Téma 4.: Pravděpodobnostní funkce, hustoty a distribuční funkce v systému STATISTICA, výpočet pravděpodobností pomocí distribučních funkcí

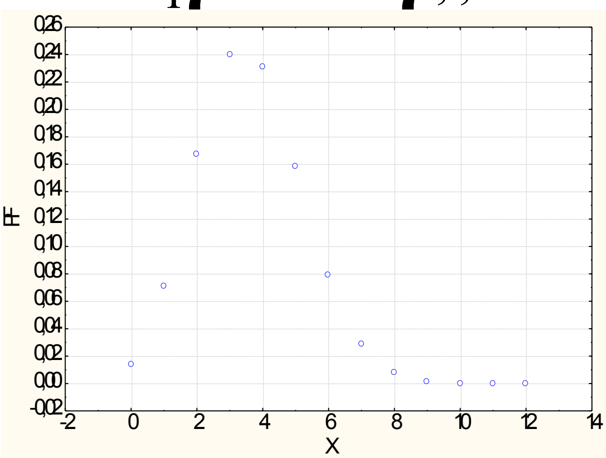
Systém STATISTICA vytváří grafy hustot a distribučních funkcí mnoha spojitých rozložení, umí stanovit hodnotu distribuční funkce či počítat 1 - hodnota distribuční funkce. Slouží k tomu Pravděpodobnostní kalkulátor v menu Statistiky. Hodnoty pravděpodobnostních funkcí, hustot a distribučních funkcí lze počítat též pomocí funkcí implementovaných v položce „Dlouhé jméno“ proměnné.
Zaměříme se na binomické rozložení, Poissonovo rozložení, rovnoměrné spojité rozložení, exponenciální rozložení a normální rozložení.

Binomické rozložení Bi(n, ϑ)
Náhodná veličina X udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ. Píšeme X ~ Bi(n, ϑ).

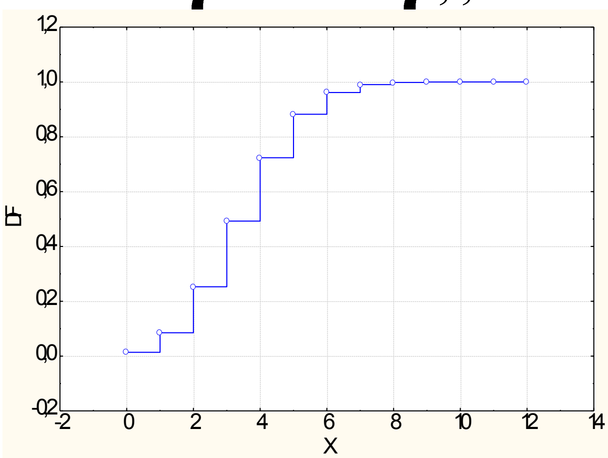
$$\pi(x) = \begin{cases} \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} & \text{pro } x = 0, \ldots, n \\ 0 & \text{jinak} \end{cases}, \quad \Phi(x) = \sum_{t=0}^{x}\binom{n}{t}\vartheta^t(1-\vartheta)^{n-t}$$

Kreslení grafů funkcí π a Φ v systému STATISTICA
Ukážeme si, jak získat grafy pravděpodobnostní a distribuční funkce náhodné veličiny $X \sim Bi(12; 0,3)$. Vytvoříme nový datový soubor o 3 proměnných a 13 případech. První proměnnou nazveme X a uložíme do ní hodnoty 0, 1, ..., 12 (do Dlouhého jména napíšeme =v0-1). Druhou proměnnou nazveme PF a uložíme do ní hodnoty pravděpodobnostní funkce (do Dlouhého jména napíšeme příkaz =Binom(x;0,3;12)). Třetí proměnnou nazveme DF a uložíme do ní hodnoty distribuční funkce (do Dlouhého jména napíšeme příkaz =IBinom(x;0,3;12)).
Graf pravděpodobnostní funkce: Grafy – Bodové grafy – Proměnné X, PF – OK – vypneme Lineární proložení – OK.
Graf distribuční funkce: Grafy – Bodové grafy – Proměnné X, DF – OK – vypneme Lineární proložení – OK – 2x klikneme na pozadí grafu – Graf:Obecné – zaškrtneme Spojnice – Typ spojnice: Schod – OK.

