



Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

TECHNICKÁ ZPRÁVA K METODOLOGII

ANALÝZY TRENDŮ V NEROVNOSTECH VZDĚLÁVACÍ SOUSTAVY
V ČESKÝCH REGIONECH

ANALÝZA NÁRODNÍCH DAT ČASOVÉ POVAHY Z OBLASTI
VZDĚLÁVÁNÍ PILOTÁŽ I

ODBOR ŠKOLSKÉ STATISTIKY A ANALÝZ

Sekce informatiky, statistiky a analýz

MŠMT, Praha



OBSAH

Úvod	4
1 Víceúrovňová analýza: Faktory související s nerovnostmi ve vzdělávací soustavě	5
1.1 Identifikace proměnných v modelu.....	5
2 Metodický postup a specifikace víceúrovňového modelu	8
2.1 Kontrola kvality dat a distribucí proměnných	8
2.2 Identifikace odlehlých hodnot.....	8
2.3 Korelační analýza nezávislých proměnných	9
2.4 Analýza kontingenčních tabulek.....	10
2.5 Analýza hlavních komponent (PCA)	10
2.6 Konstrukce kompozitních indexů	11
2.7 Krokový výběr regresorů na základě informačního kritéria	12
2.8 Typy směrodatných chyb a předpoklady modelů	12
2.9 Ověření hierarchické struktury dat pomocí nulového	14
víceúrovňového modelu.....	14
2.10 Posouzení bivariátních vztahů a vhodnosti lineárního modelu	15
2.11 Regresní modely se smíšenými efekty a hierarchickou strukturou dat.....	15
3 Analýza dat a výsledky	17
3.1 Explorační analýzy dat (EDA).....	17
3.2 Identifikace odlehlých hodnot.....	18
3.3 Korelační a asociační analýza nezávislých proměnných.....	18
3.4 Návrh kompozitních proměnných na základě analýzy hlavních komponent	23
3.5 Krokový výběr regresorů na základě informačního kritéria	25
3.5.1 Krokový výběr mezi základními školami.....	25
3.5.2 Krokový výběr mezi středními školami	26
3.5.3 Společné závěry	27
3.6 Ověření hierarchické struktury dat (nulový model, ICC).....	27
3.7 Posouzení bivariátních vztahů a vhodnosti lineárního modelu	29
3.8 Víceúrovňová regresní analýza se zohledněním regionální struktury.....	33
3.8.1 Diagnostika	34
4 Příloha.....	36
4.1 Víceúrovňová analýza: Faktory související s nerovnostmi ve vzdělávací soustavě.....	36
4.2 Vícerozměrné statistické analýzy umožňující charakterizovat prediktory nerovnosti na úrovni dospělé populace	48
5 Seznamy.....	55

5.1	Seznam tabulek	55
5.2	Seznam grafů	55

ÚVOD

Tento dokument představuje technickou a metodologickou přílohu ke zprávě *Analýza národních dat časové povahy z oblasti vzdělávání pilotáž I*. Zatímco analytická zpráva je koncipována s důrazem na srozumitelnost pro široké spektrum aktérů vzdělávací politiky a soustředí se primárně na interpretaci zjištění a jejich strategické implikace, cílem této technické zprávy je poskytnout detailní, transparentní a replikovatelný popis použitých statistických a analytických postupů.

Dokument je určen především odborné veřejnosti, datovým analytikům, výzkumníkům a akademikům. Podrobně dokumentuje celý proces zpracování dat od fáze explorační analýzy (EDA) a čištění datových sad, přes řešení problematiky chybějících a odlehlých hodnot, až po diagnostiku multikolinearity a redukci dimenzionality pomocí analýzy hlavních komponent (PCA).

Stěžejní část zprávy je věnována specifikaci vícerozměrných statistických modelů. Detailně popisuje proces krokového výběru regresorů na základě informačních kritérií (AIC, BIC), testování předpokladů lineární regrese a volbu vhodných typů směrodatných chyb (včetně cluster-robustních odhadů). Zvláštní pozornost je věnována matematické formulaci a zdůvodnění použití víceúrovňových (multilevel) regresních modelů se smíšenými efekty, které byly nezbytné pro korektní zachycení hierarchické struktury dat (školy vnořené do regionů) a pro oddělení vlivu institucionálních charakteristik od socioekonomického kontextu území.

1 VÍCEÚROVŇOVÁ ANALÝZA: FAKTORY SOUVISEJÍCÍ S NEROVNOSTMI VE VZDĚLÁVACÍ SOUSTAVĚ

Deskriptivní analýza poskytla základní vhled do problematiky nerovností a umožnila zmapovat rozdíly mezi školami a regiony. Pro hlubší pochopení vztahů je však nutné využít vícerozměrné statistické metody, které umožňují podrobněji zkoumat souvislosti mezi jednotlivými **nezávislými proměnnými** (nazývanými též regresory, prediktory, kovariátami či vysvětlujícími proměnnými) a definovanou **závislou proměnnou** (též regresandem, cílovou, vysvětlovanou či predikovanou proměnnou). Vliv ostatních proměnných je statisticky kontrolován.

Kapitola nejprve vymezuje použité proměnné a metodický postup, zahrnující kontrolu kvality dat, konstrukci kompozitních indexů a výběr regresorů. Následně je ověřena hierarchická struktura dat pomocí nulového víceúrovňového modelu (ICC) a posouzena vhodnost lineární specifikace.

Základní analytickou metodou zvolenou pro tuto část je **lineární regresní analýza**. Její volba je opodstatněná tím, že sledované výsledky lze vyjádřit jako **číselné hodnoty na plynulé škále** (například podíl úspěšných žáků se může pohybovat od 0 % do 100 %). To umožňuje odhadnout sílu a směr vztahu mezi jednotlivými nezávislými proměnnými (prediktory) a sledovaným výsledkem.

Vzhledem k tomu, že některé nezávislé proměnné vstupují do analýzy na úrovni školy a jiné na úrovni regionu (ORP), je standardní jednoúrovňový model nedostatečný. Hierarchická povaha dat, kdy školy spadají do jednotlivých regionů, vyžadovala ve většině případů použití **víceúrovňové (multilevel) specifikace** se smíšenými efekty. To umožnilo oddělit vliv charakteristik na úrovni školy (např. kvalifikovanost učitelů) od vlivu charakteristik na úrovni regionu (např. míra nezaměstnanosti v ORP). Model tedy dokáže přesněji určit, jaká část rozdílů ve výsledcích souvisí s charakteristikami samotné školy a jaká část s vnějším socioekonomickým kontextem.

1.1 IDENTIFIKACE PROMĚNNÝCH V MODELU

Pro účely regresního modelování byly definovány závislé proměnné a dvě úrovně nezávislých proměnných, které mohou výsledky ovlivňovat. Níže uvedený přehled nepředstavuje kompletní seznam vstupních proměnných, ale naopak tematické okruhy, z nichž byly konkrétní regresory vybírány. Uvažováno bylo celkem 220 školních indikátorů, jejichž konečný výběr byl postupně upřesňován podle metodických a statistických kritérií. Tematické členění napomáhá přehlednému zachycení klíčových oblastí fungování školy, tzn. její velikost a stabilitu, složení žáků, personální a materiální podmínky i podpůrné kapacity. Členění rovněž napomáhá sledovat jejich vztah k výsledkům školy.

Při výběru konkrétního **referenčního roku** jsme hledali kompromis mezi maximální aktuálností dat a jejich co nejvyšší mírou pokrytí (vyplněnosti). Cílem bylo pracovat s nejaktuálnějšími dostupnými údaji, avšak zároveň minimalizovat rozsah chybějících hodnot u klíčových ukazatelů. Z tohoto důvodu jsou **základní školy** analyzovány až za **rok 2024**, kdežto u **středních škol** je použito již **období roku 2023**. Pro rok 2024 totiž nebyla v době zpracovávání k dispozici kompletní a metodicky srovnatelná data z CZVV představující významný vstup do sledovaných výsledkových ukazatelů.

Závislé proměnné (úroveň školy)

Jako závislé proměnné, tedy hlavní ukazatele výsledků vzdělávání, byly zvoleny následující:

- Hodnocení školy: Průměrné hodnocení podmínek, průběhu a výsledků vzdělávání získaná z inspekční činnosti

- Výsledky u jednotné přijímací zkoušky (JPZ): Podíl žáků devátých tříd dané ZŠ, kteří u JPZ dosáhli alespoň celorepublikového mediánu souhrnného výsledku z českého jazyka a matematiky dohromady.
- Úspěšnost dokončení středního vzdělání: Podíl žáků dané SŠ, kteří v daném roce úspěšně ukončili studium maturitní zkouškou nebo získáním výučního listu.

Okruhy nezávislých proměnných (prediktorů)

Do modelů vstupovaly dvě skupiny nezávislých proměnných, které odpovídají hierarchické struktuře dat. Jedná se o charakteristiky na úrovni školy a charakteristiky na úrovni regionu (ORP).

Charakteristiky na úrovni školy (Level 1) zahrnují ukazatele popisující konkrétní školu, její žáky a pedagogy. Jde o proměnné z tematických oblastí popsanych v kapitole 3, například:

- základní charakteristiky školy (zřizovatel, počty žáků, počty tříd, klasifikace převládajícího typu střední školy),
- sociodemografická struktura žáků (podíly dívek, počty žáků s jiným než českým státním občanstvím, počty odkladů povinné školní docházky),
- Inkluze a žáci se SVP (podíly žáků se SVP, podíly nadaných žáků),
- mobilita žáků v systému vzdělávání a opakování ročníků (předčasné odchody žáků bez další vzdělávací trajektorie, přechody žáků na jinou školu, postupy do vyššího navazujícího ročníku, podíl žáků opakujících ročník),
- pedagogický sbor a poradenské kapacity (podíl kvalifikovaných učitelů, úvazky pedagogických pracovníků, tj. asistentů pedagoga, výchovných poradců, metodiků prevence),
- podpůrné kapacity (úvazky školních psychologů, sociálních pedagogů, čerpání šablon na financování personálního zajištění podpůrných profesí ve školách),
- materiální a technické vybavení (počet IT vybavení přístupných žákům),
- volnočasové a podpůrné aktivity v ZŠ (dostupnost a kapacita školních družin a školních klubů, rozsah nabízených činností).

Proměnné z těchto tematických oblastí společně vytvářejí rámec pro porozumění fungování školy. To znamená, že zachycují strukturu a stabilitu školy, složení a potřeby žáků, personální a materiální podmínky, dostupnost podpůrných profesí, vzdělávací výsledky i kvalitu řízení a rozvoje školy. Zakomponování těchto proměnných do modelu dává prostor pro sledování konkrétních charakteristiky uvnitř školy souvisejících s jejími výsledky nebo kvalitou.

Charakteristiky na úrovni regionu – ORP (Level 2) zahrnují socioekonomické a demografické ukazatele za příslušné ORP, které vymezují širší kontext působení školy. Jde o proměnné podrobně popsané v analytické zprávě, tedy:

- podíl nezaměstnaných osob,
- podíl osob v exekuci,
- podíl vyplácených přídavků na dítě,
- podíl osob bez dosaženého vzdělání s maturitou,
- podíl osob ve věku 65 a více let,
- meziroční přírůstek/úbytek obyvatel.

Vedle těchto měřitelných charakteristik je pro každou obec s rozšířenou působností (ORP) do modelu začleněn rovněž tzv. náhodný intercept. Tento statistický prvek modeluje specifickou výchozí úroveň výsledků pro každé ORP. Zachycuje a statisticky kontroluje vliv latentních regionálních specifik, jako je například lokální kultura nebo historický vývoj. Tato specifika mohou ovlivňovat průměrnou úroveň výsledků všech škol v daném ORP, a to nad rámec sledovaných socioekonomických ukazatelů.

2 METODICKÝ POSTUP A SPECIFIKACE VÍCEÚROVŇOVÉHO MODELU

2.1 KONTROLA KVALITY DAT A DISTRIBUCÍ PROMĚNNÝCH

Před zahájením vlastní analytické části byla provedena systematická kontrola kvality vstupních dat se zaměřením na jejich úplnost, věcnou správnost a rozložení hodnot jednotlivých proměnných. Zvláštní pozornost byla věnována podílovým a poměrovým ukazatelům, u nichž může zejména u menších škol docházet k výrazné nestabilitě hodnot. Kontrola dat umožnila identifikovat extrémní pozorování, nevhodně kódované chybějící hodnoty i proměnné, u nichž bylo vhodné zvážit transformaci nebo omezení analyzovaného vzorku. Tento krok vedl k redukci jak některých proměnných, tak i škol, které by mohly nepřiměřeně zkreslovat výsledky následné analýzy.

Výsledkem **explorační analýzy dat** je souhrnný přehled poskytující základní informace o struktuře, kvalitě a rozdělení proměnných ve vstupním datovém souboru. Pro každou proměnnou je uveden její datový typ, počet chybějících hodnot vzhledem k analyzovanému souboru škol, počet unikátních hodnot a grafické zobrazení empirického rozložení pomocí histogramů a boxplotů. Dále jsou prezentovány vybrané popisné statistiky, konkrétně:

1. minimum,
2. maximum,
3. výběrový průměr $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$,
4. medián $x_{0,50}$,
5. mezikvartilové rozpětí IQR = $x_{0,75} - x_{0,25}$ (rozdíl horního a dolního kvartilu),
6. výběrová směrodatná odchylka $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$,
7. variační koeficient (v %) $CV = \left(\frac{s}{\bar{x}}\right) \cdot 100$,
8. modus.

Kvantily x_α jsou určeny dle následujícího pravidla:

$$n \cdot \alpha = c \begin{cases} \text{je celé číslo, pak } x_\alpha = \frac{x_{(c)} + x_{(c+1)}}{2}, \\ \text{jinak, pak } x_\alpha = x_{(c)} \end{cases}, \text{ kde } n \text{ značí počet prvků.}$$

Uvedené statistiky a grafické výstupy umožňují rychlou orientaci v rozsahu, variabilitě a tvaru rozdělení dat a tvoří nezbytný základ pro navazující analytické postupy.

2.2 IDENTIFIKACE ODLEHLÝCH HODNOT

Identifikace odlehlých hodnot byla provedena na základě numerických ukazatelů vybraných z předchozí explorační analýzy dat (dále jen EDA). Do této fáze byly zahrnuty pouze kvantitativní proměnné splňující požadavky na dostatečnou variabilitu a datovou kvalitu.

Pro každou numerickou proměnnou bylo určeno, zda je ředitelství v daném ukazateli považováno za extrémně odlehlé, a to v případě, že jeho hodnota leží mimo mezní interval stanovený metodou IQR, tj.

$$(IQR - 3 * x_{0,25}; IQR + 3 * x_{0,25}).$$

Pozornost je věnována zejména tomu, **v kolika ukazatelích je dané ředitelství klasifikováno jako extrémní**, tedy jak často jeho hodnoty překračují stanovené mezní intervaly, a současně tomu, **zda je ředitelství z hlediska dostupných dat dostatečně informačně bohaté** pro další analýzu. Kritéria pro vyřazení ředitelství z další analýzy byla stanovena kombinací vysokého počtu odlehlých hodnot a zvýšeného podílu chybějících dat.

2.3 KORELAČNÍ ANALÝZA NEZÁVISILÝCH PROMĚNNÝCH

Pro lepší pochopení vztahů mezi prediktory a identifikaci potenciální multikolinearity byla provedena korelační analýza nezávislých proměnných. Ta umožnila odhalit skupiny ukazatelů, které spolu silně souvisejí a pravděpodobně zachycují podobné aspekty fungování školy. Výsledky korelační analýzy sloužily jako podklad pro výběr reprezentativních proměnných a pro rozhodnutí o případné redukci dimenze před regresním modelováním.

Výsledky korelační analýzy jsou prezentovány **korelační maticí** numerických proměnných. Řádky i sloupce tabulky odpovídají totožným proměnným a jednotlivé buňky obsahují hodnoty **Pearsonova korelačního koeficientu** vyjadřující míru lineární závislosti mezi dvojicemi ukazatelů:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

kteřý nabývá hodnot v intervalu $[-1,1]$. Hlavní diagonála matice obsahuje hodnotu 1, tedy korelaci proměnné se sebou samou. Pro zvýšení přehlednosti a omezení redundantní informace je v tabulce zobrazena pouze horní trojúhelníková část korelační matice, zatímco spodní trojúhelník je ponechán prázdný.

Zároveň jsou v tabulce ponechány pouze **statisticky významné korelační koeficienty** na 5 % hladině významnosti; nesignifikantní vztahy jsou potlačeny. Statistická významnost korelačních koeficientů byla posuzována testem nulové hypotézy $H_0: \rho = 0$, tj. že mezi proměnnými neexistuje lineární závislost, proti alternativní hypotéze $H_0: \rho \neq 0$. Nesignifikantní korelace přitom neznamenají absenci vztahu mezi proměnnými, ale pouze nedostatek důkazů pro existenci lineární závislosti odlišné od nuly, což může souviset například i s omezenou hustotou vzorku u vybraných proměnných. Toto zobrazení umožňuje rychlou identifikaci relevantních lineárních vazeb bez zatížení čtenáře množstvím slabých či statisticky neprůkazných vztahů.

Součástí tabulky jsou rovněž doplňkové souhrnné ukazatele vypočtené pro jednotlivé proměnné. Konkrétně je uveden **počet nesignifikantních korelací**, který poskytuje orientační informaci o tom, do jaké míry daná proměnná vykazuje pouze slabé nebo statisticky neprůkazné vztahy s ostatními ukazateli. Tento ukazatel slouží jako podpůrná informace při posuzování informační přidané hodnoty proměnné, přičemž nesignifikantní korelace samy o sobě nepředstavují problém z hlediska multikolinearity.

Hlavním analytickým cílem korelační analýzy je identifikace silně korelovaných proměnných, které by neměly současně vstupovat do regresních modelů z důvodu rizika multikolinearity. K tomuto posouzení slouží doplňkové ukazatele **počtu silných korelací**, definovaných jako korelace s absolutní hodnotou koeficientu vyšší než 0,8, a **průměrné absolutní hodnoty korelačního koeficientu**. Tyto informace poskytují podklad pro výběr reprezentativních proměnných a případnou redukci dimenze před vlastním regresním modelováním.

Pearsonův korelační koeficient předpokládá lineární vztah mezi spojitými proměnnými, absenci výrazných odlehlých hodnot a **přibližnou normalitu dat**. V případech porušení těchto předpokladů by byly vztahy ověřovány také pomocí robustnějších alternativ založených na pořadí.

2.4 ANALÝZA KONTINGENČNÍCH TABULEK

Analýza kontingenčních tabulek (Asociační analýza) slouží jako doplněk ke korelační analýze s cílem posoudit vztahy mezi binárními a kategoriálními nezávislými proměnnými a identifikovat potenciálně redundantní ukazatele před regresním modelováním. Analýza umožňuje zachytit statistické závislosti mezi kvalitativními proměnnými, které nelze vhodně hodnotit pomocí korelačních koeficientů.

Míra statistické závislosti mezi dvojicemi kategoriálních proměnných byla testována pomocí **chí-kvadrát testu nezávislosti**. Test nulové hypotézy H_0 : proměnné jsou stochasticky nezávislé, byl proveden na 5 % hladině významnosti. Testová statistika má tvar

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(o_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}},$$

Kde o_{ij} označuje pozorovanou četnost v buňce kontingenční tabulky a $\hat{e}_{ij} = \frac{n_{i \cdot} n_{\cdot j}}{n}$ je očekávaná četnost za předpokladu nezávislosti.

Síla zjištěné asociace byla kvantifikována pomocí **Cramérova koeficientu** $V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}}$, kde n je celkový počet pozorování a $k = \min(r, c)$ je menší z počtů řádků a sloupců kontingenční tabulky. Koeficient V nabývá hodnot v intervalu $(0,1)$, přičemž vyšší hodnoty indikují silnější asociaci.

Výsledky analýzy jsou prezentovány analogicky jako u korelační analýzy formou **asociační matice**, v níž jsou v jednotlivých buňkách uvedeny hodnoty Cramérova V . Pro zvýšení přehlednosti je zobrazena pouze horní trojúhelníková část matice a ponechány jsou pouze statisticky významné asociace; nesignifikantní vztahy jsou potlačeny. Hlavní diagonála matice obsahuje hodnotu 1, odpovídající asociaci proměnné se sebou samou.

Součástí výstupu jsou rovněž **doplňkové souhrnné ukazatele**, zejména počet nesignifikantních asociací, počet silných asociací (ty s hodnotou koeficientu vyšší než 0,3) a průměrná hodnota Cramérova V pro jednotlivé proměnné. Navíc je uveden i počet **porušení podmínek dobré aproximace**, tj. případů, kdy nebylo splněno kritérium, že alespoň 80 % očekávaných četností dosahuje hodnoty 5 a více a zbývající četnosti jsou alespoň vyšší než 2; pro tyto případy se testové statistiky nepočítaly.

2.5 ANALÝZA HLAVNÍCH KOMPONENT (PCA)

U tematicky příbuzných skupin proměnných byla využita analýza hlavních komponent jako nástroj pro redukcii dimenze a konstrukci kompozitních indexů. Tento postup umožnil shrnout informace z většího počtu korelovaných ukazatelů do menšího počtu komponent, které reprezentují obecnější charakteristiky školy, například úroveň velikosti školy, míra zastoupení dívek.

Vstupní datová matice $X \in \mathbf{R}^{n \times p}$ byla tvořena numerickými proměnnými. Před samotnou analýzou byly proměnné centrovány a standardizovány, čímž vznikla matice Z , aby byl eliminován vliv rozdílných měřítek. analýza byla provedena nad korelační maticí $\mathbf{R} = \frac{1}{n-1} \mathbf{Z}^T \mathbf{Z}$.

Hlavní komponenty byly odvozeny řešením vlastní úlohy

$$\mathbf{R} \mathbf{a}_k = \lambda_k \mathbf{a}_k,$$

kde λ_k jsou vlastní čísla a \mathbf{a}_k odpovídající vlastní vektory. Jednotlivé komponenty jsou navzájem ortogonální a seřazené podle velikosti vlastních čísel, přičemž rozptyl k -té hlavní komponenty je roven $\text{Var}(PC_k) = \lambda_k$. Podíl vysvětlené variability byl určen jako $\lambda_k / \sum_{j=1}^p \lambda_j$.

