НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Фізико-математичний факультет

Кафедра математичного аналізу та теорії ймовірності

«На правах рукопису»

УДК 517.237

До захисту допущено:

Завідувач кафедри

Олег Клесов

«___»____________20__ р.

Магістерська дисертація

на здобуття ступеня магістра

за освітньо-професійною програмою «Страхова та фінансова математика»

зі спеціальності 111 «Математика»

на тему: «Моделювання репрезентативних вибірок для MIRT моделей»

Виконав:
студент ІІ курсу, групи ОМ-21мп
Лесик Денис Віталійович

____

Науковий керівник:
кандидат фізико-математичних наук,
доцент
Круглова Наталія Володимирівна

____

Рецензент:
Доцент кафедри інформаційних
систем та технологій
кандидат технічних наук
НТУУ «КПІ»

Богданова Наталія Володимирівна ___________

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.
Студент ______________

Київ – 2024 року

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-математичний факультет

Кафедра математичного аналізу та теорії ймовірності

Рівень вищої освіти – другий (магістерський) за освітньо-науковою програмою

Спеціальність – 111 «Математика»

Освітньо-наукова програма «Страхова та фінансова математика»

ЗАТВЕРДЖУЮ

Завідувач кафедри

Олег Клесов

Завдання

на магістерську дисертацію студенту

Лесик Денис Віталійович

1. Тема дисертації «Моделювання репрезентативних вибірок для MIRT моделей», науковий керівник дисертації кандидат фізико-математичних наук, доцент Круглова Наталія Володимирівна, затверджені наказом по університету від « 13 » листопада 2023 р. № 5250-с
2. Термін подання студентом дисертації 11.01.2023
3. Об’єктом дослідження є моделі MIRT для статистичного аналізу тестів.
4. Предметом дослідження є вибір розмірності моделі MIRT.
5. Перелік завдань, які потрібно виконати:

   1) Ознайомитися з літературою та дослідити основні означення та поняття.

2) Змоделювати вибірки матриць первинних балів для компенсаторних двовимірних і трьохвимірних 2PL моделей.
3) Опанувати основні моделі та методи MIRT.
4) Дослідити методи EFA вибору розмірності моделі: PA, EKC, HULL.
5) Обрати алгоритми оцінювання латентних параметрів моделей.
6) Дослідити методи перевірки адекватності моделей MIRT.
7) Дослідити програмну реалізацію відібраних алгоритмів та методів у середовищі R.
8) На підставі відібраних алгоритмів і програм провести статистичний аналіз для змодельованих репрезентативних вибірок матриць первинних балів для компенсаторних 2PL MIRT-моделей.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу: 19 слайдів.
7. Дата видачі завдання 07.09.2023.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Ознайомитися з літературою та дослідити основні означення та поняття.	07.09.2023 - 21.09.2023	виконано
2.	Опанувати основні моделі та методи MIRT.	22.09.2023 - 13.10.2023	виконано
3.	Згенерувати репрезентативні вибірки вибірки матриць первинних балів для компенсаторних двовимірних і трьохвимірних 2PL моделей	14.10.2023- 04.11.2023	виконано
4.	Дослідити методи EFA вибору розмірності моделі: PA, EKC, HULL.	14.10.2023 - 04.11.2023	виконано

5.	Обрати алгоритми оцінювання латентних параметрів моделей.	05.11.2023 - 25.11.2023	виконано
6.	Дослідити методи перевірки адекватності моделей MIRT.	26.11.2023 - 16.12.2023	виконано
7.	Дослідити програмну реалізацію відібраних алгоритмів та методів у середовищі R.	17.12.2023 - 02.01.2024	виконано
8.	На підставі відібраних алгоритмів і програм провести статистичний аналіз змодельованих репрезентативних вибірок матриць первинних балів для компенсаторних 2PL MIRT-моделей.	02.01.2024- 08.01.2024	виконано

Студент Лесик Денис

Науковий керівник Наталія Володимирівна

Реферат

Магістерська дисертація: 52 сторінки, 24 першоджерела, 19 слайдів презентації, електронні додатки. Робота складається зі вступу, п'яти розділів, висновків та списку використаної літератури.

В дисертаційній роботі досліджується алгоритми генерації матриць первинних балів MIRT моделей, вибір розмірності моделі MIRT для аналізу згенерованих репрезентативних вибірок матриць первинних балів для компенсаторних 2PL MIRT-моделей. Основною метою дисертаційного дослідження є вибір розмірності моделі MIRT для аналізу згенерованих репрезентативних вибірок матриць первинних балів. Об'єктом дослідження є моделі MIRT для аналізу згенерованих репрезентативних вибірок матриць первинних балів. Предметом дослідження є алгоритми генерації матриць первинних балів, вибір розмірності моделі MIRT.

Перший розділ містить теоретичні відомості з основ статистичного аналізу педагогічних тестів.

Другий розділ містить математичні методи EFA попереднього визначення розмірності моделей MIRT.

Третій розділ містить методи оцінювання латентних параметрів моделей MIRT, які використовуються в роботі.

Четвертий розділ містить методи перевірки адекватності моделі.

П'ятий розділ містить статистичний аналіз результатів згенерованих репрезентативних вибірок матриць первинних балів для компенсаторних 2PL MIRT-моделей, опис параметрів згенерованих матриць первинних балів, порівняльний аналіз еталонних параметрів моделей із оціненими.

