

Statistické zpracování dat 1.přednáška

Mgr. Radmila Krkošková, Ph.D.



**SLEZSKÁ
UNIVERZITA**

OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Podmínky absolvování předmětu:



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Během zimního semestru se konají 3 průběžné testy: v 5., 8. a 11. výukovém týdnu v rámci seminářů.
- **19. týden – samostudium (videopřednáška + seminář)
(týden od 20. 11. 2023)**
- Za průběžné testy je možno získat až 30 bodů, které se započítávají do výsledné známky z předmětu Statistické zpracování dat.
- Závěrečný zkouškový písemný test obsahuje teoretické i praktické otázky (řešení příkladů na PC) a je hodnocen 0 až 70 body. K úspěšnému absolvování předmětu je zapotřebí získat alespoň 60 bodů ze 100 bodů (TEST + ZKOUŠKA)



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Klasifikace:

- 0 až 59 b. F
- 60 až 64 b E
- 65 až 69 b. D
- 70 až 79 b. C
- 80 až 89 b. B
- 90 až 100 b. A



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Statistická jednotka



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Objekty statistického zkoumání –
statistické jednotky

= to co zkoumá statistika

Příklady: zákazníci, zaměstnanci, firmy, organizace určitého typu: prodejny potravin, supermarkety (např. Hypernova), studenti SU OPF, voliči, výrobky (např. televizory, počítače aj.), události (uzávěrky, úrazy, vrhy hrací kostkou apod.)



1. Úkol statistiky: zpřehlednění dat

Cíl: analýza informací a odhalení zákonitostí skrytých v datech

2 přístupy:

1. **Popisná statistika (charakteristiky)**
2. **Induktivní statistika (pravděpodobnost rozdělení)**



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Vymezení statistické jednotky



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Statistická jednotka je vymezena alespoň ze 3 hledisek:

- **věcné hledisko**
(např. student VŠ mužského pohlaví)
- **prostorové hledisko**
(např. student SU OPF v Karviné)
- **časové hledisko** (např. v letošním školním roce student 1. ročníku)



Statistický soubor



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Statistický soubor = souhrn statistických jednotek **stejného** vymezení (věcného, prostorového, časového)

- Statistický soubor, který obsahuje *všechny* statistické jednotky daného vymezení - *základní soubor* (též *populační soubor* nebo krátce *populace*)
- Vybraná část základního souboru - *výběrový soubor*, též *vzorek*



Statistický znak



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Statistické znaky = vlastnosti statistických jednotek statistických souborů

- **znaky kvalitativní** (někdy též slovní, textové nebo alfanumerické)

Příklady: pohlaví zákazníka, typ podniku, bydliště voliče, barva výrobku, chuť nápoje, spokojenost zákazníka apod.

- **znaky kvantitativní** (též číselné, metrické, měřitelné)

Příklady: tržby firmy za měsíc, cena výrobku, počet zákazníků za den, HDP státu v USD, výsledky vrhu hrací kostkou apod.



Kvalitativní znaky



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Kvalitativní znaky členíme na:

- **nominální znaky** (též jmenovité)

Příklad : kategoriemi znaku „pohlaví zákazníka“ jsou „Muž“ a „Žena“
– **kategorie jsou rovnocenné**

- **ordinální znaky** (též pořadové)

Příklad : kategoriemi znaku „spokojenost zákazníka“ mohou být 3 výrazy „nízká“, „průměrná“ a „vysoká“, neboli 3 kódy „1“, „2“ a „3“
nejedná se o kvantitativní (číselný) znak !!!

Kategorie nejsou rovnocenné, lze je uspořádat



Kvantitativní znaky



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Kvantitativní znaky členíme na:

- **diskrétní znaky** (mají konečný nebo nekonečný počet hodnot, nabývají izolovaných číselných hodnot)

Příklady:

1. Počet zákazníků v prodejně za den - nabývá hodnot 0, 1, 2, 3, ... atd., není shora omezen (alespoň teoreticky)
- **nekonečný** diskrétní znak
2. Počet ok na hrací kostce je omezený, konkrétně nabývá hodnot 1, 2, ..., 6, - **konečný** diskrétní znak



