Téma 7.: Číselné charakteristiky diskrétních náhodných veličin, M-L věta

Výpočet střední hodnoty a rozptylu diskrétní náhodné veličiny

Střední hodnota: $E(X) = \sum_{x=-\infty}^{\infty} x\pi(x)$, **rozptyl:** $D(X) = \sum_{x=-\infty}^{\infty} x^2\pi(x) - \left[\sum_{x=-\infty}^{\infty} x\pi(x)\right]^2$

Střední hodnota a rozptyl lineární kombinace: E(a+bX) = a + bE(X), D(a+bX) = b^2D(X)

Příklad 1.: Postupně se zkouší spolehlivost čtyř přístrojů. Další se zkouší jen tehdy, když předchozí je spolehlivý. Každý z přístrojů vydrží zkoušku s pravděpodobností 0,8. Náhodná veličina X udává počet zkoušených přístrojů. Vypočtěte střední hodnotu a rozptyl náhodné veličiny X.

Řešení:

X nabývá hodnot 1, 2, 3, 4 a její pravděpodobnostní funkce je $\pi(1) = 0{,}2$, $\pi(2) = 0{,}8*0{,}2 = 0{,}16$, $\pi(3) = 0{,}8^2*0{,}2 = 0{,}128$, $\pi(4) = 0{,}8^3*0{,}2 + 0{,}8^4 = 0{,}512$, $\pi(0) = 0$ jinak

E(X) = 1*0,2 + 2*0,16 + 3*0,128 + 4*0,512 = 2,952

$D(X) = 1^2*0{,}2 + 2^2*0{,}16 + 3^2*0{,}128 + 4^2*0{,}512 - 2{,}952^2 = 1{,}4697$

Postup ve STATISTICE:

Otevřeme nový datový soubor o dvou proměnných X a cetnost a čtyřech případech. Do proměnné X napíšeme 1, 2, 3, 4, do proměnné cetnost napíšeme 200, 160, 128, 512 (hodnoty váhové proměnné musí být celá čísla).

Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – zavedeme proměnnou vah cetnost – OK - Proměnné X – OK – Detailní výsledky - zaškrtneme Průměr, Rozptyl – Výpočet.

	Popisné statistiky (Tabulka1)		
Proměnná	N platných	Průměr	Rozptyl
X	1000	2,952000	1,471167

Rozptyl však musíme upravit, musíme ho vynásobit číslem 999/1000. Do výstupní tabulky tedy přidáme za proměnnou Rozptyl novou proměnnou a do jejího Dlouhého jména napíšeme =v3*999/1000

	Popisné statistiky (Tabulka1)			
Proměnná	N platných	Průměr	Rozptyl	NProm
X	1000	2,952000	1,471167	1,469696

Příklad 2.: Dodavatel se zajímá o celkovou cenu projektu. Odhaduje, že materiál bude stát 25 000 Kč a práce 900 Kč denně. Doba trvání projektu (ve dnech) je popsána náhodnou veličinou X, která může nabývat hodnot 10, 11, 12, 13, 14 s pravděpodobnostmi 0,1, 0,3, 0,3, 0,2 a 0,1. Vypočtěte střední hodnotu a směrodatnou odchylku celkové ceny projektu.

Řešení:

Nejprve vypočteme střední hodnotu a rozptyl doby trvání projektu:

$$E(X) = \sum_{x=10}^{14} x\pi(x) = 10 \cdot 0{,}1 + 11 \cdot 0{,}3 + 12 \cdot 0{,}3 + 13 \cdot 0{,}2 + 14 \cdot 0{,}1 = 11{,}9$$

$$D(X) = \sum_{x=10}^{14} x^2\pi(x) - [E(X)]^2 = 10^2 \cdot 0{,}1 + 11^2 \cdot 0{,}3 + 12^2 \cdot 0{,}3 + 13^2 \cdot 0{,}2 + 14^2 \cdot 0{,}1 - 11{,}9^2 = 1{,}29$$

