

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



HODNOTENIE KVALITY ŠKOLY

DIPLOMOVÁ PRÁCA

2014

Lukáš Ivica

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

HODNOTENIE KVALITY ŠKOLY

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomická a finančná matematika

Študijný odbor: 1114 Aplikovaná matematika

Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky

Školiteľ: Mgr. Zuzana Juščáková, PhD.

Bratislava 2014

Lukáš Ivica



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Lukáš Ivica
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Hodnotenie kvality školy / *Assessment of School Quality*

Cieľ: Diplomová práca bude zameraná na nájdenie matematického modelu hodnotenia kvality práce školy v zmysle zásluhy školy na progrese žiaka v slovenskom jazyku a literatúre a matematike, ktorý by bol aplikovateľný aj po rozšírení národných testovaní z ďalších predmetov. Cieľom výskumu bude analyzovať využitie regresných modelov pre výpočet tzv. „pridanej hodnoty“ školy na dostupných dátach s pokusom o zahrnutie kontextuálnych premenných do modelu a skúmanie aplikácie hierarchických modelov v tejto oblasti.

Vedúci: Mgr. Zuzana Juščáková, PhD.

Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky

Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 25.01.2013

Dátum schválenia: 04.02.2013

prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Pod'akovanie

Touto cestou by som sa rád poďakoval svojej školiteľke Mgr. Zuzane Juščákovéj, PhD. za ochotu, vedenie a rady, ktoré mi pomohli pri písaní tejto práce. Taktiež ďakujem Mgr. Jánovi Somorčíkovi, PhD. za podnetné konzultácie a svojej rodine a priateľom za ich trpezlivosť a podporu.

Abstrakt

IVICA, Lukáš: Hodnotenie kvality školy [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Zuzana Juščáková, PhD., Bratislava, 2014, 69 s.

Cieľom práce je analyzovať využitie lineárnych regresných modelov a hierarchických modelov pre výpočet tzv. „pridanej hodnoty“ školy popisujúcej zásluhy školy na pokroku žiaka pre dostupné dáta v predmetoch slovenský jazyk a literatúra a matematika. Práca sa zameriava na skúmanie aplikácie týchto modelov a porovnanie vhodnosti ich využitia v oblasti edukometrie.

Kľúčové slová: pridaná hodnota; prostá pridaná hodnota; hierarchické modely, edukometria

Abstract

IVICA, Lukáš: Assessment of School Quality [Master thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; supervisor: Mgr. Zuzana Juščáková, PhD., Bratislava, 2014, 69 p.

The aim of this paper is to analyze the use of linear regression models and hierarchical models for the calculation of the so-called "value added" describing the merits of school to pupil's progress using data available in the subjects of Slovak language and literature and Mathematics. The work focuses on exploring applications of these models and the appropriateness of their use in educometrics.

Key words: value added; simple value added; hierarchical models, educometrics

Obsah

Zoznam skratiek	8
Zoznam obrázkov	9
Zoznam tabuliek	10
Úvod	11
1 Súčasný stav merania výsledkov edukácie	12
1.1 Objektívne hodnotenie kvality školy	12
1.2 Definícia pridanej hodnoty	13
1.3 Použitie kontextuálnych premenných pri modelovaní pridanej hodnoty	14
1.4 Hodnotenie škôl na Slovensku	15
1.5 Stav medzinárodného výskumu a realizácie modelov pridanej hodnoty	16
1.5.1 Modely pridanej hodnoty v Anglicku	17
1.5.2 Modely pridanej hodnoty v USA	18
1.5.3 Modely pridanej hodnoty v iných štátoch	19
2 Modely pridanej hodnoty	20
2.1 Jednoduchý verzus komplexný model	20
2.2 Základné modely	20
2.2.1 Klasické lineárne regresné modely	21
2.2.1.1 Predpoklady klasického lineárneho regresného modelu	22
2.2.1.2 Odľahlé hodnoty	25
2.2.2 Hierarchické modely	25
2.2.3 Modely s pevnými efektami	27
3 Empirický prístup	28
3.1 Výber dát	29
3.2 Transformácia skóre rozdielných skupín	31
3.3 Kvantilová normalizácia dát	33
3.4 Metóda 1 – prostý regresný model	35
3.5 Metóda 2 – viacnásobná regresia	41
3.5.1 Kontextuálne premenné	44

3.6	ICC koeficient	45
3.7	Metóda 3 – hierarchický model s náhodným priesečníkom . . .	47
3.8	Metóda 4 – model s pevnými efektami	48
4	Interpretácia a porovnanie výsledkov	49
4.1	Školy s malým výberom žiakov	50
4.2	Porovnanie bodových odhadov jednotlivých metód	50
4.3	Výpočet intervalov spoľahlivosti	51
	Záver	58
	Zoznam použitej literatúry	60
	Appendix	65
A.1		65
A.2		65
A.3		66
A.4		67
A.5		67

Zoznam skratiek

AIC - Akaike Information Criterion (Akaikeho informačné kritérium)
BLUP - Best Linear Unbiased Predictor
DVAAS - Dallas Value-Added Accountability System
EČMS - Externá časť maturitnej skúšky
ESCS - socioekonomický a kultúrny status žiaka
EVAAS - Educational Value-Added Assessment System
GYM - gymnáziá
HS - hrubé skóre
ICC - Intraclass correlation coefficient (Vnútrotriedny koeficient korelácie)
IRT - Item Response Theory (Teória odpovede na položku)
IBE - Instytut Badań Edukacyjnych
IS - interval spoľahlivosti
MAT - matematika
ML - Maximum likelihood
MM - Malá maturita
NÚCEM - Národný ústav certifikovaných meraní vzdelávania
OECD - Organisation for Economic Co-operation and Development (Organizácia pre ekonomickú spoluprácu a rozvoj)
OLS - Ordinary least squares
PH - pridaná hodnota
ptil - percentilové poradie
SJL - slovenský jazyk a literatúra
SOŠ - stredné odborné školy
T9 - Testovanie deviatakov
TVAAS - Tennessee Value-added Assessment System

Zoznam obrázkov

1	Vzťah medzi úspešnosťou vstupných a výstupných meraní (MAT vľavo, SJL vpravo)	30
2	Rozdelenia vstupných (vľavo) a výstupných (vpravo) meraní pre MAT (hore) a SJL (dole)	31
3	Rozdelenie podľa typu školy (MAT)	38
4	Rozdelenie podľa typu školy (SJL)	38
5	Zobrazenie kvantilov rozdelenia rezíduí k očakávaným hodnotám kvantilov normálneho rozdelenia - SOŠ MAT	43
6	Vychýlenie regresnej priamky spôsobené rozdielnosťou skupín na úrovni vyšších celkov	46
7	Regresné priamky po odstránení malých škôl (MAT vľavo a SJL vpravo)	47
8	Porovnanie bodových odhadov PH škôl vypočítaných jednotlivými modelmi pre MAT	52
9	Porovnanie bodových odhadov PH škôl vypočítaných jednotlivými modelmi pre SJL	53
10	IS po Bonferroniho úprave pre odhady PH škôl jednotlivých modelov pre MAT	56
11	IS po Bonferroniho úprave pre odhady PH škôl jednotlivých modelov pre SJL	57
12	Rozdelenie výstupných meraní pre skupiny A a B z MAT . . .	67

Zoznam tabuliek

1	Predpoklady klasického lineárneho regresného modelu	23
2	Porovnanie rozdielov rozdelení variánt testov A a B	33
3	Hodnoty koeficientov R a Eta^2 pre hrubé skóre a skóre transformované kvantilovou normalizáciou	34
4	Hodnoty korelačného koeficientu pre rôzne „variance stabilizing“ transformácie	37
5	Testovanie potreby rozdelenia dát podľa typu školy	39
6	Porovnanie lineárnosti vzťahu vstupných a výstupných meraní pre hrubé skóre a percentilové poradie	40
7	Hodnoty korelácie medzi výsledkami rôznych prostých modelov (MAT)	41
8	Hodnoty korelácie medzi výsledkami rôznych prostých modelov (SJL)	41
9	Výber modelu podľa kritérií R^2 a AIC	42
10	Kritériá R^2 a AIC pre rôzne viacnásobné modely	44
11	ICC pre dáta rozdelené podľa typu školy	47
12	AIC pre modely definované na s. 43 s náhodne modelovaným priesečníkom a_0	48
13	AIC pre modely definované na s. 43 s pevnými efektami vo forme dummy premenných	49
14	[P1] Linearita	68
15	[P7] Normalita rezíduí	69
16	[P6] Homoskedasticita ^b	69

Úvod

V dnešnej dobe, kedy hrá kvalitné vzdelanie dôležitú rolu v živote každého človeka, môžeme celosvetovo pozorovať narastajúcu snahu o zlepšenie vzdelávacieho systému. Neodmysliteľnou súčasťou tohto procesu je aj zlepšovanie evaluácie kvality školy, keďže precízne informácie o školách dávajú priestor k správne usmerneniu vzdelávacej politiky a rozvoju slabších oblastí vzdelávacieho systému. S rozširovaním požiadaviek kladených na vývoj oblasti edukometrie v posledných desaťročiach zaujalo dôležité miesto hodnotenie kvality školy na základe konceptu tzv. „pridanej hodnoty“.

Cieľom tejto práce je skúmať možnosť využitia tohto konceptu v rámci hodnotenia kvality stredných škôl na Slovensku. Historickým vývojom edukometrie a analýzou jej súčasného stavu sa na úvod snažíme čitateľa uviesť do problematiky. Následne predstavíme modely „pridanej hodnoty“ použité v tejto práci. Pochopenie teoretického pozadia týchto modelov a k nim prislúchajúcich predpokladov tvorí odrazový mostík pre empirickú časť práce zahŕňajúcu metodiku modelovania PH a s ňou súvisiace technické komplikácie. Na záver zhodnotíme, či je na základe dostupných údajov možné vytvoriť matematický model na hodnotenie kvality práce školy v zmysle zásluhy školy na pokroku žiaka a akým spôsobom je vhodné výsledky tohto modelu interpretovať.

1 Súčasný stav merania výsledkov edukácie

Narastajúci záujem o kvalitu školstva podnecuje otázku, nakoľko túto kvalitu vieme merať, či vôbec máme na Slovensku možnosti takýchto meraní a akým smerom by sa mal zamerať rozvoj hodnotenia kvality edukácie. V prvej kapitole tejto práce si predstavíme problematiku hodnotenia edukácie, pričom sa oboznámime so stavom meraní na Slovensku a vo svete.

1.1 Objektívne hodnotenie kvality školy

Využitie objektívneho hodnotenia kvality školy sa zacielfuje na 3 oblasti: opatrenia na zlepšenie školy, systém školskej sledovanosti a možnosť výberu školy. Je zložitá zavádzať opatrenia na zlepšenie školy z ohľadom na budúcnosť bez dôkladnej analýzy prítomnosti. V tomto zmysle objektívne hodnotenie školy pomáha identifikovať oblasti, na ktoré sa treba zamerať, kde je možnosť rásť a na tie, ktoré fungujú kvalitne. Zlepšenie hodnotenia školy zároveň prospieva systému školskej sledovanosti poskytujúcemu informáciu verejnosti, kde sa prirodzene kladie otázka, či je hodnotenie spravodlivé voči školám a či pracuje s dostatočne objektívnymi informáciami. Výber školy možno chápať ako voľbu rodiča zapísať dieťa do školy. V súčasnosti už vzdelávanie žiaka nie je natoľko predurčené miestom, kde žiak žije, čo ponúka väčšiu možnosť pri výbere školy a preto objektívna informácia o ich kvalite slúži ako možnosť škôl zlepšiť svoju kredibilitu v očiach rodičov [OECD, 2008].

Potreba objektívneho hodnotenia školy nastolila problém, ako ho dosiahnuť. Na základe akých charakteristík posudzovať kvalitu školy? Mohli by týmto meradlom byť známky žiakov na vysvedčení a či maturitný prospech alebo úspešnosť absolventa? Na základe analýz NÚCEM sa na Slovensku ukazuje, že vzájomné porovnávanie škôl na základe známkov udelených študentom na vysvedčení nie je spoľahlivé a preto známka nemôže byť indikátorom spoľahlivosti školy [Juščáková, 2013]. Na posudzovanie úspešnosti absolventa nie je dostatok dát a toto posudzovanie je vhodné skôr z dlhodobého hľadiska, v krátkodobom horizonte pre operatívne intervencie poslúžiť vo všeobecnosti nemôže. Preto v súčasnosti úspešnosť strednej školy posudzujeme hlavne na základe výstupných charakteristík zo záverečných maturitných testov.

Doteraz získané poznatky ako aj skúsenosti zo zahraničia ale ukazujú, že

do hodnotenia školy je nevyhnutné zohľadniť aspekt vstupných vlastností žiaka na začiatku štúdia. Potenciál a schopnosti žiakov tvoria základ pre budúce napredovanie. Preto aj fakt, že žiak s nadpriemernými predpokladmi pre štúdium dosiahol nadpriemerné výsledky nemožno pripisovať iba zásluhe samotnej školy. Komplexnosť problému určuje snahu nájsť vhodné charakteristiky, ktoré by boli schopné merateľným spôsobom popísať, ako sú vstupné zdroje premenené do výstupov. Preto je žiaduce nájsť meranie, ktoré by bolo akýmsi vyjadrením pokroku žiaka za čas strávený štúdiom na škole.

Ako uvádza Ryška „samotné školstvo je veľmi zložitý systém plný mnohonásobných a zväčša ťažko opísateľných interakcií. Tak ani výstupy nie sú prehľadnou množinou prvkov, ktoré by bolo možno ľahko identifikovať a ktorých charakteristiky by boli ľahko merateľné. Každý žiak vstupuje do školy s rozdielnymi schopnosťami a poznatkami a každý má dané rozdielne vplyvy na úspešné napredovanie. Zároveň na priebeh vzdelania žiaka vplyvajú rôznorodé vonkajšie faktory ako je rodina a socioekonomický status, mimoškolské aktivity, so zvyšujúcim sa vekom aj skupina priateľov, atď. Určiť výstupné charakteristiky je teda zložitou funkciou mnohých ťažko opísateľných vplyvov. Z tohto hľadiska sa tzv. „koncept pridanej hodnoty“ rysuje ako možnosť vyrovnáť sa aspoň s časťou problémov pri určovaní výstupných charakteristík a tak umožniť aspoň do určitej miery zistiť skutočný prínos školy.“ [Ryška, 2009, s. 9] Čo je to ale pridaná hodnota?

1.2 Definícia pridanej hodnoty

Základ analýzy pridanej hodnoty v edukometrii spočíva v informovaní o pokroku žiaka za obdobie strávené štúdiom v škole s ohľadom na predchádzajúce vedomosti a schopnosti. V rámci stanov OECD pridanú hodnotu a jej modelovanie definujeme nasledovne:

„Pridaná hodnota je príspevok školy k pokroku žiakov k predpísaným cieľom vzdelania (napr. kognitívny výkon), pričom je tento príspevok očistený od faktorov, ktoré prispievajú ku vzdelávaciemu pokroku žiaka a škola na ne nemá vplyv.“

„Za modelovanie pridanej hodnoty považujeme triedu štatistických modelov, ktoré odhadujú príspevky školy k pokroku žiaka v rámci uvedených alebo predpísaných cieľov vzdelávania (napr.

kognitívny výkon) meraných aspoň dvakrát počas doby vzdelávania.“ [OECD, 2008, s. 17]

Práve posudzovanie na základe dvoch alebo viacerých meraní dáva modelovaniu pridanej hodnoty oproti ostatným meradlám školských výsledkov väčšiu objektivitu, keďže modely, ktoré nezahŕňajú dve merania majú väčší problém postrehnúť účasť nezistených faktorov prispievajúcich k prospechu žiaka. „Školský výkon možno ľahšie izolovať od vedľajších faktorov pri analyzovaní pokroku študentov ako pri spoliehaní sa na výsledky študentov v jednom časovom bode.“ [OECD, 2008, s. 29]

Uvedené definície ale nešpecifikujú, akým spôsobom sa má pokrok žiaka merať, čo vedie k rôznorodým prístupom modelovania pridanej hodnoty (ďalej PH). Zároveň netreba zabúdať, že modely PH možno zostaviť len na základe údajov o výkonoch v oblasti vedomostí či schopností žiakov merateľných hlavne testovým spôsobom. Takouto formou v súčasnosti vieme popísať len časť výchovno-vzdelávacieho procesu zameranú hlavne na gramotnosť žiakov v partikulárnom vyučovanom predmete (napr. SJL, MAT). Táto ale netvorí jedinú zložku príspevku školy a preto treba mať na zreteli, aby „zameranosť na výkon vo vzdelávaní a preceňovanie štatistických výsledkov (výsledkov PH) neodsunuli do úzadia ďalšie aspekty vplyvu vzdelávania a pôsobenia školy na vývoj jednotlivca, regiónu a spoločnosti.“ [Juščáková, 2012]

1.3 Použitie kontextuálnych premenných pri modelovaní pridanej hodnoty

Okrem prínosu školy k pokroku žiaka počas štúdia prispievajú aj iné faktory nezávislé od školy, ktoré vieme nejakým spôsobom merať. Tieto faktory nazývame kontextuálne premenné. Sú to napríklad socioekonomický a kultúrny status žiaka (ESCS), veľkosť rodiny, motivácia k učeniu, prístup k študijným materiálom, vrodené schopnosti, rôzne formy samostatného vzdelávania a mnohé iné. Je otázne, nakoľko je potrebné a vhodné tieto premenné zahrnúť do modelovania PH a tiež nakoľko sú tieto premenné dostatočne merateľné pre využitie pri modelovaní.

Pre pochopenie problematiky si uveďme príklad. Povedzme, že škola s nadpriemerným počtom žiakov pochádzajúcich z rodín so slabším ESCS bude mať nižšiu pridanú hodnotu ako ostatné školy z dôvodu, že jej žiaci majú

nižší potenciál pre zvládnutie štúdia (vplyvom vrodených schopností, zníženej koncentrácie a prípravy pri domácich úlohách alebo z iných dôvodov). Empiricky sa ukáže vplyv ESCS na pokrok žiaka ako štatisticky významný. Po zahrnutí ESCS do modelovania PH môže zrazu model pre spomínanú školu vykazovať veľmi dobré výsledky (PH školy počítame na základe PH žiakov danej školy špecificky podľa použitého modelu). Zo štatistického hľadiska bolo zahrnutie kontextuálnych premenných do modelu správne, vysoká PH znamená, že škola, na to s akými žiakmi pracuje, si počína nadpriemerne dobre. Ale z teoretického hľadiska, ak sú rozhodnutia menežmentu robené na základe merania PH, nízka PH školy indikuje potrebu dodatočného financovania. V prípade spomínanej školy by dodatočné finančné prostriedky neboli poskytnuté, aj keď zo znalosti pozadia školy vyplýva, že tam potreba financovania existuje. Preto určenie vhodných kontextuálnych premenných a ich správna interpretácia a zahrnutie do modelu predstavujú veľmi zložitú problematiku. Najideálnejšie pre optimálny výber by bolo rozhodovanie sa na základe detailných analýz výsledkov meraní PH s informáciou o kontextuálnych premenných aj bez nej.

