Téma 5.: Hustoty a distribuční funkce v systému STATISTICA, výpočet kvantilů

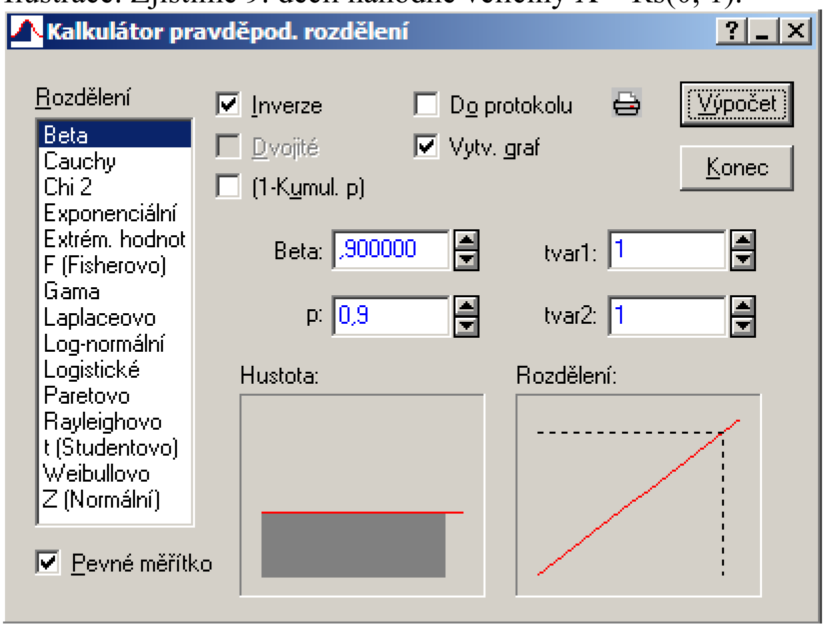
Systém STATISTICA vytváří grafy hustot a distribučních funkcí mnoha spojitých rozložení, počítá kvantily těchto rozložení a pro daný kvantil umí stanovit hodnotu distribuční funkce či počítat 1 - hodnota distribuční funkce. Slouží k tomu Pravděpodobnostní kalkulátor v menu Statistiky. Kvantily či hodnoty distribučních funkcí lze počítat též pomocí funkcí implementovaných v položce „Dlouhé jméno" proměnné – viz dále.
Zaměříme se na rovnoměrné spojité rozložení, normální rozložení a rozložení z něj odvozená.

Rovnoměrné spojité rozložení Rs(a, b)
Hustota pravděpodobnosti náhodné veličiny X je konstantní na intervalu (a, b) a plocha pod křivkou hustoty tvoří obdélník.
Píšeme X ~ Rs(a, b).

$$\varphi(x)=\begin{cases}\dfrac{1}{b-a} \text{ pro } x\in(a,b)\\ 0 \text{ jinak}\end{cases},\quad \Phi(x)=\begin{cases}0 \text{ pro } x\le a\\ \displaystyle\int_a^x \frac{1}{b-a}\,dt=\frac{x-a}{b-a} \text{ pro } x\in(a,b)\\ 1 \text{ pro } x\ge b\end{cases}$$

