Téma 2: Výpočet číselných charakteristik jednorozměrného a dvourozměrného datového souboru

Úkol 1.: 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 tlačíto se závažím – zaškrtneme Stav zapnuto, vybereme proměnnou vah četnost – OK - Proměnné X, Y – OK – Detailní výsledky – zaškrtneme Modus.

Promě	Popisné statistiky Modu	 Četn modi
X	1,000	4:
Y	1,000	4:

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) Chi-kva	 sv	 p
Pearsonův chi-k	6,420	df=	p=,16
M-V chi-kvadr.	7,075	df=	p=,13
Fí	,2533		
Kontingenční ko	,2456		
Cramér. V	,1791		

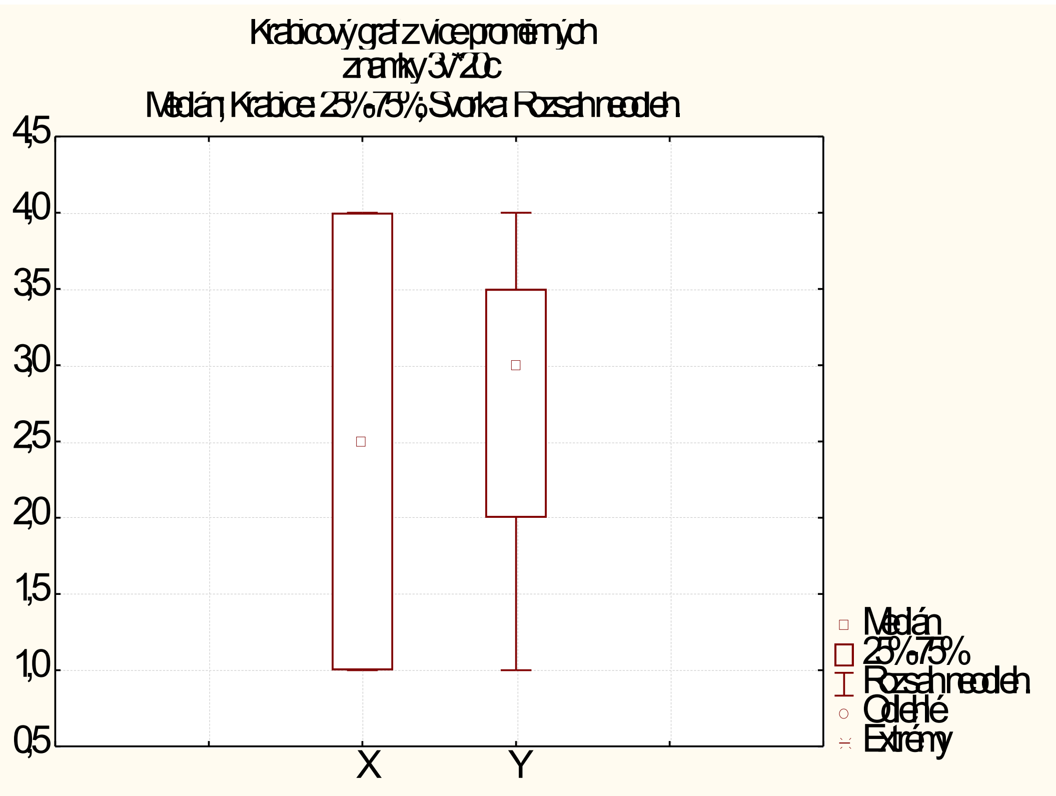
Na posledním řádku najdeme, že Cramérův koeficient nabývá hodnoty 0,179, tedy mezi způsobem zásobování bramborami a bydlištěm domácnosti exisuje jen slabá závislost – viz následující tabulka:

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 2.: 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.

