Téma 2.: Jednorozměrné a dvourozměrné intervalové rozložení četností, číselné charakteristiky nominálních a ordinálních znaků

Příklad na intervalové zpracování četností: U 60 vzorků oceli byly zjišťovány hodnoty meze plasticity a meze pevnosti v kpcm^{-2} (viz skripta Popisná statistika, př. 2.5). Datový soubor se jmenuje **ocel.sta**. Proveďte intervalové zpracování četností.

Úkol 1.: Načtěte soubor **ocel.sta**. Proměnným X a Y vytvořte návěští „mez plasticity“ a „mez pevnosti“.

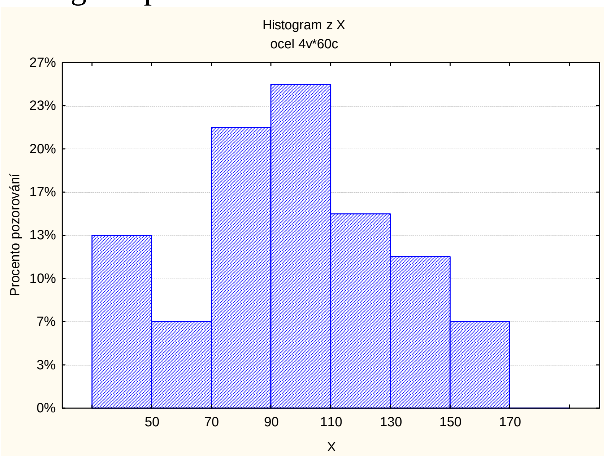
Úkol 2.: Pro X a Y použijeme intervalové zpracování četností. Podle Sturgesova pravidla ($r \approx 1 + 3{,}3 \log n$, r – počet třídicích intervalů, n – rozsah soubor) je optimální počet třídicích intervalů 7. Musíme zjistit minimum a maximum, abychom vhodně stanovili třídicí intervaly.
Návod: Statistiky - Základní statistiky/tabulky – Popisné statistiky - OK - Proměnné X,Y – OK – Detailní výsledky – ponecháme zaškrtnuté Minimum&maximum – Výpočet.

	Popisné statistiky (ocel.sta)	
Proměnná	Minimum	Maximum
X	33,00000	160,0000
Y	52,00000	189,0000

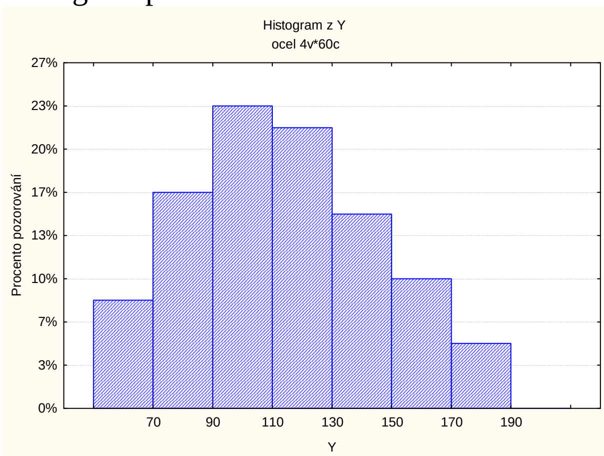
Pro X je minimum 33 a maximum 160, tedy vhodná volba třídicích intervalů je (30,50>, (50,70>, ..., (150,170>, pro Y je minimum 52 a maximum 189, tedy třídicí intervaly zvolíme (50,70>, (70,90>, ... , (170,190>.

Úkol 3.: 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: 30, Krok: 20, Maximum: 170 OK – Osa Y %. Po vykreslení histogramu lze 2 x klepnout na pozadí grafu a ve volbě Všechny možnosti měnit různé vlastnosti grafu.

