Studijní text

Název předmětu: PRAVDĚPODOBNOST A STATISTIKA

Garant předmětu: RNDr. Marek Sedlačík, Ph.D.

Téma: Zákon velkých čísel a limitní věty

Obsah

1 Limitní věty teorie pravděpodobnosti **2**
1.1 Zákon velkých čísel 2
1.2 Centrální limitní věta 3
1.3 Příklady 5

2 Výběrová šetření **6**
2.1 Druhy výběrového šetření 6
2.2 Náhodný výběr 7
2.3 Výběrové charakteristiky 8
2.4 Statistická indukce 11

Literatura **11**

Úkoly pro samostatnou práci **13**

1 Limitní věty teorie pravděpodobnosti

1.1 Zákon velkých čísel

- Jestliže opakujeme nezávisle nějaký pokus, můžeme z pozorovaných hodnot sestavit rozdělení relativních četností a informaci o tomto rozdělení shrnout do charakteristik.
- Toto rozdělení resp. charakteristiky nazveme **empirickým rozdělením** resp. **empirickými charakteristikami**.
- Při dodržování jistých podmínek můžeme očekávat, že empirické rozdělení (resp. charakteristiky) se bude blížit teoretickému rozdělení (resp. charakteristikám) a to tím více, čím větší bude počet realizovaných pokusů.
- Je třeba si uvědomit, že přibližování empirických hodnot k hodnotám teoretickým nemá charakter matematické konvergence, ale konvergence pravděpodobnostní.
- Pravděpodobnostní konvergencí rozumíme skutečnost, že při vzrůstajícím počtu pokusů se pravděpodobnost větších odchylek empirických hodnot od teoretických stále zmenšuje.

Definice 1.1 Jestliže pro posloupnost náhodných veličin $X_1, X_2, \ldots, X_n, \ldots$ platí vztah

$$\lim_{n\to\infty} P(|X_n - c| < \epsilon) = 1, \epsilon > 0,$$

říkáme, že posloupnost $\{X_n\}$ **konverguje podle pravděpodobnosti** ke konstantě c. Pravděpodobnostní konvergence se označuje

$$X_n \xrightarrow{P} c.$$

Věta 1.1 Čebyševova nerovnost
Pro libovolnou náhodnou veličinu X se střední hodnotou $E(X)$, konečným rozptylem $D(X)$ a pro každé $\epsilon > 0$ platí

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{D(X)}{\epsilon^2}.$$

Čebyševova nerovnost se uplatňuje především v oblasti teorie, je možno ji však použít i pro odhad určitých pravděpodobností u náhodných veličin, jejichž rozdělení jsou neznámá.

Věta 1.2 Bernoulliova věta
Jestliže náhodná veličina X je počet výskytů jevu v n nezávislých pokusech a π je stálá pravděpodobnost, že tento jev nastane v jednom pokuse, potom pro každé $\epsilon > 0$ platí

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - \pi\right| < \epsilon\right) = 1.$$

1.2 Centrální limitní věta

Normální rozdělení je limitním rozdělením, k němuž se za určitých podmínek blíží řada jiných rozdělení. Náhodná veličina X, která vznikla jako součet velkého počtu vzájemně nezávislých veličin $X_1, X_2, \ldots, X_n$, má za velmi obecných podmínek přibližně normální rozdělení. Nejjednodušším tvarem CLV je *Moivre-Laplaceova věta*, jistým zobecněním *Lévy-Lindebergova věta*.

Věta 1.3 Moivre-Laplaceova věta
Nechť náhodná veličina X má binomické rozdělení s parametry n a π, tj. $X \sim B(n, \pi)$ [1], střední hodnotu $E(X) = n\pi$ a rozptyl $D(X) = n\pi(1-\pi)$. Pro normovanou náhodnou veličinu

$$U = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}}$$

platí limitní vztah $\lim\limits_{n\to\infty} P(U \leq u) = \Phi(u)$, kde $\Phi(u)$ je distribuční funkce rozdělení $N(0,1)$.

