# I Úvod do testování hypotéz

## I.I Co je to statistická hypotéza?

Pomocí testování statistických hypotéz chceme ověřit, na základě náhodného výběru, zda-li základní statistický soubor, z nějž náhodný výběr pochází, má nějakou vlastnost. Samotná definice je dosti obecná.

**Nový pojem: Statistická hypotéza**

Statistická hypotéza je jakékoliv tvrzení o rozdělení nějaké náhodné veličiny.

Co všechno se do této formulace vejde? Tak například:

- náhodný výběr pochází z normálního (nebo jakéhokoliv jiného) rozdělení.
- hodnota neznámho parametru $\vartheta$ je nějaká konkrétní hodnota (náhodný výběr pochází z rozdělení $L(\vartheta)$).
- dva náhodné výběry pochází z téhož rozdělení.
- dva náhodné výběry pochází z rozdělení, které mají stejnou střední hodnotu (rozptyl, šikmost, ...).

## I.II Formulace statistické hypotézy

Základem správně provedeného testu je dobře formulovaná dvojce hypotéz. Rozlišujeme nulovou hypotézu $H_0$ a její logický opak alternativní hypotézu $H_1$.

Nulová hypotéza se obvykle vyjadřuje pomocí rovnosti. Obvykle vyjadřuje, že data pocházejí z nějakého rozdělení, hodnota parametru je nějaká konkrétní hodnota, dva náhodné výběry pocházejí z téhož rozdělení apod.

Opakem (negací) je pak alternativní hypotéza, která vyjadřuje, že data z daného rozdělení nepochází, hodnota parametru není dané číslo, náhodné výběry nepochází z téhož rozdělení atd.

Obvykle dvojci nulové a alternatiní hypotézy zapisujeme matematicky pro samotné testování, ale také ji popíšeme v kontextu daného problému, například:

- $H_0$ : Neexistuje statisticky významný rozdíl mezi platy mužů a žen v ČR.
- $H_1$ : Mezi platy mužů a žen v ČR existuje statisticky významný rozdíl.

Takto je zapsána nulova a alternativní hypotéza srozumitelně. Pokud náhodnou veličinu, která popisuje platy mužů v ČR označíme $X$ a platy žen $Y$, statisticky tutéž dvojci hypotéz zapíšeme:

- $H_0: \ E(X) = E(Y)$
- $H_1: \ E(X) \neq E(Y)$

*Užitečná poznámka.* O takto formulované alternativní hypotéze mluvíme jako o oboustranné. Pokud ale máme důvod se domnívat, že platy žen jsou menší než platy mužů, mohli bychom volit složenou nulovou hypotézu $H_0: \ E(X) \leq E(Y)$ oproti jednoustranné alternativní hypotézu $H_1: \ E(X) > E(Y)$. Vidíte podobnost s konstrukcí oboustranných a jednostranných intervalů spolehlivosti?

## I.III Chyby

Žádný test není spolehlivý na 100%. Ukážeme si na příkladu, jak můžeme během testování statistických hypotéz dojít ke špatnému závěru, ačkoliv samotná metodika testování je správně.

**Motivační příklad.** Při použití těhotenského testu se testuje nulová hypotéza:

- $H_0$ : Jsem těhotná.

oproti alternativní hypotéze:

- $H_1$ : Nejsem těhotná.

Možná vás zajímá, jak tuto hypotézu napsat matematickým zápisem. Na to je třeba znalost principu fungování těhotenského testu. Po početí se začne tvořit hormon hCG (human chorionic gonadotropin), na jehož přítomnost v moči reagují chemikálie, kterými je napuštěn papírek v těhotenském testu. Výrobce vždy udává, jak vysokou hladinu $\vartheta_0$ v moči je schopný test detekovat. Pokud tedy označíme koncentraci hCG v moči jako $\vartheta$, pak testujeme složenou nulovou hypotézu

- $H_0: \ \vartheta \leq \vartheta_0,$

oproti alternativní hypotéze

- $H_0: \ \vartheta > \vartheta_0.$

Jenže občas se stane chyba. Asi v 5% případů je výsledek testu falešně negativní. To znamená, že ženě, která je těhotná test ukáže, že těhotná není. Této chybě se říká chyba I. druhu.

