

**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Bc. Michaela Vařejková

**Modely pro vyvažování kognitivních
testů**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Patrícia Martinková, Ph.D.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2023

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Ráda bych poděkovala zejména vedoucí mé práce, RNDr. Patřicií Martinkové, Ph.D., za všechny cenné rady a připomínky, za čas, který mi věnovala, a hlavně za velkou trpělivost. Také bych chtěla poděkovat konzultantovi práce, RNDr. Matúšovi Maciakovi, Ph.D. Práce byla navíc podpořena grantovým projektem GAČR 21-03658S.

Název práce: Modely pro vyvažování kognitivních testů

Autor: Bc. Michaela Vařejková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Patrícia Martinková, Ph.D., Ústav informatiky AV ČR

Abstrakt: Tato práce se věnuje statistickým metodám pro vyvažování kognitivních testů, což je proces používaný k transformaci skóre z více verzí testu tak, aby byla mezi sebou porovnatelná. Teoretická část práce je rozčleněna do tří kapitol, kde každá se zabývá jedním z možných přístupů k procesu vyvažování. V první kapitole jsou představeny tradiční vyvažovací metody, ve druhé kapitole metody vyvažování za pomoci jádrových odhadů a ve třetí kapitole metody vyvažování pomocí modelů teorie odpovědi na položku. Závěrečná část práce je věnována empirické studii, ve které je ilustrováno použití vyvažovacích metod na konkrétním datovém souboru. Jedná se o odpovědi na dvě verze znalostního testu z matematiky, které byly zadávány studentům 4. ročníků základních škol v České republice v rámci mezinárodního šetření TIMSS v roce 2015.

Klíčová slova: vyvažování kognitivních testů, modely teorie odpovědi na položku (IRT modely), jádrové odhady.

Title: Models for equating in cognitive tests

Author: Bc. Michaela Vařejková

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Patrícia Martinková, Ph.D., Institute of Computer Science of the Czech Academy of Sciences

Abstract: This thesis focuses on statistical methods for equating cognitive tests, which is the process of transforming scores from multiple test versions to ensure their comparability. Divided into three chapters, the theoretical part of the thesis addresses different approaches to test equating. The first chapter presents traditional equating methods, the second explores kernel equating methods, while the third covers equating methods using Item Response Theory models. The concluding part of the thesis showcases an empirical study demonstrating the application of equating methods on a real dataset. This dataset contains responses to two versions of a math test taken by fourth-grade students in the Czech Republic as part of the 2015 TIMSS international survey.

Keywords: test equating, item-response theory (IRT) models, kernel smoothing.

Obsah

Úvod	2
1 Tradiční vyvažovací metody	5
1.1 Vyvažování ve střední hodnotě	5
1.2 Lineární vyvažování	6
1.3 Ekvipercentilové vyvažování	7
2 Vyvažování pomocí jádrových odhadů	9
2.1 Design ekvivalentních skupin	9
2.1.1 Vyhlazení pravděpodobností skóre	10
2.1.2 Odhad pravděpodobností skóre	12
2.1.3 Spojitá aproximace skóre	12
2.1.4 Vyvažování	16
2.1.5 Výpočet standardních chyb vyvažování	17
2.2 Design neekvivalentních skupin	23
2.2.1 Poststratifikační metoda	25
2.2.2 Metoda řetězového vyvažování	31
3 Vyvažování pomocí modelů teorie odpovědi na položku	37
3.1 IRT modely	37
3.2 Metody odhadu vyvažovacích konstant	39
3.2.1 Momentová metoda využívající průměr	40
3.2.2 Momentová metoda využívající průměr a směrodatnou od- chylku	41
3.2.3 Haebárova metoda	41
3.2.4 Stocking-Lordova metoda	42
3.3 Vyvažování celkových skóre	43
3.3.1 Vyvažování skutečných skóre	43
3.3.2 Vyvažování pozorovaných skóre	44
4 Praktická část	46
4.1 Datový soubor	46
4.2 Tradiční vyvažovací metody	48
4.2.1 Vyvažování ve střední hodnotě	48
4.2.2 Lineární vyvažování	48
4.2.3 Ekvipercentilové vyvažování	48
4.3 Vyvažování pomocí jádrových odhadů	50
4.3.1 Design ekvivalentních skupin	50
4.3.2 Design neekvivalentních skupin s kotvícími položkami	57
4.4 Vyvažování pomocí modelů teorie odpovědi na položku	62
Závěr	65
Seznam použité literatury	67

Úvod

V různých odvětvích vzdělávání je častým cílem srovnávat výsledky ze znalostních testů, které byly zadávány v jiných zkušebních termínech. Přestože jsou znalostní testy navrhovány tak, aby měly stejné parametry a byly co nejvíce srovnatelné, jistému rozdílu v obtížnostech zadaných verzí se nedá vždy vyhnout. Výsledky těchto testů pak nelze férově porovnávat, dosažené skóre 100 bodů v prvním zkušebním termínu nemusí odpovídat stejnému skóre v testu z druhého zkušebního termínu. Tento problém by byl vyřešen zadáním identického znalostního testu v obou termínech. V takovém případě by ale mohlo dojít k vyzrazení otázek a někteří studenti by tak mohli být zvýhodněni. Z tohoto důvodu je proto běžné vytvářet více variant testu. Pomocí statistických vyvažovacích metod (*angl. equating methods*) se pak snažíme dosáhnout srovnatelnosti výsledků dvou nebo více verzí znalostního testu tak, aby skóre v těchto různých formách testu mohlo být použito zcela zaměnitelně a aby respondenti, kteří absolvovali obtížnější test, nebyli znevýhodněni.

Oblast vzdělávání a testování je asi nejčastější oblast, kde se vyvažovací metody využívají. Setkat se s nimi ale můžeme i v jiných oblastech, kde měříme nějaký atribut pomocí více verzí zadávaného testu. Může se jednat například i o výkonnostní testy ve sportu či o měření postojů nebo měření zdravotního stavu.

V této práci se budeme zabývat statistickými přístupy k vyvažování testů. Popíšeme vybrané metody, jejich předpoklady a statistické procesy, z kterých vychází. V práci zavedeme i pojem vyvažovací funkce, kterou používáme pro samotný přepočítání skóre. S její pomocí pak pro dané skóre na škále jedné verze testu získáme jeho ekvivalent na škále druhé verze. Detailně se pak zaměříme hlavně na vyhodnocení celého procesu a zavedeme míru, která udává nepřesnost procesu vycházející z výběrové chyby. V závěru bude vše doplněno o ilustraci na reálném datovém souboru.

Uváděné vyvažovací metody upravují rozdíly v obtížnosti, nikoli rozdíly v obsahu. Je důležité, aby všechny varianty testu byly obsahově stejné a aby měřily stejný atribut. V literatuře (Dorans a Holland (2000), Davier (2011), Kolen a Brennan (2014)) se často uvádí celkem 5 předpokladů, které by měly být splněny, aby výsledná přepočtená skóre mohla být považována za zcela srovnatelná. Jedná se o následující předpoklady:

1. *Stejný měřený konstrukt* Všechny verze testu by měly měřit stejný konstrukt (znalost, dovednost) a být obsahově stejné.
2. *Stejná reliabilita testů* Reliabilita všech testů by měla být srovnatelná.
3. *Symetrie vyvažovací funkce* Pokud vyvažovací funkce φ převádí skóre testu X na škálu skóre Y , potom funkce φ^{-1} je vyvažovací funkcí pro převod skóre testu Y na skóre testu X (definice pojmu vyvažovací funkce bude zavedena později v první kapitole této práce).
4. *Předpoklad rovnocennosti testů* Pro žádného studenta není důležité, která verze testu je mu zadána.

5. *Invariance vůči populaci* Výsledný tvar vyvažovací funkce by měl být stejný, i kdyby byl k jejímu odvození použit jiný vzorek populace či subpopulace.

Téměř žádný z výše uvedených předpokladů se nedá ovšem prakticky ověřit, měly by proto sloužit spíše jako neformální požadavky, které je dobré zohlednit. V případě, že máme podezření nebo nějaké indicie, že by některý z nich mohl být porušen, neměli bychom skóre považovat za zaměnitelná. Tato práce se bude ale zabývat výhradně statistickým popisem procesu vyvažování. Rozhodnutí o tom, zda je vhodné vyvažování použít a do jaké míry jsou jeho výsledky platné, je potřeba konzultovat s odborníkem na danou oblast. Vždy také záleží na konkrétní situaci.

Designy sběru dat

Situace, ve kterých chceme vyvažování testů použít, se mohou lišit, a tomu pak odpovídají různé designy sběru dat (von Davier a kol. (2004)). Mezi tři nejběžnější designy patří design jedné skupiny (*angl. single-group*, SG), design ekvivalentních skupin (*angl. equivalent-groups*, EG) a design neekvivalentních skupin s kotvícími položkami (*angl. non-equivalent groups with anchor test*, NEAT).

Design jedné skupiny

U SG designu máme jen jednu skupinu respondentů vybranou z populace T , ve které jsou všem respondentům zadány všechny verze testu. Výhodou tohoto designu je, že zde dochází k relativně malým chybám měření. Na druhou stranu velkou nevýhodou je fakt, že zkoušeným osobám musí být zadáno velké množství otázek a celková doba trvání testu je dlouhá. Tím může docházet k únavě osob, která může mít na měření vliv (v později zadaných testech mohou mít zkoušení horší skóre ne kvůli vyšší obtížnosti, ale právě kvůli únavě). Tento typ designu není však příliš rozšířený hlavně kvůli jeho nepraktičnosti.

Design ekvivalentních skupin

Dalším typem designu je EG design. V tomto případě jsou dvě různé verze testu zadány dvěma skupinám zkoušených osob. Předpokládáme, že skupiny jsou ekvivalentní a byly náhodně vybrány ze stejné populace T (proto se někdy tomuto designu říká také design náhodných skupin (*angl. random-groups*, RG)). Výhodou tohoto designu je značné zkrácení doby trvání testu a eliminování efektu únavy. Nevýhodou je ale právě předpoklad ekvivalentních skupin, který při porušení vede k větším chybám.

Design neekvivalentních skupin s kotvícími položkami

Poslední z uvedených designů je NEAT design. Ten můžeme použít v případě, kdy máme dvě neekvivalentní skupiny osob. Každé skupině jsou pak zadány rozdílné testy, které však obsahují sadu stejných otázek. Na jejich základě jsou odhadnuty rozdíly ve skupinách, a ty tak nezasahují do rozdílů v obtížnosti mezi různými verzemi testu.

Možných designů sběru dat existuje mnohem více, nejsou však již tolik rozšířené, a proto je zde neuvádíme. V práci se zaměříme především na EG a NEAT design, protože se jedná v praxi o nejběžněji používané designy.

1. Tradiční vyvažovací metody

Jedním z klasických přístupů ve vyvažování více verzí testu je použití celkového dosaženého skóre. Při formulaci problému vycházíme zejména z knihy González a Wiberg (2017). Na bodové stupnice testů se můžeme dívat jako na prostory elementárních jevů náhodných veličin, které reprezentují dosažené skóre v testech. Problém vyvažování je pak nalezení vhodné transformace skóre z jedné stupnice na jejich ekvivalenty na stupnici druhé. Pro jednoduchost budeme uvažovat pouze dvě verze testu, verzi X a Y. Nechť X značí náhodnou veličinu udávající celkové skóre ve verzi X, Y pak skóre ve verzi Y. Počet účastníků každého testu se může obecně lišit, předpokládejme, že verzi X vyplnilo N respondentů, verzi Y vyplnilo M respondentů. Skóre získaná z první verze testu tak značíme X_1, \dots, X_N , skóre z druhé verze Y_1, \dots, Y_M . Platí $X_i \sim F$, $i = 1, \dots, N$ a $Y_i \sim G$, $i = 1, \dots, M$, kde F a G jsou distribuční funkce náhodných veličin X a Y . Cílem vyvažování testů je modelovat vztah skóre v jedné verzi testu s odpovídajícím skóre druhé verze. Na základě toho, jaké distribuční charakteristiky k modelování využijeme, pak rozlišujeme tři tradiční vyvažovací metody, a to vyvažování ve střední hodnotě, lineární vyvažování a vyvažování na základě percentilů, tzv. ekvippercentilové vyvažování (*angl. equipercetile equating*). Čerpáme především z knih Kolen a Brennan (2014), González a Wiberg (2017) a Davier (2011). Metody budou v této kapitole ilustrovány na EG designu, lze je ovšem rozšířit i na NEAT design.

1.1 Vyvažování ve střední hodnotě

Tato vyvažovací metoda je ze všech uváděných tradičních metod nejjednodušší a slouží spíše jako ilustrativní. V tomto případě předpokládáme, že distribuční funkce náhodných veličin X a Y mají shodný tvar a liší se pouze v parametru posunutí. Označme μ_X a μ_Y střední hodnoty náhodných veličin X a Y , realizace těchto veličin jako x a y . Položme

$$x - \mu_X = y - \mu_Y.$$

Za shodná považujeme ta skóre, která mají stejnou vzdálenost od středu, tj. od střední hodnoty příslušného pravděpodobnostního rozdělení. Vyjádřením y dostaneme

$$\varphi_M(x; \mu_X, \mu_Y) = y = x - \mu_X + \mu_Y, \quad (1.1)$$

kde funkce $\varphi_M(x)$ značí transformaci skóre X na škálu skóre Y pomocí středních hodnot a nazýváme ji vyvažovací funkcí.

Zřejmou vlastností náhodné veličiny Y a transformované veličiny $\varphi_M(X)$ je stejná střední hodnota, neboť platí

$$\mathbb{E}(\varphi_M(X)) = \mathbb{E}(X) - \mu_X + \mu_Y = \mu_X - \mu_X + \mu_Y = \mathbb{E}(Y),$$

kde $\mathbb{E}(\cdot)$ značí operátor střední hodnoty.

V praxi pak odhad vyvažovací funkce φ_M získáme nahrazením středních hodnot μ_X a μ_Y jejich odhady, v tomto případě výběrovými průměry \bar{X}_N a \bar{Y}_M , kde

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i \quad \text{a} \quad \bar{Y}_M = \frac{1}{M} \sum_{i=1}^M Y_i.$$

Dostaneme tak odhad vyvažovací funkce φ_M ve tvaru

$$\hat{\varphi}_M(x) = x - \bar{X}_N + \bar{Y}_M. \quad (1.2)$$

1.2 Lineární vyvažování

Zatímco u vyvažování ve střední hodnotě jsme předpokládali, že rozdíl v obtížnosti mezi verzemi znalostních testů musí být pro všechna skóre konstantní, u lineárního vyvažování již připouštíme variabilitu napříč stupnicí. Proto při vyvažování požadujeme shodnost nejenom středních hodnot, ale také směrodatných odchylek celkového skóre testů. Předpokládáme zde, že distribuční funkce F a G se liší pouze v parametru posunutí a parametru měřítka. Za shodná uvažujme ta skóre, pro která platí

$$\frac{x - \mu_X}{\sigma_X} = \frac{y - \mu_Y}{\sigma_Y},$$

kde μ_X , μ_Y jsou střední hodnoty po řadě náhodných veličin X a Y , σ_X a σ_Y jejich směrodatné odchylky. Vyjádřením y dostaneme

$$\varphi_L(x; \mu_X, \mu_Y, \sigma_X, \sigma_Y) = y = \frac{\sigma_Y}{\sigma_X} x + \left[\mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X \right], \quad (1.3)$$

kde vyvažovací funkce $\varphi_L(x)$ je lineární transformací skóre X na škálu skóre Y .

V tomto případě platí, že se rovnají první dva momenty náhodných veličin Y a $\varphi_L(X)$:

$$\mathbb{E}(\varphi_L(X)) = \frac{\sigma_Y}{\sigma_X} \mathbb{E}(X) + \left[\mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X \right] = \frac{\sigma_Y}{\sigma_X} \mu_X + \left[\mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X \right] = \mu_Y = \mathbb{E}(Y),$$

$$\text{var}(\varphi_L(X)) = \frac{\sigma_Y^2}{\sigma_X^2} \text{var}(X) = \frac{\sigma_Y^2}{\sigma_X^2} \sigma_X^2 = \sigma_Y^2 = \text{var}(Y),$$

kde $\text{var}(\cdot)$ značí operátor rozptylu.

Odhad vyvažovací funkce φ_L získáme nahrazením skutečných parametrů v rovnici (1.3) jejich odhady, tj.

$$\hat{\varphi}_L(x) = \frac{S_{Y,M}}{S_{X,N}} x + \left[\bar{Y}_M - \frac{S_{Y,M}}{S_{X,N}} \bar{X}_N \right], \quad (1.4)$$

kde

$$S_{X,N} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_N)^2} \quad \text{a} \quad S_{Y,M} = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (Y_i - \bar{Y}_M)^2}.$$

1.3 Ekvipercentilové vyvažování

Tato metoda je ze všech tří uváděných metod nejobecnější a neklade žádné předpoklady na tvary distribučních funkcí F a G . V tomto případě považujeme za shodná ta skóre x a y , která splňují

$$F(x) = u = G(y), \quad u \in (0,1).$$

V případě diskrétních rozdělení náhodných veličin X a Y ale nemusí vždy existovat bod, ve kterém by distribuční funkce dosahovaly přesně zvolené hodnoty u . To je problém, který je potřeba řešit a ke kterému se později vrátíme. Nyní pro jednoduchost předpokládejme, že rozdělení veličin X a Y je spojitě a že distribuční funkce F a G jsou spojitě a striktně rostoucí. Potom vyjádřením y dostaneme

$$\varphi_E(x) = y = G^{-1}(F(x)). \quad (1.5)$$

Funkci $\varphi_E(x)$ pak nazýváme ekvipercenilovou vyvažovací funkcí.

Metoda ekvipercenilového vyvažování je také zobecněním metody lineárního vyvažování v tom smyslu, že nyní se rovnají všechny momenty náhodné veličiny Y a transformované veličiny $\varphi_E(X)$. To si zformalizujeme v následujících dvou tvrzeních.

Tvrzení 1. *Nechť $F(x)$ je spojitá a striktně rostoucí distribuční funkce náhodné veličiny X . Potom má veličina $U = F(X)$ rovnoměrné rozdělení na intervalu $(0,1)$.*

Důkaz. Označme F_U distribuční funkci náhodné veličiny $U = F(X)$. Pak pro všechna $u \in (0,1)$ máme

$$F_U(u) = P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u.$$

Protože distribuční funkce jednoznačně určuje pravděpodobnostní rozdělení, má veličina $U = F(X)$ rovnoměrné rozdělení na intervalu $(0,1)$. □

Tvrzení 2. *Nechť $G(y)$ je spojitá a striktně rostoucí distribuční funkce náhodné veličiny Y , $G^{-1}(u)$, $u \in (0,1)$ její inverzní funkce. Dále nechť veličina U má rovnoměrné rozdělení na intervalu $(0,1)$. Potom $V = G^{-1}(U)$ je dobře definovaná a má stejné pravděpodobnostní rozdělení jako náhodná veličina Y .*

Důkaz. Označme F_V distribuční funkci náhodné veličiny V . Potom pro $v \in \mathbb{R}$ platí

$$F_V(v) = P(V \leq v) = P(G^{-1}(U) \leq v) = P(U \leq G(v)) = G(v)$$

a tudíž Y a V mají stejné pravděpodobnostní rozdělení. □

Když za U dosadíme $F(X)$, dostaneme, že $V = G^{-1}(F(X))$ má stejné rozdělení jako veličina Y . V našem případě tedy $\varphi_E(X) = G^{-1}(F(X))$ má stejné rozdělení jako Y , rovnají se tak všechny jejich momenty.

Odhad vyvažovací funkce φ_E z rovnice (1.5) opět dostaneme nahrazením skutečných distribučních funkcí jejich odhady, tj.

$$\hat{\varphi}_E(x) = \hat{G}^{-1}(\hat{F}(x)). \quad (1.6)$$

2. Vyvažování pomocí jádrových odhadů

V této kapitole se budeme zabývat metodou vyvažování pomocí jádrových odhadů (*angl. kernel equating*). V praxi se jedná o velmi často využívanou metodu pro vyvažování testů, zejména pro její flexibilitu. V této práci jí dáme nejvíce prostoru a ze všech uváděných metod se jí budeme věnovat nejpodrobněji. Vyčázejte budeme z knihy von Davier a kol. (2004), ve které je tato metoda popsána uceleně pro všechny designy sběru dat. Podrobný popis jednotlivých kroků pak uvádíme dle článků Holland a Thayer (1989), Holland a Thayer (2000), Holland a Thayer (1987) a Holland a kol. (1989).

Metodu vyvažování pomocí jádrových odhadů můžeme rozdělit na 5 dílčích kroků:

1. Vyhlcení pravděpodobností skóre (*angl. pre-smoothing*)
2. Odhad pravděpodobností skóre
3. Spojitá aproximace skóre (*angl. continuization*)
4. Vyvažování
5. Výpočet standardních chyb vyvažování

Postupně se detailně zaměříme na všechny jednotlivé kroky a popíšeme, jaké statistické metody jsou zde používány. Začneme popisem pro design dvou ekvivalentních skupin (EG), následně představíme metodu pro design neekvivalentních skupin s kotvícími položkami (NEAT).

2.1 Design ekvivalentních skupin

Design ekvivalentních skupin volíme v případě, kdy můžeme předpokládat, že respondenti, kteří vyplňovali test X , i respondenti, kteří vyplňovali test Y , byli vybráni ze stejné homogenní populace. Tuto populaci značíme v celé práci T a nazýváme ji cílovou populací (*angl. target population*). Všechny výsledky jsou potom platné a měly by se interpretovat vůči této populaci. Protože o žádné jiné populaci v této části nemluvíme, nebudeme to explicitně zmiňovat a považujeme to za automatické.

Začneme zavedením používaného značení. Mějme 2 náhodné veličiny X a Y reprezentující dosažené celkové skóre ve 2 různých verzích kognitivního či znalostního testu. K dispozici máme 2 sady dat, X_1, \dots, X_N a Y_1, \dots, Y_M , kde každá realizace náhodné veličiny značí skóre z testu u jednoho konkrétního respondenta. Pozorování jsou vždy v jedné sadě stejně rozdělená a mezi sebou nezávislá. Nechtě maximální dosažené celkové skóre v testu X je $J - 1$, v testu Y je to $K - 1$. Dále předpokládejme, že X_i a Y_i mohou nabývat pouze celočíselných hodnot, ale zato všech z dané škály. Platí tedy, že $X_i \in \{0, 1, \dots, J - 1\}$, $i = 1, \dots, N$

a $Y_i \in \{0, 1, \dots, K - 1\}$, $i = 1, \dots, M$. Označme $x_j = j - 1$, $y_k = k - 1$, $j = 1, \dots, J$, $k = 1, \dots, K$, a to hlavně z důvodu lepší indexace v pozdějších částech práce.

Z těchto dostupných dat pak můžeme vytvořit vektory četností $(n_1, \dots, n_J)^\top$ a $(m_1, \dots, m_K)^\top$, kde

$$\begin{aligned} n_j &= \text{počet respondentů, kteří dosáhli v testu X celkového skóre } x_j, \\ m_k &= \text{počet respondentů, kteří dosáhli v testu Y celkového skóre } y_k, \end{aligned} \quad (2.1)$$

pro $j = 1, \dots, J$, $k = 1, \dots, K$.

Platí, že $\sum_{j=1}^J n_j = N$ a $\sum_{k=1}^K m_k = M$, kde N je celkový počet respondentů, kteří absolvovali test X, podobně M je počet respondentů, kteří absolvovali test Y.

Dále označme pravděpodobnosti jednotlivých skóre jako

$$r_j = P(X = x_j), \quad s_k = P(Y = y_k), \quad (2.2)$$

kde $j \in \{1, \dots, J\}$, $k \in \{1, \dots, K\}$. Zde platí $\sum_{j=1}^J r_j = 1$ a $\sum_{k=1}^K s_k = 1$, jelikož se jedná o pravděpodobnosti. Také předpokládejme, že $r_j > 0 \ \forall j$ a $s_k > 0 \ \forall k$.

Díky nezávislosti pozorování pro jednotlivé studenty tak můžeme uvažovat, že vektory $\mathbf{n} = (n_1, \dots, n_J)^\top$ a $\mathbf{m} = (m_1, \dots, m_K)^\top$ pochází z multinomického rozdělení.

Předpoklad. Vektory $\mathbf{n} = (n_1, \dots, n_J)^\top$ a $\mathbf{m} = (m_1, \dots, m_K)^\top$ jsou nezávislé a každý má multinomické rozdělení, $\mathbf{n} \sim \text{Mult}_J(N, \mathbf{r})$, $\mathbf{m} \sim \text{Mult}_K(M, \mathbf{s})$, kde $\mathbf{r} = (r_1, \dots, r_J)^\top$ a $\mathbf{s} = (s_1, \dots, s_K)^\top$.

Právě pravděpodobnosti r_j a s_k potřebujeme odhadnout, protože budou hrát důležitou roli v dalších krocích.

2.1.1 Vyhazení pravděpodobností skóre

Jednou z možností, jak odhadnout pravděpodobnosti r_j a s_k , je použít relativní četnosti, tj.

$$\hat{r}_j = \frac{n_j}{N} \quad \text{a} \quad \hat{s}_k = \frac{m_k}{M}.$$

V případě multinomického rozdělení se jedná o nestranné a konzistentní odhady parametrů r_j a s_k . Problém by mohl nastat v tom případě, kdy nějaká z pozorovaných četností n_j nebo m_k bude nulová, a tedy i příslušný odhad pravděpodobnosti \hat{r}_j nebo \hat{s}_k bude nulový. To by bylo v rozporu s předpokladem $r_j > 0 \ \forall j$, $s_k > 0 \ \forall k$, tedy tím, že je možné, aby respondent získal v testu celkové skóre od 0 až do maximálního počtu bodů. To je také jeden z důvodů, proč se často získávají odhady pravděpodobností r_j a s_k pomocí log-lineárních, případně i dalších modelů. Získané odhady pak budou nejenom splňovat pozitivitu pravděpodobností, ale také budou hladší a dospějeme tím také ke snížení výběrové variability. My se zaměříme na modely popsané v Holland a Thayer (2000),

Holland a Thayer (1987) a Holland a Thayer (1989).

