

Cvičení 3: Výpočet číselných charakteristik intervalových a poměrových znaků

Úkol 1.: Otevřeme datový soubor ocel.sta, který obsahuje údaje o mezi plasticity (znak X, v kp cm^{-2}) a mezi pevnosti (znak Y, v kp cm^{-2}) 60 vzorků oceli.

a) Pro mez plasticity a mez pevnosti vypočteme aritmetický průměr, směrodatnou odchylku, rozptyl, koeficient variace, šikmost a špičatost.

b) Vypočteme kovarianci a Pearsonův koeficient korelace meze plasticity a meze pevnosti.

Návod:

ad a) Statistika – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnné X, Y – OK – Detailní výsledky - zaškrtneme Průměr, Směrodat. odchylka, Rozptyl, Variační koeficient, Šikmost, Špičatost – Výsledky.

Popisné statistiky (ocel)						
Proměnná	Průměr	Rozptyl	Sm.odch.	Koef.prom.	Šikmost	Špičatost
X	95,8833	1070,240	32,71453	34,11910	-0,046758	-0,605826
Y	114,4000	1075,125	32,78911	28,66181	0,297889	-0,592621

Upozornění: Systém STATISTICA počítá rozptyl podle vzorce $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2$, proto

výsledek musíme vynásobit $\frac{n-1}{n}$.

Ve výstupní tabulce přidáme za proměnnou Rozptyl tři nové proměnné nazvané rozptyl, směr. odch. a koef. variace. Do Dlouhého jména proměnné rozptyl napíšeme $=v2*59/60$, do Dlouhého jména proměnné směr. odch. napíšeme $=\text{sqrt}(v3)$ a do Dlouhého jména proměnné koef. variace napíšeme $=100*v4/v1$.

Proměnná	Průměr	Rozptyl	rozptyl $=v2*59/60$	směr. odch. $=\text{sqrt}(v3)$	koef. variace $=100*v4/v1$	Sm.odch.	Koef.prom.	Šikmost	Špičatost
X	95,8833	1070,240	1052,40306	32,4407623	33,8335779	32,71453	34,11910	-0,046758	-0,605826
Y	114,4000	1075,125	1057,20667	32,5147146	28,4219533	32,78911	28,66181	0,297889	-0,592621

ad b) Statistika – Základní statistiky/tabulky – Korelační matice – OK – 1 seznam proměnných – X, Y – OK, na záložce Možnosti zrušíme volbu Včetně průměrů a sm. odch. – Výpočet.

Korelace (ocel)		
Označ. korelace jsou významné na hlad. $p < ,05000$		
N=60 (Celé případy vynechány u ChD)		
Proměnná	X	Y
X	1,00	0,93
Y	0,93	1,00

Vidíme, že mezi X a Y existuje silná přímá lineární závislost.

Kovariance se počítá složitěji. Statistika – Vícenásobná regrese - Proměnné Nezávislá X, Závislá Y – OK – OK – Residua/předpoklady/předpovědi – Popisné statistiky – Další statistiky - Kovariance.

Proměnná	Kovariance (ocel)	
	X	Y
X	1070,240	1002,471
Y	1002,471	1075,125

Vysvětlení: Na hlavní diagonále jsou rozptyly proměnných X, Y, mimo hlavní diagonálu je kovariance. STATISTICA však ve vzorci pro výpočet kovariance nepoužívá $1/n$, ale $1/(n-1)$. Získanou kovarianci přepočítáme: k výstupní tabulce přidáme novou proměnnou, kterou vložíme za proměnnou v2. Do jejího Dlouhého jména napíšeme $=v2*59/60$. Dostaneme tabulku:

Proměnná	X	Y	NProm
X	1070,240	1002,471	985,7633
Y	1002,471	1075,125	1057,207

Úkol 2.: 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

Vypočtete průměr, rozptyl, šikmost a špičatost počtu správně zodpovězených otázek ve všech třech skupinách. Nakreslete sloupkové diagramy absolutních četností.

Návod: Načtete datový soubor body_ve_3_sk.sta. V 1. sloupci jsou varianty znaku X (tj. 0 až 10), v dalších sloupcích pak absolutní četnosti.