Kvantitativní znaky



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **spojité znaky** (mají **vždy** nekonečný počet hodnot, nabývají všech možných číselných hodnot z určitého číselného intervalu)

Příklady:

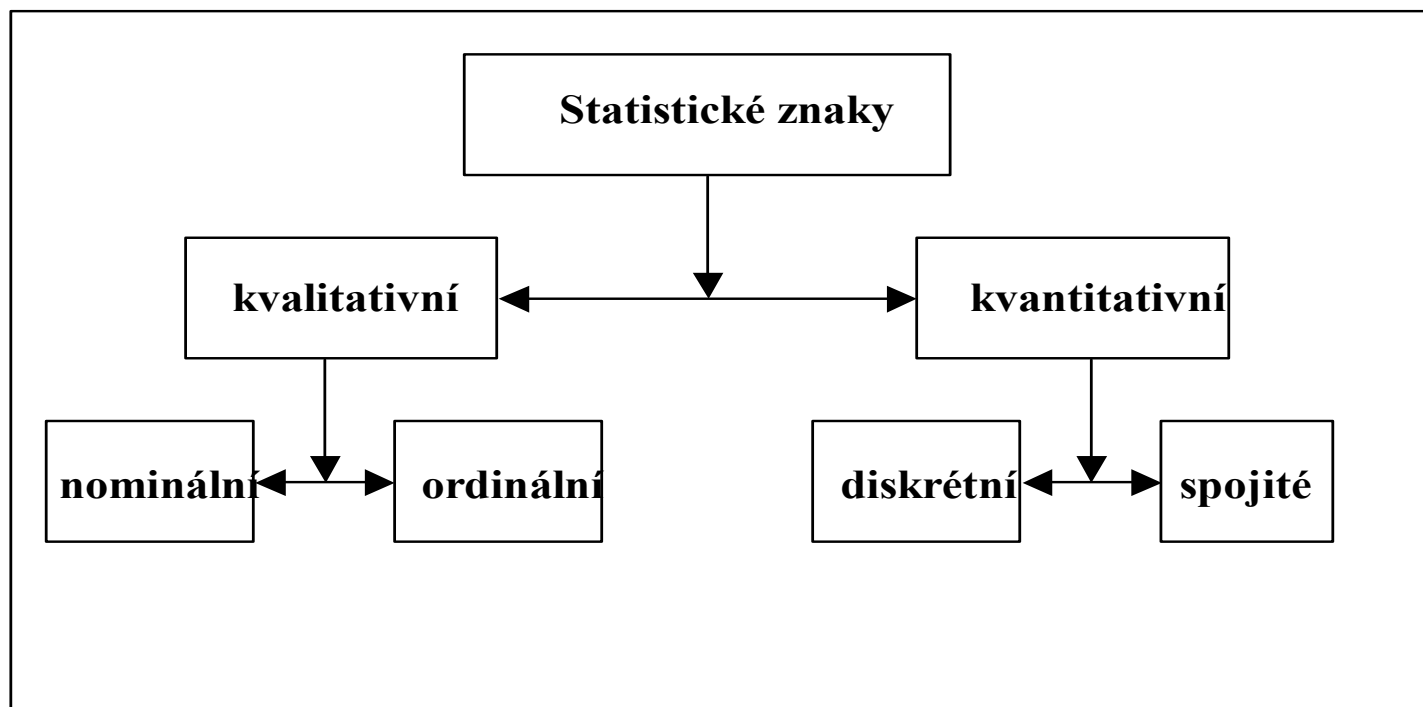
1. Cena výrobku, doba životnosti výrobku - nabývá hodnot z intervalu $(0, +\infty)$ není shora omezen (alespoň teoreticky)
2. Hmotnost výrobku, rozměry výrobku apod. - nabývá hodnot z intervalu $[a, b]$, je omezený