Řekneme, že vektor $\mathbf{r} = (r_1, \dots, r_J)^\top$ splňuje log-lineární model, pokud pro jeho jednotlivé složky platí

$$\log(r_j) = \alpha + u_j + \mathbf{b}_j^\top \boldsymbol{\beta}, \quad (2.3)$$

kde u_j je známá konstanta, \mathbf{b}_j^\top je řádkový vektor známých konstant, $\boldsymbol{\beta}$ je sloupcový vektor parametrů, které odhadujeme, a α je konstanta, která zajišťuje, aby součet všech r_j , $j = 1, \dots, J$, se rovnal 1. Vyjádřením dostaneme

$$\alpha = -\log \left(\sum_{j=1}^J \exp(u_j + \mathbf{b}_j^\top \boldsymbol{\beta}) \right). \quad (2.4)$$

Význam konstanty u_j je specifikovat pravděpodobnosti r_j v případě, kdy $\boldsymbol{\beta} = \mathbf{0}$. Častou volbou může být $u_j = 0$. Délku vektorů \mathbf{b}_j a $\boldsymbol{\beta}$ označme p_r , jedná se o počet parametrů potřebných k odhadnutí r_j .

Model můžeme zapsat také v maticovém zápisu, který má tvar

$$\log(\mathbf{r}) = \alpha + \mathbf{u} + \mathbf{B}^\top \boldsymbol{\beta}, \quad (2.5)$$

kde \mathbf{u} je vektor o rozměrech $J \times 1$, $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_J)$ matice o rozměrech $p_r \times J$ a $\boldsymbol{\beta}$ a α jsou stejné jako v předchozím případě.

Jedním z příkladů log-lineárního modelu, který se často používá, je momentový model. Ten využívá jako vektory \mathbf{b}_j mocniny skóre x_j , tedy

$$\log(r_j) = \alpha + u_j + \sum_{i=1}^{p_r} x_j^i \beta_i, \quad j = 1, \dots, J. \quad (2.6)$$

Výhodou log-lineárního modelu je to, že nám zajistí, aby se rovnaly a -té výběrové momenty a momenty rozdělení, které je určené odhadnutými pravděpodobnostmi. V případě momentového modelu z rovnice (2.6) tedy platí

$$\sum_{j=1}^J x_j^a \hat{r}_j = \sum_{j=1}^J x_j^a \frac{n_j}{N}, \quad (2.7)$$

kde $a \in \{1, \dots, p_r\}$ a \hat{r}_j splňuje $\log(\hat{r}_j) = \alpha + u_j + \sum_{i=1}^{p_r} x_j^i \hat{\beta}_i$, kde $\hat{\beta}_i$ je maximálně věrohodný odhad β_i , $i = 1, \dots, p_r$. Ten získáme vyřešením rovnic

$$\frac{\partial l_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0},$$

kde $l_N(\boldsymbol{\beta})$ značí logaritickou věrohodnost, která má v tomto případě tvar

$$\begin{aligned} l_N(\boldsymbol{\beta}) &= \sum_{j=1}^J n_j \log(r_j) + C \\ &= \sum_{j=1}^J n_j (\alpha + u_j + \mathbf{b}_j^\top \boldsymbol{\beta}) + C, \end{aligned} \quad (2.8)$$

kde $\mathbf{b}_j = (x_j^1, \dots, x_j^{p_r})^\top$ a C je konstanta.

Ve většině případů neexistuje explicitní řešení těchto rovnic a maximálně věrohodné odhady musí být nalezeny pomocí numerických iterativních metod. Mezi nejčastěji používané metody patří například Newton-Raphsonova metoda nebo Fisherova skórovací metoda.

Další možné varianty log-lineárních modelů nalezneme v Holland a Thayer (2000, str. 138 – 142).

Při volbě vhodného modelu, který co nejlépe popisuje data, se můžeme řídit více kritérii. Jedná se o obecný problém výběru modelu, ke kterému existuje řada metod a přístupů. Vždy záleží na konkrétní situaci, struktuře dat, zvoleném designu a podobně. Pro grafické znázornění si můžeme vykreslit původní pozorované frekvence n_j a porovnat je s odhadnutými frekvencemi $N\hat{r}_j$. Pro formálnější přístup můžeme spočítat hodnoty deviance, případně vhodně použít test založený na poměru věrohodností (např. v případě, kdy testujeme, jestli je vhodné do momentového modelu zařadit ještě vyšší momenty). V knize Holland a Thayer (2000) se také uvádí jako jeden z diagnostických nástrojů použití Freeman-Tukeyových reziduí. Ta jsou definována jako

$$FT_j = \sqrt{n_j} + \sqrt{n_j + 1} - \sqrt{4N\hat{r}_j + 1}. \quad (2.9)$$

V případě, že model dobře popisuje data a četnosti mají opravdu Poissonovské rozdělení, mají Freeman-Tukeyova rezidua normální rozdělení se střední hodnotou 0 a rozptylem 1. Pro doplnění lze také spočítat hodnoty některých informačních kritérií, například Akaikeho nebo Bayesovského informačního kritéria.

Po nalezení vhodného modelu pak získáme odhad vektoru pravděpodobností $\hat{\mathbf{r}} = (\hat{r}_1, \dots, \hat{r}_J)$ a zcela obdobně také odhad vektoru $\hat{\mathbf{s}} = (\hat{s}_1, \dots, \hat{s}_K)$.

2.1.2 Odhad pravděpodobností skóre

V případě designu dvou ekvivalentních skupin je tento krok zcela zřejmý. Získané odhady pravděpodobností \hat{r}_j a \hat{s}_k v předchozí části jsou přímo odhady v cílové populaci T , které potřebujeme. Žádný dodatečný krok tedy již není potřeba. V komplikovanějších designech, především v případě neekvivalentních skupin, se pak jedná již o náročnější postup, který vyžaduje i další doplňující předpoklady, aby bylo možné jej provést.

2.1.3 Spojitá aproximace skóre

Když už máme odhady pravděpodobností skóre $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$, dalším krokem je získat odhady distribučních funkcí náhodných veličin X a Y . Vzhledem k tomu, že veličina X může nabývat pouze celočíselných hodnot $0, \dots, J - 1$, veličina Y pouze hodnot $0, \dots, K - 1$, pravděpodobnostní rozdělení těchto náhodných veličin je diskrétní. Jejich distribuční funkce mají tvar

$$\begin{aligned} F(x) &= P(X \leq x) = \sum_{j:j \leq x} r_j, \\ G(y) &= P(Y \leq y) = \sum_{k:k \leq y} s_k. \end{aligned} \quad (2.10)$$

Odhady bychom získali dosazením \hat{r}_j a \hat{s}_k . Jak už jsme ale zmínili v sekci 1.3, v případě diskretních rozdělání zde nastává problém, že nemusí vždy existovat inverzní funkce G^{-1} a vyvažovací funkce $\varphi_E(x) = G^{-1}(F(x))$ by nebyla řádně definována (což platí i pro vyvažovací funkci $\varphi_E(y) = F^{-1}(G(y))$). Z tohoto důvodu je potřeba distribuční funkce F a G aproximovat spojitými verzemi. Použitá metoda je popsána v Holland a Thayer (1989) a von Davier a kol. (2004). Myšlenka je nahradit diskretní náhodné veličiny X a Y jejich spojitými aproximacemi a místo F a G pak použít jejich distribuční funkce, které už budou spojitě a jejich inverzní funkce tak bude vždy dobře definována.

Nechť X je diskretní náhodná veličina, V je spojitá náhodná veličina se standardizovaným normálním rozdělením $N(0,1)$, která je na X nezávislá, a necht $h_X > 0$. Označme μ_X střední hodnotu veličiny X , σ_X^2 její rozptyl. Potom uvažujme novou náhodnou veličinu $X + h_X V$. Tato náhodná veličina už je spojitá (protože V je spojitá) a má stejnou střední hodnotu jako veličina X , neboť platí

$$\mathbb{E}(X + h_X V) = \mathbb{E}X + h_X \mathbb{E}V = \mu_X.$$

Rozptyl těchto veličin se však již nerovná, neboť

$$\text{var}(X + h_X V) = \sigma_X^2 + h_X^2 \neq \sigma_X^2 \text{ pro } h_X > 0.$$

To nás vede k lineární transformaci veličiny $X + h_X V$, kterou definujeme jako

$$\tilde{X}(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X, \quad (2.11)$$

kde $a_X = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + h_X^2}}$.

Takto definovaná náhodná veličina už má stejnou střední hodnotu i rozptyl jako veličina X , protože platí

$$\mathbb{E}(\tilde{X}(h_X)) = \mathbb{E}[a_X(X + h_X V) + (1 - a_X)\mu_X] = a_X\mu_X + \mu_X - a_X\mu_X = \mu_X$$

a

$$\begin{aligned} \text{var}(\tilde{X}(h_X)) &= \text{var}[a_X(X + h_X V) + (1 - a_X)\mu_X] = \text{var}[a_X(X + h_X V)] = \\ &= a_X^2(\sigma_X^2 + h_X^2) = \frac{\sigma_X^2}{\sigma_X^2 + h_X^2}(\sigma_X^2 + h_X^2) = \sigma_X^2. \end{aligned}$$

Právě distribuční funkci této nově definované veličiny $\tilde{X}(h_X)$ budeme považovat za spojitou aproximaci distribuční funkce F a budeme s ní dále pracovat.

Věta 3. (von Davier a kol. (2004), str. 58)

Nechť $\tilde{X}(h_X)$ je náhodná veličina definována vztahem (2.11) a necht $F_{\tilde{X}}$ je její distribuční funkce. Potom $F_{\tilde{X}}$ je spojitá a má tvar

$$F_{\tilde{X}}(x) = \sum_{j=1}^J r_j \Phi\left(\frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}\right), \quad (2.12)$$

kde Φ je distribuční funkce standardizovaného normálního rozdělení $N(0,1)$.

Důkaz.

$$\begin{aligned}
F_{\tilde{X}}(x) &= P\left(\tilde{X}(h_X)\right) \leq x \\
&= P\left(a_X(X + h_X V) + (1 - a_X)\mu_X \leq x\right) \\
&= P\left(V \leq \frac{x - a_X X - (1 - a_X)\mu_X}{a_X h_X}\right) \\
&= \sum_{j=1}^J r_j P\left(V \leq \frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X} \mid X = x_j\right) \\
&= \sum_{j=1}^J r_j \Phi\left(\frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}\right).
\end{aligned}$$

Vidíme, že se jedná o vážený průměr spojitých distribučních funkcí normálního rozdělení, $F_{\tilde{X}}(x)$ je tedy také spojitá. \square

Odhad této distribuční funkce získáme dosazením odhadnutých pravděpodobností \hat{r}_j do vzorce (2.12), tj.

$$\hat{F}_{\tilde{X}}(x) = F_{\tilde{X}}(x; \hat{\mathbf{r}}) = \sum_{j=1}^J \hat{r}_j \Phi\left(\frac{x - \hat{a}_X x_j - (1 - \hat{a}_X)\hat{\mu}_X}{\hat{a}_X h_X}\right), \quad (2.13)$$

kde

$$\hat{\mu}_X = \sum_{j=1}^J x_j \hat{r}_j, \quad \hat{\sigma}_X^2 = \sum_{j=1}^J \hat{r}_j (x_j - \hat{\mu}_X)^2, \quad \hat{a}_X = \sqrt{\frac{\hat{\sigma}_X^2}{\hat{\sigma}_X^2 + h_X^2}}.$$

V tomto případě jsme jako aproximaci veličiny X zvolili $\tilde{X}(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X$, kde V měla normální rozdělení $N(0,1)$. Obecně se však nemusí jednat pouze o normální rozdělení, ale můžeme volit i nějaké jiné rozdělení, které je spojitě a symetrické kolem nuly. Nechť $EV = 0$, $\text{var}(V) = \sigma_V^2$ a K_V je distribuční funkce V . Potom by $F_{\tilde{X}}$ měla tvar

$$F_{\tilde{X}}(x) = \sum_{j=1}^J r_j K_V\left(\frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}\right), \quad (2.14)$$

kde $a_X = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + \sigma_V^2 h_X^2}}$.

Typickými dalšími volbami jsou logistické a rovnoměrné rozdělení, s distribučními funkcemi

$$K_V(x) = \frac{1}{1 + e^{-\frac{x}{s}}}, \quad s \in \mathbb{R}, \quad s > 0$$

pro logistické rozdělení a

$$K_V(x) = \begin{cases} 0 & \text{pro } x < -b \\ \frac{x+b}{2b} & \text{pro } -b \leq x \leq b, \\ 1 & \text{pro } x > b \end{cases} \quad b \in \mathbb{R}, \quad b > 0$$

pro rovnoměrné rozdělení.

Diskuze věnována použití těchto rozdělení je uvedena například v knize Davier (2011), v kapitole 10. My budeme v této práci dále uvažovat už jen náhodnou veličinu V se standardizovaným normálním rozdělením, a tedy distribuční funkci veličiny $\widetilde{X}(h_X)$ ve tvaru (2.12).

Výsledná distribuční funkce $F_{\widetilde{X}}$ také záleží na volbě parametru h_X . Asi nejjednodušším přístupem je h_X zvolit jako pevně danou, předem specifikovanou hodnotu. Samozřejmě čím menší h_X bude, tím blíže bude rozdělení $\widetilde{X}(h_X)$ k původnímu rozdělení X . Pokud však bude h_X příliš malé, nemusí být hustota $f_{\widetilde{X}} = F'_{\widetilde{X}}$ dostatečně vyhlazená a může obsahovat mnoho nerovností. Pro nalezení vhodného h_X je proto potřeba nalézt balanc mezi těmito dvěma vlastnostmi a nemusí být hned zřejmé, jaká hodnota je optimální. Jednou z možností, jak vhodné h_X nalézt, je například použití penalizačních funkcí a jejich minimalizace. Uvažujme

$$\text{PEN}_1(h_X) = \sum_{j=1}^J (\widehat{r}_j - \widehat{f}_{\widetilde{X}}(x_j))^2, \quad (2.15)$$

kde

$$\widehat{f}_{\widetilde{X}}(x) = \sum_{j=1}^J \widehat{r}_j \phi \left(\frac{x - \widehat{a}_X x_j - (1 - \widehat{a}_X) \widehat{\mu}_X}{\widehat{a}_X h_X} \right) \frac{1}{\widehat{a}_X h_X}$$

a ϕ je hustota standardizovaného normálního rozdělení $N(0,1)$. V tomto případě se díváme, jak se liší hodnota hustoty veličiny $\widetilde{X}(h_X)$ v bodě x_j od odhadnuté pravděpodobnosti \widehat{r}_j . Pokud odhadnutá hustota kopíruje požadovaný tvar vykreslených pravděpodobností (představíme-li si je jako histogram, na ose x bodovou škálu $0, \dots, J-1$ a na ose y hodnotu pravděpodobnosti), potom by tyto hodnoty měly být podobné. Za optimální h_X pak zvolíme tu hodnotu, která funkci $\text{PEN}_1(h_X)$ minimalizuje.

Volit můžeme také další penalizační funkce, které chceme minimalizovat, případně i jejich kombinace. Další funkce uvedená v knize von Davier a kol. (2004) je například funkce, která navíc penalizuje ty hodnoty h_X , které umožňují derivaci hustoty $f'_{\widetilde{X}}$ měnit znaménka v okolí bodů x_j , tedy zabraňuje, aby výsledná hustota měla příliš mnoho lokálních maxim. Penalizační funkce má tvar

$$\text{PEN}_2(h_X) = \sum_{j=1}^J A_j, \quad (2.16)$$

kde A_j nabývá hodnoty 1, pokud platí $\left[\left(\widehat{f}'_{\widetilde{X}}(x_j - \omega) > 0 \right) \cup \left(\widehat{f}'_{\widetilde{X}}(x_j + \omega) < 0 \right) \right]$ nebo $\left[\left(\widehat{f}'_{\widetilde{X}}(x_j - \omega) < 0 \right) \cup \left(\widehat{f}'_{\widetilde{X}}(x_j + \omega) > 0 \right) \right]$, v ostatních případech bude nabývat hodnoty 0. Konstantu ω typicky volíme jako $\omega = \frac{1}{4}$. Tyto dvě výše zmíněné penalizační funkce můžeme kombinovat, tj. volit

$$\text{PEN}(h_X) = \text{PEN}_1(h_X) + \kappa \text{PEN}_2(h_X), \quad (2.17)$$

kde κ je konstanta, například $\kappa = 1$. První část penalizační funkce nám zajišťuje požadovaný tvar rozdělení a druhá část zajišťuje, aby bylo dostatečně hladké.

Možných metod k nalezení vhodné hodnoty parametru h_X je ale celá řada. Další může být například normální referenční pravidlo neboli tzv. Silvermenovo pravidlo (Andersson a von Davier (2014)), metoda krosvalidace (Liang a von Davier (2014)) nebo tzv. metoda dvojího vyhlazení (Häggström a Wiberg (2014)).

V celé sekci jsme pracovali jen s náhodnou veličinou X , zcela totožně však vše uvažujeme i pro náhodnou veličinou Y . Pro úplnost uvedme nejdůležitější značení a vzorce. Uvažovat budeme spojitou náhodnou veličinou $\tilde{Y}(h_Y) = a_Y(Y + h_Y W) + (1 - a_Y)\mu_Y$, kde $\mathbb{E}Y = \mu_Y$, $\text{var}(Y) = \sigma_Y^2$, $h_Y > 0$, W je nezávislá na Y a má standardizované normální rozdělení $N(0,1)$, $a_Y = \sqrt{\frac{\sigma_Y^2}{\sigma_Y^2 + h_Y^2}}$. Distribuční funkce veličiny $\tilde{Y}(h_Y)$ má tvar

$$G_{\tilde{Y}}(y) = \sum_{k=1}^K s_k \Phi \left(\frac{y - a_Y y_k - (1 - a_Y)\mu_Y}{a_Y h_Y} \right), \quad (2.18)$$

její odhad

$$\hat{G}_{\tilde{Y}}(y) = G_{\tilde{Y}}(y; \hat{\mathbf{s}}) = \sum_{k=1}^K \hat{s}_k \Phi \left(\frac{y - \hat{a}_Y y_k - (1 - \hat{a}_Y)\hat{\mu}_Y}{\hat{a}_Y h_Y} \right), \quad (2.19)$$

kde

$$\hat{\mu}_Y = \sum_{k=1}^K y_k \hat{s}_k, \quad \hat{\sigma}_Y^2 = \sum_{k=1}^K \hat{s}_k (y_k - \hat{\mu}_Y)^2, \quad \hat{a}_Y = \sqrt{\frac{\hat{\sigma}_Y^2}{\hat{\sigma}_Y^2 + h_Y^2}}.$$

2.1.4 Vyvažování

V předchozím kroku jsme získali odhady spojitých distribučních funkcí $\hat{F}_{\tilde{X}}$ a $\hat{G}_{\tilde{Y}}$. Ty už přímo využijeme k výpočtu odhadu vyvažovací funkce, a to dosazením do vzorce (1.5), kde jimi nahradíme distribuční funkce F a G . Metoda vyvažování pomocí jádrových odhadů je tak jedním ze speciálních případů ekvipercenilového vyvažování.

Vyvažovací funkce metody vyvažování za pomoci jádrových odhadů (kterou nyní indexujeme K , odkazující na *kernel*) má tvar

$$\varphi_K(x) = G_{\tilde{Y}}^{-1} \left(F_{\tilde{X}}(x) \right). \quad (2.20)$$

Její odhad pak dostaneme dosazením odhadů $\hat{F}_{\tilde{X}}$ a $\hat{G}_{\tilde{Y}}$ (viz. rovnice (2.13) a (2.19)), tj.

$$\hat{\varphi}_K(x) = \varphi_K(x; \hat{F}_{\tilde{X}}, \hat{G}_{\tilde{Y}}) = \hat{G}_{\tilde{Y}}^{-1} \left(\hat{F}_{\tilde{X}}(x) \right). \quad (2.21)$$

2.1.5 Výpočet standardních chyb vyvažování

Posledním krokem v metodě vyvažování pomocí jádrových odhadů je vyhodnocení přesnosti celého procesu a výpočet standardních chyb. V tomto případě definujeme standardní chybu vyvažování jako odmocninu z rozptylu odhadu vyvažovací funkce, tedy

$$\text{SEE}_Y(x) = \sqrt{\text{var}(\hat{\varphi}_K(x))}. \quad (2.22)$$

K tomu, abychom mohli SEE_Y spočítat, tedy potřebujeme znát $\text{var}(\hat{\varphi}_K)$. Protože odhad vyvažovací funkce $\hat{\varphi}_K$ je definován jako

$$\hat{\varphi}_K(x) = \hat{G}_{\hat{Y}}^{-1} \left(\hat{F}_{\hat{X}}(x) \right),$$

kde $\hat{F}_{\hat{X}}$ a $\hat{G}_{\hat{Y}}$ jsou diferencovatelné funkce parametrů $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$, můžeme výpočet provést pomocí delta metody (její znění bude uvedeno ve Větě 5). Nejprve však musíme znát rozdělení odhadů $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$, konkrétně jejich varianční matice. Vraťme se tedy k sekci 2.1.1 a k tomu, jak byly odhady zkonstruovány. Postup provedeme pro vektor \mathbf{r} , pro \mathbf{s} by byl opět zcela obdobný.

Pravděpodobnosti \mathbf{r} byly odhadnuty pomocí log-lineárního modelu

$$\log(\mathbf{r}) = \alpha + \mathbf{u} + \mathbf{B}^\top \boldsymbol{\beta},$$

který je podrobně popsán v (2.5). Odhadnuté pravděpodobnosti $\hat{r}_j, j = 1, \dots, J$, tak mají tvar

$$\hat{r}_j = \frac{\exp(u_j + \mathbf{b}_j^\top \hat{\boldsymbol{\beta}})}{\sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \hat{\boldsymbol{\beta}})}, \quad (2.23)$$

kde $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_{p_r})^\top$ jsou maximálně věrohodné odhady. Pravděpodobnost \hat{r}_j je tak diferencovatelnou funkcí parametru $\hat{\boldsymbol{\beta}}$, pro zjištění asymptotického rozdělení můžeme proto opět využít delta metodu.

Asymptotické rozdělení $\hat{\boldsymbol{\beta}}$

Asymptotické rozdělení $\hat{\boldsymbol{\beta}}$ odvodíme díky známé teorii maximálně věrohodných odhadů. Ta nám říká, že při splnění podmínek regularity platí

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{p_r}(\mathbf{0}_{p_r}, I^{-1}(\boldsymbol{\beta})), \quad (2.24)$$

kde $I(\boldsymbol{\beta})$ je Fisherova informační matice. Jako její odhad pak můžeme například použít pozorovanou informační matici $I_N(\hat{\boldsymbol{\beta}})$, kde

$$I_N(\boldsymbol{\beta}) = -\frac{1}{N} \left[\frac{\partial^2 l_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right], \quad (2.25)$$

kde $l_N(\boldsymbol{\beta})$ je logaritmičká věrohodnost definovaná v (2.8). K výpočtu maximálně věrohodného odhadu potřebujeme znát první derivaci, k výpočtu varianční matice i druhou derivaci logaritmičké věrohodnosti. Jejich tvary uvedeme v následující větě.

Věta 4. Pro log-lineární model definovaný v (2.5) a logaritmicou věrohodnost (2.8) platí

$$(a) \quad \frac{\partial l_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{B}(\mathbf{n} - N\mathbf{r}),$$

$$(b) \quad \frac{\partial^2 l_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = -N(\mathbf{B}(\text{diag}(\mathbf{r}) - \mathbf{r}\mathbf{r}^\top) \mathbf{B}^\top),$$

kde \mathbf{n} je vektor četností, \mathbf{r} vektor pravděpodobností, N počet pozorování a \mathbf{B} je matice známých konstant definovaná v modelu (2.5).

Důkaz. V důkazu odvodíme výrazy derivováním věrohodnosti a následným zápisem v maticové formě. Podobně je důkaz uveden i v Holland a Thayer (1987, str. 5).

(a) Logaritmicou věrohodnost má tvar $l_N = \sum_{j=1}^J n_j \log(r_j) + C$ a platí, že

$$\frac{\partial l_N}{\partial \beta_k} = \sum_{j=1}^J n_j \frac{\partial \log(r_j)}{\beta_k}. \quad (2.26)$$

Dále platí $r_j = \frac{\exp(u_j + \mathbf{b}_j^\top \boldsymbol{\beta})}{\sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \boldsymbol{\beta})}$, a odtud

$$\frac{\partial \log(r_j)}{\beta_k} = b_{jk} - \frac{\sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \boldsymbol{\beta}) b_{ik}}{\sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \boldsymbol{\beta})} = b_{jk} - \sum_{i=1}^J b_{ik} r_i.$$

Dosazením do (2.26) dostaneme

$$\frac{\partial l_N}{\partial \beta_k} = \sum_{j=1}^J n_j \left(b_{jk} - \sum_{i=1}^J b_{ik} r_i \right) = \sum_{j=1}^J n_j b_{jk} - \sum_{j=1}^J n_j \sum_{i=1}^J b_{ik} r_i = \sum_{j=1}^J b_{jk} (n_j - N r_j).$$

V maticové formě pro vektor $\boldsymbol{\beta}$ máme

$$\frac{\partial l_N}{\partial \boldsymbol{\beta}} = \mathbf{B}(\mathbf{n} - N\mathbf{r}).$$

(b) Druhou derivaci logaritmicou věrohodnosti spočteme jako

$$\frac{\partial^2 l_N}{\partial \beta_r \partial \beta_k} = \frac{\partial}{\partial \beta_r} \left(\sum_{j=1}^J b_{jk} (n_j - N r_j) \right) = - \sum_{j=1}^J b_{jk} N \frac{\partial r_j}{\partial \beta_r}. \quad (2.27)$$

Platí, že

$$\begin{aligned} \frac{\partial r_j}{\partial \beta_r} &= \frac{\exp(u_j + \mathbf{b}_j^\top \boldsymbol{\beta}) (b_{jr} \sum_{l=1}^J \exp(u_l + \mathbf{b}_l^\top \boldsymbol{\beta}) - \sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \boldsymbol{\beta}) b_{ir})}{\left(\sum_{l=1}^J \exp(u_l + \mathbf{b}_l^\top \boldsymbol{\beta}) \right)^2} \\ &= \frac{\exp(u_j + \mathbf{b}_j^\top \boldsymbol{\beta})}{\sum_{l=1}^J \exp(u_l + \mathbf{b}_l^\top \boldsymbol{\beta})} \left(b_{jr} - \frac{\sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \boldsymbol{\beta}) b_{ir}}{\sum_{l=1}^J \exp(u_l + \mathbf{b}_l^\top \boldsymbol{\beta})} \right) \\ &= r_j \left(b_{jr} - \sum_{i=1}^J b_{ir} r_i \right). \end{aligned} \quad (2.28)$$

Dosazením do (2.27) získáme

$$\begin{aligned}\frac{\partial^2 l_N}{\partial \beta_r \partial \beta_k} &= - \sum_{j=1}^J b_{jk} N \left[r_j \left(b_{jr} - \sum_{i=1}^J b_{ir} r_i \right) \right] \\ &= -N \left[\sum_{j=1}^J b_{jk} r_j b_{jr} - \left(\sum_{j=1}^J b_{jk} r_j \right) \left(\sum_{i=1}^J b_{ir} r_i \right) \right].\end{aligned}$$

V maticovém zápisu tak můžeme vyjádřit druhou derivaci ve tvaru

$$\frac{\partial^2 l_N}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = -N(\mathbf{B} (\text{diag}(\mathbf{r}) - \mathbf{r} \mathbf{r}^\top) \mathbf{B}^\top).$$

□

Když označíme matici $\mathbf{R} = \text{diag}(\mathbf{r}) - \mathbf{r} \mathbf{r}^\top$, dostaneme, že platí

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{p_r} \left(\mathbf{0}_{p_r}, (\mathbf{B} \mathbf{R} \mathbf{B}^\top)^{-1} \right). \quad (2.29)$$

Asymptotické rozdělení $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$

Nyní už tedy známe tvar asymptotické varianční matice vektoru $\hat{\boldsymbol{\beta}}$ a z něj můžeme odvodit asymptotickou varianční matici vektoru $\hat{\mathbf{r}}$ pomocí delta metody.

