

# DESIGN EXPERIMENTU A TESTOVÁNÍ HYPOTÉZ

Martin Dostál

Honeywell International - Aerospace Advanced Technology Europe  
Masarykova Univerzita v Brně, Fakulta informatiky



# Experimentální výzkum

- popsat, pochopit, vysvětlit a předpovědět vztahy mezi proměnnými
- experiment
- proměnné
- metody - deskriptivní, inferenční

# Proměnné

- nezávislé (independent)
  - jsou nastavovány experimentátorem nebo se mění vnějším vlivem (obtížné pro experiment) - určujeme
  - často mají stanovené hladiny
  - jeden experiment může mít (často má) i více nezávislých proměnných, analýza je však složitější
  - sledujeme vliv nezávislých proměnných na závislé proměnné
- závislé (dependent)
  - hodnoty závislých proměnných se mohou měnit v závislosti na nezávislých proměnných - měříme
- sledované (controlled)
  - během experimentu zůstávají konstantní, ale je potřeba je sledovat (definovat)
- vnější (extraneous)
  - proměnné, které nejsou přímo nezávislými, ale stojí je za to sledovat, protože by mohly mít vliv na výsledek

**SLEDOVANÁ PROMĚNNÁ JE OXYMORON :)**

# Úlohy a scénáře

- Validační experimenty se mohou skládat z několika dílčích úhol (experimentů)
- Zpravidla nad stejnou skupinou subjektů
- Úlohy (tasks/exercises)
  - experiment s danou úlohou - experiment je potřeba promyslet a popsat (proměnné, hypotézy, metriky,...)
  - úlohy se skládají ze scénářů (scenarios)
  - příprava scénářů (review, dry-run)
  - důsledná příprava a popis scénářů!
  - používejte záznam, aby jste byli schopni cokoliv rekonstruovat
- Baseline a test-condition(s)
- Control group/Experiment group

# Design experimentu

- nezávislá proměnná
- two-group design
  - dvě hladiny nezávislé proměnné
  - typicky s/bez (viz control group),
  - nebo hladina X nebo hladina Y
- multiple-group design
  - více než dvě hladiny nezávislé proměnné
- faktorový design

# Design experimentu

- faktorový design
- faktory jsou nezávislé proměnné
- můžeme jich mít více v jednom experimentu
- faktoriálním designem rozumíme, že zkoumáme všechny kombinace hladin všech faktorů
- můžeme zkoumat nejen vliv jednotlivých proměnných, ale také jejich vzájemnou interakci
- například 3x2x4 design obsahuje tři nezávislé proměnné, první má tři, druhá dvě a třetí čtyři hladiny

# Příklad

- nezávislé proměnné:
  - vstupní zařízení: myš/trackball, dvě hladiny
  - velikost tlačítka: malá/velká, dvě hladiny
  - pozice tlačítka: levý horní roh, pravý horní roh, levý dolní roh, pravý dolní roh, celkem čtyři hladiny
- závislé proměnné
  - čas pro zasažení cíle
- sledované proměnné
  - světelné podmínky
  - počáteční vzdálenost kurzoru od cíle (toto je dobrý příklad na zvažení zda toto nemá být sledovaná/kontrolovaná proměnná)
- faktoriální design: budeme zkoumat všechny kombinace
- experiment má design 2x2x4

# Skupiny subjektů

- Design mezi skupinami (Between-subject design)
- Uvnitř skupin (Within-subject design)
  - pozor na pořadí (order effect)
- Smíšený design (Mixed-design)



# (před)zpracování dat

- import
- data nejprve „prohlédneme” - kontrola je důležitá
- upravíme/opravíme strukturu (škály, formáty,...) - překlepy, nepřesnosti
- zjistíme odlehlá pozorování

# Testování hypotéz



**POZOR, NAŠE ZPRACOVÁNÍ TÉTO PROBLEMATIKY JE S  
OHLEDEM NA PROSTOR ZNAČNĚ POVRCHNÍ. STATISTIKU JE  
TŘEBA STUDOVAT VE VĚTŠÍM DETAILU.**

**TRÉBNÍ SYNDROMU NE NEJŠIRÍM DETAILU.**

# Testování hypotéz

- úvod
- chyba prvního a druhého druhu
- inferenční statistika
- rozdělení pravděpodobnosti
- statistická významnost, hladina významnosti
- parametrické a neparametrické metody