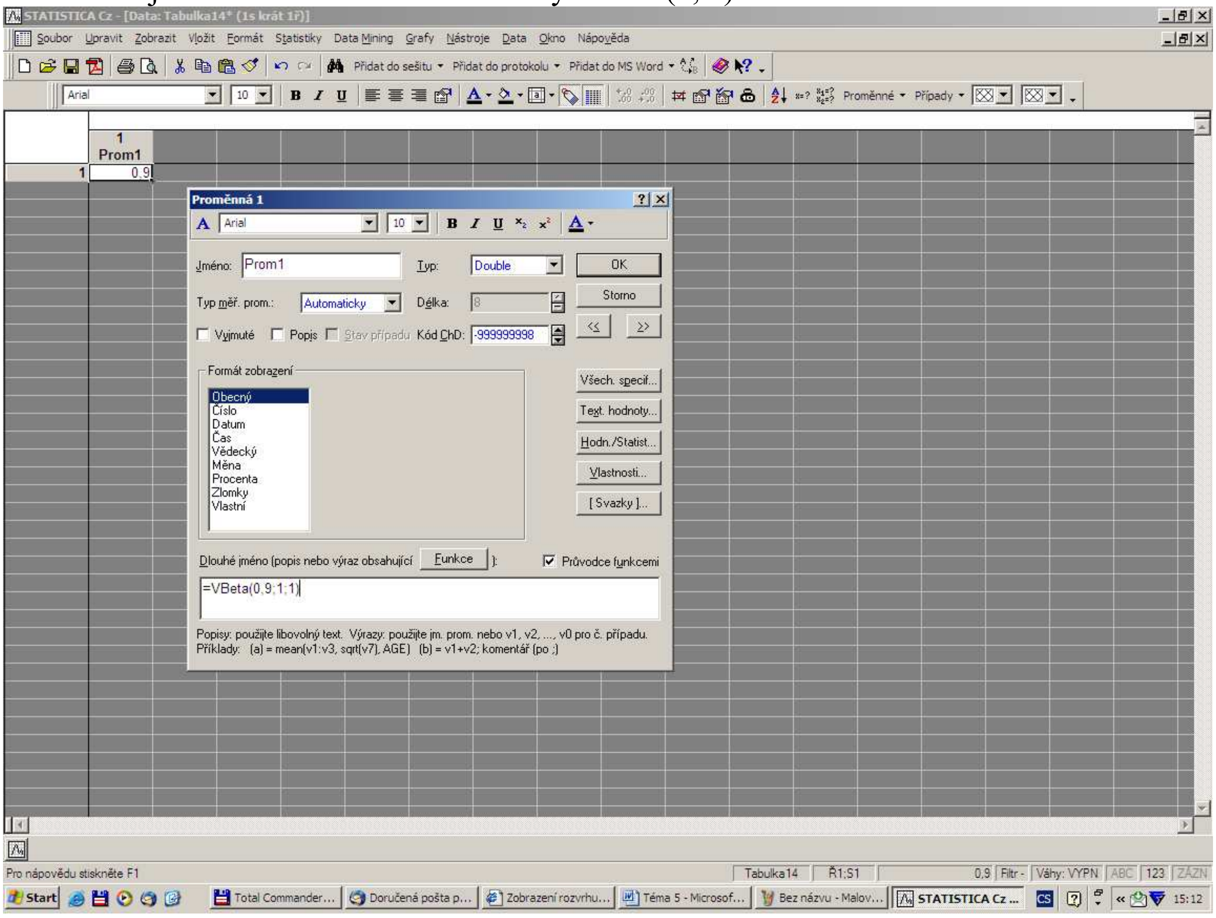
STATISTICA umí pracovat pouze s rozložením Rs(0,1), které je speciálním případem beta rozložení s parametry 1, 1. (Poučení o beta rozložení – viz např. Jiří Anděl: Matematická statistika. SNTL/ALFA, Praha 1978.). Náhodnou veličinu X ~ Rs(a, b) musíme transformovat na náhodnou veličinu Y ~ Rs(0, 1) pomocí vztahu: $Y=\dfrac{X-a}{b-a}$.

Použití systému STATISTICA:
První možnost: Statistiky – Pravděpodobnostní kalkulátor – Rozdělení – Beta – tvar 1 - napíšeme 1, tvar 2 – napíšeme 1. STATISTICA vykreslí graf hustoty a distribuční funkce rozložení Rs(0,1). Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Výpočet se v okénku Beta objeví hodnota tohoto kvantilu. Pokud zaškrtneme volbu Vytv. graf a klikneme na Výpočet, dostaneme v okně grafů graf hustoty a distribuční funkce.
Ilustrace: Zjistíme 9. decil náhodné veličiny X ~ Rs(0, 1).



V okénku Beta vidíme, že 9. decil je 0,9. Šedá plocha pod grafem hustoty má velikost 0,9 a hodnota distribuční funkce v bodě 0,9 je 0,9 (značeno šrafovaně).

