

Příklady z popisné statistiky

Příklad 1.:

a) V následující tabulce jsou uvedeny počty správně vyřešených příkladů u přijímací zkoušky z matematiky a jejich absolutní četnosti.

x_i	0	1	2	3	4
n_i	5	10	16	18	13

Sestavte variační řadu a nakreslete graf četnostní funkce a empirické distribuční funkce. (Návod: definice 2.4., příklad 2.5.)

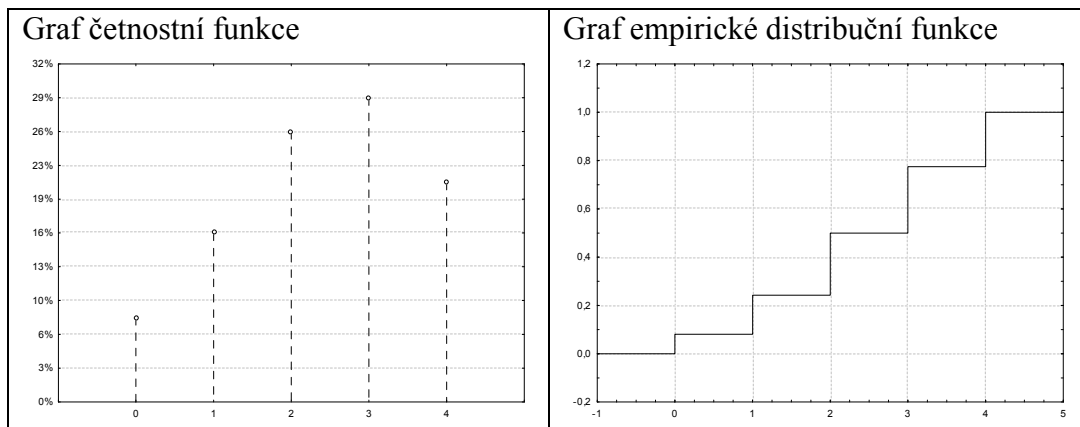
b) Z datového souboru 1,4 9,9 0,2 9,9 9,6 4,1 2,3 0,9 4,8 7,6 1,9 1,0 3,1 8,1 4,5 3,9 0,3 2,8 0,5 3,6 vypočítejte medián a kvartilovou odchylku. (Návod: definice 3.4., příklad 3.5.)

c) Hodnoty znaku X mají aritmetický průměr -1 a rozptyl 0,5. Najděte aritmetický průměr a rozptyl hodnot znaku $Y = -2 + 5X$. (Návod: věta 3.18. (a), příklad 3.19.)

Řešení:

ad a) Variační řada

$x_{[j]}$	n_j	p_j	N_j	F_j
0	5	5/62	5	5/62
1	10	10/62	15	15/62
2	16	16/62	31	31/62
3	18	18/62	49	49/62
4	13	13/62	62	1



ad b) Soubor uspořádáme podle velikosti:

0,2 0,3 0,5 0,9 1,0 1,4 1,9 2,3 2,8 3,1 3,6 3,9 4,1 4,5 4,8 7,6 8,1 9,6 9,9 9,9
Rozsah souboru $n = 20$.

$$\text{Výpočet mediánu: } n\alpha = 20 \cdot 0,5 = 10, \quad x_{0,50} = \frac{x_{(10)} + x_{(11)}}{2} = \frac{3,1 + 3,6}{2} = 3,35.$$

$$\text{Výpočet dolního kvartilu: } n\alpha = 20 \cdot 0,25 = 5, \quad x_{0,25} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{1,0 + 1,4}{2} = 1,2.$$

$$\text{Výpočet horního kvartilu: } n\alpha = 20 \cdot 0,75 = 15, \quad x_{0,75} = \frac{x_{(15)} + x_{(16)}}{2} = \frac{4,8 + 7,6}{2} = 6,2.$$

$$\text{Výpočet kvartilové odchylky: } q = x_{0,75} - x_{0,25} = 6,2 - 1,2 = 5,0.$$

$$\text{ad c) } m_1 = -1, \quad s_1^2 = 0,5, \quad m_2 = -2 + 5m_1 = -2 - 5 = -7, \quad s_2^2 = 5^2 s_1^2 = 25 \cdot 0,5 = 12,5.$$