Věta 5. (*Delta věta*)

Nechť $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n1}, \dots, \hat{\theta}_{np})^\top \in \mathbb{R}^p$ je náhodný vektor, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top \in \mathbb{R}^p$ je vektor konstant a nechť platí

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Sigma}).$$

Dále nechť $\mathbf{g} = (g_1, \dots, g_m)^\top$ je funkce z \mathbb{R}^p do \mathbb{R}^m , která je spojitě diferencovatelná na nějakém okolí bodu $\boldsymbol{\theta}$. Označme Jacobiho matici funkce \mathbf{g} v bodě \mathbf{x} jako

$$\mathbb{J}_{\mathbf{g}}(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial g_1(\mathbf{x})}{\partial x_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial g_m(\mathbf{x})}{\partial x_p} \end{pmatrix}.$$

Potom platí

$$\sqrt{n}(\mathbf{g}(\hat{\boldsymbol{\theta}}_n) - \mathbf{g}(\boldsymbol{\theta})) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_m(\mathbf{0}_m, \mathbb{J}_{\mathbf{g}}(\boldsymbol{\theta}) \boldsymbol{\Sigma} \mathbb{J}_{\mathbf{g}}(\boldsymbol{\theta})^\top).$$

Důkaz. Důkaz je uveden například ve van der Vaart (2000).

□

V našem konkrétním případě máme $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\beta}}$ a funkce \mathbf{g} je $\mathbf{g} = (g_1, \dots, g_J)^\top = (r_1, \dots, r_J)^\top$, kde

$$r_j(\boldsymbol{\beta}) = \frac{\exp(u_j + \mathbf{b}_j^\top \boldsymbol{\beta})}{\sum_{i=1}^J \exp(u_i + \mathbf{b}_i^\top \boldsymbol{\beta})}.$$

Potřebujeme tedy znát Jacobiho matici ve tvaru

$$\mathbb{J}_{\mathbf{r}}(\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial r_1(\boldsymbol{\beta})}{\partial \beta_1} & \dots & \frac{\partial r_1(\boldsymbol{\beta})}{\partial \beta_{p_r}} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_J(\boldsymbol{\beta})}{\partial \beta_1} & \dots & \frac{\partial r_J(\boldsymbol{\beta})}{\partial \beta_{p_r}} \end{pmatrix}.$$

Platí, že

$$\frac{\partial r_j}{\partial \beta_k} = r_j \left(b_{jk} - \sum_{i=1}^J b_{ik} r_i \right),$$

jak bylo uvedeno v (2.28) v důkazu Věty 4. Jacobiho matice $\mathbb{J}_{\mathbf{r}}$ má potom tvar

$$\mathbb{J}_{\mathbf{r}}(\boldsymbol{\beta}) = \text{diag}(\mathbf{r})\mathbf{B}^\top - \mathbf{r}\mathbf{r}^\top\mathbf{B}^\top = (\text{diag}(\mathbf{r}) - \mathbf{r}\mathbf{r}^\top)\mathbf{B}^\top = \mathbf{R}\mathbf{B}^\top. \quad (2.30)$$

Z delta věty tedy potom dostáváme asymptotické rozdělení odhadu $\hat{\mathbf{r}}$,

$$\sqrt{N}(\hat{\mathbf{r}} - \mathbf{r}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_J\left(\mathbf{0}_J, \mathbf{R}\mathbf{B}^\top(\mathbf{B}\mathbf{R}\mathbf{B}^\top)^{-1}\mathbf{B}\mathbf{R}^\top\right). \quad (2.31)$$

Označme asymptotickou varianční matici vektoru $\hat{\mathbf{r}}$ symbolem $\boldsymbol{\Sigma}_{\hat{\mathbf{r}}}$, kde

$$\boldsymbol{\Sigma}_{\hat{\mathbf{r}}} = \mathbf{R}\mathbf{B}^\top(\mathbf{B}\mathbf{R}\mathbf{B}^\top)^{-1}\mathbf{B}\mathbf{R}^\top. \quad (2.32)$$

Vzorec (2.32) ale není obecně vhodné použít k výpočtu varianční matice $\boldsymbol{\Sigma}_{\hat{\mathbf{r}}}$, protože rozměry matice jsou $J \times J$ a J může být v nějakých konkrétních případech dost vysoké. Potom výpočet, ale hlavně i uložení matice, by mohlo být poměrně náročné. Vhodnější způsob jak spočítat prvky matice $\boldsymbol{\Sigma}_{\hat{\mathbf{r}}}$ je uveden v následující větě. Věta je také v mírně upravené verzi uvedena v von Davier a kol. (2004, str. 52).

Věta 6. *Nechť $\hat{\mathbf{r}}$ je maximálně věrohodný odhad z log-lineárního modelu (2.5) pro \mathbf{r} , necht $\boldsymbol{\Sigma}_{\hat{\mathbf{r}}}$ je jeho varianční matice. Potom $\boldsymbol{\Sigma}_{\hat{\mathbf{r}}}$ můžeme vyjádřit jako*

$$\boldsymbol{\Sigma}_{\hat{\mathbf{r}}} = \mathbf{C}_{\mathbf{r}}\mathbf{C}_{\mathbf{r}}^\top, \quad (2.33)$$

kde $\mathbf{C}_{\mathbf{r}}$ je matice typu $J \times p_r$ a platí

- (a) $\mathbf{C}_{\mathbf{r}} = \mathbf{D}_{\sqrt{\mathbf{r}}}\mathbf{Q}$, kde \mathbf{Q} je matice, pro kterou platí
- (b) $\mathbf{Q}\mathbf{R} = (\mathbf{D}_{\sqrt{\mathbf{r}}} - \sqrt{\mathbf{r}\mathbf{r}^\top})\mathbf{B}^\top$,

kde \mathbf{Q} je matice typu $J \times p_r$ s ortogonálními sloupci, \mathbf{R} je horní trojúhelníková matice typu $p_r \times p_r$, \mathbf{B} je matice z (2.5) a $\mathbf{D}_{\sqrt{\mathbf{r}}}$ je diagonální matice, která má na diagonále prvky $\sqrt{r_j}$.

Důkaz. Důkaz věty je uveden v Holland a Thayer (1987), str. 18. □

Při použití vzorce (2.33) nám k uložení stačí matice $\mathbf{C}_{\mathbf{r}}$, která má rozměry $J \times p_r$, kde p_r je počet parametrů v modelu (2.5), tedy obvykle výrazně nižší číslo než je J .

Analogicky bychom získali asymptotickou varianční matici vektoru $\hat{\mathbf{s}}$ ve tvaru $\Sigma_{\hat{\mathbf{s}}} = \mathbf{C}_s \mathbf{C}_s^\top$.

Platí tedy, že

$$\sqrt{N}(\hat{\mathbf{r}} - \mathbf{r}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_J(\mathbf{0}_J, \mathbf{C}_r \mathbf{C}_r^\top) \quad (2.34)$$

a

$$\sqrt{M}(\hat{\mathbf{s}} - \mathbf{s}) \xrightarrow[M \rightarrow \infty]{d} \mathcal{N}_K(\mathbf{0}_K, \mathbf{C}_s \mathbf{C}_s^\top). \quad (2.35)$$

Budeme-li navíc předpokládat, že $\frac{M}{N} \rightarrow z \in (0, \infty)$, potom platí také

$$\sqrt{N}(\hat{\mathbf{s}} - \mathbf{s}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_K\left(\mathbf{0}_K, \frac{1}{z} \mathbf{C}_s \mathbf{C}_s^\top\right). \quad (2.36)$$

Z (2.34), (2.36) a díky nezávislosti $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$ dostaneme

$$\sqrt{N} \left(\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{s}} \end{pmatrix} - \begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix} \right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{J+K} \left(\begin{pmatrix} \mathbf{0}_J \\ \mathbf{0}_K \end{pmatrix}, \begin{pmatrix} \mathbf{C}_r \mathbf{C}_r^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_s \mathbf{C}_s^\top \end{pmatrix} \right). \quad (2.37)$$

Asymptotické rozdělení $\hat{\varphi}_K$

Nyní, když už známe varianční matici vektoru odhadů $(\hat{\mathbf{r}}, \hat{\mathbf{s}})^\top$, můžeme se vrátit k výpočtu standardních chyb vyvažování SEE_Y , které jsme definovali v (2.22). Ke spočtení rozptylu odhadu vyvažovací funkce $\hat{\varphi}_K$ opět použijeme delta větu, viz. Věta 5. Potřebujeme tedy znát Jacobiho matici, v tomto případě řádkový vektor délky $J + K$,

$$\mathbb{J}_{\varphi_K} = \left(\frac{\partial \varphi_K}{\partial r_1}, \dots, \frac{\partial \varphi_K}{\partial r_J}, \frac{\partial \varphi_K}{\partial s_1}, \dots, \frac{\partial \varphi_K}{\partial s_K} \right)^\top. \quad (2.38)$$

Vyvažovací funkce φ_K má tvar

$$\varphi_K(x) = G_{\tilde{Y}}^{-1} \left(F_{\tilde{X}}(x) \right),$$

podle pravidla pro derivování inverzní funkce tedy dostaneme, že

$$\frac{\partial \varphi_K}{\partial r_j}(x) = \frac{1}{G'_{\tilde{Y}}} \frac{\partial F_{\tilde{X}}(x)}{\partial r_j} \quad \text{a} \quad (2.39)$$

$$\frac{\partial \varphi_K}{\partial s_k}(x) = -\frac{1}{G'_{\tilde{Y}}} \frac{\partial G_{\tilde{Y}} \left(G_{\tilde{Y}}^{-1} \left(F_{\tilde{X}}(x) \right) \right)}{\partial s_k}, \quad (2.40)$$

kde

$$G'_{\tilde{Y}} = \frac{\partial G_{\tilde{Y}} \left(G_{\tilde{Y}}^{-1} \left(F_{\tilde{X}}(x) \right) \right)}{\partial y}.$$

K tomu je potřeba vyjádřit derivace $\frac{\partial F_{\tilde{X}}}{\partial r_j}$, $\frac{\partial G_{\tilde{Y}}}{\partial s_k}$ a $\frac{\partial G_{\tilde{Y}}}{\partial y}$. Začneme s $\frac{\partial F_{\tilde{X}}}{\partial r_j}$. Víme, že distribuční funkce $F_{\tilde{X}}$ má tvar

$$F_{\tilde{X}}(x) = \sum_{j=1}^J r_j \Phi \left(\frac{x - a_X x_j - (1 - a_X) \mu_X}{a_X h_X} \right).$$

Kromě parametrů r_j , kterými zde vážíme distribuční funkce Φ , ale na r_j závisí také μ_X i a_X (skrže σ_X^2 , $a_X = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + h_X^2}}$), kde

$$\mu_X = \sum_{j=1}^J x_j r_j \quad \text{a} \quad \sigma_X^2 = \sum_{j=1}^J r_j (x_j - \mu_X)^2.$$

V případě, kdy bychom vyhlazovací parametry h_X a h_Y volili například minimalizací penalizačního kritéria (2.17), potom by i tyto parametry byly na pravděpodobnostech \mathbf{r} a \mathbf{s} závislé. Pro odvození standardních chyb vyvažování zde ale budeme předpokládat, že h_X a h_Y jsou pevně dané konstanty.

Odvození parciální derivace $\frac{\partial F_{\tilde{X}}}{\partial r_j}$ je relativně zdlouhavé, uvedeme zde proto pouze tvrzení s výslednými tvary.

Tvrzení 7. *Nechť $F_{\tilde{X}}$ je distribuční funkce definována jako v (2.12). Potom*

$$\frac{\partial F_{\tilde{X}}}{\partial r_j}(x) = \Phi\left(\frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}\right) - M_{jX}(x) \frac{\partial F_{\tilde{X}}(x)}{\partial x}, \quad \text{kde}$$

$$\frac{\partial F_{\tilde{X}}(x)}{\partial x} = \sum_{j=1}^J r_j \phi\left(\frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}\right) \frac{1}{a_X h_X},$$

$$M_{jX}(x) = \frac{1}{2}(x - \mu_X)(1 - a_X^2) \left(\frac{x_j - \mu_X}{\sigma_X}\right)^2 + (1 - a_X)x_j.$$

Důkaz. Odvození je uvedeno v Appendixu Holland a kol. (1989), str. 34. □

Derivace $\frac{\partial G_{\tilde{Y}}}{\partial s_k}$ a $\frac{\partial G_{\tilde{Y}}}{\partial y}$ dostaneme zcela analogicky, tj.

$$\frac{\partial G_{\tilde{Y}}(y)}{\partial y} = \sum_{k=1}^K s_k \phi\left(\frac{y - a_Y y_k - (1 - a_Y)\mu_Y}{a_Y h_Y}\right) \frac{1}{a_Y h_Y} \quad \text{a}$$

$$\frac{\partial G_{\tilde{Y}}(y)}{\partial s_k} = \Phi\left(\frac{y - a_Y y_k - (1 - a_Y)\mu_Y}{a_Y h_Y}\right) - M_{kY}(y) \frac{\partial G_{\tilde{Y}}(y)}{\partial y}, \quad \text{kde}$$

$$M_{kY}(y) = \frac{1}{2}(y - \mu_Y)(1 - a_Y^2) \left(\frac{y_k - \mu_Y}{\sigma_Y}\right)^2 + (1 - a_Y)y_k,$$

kde parametry a_Y , μ_Y a σ_Y byly definovány v sekci 2.1.3.

Teď už máme vyjádřené všechny prvky vektoru \mathbb{J}_{φ_K} a známe varianční matici vektoru $(\hat{\mathbf{r}}, \hat{\mathbf{s}})^\top$ a můžeme aplikovat delta větu. Z ní dostaneme, že

$$\sqrt{N}(\hat{\varphi}_K - \varphi_K) \xrightarrow[N, M \rightarrow \infty]{d} \mathcal{N}\left(0, \mathbb{J}_{\varphi_K} \begin{pmatrix} \mathbf{C}_r \mathbf{C}_r^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_s \mathbf{C}_s^\top \end{pmatrix} \mathbb{J}_{\varphi_K}^\top\right), \quad (2.41)$$

kde matice \mathbf{C}_r a \mathbf{C}_s dostaneme z Věty 6 a vektor \mathbb{J}_{φ_K} je vyjádřen v (2.38), (2.39), (2.40) a v Tvzení 7.

Odtud už přímo dostáváme rozptyl odhadu vyvažovací funkce ve tvaru

$$\text{var}(\widehat{\varphi}_K(x)) = \frac{1}{N} \mathbb{J}_{\varphi_K} \begin{pmatrix} \mathbf{C}_r \mathbf{C}_r^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_s \mathbf{C}_s^\top \end{pmatrix} \mathbb{J}_{\varphi_K}^\top$$

a standardní chyby vyvažování jako

$$SE_{E_Y}(x) = \sqrt{\text{var}(\widehat{\varphi}_K(x))} = \sqrt{\frac{1}{N} \mathbb{J}_{\varphi_K} \begin{pmatrix} \mathbf{C}_r \mathbf{C}_r^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_s \mathbf{C}_s^\top \end{pmatrix} \mathbb{J}_{\varphi_K}^\top}.$$

Při výpočtu SE_{E_Y} nahradíme ve výrazu skutečné pravděpodobnosti \mathbf{r} , \mathbf{s} jejich odhady $\widehat{\mathbf{r}}$, $\widehat{\mathbf{s}}$ a parametr z odhadneme jako $\frac{M}{N}$.

2.2 Design neekvivalentních skupin

Doposud jsme pracovali s designem ekvivalentních skupin, tedy s případem, kdy všichni studenti (ať už vyplňovali test X nebo test Y) pocházeli ze stejné homogenní populace T . Jednalo se přímo o cílovou populaci, která nás zajímala. Všechny odhady a závěry tak byly rovnou platné vůči této populaci. V odvětví vzdělávání a testování je však zcela běžné, že skupiny studentů, kterým jsou testy zadávány, se mezi sebou liší. Příkladem u nás může být zadávání maturitních testů, které typicky probíhá každý rok ve dvou termínech, na jaře a na podzim. Zatímco na jarní termín se většinou hlásí studenti, kteří neměli problém se splněním podmínek pro připuštění k státní maturitní zkoušce, v podzimním termínu to jsou často studenti, kteří kvůli nedostatečnému prospěchu nebyli na jarní termín připuštěni, případně zkoušku opakují. Obecně horší výsledky maturitních testů v podzimním termínu tak nemusí být nutně způsobeny zadaným obtížnějším testem, ale právě převahou studentů s nižšími schopnostmi. Aby šly tyto dva efekty (obtížnost testu a abilita studentů) od sebe odlišit, zlatým standardem je použití tzv. kotvících položek testu. To je sada položek, která je obsažena ve více verzích testu, tyto položky tak zodpoví studenti z více zkušebních termínů. Na jejich základě je pak odhadnutá abilita studentů v příslušných skupinách a tento rozdíl tak nezasahuje do určování rozdílu v obtížnostech testů. Tento design nazýváme konkrétně design neekvivalentních skupin s kotvícími položkami (*angl. non-equivalent groups with anchor test*, NEAT design).

V některých situacích nelze kotvící položky použít, a to například kvůli zveřejňování všech položek testu, jako je to v případě maturitních testů v České republice. Jednou z možností, jak tuto situaci vyřešit, je nahradit skóre z kotvícího testu jinými dodatečnými informacemi o respondentech, na základě nichž bude odhadnuta jejich abilita. Musí se jednat o informace, které silně korelují s celkovým skóre testu a které jsou schopny plně vysvětlit rozdíl mezi skupinami respondentů. Typicky se volí například skóre z nějakého jiného podobného testu nebo známky z daného předmětu. Tento design potom nazýváme design neekvivalentních skupin s kovariátami (*angl. non-equivalent groups with covariates*, NEC design). Předpoklad, že tyto dodatečné informace umí vysvětlit všechny rozdíly mezi skupinami, je ale poměrně silný a obtížně dosažitelný, proto je vhodnější

volit testy s kotvícími položkami, pokud to je možné. V práci se případu, kdy kotvící test nemáme k dispozici, více věnovat nebudeme, teoretické podrobnosti lze nalézt v Wiberg a Bränberg (2015) a praktickou ukázkou na konkrétních datech například v Altintas a Wallin (2021).

Náhodné veličiny reprezentující celkové skóre ze dvou zadaných testů budeme nadále značit X a Y , náhodnou veličinu reprezentující celkové skóre z kotvících položek (na tuto sadu otázek se díváme jako na samostatný test) pak budeme značit A . Tato náhodná veličina může obecně nabývat hodnot a_l , $l = 1, \dots, L$, v našem případě $a_l = l - 1$ a maximální možné dosažené skóre z kotvícího testu je $L - 1$. Dále budeme uvažovat dvě populace studentů, populace P a Q . Obecně předpokládáme, že se tyto populace liší. Naše cílová populace, o kterou se zajímáme, bude směsí těchto dvou populací, a to

$$T = \omega P + (1 - \omega)Q, \quad (2.42)$$

kde $0 \leq \omega \leq 1$. Na T se tak můžeme dívat jako na nějakou větší populaci, která je složena z dvou vzájemně disjunktních skupin P a Q von Davier a kol. (2004, str. 35)

Skupina studentů, která vyplňovala test X , byla náhodně vybrána z populace P a skupina studentů, která vyplňovala test Y , byla zase náhodně vybraná z populace Q .

Označme pravděpodobnosti, které nám určují sdružené rozdělení (X, A) jako

$$p_{jl} = P(X = x_j, A = a_l | P). \quad (2.43)$$

Podmíněním vůči P značíme, že vzorek byl vybrán z populace P a pravděpodobnosti jsou tedy platné v této populaci. Označme

$$\mathbf{p} = (p_{11}, p_{21}, \dots, p_{J1}, p_{12}, \dots, p_{JL})^\top$$

vektor délky JL , který obsahuje jednotlivé pravděpodobnosti.

Podobně označme pravděpodobnosti sdruženého rozdělení (Y, A) jako

$$q_{kl} = P(Y = y_k, A = a_l | Q) \quad (2.44)$$

a

$$\mathbf{q} = (q_{11}, q_{21}, \dots, q_{K1}, q_{12}, \dots, q_{KL})^\top$$

vektor pravděpodobností délky KL .

K získání odhadu vyvažovací funkce v rámci tohoto designu existují dva možné přístupy, a to metoda řetězového vyvažování (*angl. chain equating*) a poststratifikační vyvažovací metoda (*angl. post-stratification equating*). Metody se liší v dodatečných předpokladech, které je potřeba u tohoto designu doplnit, proto se jejich výsledné vyvažovací funkce mohou lišit. Nejprve začneme s popisem poststratifikační metody.

2.2.1 Poststratifikační metoda

U této metody, podobně jako to bylo v případě EG designu, nejprve potřebujeme znát odhady pravděpodobnostních vektorů \mathbf{r} a \mathbf{s} , a to v naší cílové populaci T . Protože T je směs populací P a Q uvedená v (2.42), můžeme psát

$$r_j = P(X = x_j|T) = \omega r_{Pj} + (1 - \omega)r_{Qj}, \quad (2.45)$$

$$s_k = P(Y = y_k|T) = \omega s_{Pk} + (1 - \omega)s_{Qk}, \quad (2.46)$$

$$t_l = P(A = a_l|T) = \omega t_{Pl} + (1 - \omega)t_{Ql}, \quad (2.47)$$

kde

$$\begin{aligned} r_{Pj} &= P(X = x_j|P) = \sum_{l=1}^L p_{jl}, \\ s_{Qk} &= P(Y = y_k|Q) = \sum_{l=1}^L q_{kl}, \\ t_{Pl} &= P(A = a_l|P) = \sum_{j=1}^J p_{jl}, \\ t_{Ql} &= P(A = a_l|Q) = \sum_{k=1}^K q_{kl}, \\ r_{Qj} &= P(X = x_j|Q), \\ s_{Pk} &= P(Y = y_k|P). \end{aligned} \quad (2.48)$$

Pravděpodobnost r_{Pj} tedy značí pravděpodobnost, že respondent z populace P dosáhne celkového skóre x_j v testu X, pravděpodobnost r_{Qj} potom pravděpodobnost, že celkového skóre x_j v testu X dosáhne respondent z populace Q . Známe-li sdružené pravděpodobnosti celkového skóre z testu X a kotvícího testu p_{jl} z populace P (viz. rovnice (2.43)), pravděpodobnost r_{Pj} pak můžeme vyjádřit jejich součtem přes všechny možné hodnoty a_l , obdobně můžeme vyjádřit i r_{Qj} pomocí pravděpodobností q_{kl} (viz. rovnice (2.44)). Analogicky interpretujeme i zbylé pravděpodobnosti, kde pravděpodobnosti s_{Pk} , s_{Qk} se týkají celkového skóre v testu Y a pravděpodobnosti t_{Pl} , t_{Ql} celkového skóre z kotvícího testu A (v tomto případě sčítáme pravděpodobnosti p_{jl} a q_{kl} přes všechny možné hodnoty x_j , resp. y_k).

Protože test X nebyl zadán populaci Q a test Y nebyl zadán populaci P , pravděpodobnosti r_{Qj} a s_{Pk} nelze z dat odhadnout. Proto, abychom mohli vyjádřit pravděpodobnosti r_j a s_k , je potřeba přidat doplňující předpoklady. Předtím si ještě definujeme podmíněné pravděpodobnosti při dané hodnotě skóre z kotvícího testu, a to

$$\begin{aligned} r_P(x_j|a_l) &= P(X = x_j|A = a_l, P), & r_Q(x_j|a_l) &= P(X = x_j|A = a_l, Q), \\ s_Q(y_k|a_l) &= P(Y = y_k|A = a_l, Q), & s_P(y_k|a_l) &= P(Y = y_k|A = a_l, P). \end{aligned}$$

Předpoklad. (PSE1)

Předpokládáme, že

$$r_Q(x_j|a_l) = r_P(x_j|a_l),$$

tedy že pro všechny populace T (libovolnou volbu konstanty ω) je podmíněné rozdělení X při A invariantní vůči populaci.

Obdobně to předpokládáme pro podmíněné rozdělení Y .

Předpoklad. (PSE2)

Předpokládejme, že

$$s_P(y_k|a_l) = s_Q(y_k|a_l),$$

tedy že pro všechny populace T je podmíněné rozdělení Y při A invariantní vůči populaci.