Ключові слова: тести з вищої математики, MIRT, EFA, PA, EKC, HULL, алгоритми EM, MH-RM, критерії AIC, BIC, M2, RMSEA, CFI ,TLI, 2PL MIRT-модель.

Abstract

Master's Thesis: 53 pages, 24 primary sources, 19 presentation slides. The work consists of an introduction, five chapters, conclusions, and a list of references.

This dissertation investigates the algorithms for generating primary score matrices of MIRT models, the choice of the MIRT model dimension for analyzing the generated representative samples of primary score matrices for compensatory 2PL MIRT models. The main purpose of the dissertation research is to select the dimension of the MIRT model for analyzing the generated representative samples of primary score matrices. The object of research is MIRT models for analyzing generated representative samples of primary score matrices. The subject of the study is the algorithms for generating matrices of primary scores, the choice of the dimension of the MIRT model.

The first chapter provides theoretical background on the fundamentals of statistical analysis of educational tests.

The second chapter presents the mathematical methods of EFA for preliminary dimensionality estimation of MIRT models.

The third chapter describes the methods of estimating latent parameters in MIRT models used in the research.

The fourth chapter discusses the methods for assessing the adequacy of the model.

The fifth section contains a statistical analysis of the results of the generated representative samples of primary score matrices for compensatory 2PL MIRT models, a description of the parameters of the generated primary score matrices, and a comparative analysis of the reference parameters of the models with the estimated ones.

Keywords: higher mathematics tests, MIRT, EFA, PA, EKC, HULL, EM algorithms, MH-RM, AIC, BIC, M2 criteria, RMSEA, CFI, TLI, 2PL MIRT-model.

Зміст

Вступ

Актуальність магістерської дисертації зумовлена необхідністю створення набору адекватних репрезентативних матриць первинних балів іспитників. На таких вибірках викладачі можуть вчитися правильно визначати кількість компетентностей, закладених у тест; вірно підбирати MIRT модель; моделювати сценарій проведення тестування для конкретної групи студентів; прогнозувати результати ЗНО і НМТ. У MIRT аналізі дуже важливо підібрати правильну модель, щоб більш тонко вирізняти окремі риси та компетентності студентів, допрацьовувати «погані» тестові питання.

Метою роботи є створення репрезентативних вибірок матриць первинних балів для компенсаторних 2PL MIRT-моделей та проведення на них EFA, CFA, MIRT аналізу.

Моделювання результатів тестування з використанням заданих параметрів двовимірної компенсаторної моделі може бути корисним з кількох причин:

1) Моделювання дозволяє вам змінювати параметри тесту, такі як складність завдань, кількість завдань, диференціюючу здатність, кількість компетентностей. Ви можете експериментувати з різними конфігураціями тесту та оцінити їх вплив на ефективність та точність визначення компетентності студентів.

2) Моделювання дозволяє вам провести чутливість аналізу, щоб визначити, як різні параметри впливають на результати тесту. Ви можете визначити, які параметри мають найбільший вплив на точність та надійність вимірювань компетентності. Для яких рівней підготовленості студентів виникає найбільша похибка.

3) Моделювання може служити як інструмент для підготовки до реальних тестів, дозволяючи вам передбачити та визначити, як різні умови впливають на відповіді студентів. Чи треба додати у тест завдання меншої складності, чи, навпаки, посилити тест. Ви можете моделювати різні ситуації та вивчити їх вплив на результати тестування.

4) Моделювання дозволяє вам тестувати нові ідеї, стратегії оцінювання та розробляти нові тестові завдання перед їх впровадженням у реальне

тестування. Ви можете експериментувати з різними форматами та типами питань для вдосконалення якості тесту.

5) Моделювання дозволяє вам краще розуміти властивості обраної двовимірної компенсаторної моделі. Ви можете вивчити, як різні параметри взаємодіють між собою та як вони впливають на результати тестування.

І. Огляд літератури.

1. Теоретичні основи IRT

Класична теорія тестів (КТТ) [8] та Сучасна Теорія Тестів (IRT) [9] є основою сучасного статистичного аналізу тестів. Основна ідея IRT полягає у створенні певного зв'язку між ймовірністю правильної відповіді на запитання тесту та основними факторами, що описують як саме запитання, так і учасників тестування.

Цим методикам характерно, що:

• Параметри підготовки іспитників стабільні та незмінні.

• Стабільність та незмінність оцінок параметрів складності та розрізняльної здатності завдань.

• Встановлення стандартизованих шкал для кількісної оцінки всіх основних змінних.

Основні моделі IRT

IRT базується на логістичних моделях. Найбільш поширеною одновимірною дихотомічною моделлю є модель 4-PL [10].

Нехай тест складається з K завдань, які мають лише дві можливі відповіді. Його виконує N осіб. Ймовірність того, що p-й учасник тестування дасть правильну відповідь на q-те питання, визначається рівнянням:

$$P\left(Y_{pq} = 1 | \tau_p, \beta_q, \alpha_q, \delta_q, \sigma_q\right) = \delta_q + \frac{\sigma_q - \delta_p}{1 + e^{-\beta_p(\tau_p - \alpha_q)}},$$

- τ_p- показник підготовленості p -того іспитника;
- α_q - показник складності завдання ;
- $\beta_q > 0$ - дискримінативність тесту;
- δ_q- найменший параметр, який визначає асимптотичну поведінку характеристичної функції;

- σ_q- максимальний параметр, що визначає асимптотичну поведінку характеристичної функції q -того завдання.