Interpretace komponent vycházela z **komponentních vah (loadings)**, které určují směr a sílu vztahu mezi původními proměnnými a jednotlivými komponentami. Komponentní váhy byly definovány jako $w_{jk} = a_{jk} \sqrt{\lambda_k}$, kde a_{jk} je prvek vlastního vektoru odpovídající k -té hlavní komponentě a λ_k je její vlastní číslo. Směr vztahu byl posuzován podle znaménka komponentních vah w_{jk} .

Pro posouzení významu proměnných byly dále využity **procentuální příspěvky proměnných (contribution)**, vyjadřující relativní podíl proměnné na tvorbě dané komponenty, definované vztahem $\frac{w_{jk}^2}{\sum_{j=1}^p w_{jk}^2} \cdot 100$.

Analýza tak umožnila identifikovat hlavní latentní dimenze dat a na jejich základě konstruovat další vhodné kompozitní proměnné.

2.6 KONSTRUKCE KOMPOZITNÍCH INDEXŮ

Kompozitní indexy byly konstruovány dvěma alternativními postupy v závislosti na povaze a struktuře vstupních proměnných.

Agregační přístup (standardizace + aritmetický průměr)

V první variantě byly všechny relevantní numerické ukazatele standardizovány pomocí z-skóre:

$$Z(X_{ij}) = \frac{X_{ij} - \mu_j}{\sigma_j},$$

kde X_{ij} je hodnota proměnné j u jednotky i , μ_j je průměr proměnné j , σ_j její směrodatná odchylka. Pokud byly proměnné sdruženy do tematických subbloků, byl nejprve vypočten jejich průměr: $S_{ik} = \frac{1}{m_k} \sum_{j \in k} Z(X_{ij})$, kde m_k je počet proměnných v bloku k . Celkový kompozitní index byl poté stanoven jako aritmetický průměr standardizovaných proměnných (resp. subbloků):

$$Index_i = \frac{1}{K} \sum_{k=1}^K S_{ik}$$

Pro zvýšení interpretovatelnosti byl index následně převeden na kategoriální proměnnou pomocí kvantilového dělení.

PCA přístup (první hlavní komponenta)

Ve druhé variantě byl index odvozen ze skóre první hlavní komponenty:

$$PC1_i = \sum_{j=1}^p a_j Z(X_{ij}),$$

kde a_j představují váhy (prvky vlastního vektoru) odpovídající největší vlastní hodnotě korelační matice. Pro lepší interpretovatelnost byl výsledný index následně transformován:

- **Min-max normalizací:** $Index_i^{norm} = \frac{PC1_i - \min(PC1)}{\max(PC1) - \min(PC1)}$
- nebo převodem na **percentilové pořadí**.

Oba přístupy vedou k vytvoření souhrnného ukazatele zachycujícího latentní dimenzi sledovaného jevu. Zatímco první metoda pracuje s rovnoměrnými vahami, PCA přístup určuje váhy datově na základě maximální vysvětlené variability.

2.7 KROKOVÝ VÝBĚR REGRESORŮ NA ZÁKLADĚ INFORMAČNÍHO KRITÉRIA

Výběr vysvětlujících proměnných byl proveden pomocí krokového algoritmu výběru modelu (zpětné a obousměrné selekce). Tento postup iterativně porovnává kandidátní modely lišící se přidáním nebo odebráním jednotlivých regresorů a vybírá model s nejnižší hodnotou **Akaikého informačního kritéria** (AIC):

$$AIC = -2\ell(\hat{\theta}) + k \cdot p,$$

kde $\ell(\hat{\theta})$ je logaritmus věrohodnostní funkce, p je počet odhadovaných parametrů modelu a k je penalizační konstanta ($k = 2$ odpovídá standardnímu Akaikého informačnímu kritériu). AIC klade relativně slabší penalizaci na složitost modelu, a proto má tendenci upřednostňovat komplexnější modely s lepší predikční schopností.

Zpětná selekce (backward) postupně odstraňuje statisticky méně přínosné regresory z plného modelu, zatímco **obousměrná selekce** (stepwise) kombinuje odečítání i zpětné přidávání proměnných a v každém kroku hledá model s nejlepší hodnotou AIC.

Multikolinearita mezi vysvětlujícími proměnnými byla posuzována pomocí **variačního inflačního faktoru** (VIF), který kvantifikuje míru nárůstu rozptylu odhadu regresního koeficientu vlivem lineární závislosti mezi regresory. Vysoké hodnoty VIF byly interpretovány jako indikace redundantní informace a vedly k redukci modelu. VIF je definován jako

$$VIF_j = \frac{1}{1 - R_j^2},$$

kde R_j^2 představuje koeficient determinace z pomocné regrese j -té vysvětlující proměnné na všechny ostatní regresory v modelu. Doplnkově byla sledována i **tolerance**:

$$\text{Tolerance}_j = 1 - R_j^2 = \frac{1}{VIF_j},$$

kteřá vyjadřuje podíl variability proměnné, jenž není vysvětlen ostatními regresory. Nízké hodnoty VIF a vysoké hodnoty tolerance indikují absenci problematické multikolinearity. Heuristicky hodnoty $VIF > 5$ jsou považovány za indikaci potenciální multikolinearity, zatímco $VIF > 10$ signalizují závažný problém.

U uspořádaných kategoriálních proměnných byly využity **kontrastní vektory**, které umožňují testování lineárních a vyšších-řákových trendů napříč kategoriemi. Tento přístup poskytuje jemnější vhled do struktury vztahu než standardní referenční kódování, avšak v případě statisticky nevýznamných vyšších-řákových složek byla preferována jednodušší numerická reprezentace proměnné.

Algoritmus byl aplikován na data zredukovaná na základě dřívějších dílčích analýz. Krokový výběr regresorů byl využit jako explorativní nástroj pro další zúžení množiny kandidátních proměnných před zařazením nezávislých proměnných vázaných na ORP.

2.8 TYPY SMĚRODATNÝCH CHYB A PŘEDPOKLADY MODELŮ

Obecně se v regresní analýze používají tyto typy směrodatných chyb:

- **Klasické** (OLS): Předpokládají homoskedasticitu a nezávislost chyb; nejefektivnější při splnění modelových předpokladů.

- **Heteroskedasticity-robustní (HC):** Nevyžadují konstantní rozptyl chyb; odolné vůči heteroskedasticitě.
- **Klastrované (cluster-robustní):** Umožňují závislost chyb uvnitř předem definovaných skupin (klastrů), zatímco předpokládají nezávislost mezi klastry. Jsou vhodné například pro data s hierarchickou strukturou (školy v regionech).
- **Autokorelované (HAC):** Robustní vůči heteroskedasticitě i časové závislosti.

Směrodatné chyby byly v této části odhadnuty heteroskedasticity-robustním odhadem typu **HC1**, který je vhodný pro velké vzorky a poskytuje stabilní korekci směrodatných chyb při možné heteroskedasticitě. U lineárního regresního modelu s heteroskedasticitně robustními směrodatnými chybami se část klasických předpokladů zmírňuje, ale ne všechny mizí. Ve finální modelaci s hierarchickou strukturou dat byly proto použity cluster-robustní směrodatné chyby typu **CR2**, které zohledňují shlukování pozorování v rámci vyšších územních jednotek a poskytují konzistentní inferenci i při porušení předpokladů o nezávislosti reziduí.

Předpoklady lineárního regresního modelu podle typu směrodatných chyb:

1. **Správná specifikace modelu:** Lineární vztah v parametrech (koeficienty), žádné systematicky vynechané proměnné.
2. **Exogenita regresorů** $\mathbb{E}(\varepsilon | \mathbf{X}) = 0$: Chyby nejsou korelované s vysvětlujícími proměnnými.
3. **Nezávislost pozorování:** Žádná autokorelace mezi jednotkami.
4. **Absence perfektní multikolinearity:** Regresory lineárně nezávislé.
5. **Dostatečně velký vzorek a asymptotika.**
6. **Homoskedasticita reziduí** $\text{Var}(\varepsilon_i | X_i) = \sigma^2$ pro všechna i : Chyby mají stejnou variabilitu bez ohledu na úroveň X .
7. **Normalita reziduí.**

Předpoklad	OLS (klasické SE)	HC (robustní SE)	Klastrované SE
Správná specifikace modelu	☑	☑	☑
Exogenita regresorů	☑	☑	☑
Nezávislost pozorování	☑	☑	☑ (mezi klastry)
Nekorelovanost proměnných	☑	☑	☑
Homoskedasticita	☑	✗	✗
Normalita reziduí	⚠ (malé vzorky)	✗	✗
Velký vzorek (asymptotika)	⚠	☑	☑ (dost klastrů)

Legenda: ☑ vyžadováno, ✗ nevyžadováno, ⚠ závisí na kontextu

2.9 OVĚŘENÍ HIERARCHICKÉ STRUKTURY DAT POMOCÍ NULOVÉHO VÍCEÚROVŇOVÉHO MODELU

Pro posouzení, zda jsou sledované školní ukazatele systematicky ovlivněny vyšší úrovní agregace (např. ORP nebo krajem), byl využit nulový víceúrovňový model s náhodným interceptem. Tento přístup umožňuje rozlišit variabilitu, která je dána rozdíly mezi jednotlivými školami, od variability spojené s příslušností ke konkrétnímu regionu.

Základní nulový víceúrovňový model lze zapsat ve tvaru:

$$y_{ij} = \beta_0 + u_j + \varepsilon_{ij},$$

kde y_{ij} představuje hodnotu sledovaného ukazatele pro školu i ve skupině j , β_0 je celkový průměr, u_j je **náhodný efekt skupiny** a ε_{ij} reziduální složka na úrovni školy. Náhodný efekt u_j a rezidua jsou modelovány jako náhodné proměnné s nulovou střední hodnotou a rozptyly σ_u^2 a σ_ε^2 .

V případě více hierarchických úrovní byl model rozšířen o více náhodných interceptů, například pro ORP a kraj:

$$y_i = \beta_0 + u_{\text{ORP}(i)} + v_{\text{kraj}(i)} + \varepsilon_i,$$

kde jednotlivé náhodné efekty reprezentují nezávislé zdroje variability na různých úrovních regionální struktury.

Pro kvantifikaci významu hierarchické struktury dat byla využita **intratřídní korelace (ICC)**, která vyjadřuje podíl celkové variability vysvětlovaný rozdíly mezi skupinami. V případě jedné úrovně seskupení je ICC definována jako

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}.$$

Při více úrovních hierarchie je celkový podíl variability připadající na skupinovou strukturu dán součtem rozptylů jednotlivých náhodných efektů vztaheným k celkové variabilitě:

$$\text{ICC}_{\text{hier}} = \frac{\sum_k \sigma_k^2}{\sum_k \sigma_k^2 + \sigma_\varepsilon^2}.$$

Vyšší hodnoty ICC indikují silnější seskupování pozorování a porušení předpokladu nezávislosti, což podporuje použití víceúrovňového modelování.

Statistická významnost náhodných efektů byla ověřena porovnáním nulového víceúrovňového modelu s odpovídajícím jednoúrovňovým nulovým modelem bez náhodných efektů. Toto porovnání je založeno na **testu poměru věrohodností** (Likelihood Ratio Test), který vychází z rozdílu logaritmických věrohodností obou modelů:

$$\chi^2 = -2(\ell_0 - \ell_1), \quad \chi^2 \sim \chi^2(1),$$

kde ℓ_0 označuje logaritmickou věrohodnost jednoúrovňového modelu a ℓ_1 logaritmickou věrohodnost víceúrovňového modelu. Nulová hypotéza testu má tvar $H_0: \sigma_u^2 = 0$, tj. že rozptyl náhodného efektu je nulový a zařazení náhodného interceptu nevede ke zlepšení shody modelu s daty. Test posuzuje, zda zařazení náhodných efektů vede ke statisticky významnému zlepšení shody modelu s daty.

Kombinace hodnot ICC a výsledků testu významnosti náhodných efektů poskytuje ucelené zdůvodnění pro použití víceúrovňového přístupu a umožňuje posoudit, na které úrovni regionální struktury se rozdíly mezi školami nejvýrazněji projevují.

2.10 POSOUZENÍ BIVARIÁTNÍCH VZTAHŮ A VHODNOSTI LINEÁRNÍHO MODELU

Před finální specifikací regresních modelů byly bivariátní vztahy mezi vybranými vysvětlujícími proměnnými a závislými proměnnými orientačně posouzeny **vizuálně pomocí bodových grafů (scatterplotů)**. Cílem tohoto kroku bylo ověřit předpoklad přibližné linearit vztahů a identifikovat případné nelinearity či vliv extrémních hodnot. Vizualizace byly doplněny o hladké trendové křivky, které usnadnily posouzení celkového tvaru vztahu. Na základě těchto zjištění byly u vybraných proměnných zváženy případné transformace, redukce pozorování nebo omezení jejich interpretace v rámci lineárního modelu.

2.11 REGRESNÍ MODELY SE SMÍŠENÝMI EFEKTY A HIERARCHICKOU STRUKTUROU DAT

Analýza byla provedena pomocí **lineárních regresních modelů se smíšenými efekty** (linear mixed-effects models), které umožňují současně modelovat vliv proměnných na úrovni jednotlivých škol (fixní efekty) a zohlednit hierarchickou strukturu dat prostřednictvím náhodných efektů na vyšší úrovni (např. ORP a kraje).

Obecný tvar modelu lze zapsat jako:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon},$$

kde \mathbf{y} je vektor pozorovaných hodnot závislé proměnné, \mathbf{X} je designová matice fixních efektů s parametry $\boldsymbol{\beta}$, \mathbf{Z} je designová matice náhodných efektů s vektorem náhodných složek \mathbf{u} , $\boldsymbol{\varepsilon}$ je vektor reziduí.

Předpokládá se:

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}), \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}),$$

kde \mathbf{G} a \mathbf{R} jsou kovarianční matice náhodných efektů a reziduí.

V analýze je uvažován zejména **náhodný intercept**, který umožňuje, aby se základní úroveň závislé proměnné lišila mezi jednotlivými územními jednotkami.

Odhad parametrů: ML vs. REML

Parametry modelu byly odhadovány pomocí metody **restricted maximum likelihood (REML)**, která poskytuje méně zatížené odhady rozptylových komponent než klasická maximální věrohodnost (ML), zejména u modelů s menším počtem skupin.

Metoda **ML** maximalizuje plnou věrohodnost $L(\boldsymbol{\beta}, \mathbf{G}, \mathbf{R} | \mathbf{y})$, zatímco **REML** maximalizuje věrohodnost transformovaných reziduí, čímž eliminuje vliv odhadu fixních efektů na odhad rozptylů.

V praxi byl postup následující:

- **REML** bylo použito pro finální odhady parametrů,
- **ML** bylo použito při porovnávání vnořených modelů s odlišnou specifikací fixních efektů.

Porovnání vnořených modelů

Modely byly odhadovány postupně. Nejprve byl pro každou závislou proměnnou specifikován nulový (prázdný) model obsahující pouze náhodný intercept na úrovni ORP. Tento model umožnil kvantifikovat podíl variability výsledků, který lze přičíst rozdílům mezi regiony, a ověřit oprávněnost použití víceúrovňového přístupu. Následně byly do modelů postupně přidávány proměnné na úrovni školy a na úrovni regionu.

Při postupném rozšiřování modelu o další regresory byly vnořené modely porovnávány pomocí **likelihood ratio testu (ANOVA)**:

$$\chi^2 = -2(\ell_{\text{jednodušší}} - \ell_{\text{rozšířený}}), \quad \chi^2 \sim \chi^2(k_{\text{rozšířený}} - k_{\text{jednodušší}}),$$

kde ℓ značí logaritmus věrohodnosti. Nulová hypotéza testu (H_0) předpokládá, že rozšířený model neposkytuje statisticky významně lepší shodu s daty než model jednodušší, tj. že přidané regresory nemají nulový společný efekt. Tento test umožňuje posoudit, zda přidání nové proměnné vede k statisticky významnému zlepšení modelu.

Současně byla sledována **informační kritéria**:

- **Akaikovo**: $AIC = -2\ell + 2k$, které preferuje modely s lepší predikční schopností;
- **Bayesovo**: $BIC = -2\ell + k \cdot \log(n)$, kde k je počet parametrů a n počet pozorování, které silněji penalizuje složitost modelu a preferuje úspornější specifikaci a v praxi doporučovanější více.

Interpretace regresních koeficientů

Nestandardizované koeficienty $\hat{\beta}_j$ vyjadřují očekávanou změnu závislé proměnné při jednotkové změně vysvětlující proměnné X_j , při konstantních hodnotách ostatních proměnných a náhodných efektů.

Pro srovnání velikosti efektů napříč proměnnými byly reportovány také **standardizované koeficienty**:

$$\beta_j^* = \hat{\beta}_j \frac{SD(X_j)}{SD(Y)},$$

kteřé vyjadřují změnu závislé proměnné v jednotkách směrodatné odchylky při změně vysvětlující proměnné o jednu směrodatnou odchylku. Standardizované efekty slouží primárně k **porovnání relativní síly vztahů**, nikoli k přímé věcné interpretaci v původních jednotkách. Standardizované koeficienty indikátorových proměnných vyjadřují změnu závislé proměnné při zvýšení indikátoru o jednu směrodatnou odchylku, nikoli při přechodu mezi kategoriemi 0-1 jako v případě nestandardizovaných koeficientů.

2.11.1.1 Vysvětlená variabilita: marginální a podmíněné R^2

K hodnocení vysvětlené variability byly použity koeficienty determinace pro smíšené modely:

- **marginální R^2** (R_m^2), vyjadřující podíl variability vysvětlený fixními efekty,
- **podmíněné R^2** (R_c^2), zahrnující fixní i náhodné efekty.

Rozdíl mezi R_m^2 a R_c^2 poskytuje informaci o tom, jakou roli v modelu hraje hierarchická struktura dat.

3 ANALÝZA DAT A VÝSLEDKY

Tato kapitola prezentuje **postup analýzy dat a hlavní výsledky víceúrovňové modelovací analýzy** zaměřené na charakteristiky základních a středních škol a jejich vzdělávací výsledky. Před odhadem víceúrovňových regresních modelů byla provedena explorační a diagnostická analýza proměnných zahrnující kontrolu kvality dat a distribucí, korelační analýzu prediktorů a posouzení multikolinearity. U tematicky příbuzných bloků ukazatelů byla zvážena redukce dimenze pomocí analýzy hlavních komponent za účelem konstrukce kompozitních indexů. Následoval krokový výběr regresorů a posouzení vhodnosti lineární modelové specifikace. Oprávněnost víceúrovňového přístupu byla ověřena nulovým modelem a výpočtem intratřídní korelace (ICC). Tím je vytvořen základ pro vlastní víceúrovňovou regresní analýzu se zohledněním regionálního kontextu.

Tabulkové a grafické výstupy jsou **zařazeny přímo do textu pouze tehdy, jsou-li nezbytné pro interpretaci výsledků**. Ostatní podpurné výstupy, včetně detailních tabulek a doplňkových grafů, jsou k dispozici v přílohách nebo samostatných excelových souborech **Kapitola 4 - Doplňkové výstupy (ZŠ).xlsx** a **Kapitola 4 - Doplňkové výstupy (SŠ).xlsx**, které slouží jako doplňující materiál k této kapitole.

3.1 EXPLORAČNÍ ANALÝZY DAT (EDA)

Listy ,EDA_ZŠ' a ,EDA_SŠ' v souborech Doplňkové výstupy představují souhrnné tabulkové výstupy explorační analýzy dat a poskytují základní přehled o struktuře, kvalitě a rozdělení jednotlivých proměnných ve vstupních datových souborech.

Z přiložených tabulkových výstupů je možné identifikovat proměnné s **vyšším podílem chybějících hodnot**. Jedná se především o **inverzní podílové ukazatele založené na tom**, že ve jmenovateli stojí menší z dvojice vstupních proměnných. Komplikace vznikají ve chvíli, kdy jmenovatel nabývá nulové hodnoty. Zatímco při nulové hodnotě čitatele je výsledný podíl roven nule, v případě nulového jmenovatele jde o limitní situaci, která je v datech nahrazena prázdnou hodnotou. U středních škol se tato problematika týká rovněž podílů vztahených k počtu žáků se SVP, přičemž 46 středních škol nevykazuje žádné takové žáky.

Obdobný aspekt nastává také u některých **podílů dívek**, které vyjadřují jejich procentuální zastoupení vůči celkovému počtu dětí v dané skupině. Pokud je však celkový počet dětí nulový (např. počet dětí ve školní družině na 2. stupni ZŠ, počet žáků opakujících ročník), dochází opět k nahrazení mezní hodnoty prázdnou. Všechny takto problematické proměnné byly z analýzy vyřazeny, aby nedocházelo k nežádoucí redukci vstupních dat.

Výjimkou, která zůstává v analýze zachována, je proměnná **podíl žáků ukončujících PŠD v 7. a 8. ročníku ZŠ ke všem žákům ukončujícím PŠD**. Ve sledovaném roce 2024 totiž u 32 ředitelství žádný žák PŠD neukončil, tudíž proměnná definovaná jako podíl vůči celkovému počtu ukončujících žáků zůstává prázdná. Její ponechání v analýze je však z věcného hlediska odůvodněné.

V celém datovém souboru základních škol se nacházejí pouze dvě **kategoriální proměnné**, a to označení zřizovatele a souhrnný ukazatel kombinující využívání projektové podpory (šablon) na zajištění podpurných profesí a personální zajištění podpurných profesí dle výkazů. Ve sledovaném výběru převažují úplné základní školy zřizované obcí, které využívají 2 a více šablon na podporu podpurných profesí. Střední školy jsou navíc klasifikovány podle typu (gymnázium, odborná škola, smíšená škola). V souboru středních škol převažují odborné školy (cca 62 %), po nichž následují gymnázia (cca 30 %). Detailní typologie dále rozlišuje čtyřletá a víceletá gymnázia, odborné školy s maturitou a učiliště. Pro účely modelování je však tato typologie

nahrazena dvěma logickými – proměnnými: indikátorem víceletého typu a indikátorem přítomnosti oborů s výučním listem. Tyto proměnné zachycují hlavní rozdíly mezi akademicky a profesně orientovanými vzdělávacími drahami a zároveň zjednodušují parametrizaci modelu.