Druhá možnost: Výpočet kvantilu či hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V položce „Dlouhé jméno“ této proměnné použijeme funkci VBeta(x;1;1), kde x odpovídá tomu $\alpha \in (0{,}1)$, pro které chceme kvantil počítat. Pro výpočet hodnoty distribuční funkce v bodě x použijeme funkci IBeta(x;1;1).
Ilustrace: Zjistíme 9. decil náhodné veličiny X ~ Rs(0, 1).



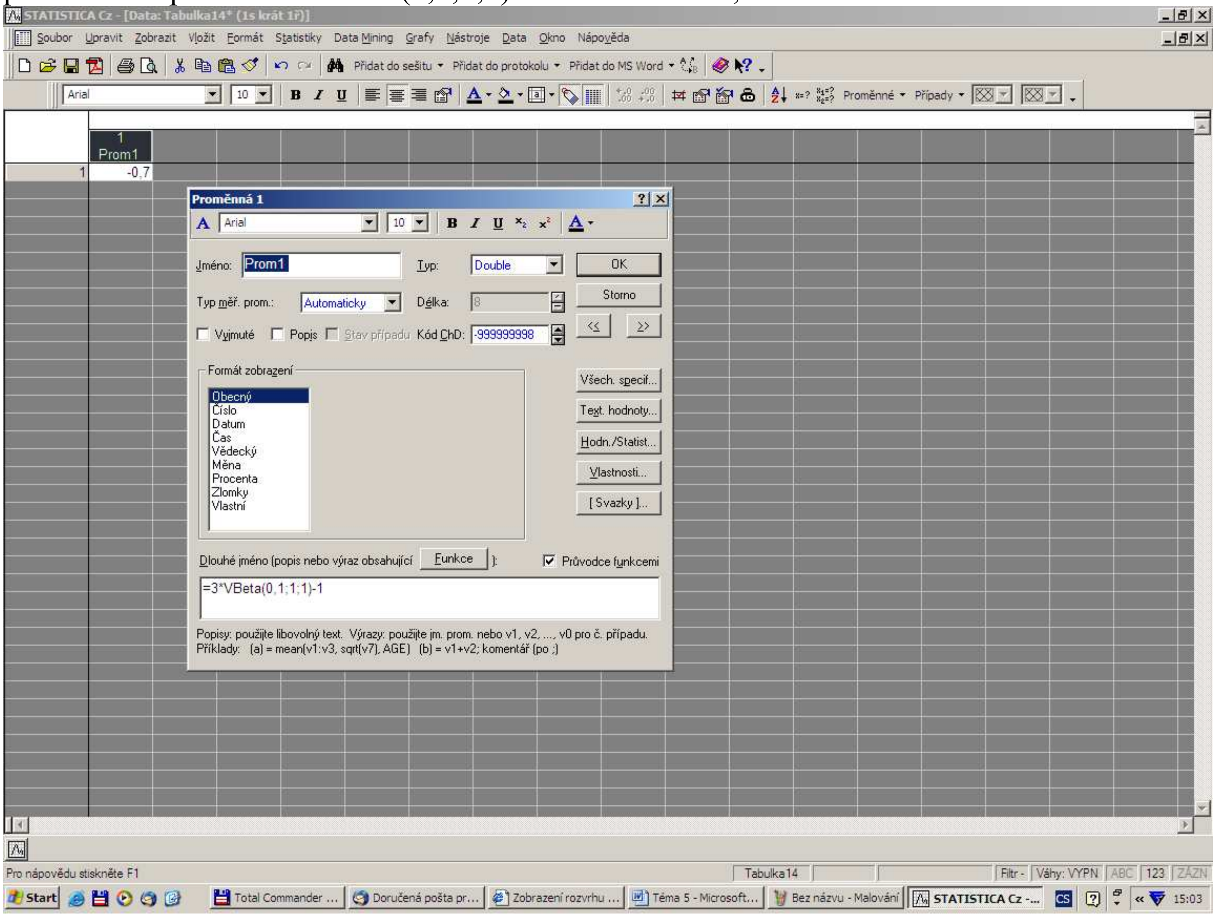
Příklad 1.: Stanovte 1. decil náhodné veličiny X ~ Rs(-1, 2).
Řešení:

$$\varphi(x) = \begin{cases} \frac{1}{3} \text{ pro } x \in (-1{,}2) \\ 0 \text{ jinak} \end{cases}, \quad \Phi(x) = \begin{cases} 0 \text{ pro } x \le -1 \\ \int_{-1}^{x} \frac{1}{3}\,dt = \frac{x+1}{3} \text{ pro } x \in (-1{,}2) \\ 1 \text{ pro } x \ge 2 \end{cases}$$

Pro 1. decil $K_{0,10}(X)$ platí: $0{,}10 = \Phi\left(K_{0,10}(X)\right) = \frac{K_{0,10}(X)+1}{3} \Rightarrow K_{0,10}(X) = 3 \cdot 0{,}10 - 1 = -0{,}7$.

Návod na výpočet pomocí systému STATISTICA: Náhodnou veličinu X ~ Rs(-1, 2) transformujeme na náhodnou veličinu Y ~ Rs(0, 1): $Y = \frac{X+1}{3}$. Pro 1. decil $K_{0,10}(X)$ platí: $K_{0,10}(X) = 3K_{0,10}(Y) - 1$.

Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do dlouhého jména této proměnné napíšeme =3*VBeta(0,1;1;1)-1. Dostaneme -0,7.



Příklad 2.: Na automatické lince se plní láhve mlékem. Působením náhodných vlivů množství mléka kolísá v intervalu(980 ml, 1020 ml). Každé množství mléka v tomto intervalu považujeme za stejně možné. Jaká je pravděpodobnost, že v náhodně vybrané láhvi bude aspoň 1010 ml mléka?

Řešení:

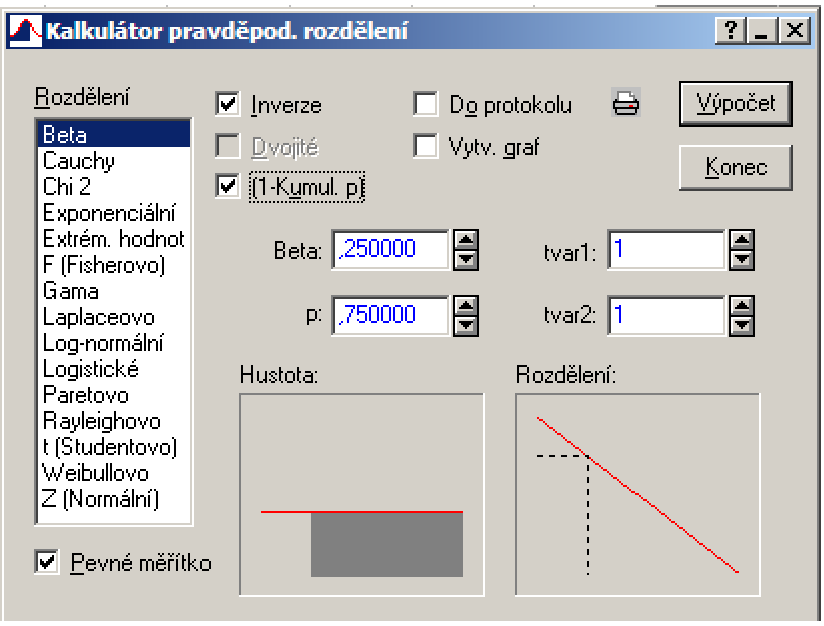
X – množství mléka v náhodně vybrané láhvi, X ~ Rs(980, 1020),

$$\varphi(x) = \begin{cases} \frac{1}{40} \text{ pro } x \in (980, 1020) \\ 0 \text{ jinak} \end{cases}, \quad P(X \geq 1010) = \int_{1010}^{1020} \frac{1}{40} dx = \frac{1}{40}[x]_{1010}^{1020} = \frac{10}{40} = 0{,}25$$