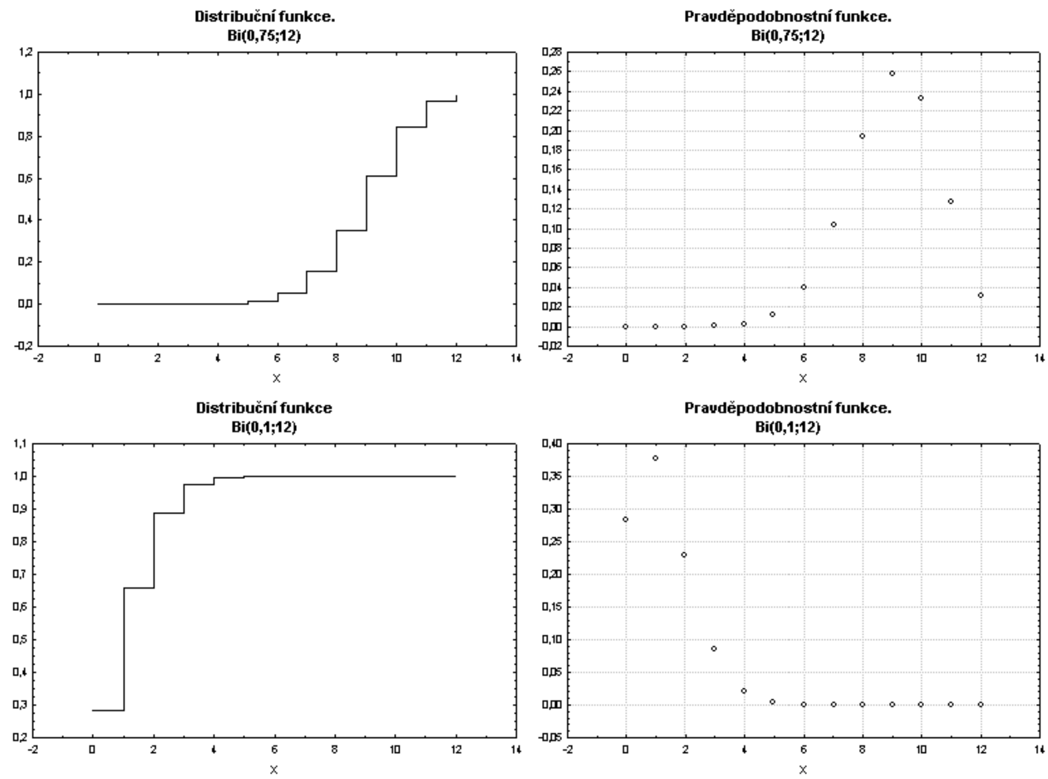
Graf funkce $p(x)$ rozložení $Bi(12; 0,3)$



Graf funkce $F(x)$ rozložení $Bi(12; 0,3)$



Analogickým způsobem můžeme získat grafy pravděpodobnostních distribučních funkcí binomického rozložení pro různá n a ϑ a sledovat vliv těchto parametrů na vzhled grafů.



Poissonovo rozložení Po(λ)

Náhodná veličina X udává počet událostí, které nastanou v jednotkovém časovém intervalu (resp. v jednotkové oblasti), přičemž k událostem dochází náhodně, jednotlivě a vzájemně nezávisle. Parametr λ je střední počet těchto událostí. Píšeme $X \sim Po(\lambda)$.

$$\pi(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{pro } x = 0, 1, \ldots \\ 0 & \text{jinak} \end{cases}, \quad \Phi(x) = \sum_{t \le x} \frac{\lambda^t}{t!} e^{-\lambda}$$

Kreslení grafů funkcí π a Φ v systému STATISTICA

Při tvorbě grafů pravděpodobnostní a distribuční funkce náhodné veličiny s Poissonovým rozložením, např. $X \sim Po(5)$, postupujeme podobně jako u binomického rozložení, ale v datovém souboru bude 16 případů a použijeme funkce Poisson(x;5) a IPoisson(x;5).

Příklad 1.: Při provozu balicího automatu vznikají během směny náhodné poruchy, které se řídí rozložením Po(2). Jaká je pravděpodobnost, že během směny dojde k aspoň jedné poruše?
Řešení: X – počet poruch během směny, X ~ Po(2), $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) =$
$= 1 - \frac{2^0}{0!} e^{-2} = 0{,}8647$.

Návod na výpočet pomocí systému STATISTICA: Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do Dlouhého jména této proměnné napíšeme =1-IPoisson(0;2). Dostaneme výsledek 0,8647.