Moivreova-Laplaceova věta tedy říká, že při dostatečně velkém počtu nezávislých pokusů konverguje binomické rozdělení k rozdělení normálnímu. Aproximace je vhodná, jestliže

$$n\pi(1-\pi) > 9 \quad \text{a} \quad \frac{1}{n+1} < \pi < \frac{n}{n+1}.$$

Věta 1.4 Moivre-Laplaceova věta pro podíl
Nechť náhodná veličina X je součet n nezávislých alternativních veličin $X_1, X_2, \ldots, X_n$, které mají stejný parametr π, $0 < \pi < 1$, a náhodná veličina $\frac{X}{n}$ má střední hodnotu $E\left(\frac{X}{n}\right) = \pi$ a rozptyl $D\left(\frac{X}{n}\right) = \frac{\pi(1-\pi)}{n}$. Potom pro normovanou náhodnou veličinu

$$U = \frac{\frac{X}{n} - \pi}{\sqrt{\pi(1-\pi)}}\sqrt{n}$$

platí limitní vztah $\lim\limits_{n\to\infty} P(U \leq u) = \Phi(u)$, kde $\Phi(u)$ je distribuční funkce normovaného normálního rozdělení $N(0,1)$.

Věta 1.5 Lévy-Lindebergova věta
Nechť náhodná veličina X je součtem n nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$, se stejným rozdělením s konečnou střední hodnotou $E(X_i) = \mu$ a konečným rozptylem $D(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$. [1] Pro normovanou náhodnou veličinu

$$U = \frac{X - n\mu}{\sqrt{n\sigma^2}}$$

platí limitní vztah $\lim\limits_{n\to\infty} P(U \leq u) = \Phi(u)$, kde $\Phi(u)$ je distribuční funkce rozdělení $N(0,1)$.

[1] X je součtem n nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$, z nichž každá má alternativní rozdělení $A(\pi)$

[1] Střední hodnota a rozptyl náhodné veličiny X jsou vzhledem k nezávislosti a stejnému rozdělení veličin X_i rovny $E(X) = n\mu$ a $D(X) = n\sigma^2$

Věta 1.6 Lévy-Lindebergova věta pro průměr
Nechť náhodná veličina $\overline{X}$ je průměr n nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$, které mají libovolný identický zákon rozdělení s konečnou střední hodnotou $E(X_i) = \mu$ a konečným rozptylem $D(X_i) = \sigma^2$. Potom pro střední hodnotu a rozptyl náhodné veličiny $\overline{X}$ platí $E(\overline{X}) = \mu$ a $D(\overline{X}) = \frac{\sigma^2}{n}$ a pro normovanou náhodnou veličinu

$$U = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$$

platí limitní vztah $\lim\limits_{n\to\infty} P(U \leq u) = \Phi(u)$, kde $\Phi(u)$ je distribuční funkce normovaného normálního rozdělení $N(0,1)$.

Pro výběrový úhrn M platí:

- $M = \sum\limits_{i=1}^{n} X_i \sim as.N(n\mu, n\sigma^2), E(M) = n\mu, D(M) = n\sigma^2$
- $U = \frac{M - E(M)}{\sqrt{D(M)}} = \frac{M - n\mu}{\sqrt{n\sigma^2}} \sim as.N(0,1)$
- $P(M \leq m) = F(m) \approx \Phi\left(\frac{m - n\mu}{\sqrt{n\sigma^2}}\right)$
- $P\left(-u_{1-\alpha/2} < \frac{m - n\mu}{\sqrt{n\sigma^2}} < u_{1-\alpha/2}\right) = 1 - \alpha$

Pro výběrový průměr $\overline{X}$ platí:

- $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i \sim as.N(\mu, \frac{\sigma^2}{n}), E(\overline{X}) = \mu, D(\overline{X}) = \frac{\sigma^2}{n}$
- $U = \frac{\overline{X} - E(\overline{X})}{\sqrt{D(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim as.N(0,1)$
- $P(\overline{X} \leq \overline{x}) = F(\overline{x}) \approx \Phi\left(\frac{\overline{x} - \mu}{\sigma}\sqrt{n}\right)$
- $P\left(-u_{1-\alpha/2} < \frac{\overline{x} - \mu}{\sigma}\sqrt{n} < u_{1-\alpha/2}\right) = 1 - \alpha$

