2. cvičení: Základní pojmy matematické statistiky. Diagnostické grafy.

Příklad 1.: Odvoďte hustotu náhodného výběru z normálního rozložení N(μ, σ^2).
Výsledek:
Náhodný vektor (X_1, …, X_n)’ má hustotu
$\varphi(x_1,\dots,x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2}$, což je hustota n – rozměrného normálního rozložení s vektorem středních hodnot $\boldsymbol{\mu}$ = (μ, …, μ)’ a varianční maticí $\sigma^2\mathbf{I}$.

Příklad 2.: Nechť X_1, ..., X_n je náhodný výběr z rozložení se střední hodnotou μ a rozptylem σ^2. Nechť $n \geq 2$.

a) Vypočtěte střední hodnotu a rozptyl výběrového průměru $M = \frac{1}{n}\sum_{i=1}^{n} X_i$.

b) Vypočtěte střední hodnotu výběrového rozptylu $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2$

Výsledky:
ad a) E(M) = μ, ad b) E(S^2) = σ^2

Příklad 3.: Odvození střední hodnoty a rozptylu výběrové distribuční funkce
Nechť X_1, ..., X_n je náhodný výběr z rozložení s distribuční funkcí $\Phi(x)$. Nechť $n \geq 2$.
Pro libovolné, ale pevně zvolené reálné x vypočtěte střední hodnotu a rozptyl výběrové distribuční funkce $F_n(x) = \frac{1}{n}\mathrm{card}\{i; X_i \leq x\}$.
Výsledky:
$E(F_n(x)) = \Phi(x)$, $D(F_n(x)) = \Phi(x)(1- \Phi(x))/n$.

Příklad 4.: Odvození střední hodnoty výběrové kovariance
Nechť (X_1,Y_1), ..., (X_n,Y_n) je náhodný výběr z dvourozměrného rozložení s vektorem středních hodnot (μ_1, μ_2) a kovariancí σ_{12}. Vypočtěte střední hodnotu výběrové kovariance
$S_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_1)(Y_i - M_2)$.
Výsledek: $E(S_{12}) = \sigma_{12}$

Příklad 5.: Ve 12 náhodně vybraných prodejnách ve městě byly zjištěny následující ceny určitého výrobku (v Kč): 102, 99, 106, 103, 96, 98, 100, 105, 103, 98, 104, 107. Těchto 12 hodnot považujeme za realizace náhodného výběru X_1, ..., X_{12} z rozložení, které má střední hodnotu μ a rozptyl σ^2.
a) Vypočtěte realizaci výběrového průměru a výběrového rozptylu.
b) Najděte výběrovou distribuční funkci $F_{12}(x)$ a nakreslete její graf.
Výsledky: m = 101,75 Kč, s^2 = 12,39 Kč^2
Hodnoty a graf výběrové distribuční funkce

$$x < 96: F_{12}(x) = 0$$

$$96 \le x < 98: F_{12}(x) = \frac{1}{12} = 0{,}08\overline{3}$$

$$98 \le x < 99: F_{12}(x) = \frac{3}{12} = 0{,}25$$

$$99 \le x < 100: F_{12}(x) = \frac{4}{12} = 0{,}\overline{3}$$

$$100 \le x < 102: F_{12}(x) = \frac{5}{12} = 0{,}41\overline{6}$$

$$102 \le x < 103: F_{12}(x) = \frac{6}{12} = 0{,}5$$

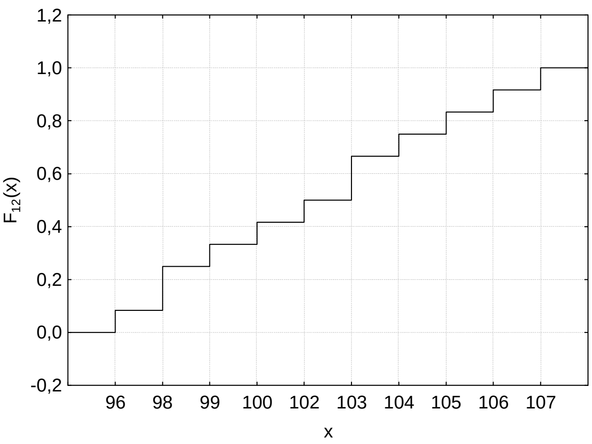
$$103 \le x < 104: F_{12}(x) = \frac{8}{12} = 0{,}\overline{6}$$