Příklad 2.: Znak X udává délku praxe (v letech) a znak Y výšku prémie (v Kč) zaměstnanců jisté firmy. Dvourozměrné rozložení četností je dáno kontingenční tabulkou:

x	y						
	1250	1750	2250	2750	3250	3750	4250
12,5	5	3	0	0	0	0	0
17,5	2	4	4	0	0	0	0
22,5	0	1	6	7	4	0	0
27,5	0	0	1	3	7	1	0
32,5	0	0	0	1	10	5	1

- Sestavte kontingenční tabulky sloupcově a řádkově podmíněných relativních četností. (definice 2.7., příklad 2.9.)
- Kolik procent pracovníků s délkou praxe 22,5 roku má prémie nanejvýš 2250 Kč? (příklad 2.9.)
- Jaká je průměrná výše prémie? (definice 3.20.)
- Stanovte modus a medián výše prémie. (definice 3.3, definice 3.4.)

Upozornění: Výsledky udávejte na tři desetinná místa.

Řešení:

Nejprve doplníme tabulku o marginální četnosti.

x	y							$n_{j.}$
	1250	1750	2250	2750	3250	3750	4250	
12,5	5	3	0	0	0	0	0	8
17,5	2	4	4	0	0	0	0	10
22,5	0	1	6	7	4	0	0	18
27,5	0	0	1	3	7	1	0	12
32,5	0	0	0	1	10	5	1	17
$n_{.k}$	7	8	11	11	21	6	1	65

ad a)

Kontingenční tabulka sloupcově podmíněných relativních četností: $p_{j(k)} = \frac{n_{jk}}{n_{.k}}$

x	y						
	1250	1750	2250	2750	3250	3750	4250
12,5	0,714	0,375	0	0	0	0	0
17,5	0,286	0,500	0,364	0	0	0	0
22,5	0	0,125	0,545	0,636	0,190	0	0
27,5	0	0	0,091	0,273	0,333	0,167	0
32,5	0	0	0	0,091	0,477	0,833	1,000

Kontingenční tabulka řádkově podmíněných relativních četností: $p_{(j)k} = \frac{n_{jk}}{n_{.j}}$

x	y						
	1250	1750	2250	2750	3250	3750	4250
12,5	0,625	0,375	0	0	0	0	0
17,5	0,200	0,400	0,400	0	0	0	0
22,5	0	0,056	0,333	0,389	0,222	0	0
27,5	0	0	0,083	0,250	0,584	0,083	0
32,5	0	0	0	0,059	0,588	0,294	0,059

ad b) Ve 3. řádku kontingenční tabulky řádkově podmíněných relativních četností sečteme čísla v 1., 2. a 3. sloupci: $0 + 0,056 + 0,333 = 0,389$. Hledaný údaj je tedy 38,9%.

ad c) $m = (7 \cdot 1250 + 8 \cdot 1750 + 11 \cdot 2250 + 11 \cdot 2750 + 21 \cdot 3250 + 6 \cdot 3750 + 4250) / 65 = 172750 / 65 = 2657,70$ Kč

ad d) Medián $y_{0,50} = y_{(33)} = 2750$ Kč, modus = 3250 Kč

Příklad 3.: Je dán datový soubor 12 1,1 6,3 3,9 11 5,8 2,5 8 4,1 2 9,5 6,6 1,7 3,4 4,9 3 10,3 2,2 5,4 15,5. Stanovíme třídící intervaly $(1, 2)$, $(2, 4)$, $(4, 7)$, $(7, 11)$, $(11, 16)$.

