



**LIMITY APLIKACE
STATISTICKÝCH TESTŮ VÝZNAMNOSTI
V PEDAGOGICKÉM VÝZKUMU:
SEMINÁŘ PRO NESTATISTIKY**

Seminář ČAPV 2018

Upozornění

- Prezentace obsahuje pouze doprovodný text k semináři pro doktorandy v rámci konference ČAPV 2018.
- Hlavní náplní semináře byly ukázky ve statistickém softwaru a jejich diskuse, přičemž tyto aspekty prezentace neobsahuje.

Cíl semináře

- V rámci semináře se pokusíme ukázat na jednoduchých příkladech z pedagogického výzkumu limity aplikace statistických testů významnosti.
- Seminář volně navazuje na článek Soukup, P., & Kočvarová, I. (2016). Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu. *Pedagogická orientace*, 26(3), 512–536.

Obsah semináře

- Zaměříme se na následující témata:
 1. proč užíváme (až nadužíváme) statistické testy významnosti;
 2. proč se mají testy aplikovat na reprezentativní výběry;
 3. proč záleží na velikosti výběru při aplikaci testů;
 4. proč vede užívání testů k černobílé interpretaci výsledků výzkumu;
 5. proč „statistická významnost“ není totéž co „významnost“.

Limity semináře

- V rámci semináře budeme pracovat bez
 - dostatku času;
 - podrobností;
 - vzorečků;
 - hromadného klikání ve statistickém softwaru;
 - statistika / statističky.

Proč užíváme (až nadužíváme) statistické testy významnosti

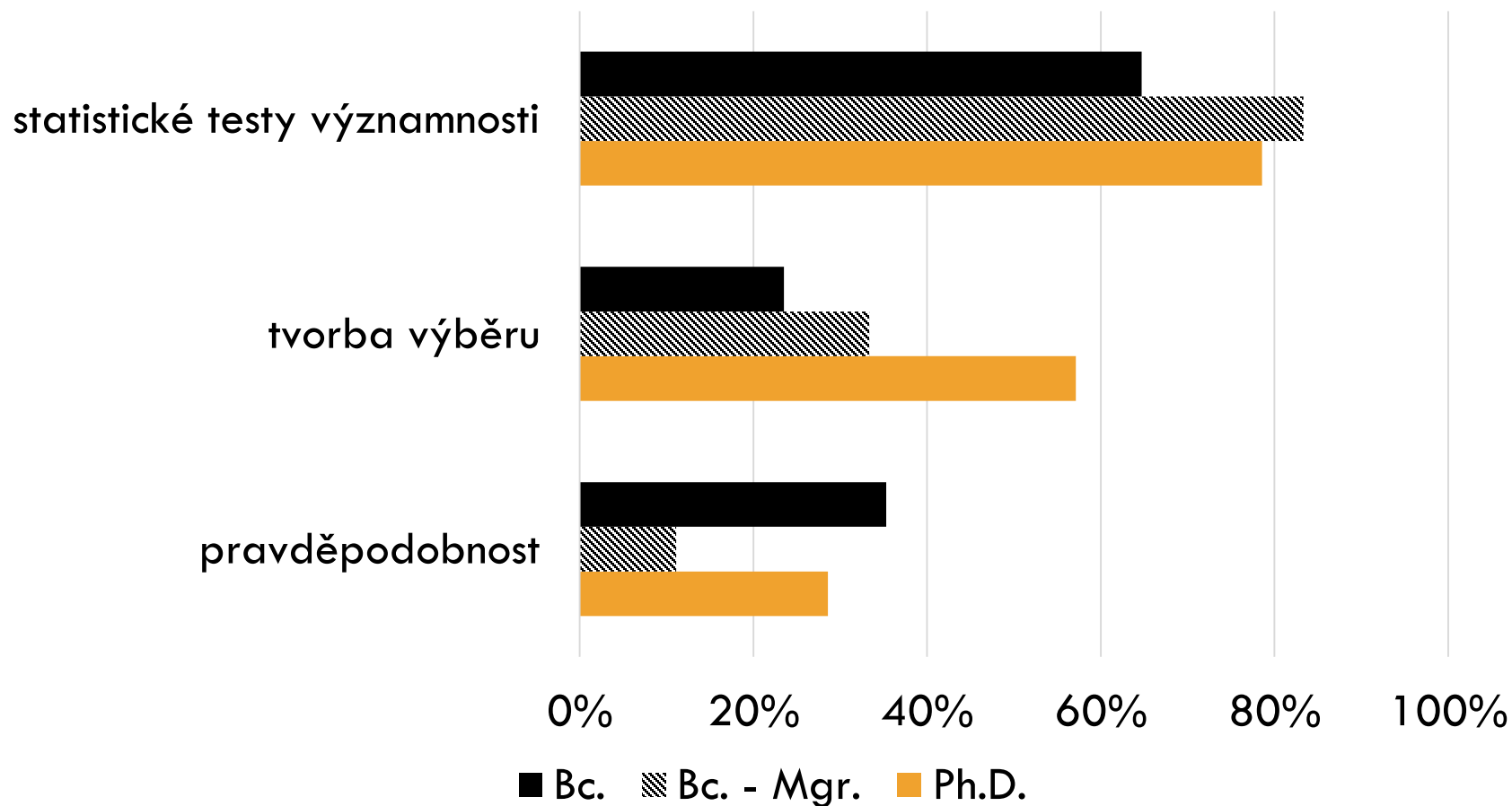
1. 3 důvody nesprávného užívání a nadužívání
2. Základní podmínky vhodné aplikace