$$104 \le x < 105: F_{12}(x) = \frac{9}{12} = 0{,}75$$

$$105 \le x < 106: F_{12}(x) = \frac{10}{12} = 0{,}8\overline{3}$$

$$106 \le x < 107: F_{12}(x) = \frac{11}{12} = 0{,}91\overline{6}$$

$$x \ge 107: F_{12}(x) = 1$$



Výpočet pomocí systému STATISTICA:
Otevřeme nový datový soubor o jedné proměnné (nazveme ji X) a 12 případech. Do proměnné X napíšeme zjištěné ceny.

Výpočet realizace výběrového průměru a výběrového rozptylu:
Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnné X – OK – Detailní výsledky – vybereme Průměr a Rozptyl – Výpočet. Dostaneme tabulku:

	Popisné statistiky (Tabulka15)	
Proměnná	Průměr	Rozptyl
X	101,7500	12,38636

Výpočet hodnot výběrové distribuční funkce:
Statistiky – Základní statistiky/tabulky – Tabulky četností – OK – Proměnné X – OK – Možnosti – ponecháme zaškrtnuté pouze Kumulativní relativní četnosti – Výpočet.

Kreslení grafu výběrové distribuční funkce:
Ke vzniklé tabulce přidáme jeden případ před první případ (do sloupce Kategorie napíšeme 95, do sloupce Kumulativní rel. četnost napíšeme 0) a jeden případ za poslední případ (do sloupce Kategorie napíšeme 107, do sloupce Kumulativní rel. četnost napíšeme 100). Proměnnou Kumulativní rel. četnost podělíme 100: do jejího Dlouhého jména napíšeme = v2/100.
Nastavíme se kurzorem na proměnnou Kumulativní rel. četnost, klikneme pravým tlačítkem – Grafy bloku dat – Spojnicový graf: celé sloupce. Ve vytvořeném grafu odstraníme značky, spojnici změníme na schodovitou a upravíme měřítko na vodorovné ose od 1 do 12.

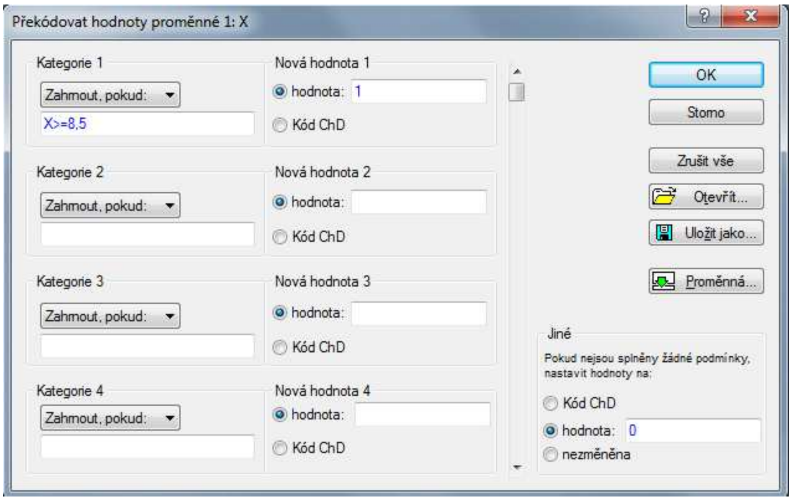
Příklad 6.: Přírůstky cen akcií v % na burze v New Yorku u 10 náhodně vybraných společností dosáhly těchto hodnot: 10, 16, 5, 10, 12, 8, 4, 6, 5, 4.
Odhadněte střední hodnotu a směrodatnou odchylku růstu cen akcií a dále odhadněte pravděpodobnost růstu cen akcií aspoň o 8,5 %.
Pomocí systému STATISTICA nakreslete krabicový diagram a NP plot.
Výsledky: Průměrný růst cen akcií odhadujeme na 8 % se směrodatnou odchylkou 3,97 %. Dále, u 40 % akcií vzrostla cena aspoň o 8,5 %.

