

STATISTIKA 2

Statistika a statistické zpracování dat

Blok 2-2

Jiří Fišer

22. listopadu 2024

Základní diskrétní rozdělení pravděpodobností

- Alternativní rozdělení
- Binomické rozdělení
- Hypergeometrické rozdělení
- Diskrétní rovnoměrné rozdělení
- Poissonovo rozdělení

Binomické rozdělení $X \sim \text{Bi}(n, p)$

Model: počet úspěchů, které nastanou při provedení n nezávislých dichotomických pokusů

Příklady

- počet děvčat v rodině s n dětmi
- počet zmetků mezi n výrobky
- počet úspěšných zkoušek přístroje z celkem n zkoušek

Parametry rozdělení

- n ... počet nezávislých dichotomických pokusů
- p ... pravděpodobnost úspěchu v jednom pokusu

Binomické rozdělení $X \sim \text{Bi}(n, p)$

Pravděpodobnostní funkce (Bernoulliho schéma)

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \dots, n, \quad n \geq 1, \quad p \in (0, 1)$$

Distribuční funkce

$$F(x; p, n) = \begin{cases} 0 & x < 0 \\ \sum_{k \leq x} \binom{n}{k} p^k (1 - p)^{n-k} & 0 \leq x < n \\ 1 & x \geq n \end{cases}$$

Excel: `BINOM.DIST(k;n;p;logická proměnná)`

- logická proměnná = 0 pro výpočet $P(X = k)$
- logická proměnná = 1 pro výpočet $F(k)$

Binomické rozdělení $X \sim \text{Bi}(n, p)$

Číselné charakteristiky

$$E(X) = np, \quad D(X) = np(1 - p), \quad \sigma_X = \sqrt{np(1 - p)}$$

$p = 1/2$ symetrické rozdělení, jinak asymetrické

Příklad (Barva květů hrachu)

*Je křížen bělokvětý hrách s fialovým, přičemž předpokládáme, že rostliny, na nichž je pokus prováděn, nebyly dosud kříženy. Podle pravidel dědičnosti lze očekávat, že $\frac{3}{4}$ nově vzniklých rostlin (potomků) pokvetou fialově a $\frac{1}{4}$ bíle. Zatím vzklíčilo 10 nových rostlin. **Jaká je pravděpodobnost, že 9 pokvete fialově?***

Nějaký nápad?

Příklad (Barva květů hrachu)

Je křížen bělokvětý hrách s fialovým, přičemž předpokládáme, že rostliny, na nichž je pokus prováděn, nebyly dosud kříženy. Podle pravidel dědičnosti lze očekávat, že $\frac{3}{4}$ nově vzniklých rostlin (potomků) pokvetou fialově a $\frac{1}{4}$ bíle. Zatím vzklíčilo 10 nových rostlin. **Jaká je pravděpodobnost, že 9 pokvete fialově?**

Řešení

X ... počet fialově kvetoucích rostlin z 10 rostlin

$$X \sim \text{Bi}(10; 0,75)$$

$$P(X = 9) = \binom{10}{9} 0,75^9 0,25^1 = 0,188$$

Excel: `BINOM.DIST(9;10;3/4;0)`

Hypergeometrické rozdělení $X \sim \text{Hg}(N, M, n)$

Model: počet úspěchů, které nastanou při provedení n závislých dichotomických pokusů (výběr bez vracení)

Příklady

- počet vybraných zmetků, vybíráme-li n výrobků z N výrobků, přičemž M z nich jsou zmetky
- počet vybraných nenaučených otázek, losujeme-li n otázek z N , přičemž na M z nich nejsme naučeni

Parametry rozdělení

- n ... počet vybraných jednotek (závislých dichotomických pokusů)
- N ... počet všech jednotek
- M ... počet jednotek, které mají sledovanou vlastnost

Hypergeometrické rozdělení $X \sim \text{Hg}(N, M, n)$

Pravděpodobnostní funkce

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, n, \quad N \geq n \geq 1, \quad 0 \leq M \leq N$$