Součástí obou datových souborů je rovněž **15 logických proměnných** typu ano/ne reprezentujících využívání šablon a dostupnost vybraných pozic. Z rozložení hodnot vyplývá, že na školách jsou relativně zřídka využívány šablony na podporu pozic cizojazyčného asistenta, kariérového poradce, školního psychologa a sociálního pedagoga.

Nejvyšší variabilitu vykazují proměnné týkající se zastoupení podpůrných profesí, podílu žáků se SVP a počtu speciálních tříd. Výraznou variabilitu dále vykazují inverzní podílové ukazatele a proměnné řádově vyššími hodnotami, jako jsou celkové počty žáků. V prostředí vzdělávací soustavy jde však o očekávaný jev.

Proměnné charakterizované nízkou variabilitou se vztahují k méně častým jevům, např. počtu nadaných žáků a speciálních tříd u základních škol. U základních a středních pak u podpůrných profesí, metodiků prevence, výchovných poradců či žáků opakujících ročník.

3.2 IDENTIFIKACE ODLEHLÝCH HODNOT

Listy ‚outliers_ZŠ‘ a ‚outliers_SŠ‘ v souborech Doplnkové výstupy představují **souhrnný přehled extrémních hodnot jednotlivých ředitelství** napříč sledovanými numerickými ukazateli.

Šest ředitelství s činností **základních škol** vykazuje extrémní hodnoty cca v jedné pětině sledovaných ukazatelů. Z další analýzy byla vyloučena všechna ředitelství, která vykazala více než 15 hodnot výrazně se odchylojících od běžného rozmezí (outlierů). Dále byla vyřazena tři ředitelství, která vykazala více než 9 odlehlých hodnot a zároveň u nich chybí přibližně 7 % údajů. Taková kombinace by mohla negativně ovlivnit kvalitu a informační hodnotu vstupních dat. Celkem bylo **vyloučeno 11 ředitelství s činností základních škol**, která jsou na odpovídajícím listu vyznačena červeně.

Analogický postup byl aplikován rovněž u **ředitelství s činností středních škol**. Z další analýzy byly vyloučeny všechna ředitelství s počtem odlehlých hodnot vyšším nebo rovným 15 (outlierů). Dále byla vyřazena čtyři ředitelství, která vykazala více než 9 odlehlých hodnot a u nichž se současně podíl chybějících údajů pohyboval v rozmezí 16–18 %. Tato kombinace by mohla opět významně snížit spolehlivost vstupních dat. Celkem tedy bylo **vyloučeno 11 ředitelství s činností středních škol**, která jsou na příslušném listu vyznačena červeně.

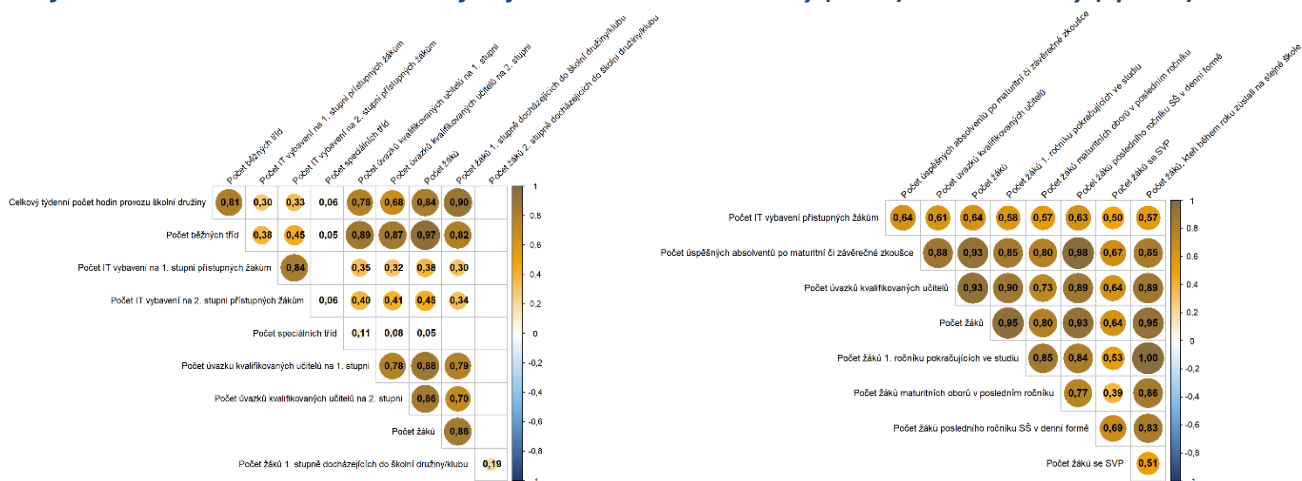
V následujících částech je již pracováno s datovým souborem upravenými v souladu s výše uvedeným postupem, tedy bez ředitelství, která byla z analýzy vyřazena.

3.3 KORELAČNÍ A ASOCIAČNÍ ANALÝZA NEZÁVISLÝCH PROMĚNNÝCH

Analýza lineárních závislostí

Listy ‚correl_matrix_ZŠ‘ a ‚correl_matrix_SŠ‘ obsahují **korelační matice numerických proměnných** vztahujících se k základním a středním školám. V jednotlivých buňkách jsou uvedeny Pearsonovy korelační koeficienty vyjadřující míru lineární závislosti mezi dvojicemi ukazatelů. V částech 3.1 a 3.2 byly identifikovány a ošetřeny odlehlé a extrémní hodnoty. Při ověřování vazeb mezi proměnnými proto nebylo nutné využívat robustnějších alternativních postupů se zvýšenou odolností vůči extrémním hodnotám.

Graf 1: Korelační matice ukazatelů definující velikost základní školy (vlevo) a střední školy (vpravo)



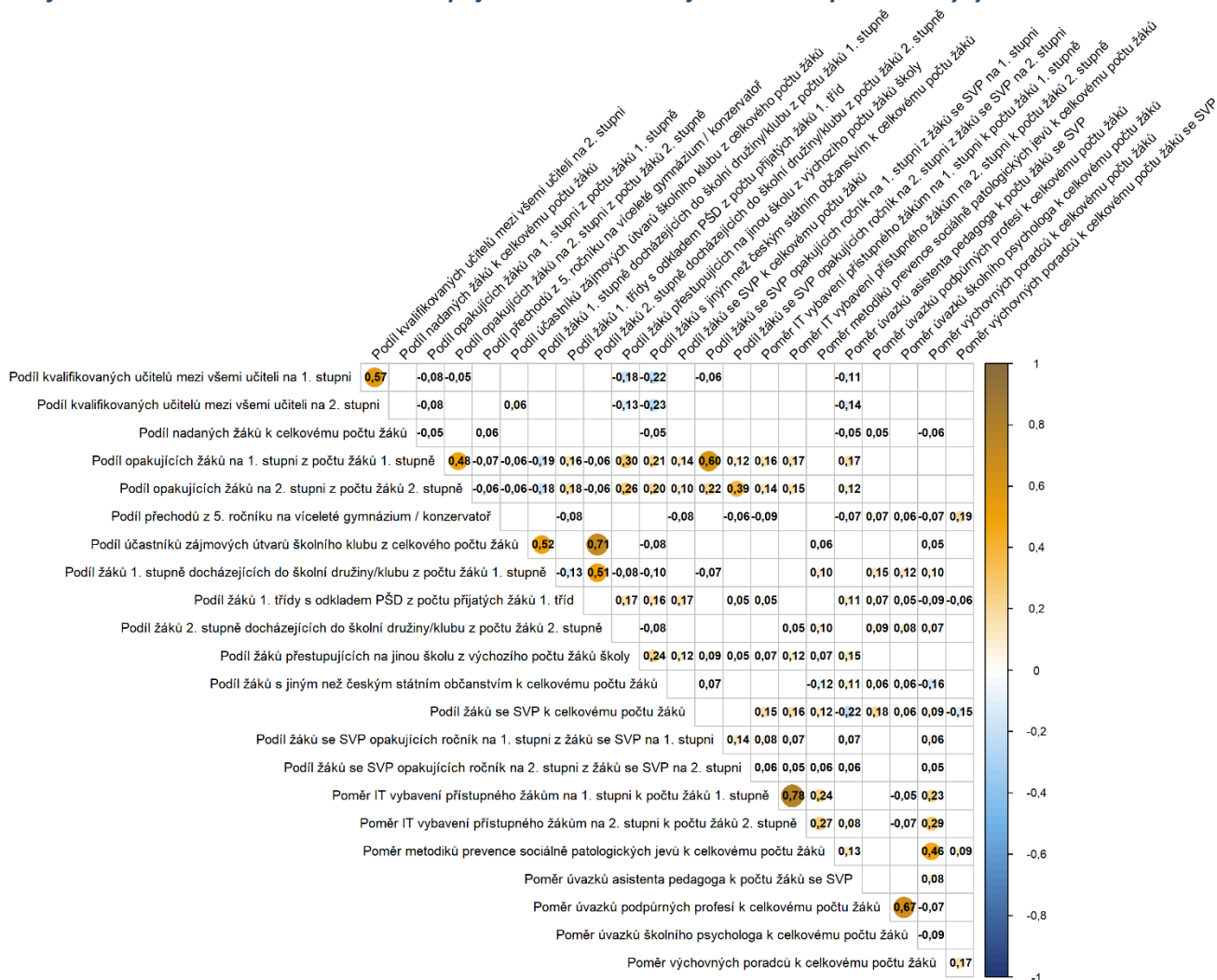
Zdroj: Výkaz R13, M03, Z02, M08

Obecně převládají **kladné korelace**. Nejsilnější vazby jsou dle Graf 1 patrné mezi charakteristikami týkajícími se velikosti školy (např. počty tříd a žáků, učitelů, IT vybavenosti atd.), což odráží jejich společný strukturální základ. Větší školy mají větší počet tříd, žáků, pedagogů, větší nároky na provoz školní družiny či na vybavenost školy. Na základě těchto zjištění byla **velikost školy** popsána pomocí souhrnného ukazatele, který kombinuje informace o standardizovaných počtech žáků, absolventů, pedagogů, tříd, IT vybavenosti, účastníků školní družiny či školního klubu a jejím provozu. Tento kompozitní index následně rozdělil školy do pěti velikostních kategorií (malá, menší, střední, větší, velká škola) na základě kvantilového dělení.

Silné korelace se objevují mezi podílovými ukazateli a jejich příslušnými absolutními hodnotami. Pro další analýzu se proto ukázalo vhodnější **pracovat s podílovými ukazateli**, které vykazují nižší korelace s ostatními proměnnými.

Slabší až střední korelace vykazují proměnné popisující specifické skupiny žáků (se SVP, nadaní, ve speciálních třídách), které nejsou tak úzce svázané s celkovou velikostí školy. Navíc jsou často tyto proměnné řídké a opět bude vhodné je vyřadit z dalších analýz.

Graf 2: Korelační matice ukazatelů vstupujících do hlavních zjištění deskriptivní analýzy ZŠ

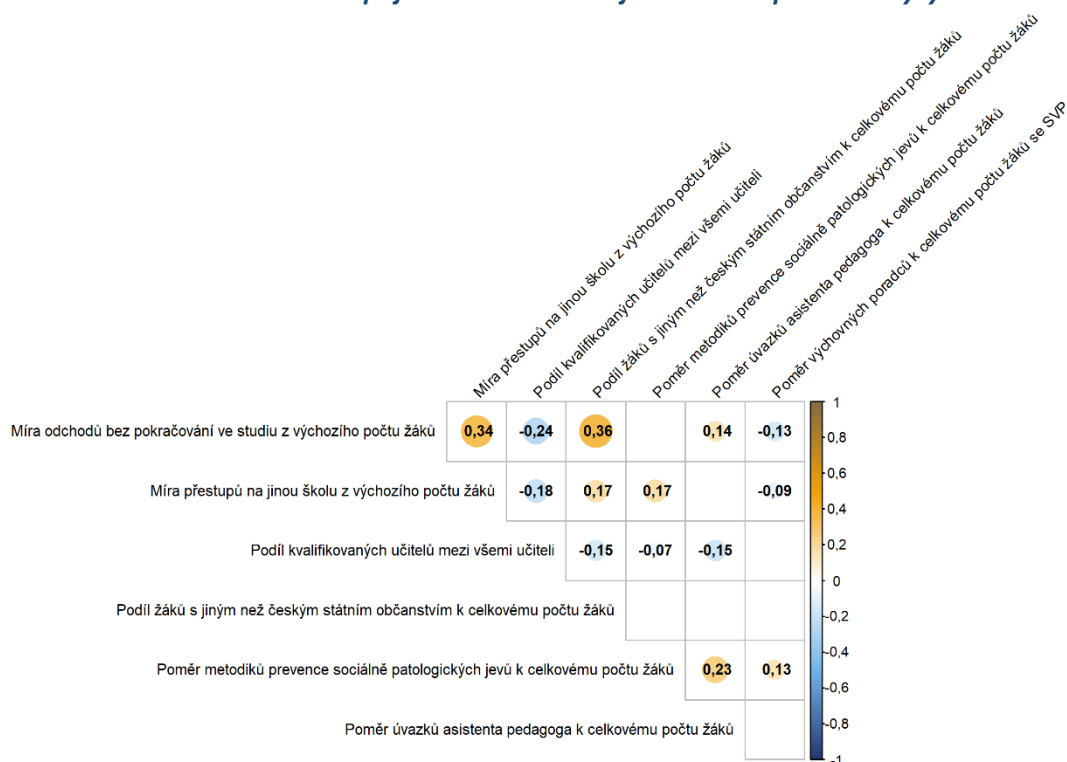


Zdroj: Výkaz R13, M03, Z02, školní matrika

Z proměnných týkajících se **základních škol** využitých k popisu zjištění v deskriptivní analýze je patrná v Graf 2 silná korelace mezi **ukazateli vztahujícími se k 1. a 2. stupni ZŠ**. Nejvýrazněji se tato vazba projevuje v oblasti dostupnosti IT vybavení, pedagogického zastoupení, účasti žáků ve školní družině či školním klubu a výskytu žáků se SVP opakujících ročník. Kvůli riziku multikolinearity je v dalších analýzách vhodné pracovat vždy pouze s jedním stupněm školy.

Obdobný problém vzniká i při souběžném zařazení **souhrnných ukazatelů** podpůrných profesí a zároveň jejich jednotlivých složek (např. školních psychologů, školních speciálních pedagogů). Dílčí profese totiž tvoří součást celkového součtu a jejich společné použití by vedlo k redundantní informaci.

Graf 3: Korelační matice ukazatelů vstupujících do hlavních zjištění deskriptivní analýzy SŠ



Zdroj: Výkaz R13, M08, školní matrika

Analogicky sestrojený Graf 3 ukazuje u **středních škol** pouze několik středně silných korelací, zejména mezi ukazateli fluktuace žáků (např. míra přechodů mezi školami a míra předčasného ukončení studia) a podílem žáků s jiným než českým státním občanstvím. V detailnějším korelačním grafu na listu „correl_matrix_SŠ“ se objevují výrazné, strukturální vazby mezi ukazateli o průchodu vzdělávacím systémem žáků. Nejzřetelnější jsou plná pozitivní vazba mezi pokračováním v 1. ročníku a přestupy na stejnou školu a plná negativní vazba mezi pokračováním v 1. ročníku a mírou předčasného ukončení studia. Tyto vztahy vyplývají z konstrukce ukazatelů a zcela jí odpovídají.

Do regresních modelů nebudou zahrnuty:

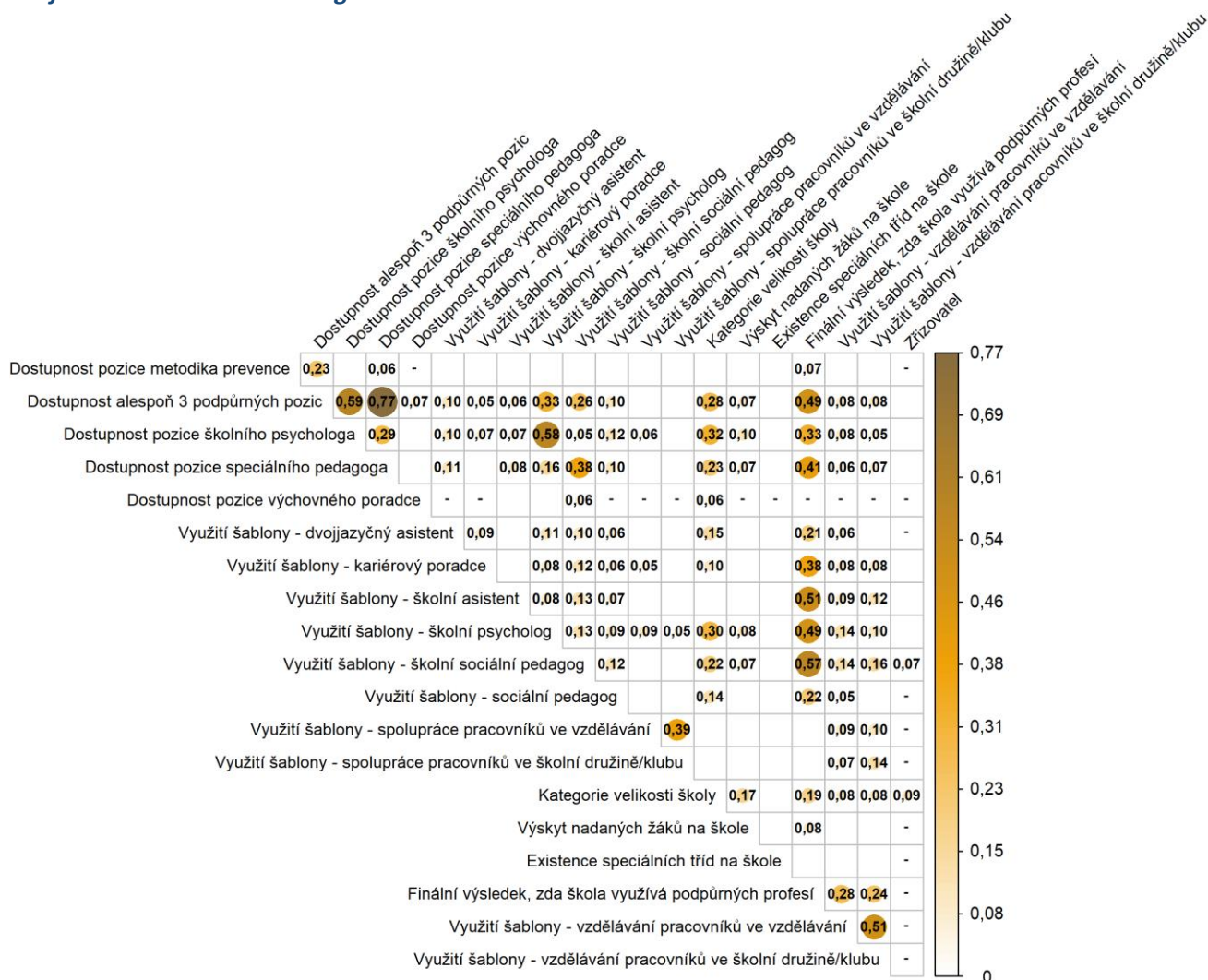
- číselné ukazatele úzce související s velikostí školy,
- absolutní počty vstupující do poměrových či podílových ukazatelů,
- proměnné vztahující se k méně početně zastoupenému stupni školy,
- ukazatele s nízkým počtem výskytů (např. nadaní žáci, speciální třídy, školní psychologové),
- proměnné s velmi nízkou variabilitou napříč vzorkem (např. metodici prevence, výchovní poradci),
- detailně členěné podíly dívek a žen, jejichž vyšší míra disagregace nepřináší dodatečnou informační hodnotu,
- jednotlivé položky tvořící souhrnné ukazatele, pokud je současně zahrnut jejich agregovaný ekvivalent,
- podílové ukazatele založené na různých jmenovatelích.

V případech velmi nízkého výskytu daného jevu byla místo původní kvantitativní proměnné zvažována náhrada logickou (binární) proměnnou zachycující samotnou přítomnost či nepřítomnost dané charakteristiky.

A u podílových ukazatelů založených na různých jmenovatelích byl v modelu ponechán vždy pouze jeden reprezentativní podíl.

Analýza kategoriálních závislostí

Graf 4: Asociační matice kategoriálních ukazatelů základních škol



Zdroj: Výkaz R13, M03, čerpání šablon

Grafy vztahů kategoriálních proměnných (Graf 4 pro ZŠ i Graf 10 pro SŠ v příloze) potvrzují, že prvky školní podpory žáků tvoří provázaný celek, nikoli izolované jevy. Vzhledem k tomu, že závěry jsou pro základní i střední školy v zásadě obdobné, je Graf 10 pro SŠ zařazen do přílohy materiálu. Specifika, která střední školy vykazují nad rámec ZŠ, jsou popsána přímo v textu níže. Školy s vyšší dostupností podpůrných profesí zároveň častěji využívají šablony. Tyto vazby poukazují na **riziko redundance a multikolinearity**, a proto je v navazujících modelech vhodné pracovat spíše se **souhrnnými indikátory** nebo volit **reprezentativní proměnné** namísto detailního rozkladu jednotlivých složek.

Zřetelná je opakovaně silná vazba mezi **souhrnným ukazatelem** dostupnosti **podpůrných profesí a jeho dílčími komponentami**, zejména dostupností školního psychologa a speciálního pedagoga (v případě střední školy i kariérového poradce). Výrazné asociace se rovněž objevují mezi **dostupností konkrétních pozic a využíváním odpovídajících šablon**, což naznačuje, že formální podpora a personální kapacity se na školách navzájem posilují. Opět je patrné, že **velikost školy** je výrazně provázána s dostupností podpůrných profesí (např. školní psycholog, speciální pedagog) a s mírou využívání šablon. U ZŠ se tento vztah projevuje rovněž s evidencí

nadaných žáků, v případě SŠ poté s existencí maturitních oborů, což značí, že tyto charakteristiky ve výrazné míře odrážejí institucionální kapacitu školy. Silné vazby jsou patrné také mezi indikátory podpory **pracovníků ve vzdělání a vychovatelů ve školní družině a školním klubu**. Výrazné jsou jak v oblasti vzdělávání, tak i v oblasti spolupráce.