Z předpokladu (PSE1) dostaneme vyjádření pro r_{Qj} ve tvaru

$$\begin{aligned} r_{Qj} &= \sum_{l=1}^L r_Q(x_j|a_l)t_{Ql} = \sum_{l=1}^L r_P(x_j|a_l)t_{Ql} = \sum_{l=1}^L P(X = x_j|A = a_l, P)t_{Ql} \\ &= \sum_{l=1}^L \frac{P(X = x_j, A = a_l|P)}{P(A = a_l|P)} t_{Ql} = \sum_{l=1}^L \frac{p_{jl}}{t_{Pl}} t_{Ql}. \end{aligned} \quad (2.49)$$

Vyjádření r_{Qj} ve tvaru (2.49) pak můžeme dosadit do (2.45), odkud dostaneme

$$\begin{aligned} r_j &= \omega r_{Pj} + (1 - \omega)r_{Qj} = \omega r_{Pj} + (1 - \omega) \sum_{l=1}^L \frac{p_{jl}}{t_{Pl}} t_{Ql} \\ &= \omega \sum_{l=1}^L p_{jl} + (1 - \omega) \sum_{l=1}^L \frac{p_{jl}}{t_{Pl}} t_{Ql} \\ &= \sum_{l=1}^L \left[\omega + \frac{(1 - \omega)t_{Ql}}{t_{Pl}} \right] p_{jl}. \end{aligned}$$

Vyjádřením pro všechny r_j současně, $j = 1, \dots, J$, dostaneme vektorový zápis ve tvaru

$$\mathbf{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_J \end{pmatrix} = \sum_{l=1}^L \left[\omega + \frac{(1 - \omega)t_{Ql}}{t_{Pl}} \right] \mathbf{p}_l, \quad \text{kde } \mathbf{p}_l = \begin{pmatrix} p_{1l} \\ \vdots \\ p_{Jl} \end{pmatrix}. \quad (2.50)$$

Analogicky vektor \mathbf{s} vyjádříme ve tvaru

$$\mathbf{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_K \end{pmatrix} = \sum_{l=1}^L \left[(1 - \omega) + \frac{\omega t_{Pl}}{t_{Ql}} \right] \mathbf{q}_l, \quad \text{kde } \mathbf{q}_l = \begin{pmatrix} q_{1l} \\ \vdots \\ q_{Kl} \end{pmatrix}. \quad (2.51)$$

Předpokládáme, že konstantu ω známe, a navíc platí, že $t_{Pl} = \sum_{j=1}^J p_{jl}$ a $t_{Ql} = \sum_{k=1}^K q_{kl}$. Jediné, co potřebujeme odhadnout, jsou tak pravděpodobnosti p_{jl} a q_{kl} . K tomu stejně jako v případě designu ekvivalentních skupin použijeme vyhlazení pomocí log-lineárního modelu, v tomto případě však jeho dvourozměrnou variantu. Nejprve však zavedeme používané značení.

Pozorovaná data, která máme v designu neekvivalentních skupin s kotvícími položkami k dispozici, jsou dvojice (X_i, A_i) , $i = 1, \dots, N$ pro respondenty z populace P a dvojice (Y_i, A_i) , $i = 1, \dots, M$ pro respondenty z populace Q. Z nich pak opět vytvoříme četnosti

$$\begin{aligned} n_{jl} &= \text{počet studentů s } X = x_j \text{ a } A = a_l, \\ m_{kl} &= \text{počet studentů s } Y = y_k \text{ a } A = a_l, \end{aligned}$$

$j = 1, \dots, J$, $k = 1, \dots, K$ a $l = 1, \dots, L$. Platí, že

$$\sum_{j=1}^J \sum_{l=1}^L n_{jl} = N \quad \text{a} \quad \sum_{k=1}^K \sum_{l=1}^L m_{kl} = M.$$

Vektor $\mathbf{n} = (n_{11}, \dots, n_{JL})^\top$ je tedy vektor o délce JL obsahující četnosti dvousložkového vektoru (X, A) . Podobně vektor $\mathbf{m} = (m_{11}, \dots, m_{KL})^\top$ je vektor o délce KL s četnostmi vektoru (Y, A) . Opět budeme předpokládat, že vektory pochází z multinomického rozdělení.

Předpoklad. Vektory $\mathbf{n} = (n_{11}, \dots, n_{JL})^\top$ a $\mathbf{m} = (m_{11}, \dots, m_{KL})^\top$ jsou mezi sebou nezávislé a mají multinomické rozdělení, tj.

$$\mathbf{n} \sim \text{Mult}_{JL}(N, \mathbf{p}), \quad \mathbf{m} \sim \text{Mult}_{KL}(M, \mathbf{q}),$$

kde $\mathbf{p} = (p_{11}, \dots, p_{J1}, p_{12}, \dots, p_{JL})^\top$ a $\mathbf{q} = (q_{11}, \dots, q_{K1}, q_{12}, \dots, q_{KL})^\top$.

Relativní četnosti mají tvar

$$\hat{p}_{jl} = \frac{n_{jl}}{N} \quad \text{a} \quad \hat{q}_{kl} = \frac{m_{kl}}{M}$$

a v případě velkých N a M by šly použít jako odhady pravděpodobností p_{jl} a q_{kl} . Ze stejných důvodů jako v designu ekvivalentních skupin však pravděpodobnosti většinou nejprve vyhladíme pomocí log-lineárního modelu.

2.2.1.1 Vyhlazení pravděpodobností skóre

Log-lineární model pro pravděpodobnosti \mathbf{p} můžeme zapsat analogicky jako v (2.5), tedy

$$\log(\mathbf{p}) = \alpha + \mathbf{u} + \mathbf{B}^\top \boldsymbol{\beta}, \quad (2.52)$$

kde α je normalizační konstanta, \mathbf{u} je vektor o délce JL , \mathbf{B} je matice známých konstant typu $p_p \times JL$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_p})^\top$ a p_p je počet parametrů potřebných k odhadnutí \mathbf{p} .

Nejčastěji uvažovaným modelem je momentový (polynomiální) model, který je opět obdobou modelu (2.6) popsáném v sekci 2.1.1. V tomto případě má obecný tvar

$$\log(p_{jl}) = \alpha + u_{jl} + \sum_{d=1}^D \beta_{x,d} x_j^d + \sum_{e=1}^E \beta_{a,e} a_l^e + \sum_{f=1}^F \sum_{g=1}^G \beta_{x_a,fg} x_j^f a_l^g. \quad (2.53)$$

Platí, že $D + E + F + G = p_p$, kde p_p je počet prvků vektoru β . Například v případě $D = E = 2$ a $F = G = 1$ dostáváme

$$\log(p_{jl}) = \alpha + u_{jl} + \beta_{x,1}x_j + \beta_{x,2}x_j^2 + \beta_{a,1}a_l + \beta_{a,2}a_l^2 + \beta_{xa,11}x_ja_l.$$

Příslušný sloupec matice \mathbf{B} , který odpovídá indexu jl , má tvar $(x_j, x_j^2, a_l, a_l^2, x_ja_l)$ a $\beta = (\beta_{x,1}, \beta_{x,2}, \beta_{a,1}, \beta_{a,2}, \beta_{xa,11})^\top$.

Odhad vektoru β dostaneme pomocí metody maximální věrohodnosti, logaritmická věrohodnost zde má tvar

$$l_N(\beta) = \sum_{j=1}^J \sum_{l=1}^L n_{jl} \log(p_{jl}) + C,$$

kde C je konstanta.

Odhady pravděpodobností p_{jl} dostaneme dosazením vektoru $\hat{\beta}$ do rovnice (2.53) a vyjádřením p_{jl} , odhad vektoru \mathbf{p} je ve tvaru $\hat{\mathbf{p}} = (\hat{p}_{11}, \dots, \hat{p}_{J1}, \hat{p}_{12}, \dots, \hat{p}_{JL})^\top$.

Protože log-lineární model (2.52) je ve stejném tvaru jako v sekci 2.1.1 při designu ekvivalentních skupin, výpočet varianční matice vektoru $\hat{\mathbf{p}}$ proběhne úplně stejně (sekce 2.1.5). Výsledná varianční matice $\Sigma_{\hat{\mathbf{p}}}$ má tvar uvedený v (2.32), jen místo vektoru pravděpodobností \mathbf{r} máme vektor sdružených pravděpodobností \mathbf{p} , tedy

$$\Sigma_{\hat{\mathbf{p}}} = \mathbf{P}\mathbf{B}^\top(\mathbf{B}\mathbf{P}\mathbf{B}^\top)^{-1}\mathbf{B}\mathbf{P}^\top,$$

kde

$$\mathbf{P} = \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top.$$

Rozměry matice $\Sigma_{\hat{\mathbf{p}}}$ jsou $JL \times JL$. Varianční matici $\Sigma_{\hat{\mathbf{p}}}$ opět můžeme faktorizovat (viz. Věta 6) a platí

$$\Sigma_{\hat{\mathbf{p}}} = \mathbf{C}_{\mathbf{p}}\mathbf{C}_{\mathbf{p}}^\top, \quad (2.54)$$

kde matice $\mathbf{C}_{\mathbf{p}}$ je typu $JL \times p_p$.

Analogicky získáme odhad vektoru \mathbf{q} jako $\hat{\mathbf{q}} = (\hat{q}_{11}, \dots, \hat{q}_{K1}, \hat{q}_{12}, \dots, \hat{q}_{KL})^\top$, a jeho varianční matici $\Sigma_{\hat{\mathbf{q}}}$ ve tvaru

$$\Sigma_{\hat{\mathbf{q}}} = \mathbf{C}_{\mathbf{q}}\mathbf{C}_{\mathbf{q}}^\top, \quad (2.55)$$

kde matice $\mathbf{C}_{\mathbf{q}}$ má rozměry $KL \times p_q$ a p_q je počet parametrů potřebných k odhadnutí pravděpodobností \mathbf{q} .

2.2.1.2 Odhad pravděpodobností skóre

V předchozí sekci jsme ukázali, jak pomocí log-lineárních modelů získat odhady sdružených pravděpodobností celkových skóre a skóre z kotvícího testu, p_{jl} a q_{kl} , definovaných jako

$$\begin{aligned} p_{jl} &= P(X = x_j, A = a_l | P), \\ q_{kl} &= P(Y = y_k, A = a_l | Q). \end{aligned}$$

My ovšem potřebujeme odhadnout marginální pravděpodobnosti celkových skóre v cílové populaci T , $r_j = P(X = x_j|T)$ a $s_k = P(Y = y_k|T)$. Jejich vyjádření právě pomocí sdružených pravděpodobností p_{jl} a q_{kl} jsme odvodili v rovnicích (2.50) a (2.51). Vektory odhadů marginálních pravděpodobností $\hat{\mathbf{r}} = (\hat{r}_1, \dots, \hat{r}_J)^\top$ a $\hat{\mathbf{s}} = (\hat{s}_1, \dots, \hat{s}_K)^\top$ tak dostaneme dosazením odhadů \hat{p}_{jl} a \hat{q}_{kl} do rovnic (2.50) a (2.51), tj.

$$\hat{\mathbf{r}} = \begin{pmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_J \end{pmatrix} = \sum_{l=1}^L \left[\omega + \frac{(1-\omega) \sum_{k=1}^K \hat{q}_{kl}}{\sum_{j=1}^J \hat{p}_{jl}} \right] \begin{pmatrix} \hat{p}_{1l} \\ \vdots \\ \hat{p}_{Jl} \end{pmatrix} \quad (2.56)$$

a

$$\hat{\mathbf{s}} = \begin{pmatrix} \hat{s}_1 \\ \vdots \\ \hat{s}_K \end{pmatrix} = \sum_{l=1}^L \left[(1-\omega) + \frac{\omega \sum_{j=1}^J \hat{p}_{jl}}{\sum_{k=1}^K \hat{q}_{kl}} \right] \begin{pmatrix} \hat{q}_{1l} \\ \vdots \\ \hat{q}_{Kl} \end{pmatrix}. \quad (2.57)$$

2.2.1.3 Spojitá aproximace skóre

Tento ani následující krok se nijak neliší v závislosti na zvoleném designu sběru dat, je tedy zcela totožný jako v sekci 2.1.3. Zopakujeme tedy jen důležité výsledky. Máme-li vektory odhadů $\hat{\mathbf{r}} = (\hat{r}_1, \dots, \hat{r}_J)^\top$ a $\hat{\mathbf{s}} = (\hat{s}_1, \dots, \hat{s}_K)^\top$, pak odhady spojitých distribučních funkcí, které nám aproximují původní diskrétní distribuční funkce F a G , mají tvar

$$\hat{F}_{\tilde{X}}(x) = \sum_{j=1}^J \hat{r}_j \Phi \left(\frac{x - \hat{a}_X x_j - (1 - \hat{a}_X) \hat{\mu}_X}{\hat{a}_X h_X} \right)$$

a

$$\hat{G}_{\tilde{Y}}(x) = \sum_{k=1}^K \hat{s}_k \Phi \left(\frac{x - \hat{a}_Y y_k - (1 - \hat{a}_Y) \hat{\mu}_Y}{\hat{a}_Y h_Y} \right),$$

kde podrobnosti jsou uvedeny v sekci 2.1.3 u designu dvou ekvivalentních skupin.

2.2.1.4 Vyvažování

Stejně tak i tento krok je totožný pro všechny uvažované designy sběru dat. Máme-li $\hat{F}_{\tilde{X}}$ a $\hat{G}_{\tilde{Y}}$ odhady distribučních funkcí, odhad vyvažovací funkce metodou vyvažování pomocí jádrových odhadů má tvar

$$\hat{\varphi}_K(x) = \hat{G}_{\tilde{Y}}^{-1} \left(\hat{F}_{\tilde{X}}(x) \right).$$

2.2.1.5 Výpočet standardních chyb vyvažování

Výpočet standardních chyb procesu vyvažování je už v tomto designu mírně komplikovanější. Zatímco v případě dvou ekvivalentních skupin jsme pomocí log-lineárního modelu získali přímo vektory odhadů marginálních pravděpodobností $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$, v případě designu neekvivalentních skupin s kotvícími položkami jsme odhadli jen sdružené pravděpodobnosti $\hat{\mathbf{p}}$ a $\hat{\mathbf{q}}$. Pravděpodobnosti \mathbf{r} a \mathbf{s} jsou pak funkcemi těchto parametrů \mathbf{p} , \mathbf{q} a spočteme je jako jejich nelineární transformace, vzorce jsou uvedeny v (2.50) a (2.51). K získání varianční matice vektoru $(\hat{\mathbf{r}}, \hat{\mathbf{s}})^\top$ tedy znovu aplikujeme delta metodu, viz Věta 5.

Asymptotické rozdělení $\hat{\mathbf{p}}$ a $\hat{\mathbf{q}}$

Ze sekcí 2.2.1.1 a 2.1.5 víme, že

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{JL}(\mathbb{0}_{JL}, \mathbf{C}_{\mathbf{p}}\mathbf{C}_{\mathbf{p}}^{\top}) \quad (2.58)$$

a

$$\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{KL}\left(\mathbb{0}_{KL}, \frac{1}{z}\mathbf{C}_{\mathbf{q}}\mathbf{C}_{\mathbf{q}}^{\top}\right). \quad (2.59)$$

Díky nezávislosti $\hat{\mathbf{p}}$ a $\hat{\mathbf{q}}$ dostaneme

$$\sqrt{N}\left(\begin{pmatrix} \hat{\mathbf{p}} \\ \hat{\mathbf{q}} \end{pmatrix} - \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}\right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{JL+KL}\left(\begin{pmatrix} \mathbb{0}_{JL} \\ \mathbb{0}_{KL} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_{\mathbf{p}}\mathbf{C}_{\mathbf{p}}^{\top} & \mathbf{0} \\ \mathbf{0} & \frac{1}{z}\mathbf{C}_{\mathbf{q}}\mathbf{C}_{\mathbf{q}}^{\top} \end{pmatrix}\right).$$

Asymptotické rozdělení $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$

Protože pravděpodobnosti \mathbf{r} a \mathbf{s} jsou funkcemi parametrů \mathbf{p} a \mathbf{q} , z delta věty opět dostaneme, že

$$\sqrt{N}\left(\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{s}} \end{pmatrix} - \begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}\right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{J+K}\left(\mathbb{0}_{J+K}, \mathbb{J}_{DF}\begin{pmatrix} \mathbf{C}_{\mathbf{p}}\mathbf{C}_{\mathbf{p}}^{\top} & \mathbf{0} \\ \mathbf{0} & \frac{1}{z}\mathbf{C}_{\mathbf{q}}\mathbf{C}_{\mathbf{q}}^{\top} \end{pmatrix}\mathbb{J}_{DF}^{\top}\right), \quad (2.60)$$

kde \mathbb{J}_{DF} je Jacobiho matice funkcí \mathbf{r} a \mathbf{s} o rozměrech $(J+K) \times (JL+KL)$, která má tvar

$$\mathbb{J}_{DF} = \begin{pmatrix} \frac{\partial \mathbf{r}}{\partial \mathbf{p}} & \frac{\partial \mathbf{r}}{\partial \mathbf{q}} \\ \frac{\partial \mathbf{s}}{\partial \mathbf{p}} & \frac{\partial \mathbf{s}}{\partial \mathbf{q}} \end{pmatrix} = \begin{pmatrix} \frac{\partial r_1}{\partial p_{11}} & \cdots & \frac{\partial r_1}{\partial p_{JL}} & \frac{\partial r_1}{\partial q_{11}} & \cdots & \frac{\partial r_1}{\partial q_{KL}} \\ \vdots & & & & & \\ \frac{\partial r_J}{\partial p_{11}} & & \cdots & & & \vdots \\ \frac{\partial s_1}{\partial p_{11}} & & & \cdots & & \vdots \\ \vdots & & & & & \\ \frac{\partial s_K}{\partial p_{11}} & \cdots & \cdots & & \frac{\partial s_K}{\partial q_{KL}} \end{pmatrix}.$$

Derivováním prvků \mathbf{r} a \mathbf{s} podle prvků \mathbf{p} a \mathbf{q} tak dostaneme jednotlivé elementy matice \mathbb{J}_{DF} . Parciální derivace mají tvar

$$\frac{\partial r_i}{\partial p_{hl}} = -p_{il} \frac{(1-\omega) \sum_{k=1}^K q_{kl}}{\left(\sum_{j=1}^J p_{jl}\right)^2} + \mathbb{1}_{[i=h]} \left[\omega + \frac{(1-\omega) \sum_{k=1}^K q_{kl}}{\sum_{j=1}^J p_{jl}} \right], \quad (2.61)$$

$$\frac{\partial r_i}{\partial q_{hl}} = \frac{p_{il}(1-\omega)}{\sum_{j=1}^J p_{jl}}, \quad (2.62)$$

kde $i = 1, \dots, J$, $h = 1, \dots, J$, $l = 1, \dots, L$ a

$$\frac{\partial s_i}{\partial p_{hl}} = \frac{q_{il}\omega}{\sum_{k=1}^K q_{kl}}, \quad (2.63)$$

$$\frac{\partial s_i}{\partial q_{hl}} = -q_{il} \frac{\omega \sum_{j=1}^J p_{jl}}{\left(\sum_{k=1}^K q_{kl}\right)^2} + \mathbb{1}_{[i=h]} \left[(1-\omega) + \frac{\omega \sum_{j=1}^J p_{jl}}{\sum_{k=1}^K q_{kl}} \right], \quad (2.64)$$

kde $i = 1, \dots, K$, $h = 1, \dots, K$, $l = 1, \dots, L$.

Asymptotické rozdělení $\hat{\varphi}_K$

Z asymptotického rozdělení odhadů $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$ (rovnice 2.60) následně získáme asymptotické rozdělení vyvažovací funkce $\hat{\varphi}_K$, a to zcela obdobně jako v případě designu dvou ekvivalentních skupin, jak bylo uvedeno v sekci 2.1.5. Platí, že

$$\sqrt{N}(\hat{\varphi}_K - \varphi_K) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}\left(0, \mathbb{J}_{\varphi_K} \mathbb{J}_{DF} \begin{pmatrix} \mathbf{C}_p \mathbf{C}_p^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_q \mathbf{C}_q^\top \end{pmatrix} \mathbb{J}_{DF}^\top \mathbb{J}_{\varphi_K}^\top\right), \quad (2.65)$$

kde tvar \mathbb{J}_{φ_K} je uveden v (2.38), jednotlivé prvky pak v rovnicích (2.39), (2.40), a v Tvzení 7.

Odsud dostáváme, že

$$\text{var}(\hat{\varphi}_K(x)) = \frac{1}{N} \mathbb{J}_{\varphi_K} \mathbb{J}_{DF} \begin{pmatrix} \mathbf{C}_p \mathbf{C}_p^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_q \mathbf{C}_q^\top \end{pmatrix} \mathbb{J}_{DF}^\top \mathbb{J}_{\varphi_K}^\top$$

a standardní chyby vyvažování tak můžeme vyjádřit jako

$$SEE_Y(x) = \sqrt{\text{var}(\hat{\varphi}_K(x))} = \sqrt{\frac{1}{N} \mathbb{J}_{\varphi_K} \mathbb{J}_{DF} \begin{pmatrix} \mathbf{C}_p \mathbf{C}_p^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbf{C}_q \mathbf{C}_q^\top \end{pmatrix} \mathbb{J}_{DF}^\top \mathbb{J}_{\varphi_K}^\top}. \quad (2.66)$$

Při výpočtu standardních chyb opět ve výrazu nahradíme všechny skutečné parametry jejich odhady.

2.2.2 Metoda řetězového vyvažování

Druhou metodou, kterou lze použít v designu neekvivalentních skupin s kotvícími položkami, je metoda řetězového vyvažování. Ta se od předchozí poststratifikační metody liší nejenom v předpokladech, ale také v tom, že v ní neprobíhá všech 5 výše zmíněných kroků jako v předešlých případech. Konkrétně se zde neodhadují pravděpodobnosti \mathbf{r} a \mathbf{s} v cílové populaci T , ale jen v subpopulacích P a Q . Metoda pak probíhá ve dvou fázích, kdy v první fázi jsou skóre z testu X transformovány na skóre z testu A (test složený z kotvících položek) v populaci P a ve druhé fázi jsou skóre z testu A transformovány na škálu skóre z testu Y v populaci Q .

Pracovat budeme s následujícími distribučními funkcemi,

$$\begin{aligned} F(x) &= P(X \leq x|T), & F_P(x) &= P(X \leq x|P), \\ G(y) &= P(Y \leq y|T), & G_Q(y) &= P(Y \leq y|Q), \\ H(a) &= P(A \leq a|T), & H_P(a) &= P(A \leq a|P), \\ & & H_Q(a) &= P(A \leq a|Q). \end{aligned}$$

Uvažovat budeme opět dva dodatečné předpoklady.

Předpoklad. (CE1)

Předpokládejme, že vyvažovací funkce z X do A je invariantní vůči populaci, tj.

$$H_P^{-1}(F_P(x)) = H^{-1}(F(x)), \quad (2.67)$$

pro jakoukoliv cílovou populaci T ve tvaru $T = \omega P + (1 - \omega)Q$.

Za předpokladu (CE1) tak dostaneme explicitní vyjádření pro distribuční funkci náhodné veličiny X v populaci T ,

$$F(x) = H \left(H_P^{-1} (F_P(x)) \right). \quad (2.68)$$

Obdobný předpoklad zavedeme i pro vyvažovací funkci z A do Y .

Předpoklad. (CE2)

Předpokládejme, že vyvažovací funkce z A do Y je invariantní vůči populaci, tj.

$$G_Q^{-1} (H_Q(a)) = G^{-1} (H(a)), \quad (2.69)$$

pro jakoukoliv cílovou populaci T ve tvaru $T = \omega P + (1 - \omega)Q$.

Z druhého předpokladu pak dostaneme vyjádření pro distribuční funkci náhodné veličiny Y v populaci T , kterou můžeme vyjádřit ve tvaru

$$G(y) = H \left(H_Q^{-1} (G_Q(y)) \right) \quad (2.70)$$

a její inverzní funkci jako

$$G^{-1}(u) = G_Q^{-1} \left(H_Q \left(H^{-1}(u) \right) \right). \quad (2.71)$$

Nyní už můžeme definovat vyvažovací funkci metody řetězového vyvažování, a to ve tvaru

$$\varphi_{CE}(x) = G^{-1} (F(x)). \quad (2.72)$$

Dosadíme-li do rovnice vyjádření z (2.68) a (2.71), dostaneme výsledný tvar

$$\begin{aligned} \varphi_{CE}(x) &= G_Q^{-1} \left(H_Q \left(H^{-1} \left(H \left(H_P^{-1} (F_P(x)) \right) \right) \right) \right) \\ &= G_Q^{-1} \left(H_Q \left(H_P^{-1} (F_P(x)) \right) \right). \end{aligned} \quad (2.73)$$

Všechny distribuční funkce z tohoto vyjádření pak už umíme z dostupných dat odhadnout. Potřebujeme k tomu znát jen odhady sdružených pravděpodobností $p_{jl} = P(X = x_j, A = a_l|P)$, $j = 1, \dots, J$, $l = 1, \dots, L$ a $q_{kl} = P(Y = y_k, A = a_l|Q)$, $k = 1, \dots, K$, $l = 1, \dots, L$. Ty můžeme získat opět pomocí log-lineárního modelu, totožně jako je uvedeno v sekci 2.2.1.1. Máme tedy odhady dvou vektorů pravděpodobností $\hat{\mathbf{p}} = (\hat{p}_{11}, \dots, \hat{p}_{J1}, \hat{p}_{12}, \dots, \hat{p}_{JL})^\top$ a $\hat{\mathbf{q}} = (\hat{q}_{11}, \dots, \hat{q}_{K1}, \hat{q}_{12}, \dots, \hat{q}_{KL})^\top$. Tyto odhady pak využijeme k odhadům marginálních pravděpodobností X a A v populaci P a marginálních pravděpodobností Y a A v populaci Q , definovaných jako v (2.48),

$$\begin{aligned} r_{Pj} &= P(X = x_j|P) = \sum_{l=1}^L p_{jl}, \\ s_{Qk} &= P(Y = y_k|Q) = \sum_{l=1}^L q_{kl}, \\ t_{Pl} &= P(A = a_l|P) = \sum_{j=1}^J p_{jl}, \\ t_{Ql} &= P(A = a_l|Q) = \sum_{k=1}^K q_{kl}. \end{aligned}$$

Máme tak k dispozici odhady čtyř vektorů,

$$\hat{\mathbf{r}}_P = (\hat{r}_{P1}, \dots, \hat{r}_{PJ})^\top = \left(\sum_{l=1}^L \hat{p}_{1l}, \dots, \sum_{l=1}^L \hat{p}_{Jl} \right)^\top, \quad (2.74)$$

$$\hat{\mathbf{s}}_Q = (\hat{s}_{Q1}, \dots, \hat{s}_{QK})^\top = \left(\sum_{l=1}^L \hat{q}_{1l}, \dots, \sum_{l=1}^L \hat{q}_{Kl} \right)^\top, \quad (2.75)$$

$$\hat{\mathbf{t}}_P = (\hat{t}_{P1}, \dots, \hat{t}_{PL})^\top = \left(\sum_{j=1}^J \hat{p}_{j1}, \dots, \sum_{j=1}^J \hat{p}_{jL} \right)^\top, \quad (2.76)$$

$$\hat{\mathbf{t}}_Q = (\hat{t}_{Q1}, \dots, \hat{t}_{QL})^\top = \left(\sum_{k=1}^K \hat{q}_{k1}, \dots, \sum_{k=1}^K \hat{q}_{kL} \right)^\top. \quad (2.77)$$

Protože stejně jako v předešlém případě potřebujeme, aby byla dobře definována inverzní funkce distribuční funkce G v rovnici (2.72) pro vyvažovací funkci (symetricky pro distribuční funkci F , kdybychom převáděli skóre z testu Y na škálu skóre z testu X), budeme opět uvažovat spojité aproximace všech distribučních funkcí uvedených v (2.73). Postup je totožný jako je uvedeno v sekci 2.1.3, jen místo pravděpodobností r_j budeme dosazovat pravděpodobnosti r_{Pj} , místo pravděpodobností s_k pravděpodobnosti s_{Qk} apod. Pro názornost uvedeme tvar jedné z distribučních funkcí, ostatní by vypadaly analogicky.

