Téma 3: Číselné charakteristiky intervalových a poměrových znaků, odhad pravděpodobnosti pomocí relativní četnosti

Úkol 1.: Otevřete datový soubor ocel.sta.
a) Pro mez plasticity a mez pevnosti vypočtěte aritmetický průměr, směrodatnou odchylku, rozptyl, koeficient variace, šikmost a špičatost. Výsledky porovnejte s údaji vypočtenými na přednášce (m_1 = 95,9, m_2 = 114,4, s_1^2 = 1052,4, s_2^2 = 1057,2, s_1 = 32,44, s_2 = 32,51, cv_1 = 0,338, cv_2 = 0,284).
b) Vypočtěte Pearsonův koeficient korelace meze plasticity a meze pevnosti (r_{12} = 0,9345). Nakreslete dvourozměrný tečkový diagram. Dále vypočtěte kovarianci a výsledek porovnejte s výsledkem vypočteným na přednášce (s_{12} = 985,76).

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)					
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

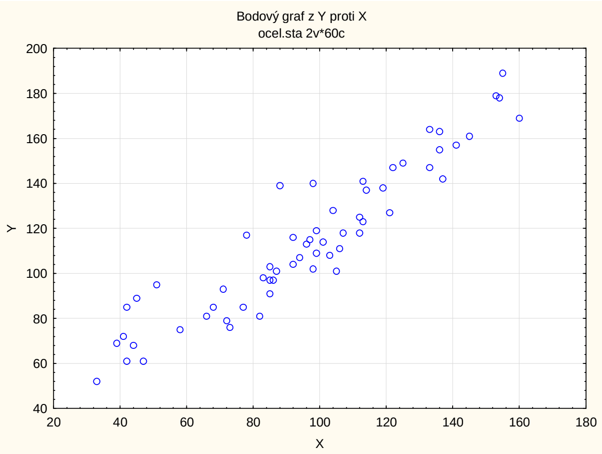
Vysvětlení: Rozptyl a směrodatná odchylka vyjdou ve STATISTICE jinak než bylo uvedeno na přednášce, protože STATISTICA ve vzorci pro výpočet rozptylu nepoužívá 1/n, ale 1/(n-1, jde o tzv. výběrový rozptyl). Koeficient variace (v tabulce označený jako Koef. Prom.) je udán v procentech.

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 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.

Vytvoření dvourozměrného tečkového diagramu: Grafy – Bodové grafy – vypneme Lineární proložení – Proměnné X, Y – OK – OK.



Kovariance se počítá složitěji. Statistiky – 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. Kovariance vyjde ve STATISTICE jinak než bylo uvedeno na přednášce, protože ve STATISTICE se ve vzorci pro výpočet kovariance nepoužívá 1/n, ale 1/(n-1), jde o tzv. výběrovou kovarianci).

Úkol 2.: U 30 náhodně vybraných domácností byly sledovány měsíční výdaje za potraviny (udané v tisících Kč)

výdaje za potraviny	(1,4; 2>	(2; 2,6>	(2,6; 3,2>	(3,2; 3,8>	(3,8; 4,4>	(4,4; 5>
počet domácností	5	7	10	6	1	1

Vypočtěte průměr a směrodatnou odchylku výše měsíčních výdajů za potraviny.

Návod:
Vytvoříme nový datový soubor se šesti případy a dvěma proměnnými X a četnost. Do proměnné X uložíme středy třídicích intervalů (tj. 1,7 2,3 2,9 3,5 4,1 4,7) a do proměnné četnost absolutní četnosti jednotlivých třídicích intervalů.
Statistiky – Základní statistiky/tabulky – Popisné statistiky – zapneme proměnnou vah četnost – Proměnné X – OK – na záložce Detailní výsledky zvolíme Průměr, Rozptyl – Výpočet:

Proměnná		
	Průměr	Rozptyl
X	2,780000	0,463448

Vidíme, že průměrné měsíční výdaje za potraviny činí 2 780 Kč. Rozptyl však musíme upravit, neboť STATISTICA poskytuje výběrový rozptyl. Do Dlouhého jména proměnné Rozptyl napíšeme =5*Rozptyl/6 – OK a do vzniklé tabulky přidáme ještě jednu proměnnou Sm. odchylka. Do jejího Dlouhého jména napíšeme =sqrt(Rozptyl). Výsledná tabulka:

Proměnná	Průměr	Rozptyl =5*v2/6	Sm. odchylka =sqrt(v2)
X	2,780000	0,463448	0,68077035[illegible]

Úkol 3.: 60 náhodně vybraných manželských párů bylo dotázáno na průměrný čistý měsíční příjem manžela (znak X) a také manželky (znak Y). Hodnoty znaku X i hodnoty znaku Y byly roztříděny do 7 třídicích intervalů. Simultánní absolutní četnosti dvourozměrných třídicích intervalů byly zaznamenány do kontingenční tabulky. Meze třídicích intervalů jsou uvedeny v tisících Kč.

$\langle u_j, u_{j+1} \rangle$ \ $\langle v_k, v_{k+1} \rangle$, n_{jk}	$\langle 10,14 \rangle$	$\langle 14,18 \rangle$	$\langle 18,22 \rangle$	$\langle 22,26 \rangle$	$\langle 26,30 \rangle$	$\langle 30,34 \rangle$	$\langle 34,38 \rangle$	$n_{j.}$
$\langle 13,18 \rangle$	6	7	2	0	0	0	0	
$\langle 18,23 \rangle$	4	5	5	0	0	0	0	
$\langle 23,28 \rangle$	2	2	2	3	0	1	0	
$\langle 28,33 \rangle$	1	0	1	1	2	2	1	
$\langle 33,38 \rangle$	0	0	0	3	1	0	1	
$\langle 38,43 \rangle$	0	0	1	1	1	0	1	
$\langle 43,48 \rangle$	0	0	0	0	1	1	2	
$n_{.k}$								60

Vypočtěte průměr, směrodatnou odchylku, koeficient variace obou znaků a jejich koeficient korelace. Nakreslete dvourozměrný tečkový diagram.