Důvody nesprávného užívání a nadužívání:

1. neznalost

- obecně slabá statistická vzdělanost v oblasti pedagogiky / sociálních věd
- problematika je zjednodušována v rámci metodologického kurikula

Výuka kvantitativní analýzy v pedagogice (2017)



Důvody nesprávného užívání a nadužívání:

2. tradice

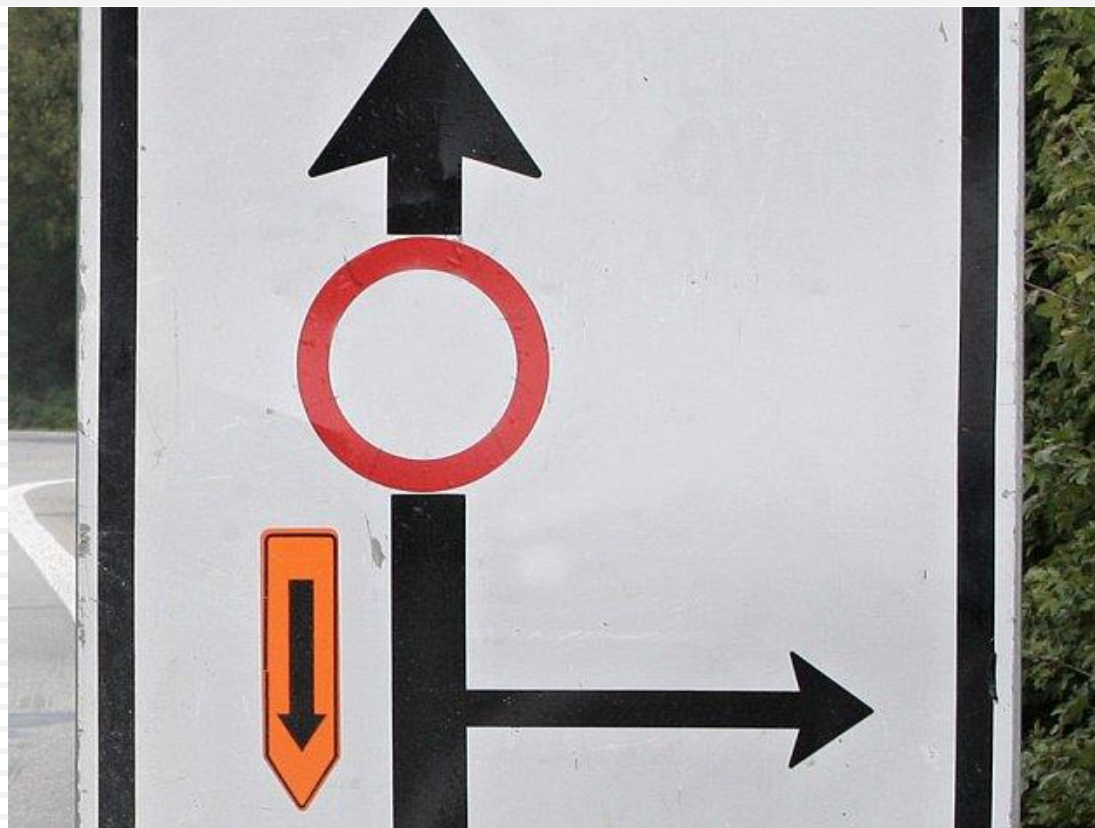
- Testování je obecně přijímáno a často vyžadováno recenzenty, školiteli.
- Soukup (2016) analyzoval 84 článků v časopisech Pedagogika a Pedagogická orientace za 10 let.
Závěr: „Více než polovina těchto textů používá koncepci statistické významnosti, i když to prezentovaná data neumožňují.“ (s. 182)

