

METODOLOGICKÉ POŽADAVKY NA VÝZKUMNÉ STUDIE

Výzkumné studie mají přinášet nová konkrétní zjištění získaná specifickými výzkumnými metodami. Prezentovaná zjištění by měla vždy vyplývat z realizovaného vlastního empirického šetření. Součástí výzkumné studie by měly být informace, které tato zjištění dokládají. Výzkumná studie by měla být zřetelně členěna na následující části: formulace problému a výzkumných hypotéz, přehled relevantní literatury, použité výzkumné metody a postupy, způsoby vyhodnocení empirických dat, výsledky a jejich interpretace, diskuse a závěry. V následujícím text uvádíme nejvýznamnější body, které by měl/a mít autor /autorka výzkumné studie na paměti s tím, že konkrétní ukazatele se mohou lišit např. vzhledem k použitému software.

Kvalitu výzkumné studie určuje především míra objektivity a reliability (nejčastěji používanou mírou vnitřní konsistence škály, resp. výzkumné metody, je **Cronbachův** standardizovaný **koeficient α** - alfa) prezentovaných výsledků a jejich ověřitelnost. Předpokladem pro publikaci studie, ve které byla použita **nová** nebo **adaptovaná metoda**, technika či nástroj, je **uvedení vlastností** včetně co nejpodrobnějšího popisu parametrů její standardizace - reliability, validity a dalších, například výsledky **položkové analýzy**.

K vyhodnocení empirických dat je třeba použít standardní metody **deskriptivní** statistiky (popisy rozložení dat) a **induktivní** statistiky, podle charakteru dat a úrovně měření (nominální, pořadová, intervalová, eventuálně i poměrová) buď **parametrické**, jsou-li splněny **podmínky** pro jejich použití, nebo **neparametrické**.

1. Je třeba rozlišovat **výzkumný soubor** a **vzorek**. Adekvátní výběr vzorku je nutnou podmínkou pro usuzování na všeobecně platné skutečnosti. Adekvátní výběr znamená **reprezentativní výběr**. Výběr osob je třeba podrobně popsat.

2. Získaná data je nutno otestovat z hlediska **normality** jejich **rozložení**. Normální rozložení dat je nutnou podmínkou pro použití parametrických testů. Soubory dat je třeba **popsat** pomocí měř **centrální tendence** (průměr, medián, modus) a měř **variability** (směrodatná odchylka, rozptyl, případně variační šíře).

3. Předmětem bivariační rozdílové statistiky bývá **vztah mezi dvěma proměnnými**. Data bývají získána na dvou nebo více souborech, případně opakovaným měřením u téhož souboru. Nejčastěji používaným testem pro srovnání těchto souborů je **t-test**. Je třeba uvést, zda jde o t-test pro **závislé** či **nezávislé** výběry, v případě t-testu pro nezávislé výběry otestovat **F-testem**, zda je třeba použít t-test pro **homogenní** či **heterogenní varianci**. Nejsou-li splněny předpoklady pro použití t-testu, je třeba použít některou neparametrickou metodu pro testování rozdílu mezi dvěma soubory, konkrétně mediánový test, znaménkový test, Wilcoxonův test, Wilcoxonův-Whiteův test, Mann-Whitneyho U-test nebo Kolmogorův-Smirnovův test. U všech testů je třeba **uvést hodnotu testového kritéria** a pravděpodobnost chyby prvního druhu, resp. **hladinu pravděpodobnosti**, kterou vypočítaná hodnota překračuje.

Není korektní používat t-testy opakovaně, například při srovnávání více než dvou skupin. V takovém případě je třeba vždy použít **jednoduchou analýzu rozptylu (ANOVA)**.