Návod na výpočet pomocí systému STATISTICA: Abychom mohli použít systém STATISTICA, musíme náhodnou veličinu X ~ Rs(980, 1020) transformovat na náhodnou veličinu Y ~ Rs(0, 1): $Y = \frac{X-980}{40}$. Pak

$$P(X \geq 1010) = P\left(\frac{X-980}{40} \geq \frac{1010-980}{40}\right) = P(Y \geq 0{,}75) = \int_{0{,}75}^{1} dy = 0{,}25$$

První možnost: Statistiky – Pravděpodobnostní kalkulátor – Rozdělení – Beta – tvar 1 - napíšeme 1, tvar 2 – napíšeme 1, do okénka Beta napíšeme 0,75, zaškrtneme 1-Kumul. p a v okénku p se objeví 0,25.



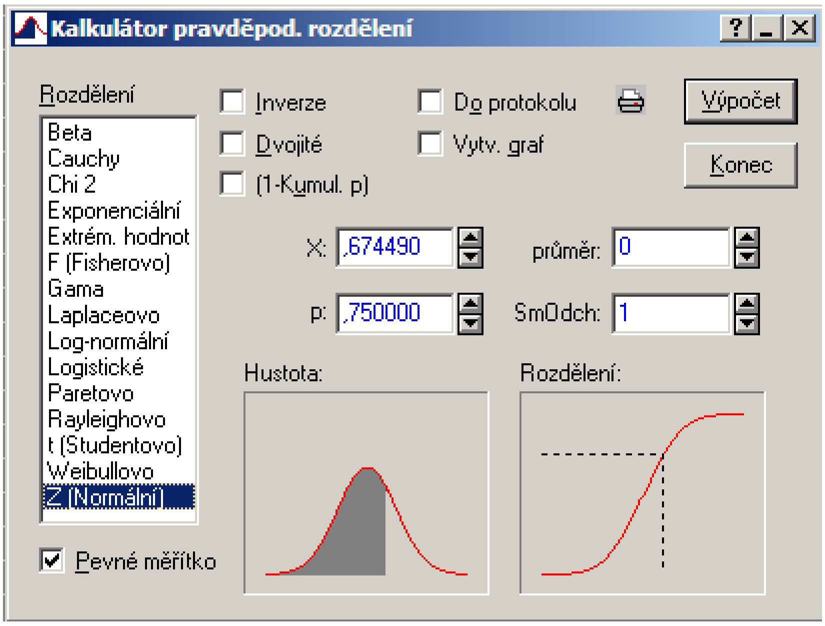
Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do dlouhého jména této proměnné napíšeme =1-IBeta(0,75;1;1). Dostaneme výsledek 0,25.

Normální rozložení N(μ, σ^2)

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

Použití systému STATISTICA:
První možnost: Ve volbě Rozdělení vybereme Z (Normální), do okénka průměr napíšeme hodnotu μ a do okénka Sm. Odch. napíšeme hodnotu σ. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Výpočet se v okénku X objeví hodnota tohoto kvantilu.
Druhá možnost: Výpočet kvantilu či hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V položce „Dlouhé jméno“ této proměnné použijeme funkci VNormal(x;mu;sigma), kde x odpovídá tomu $\alpha\in(0,1)$, pro které chceme kvantil počítat. Pro výpočet hodnoty distribuční funkce v bodě x použijeme funkci INormal(x;mu;sigma).

Příklad 3.: Nechť U ~ 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 případu.
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 4.: 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 případu.
Do dlouhého jména této proměnné napíšeme =VNormal(0,25;3;sqrt(5)). Dostaneme 1,491795.

