Statistická síla

- princip testování hypotéz (opakování)
- chyby I. a II. druhu
- statistická síla
- požadovaná velikost výběru

Statistická síla

- pravděpodobnost, že zamítneme nulovou hypotézu, která neplatí
- tj. že najdeme (statisticky významný) rozdíl, když tento rozdíl existuje

Statistická síla

- **příklad**: srovnáváme účinnost léčby úzkostných poruch
- dva typy léčby – farmakoterapie (A) a psychoterapie (B)

Testování hypotéz

- náhodně vybereme z populace pacientů s úzkostnou poruchou vzorek pacientů
- náhodně zvolená polovina z nich se podrobí farmakoterapii, druhá polovina psychoterapii
- po léčbě změříme u obou skupin standardizovaným nástrojem míru úzkosti

Testování hypotéz

- jaká bude nulová hypotéza v této studii?
- **nulová hypotéza**: průměrná míra úzkosti u pacientů s terapií A je stejná jako průměrná míra úzkosti u pacientů s terapií B
- $\mu_A = \mu_B$

Testování hypotéz

- pro porovnání průměrů vzorku A a B můžeme použít t-test (pro nezávislé výběry)
- $t = (\bar{x}_A - \bar{x}_B) / s$
- hodnotu t vyhledáme v tabulkách t-rozdělení (pro příslušný počet stupňů volnosti)

Testování hypotéz

- pokud se **t** blíží nule (tj. mezi průměry vzorků A a B není velký rozdíl), pak <u>nezamítneme nulovou hypotézu</u> – vyvodíme, že ani mezi průměry populace A a B není rozdíl
- pokud je **t** od nuly vzdáleno, pak <u>nulovou hypotézu zamítneme</u> a vyvodíme, že populační průměry se liší

Testování hypotéz

- jaké mohou být výsledky testování hypotéz?

Testování hypotézy

skutečnost → *rozhodnutí* ↓	nulová hypotéza **platí**	nulová hypotéza **neplatí**
zamítneme nulovou hypotézu	**chyba I. druhu**	správné rozhodnutí
nezamítneme nulovou hypotézu	správné rozhodnutí	**chyba II. druhu**

Testování hypotéz

- předpokládejme, že nulová hypotéza platí (tj. účinnost farmakoterapie a psychoterapie je stejná)
- 2 možnosti:
  - průměry vzorku A a B jsou velice podobné – t je blízké nule a tak **správně** nezamítneme nulovou hypotézu
  - nebo se průměry vzorku A a B liší v takové míře, že se dopustíme chyby I. druhu

Chyba I. druhu

- je možné (i když málo pravděpodobné), že vzorky z populací o stejném průměru mohou mít velice rozdílné průměry
- v tomto případě bychom nulovou hypotézu zamítli nesprávně a vyvodili, že průměry populací A a B jsou odlišné

Chyba I. druhu

- pravděpodobnost takové chyby se označuje hladina významnosti (α)
- její úroveň stanovuje výzkumník (velice často na 5%, příp. 1%)
- jde vlastně o <u>pravděpodobnost</u>, že získáme tuto hodnotu **t** (=rozdíl mezi průměry vzorků), <u>pokud by nulová hypotéza platila</u>

Testování hypotéz

- předpokládejme, že nulová hypotéza **neplatí**, terapie A není stejně účinná jako terapie B (tj. je rozdíl v míře úzkosti u pacientů z populace A a B)
- opět dvě možnosti
  - najdeme rozdíly mezi průměry vzorků – **t** je dostatečně velké a nulovou hypotézu tak správně zamítneme
  - mezi průměry vzorků není dostatečně velký rozdíl a dopustíme se chyby II. druhu

Testování hypotézy

skutečnost → *rozhodnutí* ↓	nulová hypotéza **platí**	nulová hypotéza **neplatí**
zamítneme nulovou hypotézu	**chyba I. druhu**	správné rozhodnutí
nezamítneme nulovou hypotézu	správné rozhodnutí	**chyba II. druhu**

Chyba II. druhu

- průměry populace se liší, ale přesto se může stát, že průměry vzorků budou velice podobné
- v tom případě **nesprávně nezamítneme** nulovou hypotézu a vyvodíme, že terapie jsou podobně účinné
- pravděpodobnost této chyby se označuje β

Testování hypotézy

skutečnost → *rozhodnutí* ↓	nulová hypotéza **platí**	nulová hypotéza **neplatí**
zamítneme nulovou hypotézu	**chyba I. druhu** (α)	správné rozhodnutí ($1-\beta$)
nezamítneme nulovou hypotézu	správné rozhodnutí ($1-\alpha$)	**chyba II. druhu** (β)

Statistická síla

- pravděpodobnost, že správně zamítneme nulovou hypotézu, která neplatí,
  je rovna 1 - β
- jde o tzv. **sílu testu** (power) – **schopnost zachytit rozdíl, který existuje**
- cílem je dosáhnout síly >0.8 nebo 0.9

Statistická síla

- 4 faktory jsou při testování hypotéz vzájemně provázány:
  - hladina významnosti
  - síla testu
  - velikost účinku
  - rozsah výběrového souboru
- pokud známe 3 z nich, dá se vypočítat zbylý parametr