U **středních škol** (viz Graf 10 v příloze) se nejsilnější vazby projevují mezi samotnými kategoriemi školy (mezi základní a detailní typologií dle vzdělávací dráhy) a mezi typem školy a přítomností oborů s výučním listem, nebo víceletých oborů. Tyto vztahy mají strukturální charakter. Z hlediska dalšího modelování je z tohoto důvodu vhodnější propojení **základní klasifikace školy** s logickými třídícími proměnnými zachycujícími existenci víceletých oborů či oborů s výučním listem.

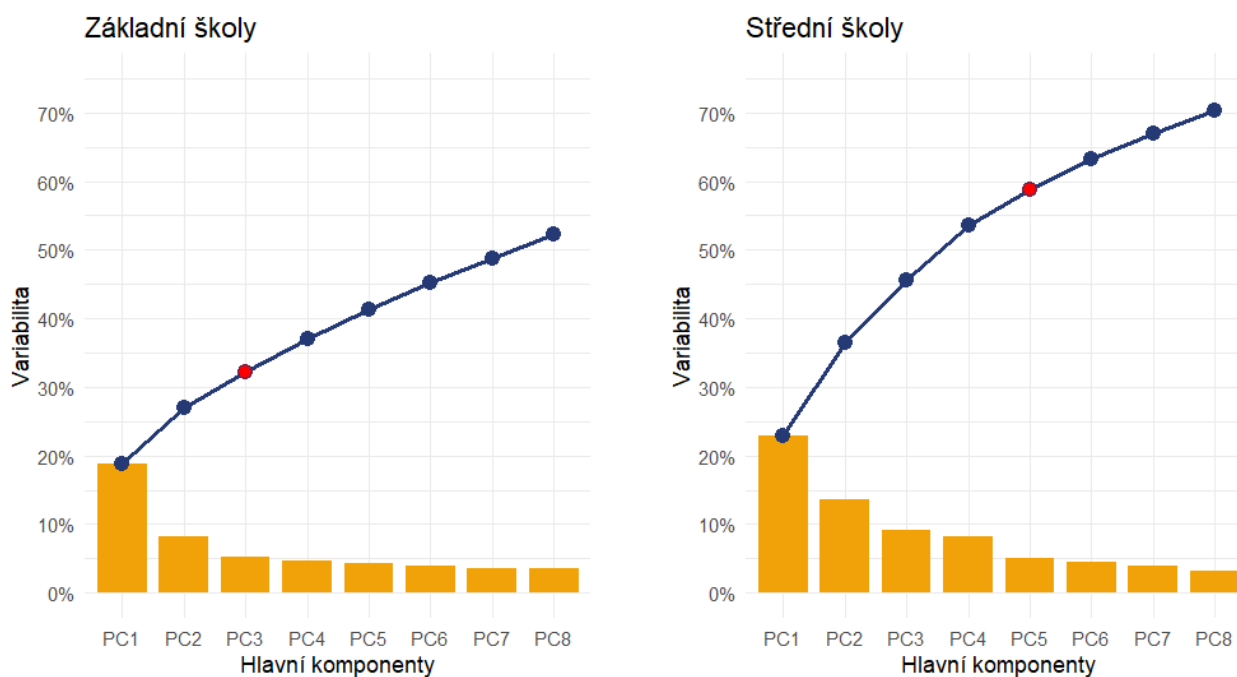
U vybraných proměnných (např. typ zřizovatele, dostupnost pozice výchovného poradce) **nebyly splněny podmínky dobré aproximace** chí-kvadrát testu, a to zejména z důvodu vysokého počtu kategorií či velmi řídkých kombinací hodnot. Pozice výchovného poradce chybí jen výjimečně a pouze zřídka se tato absence pojí s nedostatkem dalších podpůrných pozic či s nevyužíváním šablon pro zajištění financování pozic ve školách. U středních škol se navíc pouze výjimečně vyskytují speciální třídy nebo čerpání šablon pro zajištění financování pozice dvojjazyčného asistenta, což dále přispívá k řídkosti dat a omezuje použitelnost testu.

Celkově se tedy jako vhodné jeví **integrované souhrnné ukazatele**, které zachycují, zda škola využívá podpůrné profese a zda se zapojuje do programů podpory vzdělávání a spolupráce pracovníků, v nichž se personální dostupnost, využívání šablon a profesní rozvoj pracovníků vzájemně posilují.

3.4 NÁVRH KOMPOZITNÍCH PROMĚNNÝCH NA ZÁKLADĚ ANALÝZY HLAVNÍCH KOMPONENT

Analýza hlavních komponent (PCA) byla aplikována na celý datový soubor základních a středních škol. Cílem bylo identifikovat latentní dimenze a na jejich základě navrhnout nové **kompozitní proměnné**, které shrnují informaci obsaženou ve větším počtu vzájemně korelovaných ukazatelů.

Graf 5: Paretův diagram vysvětlené a kumulativní variance hlavních komponent (ZŠ vlevo, SŠ vpravo)



Prvních šest hlavních komponent zachycuje cca **45 % celkové variability dat základních škol**, zatímco prvních pět komponent u **středních škol vysvětluje 59 % variability** (viz Graf 5). Komponenty jsou řazeny dle podílu vysvětlené variance. Jejich stručná charakteristika je uvedena níže:

Latentní dimenze vzdělávací struktury základních škol:

1. **Velikost a organizační charakteristiky školy:** Komponenta shrnuje ukazatele velikosti školy a jejího provozního rozsahu, zejména počty žáků, tříd, učitelů, žáků se SVP, odkladů povinné školní docházky a účastníků školní družiny či školních klubů.
2. **Průchodnost vzdělávací dráhy a přechod na SŠ:** Komponenta zachycuje schopnost školy dovést žáky k reálné účasti na jednotné přijímací zkoušce na střední školu v kontrastu s kumulací opakování ročníků, podpůrných opatření a předčasných odchodů.
3. **Rozsah mimoškolních a volnočasových aktivit školy:** Komponenta popisuje míru zapojení žáků do školní družiny a zájmových aktivit na 1. i 2. stupni ZŠ základní školy, včetně žáků se SVP.
4. **Institucionální uspořádání speciálního vzdělávání:** Komponenta shrnuje rozsah speciálních tříd a podíl žáků vzdělávaných ve speciálních třídách jako strukturální charakteristiku školy.
5. **Selektivní přechody a poradenské kapacity školy:** Komponenta vyjadřuje míru selektivních přechodů na víceletá gymnázia v kombinaci s kapacitou školního poradenského zázemí.
6. **Sociální a organizační profil školy (gender a podpora):** Komponenta zachycuje vztah mezi genderovou strukturou žáků a mírou odborné podpory poskytované školou.

Latentní dimenze vzdělávací struktury středních škol:

1. **Velikost a průchodnost vzdělávací dráhy:** Komponenta spojuje velikost školy, počet absolventů a žákovské přechody s personálním a materiálním zázemím, čímž vystihuje celkovou kapacitu a stabilitu vzdělávací dráhy na SŠ.
2. **Průchodnost a stabilita vzdělávací dráhy:** Komponenta zachycuje napětí mezi pokračováním a úspěšným zakončením studia na jedné straně a jeho předčasným ukončováním v dané škole na straně druhé, tedy celkovou míru stability vzdělávací dráhy a dokončování vzdělávání na SŠ.
3. **Genderový profil školy:** Komponenta reprezentuje genderovou strukturu žáků a pedagogického sboru v návaznosti na podíl absolventek, dívek i učitelek, a vystihuje tak genderově podmíněný charakter vzdělávací dráhy na SŠ.
4. **Rozsah podpůrných profesí:** Komponenta vyjadřuje míru personálního zajištění podpůrných a poradenských pozic ve vztahu k počtu žáků, tedy institucionální kapacitu školy v oblasti odborné podpory.
5. **Speciální vzdělávání a podpůrné kapacity:** Komponenta vyjadřuje protiklad mezi strukturálně vymezeným speciálním vzděláváním (speciální třídy, asistenti pedagoga) a širší poradenskou a podpůrnou kapacitou školy (školní psychologové, ostatních podpůrných profesí, vyšší zastoupení žáků cizinců).

Výstupy PCA nevedly, s výjimkou kategoriální proměnné **velikosti školy**, ke konstrukci dalších kompozitních proměnných, jelikož interpretované dimenze jsou v analýze účelně reprezentovány konkrétními číselnými ukazateli, nebo by jejich výskyt byl příliš řídký. **Genderový profil školy** lze zcela nahradit jednou vybranou podílovou proměnnou dívek/žen. Pouze u středních škol je vhodné zvážit konstrukci kompozitní proměnné

„**Průchodnost a stabilita vzdělávací dráhy**“¹, která by sdružila více silně korelovaných ukazatelů zachycujících pokračování, dokončování a předčasné ukončování vzdělávání, a to i při jejich protisměrném působení.

3.5 KROKOVÝ VÝBĚR REGRESORŮ NA ZÁKLADĚ INFORMAČNÍHO KRITÉRIA

Krokový výběr vysvětlujících proměnných byl využit jako explorativní nástroj pro postupné **zúžení množiny kandidátních regresorů** a nalezení kompromisu mezi složitostí modelu a jeho vysvětlovací a predikční schopností. Tento postup byl aplikován opakovaně pro různé regresandy, přičemž dále jsou podrobně popsány výstupy pro všechny. Výběr proměnných byl proveden pomocí krokového algoritmu na základě Akaikeho informačního kritéria, a to ve variantě zpětné selekce (backward) i obousměrné selekce (stepwise).

Odhady parametrů byly ve všech modelech doplněny o heteroskedasticitně robustní směrodatné chyby (HC1) a interpretace výsledků je založena na bodových odhadech regresních koeficientů a jejich 95 % intervalech spolehlivosti.

Ve všech případech uvažovaných vysvětlovaných proměnných vedly obě selekční strategie ke shodné množině výsledných regresorů, což naznačuje stabilitu identifikované specifikace modelů. Výsledky jsou shrnuty v Tabulka 3 (ZŠ) a Tabulce 4 (SŠ), které uvádí bodové odhady regresních koeficientů, jejich 95% intervaly spolehlivosti, statistickou významnost parametrů a základní charakteristiky modelů.

3.5.1 Krokový výběr mezi základními školami

Výchozím bodem u základních škol byl plný lineární regresní model zahrnující přibližně 40 potenciálních vysvětlujících proměnných a v průměru 1 785 pozorování. Odhady parametrů výsledných modelů jsou uvedeny v Tabulka 3 v Příloze.

Model vysvětlující **průměrné hodnocení podmínek, průběhu a výsledků vzdělávání** dosahuje relativně nižší vysvětlovací schopnosti ($R^2 \approx 13\%$) a zahrnuje především proměnné související s institucionálním, poradenským a podpůrným zázemím školy. Naopak modely zaměřené na **úspěšnost žáků u jednotné přijímací zkoušky** vykazují podstatně vyšší vysvětlovací schopnost ($R^2 \approx 38\%$ a 55%) a akcentují strukturální charakteristiky školy, studijní trajektorie žáků a míru jejich zapojení do přijímacího řízení. Zajímavé jsou **rozdílné efekty kvalifikace učitelů** na 1. a 2. stupni ZŠ, které se v modelech úspěšnosti žáků promítají diferencovaně podle jejich typu a vazby na populaci žáků se SVP.

Zatímco model vztažený k **podílu úspěšných žáků ze zúčastněných** lépe zachycuje faktory související s průchodem žáků vzdělávací dráhou (podíly žáků opakujících ročník), model vztažený **ke všem žákům** je výrazně ovlivněn velikostí školy a strukturou její populační a podpůrné základny (podíl dívek, nadaných žáků, výchovných poradců, míra účasti na JPZ).

Výsledné modely sdílejí společné jádro strukturálních a personálních charakteristik školy, zejména ukazatele **velikosti školy, zřizovatele, IT vybavenosti, složení žakovské populace** (podíl žáků se SVP, podíl cizinců),

¹ **Index průchodnosti a stability vzdělávací dráhy** je lineární kombinací standardizovaných ukazatelů studijní kontinuity ve tvaru:

$$\text{Index průchodnosti a stability vzdělávací dráhy}_i = -0,469 \cdot Z(\text{Podíl odchodů bez navázání}_i) + 0,469 \cdot Z(\text{Podíl pokračujících po 1. ročníku}_i) + 0,416 \cdot Z(\text{Podíl maturantů}_i) - 0,354 \cdot Z(\text{Podíl opakujících}_i) - 0,263 \cdot Z(\text{Podíl přestupů na jinou školu}_i) + 0,333 \cdot Z(\text{Podíl zúčastněných maturity}_i) + 0,286 \cdot Z(\text{Podíl absolventů}_i).$$

Záporné váhy jsou přiřazeny indikátorům destabilizace (např. odchody bez návaznosti, opakování ročníku, přestupy na jinou školu), zatímco kladné váhy charakterizují plynulý průchod vzdělávací dráhou a úspěšné ukončení studia.

personálního zajištění (asistenti pedagoga, kvalifikovanost učitelů) a indikátory **studijní mobility** (přestupy, přechody na víceletá gymnázia, účast u JPZ).

Model průměrného hodnocení ZŠ v daných oblastech navíc akcentuje proměnné související se systémem podpory a poradenských služeb školy a využíváním systémové podpory. Naopak modely úspěšnosti u JPZ zdůrazňují strukturální charakteristiky školy **na 1. stupni ZŠ, podíly dívek** a indikátory zapojení žáků do mimoškolních aktivit či např. přechody žáků 5. tříd ZŠ na víceletá gymnázia a konzervatoře.

Kompletní seznam regresorů v členění na specifické a společné pro jednotlivé modely je uveden v Příloze v Seznamu 1.

3.5.2 Krokový výběr mezi středními školami

Výchozím bodem u **středních škol** byl plný lineární regresní model zahrnující 32 potenciálních vysvětlujících proměnných při zhruba 934–948 pozorování (v závislosti na konkrétní vysvětlované proměnné). Odhady parametrů výsledných modelů jsou uvedeny v Tabulce 4 v Příloze.

Model vysvětlující průměrné hodnocení podmínek, průběhu a výsledků vzdělávání na SŠ vykazuje spíše nižší vysvětlovací schopnost (R^2 v řádu nižších desítek procent) a akcentuje především institucionální charakteristiky školy (typ zřizovatel, velikost školy), personální zajištění a podpůrné mechanismy. Naopak modely zaměřené na podíl žáků úspěšně ukončujících vzdělávání či maturantů dosahují vyšší vysvětlovací schopnosti (R^2 až 68 %) a zdůrazňují kromě zřizovatele rovněž strukturální složení žákovské populace, indikátory stability vzdělávací dráhy (průchodnost, opakování ročníků, odchody bez návaznosti) a míru účasti žáků na maturitní zkoušce.

Specifikace jednotlivých modelů se liší podle konstrukce vysvětlované proměnné. Rozdíly vyplývají z toho, zda je úspěšnost vztahována ke všem absolventům, pouze k maturantům, příp. k celkové populaci žáků, nebo pouze k přihlášeným ke zkoušce. Modely vztažené k maturantům ukazují negativní efekt vyššího podílu dívek, a naopak pozitivní efekt vyššího podílu cizinců. Současně silněji zachycují vliv podpůrných mechanismů školy, zejména využívání šablon, dostupnosti podpůrných profesí, pozic výchovných poradců a metodiků prevence, účasti žáků u maturitní zkoušky, zastoupení žáků se SVP, a ukazatelů průchodnosti vzdělávacím systémem (podíl žáků opakujících ročník). Oproti tomu model vztažený ke všem absolventům více reflektuje širší strukturální charakteristiky školy, jako je typ zřizovatele, podíl kvalifikovaných učitelů, zastoupení asistentů pedagoga a školních psychologů či podíl maturantů volících si maturitní zkoušku z matematiky.

Výsledné modely jsou založeny na shodném okruhu strukturálních a personálních charakteristik střední školy, zejména ukazatele typ zřizovatele, IT vybavenosti, složení žákovské populace (podíl dívek v posledním ročníku, podíl cizinců), personálního zabezpečení (podíl kvalifikovaných učitelů) a indikátory stabilitu vzdělávací dráhy (míra odchodů bez pokračování ve studiu).

Kompletní přehled regresorů v členění na společné a specifické pro jednotlivé modely je uveden v Příloze (Seznam 2).

U modelů zaměřených na míry úspěšnosti maturantů je patrná nestandardně nízká hodnota interceptu. Tento výsledek je problematický zejména vzhledem k povaze regresandu, který nabývá hodnot na omezeném intervalu 0 až 100 %. Lineární model v tomto případě neomezuje predikované hodnoty na přípustný interval, tudíž může vést k interpretacím mimo realistický rozsah. To naznačuje, že klasická lineární specifikace zde není vhodná.

Z metodologického hlediska je vhodnější využít víceúrovňový zobecněný lineární model, např. s beta rodinou rozdělení (pro spojitý regresand na otevřeném intervalu 0–1) nebo s binomickou rodinou rozdělení. Cca jedna čtvrtina škol dosahuje hodnoty 1 (100% úspěšnost) v případě modelu úspěšnosti maturantů vztažené k přihlášeným. Vzhledem k tomu, že klasická beta regrese předpokládá otevřený interval (0, 1), není tato

specifikace adekvátní. Statisticky korektnějším řešením je proto binomický model, který pracuje s počty úspěchů a neúspěchů a respektuje hranice intervalu bez nutnosti transformací. Model míry úspěšnosti všech absolventů takto omezen není. Odhad interceptu se pohybuje v realistickém rozsahu a neindikuje systematickou extrapolaci mimo definiční obor regresandu. Z tohoto důvodu bude v navazující analýze dále pracováno s touto specifikací modelu.

3.5.3 Společné závěry

V průběhu selekce byla věnována pozornost také **multikolinearitě** vysvětlujících proměnných. Proměnné *Využití šablon – podpůrné profese a Finální výsledek, zda škola využívá podpůrných profesí* vykazovaly vysoké hodnoty variačního inflačního faktoru ($VIF > 7$), tzn. indikující silnou multikolinearitu mezi těmito proměnnými. Z důvodu zachování interpretovatelnosti modelu byla ponechána pouze proměnná *Finální výsledek, zda škola využívá podpůrných profesí*, která lépe vystihuje sledovaný mechanismus podpory. U modelů středních škol byla zvýšená hodnota variačního inflačního faktoru ($VIF > 5$) identifikována rovněž u *Indexu průchodnosti a stability vzdělávací dráhy*. Tento ukazatel vykazoval výraznou lineární vazbu zejména na proměnnou *Míra odchodů bez navázání dalšího studia*, což je vzhledem k jejich obsahové provázanosti očekávatelné. Kvůli této kolinearitě byl při finální specifikaci modelu zvolen pouze jeden z těchto strukturálně blízkých indikátorů, aby nedocházelo ke zkreslení odhadů parametrů.

Proměnná *Velikost školy* byla původně modelována jako uspořádaná kategoriální proměnná. V modelové diagnostice se ukázalo, že vyšší – řádové kontrastní složky (zejména kvadratický trend) nejsou statisticky významné. Z tohoto důvodu byla proměnná **převedená na ordinální numerickou škálu** (integer), čímž došlo ke zjednodušení modelu bez ztráty vysvětlující schopnosti. Současně byla škála proměnné rozšířena z původních tří na pět kategorií. Toto členění má lepší statistické vlastnosti, zejména stabilnější odhady parametrů a vyšší rozlišovací schopnost, a umožňuje zjednodušit model bez ztráty vysvětlující schopnosti.

Pro zvýšení přehlednosti a stability modelu budou ve finálním modelování některé kategoriální proměnné zestručněny. Zřizovatel školy bude reprezentován binární proměnnou indikující, zda je škola soukromá (zřizovatel privátní sektor) / veřejná (zřizovatel obec nebo kraj), státní (zřizovatel obec nebo kraj) / nestátní (zřizovatel privátní sektor nebo církev). Proměnnou *Finální výsledek, zda škola využívá podpůrných profesí* bude vhodné zredukovat ze čtyř na dvě kategorie podle toho, zda škola využívá / nevyužívá možnost zajištění financování podpůrných profesí na školách z tzv. šablon. Personální zastoupení podpůrných profesí je zohledněno numerickým ukazatelem poměru podpůrných profesí.

3.6 OVĚŘENÍ HIERARCHICKÉ STRUKTURY DAT (NULOVÝ MODEL, ICC)

Záhlaví Tabulka 1 a Tabulka 2 shrnují základní charakteristiky nulových víceúrovňových modelů pro jednotlivé vysvětlované proměnné a úrovně hierarchického členění dat. Uvádí název vysvětlované proměnné, úroveň seskupení (ORP, kraj nebo jejich kombinaci), počet pozorování a skupin, hodnotu intratřídní korelace (ICC v %). Sloupec statistické významnosti náhodných efektů udává výsledek testu významnosti náhodného interceptu, založeného na porovnání nulového modelu s náhodnými efekty a odpovídajícího jednoúrovňového modelu bez náhodných efektů pomocí testu poměru věrohodností (LRT).

Hodnota **intratřídní korelace** (ICC) vyjadřuje podíl celkové variability vysvětlované proměnné, který je připisován rozdílům mezi skupinami (např. ORP nebo kraji) v nulovém modelu, tedy před zahrnutím vysvětlujících proměnných. Udává tedy, jak silně jsou pozorování ve stejné skupině vzájemně podobná a nakolik je vhodné zohlednit hierarchickou strukturu dat pomocí víceúrovňového modelu.

