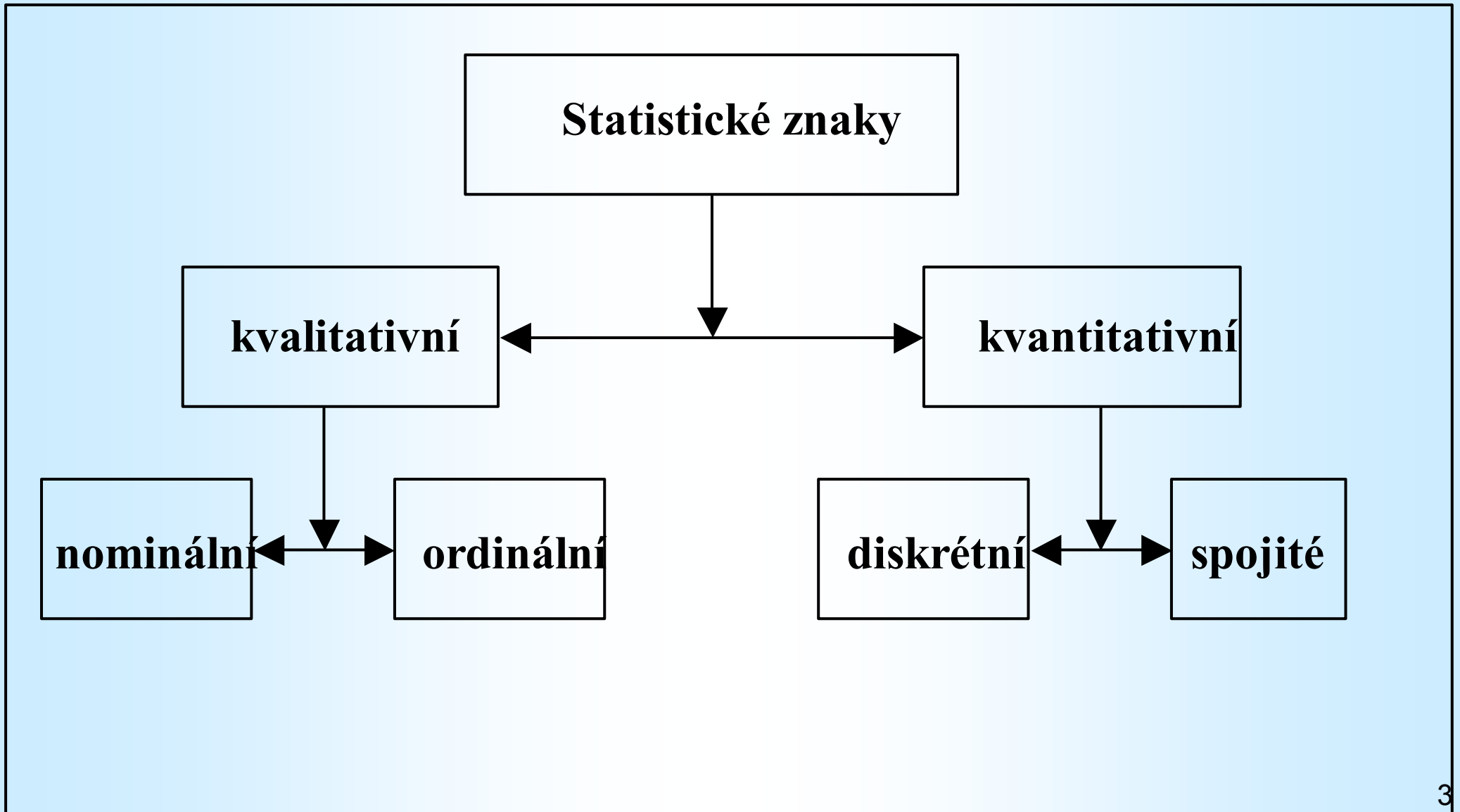


II. Popisná statistika: kvalitativní a kvantitativní znaky

Hlavní úkol statistiky: zpřehlednění dat

- **Cíl:** analýza informací a odhalení zákonitostí skrytých v datech
- 2 přístupy:
 1. **Popisná statistika**
 2. Induktivní statistika
- **Ad 1:** kategorizace, charakteristiky - nyní
- **Ad 2:** předpoklady o vzniku dat (pravděpodobnostní rozdělení) - později

Statistický znak



Příklad 1. Firma

Výběrový soubor, statistické jednotky: zaměstnanci

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

Metody zpřehlednění dat

- **Histogram četnosti** - kvalitativní znaky (nominální, ordinální)
- **Charakteristiky polohy** - kvalitativní znaky (nominální, ordinální)
- **Histogram četnosti** - kvantitativní znaky (kardinální)
- **Charakteristiky polohy** - kvantitativní znaky
- **Charakteristiky variability** - kvantitativní znaky

