Popisná statistika

Základní statistické pojmy

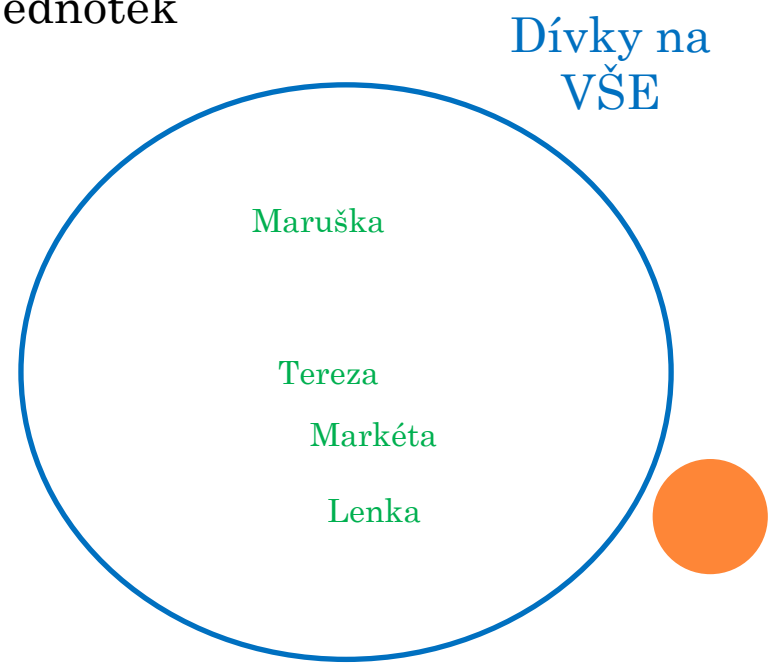
Jev hromadný
Hromadná pozorování – výsledek hromadný jev
„soustředění se“ na určitou vlastnost(i) – „ukáže“ po více pokusech
Zjistit souvislosti v prostoru a čase

Statistický soubor
Určitá množina prvků, charakteristická určitými vlastnostmi
Dívky na VŠE

EKOFUN

Statistická jednotka
Statistický soubor se skládá ze statistických jednotek
Konkrétní dívka na VŠE (Maruška)

Statistické znaky
Vlastnosti statistických jednotek
(výška, váha, věk, velikost poprsí)



Přístup ke statistickému souboru:

1) Základní soubor (populace)
2) Výběrový soubor

Vyčerpávající statistické šetření
Dílčí (výběrové) statistické šetření

Základní soubor
Počet všech žáků na VŠE, počet všech studentů ekonomie atd.
Zpravidla značně rozsáhlý – problém při analýze (nákladné)

EKOFUN

Výběrový soubor
Vybereme pouze část jednotek ze základního souboru
Snaha o co nejvyšší „reprezentativnost“
Následně provádíme úsudky o základním souboru

Výběrová chyba – výběrový soubor ≠ základní soubor

Znaky rozdělujeme na:

Kvantitativní – schopni vyjádřit číslem, má smysl odčítat, sčítat atd.
- věk, výška, váha, HDP,

Kvalitativní – nezjistíme měřitelné hodnoty
Maximálně budeme schopni porovnání (lepší, horší) – NE VŽDY!!!
- status, vzdělání, pohlaví

EKOFUN

Kvantitativní znaky spojité a nespojité
Nespojité – nabývají pouze určitých číselných hodnot
- velikost poprsí 1,2,3,4 atd.

Spojité – mohou nabývat v určitém intervalu libovolný počet hodnot
- výška, váha, příjem

Klasifikace proměnných

Nominální (jmenné)
Nelze jednoznačně určit pořadí (národnost, pohlaví)

Ordinální (pořadové)
Má význam seřadit od nejmenší po největší a porovnat
Podíl - nesmyslné
Vzdělání

EKOFUN

Metrické (měřitelné)
Lze seřadit od nejmenší po největší
Lze přesně stanovit o kolik

Rozdělení četností

Na statistických jednotkách sledujeme pouze jeden statistický znak (dívky a velikost poprsí)
Data uspořádáme od „nejmenší do největší“ hodnoty (rostoucí posloupnost)
Od nejmenšího poprsí po největší (1,2,3,4,5,6)
Počty příslušných statistických jednotek (počet dívek) – četnost

$$p_1 = \frac{5}{18} = 0{,}278 \qquad p_4 = \frac{1}{18} = 0{,}056$$

Velikost poprsí	Četnost		Kumulativní četnost	
	absolutní	relativní	absolutní	relativní
1	5	0,278	5	0,278
2	8	0,444	13	0,722
3	4	0,222	17	0,944
4	1	0,056	18	1,000
Celkem	18	1	x	x

Tabulka rozdělení četností

Relativní vs. Absolutní četnost

Absolutní četnost (4 dívky mají C)
Relativní četnost
Při porovnávání různých rozdělení VŠE, ČZU
Získáme rozdílné absolutní četnosti
Převedeme na relativní četnost – můžeme porovnávat, kdo je na tom lépe

Například 44,4% dívek ve zkoumaném souboru má velikost B

Vzorek: VŠE 2000, ČZU 1000
$VŠE_C$=500, $ČZU_C$=300
500/2000=0,25 300/1000=0,33

Kumulativní absolutní četnost
Kumulativní relativní četnost – 72,2% dívek má velikost do B

$$p_i = \frac{n_i}{\sum_{i=1}^{k} n_i} \qquad \sum_{i=1}^{k} p_i = 1$$

Intervalové rozdělení četností

Mnoho variant diskrétní/spojitá
příjem, výška a váha studentů

(1000,1050,989,876,1200,1250, 1101,….)

Problém s přehledností
Používáme intervaly četností (od,do)

EKOFUN

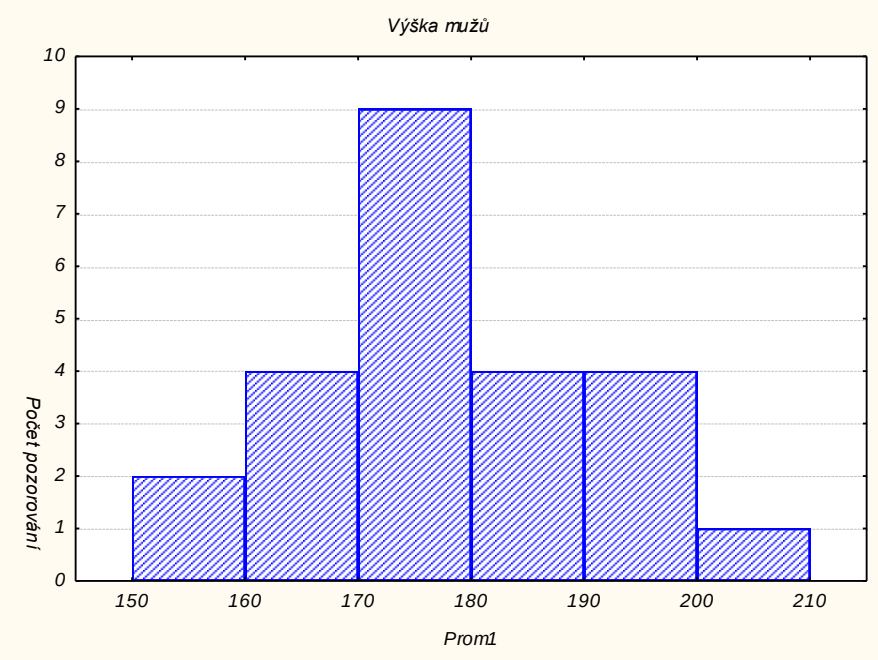
Problém
Kolik intervalů – výška mužů je od cca 150-220cm
Kolik intervalů je „správně“?

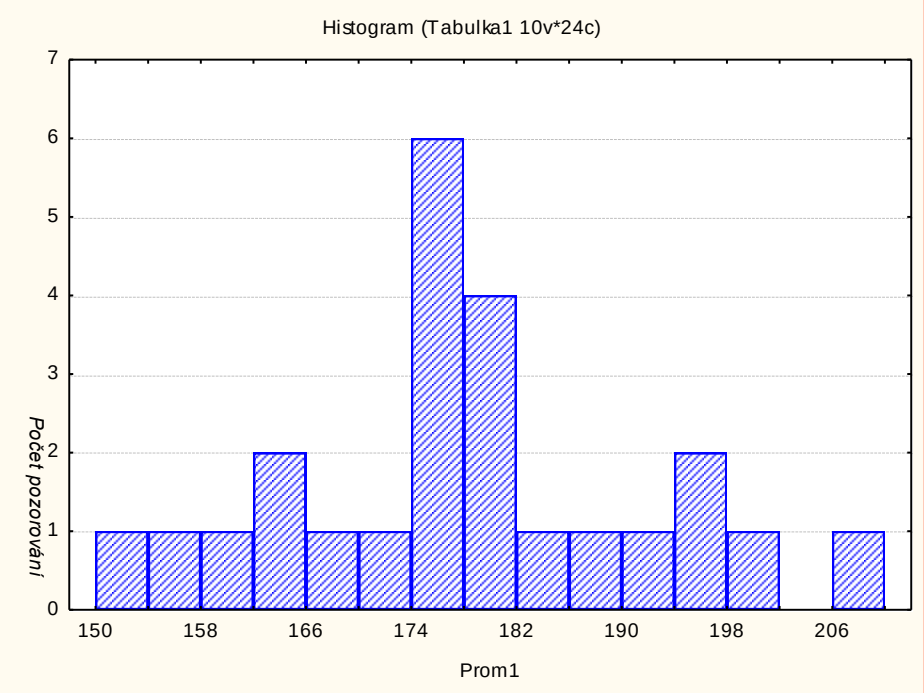
Příliš malý počet – příliš hrubé – nepostřehneme určité zákonitosti
Příliš velký počet – nepřehledné -nevyniknou zákonitosti

Odmocninové pravidlo $k \approx \sqrt{n}$

Sturgesovo pravidlo $k \approx 1 + 3{,}3 log_{10} n$

$$\text{šířka intervalu} = \frac{max.\,hodnota - min.\,hodnota}{\text{počet intervalů } k}$$





Výška mužů v cm.