Tabulka 1: Statistik ověřování významnosti náhodných efektů na úrovni ORP a krajů v rámci ZŠ

Vysvětlovaná proměnná	Skupiny	Počet pozorování	Počet skupin	ICC (v %)	Statistická významnost náhodných efektů
Průměrná úroveň hodnocení v daných oblastech, získaná z inspekční činnosti	ORP	1 782	227	8,83 %	***
	kraj	1 782	14	12,65 %	***
	ORP + kraj	1 782	227 a 14	14,16 %	***
Podíl úspěšnějších žáků u JPZ vzhledem k účastněným	ORP	1 788	227	32,24 %	***
	kraj	1 788	14	22,98 %	***
	ORP + kraj	1 788	227 a 14	32,67 %	***
Podíl úspěšnějších žáků u JPZ vzhledem ke všem žákům	ORP	1 787	227	33,05 %	***
	kraj	1 787	14	24,21 %	***
	ORP + kraj	1 787	227 a 14	33,48 %	***

Statistická významnost je vyznačena hvězdičkami: * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

Výsledky nulových víceúrovňových modelů u základních škol jednoznačně potvrzují statistickou významnost náhodných efektů jak na úrovni ORP, tak na úrovni krajů u všech vysvětlovaných proměnných. Nezanedbatelné hodnoty intratřídní korelace (ICC), zejména u ukazatelů úspěšnosti žáků u JPZ, ukazují, že významná část variability výsledků je systematicky vázána na regionální kontext. Zařazení hierarchické struktury dat do regresních modelů je proto metodicky opodstatněné a nezbytné pro korektní odhad parametrů a jejich směrodatných chyb. Zároveň se ukazuje, že úspěšnost žáků u JPZ se výrazněji liší mezi jednotlivými ORP, tedy na lokální úrovni, zatímco u hodnocení školy ČŠI se silněji uplatňují rozdíly mezi kraji, tedy širší regionální prostředí.

Tabulka 2: Statistik ověřování významnosti náhodných efektů na úrovni ORP a krajů v rámci SŠ

Vysvětlovaná proměnná	Skupiny	Počet pozorování	Počet skupin	ICC (v %)	Statistická významnost náhodných efektů
Průměrná úroveň hodnocení v daných oblastech, získaná z inspekční činnosti	ORP	936	214	16,03 %	***
	kraj	936	14	18,71 %	***
	ORP + kraj	936	214 a 14	18,71 %	***
Míra úspěšného ukončení (z žáků posledního ročníku)	ORP	948	213	0,00 %	
	kraj	948	14	0,00 %	
	ORP + kraj	948	213 a 14	0,00 %	
	ORP	934	214	1,17 %	

Míra úspěšnosti maturantů (z žáků posledního ročníku)	kraj	934	14	1,57 %	**
	ORP + kraj	934	214 a 14	1,57 %	**
Míra úspěšnosti maturantů (z přihlášených)	ORP	934	214	1,88 %	
	kraj	934	14	1,56 %	**
	ORP + kraj	934	214 a 14	1,87 %	**

Statistická významnost je vyznačena hvězdičkami: * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

Výsledky nulových víceúrovňových modelů pro střední školy ukazují odlišný obraz než v případě základních škol. U průměrné úrovně hodnocení je patrná významná regionální struktura, přičemž ICC dosahuje cca 16–19 % a náhodné efekty na úrovni ORP i krajů jsou statisticky významné. Hodnocení středních škol je tedy výrazně ovlivněno regionálním kontextem.

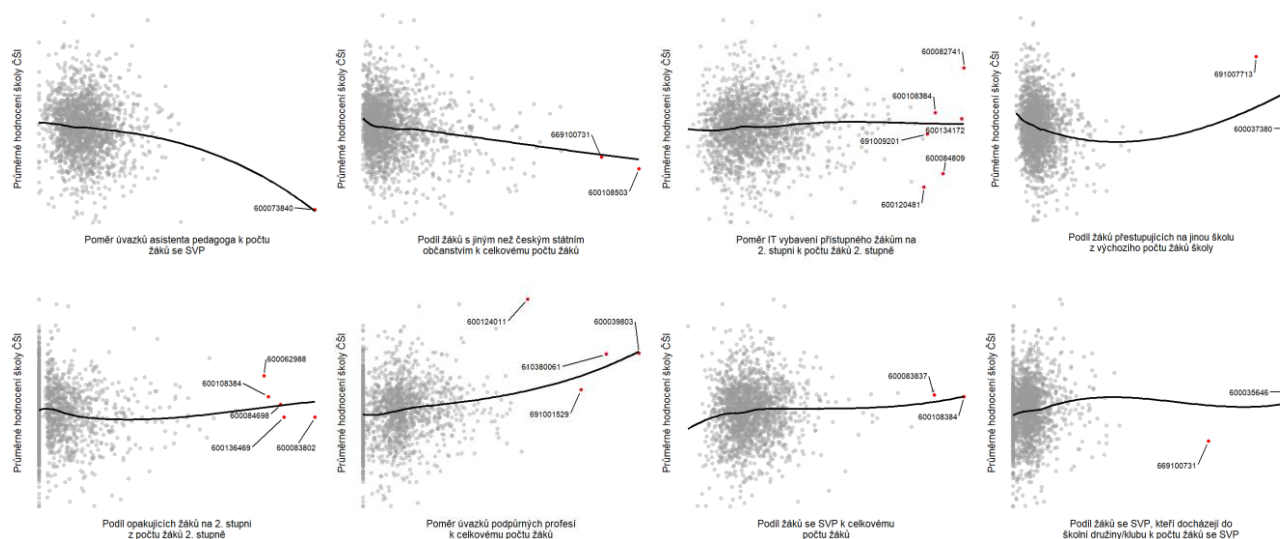
Naopak u ukazatelů úspěšného ukončení studia všemi absolventy a úspěšnosti maturantů jsou hodnoty ICC velmi nízké (v řádu jednotek procent, resp. nulové). To naznačuje, že většina variability těchto výsledků je vysvětlována rozdíly mezi jednotlivými školami, nikoli systematickými rozdíly mezi regiony. Statisticky významnější slabý regionální efekt se projevuje pouze na úrovni krajů u ukazatelů úspěšnosti u maturitní zkoušky.

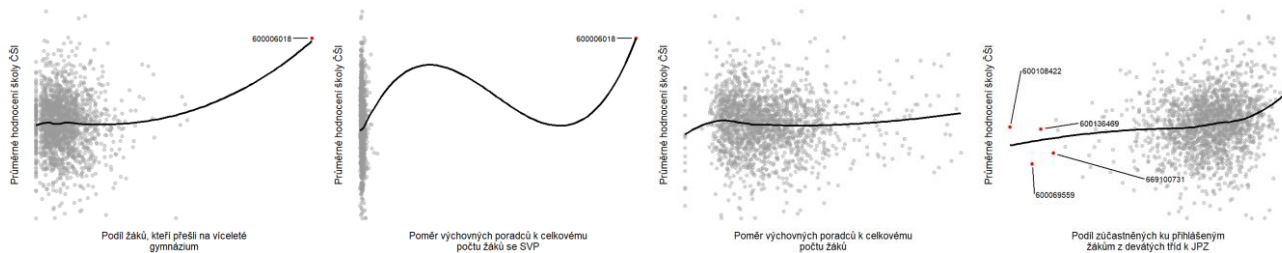
Hierarchická struktura dat je tak u středních škol zásadní zejména pro modelování hodnocení školy, zatímco u indikátorů vzdělávací úspěšnosti je regionální složka variability podstatně slabší.

3.7 POSOUZENÍ BIVARIÁTNÍCH VZTAHŮ A VHODNOSTI LINEÁRNÍHO MODELU

Grafy níže slouží k orientačnímu vizuálnímu posouzení tvaru vztahů mezi vybranými vysvětlujícími proměnnými a závislou proměnnou.

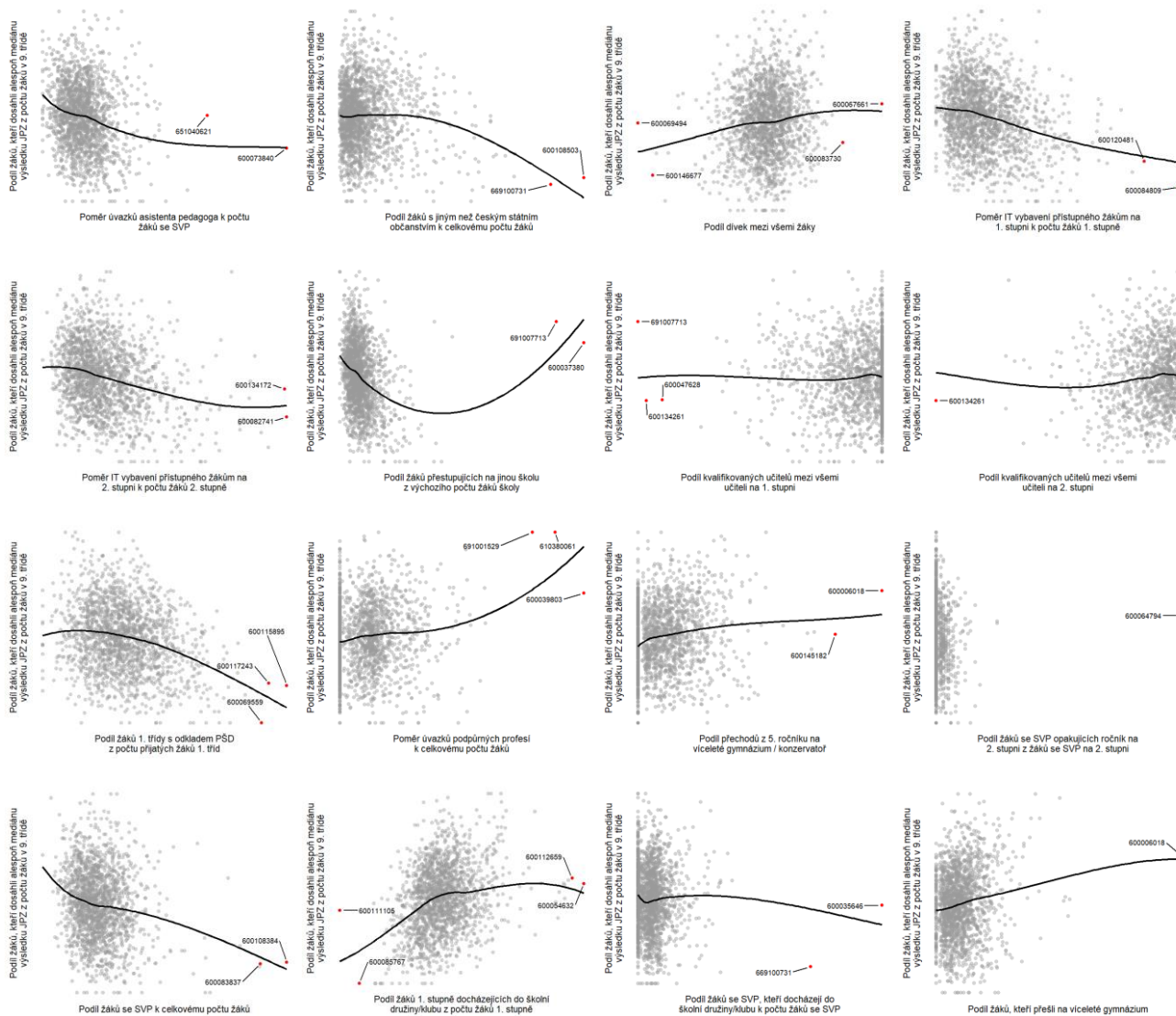
Graf 6: Bivariátní vztahy mezi významnými charakteristikami ZŠ a průměrného hodnocení podmínek, průběhu a výsledků vzdělávání

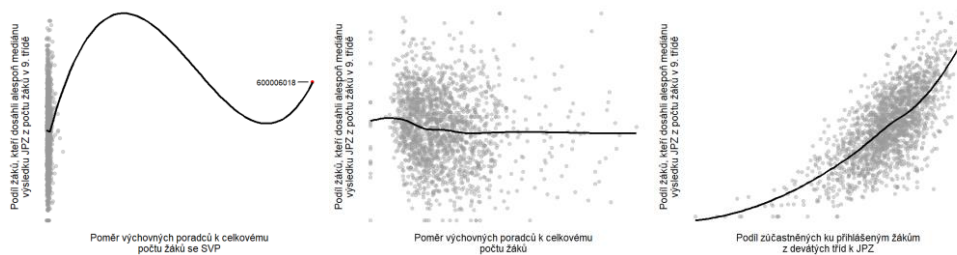




V případě **průměrného hodnocení podmínek, průběhu a výsledků vzdělávání** základních škol v Graf 6 je ve většině případů patrný spíše slabý až střední monotónní vztah, často cca lineární v centrální části dat, s možnými nelinearitami nebo změnami sklonu v extrémech rozdělení. To obecně podporuje použití lineárního modelu, avšak upozorňuje na potenciální citlivost odhadů na krajní hodnoty některých ukazatelů.

Graf 7: Bivariátní vztahy mezi významnými charakteristikami ZŠ a podílem úspěšnějších žáků u JPZ vzhledem ke všem žákům



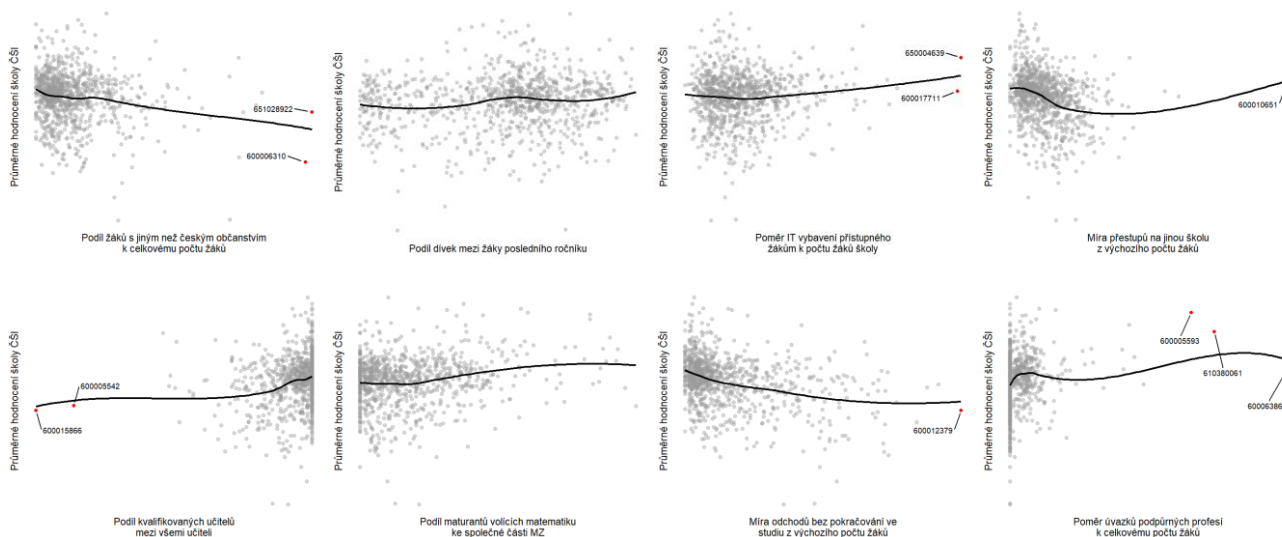


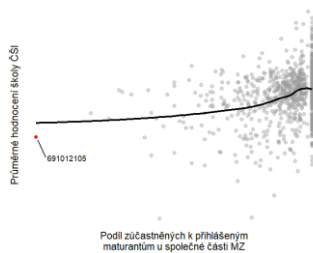
Vztahy mezi jednotlivými významnými regresory a **podílem nadmediánových žáků u JPZ** ve srovnání s hodnocením školy v daných oblastech častěji vykazují systematický trend. Např. u míry účasti přihlášených žáků na JPZ se projevuje výrazná monotónní závislost, ale i u jiných indikátorů selekce a vzdělávací dráhy.

U hodnocení školy v daných oblastech a výsledků JPZ se v některých případech ukazuje **citlivost na extrémní pozorování**, která mají tendenci „přitahovat“ vyhlazenou křivku. V datech se objevují školy s velmi extrémními hodnotami některých vysvětlujících proměnných, jež mohou výrazně ovlivňovat tvar LOESS křivky i odhad regresních koeficientů. Tato pozorování nejsou nutně chybná, avšak reprezentují specifické případy, které je vhodné v následné analýze buď explicitně ošetřit (např. pomocí robustních metod), z analýzy zcela vyčlenit, anebo je alespoň interpretovat se zvýšenou opatrností. Z analýzy základních škol bylo vyřazeno 7 ředitelství, která výrazně ovlivňovala tvar LOESS křivek.

U řady podílových ukazatelů je rovněž patrná **výrazná kumulace hodnot v bodě nula**, což je typické pro indikátory založené na řídkých jevech nebo malých absolutních počtech a může to dále ovlivňovat tvar pozorovaných vztahů.

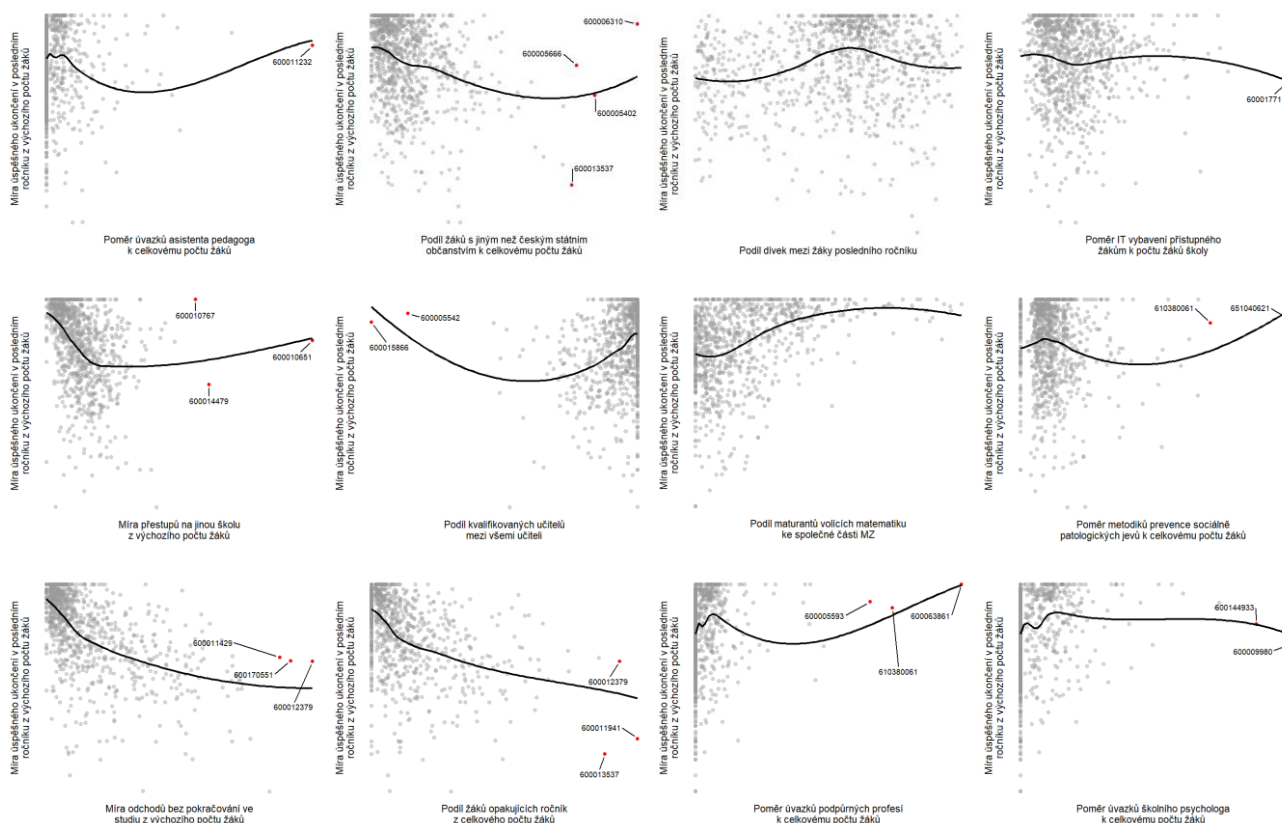
Graf 8: Bivariátní vztahy mezi významnými charakteristikami SŠ a průměrného hodnocení podmínek, průběhu a výsledků vzdělávání

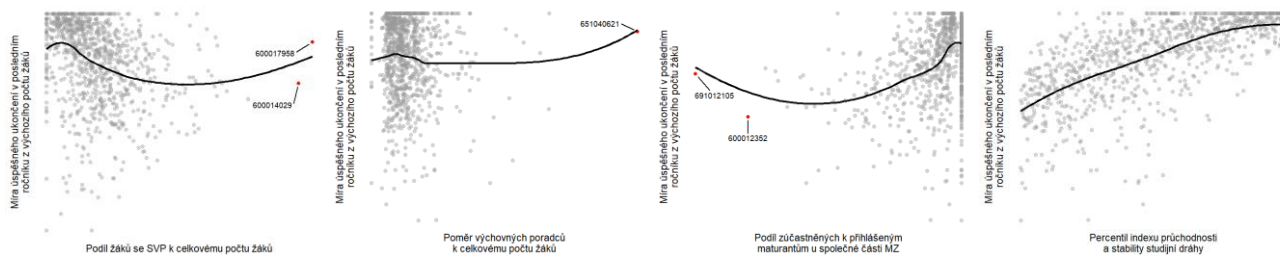




U **středních škol** se ve vztahu k **průměrnému hodnocení podmínek, průběhu a výsledků vzdělávání**, obdobně jako u základních škol, projevuje převážně slabý až středně silný monotónní trend. Ve středu rozdělení jsou vztahy převážně lineární, zatímco v krajních hodnotách se projevují známky nelinearity nebo zvýšené variability. To podporuje použití lineárního modelu a zároveň však ukazuje možnou citlivost odhadů na školy s extrémními hodnotami, zejména u ukazatelů souvisejících s personálním zabezpečením. Ve srovnání s výsledkovými ukazateli zde silněji vystupuje regionální faktor (viz vyšší ICC u hodnocení SŠ). Tento fakt může naznačovat, že průměrné hodnocení školy v daných oblastech, získaná z inspekční činnosti je do určité míry ovlivněno širším institucionálním prostředím kraje či ORP.