Důvody nesprávného užívání a nadužívání:

3. mechanická aplikace

- Aplikace testů je ve statistickém softwaru otázkou několika kliknutí, hodnoty statistické významnosti lze do výstupů implementovat automaticky.
- Jakékoli šetření využívající kvantitativní analýzu dat mohou „zdobit“ p-hodnoty či hvězdičky.

Jenže...



Podmínky vhodné aplikace

- Test statistické významnosti testuje statistickou významnost výsledku, tedy jeho **zobecnitelnost** z výběru na populaci na zvolené hladině významnosti.
- O aplikaci testu má smysl uvažovat, pokud máme
 - **reprezentativní** výběr z populace,
 - který **ZÁROVEŇ** splňuje nároky na potřebnou **velikost** s ohledem na přijatelnou **výběrovou chybu** a na požadovanou **sílu testu**,
 - a pokud splňujeme ještě další podmínky.

Proč se mají testy aplikovat na reprezentativní výběry

1. Co je reprezentativita
2. Ukázka tvorby reprezentativního výběru

Reprezentativita výběru

- Reprezentativita „znamená takovou kvalitu výběru a dalších postupů metodiky výběrového šetření, díky níž dotázaný soubor jednotek získá rozložení charakteristik, jež můžeme považovat za shodné s populací, ze které vybraný soubor pochází“ (Krejčí, 2011, s. 3).
- Reprezentativita závisí především na
 - ▣ **způsobu tvorby výběru;**
 - ▣ **velikosti výběru.**

Reprezentativní výběr



Ukázka 1: výběr z populace

- Opora výběru: data z šetření AES 2016
- N = 12 272

Vybrané charakteristiky	Zastoupení
Muži	48 %
Pražáci	6 %
Vysokoškoláci (ISCED 7 - Mgr.)	12 %

Proč záleží na velikosti výběru při aplikaci testů

1. Velikost pravděpodobnostního výběru souvisí s **výběrovou chybou**
2. Velikost pravděpodobnostního výběru souvisí se **sílou testu**

Výběrová chyba

- Výběrová chyba vyjadřuje odchylku mezi populací a výběrem.
- Při tvorbě výběru se musíme rozhodnout, jakou odchylku jsme ochotni připustit, obvykle 5 %.

