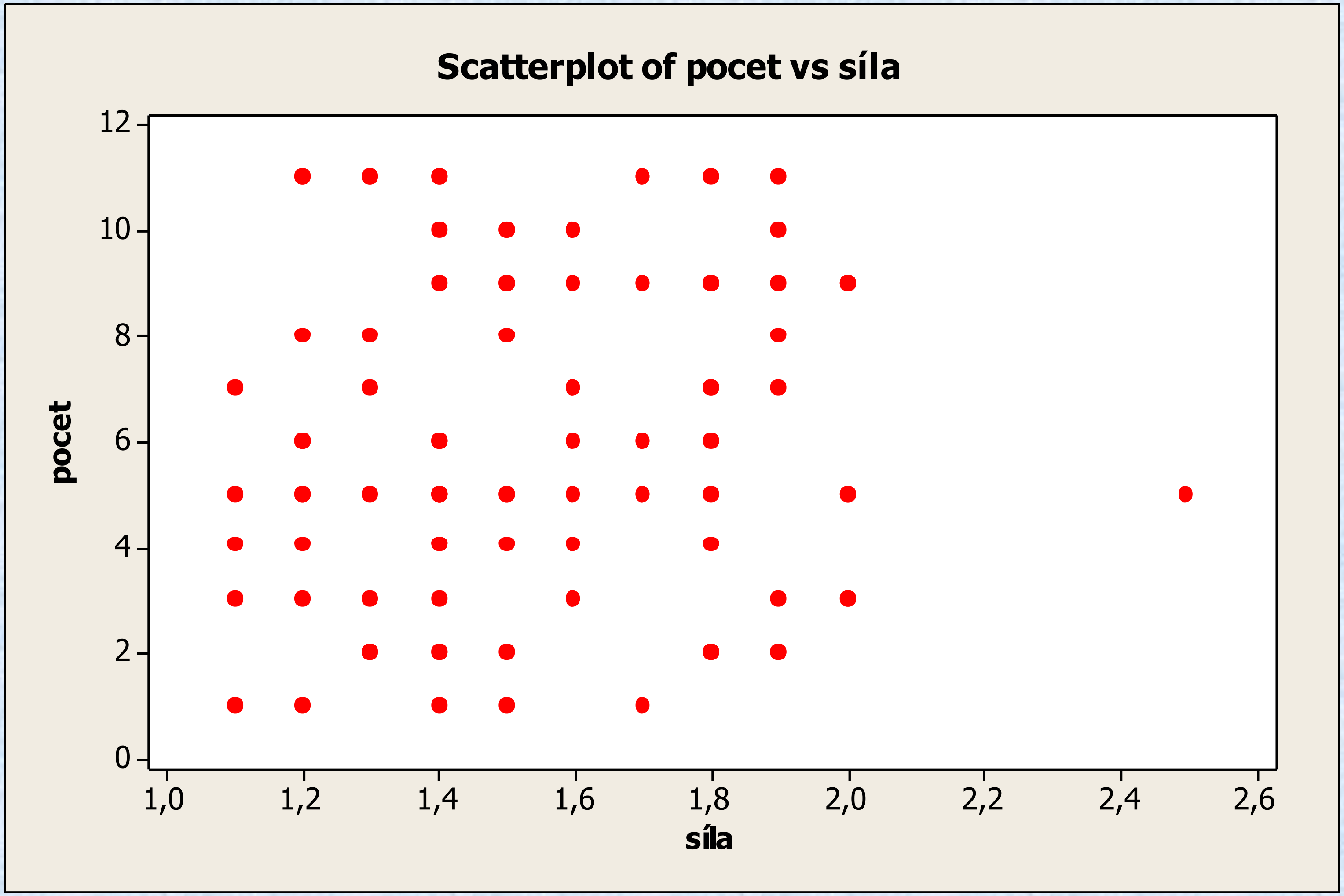
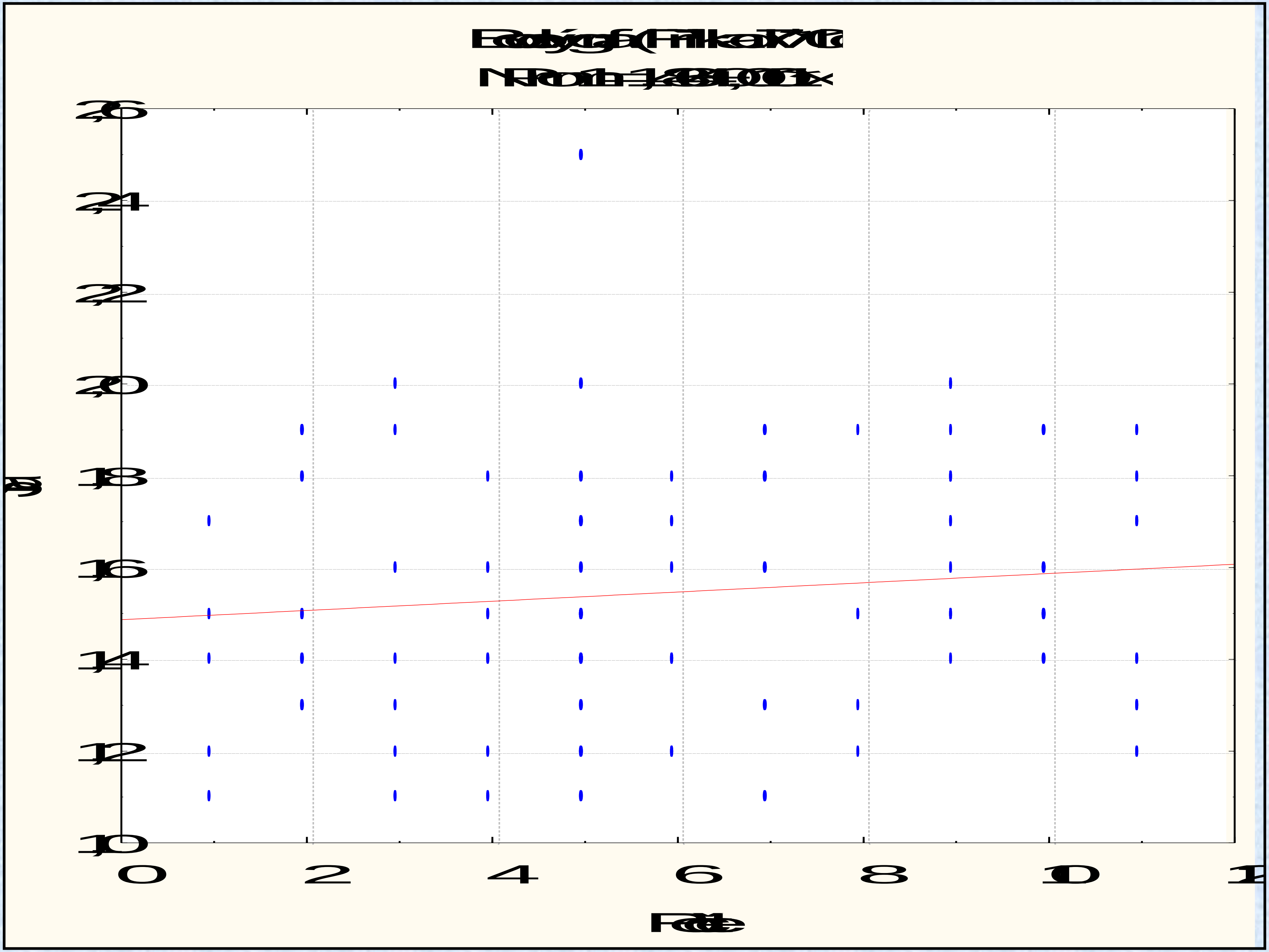
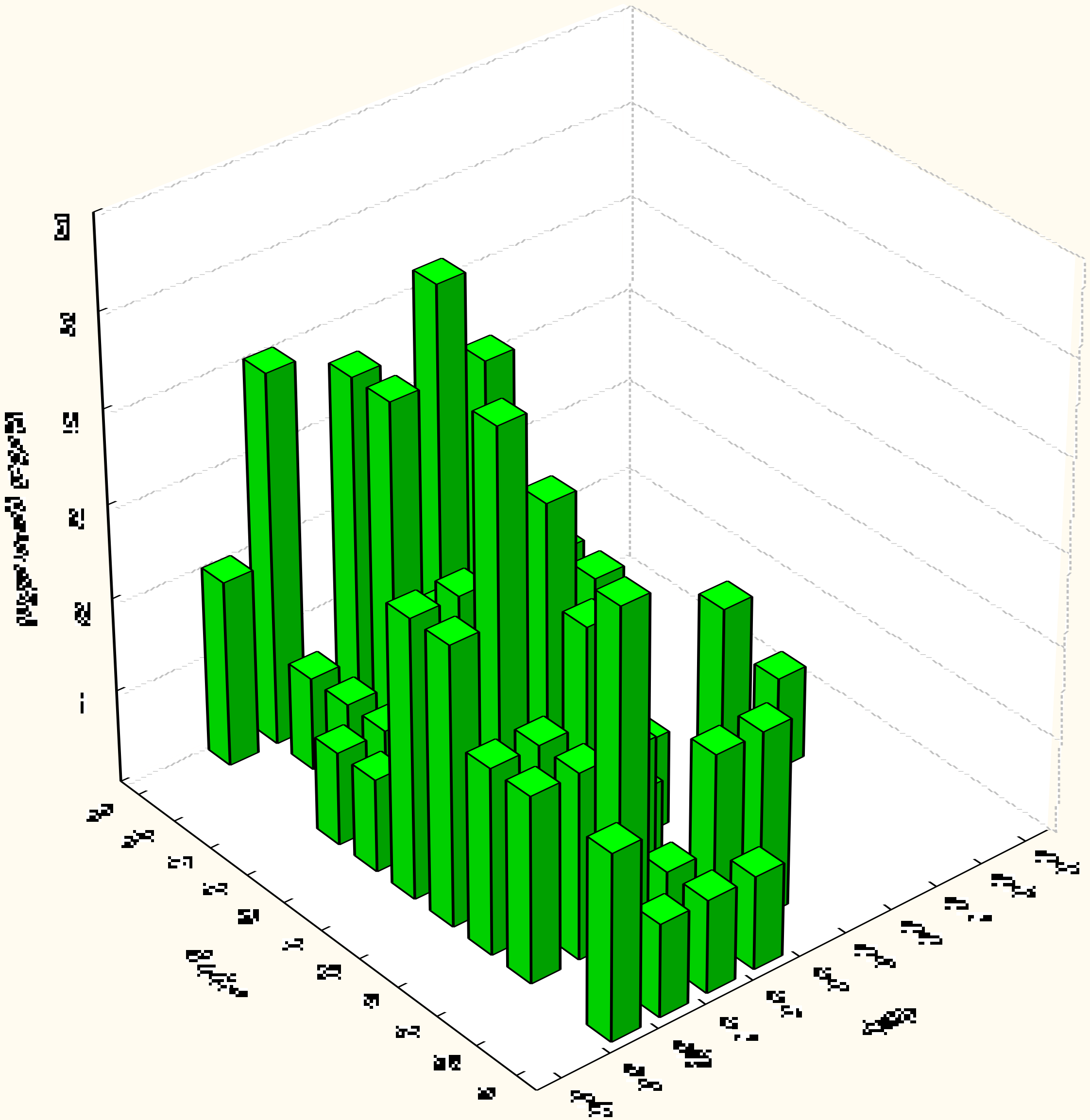