Graf 9: Bivariátní vztahy mezi významnými charakteristikami SŠ a míry úspěšného zakončení studia absolventy





Vztahy mezi jednotlivými regresory a **mírou úspěšného zakončení studia všemi absolventy** jsou obecně slabší a méně konzistentní než u výsledků JPZ na úrovni ZŠ. U většiny ukazatelů je patrný pouze mírný monotónní trend, často doprovázený výraznějším rozptylem hodnot. To odpovídá velmi nízkým hodnotám ICC u těchto proměnných. Variabilita úspěšnosti je tedy spíše dána rozdíly mezi jednotlivými školami než regionálním kontextem.

Podobně jako u ZŠ se i zde objevují extrémní hodnoty některých podílových ukazatelů, které mohou ovlivňovat tvar vyhlazených křivek. U indikátorů založených na relativně malých skupinách žáků nebo méně zastoupených školských profesí se navíc kumulují hodnoty v blízkosti nuly či jedné, což zvyšuje nestabilitu odhadů v krajních částech rozdělení. Při interpretaci regresních modelů je proto nutné tyto jevy zohlednit a v případě potřeby je ošetřit v následné analýze. Z analýzy středních škol bylo vyřazeno také 7 ředitelství, která výrazně ovlivňovala tvar LOESS křivek.

Interpretaci dále komplikuje skutečnost, že přibližně **12 % pozorování dosahuje plné úspěšnosti** (tj. maximální hodnoty 1). Tato kumulace na horní hranici škály částečně narušuje linearitu vztahu ve směru závislé proměnné a může ovlivnit tvar LOESS křivek i odhady regresních koeficientů. Výsledky je proto vhodné interpretovat s ohledem na tuto distribuci závislé proměnné a případně zvážit práci s modely vhodnými pro intervalově omezené proměnné.

3.8 VÍCEÚROVŇOVÁ REGRESNÍ ANALÝZA SE ZOHLEDNĚNÍM REGIONÁLNÍ STRUKTURY

Vícerozměrná analýza byla provedena pomocí dvouúrovňových lineárních regresních modelů, v nichž první úroveň (Level 1) tvoří jednotlivé školy a druhou úroveň (Level 2) správní obvody obcí s rozšířenou působností (ORP včetně krajské úrovně). Zvolená specifikace reflektuje hierarchickou strukturu dat, kdy školy působí v odlišných regionálních kontextech a jejich pozorování nelze považovat za statisticky nezávislá.

Obecnou podobu obou uvažovaných modelů lze vyjádřit následovně:

$$y_{ij} = \beta_0 + \beta_1 x_{1,ij} + \beta_2 x_{2,ij} + \dots + \beta_p x_{p,ij} + u_j + \varepsilon_{ij},$$

kde y_{ij} představuje hodnotu závislé proměnné pro i -tou školu v j -tém regionu (průměrnou úroveň hodnocení školy v daných oblastech, získaná z inspekční činnosti, resp. podíl žáků s nadmediánovým výsledkem u JPZ), β_0 je globální intercept, $x_{k,ij}$ jsou vysvětlující proměnné na úrovni školy a regionu, β_k odpovídající koeficienty fixních efektů, u_j náhodný intercept zachycující systematické rozdíly mezi regiony a ε_{ij} reziduální složka. Předpokládá se, že $u_j \sim \mathcal{N}(0, \sigma_u^2)$ a $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. U našich modelů s cluster-robustními (CR2) směrodatnými chybami není vyžadován tento silný předpoklad normality ani homoskedasticity reziduí; inferenční závěry jsou robustní vůči jejich porušení za předpokladu nezávislosti mezi klastry.

Proměnné na úrovni školy byly do modelů zahrnuty jako fixní efekty, přičemž jejich výběr vycházel z tematických okruhů vymezených v podkapitole 3.5 a z dostupnosti kvalitních dat. Cílem nebylo konstruovat maximálně saturovaný model, ale identifikovat klíčové faktory, které jsou statisticky významně spojeny se sledovanými vzdělávacími výsledky a dávají smysl.

Také proměnné na regionální úrovni (ORP) byly zahrnuty jako fixní efekty, zatímco náhodný intercept umožnil zachytit systematické rozdíly mezi regiony, jež nejsou plně vysvětleny zahrnutými socioekonomickými ukazateli. Tento přístup přispívá ke snížení rizika zkreslení odhadů v důsledku opomenutých regionálně podmíněných faktorů.

U čtvrtého modelu, který analyzuje míru úspěšnosti dokončení studia na střední škole, odhad meziregionální složky variance nebyl statisticky významný (viz Tabulka 2 v Odstavci 3.6). Hierarchická struktura tedy v tomto případě nepřinesla dodatečnou vysvětlující informaci nad rámec zahrnutých fixních efektů. Z tohoto důvodu byl finální model odhadnut jako klasický lineární regresní model, přičemž pro zajištění robustnosti inferenčních závěrů byly použity heteroskedasticity-robustní (HC1) směrodatné chyby. Model tak nabývá tvaru klasického lineárního regresního modelu:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_p x_{p,i} + \varepsilon_i,$$

kde $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Tento model předpokládá nezávislost pozorování a neobsahuje složku u_j zachycující regionální heterogenitu.

Podílové ukazatele jsou v modelech vyjádřeny procentuálně, což usnadňuje interpretaci výsledků. Regresní koeficienty tedy vyjadřují změnu závislé proměnné při změně příslušného prediktoru o jeden procentní bod. Pro zachování přehlednosti a srozumitelnosti nebyly aplikovány žádné transformace proměnných, nelineární specifikace ani interakční efekty.

Tabulka 5, Tabulka 7, příp. listy „regression_ZŠ“ a „regression_SŠ“ shrnují výsledky čtyř regresních modelů základních a středních škol. Tabulka je členěna do dvou sloupcových bloků odpovídajících jednotlivým modelům, zatímco řádky představují zahrnuté vysvětlující proměnné. Každý řádek obsahuje odhad regresního koeficientu doplněný 95 % intervalem spolehlivosti a označením statistické významnosti na pětiprocentní hladině významnosti. Spodní část tabulky poskytuje souhrnné charakteristiky modelů, včetně informací o vysvětlené variabilitě a podílu variance připadající na náhodné efekty u víceúrovňových modelů.

Standardizované efekty jsou v tabulce barevně zvýrazněny, přičemž intenzita zbarvení odpovídá jejich relativní velikosti. Toto vizuální odlišení umožňuje rychlou orientaci v tabulce a usnadňuje porovnání síly a směru jednotlivých vztahů napříč oběma modely, nezávisle na měrných jednotkách použitých proměnných.

3.8.1 Diagnostika

Diagnostika multikolinearity v Tabulka 6 (základní školy) a Tabulka 8 (střední školy) neodhalila problematické vztahy mezi vysvětlujícími proměnnými (všechny hodnoty VIF < 3,2).

Diagnostické grafy (Graf 11 pro průměrné hodnocení ZŠ v daných oblastech, získaná z inspekční činnosti, Graf 12 pro výsledky u JPZ, Graf 13 pro průměrné hodnocení SŠ v daných oblastech, získaná z inspekční činnosti) nenaznačují závažné porušení předpokladů smíšeného modelu. Rozdělení náhodných efektů je přibližně symetrické a odpovídá předpokladu normality. Rezidua nevykazují výrazné systematické odchylky ani dominantní vliv jednotlivých pozorování. Model zároveň dobře reprodukuje celkové rozdělení pozorovaných hodnot, což potvrzuje adekvátní shodu mezi daty a modelovými predikcemi. Vzhledem k použití cluster-robustních (CR2) směrodatných chyb nejsou drobné odchylky od homoskedasticity či normality reziduí považovány za problematické. U smíšených modelů jsou lineární vztahy posuzovány podmíněně na náhodných

efektech, takže drobná zakřivení v reziduích nejsou neočekávaná a v tomto případě grafy nenaznačují žádný systematický vzorec, který by zpochybňoval adekvátnost modelu.

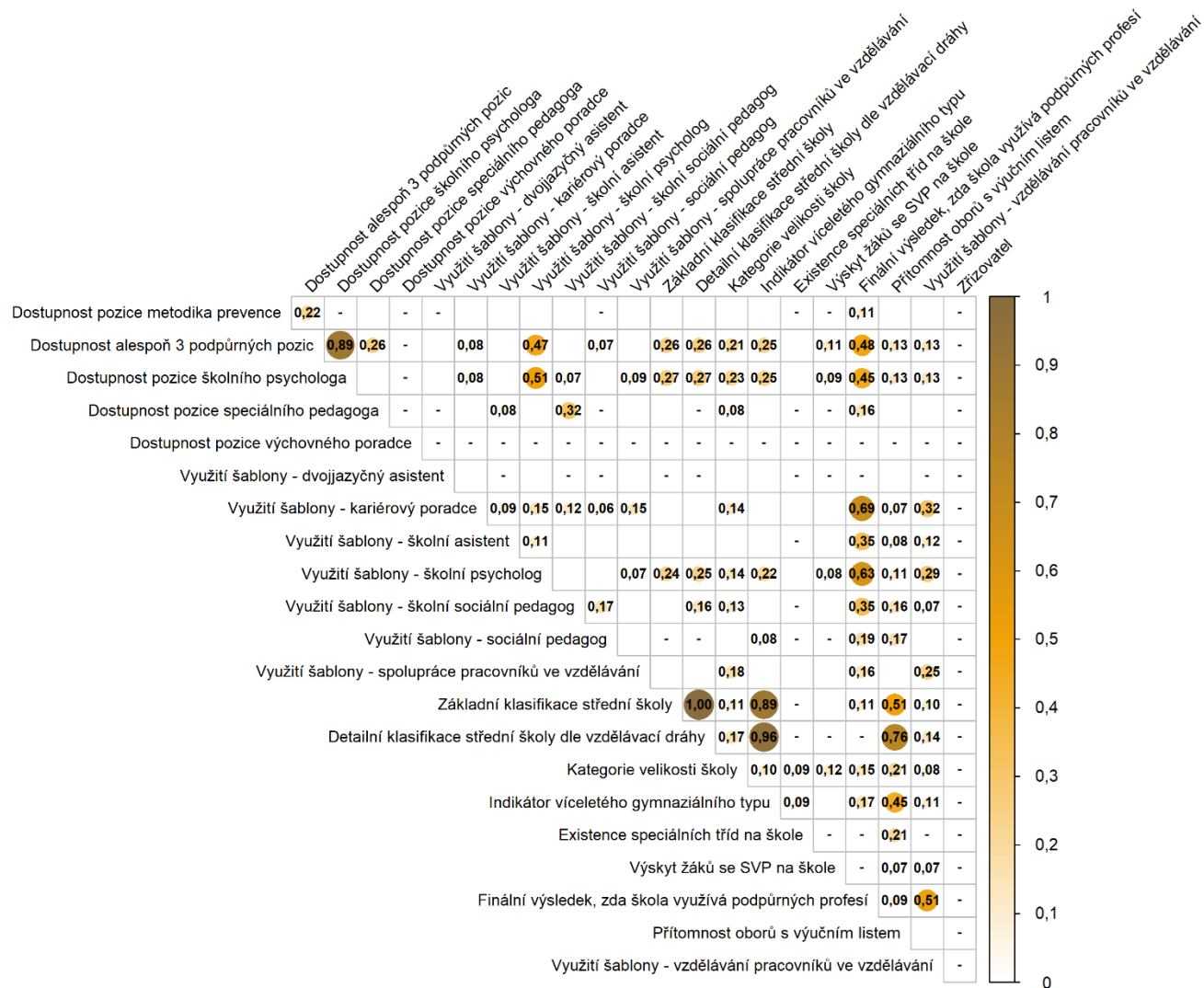
U modelu podílu úspěšného ukončení studia (Graf 14 pro úspěšnost dokončení středního vzdělání), odhadovaného pomocí klasického lineárního modelu s heteroskedasticity-robustními (HC1) směrodatnými chybami, diagnostika neodhaluje zásadní problémy. Je patrný mírný náznak heteroskedasticity a lehké odchylky od normality reziduí, ty však nejsou vzhledem k použití robustní inferenční procedury považovány za metodologicky problematické. Nebyly identifikovány dominantní vlivné body ani problematická multikolinearita.

Posterior predictive kontrola ukazuje, že lineární model méně přesně reprodukuje pravý okraj distribuce, což souvisí s omezeným rozsahem podílové proměnné (0–100) a vyšším zastoupením plné úspěšnosti. Tato odchylka však není zásadní pro účel modelu, kterým je identifikace vztahů mezi proměnnými a odhad jejich asociací. Interpretace regresních koeficientů tím není významně dotčena, zejména vzhledem k použití heteroskedasticity-robustních (HC1) směrodatných chyb.

4 PŘÍLOHA

4.1 VÍCEÚROVŇOVÁ ANALÝZA: FAKTORY SOUVISEJÍCÍ S NEROVNOSTMI VE VZDĚLÁVACÍ SOUSTAVĚ

Graf 10: Asociační matice kategoriálních ukazatelů středních škol



Zdroj: Výkaz R13, M08, čerpání šablon

Tabulka 3: Výsledky lineárního regresního modelu vysvětlujícího průměrné hodnocení podmínek, průběhu a výsledků vzdělávání na ZŠ a podíl úspěšnějších žáků u JPZ

	Model dle vysvětlované proměnné		
	Průměrná úroveň hodnocení v daných oblastech, získaná z inspekční činnosti	Podíl úspěšnějších žáků u JPZ vzhledem k zúčastněným	Podíl úspěšnějších žáků u JPZ vzhledem ke všem žákům
Regresory	Odhad koeficientu [95% interval spolehlivosti] Statistická významnost		
(Konstanta)	2,60 [2,45; 2,76] ***	19,46 [9,66; 29,27] ***	-21,50 [-34,37; -8,63] ***
Církevní zřizovatel	-0,01 [-0,10; 0,07]	10,05 [5,15; 14,94] ***	12,16 [6,87; 17,45] ***
Kategorie velikosti školy	0,03 [0,02; 0,05] ***	1,01 [0,16; 1,85] **	0,83 [-0,04; 1,69] *
Podíl dívek mezi všemi žáky			0,14 [-0,05; 0,33] .
Podíl kvalifikovaných učitelů mezi všemi učiteli na 1. stupni		-0,11 [-0,18; -0,04] ***	-0,10 [-0,16; -0,04] ***
Podíl kvalifikovaných učitelů mezi všemi učiteli na 2. stupni	0,00 [0,00; 0,00] **	0,11 [0,04; 0,19] ***	0,10 [0,03; 0,16] ***
Podíl opakujících žáků na 1. stupni z počtu žáků 1. stupně	-0,01 [-0,03; 0,00] *	-0,97 [-1,89; -0,04] **	
Podíl opakujících žáků na 2. stupni z počtu žáků 2. stupně	0,01 [-0,00; 0,02]	-0,91 [-1,69; -0,13] **	
Podíl přechodů z 5. ročníku na víceleté gymnázium / konzervatoř			-0,10 [-0,23; 0,04] .
Podíl zúčastněných ku přihlášeným žákům z devátých tříd k JPZ	0,00 [0,00; 0,00] ***	0,46 [0,40; 0,53] ***	0,81 [0,76; 0,87] ***
Podíl žáků 1. stupně docházejících do školní družiny/klubu z počtu žáků 1. stupně		0,11 [0,07; 0,16] ***	0,10 [0,06; 0,14] ***
Podíl žáků 1. třídy s odkladem PŠD z počtu přijatých žáků 1. tříd	-0,00 [-0,00; 0,00] *	-0,14 [-0,21; -0,06] ***	-0,10 [-0,17; -0,04] ***
Podíl žáků přestupujících na jinou školu z výchozího počtu žáků školy	-0,01 [-0,01; -0,00] **	-0,70 [-1,17; -0,23] ***	-0,52 [-0,92; -0,13] **
Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků	0,00 [-0,00; -0,00] ***	0,11 [0,02; 0,20] **	0,11 [0,03; 0,18] ***
Podíl žáků se SVP k celkovému počtu žáků	0,00 [0,00; 0,00] ***	-0,20 [-0,32; -0,07] ***	-0,18 [-0,29; -0,06] ***
Podíl žáků se SVP opakujících ročník na 2. stupni z žáků se SVP na 2. stupni		0,20 [0,03; 0,37] **	0,17 [0,06; 0,29] ***
Podíl žáků se SVP, kteří docházejí do školní družiny/klubu k počtu žáků se SVP	0,00 [0,00; 0,00] ***		-0,03 [-0,05; -0,00] **
Podíl žáků, kteří přešli na víceleté gymnázium	-0,01 [-0,02; 0,00] *	0,83 [0,25; 1,40] ***	1,45 [0,65; 2,24] ***
Poměr IT vybavení přístupného žákům na 1. stupni k počtu žáků 1. stupně			-0,03 [-0,07; 0,00] *
Poměr IT vybavení přístupného žákům na 2. stupni k počtu žáků 2. stupně	0,00 [0,00; 0,00] ***	-0,08 [-0,10; -0,06] ***	-0,04 [-0,08; -0,01] **
Poměr metodiků prevence sociálně patologických jevů k celkovému počtu žáků	0,08 [0,02; 0,14] ***		
Poměr úvazků asistenta pedagoga k počtu žáků se SVP	-0,00 [-0,00; 0,00] .	-0,22 [-0,32; -0,11] ***	-0,17 [-0,26; -0,07] ***
Poměr úvazků podpůrných profesí k celkovému počtu žáků	0,10 [0,04; 0,17] ***	6,04 [2,37; 9,70] ***	6,32 [2,87; 9,78] ***
Poměr výchovných poradců k celkovému počtu žáků	-0,06 [-0,13; 0,01] *		2,76 [-1,01; 6,52] .

Poměr výchovných poradců k celkovému počtu žáků se SVP	0,00 [0,00; 0,00] ***		-0,07 [-0,10; -0,04] ***
Rozsah využití podpůrných šablon na rozvoj a spolupráci pracovníků ve vzdělávání	0,02 [0,00; 0,03] ***		
Soukromý zřizovatel	0,26 [0,16; 0,35] ***	11,42 [4,37; 18,47] ***	10,14 [2,22; 18,06] **
Škola nevyužívá podporu a dle výkazu nemá 3 podpůrné pozice	-0,01 [-0,04; 0,02]		
Škola nevyužívá podporu, ale dle výkazu má alespoň 3 podpůrné pozice	-0,05 [-0,10; -0,01] **		
Škola využívá podporu (1 šablona)	0,00 [-0,02; 0,03]		
Výskyt nadaných žáků na škole	0,03 [-0,00; 0,06] *		1,50 [-0,12; 3,12] *
Počet pozorování	1 780	1 785	1 786
Počet regresorů	24	18	23
Koeficient determinace R^2	13,10 %	38,30 %	54,50 %
Typ směrodatných chyb (SE)	HC1	HC1	HC1

Statistická významnost je vyznačena hvězdičkami: . $p < 0,16$; * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

Seznam 1: Přehled regresorů vybraných krokovou selekcí (ZŠ)

Regresory společné pro všechny modely

- Poměr úvazků asistenta pedagoga k počtu žáků se SVP
- Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků
- Poměr IT vybavení přístupného žákům na 2. stupni k počtu žáků 2. stupně
- Podíl žáků přestupujících na jinou školu z výchozího počtu žáků školy
- Podíl kvalifikovaných učitelů mezi všemi učiteli na 2. stupni
- Podíl žáků 1. třídy s odkladem PŠD z počtu přijatých žáků 1. tříd
- Poměr úvazků podpůrných profesí k celkovému počtu žáků
- Podíl žáků se SVP k celkovému počtu žáků
- Podíl žáků, kteří přešli na víceleté gymnázium
- Podíl zúčastněných ku přihlášeným žákům z devátých tříd k JPZ
- Kategorie velikosti školy
- Zřizovatel

Regresory specifické pro model průměrného hodnocení ZŠ v klíčových oblastech, získaná z inspekční činnosti (neobjevují se v modelech JPZ)

- Poměr metodiků prevence sociálně patologických jevů
- Rozsah využití podpůrných šablon na vzdělávání a spolupráci
- Kategorie využívání podpory (šablony / podpůrné pozice)

Regresory specifické pro modely úspěšnosti u JPZ (neobjevují se v modelu hodnocení ZŠ v klíčových oblastech, získaná z inspekční činnosti)

- Podíl dívek mezi všemi žáky
- Poměr IT vybavení přístupného žákům na 1. stupni k počtu žáků 1. stupně
- Podíl kvalifikovaných učitelů mezi všemi učiteli na 1. stupni
- Podíl přechodů z 5. ročníku na víceleté gymnázium / konzervatoř
- Podíl žáků se SVP opakujících ročník na 2. stupni z žáků se SVP na 2. stupni

Tabulka 4: Výsledky lineárního regresního modelu vysvětlujícího průměrné hodnocení podmínek, průběhu a výsledků vzdělávání a podíl úspěšného zakončení studia SŠ

Regresory	Regresandy			
	Průměrná úroveň hodnocení školy v klíčových oblastech, získaná z inspekční činnosti	Míra úspěšného ukončení (z žáků posledního ročníku)	Míra úspěšnosti maturantů (z žáků posledního ročníku)	Míra úspěšnosti maturantů (z přihlášených)
	Odhad koeficientu [95% interval spolehlivosti] Statistická významnost			
(Konstanta)	1,845 [1,387; 2,302] ***	81,69 [73,95; 89,43] ***	-21,90 [-42,98; -0,81] **	-16,14 [-34,35; 2,07] *
Církevní zřizovatel	0,032 [-0,072; 0,136]	-3,65 [-6,94; -0,36] **		-2,19 [-4,62; 0,23] *
Jiný ústřední orgán státní správy jako zřizovatel		-4,54 [-9,85; 0,78] *		-2,11 [-4,93; 0,72] .
Kategorie velikosti školy	0,024 [0,013; 0,035] ***			
Krajský zřizovatel	-0,006 [-0,095; 0,083]	-4,13 [-7,02; -1,23] ***		-3,33 [-5,03; -1,63] ***
Míra odchodů bez pokračování ve studiu z výchozího počtu žáků	-0,010 [-0,016; -0,005] ***	-0,36 [-0,59; -0,12] ***	-0,39 [-0,67; -0,12] ***	-0,37 [-0,60; -0,15] ***
Míra přestupů na jinou školu z výchozího počtu žáků	-0,009 [-0,017; -0,001] **		0,30 [0,07; 0,53] **	
Percentil indexu průchodnosti a stability vzdělávací dráhy		0,15 [0,12; 0,17] ***	0,12 [0,08; 0,15] ***	0,08 [0,06; 0,10] ***
Podíl dívek mezi žáky posledního ročníku	0,000 [-0,000; 0,001]	0,01 [-0,02; 0,03]	-0,04 [-0,06; -0,02] ***	-0,04 [-0,05; -0,02] ***
Podíl kvalifikovaných učitelů mezi všemi učiteli	0,002 [0,000; 0,004] **	0,05 [-0,03; 0,12]		0,04 [-0,00; 0,08] *
Podíl maturantů volících matematiku ke společné části	0,001 [-0,000; 0,002] *	0,06 [0,04; 0,08] ***	0,02 [-0,00; 0,04] .	
Podíl zúčastněných k přihlášeným maturantům u společné části MZ	0,008 [0,004; 0,013] ***		1,13 [0,91; 1,35] ***	1,12 [0,92; 1,31] ***
Podíl žáků opakujících ročník z celkového počtu žáků			-0,25 [-0,74; 0,25]	0,40 [0,02; 0,79] **
Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků	-0,004 [-0,010; 0,001] .	-0,32 [-0,80; 0,15]	0,17 [0,03; 0,31] **	0,17 [0,06; 0,27] ***
Podíl žáků se SVP k celkovému počtu žáků			-0,09 [-0,19; 0,01] *	-0,14 [-0,23; -0,05] ***
Poměr IT vybavení přístupného žákům k počtu žáků školy	0,001 [-0,000; 0,001] *	0,02 [-0,01; 0,06] .	0,02 [-0,00; 0,04] *	0,03 [0,01; 0,05] ***
Poměr metodiků prevence sociálně patologických jevů k celkovému počtu žáků			-2,82 [-5,04; -0,60] **	-1,97 [-4,07; 0,13] *
Poměr úvazků asistenta pedagoga k celkovému počtu žáků		1,17 [-0,41; 2,75] .		
Poměr úvazků podpůrných profesí k celkovému počtu žáků	0,195 [0,091; 0,299] ***			2,19 [0,01; 4,37] **
Poměr úvazků školního psychologa k celkovému počtu žáků		4,83 [0,63; 9,03] **		
Poměr výchovných poradců k celkovému počtu žáků			1,79 [-0,61; 4,20] .	