Příklad 5.: Výsledky u přijímacích zkoušek na jistou VŠ jsou normálně rozloženy s parametry $\mu = 550$ bodů, $\sigma = 100$ bodů. S jakou pravděpodobností bude mít náhodně vybraný uchazeč aspoň 600 bodů?
Řešení:
X – výsledek náhodně vybraného uchazeče, $X \sim N(550, 100^2)$, $P(X \geq 600) = 1 - P(X \leq 600) + P(X = 600) = 1 - P(X \leq 600) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{600-\mu}{\sigma}\right) = 1 - P\left(U \leq \frac{600-550}{100}\right) = 1 - \Phi(0,5)$
$= 1 - 0,69146 = 0,30854$.
Návod na výpočet pomocí systému STATISTICA:
První možnost: Do okénka průměr napíšeme 550, do okénka Sm. Odch. napíšeme 100, do okénka X napíšeme 600, zaškrtneme 1-Kumul. p a v okénku p se objeví 0,308538.
Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.
Do dlouhého jména této proměnné napíšeme =1-INormal(600;550;100). Dostaneme 0,3085.

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

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

Použití systému STATISTICA:
První možnost: Ve volbě Rozdělení vybereme Chi 2 a do okénka sv. napíšeme patřičný počet stupňů volnosti. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Výpočet se v okénku Chi 2 objeví hodnota tohoto kvantilu.
Druhá možnost: Výpočet kvantilu či hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V položce „Dlouhé jméno“ této proměnné použijeme funkci VChi2(x;ný), kde x odpovídá tomu $\alpha \in (0,1)$, pro které chceme kvantil počítat a ný je počet stupňů volnosti. Pro výpočet hodnoty distribuční funkce v bodě x použijeme funkci IChi2(x;ný).

Příklad 6.: 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).

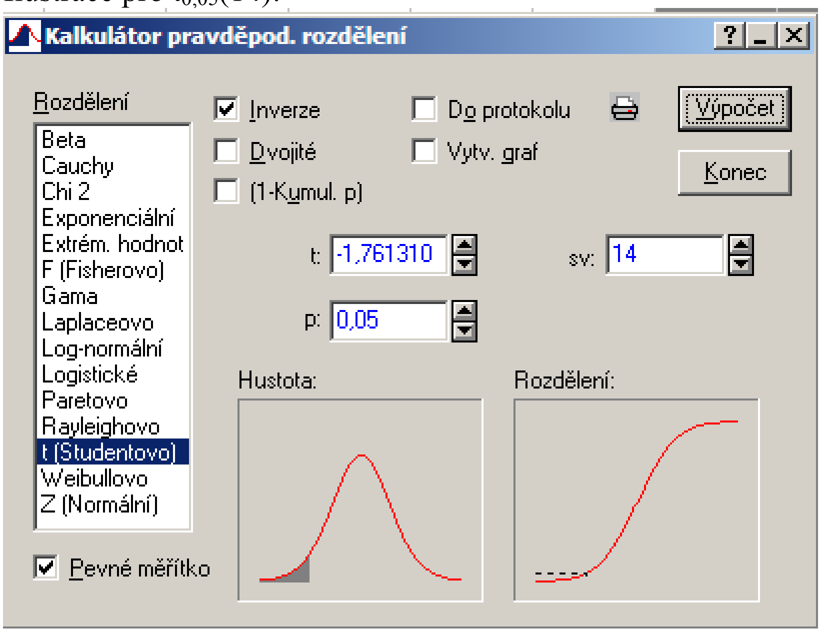
Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do dlouhého jména této proměnné napíšeme =VChi2(0,025;25). Dostaneme 13,1197.

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 $X = \dfrac{X_1}{\sqrt{\dfrac{X_2}{n}}} \sim t(n)$. Vyjádření hustoty je příliš složité, lze ho najít např.

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

Použití systému STATISTICA:
První možnost: Ve volbě Rozdělení vybereme t (Studentovo) a do okénka sv napíšeme patřičný počet stupňů volnosti. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Výpočet se v okénku t objeví hodnota tohoto kvantilu.
Druhá možnost: Výpočet kvantilu či hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V položce „Dlouhé jméno“ této proměnné použijeme funkci VStudent(x;sv), kde x odpovídá tomu $\alpha \in (0,1)$, pro které chceme kvantil počítat a sv je počet stupňů volnosti. Pro výpočet hodnoty distribuční funkce v bodě x použijeme funkci IStudent(x;sv).