Distribuční funkce

$$F(x; N, M, n) = \begin{cases} 0 & x < 0 \\ \sum_{k \leq x} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} & 0 \leq x < n \\ 1 & x \geq n \end{cases}$$

Číselné charakteristiky

$$E(X) = \frac{n \cdot M}{N}, \quad D(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \left(\frac{N-n}{N-1}\right)$$

Hypergeometrické rozdělení $X \sim \text{Hg}(N, M, n)$

Excel: `HYPGEOM.DIST(k;n;M;N;logická proměnná)`

- logická proměnná = 0 pro výpočet $P(X = k)$
- logická proměnná = 1 pro výpočet $F(k)$

Příklad (Vadné výrobky)

Mezi 100 výrobky je 20 zmetků. Vybereme 10 výrobků a sledujeme počet zmetků mezi vybranými výrobky. **Jaká je pravděpodobnost, že jsme vybrali právě 3 zmetky?**

Nějaký nápad?

Příklad (Vadné výrobky)

Mezi 100 výrobky je 20 zmetků. Vybereme 10 výrobků a sledujeme počet zmetků mezi vybranými výrobky. **Jaká je pravděpodobnost, že jsme vybrali právě 3 zmetky?**

Řešení

X ... počet vybraných zmetků

$N = 100$, $M = 20$, $n = 10$

$$X \sim \text{Hg}(100; 20; 10)$$

$$P(X = 3) = \frac{\binom{20}{3} \binom{80}{7}}{\binom{100}{10}} = 0,209$$

Excel: `HYPGEOM.DIST(3;10;20;100;0)`

Poissonovo rozdělení $X \sim \text{Po}(\lambda)$

Model: počet událostí, které nastanou při provedení **velkého počtu nezávislých dichotomických pokusů**, přičemž **pravděpodobnost vzniku události je velmi malá**

- výskyt jevu v daném časovém/prostorovém intervalu nezávisí co se stalo jindy/jinde
- v daném okamžiku/bodě nemohou nastat 2 jevy současně
- pro každý časový okamžik/bod je pravděpodobnost výskytu jevu v malém časovém/prostorovém intervalu stejná
- λ ... intenzitu výskytu události za jednotku času/délky/objemu

Příklady

- # částic emit. za jedn. času radioaktivní látkou dané hmotnosti
- # signálů, které dojdou do telefonní ústředny za jednotku času
- # jednob. organizmů na ploše velikosti t v zorném poli mikroskopu
- # dopravních nehod za jednotku času
- # typografických chyb na stránce
- # volných dnů ve firmě během měsíce

Poissonovo rozdělení $X \sim \text{Po}(\lambda)$

Pravděpodobnostní funkce

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots, \quad \lambda > 0$$

Distribuční funkce

$$F(x; \lambda) = \begin{cases} 0 & x < 0 \\ \sum_{k \leq x} \frac{\lambda^k}{k!} e^{-\lambda} & x \geq 0 \end{cases}$$

Číselné charakteristiky

$$E(X) = \lambda, \quad D(X) = \lambda, \quad \sigma_X = \sqrt{\lambda}$$

Excel: `POISSON.DIST(k;λ;logická proměnná)`

- logická proměnná = 0 pro výpočet $P(X = k)$
- logická proměnná = 1 pro výpočet $F(k)$

Příklad (Poruchy obráběcího stroje)

Při práci obráběcího stroje dochází náhodně k výpadkům. Průměrně jsou 2 výpadky za 24 hodin.

- a) *Jaká je pst, že za 24 hodin dojde alespoň k jednomu výpadku?*
- b) *Jaká je pst, že za týden nebudou více než 3 výpadky?*

Nějaký nápad?

Příklad (Poruchy obráběcího stroje)

Při práci obráběcího stroje dochází náhodně k výpadkům. Průměrně jsou 2 výpadky za 24 hodin.