	Popisné statistiky (znamky			
Proměi	Médiá	Spod kvarti	Horn kvarti	Kvartilo rozpět
X	2,500	1,000	4,000	3,000
Y	3,000	2,000	3,500	1,500

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 (znaml ChD vynechány párově Označ. korelace jsou významr	
Proměi	X	Y
X	1,000	0,688
Y	0,688	1,000

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:

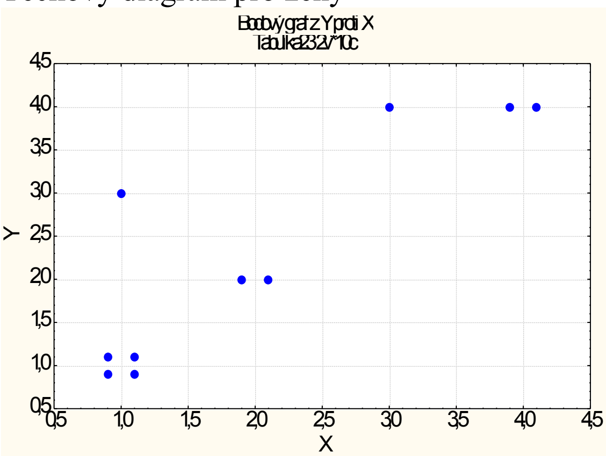
	Spearmanovy korelace (znaml ChD vynechány párově Označ. korelace jsou významr Zhrnout podmínku: Z=0	
Proměi	X	Y
X	1,000	0,860
Y	0,860	1,000

Pro muže:

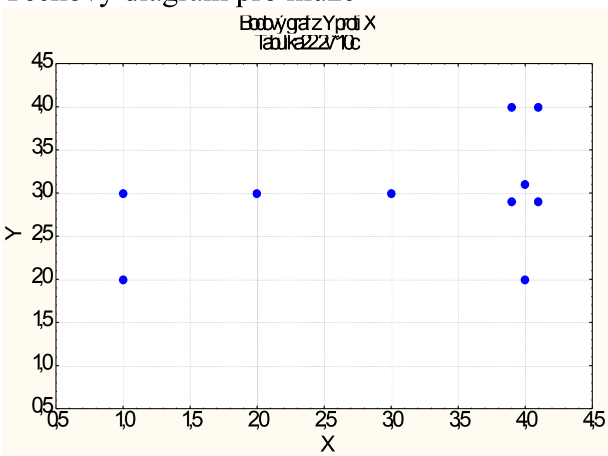
	Spearmanovy korelace (znaml ChD vynechány párově Označ. korelace jsou významr Zhrnout podmínku: Z=1	
Proměi	X	Y
X	1,000	0,373
Y	0,373	1,000

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. Je to zřetelně vidět na dvourozměrných tečkových diagramech.

Tečkový diagram pro ženy



Tečkový diagram pro muže



Význam hodnot Spearmanova (i Pearsonova) koeficientu korelace je popsán v tabulce:

Absolutní hodnota korelačního koeficientu	Interpretace hodnoty
0	lineární nezávislost
(0, 0,1)	velmi nízký stupeň závislosti
[0,1, 0,3)	nízký stupeň závislosti
[0,30, 0,50)	mírný stupeň závislosti
[0,50, 0,70)	význačný stupeň závislosti
[0,70, 0,90)	vysoký stupeň závislosti
[0,90, 1)	velmi vysoký stupeň závislosti
1	úplná lineární závislost

Úkol 3.: Otevřeme datový soubor ocel.sta.
a) Pro mez plasticity a mez pevnosti vypočteme aritmetický průměr, směrodatnou odchylku, rozptyl, koeficient variace, šikmost a špičatost. Výsledky porovnáme s údaji ve skriptech Popisná statistika (viz str. 30).
b) Vypočteme Pearsonův koeficient korelace meze plasticity a meze pevnosti. Dále vypočteme také kovarianci a výsledek porovnáme s výsledkem ve skriptech Popisná statistika (str. 30).