Histogram pro znak X



Histogram pro znak Y



Komentář: Rozložení četností jak pro mez plasticity, tak pro mez pevnosti je lehce nesymetrické. Navíc v histogramu pro mez plasticity je vidět, že interval od 50 do 70 má velmi malé četnostní zastoupení. Vysvětlení této skutečnosti je ovšem mimomatematická záležitost.

Úkol 4.: Proveďte zakódování hodnot proměnných X a Y do příslušných třídicích intervalů. Všem hodnotám proměnné X, které leží v intervalu (30,50>, přiřadíme hodnotu středu intervalu, tedy 40 atd. až všem hodnotám proměnné X, které leží v intervalu (150,170>, přiřadíme hodnotu 160. Analogicky pro Y, tedy všem hodnotám proměnné Y, které leží v intervalu (50,70>, přiřadíme hodnotu středu intervalu, tj. 60 atd. až všem hodnotám proměnné Y, které leží v intervalu (170,190>, přiřadíme hodnotu 180. Podmínky pro překódování jsou uloženy v tzv. inicializačních souborech nazvaných ocel_X.ini a ocel_Y.ini.
Návod: Vytvoříme dvě nové proměnné: Vložit – Přidat proměnné – 2 – Za Y – OK – přejmenujeme je na RX a RY. Nastavíme se kurzorem na RX – Data – Překódovat - Otevřít – ocel_X.ini – OK. Proměnná RX se vyplní středy třídicích intervalů pro mez plasticity. Poté se nastavíme kurzorem na RY - Data – Překódovat - Otevřít – ocel_Y.ini – OK. Proměnná RY se vyplní středy třídicích intervalů pro mez pevnosti.

Úkol 5.: Sestavte kontingenční tabulky absolutních četností (relativních četností, sloupcově a řádkově podmíněných relativních četností) dvourozměrných třídicích intervalů pro (X,Y).
Návod: Statistiky – Základní statistiky/tabulky – OK - Kontingenční tabulky – OK – Specif. tabulky - List 1 RX, List 2 RY, OK, Výpočet.

Kontingenční tabulka absolutních a relativních četností:

	RX	RY 60	RY 80	RY 100	RY 120	RY 140	RY 160	RY 180	Řádk. součty
Četnost	40	5	3	0	0	0	0	0	8
Celková četn.		8,33%	5,00%	0,00%	0,00%	0,00%	0,00%	0,00%	13,33%
Četnost	60	0	3	1	0	0	0	0	4
Celková četn.		0,00%	5,00%	1,67%	0,00%	0,00%	0,00%	0,00%	6,67%
Četnost	80	0	4	7	1	1	0	0	13
Celková četn.		0,00%	6,67%	11,67%	1,67%	1,67%	0,00%	0,00%	21,67%
Četnost	100	0	0	6	8	1	0	0	15
Celková četn.		0,00%	0,00%	10,00%	13,33%	1,67%	0,00%	0,00%	25,00%
Četnost	120	0	0	0	4	5	0	0	9
Celková četn.		0,00%	0,00%	0,00%	6,67%	8,33%	0,00%	0,00%	15,00%
Četnost	140	0	0	0	0	2	5	0	7
Celková četn.		0,00%	0,00%	0,00%	0,00%	3,33%	8,33%	0,00%	11,67%
Četnost	160	0	0	0	0	0	1	3	4
Celková četn.		0,00%	0,00%	0,00%	0,00%	0,00%	1,67%	5,00%	6,67%
Četnost	Vš.skup.	5	10	14	13	9	6	3	60
Celková četn.		8,33%	16,67%	23,33%	21,67%	15,00%	10,00%	5,00%	