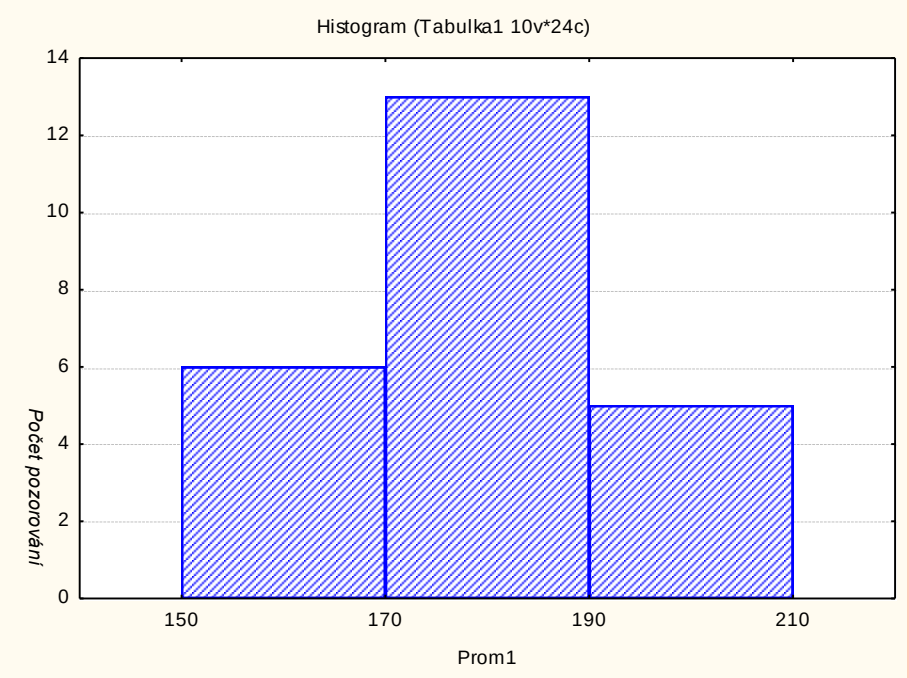
175	180	200	178
162	181	155	175
150	193	181	166
185	168	164	175
198	174	177	177
210	187	180	195

EKOFUN

$$k \approx \sqrt{n} = \sqrt{24} = 4{,}9$$

$$k \approx 1 + 3{,}3 log_{10} n = 5{,}55$$

$$\text{šířka intervalu} = \frac{max.\,hodnota - min.\,hodnota}{\text{počet intervalů } k} = 10$$



Statistické grafy

Modus

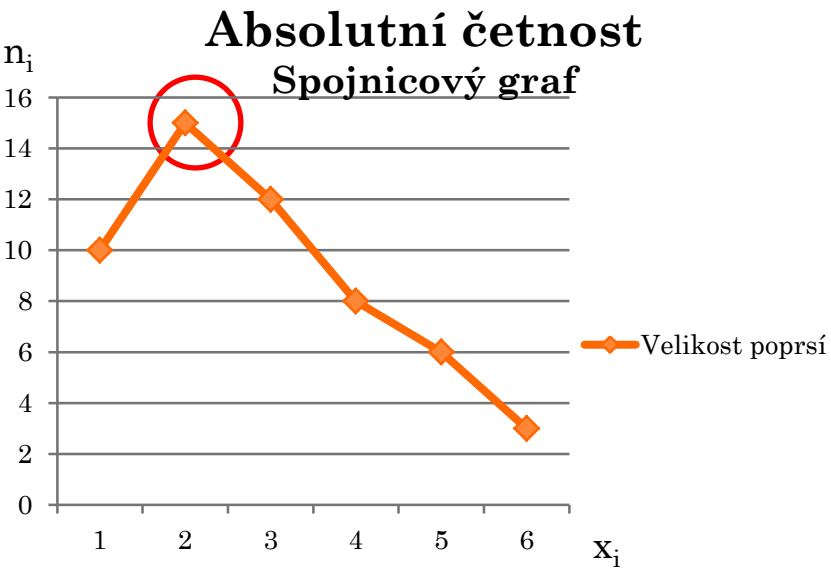
Nejčastější varianta proměnné

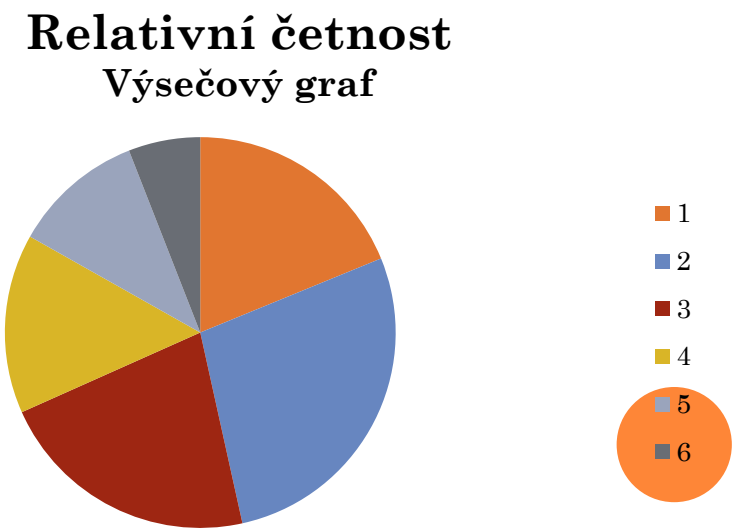
Velikost	1	2	3	4	5	6		
Ab. četnost	10	15	12	8	6	3		
Rel. četnost	0,19	0,28	0,22	0,15	0,11	0,06	∑	54

Polygon četností

Absolutní – Relativní

EKOFUN





Rozlišujeme polohu, variabilitu, tvar, počet vrcholů

Počty vrcholů

Jednovrcholová

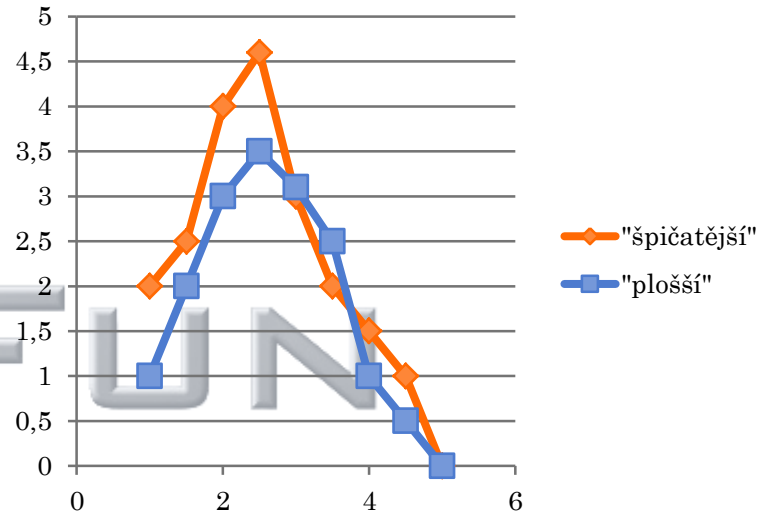
Unimodální rozdělení četností

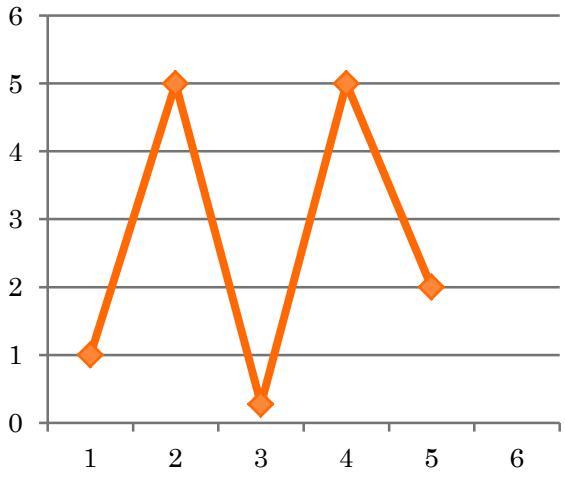
Vícevrcholové rozdělení četností

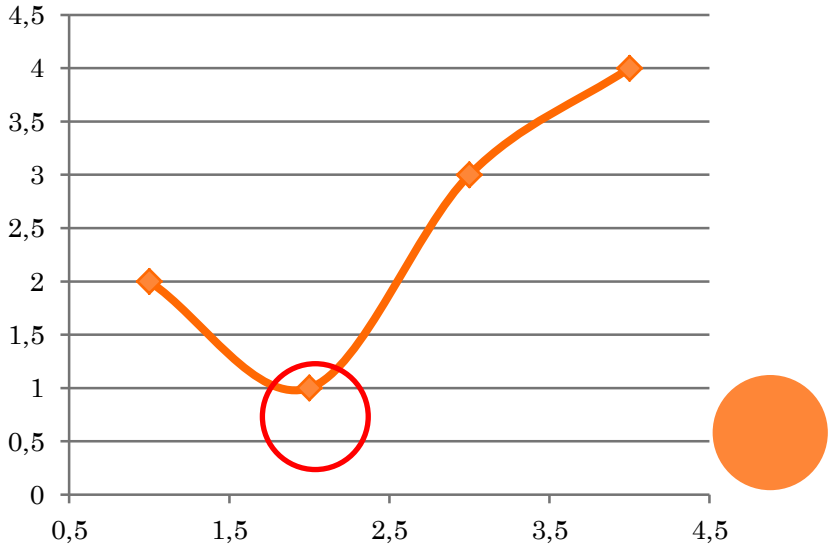
„U“ rozdělení – *antimodus*

Poloha rozdělení s nejmenší četností

Nezapomínejte popisovat osy ☺







Kvantily

Hodnota rozdělující statistický soubor hodnot na dvě části
Podle četnosti **(od nejmenší po největší)**

100p% kvantil – x_p

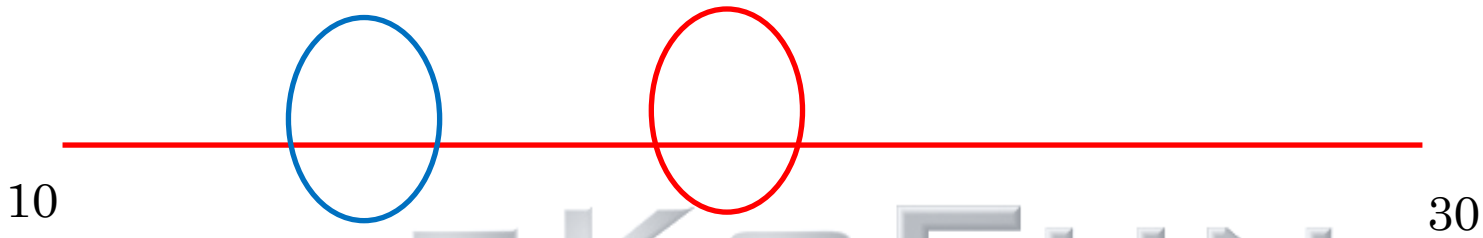
Medián x_{50}
50% kvantil – rozdělení četnosti na polovinu
Průměrná mzda v ČR je cca 23 500Kč/měs.
Medián je cca 18 000Kč/měs.
50% lidí má mzdu **DO** 18 000Kč/měs.

