KGG/STG Statistika pro geografy

4. Teoretická rozdělení

Mgr. David Fiedor
9. března 2015

Osnova

Vybraná rozdělení náhodných proměnných

- normální rozdělení
- normované normální rozdělení
- binomické rozdělení
- Poissonovo rozdělení
- χ^2 rozdělení
- Studentovo t rozdělení
- Fisher-Snedecorovo rozdělení
- exponenciální, lognormální,...

Základní charakteristiky

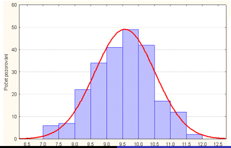
- Gaussovo rozdělení
- nejčastěji používané rozdělení spojité náhodné proměnné
- některé jiné proměnné lze převést (transformovat) na proměnnou, která má normální rozdělení
- existuje mnoho statistických procedur odvozených pro toto rozdělení
- u rozdělení s větším rozsahem lze rozdělení výběrových průměrů aproximovat normálním rozdělením
- zvonovitý tvar křivky

Frekvenční funkce normálního rozdělení

- frekvenční funkce $f(x)$ = hustota pravděpodobnosti

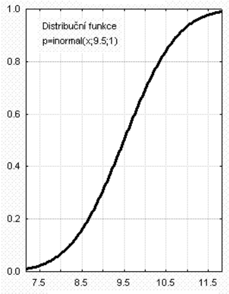
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

- parametr μ určuje, kde má křivka maximum, zároveň je tato hodnota mediánem a modem
- parametr σ určuje, jak jsou po obou stranách od hodnoty μ „vzdáleny“ inflexní body (jak je křivka roztažena do šířky)

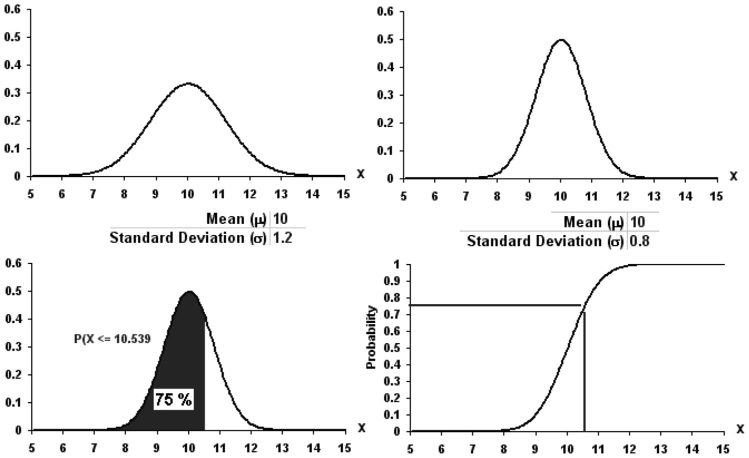


Distribuční funkce normálního rozdělení

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2}\, \mathrm{d}x$$



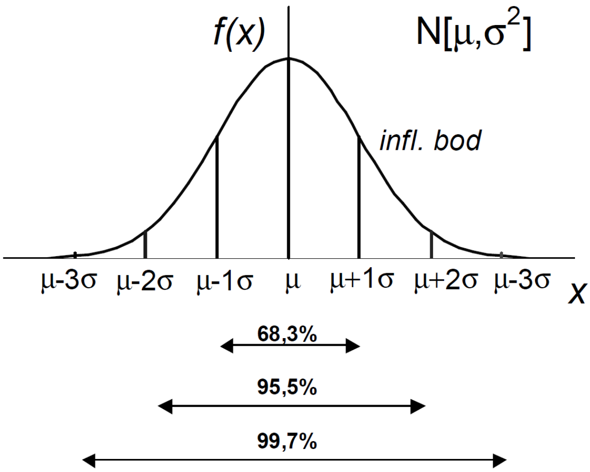
Normální rozdělení



Pravidlo „šesti sigma“

Plocha vymezená úsečkami		je rovna	v % plochy
$\bar{x} - \sigma$	$\bar{x} + \sigma$	0,683	68,3
$\bar{x} - 1,96\sigma$	$\bar{x} + 1,96\sigma$	0,950	95,0
$\bar{x} - 2,58\sigma$	$\bar{x} + 2,58\sigma$	0,990	99,0
$\bar{x} - 3\sigma$	$\bar{x} + 3\sigma$	0,997	99,7