1.4 Hodnotenie škôl na Slovensku

Decentralizácia školstva na Slovensku po páde komunizmu prispela k väčšej autonómnosti vzdelávacích inštitúcií. Nárast autonómnosti však vyžaduje zvýšenie zodpovednosti školy za vzdelávacie výsledky, s čím súvisí potreba objektívneho sebahodnotenia funkčnosti a efektívnosti. To možno dosiahnuť len v prípade, ak bude školám poskytnutý nástroj slúžiaci na hodnotenie vlastného výkonu a školy budú schopné tento nástroj používať. „Základné a stredné školy na Slovensku ale zatiaľ nemajú zákonnú povinnosť pravidelne hodnotiť výsledky svojej práce [Juščáková, 2012] a terajšia forma každoročnej správy o činnosti školy zverejňovaná na internete sa nepovažuje za objektívne kritérium.

V minulosti neexistovala na Slovensku ucelená forma pre možnosť porovnávania kvality stredných škôl na základe dátových údajov. Od roku 1998 boli vyvíjané testovania deviatakov, tzv. Monitor 9, v súčasnosti Testovanie deviatakov (T9), administrované koncom povinného deväťročného vzdelávania a externá časť maturitnej skúšky (EČMS) s cieľom:

1. objektívne prijímať na vyšší stupeň vzdelávania,

2. certifikovať ukončenie daného vzdelávacieho stupňa,
3. získať údaje pre porovnateľné a objektívne hodnotenie vedomostí žiakov v Slovenskej republike.

Do praxe boli tieto testovania oficiálne zavedené v roku 2005, pričom sa stupňujú ich zamerania hlavne na čitateľskú a matematickú gramotnosť žiakov [Juščáková & Žoltak, 2013]. V súčasnosti prebieha na rôznych úrovniach snaha o rebríčkovanie škôl, vid' INECO. Toto posudzovanie ale dostatočne nezodpovedá štandardu krajín OECD. Vo výskumnom snažení NÚCEM sa preto začína formovať prístup zvažujúci vstupno-výstupné meranie škôl. Ponúka sa možnosť použiť výsledky zo spomínaných testovaní T9 a EČMS na vyhodnocovanie tzv. „pridanej hodnoty“ škôl, pomocného meradla kvality školy. Problematike sa venuje i táto diplomová práca.

1.5 Stav medzinárodného výskumu a realizácie modelov pridanej hodnoty

Hodnotenie kvality vzdelávania časom napreduje a preto sa snaha odborníkov upriamila na jeho sprehľadnenie, zefektívnenie a odbúranie subjektívnosti. Za prvé verejne dostupné ukazovatele kvality školy historicky možno považovať školské rebríčky založené na „neupravených“ výsledkoch z testov. Ich vysoká korelovanosť so socioekonomickým postavením viedla k vytvoreniu tzv. „cross-selective“ modelov, ktoré zahŕňali regresnú analýzu priemerného výsledku školy voči relevantnému demografickému pozadiu jej žiakov. Nadstavba „cross-selective“ modelov zahŕňa hierarchickú štruktúru školy do modelovania, čím vznikli tzv. „contextualised attainment“ modely [OECD, 2008, s. 15]. Všetky vymenované modely mali ale jeden nedostatok – vychádzali z jedného časového rezu. Preto sa informácia získaná týmito modelmi nepovažovala za dostatočnú pre vyvodzovanie zodpovednosti voči školám. Riešenie sa ponúкло až rozvojom modelov pridanej hodnoty.

Modelovanie pridanej hodnoty sa dostalo do pozornosti v rámci členských krajín OECD hlavne za posledné dve desaťročia a v krajinách ako Anglicko, či niektoré štáty USA (Tennessee, Južná Karolína, Pensylvánia atď.) sa už oficiálne používa na hodnotenie edukácie. Keďže modely pridanej hodnoty výrazne prospeli systému vzdelávania v podobe zefektívňovania práce škôl,

zjednodušenia systému školskej sledovanosti, hodnotenia kvality školy, vývoja informačných systémov, menežmentu financií v školstve a prekonaniu nesprávneho ohodnotenia pre socioekonomické činitele [OECD, 2008, s. 16], OECD v súčasnosti považuje hodnotenie kvality školy na základe pridanej hodnoty za najobjektívnejší spôsob a preto vzniká iniciatíva jeho implementácie naprieč členskými krajinami organizácie. Doteraz používané modely konceptu pridanej hodnoty ako nástroja pre hodnotenie kvality školy sa však výrazne líšia, keďže v každej krajine bol koncept pridanej hodnoty upravený pre lokálne potreby. Významnú rolu zohráva vhodnosť a dostupnosť dát o školách a žiakoch a zložitosť implementácie hodnotenia škôl pomocou PH v rámci krajiny.

Ako sa líšia spôsoby modelovania PH, môže sa líšiť aj cieľ ich využitia. V európskych krajinách PH vo všeobecnosti slúži na porovnávanie kvality škôl. Naopak v Spojených štátoch amerických či Austrálii, kde sú vedené oficiálne testovania žiakov každý rok, sa PH používa skôr ako nástroj na hodnotenie efektivity učiteľov. V nasledujúcich podcelkoch si predstavíme vývoj najznámejších modelov PH v niektorých krajinách.

1.5.1 Modely pridanej hodnoty v Anglicku

Prvé modelovanie pridanej hodnoty v Anglicku sa spája s 90tymi rokmi minulého storočia. Na základe podnetov, že meranie PH by malo byť čo najjednoduchšie a najpriamejšie, ale zároveň čo najpresnejšie, bola v roku 1997 zostavená štúdia Fitz-Gibbonovou, skúmajúca možnosť uplatnenia meraní PH. Štúdia vyzdvihla 4 zásady, na ktoré by sa malo dbať pri zavádzaní celonárodného systému PH v Anglicku a to ľahkú zrozumiteľnosť, štatistickú exaktnosť, odmietnutie nadmernej záťaže škôl a zohľadnenie nákladovej efektívnosti systému. Aj keď sa prvé dve kritéria zdajú byť v rozpore, „rozsiahle štatistické overovania preukázali, že veľmi jednoduché a dostupné modely PH priniesli informácie v súlade s tými, ktoré boli získané pri komplexnejších analýzach.“ [Fitz-Gibbon, 1997, s. 11]

V roku 2002 boli prvýkrát verejne prezentované celonárodné merania PH, založenej na jednoduchej mediánovej metóde. Každá škola si mohla ľahko vypočítať PH priemerovaním PH svojich žiakov. PH žiaka predstavoval rozdiel jeho skutočnej úspešnosti na záverečnom testovaní oproti očakávanej

úspešnosti. Tá sa dala odčítať zo zverejnených mediánových tabuliek poskytujúcich informáciu o celonárodnom mediáne žiakov s rovnakými alebo podobnými výsledkami v predchádzajúcom testovaní [Ray, 2006].

Keďže mediánová metóda v skutočnosti nesledovala pokrok rovnakých žiakov, ale skôr kohort žiakov, prešlo sa v roku 2006 na modelovanie PH na báze viacúrovňových regresíí, pričom boli do modelu zahrnuté kontextuálne premenné ako pohlavie, vek, etnická skupina, materinský jazyk a ďalšie [Downess & Vindurampulle, 2007]. V súčasnosti je v Anglicku dostupný elektronický systém RAISEonline, ktorý školám poskytuje možnosť prístupu k hlbším analýzám ich PH. Výsledky tohto systému sú škálované do hodnôt v okolí 1000, pretože sa ukázalo vhodnejšie prezentovať národný priemer vo výrazne pozitívnych číslach v snahe odstrániť negatívny a demotivujúci dopad na menej úspešné školy.

1.5.2 Modely pridanej hodnoty v USA

Za prvý systém hodnotenia školstva na základe PH sa v USA považuje DVAAS (Dallas Value-Added Accountability System), ktorý bol uvedený do činnosti od roku 1984 s prestávkou neskôr v 80tych rokoch [Juščáková & Žoltak, 2013]. Tento systém zahŕňa dve fázy modelovania: v prvej fáze sa viacnásobnou regresiou určí efekt vhodného použitia premenných vzhľadom na relevantné charakteristiky žiakov, v druhej fáze sú následne tieto výsledky použité v lineárnom hierarchickom modeli s dvomi úrovňami zohľadňujúcimi študentov v rámci škôl [Webster & Mendro, 1997].

Najcitovanejším z amerických systémov PH je zrejme TVAAS (Tennessee Value-added Assessment System) známy aj ako Sandersov model. Profesor Sanders argumentuje, že efektivita učiteľa je silným prediktorom pokroku žiaka oproti iným faktorom, ako sú veľkosť triedy, socioekonomické pozadie a iné [Sanders & Horn, 1998]. Ponúka systém hodnotiaci príspevok učiteľa k výkonu žiaka sledovaním žiakovho pokroku voči svojim vlastným výsledkom za obdobie strávené v škole. Pre tento systém slúžia výsledky záverečných testovaní základných predmetov(matematika, veda, čítanie, cudzí jazyk a sociálne štúdiá) [Downess & Vindurampulle, 2007]. TVAAS systém možno považovať za základ pre EVAAS (Educational Value-Added Assessment System), komerčnú verziu systému pre hodnotenie edukácie vyvinutú profesorom

Sandersom v spolupráci s korporáciou SAS, viac Appendix 1.

Celonárodnou adaptáciou nariadenia „No child left behind“ z roku 2001 na podporu školstva bol v USA určený akýsi štandard pre hodnotenie výsledkov študentov, respektíve štáty boli vedené k potrebe rozvoja hodnotenia kvality vzdelávania, čo zvýšilo záujem o modelovanie PH. Veľa štátov USA dnes používa EVAAS systém aj keď len niekoľko z nich v celoštátnom rozsahu [Juščáková & Žoľtak, 2013].

1.5.3 Modely pridanej hodnoty v iných štátoch

V mnohých európskych štátoch je hodnotenie edukácie pomocou PH vo výskumnej fáze. Štáty ako Nórsko či Švédsko zverejňujú školské indikátory vrátane výsledkov národných testov. Keďže zverejňovanie výsledkov môže byť politicky problematické, je snaha obmedziť sporné aspekty. Napríklad v Nórsku sa výsledky malých škôl nezverejňujú a aby sa zabránilo náhodným výkyvom sú výsledky založené na informáciách z dvoch rokov [Ryška, 2009, s. 36]. V Poľsku inštitút IBE (Instytut Badań Edukacyjnych) vedie neoficiálny program „Educational value added in Poland“. Na výpočet PH jednej vlny žiakov sa používa viacnásobná regresia modelovaná z dát upravených kvantilovou normalizačnou metódou, trojročný indikátor PH školy je určený ako Bayesovský aposteriórny odhad náhodného efektu hierarchického modelu s náhodným priesečníkom pre dáta transformované IRT metódou (Item Response Theory) [Žoľtak, 2013]. V Maďarsku sa skúma využitie hierarchických modelov na výpočet PH aj so zahrnutím ESCS, pričom hodnoty žiakov chýbajúcich pri testovaní sú nahradené váženým priemerom triedy. V Austrálii „Council for Educational Research“ poskytuje školám správu zahŕňajúcu analýzy PH na základe školských dát s úmyslom pomôcť učiteľskému zboru s interpretáciou výsledkov.

2 Modely pridanej hodnoty

Druhá kapitola predstavuje modely PH použité v tejto práci. Cieľom práce je zamerať sa na porovnanie výsledkov modelov primárne hodnotiacich príspevkov školy k pokroku žiaka. Preto sa nevenujeme známym modelom stavaným na hodnotenie efektivity učiteľov (EVAAS, DVAAS modely) a modelom založeným na dostupnosti výsledkov úspešnosti žiaka v niekoľkých ročníkoch (modely analyzujúce rastovú krivku), keďže nami dostupné údaje ich aplikáciu neumožňujú. Detailný prehľad najpoužívanejších modelov na výpočet PH možno nájsť v publikáciách [Ryška, 2009, s. 106-115] a [OECD, 2008, s. 141-153].

2.1 Jednoduchý verzus komplexný model

S meraním PH sa spájajú protichodné vyjadrenia, že modely PH by mali byť „jednoduché“ a zároveň „komplexné“. Ako uvádza profesor Sanders [Sanders, 1999]: „Od modelov PH sa očakávala lepšia informácia pre dopyt verejnosti. Avšak ukázalo sa, že „lepšia informácia“ a „dopyt verejnosti“ sú nezlúčiteľné ak nie sú dostupné jednoduché a priamočiare merania PH. Jednoduchosť merania však ide proti jeho presnosti a naopak zložitosť a robustnosť modelu zamedzuje jeho ľahkú pochopiteľnosť. Preto hľadáme akýsi konsenzus medzi jednoduchosťou a komplexnosťou: spôsob ako poskytnúť čo najpresnejšie meranie kvality vzdelávania, ktoré by bolo ľahko pochopiteľné školami aj verejnosťou. Ako riešenie sa ponúka buď použitie najjednoduchších modelov za cenu ich robustnosti alebo implementácia zložitejších modelov sprevádzaných odbornou prípravou v školách ako aj verejnou informačnou kampaňou.

2.2 Základné modely

Vzhľadom k snahe o jednoduchosť modelov je primárnym záujmom tejto práce skúmať klasické lineárne regresné modely. Najprv analyzujeme tzv. „prostú“ PH ako najjednoduchší model a jej aplikáciu na dostupné dáta. Následne analyzujeme zahrnutie hierarchickej štruktúry do regresných modelov a tiež využitie modelov s pevnými efektami pre školy. Použitie uvedených modelov pramení z ich najčastejšieho využitia na meranie PH v krajinách OECD [OECD, 2008, s. 141]. V tejto časti predstavíme teoretické pozadie a predpoklady použitých modelov. Ich podrobný popis je dôležitý v zmysle

porozumenia pozadia a zo štatistického hľadiska netradičného použitia konceptu regresných modelov na hodnotenie výkonu školy.

2.2.1 Klasické lineárne regresné modely

Vo všeobecnosti sa klasické lineárne regresné modely používajú na skúmanie vzťahu medzi závislou a K nezávislými premennými. Za lineárne ich označujeme pre lineárnosť koeficientov určujúcich tento model. Teoretický zápis pre lineárnu regresiu je:

$$y = f(x_1, x_2, \dots, x_K) + \varepsilon, \quad (1)$$

kde y označuje závislú premennú a $x_1, x_2, \dots, x_K$ nezávislé premenné. Tieto premenné sú vopred určené dátami a na ich základe odhadujeme koeficienty pre lineárnu funkciu f . Disturbančný člen ε vyjadruje náhodnú chybu modelu. Tá vzniká z dôvodu, že nemôžeme očakávať, že zachytíme vstupnými premennými každý vplyv na výstup, akokoľvek zložitá bude štruktúra modelu. Zároveň ak sa snažíme modelovať veličinu, ktorá je z definície nemodelovateľná, disturbančný člen pokrýva nedokonalosť nášho modelu. Ďalší vplyv, ktorý do disturbančného člena vstupuje, je možná vzniknutá chyba merania. Regresný model (1) teda popisuje závislú premennú y ako súčet dvoch častí: deterministickej a náhodnej [Green, 2002, s. 8]. Treba si uvedomiť, že hodnotu náhodnej časti, disturbančného člena ε nepoznáme pre žiadne meranie. Jej hodnoty pre i -te meranie y_i preto odhadneme rezíduami e_i určenými ako rozdiel hodnôt $\hat{y}_i$ odhadnutých modelom a skutočných meraní y_i :

$$e_i = y_i - \hat{y}_i. \quad (2)$$

V terminológii edukometrie klasický lineárny regresný model vyjadríme ako:

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 f_{ij1} + \dots + b_n f_{ijn} + PH_{ij}, \quad (3)$$

kde index i označuje žiaka v rámci j -tej školy, $M1$ vstupné meranie, $M2$ výstupné meranie, a_0 konštantný člen, a_1 regresný člen vstupného merania, $b_1, \dots, b_n$ regresné členy kontextuálnych premenných, $f_1, \dots, f_n$ kontextuálne premenné a PH_{ij} rezíduá e_{ij} predstavujúce odhad náhodných odchýlok ε_{ij} . Rezíduá v tomto modeli predstavujú PH žiakov, keďže ich považujeme za zmenu v pokroku žiaka medzi dvoma testovaniami nepopísanú modelom. Túto zmenu pripisujeme vplyvu školy na pokrok žiaka. Z toho vyplýva, že

čím bude model dokonalejšie popisovať faktory predurčujúce žiakov pokrok mimo tých, ktoré sú ovplyvniteľné školou, tým väčšiu istotu budeme mať, že do PH určenej modelom vstupuje len vplyv školy na výsledky a nič iné.

PH školy j je následne vypočítaná ako priemer PH žiakov tejto školy:

$$\overline{PH}_j = \sum_{i=1}^{n_j} \frac{PH_{ij}}{n_j}. \quad (4)$$

Za prostú PH hodnotu budeme ďalej označovať PH určenú modelom zahŕňajúcim len vstupné a výstupné merania $M1$ a $M2$. Je to model popísaný rovnicou (3), pričom členy $b_1, \dots, b_n = 0$:

$$M2_{ij} = a_0 + a_1 M1_{ij} + PH_{ij}. \quad (5)$$

Existuje veľa metód na odhad regresných koeficientov. V našom prípade používame OLS metódu (metódu najmenších štvorcov). Podrobné vysvetlenie tejto metódy možno nájsť v literatúre [Green, 2002, s. 19-24]. Regresné koeficienty odhadujeme jej implementáciou v programe IBM SPSS Statistics 19.

2.2.1.1 Predpoklady klasického lineárneho regresného modelu

Konstruktívna forma klasických lineárnych regresných modelov a metódy odhadu ich koeficientov vo všeobecnosti vyžadujú testovanie niekoľkých predpokladov týkajúcich sa deterministického vzťahu medzi závislou a nezávislými premennými. Pre využitie tejto práce sme ich preformulovali do 7 predpokladov, ktorých prehľad možno vidieť v Tabuľke 1 [Green, 2002, s. 10]. Mnohé z nich sa vzťahujú na disturbančný člen modelu ε . Keďže v princípe skutočné hodnoty disturbancií ε_i nepoznáme, platnosť predpokladov v praxi overujeme kontrolou odhadov ε_i a teda rezíduí e_i .

P1 *Linearita modelu*. Predpoklad linearitý modelu znamená, že model sleduje lineárnu závislosť medzi vstupnými premennými a závislou premennou. To znamená, že nadobúda funkcionálnu formu:

$$y = \beta \mathbf{X} + \varepsilon, \quad (6)$$

kde $\mathbf{X}$ predstavuje maticu vstupných premenných a β vektor regresných koeficientov. V symbolike edukometrie $\mathbf{X} = (1, M1_{ij}, f_{ij1}, \dots, f_{ijn})$, $\beta = (a_0, a_1, b_1, \dots, b_n)$, $\varepsilon = PH_{ij}$ a $y = M2_{ij}$.

