


# Základy testování hypotéz

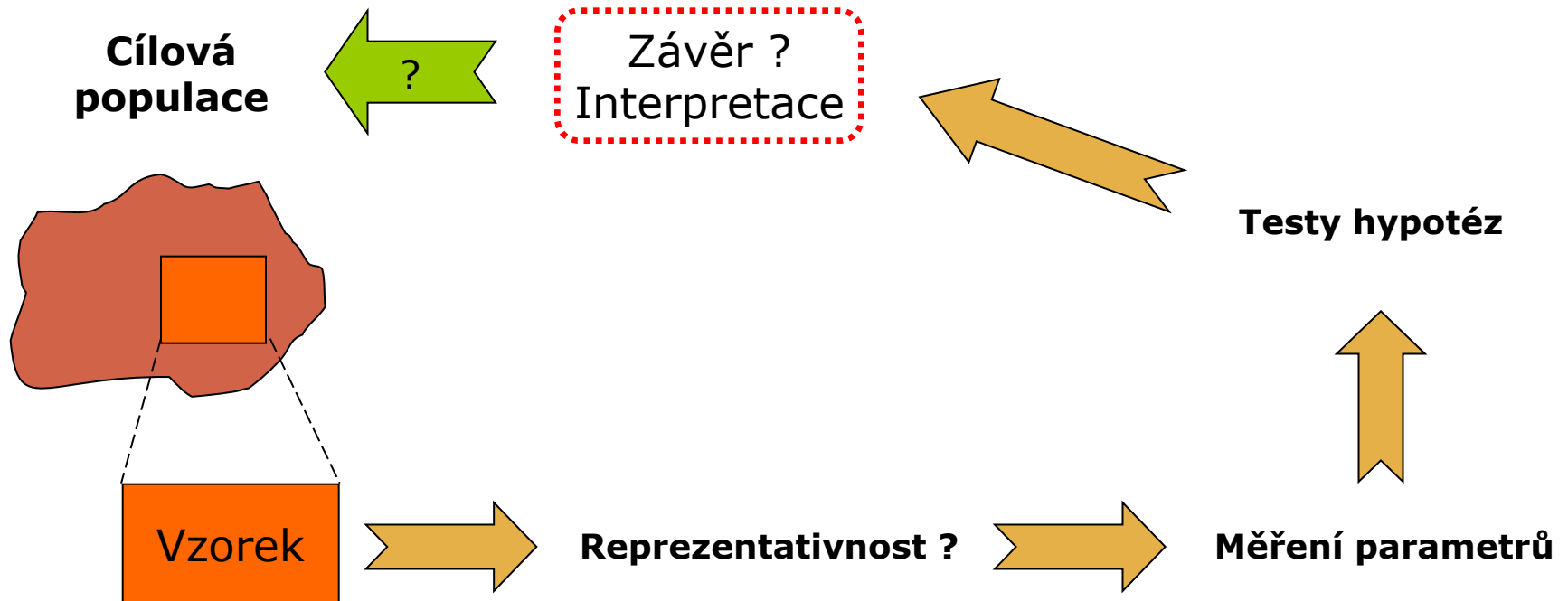


**Princip statistického testování hypotéz**  
**Pojmy statistických testů**  
**Normalita dat a její význam pro testování**

# Princip testování hypotéz



- Formulace hypotézy
- Výběr cílové populace a z ní reprezentativního vzorku
- Měření sledovaných parametrů
- Použití odpovídajícího testu  závěr testu
- Interpretace výsledků