Rozsah využití podpůrných šablon na vzdělávání a spolupráci pracovníků ve vzdělávání			-0,61 [-1,19; -0,02] **	-0,69 [-1,18; -0,20] ***
Soukromý zřizovatel	0,077 [-0,012; 0,167] *	-5,43 [-8,55; -2,32] ***		-2,71 [-4,38; -1,03] ***
Škola nevyužívá podporu a dle výkazu nemá 3 podpůrné pozice	-0,063 [-0,104; -0,022] ***		-0,97 [-2,15; 0,21] .	-1,00 [-2,09; 0,09] *
Škola nevyužívá podporu, ale dle výkazu má alespoň 3 podpůrné pozice	0,047 [-0,005; 0,100] *		-0,50 [-1,57; 0,56]	-0,68 [-1,61; 0,26] .
Škola využívá podporu (1 šablona)	-0,011 [-0,039; 0,018]		0,47 [-0,31; 1,24]	0,51 [-0,12; 1,15] .
Počet pozorování	915	928	931	933
Počet regresorů	17	14	17	20
Koeficient determinace (R ²)	24,80%	46,00%	69,70%	69,80%
Typ směrodatných chyb (SE)	HC1	HC1	HC1	HC1

Statistická významnost je vyznačena hvězdičkami: . p < 0,16; * p < 0,10; ** p < 0,05; *** p < 0,01.

Seznam 2: Přehled regresorů vybraných krokovou selekcí (SŠ)

Regresory společné pro všechny modely (objevují se alespoň ve 3 modelech ze 4)

- Míra odchodů bez pokračování ve studiu z výchozího počtu žáků
- Podíl dívek mezi žáky posledního ročníku
- Podíl kvalifikovaných učitelů mezi všemi učiteli
- Podíl maturantů volících matematiku ke společné části MZ
- Podíl zúčastněných k přihlášeným maturantům u společné části maturit
- Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků
- Poměr IT vybavení přístupného žákům k počtu žáků školy
- Kategorie využívání podpory (šablony / podpůrné pozice)
- Zřizovatel školy

Regresory specifické pro model průměrného hodnocení SŠ v klíčových oblastech, získaná z inspekční činnosti (neobjevují se v modelech úspěšnosti zakončení)

- Kategorie velikosti školy
- Míra přestupů na jinou školu z výchozího počtu žáků
- Poměr úvazků podpůrných profesí k celkovému počtu žáků

Regresory specifické pro modely úspěšného zakončení (neobjevují se v modelech JPZ)

- Percentil indexu průchodnosti a stability vzdělávací dráhy
- Podíl žáků opakujících ročník z celkového počtu žáků
- Podíl žáků se SVP k celkovému počtu žáků
- Poměr metodiků prevence sociálně patologických jevů k celkovému počtu žáků
- Poměr úvazků asistenta pedagoga k celkovému počtu žáků
- Poměr úvazků školního psychologa k celkovému počtu žáků
- Poměr výchovných poradců k celkovému počtu žáků
- Rozsah využití podpůrných šablon na vzdělávání a spolupráci pracovníků ve vzdělávání

Tabulka 5: Výsledky lineárního regresního modelu smíšených efektů vysvětlujícího průměrnou úroveň hodnocení ZŠ a podíl úspěšnějších žáků u JPZ

Regresory	Regresandy					
	Průměrná úroveň hodnocení školy v klíčových oblastech, získaná z inspekční činnosti			Podíl žáků 9. tříd s nadmediánovým výsledkem u JPZ (ze všech žáků)		
	Odhad koeficientu [95% interval spolehlivosti] Statistická významnost					
(Konstanta)	2,608	[2,468; 2,749]	***	-6,99	[-17,10; 3,13]	
Kategorie velikosti školy	0,022	[0,013; 0,032]	***	0,58	[0,14; 1,02]	**
Nestátní zřizovatel				11,71	[7,89; 15,52]	***
Podíl dívek mezi všemi žáky				0,19	[0,05; 0,34]	*
Podíl kvalifikovaných učitelů mezi všemi učiteli na 2. stupni	0,001	[-0,000; 0,002]		0,02	[-0,03; 0,07]	
Podíl nezaměstnaných osob				-1,66	[-2,48; -0,83]	***
Podíl opakujících žáků na 1. stupni z počtu žáků 1. stupně	-0,010	[-0,025; 0,006]				
Podíl osob v exekuci				-1,03	[-1,52; -0,54]	***
Podíl zúčastněných ku přihlášeným žákům z devátých tříd k JPZ	0,001	[0,000; 0,002]	*	0,70	[0,65; 0,74]	***
Podíl žáků 1. stupně docházejících do školní družiny/klubu z počtu žáků 1. stupně				0,05	[0,01; 0,09]	**
Podíl žáků 1. třídy s odkladem PŠD z počtu přijatých žáků 1. tříd				-0,10	[-0,17; -0,03]	**
Podíl žáků přestupujících na jinou školu z výchozího počtu žáků školy	-0,014	[-0,020; -0,009]	***	-0,36	[-0,67; -0,06]	*
Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků	-0,003	[-0,004; -0,002]	***	-0,14	[-0,24; -0,03]	**
Podíl žáků se SVP k celkovému počtu žáků	0,003	[0,001; 0,005]	**			
Podíl žáků se SVP opakujících ročník na 2. stupni z žáků se SVP na 2. stupni				0,26	[0,07; 0,44]	**
Podíl žáků se SVP, kteří docházejí do školní družiny/klubu k počtu žáků se SVP	0,001	[0,000; 0,001]	*			
Podíl žáků, kteří přešli na víceleté gymnázium	-0,010	[-0,021; 0,002]	.	1,36	[0,83; 1,88]	***
Poměr IT vybavení přístupného žákům na 2. stupni k počtu žáků 2. stupně	0,001	[0,001; 0,001]	***	-0,02	[-0,04; -0,00]	*
Poměr metodiků prevence sociálně patologických jevů k celkovému počtu žáků	0,070	[-0,001; 0,141]	.			
Poměr úvazků asistenta pedagoga k celkovému počtu žáků				-0,96	[-1,76; -0,16]	*
Poměr úvazků podpůrných profesí k celkovému počtu žáků	0,086	[0,028; 0,143]	**	2,63	[-0,99; 6,25]	
Rozsah využití podpůrných šablon na vzdělávání a spolupráci pracovníků ve vzdělávání	0,015	[0,004; 0,026]	**			
Soukromý zřizovatel	0,228	[0,144; 0,313]	***			
Škola využívá projektovou podporu na zajištění podpůrných pozic	0,024	[0,000; 0,048]	*			

Výskyt nadaných žáků na škole	0,034 [0,005; 0,062] *	
Směrodatná odchylka náhodného interceptu (kraj / ORP)	0,06	4,79
Směrodatná odchylka reziduí	0,19	10,01
Počet pozorování	1 774	1 780
Počet regresorů	16	16
Marginální koeficient determinace (R^2_m)	12,10 %	57,20 %
Podmíněný koeficient determinace (R^2_c)	19,30 %	65,20 %
Upravená intratřídní korelace (Adjusted ICC)	8,30 %	18,70 %
Typ směrodatných chyb (SE)	CR2	CR2

Statistická významnost je vyznačena hvězdičkami: . $p < 0,1$; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Barevné stínování odráží velikost standardizovaného efektu

Tabulka 6: Diagnostika multikolinearity vysvětlujících proměnných základních škol (VIF)

Regresory	Regresandy			
	Průměrná úroveň hodnocení školy v klíčových oblastech, získaná z inspekční činnosti		Podíl žáků 9. tříd s nadmediánovým výsledkem u JPZ (ze všech žáků)	
	Variačního inflačního faktoru (VIF) Tolerance (v %)			
Kategorie velikosti školy	1,73	57,82 %	1,40	71,61 %
Nestátní zřizovatel			1,23	81,14 %
Podíl dívek mezi všemi žáky			1,05	95,52 %
Podíl kvalifikovaných učitelů mezi všemi učiteli na 2. stupni	1,05	95,05 %	1,09	91,34 %
Podíl nezaměstnaných osob			2,13	46,94 %
Podíl opakujících žáků na 1. stupni z počtu žáků 1. stupně	1,27	78,61 %		
Podíl osob v exekuci			2,17	46,13 %
Podíl zúčastněných ku přihlášeným žákům z devátých tříd k JPZ	1,42	70,61 %	1,52	65,78 %
Podíl žáků 1. stupně docházejících do školní družiny/klubu z počtu žáků 1. stupně			1,17	85,59 %
Podíl žáků 1. třídy s odkladem PŠD z počtu přijatých žáků 1. tříd			1,18	84,40 %
Podíl žáků přestupujících na jinou školu z výchozího počtu žáků školy	1,27	78,51 %	1,27	78,51 %
Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků	1,24	80,70 %	1,30	76,98 %
Podíl žáků se SVP k celkovému počtu žáků	1,12	88,94 %		
Podíl žáků se SVP opakujících ročník na 2. stupni z žáků se SVP na 2. stupni			1,04	96,59 %
Podíl žáků se SVP, kteří docházejí do školní družiny/klubu k počtu žáků se SVP	1,08	92,32 %		
Podíl žáků, kteří přešli na víceleté gymnázium	1,05	95,31 %	1,08	92,67 %
Poměr IT vybavení přístupného žákům na 2. stupni k počtu žáků 2. stupně	1,23	81,45 %	1,20	83,52 %
Poměr metodiků prevence sociálně patologických jevů k celkovému počtu žáků	1,41	70,69 %		
Poměr úvazků asistenta pedagoga k celkovému počtu žáků			1,36	73,59 %
Poměr úvazků podpůrných profesí k celkovému počtu žáků	1,12	89,34 %	1,11	89,74 %

Rozsah využití podpůrných šablon na vzdělávání a spolupráci pracovníků ve vzdělávání	1,07	93,22 %
Soukromý zřizovatel	1,10	90,52 %
Škola využívá projektovou podporu na zajištění podpůrných pozic	1,12	88,94 %
Výskyt nadaných žáků na škole	1,05	95,29 %

Tabulka 7: Výsledky lineárního regresního modelu smíšených a fixních efektů vysvětlujícího průměrné hodnocení podmínek, průběhu a výsledků vzdělávání a míru úspěšného ukončení studia

	Regresandy	
	Průměrná úroveň hodnocení školy v klíčových oblastech, získaná z inspekční činnosti	Míra úspěšného ukončení (z žáků posledního ročníku)
Regresory	Odhad koeficientu [95% interval spolehlivosti] Statistická významnost	
(Konstanta)	1,073 [0,436; 1,710] ***	0,56 [-23,90; 25,03]
Kategorie velikosti školy	0,026 [0,015; 0,037] ***	
Míra pokračování ve studiu po 1. ročníku z výchozího počtu žáků	0,007 [0,002; 0,012] **	0,65 [0,44; 0,85] ***
Míra přestupů na jinou školu z výchozího počtu žáků	-0,009 [-0,017; -0,002] *	-0,22 [-0,52; 0,08]
Míra úspěšného ukončení v posledním ročníku z výchozího počtu žáků	0,003 [0,001; 0,005] **	
Odborná škola (převažující odborné obory)	-0,037 [-0,074; 0,001] .	-4,90 [-6,17; -3,63] ***
Podíl kvalifikovaných učitelů mezi všemi učiteli	0,002 [0,000; 0,003] *	0,07 [0,00; 0,13] *
Podíl maturantů volících matematiku ke společné části		0,03 [0,00; 0,05] *
Podíl zúčastněných k přihlášeným maturantům u společné části MZ	0,007 [0,002; 0,011] **	0,24 [0,09; 0,40] **
Podíl žáků opakujících ročník z celkového počtu žáků		-0,55 [-1,00; -0,09] *
Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků	-0,004 [-0,011; 0,003]	
Poměr IT vybavení přístupného žákům k počtu žáků školy		0,03 [0,00; 0,06] *
Poměr úvazků podpůrných profesí k celkovému počtu žáků	0,363 [0,221; 0,506] ***	
Poměr úvazků školního psychologa k celkovému počtu žáků	-0,199 [-0,408; 0,009] .	2,47 [-1,60; 6,54]
Přítomnost oborů s výučním listem		-2,29 [-3,86; -0,71] **
Smíšená škola (kombinace gymnaziálních a odborných oborů)	-0,072 [-0,125; -0,020] **	-2,94 [-4,50; -1,38] ***
Soukromý zřizovatel	0,084 [0,053; 0,116] ***	-2,18 [-3,65; -0,71] **
Škola využívá projektovou podporu na zajištění podpůrných pozic	0,038 [0,006; 0,069] *	1,01 [-0,16; 2,19] .
Typ modelu	Víceúrovňový lineární model (náhodný intercept: kraj/ORP)	Lineární regresní model (OLS)
Směrodatná odchylka náhodného interceptu (kraj / ORP)	0,08	
Směrodatná odchylka reziduí	0,17	

Počet pozorování	908	922
Počet regresorů	13	13
Koeficient determinace (R^2)		47,95 %
Marginální koeficient determinace (R^2_m)	22,20 %	
Podmíněný koeficient determinace (R^2_c)	36,80 %	
Upravená intratřídí korelace (Adjusted ICC)	18,83 %	
Typ směrodatných chyb (SE)	CR2	HC1

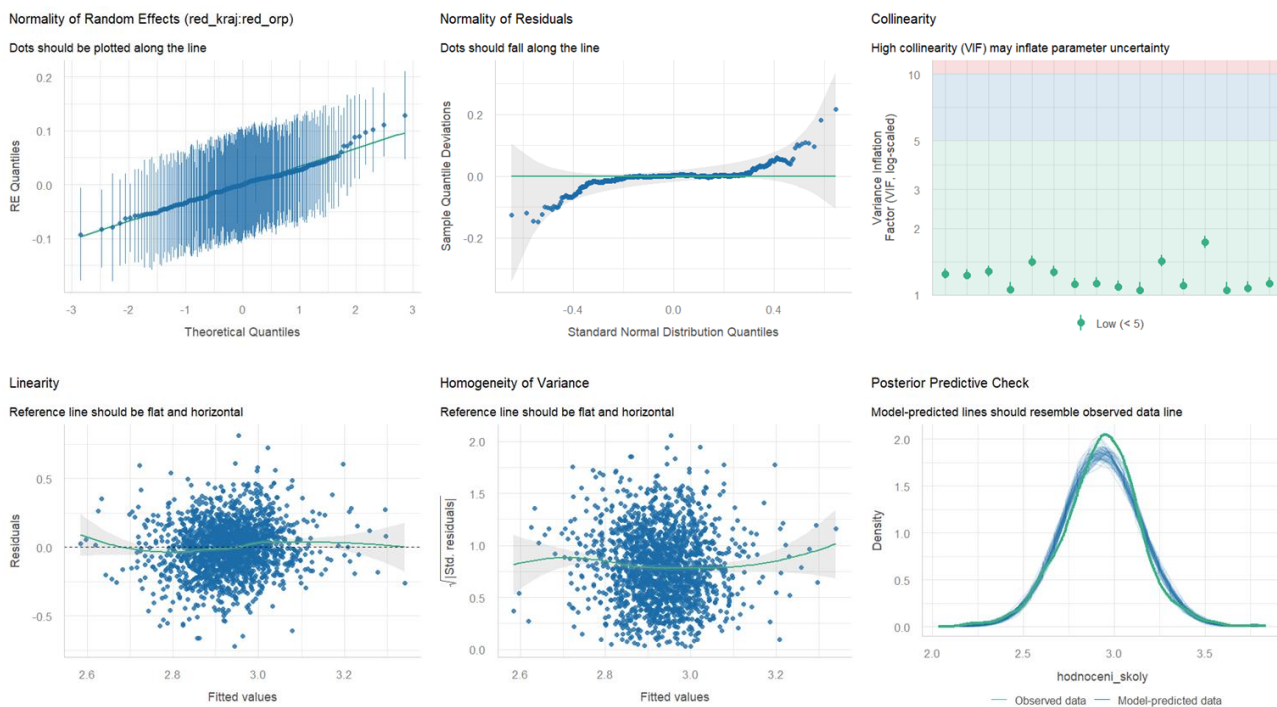
Statistická významnost je vyznačena hvězdičkami: . $p < 0,1$; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Barevné pruhy odrážejí velikost standardizovaného efektu

Tabulka 8: Diagnostika multikolinearity vysvětlujících proměnných středních škol (VIF)

Regresory	Regresandy			
	Průměrná úroveň hodnocení školy ČŠI		Míra úspěšného ukončení (z žáků posledního ročníku)	
	Variačního inflačního faktoru (VIF) / Tolerance (v %)			
Kategorie velikosti školy	1,39	71,79 %		
Míra pokračování ve studiu po 1. ročníku z výchozího počtu žáků	1,78	56,04 %	2,10	47,60 %
Míra přestupů na jinou školu z výchozího počtu žáků	1,30	76,84 %	1,30	77,17 %
Míra úspěšného ukončení v posledním ročníku z výchozího počtu žáků	1,89	52,93 %		
Podíl kvalifikovaných učitelů mezi všemi učiteli	1,14	87,81 %	1,15	86,91 %
Podíl maturantů volících matematiku ke společné části MZ			1,60	62,53 %
Podíl zúčastněných k přihlášeným maturantům u společné části MZ	1,38	72,67 %	1,37	73,04 %
Podíl žáků opakujících ročník z celkového počtu žáků			1,61	61,93 %
Podíl žáků s jiným než českým státním občanstvím k celkovému počtu žáků	1,20	83,20 %		
Poměr IT vybavení přístupného žákům k počtu žáků školy			1,11	89,80 %
Poměr úvazků podpůrných profesí k celkovému počtu žáků	3,07	32,56 %		
Poměr úvazků školního psychologa k celkovému počtu žáků	3,11	32,19 %	1,07	93,56 %
Přítomnost oborů s výučním listem			2,14	46,65 %
Soukromý zřizovatel	1,40	71,32 %	1,18	85,08 %
Škola využívá projektovou podporu na zajištění podpůrných pozic	1,05	95,34 %	1,05	95,69 %
Základní klasifikace střední školy	1,62	61,69 %	2,11	47,33 %

Graf 11: Souhrnná diagnostika smíšeného regresního modelu průměrné úrovně inspekčního hodnocení základní školy



Graf 12: Souhrnná diagnostika smíšeného regresního modelu podílu žáků 9. tříd s nadmediánovým výsledkem u JPZ ze všech žáků



Graf 13: Souhrnná diagnostika smíšeného regresního modelu průměrné úrovně inspekčního hodnocení střední školy



Graf 14: Souhrnná diagnostika regresního modelu podíl žáků dané SŠ, kteří v daném roce úspěšně ukončili studium maturitní zkouškou nebo získáním výučního listu



4.2 VÍCEROZMĚRNÉ STATISTICKÉ ANALÝZY UMOŽŇUJÍCÍ CHARAKTERIZOVAT PREDIKTORY NEROVNOSTI NA ÚROVNI DOSPĚLÉ POPULACE

Žlutým ohraničením jsou v následujících tabulkách značeny referenční kategorie.

HYPOTÉZA H1: Jedinci (respondenti) se vzděláním nižším, než střední vzdělání bez maturity mají při kontrole všech ostatních nezávislých proměnných vyšší pravděpodobnost nedosažení funkční gramotnosti než osoby s vyšším stupněm vzdělání.

