

Téma 5.: Výpočet číselných charakteristik náhodných veličin pomocí systému STATISTICA

K výpočtu kvantilů mnoha typů spojitých rozložení slouží Pravděpodobnostní kalkulátor v menu Statistika. Kvantily lze počítat též pomocí funkcí implementovaných v položce „Dlouhé jméno“ proměnné.

Normální rozložení $N(\mu, \sigma^2)$

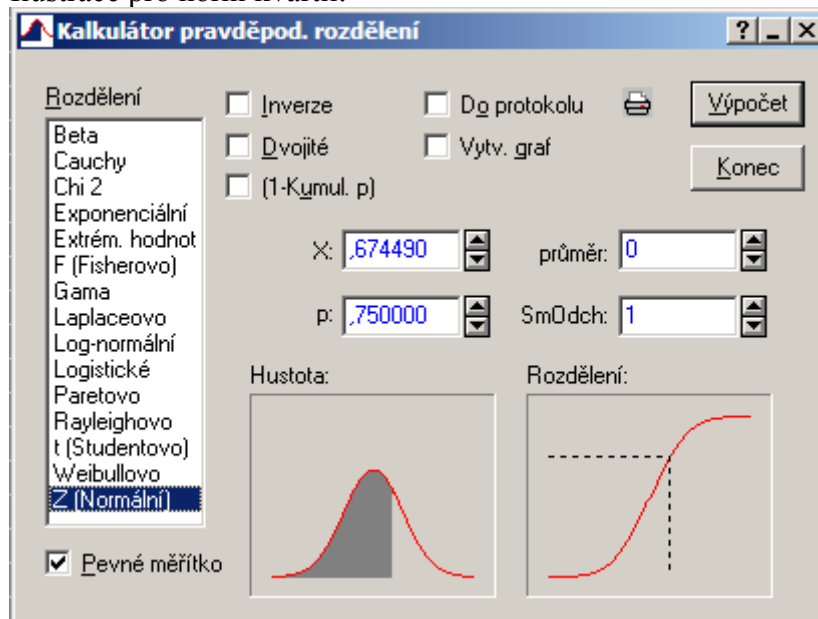
Náhodná veličina $X \sim N(\mu, \sigma^2)$ má hustotu $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x-\mu}{2\sigma^2}}$. Pro $\mu = 0, \sigma^2 = 1$ se jedná o standardizované normální rozložení, píšeme $U \sim N(0, 1)$. Hustota pravděpodobnosti má v tomto případě tvar $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$.

Příklad 1.: Nechť $U \sim N(0, 1)$. Najděte medián a horní a dolní kvartil.

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka průměr napíšeme 0, do okénka Sm. Odch. napíšeme 1, do okénka p napíšeme pro medián 0,5, pro dolní kvartil 0,25 a pro horní kvartil 0,75. V okénku X se objeví 0 pro medián, -0,67449 pro dolní kvartil a 0,67449 pro horní kvartil.

Ilustrace pro horní kvartil:



Šedá plocha pod grafem hustoty má velikost 0,75 a hodnota distribuční funkce v bodě 0,67449 je 0,75 (značeno šrafovaně).

Druhá možnost: Otevřeme nový datový soubor o třech proměnné a jednom případě.

Do dlouhého jména první proměnné napíšeme =VNormal(0,5;0;1). Dostaneme 0.

Do dlouhého jména druhé proměnné napíšeme =VNormal(0,25;0;1). Dostaneme -0,67449.

Do dlouhého jména třetí proměnné napíšeme =VNormal(0,75;0;1). Dostaneme 0,67449.

Příklad 2.: Nechť $X \sim N(3, 5)$. Najděte dolní kvartil.

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka průměr napíšeme 3, do okénka Sm. Odch. napíšeme 2,236, do okénka p napíšeme 0,25 a v okénku X se objeví 1,4918.

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případě. Do dlouhého jména této proměnné napíšeme =VNormal(0,25;3;sqrt(5)). Dostaneme 1,491795.

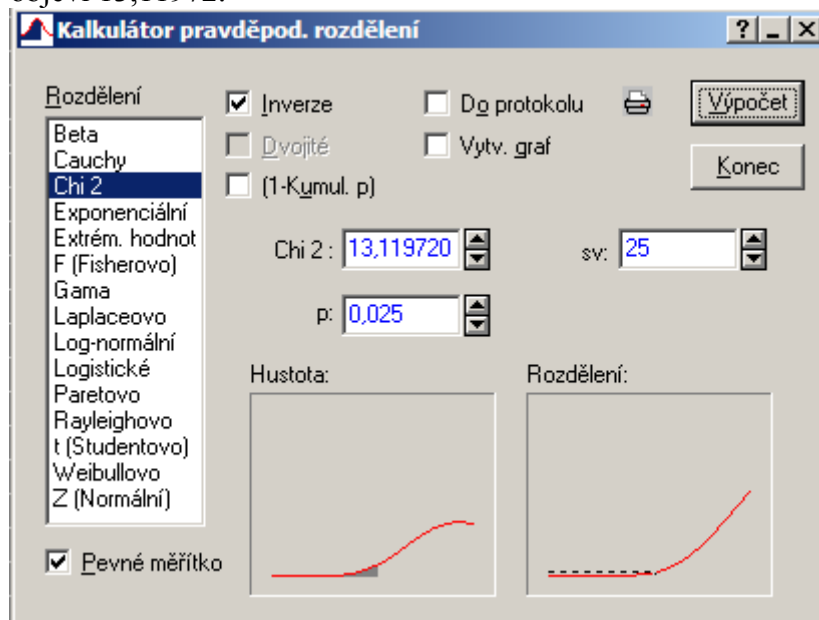
Pearsonovo rozložení chí-kvadrát s n stupni volnosti $\chi^2(n)$