Histogram četnosti

kvalitativních znaků 1

Příklad 2. Statist. soubor - Pracovníci Firmy X
Statist. znak (nominální) – *Funkce*

Tabulka:

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

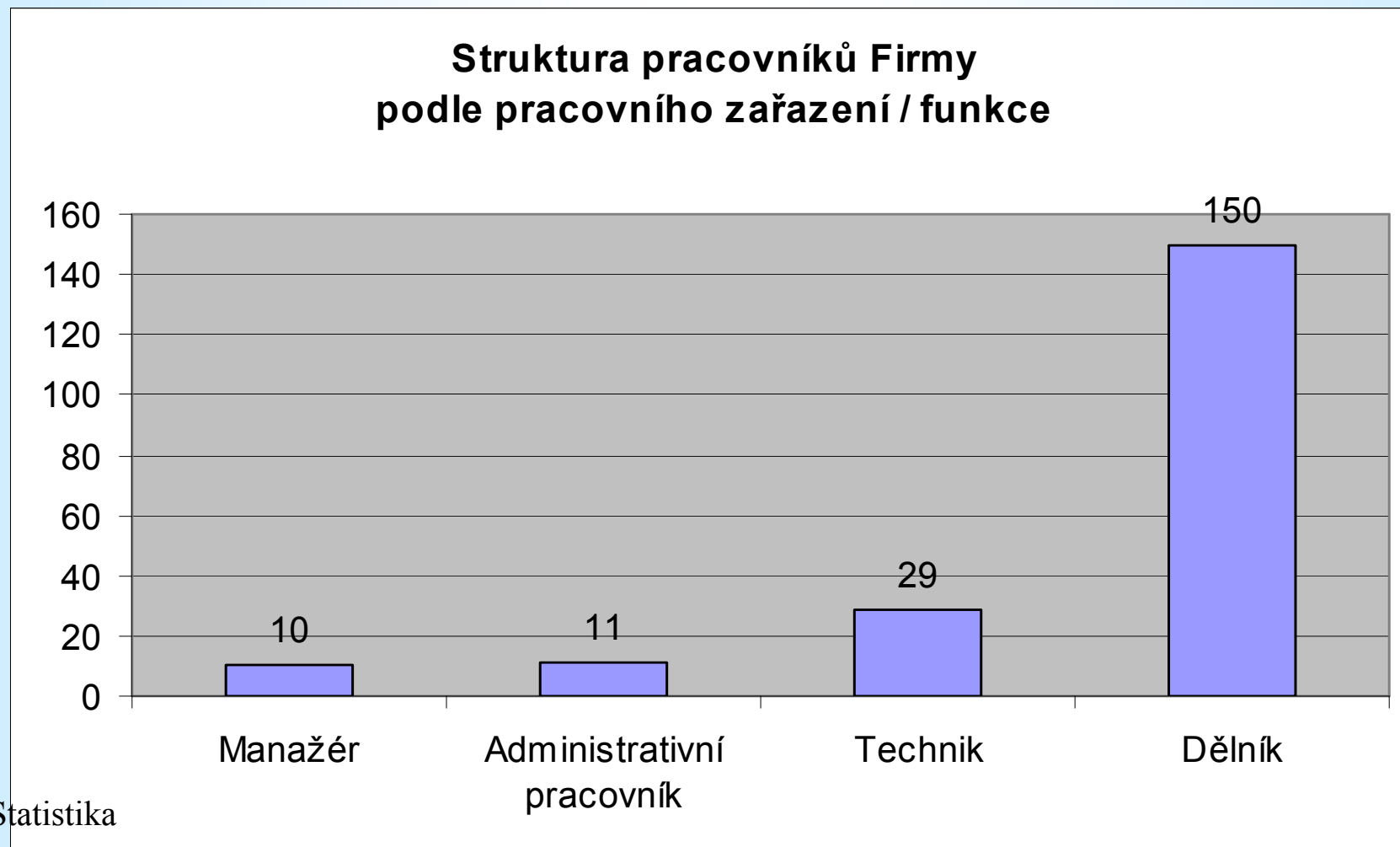
$$=100*(10/200)\%$$

Histogram četnosti

kvalitativních znaků 2 (sloupcový graf)

Stat. soubor - Pracovníci Firmy X

Stat. znak (nominální) - *Funkce*



Příklad 3. Ordinální znak **Kvalita stravy**

Kód kvality - třída	Počet	Relat. četnost
1 -velmi špatná	19	11,2%
2 - špatná	20	11,8%
3 -poměrně špatná	47	27,8%
4 -ani tak ani onak	32	18,9%
5 -poměrně dobrá	23	13,6%
6 -dobrá	16	9,5%
7 -velmi dobrá	12	7,1%
Suma	169	100,0%