Kvartily, decily, percentily
Dolní kvartil x_{25}
25% kvantil = 12 000Kč – 25% lidí má mzdu do 12 000Kč
Horní kvartil x_{75}
Decil x_{10},x_{40},x_{90}

EKOFUN

Délka prstů

12 15,3 16 26 13 22 19 17 16,3 14 11



EKOFUN

Jaký je medián – 11 hodnot
Hodnota délky prstů, které dosáhne 50% studentů
Hodnota 16 je medián daného souboru
50% chlapců má délku menší jak 16cm
A 50% chlapců má délku prstů větší jak 16cm

Jaký je 25% kvantil
Hodnota délky prstů, které dosáhne pouze 25% chlapců
25% kvantil je roven 13

Co představuje hodnota 22cm?
Jedná se o 90% kvantil
90% chlapců má délku prstů menší než 22 cm

Musíte vždy uspořádat do pořadí od nejmenší do největší

25% kvantil – dolní kvartil $\widetilde{x_{25}}$

$$n.\frac{p}{100} \leq z_P \leq n.\frac{p}{100} + 1$$

z_p - pořadové číslo jednotky, jejíž hodnota je hledaný kvantil
p- „hledaný kvantil“

1	3268	11	4129	21	4800
2	3354	12	4150	22	4820
3	3500	13	4170	23	4859
4	3559	14	4190	24	4980
5	3600	15	4300	25	5000
6	3740	16	4350	26	5010
7	3876	17	4359	27	5059
8	3890	18	4400	28	5123
9	3987	19	4589	29	5149
10	4100	20	4600	30	5200

Příklad: n=30, p=25

$$30.\frac{25}{100} \leq z_{25} \leq 30.\frac{25}{100} + 1$$

$$7{,}5 \leq z_{25} \leq 8{,}5$$

Pokud $n.\frac{p}{100}$ není celé číslo
z_p je celé číslo ležící mezi oběma stranami nerovnosti

$$z_{25} = 8 \qquad \widetilde{x_{25}} = 3890$$

50% kvantil – medián $\widetilde{x_{50}}$

$$30.\frac{50}{100} \leq z_{50} \leq 30.\frac{50}{100} + 1$$

$$15 \leq z_{50} \leq 16$$

Příklad: n=30, p=50

$$\widetilde{x_{50}} = \frac{4350 + 4300}{2} = 4325$$

Pokud $n.\frac{p}{100}$ JE celé číslo
pak odhadovaný kvantil je aritmetický průměr z hodnot $n.\frac{p}{100}, n.\frac{p}{100} + 1$

Výpočet z intervalového rozdělení

$$x_p = \frac{z_p - n_1}{n_2} . h_p + a_p \qquad z_p = n . \frac{p}{100} + 0{,}5$$

$$n_1 = 25 + 8 = 33$$

$$n_2 = 32$$

n-počet jednotek ve statistickém souboru
n_1- kumulativní četnost prvků ležící před kvantilovým intervalem
n_2- četnost intervalu, kde leží hledaný kvantil
p- „hledaný kvantil“ (25,50,75…)
h_p- délka kvantilového intervalu
a_p- dolní hranice kvantilového intervalu
z_p- pořadové číslo jednotky, jejíž hodnota je hledaný kvantil

EKOFUN

$$z_{50} = 116 . \frac{50}{100} + 0{,}5$$

$$z_{50} = 58{,}5$$

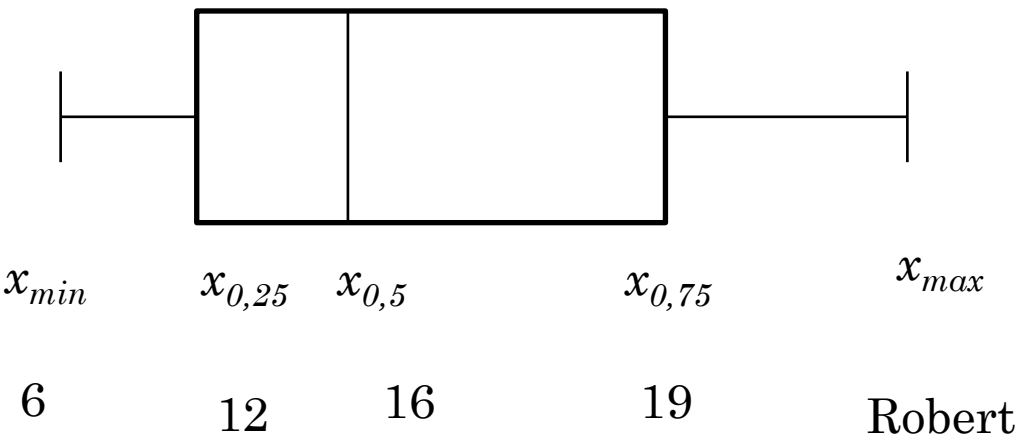
$$x_{50} = \frac{58{,}5 - 33}{32} . 400 + 7401$$

$$x_{50} = 7719{,}75$$

Interval měsíčních příjmů	Počet pracovníků	Kumulativní součty
-7000	8	8
7001-7400	25	33
7401-7800	32	65
7801-8200	26	91
8201-8600	15	106
8601-9000	6	112
9001-9400	3	115
9400 +	1	116
Celkem	116	x

Krabicové grafy

Délka prstů u chlapců



Výška mužů v cm.

175	180	200	178
162	181	155	175
150	193	181	166
185	168	164	175
198	174	177	177
210	187	180	195

150
155
162
164
166
168
174
175
175
175
177
177
178
180
180
181
181
185
187
193
195
198
200
210

Charakteristiky

Pro porovnávání několika souborů pomocí tabulek, grafů, četností
TĚŽKOPÁDNÉ

Snaha koncentrovat informace o statistickém znaku do „koncentrované podoby“

„koncentrovaná podoba“ – charakteristiky
CHARAKTERIZUJÍ ZÁKLADNÍ RYSY ZKOUMANÉHO SOUBORU

Charakterizují vlastnosti (dívek na VŠE, chlapců v Praze, velikost mezd v moravském kraji, životnost součástky….)

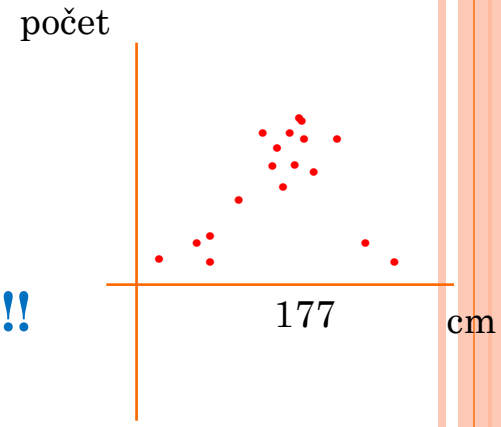
Míry polohy – střední hodnoty (průměry, medián, modus)

Míry variability (rozptyl, směrodatná odchylka, variační rozpětí atd.)

Šikmost

Špičatost

Míry polohy (charakteristiky úrovně)



Více druhů charakteristik úrovně!!!
Rozdělení četností –"zhušťování hodnot"
Měření pomocí různých druhů středních hodnot!!!
Průměry, medián, modus

Průměry

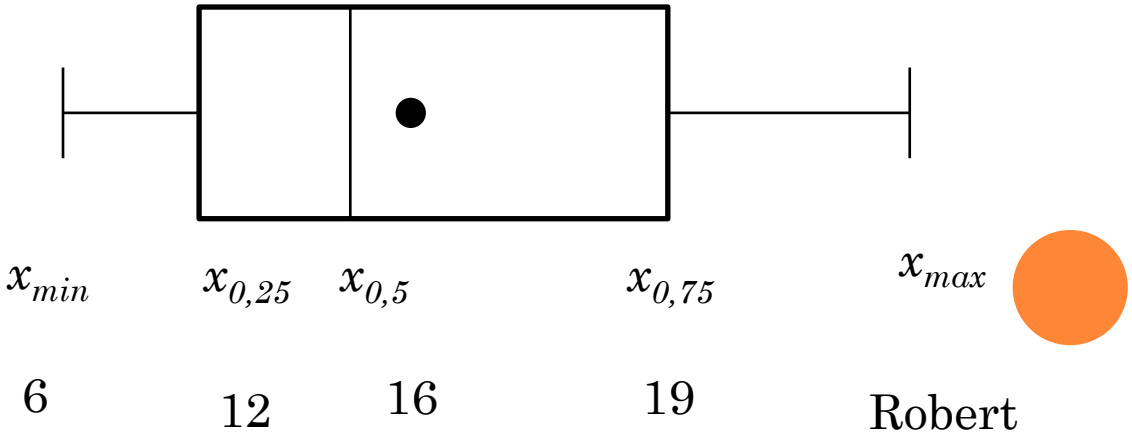
EKOFUN

Střední hodnota ze všech jednotek statistického souboru (všechny slečny)

Aritmetický, geometrický, harmonický , kvadratický

Další typy středních hodnot
Založené pouze na některých vybraných hodnotách souboru
Medián, modus

Výhoda kvantilů (medián)
Oproti průměrům
Nejsou započítány extrémy
(Robert)



Aritmetický průměr

Prostý aritmetický průměr ($\bar{x}$)

Má smysl když – má smysl sčítat hodnoty proměnných

	Počet piv
Lukáš	4
Peter 1	8
Peter 2	4
Tomáš	2
Jakub	3
Luboš	4
Samuel	10
∑	35

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{4 + 8 + 4 + 2 + 3 + 4 + 10}{7} = \frac{35}{7} = 5$$

EKOFUN

Vážený aritmetický průměr (váhy jsou četnosti)

x_i	n_i
1	10
2	20
3	8
∑	38

$$\bar{x} = \frac{x_1 . n_1 + x_2 . n_2 + \ldots + x_k x_k}{n_1 + n_2 + \ldots n_k} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$$

$$\bar{x} = \frac{1.10 + 2.20 + 3.8}{10 + 20 + 8} = \frac{74}{38} = 1{,}94$$

$$\bar{x} = \sum_{i=1}^{k} x_i . p_i$$

x_i	n_i	Relat. Čet
1	10	26,32%
2	20	52,63%
3	8	21,05%
∑	38	

$$\bar{x} = 0{,}2632.1 + 0{,}526.2 + 0{,}2105.3 = 1{,}94$$

Vlastnosti aritmetického průměru

1) Součet jednotlivých odchylek od průměru = 0

2) Aritmetický průměr konstanty = hodnotě konstanty

3) Přičtu-li k jednotlivým hodnotám znaku konstantu, zvýší se o tuto konstantu i aritmetický průměr

4) Násobím-li jednotlivé hodnoty znaku konstantou, je o tuto konstantu násoben i průměr

5) Násobím-li váhy aritmetického průměru konstantou, průměr se nezmění

6) Statistický soubor (x) – (velikost poprsí u dívek na ekonomkách), je rozdělen do *k* podsouborů ($x_1, x_2, \ldots x_k$) (VŠE, ČZU, Mendelovka, VŠB)
znám průměry každého podsouboru ($\bar{x}_{VŠE}$, $\bar{x}_{ČZU}$, $\bar{x}_{VŠB}$ ...)
znám množství pozorování v každém podsouboru ($n_{VŠE}$, $n_{ČZU}$, $n_{VŠB}$...)