- a) **Jaká je pst, že za 24 hodin dojde alespoň k jednomu výpadku?**
- b) *Jaká je pst, že za týden nebudou více než 3 výpadky?*

Řešení

X ... počet výpadků za 1 den

průměrně 2 výpadky za 24 hodin $\Rightarrow \lambda = 2$

$$X \sim \text{Po}(2)$$

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - e^{-2} = 0,865$$

Excel: **1-POISSON.DIST(0;2;0)**

Příklad (Poruchy obráběcího stroje)

Při práci obráběcího stroje dochází náhodně k výpadkům. Průměrně jsou 2 výpadky za 24 hodin.

- a) *Jaká je pst, že za 24 hodin dojde alespoň k jednomu výpadku?*
- b) **Jaká je pst, že za týden nebudou více než 3 výpadky?**

Řešení

Y ... počet výpadků za 1 týden

průměrně 2 výpadky/24 hodin
 $\Leftrightarrow$ průměrně $2 \cdot 7 = 14$ výpadků/týden

$$Y \sim \text{Po}(14)$$

$$P(Y \leq 3) = e^{-14} \left(1 + 14 + \frac{14^2}{2!} + \frac{14^3}{3!} \right) = 0,000474$$

Excel: **POISSON.DIST(3;14;1)**

2.5 Základní spojitá rozdělení pravděpodobností

- Rovnoměrné rozdělení
- Exponenciální rozdělení
- Normální rozdělení
- Studentovo rozdělení
- F-rozdělení
- Chi-kvadrát rozdělení

Rovnoměrné rozdělení $X \sim R(a, b)$

Model $X \sim R(0, d)$

- modelování **doby čekání na událost, která nastává v pravidelných intervalech** délky d
- čekání na událost zahajujeme v okamžiku, který nikterak nesouvisí s minulým ani budoucím výskytem události

Příklady

- doba čekání na autobusové zastávce na autobus, který jezdí po d minutách $R(0, d)$
- chyba při zaokrouhlení na 1 desetinné místo $R(-0,05; 0,05)$
- chyba při stanovení času z digitálních hodin v minutách $R(0; 60)$ sekund

Rovnoměrné rozdělení $X \sim R(a, b)$

Hustota

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b), \quad -\infty < a < b < \infty \\ 0 & \text{jinak} \end{cases}$$

Distribuční funkce

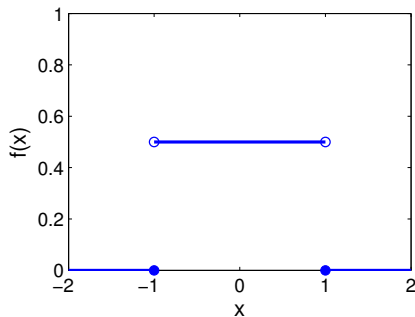
$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & x \in (a, b) \\ 1 & x \geq b \end{cases}$$

Číselné charakteristiky

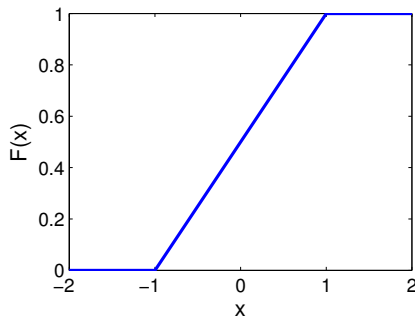
$$E(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}, \quad \sigma_X = \frac{b-a}{\sqrt{12}}$$

Rovnoměrné rozdělení $X \sim R(-1, 1)$

Hustota



Distribuční funkce



Příklad (Dodávka zboží)

*Prodejna očekává dodávku nového zboží určitý den v době od 8 do 10 hodin. Podle sdělení dodavatele je uskutečnění dodávky stejně možné kdykoliv během tohoto časového intervalu. **Jaká je pravděpodobnost, že zboží bude dodáno v době od půl deváté do tři čtvrtě na devět?***

Nějaký nápad?