Diagram ukazuje, že není vztah (korelace) mezi sílou desky a počtem vrstev.





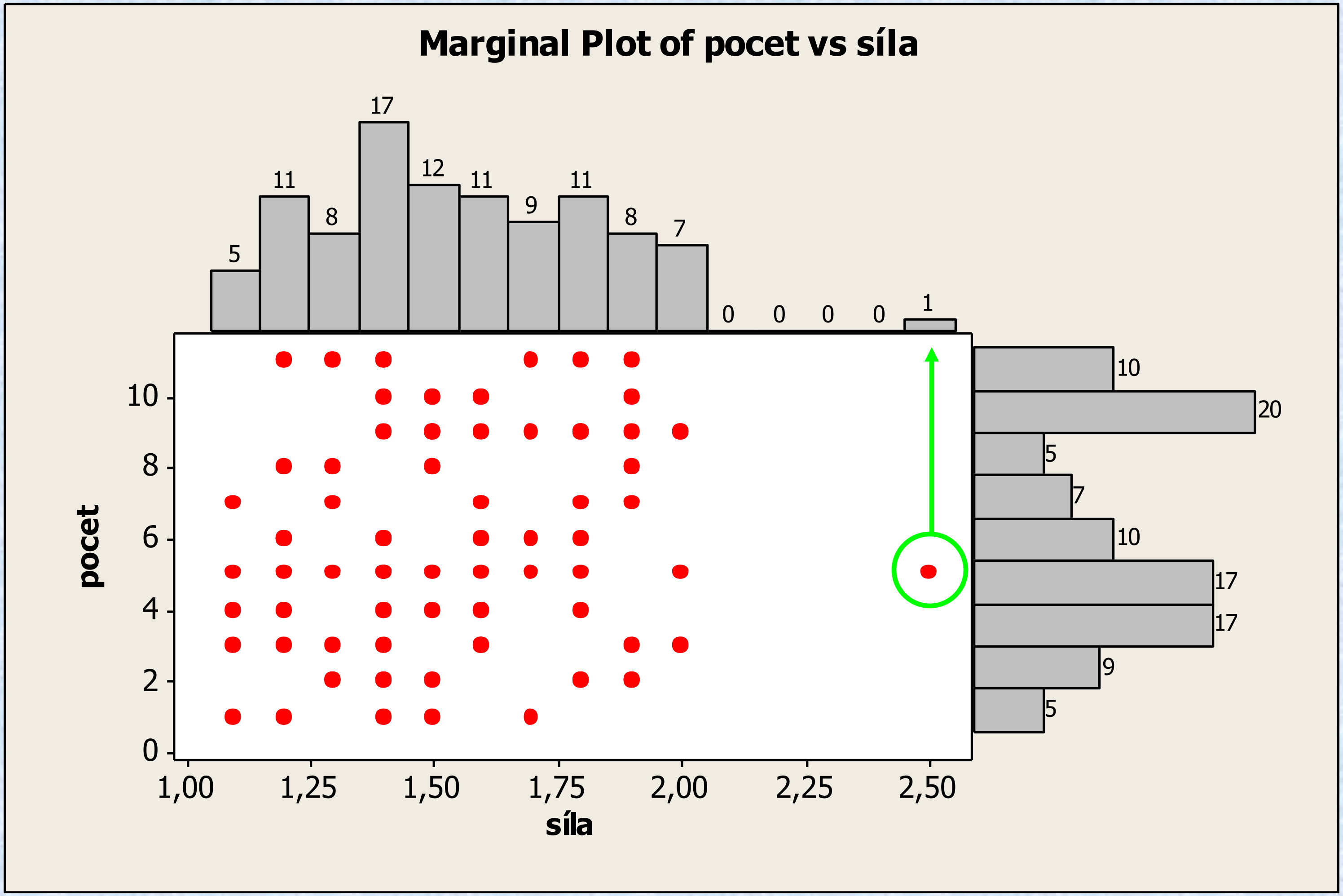
ZPRACOVÁNÍ DAT

Marginální diagram

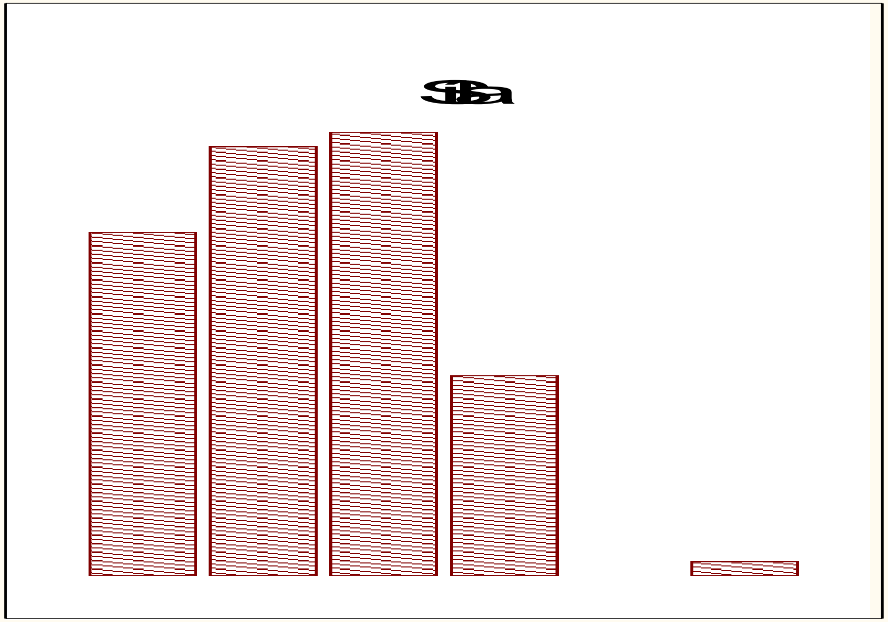
(Marginal Plot)