4. Bivariační statistika nehodnotí pouze významnost rozdílů mezi soubory dat, ale zjišťuje i **míry vztahu** mezi dvěma proměnnými, tj. míru **korelace** mezi nimi. (Testy analýzy závislosti dat se obecně dělí na korelační a regresní.) Pokud data splňují požadavky na použití parametrických testů, lze pro zjištění míry závislosti použít **Pearsonův** koeficient korelace r. Základní podmínkou jeho použití je dvourozměrné normální rozložení dat). Nejsou-li předpoklady splněny, je třeba použít některý neparametrický koeficient korelace, konkrétně

Spearmanovo ρ nebo **Kendalovo τ** . V případě vztahu mezi spojitou proměnnou a binární proměnnou je nutno použít bodově biseriální korelační koeficient. Vždy je třeba **uvést hodnotu testového kritéria** a pravděpodobnost chyby prvního druhu, resp. **hladinu pravděpodobnosti**, kterou vypočítaná hodnota překračuje.

5. **Velikost efektu** (effect size - ES) je mírou síly vztahu mezi dvěma proměnnými, která není ovlivněna velikostí vzorku. Je to užitečná deskriptivní statistika, která by měla vždy doplňovat testy inferenční statistiky. Mj. umožňuje srovnání výsledků různých výzkumů a je zásadní pro metaanalýzu. Ke zjištění velikosti efektu je nejčastěji používáno Cohenovo d , v případě neparametrických dat jeho neparametrická varianta GCLESS (generalization of common language effect size statistics). Existuje však řada měr velikosti efektu: jednak **standardizované**, jako **Cohenovo d** , (které udává velikost rozdílu mezi dvěma průměry ve směrodatných odchylkách a může být přepočítáno na korelační koeficient r), **Glassovo Δ** (delta), **Hedgeovo g** , **Cohenovo f^2** (f kvadrát), užívané při analýze rozptylu a mnohonásobné regresi, **Cramérov ϕ** (ϕ), které je nejlepší mírou velikosti efektu pro χ^2 (chí-kvadrát) test, poměr či **podíl šancí** (odds ratio) v případě dvou binárních proměnných, který je důležitý v logistické regresi, η^2 (eta-kvadrát) pro analýzu rozptylu (při statisticky nevýznamné hodnotě F je postačující mírou, při statisticky významné hodnotě F je třeba uvést Cohenovo d pro každé párové srovnání), jednak **nestandardizované** (rozdíl mezi průměry skupin nebo nestandardizované regresní koeficienty), ale mírou velikosti efektu je také bivariační lineární korelace (r), mnohonásobná korelace nebo koeficient determinace (R). Intervaly spolehlivosti pro nestandardizované míry velikosti efektu jsou uvedeny v tabulkách.

6. Vždy je třeba uvést **interval spolehlivosti** jako ukazatel reliability, tj. spolehlivosti. U většiny hodnocených sérií měření neznáme skutečnou hodnotu, např. průměru. Lze však vymezit oblast, v níž se s určitou pravděpodobností (na předem zvolené hladině významnosti) skutečná hodnota nachází. Tato oblast - **interval spolehlivosti** - je tím užší, čím jsou získané výsledky přesnější, a charakterizuje spolehlivost výsledku. Nejjednodušší způsob výpočtu využívá Studentova rozdělení, způsob navržený Dean-Dixonem pro malé soubory Lordovo rozdělení.

7. Nejpoužívanější metodou analýzy kategoriálních dat je χ^2 (chí-kvadrát) test dobré shody. Hodnoty očekávaných četností v jednotlivých kategoriích musí dosahovat minimálně 5. Je třeba uvést pozorované i očekávané hodnoty, **hodnotu testového kritéria** a **hladinu pravděpodobnosti**, kterou vypočítaná hodnota překračuje.

8. Multivariační statistické metody čili metody vícerozměrné analýzy berou na rozdíl od bivariačních v úvahu více než dvě proměnné. Nejpoužívanější metodou je **explorační faktorová analýza**. Je třeba uvést zjištěné faktory, faktorové zátěže všech proměnných, procenta variance vyčerpaná jednotlivými faktory, komunality, použitou metodu rotace, faktorové skóry, vlastní hodnoty (eigen values) a použité kritérium volby počtu faktorů.