Příklad (Dodávka zboží)

Prodejna očekává dodávku nového zboží určitý den v době od 8 do 10 hodin. Podle sdělení dodavatele je uskutečnění dodávky stejně možné kdykoliv během tohoto časového intervalu. **Jaká je pravděpodobnost, že zboží bude dodáno v době od půl deváté do tři čtvrtě na devět?**

Řešení

X ... čas příjezdu v hodinách

$$X \sim \text{Ro}(8, 10)$$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{jinak} \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{2} & x \in (8, 10) \\ 0 & \text{jinak} \end{cases}$$

Pravděpodobnost, že zboží bude dodáno v době od půl deváté do tři čtvrtě na devět

$$P(8,5 \leq X \leq 8,75) = \int_{8,5}^{8,75} \frac{1}{2} dx = \frac{1}{2}[8,75 - 8,5] = 0,125$$

Exponenciální rozdělení $X \sim \text{Ex}(\lambda)$

Model: modelování doby/vzdálenosti čekání na událost, která nemá paměť, tj. zbývající doba/vzdálenost čekání nezávisí na tom, jak dlouho už na událost čekáme

Příklady

- doba životnosti zařízení, u kterého dochází k poruše náhodně, nikoliv z důsledku opotřebení
- délka telefonního hovoru
- doba mezi 2 příchozími telefonáty

Parametr rozdělení

- $\lambda \dots$ míra rizika výskytu sledované události za jednotku času/vzdálenosti
- $\frac{1}{\lambda} \dots$ očekávaná doba/vzdálenost do výskytu sledované události

Exponenciální rozdělení $X \sim \text{Ex}(\lambda)$

Hustota

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

Distribuční funkce

$$F(x) = \begin{cases} 0 & x \leq 0, \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Číselné charakteristiky

$$E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}, \quad \sigma_X = \frac{1}{\lambda}$$

Pozor! Někdy se používá jiná parametrizace

$$X \sim \text{Ex}(\theta), \quad \theta = 1/\lambda$$

Exponenciální rozdělení $X \sim \text{Ex}(\lambda)$

Hustota

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

Distribuční funkce

$$F(x) = \begin{cases} 0 & x \leq 0, \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Číselné charakteristiky

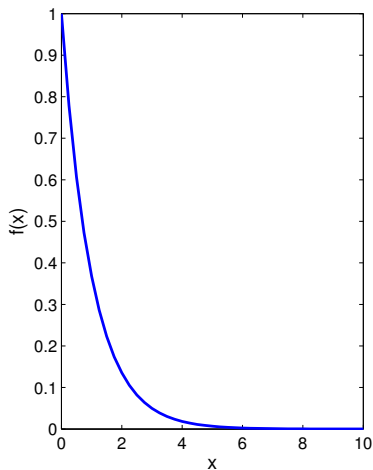
$$E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}, \quad \sigma_X = \frac{1}{\lambda}$$

Pozor! Někdy se používá jiná parametrizace

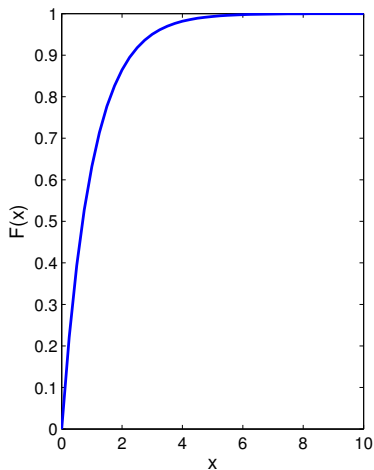
$$X \sim \text{Ex}(\theta), \quad \theta = 1/\lambda$$

Exponenciální rozdělení $X \sim \text{Ex}(1)$

Hustota



Distribuční funkce



Exponenciální rozdělení $X \sim \text{Ex}(\lambda)$

Excel: `EXPON.DIST(x;λ;logická proměnná)`

- logická proměnná = 0 pro výpočet $f(x)$
- logická proměnná = 1 pro výpočet $F(x)$

Příklad (Obsluha v restauraci)

*Doba čekání hosta na pivo v restauraci U Lva je průměrně 5 minut.
Určete*

- a) pravděpodobnost, že host bude na pivo čekat déle než 12 minut*
- b) dobu čekání hosta na pivo, během níž bude obsloužen s pravděpodobností 0,9.*

Nějaký nápad?

