

# JAK ČÍST SEKUNDÁRNÍ ANALÝZU: METODOLOGIE A POUŽITÉ METODY ANALÝZY



Agregace na úroveň krajů je prováděna jen u žáků 9. ročníku základních škol a odpovídajícího ročníku víceletých gymnázií.



V rámci sekundární analýzy 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.



Indexy a další proměnné, které jsou v rámci PISA 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 této zprávy.



Pro zjištění komplexnějších vztahů je používáno hierarchické regresní modelování na dvou úrovních (žák a škola). Regresní model je statistickou technikou, která dovoluje zjistit „efekt“ (ve smyslu asociace) konkrétní proměnné očistěné o efekt dalších proměnných, které byly do modelu zahrnuty. Hierarchické regresní modely jsou využívány oproti jednoduché lineární regresi s ohledem na hierarchickou strukturu analyzovaných datových souborů (stát → škola → třída → žák), 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. Jednodušší i složitější vztahy testované regresními modely jsou interpretovány za pomoci vizuálního znázornění grafů regresních koeficientů. Body znázorňují regresní koeficient, chybové úsečky pak méně striktní 90% konfidenční interval. Veškeré proměnné byly standardizovány do jednotky směrodatné odchylky, aby bylo možno srovnat sílu asociačního vztahu napříč proměnnými.



Jednotlivé faktory vstupují do modelů a komplexnějších multivariačních analýz jako indexy, které jsou přímo tvořené odborným týmem PISA. Jednotlivé indexy jsou v závorce uvedeny kódem z datového souboru a z technické zprávy OECD k PISA 2018.



Proměnné žákovského dotazníku vstupují do regresních modelů ve dvou úrovních. Individuální úroveň žáka (první úroveň), agregovaný průměr žákovských odpovědí na úrovni školy (průměr). Je důležité zdůraznit, že jedna proměnná může měřit odlišné jevy na úrovni žáka oproti agregované úrovni školy.



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