Vybraný znak populace	Zastoupení v populaci	Zastoupení ve výběru při 5% toleranci výběrové chyby
Muži	49 %	49 % ($\pm 2,5$ %)

Velikost prostého náhodného výběru podle tolerance výběrové chyby

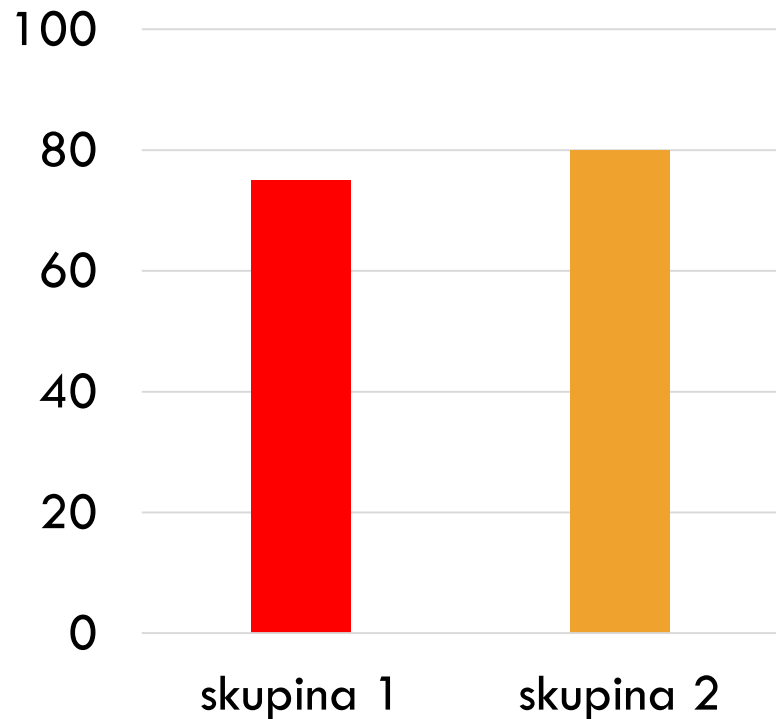
Populace	Tolerance výběrové chyby			
	5% (± 2,5 %)	3% (± 1,5 %)	2% (± 1 %)	1% (± 0,5 %)
2 000	322	696	1 091	1 655
5 000	357	879	1 622	3 288
10 000	370	964	1 936	4 899
100 000	383	1 056	2 345	8 762
1 000 000	384	1 066	2 395	9 513
10 000 000	384	1 067	2 400	9 595

Síla testu

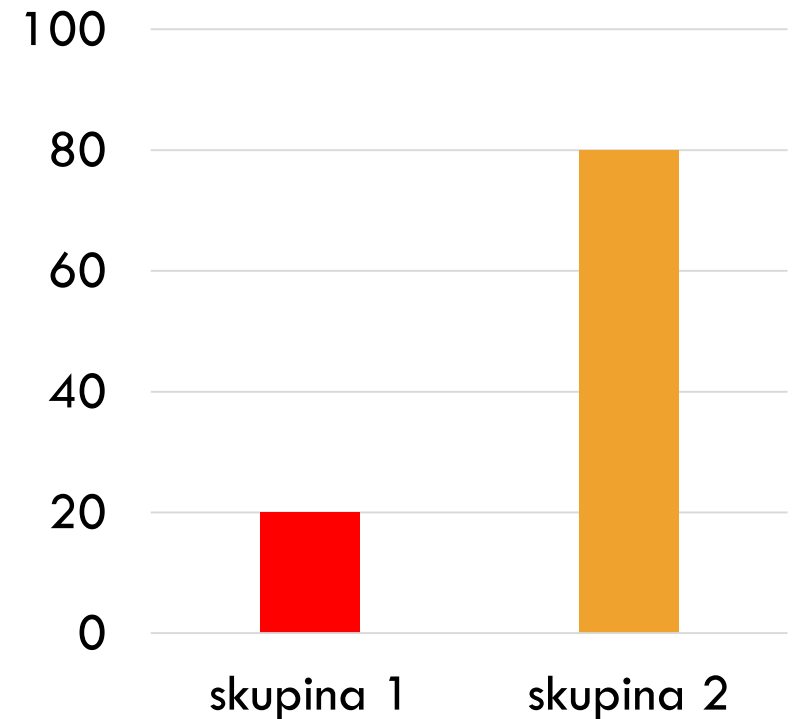
- Síla testu měří schopnost statistického testu detekovat rozdíly či souvislosti, o které se zajímáme (Hendl 2015, s. 417-422).
- Dosahuje hodnot 0 – 1. Běžně se doporučuje minimální síla testu o velikosti 0,8.
- Pro stanovení síly testu lze využít statistický software
 - ▣ při plánování výzkumu (např. Sample Power, G*Power);
 - ▣ ex post při analýze dat (např. SPSS).

