

# 6.

## Testování statistických hypotéz



### Cíl kapitoly

Údaje, se kterými sociálně hospodářská statistika pracuje jsou typické velkou mírou nejistoty. Jak jsme již uvedli v předchozích kapitolách, většina dat pochází z výběrových šetření a jsou tedy do jisté míry pouze odhadem skutečnosti. Pro seriózní kvantitativní analýzy je proto vhodné považovat výsledky šetření za hypotézy o vlastnostech zkoumaných statistických souborů. K ověření platnosti, či neplatnosti těchto hypotéz je možno užít metod statistických testů. Následující kapitola prezentuje základní východiska testování a postup, který je nutno při provádění statistického testu dodržet. Konkrétními statistickými testy se bude zabývat kapitola sedmá.



### Časová zátěž

4 hodiny (4. týden v listopadu)

Statistická  
hypotéza

Další metodou jak zhodnotit validitu údajů získaných ze statistických šetření je využití metod testování statistických hypotéz. Ve smyslu těchto metod se za **statistickou hypotézu považuje jakýkoli předpoklad o vlastnostech statistického souboru.**

Typické zdroje hypotéz

- Požadavek na potřebnou kvalitu produktu
- Hypotéza je založena na předchozí zkušenosti
- Hypotéza vychází z teorie, kterou je třeba doložit
- Hypotéza je pouhým dohadem, založeným na náhodném pozorování

Testování  
statis-  
tických  
hypotéz

Testování tudíž slouží k podobným účelům jako jsou metody intervalových a bodových odhadů. Například vyslovíme hypotézu o hodnotě aritmetického průměru základního souboru. O správnosti naší hypotézy bychom se nejspolehlivěji mohli přesvědčit na základě vyčerpávajícího šetření. Provedení takového šetření však v mnoha případech není možné (například z důvodů časové, finanční, či organizační náročnosti). Proto je k ověření hypotézy používán pouze vzorek statistických jednotek získaný pomocí výběrového šetření. Výběrový soubor, který je pomocí tohoto šetření získán je poté možno použít k rozhodnutí o správnosti, či nesprávnosti hypotézy. Tento proces je pak nazýván **statistickým testováním hypotéz.**



*Proces testování statistické hypotézy lze rozdělit do pěti kroků. Uvedené kroky představují sled úkonů, jejichž výsledkem je jednoznačný závěr v podobě zamítnutí nebo nezamítnutí hypotézy. Kapitola se zaměří na podrobný rozbor jednotlivých kroků. Příklady konkrétních statistických testů naleznete v následující kapitole. Pro hlubší porozumění je proto vhodné postupovat při studiu tak, že budete procházet obě kapitoly současně.*

Postup testování statistických hypotéz

1. Formulace testované (nulové) a alternativní hypotézy.
2. Výběr vhodného testu (testového kritéria).
3. Volba hladiny významnosti  $\alpha$  (obvykle  $0,05=5\%$ ).

4. Sestrojení kritického oboru.
5. Výpočet hodnoty testovacího kritéria.
6. Formulace výsledků testu: prokázání či neprokázání alternativní hypotézy (zamítnutí či nezamítnutí nulové hypotézy).

### Ad 1) Formulace testované (nulové) a alternativní hypotézy

Výchozím pojmem statistického testování je hypotéza. Jak jsme již uvedli hypotézou se rozumí nějaký předpoklad (tvrzení) o parametrech či tvaru rozdělení zkoumaného statistického znaku.

Např. *Aritmetický průměr statistického souboru = 3 (hypotéza o parametru).*

O správnosti takové hypotézy se můžeme přesvědčit tak, že:

- a) spočítáme tuto hodnotu (parametr) ze všech hodnot statistických znaků – velmi náročné pro velké soubory až nemožné
- b) použijeme jen vzorek statistického souboru – výběrový soubor a pro něj tento parametr určíme. Pomocí statistického testu následně rozhodneme, zda se dají závěry, které dostáváme pro výběrový soubor, zobecnit.