Модель 3-PL є окремим випадком моделі 4-PL, де $\sigma_q = 1$. У цій моделі параметр δ_q відповідає ймовірності правильного вгадування відповіді в q -тому завданні.

$$P(Y_{qp} = 1|\tau_p, \beta_q, \alpha_p, \delta_q) = \delta_q + (1 - \delta_q)\frac{e^{\beta_q(\tau_p - \alpha_q)}}{1 + e^{\beta_q(\tau_p - \alpha_q)}}$$

У модель 2-PL включено два параметри завдання: β_q, α_q, тоді як $\delta_q = 1$і $\sigma_q = 0$ для всіх завдань.

$$P(Y_{qp} = 1|\tau_p, \beta_q, \alpha_q) = \frac{e^{\beta_q(\tau_p - \alpha_q)}}{1+e^{\beta_q(\tau_p - \alpha_q)}}.$$

2.Теоретичні відомості з основ MIRT

Однією з фундаментальних передумов застосування IRT є те, що всі завдання тесту оцінюють одну й ту саму компетентність в межах певного набору здібностей. Іноді компоненти тесту можуть вимірювати різні здібності або окремі частини набору навичок чи компетентностей. Для прикладу, математичне оцінювання з алгебри може включати дві категорії запитань: знання теорем алгебри і вміння виконувати математичні операції. Перші вимагають від іспитників засвоєння теоретичного матеріалу, тоді як другі – вміння оперувати математичними символами та мати відповідну математичну техніку.

Згідно з Траубом (Traub, 1981) [11], якщо взяти до уваги всі здібності, необхідні для успішного виконання всіх завдань більшості когнітивних тестів, то більш імовірно, що існує не один вимір, а декілька. Теорія багатовимірних відповідей на завдання (MIRT) [12] була створена для визначення різних аспектів здібностей, що оцінюються за допомогою відповідних завдань, зокрема для таких ситуацій.

Метою IRT є пояснення зв'язку між особливостями завдання та базовими здібностями індивіда в тестуванні за допомогою моделей, що базуються на ймовірності. Зв'язок між групою учасників тестування та одним запитанням завжди можна пояснити за допомогою одновимірного підходу, оскільки запитання може оцінювати лише одну навичку або комбінацію пов'язаних навичок. Однак дуже важливо ретельно перевіряти припущення про одновимірність, оцінюючи взаємозв'язок між групою осіб, які проходять тестування, та різними завданнями. Якщо тест здатен оцінити кілька навичок, але учасники тестування відрізняються лише за рівнем володіння однією з них, взаємодію можна описати як таку, що має один вимір. Аналогічно, у ситуації, коли тест із завданнями оцінює або лише одну навичку, або комбінацію кількох навичок, а учасники тестування відрізняються за рівнем володіння кожною з цих навичок, взаємозв'язок також можна пояснити в одному вимірі. Тим не менш, якщо іспит має здатність розрізняти різні рівні декількох навичок, і якщо учасники тестування відрізняються за своїм досвідом у більш ніж одній з цих навичок, то необхідно пояснити кореляцію, використовуючи багатовимірну термінологію.

Важливо не припускати автоматично одновимірність, а послідовно перевіряти її існування. Практики можуть використовувати різні теоретично обґрунтовані методи для перевірки одномірності відповідей. Серед досліджень, проведених у цій галузі, можна назвати факторний аналіз МакДональда [13], який зосереджується на нелінійності, а також підхід умовних асоціацій Розенбаума та Голланда [14]. Нандакумар і Стаут [15] запропонували процедуру статистичного тестування для визначення відсутності значущої одновимірності, яку можна розглядати як альтернативний метод. Тим не менш, ці методи пропонують статистичну перевірку для визначення того, чи є певна підгрупа запитань у тесті одномірною порівняно з іншими запитаннями. Однак вони не дають кількісної оцінки загальної розмірності всього тесту.

Моделі MIRT

MIRT (Multidimensional Item Response Theory) - теорія багатовимірних відповідей на завдання) - це статистична концепція, яка використовується для визначення декількох основних змінних (також відомих як фактори) на основі відповідей респондентів на набір тестових завдань. Кожне завдання в MIRT характеризується унікальними параметрами, які вказують на його властивості, включаючи рівень складності, можливість відповісти "не знаю" та здатність до дискримінації.

Різниця між класичною моделлю IRT та MIRT полягає в тому, що MIRT дозволяє моделювати зв'язок між латентними змінними та завданнями тесту. У той час як модель IRT представляє відповіді як зв'язок між відповідями та однією латентною змінною, MIRT пропонує можливість представити відповіді як зв'язок між відповідями та декількома латентними змінними. Це дає можливість ретельніше дослідити зв'язок між кожним елементом оцінювання та кожним базовим фактором, що може бути корисним для різних дисциплін, включаючи освіту, психологію та інші суміжні галузі.

MIRT використовує багатовимірні параметри для пояснення як здібностей, так і параметрів предметів. Це означає, що кожна прихована характеристика та окремий параметр предмета пояснюється за допомогою серії чисел, де кожне число представляє певний аспект риси або параметра предмета.

Зазвичай MIRT використовують двовимірне представлення, яке передбачає опис навичок особи за допомогою двох параметрів (наприклад, здібностей до математики та мови). Аналогічно, кожен параметр запитання також описується за допомогою двох параметрів, що вказують на рівень його складності та здатність розрізняти високі та низькі результати.