P1	Linearita
P2	Multikolinearita
P3	Nenáhodnosť matice nezávislých premenných
P4	Nulová stredná hodnota rezíduí
P5	Nezávislosť rezíduí
P6	Homoskedasticita
P7	Normálne rozdelenie rezíduí

Tabuľka 1: Predpoklady klasického lineárneho regresného modelu

- P2 *Multikolinearita*. Pre výpočet regresných koeficientov je potrebné, aby boli stĺpce matice $\mathbf{X}$ lineárne nezávislé, teda aby mala matica $\mathbf{X}$ plnú hodnotu. Multikolinearita označuje situáciu, kedy medzi niektorými nezávislými premennými vstupujúcimi do modelu existuje silná lineárna závislosť a teda nezávislé premenné sa navzájom vysvetľujú. V takomto prípade nemusí byť možné odhadnúť regresné koeficienty, respektíve ich odhady môžu byť veľmi nepresné. V prípade prostej PH, kedy do modelu vstupuje len jedna premenná nie je potrebné testovať tento predpoklad. Multikolinearitu budeme pri modelovaní testovať pomocou korelačnej analýzy premenných $M1_{ij}, f_{ij1}, \dots, f_{ijn}$ a na základe VIF kritéria, viac Appendix 2.
- P3 *Nenáhodnosť matice nezávislých premenných*. Nenáhodnosť matice $\mathbf{X}$ znamená, že táto matica popisujúca deterministickú časť modelu je generovaná mechanizmom nesúvisiacim s náhodnou časťou modelu ε . Tento predpoklad pri našich dátach považujeme za splnený.
- P4 *Nulová stredná hodnota rezíduí*. Popis závislej premennej pomocou regresného modelu sme všeobecne rozdelili na dve časti, deterministickú a náhodnú. Rezíduá zastávajú odhad náhodnej časti a preto musí platiť, že vzhľadom na deterministickú časť je ich stredná hodnota rovná nule, v opačnom prípade by to znamenalo, že niektoré z deterministických meraní nesie informáciu o náhodnej časti. Formálne možno túto podmienku vyjadriť ako:

$$E[PH_{ij}] = 0. \quad (7)$$

P5 *Nezávislosť rezíduí.* Predpoklad nezávislosti rezíduí možno formálne zapísať ako:

$$\text{Cov}[PH_{ij}, PH_{kl}] = 0 \text{ pre všetky } ij \neq kl. \quad (8)$$

Pri použití klasických regresných modelov na výpočet PH možno tento predpoklad označiť za problémový. Platnosť (8) by znamenala, že ak náhodne vyberieme akýchkoľvek dvoch žiakov z dostupnej vzorky, ich odchýlky od očakávaného výkonu budú navzájom nezávislé. Je však silno pravdepodobné, že správanie žiakov pochádzajúcich z jednej školy bude ovplyvnené prostredím tejto školy a teda očakávame akúsi závislosť medzi týmito žiakmi. Pre prípad klasických regresných modelov sa rozhodneme túto závislosť čisto ignorovať. Neskoršie použitie viacúrovňových modelov súvisí hlavne so snahou ošetriť túto nepresnosť.

P6 *Homoskedasticita.* Homoskedasticita označuje konštantnosť rozptylu rezíduí. Znamená to, že pre každé rezíduum platí:

$$D[PH_{ij}] = \sigma_{PH}^2. \quad (9)$$

Konštantnosť rozptylu rezíduí v tejto práci overujeme Levenovým testom [Nordstokke & Zumbo, 2010].

P7 *Normálne rozdelenie rezíduí.* Je vhodné predpokladať, že rezíduá pochádzajú z normálneho rozdelenia pre možnosť využitia parametrickej štatistiky pre ďalšie analýzy modelu. Spolu s predpokladmi [P4], [P5] a [P6] má platiť:

$$PH_{ij} \sim \mathcal{N}(0, \sigma_{PH}^2). \quad (10)$$

Predpoklad normálneho rozdelenia rezíduí sa často označuje za nepotrebný alebo redundantný. V našom prípade ho budeme vyžadovať pre možnosť konštrukcie niektorých štatistických testov vzťahujúcich sa na štatistickú významnosť regresie a hlavne určenie intervalu spoľahlivosti pre odhadované hodnoty. Normálne rozdelenie PH budeme testovať pomocou Kolmogorov-Smirnovho a Shapiro-Wilkovho testu. Pri takom veľkom počte dát s akými pracujeme, majú štatistické testy veľkú silu a sú náchylné zamietnuť normalitu rozdelenia už pri minimálnych odchýlkach. Preto budeme používať aj kritéria vizuálnej kontroly.

Za platnosti týchto 7 predpokladov sa dá ukázať, že odhad $\hat{\beta}$ z rovnice 6 metódou OLS je najlepší nevychýlený odhad s minimálnym rozptylom (MVUE) a teda žiaden iný nevychýlený odhad nemá menší rozptyl [Hansen, 2012].

2.2.1.2 Odľahlé hodnoty

Okrem overenia predpokladov je pri analýze regresného modelu potrebné skúmať prítomnosť odľahlých hodnôt. Prítomnosť odľahlých hodnôt môže spôsobiť vychýlenosť odhadovaných koeficientov modelu a aj neplatnosť požadovaných predpokladov. Je zložité určiť kritérium rozhodovania o odľahlosti hodnôt. Ak nevieme určiť, či je odľahlá hodnota chybou vzniknutou pri zaznamenávaní údajov alebo výsledok anomálie týkajúcej sa daného merania, nemala by byť táto hodnota automaticky vylúčená z analýzy dát. Ak bol dátový bod určený správne a nemáme žiadny teoretický dôvod ho odstrániť, mal by zostať v dátach a byť súčasťou našich záverov. V prípade pochybností by praktické riešenie malo zahŕňať spustenie analýzy s odľahlou hodnotou aj bez nej s následnou prezentáciou oboch výsledkov.

Na základe [Walters, 2011, s. 70] v tejto práci určujeme odľahlé hodnoty pomocou kritéria Cookovej vzdialenosti. Vplyv odľahlej hodnoty i testujeme pomocou testovacej charakteristiky D_i :

$$D_i = e_i^2 \cdot \frac{h_i}{2(1 - h_i)}, \quad (11)$$

kde e_i označuje i -te rezíduum a h_i k nemu prislúchajúce hodnoty odrážajúce vplyv hodnoty x_i na odhadnutú regresnú priamku, viac [Markechová & al., 2011]. D_i má za platnosti nulovej hypotézy Fisherovo rozdelenie pravdepodobnosti s $(2, n - 2)$ stupňami voľnosti. Ak platí $D_i > F_{0,05}(2, n - 2)$, hodnotu x_i budeme považovať za extrémnu na hladine významnosti 5%.

Popri testovaní štatistickej významnosti odľahlých hodnôt sme pri najjednoduchších modeloch kontrolovali ich prítomnosť vizuálnym zobrazením. Vzhľadom k tomu, že pri všetkých modeloch empirickej časti bola prítomnosť odľahlých hodnôt zamietnutá, nebudeme tieto testovania v práci uvádzať.

2.2.2 Hierarchické modely

Hierarchické modely inak označované ako zmiešané viacúrovňové lineárne modely sú modely, v ktorých parametre odhadnuté na jednej úrovni vystupujú ako premenné na ďalších úrovniach. Dôvody pre využitie hierarchických modelov možno rozdeliť na teoretické a štatistické. Teoretická motivácia pre použitie hierarchických modelov premení z toho, že niektoré sady dát majú

viacúrovňový charakter. V prípade školstva napríklad vieme, že žiaci tvoria akési podskupiny v rámci škôl. Motivácia pre použitie hierarchických modelov zo štatistického hľadiska pramení z toho, že sú zovšeobecnením klasickej lineárnej regresnej analýzy. Jednou zo základných požiadaviek regresnej analýzy je nezávislosť rezíduí [P5]. Hierarchické modely svojou štruktúrou túto požiadavku, ktorú často nemáme splnenú už len kvôli dizajnu výskumu, obchádzajú [Soukup, 2006, s. 989].

Najpoužívanejší hierarchický model pri hodnotení škôl je 2-úrovňový model s náhodným priesečníkom, ktorého teoretický zápis sa uvádza nasledovne:

$$y = B_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon, \quad (12)$$

$$B_0 = \beta_0 + u, \quad (13)$$

kde (12) predstavuje model pre 1. úroveň a (13) model pre 2. úroveň. Na druhej úrovni náhodne modelujeme priesečník z prvej úrovne pre jednotlivé podskupiny. Zlúčením rovníc (12) a (13) dostávame lineárny regresný model v tvare:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + u + \varepsilon, \quad (14)$$

kde $u + \varepsilon$ predstavuje náhodnú zložku modelu. Pre tento model platia tie isté predpoklady ako pre klasický lineárny regresný model, až na spomínaný predpoklad nezávislosti rezíduí [P5], keďže zložka u v sebe nesie informáciu o závislosti náhodnej časti modelu v rámci nadcelkov [Field, 2009]. Navyše musí platiť, že náhodné premenné sú navzájom nezávislé a takisto sú nezávislé od ostatných premenných v modeli, pričom sa zachováva predpoklad normálneho rozdelenia rezíduí [P7] a teda $u \sim \mathcal{N}(0, \sigma_u^2)$ a $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

V terminológii školstva chápeme 2 úrovne modelu ako rozdelenie žiakov v rámci škôl. Model formulujeme nasledovne:

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 f_{ij1} + \cdots + b_n f_{ijn} + u_j + PH_{ij}, \quad (15)$$

$$u_j \sim \mathcal{N}(0, \sigma_u^2), \quad (16)$$

$$PH_{ij} \sim \mathcal{N}(0, \sigma_{PH}^2), \quad (17)$$

kde u_j predstavuje efekt školy j na pokrok svojich žiakov, σ_u^2 prislúchajúcu štandardnú odchýlku, PH_{ij} pridanú hodnotu žiaka i v rámci svojej školy j

a ostatné premenné majú rovnakú interpretáciu ako v rovnici klasickej lineárnej regresie (3). V prípade hierarchického modelu teda pridanú hodnotu školy j určuje náhodná časť u_j z 2. úrovne modelu (13) a nie priemer PH žiakov tejto školy. Efekt školy v tomto prípade vypočítame pomocou empirického Bayesovského odhadu. Základom empirických Bayesovských odhadov je zahrnutie údajov všetkých škôl modelovanej vzorky pomocou tzv. „zrážania“ („shrinking“). Ich odhad určíme ako škálovaný priemer náhodnej zložky regresného modelu [Leckie & Goldstein, 2009]:

$$\hat{u}_j = S_j \cdot \bar{r}_j, \quad (18)$$

$$S_j = \frac{\sigma_u^2}{\sigma_u^2 + \frac{\sigma_{PH}^2}{n_j}}, \quad (19)$$

kde σ_u^2 a σ_{PH}^2 reprezentujú disperzie prislúchajúce k náhodným zložkám modelu z rovnice (15), $\bar{r}_j$ označuje priemer náhodnej časti modelu pre školu j a teda $\sum_{i=1}^{n_j} (\frac{u_j + PH_{ij}}{n_j})$ v zmysle značenia modelu (15). Škálujúci faktor, tzv. „shrinkage“ faktor S_j posúva priemer náhodnej časti smerom k 0, respektíve k priemeru všetkých škôl. Analogicky s rastúcim počtom žiakov školy sa hodnota „shrinkage“ faktora približuje k 1, keďže disperzia odhadu PH školy sa znižuje, čo dáva vyššiu váhu odhadom generovaným väčším počtom žiakov. Škálované odhady sú teda „konzervatívne“ v zmysle zohľadňovania nedostatku informácií pri školách zastúpených malým výberom žiakov. Viac k teoretickému pozadiu výpočtu empirických Bayesovských odhadov možno nájsť v Appendixe 3.

2.2.3 Modely s pevnými efektami

Model s pevnými efektmi školy je obmenou hierarchického regresného modelu vyjadreného rovnicou (15):

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 f_{ij1} + \dots + b_n f_{ijn} + u_j + PH_{ij},$$

pričom považujeme efekt školy u_j za pevný a nie náhodný parameter. Všeobecne rozoznávame dve alternatívy modelov s pevnými efektami: model s pevnými efektami vo forme dummy premenných a model s pevnými efektami diferencovanými v rámci skupín [Kim & Lalancette, 2013, s. 14-15]. Prvý model povoľuje rozdielnosť medzi školami formou rozdielneho konštantného

člena a_0 zahrnutím dummy premennej pre označenie každej školy:

$$M2_{ij} = a_{01} + a_{02}D_2 + a_{03}D_3 + \dots + a_{0n}D_n + \\ + a_1M1_{ij} + b_1f_{ij1} + \dots + b_nf_{ijn} + PH_{ij}, \quad (20)$$

kde D_i nadobúdajú hodnotu 1 pre školu i a 0 pre ostatné školy, a_{01} meria efekt školy 1 a a_{0j} pre $j = 2, \dots, n$ sú koeficienty prislúchajúce k dummy premenným a teda merajúce rozdiel medzi efektom školy j a referenčnou školou, ktorú predstavuje škola 1. Analogicky vypočítame PH škôl $j = 2, \dots, n$ nasledovne:

$$\hat{u}_j = a_{01} + a_{0j}. \quad (21)$$

Druhý model diferencuje efekty školy j odčítaním priemerných hodnôt premenných vstupných a výstupných meraní ($\overline{M1}_j, \overline{M2}_j$) a ostatných vstupných charakteristík ($\bar{f}_{jm}$, pre $m = 1, \dots, n$) žiakov školy j . Teoretická formulácia modelu s pevnými efektami diferencovanými v rámci skupín je:

$$(M2_{ij} - \overline{M2}_j) = a_0 + a_1(M1_{ij} - \overline{M1}_j) + b_1(f_{ij1} - \bar{f}_{j1}) + \\ + \dots + b_n(f_{ijn} - \bar{f}_{jn}) + (PH_{ij} - \overline{PH}_j). \quad (22)$$

Po odčítaní priemerných hodnôt školy od každého študenta vieme klasickou OLS metódou vypočítať hodnoty regresných koeficientov. PH školy určíme ako [Clark & al., 2010]:

$$\hat{u}_j = \overline{M2}_j - \hat{a}_1\overline{M1}_j - (\hat{b}_1\bar{f}_{j1} + \dots + \hat{b}_n\bar{f}_{jn}). \quad (23)$$

Viacúrovňové modely ako hierarchický model či modely s pevnými efektami školy pôsobia oproti klasickým lineárnym regresným modelom dôveryhodnejšie a komplexnejšie. Treba si ale uvedomiť, že „veľké a komplexné modely môžu vyzeráť realistickejšie, no len za cenu nestability. Nestabilita znamená, že malé zmeny v modeli ústia do veľkých zmien vo výsledkoch analýzy.“ [Kreft & Leeuw, 1998]. Koeficienty viacúrovňových modelov budeme v tejto práci odhadovať implementáciou ML metódy (metódy maximálnej vierohodnosti) v programe IBM SPSS Statistics 19 (viac o tejto metóde možno nájsť v literatúre [Green, 2002, kapitola 17]).

3 Empirický prístup

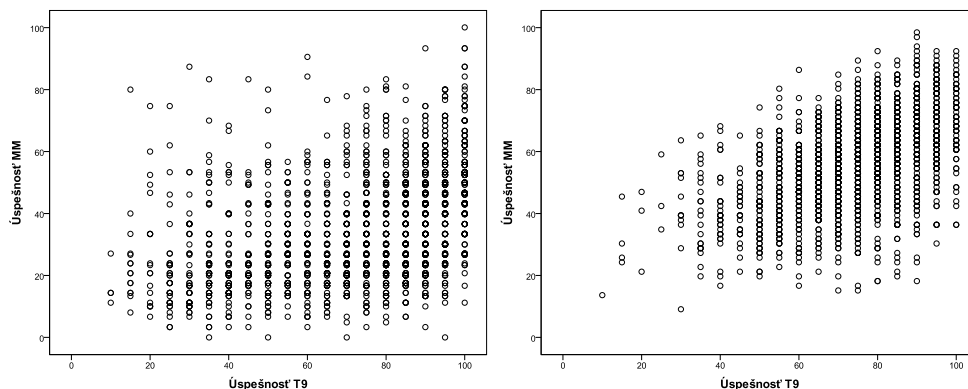
Dáta pre výskum modelovania PH máme k dispozícii z NÚCEM-u. V tejto kapitole predstavíme ich charakter ako aj analýzy a zvolenú metodiku spojenú s

ich využitím v rámci aplikácie modelov PH predstavených v predchádzajúcej kapitole.

3.1 Výber dát

K dispozícii máme dáta v dvoch predmetoch, v matematike (MAT) a v slovenskom jazyku a literatúre (SJL). Databáza pre MAT obsahuje údaje o 1669 žiakoch z 31 škôl, databáza pre SJL obsahuje údaje o 1895 žiakoch z 35 škôl. Obidve databázy obsahujú výsledky testovania T9 z roku 2010, výsledky testovania Malá maturita z roku 2013 a popisné charakteristiky žiakov, ktoré zahŕňajú typ školy (gymnázia GYM, stredné odborné školy SOŠ) a pohlavie (chlapci, dievčatá). Za výstupné merania teda neberieme výsledky EČMS, ktoré v čase písania práce pre túto vlnu žiakov ešte neboli k dispozícii (výsledky budú dostupné počas roka 2014), ale výsledky zo simulovanej maturity, tzv. Malej maturity (MM), písanej ku koncu tretieho ročníka stredných škôl s cieľom prípravy na maturitu. To znamená, že modely, ktoré skúmame v tejto práci, zachytávajú vplyv školy na pokrok žiaka za trojročné obdobie. Databázy neobsahujú žiakov so zdravotným znevýhodnením. K dispozícii máme aj databázu s rozličnými kontextuálnymi premennými pre obidva predmety, avšak často s neúplnými údajmi pre všetkých žiakov.

Prvotná štúdia dát navodzuje obavy o ich dostatočnú kvalitu a vhodnosť pre modelovanie PH jednoduchými regresnými modelmi. Všetky zobrazenia závislosti vstupných a výstupných meraní pripomínajú viac akúsi „hmlovinu“ ako útvar jednoducho modelovateľný regresiou (Obr. 1) a teda vizuálne pozorujeme veľmi slabú koreláciu medzi testovaniami T9 a MM, hlavne v MAT. Ako uvádza anglická štúdia: „Ak súčasne dostupné merania predchádzajúceho výkonu vysoko nekorelujú s nasledujúcim výkonom, nemá zmysel sledovať relatívny pokrok žiakov, t. j. pridanú hodnotu.“ [Fitz-Gibbon, 1997] V našom prípade sa hodnoty korelačného koeficientu pohybujú pod úrovňou 0,5, čo je výrazne menej v porovnaní napríklad s Anglickom, kde údaje vykazujú hodnoty od 0,52 do 0,87 [Kolková & Ivica, 2013]. Slabou koreláciou medzi vstupným a výstupným meraním vzniká veľký priestor pre nepresnosť, pričom nepresnosť regresného modelu absorbujú práve rezíduá, ktoré v našej terminológii predstavujú PH žiakov. Ako môžeme vidieť na Obr. 1 vľavo, najúspešnejší žiaci testovania T9 v MAT sa pri testovaní MM ocitli na celej škále úspešnosti. Tento fakt vyvoláva obavu o hodnovernosť výsledkov

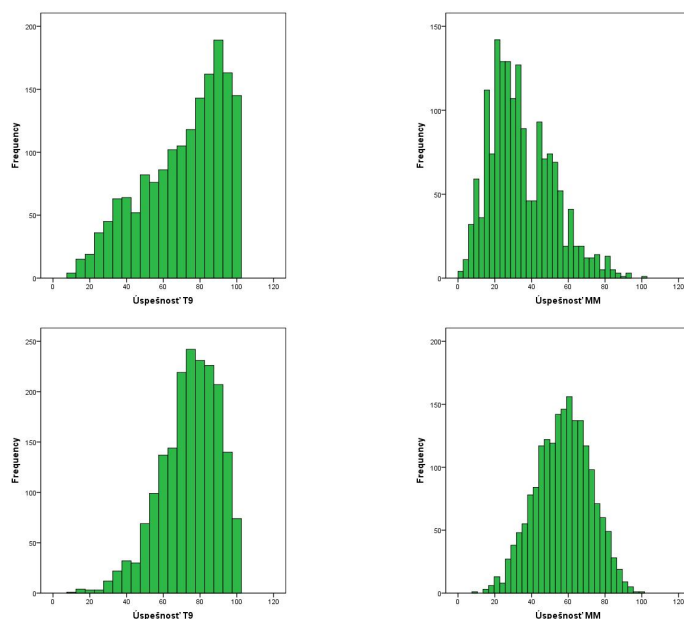


Obr. 1: Vzťah medzi úspešnosťou vstupných a výstupných meraní (MAT vľavo, SJL vpravo)

testovania T9. Tri z rôznych možných vysvetlení sú:

1. Za 3 roky na strednej škole sa mnohí pôvodne úspešní žiaci z nejakej príčiny vzdali vôle napredovať v predmete matematika (nevhodný učiteľ, zameranie sa na humanitné predmety...).
2. Testovanie T9 bolo ľahké, slabo diferencovalo žiakov, resp. nedostatočne odrážalo vedomosti žiakov ku koncu základnej školy. Tomuto faktoru napovedá aj rozdelenie testovania T9 (na Obr. 2), kde môžeme pozorovať nadmernú väčšinu žiakov v okolí vysokých hodnôt úspešnosti.
3. Pri testovaní T9 bolo možné opisovať, čo by mohlo vysvetliť fakt, prečo toľko žiakov s podpriemernými výsledkami v testovaní MM dosiahlo na testovaní T9 maximálne skóre.