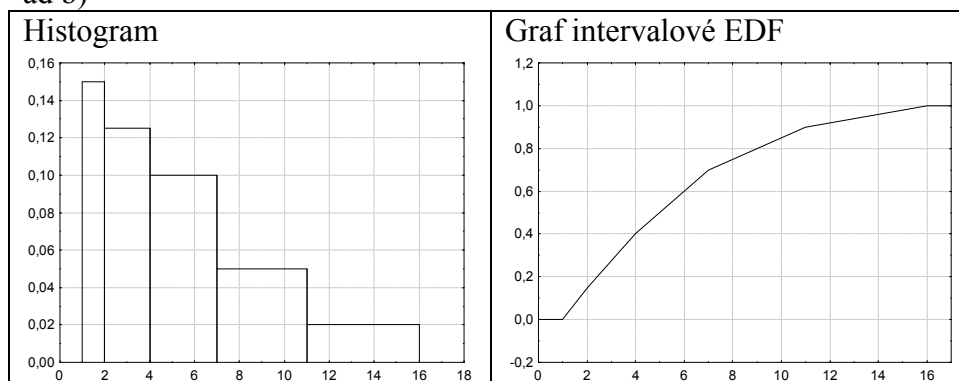
- Sestavte tabulku rozložení četností.
- Nakreslete histogram a graf intervalové empirické distribuční funkce.
- Stanovte medián datového souboru.
- Vypočtěte průměr datového souboru.

Řešení:

ad a)

(u_j, u_{j+1})	$x_{[j]}$	d_j	n_j	p_j	N_j	F_j	f_j
$(1, 2)$	1,5	1	3	$3/20=0,15$	3	$3/20=0,15$	$3/20=0,15$
$(2, 4)$	3	2	5	$5/20=0,25$	8	$8/20=0,4$	$5/40=0,125$
$(4, 7)$	5,5	3	6	$6/20=0,3$	14	$14/20=0,7$	$6/60=0,1$
$(7, 11)$	9	4	4	$4/20=0,2$	18	$18/20=0,9$	$4/80=0,05$
$(11, 16)$	13,5	5	2	$2/20=0,1$	20	$20/20=1$	$2/100=0,02$

ad b)



ad c) Medián je průměr 10. a 11. uspořádané hodnoty, tedy $x_{0,50} = (4,9 + 5,4)/2 = 5,15$

$$m = \frac{1}{n} \sum_{j=1}^r n_j x_{[j]} = \frac{1}{20} (3 \cdot 1,5 + 5 \cdot 3 + 6 \cdot 5,5 + 4 \cdot 9 + 2 \cdot 13,5) = \frac{1}{20} 115,5 = 5,775$$

nebo $m = \frac{1}{n} \sum_{i=1}^n x_i = 5,96$

Příklad 4.: V akciové společnosti je průměrná mzda 18 900 Kč. Přitom 30% pracovníků s nejnižší mzdou má průměrně 14 000 Kč. Na začátku roku dostal každý z těchto pracovníků přidáno 1 500 Kč.

- Jaká je průměrná mzda ostatních 70% pracovníků?
- Jaká je nyní průměrná mzda v celé akciové společnosti?
- O kolik procent vzrostla průměrná mzda v celé akciové společnosti?

Návod: příklad 5 z Kontrolních otázek a úkolů ve 3. kapitole

Řešení:

ad a) Označme m_2 průměrnou mzdou ostatních 70% pracovníků. Musí platit:

$$0,3 \cdot 14\,000 + 0,7 \cdot m_2 = 18\,900. \text{ Odtud } m_2 = 21\,000 \text{ Kč.}$$

ad b) Průměrná mzda v celé akciové společnosti je nyní

$$0,3 \cdot 15\,500 + 0,7 \cdot 21\,000 = 19\,350 \text{ Kč}$$

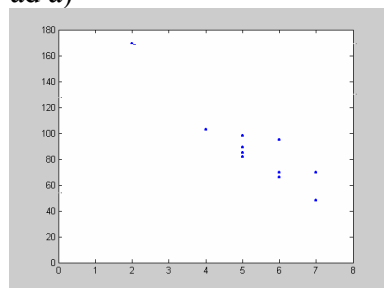
ad c) $\frac{19350}{18900} = 1,0238$, tedy průměrná mzda v celé akciové společnosti vzrostla o 2,38%.