Poznámka k opravě na spojitost
Používáme-li normální rozdělení jako aproximaci pro rozdělení diskrétní náhodné veličiny, doporučuje se, zejména u silněji asymetrických rozdělení, provést tzv. **opravu na spojitost**, která tuto aproximaci zlepšuje. Při výpočtu pravděpodobností $P(X \leq x)$ resp. $P(X \geq x)$ pomocí normální aproximace jsou totiž výsledky podhodnocené. Naopak při výpočtu pravděpodobností $P(X < x)$ resp. $P(X > x)$ pomocí normální aproximace jsou výsledky nadhodnocené. Podstata výpočtů pravděpodobností pomocí opravy na spojitost spočívá v tom, že výpočet pravděpodobnosti pro argument x provedeme přibližně pomocí argumentu $x + 0{,}5$ resp. $x - 0{,}5$. Příklady provedených korekcí ukazuje tabulka 1.

před opravou	$x < 3$	$x \leq 3$	$x = 5$	$x \geq 7$	$x > 7$
po opravě	$x < 2{,}5$	$x < 3{,}5$	$4{,}5 < x < 5{,}5$	$x > 6{,}5$	$x > 7{,}5$

Tabulka 1: Oprava na spojitost

1.3 Příklady

Příklad 1.1 Pravděpodobnost zásahu cíle při jednom výstřelu je 0,8. Jaká je pravděpodobnost, že se počet zásahů při 200 výstřelech nebude lišit od střední hodnoty počtu zásahů o více než 10 zásahů?

Řešení:

a) Binomické rozdělení:

$$E(X) = n\pi = 200 \cdot 0{,}8 = 160 \text{ a } D(X) = n\pi(1-\pi) = 200 \cdot 0{,}8 \cdot (1-0{,}8) = 32$$

$$\begin{aligned} P(150 \leq X \leq 170) &= p(150) + p(151) + \cdots + p(170) = \\ &= \tbinom{200}{150} 0{,}8^{150} \cdot 0{,}2^{50} + \tbinom{200}{151} 0{,}8^{151} \cdot 0{,}2^{49} + \cdots + \tbinom{200}{170} 0{,}8^{170} \cdot 0{,}2^{30} = 0{,}937. \end{aligned}$$

b) Pomocí Moivre-Laplaceovy věty:

$$F(x) \approx \Phi\left(\frac{x - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

$$P(150 \leq X \leq 170) = F(170) - F(149) \approx \Phi\left(\frac{170-160}{\sqrt{32}}\right) - \Phi\left(\frac{149-160}{\sqrt{32}}\right) = 0{,}936.$$

c) Pomocí Moivre-Laplaceovy věty (s opravou na spojitost):

$$F(x) \approx \Phi\left(\frac{x - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

$$\begin{aligned} P(150 \leq X \leq 170) &\approx P(149{,}5 < X < 170{,}5) = F(170{,}5) - F(149{,}5) = \\ &= \Phi\left(\tfrac{170{,}5-160}{\sqrt{32}}\right) - \Phi\left(\tfrac{149{,}5-160}{\sqrt{32}}\right) = 0{,}937. \end{aligned}$$

d) Pomocí Čebyševovy nerovnosti:

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{D(X)}{\epsilon^2}$$

$$E(X) = n\pi = 200 \cdot 0{,}8 = 160 \text{ a } D(X) = n\pi(1-\pi) = 200 \cdot 0{,}8 \cdot (1-0{,}8) = 32$$

$$P(|X - 160)| < 10) \geq 1 - \frac{32}{10^2} = 0{,}68$$

$$P(|X - 160)| < 11) \geq 1 - \frac{32}{11^2} = 0{,}736.$$

Příklad 1.2 Ve volbách dalo 52 % voličů hlas koaličním stranám. Jaká je pravděpodobnost, že při průzkumu veřejného mínění o rozsahu 2600 respondentů získala převahu opozice?

Řešení:

a) Binomické rozdělení:
$X \ldots$ počet respondentů volící opozici, $X \sim B(2600; 0{,}48)$

$$E(X) = n\pi = 2600 \cdot 0{,}48 = 1248 \text{ a } D(X) = n\pi(1-\pi) = 2600 \cdot 0{,}48 \cdot (1 - 0{,}48) = 648{,}96$$

$$\begin{aligned} P(X > 1300) &= 1 - P(X \leq 1300) = 1 - [p(0) + \cdots + p(1300)] = \\ &= 1 - \left[\tbinom{2600}{0} \cdot 0{,}48^0 \cdot 0{,}52^{2600} + \cdots + \tbinom{2600}{1300} \cdot 0{,}48^{1300} \cdot 0{,}52^{1300}\right] = \\ &= 1 - 0{,}98031 = 0{,}01969. \end{aligned}$$

