Téma 7: Výpočet číselných charakteristik náhodných veličin

Výpočet kvantilů pomocí systému STATISTICA:
První způsob: Statistica – Probability Calculator – Distributions
a) Normální rozložení
Ve volbě Distributions vybereme Z (Normal), do okénka mean napíšeme hodnotu μ a do okénka st. dev. napíšeme hodnotu σ. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Compute se v okénku X objeví hodnota tohoto kvantilu.
b) Pearsonovo rozložení chí-kvadrát s n stupni volnosti $\chi^2(n)$
Ve volbě Distributions vybereme Chi 2 a do okénka df napíšeme patřičný počet stupňů volnosti. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Compute se v okénku Chi 2 objeví hodnota tohoto kvantilu.
c) Studentovo rozložení s n stupni volnosti t(n)
Ve volbě Distributions vybereme t (Student) a do okénka df napíšeme patřičný počet stupňů volnosti. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Compute se v okénku t objeví hodnota tohoto kvantilu.
d) Fisherovo-Snedecorovo rozložení s n_1 a n_2 stupni volnosti $F(n_1, n_2)$
Ve volbě Distributions vybereme F (Fisher) a do okének df1 a df2 napíšeme počet stupňů volnosti čitatele a jmenovatele. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Compute se v okénku F objeví hodnota tohoto kvantilu.

Druhý způsob: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V Long name této proměnné použijeme funkci
a) VNormal(x;mu;sigma) pro x-kvantil normálního rozložení se střední hodnotou mu a směrodatnou odchylkou sigma
b) VChi2(x;nu) pro x-kvantil Pearsonova rozložení s nu stupni volnosti
c) Student(x;df) pro x-kvantil Studentova rozložení s df stupni volnosti
d) VF(x;nu;omega) pro x-kvantil Fisherova – Snedecorova rozložení s nu a omega stupni volnosti.

Výpočet střední hodnoty a rozptylu diskrétní náhodné veličiny
Vzorový 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

$$E(X) = \sum_{x=1}^{4} x\pi(x) = 1.0{,}2 + 2.0{,}16 + 3.0{,}128 + 4.0{,}512 = 2{,}952$$

$$D(X) = E(X^2) - [E(X)]^2 = \sum_{x=1}^{4} x^2\pi(x) - [E(X)]^2 =$$

$$= 1^2.0{,}2 + 2^2.0{,}16 + 3^2.0{,}128 + 4^2.0{,}512 - 2{,}952^2 = 1{,}4697$$

Postup ve STATISTICE:
Otevřeme nový datový soubor o čtyřech případech a pěti proměnných, které nazveme x, pi(x), x*pi(x), xkvadrat, xkvadrat*pi(x). První proměnnou naplníme hodnotami náhodné veličiny X, druhou hodnotami její pravděpodobnostní funkce. Do třetí proměnné uložíme součin $x\pi(x)$ (do Long name napíšeme =v1*v2), do čtvrté x^2 (do Long name napíšeme =v1^2), do páté součin $x^2\pi(x)$ (do Long name napíšeme v4*v2).

x	pi(x)	x*pi(x)	xkvadrat	xkvadrat*pi(x)
1	0,2	0,2	1	0,2
2	0,16	0,32	4	0,64
3	0,128	0,384	9	1,152
4	0,512	2,048	16	8,192

Výpočty E(X) a D(X) provedeme takto:

Statistics – Basic Statistics/Tables – Descriptive Statistics –Variables x*pi(x), xkvadrat*pi(x) – OK, zaškrtneme Sum - Summary

Proměnnou Sum ve workbooku transponujeme: Data – Transpose – File.
Proměnou x*pi(x) přejmenujeme na E(X) (vidíme, že E(X) = 2,952). Přidáme (ve workbooku) proměnnou D(X) a do jejího Long name napíšeme = v2-v1^2.
Vidíme, že D(X) = 1,4697.

	Descriptiv
Variable	Sum
x*pi(x)	2,95200
xkvadrat*pi(x)	10,18400

Výpočet koeficientu korelace
Vzorový příklad 2. Náhodná veličina X udává příjem manžela (v tisících dolarů) a náhodná veličina Y příjem manželky (v tisících dolarů. Je známa simultánní pravděpodobnostní funkce $\pi(x,y)$ diskrétního náhodného vektoru (X,Y): $\pi(10,10) = 0,2$, $\pi(10,20) = 0,04$, $\pi(10,30) = 0,01$, $\pi(10,40) = 0$, $\pi(20,10) = 0,1$, $\pi(20,20) = 0,36$, $\pi(20,30) = 0,09$, $\pi(20,40) = 0$, $\pi(30,10) = 0$, $\pi(30,20) = 0,05$, $\pi(30,30) = 0,1$, $\pi(30,40) = 0$, $\pi(40,10) = 0$, $\pi(40,20) = 0$, $\pi(40,30) = 0$, $\pi(40,40) = 0,05$, $\pi(x,y) = 0$ jinak. Vypočtěte koeficient korelace příjmů manžela a manželky.