V tabulce Popisné statistiky zadáme Proměnná X a klepneme na tlačítko W, abychom program upozornili, že budeme pracovat s daty zadanými pomocí absolutních četností. Zadáme Proměnná vah SK1, zaškrtneme Stav Zapnuto, OK

The screenshot shows the STATISTICA software interface. The main window displays a data table with columns for 'X' (0-10) and three groups (SK1, SK2, SK3). Two dialog boxes are open over the table:

- Základní statistiky a tabulky:** This dialog has 'Popisné statistiky' selected. Under 'Možnosti', the 'W' option (Absolutní četnosti) is checked. The 'Proměnná vah' field is set to 'SK1'.
- Váhy případů pro analýzu/graf:** This dialog has 'Použít váhy z tabulky' selected. The 'Proměnná vah' field is also set to 'SK1'. The 'Stav' section has 'Zapnuto' selected.

Ve volbě Popisné statistiky zaškrtneme Průměr, Rozptyl, Šikmost, Špičatost – Výpočet. Dále pro znak X nakreslíme sloupkový diagram. Tytéž úkoly provedeme s váhovými proměnnými SK2 a SK3.

1. skupina (X váženo pomocí SK1)

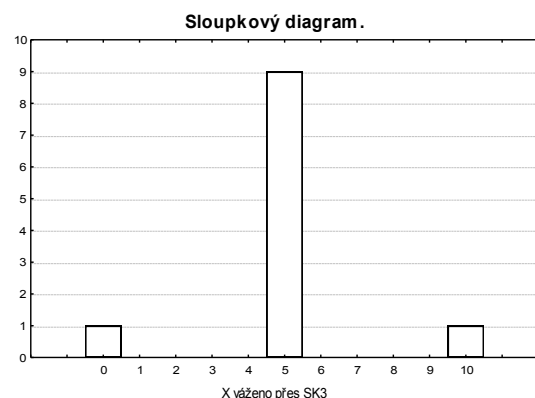
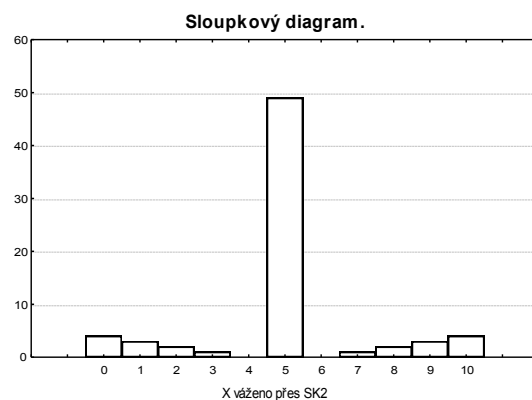
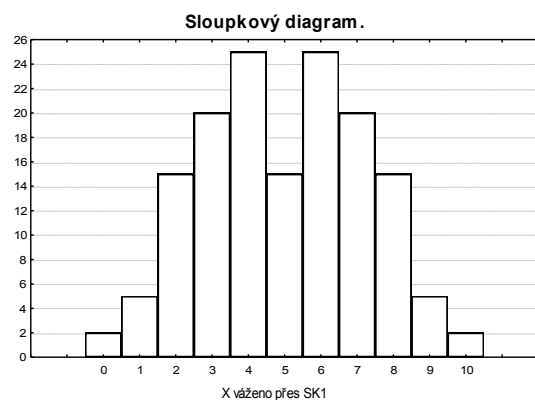
Proměnná	Průměr	Rozptyl	Šikmost	Špičatost
X	5,000000	5,000000	-0,000000	-0,759500

2. skupina (X váženo pomocí SK2)

Proměnná	Průměr	Rozptyl	Šikmost	Špičatost
X	5,000000	5,000000	-0,000000	1,291133

3. skupina (X váženo pomocí SK3)

Proměnná	Průměr	Rozptyl	Šikmost	Špičatost
X	5,000000	5,000000	0,00	5,000000



Všechny tři skupiny mají též 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.