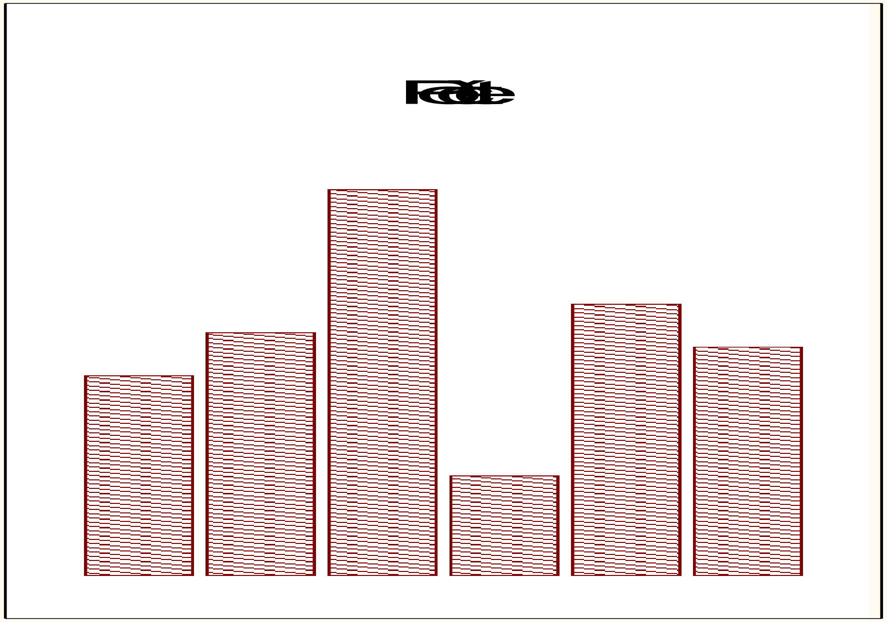
„Marginální” diagram je korelační diagram s histogramy marginálního rozdělení četností na stranách. Histogramy odpovídají výběrovému rozdělení jedné resp. druhé náhodné veličiny zakreslované na osu x, resp. na osu y.

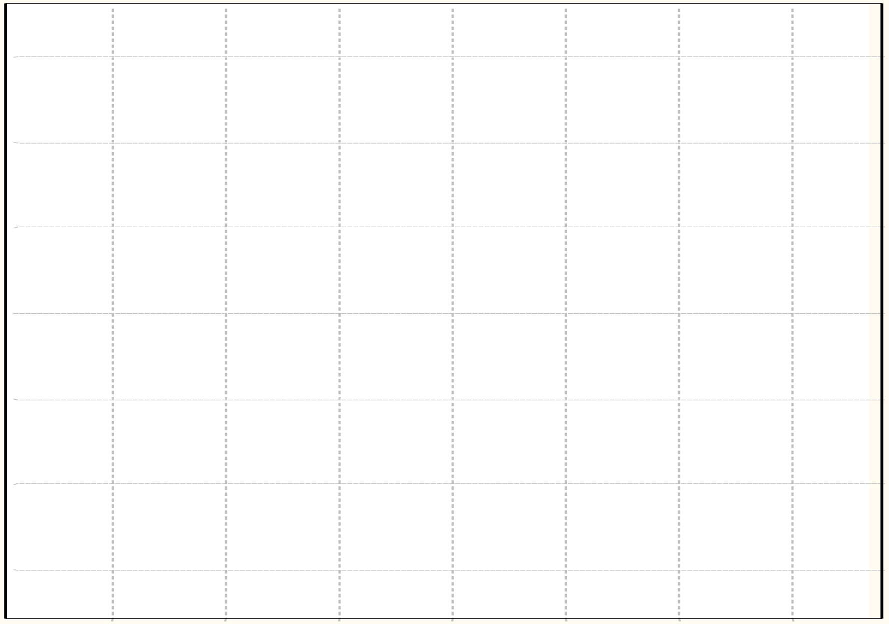
Budeme opět uvažovat sílu desek z nichž každá má odlišný počet vrstev.



Vidíme, že síla desky je soustředěna okolo 1,4 mm s jednou hodnotou ležící mimo většinu napozorovaných hodnot.







Nadstavbové diagramy pro analýzu dat

Diagram hlavních efektů

(Main Effects Plot)

Diagram hlavních efektů

použijeme když několik faktorů ovlivňuje sledovanou náhodnou veličinu (*výsledek na výstupu procesu*) je vhodné použít.

Body v tomto diagramu jsou zobrazeny průměry sledované náhodné veličiny při různých verzích (úrovních) každého z faktorů. Diagram hlavních efektů je vhodný pro jejich vzájemné porovnání.

Referenční přímka je zakreslena pro celkový průměr napozorovaných dat.

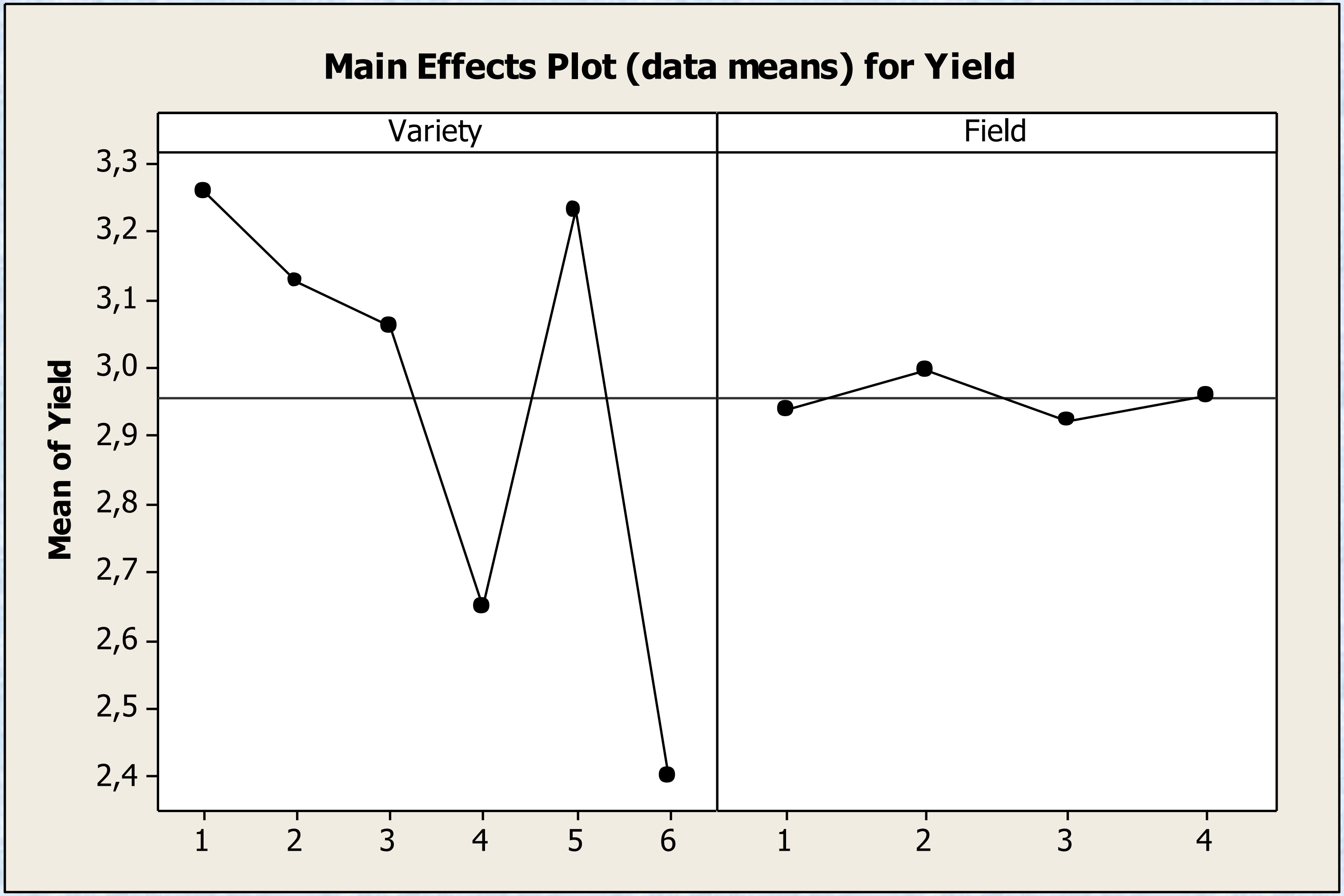
Diagram hlavních efektů a diagram interakcí jsou standardními metodami při použití metod DOE.