Návod:
Vytvoříme nový datový soubor se třemi proměnnými (nazveme je X, Y a četnost) a 7*7 = 49 případy. Do proměnné X uložíme 7x pod sebe střed 1. třídicích intervalu, pak 7x pod sebe střed 2. třídicího intervalu atd. až 7x pod sebe střed 7. třídicího intervalu, a to vždy v Kč, nikoliv v tisících Kč, tedy 15500, 20500, 25500, 30500, 35500, 40500, 45000. STATISTICA totiž při použití váhové proměnné neumožňuje práci s neceločíselnými variantami znaku. Do proměnné Y uložíme pod sebe středy všech sedmi třídicích intervalů (tj. 12000, 16000, 20000, 24000, 28000, 32000, 36000) a těchto sedm čísel uložíme 7x pod sebe. Ke každé dvojici variant znaků X a Y napíšeme příslušnou simultánní absolutní četnost.

Statistiky – Základní statistiky/tabulky – Korelační matice - zapnout proměnnou vah četnost – 1 seznam proměnných X, Y – OK.

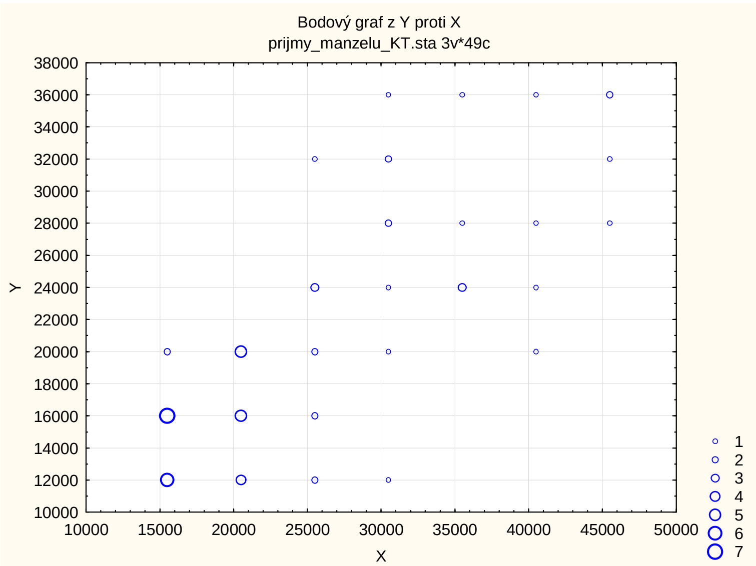
Proměnná	Korelace (prijmy_manzelu_KT.sta) Označ. korelace jsou významné na hlad. p < ,05000 N=60 (Celé případy vynechány u ChD) Průměry	Sm.odch.	X	Y
X	25666,67	9250,134	1,000000	0,754073
Y	20666,67	7516,460	0,754073	1,000000

Směrodatnou odchylku však musíme upravit. Do Dlouhého jména proměnné Sm. odch. napíšeme =sqrt(59*v2^2/60). Dále za proměnnou Sm. odch. přidáme proměnnou cv a do jejího Dlouhého jména napíšeme = v2/v1. Dostaneme tabulku:

Proměnná	Korelace (prijmy_manzelu_KT.sta) Označ. korelace jsou významné na hlad. p < ,05000 N=60 (Celé případy vynechány u ChD)				
	Průměry	Sm.odch.	cv	X	Y
X	25666,67	9172,725	0,35737891	1,000000	0,754073
Y	20666,67	7453,560	0,36065613	0,754073	1,000000

Vidíme, že průměrný příjem manželek je o 5000 Kč menší než průměrný příjem manželů. Koeficienty variace jsou téměř shodné. Koeficient korelace nabývá hodnoty 0,754, což svědčí o existenci silné přímé lineární závislosti mezi příjmy manželů a manželek.

Dvourozměrný tečkový diagram: Grafy – Bodové grafy – vypneme lineární proložení – Proměnné X, Y – OK – na záložce Detaily zvolíme Typ grafu Četnost – OK.



Úkol k samostatnému řešení: Načtěte datový soubor prijem_manzelu.sta, který obsahuje původní údaje o průměrných měsíčních příjmech manželů a manželek. Pro oba znaky vypočtěte průměr, směrodatnou odchylku, koeficient variace, koeficient korelace a porovnejte je s váženými číselnými charakteristikami.

Výsledek:

Proměnná	Korelace (prijem_manzelu.sta) Označ. korelace jsou významné na hlad. p < ,05000 N=60 (Celé případy vynechány u ChD)				
	Průměry	Sm.odch.	cv	X	Y
X	25622,17	9320,308	0,36375956	1,000000	0,797578
Y	20804,33	7502,822	0,36063745	0,797578	1,000000

Úkol 4.: Přírůstky cen akcií na burze (v %) u 10 náhodně vybraných společností dosáhly těchto hodnot: 10, 16, 5, 10, 12, 8, 4, 6, 5, 4. Odhadněte pravděpodobnost, že přírůstek ceny akcie překročí 8,5 %.

Návod: Vytvoříme nový datový soubor o dvou proměnných X a úspěch a 10 případech. Do proměnné X napíšeme přírůstky cen akcií. Nastavíme se kurzorem na proměnnou úspěch. Data – Překódovat – Kategorie 1 - X>8,5 – Nová hodnota 1 – Jiné 0. Vypočítáme průměr proměnné úspěch a zjistíme, že je 0,4.