Průměr celého souboru ($\bar{x}_{Ekonomky}$)je váženým aritmetickým průměrem dílčích průměrů a vahami je četnost podsouborů

$$\bar{x} = \frac{\sum_{i=1}^{k} \bar{x}_i . n_i}{\sum_{i=1}^{k} n_i}$$

1. $(4-5)+(8-5)+(4-5)+(2-5)+(3-5)+(4-5)+(10-5)=0$

	Počet piv
Lukáš	4
Peter 1	8
Peter 2	4
Tomáš	2
Jakub	3
Luboš	4
Samuel	10
$\sum$	35

$\bar{x}=5$

2. $$\bar{x}=\frac{4}{1}=4$$

	Počet piv
Lukáš	6
Peter 1	**10**
Peter 2	6
Tomáš	4
Jakub	5
Luboš	6
Samuel	12
$\sum$	49

+2

3. $$\bar{x}=\frac{6+10+6+4+5+6+11}{7}=\frac{49}{7}=7$$

	Počet piv
Lukáš	8
Peter 1	16
Peter 2	8
Tomáš	4
Jakub	6
Luboš	8
Samuel	20
$\sum$	70

.2

4. $$\bar{x}=\frac{8+16+8+4+6+8+20}{7}=\frac{70}{7}=10$$

$$\bar{x} = \frac{1.10 + 2.20 + 3.8}{10 + 20 + 8} = \frac{74}{38} = 1{,}94$$

x_i	n_i
1	10
2	20
3	8
∑	38

5.

$$\bar{x} = \frac{1.30 + 2.60 + 3.24}{30 + 60 + 24} = \frac{222}{114} = 1{,}94$$

x_i	n_i
1	30
2	60
3	24
∑	114

.3

6.

$$\bar{x} = \frac{1.10 + 2.13 + 3.3}{10 + 13 + 3} = \frac{45}{26} = 1{,}73$$

	1	2	3
VŠE	5	6	2
ČZU	2	4	1
JČU	3	3	0

$$\bar{x}_{VŠE} = \frac{1.5 + 2.6 + 3.2}{5 + 6 + 2} = \frac{23}{13} = 1{,}769$$

$$\bar{x}_{ČZU} = \frac{1.2 + 2.4 + 3.1}{2 + 4 + 1} = \frac{13}{7} = 1{,}857$$

$$\bar{x}_{JČU} = \frac{1.3 + 2.3 + 3.0}{3 + 3 + 0} = \frac{9}{6} = 1{,}5$$

$$\bar{x}_{Celkem} = \frac{13}{26}.1{,}769 + \frac{7}{26}.1{,}857 + \frac{6}{26}.1{,}5 = 1{,}73$$

Geometrický průměr

Má smysl když – má smysl součin hodnot proměnné (informační) (koeficienty růstu, využití u časových řad)

$$\overline{x_G} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Vážený geometrický průměr

$$\overline{x_G} = \sqrt[n]{x_1^{n1} . x_2^{n2} \dots x_k^{nk}}$$

Harmonický průměr

$$\overline{x_H} = \frac{n}{\sum_{i=1}^{k} \frac{1}{x_i}}$$

EKOFUN

Vážený harmonický průměr

$$\overline{x_H} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

Kvadratický průměr

$$\overline{x_K} = \sqrt{\frac{\sum_{i=1}^{k} x_i^2}{n}}$$

Vážený kvadratický průměr

$$\overline{x_k} = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 . n_i}{\sum_{i=1}^{k} n_i}}$$

$$\overline{x_H} \leq \overline{x_G} \leq \bar{x} \leq \overline{x_K}$$

$$\overline{x_H} \leq \overline{x_G} \leq \bar{x} \leq \overline{x_K}$$

Geometrický průměr

$$\overline{x_G} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

$$\overline{x_G} = \sqrt[5]{15.14.16.17.13} = 14{,}93$$

$$\bar{x} = 15$$

Vážený geometrický průměr

$$\overline{x_G} = \sqrt[n]{x_1^{n1}.x_2^{n2}\ldots x_k^{nk}}$$

$$\overline{x_G} = \sqrt[38]{1^{10}.2^{20}.3^{8}} = 1{,}44$$

EKOFUN

Harmonický průměr

$$\overline{x_H} = \frac{n}{\sum_{i=1}^{k} \frac{1}{x_i}}$$

$$\overline{x_H} = \frac{5}{\frac{1}{15}+\frac{1}{14}+\frac{1}{16}+\frac{1}{17}+\frac{1}{13}} = 14{,}86$$

x_i	n_i
1	10
2	20
3	8
∑	38

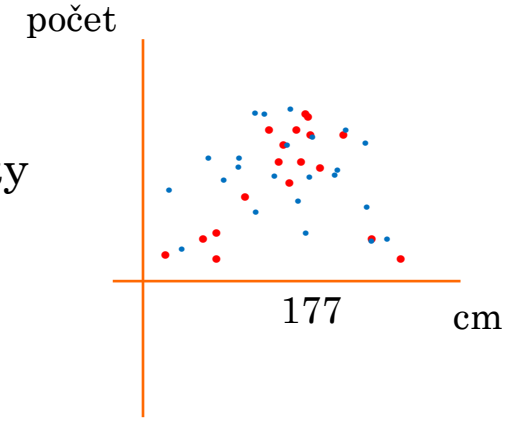
Vážený harmonický průměr

$$\overline{x_H} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

$$\overline{x_H} = \frac{38}{\frac{10}{1}+\frac{20}{2}+\frac{8}{3}} = 1{,}67$$

i	x_i
1	15
2	14
3	16
4	17
5	13

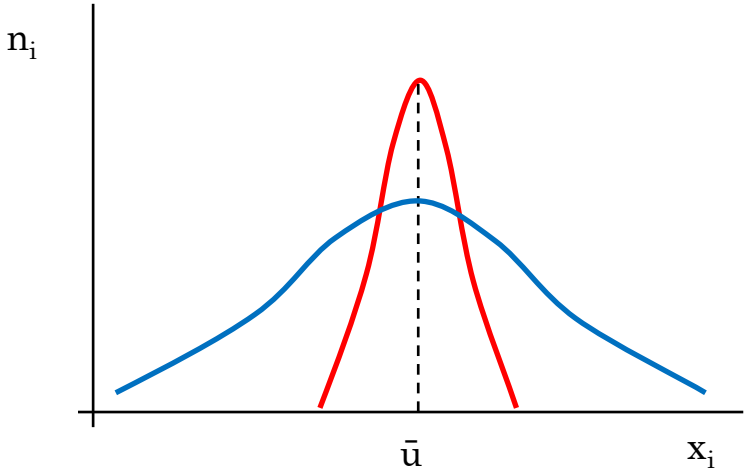
Míry variability (charakteritiky)



RŮZNÉ MÍRY VARIABILITY
Např. „jak moc lítá“ četnost kolem průměrné hodnoty
Vypovídací schopnost pro aritmetický průměr
– nižší variabilita, větší vypovídací schopnost

Míra absolutní vs. Míra relativní variability

Míra absolutní variability
Variabilita (kolísání) - ve stejných jednotkách, ve kterých je vyjadřován sledovaný znak - délka prstů v cm
Míra relativní variability
Variabilita v poměru k úrovni sledovaného znaku v souboru
Bezrozměrné číslo (cm/stopa, kg/libra)



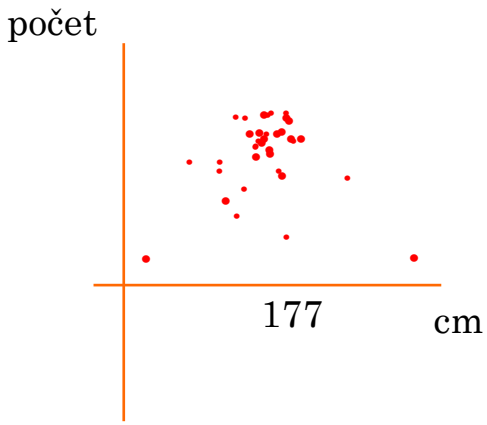
Stejný průměr

Modré rozdělení
Víc „lítá“ okolo střední hodnoty

Výška v Praze a Brně
V Praze jsou větší extrémy

Míra absolutní variability

Variační rozpětí R



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

$$R = 24 - 6 = 18$$

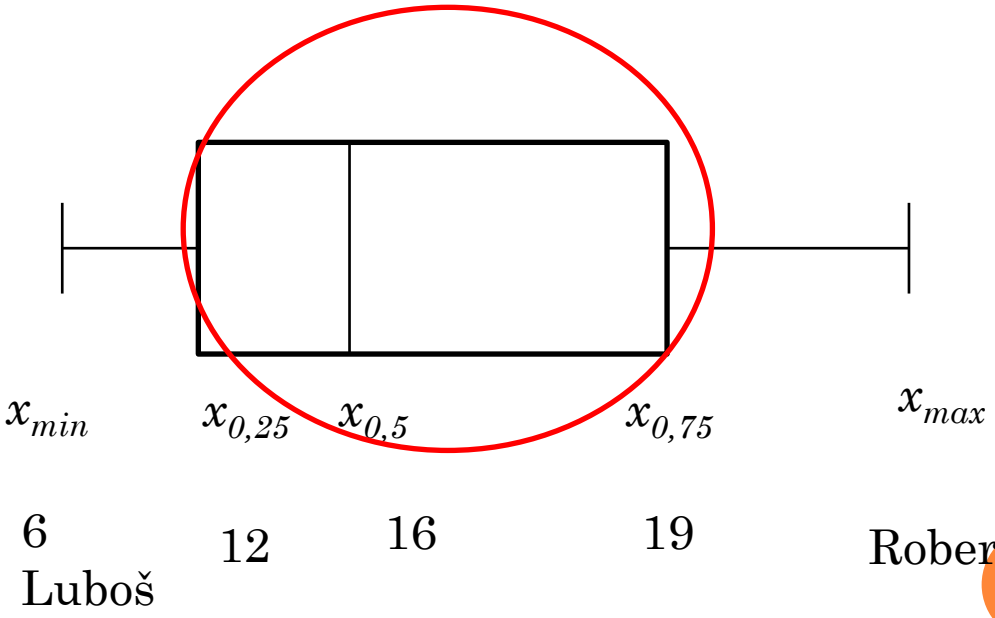
Nic neříká o variabilitě hodnot uvnitř variačního rozpětí
Problém s výskytem extrémních hodnot

EKUFUN

Kvartilové rozpětí

$$KR = \widetilde{x_{75}} - \widetilde{x_{25}}$$

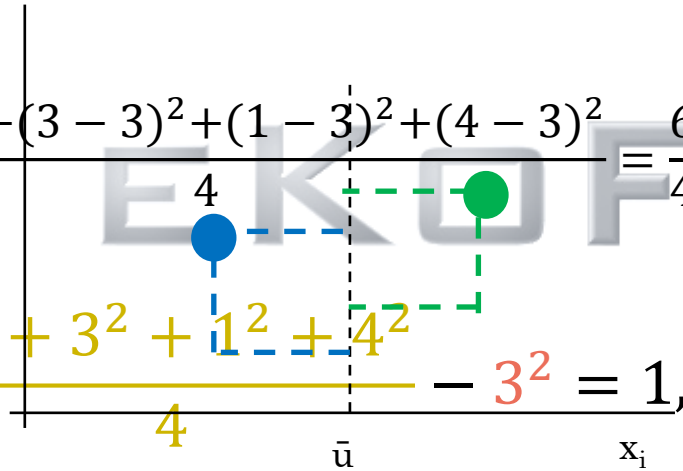
$$= 19 - 12 = 7$$



Rozptyl

Variabilita hodnot kolem aritmetického průměru
Průměr čtverců odchylek jednotlivých hodnot znaku (poprsí)
Od jejich aritmetického průměru