Spojité aproximace distribuční funkce F_P má obdobný tvar jako v (2.12),

$$F_{\tilde{X}_P}(x) = \sum_{j=1}^J r_{Pj} \Phi \left(\frac{x - a_{XP}x_j - (1 - a_{XP})\mu_{XP}}{a_{XP}h_{XP}} \right), \quad (2.78)$$

kde

$$\mu_{XP} = \sum_{j=1}^J x_j r_{Pj}, \quad \sigma_{XP}^2 = \sum_{j=1}^J r_{Pj} (x_j - \mu_{XP})^2, \quad a_{XP} = \sqrt{\frac{\sigma_{XP}^2}{\sigma_{XP}^2 + h_{XP}^2}}$$

a $h_{XP} > 0$ je vyhlazovací parametr.

Odhad této distribuční funkce dostaneme nahrazením r_{Pj} za odhady \hat{r}_{Pj} , tj.

$$\hat{F}_{\tilde{X}_P}(x) = F_{\tilde{X}_P}(x; \hat{\mathbf{r}}_P) = \sum_{j=1}^J \hat{r}_{Pj} \Phi \left(\frac{x - \hat{a}_{XP}x_j - (1 - \hat{a}_{XP})\hat{\mu}_{XP}}{\hat{a}_{XP}h_{XP}} \right), \quad (2.79)$$

kde

$$\hat{\mu}_{XP} = \sum_{j=1}^J x_j \hat{r}_{Pj}, \quad \hat{\sigma}_{XP}^2 = \sum_{j=1}^J \hat{r}_{Pj} (x_j - \hat{\mu}_{XP})^2, \quad \hat{a}_{XP} = \sqrt{\frac{\hat{\sigma}_{XP}^2}{\hat{\sigma}_{XP}^2 + h_{XP}^2}}.$$

Analogicky dostaneme spojité aproximace distribučních funkcí G_Q , H_P a H_Q , které značíme po řadě $G_{\tilde{Y}_Q}$, $H_{\tilde{A}_P}$ a $H_{\tilde{A}_Q}$.

Odhad vyvažovací funkce metody řetězového vyvažování pak získáme tak, že do rovnice (2.72) dosadíme tyto spojité aproximace a dostaneme

$$\widehat{\varphi}_{CE}(x) = G_{\widetilde{Y}_Q}^{-1} \left(H_{\widetilde{A}_Q} \left(H_{\widetilde{A}_P}^{-1} \left(F_{\widetilde{X}_P}(x) \right) \right) \right). \quad (2.80)$$

Poslední, co zbývá, je spočítat standardní chybu řetězového vyvažování. Vyvažovací funkci $\varphi_{CE}(x)$ můžeme zapsat ve tvaru

$$\varphi_{CE}(x) = \varphi_Y(\varphi_A(x)), \quad (2.81)$$

kde

$$\varphi_A(x) = H_{\widetilde{A}_P}^{-1} \left(F_{\widetilde{X}_P}(x) \right) \quad (2.82)$$

je vyvažovací funkce z X do A v populaci P a

$$\varphi_Y(a) = G_{\widetilde{Y}_Q}^{-1} \left(H_{\widetilde{A}_Q}(a) \right) \quad (2.83)$$

je vyvažovací funkce z A do Y v populaci Q .

Jacobiho matice $\varphi_A(x)$ a $\varphi_Y(a)$ tak můžeme spočítat stejně, jako je uvedeno v sekci 2.1.5, konkrétně v rovnicích (2.38), (2.39), (2.40) a v Tvrzení 7. Vždy jen místo původních parametrů $F_{\widetilde{X}}, G_{\widetilde{Y}}, \mathbf{r}$ a \mathbf{s} nahradíme příslušnými alternativami, tj. $F_{\widetilde{X}_P}, H_{\widetilde{A}_P}, \mathbf{r}_P, \mathbf{t}_P$ pro φ_A a $G_{\widetilde{Y}_Q}, H_{\widetilde{A}_Q}, \mathbf{s}_Q, \mathbf{t}_Q$ pro φ_Y . Označme tyto Jacobiho matice jako \mathbb{J}_{φ_A} a \mathbb{J}_{φ_Y} . Označíme-li dále $\varphi'_Y(a) = \frac{\partial \varphi_Y}{\partial a}(a)$, můžeme Jacobiho matici (v tomto případě vektor) vyvažovací funkce metody řetězového vyvažování vyjádřit ve tvaru

$$\mathbb{J}_{\varphi_{CE}}(x) = \left(\varphi'_Y(\varphi_A(x)) \mathbb{J}_{\varphi_A}(x), \mathbb{J}_{\varphi_Y}(\varphi_A(x)) \right). \quad (2.84)$$

Protože \mathbb{J}_{φ_A} je délky $(J + L)$ a \mathbb{J}_{φ_Y} délky $(K + L)$, vektor $\mathbb{J}_{\varphi_{CE}}$ má rozměry $1 \times (J + K + 2L)$.

K tomu, abychom spočítali standardní chybu vyvažování a tedy asymptotickou varianční matici odhadu $\widehat{\varphi}_{CE}$, potřebujeme ještě znát varianční matici vektoru $(\widehat{\mathbf{r}}_P, \widehat{\mathbf{t}}_P, \widehat{\mathbf{t}}_Q, \widehat{\mathbf{s}}_Q)^\top$.

Začneme tím, že platí

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{JL}(\mathbf{0}_{JL}, \mathbf{C}_P \mathbf{C}_P^\top),$$

jak je uvedeno v (2.58).

Dále víme, že \mathbf{r}_P a \mathbf{t}_P jsou funkcemi \mathbf{p} , tedy z delta věty (Věta 5) dostaneme, že

$$\sqrt{N} \left(\begin{pmatrix} \widehat{\mathbf{r}}_P \\ \widehat{\mathbf{t}}_P \end{pmatrix} - \begin{pmatrix} \mathbf{r}_P \\ \mathbf{t}_P \end{pmatrix} \right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{J+L} \left(\mathbf{0}_{J+L}, \mathbb{J}_{DF,P} \mathbf{C}_P \mathbf{C}_P^\top \mathbb{J}_{DF,P}^\top \right), \quad (2.85)$$

kde $\mathbb{J}_{DF,P}$ je typu $(J+L) \times JL$ a má tvar

$$\mathbb{J}_{DF,P} = \begin{pmatrix} \frac{\partial \mathbf{r}_P}{\partial \mathbf{p}} \\ \frac{\partial \mathbf{t}_P}{\partial \mathbf{p}} \end{pmatrix} = \begin{pmatrix} \frac{\partial r_{P1}}{\partial p_{11}} & \frac{\partial r_{P1}}{\partial p_{21}} & \cdots & \cdots & \frac{\partial r_{PJ}}{\partial p_{JL}} \\ \vdots & & & & \vdots \\ \frac{\partial r_{PJ}}{\partial p_{11}} & \cdots & \cdots & \cdots & \frac{\partial r_{PJ}}{\partial p_{JL}} \\ \frac{\partial t_{P1}}{\partial p_{11}} & \cdots & \cdots & \cdots & \frac{\partial t_{P1}}{\partial p_{JL}} \\ \vdots & & & & \vdots \\ \frac{\partial t_{PL}}{\partial p_{11}} & \cdots & \cdots & \cdots & \frac{\partial t_{PL}}{\partial p_{JL}} \end{pmatrix}. \quad (2.86)$$

Prvky matice $\mathbb{J}_{DF,P}$ jsou ve tvaru

$$\frac{\partial r_i}{\partial p_{hl}} = \mathbb{1}_{[i=h]}, \quad (2.87)$$

kde $i = 1, \dots, J$, $h = 1, \dots, J$, $l = 1, \dots, L$ a

$$\frac{\partial t_i}{\partial p_{hl}} = \mathbb{1}_{[i=l]}, \quad (2.88)$$

kde $i = 1, \dots, L$, $h = 1, \dots, J$, $l = 1, \dots, L$.

Obdobně z delta věty a platnosti rovnice (2.59), tedy

$$\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{KL} \left(\mathbb{0}_{KL}, \frac{1}{z} \mathbf{C}_q \mathbf{C}_q^\top \right),$$

dostaneme asymptotické rozdělení

$$\sqrt{N} \left(\begin{pmatrix} \hat{\mathbf{t}}_Q \\ \hat{\mathbf{s}}_Q \end{pmatrix} - \begin{pmatrix} \mathbf{t}_Q \\ \mathbf{s}_Q \end{pmatrix} \right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{K+L} \left(\mathbb{0}_{K+L}, \frac{1}{z} \mathbb{J}_{DF,Q} \mathbf{C}_q \mathbf{C}_q^\top \mathbb{J}_{DF,Q}^\top \right). \quad (2.89)$$

Matice $\mathbb{J}_{DF,Q}$ má analogický tvar jako matice $\mathbb{J}_{DF,P}$, a to

$$\mathbb{J}_{DF,Q} = \begin{pmatrix} \frac{\partial \mathbf{t}_Q}{\partial \mathbf{q}} \\ \frac{\partial \mathbf{s}_Q}{\partial \mathbf{q}} \end{pmatrix} \quad (2.90)$$

s prvky

$$\frac{\partial t_i}{\partial q_{hl}} = \mathbb{1}_{[i=l]}, \quad (2.91)$$

kde $i = 1, \dots, L$, $h = 1, \dots, K$, $l = 1, \dots, L$ a

$$\frac{\partial s_i}{\partial q_{hl}} = \mathbb{1}_{[i=h]}, \quad (2.92)$$

kde $i = 1, \dots, K$, $h = 1, \dots, K$, $l = 1, \dots, L$.

Protože vektory $\hat{\mathbf{p}}$ a $\hat{\mathbf{q}}$ jsou nezávislé, kombinací (2.85) a (2.89) dostaneme

$$\sqrt{N} \left(\begin{pmatrix} \hat{\mathbf{r}}_P \\ \hat{\mathbf{t}}_P \\ \hat{\mathbf{t}}_Q \\ \hat{\mathbf{s}}_Q \end{pmatrix} - \begin{pmatrix} \mathbf{r}_P \\ \mathbf{t}_P \\ \mathbf{t}_Q \\ \mathbf{s}_Q \end{pmatrix} \right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}_{J+K+2L} (\mathbb{0}_{J+K+2L}, D), \quad (2.93)$$

kde

$$D = \begin{pmatrix} \mathbb{J}_{DF,P} \mathbf{C}_p \mathbf{C}_p^\top \mathbb{J}_{DF,P}^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{z} \mathbb{J}_{DF,Q} \mathbf{C}_q \mathbf{C}_q^\top \mathbb{J}_{DF,Q}^\top \end{pmatrix}. \quad (2.94)$$

Teď už známe všechny varianční a Jacobiho matice, které potřebujeme, abychom mohli vyjádřit asymptotické rozdělení vyvažovací funkce metody řetězového vyvažování. Asymptotické rozdělení má tvar

$$\sqrt{N}(\hat{\varphi}_{CE} - \varphi_{CE}) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}\left(0, \mathbb{J}_{\varphi_{CE}} D \mathbb{J}_{\varphi_{CE}}^\top\right). \quad (2.95)$$

Platí, že

$$\text{var}(\hat{\varphi}_{CE}(x)) = \frac{1}{N} \mathbb{J}_{\varphi_{CE}} D \mathbb{J}_{\varphi_{CE}}^\top$$

a standardní chyby vyvažování mají tvar

$$SEE_Y(x) = \sqrt{\frac{1}{N} \mathbb{J}_{\varphi_{CE}} D \mathbb{J}_{\varphi_{CE}}^\top}. \quad (2.96)$$

3. Vyvažování pomocí modelů teorie odpovědi na položku

V poslední teoretické kapitole pro úplnost představíme i základní koncepty vyvažování pomocí modelů teorie odpovědi na položku, tzv. IRT modelů (*angl. item response theory, IRT models*). Protože v mnoha testovacích společnostech jsou právě IRT modely často využívány k analýze a vytváření znalostních testů, je přirozené je využít i k procesu vyvažování. Nejprve obecně představíme IRT modely a jejich předpoklady, poté jejich použití při vyvažování. Popis metod nalezneme především v Kolen a Brennan (2014) a González a Wiberg (2017).

3.1 IRT modely

Hlavní myšlenka IRT modelů spočívá v tom, že pravděpodobnost správné odpovědi na položku je funkcí parametrů této položky (např. obtížnosti) a parametrů respondenta (nějakých jeho latentních vlastností, např. inteligence). V této práci budeme předpokládat, že pravděpodobnost správné odpovědi je závislá pouze na jedné latentní vlastnosti, pracovat tedy budeme jen s jednorozměrným parametrem (tento parametr budeme nazývat *abilita* a budeme ho značit symbolem θ). V praxi se můžeme setkat i se situací, kdy je pravděpodobnost ovlivněna více latentními vlastnostmi a příslušné parametry jsou vícerozměrné. Tomuto případu se v práci věnovat nebudeme, ale v literatuře se dají najít zobecnění i pro tyto situace. Stejně tak se zaměříme pouze na modely s dichotomickými položkami, tedy s položkami, které mohou být hodnoceny jen dvěma hodnotami (většinou uvažujeme, že respondent zodpověděl otázku správně či nesprávně). Metody se však dají zobecnit na polytomní modely, kde položky mohou být hodnoceny i jiným počtem bodů.

Mějme dvě verze znalostního testu, verzi X a verzi Y. Předpokládejme, že test X se skládá z J položek a že jej vyplnilo N respondentů, test Y se skládá z K položek a byl vyplněn M respondenty. Počet kotvících položek, které jsou obsaženy v obou verzích testu, označme V . Platí, že $V < J$ a $V < K$. Dále označme X_{ij} , $i = 1, \dots, N$, $j = 1, \dots, J$ binární náhodnou veličinu, která značí odpověď i -tého respondenta na položku j v testu X. Platí, že $X_{ij} = 1$, když respondent zodpověděl položku správně, a $X_{ij} = 0$, když zodpověděl položku nesprávně. Obdobně označme Y_{ij} , $i = N + 1, \dots, N + M$, $j = J - V + 1, \dots, J + K - V$ náhodnou veličinu, která značí odpověď i -tého respondenta na položku j v testu Y. Položky indexujeme tak, že nejprve řadíme položky unikátní pro test X, následně kotvící položky, které jsou společné pro oba testy, a na závěr položky unikátní pro test Y. Kotvící položky tak mají v tomto případě indexy od $J - V + 1$ do J . Tuto indexaci využijeme především v praktické části práce, kdy práce s takto řazenými položkami bude přehlednější. Pozorovaná data, která máme k dispozici, jsou tak sady N vektorů $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,J})^\top$, $i = 1, \dots, N$ a M vektorů $\mathbf{Y}_i = (Y_{i,J-V+1}, \dots, Y_{i,J+K-V})^\top$, $i = N + 1, \dots, N + M$. Každý vektor tak nese informaci o správných a nesprávných odpovědích u konkrétního jednoho respon-

denta. Předpokládáme, že jednotlivé vektory \mathbf{X}_i a \mathbf{Y}_i jsou mezi sebou nezávislé. Dalším předpokladem, který zde zavádíme, je předpoklad lokální nezávislosti. Ten nám říká, že při pevně daném parametru θ (abilitě respondenta), jsou odpovědi respondenta na jednotlivé položky mezi sebou nezávislé. Pravděpodobnost, že respondent s danou hodnotou θ zodpoví správně položku 1 i 2, tak můžeme spočítat jako součin pravděpodobnosti, že odpoví správně na položku 1, a pravděpodobnosti, že zodpoví správně položku 2.

Podle toho, kolik parametrů položky uvažujeme, rozlišujeme různé IRT modely. Jedním z nejobecnějších modelů je čtyřparametrický logistický IRT model, který budeme značit 4PL IRT model. Za platnosti 4PL IRT modelu má pravděpodobnost, že i -tý respondent správně zodpoví j -tou položku, tvar

$$\begin{aligned} p_{ij}(\theta_i; a_j, b_j, c_j, d_j) &= P(X_{ij} = 1 | \theta_i; a_j, b_j, c_j, d_j) \\ &= c_j + (d_j - c_j) \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}, \end{aligned} \quad (3.1)$$

kde $\theta_i \in (-\infty, \infty)$ je abilita i -tého respondenta a a_j, b_j, c_j, d_j jsou parametry j -té položky. Parametr $a_j \in (-\infty, \infty)$ obvykle nazýváme parametrem diskriminace, $b_j \in (-\infty, \infty)$ parametrem obtížnosti, $c_j \in (0, 1)$ parametrem dolní asymptoty a $d_j \in (0, 1)$ parametrem horní asymptoty. Parametr c_j zde reprezentuje pravděpodobnost, že respondent správnou odpověď na položku pouze uhodne. Naopak parametr d_j značí pravděpodobnost, že respondent, který zná správnou odpověď, správně i odpoví. Doplněk $(1 - d_j)$ tedy reprezentuje pravděpodobnost, že respondent, který zná správnou odpověď, odpoví nesprávně. U znalostních testů to může být například vlivem nepozornosti.

Budeme-li předpokládat, že k vlivu nepozornosti nedochází, a že $d_j = 1 \forall j$, dostaneme pak rovnici pro tříparametrický logistický IRT model, který značíme 3PL IRT. Za platnosti 3PL IRT modelu je pravděpodobnost, že i -tý respondent správně zodpoví j -tou položku, ve tvaru

$$p_{ij}(\theta_i; a_j, b_j, c_j) = P(X_{ij} = 1 | \theta_i; a_j, b_j, c_j) = c_j + (1 - c_j) \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}. \quad (3.2)$$

Předpokládáme-li navíc, že i pravděpodobnosti uhodnutí položek jsou nulové, tedy $c_j = 0 \forall j$, dostaneme dvouparametrický 2PL IRT model a pravděpodobnost, že i -tý respondent zodpoví j -tou položku správně je

$$p_{ij}(\theta_i; a_j, b_j) = P(X_{ij} = 1 | \theta_i; a_j, b_j) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}. \quad (3.3)$$

Speciálním případem 2PL IRT modelu je i 1PL IRT model (jednoparametrický, někdy nazývaný také jako Raschův), ve kterém navíc předpokládáme, že všechny položky mají diskriminační parametr a_j roven a . Pravděpodobnost správné odpovědi na položku má potom tvar

$$p_{ij}(\theta_i; a, b_j) = P(X_{ij} = 1 | \theta_i; a, b_j) = \frac{e^{a(\theta_i - b_j)}}{1 + e^{a(\theta_i - b_j)}}. \quad (3.4)$$

Díváme-li se na pravděpodobnosti p_{ij} jako na funkce parametru θ při pevně daných parametrech položky, nazýváme je charakteristické funkce položky (*angl. item characteristic functions, ICFs*). Při jejich vykreslení na škále parametru θ je pak nazýváme charakteristické křivky položky (*angl. item characteristic curves, ICCs*).

Vhodný model volíme jak na základě dat a příslušných metod pro výběr modelů, tak na základě naší znalosti o tvorbě a průběhu znalostního testu. Tak můžeme posoudit, které předpoklady je rozumné uvažovat a které například nikoli. V této kapitole budeme pro ilustraci pracovat převážně s 2PL IRT modelem.

Abychom mohli s IRT modelem pracovat, je potřeba odhadnout jak parametry položek (a_j, b_j) , tak parametry ability respondentů (θ_i) . K tomu existují dva základní přístupy - jeden pomocí metody sdružené maximální věrohodnosti a druhý pomocí metody marginální maximální věrohodnosti. V této práci je nebudeme podrobně uvádět, jejich popis a srovnání lze například nalézt v Holland (1990). Co je ovšem potřeba zmínit, je fakt, že oba přístupy k odhadnutí parametrů musí předpokládat nějaké pravděpodobnostní rozdělení pro parametr θ . Nejčastější volbou je standardizované normální rozdělení s nulovou střední hodnotou a jednotkovým rozptylem, $N(0,1)$.

Metoda vyvažování pomocí IRT modelů se dá obecně rozdělit do 3 kroků. Prvním krokem je právě odhadnutí parametrů položek a abilit respondentů, druhým pak transformace těchto parametrů na společnou škálu. V případě designu dvou ekvivalentních skupin není transformace potřeba, protože parametry již na stejné škále jsou. Protože předpokládáme, že skupiny jsou ekvivalentní, předpoklad stejného pravděpodobnostního rozdělení parametru θ v kroku odhadování parametrů nijak nevádí. Ovšem v případě neekvivalentních skupin už je tento předpoklad problematický. Odhadneme-li parametry položek a respondentů zvlášť ve skupině, která vyplňovala test X, a ve skupině, která vyplňovala test Y, budeme v obou případech předpokládat, že θ má stejné pravděpodobnostní rozdělení, a to například $N(0,1)$. V případě neekvivalentních skupin ale víme, že rozdělení být stejné nemůže. Je proto potřeba odhadnuté parametry z jedné skupiny převést na škálu parametrů druhé skupiny, aby byly porovnatelné. Posledním krokem metody vyvažování pomocí IRT modelů je pak samotná transformace celkového skóre z jednoho testu na škálu skóre druhého testu.

3.2 Metody odhadu vyvažovacích konstant

Označme θ_{Xi} abilitu i -tého respondenta na škále parametrů z testu X a θ_{Yi} jeho abilitu na škále testu Y. Budeme-li předpokládat, že vztah mezi škálami je lineární, potom platí vztah

$$\theta_{Yi} = A\theta_{Xi} + B, \quad (3.5)$$

kde A a B jsou konstanty, které nazýváme vyvažovacími koeficienty (*angl. equating coefficients*). Vztahy mezi parametry položek pro položku j jsou ve tvaru

(Kolen a Brennan (2014, str. 178))

$$a_{Yj} = \frac{a_{Xj}}{A}, \quad (3.6)$$

$$b_{Yj} = Ab_{Xj} + B, \quad (3.7)$$

kde parametry s indexem X vždy značí parametry na škále X a obdobně parametry s indexem Y jsou parametry na škále Y .

Je snadné ukázat, že touto transformací parametry nezmění svoji interpretaci a že se jedná o vhodnou transformaci škál. Uvažujeme-li 2PL IRT model, má pravděpodobnost, že i -tý respondent zodpoví položku j v testu Y správně, tvar

$$p_{ij} = \frac{e^{a_{Yj}(\theta_{Yi} - b_{Yj})}}{1 + e^{a_{Yj}(\theta_{Yi} - b_{Yj})}}.$$

Dosadíme-li za θ_{Yi} , a_{Yj} a b_{Yj} výrazy z rovnic (3.5) - (3.7), dostaneme

$$\begin{aligned} p_{ij} &= \frac{e^{\frac{a_{Xj}}{A}(A\theta_{Xi} + B - Ab_{Xj} - B)}}{1 + e^{\frac{a_{Xj}}{A}(A\theta_{Xi} + B - Ab_{Xj} - B)}} \\ &= \frac{e^{\frac{a_{Xj}}{A}(A(\theta_{Xi} - b_{Xj}))}}{1 + e^{\frac{a_{Xj}}{A}(A(\theta_{Xi} - b_{Xj}))}} \\ &= \frac{e^{a_{Xj}(\theta_{Xi} - b_{Xj})}}{1 + e^{a_{Xj}(\theta_{Xi} - b_{Xj})}}. \end{aligned} \quad (3.8)$$

Pravděpodobnost p_{ij} je tedy ve stejném tvaru, jen parametry jsou nyní na škále X .

Nyní představíme několik vybraných metod, jak odhadnout vyvažovací koeficienty A a B , abychom mohli parametry transformovat.

3.2.1 Momentová metoda využívající průměr

Jako první představíme momentovou metodu využívající průměr (*angl. mean-mean method*). Z rovnic (3.6) a (3.7) dostaneme přímo vyjádření, že

$$A = \frac{a_{Xj}}{a_{Yj}} \quad \text{a} \quad (3.9)$$

$$B = b_{Yj} - Ab_{Xj}. \quad (3.10)$$

Protože vztahy (3.9) a (3.10) platí pro všechny položky, platí také (Kolen a Brennan, 2014, str. 180)

$$A = \frac{\mu(a_X)}{\mu(a_Y)} \quad \text{a} \quad (3.11)$$

$$B = \mu(b_Y) - A\mu(b_X), \quad (3.12)$$

kde $\mu(a_X)$ je průměr parametrů a_{Xj} , $\mu(b_X)$ je průměr parametrů b_{Xj} , $\mu(a_Y)$ a $\mu(b_Y)$ jsou průměry po řadě parametrů a_{Yj} , b_{Yj} , kde $j \in \{J - V + 1, \dots, J\}$. Využíváme zde pouze parametry kotvících položek (kotvící položky jsou indexovány od $J - V + 1$ do J), protože jen ty byly zadávány v obou verzích testu, a známe tak odhady parametrů a_j a b_j na obou škálách X i Y. To nám pak umožní získat odhady vyvažovacích konstant.

Odhad konstanty A pak získáme dosazením odhadů \hat{a}_{Xj} , \hat{a}_{Yj} do rovnice (3.11), odhad konstanty B dosazením odhadů \hat{b}_{Xj} , \hat{b}_{Yj} a odhadu konstanty \hat{A} do rovnice (3.12).

3.2.2 Momentová metoda využívající průměr a směrodatnou odchylku

Další momentovou metodou je metoda využívající směrodatné odchylky parametrů b_j , (*angl. mean-sigma method*). Z rovnice (3.7) můžeme konstantu A vyjádřit jako

$$A = \frac{b_{Yj} - B}{b_{Xj}}, \quad (3.13)$$

pro položku j a ze stejné rovnice dostaneme vyjádření pro konstantu B ve tvaru

$$B = b_{Yj^*} - Ab_{Xj^*} \quad (3.14)$$

pro položku j^* . Dosazením (3.14) do (3.13) pak dostaneme vyjádření konstanty A , kde

$$A = \frac{b_{Yj} - b_{Yj^*}}{b_{Xj} - b_{Xj^*}}. \quad (3.15)$$

Odtud dostaneme, že platí

$$A = \frac{\sigma(b_Y)}{\sigma(b_X)}, \quad (3.16)$$

kde $\sigma(b_X)$ je výběrová směrodatná odchylka parametrů b_{Xj} , $\sigma(b_Y)$ je výběrová směrodatná odchylka parametrů b_{Yj} , $j \in \{J - V + 1, \dots, J\}$.