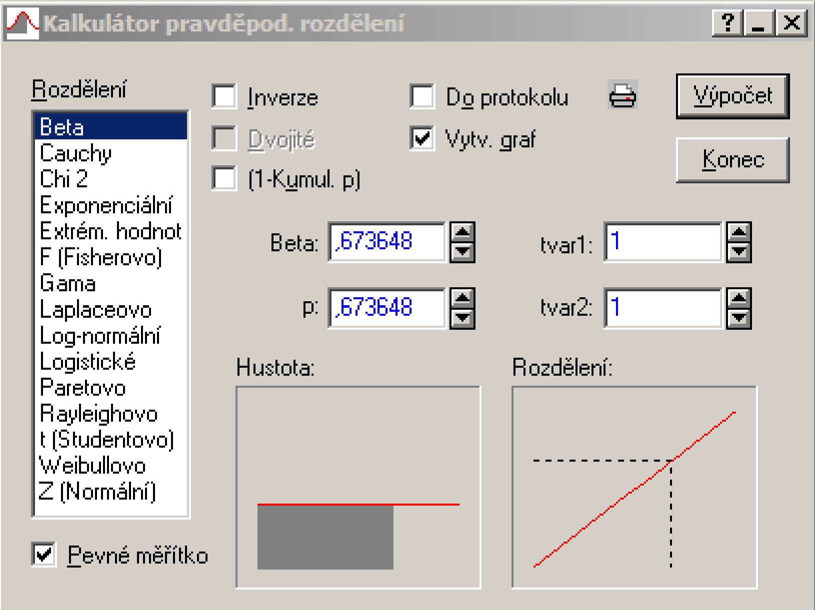
Rovnoměrné spojité rozložení Rs(a, b)

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

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

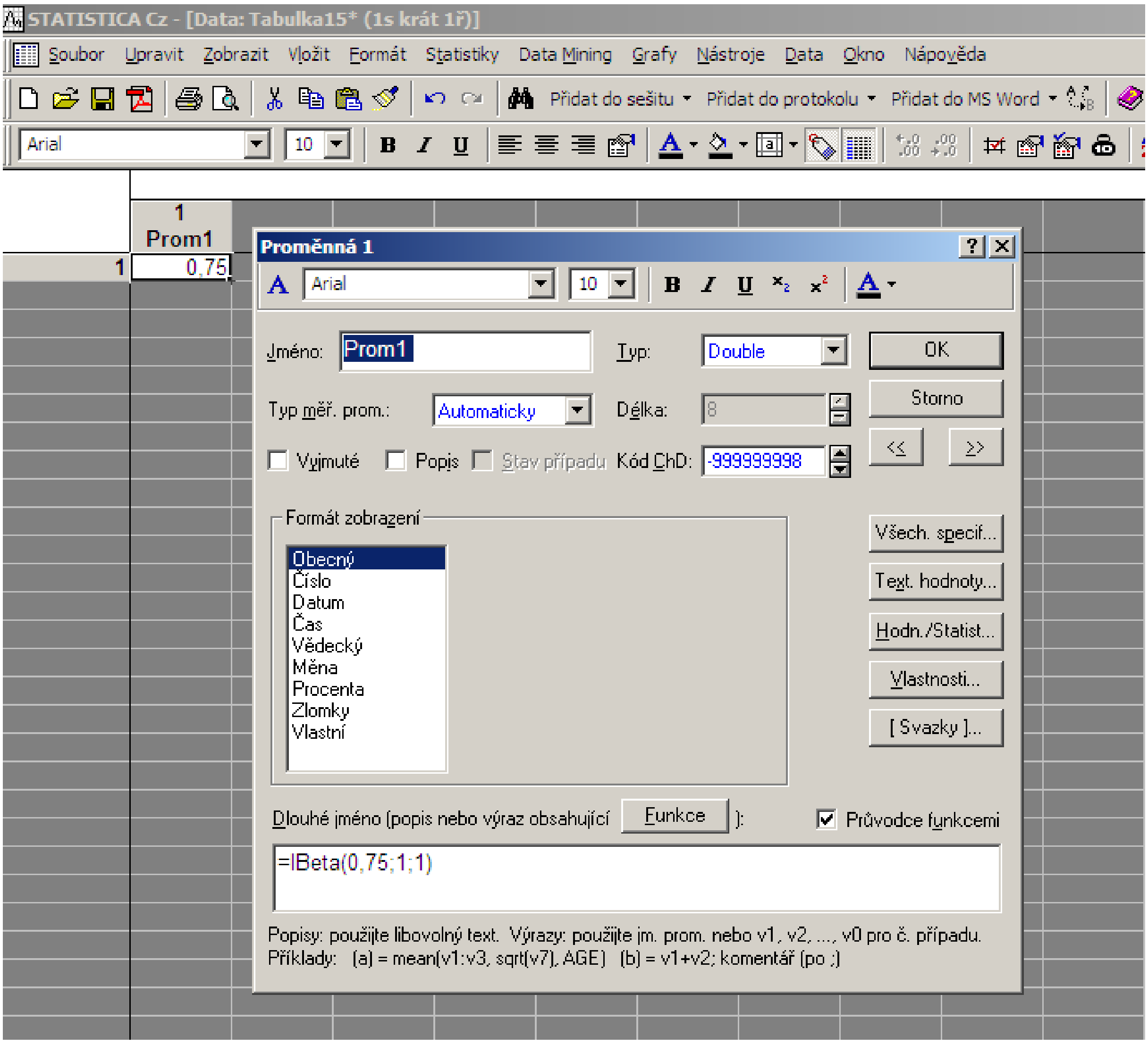
STATISTICA umí pracovat pouze s rozložením Rs(0,1), které je speciálním případem beta rozložení s parametry 1, 1. (Poučení o beta rozložení – viz např. Jiří Anděl: Matematická statistika. SNTL/ALFA, Praha 1978.). Náhodnou veličinu X ~ Rs(a, b) musíme transformovat na náhodnou veličinu Y ~ Rs(0, 1) pomocí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

