Cvičení 2.: Intervalové rozložení četností, výpočet číselných charakteristik nominálních a ordinálních znaků

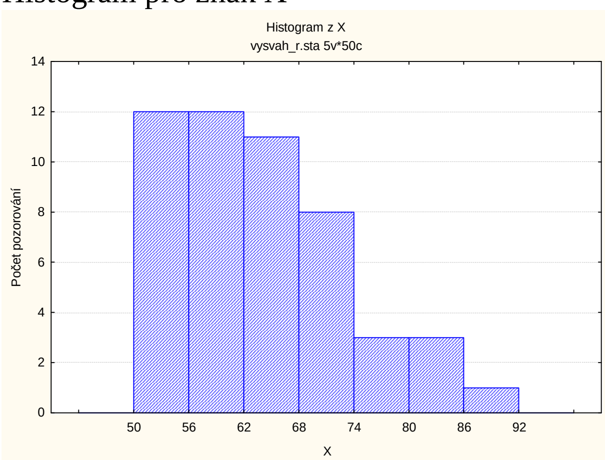
Úkol 1.: Datový soubor vysvah.sta obsahuje údaje o hmotnosti (znak X, v kg), výšce (znak Y, v cm) a pohlaví (znak Z, 0 – žena, 1 – muž) 50 náhodně vybraných studentů. Načtěte tento soubor do systému STATISTICA. Proměnným X, Y, Z vytvořte návěští „hmotnost“, „výška“ a „pohlaví“. Popište, co u znaku Z znamenají varianty 0, 1. Podle Sturgesova pravidla najděte optimální počet třídicích intervalů pro znaky X a Y a vhodně stanovíte meze třídicích intervalů.
Návod: Soubor – Otevřít – vybereme příslušný adresář se souborem vysvah.sta – Otevřít. Kurzor nastavíme na X – 2x klikneme myší – Dlouhé jméno hmotnost – OK, kurzor nastavíme na Y – 2x klikneme myší – Dlouhé jméno výška – OK, kurzor nastavíme na Z - 2x klikneme myší – Dlouhé jméno pohlaví, Text. hodnoty – 0 žena, 1 – muž - OK.
Protože případů je 50, podle Sturgersova pravidla je optimální počet třídicích intervalů 7.
Musíme zjistit minimum a maximum, abychom vhodně stanovili třídicí intervaly: Statistiky - Základní statistiky/tabulky – Popisné statistiky - OK - Proměnné X,Y – OK – Detailní výsledky – ponecháme zaškrtnuté pouze Minimum&maximum – Výpočet.

	Popisné statistiky (vysvah.sta)		
Proměnná	N platných	Minimum	Maximum
X	50	51,0000	90,0000
Y	50	160,0000	192,0000

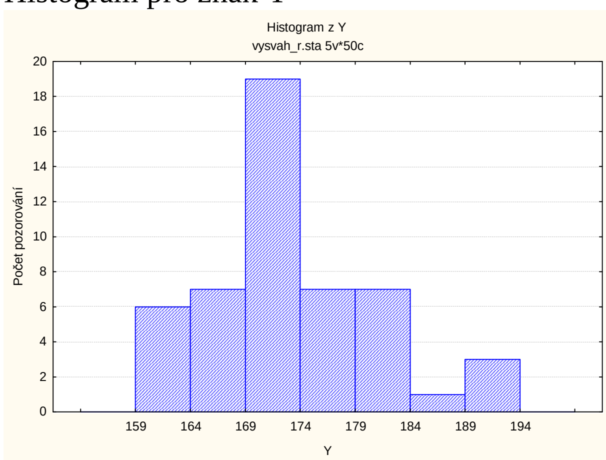
Pro X je minimum 51 a maximum 90, tedy dolní mez prvního třídicího intervalu volíme 50, horní mez posledního třídicího intervalu 92. Celkem tedy třídicí intervaly pro znak X budou: (50,56>, (56,62>, (62,68>, (68,74>, (74,80>, (80,86>, (86,92>.
Pro Y je minimum 160 a maximum 192, tedy dolní mez prvního třídicího intervalu volíme 159, horní mez posledního třídicího intervalu 194. Celkem tedy třídicí intervaly pro znak Y budou:
(159,164>, (164,169>, (169,174>, (174,179>, (179,184>, (184,189>, (189,194>.