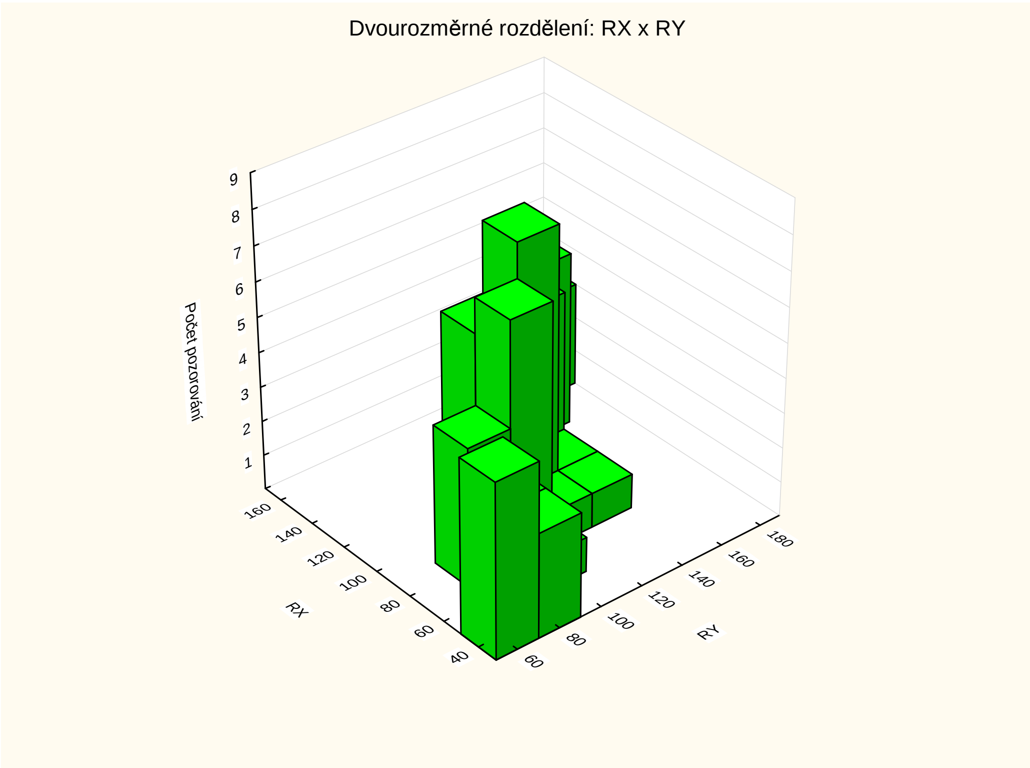
Kontingenční tabulka řádkově podmíněných relativních četností.

	RX	RY 60	RY 80	RY 100	RY 120	RY 140	RY 160	RY 180	Řádk. součty
Četnost	40	5	3	0	0	0	0	0	8
Řádk. četn.		62,50%	37,50%	0,00%	0,00%	0,00%	0,00%	0,00%	
Četnost	60	0	3	1	0	0	0	0	4
Řádk. četn.		0,00%	75,00%	25,00%	0,00%	0,00%	0,00%	0,00%	
Četnost	80	0	4	7	1	1	0	0	13
Řádk. četn.		0,00%	30,77%	53,85%	7,69%	7,69%	0,00%	0,00%	
Četnost	100	0	0	6	8	1	0	0	15
Řádk. četn.		0,00%	0,00%	40,00%	53,33%	6,67%	0,00%	0,00%	
Četnost	120	0	0	0	4	5	0	0	9
Řádk. četn.		0,00%	0,00%	0,00%	44,44%	55,56%	0,00%	0,00%	
Četnost	140	0	0	0	0	2	5	0	7
Řádk. četn.		0,00%	0,00%	0,00%	0,00%	28,57%	71,43%	0,00%	
Četnost	160	0	0	0	0	0	1	3	4
Řádk. četn.		0,00%	0,00%	0,00%	0,00%	0,00%	25,00%	75,00%	
Četnost	Vš.skup.	5	10	14	13	9	6	3	60

Kontingenční tabulka sloupcově podmíněných relativních četností:

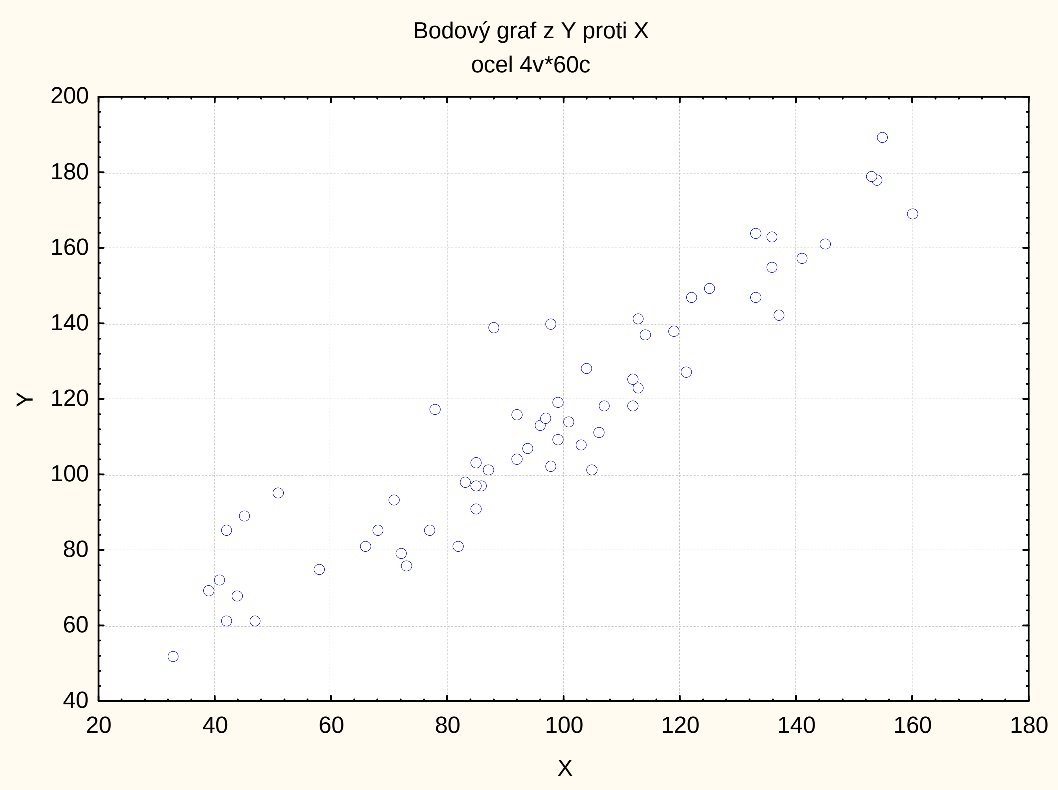
	RX	RY 60	RY 80	RY 100	RY 120	RY 140	RY 160	RY 180	Řádk. součty
Četnost	40	5	3	0	0	0	0	0	8
Sloupc. četn.		100,00%	30,00%	0,00%	0,00%	0,00%	0,00%	0,00%	
Četnost	60	0	3	1	0	0	0	0	4
Sloupc. četn.		0,00%	30,00%	7,14%	0,00%	0,00%	0,00%	0,00%	
Četnost	80	0	4	7	1	1	0	0	13
Sloupc. četn.		0,00%	40,00%	50,00%	7,69%	11,11%	0,00%	0,00%	
Četnost	100	0	0	6	8	1	0	0	15
Sloupc. četn.		0,00%	0,00%	42,86%	61,54%	11,11%	0,00%	0,00%	
Četnost	120	0	0	0	4	5	0	0	9
Sloupc. četn.		0,00%	0,00%	0,00%	30,77%	55,56%	0,00%	0,00%	
Četnost	140	0	0	0	0	2	5	0	7
Sloupc. četn.		0,00%	0,00%	0,00%	0,00%	22,22%	83,33%	0,00%	
Četnost	160	0	0	0	0	0	1	3	4
Sloupc. četn.		0,00%	0,00%	0,00%	0,00%	0,00%	16,67%	100,00%	
Četnost	Vš.skup.	5	10	14	13	9	6	3	60