Naopak, asi v 1% případů vyjde test falešně pozitivní – tedy ženě, která těhotná není, test ukáže, že těhotná je. Této chybě se říká chyba II. druhu.

**Nový pojem: Chyba I. a II. druhu**

Chyba I. druhu při testování statistické hypotézy nastane, pokud ve skutečnosti $H_0$ platí, ale test ji zamítne.
Chyba II. druhu při testování statistické hypotézy nastane, pokud ve skutečnosti $H_0$ neplatí, ale test ji nezamítne.

Pravděpodobnost, že dojde k chybě I. druhu značíme $\alpha$ – je to riziko, které volíme libovolně. Také se jí říká hladina významnosti, volíme ji libovolně, obvykle se používá 5% (např. testy provedené na hladině významnosti 5% jsou již u soudu brány jako důkazy). U příkladu s těhotenským testem je to citlivost $\vartheta_0$ – čím nižší test stanovuje, tím spíše zachytí těhotenství. Jenže, čím menší hladinu významnosti zvolíme, tím menší bude síla testu $(1 - \beta)$, kde $\beta$ je pravděpodobnost chyby II. druhu. Pravděpodobnost chyby I. druhu $\alpha$ a II. druhu $\beta$) jsou spolu provázány, takže jakmile $\alpha$ zvolíme, $\beta$ už nemůžeme ovlivnit.

## I.IV Samotné testování hypotéz

Statistické testy se dělí na dva druhy:

- parametrické – vyžadují znalost rozdělení, ze kterého náhodný výběr pochází. Jsou silnější.
- neparametrické – můžeme je použít, aniž bychom znali rozdělení, ze kterého náhodný výběr pochází, jsou ale slabší.

Ve všech testech je třeba určit testovací statistiku. Ta u parametrických testů bývá shodná s pivotovými statistikami, které již známe. V případě neparametrických testů jsou konstruované speciální statistiky. Různé testy si vysvětlíme ve zvláštních materiálech, zde je jen vysvětlen pricnip, jak fungují. Ten je důležitý pochopit.

Pro samotný průběh testu se využívá tří postupů. Všechny jsou ekvivalentní, tj. vedou ke stejnému závěru.

### Testování pomocí intervalu spolehlivosti

Chceme na hladině významnosti $\alpha$ otestovat dvojici nulové a alternativní hypotézy:

- $H_0 : \ \vartheta = \vartheta_0$,
- $H_1 : \ \vartheta \neq \vartheta_0$.

Pomocí pivotové statistiky zkonstruujeme $(1-\alpha)\cdot 100\%$ interval spolehlivosti pro parametr $\vartheta$, který označíme $V$ a nazýváme jej obor nezamítnutí nulové hypotézy. Zbylé přípustné hodnoty (obvykle $\mathbb{R} \setminus V$, popř. $\mathbb{R}_+ \setminus V$) značíme $W$ – kritický obor.

Pokud $\vartheta \in V$, pak nulovou hypotézu nezamítáme. Analogicky, pokud $\vartheta \in W$, pak nulovou hypotézu na hladině významnosti $\alpha$ zamítáme.

*Užitečná poznámka.* V případě, že nulová hypotéza je dána nerovností $H_0 : \vartheta \leq \vartheta_0$, pak konstruujeme levostranný interval spolehlivosti jako obor nezamítnutí $H_0$, naopak kritický obor $W$ je pravostranný. Analogicky pro $H_0$ zadáno opačnou nerovností.

*Interpretační poznámka.* Tento postup je velmi intuitivní – tím, že zkonstruujeme $(1-\alpha)\cdot 100\%$ interval spolehlivosti, máme všechny možné hodnoty parametru $\vartheta$, pro které bychom nulovou hypotézu nezamítnuli. Pokud je $\vartheta_0$ mezi nimi, pak $H_0$ nezmítneme, pokud $\vartheta_0$ neleží v intervalu spolehlivosti, pak $H_0$ zamítáme.