# Statistické testování – základní pojmy



➤ **Nulová hypotéza  $H_0$**

$H_0$ : sledovaný efekt je nulový

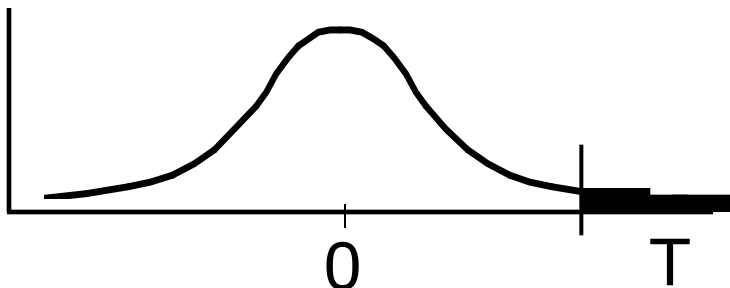
➤ **Alternativní hypotéza  $H_A$**

$H_A$ : sledovaný efekt je různý mezi skupinami

➤ **Testová statistika**

$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota} - \text{Očekávaná hodnota}}{\text{Variabilita dat}} * \sqrt{\text{Velikost vzorku}}$$

➤ **Kritický obor testové statistiky**

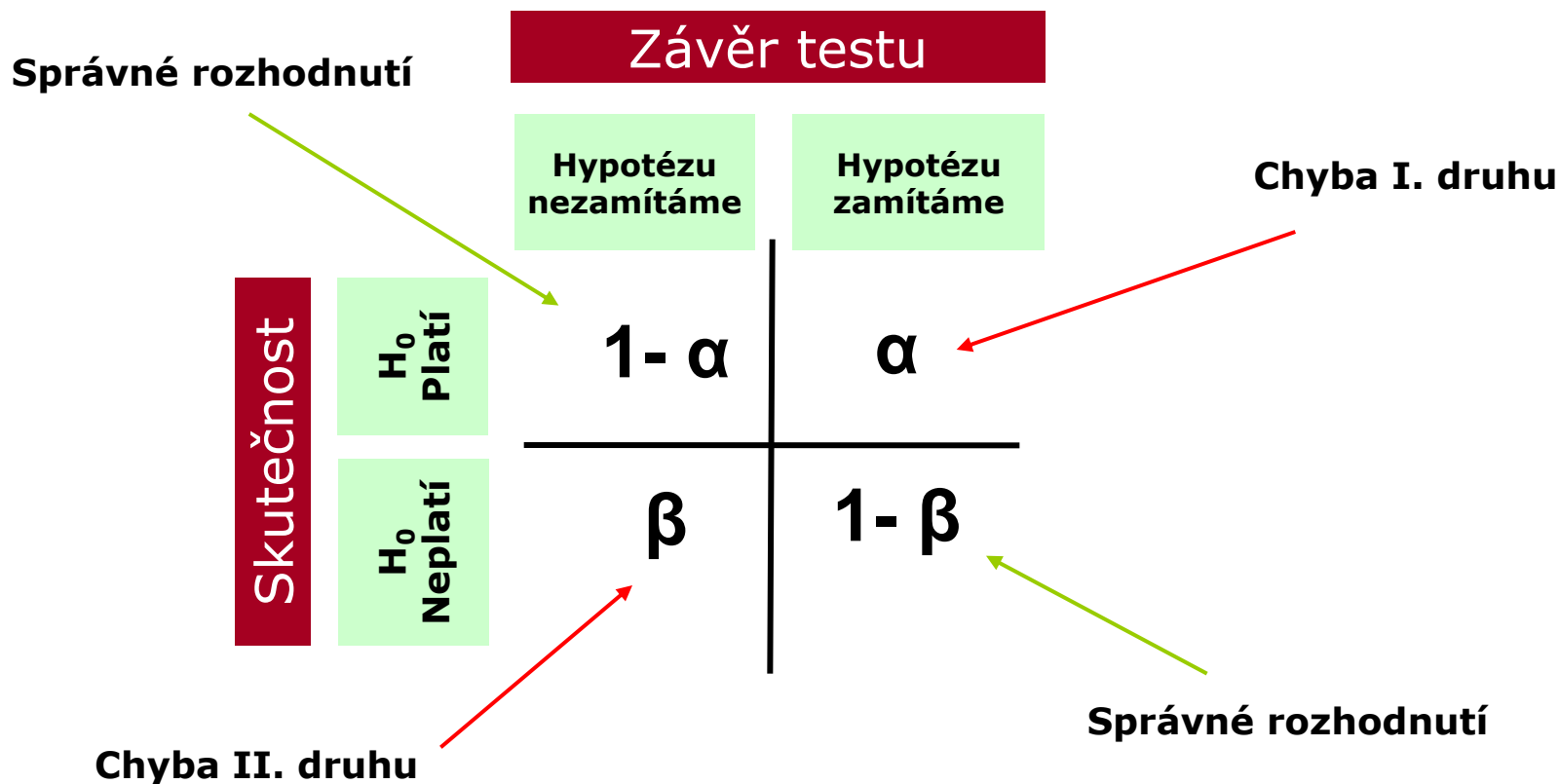


**Statistické testování odpovídá na otázku zda je pozorovaný rozdíl náhodný či nikoliv. K odpovědi na otázku je využít statistický model – testová statistika.**

# Možné chyby při testování hypotéz



- I přes dostatečnou velikost vzorku a kvalitní design experimentu se můžeme při rozhodnutí o zamítnutí/nezamítnutí nulové hypotézy dopustit chyby.

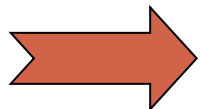


# Význam chyb při testování hypotéz



## Pravděpodobnost chyby 1. druhu

$\alpha$



**Pravděpodobnost nesprávného zamítnutí nulové hypotézy**



## Pravděpodobnost chyby 2. druhu

$\beta$



**Pravděpodobnost nerozpoznání neplatné nulové hypotézy**



## Síla testu

$1-\beta$



**Pravděpodobnostně vyjádřená schopnost rozpoznat neplatnost hypotézy**

# Způsoby testování



Testování  $H_0$  proti  $H_A$  na hladině významnosti  $\alpha$  můžeme provést třemi různými způsoby:

1. Kritický obor (označení  $W$ ) neboli obor zamítnutí  $H_0$
2. Interval spolehlivosti
3. P-hodnota

# P-hodnota



Významnost hypotézy hodnotíme dle získané tzv. **p-hodnoty**, která vyjadřuje pravděpodobnost, s jakou číselné realizace výběru podporují  $H_0$ , je-li pravdivá.