Návod:
ad a) Statistiky – 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)					
Promě	Prům	Rozpt	Sm.od	Koef.pr	Šikmo	Špičat
X	95,8	1070,	32,71	34,11	-0,046	-0,605
Y	114,4	1075,	32,78	28,66	0,297	-0,592

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). Koeficient variace (v tabulce označený jako Koef. Prom.) je udán v procentech.

ad b) Statistiky – 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 N=60 (Celé případy vynechány	
Promě	X	Y
X	1,(	0,9
Y	0,9	1,(

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

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

Proměn	Kovariance (	
	X	Y
X	1070,	1002,
Y	1002,	1075,

Vysvětlení: Na hlavní diagonále jsou rozptyly proměnných X, Y, mimo hlavní diagonálu je kovariance. Kovariance vyjde ve STATISTICE jinak než ve skriptech, protože ve STATISTICE se ve vzorci pro výpočet kovariance nepoužívá 1/n, ale 1/(n-1).

Úkol 4.: 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.

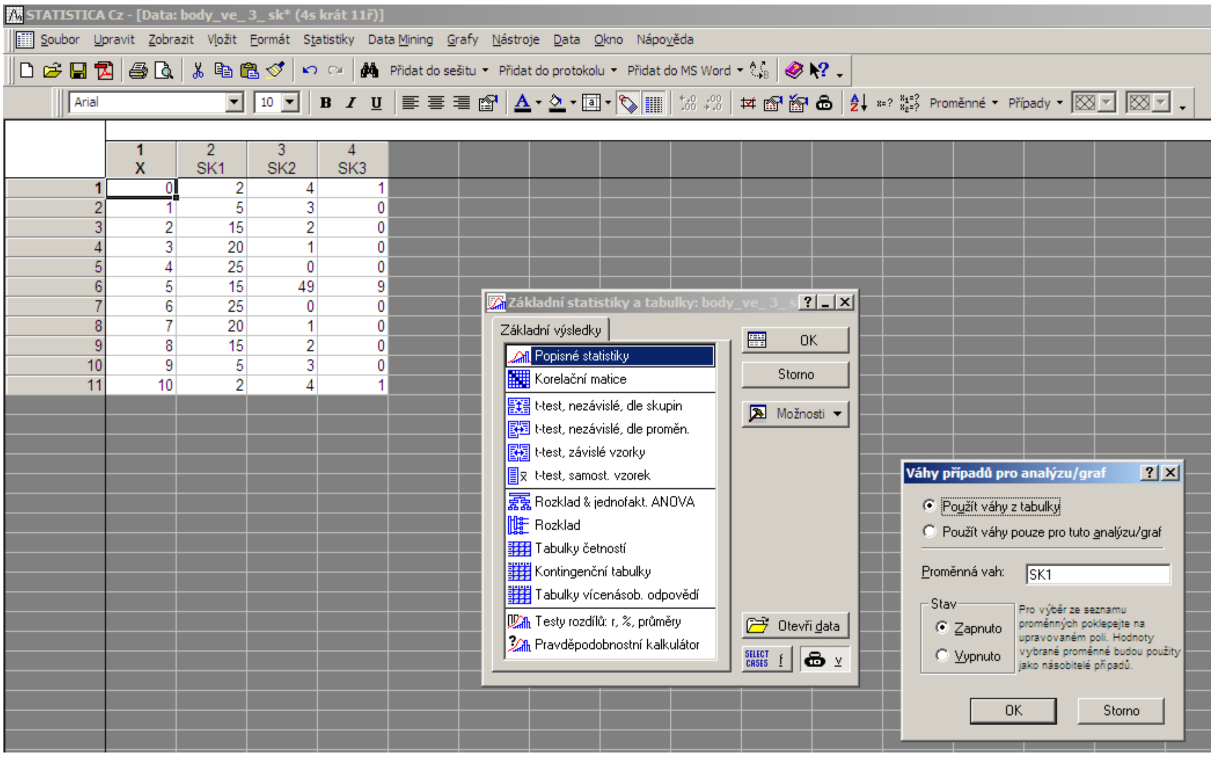
č. sk.	X										
	0	1	2	3	4	5	6	7	8	9	10
1	2	5	15	20	25	15	25	20	15	5	2
2	4	3	2	1	0	49	0	1	2	3	4
3	1	0	0	0	0	9	0	0	0	0	1

Vypočtěte průměr, rozptyl, šikmost a špičatost 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.