Використання багатовимірних параметрів дає змогу точніше відобразити взаємозв'язок між здібностями та параметрами запитання, що призводить до покращеного оцінювання успішності учнів або студентів. Проте включення кількох змінних у процедуру оцінювання та аналізу даних ускладнює завдання.

Існує дві основні моделі MIRT, які використовуються для пояснення даних про відповіді на дихотомічні завдання: компенсаторні та некомпенсаторні. Модель MC2PL, розроблена Reckase у 1985 році [12], дозволяє обчислити ймовірність отримання правильної відповіді на питання, враховуючи декілька вимірів.

Рівень підготовленності іспитника в багатовимірній моделі визначається шляхом утворення лінійної комбінації компонент k - вимірного вектору $\overline{\tau_p} = (\tau_{p1}, \tau_{p2}, \ldots, \tau_{pk})$. Припустимо що в тесті K завдань, їх виконали N студентів.

$$P\left(Y_{qp} = 1 | \bar{\tau}_p, \bar{\beta}_q, b_q\right) = \frac{e^{\bar{\beta}_q \bar{\tau}_p^{'} + b_q}}{1 + e^{\bar{\beta}_q \bar{\tau}_p^{'} + b_q}} \quad q = \overline{1, K}, p = \overline{1, N}.$$

Ми можемо розписати експоненту в даній формулі для того, щоб показати різні частини векторів β і τ.

$$\bar{\beta}_q \bar{\tau}_p' + b_q = \beta_{q1}\tau_{p1} + \beta_{q2}\tau_{p2} + \cdots + \beta_{qsk}\tau_{pk} + b_q = \sum_{w=1}^{k} \beta_{qw}\tau_{pw} + b_q$$

В даній моделі елементи вектору τ представлені лінійною функцією exp. Параметр b представляє переріз, а елементи вектору β представляють параметри нахилу. Показник вказує на визначення лінії у просторі з m вимірами. Коли показник степеня зафіксовано на постійному значенні n , всі вектори $\bar{\tau}$, що задовольняють вираз $n = \bar{\beta}_q \bar{\tau}_p' + b_q$, лежать на одній прямій. Зауважимо що вектори матимуть одну й ту ж саму ймовірність вірної відповіді в моделі.

Ми можемо записати формулу для розрахунку ймовірності точної відповіді, використовуючи двопараметричну логістичну некомпенсовану модель (MNC2-PL) [16], наступним чином:

$$P\left(Y_{qp} = 1|\bar{\tau}_p, \bar{\beta}_q, \bar{\alpha}_q\right) = \prod_{w=1}^{v} \frac{e^{\bar{\beta}_{qw}(\bar{\tau}_{pw} - \bar{\alpha}_{qw})}}{1 + e^{\bar{\beta}_{qw}(\bar{\tau}_{pw} - \bar{\alpha}_{qw})}}$$

У цьому твердженні Y_{qp} - характеристика (0,1), присвоєна питанню q,

$\bar{\beta}_{qw}$ = (β_{q1}, β_{q2}, ..., β_{qv}) - це набір параметрів дискримінації, $\bar{\alpha}_{qw}$= (α_{q1}, α_{q2}, ..., α_{qv}) - представляє ряд параметрів складності для питання q, а $\bar{\tau}_{pw}$ = (τ_{p1}, τ_{p2}, ..., τ_{pv}) - представляє ряд параметрів здібностей. Для кожного виміру ця модель включає параметри як дискримінації, так і складності. Важливо розуміти, що ця модель може бути виражена як результат об'єднання двох або більше окремих одновимірних моделей, зокрема, двопараметричної логістичної моделі [2-PL] IRT. Зокрема, коли ми маємо справу з двома вимірами, некомпенсаційна модель може бути виражена в іншій формі як:

$$P\left(Y_{qp} = 1|\bar{\tau}_p, \bar{\beta}_q, \bar{\alpha}_q\right) = \frac{e^{\beta_{q1}(\tau_{p1} - \alpha_{q1})}}{1+e^{\beta_{q1}(\tau_{p1} - \alpha_{q1})}} \frac{e^{\beta_{q2}(\tau_{p2} - \alpha_{q2})}}{1+e^{\beta_{q2}(\tau_{p2} - \alpha_{q2})}}$$

Різні компоненти моделі не дозволяють екзаменатору компенсувати брак навичок в одній сфері за рахунок виняткових результатів в іншій.

Модель M3-PL розширює модель M2-PL і враховує існування нижньої асимптоти, відмінної від нуля. Ця модель розширює трипараметричну логістичну модель UIRT шляхом включення декількох змінних. Математична формула моделі M3-PL:

$$P\left(Y_{qp} = 1|\bar{\tau}_p, \bar{\beta}_q, \delta_q, b_q\right) = \delta_q + \left(1 - \delta_q\right) \frac{e^{\bar{\beta}_q \bar{\tau}_p' + b_q}}{1+e^{\bar{\beta}_q \bar{\tau}_p' + b_q}} \qquad (1)$$

Модель M3-PL була створена з метою уточнення емпіричних даних, які спостерігаються, як показано в дослідженні Лорда (1980), і які вказують на те, що особи з низькими здібностями все ж мають шанс надати правильну відповідь у багатовимірних завданнях. Модель включає параметр, який називається нижня асимптота або параметр псевдовгадування δ_{i}, який визначає ймовірність того, що особи з дуже низькими значеннями τ виберуть правильну відповідь. Цей параметр включено, оскільки виявляється, що процес вибору правильної відповіді особами з низькими здібностями не пов'язаний з успішністю, яка вимірюється тестовим завданням.