Původní předpoklad, který jsme si určili o parametru resp. tvaru rozdělení statistického znaku označujeme zpravidla za **nulovou hypotézu** a značíme ji  $H_0$ . Proti nulové hypotéze stavíme **alternativní hypotézu** – značíme ji  $H_1$ . Může být tří typů

- a) oboustranná hypotéza  $H_1 : \mu \neq \mu_0$  – oboustranný test
- b) jednostranná hypotéza – levostranná  $H_1 : \mu < \mu_0$
- c) jednostranná hypotéza – pravostranná  $H_1 : \mu > \mu_0$

Hypotéza bývá obvykle formulována ve tvaru rovnice, tedy proměnná = hodnotě (číslu). Například  $H_0 : \mu = \mu_0$ , kde  $\mu$  je předpokládaná hodnota parametru ( $=3$ ) kterou určíme z výběrového souboru.

*Je obvyklé formulovat samotnou testovou úlohu tak, aby skutečnost, kterou chceme dokázat byla obsažena v alternativní hypotéze  $H_1$ . Tato skutečnost souvisí s tzv. chybami I. a II. druhu, resp. statistickou významností, o které se ještě zmíníme při diskusi třetího postupného kroku testování.*

Nulová hypotéza



### Ad 2) Volba testového kritéria

**Testové kritérium** je zpravidla ve tvaru nějaké funkce, která souvisí s výběrem, resp. jeho vlastnostmi. Konkrétní podoba testového kritéria je pro konkrétní výběrový soubor standardně dána v závislosti na typu testované hypotézy. Existují zvláštní testová kritéria pro testování hypotéz o průměru, rozptylu či tvaru rozdělení statistického souboru.

Testové kritérium

**Standardní statistické testy** (a z nich vyplývající nejvýznamnější testová kritéria) se rozdělují do dvou významných skupin. A to na testy:

- **parametrické** – v tomto případě je předmětem hypotézy parametr rozdělení (které známe)
- **neparametrické** – předmětem hypotézy je buď samotný tvar rozdělení nebo tvar rozdělení není nutný pro testování



Jednotlivými konkrétními typy těchto testů se budeme zabývat v následující kapitole. Zatímco parametrické testy jsou užívány především k ověřování předpokladu o jedné konkrétní vlastnosti souboru, neparametrické testy vedou zpravidla k poněkud obecnějším závěrům o zkoumaném souboru.

### Ad 3) Volba hladiny významnosti

Jelikož při statistickém testování pracujeme pouze s výběrovými soubory (obvykle porízenými náhodným výběrem), jsou výsledky těchto testů zatíženy určitou mírou nespolehlivosti. Možné nezdary se označují jako chyby a jsou označovány jako chyba I. a chyba II. druhu.

Hladina významnosti

K **chybě I. druhu** dochází v případě, že nulová hypotéza sice platí, ale my ji na základě výsledku testu zamítáme. Maximální přípustnou pravděpodobnost této chyby nazýváme **hladina významnosti** a značíme symbolem  $\alpha$  **Hladina významnosti** ( $\alpha$ ) tedy označuje pravděpodobnost, že výsledek testu bude hovořit ve prospěch zamítnutí hypotézy  $H_0$ , která je ve skutečnosti pravdivá.

Síla testu

K **chybě II. druhu** dochází, když nulová hypotéza neplatí, ale my ji nezamítneme (nepoznáme, že neplatí). Doplněk pravděpodobnosti chyby II. druhu do jedničky ( $1 - \beta$ ) se nazývá **síla testu**. Je to pravděpodobnost, že nulovou hypotézu zamítneme, když tato hypotéza neplatí, tedy pravděpodobnost, s jakou neplatnost hypotézy objevíme. Síla testu závisí na zvolené testové metodě a zejména na tom, jaké je skutečné rozdělení dat (a tedy použité statistiky), například jaké jsou skutečné hodnoty parametru.

Čtyři možné kombinace výsledku testu v porovnání se skutečností přináší následující tabulka.

Výsledek testu/skutečnost	$H_0$ se přijímá	$H_0$ se zamítá
$H_0$ je pravdivá	<b>Správné rozhodnutí</b> pravd.= $1-\alpha$ = interval spolehlivosti	<b>Chyba I. druhu</b> pravd.= $\alpha$ (hladina významnosti)
$H_0$ je nepravdivá ( $H_1$ je pravdivá)	<b>Chyba II. druhu</b> pravd.= $\beta$	<b>Správné rozhodnutí</b> pravd.= $1-\beta$ (síla testu)

Hodnotu hladiny významnosti je nutno určit ještě před začátkem testování a je dána našimi požadavky na přesnost testu. Je nepřímo úměrná tzv. síle testu ( $\beta$ ) – pravděpodobnosti tzv. chyby II. druhu.