Nechť X_1, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1)$, $i = 1, \dots, n$. Pak náhodná veličina $X = X_1^2 + \dots + X_n^2 \sim \chi^2(n)$. Vyjádření hustoty je příliš složité, lze ho najít např. v příloze A skript Marie Budíkové, Pavel Osecký, Štěpán Mikoláš: Teorie pravděpodobnosti a matematická statistika. Sběrka příkladů. MU Brno 2007.

Příklad 3.: Určete $\chi^2_{0,025}(25)$.

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka sv. napíšeme 25 a do okénka p napíšeme 0,025. V okénku Chi 2 se objeví 13,11972.



Šedá plocha pod grafem hustoty má velikost 0,025 a hodnota distribuční funkce v bodě 13,11972 je 0,025 (značeno šrafovaně).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případě. Do dlouhého jména této proměnné napíšeme =VChi2(0,025;25). Dostaneme 13,1197.

Studentovo rozložení s n stupni volnosti t(n)

Nechť X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_1 \sim N(0, 1)$, $X_2 \sim \chi^2(n)$. Pak náhodná veličina $X = \frac{X_1}{\sqrt{\frac{X_2}{n}}} \sim t(n)$. Vyjádření hustoty je příliš složité, lze ho najít např.

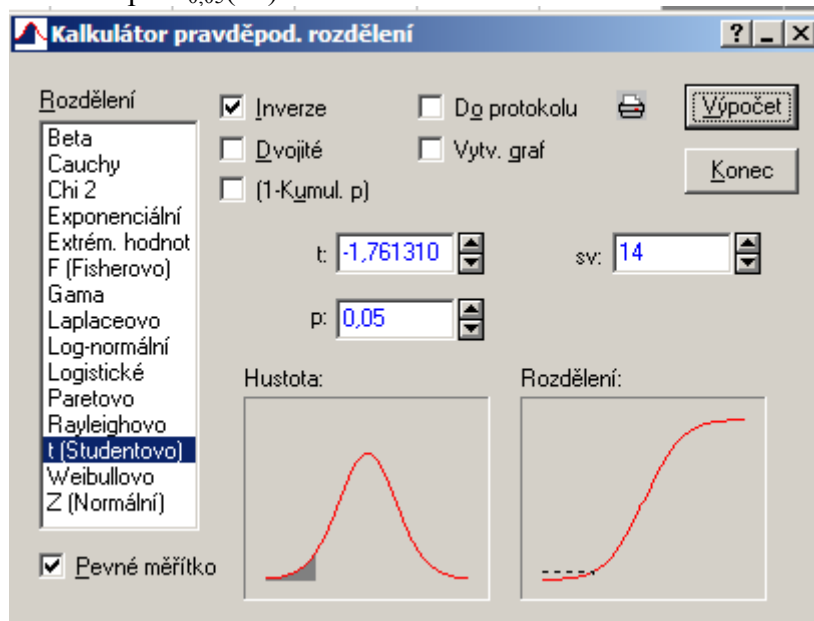
v příloze A skript Marie Budíkové, Pavel Osecký, Štěpán Mikoláš: Teorie pravděpodobnosti a matematická statistika. Sběrka příkladů. MU Brno 2007.

Příklad 4.: Určete $t_{0,99}(30)$ a $t_{0,05}(14)$.

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka sv. napíšeme 30 (resp. 14) a do okénka p napíšeme 0,99 (resp. 0,05). V okénku t se objeví 2,457262 (resp. -1,761310).

Ilustrace pro $t_{0,05}(14)$:



Šedá plocha pod grafem hustoty má velikost 0,05 a hodnota distribuční funkce v bodě -1,76131 je 0,05 (značeno šrafovaně).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případě.

Do dlouhého jména této proměnné napíšeme =VStudent(0,99;30) (resp. VStudent(0,05;14)).

Dostaneme 2,457262 (resp. -1,76131).

Fisherovo-Snedecorovo rozložení s n_1 a n_2 stupni volnosti $F(n_1, n_2)$

Nechť X_1, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \chi^2(n_i)$, $i = 1, 2$. Pak

náhodná veličina $X = \frac{X_1 / n_1}{X_2 / n_2} \sim F(n_1, n_2)$. Vyjádření hustoty je příliš složité, lze ho najít např.