Statistický znak



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Příklad 1. Firma



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

ID	Pohlaví	Věk	Stav	Vzdělání	Funkce	MzdaRok	KvalitaStravy
5060	M	65	rozvedený	SŠ	dělník	258800	4
1030	M	60	rozvedený	VŠ	manažér	630000	2
3049	M	60	ženatý	ZŠ	technik	436600	5
5047	M	60	vdovec	ZŠOU	dělník	240600	3
5061	M	60	vdovec	ZŠOU	dělník	241800	1
5087	M	60	vdovec	SŠ	dělník	239500	
5133	Z	60	vdaná	SŠ	dělník	241100	4
5177	Z	60	vdova	SŠ	dělník	239600	4
3030	Z	58	vdova	ZŠ	technik	422600	1
3014	Z	56	vdova	VŠ	technik	303600	3
5012	Z	56	vdova	ZŠOU	dělník	223100	4
5056	M	56	rozvedený	ZŠ	dělník	225200	5
5101	M	56	svobodný	ZŠOU	dělník	224600	4
5106	M	56	ženatý	ZŠOU	dělník	226100	7
5146	Z	56	vdaná	ZŠOU	dělník	224900	3
5153	M	56	rozvedený	SŠ	dělník	224500	4
5189	M	56	ženatý	ZŠOU	dělník	224600	1
5196	M	56	vdovec	ZŠOU	dělník	222800	3
1031	M	55	ženatý	VŠ	manažér	429000	
5016	M	55	rozvedený	SŠ	admin. pr.	259000	5
5021	Z	55	vdaná	ZŠOU	dělník	220200	
5062	Z	55	vdova	ZŠOU	dělník	221400	5
5107	M	55	rozvedený	ZŠOU	dělník	220500	4
5154	Z	55	vdova	ZŠOU	dělník	219200	5
5195	M	55	ženatý	ZŠOU	dělník	219400	6



Statistický znak (nominální) – Funkce



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

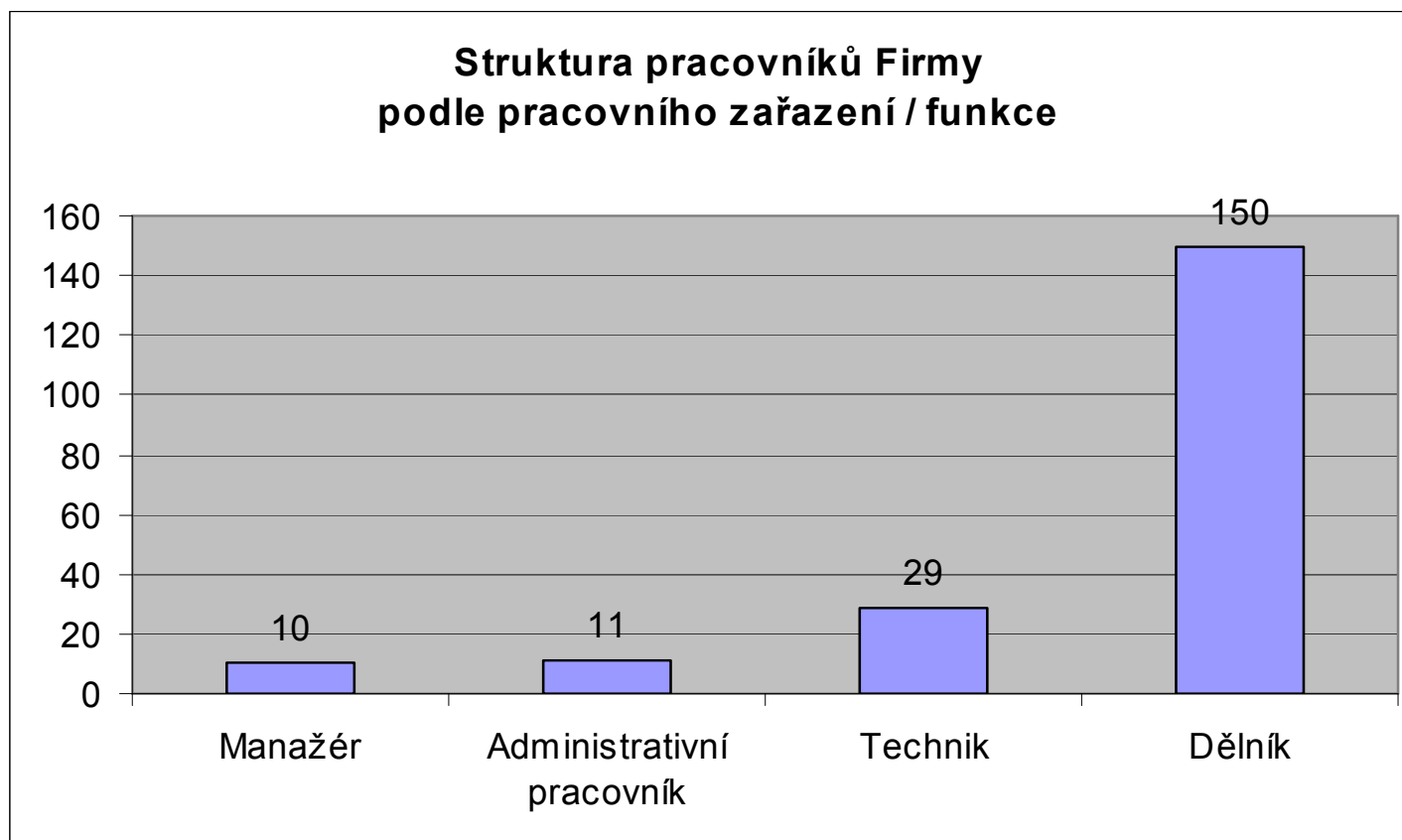
Pracovní zařazení/Funkce	Četnost	Relativní četnost
Manažér	10	5,0%
Administrativní pracovník	11	5,5%
Technik	29	14,5%
Dělník	150	75,0%
Celkem	200	100,0%