Příklad 3. Ordinální znak - **Kvalita stravy**

Hodnocení kvality stravy pracovníky Firmy

Legenda: 1 - velmi špatná

2 - špatná

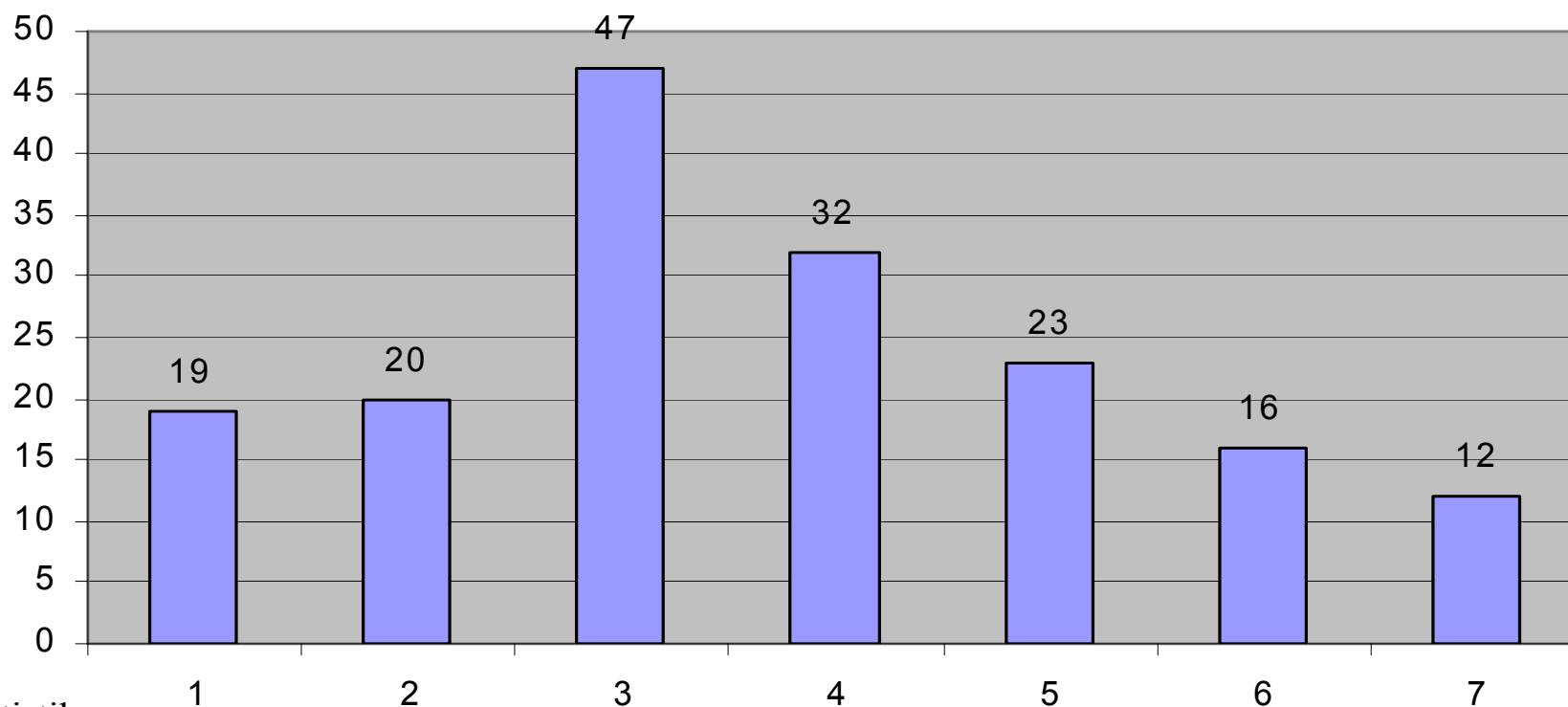
3 - poměrně špatná

4 - ani tak, ani onak

5 - poměrně dobrá

6 - dobrá

7 - velmi dobrá



Charakteristiky polohy kvalitativních znaků

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