Normální rozdělení



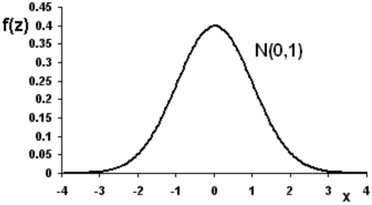
Normované normální rozdělení $N(0, 1)$

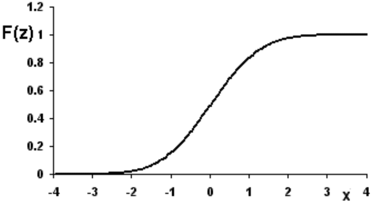
- hodnoty μ a σ se výběr od výběru liší
- frekvenční funkce má různý tvar

$$z = \frac{x - \mu}{\sigma}$$

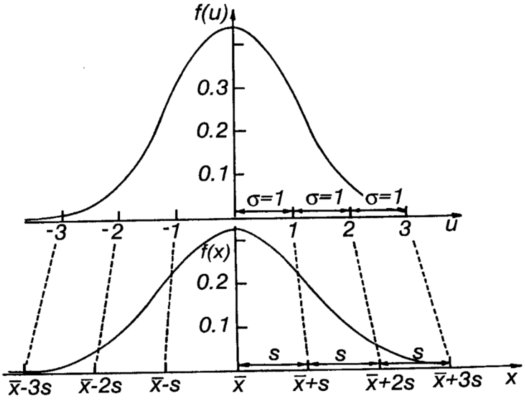
- normované normální rozdělení nezáleží na parametrech μ a σ
- frekvenční funkce $f(z) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- distribuční funkce $F(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-x^2/2} \, \mathrm{d}z$

Normované normální rozdělení $N(0, 1)$





Normované normální rozdělení



Obr. 3.9 Normování normálního rozdělení

Extremita jevů

Pravděpodobná chyba c pro normální rozdělení
$c = 0,6745s$

slovní označení extremity	symbol	meze	pravděpodobnost výskytu jevu (%)	meze	pravděpodobnost výskytu jevu (%)
extrémně podnormální	EP	< x - 3s	0,135	< x - 3c	2,150
silně podnormální	SP	x - 3s až x - 2s	2,190	x - 3c až x - 2c	8,870
podnormální	P	x -2s až x - s	13,590	x -2c až x - c	13,980
normální	O	x - s až x + s	68,270	x - c až x + c	50,000
nadnormální	N	x + s až x + 2s	13,590	x + c až x + 2c	13,980
silně nadnormální	SN	x + 2s až x + 3s	2,190	x + 2c až x + 3c	8,870
extrémně nadnormální	EN	> x + 3s	0,135	> x + 3c	2,150

Příklad - zadání

Plochu obhospodařované zemědělské půdy u sledovaného souboru farmářů modelujeme normálním rozdělením. Zjistili jsme, že parametry rozdělení $N(3\,400\ m^2, 360\,000\ m^2)$. Vypočtěte pravděpodobnost, že náhodně vybraný zemědělec bude mít:

a) méně než $4\,000\ m^2$ půdy
b) více než $4\,200\ m^2$ půdy
c) méně než $3\,000\ m^2$ půdy
d) mezi $2\,800\ m^2$ a $4\,000\ m^2$ půdy

Příklad - řešení

Lze řešit několika způsoby:

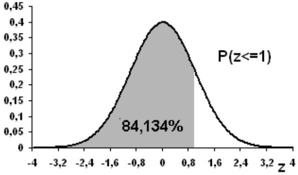
- převodem na normované normální rozdělení a využití statistických tabulek
- pomocí programu STATISTICA modelujícího hodnoty frekvenční a distribuční funkce - pravděpodobnostní kalkulátor
- pomocí funkce zadané do dlouhého jména proměnné v programu STATISTICA

Příklad - řešení a)

Transformace hodnoty x na normovanou veličinu:

$$z = \frac{x - \mu}{\sigma} = \frac{4\,000 - 3\,400}{600} = 1$$

Můžeme tedy psát: $P(x \leq 4\,000) = P(z \leq 1) = 0,84134$



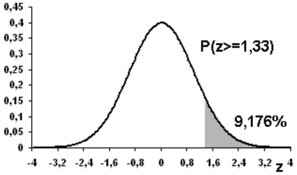
Příklad - řešení b)

