

# JAK ČÍST SEKUNDÁRNÍ ANALÝZU: METODOLOGIE A POUŽITÉ METODY ANALÝZY



V rámci sekundární analýzy TALIS 2018 byly použity metody statistické analýzy běžně využívané v pedagogických a sociálních vědách. Mezi tyto patří deskriptivní statistiky s tříděním prvního a druhého stupně, které poskytují souhrnné informace o dílčích proměnných a jsou tak základem pro provedení pokročilých analýz a sestavení komplexních modelů. Ve velké míře jsou využívány grafickou formou (graf chybových úseček) znázorněné analýzy rozdílů v průměrech (t-test, ANOVA), sloužící k porovnávání statistické podobnosti průměrů mezi skupinami žáků či učitelů.



Indexy a další proměnné, které jsou v rámci TALIS měřeny na kardinální škále, byly pro účely deskriptivní statistiky rozděleny na jednotlivé kategorie za pomoci metody *Jenks Natural Breaks*, která slučuje sousedící hodnoty do skupin (klastřů) vyznačujících se minimalizací odchylky od průměru dané skupiny, zatímco mezi jednotlivými skupinami zachovává maximální rozdíly. Hranice kategorií proto respektují přirozeně se vyskytující rozdíly mezi shluky (skupinami) dat na kardinální škále. Pro každý index jsou nicméně hranice specifické a nelze je zobecnit. Princip fungování metody *Jenks Natural Breaks* je vizuálně znázorněn v příloze zprávy.



Efekty některých faktorů na postoje učitelů mohou být ovlivněny přítomností či absencí dalších proměnných, které s těmito faktory souvisejí. Proto kombinujeme jednoduché deskriptivní techniky s komplexnějšími multivariačními statistickými modely.



Pro zjištění komplexnějších vztahů je používáno hierarchické regresní modelování na dvou úrovních (učitel a škola). Regresní model je statistickou technikou, která dovoluje zjistit „efekt“ (ve smyslu asociace) konkrétní proměnné očištěné o efekt dalších do modelu zahrnutých proměnných. Hierarchické regresní modely (HLM) jsou využívány oproti jednoduché lineární regresi s ohledem na hierarchickou strukturu analyzovaných datových souborů (škola → učitel), a to z toho důvodu, že jednoduchá lineární regrese není schopna korektně vyhodnotit statistickou významnost a vztahy mezi proměnnými v hierarchické struktuře dat (viz Gelman a Hill 2007; Gelman, Hill, a Vehtari 2020).



Modely byly testovány s náhodnou konstantou na úrovni školy, proto jsou v textu zobrazeny modely s druhou úrovní odpovídající škole. Přestože v případě postojových otázek učitelů jsou menší rozdíly mezi školami (vnitrotřídní koeficient korelace ICC v rozmezí 5–10 %), preferujeme používání HLM regresi oproti jednoduché lineární regresi.



Jednodušší i složitější vztahy testované regresními modely jsou interpretovány za pomoci vizuálního znázornění grafu regresních koeficientů. Body znázorňují regresní koeficient, chybové úsečky pak méně striktní 90% konfidenční interval. Veškeré proměnné byly standardizovány na škálu dvou směrodatných odchylek, aby bylo možno srovnat sílu asociačního vztahu napříč proměnnými. Jedná se o tzv. Gelmanovu metodu (viz Gelman 2008), kdy efekty proměnných se interpretují následovně, pokud se nezávislá proměnná změní zhruba z minima na maximální hodnotu, hodnota závislé proměnné se změní o hodnotu koeficientu. Výhodou je porovnatelnost škálových a ordinálních proměnných s kategoriálními binárními (dummy) proměnnými (Gelman 2008, s. 2867).



Jednotlivé faktory vstupují do modelů a komplexnějších multivariačních analýz jako indexy, které jsou přímo tvořené odborným týmem TALIS. Jednotlivé indexy jsou v závorce uvedeny kódem z datového souboru a z technické zprávy OECD k TALIS 2018. Je nutné mít na paměti, že se ve všech případech jedná o deklaratorní odpovědi učitelů a ředitelů. Validita takto získaných odpovědí je obecně nižší, protože v dotazníkových šetřeních mohou respondenti záměrně své výpovědi zkreslit, například z důvodu sociální konvence (tzv. social desirability bias).



Šetření TALIS 2018 bylo tvořeno odpověďmi učitelů z učitelského dotazníku a odpověďmi ředitelů škol. Veškeré proměnné tak vyjadřují subjektivní vnímání daných faktorů učiteli nebo řediteli. Odpovědi ředitelů jsou vždy na úrovni školy. Proměnné učitelského dotazníku vstupují do regresních modelů ve dvou úrovních. Individuální úroveň učitele (první úroveň), agregovaný průměr učitelských odpovědí na úrovni školy (průměr). Je důležité zdůraznit, že jedna proměnná může měřit odlišné jevy na úrovni učitele oproti agregované úrovni školy. Jedná se tak o kontextuální či kompoziční faktory školy ovlivňující postoje učitelů. Celá řada postojových konceptů je pak navzájem endogenní, to znamená, že nejsme schopni na koncepční úrovni odlišit příčinu a následek, jsme schopni pouze měřit souvislosti mezi těmito jevy. Některé postoje mohou být považovány i jako intervenující proměnné, které vztahy zprostředkovávají.