

Cvičení 5.: Bayesův vzorec, příklady na normální rozložení, výpočet číselných charakteristik

Bayesův vzorec

Jevy H_1, \dots, H_n tvoří úplný systém hypotéz, tj. navzájem se vylučují a přitom vyčerpávají všechny možnosti. Jev A s hypotézami nesouvisí.

Počítáme pravděpodobnost některé hypotézy za podmínky, že nastal jev A :

$$P(H_k/A) = \frac{P(H_k)P(A/H_k)}{P(A)}, \quad k = 1, \dots, n, \quad \text{kde } P(A) = \sum_{i=1}^n P(H_i)P(A/H_i).$$

Vzorový příklad: U jistého druhu elektrického spotřebiče se s pravděpodobností 0,01 vyskytuje výrobní vada. U spotřebiče s touto výrobní vadou dochází v záruční lhůtě k poruše s pravděpodobností 0,5. Výrobky, které tuto vadu nemají, se v záruční lhůtě porouchají s pravděpodobností 0,01. Jaká je pravděpodobnost, že výrobek, který se v záruční lhůtě porouchá, bude mít dotyčnou výrobní vadu?

Řešení:

H_1 - výrobek má dotyčnou výrobní vadu

H_2 - výrobek nemá tuto výrobní vadu

A - výrobek se v záruční době porouchá

Pak je: $P(H_1) = 0,01$, $P(H_2) = 0,99$, $P(A/H_1) = 0,5$, $P(A/H_2) = 0,01$

$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) = 0,01 \cdot 0,5 + 0,99 \cdot 0,01 = 0,0149$

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0,01 \cdot 0,5}{0,0149} = 0,3386$$

Příklady k samostatnému řešení:

1. Ve společnosti je 45% mužů a 55% žen. Výšku nad 190 cm má 5% mužů a 1% žen.

Náhodně vybraná osoba je vyšší než 190 cm. Jaká je pravděpodobnost, že je to žena?

Návod: A ... osoba měří více než 190 cm, H_1 ... osoba je žena, H_2 ... osoba je muž. Počítáme $P(H_1/A)$.

Výsledek: 0,1964

2. Potřebu smrkových sazenic kryje lesní závod produkcí dvou školky. První školka kryje 75% výsadby, přičemž ze 100 sazenic je 80 první jakosti. Druhá školka kryje výsadbu z 25%, přičemž na 100 sazenic připadá 60 první jakosti. Jaká je pravděpodobnost, že náhodně vybraná sazenice první jakosti pochází z produkce první školky?

Návod: A ... sazenice je 1. jakosti, H_1 ... sazenice pochází z 1. školky, H_2 ... sazenice pochází z 2. školky. Počítáme $P(H_1/A)$.

Výsledek: 0,8

Příklady na normální rozložení

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

Použití systému STATISTICA pro výpočet distribuční funkce:

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 distribuční funkce.

Druhá možnost: Výpočet hodnoty distribuční funkce pomocí funkcí implementovaných v položce „Dlouhé jméno“: Otevřeme nový datový soubor o jedné proměnné a jednom případě. V položce „Dlouhé jméno“ této proměnné použijeme funkci INormal(x ;mu;sigma).

Příklad 1.: 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 případě. Do dlouhého jména této proměnné napíšeme =1-INormal(600;550;100). Dostaneme 0,3085.

Příklad 2: 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

- aspoň 320 hodin?
- nejvýše 310 hodin?

Výsledek:

ad a) $P(X > 320) = 0,28434$

ad b) $P(X \leq 310) = 0,61245$

Příklad 3.: 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 ± 30 g od deklarované hmotnosti 1000 g?

Výsledek:

$P(X \notin (970, 1030)) = 1 - P(970 < X < 1030) = 0,104$

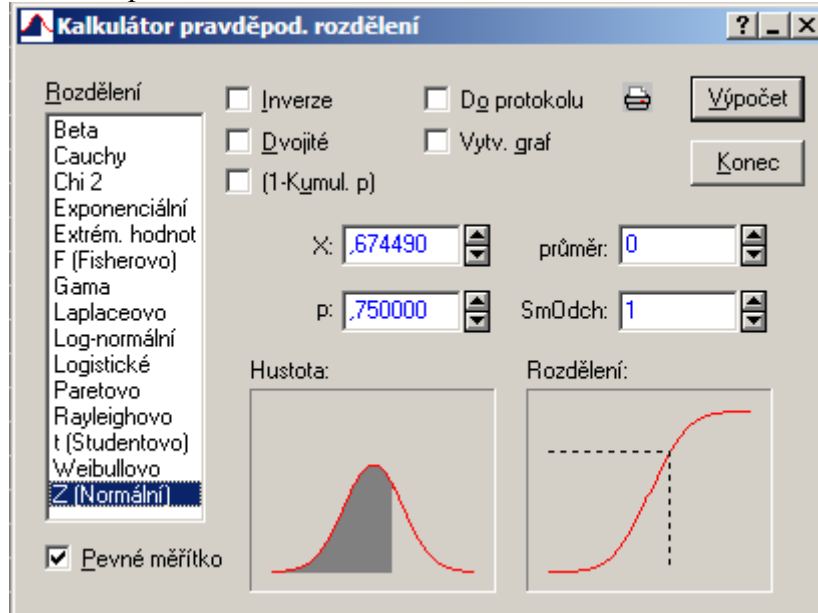
Výpočet kvantilů

Příklad 1.: Necht' $U \sim N(0, 1)$. 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.: Necht' $X \sim N(3, 5)$. 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.

