

Téma 3: Výpočet číselných charakteristik jednorozměrného datového souboru

Vzorový příklad: Pro následující datové soubory vypočtěte číselné charakteristiky.

Postup ve STATISTICE:

1. Načtěte soubor **znamky.sta**. Pro známky z matematiky a angličtiny vypočtěte medián, dolní a horní kvartil a kvartilovou odchylku. Výsledky porovnejte s údaji ve skriptech Popisná statistika (viz str. 28).

Návod: Statistics – Basic Statistics/Tables – Descriptive Statistics – OK - Variables X, Y, OK – zaškrtneme Median, Lower & upper quartiles, Quartile range – Summary.

Řešení:

Variable	Descriptive Statistics (znamky)			
	Median	Lower Quartile	Upper Quartile	Quartile Range
X	2,500000	1,000000	4,000000	3,000000
Y	3,000000	2,000000	3,500000	1,500000

Komentář: Vidíme např., že medián známek z angličtiny je 3. Znamená to, že aspoň polovina studentů má známku 3 a lepší. Podobně dolní kvartil je 2, tudíž aspoň čtvrtina studentů má známku 2 a lepší.

2. Načtěte soubor **ocel.sta**. Pro mez plasticity a mez pevnosti vypočtěte aritmetické průměry, směrodatné odchylky a rozptyly. Výsledky porovnejte s údaji ve skriptech Popisná statistika (viz str. 30).

Návod: Statistics – Basic Statistics/Tables – Descriptive Statistics – OK - Variables X, Y, OK – zaškrtneme Mean, Standard Deviation, Variance – Summary.

Vysvětlení: Rozptyl a směrodatná odchylka vyjdou ve STATISTICE jinak než ve skriptech, protože STATISTICA ve vzorci pro výpočet rozptylu nepoužívá $1/n$, ale $1/(n-1)$.

Řešení:

Variable	Descriptive Statistics (ocel)		
	Mean	Variance	Std.Dev.
X	95,8833	1070,240	32,71453
Y	114,4000	1075,125	32,78911

Komentář: Průměrná hodnota meze plasticity je o něco nižší než průměrná hodnota meze pevnosti (95,88 oproti 114,4), avšak variabilita vyjádřená směrodatnou odchylkou se liší jen nepatrně (32,71 oproti 32,79).

3. Je třeba si uvědomit, že průměr a rozptyl nepopisují rozložení četností jednoznačně. Existují datové soubory, které mají shodný průměr i rozptyl, ale přesto se jejich rozložení četností velmi liší. Tuto skutečnost dobře ilustruje následující příklad: Tři skupiny studentů o počtech 149, 69 a 11 odpovídaly při testu na 10 otázek. Znak X je počet správně zodpovězených otázek. Známe absolutní četnosti znaku X ve všech třech skupinách.

č. sk.	X										
	0	1	2	3	4	5	6	7	8	9	10
1	2	5	15	20	25	15	25	20	15	5	2
2	4	3	2	1	0	49	0	1	2	3	4
3	1	0	0	0	0	9	0	0	0	0	1

Vypočtěte průměr (mean), rozptyl (variance), šikmost (skewness) a špičatost (kurtosis) počtu správně zodpovězených otázek ve všech třech skupinách. Nakreslete sloupkové diagramy absolutních četností.

Návod: Při zadávání dat do STATISTIKY utvořte čtyři proměnné a 11 případů. V 1. sloupci budou varianty znaku X (tj. 0 až 10), v dalších sloupcích pak absolutní četnosti. Proměnné pojmenujeme X, SK1, SK2, SK3. V tabulce Descriptive Statistics zadáme Variable X a klepneme na tlačítko W, abychom program upozornili, že budeme pracovat s daty zadanými pomocí absolutních četností. Zadáme Weight variable SK1, zaškrtneme Status On, OK – zaškrtneme Mean, Variance, Skewness, Kurtosis – Summary. Dále pro znak X nakreslíme sloupkový diagram – viz úkol č. 4 v tématu „Bodové rozložení četností“. Tytéž úkoly provedeme s Weight variable SK2 a SK3.

Řešení:

1. skupina (X weightet by SK1)

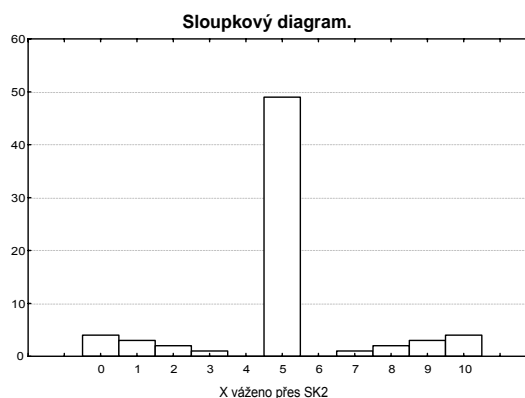
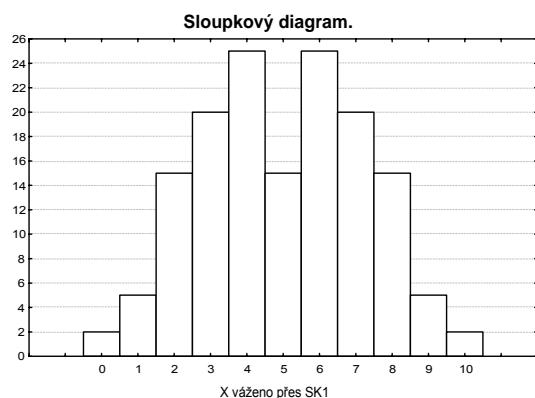
Variable	Descriptive Statistics			
	Mean	Variance	Skewness	Kurtosis
X	5,000000	5,000000	-0,000000	-0,759500

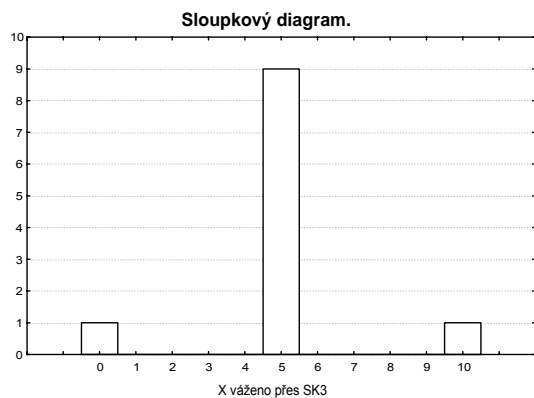
2. skupina (X weightet by SK2)

Variable	Descriptive Statistics			
	Mean	Variance	Skewness	Kurtosis
X	5,000000	5,000000	-0,000000	1,291133

3. skupina (X weightet by SK3)

Variable	Descriptive Statistics (cischar)			
	Mean	Variance	Skewness	Kurtosis
X	5,000000	5,000000	-0,000000	5,000000





Komentář: Všechny tři skupiny mají týž průměr, rozptyl a šikmost, liší se pouze ve špičatosti. Sloupkové diagramy počtu správně zodpovězených otázek v každé ze tří uvažovaných skupin mají naprosto odlišný vzhled.