Dosadíme-li směrodatné odchylky odhadů b_j na škálách X a Y do (3.16), získáme odhad konstanty A momentovou metodou využívající průměr a směrodatné odchylky. Odhad konstanty B je stejný jako u momentové metody využívající jen průměry, tedy opět vychází z rovnice (3.12).

3.2.3 Haebárova metoda

V (3.8) jsme ukázali, že pravděpodobnost, že i -tý respondent zodpoví položku j v testu Y správně, je stejná, ať už použijeme původní parametry θ_{Yi} , a_{Yj} , b_{Yj} nebo jejich transformace z (3.5)-(3.7), tj. že platí:

$$p_{ij}(\theta_{Yi}; a_{Yj}, b_{Yj}) = p_{ij}\left(A\theta_{Xi} + B; \frac{a_{Xj}}{A}, Ab_{Xj} + B\right). \quad (3.17)$$

V praxi, kdy však místo skutečných parametrů použijeme jejich odhady, už tato rovnost platit nemusí a pro všechny kotvící položky současně ani velmi pravděpodobně platit nebude. Z toho vychází i Haebarova metoda (Haebara, 1980) odhadu vyvažovacích koeficientů A a B . Ta hledá koeficienty tak, aby v nějakém smyslu minimalizovaly porušení rovnosti (3.17). Definujme součet čtverců rozdílů charakteristických křivek položky přes všechny kotvící položky pro pevné θ_i ,

$$H(\theta_i) = \sum_{j=J-V+1}^J \left[p_{ij}(\theta_i; \hat{a}_{Yj}, \hat{b}_{Yj}) - p_{ij}\left(\theta_i; \frac{\hat{a}_{Xj}}{A}, A\hat{b}_{Xj} + B\right) \right]^2. \quad (3.18)$$

Součty rozdílů $H(\theta_i)$ pak ještě sečteme přes všechny uvažované indexy i a hledáme takové A a B , které následující kritérium minimalizují,

$$H_{crit} = \sum_i H(\theta_i). \quad (3.19)$$

Vyvažovací koeficienty, které zvolené kritérium minimalizují, hledáme nějakou vhodnou iterační metodou.

Metoda se také může lišit v závislosti na tom, jaké indexy i v (3.19) uvažujeme. Můžeme počítat přes odhadnuté ability všech respondentů, přes odhadnuté ability jen těch respondentů, kterým byl zadáván např. test X , nebo také přes vybrané hodnoty θ_i vážené na základě proporcí v populaci. Podrobnosti i další přístupy nalezneme v Kolen a Brennan (2014) na straně 185.

V praktické části práce pak budeme k IRT vyvažování využívat balík *equateIRT* (Battaui, 2015), který z praktických důvodů pro nižší výpočetní náročnost (především v případě velkého počtu respondentů) a případná zobecnění minimalizuje funkci

$$H_{crit} = \frac{1}{2} \int_{-\infty}^{\infty} \sum_{j=J-V+1}^J \left[p_{ij}(\theta_i; \hat{a}_{Yj}, \hat{b}_{Yj}) - p_{ij}\left(\theta_i; \frac{\hat{a}_{Xj}}{A}, A\hat{b}_{Xj} + B\right) \right]^2 \psi(\theta_i) d\theta_i, \quad (3.20)$$

kde $\psi(\theta_i)$ je hustota standardizovaného normálního rozdělení. Integrál typicky nemá analytické řešení a obecně je aproximován pomocí Gaussova kvadraturního pravidla.

Takto nalezené konstanty A a B pak považujeme za odhady skutečných vyvažovacích koeficientů A , B .

3.2.4 Stocking-Lordova metoda

Stocking-Lordova metoda je obdobná jako Haebarova metoda, minimalizační kritérium ale využívá druhou mocninu rozdílu součtu charakteristických křivek všech kotvících položek, tj. pro pevné θ_i definujeme

$$SL(\theta_i) = \left[\sum_{j=J-V+1}^J p_{ij}(\theta_i; \hat{a}_{Yj}, \hat{b}_{Yj}) - \sum_{j=J-V+1}^J p_{ij}\left(\theta_i; \frac{\hat{a}_{Xj}}{A}, A\hat{b}_{Xj} + B\right) \right]^2. \quad (3.21)$$

Za odhady konstant A a B pak považujeme ty hodnoty A , B , které minimalizují následující kritérium,

$$SL_{crit} = \sum_i SL(\theta_i) \quad (3.22)$$

pro zvolené indexy i .

Obdobně jako u Haebarovy metody, balík *equateIRT* v softwaru R minimalizuje funkci definovanou jako

$$SL_{crit} = \frac{1}{2} \int_{-\infty}^{\infty} \left[\sum_{j=J-V+1}^J p_{ij}(\theta_i; \hat{a}_{Yj}, \hat{b}_{Yj}) - \right. \quad (3.23)$$

$$\left. - \sum_{j=J-V+1}^J p_{ij} \left(\theta_i; \frac{\hat{a}_{Xj}}{A}, A\hat{b}_{Xj} + B \right) \right]^2 \psi(\theta_i) d\theta_i, \quad (3.24)$$

kde $\psi(\theta_i)$ je hustota standardizovaného normálního rozdělení.

Když už známe odhady vyvažovacích koeficientů A a B , můžeme převést všechny parametry položek i parametry abilit respondentů na stejnou škálu pomocí transformačních rovnic (3.5)-(3.7).

3.3 Vyvažování celkových skóre

Kdybychom jako reportované skóre z testů uváděli právě hodnotu θ_i (často se uvádí jako tzv. IRT skóre), byli bychom v předcházejícím kroku již hotovi, protože bychom k transformaci na společnou škálu využili rovnici (3.5). V případě zadávání znalostních testů je ovšem typičtější udávat jako výsledné skóre testu počet správných odpovědí, nebo nějakou podobnou míru. Zde je potom potřeba provést ještě jeden dodatečný krok. Dvě běžně k tomu používané metody jsou vyvažování skutečných skóre a vyvažování pozorovaných skóre. Obě tyto metody transformují parametry položek na společnou škálu nějakou výše zmíněnou metodou za pomoci rovnic (3.6) a (3.7), transformaci parametru ability θ_i však nevyužívají. Obě se naopak dívají na podmíněné rozdělení celkových skóre při dané hodnotě parametru θ_i . Blíže si metody stručně představíme v následujících dvou sekcích, popsány jsou i v Kolen a Brennan (2014). Protože již uvažujeme parametry položek transformované na stejnou škálu, při značení už vynecháváme indexy X a Y .

3.3.1 Vyvažování skutečných skóre

Prvním přístupem, který uvedeme, je vyvažování skutečných skóre. Skutečné skóre definujeme jako podmíněnou střední hodnotu náhodné veličiny udávající počet správných odpovědí v testu, při pevně dané hodnotě θ_i , tj.

$$\tau_X(\theta_i) = \mathbb{E} \left(\sum_{j=1}^J X_{ij} | \theta_i \right) = \sum_{j=1}^J p_{ij}(\theta_i; a_j, b_j) \quad (3.25)$$

je skutečné skóre v testu X pro respondenta s abilitou θ_i a

$$\tau_Y(\theta_i) = \mathbb{E} \left(\sum_{j=J-V+1}^{J+K-V} Y_{ij} | \theta_i \right) = \sum_{j=J-V+1}^{J+K-V} p_{ij}(\theta_i; a_j, b_j) \quad (3.26)$$

je skutečné skóre respondenta s abilitou θ_i v testu Y.

Předpokládáme, že pro pevně dané θ_i se skutečné skóre $\tau_X(\theta_i)$ rovná skutečnému skóre $\tau_Y(\theta_i)$. Za platnosti tohoto předpokladu už dostaneme tvar vyvažovací funkce, jak spočítat skutečné skóre v testu Y při známé hodnotě skutečného skóre z testu X, tj.

$$\varphi_{IRT}(\tau_X, \theta_i) = \tau_Y(\tau_X^{-1}). \quad (3.27)$$

Hodnotu τ_X^{-1} v tomto případě definujeme jako to θ_i , které odpovídá skutečnému skóre τ_X při daných hodnotách parametrů a_j a b_j . V praxi skutečné hodnoty těchto parametrů neznáme, proto bychom do rovnic (3.25) a (3.26) dosazovali jejich odhady \hat{a}_j a \hat{b}_j . Stejně tak však neznáme ani skutečné skóre τ_X (příp. τ_Y), protože se jedná o nepozorovaný parametr. Vezmeme tedy odvozený vztah pro vyvažování (3.27) a místo skutečného skóre τ_X do něj dosadíme pozorované celkové skóre z testu X. Metoda tedy probíhá v následujících krocích:

1. Vezmeme pozorované celkové skóre z testu X, tj. $\sum_{j=1}^J X_{ij}$, $0 \leq \sum_{j=1}^J X_{ij} \leq J$.
2. Z rovnice $\sum_{j=1}^J X_{ij} = \sum_{j=1}^J p_{ij}(\theta_i; \hat{a}_j, \hat{b}_j)$ vyjádříme θ_i (k tomu můžeme využít například iterativní Newton-Raphsonovu metodu, protože hledáme kořen nelineární funkce).
3. Dosadíme θ_i do výrazu $\sum_{j=J-V+1}^{J+K-V} p_{ij}(\theta_i; \hat{a}_j, \hat{b}_j)$ a získáme tak ekvivalent pozorovaného celkového skóre $\sum_{j=1}^J X_{ij}$ na škále testu Y.

3.3.2 Vyvažování pozorovaných skóre

Druhým přístupem ve vyvažování pomocí IRT modelů je využít příslušný IRT model k odhadu pravděpodobnostního rozdělení celkového pozorovaného skóre v obou verzích testu a na ty následně aplikovat metodu ekvipercilového vyvažování, která byla představena v sekci 1.3. Označme $f_1(x = 1 | \theta_i) = p_{i1}$ pravděpodobnost, že respondent s abilitou θ_i odpoví na první položku v testu X správně a jeho celkové skóre bude tak po první položce rovno 1. Pravděpodobnost, že odpoví na první položku nesprávně a jeho celkové skóre bude po první položce 0 je potom $f_1(x = 0 | \theta_i) = 1 - p_{i1}$. Obecně označme $f_r(x | \theta_i)$ rozdělení celkového dosaženého skóre po prvních r položkách u respondenta s abilitou θ_i . Pro $r > 1$ je $f_r(x | \theta_i)$ definováno rekurzivně (Kolen a Brennan, 2014, str. 199)

$$\begin{aligned} f_r(x | \theta_i) &= f_{r-1}(x | \theta_i)(1 - p_{ir}), & x = 0, \\ &= f_{r-1}(x | \theta_i)(1 - p_{ir}) + f_{r-1}(x - 1 | \theta_i)p_{ir}, & 0 < x < r, \\ &= f_{r-1}(x - 1 | \theta_i)p_{ir} & x = r. \end{aligned} \quad (3.28)$$

Takto můžeme spočítat podmíněné rozdělení celkového skóre z testu X při dané hodnotě ability θ_i , $f(x|\theta_i) = f_J(x|\theta_i)$. Pokud je rozdělení parametru θ spojité, dostaneme nepodmíněné marginální rozdělení celkového skóre jako

$$f(x) = \int f(x|\theta)\psi(\theta)d\theta, \quad (3.29)$$

kde $\psi(\theta)$ je hustota θ .

Když už známe nepodmíněné rozdělení celkového skóre v testu X, obdobně můžeme spočítat i nepodmíněné rozdělení celkového skóre v testu Y, z hustot vyjádřit distribuční funkce a použít metodu ekvipercenilového vyvažování popsanou v sekci 1.3.

4. Praktická část

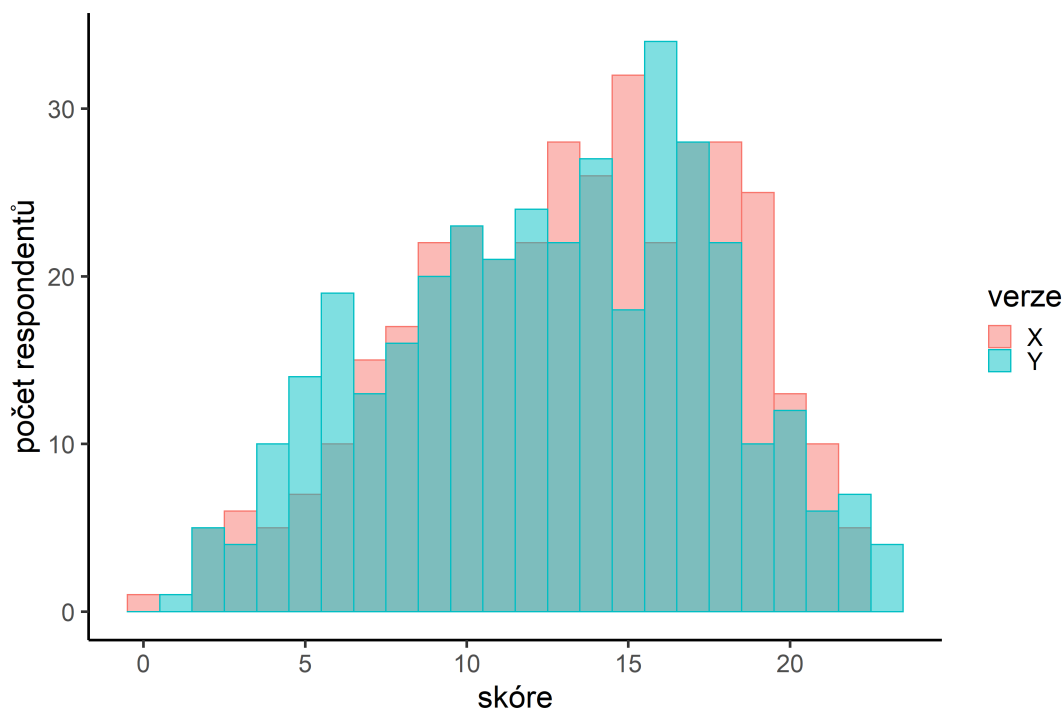
V praktické části práce si představíme výše zmíněné metody na reálném datovém souboru a porovnáme jejich výsledné vyvažovací funkce a standardní chyby vyvažování.

Pracovat budeme ve volně dostupném statistickém softwaru R (verze 4.1.0, viz R Core Team (2021)). Tradiční vyvažovací metody jsou implementovány v balíku *equate* (Albano, 2016), vyvažování pomocí jádrových odhadů v balíku *kequate* (Andersson a kol., 2013) a vyvažování pomocí modelů teorie odpovědi na položku v balíku *equateIRT* (Battauz, 2015).

4.1 Datový soubor

K analýze využijeme reálná data z mezinárodního šetření TIMSS (*Trends in International Mathematics and Science Study*), které je koordinováno Mezinárodní asociací pro hodnocení výsledků vzdělávání (*angl. The International Association for the Evaluation of Educational Achievement, IEA*). Jedná se o mezinárodní studii, která monitoruje trendy a zjišťuje úroveň vědomostí žáků v matematice a v přírodních vědách. Šetření bylo poprvé provedeno v roce 1995 a od té doby probíhá v pravidelných čtyřletých cyklech. Zapojeno je do něj více než 60 zemí světa, v každé z nich je vybrán reprezentativní vzorek přibližně 4000 až 5000 studentů z různých typů škol. Testování se účastní žáci 4. a 8. ročníků povinné školní docházky (případně jiných srovnatelných ročníků, které odpovídají věkovým kategoriím devítiletých a třináctiletých žáků). Kromě výsledků z testů z matematiky a přírodních věd jsou zaznamenávány i dodatečné informace o studentech a školách, a to pomocí dotazníků, které jsou administrovány studentům, jejich rodičům a ředitelům škol. Samotné vědomostní testy z matematiky a přírodních věd jsou v každém cyklu testování vytvářeny ve 14 různých verzích, kde vždy alespoň dvě verze mají nějakou sadu stejných položek, které můžeme využít jako kotvící položky při procesu vyvažování. V průběhu testování je pak každému studentovi náhodně přiřazena jedna ze 14 možných verzí testu.

V našem případě se budeme zabývat datovým souborem z šetření v roce 2015, které proběhlo u studentů 4. ročníků v České republice. Zaměříme se na znalostní test z matematiky, ve kterém bylo testováno celkem 5198 studentů. Protože všechny vyvažovací metody byly popsány jen pro dvě rozdílné verze testu, zůžeme si datový soubor z původních 14 verzí testu a 5198 studentů pouze na dvě verze testu, a tedy 731 studentů. Budeme-li používat stejné označení jako v celé práci, potom verzi X vyplnilo 371 žáků a verzi Y 360 žáků. Obě verze testu se skládaly z 23 otázek, které byly hodnoceny pouze binárně. Pokud student zodpověděl otázku správně, získal jeden bod, pokud odpověděl nesprávně, bod nezískal. Maximální počet bodů v obou verzích testu byl tedy 23, celkové skóre v testech mohlo nabývat 24 možných hodnot z množiny $\{0, \dots, 23\}$. Společných otázek v obou verzích testu bylo 12, možná dosažená skóre z kotvícího testu jsou tak z množiny $\{0, \dots, 12\}$. Kvůli relativně nízkému počtu otázek v obou testech zde budeme oproti teoretické části práce uvažovat jeden rozdíl, a to že



Obrázek 4.1: Histogram celkového skóre z verze testu X a z verze testu Y.

	test X	test Y
průměr	13.21	12.63
směr. odchylka	4.77	4.99
min	0	1
max	22	23

Tabulka 4.1: Základní popisné charakteristiky celkových skóre testu X a Y.

skóre z kotvících položek budeme započítávat i do celkového skóre z testu. Popis většiny kroků a metod je ovšem stejný, liší se pouze v prvním kroku vyhlazení pravděpodobností skóre. Zde musíme dát pozor na to, aby pravděpodobnosti, že respondent dosáhne nižšího celkového skóre než je jeho skóre z kotvícího testu, byly nulové.

Nejprve budeme předpokládat, že data pocházejí z designu dvou ekvivalentních skupin. Protože při testování jsou jednotlivé verze testu z matematiky přidělovány náhodně, je tedy celkem rozumné předpokládat, že se jedná o homogenní skupiny. Histogramy celkových dosažených skóre v testu X a v testu Y jsou zobrazeny na Obrázku 4.1. Vidíme zde, že se výrazně překrývají, rozdíl v obtížnosti testů X a Y tedy nejspíš nebude příliš velký. Mírně obtížnější by mohl být test Y, protože počet respondentů s vyšším celkovým dosaženým skóre se zdá být u testu X lehce vyšší. Tomu napovídají i průměry celkových skóre, kde v testu X je průměrné dosažené skóre 13.21 (směrodatná odchylka ± 4.77), v testu Y je průměrné skóre 12.63 (± 4.99), viz Tabulka 4.1.

4.2 Tradiční vyvažovací metody

Nejprve začneme tradičními vyvažovacími metodami, které byly popsány v kapitole 1. Ukážeme si všechny tři představené metody, tedy vyvažování ve střední hodnotě, lineární vyvažování i ekvipercentilové vyvažování. Stejně jako v celé teoretické části, budeme i zde vždy převádět skóre ze škály testu X na škálu testu Y , tj. pro zvolené skóre ve verzi X budeme hledat jeho ekvivalent na škále verze Y .

4.2.1 Vyvažování ve střední hodnotě

Metoda vyvažování ve střední hodnotě je velmi jednoduchá a přímočará. V tomto případě nám stačí spočítat rozdíl mezi průměrem dosaženého celkového skóre v testu X a průměrem dosaženého celkového skóre v testu Y a přičíst jej ke zvolenému skóre z testu X , ke kterému hledáme ekvivalent. Vyvažovací funkce (resp. její odhad) má tvar jako v (1.2), zde konkrétně

$$\hat{\varphi}_M(x) = x - 13.21 + 12.63 = x - 0.58.$$

Od každého skóre z testu X tak stačí odečíst hodnotu 0.58 a dostaneme jeho ekvivalentní skóre na škále testu Y . Přepočítání pro všechna možná skóre testu X vidíme v Tabulce 4.2 v druhém sloupci.

4.2.2 Lineární vyvažování

V případě metody lineárního vyvažování potřebujeme kromě průměrů spočítat ještě výběrové směrodatné odchylky celkových skóre testů X a Y . Ty pak dosadíme do rovnice (1.4) ze sekce 1.2 a dostaneme tak odhad vyvažovací funkce ve tvaru

$$\hat{\varphi}_L(x) = \frac{4.99}{4.77}x + \left[12.63 - \frac{4.99}{4.77}13.21\right] = 1.05x - 1.20.$$

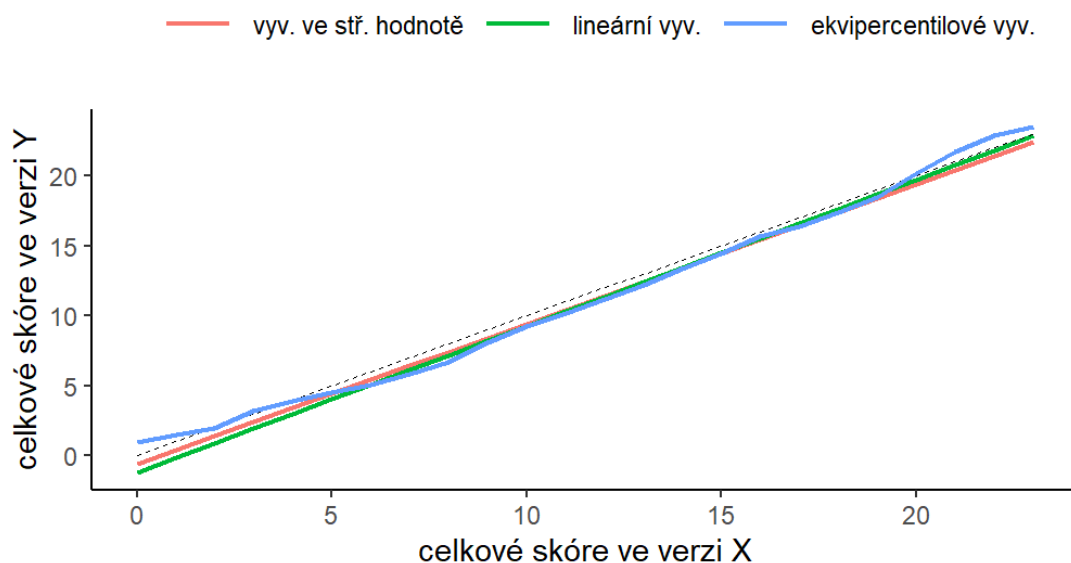
Konkrétní hodnoty pro všechny volby skóre x nalezneme v Tabulce 4.2 v třetím sloupci.

4.2.3 Ekvipercentilové vyvažování

Poslední uváděnou tradiční vyvažovací metodou bylo ekvipercentilové vyvažování. Zde má vyvažovací funkce tvar (1.5), její odhad dostaneme dosazením odhadů distribučních funkcí, viz. (1.6). V případě, kdy mají náhodné veličiny X a Y diskrétní rozdělení, je ovšem potřeba jejich distribuční funkce spojitě aproximovat, aby byla vždy dobře definována jejich inverzní funkce. Jednomu možnému přístupu, jak spojitě aproximovat diskrétní distribuční funkce, jsme se podrobně věnovali v kapitole o metodě vyvažování pomocí jádrových odhadů, konkrétně v sekci 2.1.3. Další možnou metodou je například metoda založená na percentilovém pořadí (*angl. percentile ranks*), kterou právě využívá balík *equate*, který zde používáme. Tuto metodu jsme v teoretické části nezmiňovali, protože metoda popsána v 2.1.3 je velmi podobná, ale flexibilnější. Podrobnosti k metodě percentilového pořadí můžeme nalézt v Kolen a Brennan (2014, str. 37).

celkové skóre	vyvažování ve stř. hodnotě	lineární vyvažování	ekvipercenilové vyvažování
0	-0.58	-1.20	0.99
1	0.42	-0.15	1.47
2	1.42	0.89	1.98
3	2.42	1.94	3.18
4	3.42	2.99	3.91
5	4.42	4.03	4.49
6	5.42	5.08	5.08
7	6.42	6.13	5.83
8	7.42	7.17	6.72
9	8.42	8.22	8.04
10	9.42	9.27	9.23
11	10.42	10.31	10.19
12	11.42	11.36	11.16
13	12.42	12.41	12.21
14	13.42	13.45	13.37
15	14.42	14.50	14.44
16	15.42	15.55	15.69
17	16.42	16.59	16.41
18	17.42	17.64	17.36
19	18.42	18.69	18.49
20	19.42	19.73	20.18
21	20.42	20.78	21.69
22	21.42	21.83	22.89
23	22.42	22.87	23.50

Tabulka 4.2: Ekvivalenty celkových skóre z testu X na škále testu Y pro metodu vyvažování ve střední hodnotě, lineární vyvažování a ekvipercenilové vyvažování.



Obrázek 4.2: Přepočtení celkového skóre z testu X na celkové skóre z testu Y pomocí vyvažování ve střední hodnotě, lineárního vyvažování a ekvipercenilového vyvažování.

Výsledná přepočtená skóre z verze testu X na skóre testu Y pomocí ekvipercenilového vyvažování a balíku *equate* vidíme v Tabulce 4.2, v posledním sloupci. Vykreslená všechna přepočtená skóre, získaná pomocí tradičních vyvažovacích metod, vidíme pro lepší vizualizaci také na Obrázku 4.2. Zde vidíme, že jednotlivé tradiční metody se od sebe příliš neliší. To je způsobeno také tím, že i samotné testy X a Y se příliš neliší v obtížnosti, distribuční funkce náhodných veličin X a Y , reprezentující celková skóre, tak nebudou výrazně odlišné.