	1 X	2 SK1	3 SK2	4 SK3
1	(	2	4	1
2	1	5	3	(
3	2	1	2	(
4	3	2(	1	(
5	4	2	(	(
6	5	1	4	9
7	6	2	(	(
8	7	2(	1	(
9	8	1	2	(
1(	9	5	3	(
1	1(	2	4	1

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



	1 X	2 SK1	3 SK2	4 SK3
1	0	2	4	1
2	1	5	3	0
3	2	15	2	0
4	3	20	1	0
5	4	25	0	0
6	5	15	49	9
7	6	25	0	0
8	7	20	1	0
9	8	15	2	0
10	9	5	3	0
11	10	2	4	1

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)

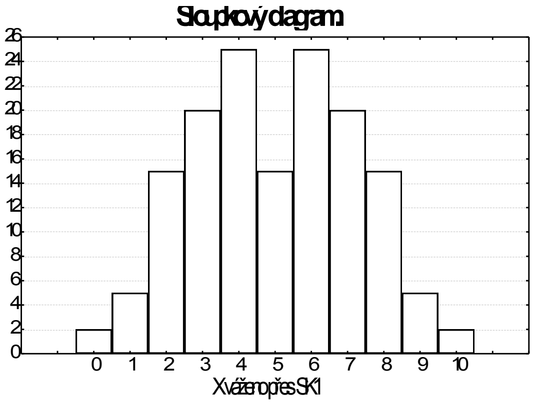
Varial	Descriptive Statistics			
	Meai	Variar	Skewn	Kurtos
X	5,000	5,000	-0,00(	-0,759

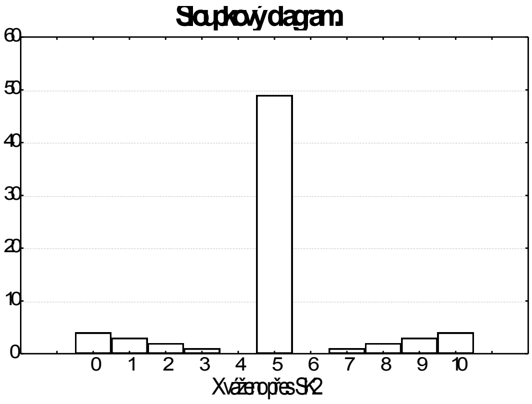
2. skupina (X váženo pomocí SK2)

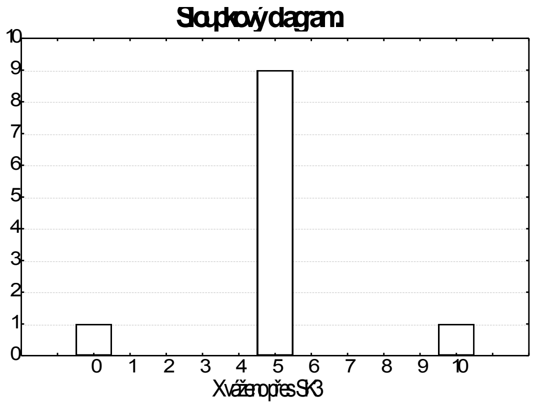
Varial	Descriptive Statistics			
	Meai	Variar	Skewn	Kurto:
X	5,000	5,000	-0,00(	1,291

3. skupina (X váženo pomocí SK3)

Varial	Descriptive Statistics (ciscr			
	Meai	Variar	Skewn	Kurto:
X	5,000	5,000	-0,00(	5,000







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.