Histogram četnosti



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Charakteristiky polohy kvalitativních znaků



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **Modus** - \hat{x} nejčastější hodnota (kategorie) kvalitativního znaku x v daném statistickém souboru

Příklad: \hat{x} = „dělník“

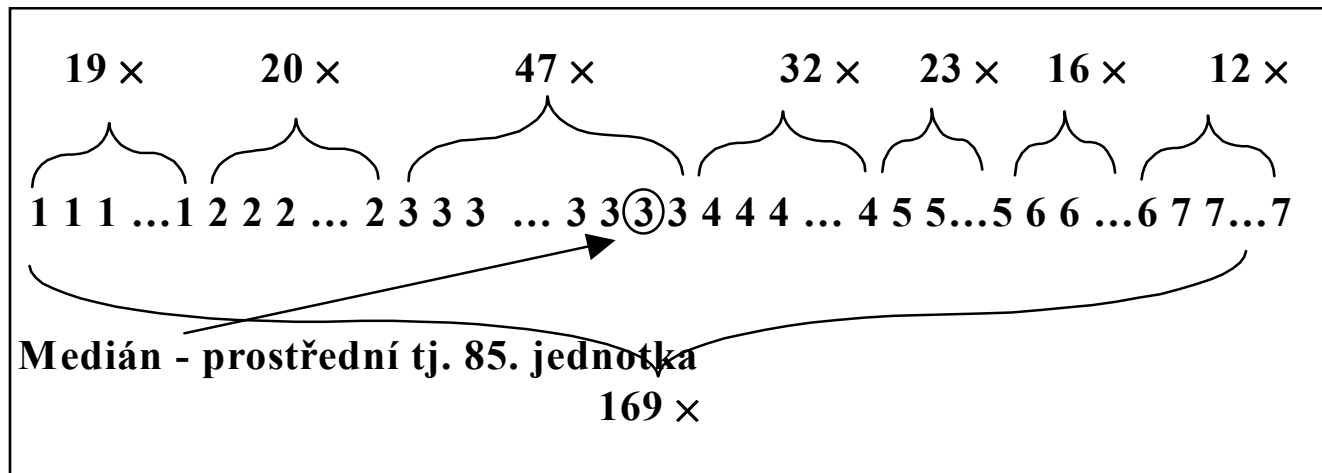
- **Medián** - \tilde{x} představuje hodnotu odpovídající prostřední jednotce v souboru jednotek uspořádaných podle ordinálního znaku x



Kvalita stravy - medián



Příklad: $\tilde{x} = 3 =$ „poměrně špatná kvalita“



Charakteristiky polohy



- **Aritmetický průměr:**
populační průměr -

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

- **výběrový průměr -**

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

- **Vážený průměr:**

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



Charakteristiky polohy



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **Medián** - \tilde{x} - prostřední hodnota
v uspořádaném souboru hodnot
(50% hodnot je menších než medián,
50% hodnot je větších, nebo stejných)
- **Modus** - \hat{x} - nejčastější hodnota
(může jich být i více)