4.3 Vyvažování pomocí jádrových odhadů

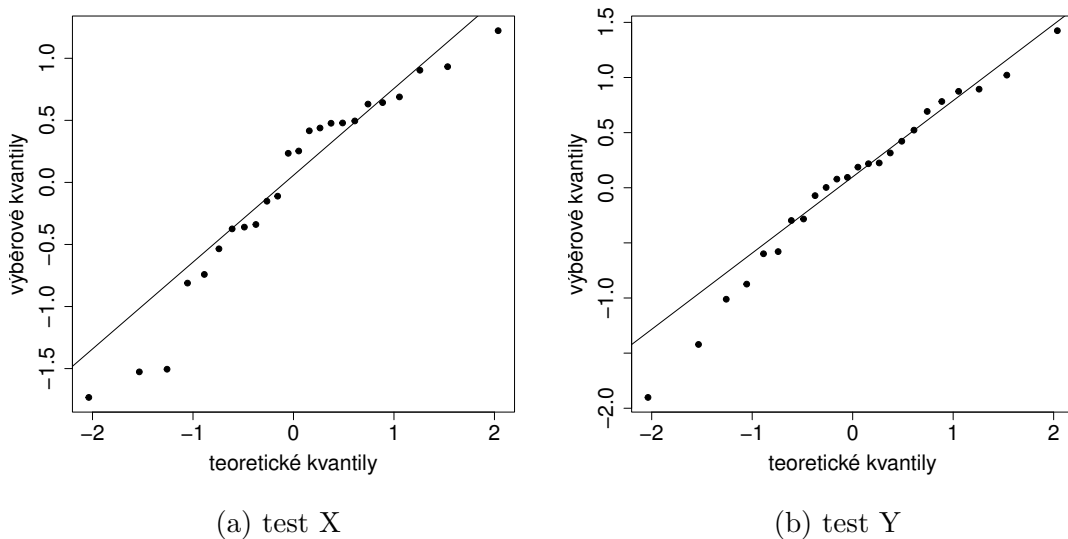
V této sekci si na zvoleném datovém souboru ukážeme metodu vyvažování pomocí jádrových odhadů, které se věnovala většina práce. Nejprve budeme stejně jako v předchozí části s tradičními vyvažovacími metodami uvažovat design dvou ekvivalentních skupin. Následně si na stejném datovém souboru pro srovnání ukážeme metodu i v případě, kdy bychom předpokládali neekvivalentní skupiny, a tedy design dvou neekvivalentních skupin s kotvícími položkami (právě proto jsme volili takový datový soubor, který kotvící položky obsahuje).

4.3.1 Design ekvivalentních skupin

Metodu vyvažování pomocí jádrových odhadů v designu dvou ekvivalentních skupin jsme podrobně popsali v sekci 2.1, a to všech 5 dílčích kroků. Začneme zde prvním z nich, vyhlazením pravděpodobností celkových skóre. K tomu nejprve potřebujeme vektory četností definovaných v (2.1), v našem konkrétním případě 24-složkové vektory $(n_1, \dots, n_{24})^\top$ a $(m_1, \dots, m_{24})^\top$. Tyto vektory četností neboli počty respondentů s danými hodnotami celkových skóre v testech X a Y nalez-

celkové skóre	počet respondentů v X	počet respondentů v Y
0	1	0
1	0	1
2	5	5
3	6	4
4	5	10
5	7	14
6	10	19
7	15	13
8	17	16
9	22	20
10	23	23
11	21	21
12	22	24
13	28	22
14	26	27
15	32	18
16	22	34
17	28	28
18	28	22
19	25	10
20	13	12
21	10	6
22	5	7
23	0	4

Tabulka 4.3: Počty respondentů v jednotlivých kategoriích celkového skóre.



Obrázek 4.3: Normální Q-Q grafy Freeman-Tukeyových reziduí pro četnosti celkových skóre v testech X a Y.

neme v Tabulce 4.3.

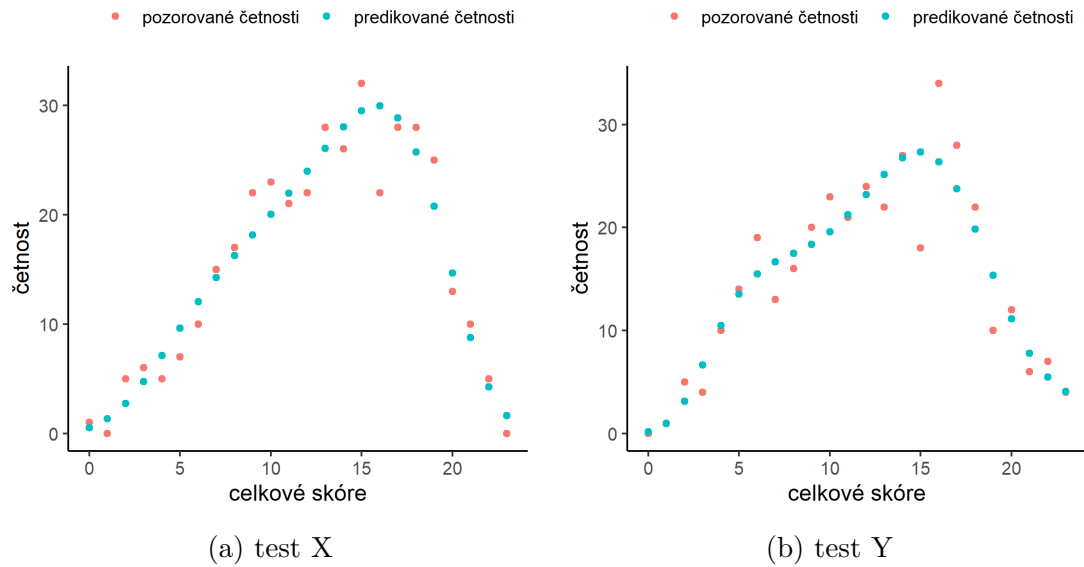
Dále potřebujeme odhady pravděpodobností $\mathbf{r} = (r_1, \dots, r_{24})^\top$, $\mathbf{s} = (s_1, \dots, s_{24})^\top$ definovaných v (2.2). Jak už jsme zmínili v sekci 2.1, místo relativních četností $\frac{n_j}{N}$ a $\frac{m_k}{M}$ použijeme odhady založené na log-lineárních modelech.

V obou případech jsme zvolili momentové modely definované rovnicí (2.6). V případě pravděpodobností \mathbf{r} jsme za nejvhodnější model zvolili model s prvními čtyřmi momenty, v případě pravděpodobností \mathbf{s} pak model s prvními pěti momenty, tedy

$$\log(r_j) = \alpha_r + \beta_{1r}x_j + \beta_{2r}x_j^2 + \beta_{3r}x_j^3 + \beta_{4r}x_j^4, \quad j = 1, \dots, 24, \quad (4.1)$$

$$\log(s_k) = \alpha_s + \beta_{1s}y_k + \beta_{2s}y_k^2 + \beta_{3s}y_k^3 + \beta_{4s}y_k^4 + \beta_{5s}y_k^5, \quad k = 1, \dots, 24. \quad (4.2)$$

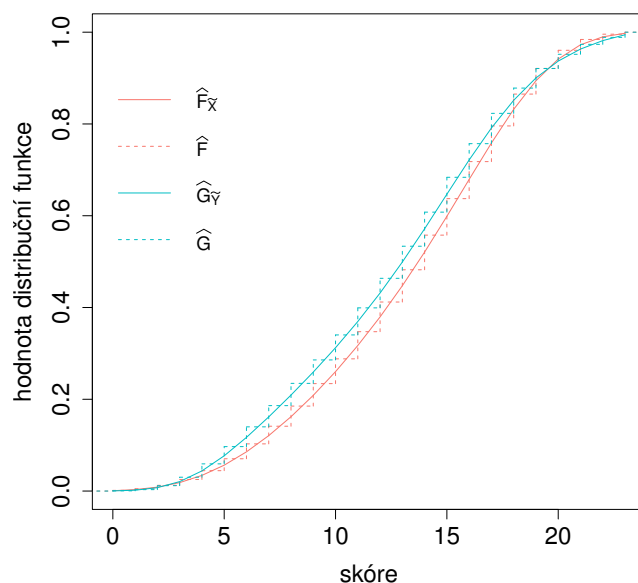
Při výběru modelu jsme v obou případech začali s modelem s prvními sedmi momenty celkového skóre a postupně jsme nejvyšší momenty odebírali. Pomocí testu poměrem věrohodnosti jsme pak vždy ověřili, že odebráním daného členu nedostaneme signifikantně horší model. V případě pravděpodobností \mathbf{r} jsme se zastavili u modelu s prvními čtyřmi momenty (p-hodnota testu poměrem věrohodnosti mezi modelem s prvními čtyřmi momenty a modelem s prvními třemi momenty je 0.002), v případě pravděpodobností \mathbf{s} jsme se zastavili u modelu s prvními pěti momenty (p-hodnota 0.034). U každého modelu jsme se pro doplnění podívali i na hodnoty informačních kritérií AIC a BIC, rozhodnutí o výběru modelu by bylo stejné. Normalitu Freeman-Tukeyových reziduí definovaných v (2.9) můžeme ověřit z grafů na Obrázku 4.3. Mírné odchýlení bodů od diagonály zde vidíme, ale není výrazné a proto můžeme modely považovat za vyhovující. Posledním diagnostickým grafem jsou vykreslené pozorované četnosti n_j a odhadnuté četnosti $N\hat{r}_j$, $j = 1, \dots, 24$ (resp. četnosti m_k a $M\hat{s}_k$, $k = 1, \dots, 24$) na Obrázku 4.4. I zde vidíme, že se jednotlivé četnosti až na pár výjimek od sebe příliš neliší a model je tak dostačující. Odhady pravděpodobností \mathbf{r} a \mathbf{s} získané



Obrázek 4.4: Pozorované a predikované četnosti celkových skóre v testu X a v testu Y v designu dvou ekvivalentních skupin.

celkové skóre	predikované pravd. \hat{r}_j	predikované pravd. \hat{s}_k
0	0.001	0.000
1	0.004	0.003
2	0.007	0.009
3	0.013	0.018
4	0.019	0.029
5	0.026	0.038
6	0.032	0.043
7	0.038	0.046
8	0.044	0.049
9	0.049	0.051
10	0.054	0.054
11	0.059	0.059
12	0.065	0.064
13	0.070	0.070
14	0.076	0.074
15	0.079	0.076
16	0.081	0.073
17	0.078	0.066
18	0.069	0.055
19	0.056	0.043
20	0.040	0.031
21	0.024	0.022
22	0.011	0.015
23	0.004	0.011

Tabulka 4.4: Predikované pravděpodobnosti \hat{r}_j a \hat{s}_k .



Obrázek 4.5: Spojité distribuční funkce $\hat{F}_{\tilde{X}}$, $\hat{G}_{\tilde{Y}}$ a po částech konstantní distribuční funkce \hat{F} , \hat{G} .

z modelů (4.1) a (4.2) vidíme v Tabulce 4.4.

Odhadnuté pravděpodobnosti $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$ nyní využijeme k získání odhadů distribučních funkcí $F_{\tilde{X}}$ a $G_{\tilde{Y}}$, které jsou spojitě aproximace po částech konstantních distribučních funkcí F a G , viz. rovnice (2.10). Tyto aproximace jsme popsali v sekci 2.1.3, konkrétní tvary distribučních funkcí $F_{\tilde{X}}$ a $G_{\tilde{Y}}$ jsou uvedeny v (2.12) a (2.18), jejich odhady v (2.13) a (2.19). Při volbě vyhlazovacích parametrů h_X a h_Y jsme v obou případech využili penalizační funkci (2.17) s parametry $\kappa = 1$ a $\omega = \frac{1}{2}$. Tím jsme dostali parametr h_X roven 0.58 a parametr h_Y roven 0.56. Vykreslené distribuční funkce $\hat{F}_{\tilde{X}}$ a $\hat{G}_{\tilde{Y}}$ vidíme na Obrázku 4.5. Pro doplnění jsou v grafu vykresleny i \hat{F} a \hat{G} , odhady distribučních funkcí původních diskretních náhodných veličin X a Y vycházející z (2.10).

Odhadnuté distribuční funkce $\hat{F}_{\tilde{X}}$ a $\hat{G}_{\tilde{Y}}$ už nyní přímo dosadíme do rovnice (2.21), ve které je uveden tvar vyvažovací funkce metody vyvažování pomocí jádrových odhadů. Z ní už můžeme pro všechna možná celková skóre z testu X spočítat jejich ekvivalenty na škále skóre z testu Y . Tento přepočít vidíme v Tabulce 4.5, v prvním sloupci.

Posledním pátým krokem metody je výpočet standardních chyb procesu vyvažování. Standardní chyba vyvažování je definována jako odmocnina z rozptylu odhadu vyvažovací funkce $\hat{\varphi}_K$, viz. rovnice (2.22). Detailní postup, jak tento rozptyl spočítat, je uveden v sekci 2.1.5. Konkrétní hodnoty standardních chyb v tomto příkladu vidíme v prvním sloupci Tabulky 4.6.

celk. skóre	EG design	NEAT design - PSE	NEAT design - CE
0	0.47	0.33	0.30
1	1.36	1.11	1.09
2	2.12	1.84	1.79
3	2.88	2.55	2.50
4	3.63	3.25	3.22
5	4.41	3.97	3.95
6	5.23	4.70	4.70
7	6.09	5.48	5.47
8	7.01	6.30	6.29
9	7.99	7.19	7.15
10	9.02	8.16	8.08
11	10.09	9.21	9.09
12	11.17	10.33	10.17
13	12.24	11.49	11.32
14	13.31	12.67	12.51
15	14.36	13.84	13.73
16	15.43	15.00	14.95
17	16.52	16.16	16.17
18	17.65	17.34	17.41
19	18.86	18.59	18.66
20	20.14	19.91	19.97
21	21.46	21.26	21.29
22	22.59	22.46	22.53
23	23.40	23.47	23.45

Tabulka 4.5: Ekvivalenty celkových skóre z testu X na škále testu Y získané metodou vyvažování pomocí jádrových odhadů v případě designu ekvivalentních skupin (EG), v případě designu neekvivalentních skupin s kotvícími položkami a poststratifikační metody (NEAT - PSE) a v případě designu neekvivalentních skupin s kotvícími položkami a metody řetězového vyvažování (NEAT - CE).

celk. skóre	EG design	NEAT design - PSE	NEAT design - CE
0	0.55	0.43	0.44
1	0.56	0.41	0.46
2	0.52	0.38	0.43
3	0.47	0.34	0.39
4	0.45	0.30	0.34
5	0.45	0.27	0.30
6	0.46	0.25	0.27
7	0.49	0.24	0.25
8	0.53	0.24	0.24
9	0.56	0.24	0.24
10	0.57	0.25	0.25
11	0.57	0.25	0.25
12	0.54	0.24	0.25
13	0.51	0.23	0.24
14	0.47	0.22	0.23
15	0.44	0.21	0.22
16	0.41	0.20	0.21
17	0.40	0.19	0.20
18	0.41	0.20	0.21
19	0.44	0.22	0.23
20	0.49	0.26	0.27
21	0.53	0.30	0.33
22	0.40	0.28	0.29
23	0.23	0.23	0.17

Tabulka 4.6: Standardní chyby vyvažování (SEE) při vyvažování skóre z testu X na skóre z testu Y pomocí jádrových odhadů v případě designu ekvivalentních skupin (EG), v případě designu neekvivalentních skupin s kotvícími položkami a poststratifikační metody (NEAT - PSE) a v případě designu neekvivalentních skupin s kotvícími položkami a metody řetězového vyvažování (NEAT - CE).

4.3.2 Design neekvivalentních skupin s kotvícími položkami

Nyní si metodu vyvažování pomocí jádrových odhadů ukážeme při designu dvou neekvivalentních skupin s kotvícími položkami. V tomto případě předpokládáme, že skupiny studentů, kteří vyplňovali test X a Y, nebyly srovnatelné a že jedna ze skupin byla schopnější. Rozdíl ve schopnostech odhadneme pomocí skóre z kotvícího testu (náhodnou veličinu reprezentující skóre z tohoto testu značíme A), který byl stejný pro obě skupiny a vyplnili ho tak všichni respondenti. V teoretické části 2.2 jsme představili dva možné přístupy, nejprve začneme poststratifikační metodou. Tato metoda se od designu ekvivalentních skupin liší především v odhadu pravděpodobností skóre \mathbf{r} a \mathbf{s} a jejich vyhlazení, další kroky jsou pak obdobné.

Poststratifikační metoda

K odhadu pravděpodobností \mathbf{p} a \mathbf{q} , definovaných v (2.43) a (2.44) opět využijeme log-lineární model založený na momentech celkového skóre, skóre z kotvícího testu a jejich vzájemné interakce. Model je blíže popsán v části 2.2.1.1 rovnicí (2.53). V našem konkrétním případě využijeme pro pravděpodobnosti p_{jl} , $j = 1, \dots, 24$, $l = 1, \dots, 13$ model ve tvaru

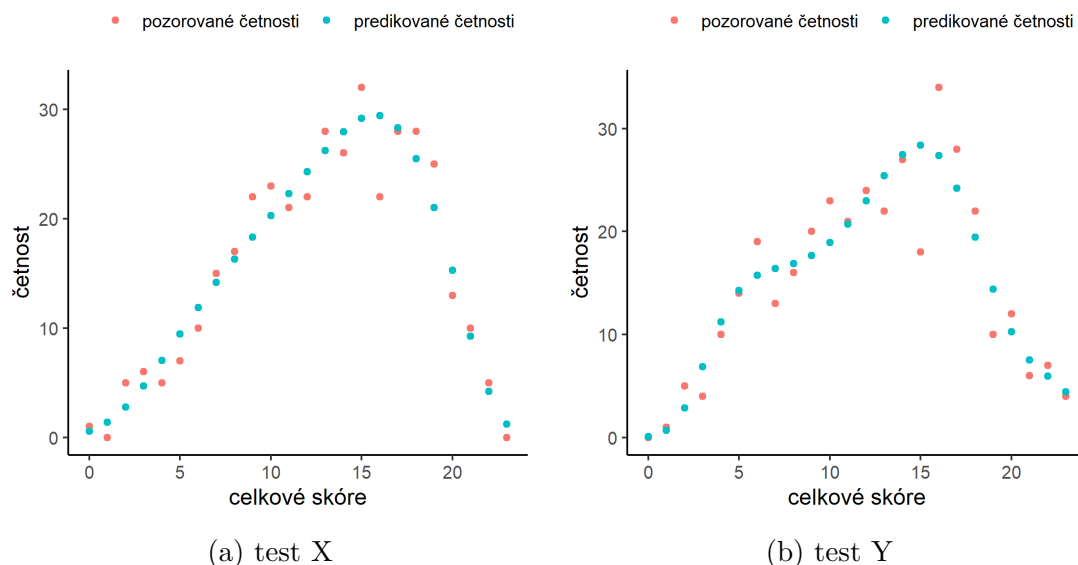
$$\begin{aligned} \log(p_{jl}) = & \alpha_p + \beta_{x,1}x_j + \beta_{x,2}x_j^2 + \beta_{x,3}x_j^3 + \beta_{x,4}x_j^4 + \beta_{a,1}a_l + \beta_{a,2}a_l^2 \\ & + \beta_{xa,11}x_ja_l + \beta_{xa,21}x_j^2a_l \end{aligned} \quad (4.3)$$

a pro pravděpodobnosti q_{kl} , $k = 1, \dots, 24$, $l = 1, \dots, 13$ volíme model tvaru

$$\begin{aligned} \log(q_{kl}) = & \alpha_q + \beta_{y,1}y_k + \beta_{y,2}y_k^2 + \beta_{y,3}y_k^3 + \beta_{y,4}y_k^4 + \beta_{a,1}a_l + \beta_{a,2}a_l^2 + \beta_{ya,11}y_ka_l \\ & + \beta_{ya,21}y_k^2a_l + \beta_{ya,12}y_ka_l^2 + \beta_{ya,22}y_k^2a_l^2 + \beta_{ya,31}y_k^3a_l + \beta_{ya,32}y_k^3a_l^2. \end{aligned} \quad (4.4)$$

V obou případech jsme vycházeli z modelu s prvními čtyřmi momenty celkového skóre, prvními čtyřmi momenty skóre z kotvícího testu a všemi jejich vzájemnými interakcemi (celkem tedy s 24 proměnnými). Postupně jsme odebírali nesignifikantní členy a modely mezi sebou testovali pomocí testu poměrem věrohodnosti. Vhodnost finálních modelů jsme pak ověřili vykreslením grafů porovnávajících pozorované a predikované četnosti celkového skóre, podobně jako v předchozím případě v designu dvou ekvivalentních skupin, viz Obrázek 4.6. Navíc jsme se zde ještě dívali i na podmíněné střední hodnoty a podmíněné směrodatné odchylky pozorovaných a predikovaných skóre, zda se příliš neliší. Výsledné grafy jsou vykresleny na Obrázku 4.7. Z příslušných grafů usuzujeme, že zvolené modely můžeme považovat za dostatečně vhodné.

Odhady pravděpodobností $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$ pak získáme dosazením odhadnutých pravděpodobností $\hat{\mathbf{p}} = (\hat{p}_{1,1}, \dots, \hat{p}_{24,13})^\top$ a $\hat{\mathbf{q}} = (\hat{q}_{1,1}, \dots, \hat{q}_{24,13})^\top$ do vzorců (2.56), (2.57). Nyní, když už máme odhady $\hat{\mathbf{r}}$ a $\hat{\mathbf{s}}$, jsou dalšími kroky spojitá aproximace skóre a samotný proces vyvažování. Tyto kroky jsou již zcela totožné jako v předchozím případě při designu dvou ekvivalentních skupin, nebudeme je zde proto více rozvádět. Výsledná přepočtená celková skóre z testu X na škálu testu Y vidíme



Obrázek 4.6: Pozorované a predikované četnosti celkových skóre v testu X a v testu Y v designu dvou neekvivalentních skupin s kotvícími položkami.

v Tabulce 4.5, v druhém sloupci.

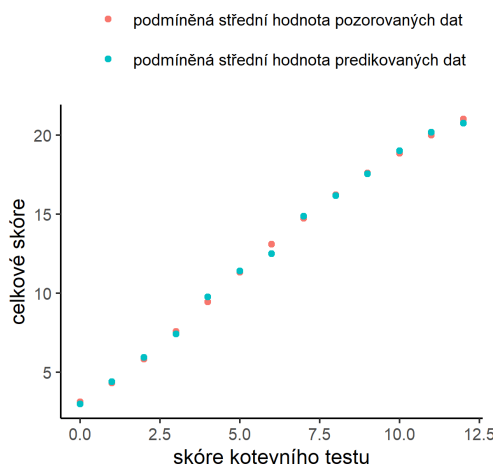
Ve druhém sloupci Tabulky 4.6 pak vidíme i standardní chyby vyvažování při designu neekvivalentních skupin získané poststratifikační metodou, teoreticky podrobně popsané v sekci 2.2.1.5.

Metoda řetězového vyvažování

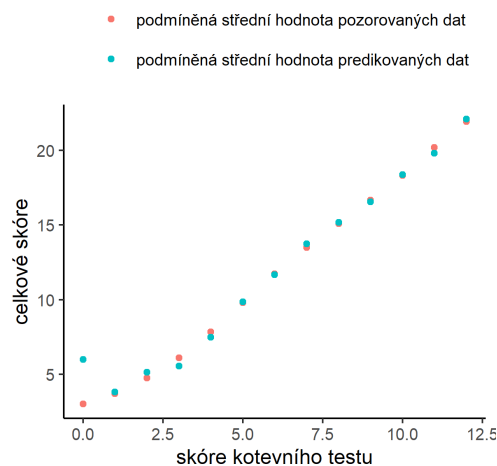
Alternativou k poststratifikační metodě je metoda řetězového vyvažování. Ta probíhá ve dvou fázích, kdy v první fázi je skóre z testu X transformováno na skóre z kotvícího testu A v populaci P a ve druhé fázi je skóre z kotvícího testu A transformováno na skóre z testu Y v populaci Q. To je možné provést za splnění předpokladu, že vyvažovací funkce z X do A a vyvažovací funkce z A do Y jsou invariantní vůči populaci, viz. předpoklady (CE1) a (CE2) na straně 31.

V prvním kroku potřebujeme znát odhady pravděpodobností p_{jl} a q_{kl} . Ty opět odhadneme log-lineárními modely, stejně jako v předchozím případě u poststratifikační metody. K odhadnutí pravděpodobností p_{jl} tedy využijeme model definovaný rovnicí (4.3), pravděpodobnosti q_{kl} odhadneme modelem (4.4).

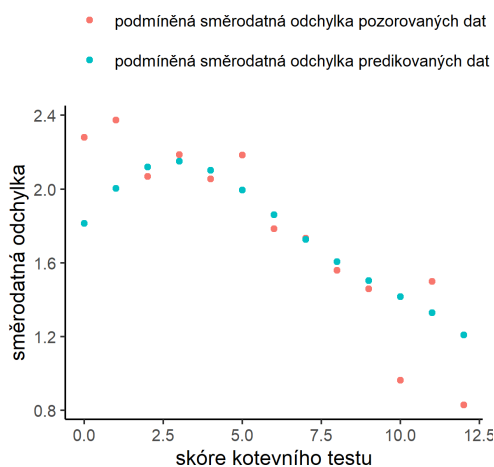
Za pomoci těchto odhadů už pak získáme i odhady pravděpodobností $\hat{\mathbf{r}}_P$, $\hat{\mathbf{s}}_Q$, $\hat{\mathbf{t}}_P$ a $\hat{\mathbf{t}}_Q$, které jsou definovány rovnicemi (2.74) - (2.77). Ty jsou klíčové pro vyjádření distribučních funkcí $\hat{F}_{\tilde{X}_P}$ (distribuční funkce celkového skóre z testu X v populaci P), $\hat{G}_{\tilde{Y}_Q}$ (distribuční funkce celkového skóre z testu Y v populaci Q), $\hat{H}_{\tilde{A}_P}$ (distribuční funkce celkového skóre z kotvícího testu A v populaci P) a $\hat{H}_{\tilde{A}_Q}$ (distribuční funkce celkového skóre z kotvícího testu A v populaci Q). Konkrétní vyjádření funkce $\hat{F}_{\tilde{X}_P}$ je uvedeno v rovnici (2.79), zbylé distribuční funkce vypadají zcela analogicky. Grafické znázornění všech čtyř distribučních funkcí vidíme



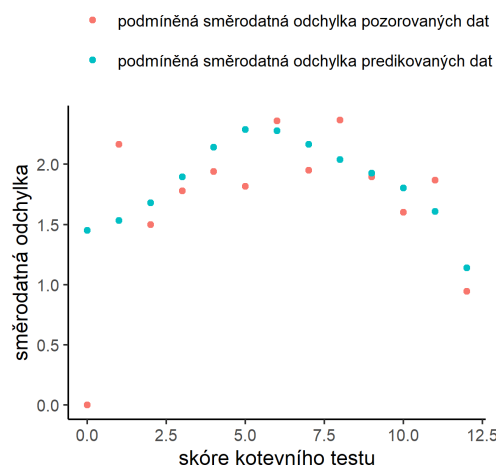
(a) Podmíněná střední hodnota celkového skóre z testu X při daném skóre z kotevního testu A.



(b) Podmíněná střední hodnota celkového skóre z testu Y při daném skóre z kotevního testu A.

