

Oponentský posudek na diplomovou práci:

M. VAŘEJKOVÁ: MODELÝ PRO VYVAŽOVÁNÍ KOGNITIVNÍCH TESTŮ

Předložená práce se zabývá takzvaným vyvažováním testů, tj. vlastně přepočítáváním výsledku testu X na škálu testu Y. Uchazečka srozumitelně a podrobně popisuje vybrané metody a v závěru je používá na skutečná data.

Úvodní kapitola stručně uvádí nejjednodušší tradiční vyvažovací metody. Druhá kapitola představuje vyvažování založené na jádrovém vyhlazení pravděpodobností získaných pomocí log-lineárních modelů (zhruba podle knihy von Davier, Holland & Thayer (2004)), zajímavý je přitom design neekvivalentních skupin s kotvícími položkami popisovaný v sekci 2.2. Uchazečka přitom u všech metod podrobně rozepisuje odvození asymptotického rozdělení odhadů vyvažovací funkce pomocí opakovaného použití delta metody (podobně jako v práci Holland & Thayer (1987)).

Následující třetí kapitola stručněji popisuje vyvažovací metody založené na IRT modelu, podobně jako například v knize Kolen & Brennan (2014). Vlastním přínosem je čtvrtá kapitola, kde jsou popisované metody ilustrované na skutečných datech.

Celá práce je zpracovaná pečlivě a nenarazil jsem na žádné překlepy, oceňuji i vzornou grafickou úpravu (až snad na několik tučněji vtištěných řádků na stranách 50, 54, 58, 60 a 61)

K obsahu mám několik obecných dotazů:

- Z celé práce mi není jasné, kdy a za jakým účelem by se vlastně mělo vyvažování testů používat. Můžete uvést nějaký příklad, kdy je vyvážení testů nezbytné?
- Proč je nutné výsledky testu X přepočítávat na test Y, místo aby se oba testy transformovaly na společnou škálu, například do intervalu  $(0, 1)$ ?
- Jsou, vzhledem k nejistotě při odhadování vyvažovací funkce, přepočítané výsledky z testu X přímo srovnatelné s výsledky v testu Y?
- Uvažovala jste i o jiných možných odhadech pravděpodobnostních rozdělení  $r$  a  $s$ ?
- Mohou při použití některých metod přepočítané výsledky testu X padnout mimo rozsah možných výsledků testu Y? Jak by se dal tento problém vyřešit? Jaká je interpretace hodnot v posledním řádku tabulky 4.5?
- Využila jste ve čtvrté kapitole existující knihovny `equate`, `kequate` a `equateIRT` nebo jste si některé metody naprogramovala sama?

Další dotazy:

- str. 9 dole**  $J - 1$  maximální dosažené nebo maximální dosažitelné skóre? Je mezi těmito dvěma pojmy nějaký rozdíl?
- str. 11 dole** Při odvození logaritmické věrohodnosti předpokládáte rozdělení multinomické (předpoklad na str. 10) nebo Poissonovo (zmíněné za vzorcem (2.9))?
- str. 22** Vzorec (2.41) uvádí asymptotický rozptyl  $\hat{\varphi}_K$  v bodě  $x$ . Bylo by možné podobně vypočítat i kovarianci  $\hat{\varphi}_K(x_1)$  a  $\hat{\varphi}_K(x_2)$ ?
- str. 39** Není vzorec (3.5) v rozporu s dříve uvedeným předpokladem  $\theta \sim N(0, 1)$ ?
- str. 40** Jak plyne (3.11) z (3.9) a (3.10)?
- str. 40** Jak plyne (3.16) z (3.15)?
- str. 58** K jaké populaci se vztahují odhady nakreslené v obrázku 4.6? Jak výsledek závisí na parametru  $\omega$ ?
- str. 59** Čím je způsobená nečekaně vysoká predikce pro nulovou hodnotu kotevního testu v obrázku 4.7(b)?

Celkově lze předloženou práci charakterizovat jako podrobnou rešerši na zadané téma, kterou díky pečlivému a přehlednému zpracování doporučuji uznat jako práci diplomovou.

Doc. RNDr. Zdeněk Hlávka, Ph.D.