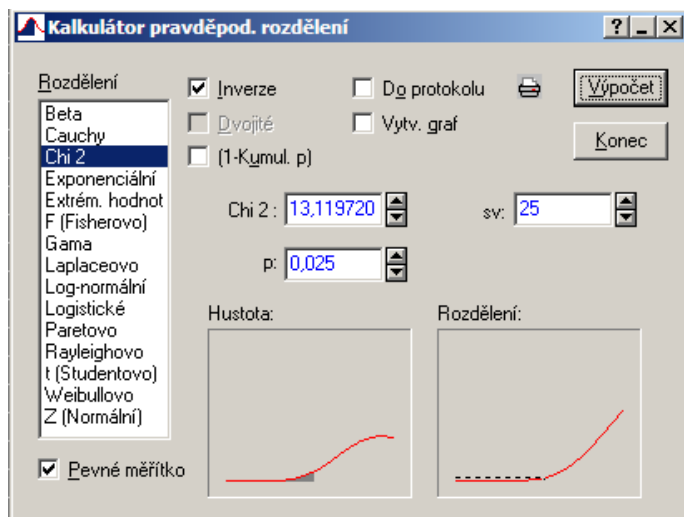
Pearsonovo rozložení chí-kvadrát s n stupni volnosti $\chi^2(n)$

Necht' X_1, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1)$, $i = 1, \dots, n$. Pak náhodná veličina $X = X_1^2 + \dots + X_n^2 \sim \chi^2(n)$.

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 šrafovane).

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

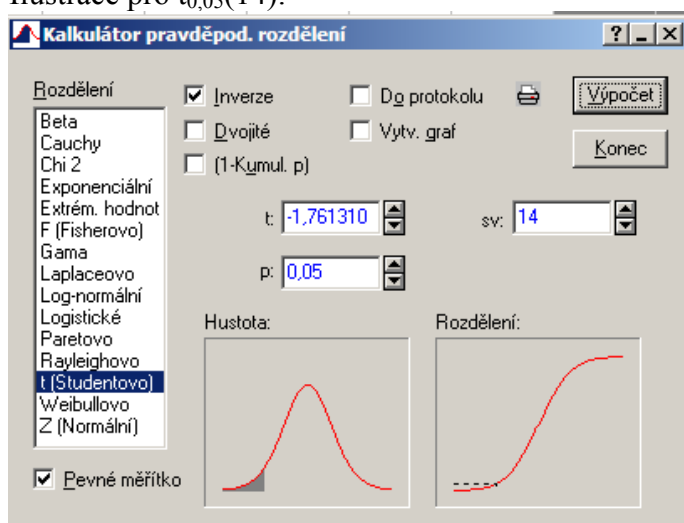
náhodná veličina $X = \frac{X_1}{\sqrt{\frac{X_2}{n}}} \sim t(n)$.

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 šrafovane).

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

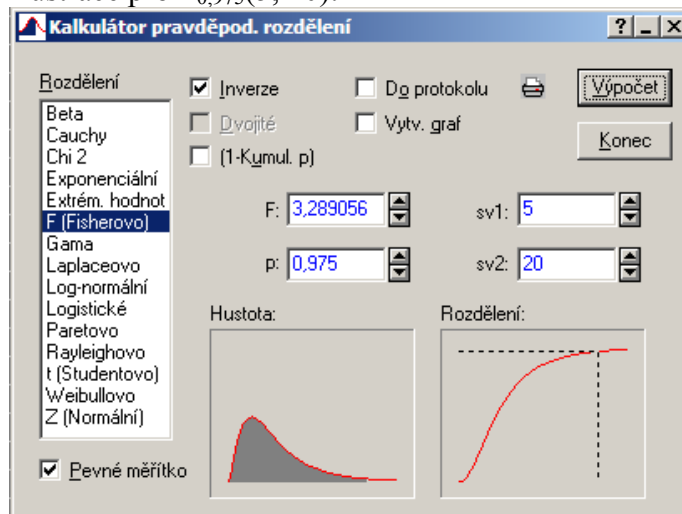
náhodná veličina $X = \frac{X_1 / n_1}{X_2 / n_2} \sim F(n_1, n_2)$.

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 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).

Výpočet střední hodnoty a rozptylu

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:

$$\pi(1) = 0,2,$$

$$\pi(2) = 0,8 \cdot 0,2 = 0,16,$$

$$\pi(3) = 0,82 \cdot 0,2 = 0,128,$$

$$\pi(4) = 0,83 \cdot 0,2 + 0,84 = 0,512,$$

$$\pi(0) = 0 \text{ jinak}$$

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

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

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 $=\sqrt{3} \cdot 999/1000$

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

Příklad 2.: Náhodná veličina X udává počet ok při hodu kostkou. Pomocí systému STATISTICA vypočítejte její střední hodnotu a rozptyl.

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

Příklad 3.: 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. Vypočítejte koeficient korelace příjmů manžela a manželky.

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