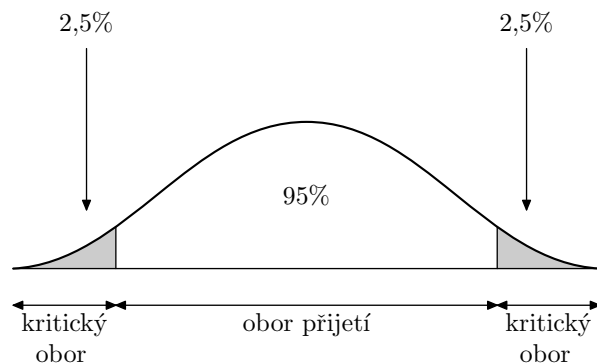
Hladina významnosti je obvykle stanovena na 0,05 (= 5%), případně na 0,01 (= 1%). Tato hodnota říká, že s 95% jistotou je zamítnutí hypotézy  $H_0$  (a platnost  $H_1$ ) správný závěr.

Nic však neříká o pravděpodobnosti (jistotě) potvrzení  $H_0$  (resp. zamítnutí  $H_1$ ). Tato pravděpodobnost je dána silou testu  $1 - \beta$  a její velikost je obtížně určitelná. Závisí na tvaru rozdělení, testovém kritériu i samotné hypotéze.

#### Ad 4) Sestrojení kritického oboru

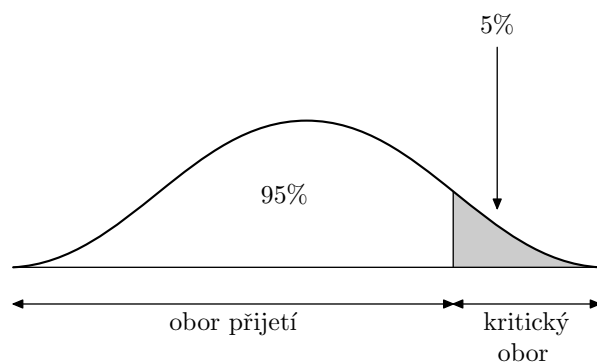
Aby bylo možno posoudit, jestli výsledky testu (hodnota testového kritéria) hovoří ve prospěch, či v neprospěch hypotézy  $H_0$ , je nutno vytvořit nějaké pomocné kritérium. Tímto kritériem je rozdělení všech možných výsledků do dvou částí (intervalů). První interval obsahuje výsledky, které hovoří ve prospěch platnosti  $H_0$  a nazývá se **obor přijetí**. Druhý interval hovoří v neprospěch přijetí  $H_0$  a tedy ve prospěch přijetí  $H_1$  a nazývá se **kritický obor**  $\langle T_{\min}; T_{\max} \rangle$ . Hranice kritického oboru tvoří kvantily některého ze standardního rozdělení náhodné veličiny.

Kritický obor



Obrázek 6.1: Kritický obor a obor přijetí (oboustranná varianta, hladina významnosti 5%).

V oboustranné variantě testu je kritický obor tvořen sjednocením dvou intervalů. Tyto intervaly jsou od oboru přijetí odděleny tzv. kritickými hodnotami. Kritické hodnoty jsou dány příslušným kvantilem rozdělení, které je pro testování použito (obsaženo v testovém kritériu). Plocha pod křivkou je pravděpodobnost a její součet pro kritický obor se musí rovnat hladině významnosti.



Obrázek 6.2: Kritický obor a obor přijetí (jednostranná varianta, hladina významnosti 5%).

V jednostranné variantě testu je kritický obor pouze jedním intervalem. Hovoří-li ve prospěch  $H_1$  vysoké hodnoty testového kritéria bude kritický obor pravostranný (viz obrázek 6.2), v opačném případě bude levostranný.