Chceme tým poukázať na fakt, že k budúcim výsledkom modelov pre PH škôl treba pristupovať veľmi opatrne a treba byť veľmi obozretný pri ich interpretácii. Výrazne odlišné rozdelenia vstupných a výstupných výsledkov testovania na Obr. 2 ukazujú, že obťažnosť testu vystihujúca vedomosti žiakov môže výrazne kolísať. Preto ak na základe týchto dát jednorázovo určíme vhodný model PH, tento model nemusí byť pre výsledky iných ročníkov dostatočujúci kvôli inej obťažnosti vstupných a výstupných meraní.



Obr. 2: Rozdelenia vstupných (vľavo) a výstupných (vpravo) meraní pre MAT (hore) a SJL (dole)

3.2 Transformácia skóre rozdielnych skupín

Pri výsledkoch skóre z MM nastáva problém ekvivalentnosti variantov testu. Žiaci, aby nemohli opisovať, boli rozdelení do dvoch skupín A a B, pričom varianty testov A, B sa líšili v časti úloh, teda neboli identické. Z toho dôvodu je potrebné zistiť, či rovnaké skóre v skupinách A a B má porovnateľnú hodnotu. Štruktúra testu z MAT bola nasledovná: test mal 30 úloh, z ktorých 14 bolo identických pre oba varianty. Identické položky budeme ďalej označovať ako kotviace položky (K), pričom predpokladáme, že tieto položky sú hodnovernou vzorkou vyjadrujúcou obťažnosť celého testu.

Pri dátach z MAT rozdelenie výsledkov skupín A a B nepochádza z normálneho rozdelenia. Na testovanie ekvivalencie skupín použijeme neparametrický Mann-Whitney test využívajúci ranky hodnôt dvoch nezávislých vzoriek [McKnight & Najab, 2010]. Je zamietnutá hypotéza, že dáta pochádzajú z rovnakého rozdelenia (pri podobnej disperzii je medián v skupine A 10 naproti 8 v skupine B, výsledky testu vid' Tabuľka 2). Analogicky je nutné transformovať výsledky jedného testu na hodnoty v škále výsledkov druhého testu

na základe kotviacich otázok. To spravíme pomocou porovnávacích metód. Najčastejšie sa používajú lineárna porovnávacia metóda a ekvipercentilová metóda [Kalen, 1988]. V našom prípade sme sa rozhodli uprednostniť lineárnu porovnávaciu metódu pred ekvipercentilovou metódou, viac Appendix 4. V prípade nerovnosti disperzií kotviacich položiek sa pre lineárne porovnávanie používajú Levinova alebo Tuckerova porovnávacia metóda. Funkčná forma týchto metód prevádzajúca hodnoty skupiny B do škálovania v skupine A je vyjadrená nasledovne:

$$l_A(x) = \frac{\sqrt{\sigma^2(A) + \gamma^2 \cdot [\sigma_K^2(A) - \sigma_K^2(B)]}}{\sigma^2(B)}[x - \mu(B)] + \mu(A), \quad (24)$$

kde $l_A(x)$ je lineárna porovnávacia funkcia pre skupinu B v škále skupiny A, σ^2 označuje disperziu danej skupiny, σ_K^2 označuje disperziu v kotviacich položkách danej skupiny, μ označuje priemer zložiek danej skupiny a γ je tzv. „expansion factor“ zadaný rozdielne pre Tuckerovu a Levinovu metódu, viac v literatúre [Chen & al., 2010].

Ak by boli štandardné odchýlky kotviacich položiek v oboch skupinách A a B ekvivalentné, Levinova metóda by dávala rovnaké hodnoty ako jednoduché lineárne porovnávanie založené na zrovnaní štandardizovaného skóre oboch testov vo forme:

$$l_A(x) = \frac{\sigma(A)}{\sigma^2(B)}[x - \mu(B)] + \mu(A). \quad (25)$$

Na testovanie rovnosti disperzií kotviacich položiek slúži Levenov test homogenity disperzií. Klasický Levenov test je založený na predpoklade, že skupiny testovaných dát sledujú normálne rozdelenie. Pri neplatnosti tohto predpokladu používame neparametrickú alternatívu Levenovho testu využívajúcu ranky testovaných dát [Nordstokke & Zumbo, 2010]. Kolmogorov-Smirnov aj Shapiro-Wilkov test zamietli pôvod dát z normálneho rozdelenia. Analogicky Levenov neparametrický test dáva p-hodnotu 0.575 potvrdzujúc H_0 o rovnosti disperzií kotviacich položiek.

Pri dátach zo SJL mal test z MM 66 otázok, z ktorých 32 bolo identických pre skupiny A aj B. Neparametrický Mann-Whitney test porovnávajúci rozdielnosť mediánov rozdelení prijíma hypotézu H_0 , že rozdielnosť nie je štatisticky významná (Tabuľka 2) a preto nie je potrebné transformovať dáta.

	MAT	SJL
Mann-Whitney U	302723,5	432204,5
Wilcoxon W	655943,5	895407,5
Asymp. Sig. (2-tailed)	,000	,164

Tabuľka 2: Porovnanie rozdielov rozdelení variánt testov A a B

3.3 Kvantilová normalizácia dát

Pri modelovaní PH je často potrebné dáta prvotne upraviť transformáciami za účelom dosiahnutia lepších „psychometrických“ vlastností (zlepšenia lineárnosti vzťahu medzi vstupnými a výstupnými testovaniami). V našom prípade skúmame transformáciu dát pomocou kvantilovej normalizácie, keďže je to jednoduchá a v danej problematike často používaná metóda (Žoltak, 2013). Z formálneho hľadiska by bol vhodnejší odhad schopností žiakov pomocou IRT metódy, ktorá implicitne nepredpokladá, že všetky položky testov sú rovnakej obťažnosti, ale z praktického hľadiska sa na začiatku vývoja modelov PH odporúča používať najjednoduchšie transformácie. Vo všeobecnosti je vhodnejšie zvyšovať komplexitu metodológie postupne a IRT modely sú rádovo komplikovanejšie ako kvantilová normalizácia. Prejdime k samotnej definícii kvantilovej normalizácie.

Nech žiak x_i patrí do množiny homogénnej populácie žiakov X . Potom hodnotu transformovaného skóre U pre x_i vypočítame ako kvantil normálneho rozdelenia jeho percentuálneho umiestnenia v rámci populácie žiakov X :

$$U(X = x_i) = \Phi^{-1} \left(\frac{N(X \leq x_i) - \frac{N(X=x_i)}{2} - 0,5}{n} \right), \quad (26)$$

kde $N(X \leq x_i)$ reprezentuje počet žiakov, ktorých žiak i predbehol, $N(X = x_i)$ reprezentuje počet žiakov, ktorí mali rovnaké skóre ako žiak i a n reprezentuje celkový počet žiakov [Barnett, 1975].

Otázkou ostáva, či hrubé skóre zo vstupného a výstupného testovania, v našom prípade testovania T9 a MM, reprezentujú navzájom zmysluplnú mieru schopností žiakov, alebo bude ľahšie interpretovať transformované skóre žiakov. Po prvé, v psychometrii je dlhodobo zaužívaná tradícia predpokladať, že charakteristiky ako kognitívne schopnosti sledujú normálne rozdelenie v

rámci homogénnej populácie. Ak predpokladáme, že testy sú mierou skôr všeobecných schopností konceptualizovaných ako spojité premenné, mali by sme predpokladať aj ich pôvod z normálneho rozdelenia. V prípade, že hrubé skóre testov nepochádza z normálneho rozdelenia, javí sa transformácia ako žiaduca. Aj keď je kvantilová normalizácia neparametrická transformácia, ktorá nelineárne mení charakter dát, vychádzajúc z myslenia založeného na tradícií je táto zmena to, čo si dáta vyžadujú pre správnu interpretáciu výsledkov. Po druhé, ak vstupné aj výstupné testy merajú podobné, ideálne rovnaké charakteristiky, očakávame, že ich prepojenie bude nadobúdať akúsi lineárnu súvislosť. Ak transformované skóre vykazuje lineárnejší vzťah ako hrubé skóre, môžeme ho považovať v istom zmysle za vhodnejšie, zohľadňujúce rozdielnosť v charaktere testov. Prvé kritérium je vždy v prospech použitia transformácie. Lineárnosť pre druhé kritérium zistíme empiricky z dát. Vzhľadom k tomu, že definícia kvantilovej normalizácie zahŕňa použitie homogénnej populácie, budeme ju aplikovať zvlášť pre populáciu žiakov SOŠ a zvlášť pre gymnáziá.

		HS		QN		
	typ školy	R	Eta ²	R	Eta ²	N
MAT	SOŠ	,242	,296	,272	,309	1062
	GYM	,354	,403	,382	,407	833
SJL	SOŠ	,387	,400	,384	,401	1895
	GYM	,384	,400	,386	,403	1895

Tabuľka 3: Hodnoty koeficientov R a Eta² pre hrubé skóre a skóre transformované kvantilovou normalizáciou

V Tabuľke 3 môžeme vidieť Pearsonov korelačný koeficient R a koeficient Eta² vyjadrujúci mieru sily sledovaného vzťahu tzv. „effect size“, pre hrubé skóre a skóre transformované kvantilovou normalizáciou (QN) zvlášť pre populácie žiakov SOŠ a GYM z obidvoch predmetov. Korelačný koeficient R sa pohybuje okolo hodnoty 0,3, čo je na základe zaužívaného pravidla, tzv. „rule of thumb“ [Mukaka, 2012], akási hranica medzi slabou a zanedbateľnou koreláciu medzi premennými. Až na model pre SOŠ v SJL kvantilová normalizácia mierne zvyšuje linearitu vzťahu medzi vstupnými a výstupnými meraniami a vo všetkých prípadoch vplýva na nárast miery sily sledovaného

vzťahu tzv. „effect size“ (Eta^2) [Levine & Hullett, 2002]. Empiricky sa javí počiatočná úprava dát kvantilovou normalizáciou ako výhodná. Keďže táto transformácia je formou nelineárnej transformácie, je vhodné skúmať, či má jej použitie výrazný vplyv na zmenu výsledkov modelu oproti výsledkom rovnakého modelu s použitím netransformovaného skóre.

3.4 Metóda 1 – prostý regresný model

Najjednoduchší regresný model na meranie PH je prostý model určený rovnicou (6):

$$M2_{ij} = a_0 + a_1 M1_{ij} + PH_{ij}.$$

Výhoda modelovania prostej PH spočíva v jej jednoduchosti. Nenáročné výpočty dávajú ľahko interpretovateľné výsledky. Na druhej strane tak jednoduchý model nemusí dostatočne hodnoverne hodnotiť realitu. Ak sú však jeho výsledky konzistentné s výsledkami zložitejších modelov je model prostej PH najvhodnejším a najjednoduchším aplikovateľným modelom pre výpočet PH školy [Fitz-Gibbon, 1997].

Prostý model aplikujeme na merania testov vo forme úspešnosti. Súčasne skúmame výsledky prostého modelu pre rozdelené sady dát podľa typu školy a pohlavia a pri predmete MAT delíme dáta aj podľa toho, ktorí žiaci v 3. ročníku stredných škôl plánovali ísť maturovať z matematiky. Pre všetky modely, až na modely delené zvlášť pre SOŠ zo SJL a zvlášť pre chlapcov zo SJL, nie sú splnené požadované predpoklady lineárnej regresie. Najčastejšie ide o výchyľky rezíduí od normálneho rozdelenia [P7] a prítomnosť heteroskedasticity [P6] (výsledky testov pre nadmerný rozsah v práci neuvádzame).

V snahe o korekciu nesplnených predpokladov sa experimentálne pokúšame o transformáciu závislej premennej (výsledky z testovaní MM). Často sa stáva, že modelovaná premenná y je akýmsi spôsobom závislá na vlastnom priemere μ . Vtedy platí $E(y) = \mu$ a $Var(y) = g(\mu)$, pričom funkcia g je známa. Teda je potrebné nájsť vhodnú tzv. „variance-stabilizing“ transformáciu $z = f(y)$ tak, aby disperzia premennej z bola približne konštantná. Taylorov rozvoj 1. stupňa pre funkciu z v okolí bodu μ dáva nasledovnú rovnosť:

$$z = f(\mu) + (y - \mu)f'(\mu), \quad (27)$$

pričom predpokladáme, že $f'(y)$ v okolí bodu μ existuje. Pre strednú hodnotu a disperziu platí:

$$E(z) \approx f(\mu) + (E(y) - \mu)f'(\mu) = f(\mu), \quad (28)$$

$$Var(z) \approx Var(yf'(\mu)) = g(\mu)(f'(\mu))^2. \quad (29)$$

Keďže sa snažíme získať konštantnú varianciu (označme ju ako k^2), tak $k^2 = g(\mu)(f'(\mu))^2$. Odtiaľ po úprave dostávame vzorec pre „variance-stabilizing“ transformáciu vo forme [Walters, 2011]:

$$f(\mu) = k \int (g(\mu))^{-1/2} d\mu. \quad (30)$$

Teoreticky teda vieme vypočítať, že ak $g(\mu) = a\mu$ pričom a predstavuje konštantu, je vhodné použiť transformáciu $z = \sqrt{y}$ alebo ak $g(\mu) = a\mu^2$, pomôže transformácia $z = \ln y$. V praxi však zvyčajne nedokážeme určiť, či je v modeli prítomná závislosť premennej y na vlastnom priemere a takisto je ťažké určiť jej existenčnú formu. Zvykne sa preto skúšať niekoľko základných „variance-stabilizing“ transformácií. Najčastejšie je používaná Box-Cox trieda transformácií [Osborne, 2010]:

$$f_\lambda(y) = \begin{cases} (y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln y, & \lambda = 0 \end{cases} \quad (31)$$

pod ktorú patria aj formy logaritmickej ($\ln y$) mocninovej (y^2), reciprokálny ($1/y$) a odmocninovej ($y^{1/2}$) transformácie. V našom prípade sme sa obmedzili len na kontrolu týchto foriem transformácií. Treba si uvedomiť, že samotné transformácie prispievajú k funkcionálnej zložitosti modelu a preto sa neodporúča používať iné ako základné hodnoty parametra λ . Aj pri základných hodnotách λ ale strácame jednoduchosť простého modelu, keďže dostávame model:

$$f(M2_{ij}) = a_0 + a_1 M1_{ij} + PH_{ij}, \quad (32)$$

kde f predstavuje nelineárnu funkciu. Potom PH žiaka i z j -tej školy vieme určiť iba ako:

$$PH_{ij} = f(M2_{ij}) - a_0 - a_1 M1_{ij}. \quad (33)$$

V takomto prípade už lineárne nepopisujeme vzťah medzi výsledkami vstupných a výstupných testov, ale akúsi nelineárnu závislosť medzi týmito testami.

	MAT					SJL				
	HS	$y^{1/2}$	$\ln y$	$1/y$	y^2	HS	$y^{1/2}$	$\ln y$	$1/y$	y^2
Celok	,394	,411	,408	,194	,336	,505	,502	,492	,446	,498
GYM	,354	,359	,355	,294	,329	,384	,377	,367	,332	,39
SOŠ	,242	,26	,259	,1	,182	,387	,384	,377	,344	,381
Chlapci	,369	,384	,379	,181	,315	,445	,443	,437	,406	,437
Dievčatá	,45	,466	,461	,218	,393	,522	,518	,507	,442	,515
Maturant	,419	,441	,452	,411	,367	-	-	-	-	-

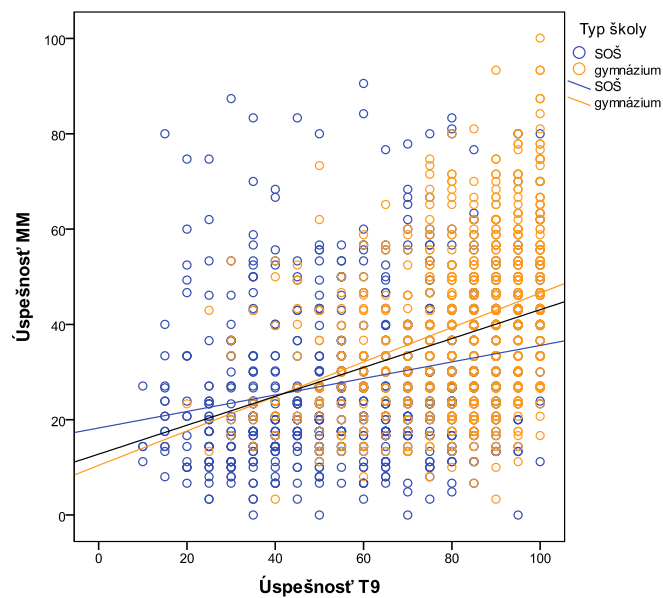
Tabuľka 4: Hodnoty korelačného koeficientu pre rôzne „variance stabilizing“ transformácie

Na základe Tabuľky 4 môžeme vidieť, že k linearite vstupno-výstupného vzťahu v prípade MAT najviac prispieva odmocninová transformácia našich dát. Len v niektorých prípadoch sa nám však podarilo korigovať nesplnené predpoklady [P6] a [P7] a v prípade SJL všeobecne transformácie neprispievajú k presnosti modelu. Prihliadnuc na tento fakt a vzhľadom k nelineárnosti „variance stabilizing“ transformácií sa nejaví ich použitie v rámci modelovania PH ako vhodné pre aktuálne dáta.