Transformace hodnoty x na normovanou veličinu:

$$z = \frac{x - \mu}{\sigma} = \frac{4\,200 - 3\,400}{600} = 1,33$$

Pravděpodobnost, že normovaná proměnná překročí hodnotu 1,33 ovšem v tabulkách není. Proto:

$P(x \geq 4\,200) = P(z \geq 1,33) = 1 - P(z \leq 1,33) = 1 - 0,90824 = 0,09176$



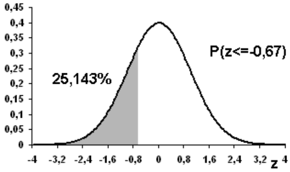
Příklad - řešení c)

Transformace hodnoty x na normovanou veličinu:

$$z = \frac{x - \mu}{\sigma} = \frac{3\,000 - 3\,400}{600} = -0,67$$

Pravděpodobnost $P(z \leq -0,67)$ *určíme na základě symetrie normovaného normálního rozdělení. Můžeme tedy psát:*

$P(z \leq -0,67) = P(z \geq 0,67) = 1 - P(z \leq 0,67) = 1 - 0,74857 = 0,25143$



Příklad - řešení d)

Transformace hodnoty x na normovanou veličinu:

$$z_1 = \frac{x-\mu}{\sigma} = \frac{2\,800 - 3\,400}{600} = -1$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{4\,000 - 3\,400}{600} = 1$$

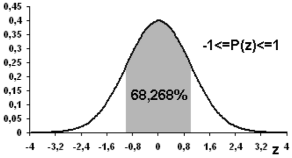
Pravděpodobnost mezi hodnotami $2\,800$ *a* $4\,000 m^2$ *u rozdělení* $N(3\,400\ m^2, 360\,000\ m^2)$ *je stejná jako plocha mezi hodnotami* -1 *a* 1 *u rozdělení* $N(0,1)$*. Tedy:*

$$P(2\,800 \leq x \leq 4\,000) = P(-1 \leq z \leq 1)$$

Příklad - řešení d)

Od plochy před hodnotou 1 odečteme plochu před hodnotou -1, proto:

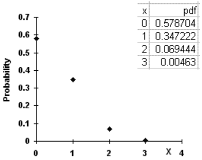
$P(-1 \leq z \leq 1) = P(z \leq 1) - P(z \leq -1) = P(z \leq 1) - [1 - P(z \leq 1)] = 0,84134 - [1 - 0,84134] = 0,68268$

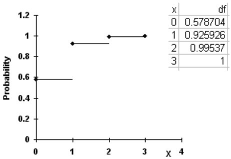


Binomické rozdělení

Příklad

Třikrát hodíme kostkou. Jaká je pravděpodobnost, že šestka nepadne, že padne jednou, dvakrát, třikrát?





Binomické rozdělení

- rozdělení diskrétní náhodné proměnné
- udává rozdělení výsledků jednoho a téhož pokusu za stejných podmínek
- výsledkem pokusu mohou být pouze dvě alternativy (A nebo B)
- pravděpodobnost varianty A značíme p, pravděpodobnost varianty B značíme q
- pokus provedeme n-krát a hledáme pravděpodobnost, že alternativa A nastane právě x-krát
- rozdělení pravděpodobnosti

$$f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} = \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$$

Binomické rozdělení

Předpoklady pro vznik náhodné proměnné s binomickým rozdělením:

- provádíme *n* pozorování nebo pokusů
- pozorování nebo pokusy jsou nezávislé - znalost výsledku v jednom pozorování nebo pokusu nám nic neříká o výsledku jiného pozorování
- výsledky pozorování nebo pokusu mohou být jenom dva, nazveme je „úspěch“ a „neúspěch“
- pravděpodobnost *p* každého „úspěchu“ je stejná pro všechna pozorování nebo pokusy

Binomické rozdělení

Základní charakteristiky binomického rozdělení

$$\mu = n \cdot p$$

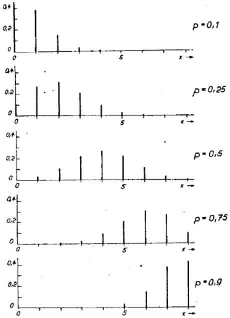
$$\sigma^2 = n \cdot p \cdot q$$

Příklady použití binomického rozdělení v geografii

- rozdělení počtu dní s určitým meteorologickým jevem za měsíc
- pravděpodobnost narození dvou chlapců v rodinách se třemi dětmi
- pravděpodobnost pozdního příchodu na jednu z přednášek statistiky
- ...