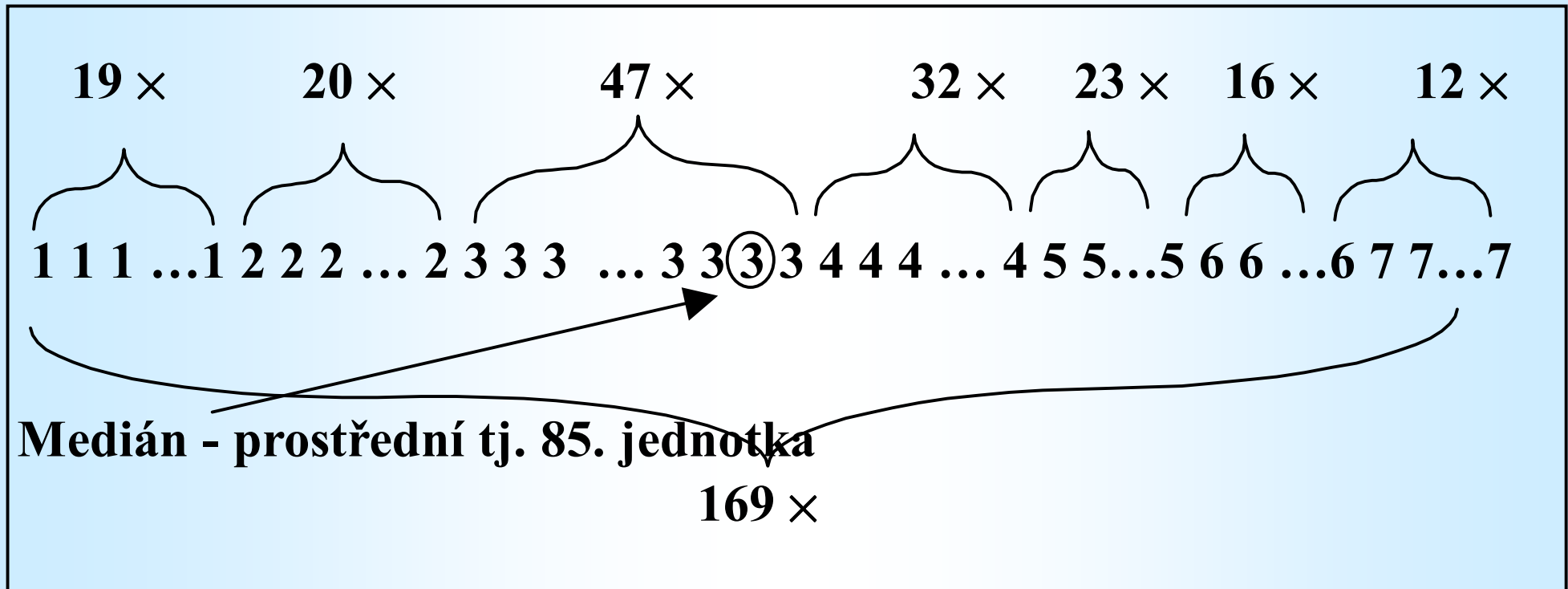
Příklad: $\hat{x} = \text{„dělník“}$

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

Soubor: Firma X, statistický znak:

„Kvalita stravy“, medián:

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



Příklad 1. Firma



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
5196	M	55	ženatý	ZŠOU	dělník	219400	6