Hladina významnosti

- čím přísněji ji stanovíme (např. 0,1%), tím nižší síla testu

Velikost vzorku

- s větším vzorkem máme větší pravděpodobnost, že existující rozdíl zachytíme

Velikost účinku

- čím je rozdíl mezi populačními průměry větší, tím větší pravděpodobnost, že najdeme i rozdíl mezi průměry vzorků
- proto nejmenší rozdíl, po kterém má smysl pátrat, je ten, který je ještě klinicky významný
- vychází i z podstaty problému - pokud porovnáváme např. lék s placebem, můžeme očekávat větší rozdíl účinku než při porovnání dvou léků

Požadovaná velikost výběru

- jedním z účelů analýzy statistické síly je určení, **jak velký musí být náš vzorek**, abychom měli dostatečnou pravděpodobnost, že zachytíme předpokládaný rozdíl
- je ovšem možné i zpětně posoudit sílu našeho testování poté, co byl výzkum proveden (příp. při metaanalýzách)

Požadovaná velikost výběru

- nejprve se musíme rozhodnout, jaký **nejmenší rozdíl** je ještě **klinicky významný**
- často se používá <u>Cohenův koeficient účinku</u> **d**
- označuje se jako tzv. **effect size** – velikost účinku
- jde např. o **standardizovaný rozdíl průměrů** (vzhledem ke směrodatné odchylce) nebo korelaci mezi nezávislou a závislou proměnnou (pak se označuje r)

Požadovaná velikost výběru

- podle Cohena je
  - d < 0.20 malý účinek (r=0.10)
  - d = 0.50 střední (r=0.243)
  - d > 0.80 velký (r=0.371)
  - závisí ale i na kontextu

Požadovaná velikost výběru

- dále musíme odhadnout variabilitu znaku v populaci (σ) – z předchozích výzkumů, pilotní studie atd.
- pak stanovit hladinu významnosti (obvykle 5%)
- a nakonec sílu testu – tj. jakou chceme mít pravděpodobnost, že pokud rozdíl existuje, že ho prokážeme? (ideálně min. 80%)

Požadovaná velikost výběru

- pro různé statistické testy se požadovaná velikost vzorku počítá různě
- existují speciální počítačové programy, statistické software mají obvykle v pokročilejších modulech tyto výpočty zabudovány
- je možné provést i ruční výpočet (s pomocí tabulky pro hodnoty δ)

Požadovaná velikost výběru

- **příklad:** pro studii srovnávání účinnosti terapií úzkostných poruch chceme vypočítat velikost výběru
- velikost účinku: jednu metodu terapie bychom upřednostnili před druhou, pokud by rozdíl v testu úzkosti byl nejméně 5 bodů
- směrodatná odchylka pro test úzkosti je 10 bodů

Požadovaná velikost výběru

- velikost účinku je pro naši studii
  d = 5/10 = 0.5
- hladina významnosti $\alpha = 0.05$
- chceme dosáhnout síly testu 0.80

Požadovaná velikost výběru

- vzorec pro test porovnávající dva průměry ze stejně velkých výběrů:

- $N = 2(\delta/d)^2$

Požadovaná velikost výběru

- $N = 2(\delta/d)^2$
- δ najdeme v tabulce (hledáme δ pro sílu testu 0.80 a $\alpha = 0.05$)
  - **δ = 2.80**
- $N = 2(2.8/0.5)^2 = 2(5,6)^2$
- $N =$ **62.72**

Požadovaná velikost výběru

- požadovaná velikost výběru je asi 63 v každé skupině, tj. celkem **126 osob**

Síla již provedeného testu

- obdobně můžeme spočítat sílu již provedeného testování – kdy víme, jaká byla velikost výběru
- kdyby byl v našem příkladu počet osob v jedné skupině 25, jaká by byla síla testu?

Síla již provedeného testu

- $N = 2(\delta/d)^2$
- $\delta = d\sqrt{N/2}$
- $\delta = 0{,}5\sqrt{25/2}$
- $\delta = 0{,}5\ (3{,}54) = \mathbf{1{,}77}$
- pro $\delta = 1.77$ a $\alpha = 0{,}05$ je síla testu asi **0,43**

Síla již provedeného testu

- při N=50 (v každé skupině 25) bychom měli pouze 43% pravděpodobnost, že najdeme rozdíl, i kdyby skutečně existoval

Požadovaná velikost výběru pro sílu testu >0.80 a 5% hladině významnosti

velikost účinku	*d*	*jedno-výběrový t-test*	*dvouvýběrový (nezávislý) t-test*
malý	,20	196	784
střední	,50	32	126
velký	,80	13	49

Výpočet síly testu ve Statistice

- program Statistica tyto výpočty provádí automaticky – stačí zadat např. hodnoty průměrů a směrodatné odchylky, hladinu významnosti, požadovanou sílu nebo skutečnou velikost výběru
- procedury jsou rozděleny podle typu testu (porovnání průměrů, korelace atd.)

Analýza síly testu ve výzkumné zprávě

- podle Cohenovy analýzy empirických studií z oblasti psychologie (z roku 1972) – průměrná síla testu jen 0,48
- jen malý počet studií obsahuje údaje o síle testu – postupně je však mezinárodní časopisy vyžadují
- Cohen zdůrazňuje význam určení alternativní hypotézy

Literatura

- Hendl: kapitola 11
- Hendl – str. 407, tabulka 11.2