Samostatná práce

Úkol 3.: U 27 dětí ve věku 9,5 – 10 let byla zjišťována tělesná výška (v cm). Máme k dispozici výsledky měření rozdělené do šesti třídicích intervalů a podle pohlaví dítěte:

Střed třídicího intervalu	Počet hochů	Počet dívek
125	1	0
130	1	2
135	3	1
140	7	5
145	1	3
150	2	1

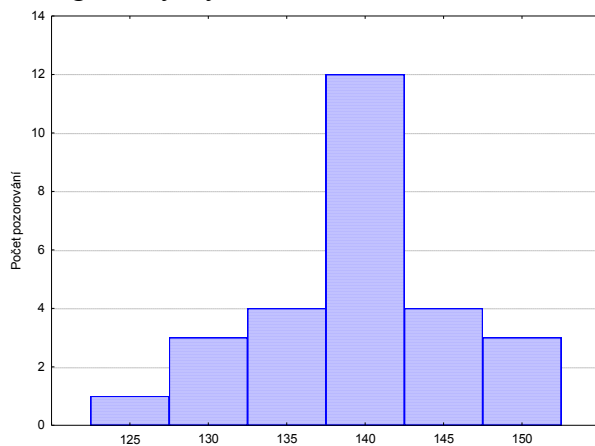
a) Vypočítejte průměr a směrodatnou odchylku výšky pro všechny děti a pak zvlášť pro hochy a zvlášť pro dívky.

b) Vytvořte histogram výšky pro všechny děti a pak zvlášť pro hochy a zvlášť pro dívky. (Data jsou uložena v souboru vysky_deti.sta)

Výsledky:

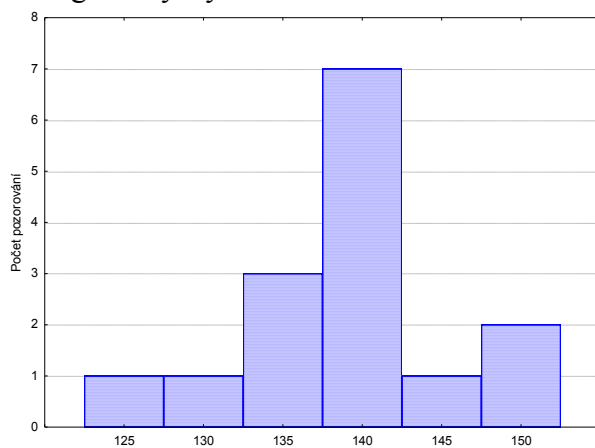
Průměrná výška všech dětí je 139,4 cm, směrodatná odchylka výšky je 6,1 cm.

Histogram výšky všech dětí:

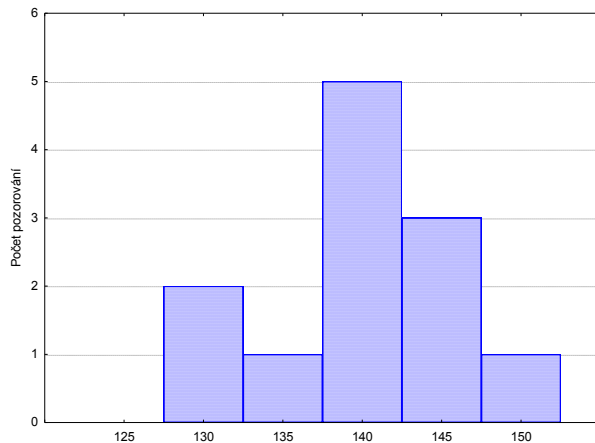


Průměrná výška hochů je 139 cm, směrodatná odchylka výšky je 6,5 cm.

Histogram výšky hochů:



Průměrná výška dívek je 140 cm, směrodatná odchylka výšky je 5,9 cm.
Histogram výšky dívek:



Úkol 4.: 5317 manželských párů bylo dotázáno na věk manžela a věk manželky. Zjištěné údaje o věku manžela (znak X) byly roztrženy do 7 třídících intervalů o délce 10 roků se středy 20, 30, ..., 80, stejně tak údaje o věku manželky (znak Y). Máme k dispozici kontingenční tabulku simultánních absolutních četností. Vypočtete koeficient korelace znaků X, Y.

(Data jsou uložena v souboru vek_manzela.sta)

X _[j]	Y _[k]						
	20	30	40	50	60	70	80
20	193	50	1	0	0	0	0
30	231	1162	108	4	0	0	0
40	12	408	977	92	4	0	0
50	1	36	320	652	66	3	0
60	0	5	37	211	358	34	1
70	0	1	6	24	105	133	10
80	0	0	1	4	10	32	25

Výsledek: $r_{12} = 0,893$