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

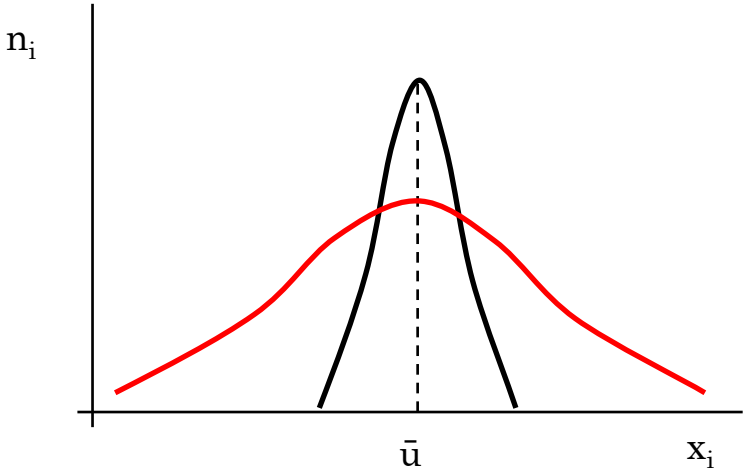


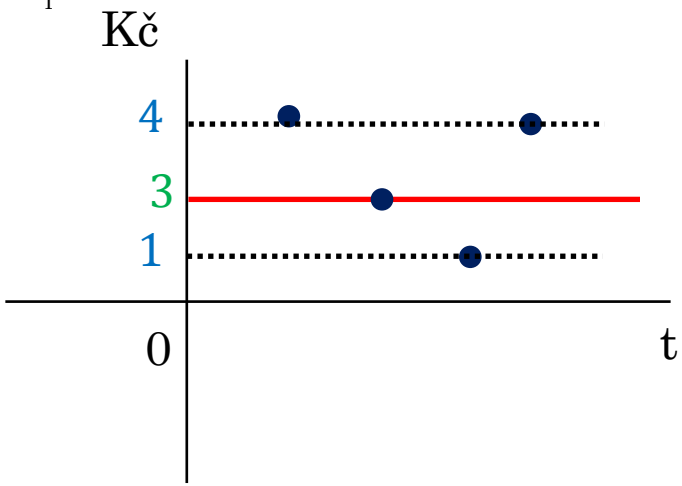
$$s_x^2 = \frac{(4-3)^2+(3-3)^2+(1-3)^2+(4-3)^2}{4} = \frac{6}{4}$$

$$s_x^2 = \frac{4^2+3^2+1^2+4^2}{4} - 3^2 = 1{,}5$$

$$s_x^2 = \overline{x^2} - \bar{x}^2$$

EKOFUN





Směrodatná odchylka

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

Problém rozptylu – čtverec
Výsledek rozptylu v jednotkách „na druhou“
cm², Kč², velikost poprsí na druhou ☺
Odmocnina z rozptylu

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

$$s_x^2 = \frac{(4-5)^2+(8-5)^2+(4-5)^2+(2-5)^2+(3-5)^2+(4-5)^2+(10-5)^2}{7} = 7{,}14$$

EKOFUN

$$s_x = \sqrt{7{,}14} = 2{,}67 \text{ piva}$$

	Počet piv
Lukáš	4
Peter 1	8
Peter 2	4
Tomáš	2
Jakub	3
Luboš	4
Samuel	10
∑	35

Vážený rozptyl

Velikost rozptylu bude záviset
Také na četnosti

$$s_x^2 = \frac{\sum_{=1}^{k}(x_i - \bar{x})^2 . n_i}{\sum_{i=1}^{k} n_i}$$

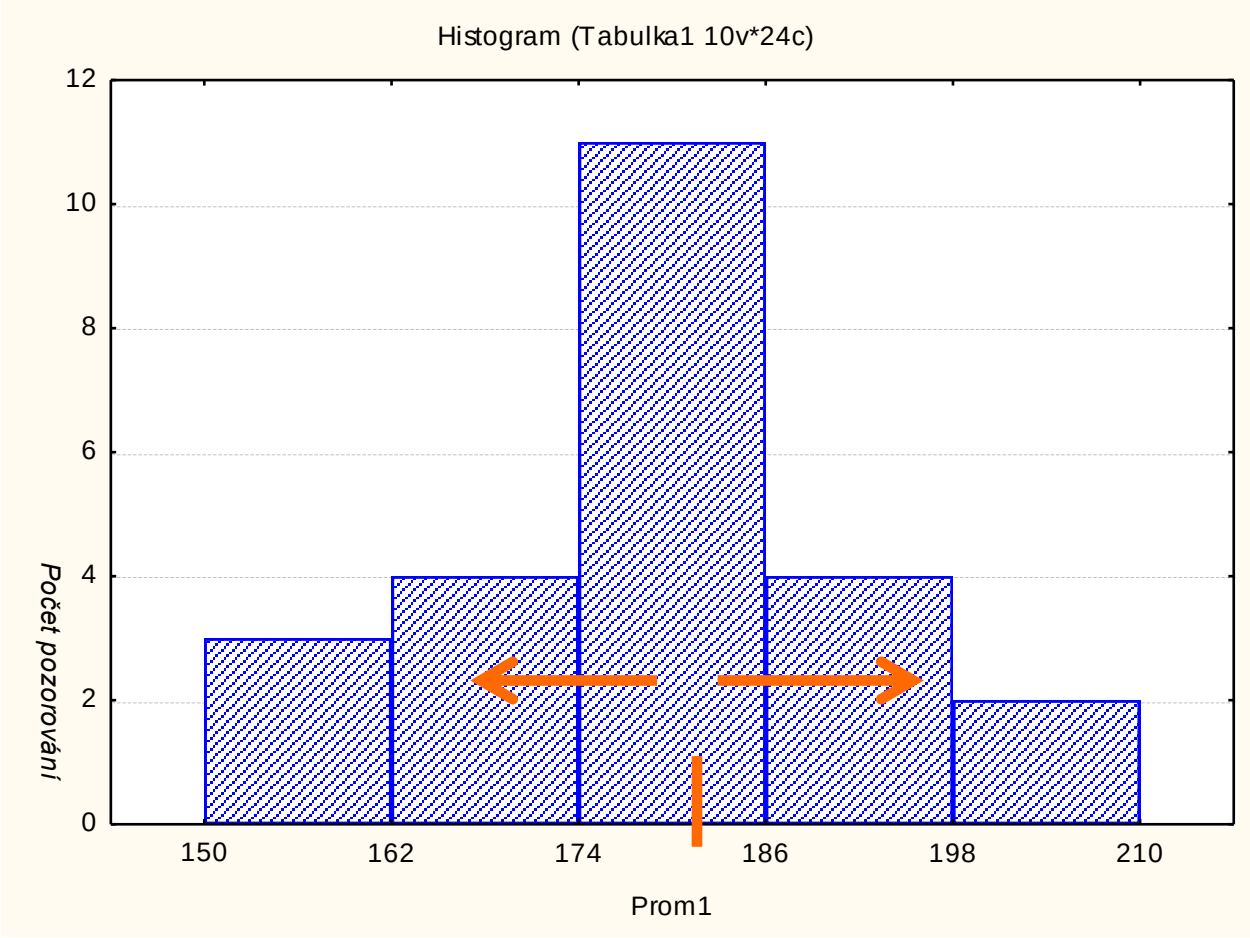
x_i	n_i
1	10
2	20
3	8
∑	38

Výška mužů v cm.

175	180	200	178
162	181	155	175
150	193	181	166
185	168	164	175
198	174	177	177
210	187	180	195

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

x-=178,53 cm
σ=13,88
Odchylka od průměrné hodnoty



Vlastnosti rozptylu

1) Rozptyl konstanty je = 0

2) Přičteme-li ke všem hodnotám znaku (poprsí) konstantu rozptyl se nezmění

3) Násobíme-li všechny hodnoty znaku konstantou rozptyl je násoben čtvercem této konstanty

EKOFUN

4) Rozptyl součtu/rozdílu dvou proměnných, je roven součtu rozptylů obou proměnných, zvětšenému (+) nebo zmenšenému (-) o dvojnásobek kovariance

$$s_Z^2 = s_{x \mp y}^2 = s_x^2 + s_y^2 \mp 2s_{xy}$$

Kovariance x,y (s_{xy}) – vyjadřuje vzájemnou závislost proměnných (x) a (y)

$$s_x^2 = \frac{(4-5)^2+(8-5)^2+(4-5)^2+(2-5)^2+(3-5)^2+(4-5)^2+(10-5)^2}{7} = 7{,}14$$

	Počet piv
Lukáš	4
Peter 1	8
Peter 2	4
Tomáš	2
Jakub	3
Luboš	4
Samuel	10
∑	35

$\bar{x} = 5$

2.

$$s_x^2 = \frac{(6-7)^2+(10-7)^2+(6-7)^2+(4-7)^2+(5-7)^2+(6-7)^2+(12-7)^2}{7} = 7{,}14$$

Pozor neplést s aritmetickým průměrem
Ten vzroste o z 5 na 7!!!