Řešení:
Náhodná veličina X i náhodná veličina Y nabývají hodnot 10, 20, 30, 40. Stanovíme hodnoty marginálních pravděpodobnostních funkcí: $\pi_1(10) = 0,25$, $\pi_1(20)=0,55$, $\pi_1(30) = 0,15$, $\pi_1(40) = 0,05$, $\pi_1(x) = 0$ jinak, $\pi_2(10) = 0,3$, $\pi_2(20) = 0,45$, $\pi_2(30) = 0,2$, $\pi_2(10) = 0,05$, $\pi_2(y) = 0$ jinak. 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:
Budeme potřebovat dva nové soubory. První pro výpočet středních hodnot a rozptylů, druhý pro výpočet kovariance a koeficientu korelace. První soubor bude mít 4 případy a 10 proměnných. Zde jsou pro výpočet středních hodnot a rozptylů použity dva soubory vzhledem k přílišné délce tabulky pro obě náhodné veličiny.

x	pi(x)	x*pi(x)	xkvadrat	xkvadrat*pi(x)
10	0,25	2,5	100	25
20	0,55	11	400	220
30	0,15	4,5	900	135
40	0,05	2	1600	80

Variable	Descriptiv Sum
x*pi(x)	20,0000
xkvadrat*pi(x)	460,0000

y	pi(y)	y*pi(y)	ykvadrat	ykvadrat*pi(y)
10	0,3	3	100	30
20	0,45	9	400	180
30	0,2	6	900	180
40	0,05	2	1600	80

Variable	Descriptiv Sum
y*pi(y)	20,0000
ykvadrat*pi(y)	470,0000

Nyní vytvoříme nový datový soubor o 16 případech a 4 proměnných, které nazveme x, y, pi(x,y) a x*y*pi(x,y). Do první proměnné napíšeme 10, 10, 10, 10, 20, 20, 20, 20, 30, 30, 30, 30, 40, 40, 40, 40 a do druhé proměnné 10, 20, 30, 40, 10, 20, 30, 40, 10, 20, 30, 40, 10, 20, 30, 40.
Do třetí proměnné zapíšeme hodnoty simultánní pravděpodobnostní funkce $\pi(x,y)$ a do čtvrté proměnné uložíme součin $xy\pi(x,y)$ (do Long name napíšeme =v1*v2*v3).

x	y	pi(x,y)	x*y*pi(x,y)
10	10	0,2	20
10	20	0,04	8
10	30	0,01	3
10	40	0	0
20	10	0,1	20
20	20	0,36	144
20	30	0,09	54
20	40	0	0
30	10	0	0
30	20	0,05	30
30	30	0,1	90
30	40	0	0
40	10	0	0
40	20	0	0
40	30	0	0
40	40	0,05	80

Statistics – Basic Statistics/Tables – Variables x*y*pi(x,y) – OK , zaškrtneme

Sum – Summary.

	Descriptiv
Variable	Sum
x*y*pi(x,y)	449,0000

Proměnnou Sum ve workbooku přejmenujeme na E(X,Y) a přidáme k ní 6 nových proměnných E(X), E(Y), D(X), D(Y), C(X,Y), R(X,Y). Do proměnných E(X), E(Y), D(X), D(Y) napíšeme vypočtené střední hodnoty a rozptyly. Do Long name proměnné C(X,Y) napíšeme=v1-vv2*v3 a do Long name proměnné R(X,Y) napíšeme =v6/sqrt(v4*v5).

	E(X,Y)	E(X)	E(Y)	D(X)	D(Y)	C(X,Y)	R(X,Y)
x*y*pi(x,y)	449	20	20	60	70	49	0,756086

Příklady k samostatnému řešení:
1. Náhodná veličina X udává počet ok při hodu kostkou. Pomocí systému STATISTICA vypočtěte její střední hodnotu a rozptyl.
(Výsledek: E(X) = 21/6 = 3,5, D(X) = 35/12 = 2,9167)

2. 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 pomocí systému STATISTICA vypočtěte $R(X_1,X_2)$.
(Výsledek: c = 0,1, E(X) = 1,4, E(Y) = 0,3, D(X) = 0,44, D(Y) = 0,41. Dosazením do vzorce pro výpočet kovariance zjistíme, že C(X,Y) = 0,18, tedy koeficient korelace R(X,Y) = $0,18/\sqrt{0,44}\sqrt{0,41} = 0,42379$)