b) Pomocí Moivre-Laplaceovy věty:

$$F(x) \approx \Phi\left(\frac{x - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

$$\begin{aligned} P(X > 1300) &= 1 - P(X \leq 1300) = 1 - F(1300) \approx \\ &\approx 1 - \Phi\left(\tfrac{1300-1248}{\sqrt{648{,}96}}\right) = 1 - 0{,}97939 = 0{,}02061. \end{aligned}$$

c) Pomocí Moivre-Laplaceovy věty (s opravou na spojitost):

$$F(x) \approx \Phi\left(\frac{x - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

$$\begin{aligned} P(X > 1300) &= 1 - P(X \leq 1300) \approx 1 - P(X < 1300{,}5) = 1 - \Phi\left(\tfrac{1300{,}5-1248}{\sqrt{648{,}96}}\right) = \\ &= 1 - 0{,}98034 = 0{,}01966. \end{aligned}$$

2 Výběrová šetření

2.1 Druhy výběrového šetření

Statistické šetření rozlišujeme

- úplné (vyčerpávající),
- neúplné (výběrové).

U výběrového šetření se snažíme o to, aby výběrový soubor dobře reprezentoval základní soubor. Za tohoto předpokladu je možné zobecnit poznatky z výběrového souboru na celý základní soubor. Výběrový soubor získáme buď **záměrným** nebo **náhodným výběrem**. Nejjednodušší náhodný výběr je tzv. **prostý náhodný výběr**, tj. přímý výběr, kde každá jednotka má stejnou pravděpodobnost výběru.

Problém výběru s vracením a bez vracení
Je-li $n/N \leq 0{,}05$ nebo základní soubor je nekonečný, hypotetický, považuje se požadavek na nezávislost za splněný, tedy mezi výběrem s vracením a bez vracení nebudeme dělat rozdíl.

Definice 2.1 Náhodné veličiny $X_1, X_2, \ldots, X_n$ jsou **nezávislé** právě tehdy, když pro libovolná čísla $x_1, x_2, \ldots, x_n \in \mathbb{R}$ platí

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \cdots P(X_n \leq x_n).$$

- Mějme náhodný vektor $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, jehož složky $X_1, X_2, \ldots, X_n$ jsou náhodné veličiny. Nechť

  $$F(\mathbf{x}) = F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

  je **sdružená distribuční funkce** a $F(x_1), F(x_2), \ldots, F(x_n)$ jsou distribuční funkce náhodných veličin $X_1, X_2, \ldots, X_n$. Náhodné veličiny $X_1, X_2, \ldots, X_n$ jsou nezávislé právě tehdy, když

  $$F(x_1, x_2, \ldots, x_n) = F(x_1) \cdot F(x_2) \cdots F(x_n).$$

- Je-li $\mathbf{X}$ náhodný vektor, jehož složky jsou diskrétní náhodné veličiny, funkce

  $$p(\mathbf{x}) = p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

  je **sdružená pravděpodobnostní funkce** a $p(x_1), p(x_2), \ldots, p(x_n)$ jsou pravděpodobnostní funkce náhodných veličin $X_1, X_2, \ldots, X_n$, pak platí:
  Náhodné veličiny $X_1, X_2, \ldots, X_n$ jsou nezávislé právě tehdy, když

  $$p(x_1, x_2, \ldots, x_n) = p(x_1) \cdot p(x_2) \cdots p(x_n).$$

- Je-li $\mathbf{X}$ náhodný vektor, jehož složky jsou spojité náhodné veličiny, funkce

  $$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$$

  je **sdružená funkce hustoty pravděpodobnosti** a $f(x_1), f(x_2), \ldots, f(x_n)$ jsou funkce hustoty pravděpodobnosti náhodných veličin $X_1, X_2, \ldots, X_n$, pak platí:
  Náhodné veličiny $X_1, X_2, \ldots, X_n$ jsou nezávislé právě tehdy, když

  $$f(x_1, x_2, \ldots, x_n) = f(x_1) \cdot f(x_2) \cdots f(x_n).$$

2.2 Náhodný výběr

U každé jednotky, která se dostane do výběrového souboru, zjistíme hodnotu zkoumaného znaku x_i ($i = 1, 2, \ldots, n$). Tuto hodnotu můžeme chápat jako jednu z možných hodnot náhodné veličiny X_i. Každá z těchto n náhodných veličin má stejné rozdělením.