Úkol 2.: Vytvořte histogram pro X a pro Y.
Návod: Grafy – Histogramy – Proměnné X – vypneme Normální proložení – Detaily – zaškrtneme Hranice – Určit hranice – zvolíme Zadejte hraniční rozmezí – Minimum = 50, Krok = 6, Maximum = 92 - OK – OK. Po vykreslení histogramu lze 2 x klepnout na pozadí grafu a ve volbě Všechny možnosti měnit různé vlastnosti grafu.
Analogicky pro Y.

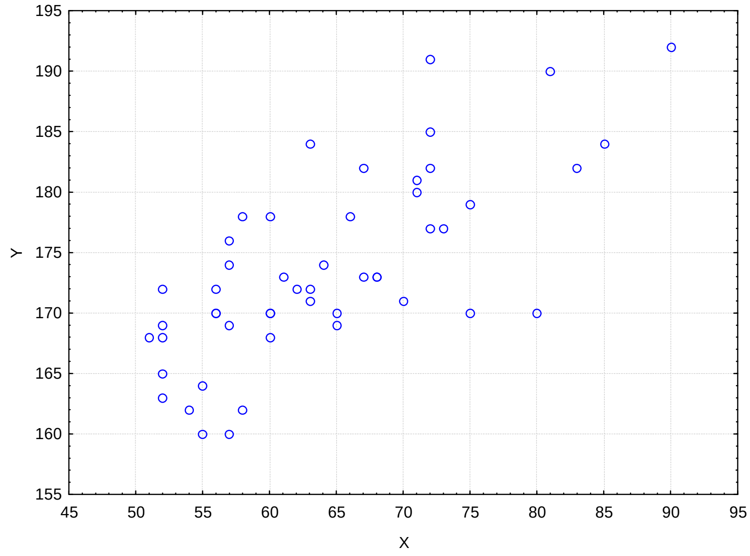
Histogram pro znak X



Histogram pro znak Y



Úkol 3.: Nakreslete dvourozměrný tečkový diagram pro (X,Y).
Návod: Grafy – Bodové grafy – Proměnné X,Y – OK - vypneme Lineární proložení – OK.



Vidíme, že mezi oběma proměnnými existuje určitý stupeň přímé lineární závislosti – s růstem hmotnosti vesměs rostou hodnoty výšky a naopak.

Samostatná práce: úkoly 1 až 3 proveďte zvlášť pro muže a zvlášť pro ženy.

Úkol 4.: U 100 náhodně vybraných domácností byl zjišťován způsob zásobování bramborami (znak X, varianty 1 = vlastní sklep, 2 = jinde, 3 = nákup) a bydliště (znak Y, varianty 1 = velké město, 2 = malé město, 3 = vesnice).

způsob zásobování	bydliště		
	velké město	malé město	vesnice
vlastní sklep	13	15	14
jinde	11	7	2
nákup	19	9	10

a) Pro oba znaky určíme modus.
b) Vypočteme Cramérův koeficient znaků X, Y.
Návod: Otevřeme nový datový soubor se třemi proměnnými X, Y, četnost a devíti případy. Do proměnné X napíšeme 3 jedničky, 3 dvojky a 3 trojky, do proměnné Y napíšeme 3 krát pod sebe 1, 2, 3 a do proměnné četnost napíšeme odpovídající simultánní absolutní četnosti dvojic variant (X, Y), tj. 13, 15, 14, 11, 7, 2, 19, 9, 10. Proměnným vytvoříme návěští a popíšeme význam jednotlivých variant.

ad a) Výpočet modu: Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – klikneme na tlačítko se závažím – zaškrtneme Stav zapnuto, vybereme proměnnou vah četnost – OK - Proměnné X, Y – OK – Detailní výsledky – zaškrtneme Modus.

Proměnná	Popisné statistiky (brambory)	
	Modus	Četnost modu
X	1,000000	42
Y	1,000000	43

Proměnná X má modus 1, tj. nejvíce domácností skladuje brambory ve vlastním sklepě a proměnná Y má také modus 1, tj. nejvíce domácností bydlí ve velkém městě.

ad b) Výpočet Cramérova koeficientu: Statistiky – Základní statistiky/tabulky – Kontingenční tabulky – OK – Specif. tabulky - List 1 X, List 2 Y - OK – na záložce Možnosti ve Statistikách 2 rozměrných tabulek zaškrtneme Fí (tabulky 2x2) & Cramérovo V & C – přejdeme na záložku Detailní výsledky – Detailní 2-rozm. tabulky.