Příklad (Obsluha v restauraci)

Doba čekání hosta na pivo v restauraci U Lva je průměrně 5 minut. Určete

- a) *pravděpodobnost, že host bude na pivo čekat déle než 12 minut*
- b) *dobu čekání hosta na pivo, během níž bude obsloužen s pravděpodobností 0,9.*

Řešení X ... doba čekání na pivo (v minutách)

průměrná doba čekání na pivo ... 5 minut

$$X \sim \text{Ex}(\lambda), \lambda = \frac{1}{5}$$

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{5}e^{-\frac{1}{5}x} & x \geq 0 \end{cases}, \quad F(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-\frac{1}{5}x} & x \geq 0 \end{cases}$$

Pravděpodobnost, že host bude na pivo čekat déle než 12 minut

$$\begin{aligned} P(X > 12) &= 1 - P(X \leq 12) = 1 - F(12) = 1 - \left(1 - e^{-\frac{1}{5} \cdot 12}\right) \\ &= e^{-\frac{12}{5}} = 0,09 \end{aligned}$$

Excel: 1-EXPON.DIST(12;1/5;1)

Doba čekání hosta na pivo, během níž bude obsloužen s pravděpodobností 0,9

$$P(0 \leq X \leq t) = 0,9$$

$$F(t) - F(0) = 0,9$$

$$1 - e^{-t/5} - 0 = 0,9$$

$$e^{-t/5} = 0,1$$

$$t = -5 \ln(0,1) \doteq 11,51 \text{ min} \doteq 11 \text{ min } 30 \text{ s}$$

Normální rozdělení $X \sim N(\mu, \sigma^2)$

Hustota

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}^1, \mu \in \mathbb{R}^1, \sigma > 0.$$

Vlastnosti

- X může nabývat jakýchkoli reálných hodnot
- hustota je symetrická kolem μ
- hustota nabývá maxima v bodě $x = \mu$, $(f(\mu) = \frac{1}{\sqrt{2\pi}\sigma})$
- normální rozdělení s parametry $\mu = 0$ a $\sigma^2 = 1$ nazýváme **normované (standardizované) normální rozdělení**.
- μ parametr polohy (střední hodnota)
- σ parametr měřítka (směrodatná odchylka)

Normální rozdělení $X \sim N(\mu, \sigma^2)$

Číselné charakteristiky

$$E(X) = \mu, \quad x_{0.5} = \mu, \quad \hat{x} = \mu, \quad D(X) = \sigma^2$$

$$\gamma_1(X) = 0 \text{ (šikmost)}, \quad \gamma_2(X) = 0 \text{ (špičatost)}$$

Excel: **NORM.DIST**($x; \mu; \sigma$; logická proměnná)

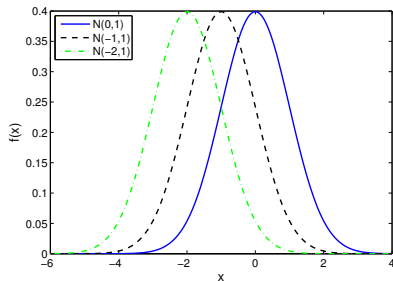
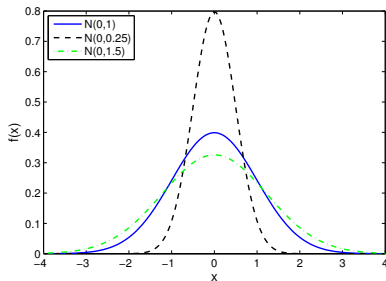
- logická proměnná = 0 pro výpočet $f(x)$
- logická proměnná = 1 pro výpočet $F(x)$