Definice 2.2 Náhodný výběr o rozsahu n je posloupnost nezávislých náhodných veličin $X_1, X_2, \ldots, X_n$ se stejným rozdělení.

Můžeme ho chápat jako vektor $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Konkrétní realizaci budeme značit $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Naměřené hodnoty $x_1, x_2, \ldots, x_n$ nazýváme pozorování nebo také vstupní (empirická) data.

- Protože jsou náhodné veličiny $X_1, X_2, \ldots, X_n$ nezávislé a mají stejné rozdělení, platí pro distribuční funkci $F(\mathbf{x})$ náhodného výběru

$$F(\mathbf{x}) = F(x_1)F(x_2)\cdots F(x_n), \quad x_i \in \mathbb{R}.$$

Příklad 2.1 Nechť $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ je náhodný výběr z rovnoměrného rozdělení na intervalu $(0, 1)$. Určete distribuční funkci $F(\mathbf{x})$ náhodného výběru.

Řešení: $X_i \sim R(0, 1)$, tedy $F(x_i) = x_i$ pro $0 < x_i < 1$,

$$F(\mathbf{x}) = F(x_1)F(x_2)\cdots F(x_n) = x_1 \cdot x_2 \cdots x_n.$$

- Pravděpodobnostní funkce $p(\mathbf{x})$ náhodného výběru v případě diskrétního rozdělení veličin $X_1, X_2, \ldots, X_n$ je

$$p(\mathbf{x}) = p(x_1)p(x_2)\cdots p(x_n), \quad x_i \in \mathbb{R}.$$

Příklad 2.2 Nechť $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ je náhodný výběr z Poissonova rozdělení s parametrem λ. Určete pravděpodobnostní funkci $p(\mathbf{x})$ náhodného výběru.

Řešení: $X_i \sim Po(\lambda)$, tedy $p(x_i) = \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$ pro $x_i = 0, 1, 2, \ldots, i = 1, 2, \ldots, n$

$$p(\mathbf{x}) = \frac{\lambda^{x_1}}{x_1!}e^{-\lambda} \cdots \frac{\lambda^{x_n}}{x_n!}e^{-\lambda} = \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \frac{1}{x_1! \cdot x_2! \cdots x_n!}.$$

- Hustota rozdělení $f(\mathbf{x})$ náhodného výběru z rozdělení s hustotou $f(x)$ je

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n),$$

kde $x_i \in \mathbb{R}, i = 1, 2, \ldots, n$.

Příklad 2.3 Nechť $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ je náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$. Určete hustotu pravděpodobnosti $f(\mathbf{x})$ náhodného výběru.

Řešení: $X_i \sim N(\mu, \sigma^2)$, tedy $f(x_i) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ pro $x_i \in \mathbb{R}$, $i = 1, 2, \ldots, n$

$$f(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

2.3 Výběrové charakteristiky

Definice 2.3 Funkce náhodných veličin $X_1, X_2, \ldots, X_n$ se nazývá **statistika**

$$T = T(X_1, X_2, \ldots, X_n) = T(\mathbf{X}).$$

- Výběrový úhrn $M = \sum_{i=1}^n X_i$
- Výběrový průměr $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$

- Výběrový rozptyl $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
- Výběrová směrodatná odchylka $S = \sqrt{S^2}$
- Základní rozptyl výběru $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n-1}{n}S^2$
- Výběrový r-tý obecný moment $M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$
- Výběrový r-tý centrální moment $M_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r$
- Výběrový koeficient šikmosti $A_3 = \frac{M_3}{M_2^{3/2}}$
- Výběrový koeficient špičatosti $A_4 = \frac{M_4}{M_2^2} - 3$

Rozdělení výběrového úhrnu

Mějme náhodný výběr $X_1, X_2, \ldots, X_n$ z rozdělení se střední hodnotou μ a rozptylem σ^2, tzn. $E(X_i) = \mu$, $D(X_i) = \sigma^2$, pro $i = 1, 2 \ldots, n$. Střední hodnota a rozptyl výběrového úhrnu jsou

$$E(M) = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E(X_i) = n\mu,$$

$$D(M) = D\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} D(X_i) = n\sigma^2.$$