Příklad 7.: Určete $t_{0,99}(30)$ a $t_{0,05}(14)$.
Návod na výpočet pomocí systému STATISTICA:
První možnost: Do okénka sv. napíšeme 25 (resp. 14) a do okénka p napíšeme 0,99 (resp. 0,05). V okénku t se objeví 2,457262 (resp. -1,761310).
Ilustrace pro $t_{0,05}(14)$:



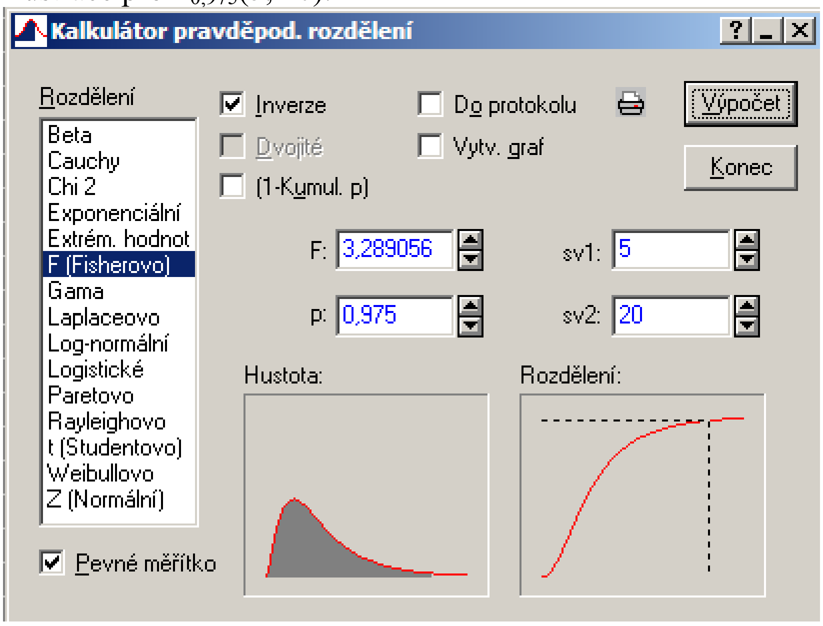
Šedá plocha pod grafem hustoty má velikost 0,05 a hodnota distribuční funkce v bodě -1,76131 je 0,05 (značeno šrafovaně).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.
Do dlouhého jména této proměnné napíšeme =VStudent(0,99;30) (resp. VStudent(0,05;14)).
Dostaneme 13,1197 (resp. -1,76131).

Fisherovo-Snedecorovo rozložení s n_1 a n_2 stupni volnosti $F(n_1, n_2)$
Nechť $X_1, ..., X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \chi^2(n_i)$, i = 1, 2. Pak náhodná veličina $X = \dfrac{X_1 / n_1}{X_2 / n_2} \sim F(n_1, n_2)$. Vyjádření hustoty je příliš složité, lze ho najít např. v příloze A skript Marie Budíková, Pavel Osecký, Štěpán Mikoláš: Teorie pravděpodobnosti a matematická statitika. Sbírka příkladů. MU Brno 2007.

Použití systému STATISTICA:
První možnost: Ve volbě Rozdělení vybereme F (Fisherovo), do okénka sv1 napíšeme počet stupňů volnosti čitatele a do okénka sv2 počet stupňů volnosti jmenovatele. Hodnotu α-kvantilu zjistíme tak, že do okénka označeného p napíšeme dané α a po kliknutí na Výpočet se v okénku F objeví hodnota tohoto kvantilu.
Druhá možnost: Výpočet kvantilu či hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případu. V položce „Dlouhé jméno“ této proměnné použijeme funkci VF(x;ný;omega), kde x odpovídá tomu $\alpha \in (0,1)$, pro které chceme kvantil počítat, ný je počet stupňů volnosti čitatele a omega je počet stupňů volnosti jmenovatele. Pro výpočet hodnoty distribuční funkce v bodě x použijeme funkci IF(x;ný;omega).

Příklad 8.: 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 šrafovaně).

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).