

## 16. Testování statistických hypotéz

### Úvod

Při práci s daty se mnohdy spokojujeme s intervalovým či bodovým odhadem parametrů populace. V mnohých případech se však uchylujeme k jinému postupu, většinou jde o případy, kdy řešíme problémy související s typem rozdělení populace, s nezávislostí výběru atd., někdy jsme prostě nuceni stanovit určité výroky, které mají statistické pozadí a studovat jejich pravdivost. Zavádíme proto pojem **statistické hypotézy**.

#### Definice 16.1

**Statistickou hypotézou** budeme rozumět výrok o náhodných veličinách. V případě, že tento výrok pojednává o parametrech populace nazýváme takovou hypotézu **parametrickou**, jestliže vyšetřujeme jiné okolnosti (např. typ rozdělení populace, nezávislost výběru, způsob výběru atd.) nazýváme takovou hypotézu **neparametrickou**.

Hypotézami parametrickými se budeme zabývat v následující 16. kapitole a neparametrickými v kapitole 17. a kapitole 18.

Obecně je možno dělit statistické hypotézy ještě mnoho dalšími způsoby např. podle počtu šetřených populací, podle toho zda jsou jednoduché nebo složené. Těmito okolnostmi se budeme zabývat víceméně okrajově.

Výrok, jehož platnost ověřujeme, nazýváme většinou **nulovou** nebo jinak **testovanou** hypotézou. Podle ustáleného označení takový výrok popisujeme symbolem  $H_0$ . Hypotéza  $H_0$  v tomto případě vyjadřuje dosavadní představy o těch skutečnostech, které vyšetřujeme (odráží stav poznání, které máme v současné době o těchto skutečnostech). Názor pochybnosti o platnosti nulové hypotézy vyjadřuje v našem modelu **alternativní** nebo jinak **výzkumná** hypotéza. Takovou hypotézu většinou stavíme tak, aby její pravdivost znamenala většinou pokrok či nějakou změnu v dosavadních šetřeních. Alternativní hypotézu většinou značíme  $H_1$  resp.  $H_A$ .

Testem statistické hypotézy rozumíme proces rozhodování, při kterém na základě náhodných výběrů provedeme rozhodnutí ve prospěch právě jedné z předložených hypotéz. Znamená to, že formulace hypotéz je prováděna tak, aby v daném okamžiku platila **právě jedna**. Z formálního hlediska představují tedy hypotézy  $H_0$  a  $H_1$  úplnou množinu neslučitelných jevů.

Abychom tedy provedli správný vyhodnocovací proces o správnosti předložených hypotéz (výroků), je nutné mít k dispozici nástroj, pomocí něhož rozhodujeme o správnosti předložených výroků.

Tento nástroj se nazývá **testovací statistika** nebo **testové kritérium**. Množina hodnot, kterých testová statistika nabývá se rozpadá na dvě disjunktní množiny, které nazýváme **obor přijetí** (testované hypotézy  $H_0$ )  $\mathcal{U}$  a **kritický obor** (obor zamítnutí hypotézy  $H_0$ )  $\mathcal{W}$ .

Kritický obor  $\mathcal{W}$  obsahuje takové hodnoty, že pravděpodobnost jejich výskytu je velmi malá. Jestliže přesto hodnota testové statistiky je prvkem množiny  $\mathcal{W}$ , zamítáme testovanou hypotézu  $H_0$  ve prospěch alternativní hypotézy  $H_1$ . Padne-li však hodnota testové statistiky do množiny  $\mathcal{U}$  říkáme, že uvedený test neprokázal nepravdivost testované hypotézy  $H_0$ .