Budeme uvažovat případ, kdy na čtyřech různých parcelách je pěstováno šest odrůd vojtěšky. Je zjišťována hmotnost - výnos (posečená hmota) každé odrůdy na každé z parcel. Zajímá nás porovnání výnosu jednotlivých odrůd. Chceme s použitím diagramu hlavních efektů posoudit výnos jednotlivých odrůd a vliv polí.

Faktor A - odrůda - má šest úrovní (1, 2, 3, 4, 5, 6)

Faktor B - pole - má čtyři úrovně (1, 2, 3, 4)

Odezvová (responsní) veličina - výnos, posečená hmota



Z diagramu je vidět, že odrůdy vojtěšky (Variety) mají větší vliv na průměrný výnos (Mean of Yield) než výběr pole (Field).

Diagram interakcí

(Interactions Plot)

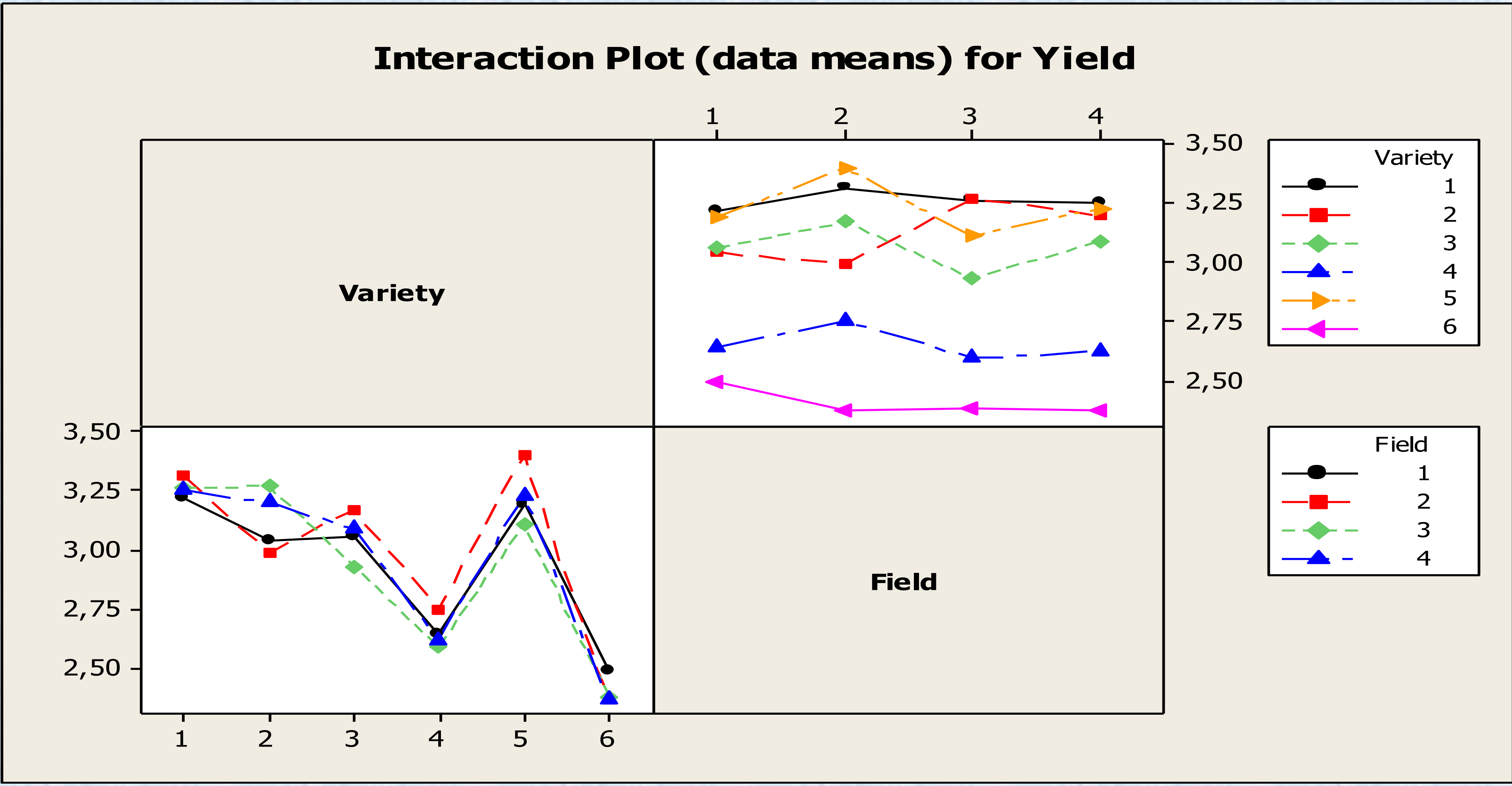
Graf interakcí je graf zobrazující průměry každé úrovně jednoho faktoru při konstantní úrovni druhého faktoru. Rovnoběžné přímky ukazují, že interakce mezi faktory neexistuje.

Budeme opět uvažovat případ, kdy na čtyřech různých polích je pěstováno šest odrůd vojtěšky. Je zjišťována hmotnost - výnos (posečená hmota) každé odrůdy na každé z parcel.

Faktor A - odrůda - má šest úrovní (1, 2, 3, 4, 5, 6)

Faktor B - pole - má čtyři úrovně (1, 2, 3, 4)

Odezvová (responsní) veličina - výnos, posečená hmota.



Zde je vidět, že interakce existuje, protože přímky nejsou rovnoběžné. Interakce znamená, že efekt jednoho faktoru závisí na úrovni druhého faktoru, tj.tyto dva faktory nejsou nezávislé.

Maticový diagram

(Draftsman Plot;

Casement Display;

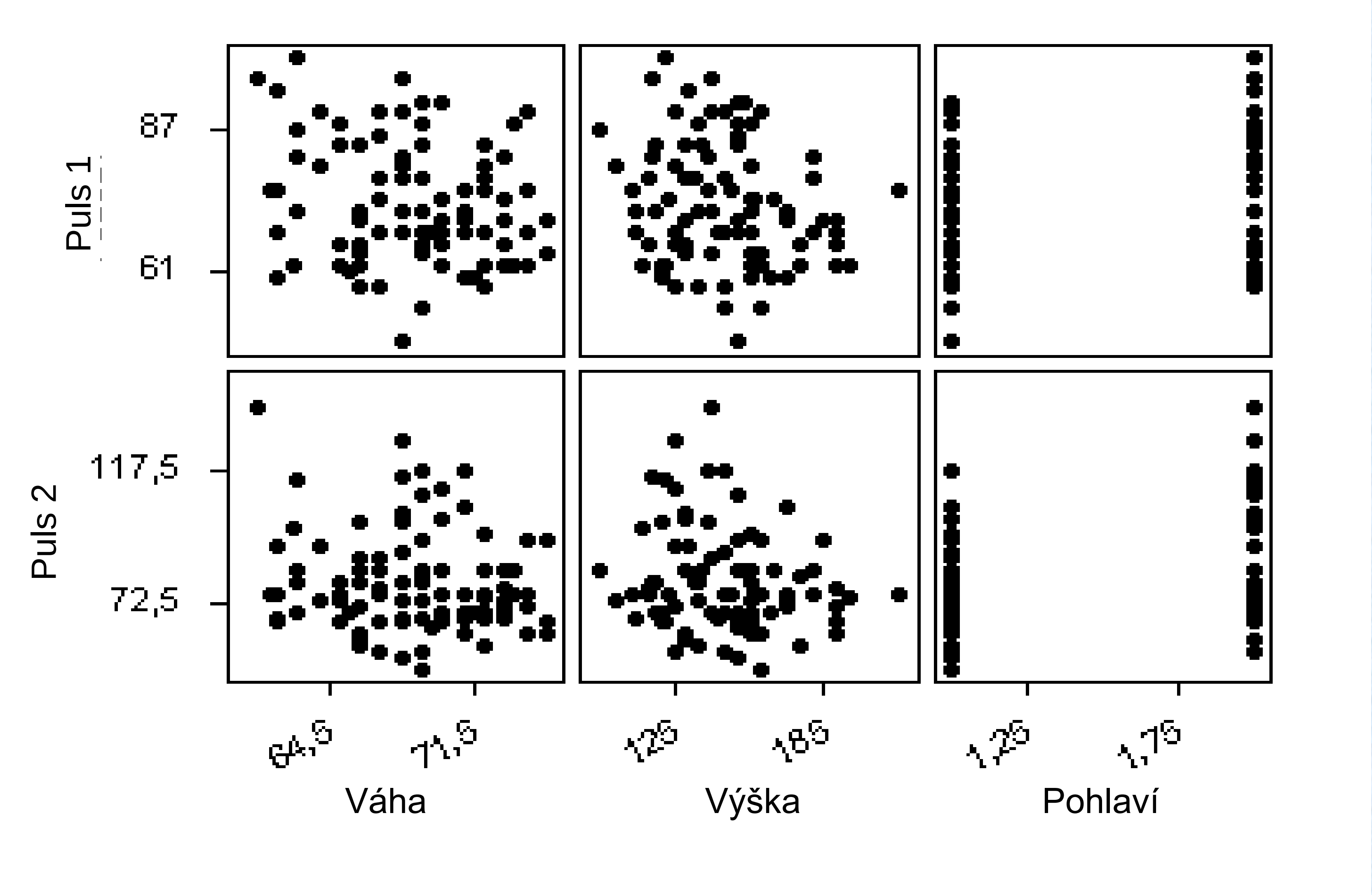
Matrix Plot)