### Testování pomocí testového kritéria

Postup je podobný, jako při testování pomocí intervalu spolehlivosti. Opět zvolíme testovací statistiku $T(\boldsymbol{X};\vartheta)$ a spočítáme testovací kritérium $T_0(\boldsymbol{x},\vartheta_0)$.

Protože známe rozdělení testovací statistiky $T(\boldsymbol{X})$ označme ho $t$, můžeme nalézt kvantily tohoto rozdělení a určit kritický obor W (obor zamítnutí nulové hypotézy).Ten má obvykle tvar

- $W = (-\infty, t_{\alpha/2}\,\rangle \cup \langle\, 1 - t_{\alpha/2}, \infty)$ pro $H_0 : \ \vartheta = \vartheta_0$ oproti $H_1 : \ \vartheta \neq \vartheta_0$.
- $W = (-\infty, t_\alpha\rangle$ pro $H_0 : \ \vartheta \geq \vartheta_0$ oproti $H_1 : \ \vartheta < \vartheta_0$.
- $W = \langle t_{1-\alpha}, \infty)$ pro $H_0 : \ \vartheta \leq \vartheta_0$ oproti $H_1 : \ \vartheta > \vartheta_0$.

Pokud $\vartheta_0 \in W$, pak nulovou hypotézu zamítáme. Naopak pro $\vartheta_0 \notin W$ nulovou hypotézu nezamítáme na hladině významnosti 0,05.

*Interpretační poznámka.* Testování pomocí intervalu spolehlivosti a tetovacího kritéria je totožné. Testujeme-li pomocí testovací statistiky, tak na náhodnou veličinu $\vartheta$ aplikujeme transformaci $T(\boldsymbol{X};\vartheta)$. Pokud testujeme pomocí intervalu spolehlivosti, tak $\alpha$-kvantily transformujeme inverzní transformací $T^{-1}$.

### Testování pomocí $p$-hodnoty

Pokud pracujeme se statistickým softwarem, obvykle nezamítne nulovou hypotézu za nás, ale uvede tzv. $p$-hodnotu.

**Nový pojem 1: $p$-hodnota**

Na základně daného náhodného výběru a testovací statistiky udává $p$-hodnota nejnižší hladinu významnosti $\alpha$, při které nulovou hypotézu zamítáme.

Pracujeme-li tedy na hladině významnosti $\alpha$, pak $H_0$ zamítáme, pokud $p < \alpha$ a naopak – při $p > \alpha$, pak $H_0$ nezamítáme.

Výhodou $p$-hodnoty je, že nezávisí na zvolené hladině významnosti $\alpha$. Například pro velmi malou $p$-hodnotu (p<0,001) nulová hypotéza takřka určitě neplatí, naopak pro $p$-hodnoty blízké jedničce nemáme důvod pochybovat o $H_0$ na základě našeho náhodného výběru.

Při testování pomocí softwaru je třeba si dávat pozor, s jakou nulovou hypotézou pracujeme (dohledejte si v nápovědě programu).

## I.V Vyslovení závěru o testované hypotéze

Vždy se ve výsledcích vyjadřujeme k nulové hypotéze, tedy „Na hladině významnosti $\alpha$ nulovou hypotézu $H_0$ nezamítáme.“ nebo naopak „Na hladině významnosti $\alpha$ nulovou hypotézu $H_0$ zamítáme ve prospěch alternativní hypotézy $H_1$“.Můžeme samozřejmě použít i jiné formulace, ale nepoužíváme „definitivních“ závěrů – vždy je riziko, že jsme se dopustili chyby I. nebo II. druhu. V závěru by se měla objevit i hladina významnosti $\alpha$.

Dále by se měl obejvit slovní komentář, co jsme vlastně testovali a co zamítnutí (nezamítnutí) nulové hypotézy znamená v kontextu problému.