Věta 2.1 Je-li $X_1, X_2, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$, má také výběrový úhrn normální rozdělení

$$M \sim N(n\mu, n\sigma^2).$$

Rozdělení výběrového průměru

Mějme náhodný výběr $X_1, X_2, \ldots, X_n$ z rozdělení se střední hodnotou μ a rozptylem σ^2, tzn. $E(X_i) = \mu$, $D(X_i) = \sigma^2$, pro $i = 1, 2 \ldots, n$. Střední hodnota a rozptyl výběrového průměru jsou

$$E(\overline{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \mu,$$

$$D(\overline{X}) = D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

Věta 2.2 Je-li $X_1, X_2, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$, má také výběrový průměr normální rozdělení

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Normováním dostaneme statistiku $Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}$, která má normované normální rozdělení $N(0,1)$. Je-li $X_1, X_2, \ldots, X_n$ náhodný výběr z rozdělení se střední hodnotou μ a rozptylem σ^2, pak ná veličina

$$Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}$$

pro $n \geq 30$ přibližně normované normální rozdělení $N(0,1)$ – viz centrální limitní věta.

Rozdělení výběrového rozptylu

Mějme náhodný výběr $X_1, X_2, \ldots, X_n$ z rozdělení se střední hodnotou μ a rozptylem σ^2, tzn. $E(X_i) = \mu$, $D(X_i) = \sigma^2$, pro $i = 1, 2 \ldots, n$. Střední hodnoty výběrového a základního rozptylu jsou

$$E(S^2) = \sigma^2, \quad E(S_n^2) = \frac{n-1}{n}\sigma^2.$$

Při odvození střední hodnoty výběrového rozptylu použijeme vztahy:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2$$

$$D(X_i) = E(X_i^2) - E(X_i)^2 \to E(X_i^2) = D(X_i) + E(X_i)^2 = \sigma^2 + \mu^2$$

$$D(\overline{X}) = E(\overline{X}^2) - E(\overline{X})^2 \to E(\overline{X}^2) = D(\overline{X}) + E(\overline{X})^2 = \frac{\sigma^2}{n} + \mu^2$$

$$\begin{aligned} E(S_n^2) &= E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - E(\overline{X}^2) = \\ &= \frac{1}{n}\sum_{i=1}^{n}E(X_i^2) - E(\overline{X}^2) = \frac{1}{n}n(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \\ &= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \end{aligned}$$

$$E(S^2) = E\left(\frac{n}{n-1}S_n^2\right) = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2$$

Věta 2.3 Mějme náhodný výběr z normálního rozdělení se střední hodnotou μ a rozptylem σ^2. Náhodná veličina

$$\chi^2 = \frac{n-1}{\sigma^2}S^2$$

má χ^2-rozdělení s $n-1$ stupni volnosti.

Předpokládejme, že máme výběr z normálního rozdělení se střední hodnotou μ a rozptylem σ^2. Víme, že $Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$ a $\chi^2 = \frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$. Náhodná veličina

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{n-1}}} = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}\cdot\frac{\sqrt{n-1}}{\sqrt{\frac{n-1}{\sigma^2}S^2}} = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}\cdot\frac{\sigma}{S} = \frac{\overline{X}-\mu}{S}\sqrt{n}$$

má Studentovo t-rozdělení s $n-1$ stupni volnosti.

Věta 2.4 Mějme náhodný výběr z normálního rozdělení se střední hodnotou μ a rozptylem σ^2. Náhodná veličina

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n}$$

má Studentovo t-rozdělení s $n - 1$ stupni volnosti.

Rozdělení výběrového podílu

Předpokládejme, že rozdělení v základním souboru je alternativní s parametrem π. Náhodný výběr mohou být buď jedničky nebo nuly. Náhodnou veličina $X = X_1 + X_2 + \cdots + X_n$ potom určuje počet jedniček ve výběru (tzv. **výběrová absolutní četnost**). Podíl

$$P = \frac{X}{n}$$

bývá označován jako **výběrová relativní četnost** nebo častěji **výběrový podíl**. Předpokládejme, že máme náhodný výběr velkého rozsahu n z alternativního rozdělení s parametrem π. Náhodná veličina $P = \frac{X}{n}$ má přibližně normální rozdělení se střední hodnotou π a směrodatnou odchylkou $\sqrt{\pi(1-\pi)/n}$ – viz centrální limitní věta. Z CLV plyne, že normovaná náhodná veličina

$$Z = \frac{P - \pi}{\sqrt{\pi(1-\pi)/n}}$$

má pro velká n přibližně normální rozdělení. Aproximace je vhodná, pokud $n\pi \geq 5$ a zároveň $n(1-\pi) \geq 5$.