Výpočet pomocí systému STATISTICA:
Otevřeme nový datový soubor o jedné proměnné (nazveme ji X) a 10 případech. Do proměnné X napíšeme zjištěné přírůstky cen akcií.

Výpočet realizace výběrového průměru a výběrové směrodatné odchylky:
Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnné X – OK – Detailní výsledky – vybereme Průměr a Směrodat. odchylka – Výpočet. Dostaneme tabulku:

	Popisné statistiky (Tabulka1)	
Proměnná	Průměr	Sm.odch.
X	8,000000	3,972125

Odhad pravděpodobnosti růstu cen akcií aspoň o 8,5 %:
Překódujeme hodnoty proměnné X tak, že hodnotám větším nebo rovným 8,5 přiřadíme 1 a ostatním hodnotám 0.
Nastavíme se kurzorem na X. Data – Překódovat. Otevře se okno, které vyplníme takto:

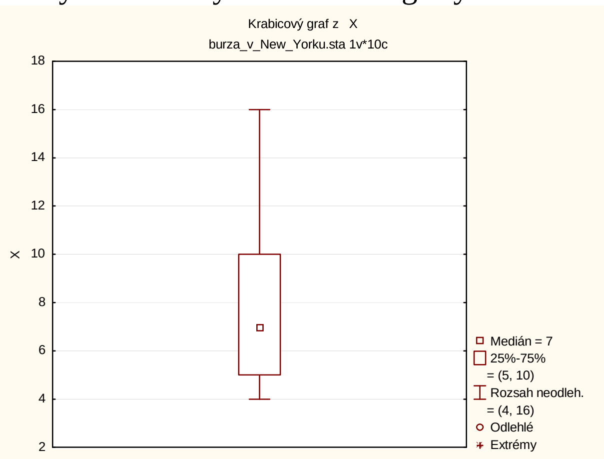


Statistiky – Základní statistiky/tabulky – Tabulky četností – OK – Proměnné X – OK –– Výpočet.

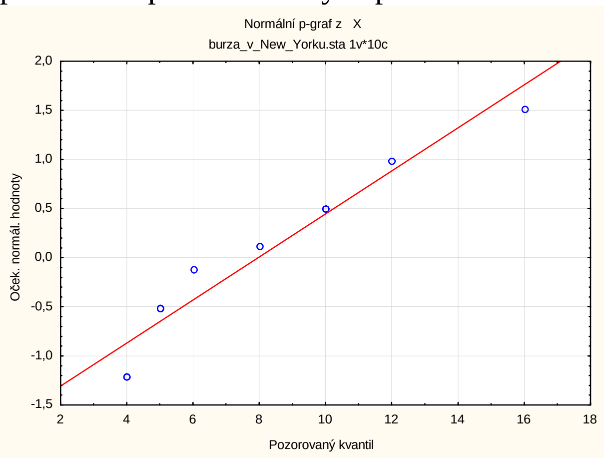
	Tabulka četností:X (Tabulka1)			
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
0	6	6	60,00000	60,0000
1	4	10	40,00000	100,0000
ChD	0	10	0,00000	100,0000

Vidíme, že u Kategorie 1 je relativní četnost 40 %.

Kreslení krabicového grafu:
Grafy – 2D Grafy – Krabicové grafy - Proměnné – Závisle proměnné X - OK.



Kreslení NP plotu:
Grafy – 2D Grafy – Normální pravděpodobnostní grafy – Proměnné X – zrušíme Neurčovat průměrnou pozici svázaných pozorování - OK.



Příklad 7.: Výpočet výběrového koeficientu korelace
Máme k dispozici výsledky testů ze dvou předmětů zjištěné u osmi náhodně vybraných studentů určitého oboru.

Číslo studenta	1	2	3	4	5	6	7	8
Počet bodů v 1. testu	80	50	36	58	42	60	56	68
Počet bodů ve 2. testu	65	60	35	39	48	44	48	61

Vypočtěte a interpretujte výběrový koeficient korelace. Pro usnadnění výpočtů máte k dispozici tyto součty:

$$\sum_{i=1}^{8} x_i = 450, \sum_{i=1}^{8} y_i = 400, \sum_{i=1}^{8} x_i^2 = 26684, \sum_{i=1}^{8} y_i^2 = 20836, \sum_{i=1}^{8} x_i y_i = 23214$$

Dále pomocí systému STATISTICA nakreslete dvourozměrný tečkový diagram s proloženou 95% elipsou konstantní hustoty pravděpodobnosti.