NORM.S.DIST(x ; logická proměnná) pro práci s $N(0, 1)$

NORM.INV($\alpha; \mu; \sigma$) výpočet α kvantilu

NORM.S.INV(α) výpočet α kvantilu pro $X \sim N(0, 1)$

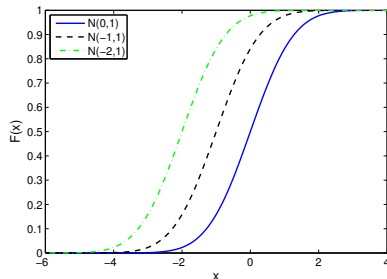
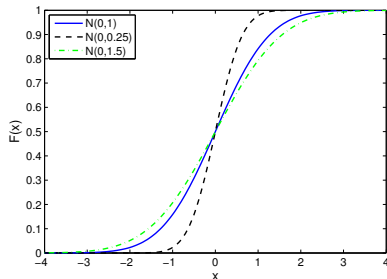
Hustota normálního rozdělení



$$X \sim N(0, \sigma^2) : f(-x) = f(x), \quad u_{1-\alpha} = -u_{\alpha}$$

$$X \sim N(\mu, \sigma^2) : f(-x + \mu) = f(x + \mu)$$

Distribuční funkce normálního rozdělení



$$X \sim N(0, \sigma^2) : F(-x) = 1 - F(x)$$

$$X \sim N(\mu, \sigma^2) : F(-x + \mu) = 1 - F(x + \mu)$$

Vztah mezi $N(\mu, \sigma^2)$ a $N(0, 1)$

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$X \sim N(0, 1), \mu \in \mathbb{R}^1, \sigma^2 > 0 \Leftrightarrow \sigma X + \mu \sim N(\mu, \sigma^2)$$

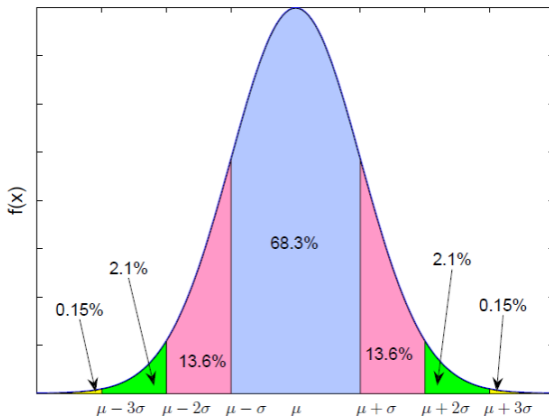
Pravidlo 3 sigma

Pro $X \sim N(\mu, \sigma^2)$ platí

$$P[\mu - \sigma < X < \mu + \sigma] \approx 68,3\%$$

$$P[\mu - 2\sigma < X < \mu + 2\sigma] \approx 95,5\%$$

$$P[\mu - 3\sigma < X < \mu + 3\sigma] \approx 99,7\%$$



Příklad (IQ testy)

Hodnota IQ se řídí normálním rozdělením $N(100, 15^2)$. Jaká je pravděpodobnost, že náhodně vybraná osoba bude mít

- a) IQ více než 130 bodů,*
- b) IQ méně než 90 bodů,*
- c) IQ v intervalu od 100 do 130 bodů?*

Nějaký nápad?

Příklad (IQ testy)

Hodnota IQ se řídí normálním rozdělením $N(100, 15^2)$. Jaká je pravděpodobnost, že náhodně vybraná osoba bude mít

- a) IQ více než 130 bodů,*
- b) IQ méně než 90 bodů,*
- c) IQ v intervalu od 100 do 130 bodů?*

Řešení. Přímý výpočet s využitím softwaru

$$P(100 < X < 130) = F(130) - F(100) = 0,9772 - \frac{1}{2} = 0,4772,$$

kde $F(x)$ je distribuční funkce $N(100, 15^2)$

Studentovo rozdělení

- Studentovo rozdělení je pojmenováno po Williamu Sealy Gossetovi, statistikovi pracujícím pro pivovar Guinness.
- Aby se vyhnul problémům s publikováním, používal pseudonym „Student“.
- V roce 1908 zveřejnil práci, která popisovala rozdělení nyní známé jako Studentovo rozdělení.
- Jeho cílem bylo vyřešit problémy s malými vzorky v průmyslu.