Příklad 5.: U 11 náhodně vybraných aut jisté značky bylo zjišťováno jejich stáří (znak X – v letech) a cena (znak Y – v tisících Kč). Výsledky: (5, 85), (4, 103), (6, 70), (5, 82), (5, 89), (5, 98), (6, 66), (6, 95), (2, 169), (7, 70), (7, 48). Pro úsporu času máte uvedeny číselné charakteristiky (zaokrouhlené na dvě desetinná místa): $m_1 = 5,28$, $m_2 = 88,63$, $s_1^2 = 2,02$, $s_2^2 = 970,85$, $s_{12} = -40,89$.

- Nakreslete dvourozměrný tečkový diagram a s jeho pomocí posuďte, zda závislost Y na X lze uspokojivě popsat regresní přímkou. Návod: viz př. 2.3.(e), př. 4.4.(b)
- Vypočítejte koeficient korelace a interpretujte ho. Návod: viz poznámka 3.17.
- Najděte rovnici regresní přímky znaku Y na znak X. (Návod: viz věta 4.3.)
- Jaký je regresní odhad ceny auta, které je staré 3 roky? Návod: viz př. 4.4.(d)

Řešení:

ad a)



ad b) $r_{12} = -0,92$. Mezi znaky X a Y existuje silná nepřímá lineární závislost. Čím starší auto, tím nižší cena.

ad c) $y = 195,31 - 20,24x$

ad d) $y = 195,31 - 3 \cdot 20,24 = 134,59$

Příklad 6.: Je dána kontingenční tabulka obsahující hodnoty simultánní četnostní funkce $p(x,y)$ vektorového znaku (X, Y) :

x	y		
	0	1	2
0	0,20	0,20	0,00
1	0,05	0,25	0,03
2	0,05	0,01	0,05
3	0,05	0,01	0,10

- Doplňte tabulku o marginální četnostní funkce $p_1(x)$, $p_2(y)$.
- Vypočtěte průměry znaků X , Y .
- Vypočtěte rozptyly znaků X , Y .
- Vypočtěte a interpretujte koeficient korelace znaků X , Y .

Řešení:

ad a)

x	y			$p_1(x)$
	0	1	2	
0	0,20	0,20	0,00	0,40
1	0,05	0,25	0,03	0,33
2	0,05	0,01	0,05	0,11
3	0,05	0,01	0,10	0,16
$p_2(y)$	0,35	0,47	0,18	1

ad b)

$$m_1 = 1 \cdot 0,33 + 2 \cdot 0,11 + 3 \cdot 0,16 = 1,03, \quad m_2 = 1 \cdot 0,47 + 2 \cdot 0,18 = 0,83,$$

ad c)

$$s_1^2 = 1^2 \cdot 0,33 + 2^2 \cdot 0,11 + 3^2 \cdot 0,16 - 1,03^2 = 1,1491,$$

$$s_2^2 = 1^2 \cdot 0,47 + 2^2 \cdot 0,18 - 0,83^2 = 0,5011,$$

ad d)

$$s_{12} = 1 \cdot 1 \cdot 0,25 + 1 \cdot 2 \cdot 0,03 + 2 \cdot 1 \cdot 0,01 + 2 \cdot 2 \cdot 0,05 + 3 \cdot 1 \cdot 0,01 + 3 \cdot 2 \cdot 0,1 - 1,03 \cdot 0,83 = 0,3051,$$

$$r_{12} = \frac{0,3051}{\sqrt{1,1491} \sqrt{0,5011}} = 0,4021$$

Mezi znaky X a Y existuje slabá přímá lineární závislost.