Zavedeme náhodnou veličinu Y, která udává celkovou cenu projektu. Je zřejmé, že $Y = 25000 + 900X$. Z vlastností střední hodnoty a rozptylu plyne:

$$E(Y) = E(25000 + 900X) = 25000 + 900E(X) = 25000 + 900 \cdot 11{,}9 = 35710 \text{ Kč}$$

$$D(Y) = D(25000 + 900X) = 900^2 D(X) = 810000 \cdot 1{,}29 = 1044900 \text{ Kč}^2$$

$$\sqrt{D(Y)} = \sqrt{1044900} = 1022{,}20 \text{ Kč}$$

Postup ve STATISTICE:
Otevřeme nový datový soubor o dvou proměnných X a cetnost a pěti případech. Do proměnné X napíšeme 10, 11, 12, 13, 14, do proměnné cetnost napíšeme 1, 3, 3, 2, 1.
Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – zavedeme proměnnou vah cetnost – OK - Proměnné X – OK – Detailní výsledky - zaškrtneme Průměr, Rozptyl – Výpočet.

	Popisné statistiky (Tabulka16)		
Proměnná	N platných	Průměr	Rozptyl
X	10	11,90000	1,433333

Rozptyl však musíme upravit, musíme ho vynásobit číslem 9/10. Do výstupní tabulky tedy přidáme za proměnnou Rozptyl novou proměnnou a do jejího Dlouhého jména napíšeme =v3*9/10.

	Popisné statistiky (Tabulka16)			
Proměnná	N platných	Průměr	Rozptyl	NProm
X	10	11,90000	1,433333	1,29

Pro výpočet střední hodnoty a směrodatné odchylky celkové ceny projektu přidáme k této tabulce dvě nové proměnné, které nazveme E(Y) a sqrtD(Y). Do Dlouhého jména proměnné E(Y) napíšeme =25000+900*v2 a do Dlouhého jména proměnné sqrtD(Y) napíšeme =900*sqrt(v4).

Výsledná tabulka:

	Popisné statistiky (Tabulka16)					
Proměnná	N platných	Průměr	Rozptyl	NProm	E(Y)	sqrtD(Y)
X	10	11,90000	1,433333	1,29	35710	1022,2035

Příklad k samostatnému řešení: Náhodná veličina X udává počet ok při hodu kostkou. Pomocí systému STATISTICA vypočtěte její střední hodnotu a rozptyl.
Výsledek: E(X) = 3,5, D(X) = 2,9167

Výpočet koeficientu korelace diskrétních náhodných veličin

Koeficient korelace: $R(X_1, X_2) = \dfrac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$, kde

kovariance $C(X_1, X_2) = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 x_2 \pi(x_1, x_2) - \sum_{x_1=-\infty}^{\infty} x_1 \pi_1(x_1) \sum_{x_2=-\infty}^{\infty} x_2 \pi_2(x_2)$

Příklad 3.: Náhodná veličina X udává roční příjem manžela (v tisících dolarů) a náhodná veličina Y roční příjem manželky (v tisících dolarů). Je známa simultánní pravděpodobnostní funkce $\pi(x,y)$ diskrétního náhodného vektoru (X,Y):

Příjem manžela	Příjem manželky			
	10	20	30	40
10	0,2	0,04	0,01	0
20	0,1	0,36	0,09	0
30	0	0,05	0,1	0
40	0	0	0	0,05

Vypočtěte koeficient korelace příjmů manžela a manželky.
Řešení:
Stanovíme hodnoty marginálních pravděpodobnostních funkcí $\pi_1(x)$, $\pi_2(y)$:

Příjem manžela	Příjem manželky				$\pi_1(x)$
	10	20	30	40	
10	0,2	0,04	0,01	0	0,25
20	0,1	0,36	0,09	0	0,55
30	0	0,05	0,1	0	0,15
40	0	0	0	0,05	0,05
$\pi_2(y)$	0,3	0,45	0,2	0,05	1

Spočteme E(X) = 20, E(Y) = 20, D(X) = 60, D(Y) = 70. Dosazením do vzorce pro výpočet kovariance zjistíme, že C(X,Y) = 49, tedy koeficient korelace $R(X,Y) = 49/\sqrt{60}\sqrt{70} = 0,76$.