Binomické rozdělení

Rozdělení pravděpodobnosti binomického rozdělení pro $n = 8$ a různé hodnoty pravděpodobnosti p:



Příklad

Znalostní test sestává z otázek s několika volitelnými odpověďmi. Předpokládejme, že student má pravděpodobnost $0,55$, že správně odpoví na otázku náhodně zvolenou ze sestavy otázek. Správnost odpovědí na specifickou otázku nezáleží na ostatních otázkách. Test obsahuje 20 otázek. Odhadněte pravděpodobnost, s jakou student odpoví aspoň 14 otázek správně.

Řešení

$$f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} =$$

$$\binom{20}{14} \cdot 0,55^{14} \cdot 0,45^6 + \cdots + \binom{20}{20} \cdot 0,55^{20} \cdot 0,45^0 = 0,1299$$

Poissonovo rozdělení

- rozdělení diskrétní náhodné proměnné
- pojmenováno po francouzském matematikovi S. D. Poissonovi (1781–1840)
- náhodná veličina může nabývat hodnot $x = 0, 1, 2, 3, \ldots$ (kolikrát daný jev nastal v určitém časovém okamžiku) a to s pravděpodobnostní funkcí

$$f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

- platí $\mu = \sigma^2 = \lambda$, kde $\lambda = n \cdot p$ je jediným parametrem tohoto rozdělení

Poissonovo rozdělení

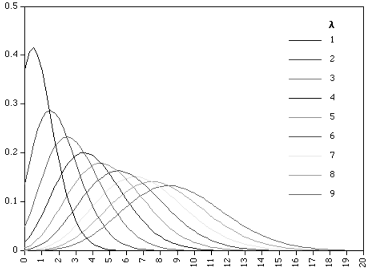
Předpoklady pro použití tohoto rozdělení:

- pravděpodobnost výskytu jedné události v daném intervalu (čase nebo prostoru) je úměrná délce tohoto intervalu
- události se vyskytují nezávisle jak ve stejném intervalu, tak mezi po sobě jdoucími intervaly
- událost může nastat v kterémkoliv okamžiku
- výskyt dvou či většího počtu událostí během krátkého časového okamžiku (ale i v malém prostoru) je prakticky nemožný

Poissonovo rozdělení

Příklady použití Poissonova rozdělení

- počet ztracených dětí v obchodním domě v určitém časovém úseku
- počet telefonních hovorů v určitém časovém úseku
- počet borovic na jednotku plochy smíšeného lesa
- počet tiskových chyb na jedné straně textu



Příklad

Předpokládejme, že průměr počtu těžkých zranění v jednom ročníku hokejové ligy je 5. Rozdělení četností zranění budeme modelovat pomocí Poissonova rozdělení. Máme zjistit, jaká je pravděpodobnost, že počet těžkých zranění bude více než 4.

Řešení

$$f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{5^x \cdot e^{-5}}{x!}$$

Dosazením pro $x = 0, 1, 2, 3, 4$ získáme $f(0) = 0,00674$, $f(1) = 0,03370$, $f(2) = 0,08425$, $f(3) = 0,14042$, $f(4) = 0,17552$. Tyto pravděpodobnosti sečteme a dostaneme hodnotu 0,44.
Hledaná pravděpodobnost, že počet zraněných přesáhne 4, má tedy hodnotu $1 - 0,44 = 0,56$.

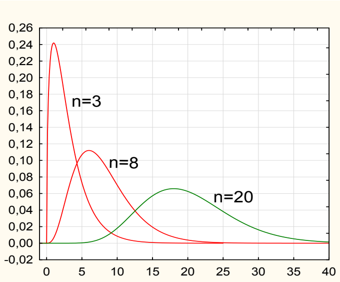
χ^2 rozdělení $\chi^2(n)$

- ze základního souboru s normovaným normálním rozdělením provedeme náhodný výběr n prvků, které označíme $x_1, x_2, \ldots, x_n$
- náhodná veličina představující součet čtverců těchto hodnot se řídí Pearsonovým χ^2 rozdělením s n stupni volnosti (stupně volnosti budeme vždy označovat ν)
- hodnota χ^2 může nabývat v různých výběrech různých hodnot v intervalu $(0, \infty)$
- jediným parametrem je n rozsah souboru a označuje počet stupňů volnosti
- s roustoucím rozsahem náhodného výběru se rozdělení blíží normálnímu rozdělení

χ^2 rozdělení

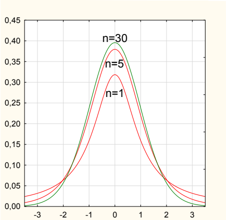
Příklady použití χ^2 rozdělení

- v teorii odhadu parametrů základního souboru
- při testování statistických hypotéz
- při ověřování předpokladu o shodě teoretického a empirického rozdělení
- ...