У випадку багатовимірних моделей рівень складності q-го завдання можна виміряти як відстань від точки $\bar{\tau}$ до точки, яка збігається з найгострішим нахилом поверхні. Знак на індикаторі показує положення точки відносно початку координат. Значення розраховується за наступною формулою:

$$\hat{B}_q = \frac{-b_q}{\sqrt{\sum_{n=1}^{k} \beta_{qn}^2}}$$

Багатовимірна диференціююча здатність - це швидкість зміни характеристичної поверхні в точці з найкрутішим нахилом, а саме в напрямку точки $\bar{\tau}$. Її значення можна обчислити за наступною формулою:

$$\hat{A}_q = \sqrt{\sum_{n=1}^{k} \beta_{qn}^2}$$

II. Математичні методи оцінки кількості вимірів у MIRT-моделях

1. Алгоритми EFA, що використовуються для визначення розмірності моделі

Дослідницький факторний аналіз (EFA) і підтверджуючий факторний аналіз (CFA) [8] є основними методами, що застосовуються в MIRT для визначення основних факторів.

Використання дослідницького факторного аналізу (ДФА) в багатофакторних моделях (MIRT) допомагає виявити основні фактори у відповідях на складні тести або запитання, які оцінюють різні виміри здібностей або характеристик.

EFA - це загальновживаний статистичний метод для дослідження глибинної структури численних спостережуваних змінних, особливо коли немає очевидного теоретичного підґрунтя для конкретної моделі. EFA використовує метод, який спирається на аналіз даних для виявлення прихованих закономірностей, але припускаючи при цьому наявність спільної факторної моделі. У цій концепції кожна змінна, що спостерігається, розглядається як комбінація зважених сум групи факторних змінних, які можуть бути корельовані між собою, а також одного визначального фактору. Спільні елементи відповідають за взаємозв'язки між спостережуваними змінними і є вирішальними факторами, що мають теоретичне значення. Аспекти, що різняться, пояснюють лише відмінності в окремих спостережуваних змінних, які розглядаються як неточності вимірювання у порівнянні зі спільними факторами.

Визначення кількості факторів, що лежать в основі, є ключовим питанням у сфері дослідницького факторного аналізу (ОДФА) [17]. Недооцінка або переоцінка кількості факторів може негативно вплинути на якість ОФА. Помилки в усіх факторних навантаженнях виникають, коли неправильно оцінюється кількість факторів, незалежно від їхньої значущості в точно визначеній моделі.

Згідно з Jöreskog (2007) [18], загальна факторна модель передбачає існування V прихованих загальних факторів $\tau_1,\dots,\tau_V$, які пояснюють мінливість K спостережуваних випадкових величин $y_1,\dots,y_K$, які були стандартизовані. Припускається, що кожна спостережувана змінна, представлена y_q, є результатом комбінації факторів $\theta_1,\dots,\theta_V$ та відмінного фактора ε_q, як і в лінійній регресії.

$$y_q = \beta_{q1}\tau_1 + \beta_{q2}\tau_2 + \ldots + \beta_{qV}\theta_V + \varepsilon_q, 1 \leq q \leq K$$

В рамках загальної факторної моделі фактори ε_q залишаються некорельованими з усіма іншими факторами $\tau_1, \ldots, \tau_V$ а також з будь-яким $\varepsilon_{q'}$ де q не дорівнює q'. Коефіцієнт β_{qp} означає вплив q -го спостережуваної змінної на фактор p. Отже, мета полягає в тому, щоб визначити спільні базові фактори, яких менше, ніж змінні, які визначають кореляції між спостережуваними змінними $y_1, \ldots, y_K$ для того, щоб $y_1, \ldots, y_K$ перестали бути корельованими при введенні латентних факторів $\tau_1, \ldots, \tau_V$.

Існує багато методів знаходження кількості факторів в EFA, більша частина яких базуються на власних значеннях, що представляють величину дисперсії, яка пояснюється кожним спільним фактором. Серед доступних методів є емпіричний критерій Кайзера, метод Халла та паралельний аналіз (PA) [17]. У зв'язку з тим, що моделювання структурних рівнянь включає в себе загальну факторну модель, щоб перевірити точність обраної кількості факторів використовують показники адекватності моделі.

Підтверджувальний факторний аналіз (CFA) [8] широко визнаний як остаточний підхід до оцінювання підтверджувальної моделі. Тоді як FA ілюструє кореляцію між спостережуваними змінними і складеними факторами, підтверджуючий факторний аналіз (CFA) демонструє кореляцію між спостережуваними змінними і базовими факторами. У сфері багаторівневих ієрархічних моделей MIRT, CFA використовують щоб виявити взаємодії між підшкалами (різними рівнями вимірювання), і декількома латентними факторами.

У підході MIRT ми представляємо кожен показник вимірювання як функцію декількох латентних факторів, які представляють різні аспекти здатності або характеристики. Включення кореляцій між латентними факторами, що використовуються для опису різних вимірів, і різні ваги, присвоєні індикаторам, щоб відобразити їхній різний внесок у латентний фактор, дозволені в CFA в рамках MIRT.