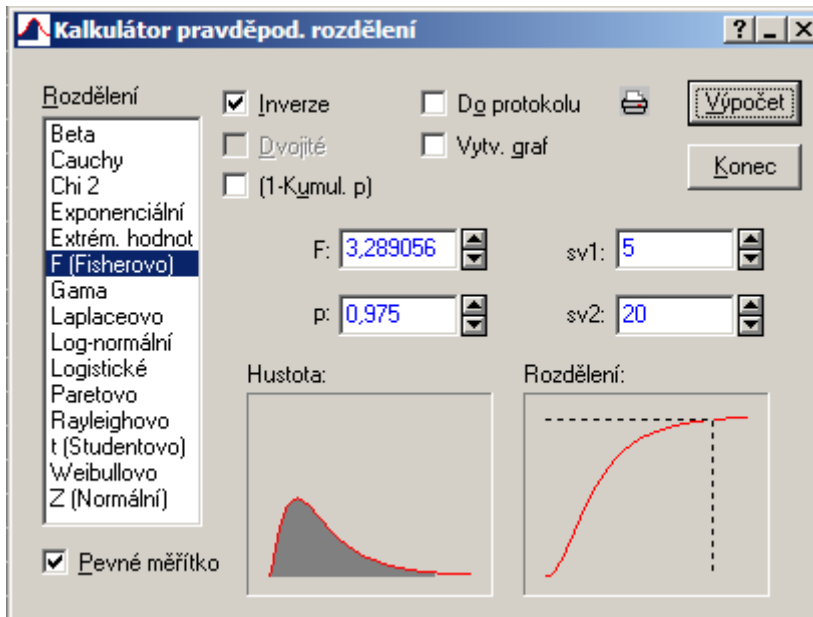
v příloze A skript Marie Budíková, Pavel Osecký, Štěpán Mikoláš: Teorie pravděpodobnosti a matematická statistika. Sbíрка příkladů. MU Brno 2007.

Příklad 5.: Určete $F_{0,975}(5, 20)$ a $F_{0,05}(2, 10)$.

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka sv1 napíšeme 5 (resp. 2), do okénka sv2 napíšeme 20 (resp. 10) a do okénka p napíšeme 0,975 (resp. 0,05). V okénku F se objeví 3,289056 (resp. 0,05156).

Ilustrace pro $F_{0,975}(5, 20)$:



Šedá plocha pod grafem hustoty má velikost 0,975 a hodnota distribuční funkce v bodě 3,289056 je 0,975 (značeno šrafovane).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a dvou případech. Do dlouhého jména první proměnné napíšeme =VF(0,975;5;20), do dlouhého jména druhé proměnné napíšeme =VF(0,05;2;10). Dostaneme 3,2891 (resp. 0,05156).

Příklad 6.: 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. Vypočítejte střední hodnotu a rozptyl náhodné veličiny X .

Řešení:

X nabývá hodnot 1, 2, 3, 4 a její pravděpodobnostní funkce je $\pi(1) = 0,2$, $\pi(2) = 0,8 \cdot 0,2 = 0,16$, $\pi(3) = 0,82 \cdot 0,2 = 0,128$, $\pi(4) = 0,83 \cdot 0,2 + 0,84 = 0,512$, $\pi(0) = 0$ jinak

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

$D(X) = 12 \cdot 0,2 + 22 \cdot 0,16 + 32 \cdot 0,128 + 42 \cdot 0,512 - 2,952^2 = 1,4697$

Postup ve STATISTICE:

Otevřeme nový datový soubor o dvou proměnných X a četnost a čtyřech případech. Do proměnné X napíšeme 1, 2, 3, 4, do proměnné četnost napíšeme 200, 160, 128, 512.

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

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

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

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

Příklad 7.: Náhodná veličina X udává počet ok při hodu kostkou. Pomocí systému STATISTICA vypočtete její střední hodnotu a rozptyl.

Výsledek: $E(X) = 3,5$, $D(X) = 2,9167$

Příklad 8.: 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. Vypočtete 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, $\pi_2(10) = 0,3$, $\pi_2(20) = 0,45$, $\pi_2(30) = 0,2$, $\pi_2(40) = 0,05$, $\pi_2(y) = 0$ 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:

Vytvoříme nový datový soubor o třech proměnných X, Y, četnost 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 četnost 20, 4, 1, 0, 10, 36, 9, 0, 0, 5, 10, 0, 0, 0, 0, 5.

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

Proměnná	Korelace (Tabulka6)			
	Průměry	Sm.odch.	X	Y
X	20,00000	7,784989	1,000000	0,756086
Y	20,00000	8,408750	0,756086	1,000000

Příklad 9.: 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 vypočtete $R(X_1, X_2)$.

Výsledek: $c = 0,1$, $R(X_1, X_2) = 0,42379$.