Studentovo t rozdělení

- spojité rozdělení pravděpodobnosti, které umožňuje i na základě souborů dat s malými rozsahy, u kterých nelze hovořit o asymptoticky normálním rozdělení, dělat přijatelné závěry
- Studentovo *t* rozdělení má jeden parametr a tím je opět rozsah souboru
- využívá se především pro testování hypotéz



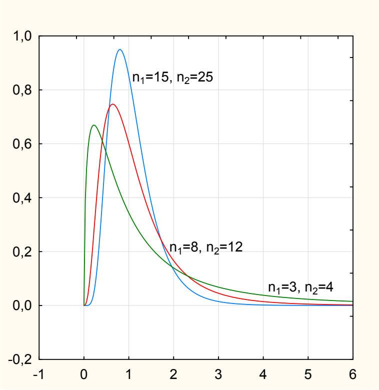
Fisher-Snedecorovo rozdělení

- uvažujme dvě nezávislé náhodné veličiny, které mají χ^2 rozdělení s ν_1 a ν_2 stupňů volnosti
- poměr

$$F = \frac{X_1\,/\,n_1}{X_2\,/\,n_2}$$

  má F rozdělení
- nabývá pouze kladných hodnot, má dva parametry ν_1 a ν_2
- používá se u testů hypotéz, při analýze rozptylu a v regresní analýze

Fisher-Snedecorovo rozdělení



Exponenciální rozdělení

- udává dobu čekání na příchod nějaké události, která se může dostavit každým okamžikem se stejnou šancí bez ohledu na dosud pročekanou dobu
- hustota pravděpodobnosti (frekvenční funkce) tohoto rozdělení má následující tvar pro $x > 0$:

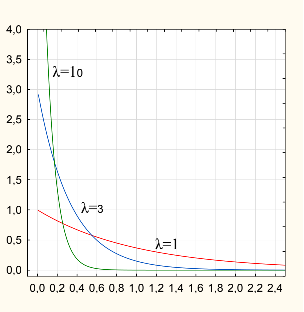
$$f(x) = \lambda e^{-\lambda x}$$

- rozdělení má jeden parametr λ - velké hodnoty indikují velký sklon hustoty pravděpodobnosti a naopak
- střední hodnota je rovna $\frac{1}{\lambda}$ a rozptyl $\frac{1}{\lambda^2}$

Exponenciální rozdělení

- hodnoty pravděpodobnosti lze určovat přímo z distribuční funkce

$$F(x) = p(X < x) = 1 - e^{-\lambda x}$$



Příklad

40 studentů dojíždí do školy z průměrné vzdálenosti 7 km. Histogram hodnot vzdálenosti vykazuje pozitivní asymetrii. Hodnota parametru exponenciálního rozdělení bude $\lambda = \frac{1}{7} = 0,143$. Jaká je pravděpodobnost, že student dojíždí ze vzdálenosti 15 km a delší?

$$p(15 > x) = 1 - p(15 < x) = 1 - \left[1 - e^{-0,143 \cdot 15}\right] = e^{-2,145} = 0,117\,068$$

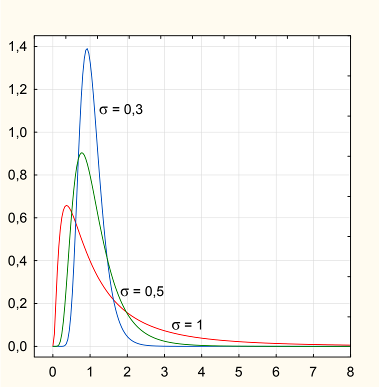
Lognormální rozdělení

- proměnná X má lognormální rozdělení pravděpodobnosti, když logaritmickou transformací $Y = lnX$ získá rozdělení normální s parametry μ a σ

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{\frac{(lnx - \mu)^2}{2\sigma^2}}$$

- používá se při rozdělení věku obyvatelstva v populaci, při stopové analýze (koncentraci stopových prvků v horninách), . . .

Lognormální rozdělení



Děkuji za pozornost...