## 6. Testování statistických hypotéz

Tyto intervaly jsou od oboru přijetí opět odděleny kritickými hodnotami. Plocha pod křivkou je pravděpodobnost a pro kritický obor se musí rovnat hladině významnosti

### Ad 5) a 6) Výpočet hodnoty testovacího kritéria, Formulace výsledků testu

Na základě hodnot statistických znaků, které obdržíme z výběrového souboru, vypočítáme hodnotu testového kritéria (tedy konkrétní číslo). Mohou nastat dva případy:

a) **Výsledek je v kritickém oboru**

S  $100 \cdot \alpha\%$  pravděpodobností (spolehlivostí) můžeme říct, že **zamítáme hypotézu  $H_0$**  a přijímáme hypotézu  $H_1$ .

Jedná se o žádoucí výsledek testu, neboť v tomto případě můžeme učinit jednoznačný závěr (neplatí  $H_0$  a platí  $H_1$ ) při pevně stanoveném riziku ( $\alpha$ , obvykle 5%) že toto tvrzení není správné.

b) **Výsledek je v oboru přijetí**

S  $100 \cdot \beta\%$  pravděpodobností můžeme říct, že zamítáme  $H_1$  a přijímáme hypotézu  $H_0$ . Protože však obvykle volíme pevnou hladinu spolehlivosti  $\alpha$  a tudíž neznáme  $\beta$  (které je větší než  $\alpha$ ), nemůžeme stanovit výsledek testu ve formě výše uvedené formulace (platí  $H_0$ ). Omezíme se pouze na vyjádření, že se **neprokázala platnost hypotézy  $H_1$**  na hladině významnosti 5%.

### Vztah mezi intervalem spolehlivosti a testováním

Jak jsme již uvedli v úvodu kapitoly, princip testování statistických hypotéz je velmi úzce spojen s metodami konstrukce intervalových odhadů. Lze konstatovat, že každá hypotéza, která leží **vně intervalu spolehlivosti** se považuje za nepravděpodobnou a tudíž ji **zamítáme**. Zamítáme ji s pravděpodobností jejíž výše odpovídá spolehlivosti, s jakou je interval konstruován.

Každá hypotéza, která leží **uvnitř intervalu spolehlivosti** se posuzuje jako možná a tudíž ji **nezamítáme**. Interval spolehlivosti je proto možno považovat za množinu přijatelných hypotéz, v teorii statistického testování označovanou jako obor přijetí. Konstrukce intervalu spolehlivosti je tedy v podstatě jednou z alternativních forem statistického testu.

„95% interval spolehlivosti = test hypotézy na 5% hladině významnosti.“



### Shrnutí kapitoly

Jakýkoli předpoklad o vlastnostech či struktuře statistického souboru je označován pojmem statistická hypotéza. Uvedený předpoklad je obvykle zatížen vysokou mírou nejistoty. Jednou z možností, jak tuto nejistotu snížit je využití metod testování statistických hypotéz. Metody se na základě kvantifikovatelných kritérií snaží rozhodnout o správnosti (nezamítnutí) či nesprávnosti (zamítnutí) hypotézy.

Standardní statistický test má poměrně ustálenou podobu, kterou je možno rozdělit do šesti kroků. Jednotlivé kroky jsou v kapitole popsány. Prvním krokem je formulace hypotézy, kterou se obvykle snažíme formulovat jako negaci toho co chceme ve skutečnosti dokázat. K otestování hypotézy je poté nutno vybrat vhodný statistický test, resp. testové kritérium a stanovit spolehlivost závěru, který vyslovíme – hladinu významnosti. Možné výsledky testu jsou rozděleny do dvou skupin – na výsledky, které hovoří ve prospěch zamítnutí původní hypotézy (jsou označovány jako kritický obor) a na výsledky, na základě kterých nemůžeme původní hypotézu zamítnout (jsou označovány jako obor přijetí). Po výpočtu testového kritéria na základě příslušných dat pak můžeme vyslovit závěr o zamítnutí, či nezamítnutí hypotézy.

### Otázky k zamyšlení

1. O čem hovoří výsledek statistického testu, kdy testové kritérium spadá do kritického oboru?
2. Vysvětlíte pojem statistické významnosti.
3. Určete, kde bude ležet kritický obor a obor přijetí, budeme-li uvažovat hladinu významnosti 1%. Výsledek se pokuste zachytit i graficky.