Tabulka 9: H1 Čtenářská gramotnost

Proměnná	Kategorie	Čtenářská gramotnost				
		B (Koeficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,32	0,73	0,14	0,02	-27%
Věková kategorie	50-60 let	-0,87	0,42	0,15	0,00	-58%
Pohlaví	Muž		1,00			0%
Pohlaví	Žena	-0,16	0,85	0,10	0,11	-
Nejvyšší dokončené vzdělání	Vyučení bez maturity		1,00			
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného	0,04	1,06	0,36	0,90	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,31	3,70	0,14	0,00	270%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	2,22	9,26	0,15	0,00	826%
Pracovní status	Pracující OSVČ		1,00			
Pracovní status	Pracující zaměstnanec	-0,05	0,96	0,19	0,80	-
Pracovní status	Nepracující (bez studentů)	-0,66	0,52	0,24	0,00	-48%
Důvěra v ostatní	Nízká		1,00			
Důvěra v ostatní	Střední	0,22	1,25	0,13	0,08	-
Důvěra v ostatní	Vysoká	0,54	1,72	0,16	0,00	72%
Vnímaná možnost ovlivnit politiku	Malá		1,00			
Vnímaná možnost ovlivnit politiku	Střední	0,22	1,24	0,13	0,10	-
Vnímaná možnost ovlivnit politiku	Velká	0,72	2,06	0,13	0,00	106%
Účast na dobrovolné činnosti	Ne		1,00			
Účast na dobrovolné činnosti	Ano	0,19	1,22	0,16	0,21	-

Tabulka 10: H1 Numerická gramotnost

Proměnná	Kategorie	Numerická gramotnost				
		B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	0,08	1,09	0,13	0,54	-
Věková kategorie	50-60 let	-0,42	0,66	0,14	0,00	-34%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,59	0,55	0,12	0,00	-45%
Nejvyšší dokončené vzdělání	Vyučení bez maturity		1,00			
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného	0,13	1,14	0,31	0,68	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,43	4,19	0,16	0,00	319%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	2,44	11,54	0,18	0,00	1054%
Pracovní status	Pracující OSVČ		1,00			
Pracovní status	Pracující zaměstnanec	-0,03	0,98	0,21	0,89	-
Pracovní status	Nepracující (bez studentů)	-0,71	0,50	0,23	0,00	-50%
Důvěra v ostatní	Nízká		1,00			
Důvěra v ostatní	Střední	0,30	1,35	0,13	0,02	35%
Důvěra v ostatní	Vysoká	0,66	1,95	0,15	0,00	95%
Vnímaná možnost ovlivnit politiku	Malá		1,00			
Vnímaná možnost ovlivnit politiku	Střední	0,14	1,15	0,13	0,29	-
Vnímaná možnost ovlivnit politiku	Velká	0,75	2,13	0,12	0,00	113%
Účast na dobrovolné činnosti	Ne		1,00			
Účast na dobrovolné činnosti	Ano	0,35	1,42	0,13	0,01	42%

Tabulka 11: H1 Schopnost řešit problémy

Proměnná	Kategorie	Schopnost řešit problémy				
		B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,28	0,76	0,14	0,04	-24%
Věková kategorie	50-60 let	-1,07	0,34	0,16	0,00	-66%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,42	0,65	0,12	0,00	-35%
Nejvyšší dokončené vzdělání	Vyučení bez maturity		1,00			
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného	0,46	1,60	0,40	0,26	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,31	3,73	0,22	0,00	273%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	2,21	9,16	0,19	0,00	816%
Pracovní status	Pracující OSVČ		1,00			
Pracovní status	Pracující zaměstnanec	-0,03	0,98	0,19	0,88	-
Pracovní status	Nepracující (bez studentů)	-0,65	0,52	0,21	0,00	-48%
Důvěra v ostatní	Nízká		1,00			
Důvěra v ostatní	Střední	-0,02	0,98	0,13	0,86	-
Důvěra v ostatní	Vysoká	0,49	1,65	0,17	0,00	65%
Vnímaná možnost ovlivnit politiku	Malá		1,00			
Vnímaná možnost ovlivnit politiku	Střední	-0,02	0,98	0,13	0,86	-
Vnímaná možnost ovlivnit politiku	Velká	0,74	2,11	0,13	0,00	111%
Účast na dobrovolné činnosti	Ne		1,00			
Účast na dobrovolné činnosti	Ano	0,15	1,16	0,15	0,32	-

HYPOTÉZA H2: Zaměstnanci pracující v manuálních profesích s nižšími nároky na kvalifikaci mají při kontrole všech ostatních nezávislých proměnných vyšší pravděpodobnost nedosažení funkční gramotnosti než zaměstnanci pracující v jiných typech profesí.

Tabulka 12: H2 Čtenářská gramotnost

Čtenářská gramotnost						
Proměnná	Kategorie	B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,29	0,75	0,16	0,07	-
Věková kategorie	50-60 let	-0,73	0,48	0,19	0,00	-52%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,27	0,76	0,12	0,03	-24%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	0,08	1,15	0,52	0,87	15%
Nejvyšší dokončené vzdělání	Střední s maturitou	1,20	3,57	0,56	0,03	257%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	1,99	7,90	0,58	0,00	690%
Hlavní třída zaměstnání	Kvalifikovaní manuální pracovníci		1,00			
Hlavní třída zaměstnání	Zákonodárci a řídicí pracovníci	0,63	1,93	0,45	0,17	-
Hlavní třída zaměstnání	Specialisté	0,89	2,47	0,33	0,01	147%
Hlavní třída zaměstnání	Techničtí a odborní pracovníci	0,54	1,72	0,28	0,06	-
Hlavní třída zaměstnání	Úředníci	0,46	1,60	0,31	0,14	-
Hlavní třída zaměstnání	Pracovníci ve službách a prodeji	-0,01	0,99	0,26	0,96	-
Hlavní třída zaměstnání	Pomocní a nequalifikovaní pracovníci	-0,59	0,56	0,31	0,06	-
Ekonomické odvětví	Zdravotnictví		1,00			
Ekonomické odvětví	Primární sektor a energetika	-0,39	0,68	0,43	0,36	-
Ekonomické odvětví	Výroba a stavebnictví	0,19	1,22	0,25	0,44	-
Ekonomické odvětví	Obchod, doprava, pohostinství	0,28	1,33	0,26	0,27	-
Ekonomické odvětví	Ostatní služby	0,56	1,78	0,28	0,04	78%
Ekonomické odvětví	Školství	-0,22	0,81	0,30	0,47	-

Tabulka 13: H2 Numerická gramotnost

Proměnná	Kategorie	Numerická gramotnost				
		B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	50-60 let	-0,23	0,79	0,19	0,21	-
Věková kategorie	35-49 let	0,12	1,13	0,16	0,46	-
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,71	0,49	0,17	0,00	-51%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	-0,16	0,86	0,38	0,66	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,10	3,03	0,38	0,00	203%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	1,85	6,43	0,42	0,00	543%
Hlavní třída zaměstnání	Kvalifikovaní manuální pracovníci		1,00			
Hlavní třída zaměstnání	Zákonodárci a řídící pracovníci	0,73	2,09	0,36	0,04	109%
Hlavní třída zaměstnání	Specialisté	0,85	2,36	0,29	0,00	136%
Hlavní třída zaměstnání	Techničti a odborní pracovníci	0,59	1,81	0,26	0,03	81%
Hlavní třída zaměstnání	Úředníci	0,43	1,55	0,27	0,11	-
Hlavní třída zaměstnání	Pracovníci ve službách a prodeji	-0,01	1,00	0,24	0,96	-
Hlavní třída zaměstnání	Pomocní a nekvalifikovaní pracovníci	-0,52	0,60	0,33	0,11	-
Ekonomické odvětví	Zdravotnictví		1,00			
Ekonomické odvětví	Primární sektor a energetika	-0,25	0,81	0,47	0,59	-
Ekonomické odvětví	Výroba a stavebnictví	0,32	1,40	0,30	0,28	-
Ekonomické odvětví	Obchod, doprava, pohostinství	0,39	1,50	0,32	0,22	-
Ekonomické odvětví	Ostatní služby	0,60	1,85	0,29	0,04	85%
Ekonomické odvětví	Školství	-0,20	0,85	0,39	0,61	-

Tabulka 14: H2 Schopnost řešit problémy

Proměnná	Kategorie	Schopnost řešit problémy				
		B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,23	0,80	0,17	0,18	-
Věková kategorie	50-60 let	-0,93	0,40	0,19	0,00	-60%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,51	0,60	0,14	0,00	-40%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	-0,42	0,66	0,45	0,35	-
Nejvyšší dokončené vzdělání	Střední s maturitou	0,69	2,05	0,50	0,17	-
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	1,33	3,90	0,52	0,01	290%
Hlavní třída zaměstnání	Kvalifikovaní manuální pracovníci		1,00			
Hlavní třída zaměstnání	Zákonodárci a řídící pracovníci	0,69	2,06	0,43	0,10	-
Hlavní třída zaměstnání	Specialisté	1,01	2,77	0,30	0,00	177%
Hlavní třída zaměstnání	Techničti a odborní pracovníci	0,57	1,80	0,32	0,07	-
Hlavní třída zaměstnání	Úředníci	0,49	1,64	0,30	0,11	-
Hlavní třída zaměstnání	Pracovníci ve službách a prodeji	0,10	1,12	0,32	0,75	-
Hlavní třída zaměstnání	Pomocní a nekvalifikovaní pracovníci	-0,49	0,63	0,40	0,22	-
Ekonomické odvětví	Zdravotnictví		1,00			
Ekonomické odvětví	Primární sektor a energetika	-0,17	0,87	0,47	0,71	-
Ekonomické odvětví	Výroba a stavebnictví	0,44	1,57	0,29	0,13	-
Ekonomické odvětví	Obchod, doprava, pohostinství	0,59	1,82	0,33	0,07	-
Ekonomické odvětví	Ostatní služby	0,76	2,16	0,31	0,01	116%
Ekonomické odvětví	Školství	-0,13	0,90	0,36	0,73	-

H3: Jedinci (respondenti), kteří se neúčastní dalšího vzdělávání (formálního či neformálního), mají při kontrole všech ostatních nezávislých proměnných vyšší pravděpodobnost nedosažení funkční gramotnosti než jedinci, kteří se dalšího vzdělávání účastní.

Tabulka 15: H3 Čtenářská gramotnost

Čtenářská gramotnost						
Proměnná	Kategorie	B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,30	0,74	0,13	0,02	-26%
Věková kategorie	50-60 let	-0,78	0,46	0,15	0,00	-54%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,22	0,80	0,10	0,03	-20%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	0,04	1,06	0,35	0,91	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,38	4,06	0,34	0,00	306%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	2,36	10,74	0,34	0,00	974%
Účast na dalším vzdělávání v posledních 12 měsících	Účastnil/a se		1,00			
Účast na dalším vzdělávání v posledních 12 měsících	Neúčastnil/a se	-0,83	0,44	0,25	0,00	-56%
Účast na dalším vzdělávání a pracovní status	Účastnil/a se, pracující		1,00			
Účast na dalším vzdělávání a pracovní status	Účastnil/a se, nepracující (bez studentů)	-0,17	0,86	0,35	0,63	-
Zaměření dalšího vzdělávání	Účastnil/a se, pracovní		1,00			
Zaměření dalšího vzdělávání	Účastnil/a se, nepracovní	0,30	1,36	0,25	0,23	-
Učení se v práci	Velmi časté		1,00			
Učení se v práci	Středně časté	0,25	1,29	0,17	0,15	-
Učení se v práci	Méně časté	0,26	1,30	0,19	0,17	-

Tabulka 16: H3 Numerická gramotnost

Numerická gramotnost						
Proměnná	Kategorie	B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	0,09	1,10	0,13	0,46	-
Věková kategorie	50-60 let	-0,32	0,73	0,13	0,02	-27%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,66	0,52	0,12	0,00	-48%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	-0,03	0,97	0,29	0,91	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,44	4,24	0,29	0,00	324%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	2,55	12,90	0,30	0,00	1190%
Účast na dalším vzdělávání v posledních 12 měsících	Účastnil/a se		1,00			
Účast na dalším vzdělávání v posledních 12 měsících	Neúčastnil/a se	-0,73	0,48	0,21	0,00	-52%
Účast na dalším vzdělávání a pracovní status	Účastnil/a se, pracující		1,00			
Účast na dalším vzdělávání a pracovní status	Účastnil/a se, nepracující (bez studentů)	-0,11	0,90	0,32	0,73	-
Zaměření dalšího vzdělávání	Účastnil/a se, pracovní		1,00			
Zaměření dalšího vzdělávání	Účastnil/a se, nepracovní	0,32	1,40	0,24	0,18	-
Učení se v práci	Velmi časté		1,00			
Učení se v práci	Středně časté	0,41	1,51	0,17	0,01	51%
Učení se v práci	Méně časté	0,35	1,43	0,16	0,02	43%

Tabulka 17: H3 Schopnost řešit problémy

Proměnná	Kategorie	Schopnost řešit problémy				
		B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,26	0,77	0,13	0,05	-
Věková kategorie	50-60 let	-1,01	0,37	0,16	0,00	-63%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,47	0,63	0,12	0,00	-37%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	-0,41	0,67	0,38	0,28	-
Nejvyšší dokončené vzdělání	Střední s maturitou	0,95	2,64	0,37	0,01	164%
Nejvyšší dokončené vzdělání	VOŠ a vysokoškolské	1,93	7,00	0,38	0,00	600%
Účast na dalším vzdělávání v posledních 12 měsících	Účastnil/a se		1,00			
Účast na dalším vzdělávání v posledních 12 měsících	Neúčastnil/a se	-0,73	0,49	0,25	0,00	-51%
Účast na dalším vzdělávání a pracovní status	Účastnil/a se, pracující		1,00			
Účast na dalším vzdělávání a pracovní status	Účastnil/a se, nepracující (bez studentů)	-0,09	0,93	0,36	0,81	-
Zaměření dalšího vzdělávání	Účastnil/a se, pracovní		1,00			
Zaměření dalšího vzdělávání	Účastnil/a se, nepracovní	0,04	1,05	0,23	0,86	-
Učení se v práci	Velmi časté		1,00			
Učení se v práci	Středně časté	0,33	1,41	0,21	0,12	-
Učení se v práci	Méně časté	0,22	1,25	0,22	0,32	-

H4: Jedinci, kteří vnímají vyšší možnost ovlivnit politická rozhodnutí, mají při kontrole všech ostatních nezávislých proměnných vyšší pravděpodobnost dosažení funkční gramotnosti než jedinci, kteří vnímají svou možnost ovlivnit politiku jako nízkou.

Tabulka 18: H4 Čtenářská gramotnost

Proměnná	Kategorie	Čtenářská gramotnost				
		B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,23	0,79	0,14	0,09	-
Věková kategorie	50-60 let	-0,82	0,44	0,14	0,00	-56%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,29	0,75	0,10	0,00	-25%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	0,04	1,06	0,36	0,90	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,37	4,02	0,35	0,00	302%
Nejvyšší dokončené vzdělání	Vysokoškolské bakalářské včetně VOŠ	2,36	10,75	0,35	0,00	975%
Vnímaná možnost ovlivnit politiku	Malá		1,00			
Vnímaná možnost ovlivnit politiku	Střední	0,83	2,30	0,13	0,00	130%
Vnímaná možnost ovlivnit politiku	Velká	0,26	1,30	0,13	0,05	30%

Tabulka 19: H4 Numerická gramotnost

Numerická gramotnost						
Proměnná	Kategorie	B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	0,17	1,19	0,13	0,19	-
Věková kategorie	50-60 let	-0,36	0,70	0,13	0,01	-30%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,72	0,49	0,12	0,00	-51%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	-0,01	0,99	0,31	0,97	-
Nejvyšší dokončené vzdělání	Střední s maturitou	1,43	4,21	0,31	0,00	321%
Nejvyšší dokončené vzdělání	Vysokoškolské bakalářské včetně VOŠ	2,54	12,72	0,31	0,00	1172%
Vnímaná možnost ovlivnit politiku	Malá		1,00			
Vnímaná možnost ovlivnit politiku	Střední	0,89	2,44	0,12	0,00	144%
Vnímaná možnost ovlivnit politiku	Velká	0,19	1,21	0,12	0,13	-

Tabulka 20: H4 Schopnost řešit problémy

Schopnost řešit problémy						
Proměnná	Kategorie	B (Koefficient)	Exp(B) / Odds Ratio	S.E. (Chyba)	Signifikance (p)	Změna šance (%)
Věková kategorie	16-34 let		1,00			
Věková kategorie	35-49 let	-0,20	0,82	0,14	0,14	-
Věková kategorie	50-60 let	-1,02	0,36	0,15	0,00	-64%
Pohlaví	Muž		1,00			
Pohlaví	Žena	-0,54	0,59	0,12	0,00	-41%
Nejvyšší dokončené vzdělání	Základní včetně nedokončeného		1,00			
Nejvyšší dokončené vzdělání	Vyučení bez maturity	-0,39	0,68	0,40	0,33	-
Nejvyšší dokončené vzdělání	Střední s maturitou	0,94	2,61	0,38	0,01	161%
Nejvyšší dokončené vzdělání	Vysokoškolské bakalářské včetně VOŠ	1,90	6,77	0,39	0,00	577%
Vnímaná možnost ovlivnit politiku	Malá		1,00			
Vnímaná možnost ovlivnit politiku	Střední	0,84	2,32	0,12	0,00	132%
Vnímaná možnost ovlivnit politiku	Velká	0,30	1,35	0,16	0,06	-

5 SEZNAMY

5.1 SEZNAM TABULEK

Tabulka 1: Statistik ověřování významnosti náhodných efektů na úrovni ORP a krajů v rámci ZŠ	28
Tabulka 2: Statistik ověřování významnosti náhodných efektů na úrovni ORP a krajů v rámci SŠ	28
Tabulka 3: Výsledky lineárního regresního modelu vysvětlujícího průměrnou úroveň hodnocení ZŠ a podíl úspěšnějších žáků u JPZ	37
Tabulka 4: Výsledky lineárního regresního modelu vysvětlujícího průměrnou úroveň hodnocení SŠ a podíl úspěšného zakončení studia	39
Tabulka 5: Výsledky lineárního regresního modelu smíšených efektů vysvětlujícího průměrnou úroveň hodnocení ZŠ a podíl úspěšnějších žáků u JPZ	42
Tabulka 6: Diagnostika multikolinearity vysvětlujících proměnných základních škol (VIF)	43
Tabulka 7: Výsledky lineárního regresního modelu smíšených a fixních efektů vysvětlujícího průměrnou úroveň hodnocení SŠ a míru úspěšného ukončení studia	44
Tabulka 8: Diagnostika multikolinearity vysvětlujících proměnných středních škol (VIF)	45
Tabulka 9: H1 Čtenářská gramotnost	48
Tabulka 10: H1 Numerická gramotnost	49
Tabulka 11: H1 Schopnost řešit problémy	49
Tabulka 12: H2 Čtenářská gramotnost	50
Tabulka 13: H2 Numerická gramotnost	51
Tabulka 14: H2 Schopnost řešit problémy	51
Tabulka 15: H3 Čtenářská gramotnost	52
Tabulka 16: H3 Numerická gramotnost	52
Tabulka 17: H3 Schopnost řešit problémy	53
Tabulka 18: H4 Čtenářská gramotnost	53
Tabulka 19: H4 Numerická gramotnost	54
Tabulka 20: H4 Schopnost řešit problémy	54

5.2 SEZNAM GRAFŮ

Graf 1: Korelační matice ukazatelů definující velikost základní školy (vlevo) a střední školy (vpravo)	19
Graf 2: Korelační matice ukazatelů vstupujících do hlavních zjištění deskriptivní analýzy ZŠ	20
Graf 3: Korelační matice ukazatelů vstupujících do hlavních zjištění deskriptivní analýzy SŠ	21
Graf 4: Asociační matice kategoriálních ukazatelů základních škol	22
Graf 5: Paretův diagram vysvětlené a kumulativní variance hlavních komponent (ZŠ vlevo, SŠ vpravo)	23
Graf 6: Bivariátní vztahy mezi významnými charakteristikami ZŠ a průměrnou úrovní hodnocení ČŠI	29
Graf 7: Bivariátní vztahy mezi významnými charakteristikami ZŠ a podílem úspěšnějších žáků u JPZ vzhledem ke všem žákům	30
Graf 8: Bivariátní vztahy mezi významnými charakteristikami SŠ a průměrnou úrovní hodnocení ČŠI	31
Graf 9: Bivariátní vztahy mezi významnými charakteristikami SŠ a míry úspěšného zakončení studia absolventy	32
Graf 10: Asociační matice kategoriálních ukazatelů středních škol	36
Graf 11: Souhrnná diagnostika smíšeného regresního modelu průměrné úrovně inspekčního hodnocení základní školy	46
Graf 12: Souhrnná diagnostika smíšeného regresního modelu podílu žáků 9. tříd s nadmediánovým výsledkem u JPZ ze všech žáků	46
Graf 13: Souhrnná diagnostika smíšeného regresního modelu průměrné úrovně inspekčního hodnocení střední školy	47
Graf 14: Souhrnná diagnostika regresního modelu podíl žáků dané SŠ, kteří v daném roce úspěšně ukončili studium maturitní zkouškou nebo získáním výučního listu	47

Kolektiv autorů projektu Datově-analytická podpora pro hodnocení a řízení vzdělávací soustavy ČR, 2026

Materiál je pod licencí Creative Commons CC BY SA 4.0

Uveďte původ – Zachovejte licenci 4.0 Mezinárodní.



Zpracovali zaměstnanci
oddělení analytické podpory a projektových výstupů
oddělení národních analýz
v odboru školské statistiky a analýz
v sekci informatiky, statistiky a analýz
na Ministerstvu školství, mládeže a tělovýchovy.

Dokument vznikl v rámci projektu s názvem:
Datově-analytická podpora pro hodnocení a řízení vzdělávací soustavy ČR
Registrační číslo projektu: CZ.02.02.XX/00/22_005/0002901

V případě dotazů nás kontaktujte:

ips.data@msmt.gov.cz



Spolufinancováno
Evropskou unií



edu.cz