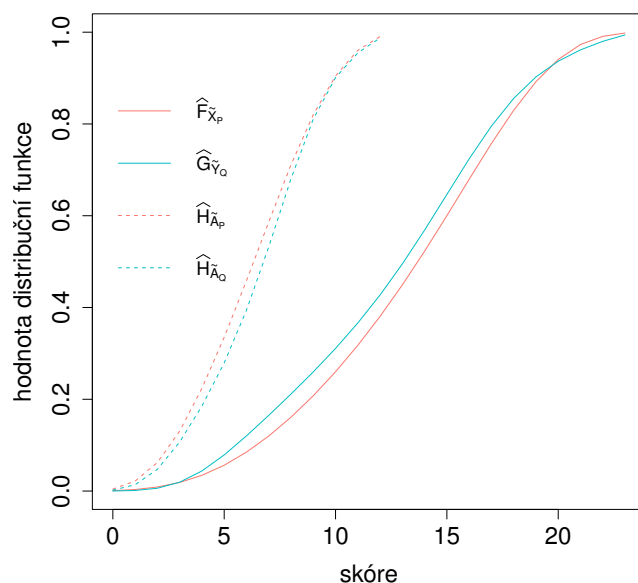


(c) Podmíněná směrodatná odchylka celkového skóre z testu X při daném skóre z kotevního testu A.



(d) Podmíněná směrodatná odchylka celkového skóre z testu Y při daném skóre z kotevního testu A.

Obrázek 4.7: Podmíněné střední hodnoty a podmíněné směrodatné odchylky celkových skóre z testů X a Y při daných hodnotách z kotevního testu A.

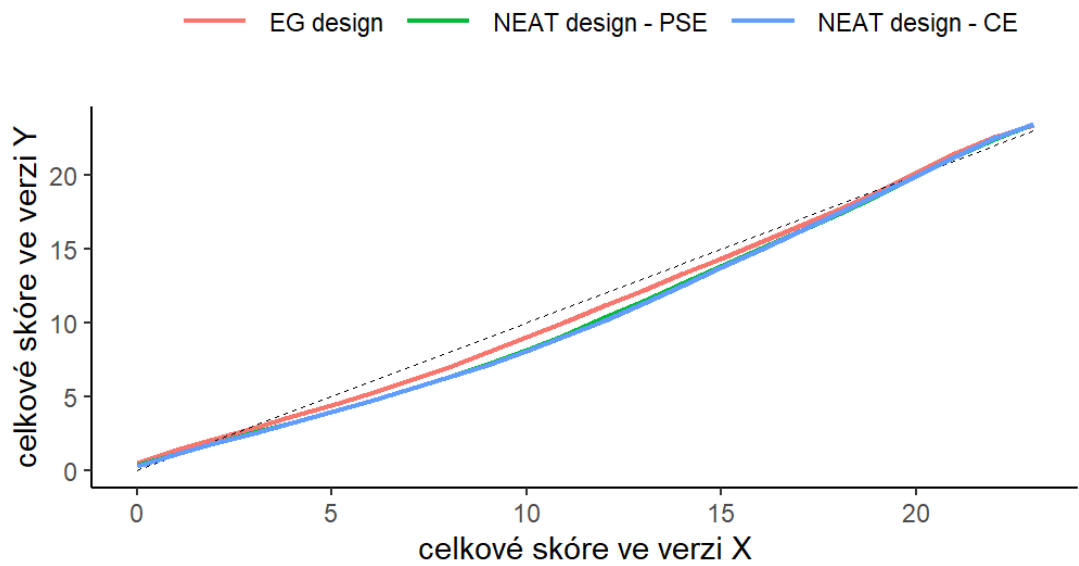


Obrázek 4.8: Spojité distribuční funkce $\widehat{F}_{\widetilde{X}_P}$, $\widehat{G}_{\widetilde{Y}_Q}$, $\widehat{H}_{\widetilde{A}_P}$ a $\widehat{H}_{\widetilde{A}_Q}$.

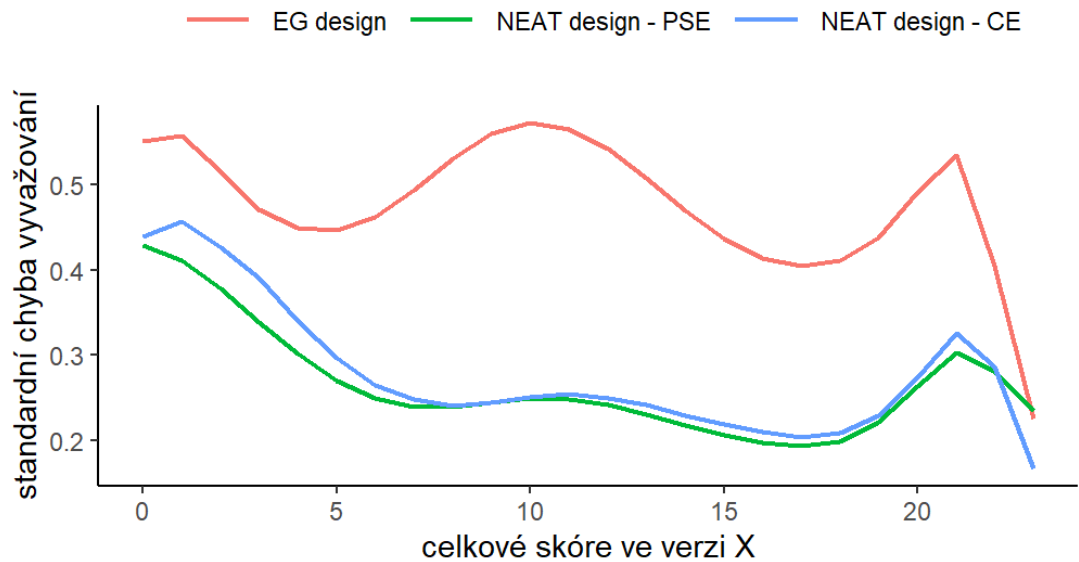
na Obrázku 4.8.

Nyní už máme všechny odhady, které potřebujeme dosadit do vyvažovací funkce metody řetězového vyvažování, která je definována rovnicí (2.80). Přepočtená skóre z testu X na škálu testu Y jsou pak uvedena v Tabulce 4.5 v posledním sloupci, standardní chyby vyvažování jsou v Tabulce 4.6, opět v posledním sloupci. Výpočet standardních chyb byl podrobně popsán v sekci 2.2.2, na stranách 34 až 36.

Hodnoty z Tabulky 4.5 a Tabulky 4.6 jsou pro přehlednost vykresleny i do grafů na Obrázku 4.9 a Obrázku 4.10. Na Obrázku 4.9 jsou vykresleny přepočtená skóre testu X na škálu testu Y a dobře zde vidíme, že metody opět dávají velmi podobné výsledky a že se na těchto konkrétních datech příliš neliší. Především pak postratifikační metoda a řetězové vyvažování, které navíc vychází ze stejného předpokladu neekvivalentních skupin. Tyto dvě metody mají také velmi podobné hodnoty standardních chyb vyvažování, viz. Obrázek 4.10. Mírně vyšší hodnoty standardních chyb má obecně metoda vyvažování za pomoci jádrových odhadů, která předpokládá ekvivalentní skupiny. Předpokládat neekvivalentní skupiny bude v tomto případě tedy asi lehce vhodnější, výběr konkrétního přístupu (postratifikační metoda nebo řetězové vyvažování) ale velkou roli hrát nebude.



Obrázek 4.9: Přepočítání celkového skóre z testu X na celkové skóre z testu Y metodami vyvažování pomocí jádrových odhadů, za předpokladu dvou ekvivalentních skupin (EG design), za předpokladu dvou neekvivalentních skupin pomocí postratifikační metody (NEAT design - PSE) a za předpokladu dvou neekvivalentních skupin pomocí řetězového vyvažování.



Obrázek 4.10: Standardní chyby vyvažování pro metody vyvažování pomocí jádrových odhadů, za předpokladu dvou ekvivalentních skupin (EG design), za předpokladu dvou neekvivalentních skupin pomocí postratifikační metody (NEAT design - PSE) a za předpokladu dvou neekvivalentních skupin pomocí řetězového vyvažování.

4.4 Vyvažování pomocí modelů teorie odpovědi na položku

Poslední metodou vyvažování, kterou jsme popsali v teoretické části práce, je metoda vyvažování pomocí modelů teorie odpovědi na položku, neboli IRT vyvažování. Ta probíhá ve třech krocích. Nejprve odhadneme parametry položek a parametry respondentů pomocí vhodně zvoleného IRT modelu, v dalším kroku převedeme všechny parametry na stejnou škálu a na závěr transformujeme samotná celková skóre.

Celkově máme 731 respondentů, 371 jich vyplňovalo test X a 360 test Y. Oba testy X i Y obsahovaly 23 položek - 11 unikátních pro každý test a 12 společných, tzv. kotvících položek. Indexy $1, \dots, 11$ označíme unikátní položky testu X, indexy $12, \dots, 23$ společné kotvící položky a indexy $24, \dots, 34$ opět unikátní položky pro test Y. K dispozici tak máme 371 vektorů $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,23})^\top$, $i = 1, \dots, 371$ a 360 vektorů $\mathbf{Y}_i = (Y_{i,12}, \dots, Y_{i,34})^\top$, $i = 372, \dots, 731$. Proměnná X_{ij} , $i = 1, \dots, 371$, $j = 1, \dots, 23$ je binární a značí, zda i -tý respondent odpověděl na položku j v testu X správně ($X_{ij} = 1$), nebo nesprávně ($X_{ij} = 0$). Stejný význam mají i binární proměnné Y_{ij} , $i = 372, \dots, 731$, $j = 12, \dots, 34$, jen se jedná o respondenty a jejich odpovědi v testu Y.

K odhadnutí parametrů položek a parametrů respondentů jsme zvolili dvouparametrický 2PL IRT model, viz. rovnice (3.3). Uvažovali jsme také jednoparametrický (viz. rovnice (3.4)) a tříparametrický (viz. rovnice (3.2)) model, ale na základě informačních kritérií jsme za nejvhodnější zvolili právě 2PL IRT model. Všechny IRT modely, které používáme, jsou implementovány v balíku *mirt* (Chalmers, 2012). Ten k odhadu parametrů položek používá metodu marginální maximální věrohodnosti a k maximalizaci věrohodnosti využívá EM algoritmus. Navíc předpokládá, že parametry ability respondentů θ_i se řídí standardizovaným normálním rozdělením $N(0,1)$.

Metody odhadu vyvažovacích koeficientů

Když už máme odhady parametrů položek a_j a b_j (konkrétně máme odhady $\hat{a}_{X1}, \dots, \hat{a}_{X23}, \hat{a}_{Y12}, \dots, \hat{a}_{Y34}$ a $\hat{b}_{X1}, \dots, \hat{b}_{X23}, \hat{b}_{Y12}, \dots, \hat{b}_{Y34}$) a odhady ability respondentů θ_i ($\hat{\theta}_{X1}, \dots, \hat{\theta}_{X371}, \hat{\theta}_{Y372}, \dots, \hat{\theta}_{Y731}$), můžeme všechny parametry převést na jednu společnou škálu. Vztahy mezi jednotlivými parametry na obou škálách jsou uvedeny v rovnicích (3.5), (3.6) a (3.7). Jediné, co ještě potřebujeme znát, jsou vyvažovací koeficienty, resp. jejich odhady.

Jako první jsme si v teoretické části představili momentovou metodu odhadu

využívající průměr. V našem případě touto metodou získáme odhady

$$\hat{A} = \frac{\mu(\hat{a}_X)}{\mu(\hat{a}_Y)} = \frac{\frac{1}{12} \sum_{j=12}^{23} \hat{a}_{Xj}}{\frac{1}{12} \sum_{j=12}^{23} \hat{a}_{Yj}} \doteq \frac{1.026}{1.003} \doteq 1.024 \quad \text{a}$$

$$\begin{aligned} \hat{B} &= \mu(\hat{b}_Y) - \hat{A} \mu(\hat{b}_X) = \frac{1}{12} \sum_{j=12}^{23} \hat{b}_{Yj} - \hat{A} \frac{1}{12} \sum_{j=12}^{23} \hat{b}_{Xj} \\ &\doteq -0.225 + 1.024 \times 0.151 \doteq -0.101. \end{aligned}$$

Další metodou je momentová metoda, která využívá kromě průměru i směrodatnou odchylku parametrů $b_j, j \in \{12, \dots, 23\}$. Odhad vyvažovací konstanty A vychází z rovnice (3.16) a dostáváme

$$\begin{aligned} \hat{A} &= \frac{\sigma(\hat{b}_Y)}{\sigma(\hat{b}_X)} = \frac{\sqrt{\frac{1}{11} \sum_{j=12}^{23} (\hat{b}_{Yj} - \mu(\hat{b}_Y))^2}}{\sqrt{\frac{1}{11} \sum_{j=12}^{23} (\hat{b}_{Xj} - \mu(\hat{b}_X))^2}} \\ &\doteq \frac{1.024}{0.910} \doteq 1.125. \end{aligned}$$

Vyvažovací konstanta B má stejný tvar jako v předchozím případě a opět vychází z rovnice (3.12), nyní dostáváme

$$\hat{B} = \mu(\hat{b}_Y) - \hat{A} \mu(\hat{b}_X) \doteq -0.225 + 1.125 \times 0.151 \doteq -0.085.$$

Další metodou, kterou jsme uváděli v teoretické části práce, je Haebarova metoda. Ta se dívá na druhé mocniny rozdílů charakteristických křivek položek a za vyvažovací koeficienty A a B volí hodnoty, které tyto rozdíly přes všechny kotvící položky a ability respondentů minimalizují. Konkrétně získáme konstanty A, B minimalizací funkce uvedené v (3.20). V našem případě tak získáváme odhady $\hat{A} = 1.099$ a $\hat{B} = -0.155$.

Poslední zmíněnou metodou byla Stocking-Lordova metoda, která je obdobná jako Haebarova. Opět je založena na rozdílu charakteristických křivek položek, jen se v tomto případě díváme na druhou mocninu rozdílu součtu všech kotvících položek. Minimalizací přes všechny hodnoty ability, konkrétně tedy minimalizací funkce uvedené v (3.23), dostaneme odhady $\hat{A} = 1.058$ a $\hat{B} = -0.153$.

Metoda	\hat{A}	\hat{B}
momentová - průměr	1.024	- 0.101
momentová - průměr a sm. odchylka	1.125	- 0.085
Haebarová	1.099	- 0.155
Stocking-Lordova	1.058	- 0.153

Tabulka 4.7: Odhady vyvažovacích koeficientů \hat{A} a \hat{B} získané různými metodami.

celk. sk.	mom. m. - průměr		mom. m. - pr. a sm. odch.		Haebara		Stocking-Lord	
	skut. sk.	pozor. sk.	skut. sk.	pozor. sk.	skut. sk.	pozor. sk.	skut. sk.	pozor. sk.
0		0.40		0.63		0.64		0.56
1	1.12	1.56	1.45	1.88	1.43	1.92	1.28	1.76
2	2.32	2.73	2.81	3.17	2.81	3.25	2.59	3.03
3	3.60	3.94	4.16	4.44	4.20	4.55	3.95	4.31
4	4.90	5.15	5.47	5.62	5.56	5.75	5.30	5.55
5	6.18	6.33	6.70	6.74	6.84	6.91	6.60	6.72
6	7.40	7.47	7.85	7.81	8.03	8.01	7.81	7.83
7	8.54	8.54	8.91	8.82	9.12	9.05	8.94	8.90
8	9.61	9.57	9.89	9.79	10.13	10.04	9.98	9.91
9	10.61	10.55	10.81	10.70	11.07	10.97	10.96	10.88
10	11.56	11.50	11.67	11.58	11.95	11.87	11.88	11.80
11	12.46	12.41	12.49	12.42	12.79	12.72	12.75	12.69
12	13.33	13.29	13.28	13.24	13.59	13.55	13.58	13.55
13	14.17	14.15	14.04	14.04	14.36	14.36	14.38	14.38
14	14.99	15.01	14.80	14.83	15.12	15.16	15.17	15.20
15	15.80	15.86	15.55	15.63	15.88	15.95	15.95	16.02
16	16.63	16.71	16.32	16.43	16.64	16.75	16.74	16.83
17	17.46	17.55	17.11	17.25	17.42	17.55	17.53	17.65
18	18.31	18.42	17.93	18.09	18.22	18.37	18.35	18.48
19	19.18	19.30	18.79	18.96	19.05	19.22	19.19	19.34
20	20.07	20.19	19.69	19.85	19.91	20.10	20.05	20.21
21	20.97	21.10	20.62	20.77	20.80	20.99	20.93	21.11
22	21.88	22.03	21.62	21.71	21.74	21.91	21.84	22.02
23	23.00	22.98	23.00	22.69	23.00	22.87	23.00	22.96

Tabulka 4.8: Ekvivalenty celkových skóre z testu X na škále testu Y získané metodou vyvažování pomocí IRT modelů. Vyvažovací koeficienty byly odhadnuty momentovou metodou využívající průměr (mom. m - průměr), momentovou metodou využívající průměr a směrodatnou odchylku (mom. m - pr. a sm. odch.), Haebarovou a Stocking-Lordovou metodou. Ekvivalenty celkového skóre z testu X na škálu testu Y byly transformovány pomocí metody vyvažování celkových skóre i metody vyvažování pozorovaných skóre.

Všechny odhady vyvažovacích konstant získané uvedenými metodami najdeme pro snazší porovnání společně v Tabulce 4.7.

Dosažením odhadů vyvažovacích konstant \hat{A} a \hat{B} do rovnic (3.6) a (3.7) tak už můžeme převést všechny parametry položek na společnou škálu. Pomocí rovnice (3.5) bychom na stejnou škálu převedli i odhadnuté parametry θ_i a mohli je tak mezi sebou porovnat. V sekcích 3.3.1 a 3.3.2 jsme ale ještě představili dvě metody vyvažování pro celková dosažená skóre. Pro úplnost tak uvádíme i Tabulku 4.8, ve které vidíme ekvivalenty celkových skóre z testu X na škále testu Y získané metodou vyvažování skutečných skóre (sekce 3.3.1) i metodou vyvažování pozorovaných skóre (sekce 3.3.2). Ekvivalenty skóre jsou uvedeny za použití všech čtyř zmíněných metod pro odhadování vyvažovacích koeficientů pro transformaci parametrů položek, celkově tedy máme tabulku s 8 sloupci odhadnutých skóre.

Závěr

V této práci jsme se zabývali metodami pro vyvažování kognitivních i dalších testů. Vyvažování testů je statistický proces používaný k upravení a transformaci skóre z více verzí testu tak, aby bylo možné je férově porovnávat napříč skupinami. Proces vyvažování je nezbytnou součástí při vyhodnocování a interpretaci skóre z testů, a to především v případě, kdy lze očekávat, že se různé verze testu mohou lišit v obtížnosti.

Existují různé metody, kterými lze k vyvažování testů přistupovat. V první kapitole jsme představili tři tradiční metody, a to vyvažování ve střední hodnotě, lineární vyvažování a ekvipercentilové vyvažování. Metoda vyvažování ve střední hodnotě je velmi striktní a má silné předpoklady, proto se v praxi téměř nepoužívá a slouží spíše jen jako ilustrativní. Výrazně rozšířenější a už běžně používané metody jsou lineární a ekvipercentilové vyvažování.

Jako na jeden ze speciálních případů ekvipercentilového vyvažování se můžeme dívat na vyvažování pomocí jádrových odhadů. Této metodě jsme se detailně věnovali ve druhé kapitole práce. Zde jsme podrobně popsali všech pět dílčích kroků, a to pro dva různé designy sběru dat - pro design dvou ekvivalentních skupin a pro design dvou neekvivalentních skupin s kotvícími položkami. Zaměřili jsme se především na popis a odvození standardních chyb vyvažování v obou designech. Za pomoci delta věty jsme ukázali asymptotické rozdělení všech potřebných odhadů a na závěr i rozdělení samotného odhadu vyvažovací funkce. Pro oba výše zmíněné designy sběru dat jsme vyjádřili i konkrétní prvky jednotlivých matic, které potřebujeme k výsledné varianční matici odhadu vyvažovací funkce, a to podrobněji než je běžně uváděno v literatuře věnující se tématu vyvažování.

Ve třetí kapitole této práce jsme představili metodu vyvažování pomocí modelů teorie odpovědi na položku, tzv. IRT modelů. Tato metoda je v praxi také velmi rozšířená a spočívá především v transformaci parametrů položek na společnou škálu. To lze provést za pomoci vyvažovacích koeficientů, jejichž možné metody odhadu byly také představeny v této kapitole. Na závěr byly zmíněny i metody pro případné vyvažování celkových skóre testů.

Poslední čtvrtá kapitola byla věnována praktické části, ve které ilustrujeme použití všech výše zmíněných metod na reálném datovém souboru ve statistickém softwaru R. K dispozici jsme měli data z mezinárodního šetření TIMSS v roce 2015, které probíhalo v České republice u studentů 4. ročníků základních škol. Porovnávali jsme dvě různé verze znalostního testu z matematiky a aplikovali na ně vyvažovací metody popsané v teoretické části. Protože se ovšem jedná o velkou mezinárodní studii a při vytváření znění testů je kladen velký důraz na to, aby byly jednotlivé verze testu co nejsrovnatelnější, různé verze se zde v obtížnosti příliš neliší. Z toho důvodu na tomto konkrétním datovém souboru nevidíme výrazné rozdíly ani mezi uvedenými vyvažovacími metodami. Všechny jejich rozdíly, především ve zvolených a uvažovaných předpokladech, jsou však popsány v teoretické části práce.

Celkově si práce kladla za cíl představit problém vyvažování testů a jeho důležitost v procesu testování a vyhodnocování výsledků. Problém vyvažování testů jsme zde formulovali jako statistický problém, s jasně definovaným značením a matematickým zápisem. Častým problémem v literatuře, která se věnuje tématu vy-

važování testů, je právě jistá vágnost a nepřesnost ve formulaci problému a předpokladů. Dále jsme pak popsali většinu nejčastěji používaných metod, všechny jejich předpoklady a myšlenky, ze kterých vychází. Nejpodrobněji jsme se potom věnovali metodě vyvažování za pomoci jádrových odhadů, kterou považujeme za jednu z nejmodernějších a nejflexibilnějších metod k problému vyvažování testů. Zde jsme se zaměřili na podrobný popis jednotlivých kroků, především pak na odvození standardních chyb vyvažování pro dva nejběžněji používané designy sběru dat, design dvou ekvivalentních skupin a design dvou neekvivalentních skupin s kotvicími položkami. Výpočet standardních chyb vyvažování je pro všechny designy současně uveden v knize von Davier a kol. (2004), ale v této práci jsme pro rozšíření vyšli přímo z asymptotického rozdělení maximálně věrohodných odhadů, souhrnně odvodili asymptotické rozdělení všech dalších potřebných odhadů a ukázali, jak pomocí delta věty získat standardní chyby vyvažování. Pro dva výše zmíněné designy sběru dat jsme vyjádřili i konkrétní tvary jednotlivých variančních matic a jejich prvků. V praktické části pak byly ilustrovány všechny metody na reálných datech, byly porovnány a diskutovány jejich předpoklady.

Seznam použité literatury

- ALBANO, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, **74**(8), 1–36. doi: 10.18637/jss.v074.i08.
- ALTINTAS, O. a WALLIN, G. (2021). Equality of admission tests using kernel equating under the non-equivalent groups with covariates design. *International Journal of Assessment Tools in Education*, page 729–743. doi: 10.21449/ijate.976660.
- ANDERSSON, B., BRÄNBERG, K. a WIBERG, M. (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software*, **55**(6), 1–25. URL <https://doi.org/10.18637/jss.v055.i06>.
- ANDERSSON, B. a VON DAVIER, A. A. (2014). Improving the bandwidth selection in kernel equating. *Journal of Educational Measurement*, **51**(3), 223–238. doi: 10.1111/jedm.12044.
- BATTAUZ, M. (2015). equateIRT: An R package for irt test equating. *Journal of Statistical Software*, **68**(7), 1–22. doi: 10.18637/jss.v068.i07.
- CHALMERS, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, **48**(6), 1–29. doi: 10.18637/jss.v048.i06.
- DAVIER, A. A. (2011). *Statistical models for test equating, scaling, and linking*. Springer New York.
- DORANS, N. J. a HOLLAND, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, **37**(4), 281–306. doi: 10.1111/j.1745-3984.2000.tb01088.x.
- GONZÁLEZ, J. a WIBERG, M. (2017). *Applying test equating methods: Using R*. Springer.
- HAEBARA, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, **22**(3), 144–149. doi: 10.4992/psycholres1954.22.144.
- HOLLAND, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, **55**(4), 577–601. doi: 10.1007/bf02294609.
- HOLLAND, P. W. a THAYER, D. T. (1987). Notes on the use of log-linear models for fitting discrete probability distributions. *ETS Research Report Series*, **1987**(2), i–40. doi: 10.1002/j.2330-8516.1987.tb00235.x.
- HOLLAND, P. W. a THAYER, D. T. (1989). The kernel method of equating score distributions. *ETS Research Report Series*, **1989**(1), i–45. ISSN 23308516. doi: 10.1002/j.2330-8516.1989.tb00333.x. URL <https://onlinelibrary.wiley.com/doi/10.1002/j.2330-8516.1989.tb00333.x>.

- HOLLAND, P. W. a THAYER, D. T. (2000). Univariate and bivariate log-linear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, **25**(2), 133–183. doi: 10.2307/1165330.
- HOLLAND, P. W., KING, B. F. a THAYER, D. T. (1989). The standard error of equating for the kernel method of equating score distributions. *ETS Research Report Series*, **1989**(1), i–52. doi: 10.1002/j.2330-8516.1989.tb00332.x.
- HÄGGSTRÖM, J. a WIBERG, M. (2014). Optimal bandwidth selection in observed-score kernel equating. *Journal of Educational Measurement*, **51**(2), 201–211. doi: 10.1111/jedm.12042.
- KOLEN, M. J. a BRENNAN, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. New York, third edition.
- LIANG, T. a VON DAVIER, A. A. (2014). Cross-validation. *Applied Psychological Measurement*, **38**(4), 281–295. doi: 10.1177/0146621613518094.
- R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- VAN DER VAART, A. W. (2000). *Asymptotic statistics*. Cambridge Univ. Press.
- VON DAVIER, A. A., HOLLAND, P. W. a THAYER, D. T. (2004). *The Kernel Method of Test Equating*. Springer New York, New York, NY. ISBN 9780387019857. doi: 10.1007/b97446. URL <http://link.springer.com/10.1007/b97446>.
- WIBERG, M. a BRÄNBERG, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement*, **39**(5), 349–361. doi: 10.1177/0146621614567939.