Při takovémto postupu samozřejmě mohou nastat chyby v našich rozhodnutích. Zásadou je takové chyby studovat a pomocí jejich výskytu potom oceňovat vhodnost či

nevhodnost použité hypotézy. Jestliže chybně zamítneme hypotézu  $H_0$  vytvoříme **chybu I. druhu**, pravděpodobnosti jejího výskytu je označována jako  $\alpha$  (někdy jako  $1 - \alpha$ ), tomuto číslu se pak říká **hladina významnosti**. V praktických úlohách se vyskytují převážně hodnoty  $\alpha = 0,95$  (resp.  $0,05$ ) – říkáme, že tato hladina je **významná**; pro hladinu  $\alpha = 0,99$  prohlašujeme, že hladina je **velmi významná** a konečně pro hodnotu  $\alpha = 0,99$  je hladina **vysoce významná**.

Kromě chyby prvního druhu se můžeme dopustit při práci s hypotézami i chyby, kdy nezamítneme hypotézu  $H_0$ , i když tato hypotéza neplatí. Takovéto chybě se říká **chyba II. druhu**, pravděpodobnost výskytu takové chyby označujeme  $\beta$ . Doplněk hodnoty  $\beta$  do jedné je  $1 - \beta$  je tedy pravděpodobnost, že hodnota testovací statistiky správně spadne do kritického oboru  $W$ . Platí tedy následující (symbolem  $T$  označujeme testovací statistiku)

$$P(T \in W | H_0) = \alpha \quad (16.1)$$

$$P(T \in V | H_1) = \beta \quad (16.2)$$

$$P(T \in W | H_1) = 1 - \beta \quad (16.3)$$

Pro posouzení kvality testovacího postupu je důležitá **silofunkce**, vyjadřuje průběh pravděpodobnosti hodnoty  $1 - \beta$  při různých hodnotách  $\alpha$ . Nepodaří-li se nám prokázat, že hodnota testové statistiky je v kritickém oboru, musíme se spokojit s tvrzením, že daná data nejsou v rozporu s tvrzením hypotézy  $H_0$ .

#### **Popišme si tedy dále vlastní postup při práci s testováním statistických hypotéz:**

- I. Na základě znalostí matematického problému stanovujeme hypotézu  $H_0$ . Tuto hypotézu budeme považovat za pravdivou do okamžiku, kdy vybraná data neprokážou opak.
- II. Stanovení alternativní hypotézy  $H_1$ . Jde o tvrzení, které je v rozporu s hypotézou  $H_0$ . Alternativní hypotéza je považována za pravdivou po vyvrácení hypotézy  $H_0$ .
- III. Volba testovací statistiky  $T$ . Jde o náhodnou veličinu na základě níž rozhodneme o pravdivosti jednotlivých hypotéz.
- IV. Stanovení hladiny významnosti testu, stanovíme pravděpodobnost chyby **I. druhu**
- V. Stanovení kritického oboru  $W$  testovací statistiky  $T$ . Konkrétní stanovení kritického oboru závisí na hladině významnosti  $\alpha$
- VI. Ověření předpokladů testu. Některé testy vyžadují splnění určitých podmínek, které musíme pro naše konkrétní data ověřit
- VII. Vlastní provedení náhodného výběru z populace by mělo být provedeno korektně. Podmínky takového výběru by měli být podrobně ověřeny. Data poté musíme přesně interpretovat při výpočtu v testovací statistice.
- VIII. Po tomto procesu je zapotřebí rozhodnout o dvou možných výrocích:
  - a. Hypotéza  $H_0$  se zamítá a alternativní hypotéza se přijímá. Takovýto výrok připouštíme tehdy, když hodnota testovací statistiky leží v kritickém oboru  $W$ . V tomto případě je zapotřebí konstatovat, že je prokázána pravdivost alternativní hypotézy  $H_1$  nebo, že rozdíl mezi předpokládanou a vypočítanou hodnotou se ukázal na dané hladině významnosti jako statisticky významný.

- a. Hypotézu  $H_0$  nezamítáme v případě, je-li hodnota testovací statistiky v oboru přijetí  $\mathcal{D}$ . Většinou konstatujeme, že na dané hladině významnosti nebyla testovaná hypotéza  $H_0$  zamítnuta..