Definice

Definition

Studentovo rozdělení s ν (řecké písmeno „ný“) stupni volnosti je užitečné při odhadu střední hodnoty normální populace na základě malého vzorku, pokud směrodatná odchylka populace není známa. Hustota Studentova rozdělení je dána vzorcem:

$$f(x; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

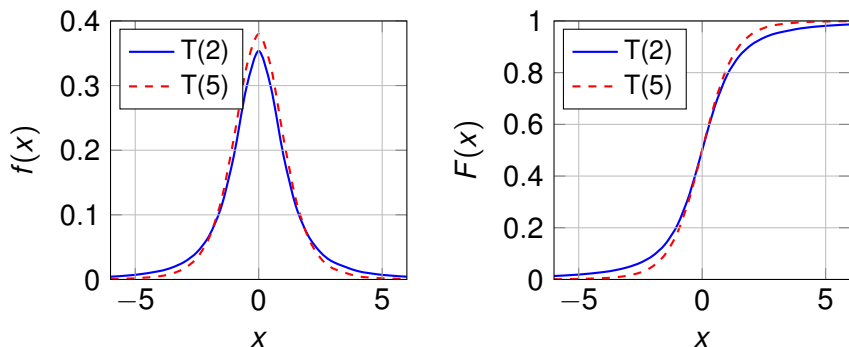
kde ν je počet stupňů volnosti a Γ je gama funkce.

Základní číselné charakteristiky

- **Střední hodnota:** 0 (pro $\nu > 1$)
- **Rozptyl:** $\frac{\nu}{\nu - 2}$ (pro $\nu > 2$)
- **Asymptotické chování:** Pro velké ν se Studentovo rozdělení blíží normálnímu rozdělení.

Grafy hustot a distribuční funkce

Grafy znázorňující hustotu a distribuční funkci Studentova rozdělení pro $\nu = 2$ a $\nu = 5$ stupňů volnosti:



Obrázek: Grafy hustot a distribučních funkcí Studentova rozdělení pro 2 a 5 stupňů volnosti

Excelovské funkce

Pro práci se Studentovým rozdělením lze v Excelu použít následující funkce:

- **Hustota pravděpodobnosti (PDF):** `T.DIST(x;  $\nu$ ; FALSE)`
- **Distribuční funkce (CDF):** `T.DIST(x;  $\nu$ ; TRUE)`
- **Kvantilová funkce:** Funkce `T.INV(p;  $\nu$ )` vrací kvantil pro danou pravděpodobnost p a ν stupni volnosti.

Procvičení

Použijte vhodné excelovské funkce k procvičení práce s rozdělením:

- 1 Vypočítejte hodnoty hustoty pravděpodobnosti pro Studentovo rozdělení s $\nu = 8$ a následující hodnoty $x = -2, -1, 0, 1, 2$.
- 2 Vypočítejte hodnoty distribuční funkce pro Studentovo rozdělení s $\nu = 8$ a stejné hodnoty $x = -2, -1, 0, 1, 2$.
- 3 Pomocí funkce `T.INV()` najděte kvantily pro pravděpodobnosti $p = 0,05; 0,5; 0,95$ při $\nu = 8$. Co nám ty výsledky říkají?
- 4 Vytvořte přibližné grafy hustoty a distribuční funkce pro Studentovo rozdělení v Excelu pomocí výpočtu jejich hodnot v dostatečně husté síti bodů na ose x . Můžete opět použít $\nu = 8$ a $x = -6; -5,5; -5; \dots; 5,5; 6$.

Alternativně použijte GeoGebru.

Podobně projděte a procvičte další dvě rozdělení ze skript:

- F-rozdělení
- Chi-kvadrát