Postup ve STATISTICE:
Vytvoříme nový datový soubor o třech proměnných X, Y, cetnost a 16 případech. Do proměnné X napíšeme 10, 10, 10, 10, 20, 20, 20, 20, 30, 30, 30, 30, 40 ,40 40, 40 , do proměnné Y 4x pod sebe 10, 20, 30, 40 a do proměnné cetnost 20, 4, 1, 0, 10, 36, 9, 0, 0, 5, 10, 0, 0, 0, 0, 5.
Statistiky - Základní statistiky/tabulky – zavedeme proměnnou vah cetnost – OK - Korelační matice – OK – 1 seznam proměnných – X, Y – OK.

	Korelace (Tabulka6) Označ. korelace jsou významné na hlad. p < ,05000 N=100 (Celé případy vynechány u ChD)			
Proměnná	Průměry	Sm.odch.	X	Y
X	20,00000	7,784989	1,000000	0,756086
Y	20,00000	8,408750	0,756086	1,000000

Příklad k samostatnému řešení: Diskrétní náhodný vektor (X_1,X_2) má simultánní pravděpodobnostní funkci s hodnotami $\pi(0,-1) = c$, $\pi(0,0) = \pi(0,1) = \pi(1,-1) = \pi(2,-1) = 0$, $\pi(1,0) = \pi(1,1) = \pi(2,1) = 2c$, $\pi(2,0) = 3c$, $\pi(x,y) = 0$ jinak. Určete konstantu c a vypočtěte $R(X_1,X_2)$.
Výsledek: $c = 0,1$, $R(X_1,X_2) = 0,42379$.

Aplikace Moivreovy - Laplaceovy věty

Příklad 4.: Pravděpodobnost úspěchu při jednom pokusu je 0,3. S jakou pravděpodobností lze tvrdit, že počet úspěchů ve 100 pokusech bude v mezích od 20 do 40?

Y_{100} – počet úspěchů v posloupnosti n = 100 opakovaných nezávislých pokusů, pravděpodobnost úspěchu ϑ = 0,3, $Y_{100} \sim$ Bi(100, 0,3), E(Y_{100}) = $n\vartheta$ = 30, D(Y_{100}) = $n\vartheta(1-\vartheta)$ = 21.

Aproximativní výpočet:
Vidíme, že podmínky dobré aproximace jsou splněny, protože
$\frac{1}{101} < 0{,}3 < \frac{100}{101}$ a $100 \cdot 0{,}3 \cdot (1 - 0{,}3) = 21 > 9$.

$$P(20 \le Y_{100} \le 40) = P\left(\frac{19-30}{\sqrt{21}} < \frac{Y_{100}-30}{\sqrt{21}} \le \frac{40-30}{\sqrt{21}}\right) \approx \Phi\left(\frac{10}{\sqrt{21}}\right) - \Phi\left(-\frac{11}{\sqrt{21}}\right) = 0{,}9773,$$

kde Φ(x) je distribuční funkce rozložení N(0,1).

Přesný výpočet:

$$P(20 \le Y_{100} \le 40) = P(19 < Y_{100} \le 40) = \Phi(40) - \Phi(19) = 0{,}978614,$$

kde Φ(x) je distribuční funkce rozložení Bi(100, 0,3).

Postup ve STATISTICE:
Otevřeme nový datový soubor se dvěma proměnnými a jedním případem. Nastavíme se kurzorem na 1. sloupec.
Do Dlouhého jména první proměnné napíšeme
=INormal(10/sqrt(21);0;1)- INormal(-11/sqrt(21);0;1) OK.
(Funkce INormal(x;mu;sigma) poskytuje hodnotu distribuční funkce v bodě x normálního rozložení se střední hodnotou mu a směrodatnou odchylkou sigma.)
Nastavíme se kurzorem na 2. sloupec. Do Dlouhého jména druhé proměnné napíšeme =IBinom(40;0,3;100)- IBinom(19;0.3;100). (Funkce IBinom(x;p;n) poskytuje hodnotu distribuční funkce v bodě x binomického rozložení s parametry p a n.)

	1 Prom1	2 Prom2
1	0,97726318	0,97861426

Příklad k samostatnému řešení: Pravděpodobnost, že určitý typ výrobku bude mít výrobní vadu, je 0,05. Jaká je pravděpodobnost, že ze série 1000 výrobků bude mít výrobní vadu nejvýše 70 výrobků? (Při aproximativním výpočtu nezapomeňte na ověření podmínek dobré aproximace.)
Výsledek: ≈0,99815, =0,9977