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). Pokud zaškrtneme volbu Vytv. graf a klikneme na Výpočet, dostaneme v okně grafů graf hustoty a distribuční funkce.
Ilustrace: Vytvoření grafů hustoty a distribuční funkce náhodné veličiny X ~ Rs(0, 1).



Šedá plocha pod grafem hustoty má velikost 0,673648 (implicitní volba, lze samozřejmě měnit od do 1), hodnota distribuční funkce v bodě 0,673648 je 0,673648 (značeno šrafovaně).

Druhá možnost: Výpočet 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 IBeta(x;1;1), kde $x \in \langle 0, 1 \rangle$ je argument distribuční funkce.
Ilustrace: Zjistíme hodnotu distribuční funkce náhodné veličiny X ~ Rs(0, 1) v bodě 0,75.



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:

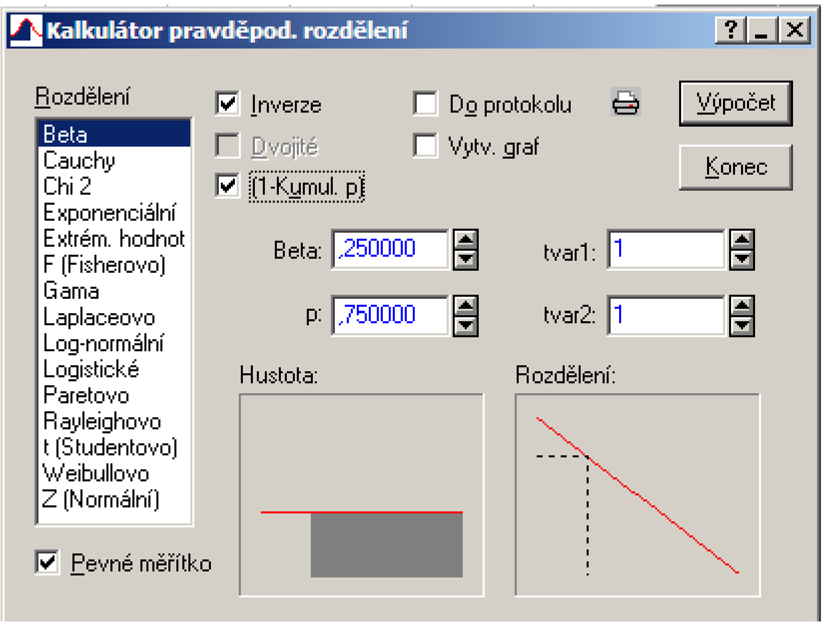
X – množství mléka v náhodně vybrané láhvi, X ~ Rs(980, 1020),

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

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.