2.4 Statistická indukce

Proces získávání závěrů o povaze celé populace pomocí výběru nazýváme ve statistice *induktivní statistické usuzování*. Tento proces je se všemi důležitými souvislostmi graficky znázorněn na obrázku 1.

Literatura

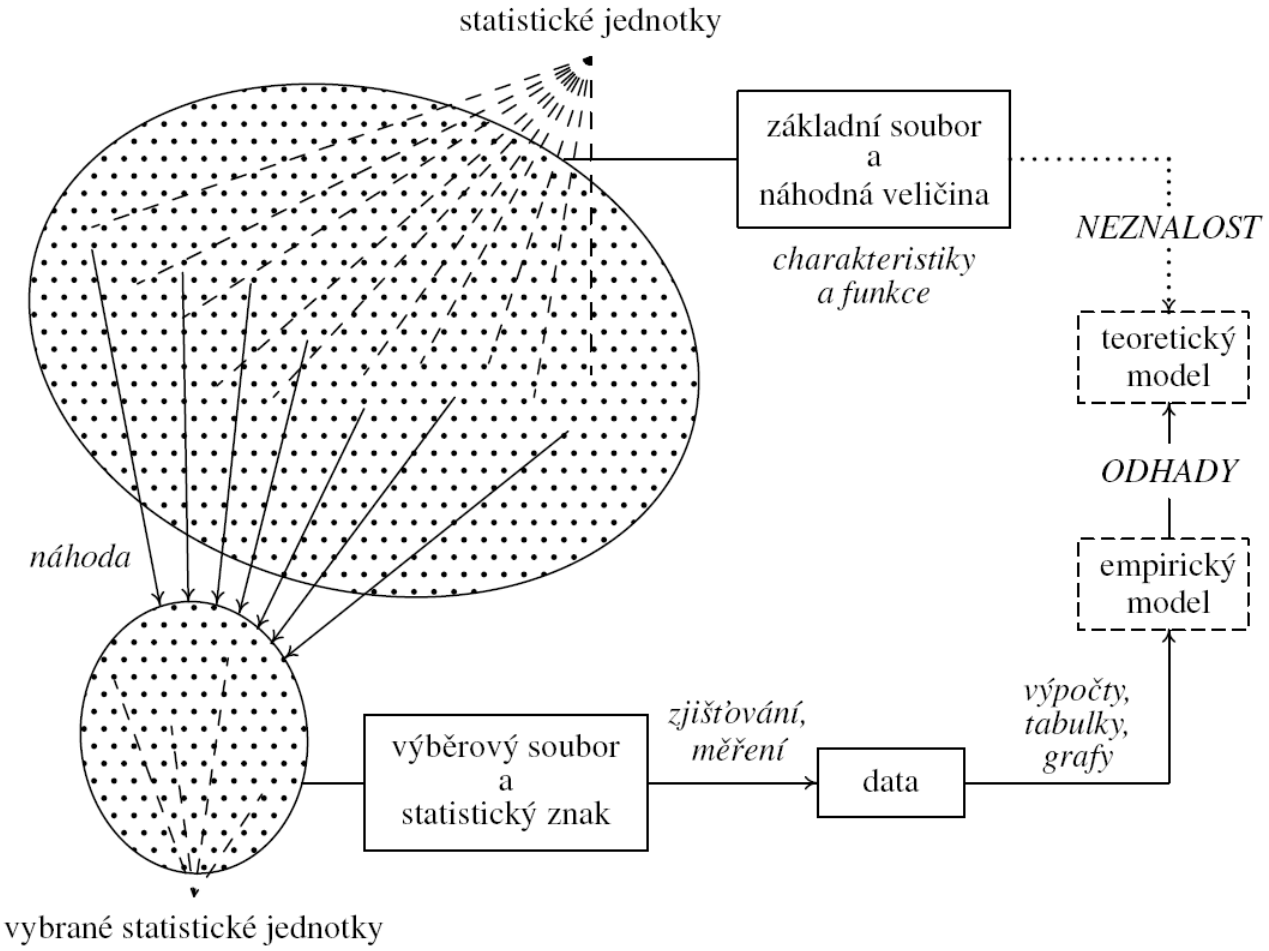
Základní

MANN, P.S. Introductory Statistics. 6th edition. Hoboken: Wiley, 2007. ISBN 978-0-471-75530-2.