Maticový diagram (*Matrix plot*)
Projekční diagram (*Draftsman diagram*)
Okenní diagram (*Casement Display*)

je dvourozměrná matice složená z jednotlivých grafů.

Maticové diagramy jsou velice vhodné pro sledování vztahu dvou proměnných mezi několika najednou sledovanými proměnnými.

Budeme sledovat srdeční tep u vyšetřovaných osob vzhledem k pohlaví (1, 2), váze a výšce před a po určité zátěži (Puls 1 a Puls 2).



Jedná se v podstatě o bodové diagramy zpracované pro každou ze vstupních veličin. V tomto případě je zřejmé, že pohlaví 2 má vyšší puls než pohlaví 1

„ Multi-vari “ diagramy

(Multi-vari Charts)

Multi - Vari diagramy

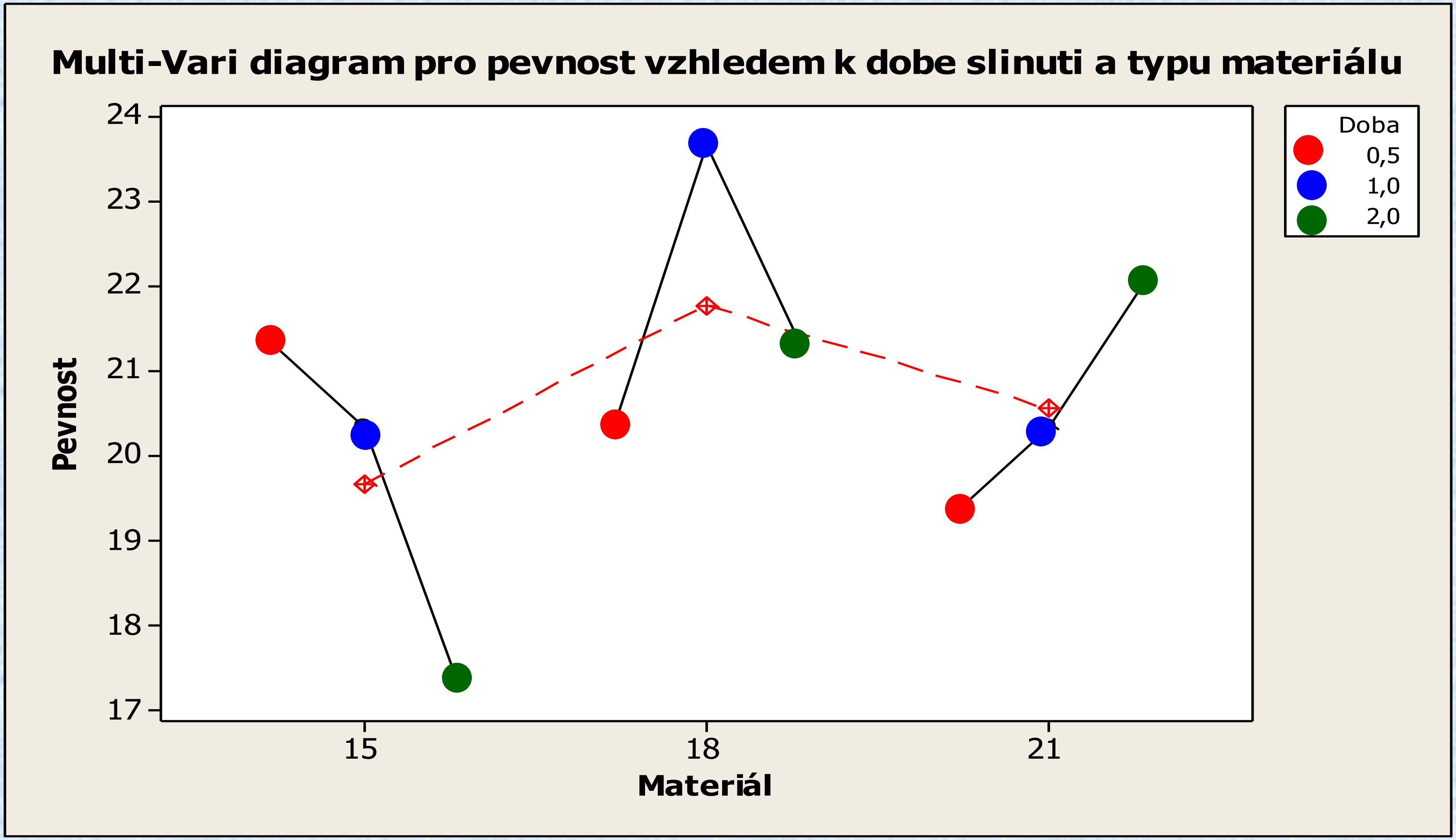
umožňují prezentovat grafickou formou analýzu rozptylu, poskytují grafickou alternativu k **numerické analýze rozptylu**.

Je to užitečný nástroj k identifikaci hlavních zdrojů variability.

Umožňuje analyzovat vlivy diskrétních proměnných X na spojité proměnné Y.

Je třeba vyhodnotit vliv doby slinutí třech různých kovových materiálů (15, 18, 21) na jejich pevnost. Pevnost byla měřena na třech vzorcích pro každý materiál a pro dobu slinutí: 0,5; 1 a 2 hodiny.

Dříve, než by se prováděla celá rozsáhlá číselná analýza, je vhodné se podívat pomocí Multi-Vari diagramu na napozorovaná data, zda ***nevykazují zřejmé efekty či interakce.***