P-hodnotu porovnáme s  $\alpha$  (**hladina významnosti**, stanovujeme ji na 0,05, tzn., že připouštíme 5% chybu testu, tedy, že zamítneme  $H_0$ , ačkoliv ve skutečnosti platí).

P-hodnotu získáme při testování hypotéz ve statistickém softwaru.

P-hodnota vyjadřuje pravděpodobnost za platnosti  $H_0$ , s níž bychom získali stejnou nebo extrémnější hodnotu testové statistiky.

# Parametrické vs. neparametrické testy



## Parametrické testy

- Mají předpoklady o rozložení vstupujících dat (např. normální rozložení)
- Při stejném N a dodržení předpokladů mají vyšší sílu testu než testy neparametrické
- Pokud nejsou dodrženy předpoklady parametrických testů, potom jejich síla testu prudce klesá a výsledek testu může být zcela chybný a nesmyslný

## Neparametrické testy

- Nemají předpoklady o rozložení vstupujících dat, lze je tedy použít i při asymetrickém rozložení, odlehlých hodnotách, či nedetekovatelném rozložení
- Snížená síla těchto testů je způsobena redukcí informační hodnoty původních dat, kdy neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí



# One-sample vs. two sample testy



## Jedno-výběrové testy (one-sample)

- Srovnávají jeden vzorek (one sample, jednovýběrové testy) s referenční hodnotou (popřípadě se statistickým parametrem cílové populace)
- V testu je tedy srovnáváno rozložení hodnot (vzorek) s jediným číslem (referenční hodnota, hodnota cílové populace)
- Otázka položená v testu může být vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek

## Dvou-výběrové testy (two-sample)

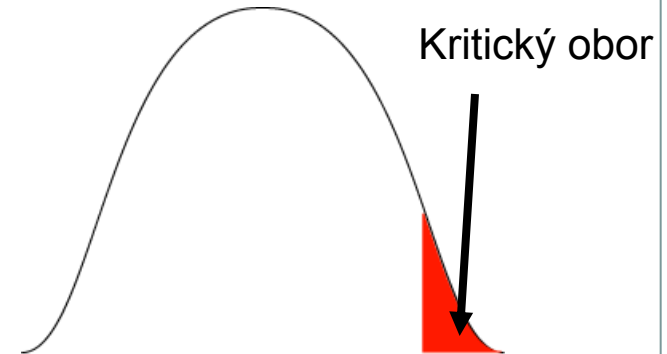
- Srovnávají navzájem dva vzorky (two sample, dvouvýběrové testy)
- V testu jsou srovnávány dvě rozložení hodnot
- Otázka položená v testu může být opět vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek
- Kromě testů pro dvě skupiny hodnot existují samozřejmě i testy pro více skupin dat

# One-tailed vs. Two-tailed tests



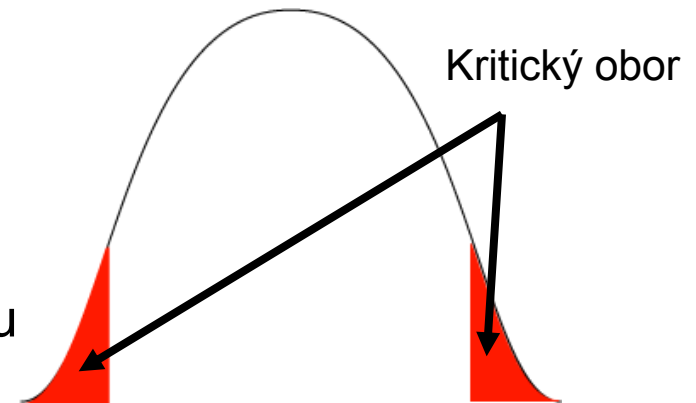
## One – tailed testy

- Hypotéza testu je postavena asymetricky, tedy ptáme se na **větší než/ menší než**
- Test může mít pouze dvojí výstup – jedna z hodnot je větší (menší) než druhá a všechny ostatní případy



## Two – tailed testy

- Hypotéza testu se ptá na otázku **rovná se/nerovná se**
- Test může mít trojí výstup – **menší - rovná se – větší než**
- Situace **nerovná se** je tedy souhrnem dvou možných výstupů testu (**menší+větší**)

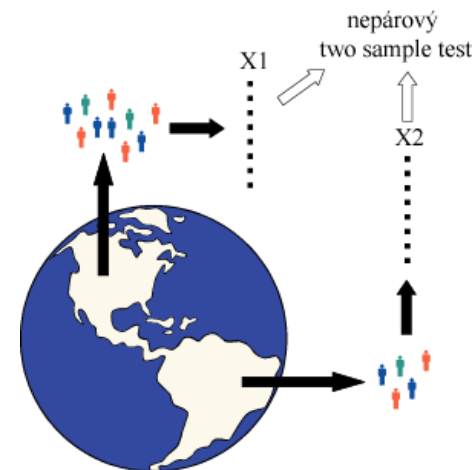


# Nepárový vs. párový design



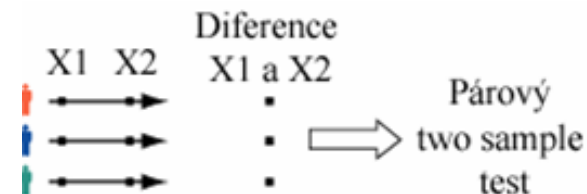
## Nepárový design

- Skupiny srovnávaných dat jsou na sobě zcela nezávislé (též nezávislý, independent design), např. lidé z různých zemí, nezávislé skupiny pacientů s odlišnou léčbou atd.
- Při výpočtu je nezbytné brát v úvahu charakteristiky obou skupin dat