Для оцінки параметрів MIRT з більшою впевненістю можна використовувати байєсівський метод оцінки або метод максимальної правдоподібності (ML) [10]. Більшість досліджень зазвичай покладаються на метод максимальної правдоподібності, однак метод Байєса може бути кращим у ситуаціях, коли необхідно врахувати невизначеність параметрів або коли дослідження базується на невеликому обсязі вибірки.

Метод ML [10] широко використовується для оцінювання параметрів у CFA і користується великою популярністю. Суть цього полягає в тому, що статистичний процес підвищує ймовірність того, що побудована нами модель точно відображає спостережувані дані.

При проведенні факторного аналізу з використанням методу максимальної правдоподібності на першому етапі ми будуємо модель, яка включає як латентні спільні фактори, так і їхні спостережувані індикатори. Ми прагнемо підвищити точність моделі, використовуючи статистичну процедуру максимальної правдоподібності для визначення значень параметрів моделі, таких як спільні фактори та їх навантаження. Такий підхід підвищує ймовірність того, що наша модель ефективно пояснює спостережувані дані.

Простіше кажучи, метод максимальної правдоподібності дозволяє нам визначити значення параметрів моделі, які найкраще відповідають нашим спостереженням. Водночас, ми маємо можливість використовувати різні стандарти для оцінки достатності моделі, включаючи індекси доброякісності, які дозволяють нам оцінити, наскільки добре модель узгоджується зі спостереженнями.

ML метод є загальновизнаним і високоефективним в CFA, хоча на нього можуть впливати попущені дані та значні викиди, що, в свою чергу, впливає на його точність і надійність. В такому разі можна використовувати альтернативні підходи, наприклад, метод найменших квадратів (МНК) [10].

Підвищити рівень довіри до CFA можна за допомогою байєсівських методів оцінювання [10], які дозволяють оцінювати як параметри моделі, так і інформативні попередні розподіли, пов'язані з цими параметрами. Ці методи застосовують байєсівське правило, щоб переглянути оцінки параметрів моделі, на відміну від вживаних підходів до оцінювання, таких як метод максимальної правдоподібності (ML) або ж метод моментів (MM), з метою врахування впливу нової інформації, що призводить до підвищення рівня довіри до оцінок.

Методи для встановлення розмірності моделі у MIRT

У дослідницькому факторному аналізі ("EFA") переважна кількість критеріїв, які використовуються щоб знайти кількість факторів, залежать від власних чисел [1]. Для того, щоб краще зрозуміти зв'язок між загальною факторною моделлю і власними числом, можна представити останню у вигляді матриці. Для матриці навантажень розмірністю $K \times V$ позначеної як Λ, та $\bar{Y} = (y_1, \ldots, y_K)', \bar{\tau} = (\tau_1, \ldots, \tau_V)', \bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_K)'$,

$$\bar{Y} = \Lambda\bar{\tau} + \bar{\varepsilon} \tag{1}$$

Одним із способів представлення кореляційної матриці явних змінних Y є використання наступного виразу:

$$C = E(\bar{Y}\bar{Y}') \tag{2}$$

де E - математичне сподівання, за умови, що спостережувані змінні були стандартизовані. Згідно з рівнянням 1,

$$\bar{Y}\bar{Y}' = (\Lambda\bar{\tau} + \bar{\varepsilon})(\Lambda\bar{\tau} + \bar{\varepsilon})' \tag{3}$$

$$= \Lambda\bar{\tau}'\Lambda' + \Lambda\bar{\tau}\bar{\varepsilon}' + \bar{\varepsilon}\bar{\tau}'\Lambda' + \bar{\varepsilon}\bar{\varepsilon}' \tag{4}$$

Застосуємо таке позначення $\Phi = E(\bar{\tau}\bar{\tau}')$ для представлення кореляційної матриці між спільними факторами $\bar{\tau}$, тоді як $\Delta = E(\bar{\varepsilon}\bar{\varepsilon}')$ використовується для коваріаційної матриці унікальних факторів $\bar{\varepsilon}$. Враховуючи, що $\bar{\tau}$ та $\bar{\varepsilon}$ незалежні, модель представляє кореляційну матрицю наступним чином:

$$C = \Lambda \Phi \Lambda' + \Delta \tag{5}$$

Причиною того, що в рівнянні 5, Δ - діагональна матриця , є те, що модель із загальними факторами передбачає незалежність усіх окремих факторів ε_q, $\varepsilon_{q'}$, для $q \neq q'$. Крім того, ці фактори становлять частину дисперсії y_q, , яку можна віднести до латентних факторів. Модель із загальними факторами точно обчислює матрицю Λ за допомогою обертання.

$$\tilde{C}_B \approx \Lambda \Lambda' \tag{6}$$

Після заміни унікальних значень, матриця $\tilde{C}_B$ виводиться з рівняння 6, в яке підставляються діагональні елементи С. Одним із способів підвищення достовірності є визначення навантажень Λ, які пропорційні власним векторам $\tilde{C}_B$ [18], шляхом застосування МНК для (6). Власними векторами є вектори, які мають такі характеристики:

$$A\sigma = \lambda\sigma, \sigma \neq 0 \tag{7}$$

У випадку симетричних матриць, які є додатньо визначеними, таких як R_B матриці або ж матриці, які є коваріаційними, існує рівно K невід'ємних власних значень, які не обов'язково можуть бути різними. Важливо, що дисперсію, яка припадає на p-й фактор, можна визначити, дослідивши p-те найбільше власне значення матриці $\tilde{C}_B$.