# Hypotéza

- statistická hypotéza je tvrzení o parametrech (parametrické testy) pozorované náhodné veličiny pocházející se základního souboru nebo o tvaru rozdělení znaku v základním souboru (neparametrické testy) na základě pozorované náhodné veličiny

# Hypotézy

- cílem testování hypotéz je rozhodnout o „platnosti“ (neformálně použito) určité hypotézy
- výsledkem je **zamítnutí** (reject) nebo **nezamítnutí** (fail-to-reject) hypotézy
- pohybujeme se v **nejistotě**, avšak na určité hladině významnosti  $\alpha$
- nulová hypotéza:  $H_0$  (null hypothesis)
  - je obvykle opakem toho co je očekávným závěrem
- alternativní hypotéza:  $H_A$  (alternative hypothesis)
  - alternativní hypotéza může být
    - oboustranná (simple hypothesis)
    - levostranná (left-sided, left-tailed)
    - pravostranná (right-sided, right-tailed)

# Nulová hypotéza

- typicky formulovaná jako „není rozdíl mezi ...”
- interpretace: rozdíly mezi skupinami nejsou větší než odpovídá náhodné variabilitě
- **nezamítnutí** nulové hypotézy se nerovná jejímu **potvrzení**

# Postup

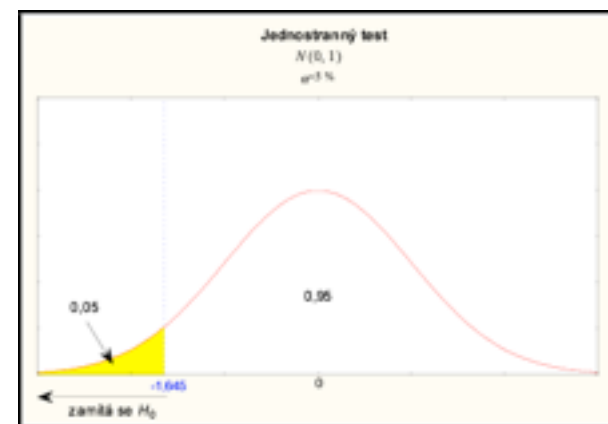
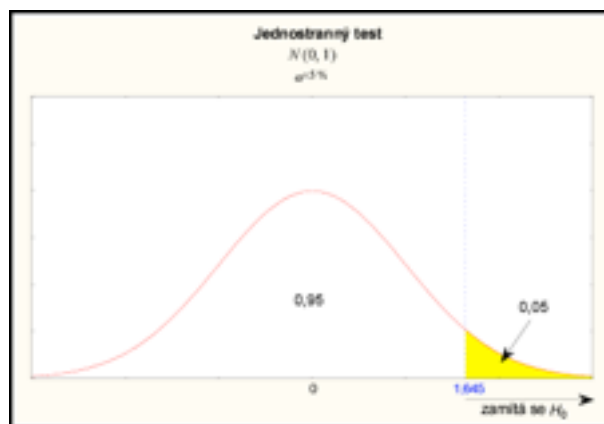
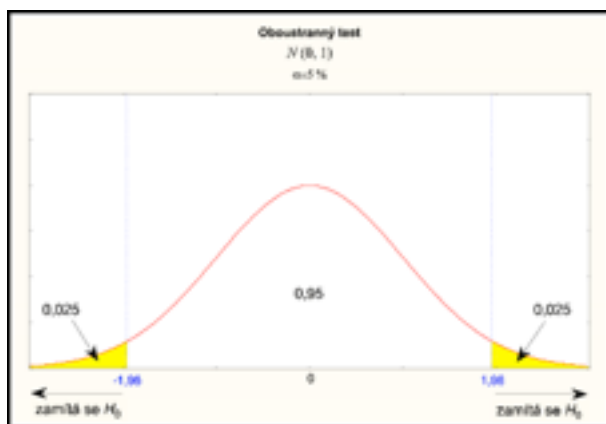
- stanovíme **nulovou hypotézu**, předpokládáme že platí
- stanovíme **alternativní hypotézu** (oboustranná či jiná?)
- stanovíme **hladinu významnosti  $\alpha$**
- ověříme **předpoklady** metody pro testování hypotézy
- získáme **testovou statistiku**
- **rozhodneme** o (ne)zamítnutí nulové hypotézy