Exponenciální rozložení Ex(λ)

Náhodná veličina X udává dobu čekání na příchod nějaké události, která se může dostavit každým okamžikem se stejnou šancí bez ohledu na dosud pročekanou dobu. (Jde o tzv. čekání bez paměti.) Přitom $\frac{1}{\lambda}$ vyjadřuje střední dobu čekání. Náhodná veličina X ~ Ex(λ) má hustotu

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{jinak} \end{cases}.$$

Použití systému STATISTICA:

První možnost: Ve volbě Rozdělení vybereme Exponenciální, do okénka lambda napíšeme hodnotu parametru λ. Hodnotu distribuční funkce v bodě x zjistíme tak, že do okénka označeného X napíšeme dané x a po kliknutí na Výpočet se v okénku p objeví hodnota distriubuční funkce.

Druhá možnost: Výpočet 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 IExpon(x;lambda).

Příklad 3.: Doba do ukončení opravy v opravně obuvi je náhodná veličina, která se řídí exponenciálním rozložením se střední dobou opravy 3 dny. Jaká je pravděpodobnost, že oprava bude ukončena do dvou dnů?

Řešení: X ~ Ex(1/3), $P(X \le 2) = \int_0^2 \frac{1}{3} e^{-\frac{x}{3}} dx = \left[-e^{-\frac{x}{3}} \right]_0^2 = 1 - e^{-\frac{2}{3}} = 0,4866$

Návod na výpočet pomocí systému STATISTICA:

První možnost: Do okénka lambda napíšeme 0,3333, do okénka exp. napíšeme 2 a po kliknutí na Výpočet se v okénku p objeví 0,4866.

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu. Do dlouhého jména této proměnné napíšeme =IExpon(2;1/3). Dostaneme 0,4866.

Příklad 4.: Doba (v hodinách), která uplyne mezi dvěma naléhavými příjmy v jisté nemocnici, se řídí exponenciálním rozložením se střední dobou čekání 2 h. Jaká je pravděpodobnost, že uplyne více než 5 h bez naléhavého příjmu?
Výsledek: $X \sim Ex(1/2)$, $P(X > \; = \;)8$

Normální rozložení $N(\mu, \sigma^2)$

Náhodná veličina $X \sim N(\mu, \sigma^2)$ má hustotu $\varphi = \frac{1}{\ldots \pi} e^{-\frac{(x-\mu)}{\sigma}\ldots}$. 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 $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{\ldots}}$.

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 distribuční funkce v bodě x zjistíme tak, že do okénka označeného X napíšeme dané x a po kliknutí na Výpočet se v okénku p objeví hodnota disriubuční funkce.
Druhá možnost: Výpočet 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 INormal(x;mu;sigma).

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, $X \sim N(550, 100^2)$, $P(X \geq 600) = 1 - P(X \leq 600) + P(X = 600) = 1 - P(X \leq 600) = 1 - P\left(\frac{X - \ldots}{\sigma} \leq \frac{600 - \ldots}{\sigma}\right) = 1 - P\left(U \leq \frac{600 - 550}{100}\right) = 1 - \Phi(0{,}5)$ $= 1 - 0{,}69146 = 0{,}30854$.
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.

Příklad 6: Životnost baterie v hodinách je náhodná veličina, která má normální rozložení se střední hodnotou 300 hodin a směrodatnou odchylkou 35 hodin. Jaká je pravděpodobnost, že náhodně vybraná baterie bude mít životnost
a) aspoň 320 hodin?
b) nejvýše 310 hodin?
Výsledek:
ad a) $P(X > 2\ldots = \ldots 284\ldots$
ad b) $P(X < 1\ldots = \ldots 512\ldots$

Příklad 7.: Na výrobní lince jsou automaticky baleny balíčky rýže o deklarované hmotnosti 1000 g. Působením náhodných vlivů hmotnost balíčků kolísá. Lze ji považovat za náhodnou veličinu, která se řídí normálním rozložením se střední hodnotou 996 g a směrodatnou odchylkou 18 g. Jaká je pravděpodobnost, že náhodně vybraný balíček rýže neprojde výstupní kontrolou, jestliže je povolená tolerance $\pm$ [illegible] g od deklarované hmotnosti 1000 g?

Výsledek:

$$P(X \notin (970, 1030)) = 1 - P(970 < X < 1030) = 0{,}10\ldots$$