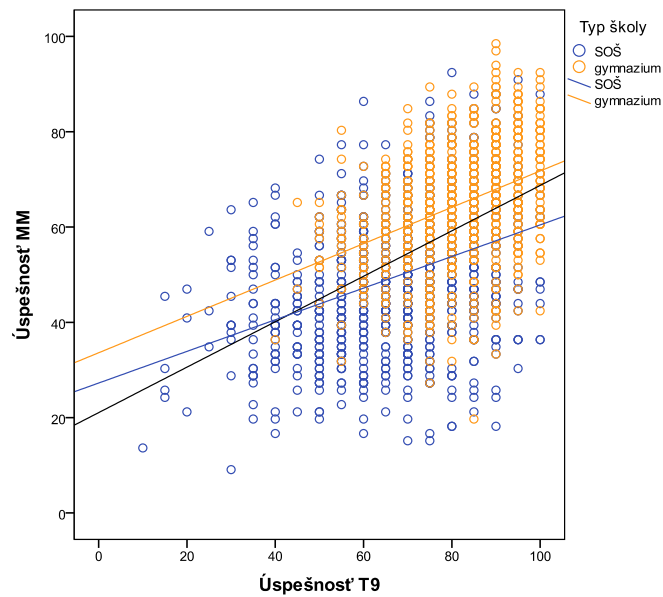
Neúplné výsledky pri použití hrubého skóre nás vedú k aplikácii prostého modelu na dáta rozdelené podľa typu školy a počiatočne upravené kvantilovou normalizáciou. Pre meranie prostej PH podľa typu školy zvlášť pre gymnáziá a SOŠ sú 2 dôvody:

1. potreba homogénnej populácie žiakov pre kvantilovú úpravu dát,
2. teoretická a empirická rozdielnosť týchto dvoch skupín.

Empirickú rozdielnosť pozorujeme priamo z dát. Na Obr. 3 a 4 sa regresné priamky pre jednotlivé typy škôl výrazne odlišujú. V prípade SJL pozorujeme rozdielnosť priesečníkov s osou y , v prípade MAT sú odlišné aj priesečníky aj sklon regresných priamok. Odlišnosť testujeme na základe významnosti regresných členov.



Obr. 3: Rozdelenie podľa typu školy (MAT)



Obr. 4: Rozdelenie podľa typu školy (SJL)

Pre výpočet odchýlok regresných koeficientov a ich štatistickej význam-

	MAT				SJL			
	B	Bias ^a	S. E. ^a	Sig. ^a	B	Bias ^a	S. E. ^a	Sig. ^a
(Constant)	18,261	-0,15	1,582	,001	26,24	-0,010	1,415	,001
Úspešnosť T9	0,173	0,002	0,025	,001	0,347	0,000	0,020	,001
Typ školy	-7,843	0,240	3,000	,012	10,208	-0,021	0,610	,001
Sklon	0,188	-0,003	0,041	,001				

a. Bootstrap results based on 1000 bootstrap samples

Tabuľka 5: Testovanie potreby rozdelenia dát podľa typu školy

nosti používame bootstrap metódu [Fox, 2002], keďže empiricky sa ukázal predpoklad normálneho rozdelenia rezíduí, na základe ktorého je odvodená testovacia štatistika pre odhadované parametre, ako neplatný. Bootstrap metóda je založená na nahradení vzorky výberu populáciou generovanou permutáciami s vrátením z pôvodnej vzorky. V prípade bootstrapu regresných koeficientov postupujeme podľa literatúry, pričom uvažujeme fixované nezávislé premenné, viac [Fox, 2008, s. 597]. V Tabuľke 5 vidno, že všetky testované koeficienty sú štatisticky významné, čo potvrdzuje empirickú potrebu modelovania prostej PH zvlášť pre gymnáziá a SOŠ. Po rozdelení máme 818 gymnazistov a 851 žiakov SOŠ pre MAT a 833 gymnazistov a 1062 žiakov SOŠ pre SJL.

Vzhľadom k regresným predpokladom sa javí použitie meraní upravených kvantilovou normalizáciou výhodnejšie ako využitie hrubého skóre pre modelovanie prostej PH, keďže v ich prípade sú všetky požadované predpoklady splnené, viac Appendix 5.

Iný metodologický prístup spočíva v použití percentilového poradia žiakov namiesto úspešnosti v testovaní. Využitie percentilového poradia nezohľadňuje veľkosť rozdielov medzi žiakmi a čisto určuje umiestnenie žiaka voči zvyšku populácie. Ako percentilové poradie pre žiaka i rozumieme percento ľudí s výsledkami ostro horšími ako sú výsledky tohto žiaka, teda všetkých, ktorých predbehol. Tabuľka 6 zobrazuje porovnanie korelácií medzi vstupnými a výstupnými meraniami pre hrubé skóre (HS) a percentilové poradie meraní (ptil). Všeobecne badať vyššiu koreláciu pri výsledkoch pre percentilové poradie. Kontrola regresných predpokladov pre prostý model PH s

	MAT		SJL	
	ptil	HS	ptil	HS
SOŠ	,292	,242	,379	,387
GYM	,375	,354	,387	,384

Tabuľka 6: Porovnanie lineárnosti vzťahu vstupných a výstupných meraní pre hrubé skóre a percentilové poradie

použitím percentilového poradia, ale v prípade celkového aj samostatných modelov pre obidva predmety (SJL aj MAT), poukazuje na prítomnosť heteroskedasticity a zamietá normalitu rozdelenia rezíduí. Súčasne sa ukazuje kvadratický trend vo vzťahu medzi vstupnými a výstupnými meraniami ako štatisticky významný. V záujme zachovania jednoduchosti a linearít modelu nebudeme ďalej analyzovať dáta vo forme percentilového poradia.

Tabuľky 7 a 8 zobrazujú Kendallovo tau [Abdi, 2007], korelačný koeficient vyjadrujúci mieru podobnosti poradia medzi výsledkami prostých modelov s použitím hrubého skóre (HS), skóre upraveného kvantilovou normalizáciou (QN) a percentilového poradia (ptil). Informačne pridávame aj koreláciu s výsledkami modelu s použitím percentilového poradia rozšíreného o kvadratický člen ($ptil^2$). Vo všetkých prípadoch pozorujeme úroveň korelácie nad 0,95. Pearsonov korelačný koeficient R nadobúda vo všetkých prípadoch hodnoty nad 0,985, čo znamená, že rozdiely v poradí sú minimálne v zmysle rozdielnosti hodnôt vypočítaných modelom. Použitie hrubého skóre alebo jeho úprav vo forme percentilového poradia či kvantilovej normalizácie našich dát ústí do veľmi podobných výsledkov pri modelovaní PH. Analogicky sa javí modelovanie prostej PH s dátami počiatočne upravenými kvantilovou normalizáciou ako najvhodnejšie, keďže v tomto prípade boli splnené základné predpoklady regresných modelov.

GYM					SOŠ			
	HS	QN	ptil	ptil ²	HS	QN	ptil	ptil ²
HS	1	,964	,964	,964	1	1,000	,964	,964
QN		1	1,000	1,000		1	,964	,964
ptil			1	1,000			1	1,000
ptil ²				1				1

Tabuľka 7: Hodnoty korelácie medzi výsledkami rôznych prostých modelov (MAT)

GYM					SOŠ			
	HS	QN	ptil	ptil ²	HS	QN	ptil	ptil ²
HS	1	1,000	1,000	1,000	1	,964	,952	,952
QN		1	1,000	1,000		1	,951	,890
ptil			1	1,000			1	,956
ptil ²				1				1

Tabuľka 8: Hodnoty korelácie medzi výsledkami rôznych prostých modelov (SJM)

3.5 Metóda 2 – viacnásobná regresia

Model viacnásobnej regresie je určený rovnicou (3):

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 f_{ij1} + \dots + b_n f_{ijn} + PH_{ij}.$$

Tento model všeobecne predstavuje obohatenie modelu prostej PH o rôzne charakteristiky alebo kontextuálne premenné f_{ij} . V našom prípade používame charakteristiky typ školy (SOŠ=0, GYM=1) a pohlavie (chlapec=0, dievča=1) vo forme dummy premenných a tiež ich interakcie so vstupným testovaním. Vzhľadom k neúplnosti databázy obsahujúcej kontextuálne premenné skúmame ich použitie osobitne. Výhodou modelov viacnásobnej regresie je ich relatívna jednoduchosť a zvýšená komplexnosť oproti modelom prostej PH, keďže nehomogénnosť v dátach možno popísať samotným modelom bez nutnosti ich delenia.

Model	MAT		SJL	
	R ²	AIC	R ²	AIC
(34)	0,180	9183,7	0,352	9419,2
(35)	0,165	9215,6	0,285	9606,4
(36)	0,203	9139,0	0,367	9376,1
(37)	0,187	9168,9	0,352	9421,5
(38)	0,162	9220,3	0,286	9603,2
(39)	0,210	9126,7	0,366	9379,2

Tabuľka 9: Výber modelu podľa kritérií R² a AIC

Použitím hrubého skóre vyjadreného formou úspešnosti skúmame modely:

$$M2_{ij} = a_0 + a_1M1_{ij} + b_1SCHOO L_j + PH_{ij} \quad (34)$$

$$M2_{ij} = a_0 + a_1M1_{ij} + b_1SEX_j + PH_{ij} \quad (35)$$

$$M2_{ij} = a_0 + a_1M1_{ij} + b_1SCHOO L_j + b_2SEX_j + PH_{ij} \quad (36)$$

$$M2_{ij} = a_0 + a_1M1_{ij} + b_1SCHOO L_jM1_{ij} + PH_{ij} \quad (37)$$

$$M2_{ij} = a_0 + a_1M1_{ij} + b_1SEX_jM1_{ij} + PH_{ij} \quad (38)$$

$$M2_{ij} = a_0 + a_1M1_{ij} + b_1SCHOO L_jM1_{ij} + b_nSEX_jM1_{ij} + PH_{ij} \quad (39)$$

kde premenná *SCHOOL* označuje typ školy a premenná *SEX* pohlavie žiaka. Modely (34), (35) a (36) ponímajú posun medzi regresnými priamkami pre dané charakteristiky a modely (37), (38) a (39) s interakčnými členmi zas rozdielnosť sklonov pre tieto priamky. Kombinácia samotných dummy premenných s k nim prislúchajúcimi interakčnými členmi spôsobuje multokolinearitu (hodnoty VIF kritéria okolo 20) a preto tieto modely vynechávame. Najvhodnejší model pre jednotlivé predmety určíme na základe Akaikeho informačného kritéria (AIC). Toto kritérium odvodené z informačnej teórie vyvažuje presnosť modelu a penalizáciu za nárast počtu parametrov. Za najoptimálnejší model sa považuje model s najnižšou hodnotou AIC počítaného ako:

$$AIC = -2 \ln L(\hat{\theta}) + 2k, \quad (40)$$

kde k predstavuje počet odhadovaných koeficientov modelu a $L(\hat{\theta})$ maximalizovanú funkciu vierohodnosti [Bozdogan, 2000]. Tabuľka 9 zobrazuje hodnoty

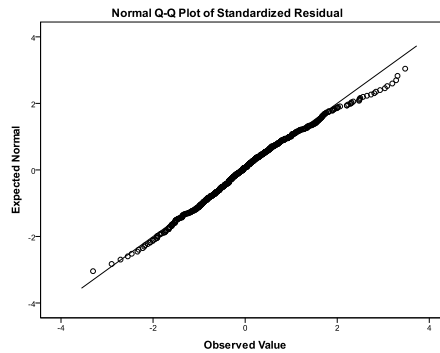
AIC pre jednotlivé modely spolu s hodnotami R^2 . Na základe uvedených kritérií sa javí model s typom školy a pohlavím ako dummy premennými (36) za najvhodnejší pri modelovaní SJL. Pre tento model sú zároveň splnené všetky regresné predpoklady. Aj keď sa v prípade MAT ukazuje model s oboma interakciami vstupného testovania a dummy premenných (39) ako najvhodnejší, pre tento model ani ostatné modely v MAT nie sú splnené predpoklady normálneho rozdelenia rezíduí a homogenity ich disperzií. Analogicky dáta počiatočne upravíme kvantilovou normalizáciou tak ako v prípade prostého modelu. Sady dát rozdelíme podľa typu školy vzhľadom k potrebe homogénnej populácie pre kvantilovú normalizáciu. Skúmame modely:

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 SEX_j + PH_{ij}, \quad (41)$$

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 SEX_j M1_{ij} + PH_{ij}, \quad (42)$$

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 SEX_j + b_2 SEX_j M1_{ij} + PH_{ij}, \quad (43)$$

pričom zahrňame aj model s obidvoma interakčnými členmi (43), keďže v tomto prípade neústi v prítomnosť multikolinearity. Rovnako ako v prípade prostej regresie modely s použitím dát upravených kvantilovou normalizáciou už modely (41), (42), (43) spĺňajú základné regresné predpoklady (v prípade SOŠ v MAT testy zamietajú normalitu rozdelenia rezíduí, ale ako môžeme vidieť na Obr. 5, tieto výchylky nie sú veľké). Na základe kritérií R^2 a AIC (Tabuľka 10) sa ukazuje model s dummy premennou pohlavie ako najvhodnejší.



Obr. 5: Zobrazenie kvantilov rozdelenia rezíduí k očakávaným hodnotám kvantilov normálneho rozdelenia - SOŠ MAT

Model	MAT				SJL			
	R^2	AIC	R^2	AIC	R^2	AIC	R^2	AIC
(41)	,431	2800	,396	2227,7	,299	2330	,415	2170,5
(42)	,385	2847,1	,391	2229,9	,275	2342,8	,382	2196,1
(43)	,431	2801,2	,391	2229,8	,299	2332	,415	2172,5

Tabuľka 10: Kritériá R^2 a AIC pre rôzne viacnásobné modely

3.5.1 Kontextuálne premenné

Databáza s kontextuálnymi premennými obsahuje výstupy odpovedí na dotazníky a merania rôznych schopností žiakov. Sú to údaje o všeobecných a priestorových schopnostiach žiakov, študijných predpokladoch, motivácii študentov, socioekonomický a kultúrny status žiaka (ESCS) a rôzne iné. Je zložitá posúdiť, ktoré kontextuálne premenné sú naozaj vhodné pre modelovanie PH. Škola nesmie mať za merané obdobie na tieto premenné žiaden vplyv, inak by koncept rezíduí nepokrýval celý príspevok školy k pokroku žiaka. Prvotný výber kontextuálnych premenných vykonávame na základe dvoch kritérií:

1. Premenná nie je ovplyvniteľná pôsobením školy za merané obdobie.
2. Premenná aspoň minimálne koreluje s úspešnosťou žiakov v testovaní MM.

Po zohľadnení týchto kritérií ďalej analyzujeme premenné súvisiace so vstupnými testovaniami všeobecných schopností a študijných predpokladov a s ESCS, pre ktoré máme po odstránení chýbajúcich údajov k dispozícii databázy s nasledovnými počtami: 474 žiakov SOŠ a 582 žiakov GYM v MAT a 625 žiakov SOŠ a 591 žiakov v SJL. Všeobecne sa zvyknú používať rôzne formy nahrádzania chýbajúcich údajov, ale vzhľadom k faktu, že chýbajúce údaje predstavujú viac ako tretinu pôvodnej databázy, ich použitie by nebolo efektívne. Na druhej strane môže odstránenie chýbajúcich údajov viesť ku skresleným odhadom vzhľadom k možným systematickým rozdielom medzi žiakmi s chýbajúcimi hodnotami a ostatnými žiakmi [Kim & Lalancette, 2013, s. 33]. Uvedené kontextuálne faktory sú vyjadrené pomocou 16 subindexov, z ktorých vyberáme vhodné nezávislé premenné na základe testovania vnorených modelov (viac [Walters, 2011, kapitola 2]). Empiricky sa

ukazuje zložité jednoznačne určiť indexy všeobecne vhodné pre modelovanie PH, vzhľadom k faktu, že štatisticky významné indexy z jedného modelu sú nadbytočné v ďalšom modeli. Analogicky môžu platiť niektoré z tvrdení:

1. Merania kontextuálnych premenných nie sú dostatočné, resp. nehodnoverne odrážajú realitu.
2. Kontextuálne charakteristiky sú v niektorých prípadoch odzrkadlené už vo vstupnom testovaní (k rovnakému záveru prichádzajú Ballou, Sanders a Wright [Ballou & al., 2004]).
3. Miera závislosti výsledkov testovaní na kontextuálnych premenných sa líši pre rozdielne typy škôl (k rovnakému záveru prichádza aj Ryška pri zohľadnení ESCS [Ryška, 2009, s. 56-61]).
4. Uvedené kontextuálne charakteristiky v skutočnosti nemajú vplyv na výkon žiakov v jednotlivých testovaniach.

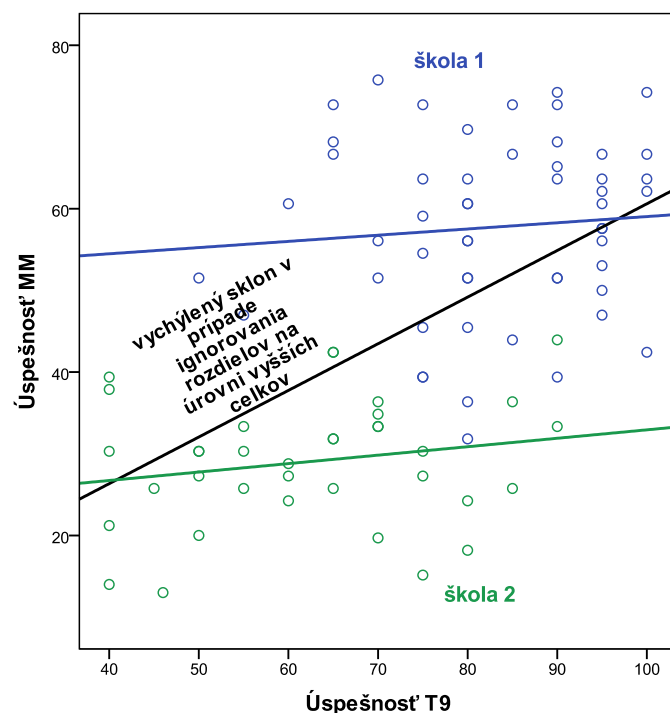
Bez ohľadu na aktuálne dáta, zvolené kontextuálne premenné evidentne súvisia s kvalitou vzdelávacieho procesu, čo môže spôsobovať korelovanú disperziu odhadov PH škôl (rozpor s [P5]) a preto by mala byť táto oblasť podrobená hlbšiemu výskumu.