	Počet piv
Lukáš	6
Peter 1	**10**
Peter 2	6
Tomáš	4
Jakub	5
Luboš	6
Samuel	12
∑	49

+2

3.

$$s_x^2 = \frac{(8-10)^2+(16-10)^2+(8-10)^2+(4-10)^2+(6-10)^2+(8-10)^2+(20-10)^2}{7} = 28{,}57$$

4.7,14=28,56

Kvadrát konstanty je 2^2

	Počet piv
Lukáš	8
Peter 1	16
Peter 2	8
Tomáš	4
Jakub	6
Luboš	8
Samuel	20
∑	35

.2

5) Statistický soubor (dívky na VŠE)
o rozsahu (n) statistických jednotek (počet dívek na VŠE)
je rozdělen do (k) dílčích podsouborů (FFÚ,FMV,FPH,NF,FIS)
A my známe dílčí: rozptyly (s_{ix}^2), průměry (x_i) a četnosti (n_i)

Rozptyl celého souboru (všech dívek na VŠE)
Je dán součtem rozptylu skupinových průměrů (FFÚ,FMV..)
A průměru z skupinových rozptylů

Celkový rozptyl =rozptyl z průměrů + průměr z rozptylů

$$s_x^2 = s_{\bar{x}}^2 + \overline{s^2}$$

Rozptyl skupinových průměrů – meziskupinová variabilita

Průměr ze skupinových rozptylů – vnitroskupinová variabilita

Cílem není nic jiného než spočítat rozptyl poprsí na celé VŠE
Díky znalosti údajů z jednotlivých fakult

4) Cenný papír A
Rozptyl = 10

Cenný papír B
Rozptyl = 4

Cov(A,B)=-3

$$s_Z^2 = s_{x\mp y}^2 = s_x^2 + s_y^2 \mp 2s_{xy}$$

$$s_Z^2 = 10 + 4 + 2.(-3) = 8$$

	1	2	3
VŠE	5	6	2
ČZU	2	4	1
JČU	3	3	0

$$\bar{x}_{VŠE} = 1{,}769$$

$$\bar{x}_{ČZU} = 1{,}857$$

$$\bar{x}_{JČU} = 1{,}5$$

$$s_x^2 = \frac{5.(1-1{,}769)^2 + 6.(2-1{,}769)^2 + 2.(3-1{,}769)^2}{5+6+2} = 0{,}485$$

$$s_x^2 = \frac{2.(1-1{,}857)^2 + 4.(2-1{,}857)^2 + 1.(3-1{,}857)^2}{2+4+1} = 0{,}408$$

$$s_x^2 = \frac{3.(1-1{,}5)^2 + 3.(2-1{,}5)^2}{3+3} = 0{,}25$$

Kovariance

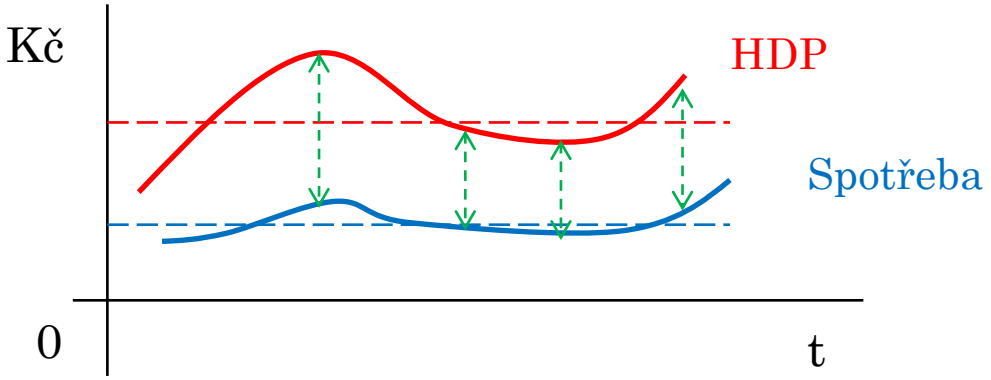
Charakterizuje vzájemnou LINEÁRNÍ závislost proměnných x,y
Vývoj HDP a spotřeby
Cena akcie energetické společnosti a společnosti na těžbu uhlí
Úroková míra a investice

(-∞,∞)

EKOFUN

Pozitivní kovariance – pozitivní vztah- roste HDP roste spotřeba
Negativní kovariance – negativní vztah – roste úroková míra klesají investice
Nulová kovariance – LINEÁRNÍ nezávislost

$$s_{xy} = \frac{1}{n}.\sum(x_i - \bar{x}).(y_i - \bar{y}) \qquad = \frac{\sum x_i.y_i}{n} - \bar{x}\bar{y} = \overline{xy} - \bar{x}\bar{y}$$



$$s_{xy} = \frac{1}{n}.\sum(x_i - \bar{x}).(y_i - \bar{y}) = \frac{\sum x_i.y_i}{n} - \bar{x}\bar{y} = \overline{xy} - \bar{x}\bar{y}$$

$$cov(X,Y) = E\{[X - E(X)].[Y - E(X)]\}$$

$$s_{xy} = \frac{(10 - 11{,}4).(7 - 7{,}8) + (13 - 11{,}4).(9 - 7{,}8) + (14 - 11{,}4).(11 - 7{,}8) + (11 - 11{,}4).(8 - 7{,}8) + (9 - 11{,}4).(4 - 7{,}8)}{5}$$

$$s_{xy} = cov(X,Y) = 4{,}08$$

EKOFUN

Mezi HDP a spotřebou existuje pozitivní vztah
Korelační koeficient pro „lepší“ interpretaci
(v regresní a korelační analýze)

$\overline{HDP} = 11{,}4$ $\overline{Spotřeba} = 7{,}8$

Rok	HDP	Spotřeba
2000	10	7
2001	13	9
2002	14	11
2003	11	8
2004	9	4

Kvartilové odchylky

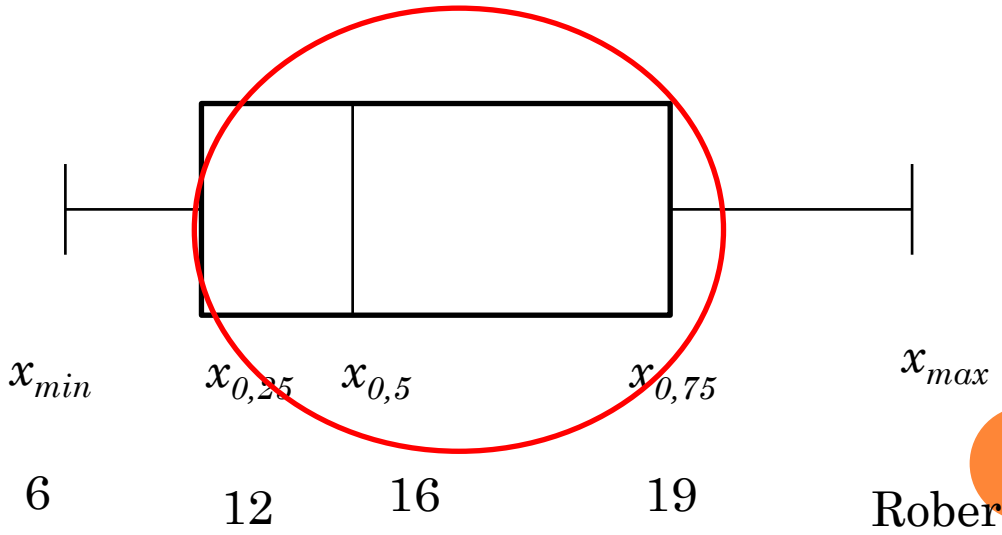
Absolutní míra variability

$$Q = \frac{\widetilde{x_{75}} - \widetilde{x_{25}}}{2} \quad = \frac{19 - 12}{2} = 3{,}5$$

Decilové a percentilové odchylky

EKOFUN

$$Q = \frac{\widetilde{x_{30}} - \widetilde{x_{20}}}{2}$$



Robert

Míra relativní variability

Různé jednotky – problém při porovnávání
(Měny, míry délky, váhy atd.)

Variační koeficient

$$V_x = \frac{s_x}{\bar{x}}$$

Poměr směrodatné odchylky ku aritmetickému průměru
Výsledek násobíme 100 a získáme vyjádření v procentech

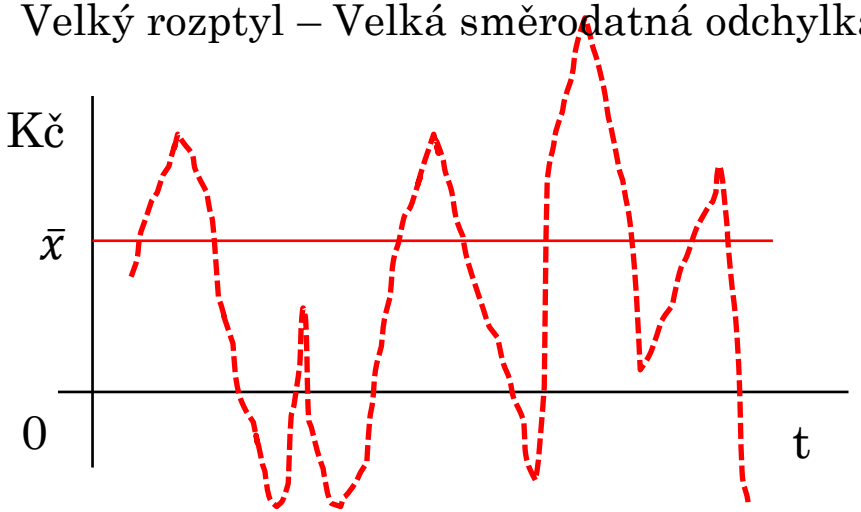
$\uparrow s_x$ - $\uparrow V_x$

EKOFUN

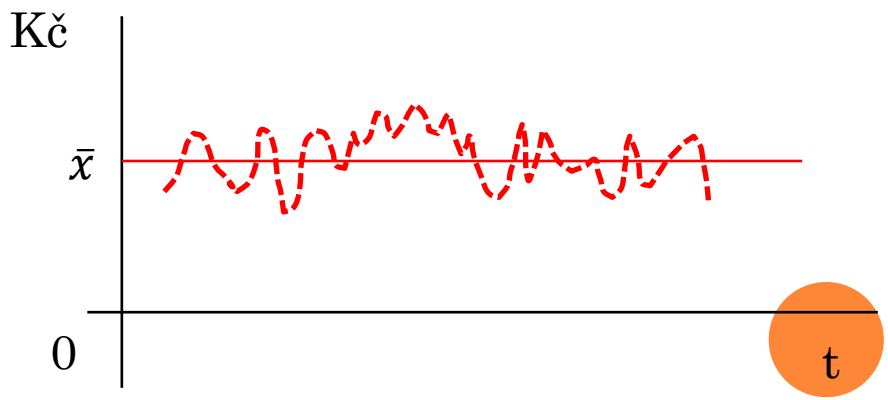
Variační koeficient **větší jak 50%** - znak značné nesourodosti

(-∞,∞)