PCA часто використовується як замінник EFA при проведенні факторного аналізу, спираючись на спільну факторну модель. Тим не менш, PCA - це метод що не враховує відмінну варіацію, тим самим скорочує кількість даних. У психологічних дослідженнях EFA зазвичай застосовується, коли на меті є спроба виявлення основних закономірностей, які відображають взаємозв'язки між змінними з урахуванням випадкової похибки. Цей підхід вважається більш практичним у таких сценаріях. Основна відмінність полягає в обчисленні власних значень PCA з використанням кореляційної матриці C замість C_B. Тим не менш, кореляційна матриця C та її відповідні власні значення повністю визначаються гіпотетичною загальною факторною моделлю. Існують докази того, що в практичних сценаріях аналіз головних компонент (PCA) та дослідницький факторний аналіз (EFA) можуть давати подібні результати.

Незалежно від того, які критерії використовують щоб визначити кількість факторів, існують певні обставини, які можуть завдання визначення правильної кількості факторів зробити тяжчим, або ж навпаки, легшим. Особливо, коли насиченість факторами є низькою (наприклад, через низьке факторне навантаження, невелику кількість показників у факторі або високу кореляцію між факторами), визначення точної кількості факторів стає більш складним завданням, оскільки власні значення, пов'язані з фактичними факторами, є дуже схожими на решту власних значень [1].

Враховуючи, що власні значення можуть бути отримані із загальної факторної моделі (як показано в рівняннях 5 і 6), ми маємо можливість безпосередньо оцінити вплив різних факторних моделей на очікувані власні значення. Власні значення можна обчислити з кореляційної матриці C гіпотетичної загальної факторної моделі з V факторами, K показниками на фактор, стандартизованими факторними навантаженнями l та міжфакторною кореляцією ρ.

$$\lambda_1 = 1 + (K-1)w^2 + (V-1)K\rho w^2$$

$$\lambda_2 = \ldots = \lambda_V = 1 + (K-1)w^2 - K\rho w^2 \tag{8}$$

$$\lambda_{V+1} = \ldots = \lambda_{KV} = 1 - w^2$$

Аналогічно, ми можемо обчислити власні значення C_B [1], як

$$\lambda_1 = Kw^2 + (V-1)K\rho w^2$$

$$\lambda_2 = \ldots = \lambda_V = Kw^2 - K\rho w^2 \quad (9)$$

$$\lambda_{V+1} = \ldots = \lambda_{KV} = 0$$

Власні значення, отримані за допомогою вибірки, відрізнятимуться від власних значень генеральної сукупності, зазначених у рівняннях 8 і 9, що означає зменшення визначеності. Брекен та ван Ассен [1] продемонстрували, що мінливість вибірки впливає на власні значення трьома різними способами. По-перше, використання випадкових кореляцій у кореляційній матриці вибірки під час вилучення факторів призводить до збільшення власних значень $\lambda_1, \ldots, \lambda_{[V/2]}$ (які представляють першу половину власних значень, пов'язаних з дійсними факторами). Крім того, власні значення $\lambda_{[V/2]}, \ldots, \lambda_V$, пов'язані з істинними факторами, поступово зменшуються, оскільки первинні фактори відповідають за більшу частину відмінностей. Крім того, початкова частина решти власних значень $(\lambda_{V+1}, \ldots, \lambda_{[V+(K-V)/2]})$ знову зростає. Подібно до власних значень $\lambda_1, \ldots, \lambda_{[V/2]}$, додаткові фактори також включають решту випадкових кореляцій, які не були враховані попередніми факторами, тим самим підвищуючи рівень довіри. Як наслідок, набір власних значень стає менш визначеним, оскільки розрив між λ_V та λ_{V+1} зменшується.

Загалом, це демонструє ефективність методів екстракції, які використовують власні значення вибірки, в ситуаціях, коли є багато показників зі значними навантаженнями та обмеженими кореляціями між факторами. Серйозні проблеми виникають, особливо при обмеженому обсязі вибірки, коли навантаження показників є низькими, на один фактор припадає мало показників, а кореляції між факторами є сильними.

Підходи до визначення кількості факторів.

EFA зазвичай застосовують у ситуаціях, коли існує невизначеність щодо кількості основних латентних факторів, які пояснюють спостережувані змінні, а отже, підвищується рівень достовірності. У цьому розділі ми надамо стислий огляд традиційних і сучасних підходів, що використовуються для визначення належної кількості факторів, які слід залишити в дослідницькому факторному аналізі (EFA).

2. Паралельний аналіз

В 1965 році Хорн [2] представив техніку паралельного аналізу (PA) [2], яка допомагає визначити відповідну кількість факторів, які слід залишити. Цей метод ґрунтується на генеруванні випадкових величин. PA порівнює власні значення в кореляційній матриці, що аналізується, з власними значеннями, отриманими з нормальних некорельованих змінних. Для того, щоб отримати "очікувані" власні значення, використовується процедура моделювання за методом Монте-Карло для відтворення вихідних даних з аналогічним розміром вибірки та кількістю змінних. Метод, що використовується у факторному аналізі, відомий як PA, дотримується подібної процедури, з тією лише різницею, що замість діагоналі кореляційної матриці використовуються квадрати множинних кореляцій. Таке коригування робиться як початковий крок в апроксимації загальних змінних в EFA, тим самим зберігаючи цілісність і суть вихідного процесу. У минулому важливість фактора оцінювалася шляхом порівняння його власного значення із середнім власним значенням даних, які не мали кореляції. Зараз доцільно використовувати власне значення, пов'язане з певним процентилем, наприклад, 95-м процентилем, при розгляді розподілу власних значень, отриманих на основі випадкових даних [4].