3.6 ICC koeficient

Nezávislosť rezíduí [P5], jeden zo základných predpokladov lineárnej regresie, je predpoklad, ktorého platnosť môžeme spochybníť na základe očakávania prítomnosti hierarchickej štruktúry v modelovaných dátach. Je logické predpokladať, že prostredie každej školy a iné vplyvy, ktoré nevieme definovať, budú spôsobovať rozdielnosť medzi výsledkami žiakov rozdielnych škôl. Príklad, ako hodnoty priemerných úrovní premenných líšiacich sa na úrovni vyšších celkov môžu spôsobiť vychýlenosť odhadnutých koeficientov regresných modelov, zobrazuje Obr. 6. Na meranie podielu medziskupinovej disperzie oproti vnútroskupinovej disperzii slúži tzv. vnútrotriedny koeficient korelácie, všeobecne známy ako Intraclass correlation coefficient (ICC). ICC vypočítame ako:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}, \quad (44)$$

kde σ_e^2 prislúcha úrovni rozptylu jednotlivcov a σ_u^2 úrovni rozptylu vyšších celkov [Soukup, 2006, s. 994]. Výrazné hodnoty tohto koeficientu indikujú

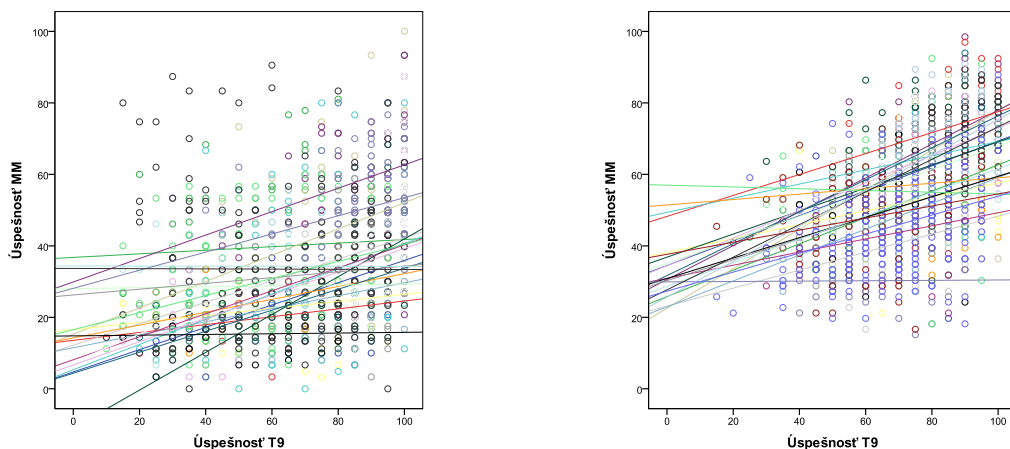


Obr. 6: Vychýlenie regresnej priamky spôsobené rozdielnosťou skupín na úrovni vyšších celkov

potrebu zahrnutia hierarchickej štruktúry do modelovania PH. Pre výpočet hodnôt ICC skúmame tzv. „NULL“ model (najjednoduchší hierarchický model):

$$M2_{ij} = \beta_0 + u_j + e_{ij}, \quad (45)$$

kde u reprezentuje náhodnú zložku na úrovni skupín a e na úrovni jednotlivcov. Rozptyl týchto zložiek vstupuje do výpočtu ICC. Z výpočtov ICC vylúčime školy s počtom žiakov menším ako 30 z dôvodu, že pri malom výbere žiakov majú extrémne merania nadmerný vplyv na výsledky [Kolková & Ivica, 2013]. Vypočítané hodnoty ICC určujú, že v prípade MAT 32 % a v prípade SJL 44 % rozdielov vo výkonoch možno pripísať rozdielom medzi jednotlivými školami a nie rozdielom v rámci populácie žiakov. Rozdiely v sklonoch a priesečníkoch regresných priamok škôl poukazujúce na potrebu modelovania PH pomocou hierarchických modelov môžeme jasne vidieť na Obr. 7. V zmysle neskoršieho delenia dát na SOŠ a GYM počítame ICC aj zvlášť pre tieto súbory. Výsledky zobrazené v Tabuľke 11 potvrdzujú potrebu



Obr. 7: Regresné priamky po odstránení malých škôl (MAT vľavo a SJL vpravo)

	GYM	SOŠ
MAT	35%	16%
SJL	19%	32%

Tabuľka 11: ICC pre dáta rozdelené podľa typu školy

zahrnutia hierarchickej štruktúry do modelovania PH.

3.7 Metóda 3 – hierarchický model s náhodným priesečníkom

V rámci hierarchických modelov sa v tejto práci zaoberáme dvojúrovňovým modelom s náhodným priesečníkom určeným rovnicou 15:

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 f_{ij1} + \dots + b_n f_{ijn} + u_j + PH_{ij}.$$

Dôvod, prečo do modelovania nezahrňame viac úrovní, je nadmerná zložitosť takýchto modelov. Teoreticky vieme, že hierarchická štruktúra škôl predstavuje združenie žiakov v rámci tried a tried v rámci škôl. Nás ale v zmysle konceptu merania PH školy nebude zaujímať rozdielnosť podskupín žiakov v rámci škôl, aj keď aj tie môžu byť objektom pozorovania (napríklad v rámci anglického systému RAISEOnline [Ray, 2006]).

	MAT		SJL	
Model	SOŠ	GYM	SOŠ	GYM
(41)	2240,9	1846,0	2564,4	2117,2
(42)	2258,0	1880,4	2568,9	2118,1
(43)	2242,6	1848,0	2566,4	2117,3

Tabuľka 12: AIC pre modely definované na s. 43 s náhodne modelovaným priesečníkom a_0

Modely konštruujeme zvlášť pre SOŠ a GYM z dôvodu teoretických požiadaviek na počiatočnú úpravu dát kvantilovou normalizáciou (nesplnené základné regresné predpoklady) a jednoduchosti porovnávania s ostatnými modelmi. Na základe AIC kritéria zobrazeného v Tabuľke 12) vyberáme ako v prípade viacnásobnej regresie model s charakteristikou pohlavie žiaka:

$$M2_{ij} = a_0 + a_1 M1_{ij} + b_1 SEX_j + u + PH_{ij}. \quad (46)$$

Už spomínaná výhoda hierarchických modelov spočíva v presnosti odhadovania efektov škôl aj pri malej vzorke dát vďaka „zrážaniu“ odhadov k priemeru populácie, čím sa zvyšuje ich štatistická efektivita. Na druhej strane si treba uvedomiť, že toto „zrážanie“ môže niekedy spôsobiť výchylky v modeli [Kim & Lalancette, 2013, s. 18]. Preto je vhodné porovnávať odhady hierarchických modelov s modelmi s pevnými efektami kde uvedený problém nenastáva.

3.8 Metóda 4 – model s pevnými efektami

Ako napovedá už samotný názov, modely s pevnými efektami považujú PH školy za pevné parametre na rozdiel od hierarchických modelov, kde sa očakávajú PH školy vo forme náhodných premenných s normálnym rozdelením. Na druhej strane sa odhady efektov škôl týchto modelov oproti odhadom hierarchických modelov môžu z roka na rok výrazne líšiť, keďže nevyužívajú proces tzv. „zrážania“ [OECD, 2008, s. 145]. V práci používame model s pevnými efektami vo forme dummy premenných vyjadrených rovnicou (20):

$$M2_{ij} = a_{01} + a_{02}D_2 + a_{03}D_3 + \dots + a_{0n}D_n + \\ + a_1 M1_{ij} + b_1 f_{ij1} + \dots + b_n f_{ijn} + PH_{ij},$$

pričom pre model platia rovnaké predpoklady ako v prípade klasických lineárnych regresných modelov. Sadu dát delíme podľa typu školy pre možnosť

	MAT		SJL	
Model	SOŠ	GYM	SOŠ	GYM
(41)	2216,2	1810,4	2524,5	2094,4
(42)	2234,2	1845,4	2527,9	2095,3
(43)	2217,8	1812,3	2526,5	2094,2

Tabuľka 13: AIC pre modely definované na s. 43 s pevnými efektami vo forme dummy premenných

počiatočnej úpravy dát kvantilovou normalizáciou, keďže sa ukázal základný predpoklad normality rezíduí [P7] ako neplatný. Kim a Lalancette argumentujú [Kim & Lalancette, 2013], že prítomnosť veľkého počtu dummy premenných v modeli vytvára možnosť prítomnosti multikolinearity [P2], v prípade našich dát sa táto skutočnosť ale nepotvrdila. Analogicky vyberáme vstupné charakteristiky modelu na základe AIC. Najvhodnejší sa ako v predchádzajúcich prípadoch javí model s charakteristikou o pohlaví žiaka (na základe hodnôt AIC v Tabuľke 13 je pre GYM v SJL vhodnejší model s dodatočným zahrnutím interakčného člena, ale pre jednoduchosť a malú rozdielnosť výsledkov tento fakt zanedbávame).

4 Interpretácia a porovnanie výsledkov

Problematika správnej interpretácie výsledkov PH je veľmi citlivou oblasťou. Je dôležité, aby numerické hodnoty edukometrie nezakryli ich teoretické pozadie. Pri nerozvážnej publikácii akýchkoľvek výsledkov sa môže ľahko stať, že sa na základe nesprávneho pochopenia meraní vyvodí rozhodnutia, ktoré budú mať negatívny dopad na školstvo. Ako uvádza Raudenbush: „Výsledky modelovania PH môžu verejne slúžiť nanajvýš záujmu rodičov pri výbere školy a nie funkcionárom vyvodzujúcim zodpovednosť učiteľov a škôl za vzdelávaciu prax. Nemožno poprieť, že výsledky modelovania PH v kombinácii s ďalšími informáciami majú potenciál iniciovať potrebnú verejnú diskusiu ako zlepšiť vzdelávanie. Avšak nemôžu slúžiť ako priamy dôkaz vplyvu vzdelávacej praxe.“ [Raudenbush, 2004, s. 128]

4.1 Školy s malým výberom žiakov

Zohľadniť zastúpenie škôl veľkosťou výberu ich žiakov pred porovnávaním výsledkov je veľmi dôležité. Hodnotiť príspevok školy k pokroku žiakov na základe výberu 10 žiakov zo školy s populáciou cez 1000 žiakov by zrejme nebolo objektívne a mohli by sme očakávať vychýlenosť meraní od skutočných hodnôt. Tento fakt implikuje potrebu spodnej hranice pre minimálny počet žiakov potrebný pre merania. Všeobecne je pri interpretácii výsledkov PH zaužívané pravidlo, že za hodnovernú vzorku žiakov pre školu sa pokladá výber aspoň 30 žiakov [Kolková & Ivica, 2013]. V prípade dostupnej vzorky máme po odstránení škôl s malým výberom žiakov v SJL odhady pre 25 škôl (14 SOŠ a 11 GYM) a v MAT odhady pre 22 škôl (11 SOŠ a 11 GYM). Dôvod, prečo sme nevenovali pozornosť veľkosti výberu žiakov škôl pred výpočtom PH žiakov, vychádzal pri lineárnych regresných modeloch z predpokladu nezávislosti rezíduí [P5]. Tento predpoklad je ekvivalentný s homogenitou populácie žiakov a teda každé jedno meranie bolo relevantné pri odhade koeficientov regresie celej populácie. V prípade hierarchických modelov, kedy tento predpoklad nebol splnený, zohľadňujeme veľkosť výberu zrážaním odhadov PH škôl k priemeru všetkých meraní.

4.2 Porovnanie bodových odhadov jednotlivých metód

Na obr. 8 a obr. 9 môžeme vidieť zobrazenie bodových odhadov PH škôl vypočítaných jednotlivými modelmi. Z dôvodu použitia skóre počiatočne upraveného kvantilovou normalizáciou nadobúdajú odhady hodnoty z intervalu $\langle -2, 2 \rangle$. Zaužívaným krokom pri verejnej publikácii odhadov je zmena škály výsledkov do nezáporných čísel (v Anglicku sú výsledky centrovane okolo hodnoty 1000 [Ray, 2006, s. 30]). Tento krok, ktorý vynechávame, sa opodstatňuje argumentom, že záporné čísla môžu byť verejnosťou nesprávne vyložené v zmysle zhoršovania sa progresu daných škôl. Treba si však uvedomiť, že bez ohľadu na použitý model je veľkosť číselných hodnôt odhadov irelevantná, zaujímavá je ich rozdielnosť voči ostatným odhadom. Silu súvislosti medzi jednotlivými odhadmi rôznych modelov porovnávame pomocou Kendallovho tau. Až na SOŠ v SJL korelácia nadobúda hodnoty nad 0,9, čo hovorí o relatívnom zachovaní poradia v rámci súboru odhadov škôl. V prípade SOŠ v SJL, kde máme k dispozícii viacej odhadov (14 oproti 11 v ostatných prípadoch) a preto by sme očakávali väčšiu stabilitu výsledkov, model viacnásobnej regresie vykazuje zmeny oproti ostatným modelom. Ako

neskôr uvidíme, tieto zmeny nie sú štatisticky významné, avšak upozorňujú na fakt, že interpretácia výsledkov čisto na základe bodových odhadov by mohla pre jednotlivé modely vyústiť do rozličných záverov.

4.3 Výpočet intervalov spoľahlivosti

Bodové odhady PH škôl môžu byť často nepresné a preto sa zvykne výsledky prezentovať formou rozdielnosti intervalov spoľahlivosti (IS) pre vypočítané odhady od národného priemeru (priemer všetkých odhadov) [OECD, 2008, s. 60-61]. Kritika IS sa spája s mylnou interpretáciou týchto intervalov. Ak zvolíme hladinu významnosti $\alpha = 0,05$, potom $100(1 - \alpha)\%$ IS pre odhad PH školy nevyjadruje 95% pravdepodobnosť, že skutočná PH školy padne do daného intervalu, ale istotu, že ak by sme opakovane vykonávali odhady PH škôl pre rôzne sady dát popisujúce tieto školy, tak v 95% prípadov by IS pokryl neznámy parameter, teda skutočnú PH školy.

Pre modely prostej PH a modely s použitím viacnásobnej regresie počítame PH škôl podľa rovnice (4):

$$\overline{PH}_j = \sum_{i=1}^{n_j} \frac{PH_{ij}}{n_j}.$$

Na základe [P7] platí $PH_{ij} \sim \mathcal{N}(0, \sigma_{PH}^2)$. Analogicky z pravdepodobnostného rozdelenia ľahko odvodíme interval spoľahlivosti na α hladine významnosti:

$$\overline{PH}_j \pm u_{(1-\frac{\alpha}{2})} \frac{\sigma_{PH}}{\sqrt{n_j}}, \quad (47)$$

kde $u_{(1-\frac{\alpha}{2})}$ predstavuje hodnotu $100(1 - \frac{\alpha}{2})\%$ kvantilu normálneho rozdelenia.

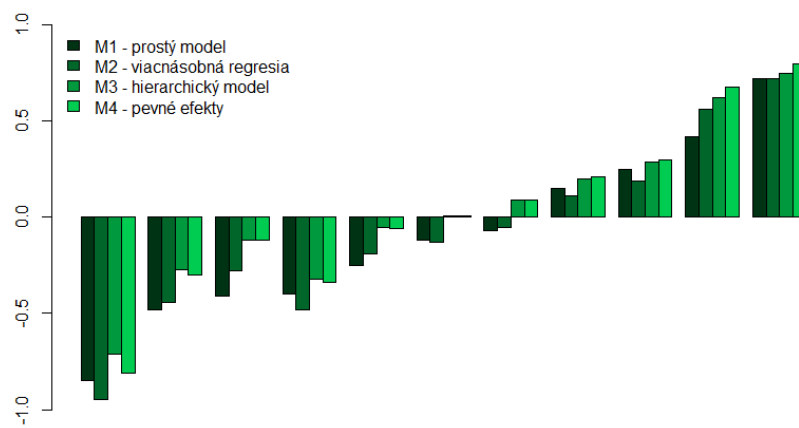
Pre 2-úrovňový hierarchický model s náhodným priesečníkom očakávame, že pre posteriórne Bayesovské rozdelenie platí (Appendix 3):

$$u|\bar{r}_j \sim \mathcal{N}(S_j \bar{r}_j, S_j \frac{\sigma_{PH}^2}{n_j}) = \mathcal{N}(S_j \bar{r}_j, \frac{1}{\frac{n_j}{\sigma_{PH}^2} + \frac{1}{\sigma_u^2}}).$$

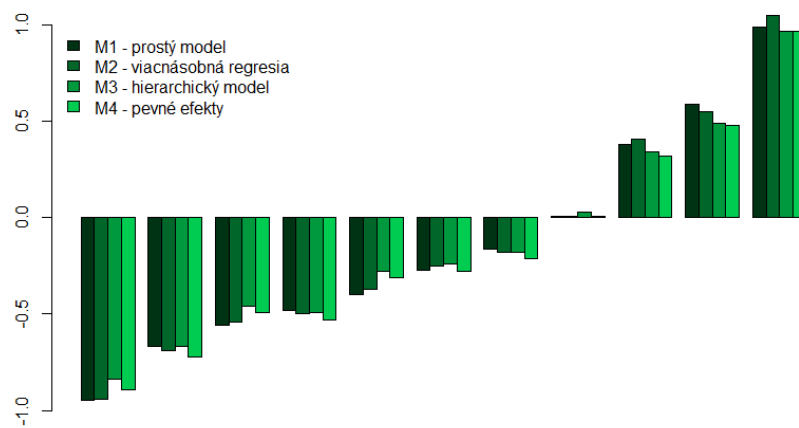
Teda na α hladine významnosti počítame interval spoľahlivosti pre BLUP odhad PH škôl $\hat{u}_j$ ako [Leckie & Goldstein, 2009, s. 839]:

$$\hat{u}_j \pm u_{(1-\frac{\alpha}{2})} \sqrt{\frac{1}{\frac{n_j}{\sigma_{PH}^2} + \frac{1}{\sigma_u^2}}}. \quad (48)$$

SOS

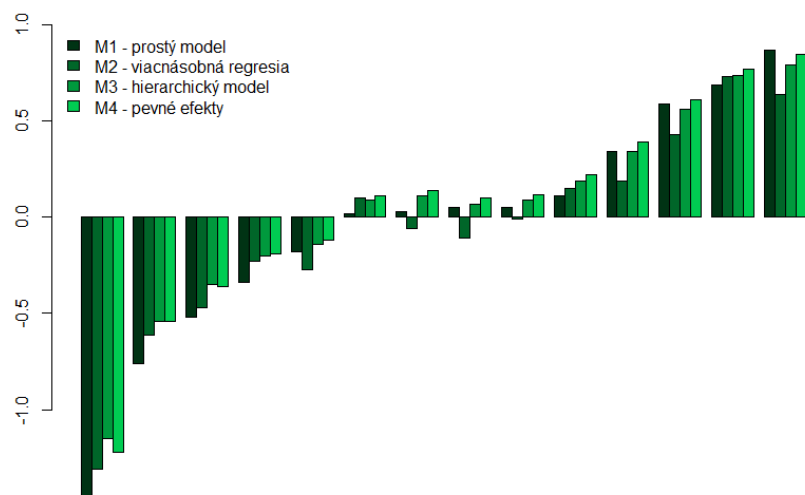


GYM

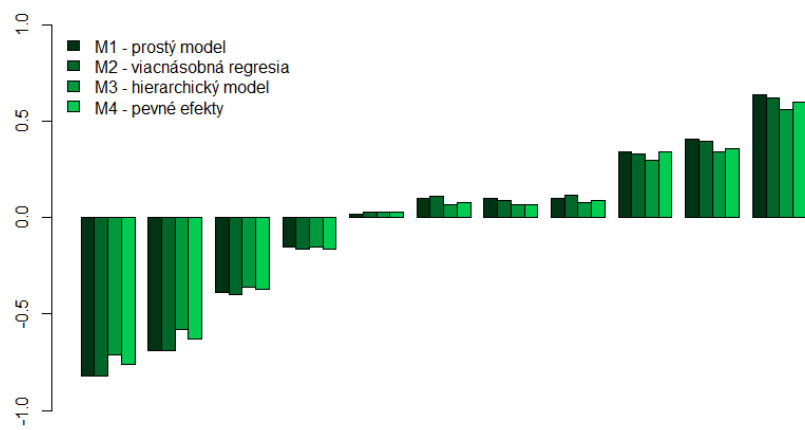


Obr. 8: Porovnanie bodových odhadov PH škôl vypočítaných jednotlivými modelmi pre MAT

SOS



GYM



Obr. 9: Porovnanie bodových odhadov PH škôl vypočítaných jednotlivými modelmi pre SJL

Pre viacúrovňové modely s pevnými efektami počítame PH školy j ako súčet priesečníku a_{01} a prislúchajúceho dummy koeficientu a_{0j} z rovnice (20). Dá sa ukázať, že pre regresné koeficienty platí (viac [Walters, 2011, s. 34]):

$$x_j^T \hat{a} \sim \mathcal{N}(x_j^T a, \sigma_{PH}^2 x_j^T (\mathbf{X}^T \mathbf{X})^{-1} x_j^T), \quad (49)$$

kde $a = (a_{01}, \dots, a_{0n}, a_1, b_1, \dots, b_k)^T$ predstavuje vektor regresných koeficientov z rovnice (20), $\hat{a}$ vektor ich odhadov a x_j vektor určujúci koeficienty potrebné pre výpočet PH školy j (napríklad $x_3 = [1, 0, 1, 0, 0, \dots, 0]$). Analogicky vieme na α hladine významnosti odvodiť IS pre odhady PH škôl ako:

$$x_j^T \hat{a} \pm t_{n-p, (1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}_{PH}^2 x_j^T (\mathbf{X}^T \mathbf{X})^{-1} x_j^T}, \quad (50)$$

kde $t_{n-p, (1-\frac{\alpha}{2})}$ je hodnota $100(1 - \frac{\alpha}{2})\%$ kvantilu Studentovho t-rozdelenia s $n-p$ stupňami voľnosti (n = počet meraní a p = počet koeficientov modelu).