9. Při hledání vztahu mezi prediktorem a kritériem se doporučuje použít **vícenásobnou lineární regresi**. Jako při použití lineární regresní analýzy je třeba uvést regresní rovnice, R^2 (proporce společné variance), β (standardní regresní koeficienty) a r (korelace mezi skutečnými a odhadnutými skóry).

10. Metody analýzy rozptylu – ANOVA a MANOVA patří k velmi frekventovaným. Vždy je nutno uvést **součty čtverců, počty stupňů volnosti, podíly čtverců, hodnoty F-testů a hladiny pravděpodobnosti**. Hladiny významnosti u paralelních testů je třeba upravit podle Bonferroniho korekce. V případě MANOVA je třeba uvést i hodnoty Wilksovy Λ (lambda). Analogicky jako u všech analýz variance je třeba postupovat při použití analýzy kovariance.

11. **Logistická regresní analýza** není výrazně závislá na předpokladech multinormálního rozložení a homogenního rozložení variance. Logistický regresní model musí uvést hodnoty B, S.E., Wald, stupňů volnosti, významnost, R (mnohonásobný koeficient determinace - logistický ekvivalent koeficientu parciální korelace) a Exp(B). Analogicky je třeba postupovat při použití mnohonásobné regresní analýzy a hierarchické regresní analýzy.

12. **Shluková analýza** (cluster analysis) je založena na využití měr nepodobnosti, resp. podobnosti. Je třeba uvést matice měření, matice vzdáleností, upravené matice vzdáleností a dendrogramy.

13. **Analýza hlavních komponent** (PCA – principal component analysis) se zabývá možnostmi redukce počtu proměnných, které popisují variabilitu všech proměnných a vztahy mezi nimi. Je třeba uvést tabulky pro vícerozměrnou analýzu, korelační matice a scree graf, zobrazující svislost vysvětlené variability na počtu hlavních komponent.

14. **Kanonická korelační analýza** zobecňuje mnohonásobnou lineární regresní analýzu a má společné některé rysy s metodou hlavních komponent. Je třeba uvést korelační matice.

15. **Diskriminační analýza** posuzuje pomocí mnohorozměrných dat odlišnosti určité populace a možnosti nalezení predikční rovnice pro klasifikaci objektů do skupin. Je třeba uvést rovnice diskriminačních funkcí, porovnání různých klasifikačních postupů a jejich výsledky.

16. **Mnohorozměrné škálování** (MDS – multidimensional scaling) souvisí s redukcí dimenzionality dat. Je třeba uvést tabulku aktuálních a predikovaných nepodobností.

17. **Konfirmační faktorová analýza** patří mezi metody modelování pomocí latentních proměnných. Hypotéza o matici zátěží je vyjádřena pomocí předpokládaného ideálního tvaru matice, kde jedničky představují parametry, jež mají být odhadnuty a nuly pevné omezující podmínky. Je třeba uvést indexy shody - např. chí-kvadrát, hladinu pravděpodobnosti, Tucker-Lewisův index (TLI), komparativní index shody (CFI) a odmocninu z průměrného čtverce chyby odhadu (RMSEA)

18. Modelování pomocí **strukturních rovnic** (SEM) kombinuje faktorovou analýzu, mnohonásobnou regresi a úsekovou analýzu. Je třeba uvést příslušné indexy shody, např. chí-kvadrát, stupně volnosti, hladinu pravděpodobnosti, RMSEA (Root Mean Square Error Of Approximation), interval spolehlivosti a SRMR (Standardized Root Mean Square Error Residual).

Analýza s využitím **modelu LISREL** (Linear structural relations analysis) se zabývá analýzou vztahů mezi manifestními a latentními proměnnými. Může být použita např. při realizaci výpočtů konfirmační faktorové analýzy. Je třeba uvést vypočítané hodnoty adjustovaného χ^2 (chí-kvadrátu), GFI (Goodness of Fit Index) a RMS (Root Mean Square Residual).

19. Doporučuje se využívání standardních statistických balíků pro analýzu dat ve společenských vědách (SPSS, BMDP, Statistica).