V další části využijeme a upravíme příklad 2.3, který je uvedený ve skriptech Praktikum k výuce matematické statistiky II: Testování hypotéz, VŠE Praha, 2000.

### Příklad 16.1

Předpokládejme, že výsledky experimentů souhlasily s náhodnou veličinou  $X$  s normálním rozdělením se střední hodnotou  $\mu$  a rozptylem 64. Dosud uznávaná hodnota  $\mu$  je 25. Při vyhodnocování hodnot experimentů došlo k určitým změnám a nyní se zdá, že hodnota  $\mu$  vzrostla a je větší než 25. Úkolem je na základě  $n$  nezávislých experimentů rozhodnout, zda platí hypotéza  $H_0 : \mu \leq 25$  nebo alternativní hypotéza  $H_1 : \mu > 25$ .

K tomu, abychom mohli některou z uvedených hypotéz zamítnout, je třeba nalézt vhodnou statistiku. Již z předchozích přednášek víme, že nejlepším bodovým odhadem  $\mu$  je

aritmetický průměr  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ , jako testovou statistiku použijeme tedy realizaci tohoto průměru. Provedeme nyní řešení našeho problému třemi různými způsoby.

### Varianta 1.

Provedli jsme náhodný výběr o rozsahu 40 prvků. Stanovíme hypotézu  $H_0 : \mu \leq 25$  a alternativní hypotézu  $H_1 : \mu > 25$ . Kritickým oborem bude stanovena množina

$$W = \left\{ (x_1, \dots, x_n); \sum_{i=1}^n x_i > 40 \cdot 25 = 1000 \right\}.$$

Zjistíme hodnotu silofunkce  $K(\beta) = 1 - \beta(\mu)$ . Silofunkce udává pro různé hodnoty  $\mu$  pravděpodobnost, že výběrový průměr bude větší než 25. Při náhodném výběru z populace popsané  $N(\mu, 64)$  víme, že výběrový průměr je typu normální rozdělení  $N(\mu, \frac{64}{40} = \frac{8}{5})$ .

Můžeme tedy hodnotu  $K(\mu)$  přímo spočítat :

$$K(\mu) = P(\bar{X} > 25) = P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{8}{5}}} > \frac{25 - \mu}{\sqrt{\frac{8}{5}}}\right) = P\left(Z > \frac{25 - \mu}{\sqrt{\frac{8}{5}}}\right).$$

V dále uvedené tabulce jsou zaznamenány hodnoty  $K(\mu)$  :

| $K(\mu)$ | $\mu$ |
|----------|-------|
| 0,008853 | 22    |
| 0,056923 | 23    |
| 0,214598 | 24    |

| $K(\mu)$ | $\mu$ |
|----------|-------|
| 0,5      | 25    |
| 0,785402 | 26    |
| 0,943077 | 27    |
| 0,991147 | 28    |

Všimněme si, že pro  $\mu = 25$  je hypotéza  $H_0$  pravdivá, ale pravděpodobnost jejího zamítnutí je 50%. To je příliš vysoká pravděpodobnost zamítnutí pravdivé hypotézy !

### Varianta 2.

Alternativní hypotézu  $H_1 : \mu > 25$  přijmeme za podmínky, je – li výběrový průměr větší než 27. V tomto případě bude hodnota silofunkce  $K(\mu)$  rovna

$$K(\mu) = P(\bar{X} > 27) = 1 - \Phi\left(\frac{27 - \mu}{\sqrt{\frac{8}{5}}}\right) \quad (16.4)$$

Po dosazení stejných hodnot jako v minulé variantě získáváme následující tabulku:

| $K(\mu)$ | $\mu$ |
|----------|-------|
| 3,86E-05 | 22    |
| 0,000783 | 23    |
| 0,008853 | 24    |
| 0,056923 | 25    |
| 0,214598 | 26    |
| 0,5      | 27    |
| 0,785402 | 28    |

Pro hodnotu  $\mu = 25$  je nyní pravděpodobnost zamítnutí ( pravdivé ) hypotézy  $H_0$  malá na úrovni 5,7% , ale pro hodnotu 26 ( nepravdivé ) je pravděpodobnost zamítnutí jen 21,5% a to je trochu příliš.