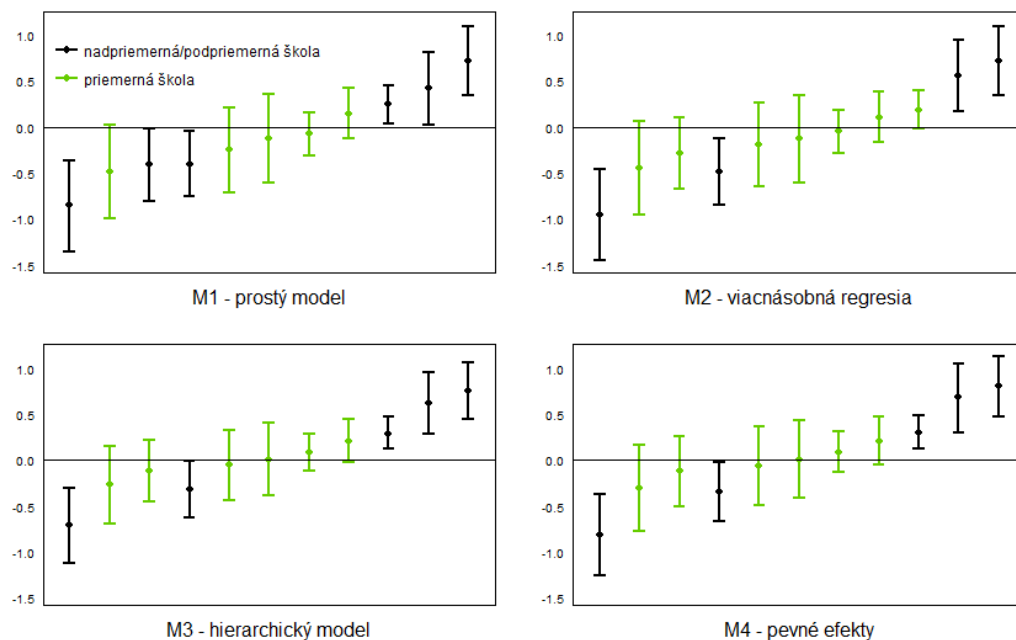
Vypočítané IS nezávisle od seba testujú pre odhady PH škôl rozdielnosť od národného priemeru. Často sa stretáme s prezentáciou týchto intervalov formou vzájomného porovnávania v rámci jedného grafu. Treba si ale uvedomiť, že tento postup nie je správny vzhľadom k tomu, že nastáva scenár tzv. simultánneho testovania. Pre jednoduché vysvetlenie tohto pojmu uvedieme príklad. Predstavme si, že vykonáme 100 simultánnych hypotéz. Klasickým postupom zamietame individuálnu nulovú hypotézu, ak je zodpovedajúca p-hodnota menšia alebo rovná ako hladina významnosti testu, povedzme $\alpha = 0,05$. Potom pre ktorúkoľvek platnú nulovú hypotézu je pravdepodobnosť mylného zamietnutia tejto hypotézy rovná 0,05. Ale „kombinovaná“ pravdepodobnosť mylného zamietnutia môže byť oveľa väčšia. Pre ilustráciu predpokladajme, že všetkých 100 nulových hypotéz je pravdivých a jednotlivé výsledky testov sú na sebe nezávislé. Potom môžeme očakávať, že počet mylných zamietnutí H_0 je $100 \cdot 0,05 = 5$. Šanca, že aspoň jedno zamietnutie je mylné je teda $1 - 0,95^{100} = 0,994$ [Afshartous & Wolf, 2007].

Ak chceme súčasne prezentovať IS, povedzme v jednom grafe, je treba pri ich výpočte použiť jednu z metód simultánneho porovnávania [Raftery & al., 2002]. To znamená, že tieto intervaly budú širšie ako v prípade samostatného testovania. Jednou z najznámejších metód je Bonferroniho úprava. Bonferroni tvrdí, že ak pre k odhadov vypočítame IS na $\frac{\alpha}{k}$ hladine významnosti, potom ich celková spoľahlivosť sa rovná aspoň $100(1 - \alpha)\%$. Výhodou tejto

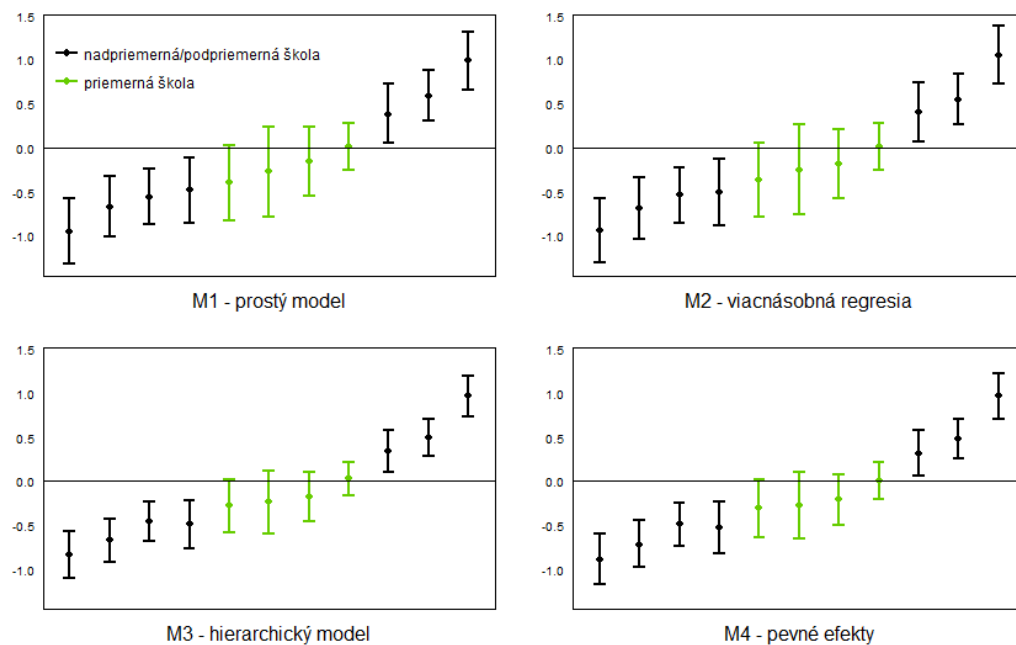
úpravy je fakt, že je ľahko aplikovateľná pre široké spektrum štatistických problémov zahŕňajúcich simultánne IS. V prípade veľkého počtu intervalov sa ale stáva výrazne konzervatívnou. Vzhľadom k nedôvere v dostatočnú vhodnosť dát pre modelovanie PH (slabá miera korelácie medzi vstupným a výstupným testovaním, úvod 3. kapitoly) a k faktu, že berieme do úvahy len odhady založené na vzorke aspoň 30 žiakov, sa uspokojíme s konzervatívnosťou Bonferroniho IS.

Empiricky zisťujeme, že IS pre odhady viacúrovňových modelov (M3 a M4) sú v priemere o 30% užšie (20% po Bonferroniho úprave) ako IS pre modely odhadnuté klasickou lineárnou regresiou (M1 a M2). Najužšie IS dostávame pre odhady hierarchického modelu M3, keďže odhady tohto modelu sú v prípade nedostatku dát „zrážané“ k národnému priemeru (odhady sú efektívnejšie ako pri modeloch s pevnými efektami, keďže majú menšiu strednú štvorcovú chybu). Ako však môžeme vidieť na Obr. 10 a na Obr. 11, kde sú zobrazené odhady PH škôl s prislúchajúcimi Bonferroniho IS, metódy M3 a M4 dávajú identické výsledky z pohľadu rozdielnosti odhadov od národného priemeru. Ukazuje sa, že v prípade GYM sú taktiež v oboch predmetoch výsledky všetkých modelov identické. Rozdielnosť pozorujeme len pri SOŠ. Dôležité je pozorovanie, že v prípade SOŠ v SJL, kedy sa bodové odhady modelu viacnásobnej regresie M2 líšili od odhadov ostatných modelov, sa táto rozdielnosť výsledkov pohybuje v rámci nepresnosti odhadov, resp. nebadáť výraznú rozdielnosť medzi IS modelu M2 a IS ostatných modelov. Všeobecne v zmysle zamietnutia rozdielnosti odhadov od národného priemeru pozorujeme medzi jednotlivými modelmi len minimálne rozdiely a teda je pre dostupnú vzorku dát najvhodnejšie modelovať PH škôl najjednoduchším prostým modelom (M1). Tento jav ale môže byť skreslený veľkosťou analyzovanej vzorky a preto ho treba pri jej zväčšení verifikovať.

SOŠ

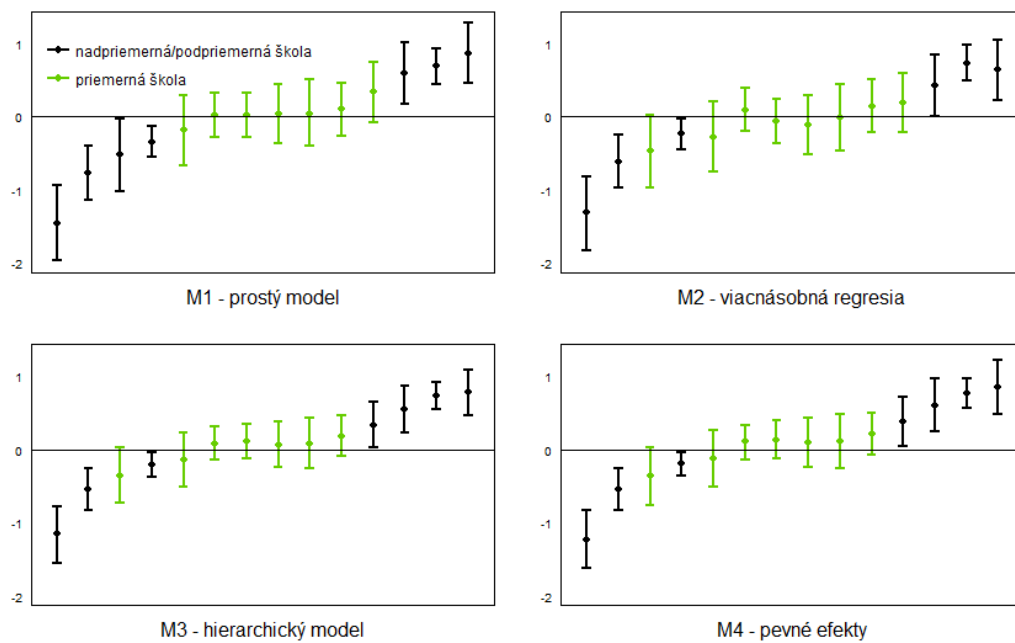


GYM

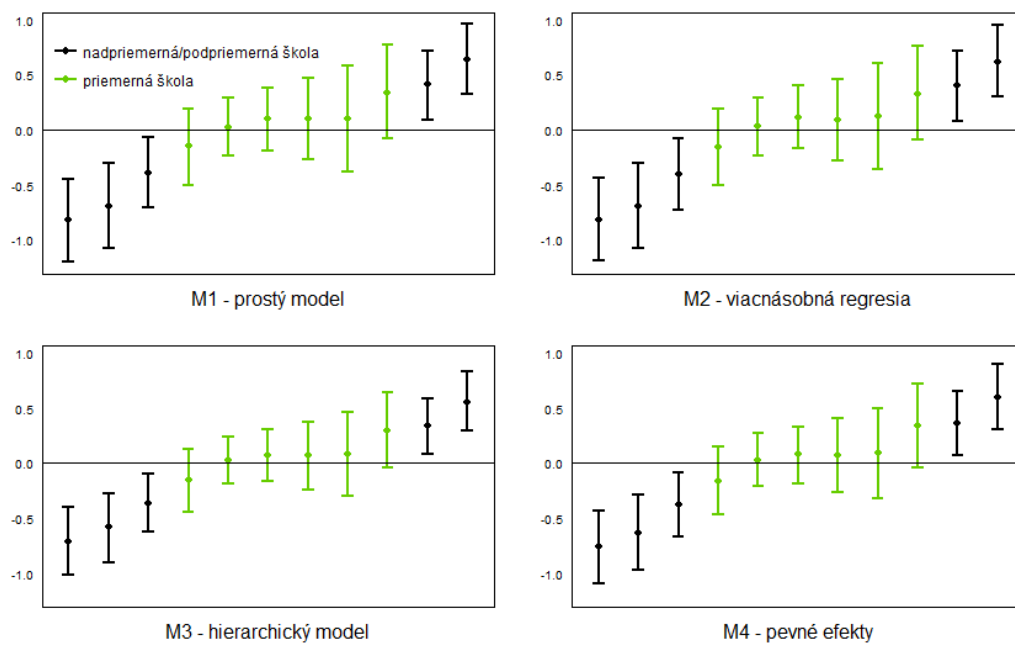


Obr. 10: IS po Bonferroniho úprave pre odhady PH škôl jednotlivých modelov pre MAT

SOŠ



GYM



Obr. 11: IS po Bonferroniho úprave pre odhady PH škôl jednotlivých modelov pre SJL

Záver

V diplomovej práci sme prezentovali empirický výskum v oblasti edukometrie s cieľom nájsť vhodný model pre výpočet tzv. pridanej hodnoty ako meradla kvality školy. Na základe výskumu ale nevieme dať na otázku výberu optimálneho modelu jednoznačnú odpoveď.

V práci sme predstavili 4 modely: prostý regresný model, model s použitím viacnásobnej regresie, hierarchický model s náhodným priesečníkom a viacúrovňový model s pevnými efektami. Ich teoretické pozadie sa zakladá na platnosti k nim prislúchajúcich predpokladov, pre splnenie ktorých sa ukázalo potrebné empirické dáta počiatočne transformovať kvantilovou normalizáciou. Pre túto nelineárnu transformáciu sme sa rozhodli z dôvodu jej použitia v iných krajinách OECD a na základe faktu, že nemala vplyv na zmeny v samotných výsledkoch pridaných hodnôt škôl. Pri zbere ďalších dát v budúcnosti odporúčame využitie tejto transformácie hlbšie analyzovať. Vzhľadom k rozdielnosti bodových odhadov jednotlivých modelov nedoporučujeme, aby bol zverejňovaný detailný rebríček škôl, kým nebudú viacročné skúsenosti s odhadovaním pridanej hodnoty. Spravovanie školskej siete, masmédiá a spoločenská situácia by mohli na základe prvotných reportov urobiť vážne a chybné rozhodnutia. Vhodnejšie je výsledky jednotlivých modelov publikovať formou štatistickej významnosti rozdielnosti ich bodových odhadov od národného priemeru. Čo sa týka výberu modelu, najsilnejším argumentom v prospech použitia prostého regresného modelu je jeho jednoduchosť. Všeobecne platí, že ak sa výsledky tohto modelu výrazne nelíšia od alternatívnych modelov, je v praxi preferovaný. Zo štatistického hľadiska je jeho nadstavbou model viacnásobnej regresie. V praxi sa ale ukázalo zložité určiť, ktoré kontextuálne premenné sú objektívne vhodné pre konštrukciu tohto modelu. Keďže koeficient ICC dokázal rozdielnosť žiakov v zmysle viacúrovňovej štruktúry, je potrebné zohľadniť výsledky viacúrovňových modelov. Keďže hierarchický model zohľadňuje pri odhadovaní pridaných hodnôt škôl veľkosť ich výberu a analogicky majú odhady užšie intervaly spoľahlivosti ako pri ostatných modeloch, považujeme tento model za najoptimálnejší.

Pozorovanie, že medzi výsledkami prostého a hierarchického modelu sú len nepatrné rozdiely, nás vedie k záveru, že modelovanie pridanej hodnoty najjednoduchším prostým modelom by mohlo byť postačujúce. V záujme ko-

rektnosti odporúčame proces modelovania navrhnutý v tejto práci v budúcnosti overiť na väčšom výbere škôl. Ak nárast vzorky spôsobí rozdiely medzi výsledkami týchto modelov, odporúčame použitie komplexnejšieho hierarchického modelu s náhodným priesečníkom. Dôležité je, aby výsledky boli verejnosti, ako sme už uviedli, reportované vo forme odchýlky od národného priemeru, resp. kategorizácie podpriemer, priemer a nadpriemer.

Literatúra

- ABDI H., 2007. *The Kendall Rank Correlation Coefficient*. Neil Salkind (Ed.), Encyclopedia of Measurement and Statistics, Thousand Oaks (CA): Sage, 2007. dostupné na internete (30.3.2014): <https://www.utdallas.edu/~herve/Abdi-KendallCorrelation2007-pretty.pdf>
- AFSHARTOUS D., WOLF M., 2007. *Avoiding ‘data snooping’ in multilevel and mixed effects models*. Zurich: Royal Statistical Society, Vol. 170, No. 4, 1035–1059, 2007. dostupné na internete (30.3.2014): http://moya.bus.miami.edu/~dafshartous/jrss-a_2007.pdf
- BALLOU D., SANDERS W., WRIGHT P., 2004. *Controlling for student background in value-added assessment of teachers*. Journal of Educational and Behavioral Statistics, Vol. 29, No. 1, 37–66., 2004.
- BARNETT V., 1975. *Probability Plotting Methods and Order Statistics*. Applied Statistics, Vol. 24, No. 1, 95–108, 1975.
- BOZDOGAN H., 2000. *Akaike’s Information Criterion and Recent Developments in Information Complexity*. Journal of Mathematical Psychology, Vol. 44, 62–91, 2000.
- CHEN H., CUI Z., ZHU R., GAO X., 2010. *Evaluating the Effects of Differences in Group Abilities on the Tucker and the Levine Observed-Score Methods for Common-Item Nonequivalent Groups Equating*. Iowa: ACT Research Report Series, 2010. dostupné na internete (19.3.2014): http://www.act.org/research/researchers/reports/pdf/ACT_RR2010-1.pdf
- CLARK P., CRAWFORD C., STEELE F., VIGNOLES A., 2010. *The Choice Between Fixed and Random Effects Models: Some Considerations For Educational Research*. University of London: Working Paper No. 10/240, 2010. dostupné na internete (16.3.2014): <http://www.bristol.ac.uk/cmpo/publications/papers/2010/wp240.pdf>
- DOWNESS D., VINDURAMPULLE O., 2007. *Value-added measure for school improvement*. Melbourne: Department of Education and Early Childhood Development, 2007. dostupné na internete (6.3.2014): <https://www.eduweb.vic.gov.au/edulibrary/public/publ/research/publ/value-added-measures-report.pdf>

- FIELD A., 2009. *Discovering Statistics Using SPSS*. 3rd Edition. London: SAGE Publications Inc., 2009.
- FITZ-GIBBON C. T., 1997. *Feasibility studies for a national system of value-added indicators*. CEM, University of Durham: The Value Added National Project, February 1997. dostupné na internete (6.3.2014): <http://www.cem.org/attachments/publications/CEMWeb009%20Feasibility%20Study%20Nat%20System%20VA%20Indicators.pdf>
- FOX J., 2002. *Bootstrapping Regression Models*. Appendix to An R and S-PLUS Companion to Applied Regression, 2002. dostupné na internete (19.3.2014): <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-bootstrapping.pdf>
- FOX J., 2008. *Applied Regression Analysis and Generalized Linear Models, Chapter 21*. Sage Publications, 2008. dostupné na internete (19.3.2014): http://www.sagepub.com/upm-data/21122_Chapter_21.pdf
- GLICKMAN M. E., VAN DYK D. A., 2007. *Basic Bayesian Methods*. dostupné na internete (6.3.2014): <http://www.glicko.net/research/glickman-vandyk.pdf>
- GREEN W. H., 2002. *Econometric Analysis*. 5th Edition, New Jearsy: Paerson Education, 2002. ISBN 0-13-066189-9.
- HANSEN B. E., 2012. *Econometrics*. University of Wisconsin, 2012.
- JORDAN M. I., 2010. *The Conjugate Prior for the Normal Distribution*. dostupné na internete (16.3.2014): <http://www.cs.berkeley.edu/~jordan/courses/260-spring10/lectures/lecture5.pdf>
- JUŠČÁKOVÁ Z., 2012. *Indikátory kvality vzdelávania a autoevaluácia školy*. Bratislava: Didaktika, s. 2-7, 2012.
- JUŠČÁKOVÁ Z., 2013. *Indikátory kvality vzdelávania v hodnotení škôl II*. Bratislava: NÚCEM, 2013.
- JUŠČÁKOVÁ Z., ŽOLTAK T., 2013. *History of VAM*. prezentácia, International Workshop VAM: Vrátna dolina 11.6. – 13.6.2013, 2013.