Histogram četnosti

kvantitativních znaků

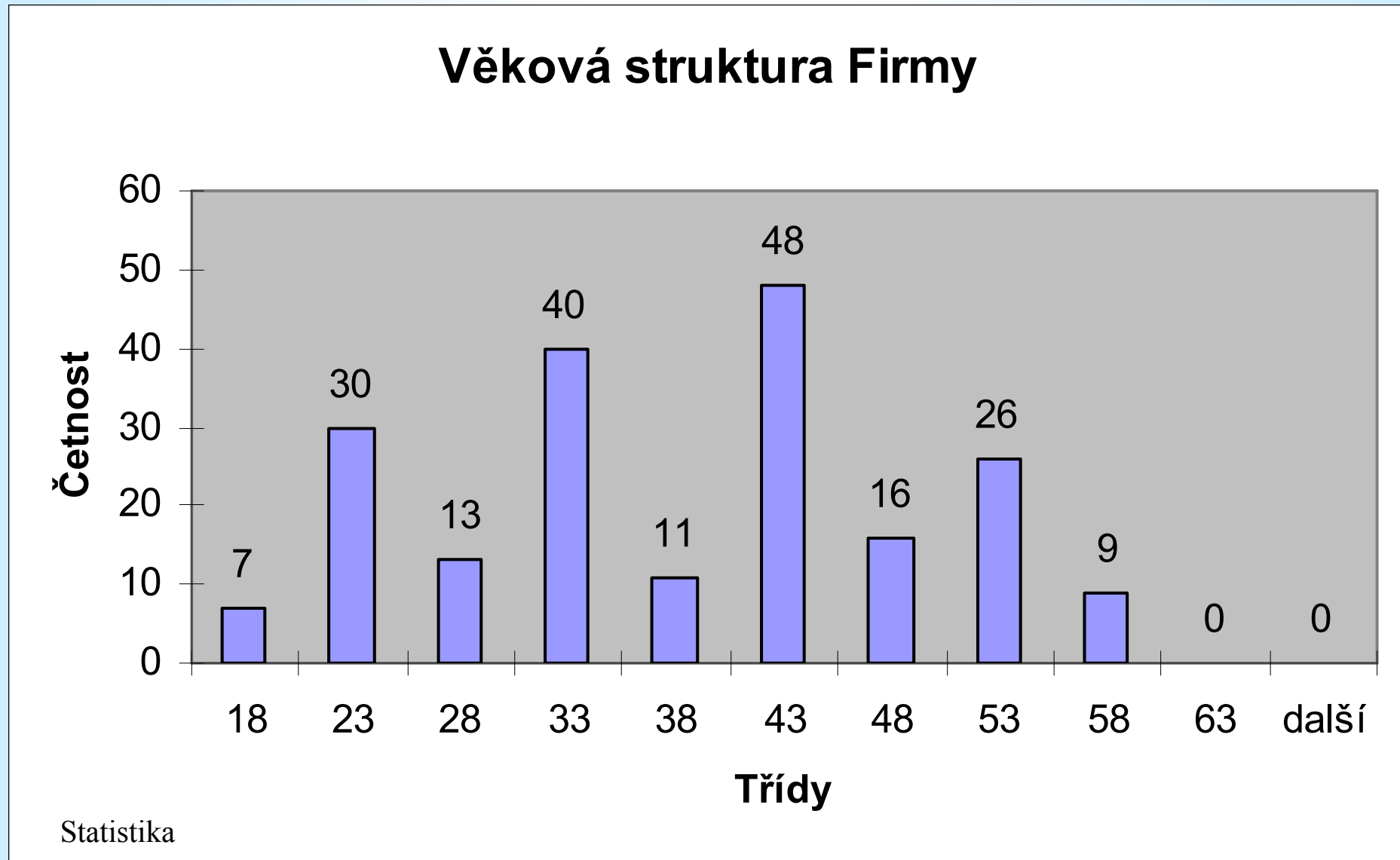
Příklad 4. Věk pracovníků Firmy

<i>Třídy</i>	<i>Četnost</i>	<i>Kumulat. četnost</i>	<i>Relat. četnost</i>	<i>Kumul. rel. četnost</i>
18	7	7	0,035	0,035
23	30	37	0,150	0,185
28	13	50	0,065	0,250
33	40	90	0,200	0,450
38	11	101	0,055	0,505
43	48	149	0,240	0,745
48	16	165	0,080	0,825
53	26	191	0,130	0,955
58	9	200	0,045	1,000
63	0	200	0,000	1,000
další	0	0	0,000	0,000

Histogram četnosti

kvantitativních znaků

Příklad 4. Věk pracovníků Firmy



Jak určit počet tříd v histogramu?

- U kvalitativních znaků: **třída = kategorie**
- U kvantitativních znaků:

Sturgersovo pravidlo:

$$N = \text{ZAOKROUHLIT}(3,3 \log_{10} n) + 1$$

N – počet tříd, n – počet dat

$$\text{Šířka třídy} \approx (\text{MAX} - \text{MIN})/N$$

Příklad 5:

Firma X, *Mzda za rok*

Počet tříd:

$$n = 200 \Rightarrow$$

$$N = \text{ZAOKROUHLIT}(3,3 * 2,301) + 1 = 9$$

Šířka tříd:

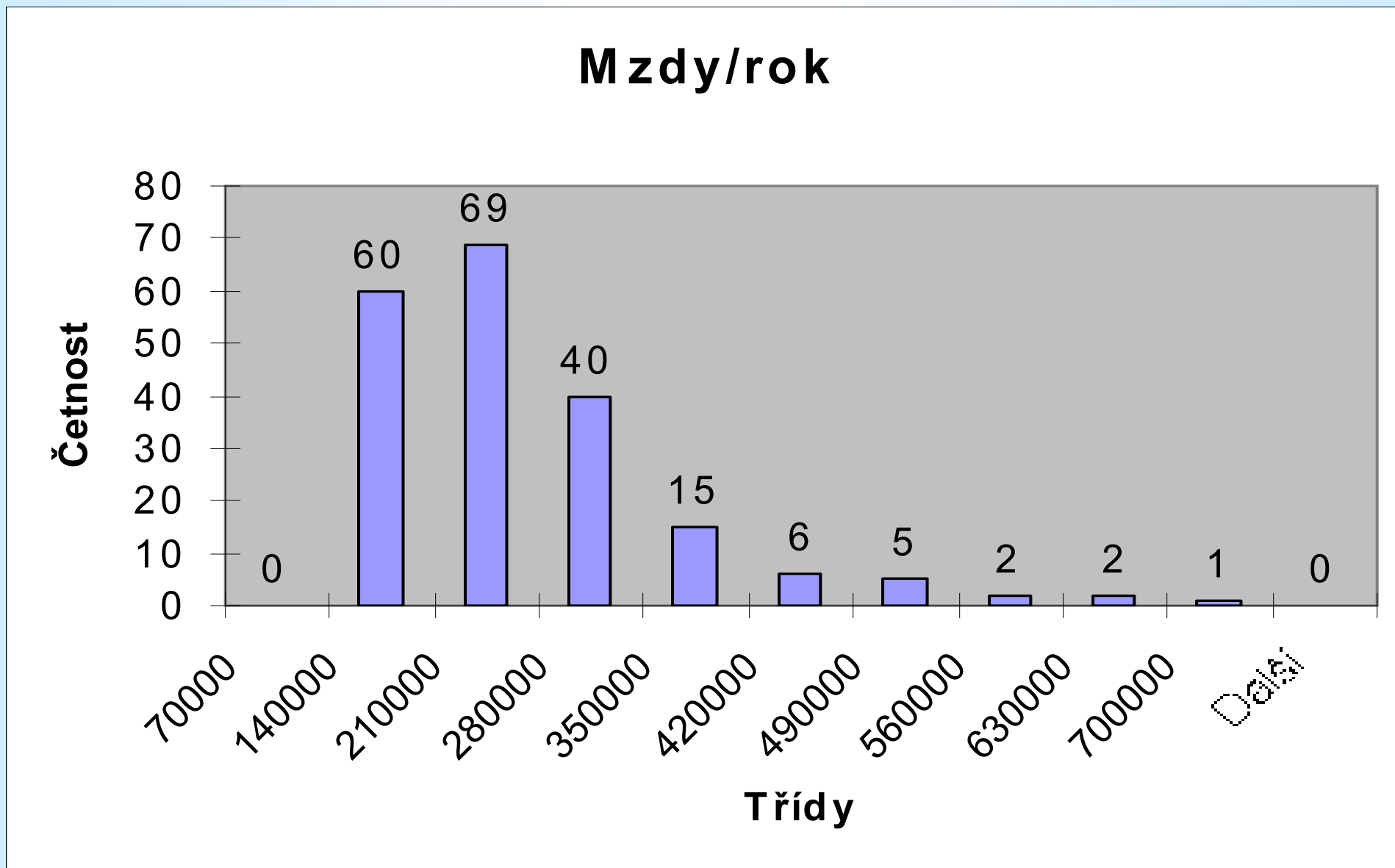
$$\text{max} = 657\ 000, \text{min} = 71\ 000 \Rightarrow$$

$$d = (657000 - 71000) / 9 = 65\ 111 \approx 70\ 000$$

Příklad 5. Roční mzda

<i>Třídy</i>	<i>Četnost</i>	<i>Kumulat. četnost</i>	<i>Relat. četnost</i>	<i>Kumul. rel. četnost</i>
70000	0	0	0,000	0,000
140000	60	60	0,300	0,300
210000	69	129	0,345	0,645
280000	40	169	0,200	0,845
350000	15	184	0,075	0,920
420000	6	190	0,030	0,950
490000	5	195	0,025	0,975
560000	2	197	0,010	0,985
630000	2	199	0,010	0,995
700000	1	200	0,005	1,000
Další	0	200	0,000	1,000

Příklad 5. Roční mzda ve firmě



Charakteristiky polohy 1

- **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 2

- **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)
- Též výběrový medián a výběrový modus

Vzorek 9 jednotek

Příklad 6.

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

PRŮMĚR	39,92	Populace 200 jednotek	201822,00	21
MEDIÁN	42,00		178700,00	
MODUS	45,00		98400,00	

Statistika

Příklad 6.

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 medián:

$$\tilde{x} = 18$$

Výběrový modus:

$$\hat{x} = 18$$

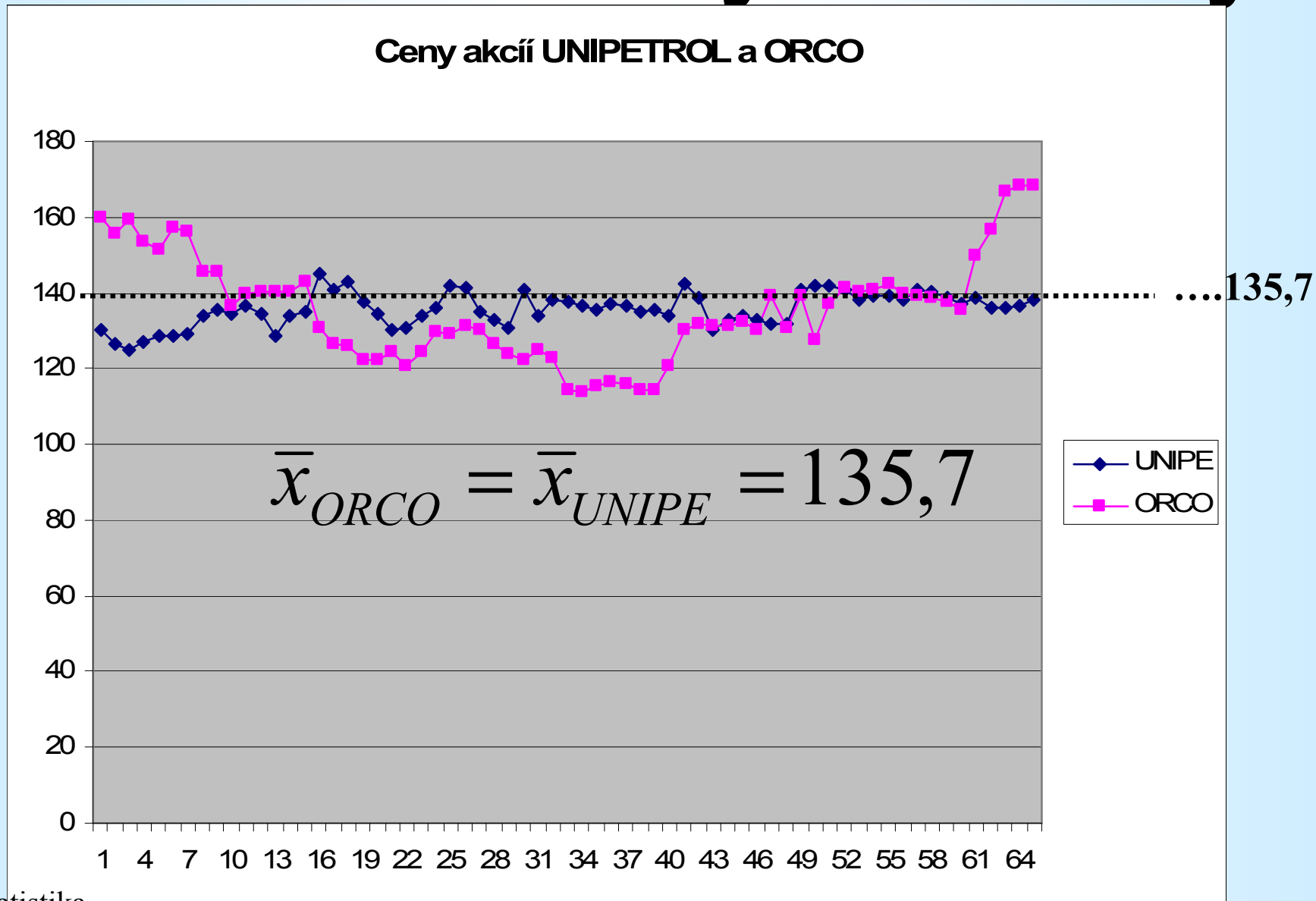
Populační charakteristiky:

$$\mu = 39,9 \quad \tilde{x} = 42 \quad \hat{x} = 45$$

Průměrná měsíční mzda

pracovník	2010	2011
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

Charakteristiky variability



Populační charakteristiky variability

- **Rozpětí:** $R = \text{MAX } x_i - \text{MIN } x_i$

- **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

zachycují vztah variability k průměru

- **Variační koeficient (populační):**

$$V = \frac{\sigma}{\mu}$$

- **Výběrový variační koeficient:**

$$v = \frac{s}{\bar{x}}$$

V Excelu nejsou funkce! Je třeba počítat podle vzorce!

Příklad 7. Akcie UNIPE vers. ORCO

$$\bar{x}_{UNIPE} = 135,7 \quad s = 2,09 \quad \Rightarrow \quad V_{UNIPE} = 2,09/135,7 = 0,015$$

tj. riziko = 1,5%

$$\bar{x}_{ORCO} = 135,7 \quad s = 3,72 \quad \Rightarrow \quad V_{ORCO} = 3,72/135,7 = 0,027$$

tj. riziko = 2,7%

Akcie UNIPE jsou méně riziková než ORCO!

Konkrétně: V uvedeném období jsou akcie UNIPE
1,8 krát méně rizikové než ORCO!

Šikmost

- **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
aritmetický průměr = medián, tj.