Obecně platí, že čím menší je předpokládaný zjištěný efekt (rozdíl, souvislost), tím větší výběr musíme volit.

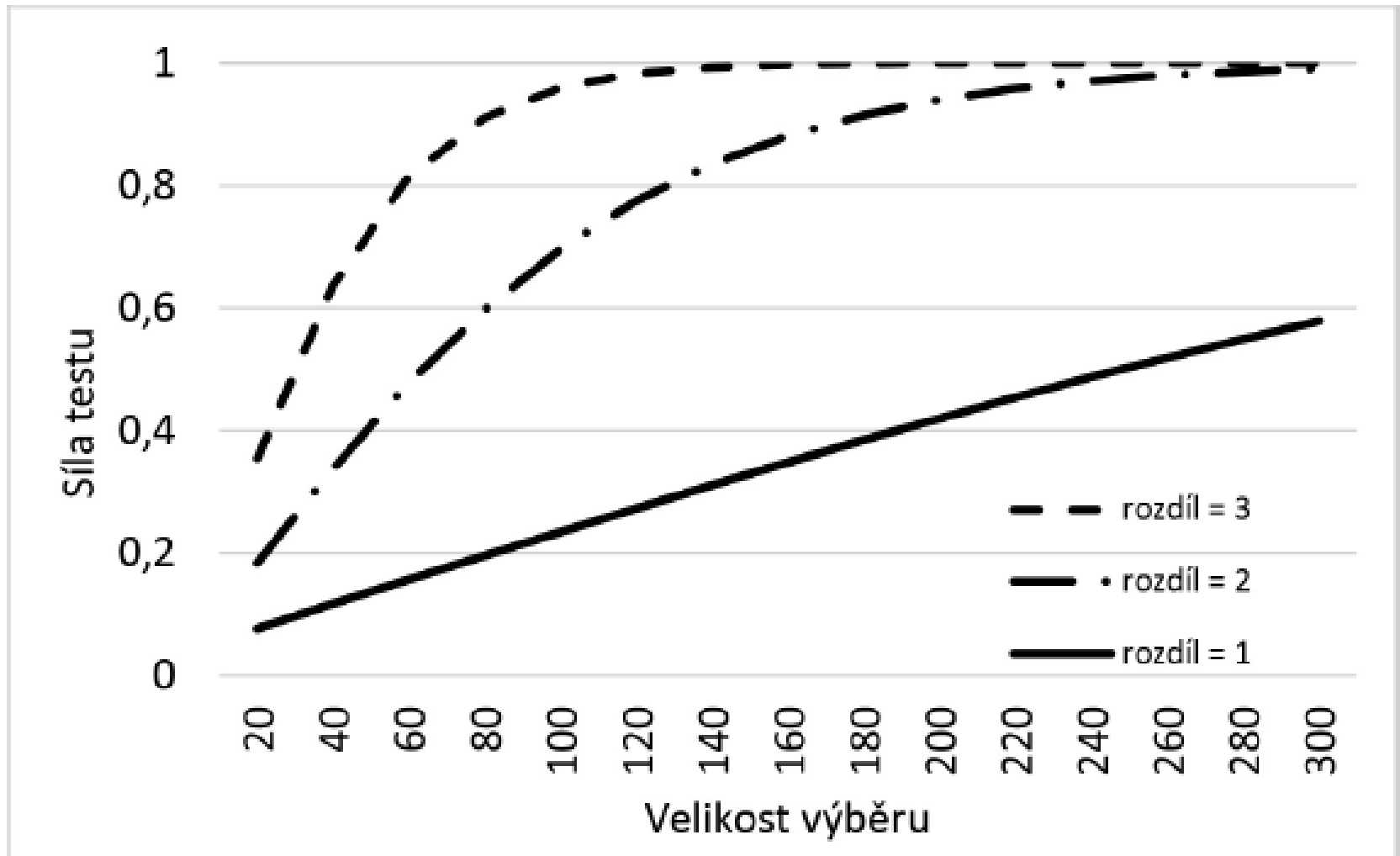
MALÝ ROZDÍL



VELKÝ ROZDÍL



Velikost síly testu pro dvouvýběrový t-test s ohledem na velikost výběru pro prostý náhodný výběr ($\alpha = 0,05$)