## Párový design

- Mezi objekty v srovnávaných skupinách existuje vazba, daná např. člověkem před a po operaci, reakce stejného kmene krys atd.
- Vazba může být buď přímo dána nebo pouze předpokládána (v tom případě je nutné ji ověřit)
- Test je v podstatě prováděn na diferencích skupin, nikoliv na jejich původních datech



# Statistické testy a normalita dat



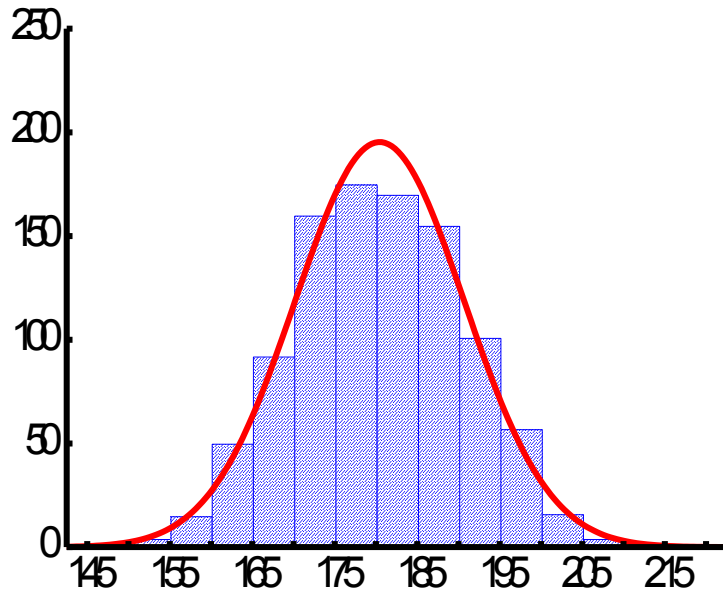
- Normalita dat je jedním z předpokladů tzv. parametrických testů (testů založených na předpokladu nějakého rozložení) – např.  $t$ -testy
- Pokud data nejsou normální, neodpovídají ani modelovému rozložení, které je použito pro výpočet ( $t$ -rozložení) a test tak může lhát
- Řešením je tedy:
  - Transformace dat za účelem dosažení normality jejich rozložení
  - Neparametrické testy – tyto testy nemají žádné předpoklady o rozložení dat

Typ srovnání	Parametrický test	Neparametrický test
2 skupiny dat nepárově:	Nepárový $t$ -test	Mann Whitney test
2 skupiny dat párově:	Párový $t$ -test	Wilcoxon test, znaménkový test
Více skupin nepárově:	ANOVA	Kruskal- Wallis test
Korelace:	Pearsonův koeficient	Spearmanův koeficient

# Testy normality



- Testy normality pracují s nulovou hypotézou, že není rozdíl mezi zpracovávaným rozložením a normálním rozložením. Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.



## • Test dobré shody

V testu dobré shody jsou data rozdělena do kategorií (obdobně jako při tvorbě histogramu), tyto intervaly jsou normalizovány (převedeny na normální rozložení) a podle obecných vzorců normálního rozložení jsou k nim dopočítány očekávané hodnoty v intervalech, pokud by rozložení bylo normální. Pozorované normalizované četnosti jsou poté srovnány s očekávanými četnostmi pomocí  $\chi^2$  testu dobré shody. Test dává dobré výsledky, ale je náročný na  $n$ , tedy množství dat, aby bylo možné vytvořit dostatečný počet tříd hodnot.

## • Kolmogorov Smirnov test

Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložením. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. Měl by být počítán pouze v případě, že známe průměr a směrodatnou odchylku hypotetického rozložení, pokud tyto hodnoty neznáme, měla by být použita jeho modifikace – Lilieforsův test.

## • Shapiro-Wilk's test

Jde o neparametrický test použitelný i při velmi malých  $n$  (10) s dobrou silou testu, zvláště ve srovnání s alternativními typy testů, je zaměřen na testování symetrie.