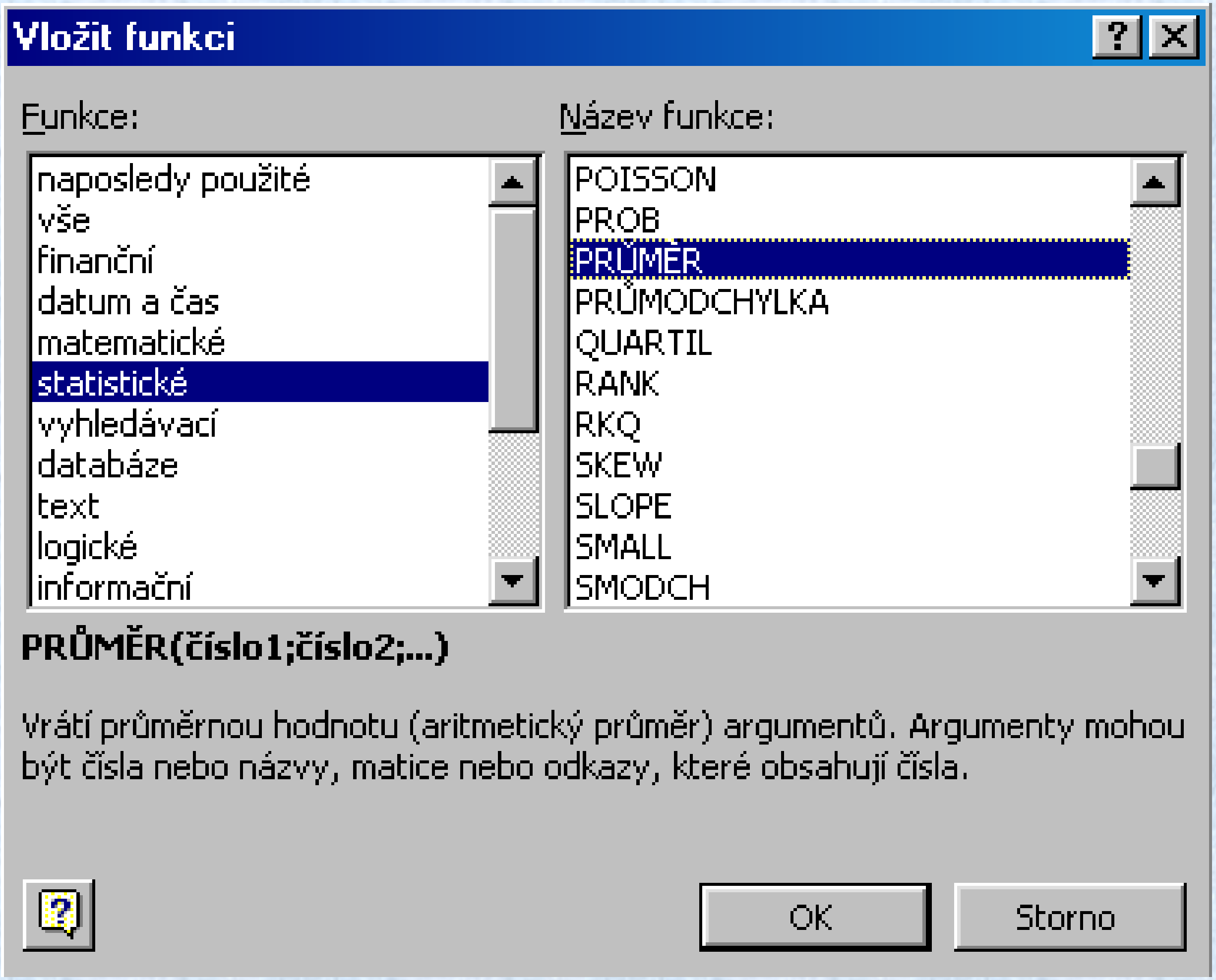
Můžeme vidět, že pro každý typ materiálu budeme mít jinou křivku doby slinutí.

ZPRACOVÁNÍ DAT

Výpočty s podporou Microsoft Excel

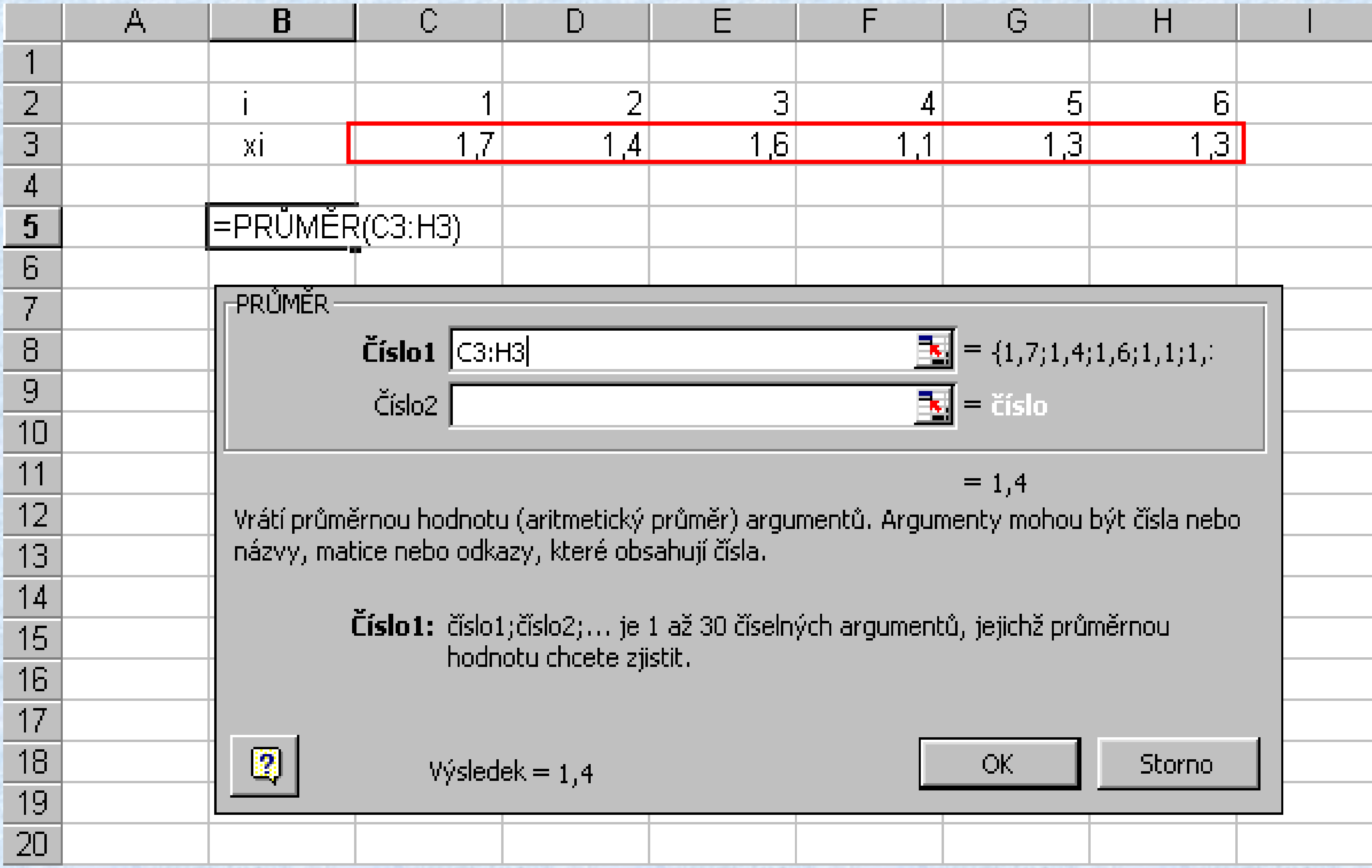
Výpočty s podporou Microsoft Excel

Použití statistických funkcí:



Příklad: výpočet výběrového průměru pomocí funkce:

PRŮMĚR(číslo1;číslo2;...)



Příklad: Výpočet výběrového průměru pomocí funkce

PRŮMĚR(číslo1;číslo2;...) - **výsledek**

B5 = =PRŮMĚR(C3:H3)

	A	B	C	D	E	F	G	H
1								
2		i	1	2	3	4	5	6
3		xi	1,7	1,4	1,6	1,1	1,3	1,3
4								
5		1,4						
6								

Vybrané funkce popisné statistiky:

Výběrový průměr - **PRŮMĚR**(číslo1; číslo2; …)

Výběrový medián - **MEDIA**N(číslo1; číslo2; …)

Výběrový modus - **MODE**(číslo1; číslo2; …)

Směrodatná odchylka stat. souboru - **SMODCH**(číslo1; číslo2; …)

Výběrová směrodatná odchylka - **SMODCH.VÝBĚR**(číslo1; číslo2; …)