Výsledek:
$r_{12} = 0{,}6668$, mezi výsledky obou testů existuje středně silná přímá lineární závislost.

Výpočet pomocí systému STATISTICA:
Otevřeme nový datový soubor o dvou proměnných X a Y a osmi případech. Do proměnných X a Y zapíšeme hodnoty testů.
Výpočet výběrové kovariance: Statistiky – Vícerozměrná regrese – Proměnné – Závisle proměnná Y, nezávisle proměnná X – OK – OK – Residua/předpoklady/předpovědi – Popisné statistiky – Další statistiky – Kovariance. Dostaneme tabulku:

	Kovariance (dva testy.sta)	
Proměnná	X	Y
X	195,9286	102,0000
Y	102,0000	119,4286

Vidíme, že výběrová kovariance veličin X, Y se realizuje hodnotou 102. (Výběrový rozptyl proměnné X resp. Y nabyl hodnoty 195,93 resp. 119,43.)

Výpočet výběrového koeficientu korelace: V menu Další statistiky vybereme Korelace.

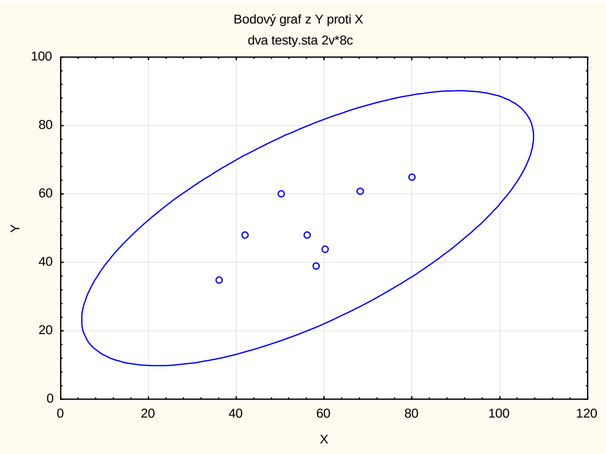
	Korelace (dva testy.sta)	
Proměnná	X	Y
X	1,000000	0,666802
Y	0,666802	1,000000

Výběrový koeficient korelace veličin X, Y nabyl hodnoty 0,6668, tedy mezi veličinami x, Y existuje středně silná přímá lineární závislost.

Upozornění: Výběrový koeficient korelace lze pomocí systému STATISTICA vypočítat i jiným způsobem: Statistiky – Základní statistiky/tabulky – Korelační matice – OK – 1 seznam proměnných – X, Y – OK – Výpočet. Ve výsledné tabulce máme též realizace výběrových průměrů a směrodatných odchylek.

	Korelace (dva testy.sta) Označ. korelace jsou významné na hlad. p < ,05000 N=8 (Celé případy vynechány u ChD)			
Proměnná	Průměry	Sm.odch.	X	Y
X	56,25000	13,99745	1,000000	0,666802
Y	50,00000	10,92834	0,666802	1,000000

Kreslení dvourozměrného tečkového diagramu s 95% elipsou konstantní hustoty pravděpodobnosti:
Grafy – 2D Grafy - Bodové grafy. Vypneme lineární proložení. Zadáme Proměnné – X – Y – OK.
Dostaneme dvourozměrný tečkový diagram pro vektorovou proměnnou (X, Y).
Nyní do diagramu zakreslíme 95% elipsu konstantní hustoty pravděpodobnosti: 2x klikneme na pozadí grafu a otevře se okno s názvem Vš. možnosti.
Vybereme Graf: Elipsa, zvolíme Přidat novou elipsu.
Po vykreslení elipsy změníme měřítko: na vodorovné ose bude minimum 0, maximum 120, na svislé ose bude minimum 0, maximum 100. (Stačí 2x kliknout na číselný popis osy a na záložce Měřítka vybrat manuální mód.)


Bodový graf z Y proti X
dva testy.sta 2v*8c
100
80
60
40
20
0
Y
0
20
40
60
80
100
120
X