Úkol 6.: Vytvořte stereogram pro (RX,RY).
Návod: Statistiky – Základní statistiky/tabulky – Kontingenční tabulky – OK – Specif. tabulky - List 1 RX, List 2 RY – OK – OK – Detailní výsledky – zaškrtneme 3D histogramy. Ve výsledném grafu 2x klikneme na pozadí, vybereme Graf – Vzhled – Mezery mezi sloupci – pro X zvolíme 0 a pro Y také zvolíme 0.



Upozornění: V našem pojetí je výška jxk-tého kvádru ve stereogramu rovna četnostní hustotě jxk-tého dvourozměrného třídicího intervalu, avšak systém STATISTICA vytváří stereogram tak, že výška jxk-tého kvádru je rovna absolutní četnosti jxk-tého dvourozměrného třídicích intervalu.

Úkol 7.: 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 hodnot meze plasticity vesměs rostou hodnoty meze pevnosti a naopak.

Úkol 8.: 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		

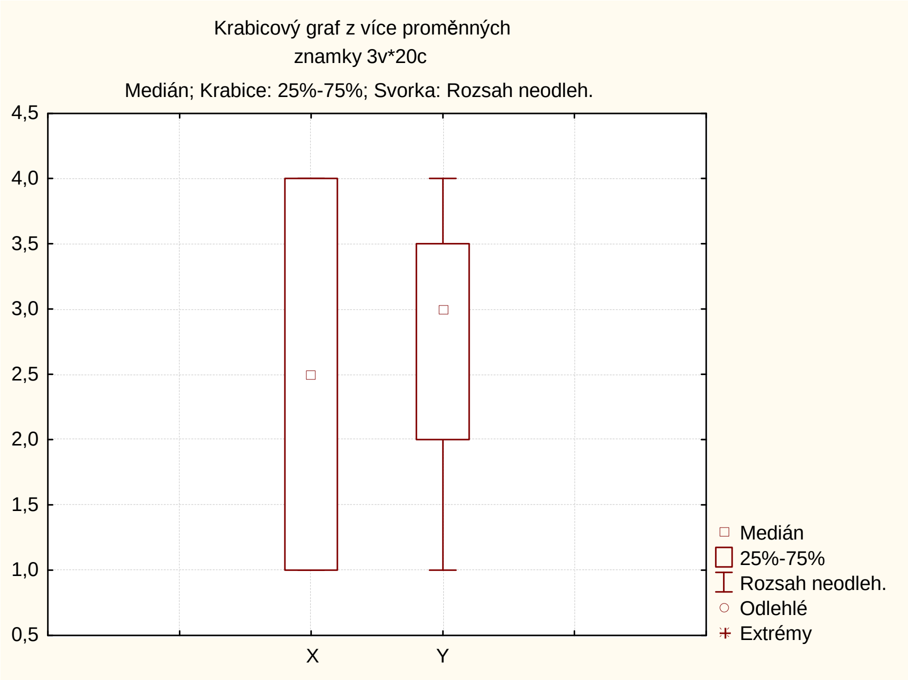
Na posledním řádku najdeme, že Cramérův koeficient nabývá hodnoty 0,179, tedy mezi způsobem zásobování bramborami a bydlištěm domácnosti existuje jen slabá závislost – viz následující tabulka:

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 9.: Otevřeme datový soubor znamky.sta. Pro známky z matematiky a angličtiny vypočteme medián, dolní a horní kvartil, kvartilovou odchylku a vytvoříme krabicový diagram.
Návod:
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.


Krabicový graf z více proměnných
znamky 3v*20c
Medián; Krabice: 25%-75%; Svorka: Rozsah neodleh.
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
X
Y
Medián
25%-75%
Rozsah neodleh.
Odlehlé
Extrémy