MOUČKA, J., RÁDL, P. Matematika pro studenty ekonomie. 1. vyd. Grada 2010. ISBN 978-80-247-3260-2.

NEUBAUER, J., SEDLAČÍK, M., KŘÍŽ, O. Základy statistiky – Aplikace v technických a ekonomických oborech. Grada 2012.ISBN: 978-80-247-4273-1.

ŘEZANKOVÁ, H. Analýza dat z dotazníkových šetření. 2. vydání, Professional Publishing, 2010. ISBN: 9788074310195.



Obrázek 1: Populace, výběr, teoretický a empirický model

Doporučená

AGRESTI, A. Categorical Data Analysis. Second Edition. Wiley 2002. ISBN: 0-471-36093-7.

ANDĚL, J. Statistické metody. 3. vydání. Praha: Matfyzpress, 2003. ISBN 80-86732-08-8.

ANDĚL, J. Základy matematické statistiky. 2. vyd. Praha: Matfyzpress, 2007, 358 s. ISBN 978-80-7378-001-2.

VÁGNER, M. Integrální počet funkcí jedné proměnné. 1. vydání. Brno: UO, 2005,126 s. ISBN 80-7231-025-9.

VÁGNER, M., KAŠTÁNKOVÁ, V. Posloupnosti a řady. 1. vydání. Brno: UO, 2006. ISBN 80-7231-131-X.

Úkoly pro samostatnou práci

Limitní věty teorie pravděpodobnosti

1. Dlouhodobým pozorováním bylo zjištěno, že doba potřebná k nalezení a odstranění poruchy stroje má střední hodnotu 30 minut a směrodatnou odchylku 12 minut. Určete
   a) dobu, kterou si vyžádá nalezení a odstranění poruchy u 40 strojů, požadujeme-li, aby tato doba nebyla s pravděpodobností 0,95 překročena,
   b) pravděpodobnost, že průměrná doba nalezení a odstranění poruchy u 40 strojů nepřekročí 32 minut.

2. Z chovného rybníka bylo vyloveno 15 kaprů a po zvážení byli puštěni zpět. Na základě naměřených hmotností byla odhadnuta střední hodnota 2,2 kg a směrodatná odchylka 0,6 kg. Předpokládejme, že hmotnost kaprů se řídí normálním rozdělením. Do rybníka bylo vysazeno 1500 ks kapřího plůdku a počítá se s 10% úmrtností. Jaká je pravděpodobnost, že
   a) náhodně vylovený kapr bude vážit méně než 2 kg,
   b) při výlovu celého rybníka získáme alespoň 3000 kg kaprů?

Řešení:

1. a) 1324; b) 0,853;
2. a) 0,371; b) 0,087.

Výběrová šetření

1. Rektorát vysoké školy připravuje průzkum názorů studentů na kvalitu výuky. Počet všech studentů je 1850, velikost výběrového souboru se předpokládá přibližně 50. Navrhněte konstrukci výběrového souboru.

2. V lednu provedla dopravní policie rozsáhlou akci, při které se zjišťovalo, zda vozidla mají „obuté“ zimní pneumatiky. Ze všech projíždějících automobilů se kontrole podrobilo každé desáté vozidlo. Celkem bylo kontrolováno 1463 vozidel, z nichž 97 nemělo zimní pneumatiky.

a) Která množina aut tvoří základní a která výběrový soubor?
b) Jaký podíl kontrolovaných vozidel neměl v pořádku pneumatiky?
c) Jaký náhodný výběr zde byl provedený?
d) Jakou pravděpodobnost mělo náhodně vybrané vozidlo, že bude kontrolované?

Řešení:

1. Mechanický výběr s krokem 40 nebo oblastní výběr.
2. a) základní soubor – všechna vozidla, výběrový soubor – kontrolovaná vozidla; b) 6,63 %; c) mechanický, s krokem 10; d) $0{,}10$.