7. Statistické testování

Princip testování
Chyby
p-hodnota

Vytvořil Institut biostatistiky a analýz, Masarykova univerzita
J. Jarkovský, L. Dušek, J. Kalina

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-Whitneyův test
2 skupiny dat párově:	Párový t-test	Wilcoxonův test znaménkový test
Více skupin nepárově:	ANOVA	Kruskal- Wallisův test
Korelace:	Pearsonův koeficient	Spearmanův koeficient Kendallův koeficient

Shrnutí statistických testů

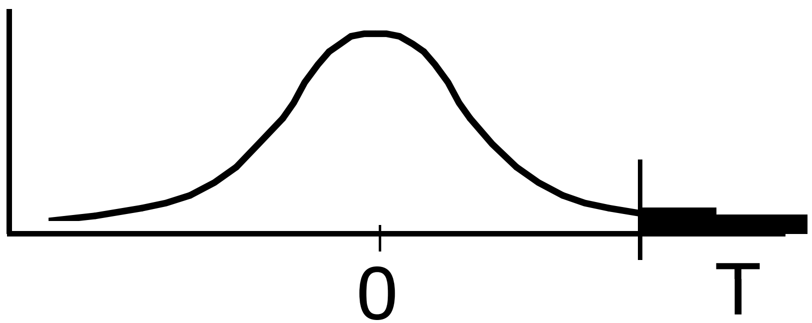
Typ srovnání	Nulová hypotéza	Parametrický test	Neparametrický test
1 skupina dat vs. etalon	**Střední hodnota je rovna hodnotě etalonu.**	jednovýběrový t-test	Wilcoxonův test; znaménkový test
2 skupiny dat nepárově	**Obě skupiny hodnot pochází ze stejného rozdělení.**	nepárový t-test	Mann-Whitneyův test
2 skupiny dat párově	**Zkoumaný efekt mezi páry hodnot je nulový.**	Párový t-test	Wilcoxonův test; znaménkový test
shoda rozdělení	**rozdělení dat ve skupině odpovídá teoretickému (vybranému) rozdělení.**	Shapiro-Wilkův test; Kolmogorovův-Smirnovův test; Lilieforsův test	χ2 test, test dobré shody
homoskedasticita (shoda rozptylů)	**rozptyl obou (všech) skupin je shodný.**	Levenův test	
více skupin nepárově	**Zkoumaný efekt mezi skupinami hodnot je nulový.**	ANOVA	Kruskal- Wallisův test
korelace	**Neexistuje (příčinná, důsledková) vazba mezi skupinami hodnot.**	Pearsonův koeficient	Spearmanův koeficient; Kendallův koeficient

Statistické testování – základní pojmy

- **Nulová hypotéza H_O** — H_O: sledovaný efekt je nulový
- **Alternativní hypotéza H_A** — H_A: sledovaný efekt je různý mezi skupinami
- **Testová statistika**

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

- **Kritický obor testové statistiky**



Statistické testování odpovídá na otázku zda je pozorovaný rozdíl náhodný či nikoliv. K odpovědi na otázku je využit 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.**

Správné rozhodnutí

Chyba I. druhu

		Závěr testu	
		Hypotézu nezamítáme	**Hypotézu zamítáme**
Skutečnost	**H_0 Platí**	**$1-\alpha$**	**α**
	H_0 Neplatí	**β**	**$1-\beta$**

Chyba II. druhu

Správné rozhodnutí

Význam chyb při testování hypotéz

Pravděpodobnost chyby 1. druhu

Před výpočtem testu si stanovujeme maximální přípustnou pravděpodobnost. Obvykle 5 %.

α → **Pravděpodobnost nesprávného zamítnutí nulové hypotézy**

Pravděpodobnost chyby 2. druhu

Nemůžeme ovlivnit jinak než výběrem testu.

β → **Pravděpodobnost nerozpoznání neplatné nulové hypotézy**

Síla testu

Síla testu je vlastností testu – parametrické testy mají vyšší sílu než neparametrické.

$1-\beta$ → **Pravděpodobnostně vyjádřená schopnost rozpoznat neplatnost hypotézy**

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 α (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.

- Je-li p-hodnota $\leq \alpha$, pak H_0 zamítáme na hladině významnosti α a přijímáme H_A.
- Je-li p-hodnota $> \alpha$, pak H_0 nezamítáme na hladině významnosti α.

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

Jednovýběrové testy (one-sample)

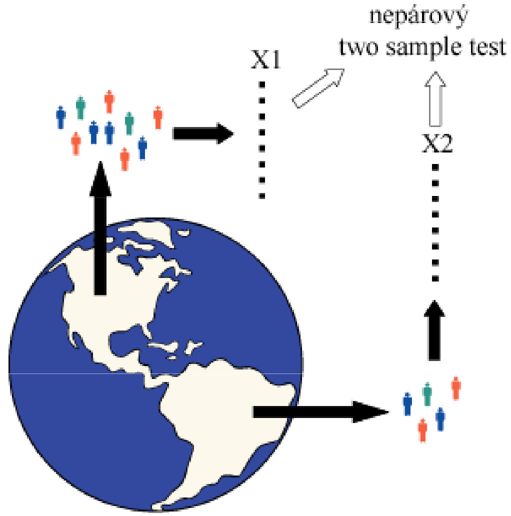
- Srovnávají jeden vzorek (one sample, jednovýběrové testy) s referenční hodnotou (popřípadě se statistickým parametrem cílové populace).
- V testu je tedy srovnáváno rozložení hodnot (vzorek) s jediným číslem (referenční hodnota, hodnota cílové populace).
- Otázka položená v testu může být vztažena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

Dvouvýběrové 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.

Nepárový vs. párový design

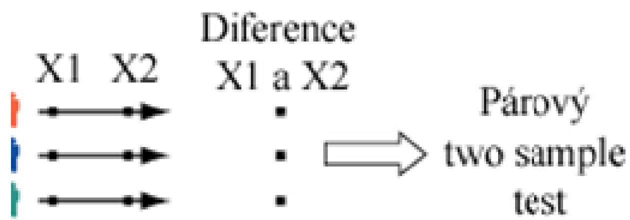
Nepárový design



- 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



Normalita dat

Normální rozdělení pravděpodobnosti je definováno rovnicí:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Kde *f(x)* značí hustotu pravděpodobnosti, *μ* značí střední hodnotu (aritmetický průměr), *σ* značí směrodatnou odchylku a *x* hodnotu zkoumané veličiny.

Dosazením *s* za *σ* a *$\bar{x}$* za *μ* získáme křivku idealizovaného rozdělení pro daný výběr.