Velký rozptyl – Velká směrodatná odchylka



Malý rozptyl – malá směrodatná odchylka



Proč to děláme

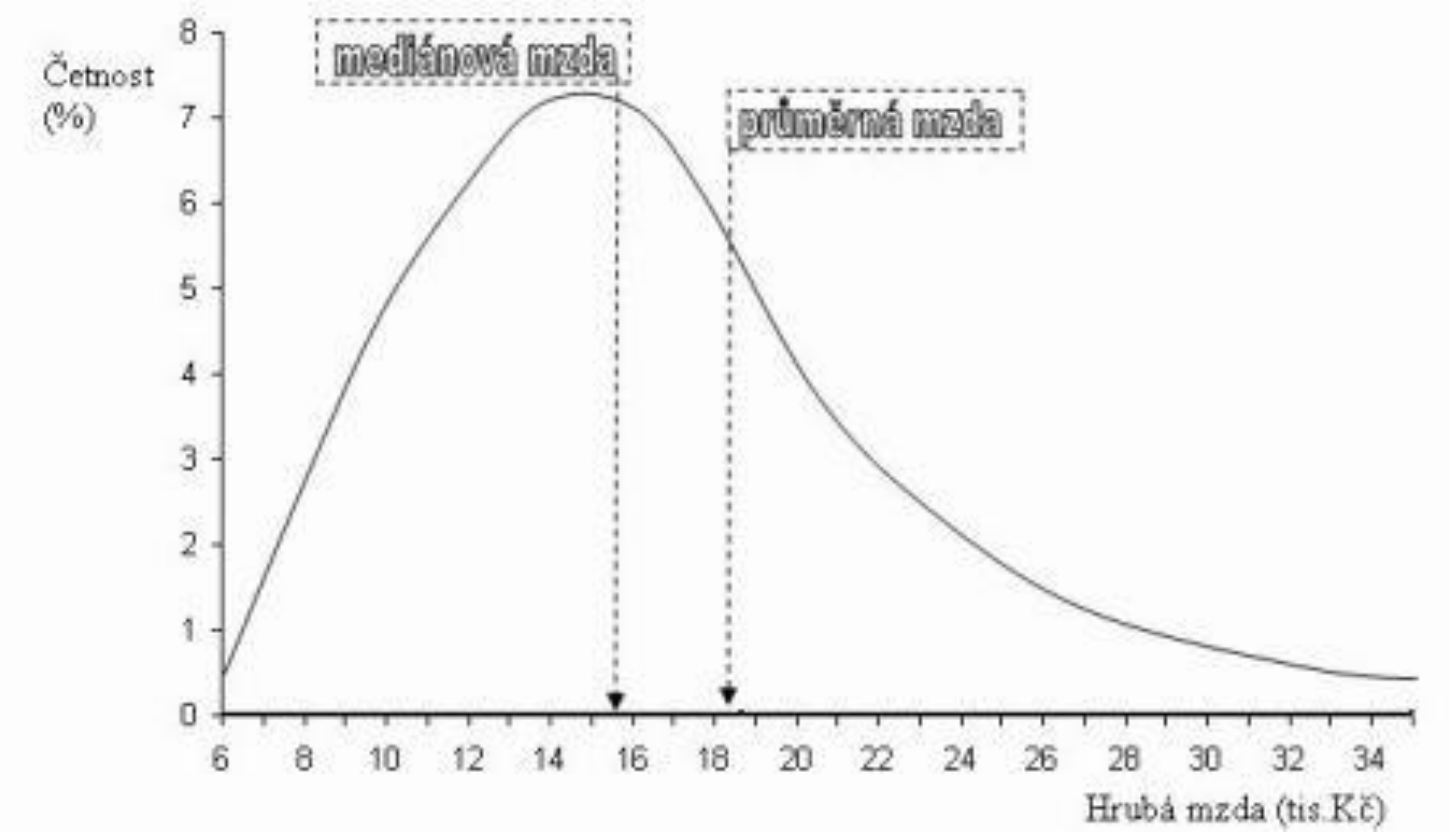
Základní soubor – zkoumat příliš drahé

Výběrový soubor
Vybereme několik jednotek
Popíšeme pomocí popisné statistiky

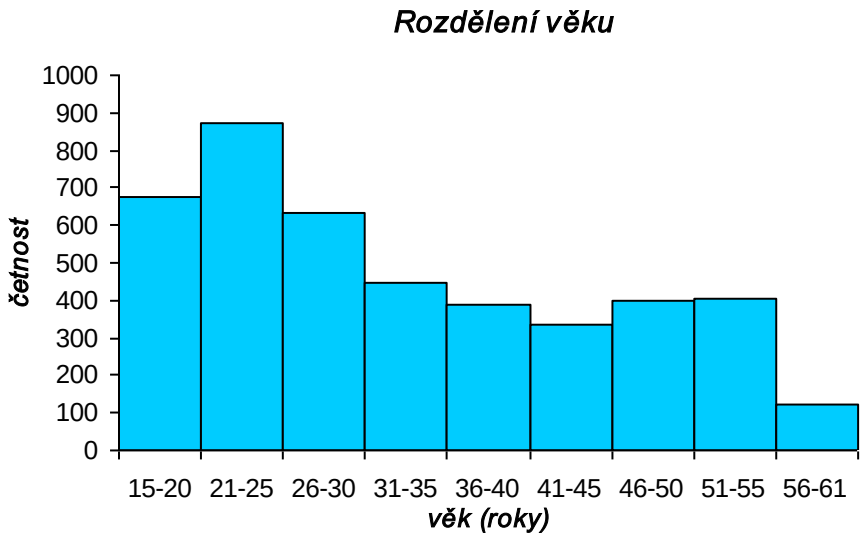
EKOFUN

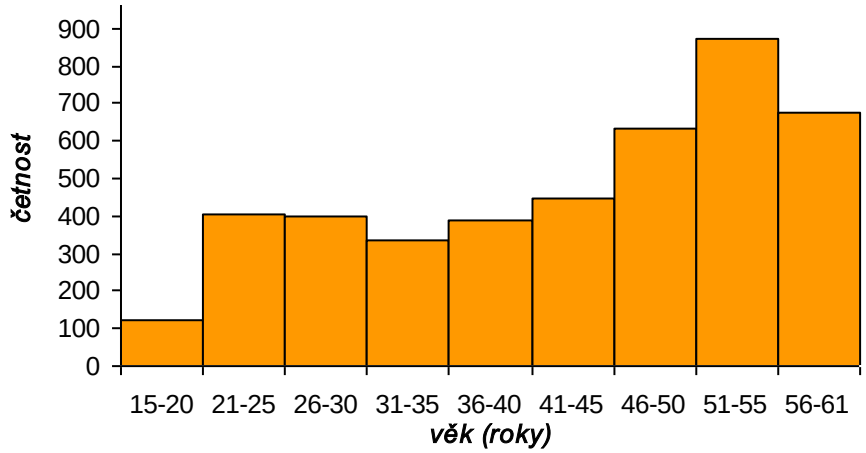
Budeme usuzovat jak se výběrový soubor podobá základnímu souboru
Testování hypotéz

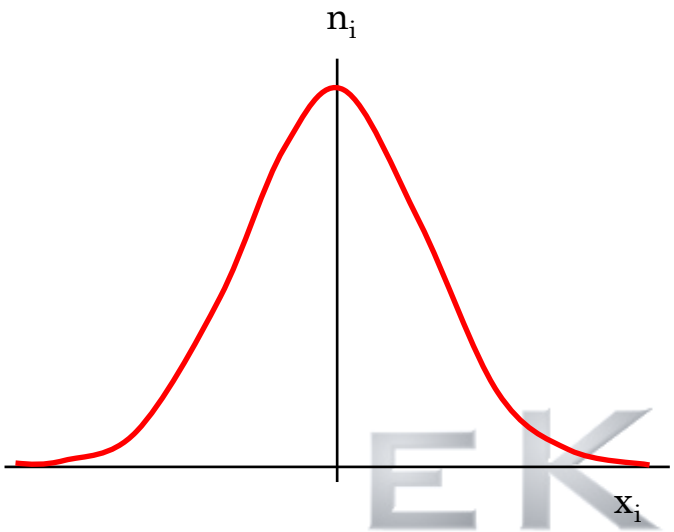
Budeme se pokoušet popsat vzájemné vztahy
Regresní a korelační analýzy

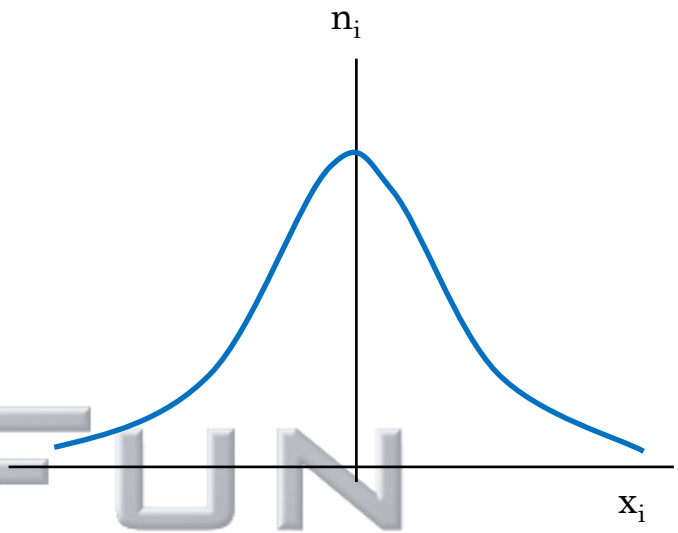

Četnost (%)
mediánová mzda
průměrná mzda
0
1
2
3
4
5
6
7
8
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
Hrubá mzda (tis.Kč)