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

Šikmost

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

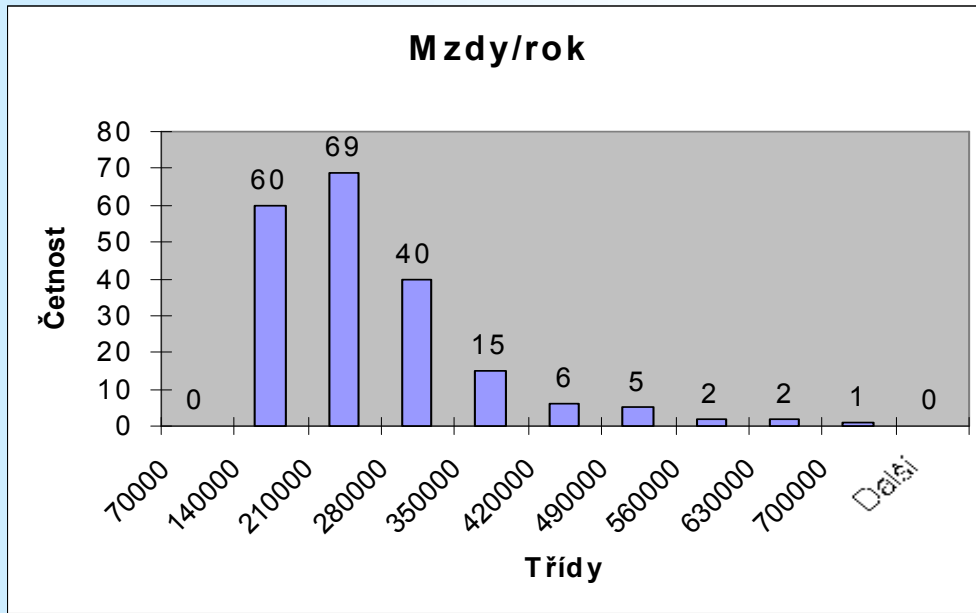
- Šikmost je menší než 0 (záporná), když je graf četnosti zešikmen *doprava*
- Šikmost je větší než 0 (kladná), když je graf zešikmen *doleva*

V Excelu funkce: SKEW

Příklad 8.

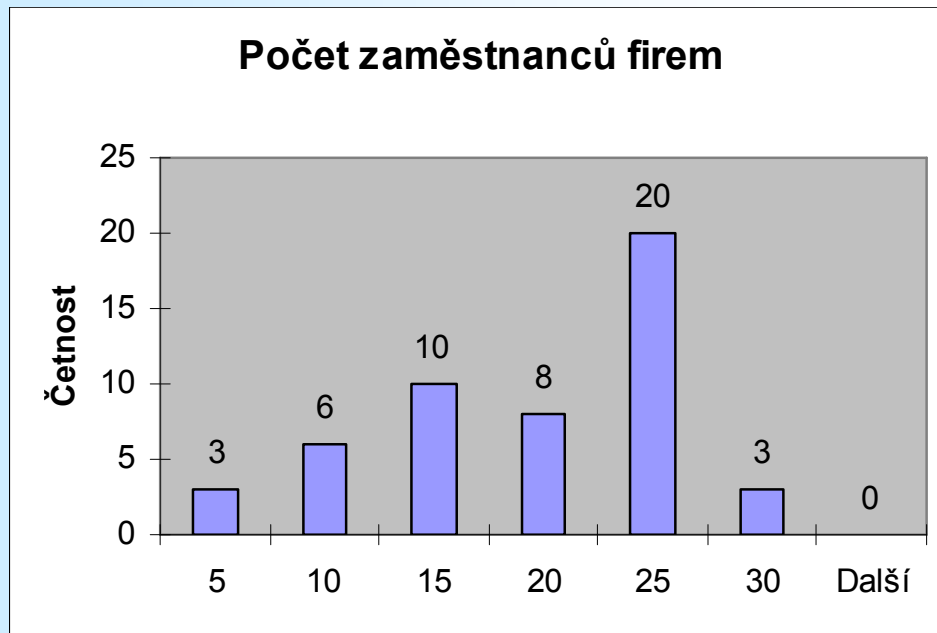
ID	Pohlavi	Vek	Stav	Vzdělání	Funkce	MzdaRok	KvalitaStravy	KvalitaPracPod
5093	M	18	ženatý	ZŠ	dělník	71000	3	1
5135	M	18	svobodný	ZŠ	dělník	72400		
5049	M	18	svobodný	SŠ	dělník	72900	6	4
5179	Z	18	svobodná	ZŠ	dělník	73100		
5002	M	18	svobodný	SŠ	dělník	73400	5	7
2019	M	19	svobodný	SŠ	adm.prac.	97500	7	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		
SMODCH		11,10				101549,08		
VAR.KOEF.		0,28				0,50		
SKREW		-0,56				0,99		

Příklad 9.



$$Sk = 0,99 > 0$$

- graf je sešikmen
(„sešlápnut“)
doleva



$$Sk = - 0,51 < 0$$

- graf je sešikmen
doprava

Shrnutí

- ⑩ Statistické znaky
- ⑩ Metody zpřehlednění dat
- ⑩ Histogram četnosti kvalitativního znaku
- ⑩ Charakteristiky polohy kvalitativního znaku: modus a medián
- ⑩ Histogram četnosti kvantitativního znaku

Shrnutí - dokončení

- ⑩ Sturgersovo pravidlo ke stanovení počtu tříd
- ⑩ Charakteristiky polohy kvantitativního znaku: průměry (střední hodnoty), modus a medián
- ⑩ Charakteristiky variability kvantitativního znaku: rozptyl a směrodatná odchylka
- ⑩ Variační koeficient a jeho použití
- ⑩ Šikmost a tvar grafu