### Varianta 3.

V této části budeme postupovat obráceně a určíme přímo hodnoty silofunkce  $K(\mu)$  v některých důležitých bodech a z těchto skutečností budeme chtít určit kritický obor  $W$  a rozsah výběru. Necht' tedy  $W = \left\{ (x_1, \dots, x_n); \sum_{i=1}^n x_i > n.c \right\}$ , kde hodnoty  $n$  a  $c$  jsou v dané chvíli neznámé a budeme je dále počítat. Za těchto podmínek má silofunkce hodnotu

$$K(\mu) = P(\bar{X} > c) = 1 - \Phi\left(\frac{c - \mu}{\sqrt{\frac{64}{n}}}\right) \quad (16.5)$$

Chtějme nyní , aby pro hodnotu  $\mu = 25$  byla síla testu rovna 0,05 a zároveň aby pro hodnotu  $\mu = 26$  byla síla testu 0,9. Získáme tedy po dosazení následující soustavu:

$$1 - \Phi \left( \frac{c - 25}{\sqrt{\frac{64}{n}}} \right) = 0,05 ,$$

$$1 - \Phi \left( \frac{c - 26}{\sqrt{\frac{64}{n}}} \right) = 0,90$$

Nyní využijeme kvantily  $N(0,1)$  a soustava se převede na

$$\frac{c - 25}{\sqrt{\frac{64}{n}}} = 1,64$$

$$\frac{c - 26}{\sqrt{\frac{64}{n}}} = -1,28$$

Řešením této soustavy jsou hodnoty  $n = 545,7$  a hodnota  $c = 25,56$ . Shrňme – li nyní získané výsledky měl by mít náš výběr rozsah aspoň 546 prvků a kritický obor by měl být omezen číslem 25,56.

### Příklad 16.2

Předávací cena mezi výrobcem a prodávajícím byla odvozena mimo jiné z toho, že podíl reklamovaných výrobků v záruční době byl 15%. Tato situace se jednou za čas ověřuje. Náhodným výběrem určitého počtu výrobků :

- Chtějí oba ověřit, že předpokládaný podíl je stále aktuální
- Výrobce chce ukázat, že podíl je nadhodnocený a cenu chce zvýšit
- Prodejce chce prokázat, že podíl je podhodnocený a cenu chce snížit.

Formulujte v těchto případech hypotézy .

### Řešení:

Je zřejmé , že ve všech případech je hypotéza  $H_0$  stejná a rovna :

$$H_0 : \pi = 0,15$$

Hodnota  $\pi$  je parametr alternativního rozdělení. Hypotéza  $H_0$  je jednoduchá( není složená ). Provedeme nyní konstrukci alternativních hypotéz pro jednotlivé případy.

- $H_1 : \pi \neq 0,15$  . Jde o dvoustrannou hypotézu ( složenou ).
- $H_1 : \pi < 0,15$  . Jde o jednostrannou hypotézu levostrannou ( složenou ).
- $H_1 : \pi > 0,15$  . Jde o jednostrannou hypotézu pravostrannou ( složenou ).

### Příklad 16.3

Pokračujme v našem příkladě. Výrobce předpokládá, že podíl zboží v reklamaci je 15%. Prodejce tvrdí, že v předcházejícím roce byl tento podíl 21% a chce proto cenu snížit. Nakonec se dohodnou , že vyberou 10 výrobků, jestliže budou ve výběru nejvýše 2 reklamované výrobky nebudou cenu měnit. V opačném případě se cena sníží. Jakých chyb a s jakou pravděpodobností se mohou oba dopustit?