Příklad: *Rozdělení věku nezaměstnaných (Příbram 2002)*

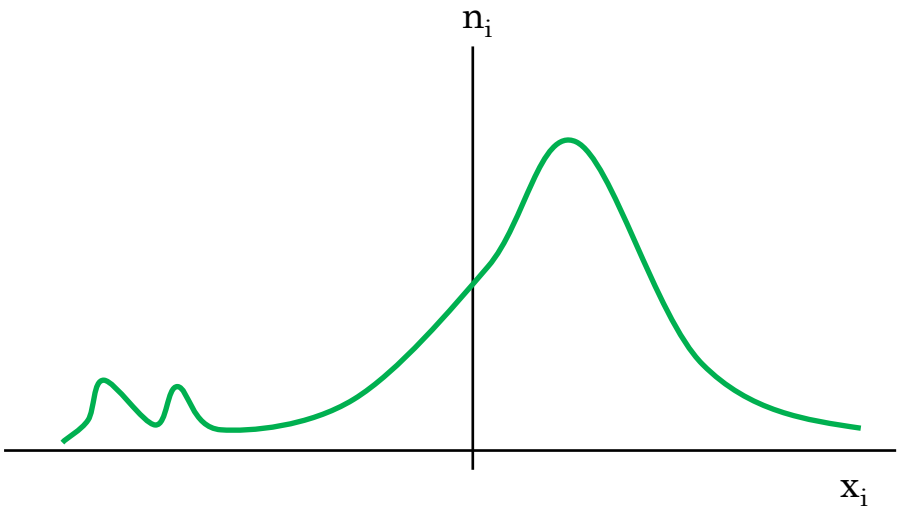


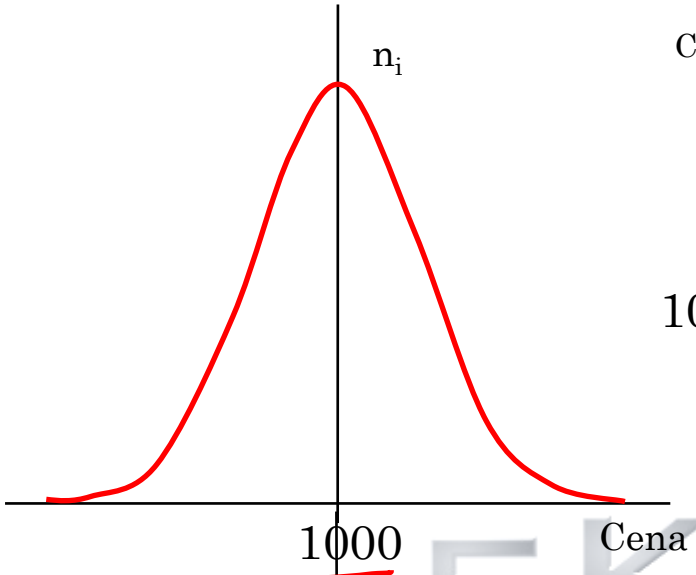




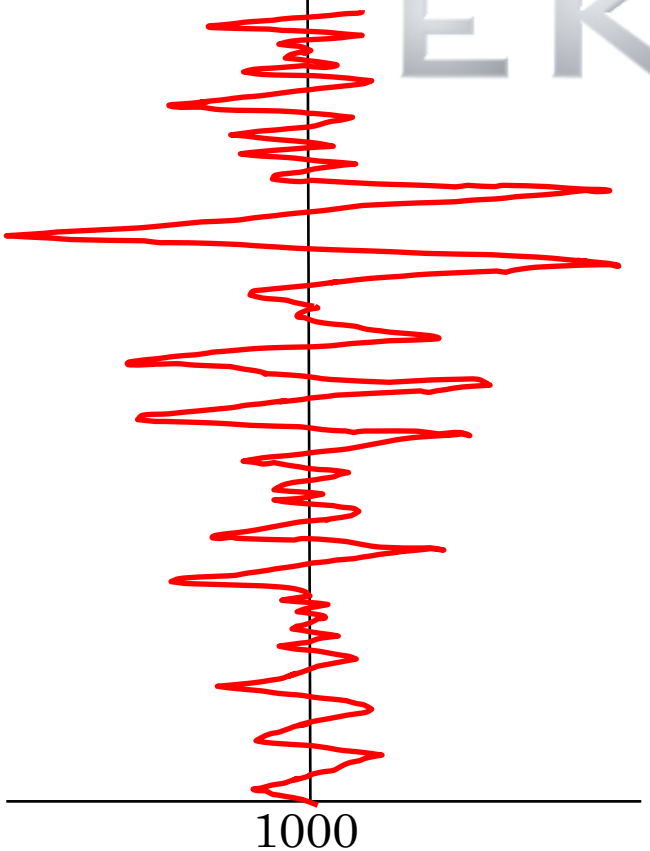


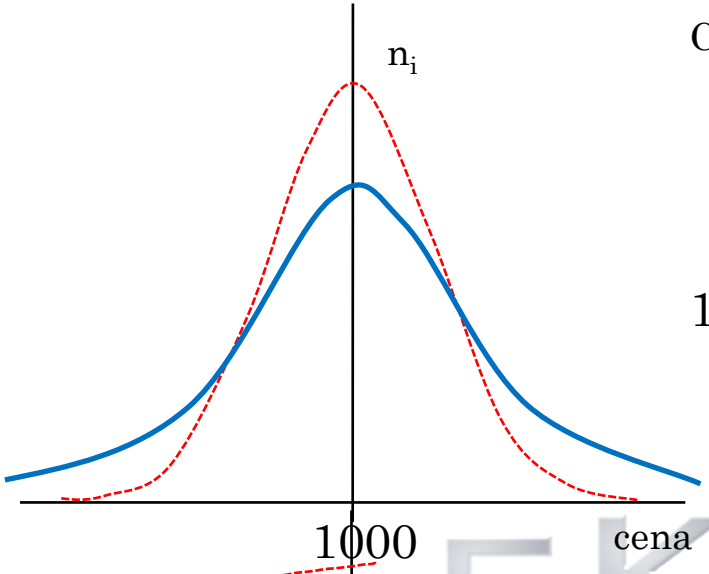
EKOFUN

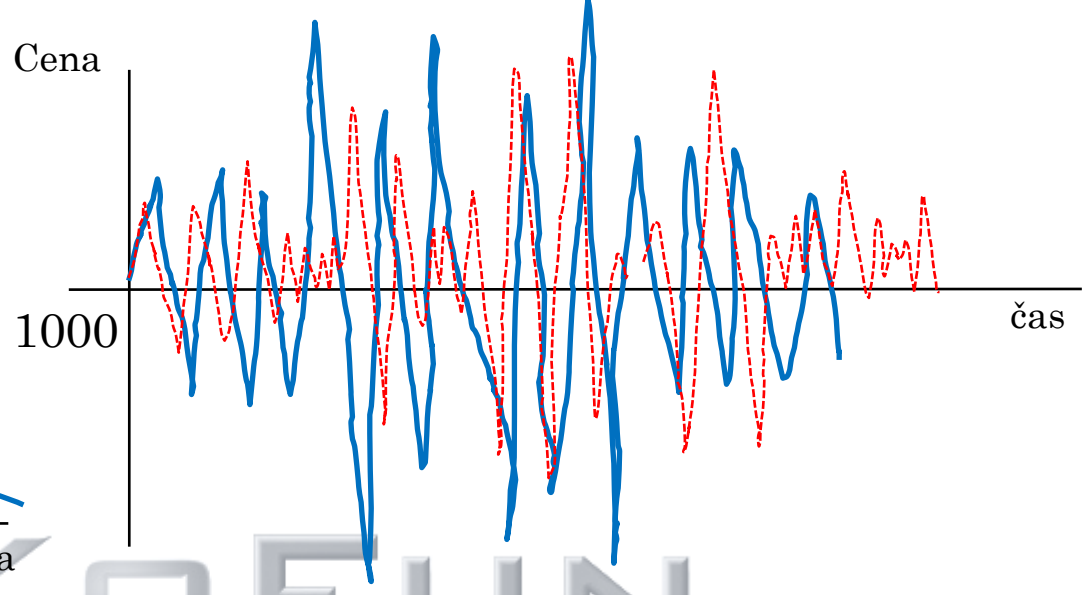




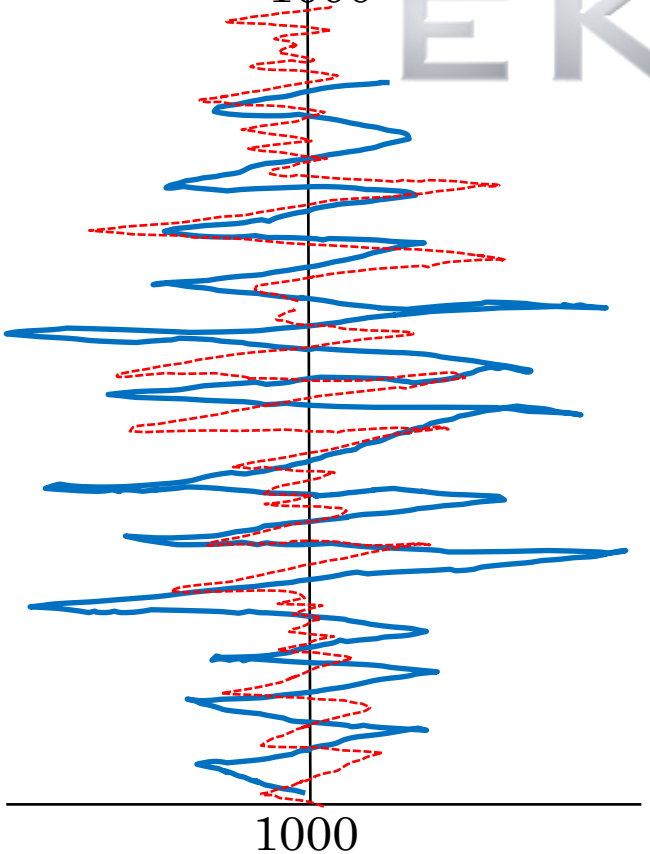
EKOFUN

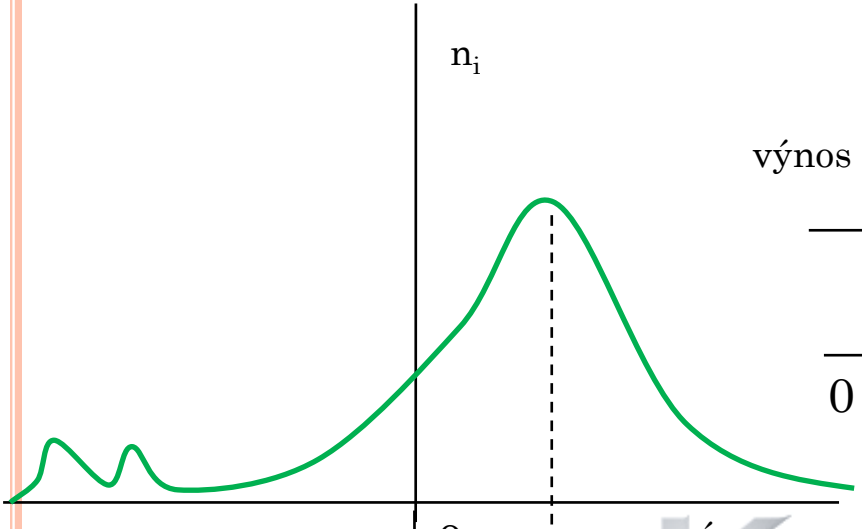




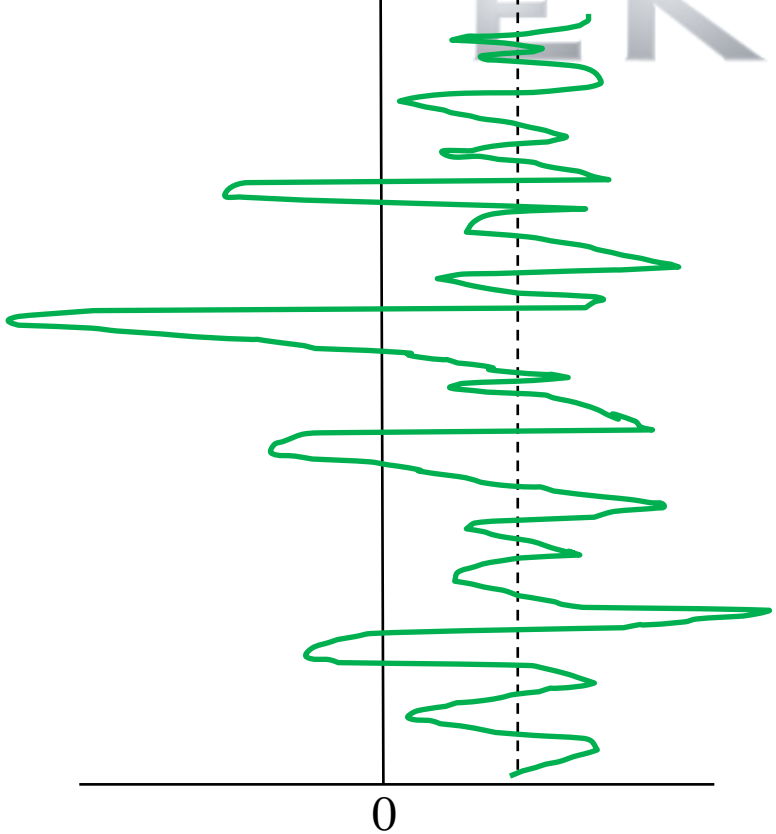


EKOFUN





EKOFUN



Nejčastější hodna nemusí být průměr ani medián!!!