Statist.	Statist. : X(3) x Y(3) (brambory)		
	Chí-kvadr.	sv	p
Pearsonův chí-kv.	6,420286	df=4	p=,16989
M-V chí-kvadr.	7,075760	df=4	p=,13195
Fí	,2533828		
Kontingenční koeficient	,2456207		
Cramér. V	,1791687		

Na posledním řádku najdeme, že Cramérův koeficient nabývá hodnoty 0,179, tedy mezi způsobem zásobování bramborami a bydlištěm domácnosti existuje jen slabá závislost – viz následující tabulka:

Cramérův koeficient	interpretace
mezi 0 až 0,1	zanedbatelná závislost
mezi 0,1 až 0,3	slabá závislost
mezi 0,3 až 0,7	střední závislost
mezi 0,7 až 1	silná závislost

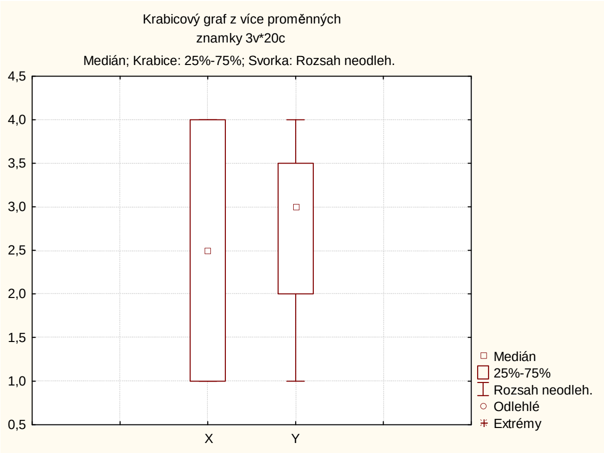
Úkol 5.: Datový soubor znamky.sta obsahuje údaje o 20 studentech 1. ročníku ekonomicky zaměřené vysoké školy. Znak X – známka z matematiky v 1. zkušebním termínu (má varianty 1, 2, 3, 4), znak Y – známka z angličtiny v 1. zkušebním termínu (má rovněž varianty 1, 2, 3, 4), znak Z – pohlaví studenta (0 – žena, 1 – muž).
Otevřeme datový soubor znamky.sta.
a) Pro známky z matematiky a angličtiny vypočteme medián, dolní a horní kvartil, kvartilovou odchylku a vytvoříme krabicový diagram.
b) Vypočteme Spearmanův korelační koeficient známek z matematiky a angličtiny pro všechny studenty, pak zvlášť pro muže a zvlášť pro ženy. Získané výsledky budeme interpretovat.

Návod:

ad a) Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnné X, Y – OK – Detailní výsledky - zaškrtneme Medián, Dolní & horní kvartily, Kvartil. rozpětí – Výpočet.

Proměnná	Popisné statistiky (znamky)			
	Medián	Spodní kvartil	Horní kvartil	Kvartilové rozpětí
X	2,500000	1,000000	4,000000	3,000000
Y	3,000000	2,000000	3,500000	1,500000

Vytvoření krabicového diagramu: Grafy – 2D Grafy – Krabicové grafy – vybereme Vícenásobný – Proměnné X, Y – OK.



ad b) Statistiky – Neparametrická statistika – Korelace – OK – Proměnné X, Y – OK – Spearman R.

Pro všechny:

Spearmanovy korelace (znamky)
ChD vynechány párově
Označ. korelace jsou významné na hl. p <,05000

Proměnná	X	Y
X	1,000000	0,688442
Y	0,688442	1,000000

Počítáme-li Spearmanův korelační koeficient pro ženy (resp. pro muže), použijeme filtr: tlačítko Select Cases – Zapnout filtr – včetně případů – některé, vybrané pomocí výrazu Z=0 (resp. Z=1).

Pro ženy:

Spearmanovy korelace (znamky)
ChD vynechány párově
Označ. korelace jsou významné na hl. p <,05000
Zhrnout podmínku: Z=0

Proměnná	X	Y
X	1,000000	0,860314
Y	0,860314	1,000000

Pro muže:

Spearmanovy korelace (znamky)
ChD vynechány párově
Označ. korelace jsou významné na hl. p <,05000
Zhrnout podmínku: Z=1

Proměnná	X	Y
X	1,000000	0,373544
Y	0,373544	1,000000

Vidíme, že nejsilnější přímá pořadová závislost mezi známkami z matematiky a angličtiny je u žen, r_S = 0,86. U mužů je tato závislost mnohem slabší, r_S = 0,37. U žen tedy dochází k tomu, že se sdružují podobné známky z obou předmětů, zatímco u mužů se projevuje spíše tendence k různým známkám.