Ukázka 2: vliv velikosti výběru na výsledek testu

- 2 skupiny žáků psaly test.
- Bylo možné získat 0 – 100 bodů.
- H_0 : Výsledky dvou skupin žáků jsou shodné.

$$\mu_1 = \mu_2$$

- Ukázka zahrnuje 3 situace
 - $n = 100$
 - $n = 400$
 - $n = 600$

Proč vede užívání testů k černobílé interpretaci výsledků výzkumu

Zamítáme H_0 ?

- I za situace, kdy splňujeme veškeré podmínky pro vhodnou aplikaci testu statistické významnosti, je nutné si uvědomit, že otázka přijetí / zamítnutí námi testované hypotézy stojí na obecně přijímané **dohodě**, kdy pracujeme obvykle na 5% či 1% hladině statistické významnosti.
- Vytváříme tak umělou hranici, která sice přináší jednoznačné vědecké závěry, avšak za cenu extrémního zjednodušení reality.

Výsledek testu na hraně

Hypotéza	Hodnota statistické signifikance ($\alpha = 0,01$)	Interpretace
H_1	0,051	Nezamítáme H_0
H_2	0,050	Nezamítáme H_0
H_3	0,049	Nezamítáme H_0

Hypotéza	Hodnota statistické signifikance ($\alpha = 0,05$)	Interpretace
H_1	0,051	Nezamítáme H_0
H_2	0,050	Zamítáme H_0
H_3	0,049	Zamítáme H_0

Proč „statistická významnost“ není totéž
co „významnost“

Statistická a věcná významnost

STATISTICKÁ

- Zobecnitelnost
- Závisí na velikosti výběru
- Testy statistické významnosti

VĚCNÁ

- Praktická důležitost
- Nezávisí na velikosti výběru
- Popisné statistiky a míry věcné významnosti (effect size measures)

(Ne) závislost výsledků na velikosti výběru

Velikost výběru	Věcný rozdíl	Hodnota statistické signifikance
n = 100	1 bod	0,214
n = 400	1 bod	0,013
n = 600	1 bod	0,002

Pozor, rozhodně neplatí, že čím větší výběr, tím lepší!

Závěr

Netestujme mechanicky

- Aplikace statistického testu je možná pouze za splnění požadavků na reprezentativitu, velikost výběru a **dalších, které nebyly v rámci prezentace zmíněny.**
- „Statistická významnost“ není „významnost“. I statisticky významný výsledek může být věcně nevýznamný (a naopak).
- Zdůrazňování statistické významnosti vede k černobílé interpretaci výsledků výzkumů.

Reference

- Hendl, J. (2015). *Přehled statistických metod: analýza a metaanalýza dat* (Páté, rozšířené vydání). Praha: Portál.
- Krejčí, J. (2011). Kvalita výběrových šetření a standardy v sociálním výzkumu. Praha, Sociologický ústav AV ČR.
- Mareš, P., Rabušic, L., & Soukup, P. (2015). *Analýza sociálněvědních dat (nejen) v SPSS*. Brno: Masarykova univerzita.
- Soukup, P. (2016). Užívání statistické a věcné významnosti v časopise *Pedagogická orientace* a *Pedagogika* v posledních deseti letech: pohled statistika. *Pedagogická orientace*, 26(2), 182–201.
- Soukup, P., & Kočvarová, I. (2016). Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu. *Pedagogická orientace*, 26(3), 512–536.



Děkuji za pozornost!

Ilona Kočvarová

Centrum výzkumu FHS UTB ve Zlíně

kocvarova@utb.cz