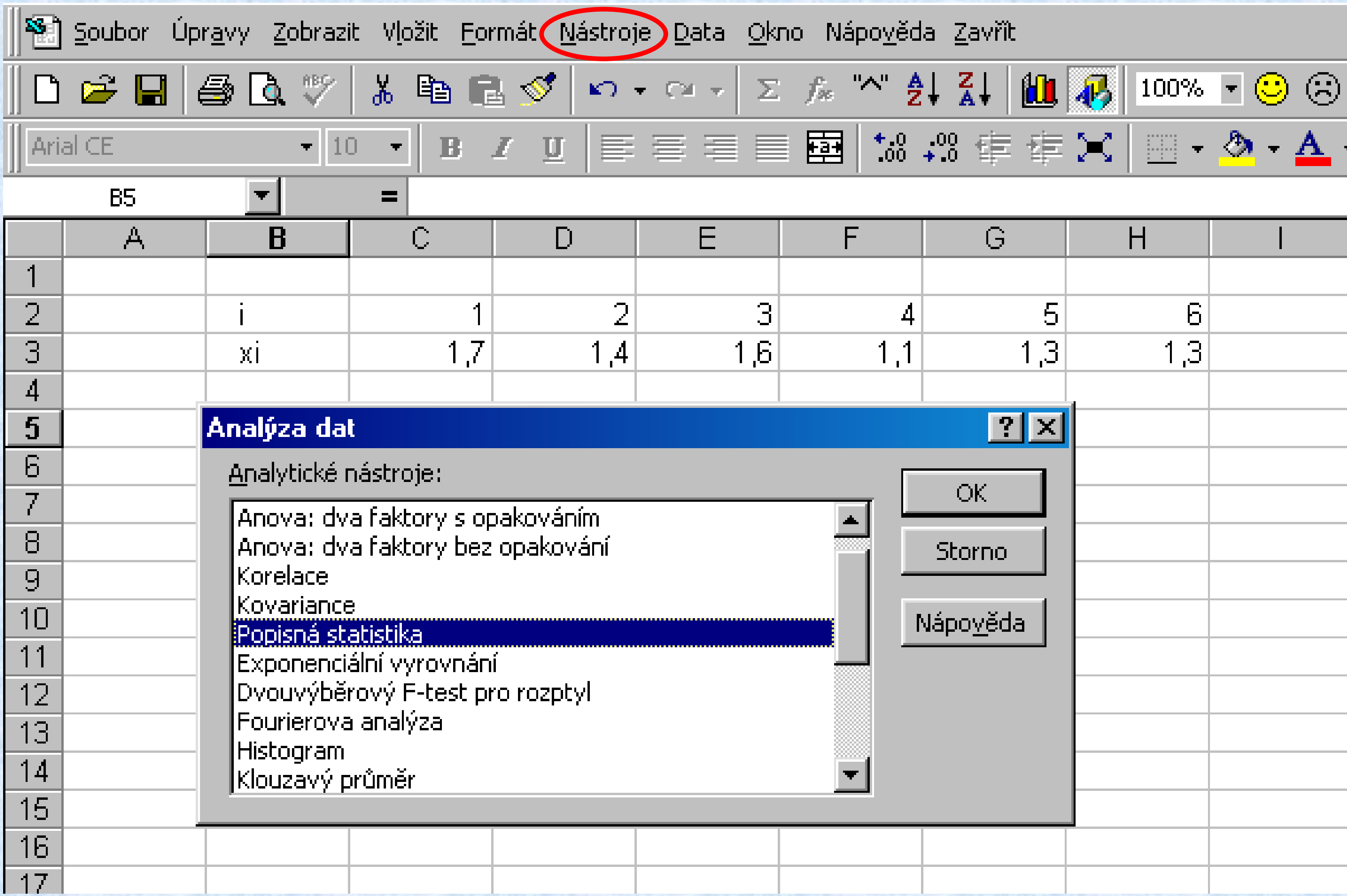
Výběrový rozptyl - **VAR.VÝBĚR**(číslo1; číslo2; …)

Maximální hodnota - **MAX**(číslo1; číslo2; …)

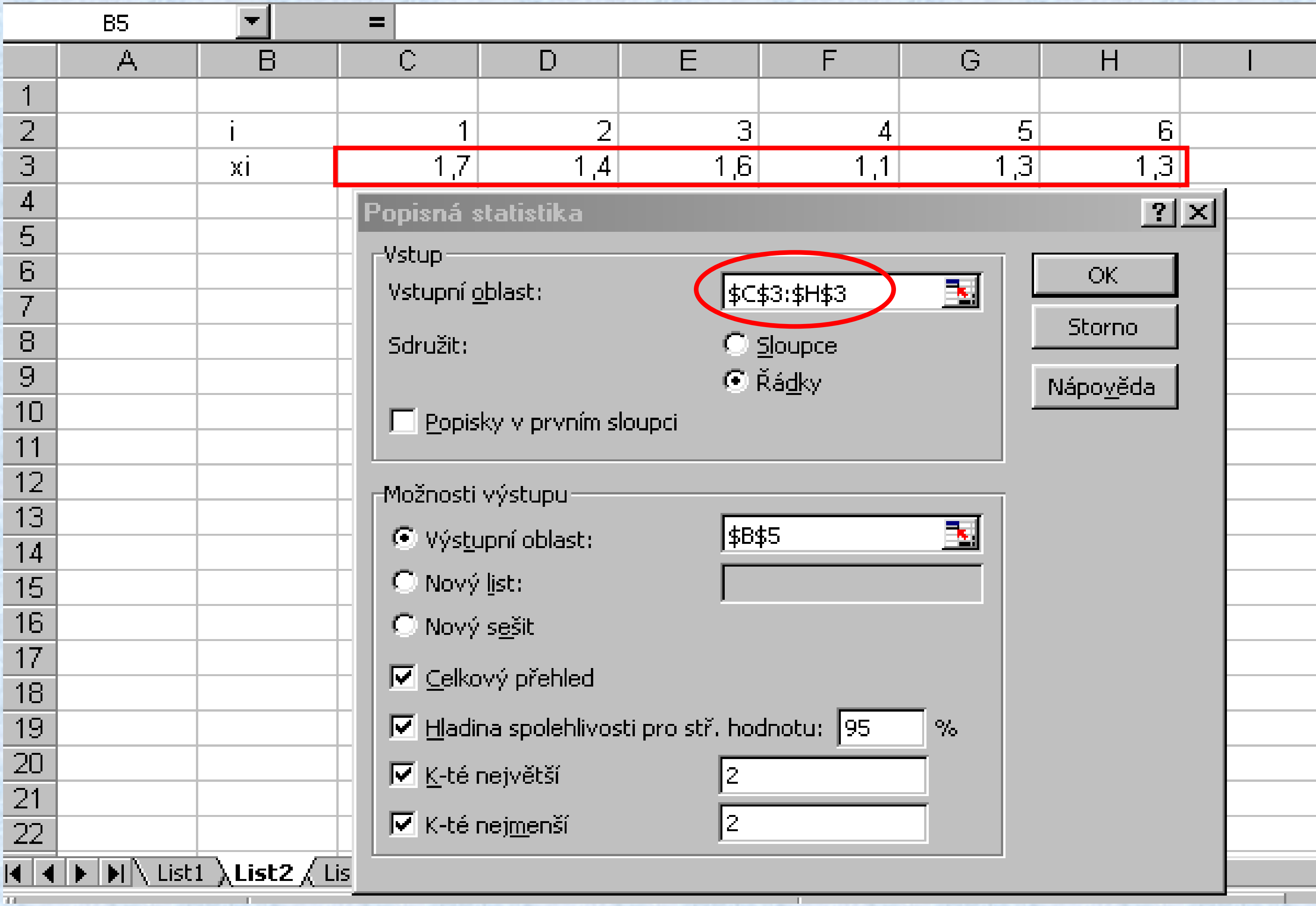
Minimální hodnota - **MIN**(číslo1; číslo2; …)

Počet hodnot - **POČET**(číslo1; číslo2; …)

Použití **nástrojů** analýzy dat:



Příklad: použití nástroje „**Popisná statistika**“ :



Příklad: **Výstup** nástroje „Popisná statistika“ :

	A	B	C	D	E	F	G	H
1								
2		i	1	2	3	4	5	6
3		xi	1,7	1,4	1,6	1,1	1,3	1,3
4								
5		*Řádek1*						
6								
7		Stř. hodnota	1,4					
8		Chyba stř. hodnoty	0,089443					
9		Medián	1,35					
10		Modus	1,3					
11		Směr. odchylka	0,219089					
12		Rozptyl výběru	0,048					
13		Špičatost	-0,78125					
14		Šikmost	0,171163					
15		Rozdíl max-min	0,6					
16		Minimum	1,1					
17		Maximum	1,7					
18		Součet	8,4					
19		Počet	6					
20		Největší (2)	1,6					
21		Nejmenší (2)	1,3					
22		Hladina spolehlivosti (95,0%)	0,229919					

Další vybrané funkce a nástroje popisné statistiky

konstrukce histogramu:

Vyhodnocení třídních četností - funkce -

ČETNOSTI (pole_dat; pole_míst)

pole_dat : pole obsahující napozorované hodnoty, jejichž četnosti chceme stanovit

pole_míst : pole obsahující horní meze třídních intervalů

Zpracování třídních četností a vytvoření histogramu - nástroj -

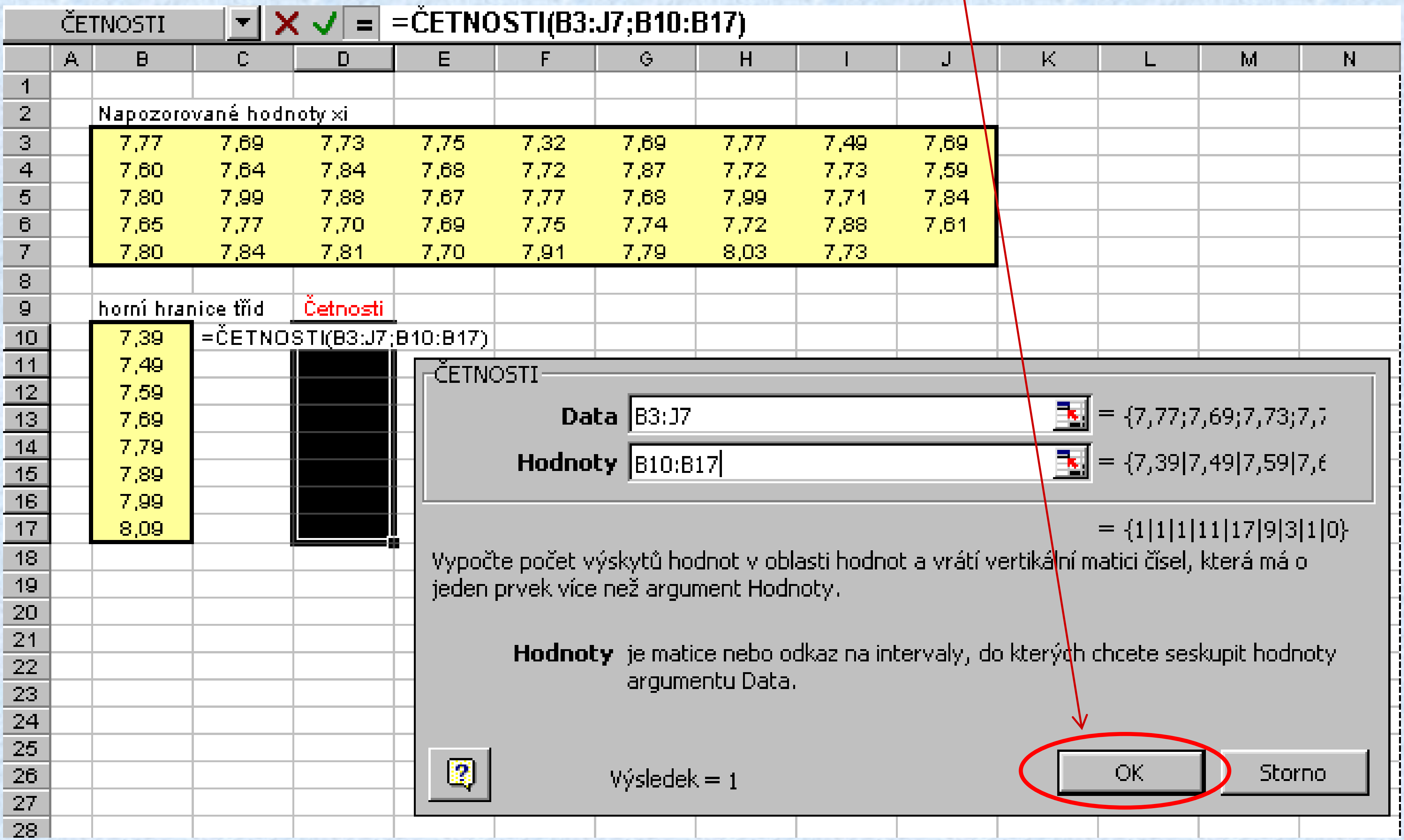
HISTOGRAM

Vyžaduje zadávat argumenty vyžádané dialogovým oknem - vstupní oblast, hranice tříd, výstupní oblast atd.

Příklad: výpočet četností pomocí funkce:

ČETNOSTI(pole_dat; pole_míst)

Zapsání výsledku je nutno potvrdit stisknutím Ctrl + Shift + Enter

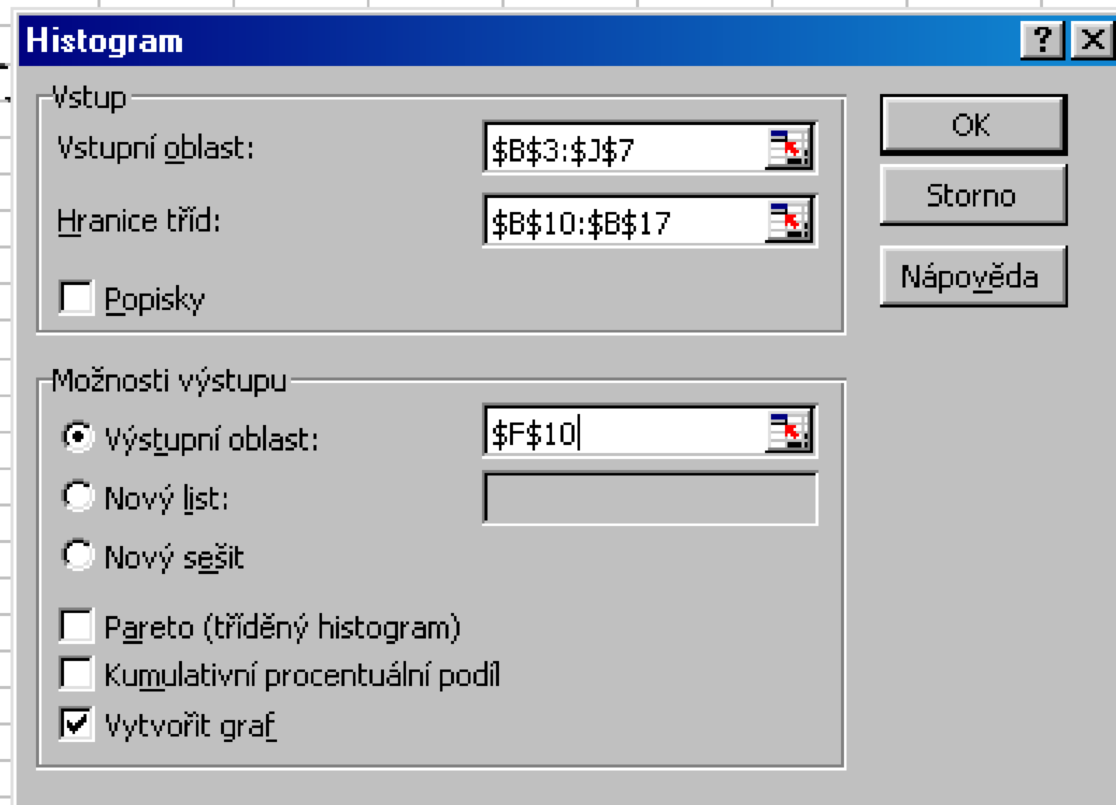


Příklad: použití nástroje „**Histogram**“ :

Napozorované hodnoty xi

7,77	7,69	7,73	7,75	7,32	7,69	7,77	7,49	7,69
7,60	7,64	7,84	7,68	7,72	7,87	7,72	7,73	7,59
7,80	7,99	7,88	7,67	7,77	7,68	7,99	7,71	7,84
7,65	7,77	7,70	7,69	7,75	7,74	7,72	7,88	7,61
7,80	7,84	7,81	7,70	7,91	7,79	8,03	7,73	

horní hranice tříd	Četnosti
7,39	1
7,49	1
7,59	1
7,69	11
7,79	17
7,89	9
7,99	3
8,09	1



Příklad: **Výstup** a) funkce „**Četnosti**“ a

b) nástroje „**Histogram**“

Pro zadané napozorované hodnoty a horní hranice tříd se

a) vyznačí oblast, kam budou četnosti zapsány a výpočet se potvrdí Ctrl + Shift + Enter;

b) vyznačí se buňka (F10) a potvrdí zakreslení grafu (který se může normálním způsobem upravovat)

Napozorované hodnoty xi

7,77	7,69	7,73	7,75	7,32	7,69	7,77	7,49	7,69
7,60	7,64	7,84	7,68	7,72	7,87	7,72	7,73	7,59
7,80	7,99	7,88	7,67	7,77	7,68	7,99	7,71	7,84
7,65	7,77	7,70	7,69	7,75	7,74	7,72	7,88	7,61
7,80	7,84	7,81	7,70	7,91	7,79	8,03	7,73	

horní hranice tříd	Četnosti
7,39	1
7,49	1
7,59	1
7,69	11
7,79	17
7,89	9
7,99	3
8,09	1

Histogram

Třídy	*Četnost*
7,39	1
7,49	1
7,59	1
7,69	11
7,79	17
7,89	9
7,99	3
8,09	1
Další	0