# p-hodnota

- p-hodnotu potřebujeme pro rozhodnutí o hypotéze (Existuje také tradiční postup založený na tzv. kritickém oboru, postup přes p-hodnotu je dnes častější díky používání software. Jde jen o jiný, úzce související postup, který vede ke stejnému výsledku)
- neformálně, p-hodnota je pravděpodobnost, s jakou testovací statistika nabývá hodnot svědčících proti testované hypotéze
- p-hodnota (p-value) je výstupem (krom jiného) mnoha statistických inferenčních metod
- p-hodnotu interpretujeme jako mezní hladinu významnosti
- pokud  $p < \alpha$ , pak zamítáme nulovou hypotézu, v opačném případě nulovou hypotézu nezamítáme
- p-hodnota není totéž co  $\alpha$ 
  - $\alpha$  je určená
  - p-hodnota je vypočtená

# Testování hypotézy v mírném detailu



- sledujeme tzv. Z-skore (nebo p-hodnotu získanou ze Z-skore) zda spadne do kritického oboru pro zamítnutí nulové hypotézy
- Příklad: průměrný čas pro kliknutí na tlačítko v celé skupině je 190ms, u populace 100 dětí 198ms, rozptyl je 1.5. Je čas u dětí statisticky významně vyšší? Nulová hypotéza: „není rozdíl že, ...”
- alt. hypotéza bude pravostranná, dále položíme  $\alpha = 0.05$
- $Z = (198-190)/1.5 = 5.33$ ,  $p < 0.001$
- závěr: zamítáme nulovou hypotézu

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

# Chyba I. a II. druhu

- $H_0$  - nulová hypotéza
- $H_A$  - alternativní hypotéza
- $\alpha$  - pravděpodobnost chyby I. druhu
- $\beta$  - pravděpodobnost chyby II. druhu

<b>Rozhodnutí</b>	<b>Skutečnost</b>	
	$H_0$ platí	$H_0$ neplatí, platí $H_A$
Nemůžeme zamítnout $H_0$ (nevýznamný výsledek)	Správné rozhodnutí	Chyba II. druhu (s pravděpodobností $\beta$ )
Zamítneme $H_0$ (významný výsledek)	Chyba I. druhu (s pravděpodobností $\alpha$ )	Správné rozhodnutí

# Chyba I. a II. druhu

- chyba prvního druhu: falešná pozitivita (false positive, producer error)
- $\alpha$  se nejčastěji volí 0.05, 0.01, 0.001
- chyba druhého druhu: falešná negativita (false negative, consumer error)
- $1 - \alpha$  označujeme jako spolehlivost testu
- $1 - \beta$  označujeme jako sílu testu
- síla testu by měla být větší než 0.8
- $\beta$  závisí na velikosti výběru (větší výběr, menší  $\beta$ ), velikostí  $\alpha$  ( $\alpha$  roste,  $\beta$  klesá a naopak) a na odchylce skutečné hodnoty parametru od předpokládané hodnoty (čím větší odchylka, tím menší  $\beta$ )

# Hypotézy a metody

- podle rozdělení pravděpodobnosti
  - parametrické (opírají se předpoklad rozdělení parametru rozdělení, typicky normálního)
  - neparametrické - nemají požadavky na rozdělení, širší aplikovatelnost znamená ale vyšší riziko chyby druhého druhu - parametrické metody jsou slabší
- podle skupin
  - nepárové
  - párové

**PRO POUŽITÍ PARAMETRICKÝCH METOD, JE TAKÉ  
MOŽNOST PROVÉST TRANSFORMACI DAT**

# Přehled základních metod

	Typ závislé proměnné		
	Intervalová/podílová data s normálním rozdělením	Intervalová/podílová data s jiným než normálním rozdělením, ordinální data	Dichotomická
<b>Porovnání dvou nepárových skupin</b>	Nepárový t-test	Mann-Whitneyho test	Fisherův test
<b>Porovnání dvou párových skupin</b>	Párový t-test	Wilcoxonův test	McNemarův test
<b>Porovnání více než dvou nepárových skupin</b>	Analýza rozptylu (ANOVA)	Kruskal-Wallis test	Chi-kvadrát test
<b>Porovnání více než dvou párových skupin</b>	Analýza rozptylu s opakovanými pozorováními (Repeated-measures ANOVA, RM-ANOVA)	Friedmanův test	Cochranův Q-test
<b>Zjištění korelace mezi dvěma proměnnými</b>	Pearsonův korelační koeficient	Spearmanův korelační koeficient	Cramerovo V