### Řešení:

Provedeme konstrukci hypotéz .

$$H_0 : \pi = 0,15$$

$$H_1 : \pi = 0,21$$

Je zřejmé, že testová statistika je rovna binomickému rozdělení  $Bi(10;\pi)$ . Jestliže má pravdu výrobce je rovna hodnota  $\pi = 0,15$  , naopak má – li pravdu prodejce je hodnota  $\pi = 0,21$ . Podle zadání je jasné, že kritický obor je množina  $W = \{3, 4, \dots, 10\}$  a obor přijetí testované hypotézy je roven  $v = \{0, 1, 2\}$ .

- 16. Za předpokladu , že má pravdu výrobce ( platí  $H_0$  ) a v náhodném výběru budou 0,1,2 výrobky vede test k správnému přijetí testované hypotézy  $H_0$ .**
- 17. Za předpokladu, že pravdu má výrobce, ale ve výběru budou více než 2 reklamované výrobky , pak test vede k chybě I. druhu**
- 18. Za předpokladu, že pravdu má prodejce a v náhodném výběru budou více než dva reklamované výrobky , vede test ke správnému zamítnutí testované hypotézy a k přijetí hypotézy alternativní.**
- 19. Za předpokladu, že pravdu má prodejce a v náhodném výběru budou nejvýše dva výrobky reklamovány , pak test vede k chybě II. druhu.**

Vidíme, že body I. a III. vedou ke správným závěrům. Spočítáme nyní dále s jakými pravděpodobnostmi se dopouštíme chyb I. a II. druhu.

Pravděpodobnost chyby I. druhu označujeme  $\alpha$  a nazýváme ji hladina významnosti.

$$\alpha = \sum_{i=3}^{10} \binom{10}{i} \cdot 0,15^i \cdot (1-0,15)^{10-i} = 0,1798$$

Pravděpodobnost chyby druhého druhu označujeme  $\beta$

$$\beta = \sum_{i=0}^2 \binom{10}{i} \cdot 0,21^i \cdot (1-0,21)^{10-i} = 0,6474 .$$

#### **Příklad 16.4**

Provedli jsme výběr s rozsahem  $n = 400$  prvků z populace normálně rozdělené. Výběrový průměr  $\bar{X} = 50$  a výběrová směrodatná odchylka  $s = 116$ . Máme rozhodnout, zda můžeme předpokládat, že střední hodnota populace je rovna 47,5.

#### **Řešení:**

Zformulujeme nejprve testovanou hypotézu.

$$H_0 : \mu = 47,5$$

Formulujeme alternativní hypotézu

$$H_1 : \mu \neq 47,5$$

Testovací statistikou bude  $t = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}}$ , která je za předpokladu normality populace

studentovo rozdělení o  $n-1$  stupních volnosti. Protože je hodnota  $n > 100$  aproximujeme studentovo rozdělení rozdělením normálním.

Stanovíme hladinu významnosti na 5%, musíme ještě najít příslušný kvantil  $N(0,1)$ . Jde o 97,5% kvantil, který má hodnotu 1,96. Tedy kritickým oborem jsou všechny hodnoty větší nebo rovny hodnotě 1,96.

Stačí nyní dosadit do testovací statistiky za známé hodnoty a za hodnotu  $\mu = 47,5$ .

$$\frac{50 - 47,5}{\frac{18}{\sqrt{400}}} = \frac{2,5}{\frac{18}{20}} = \frac{50}{18} = 2,78.$$

Tato hodnota leží v oboru zamítnutí testované hypotézy a v oboru přijetí hypotézy alternativní. Kdybychom nyní zjišťovali při jaké hodnotě by za těchto předpokladů byla poprvé přijata testovaná hypotéza, museli bychom zjistit příslušný kvantil normovaného normálního rozdělení. Jde o 99,46%.

Museli bychom tedy volit  $100\% - 99,46\% = 0,54\%$  jako hladinu významnosti. Takováto hodnota bývá nazývána v programech pro zpracování statistických testů jako **p – value**.