Příklad – průměr, modus, medián



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

ID	Pohlaví	Věk	Stav	Vzdělání	Funkce	MzdaRok	KvalitaStravy	KvalitaPracPod
5093	M	18	ženatý	ZŠ	dělník	71000	3	1
5135	M	18	svobodný	ZŠ	dělník	72400		
5049	M	18	svobodný	SŠ	dělník	72900	6	4
5179	Z	18	svobodná	ZŠ	dělník	73100		
5002	M	18	svobodný	SŠ	dělník	73400	5	7
2019	M	19	svobodný	SŠ	adm.prac.	97500	7	3
2026	M	20	rozvedený	SŠ	adm.prac.	96700		
2021	M	23	svobodný	VŠ	adm.prac.	117800		
5029	M	24	ženatý	ZŠOU	dělník	95100	3	6
...
...
3030	Z	58	vdova	ZŠ	technik	422600	1	3
5087	M	60	vdovec	SŠ	dělník	239500		
5177	Z	60	vdova	SŠ	dělník	239600	4	3
5047	M	60	vdovec	ZŠOU	dělník	240600	3	3
5133	Z	60	vdaná	SŠ	dělník	241100	4	3
5061	M	60	vdovec	ZŠOU	dělník	241800	1	3
3049	M	60	ženatý	ZŠ	technik	436600	5	5
1030	M	60	rozvedený	VŠ	manažér	630000	2	6
5060	M	62	rozvedený	SŠ	dělník	258800	4	4
PRUMĚR		39,92				201822,00		
MEDIÁN		42,00				178700,00		
MODE		45,00				98400,00		



Výběrové x populační charakteristiky



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Výběrový průměr:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = (18 + 18 + 18 + 18 + 18 + 19 + 20 + 23 + 24) / 9 = 19,6$$

Výběrový modus: $\hat{x} = 18$

Výběrový medián: $\tilde{x} = 18$

Populační charakteristiky: $\mu = 39,9$ $\tilde{x} = 42$ $\hat{x} = 45$



Průměrná měsíční mzda

prac.	2008	2009
A	10	+
B	10	+
C	10	+
D	20	20
E	20	20
F	20	20
G	20	20
H	20	20
I	20	20
J	20	20
K	20	20
L	20	20
M	20	20
N	50	50
O	50	50
Aritm.pr.	22	25
median	20	20

Aritmetický průměr roste, přestože nikomu se mzda nezvýšila, 3 pracovníci byli propuštěni (s nejnižším platem)!



Populační charakteristiky variability



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **Rozpětí:** $R = \max x_i - \min x_i$

- **Rozptyl:**
$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2$$

- **Směrodatná odchylka:**

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2}$$



Výběrové charakteristiky variability



- Výběrový rozptyl:

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

- Výběrová směrodatná odchylka:

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



Variační koeficienty



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- *Variační koeficient* $V = \frac{\sigma}{\mu}$

- *Výběrový variační koeficient* $v = \frac{s}{\bar{x}}$



Příklad



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

$$\text{Akcie 1: } \bar{x} = 100 \quad s = 60 \Rightarrow V_1 = 60/100 = 0,6$$

$$\text{Akcie 2: } \bar{x} = 600 \quad s = 420 \Rightarrow V_2 = 420/600 = 0,7$$

Akcie 1 je méně riziková, neboť variační koeficient má menší hodnotu!



- *Šikmost* vyjadřuje tvar rozdělení četnosti pomocí jediného čísla:

$$S_k = \frac{3(\mu - \tilde{x})}{\sigma}$$

- Pokud $S_k = 0$, potom je histogram četnosti **symetrický** v tom smyslu, že

$$\mu = \tilde{x}$$

Šikmost v grafu četnosti



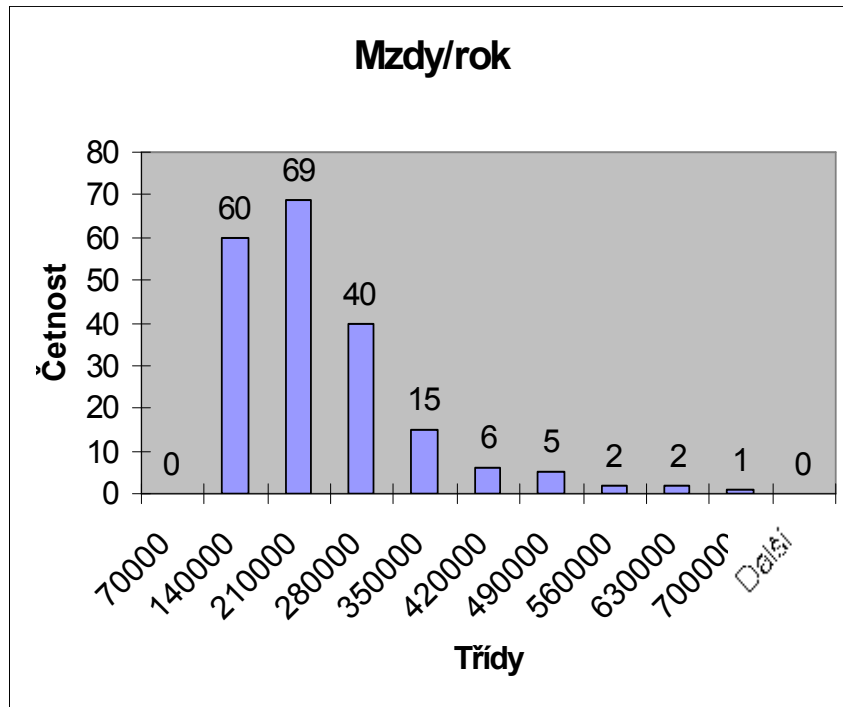
**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Šikmost je menší než 0 (záporná), jestliže je graf četnosti zešikmen doprava.

Šikmost je větší než 0 (kladná), jestliže je graf zešikmen doleva.



Příklad

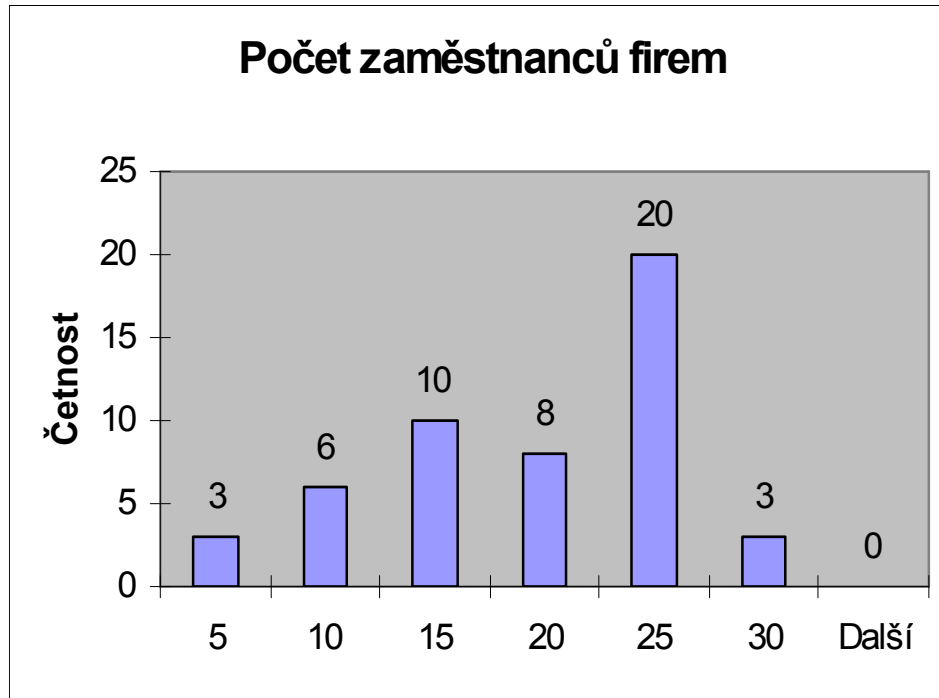


$$S_k = 0,99 > 0$$

Graf je sešikmen doleva.



Příklad



$$S_k = -0,51 < 0$$

Graf je sešikmen doprava.





Děkuji Vám za pozornost !!!