- KALEN M. J., 1988. *An NCME Instructional Module on Traditional Equating Methodology*. The American College Testing Program, 1988. dostupné na internete (19.3.2014): <http://ncme.org/linkservid/65DCF34D-1320-5CAE-6E896C93330B9C73/showMeta/0/>
- KIM H., LALANCETTE D., 2013. *Literature review on the value-added measurement in higher education*. OECD, 2013. dostupné na internete (16.3.2014): <http://www.oecd.org/edu/skills-beyond-school/Litterature%20Review%20VAM.pdf>
- KOLKOVÁ M., IVICA L., 2013. *Čo odhaduje tzv. pridaná hodnota školy? Hodnotenie kvality vzdelávania - súčasný stav a perspektívy*. Zborník príspevkov z konferencie. Bratislava: NÚCEM, 2013. ISBN 9-788-089-638-11-6.
- KREFT I., LEEUW J., 1998. *Introducing Multilevel Modeling*. Los Angeles: SAGE Publications, 1998.
- LECKIE G., GOLDSTEIN H., 2009. *The limitations of using school league tables to inform school choice*. Journal of the Royal Statistical Society Vol. 172, No. 4, 835–851, 2009.
- LEVINE T. R., HULLETT C. R. *Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research*. Human Communication Research, Vol. 28, No. 4, 612–625, 2002.
- MARKECHOVÁ D., STEHLÍKOVÁ B., TIRPÁKOVÁ A., 2011. *Základy štatistiky pre pedagógov*. Nitra: FPV UKF, 2011.
- MCKNIGHT P. E., NAJAB J., 2010. *Mann-Whitney U Test*. Corsini Encyclopedia of Psychology 1, 2010.
- MUKAKA M. M., 2012. *Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research*. Malawi Medical Journal, Vol. 24, No. 3, 69-71, 2012.
- NORDSTOKKE D. W., ZUMBO B. D., 2010. *A New Nonparametric Levene Test for Equal Variances*. Universitat de València: Psicológica, Vol. 31, No. 2, 401-430, 2010.
- O'BRIEN R. M., 2007. *A Caution Regarding Rules of Thumb for Variance Inflation Factors*. Springer: Quality & Quantity, Vol. 41, 673–690,

2007. dostupné na internete (16.3.2014): <http://link.springer.com/article/10.1007%2Fs11135-006-9018-6>
- OSBORNE J. W., 2010. *Improving your data transformations: Applying the Box-Cox transformation*. Practical Assessment, Research & Evaluation: Electronic journal, Vol. 15, No.12, 2010. dostupné na internete (19.3.2014): <http://pareonline.net/pdf/v15n12.pdf>
- OECD, 2008. *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. Paris: OECD Publications, 2008. ISBN: 978-92-64-05022-8.
- RAFTERY J. A., ABELLY M. L., BRASELTONY J. P., 2002. *Multiple Comparison Methods for Means*. Society for Industrial and Applied Mathematics: Vol. 44, No. 2, 259-278, 2002. dostupné na internete (30.3.2014): <http://sci2s.ugr.es/keel/pdf/specific/articulo/35723.pdf>
- RAPP J., 1999. *Linear and Equipercetile Methods for Equating PET*. National Institute for Testing and Evaluation, 1999. dostupné na internete (19.3.2014): <https://www.nite.org.il/files/reports/e266.pdf>
- RAUDENBUSH S. W., 2004. *What Are Value-Added Models Estimating and What Does This Imply For Statistical Practice*. Journal of Educational and Behavioral Statistics, Vol. 29, No. 1, 121-129, 2004. dostupné na internete (30.3.2014): <http://www.ssicentral.com/hlm/techdocs/value-added04.pdf>
- RAY A., 2006. *School Value Added Measures in England*. Paper for the OECD Project on the Development of Value-Added Models in Education Systems, UK: Department for Education and Skills, 2006.
- RYŠKA R., 2009. *Evaluce a přidaná hodnota ve vzdělávání*. Praha: Vydavatelství Pedagogické fakulty Univerzity Karlovy v Praze, 2009. dostupné na internete (6.3.2014): <http://www.strediskovzdelavacipolitiky.info/download/Evaluace%20a%20pridana%20hodnota%20ve%20vzdelavani.pdf>
- SANDERS W. L., 1999. *A Brief History of Educational 'Value Added': How Did We Get To Where We Are?*. School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, Vol. 10, No. 2, 233-256, 1999.

- SANDERS W. L., HORN S. P., 1998. *Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research*. Journal of Personnel Evaluation in Education Vol. 12, No. 3, 247-256, Tennessee: Kluwer Academic Publishers, 1998.
- SOUKUP P., 2006. *Proč užívat hierarchické lineární modely?*. Sociologický časopis, Vol. 42, No. 5, 987-1012, Praha: Sociologický ústav AV ČR, 2006. dostupné na internetu (16.3.2014): http://sreview.soc.cas.cz/uploads/fad08138afdabed4309104945d7a7ff2dcc967f8_613_09soukup28.pdf
- WALTERS K., 2011. *Linear Models*. Podklady k přednášce MAS6003, Sheffield: University of Sheffield, 2011.
- WEBSTER W. J., MENDRO R. L., 1997. *The Dallas Value-Added Accountability System*. Jason Millman (Ed.). Grading teachers, grading schools, Is student achievement a valid evaluation measure? s. 81-99, Dallas Public Schools, 1997. dostupné na internetu (6.3.2014): <http://www.dallasisd.org/cms/lib/TX01001475/Centricity/Shared/evalacct/research/articles/Webster-Dallas-Value-Added-Accountability-System.pdf>
- ŽOLTAK T., 2013. *Value-added indicators for Polish schools: an overview*. Poland: Educational Research Institute, 2013.

Appendix

A.1: EVAAS

Nech i označuje žiakov, j školy, ktoré navštevujú a p prechod v rámci ročníkov. Potom základný tvar EVAAS systému modelovania PH možno formulovať ako:

$$(y_{ip}, z_{ip}) = (\mu_p, \gamma_p) + \text{sum}_{k \leq p}(\theta_{ijk}, \varphi_{ijk}) + (\varepsilon_{ip}, \delta_{ip}), \quad (51)$$

kde (y_{ip}, z_{ip}) reprezentuje skóre v literárnej a matematickej gramotnosti žiaka i v prechode p v uvedenom poradí, (μ_p, γ_p) reprezentuje priemerné skóre v daných oblastiach pre celú populáciu v prechode p , $(\theta_{ijk}, \varphi_{ijk})$ reprezentuje k -ty efekt školy j pre i -teho žiaka v daných oblastiach, pričom tieto parametre sú náhodné a navzájom nezávislé. ε_{ip} a δ_{ip} sú náhodné chybové členy pre dané oblasti nezávislé od ostatných parametrov modelu, pričom predpokladáme, že platí $(\varepsilon_{ip}, \delta_{ip}) \sim \mathcal{N}((0, 0), \Sigma)$, kde kovariančná matica $\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & \text{cov}(\varepsilon_{ip}, \delta_{ip}) \\ \text{cov}(\varepsilon_{ip}, \delta_{ip}) & \sigma_\delta^2 \end{pmatrix}$ je kladne definitná [OECD, 2008, s. 149].

A.2: VIF kritérium

Proporciu disperzie v i -tej nezávislej premennej súvisiacej s ostatnými nezávislými premennými v modeli označíme ako R_i^2 . Toleranciu pre i -te nezávislé meranie vyjadríme ako $1 - R_i^2$, čo reprezentuje proporciu disperzie i -tej nezávislej premennej, ktorá nie je pridružená ostatným premenným v modeli. VIF (Variance Inflation Factor) kritérium je prevrátenou hodnotou tolerancie:

$$VIF = \frac{1}{1 - R_i^2}, \quad (52)$$

intuitívne vyjadrujúcou efekt R_i^2 na disperziu odhadu koeficientu i -tej premennej. Analogicky vysoké hodnoty VIF kritéria reprezentujú prítomnosť multikolinearity medzi nezávislou premennou i a ostatnými nezávislými premennými v modeli. V praxi sa hodnoty VIF kritéria okolo 10 považujú za hraničné, aj keď názory sa výrazne líšia naprieč literatúrou O'Brien [2007].

A.3: Posteriorne rozdelenie empirických Bayesovských odhadov

Pri odvodení empirického Bayesovského odhadu náhodných efektov škôl vychádzame z publikácií Jordan [2010] a Glickman & van Dyk [2007]. Bayesovský prístup je prístup používajúci pravdepodobnosť na vyjadrenie viery v poznatky o neznámych veličinách. Zahŕňa formuláciu pravdepodobnostného modelu s očakávaním priórneho rozdelenia pre neznáme parametre pred samotným skúmaním dát. Na základe pozorovania dát možno následne zhotoviť funkciu vierohodnosti a v kombinácii s priórnym rozdelením možno vypočítať posteriorne rozdelenie odhadovaného parametra. Posteriorne rozdelenie v istom zmysle vyčísľuje neistotu v odhadovanom parametri modelu po skúmaní dostupných údajov. V našom prípade hľadáme pre hierarchický model popísaný rovnicou (15) posteriorne rozdelenie $u|\bar{r}_j$, ktorého stredná hodnota určuje BLUP odhad $\hat{u}$. Predpoklad (16) $u \sim \mathcal{N}(0, \sigma_u^2)$ určuje priórne rozdelenie. Pre známy priemer náhodnej časti modelu $\bar{r}_j$ sa dá na základe predpokladu (17) $PH_{ij} \sim \mathcal{N}(0, \sigma_{PH}^2)$ ľahko ukázať, že platí $\bar{r}_j|u \sim \mathcal{N}(0, \frac{\sigma_{PH}^2}{n_j})$. Funkciu vierohodnosti definujeme ako $L(\bar{r}_j|u) = \Phi(u, \bar{r}_j, \frac{\sigma_{PH}^2}{n_j})$, pričom $\Phi(x, \mu, \sigma^2)$ predstavuje pravdepodobnostnú funkciu pre hustotu normálneho rozdelenia $\mathcal{N}(\mu, \sigma^2)$ v bode x . Na základe Bayesovej vety je posteriorne rozdelenie úmerné súčinu priórneho rozdelenia a funkcie vierohodnosti:

$$u|\bar{r}_j \propto u \cdot L(\bar{r}_j|u), \quad (53)$$

čo môžeme presne definovať ako integrál:

$$u|\bar{r}_j = \int \Phi(u, 0, \sigma_u^2) L(\bar{r}_j|u) du = \int \Phi(u, 0, \sigma_u^2) \Phi(u, \bar{r}_j, \frac{\sigma_{PH}^2}{n_j}) du. \quad (54)$$

Teda podmienené rozdelenie $u|\bar{r}_j$ získame ako súčin pravdepodobnosti získania $\bar{r}_j$ pre dané u a pravdepodobnosti, že dané u bude pochádzať z $\mathcal{N}(0, \sigma_u^2)$, teda z predpokladaného priórneho rozdelenia. Strednú hodnotu a disperziu tohto rozdelenia definujeme ako:

$$E(u|\bar{r}_j) = \int u \Phi(u, 0, \sigma_u^2) \Phi(u, \bar{r}_j, \frac{\sigma_{PH}^2}{n_j}) du = S_j \bar{r}_j, \quad (55)$$

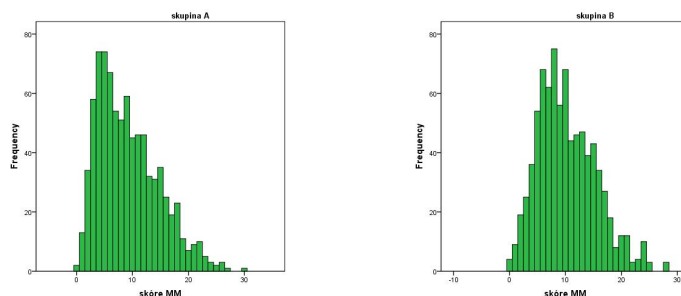
$$\begin{aligned} Var(u|\bar{r}_j) &= \int (u - E(u|\bar{r}_j))^2 \Phi(u, 0, \sigma_u^2) \Phi(u, \bar{r}_j, \frac{\sigma_{PH}^2}{n_j}) du \\ &= S_j \frac{\sigma_{PH}^2}{n_j}. \end{aligned} \quad (56)$$

Keďže priórna funkcia a aj funkcia vierohodnosti pochádzali z normálneho rozdelenia, analogicky sa dá ukázať, že aj posteriórne rozdelenie bude normálne a tak platí:

$$u|\bar{r}_j \sim \mathcal{N}(S_j\bar{r}_j, S_j\frac{\sigma_{PH}^2}{n_j}). \quad (57)$$

A.4: Ekvipercentilové vs. lineárne porovnávanie

Ekvipercenilová metóda sa považuje oproti lineárnemu porovnávaniu za robustnejšiu, keďže zohľadňuje rozdielnosť tvarov porovnávaných rozdelení. Táto metóda ale vyžaduje spojitosť rozdelenia, čo pri diskretných dátach pramení v aplikáciu vyhladzovacích metód. Robustnosť tejto metódy pôsobí proti jej šetrnosti a pri menšom počte dát vykazuje väčšiu náhodnú chybu. Preto ak dáta nevykazujú významné rozdiely tvarov rozdelenia, zvykne sa používať jednoduchšia lineárna metóda porovnávania Rapp [1999]. V prípade histogramov pre dáta z MAT vizuálne pozorujeme podobný tvar výsledkov rozdelení obidvoch skupín (Obr. 12).



Obr. 12: Rozdelenie výstupných meraní pre skupiny A a B z MAT

A.5: Kontrola predpokladov pre model prostej PH s použitím hrubého skóre a skóre upraveného kvantilovou normalizáciou

Tabuľka 14 zobrazuje porovnanie výstupov ANOVA tabulky programu SPSS Statistics 19 pre uvedené modely (pri modeloch s použitím hrubého skóre pozorujeme v prípade MAT štatistickú významnosť nelineárneho vzťahu medzi vstupnými a výstupnými premennými). Tabuľky 15 a 16 zobrazujú porovnanie výsledkov testovania normálneho rozdelenia rezíduí a homogenity variancie jednotlivých modelov.

Tabuľka 14: [P1] Linearita

		Hrubé skóre ^a				Kvantilová normalizácia						
		SS	df	MS	F	Sig.	SS	df	MS	F	Sig.	
SJL GYM	Between Groups	(Combined)	18680,77	12	1556,73	13,032	0	134,565	12	11,214	13,242	0
	Linearity		17213,28	1	17213,28	144,096	0	123,521	1	123,521	145,862	0
	Deviation ^b		1467,482	11	133,407	1,117	0,344	11,044	11	1,004	1,186	0,293
Within Groups			97954,9	820	119,457		694,404	820	0,847			
Total			116635,7	832			828,969	832				
SJL SOŠ	Between Groups	(Combined)	32465,01	18	1803,612	11,054	0	170,012	18	9,445	11,094	0
	Linearity		30334,29	1	30334,29	185,919	0	155,981	1	155,981	183,213	0
	Deviation ^b		2130,719	17	125,336	0,768	0,731	14,032	17	0,825	0,97	0,491
Within Groups			170174,2	1043	163,158		887,971	1043	0,851			
Total			202639,2	1061			1057,983	1061				
MAT GYM	Between Groups	(Combined)	36064,41	16	2254,026	9,721	0	135,015	16	8,438	9,961	0
	Linearity		27788,63	1	27788,63	119,843	0	118,557	1	118,557	139,945	0
	Deviation ^b		8275,788	15	551,719	2,379	0,002	16,458	15	1,097	1,295	0,198
Within Groups			185731,7	801	231,875		678,582	801	0,847			
Total			221796,1	817			813,597	817				
MAT SOŠ	Between Groups	(Combined)	19410,09	18	1078,338	4,449	0	80,786	18	4,488	4,892	0
	Linearity		12925,43	1	12925,43	53,322	0	62,55	1	62,55	68,18	0
	Deviation ^b		6484,655	17	381,45	1,574	0,065	18,236	17	1,073	1,169	0,284
Within Groups			201679	832	242,403		763,295	832	0,917			
Total			221089,1	850			844,08	850				

^a. Standardized Residual; ^b. From Linearity; SS - Sum of Squares; MS - Mean Square

Tabuľka 15: [P7] Normalita rezíduí

		Hrubé skóre ^a				Kvantilová normalizácia					
		Kolmogorov-Smirnov ^b		Shapiro-Wilk		Kolmogorov-Smirnov ^b		Shapiro-Wilk			
		Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	
SJL GYM		0,03	833	0,081	0,993	833	0,001	0,021	833	0,998	0,642
SJL SOŠ		0,023	1062	,200*	0,999	1062	0,613	0,025	1062	0,998	0,233
MAT GYM		0,048	818	0	0,986	818	0	0,03	818	0,998	0,616
MAT SOŠ		0,119	851	0	0,921	851	0	0,027	851	0,997	0,163

^a. Standardized Residual; ^b. Lilliefors Significance Correction; *. This is a lower bound of the true significance.

Tabuľka 16: [P6] Homoskedasticita^b

	Hrubé skóre ^a	Kvantilová normalizácia
SJL GYM	0,304	0,136
SJL SOŠ	0,174	0,207
MAT GYM	0,147	0,044
MAT SOŠ	0,633	0,349

^a. Standardized Residual; ^b. Accordingly highlighted parametric and nonparametric Levene's test.