Численні дослідження підтверджують, що метод PA підходить для визначення кількості факторів [19]. Дослідження, проведене Цвіком і Веллісером (1986) [3], показало, що PA є найнадійнішим методом порівняно з іншими розглянутими методами, демонструючи нижчий рівень варіабельності і менше піддаючись впливу різних чинників. Глорфельд (1995) [4] погоджується з цією оцінкою і стверджує, що існує мінімальне обґрунтування для вибору будь-якого іншого методу, окрім PA, при оцінці ефективності різних методів.

Формулювання алгоритму паралельного аналізу для факторного аналізу можна математично виразити наступним чином:

1. Отримати вибірку даних, що складається з K спостережень і V змінних (або факторів).

2. На основі розрахунків по вибірці даних створити кореляційну матрицю C (KxK).

3. Згенерувати n вибірок даних, які відповідають нормальному розподілу, кожна з яких містить K спостережень та V змінних. Повторити процес, застосований для вихідної вибірки, для кожної з цих вибірок, в результаті чого отримуємо k кореляційних матриць: C_1, C_2, ..., C_n.

4. Визначити власні значення кожної кореляційної матриці C_q .

5. Отримати кількість критичних власних значень із середнього значення n вибірок даних.

3. Емпіричний критерій Кайзера

Критерій Кайзера враховує випадкові коливання власних чисел при застосуванні емпіричного критерію Кайзера (EKC) [1]. На рівні окремих точок даних критерій Кайзера розглядає розподіл власних чисел для даних, які є нормально розподіленими. У нульовій моделі розподіл власних чисел наближається до розподілу Марченка-Пастура [20], коли n наближається до нескінченності. Максимальне значення цього розподілу відповідає першому власному числу λ, яке визначається формулою

$$\lambda_1 = (1 + \sqrt{\frac{V}{K}})^2$$

Для підвищення достовірності власні числа наступних факторів у наборі даних з K спостережень і V елементів коригуються на основі частки дисперсії, що пояснюється власними числами попередніх факторів. Визначення p-го власного числа подано у вигляді

$$\lambda_p = max(\frac{V - \sum_{q-0}^{p-1} \lambda_p}{V - p + 1} [1 + \sqrt{\frac{V}{K}}]^2, 1)$$

В результаті, коли попередні власні числа вищі, еталонне власне число зменшується, оскільки частка неврахованої дисперсії стає меншою. Згідно з початковим критерієм Кайзера, важливо, щоб еталонне власне число було більшим або дорівнювало одиниці.

Умови для точного визначення кількості факторів при ЕКС визначені для всіх значень $1 \leq p \leq V$, а також для загальної кількості n факторів . Умови для ортогональних факторів є відносно простими і передбачають високі значення альфа-коефіцієнта Кронбаха α_p та велику кількість спостережень K. Крім того, це використання коротких шкал та низька кореляція між факторами. Коли мова йде про корельовані фактори, вимоги ускладнюються, але є більше шансів задовольнити їх, якщо α та K високі , шкали короткі, а кореляції між факторами мінімальні. Згідно з дослідженням, проведеним Braeken та van Assen (2017) [1], було підтверджено, що ЕКС демонструє високий рівень точності (> 0,90) за сприятливих обставин, але точність значно падає (< 0,50), якщо умови не виконуються. Зокрема, коли фактори корелюють і оцінюються з використанням лише обмеженої кількості елементів з сильним навантаженням, ефективність ЕКС була вищою, ніж у традиційного PA, особливо коли в якості міри використовувався 95-й процентиль. Імітаційні дослідження зі значною кількістю факторів і множинними спостережуваними змінними також підтвердили, що ЕКС забезпечує подібні результати до реконструйованих PA і CD, підтверджуючи думку про надійність ЕКС.

Згідно з дослідженням, проведеним Braeken та van Assen (2017) [1], ефективність ЕКС виявилася досить високою, коли всі точності перевищували 0,93, але значно знижувалася, коли ці умови були протилежними (нижче 0,83). Тим не менш, для того, щоб ЕКС та інші критерії вилучення демонстрували виняткову ефективність, важливо мати доступ до конкретної інформації про структуру факторів, яка доступна лише дослідникам. Це може вважатися недоречним в рамках EFA, оскільки він зазвичай використовується для того, щоб запобігти припущенням про приховану структуру факторів.

Алгоритм емпіричного критерію Кайзера можна зобразити таким способом:

1. Визначте кореляційну матрицю стовпчиків.

2. Знайдіть власні числа цієї матриці.

3. Представте всі ці числа у порядку спадання.

4. Знайдіть кількість власних чисел, більших за 1 (зокрема, власних чисел, які перевищують середнє значення всіх власних чисел).

5. Це рекомендована кількість факторів.

4. Hull Method

Критерій HULL - це критерій, який використовується для визначення кількості факторів у факторному аналізі за допомогою методу максимальної правдоподібності (MLE) або аналізу головних компонент (PCA). Він допомагає підвищити достовірність такого визначення.

Критерій HULL ґрунтується на порівнянні фактичної кореляційної матриці між факторами з матрицею часткових кореляцій між цими факторами. Необхідно обчислити дві матриці: матрицю, що відображає кореляцію між факторами, яка спостерігається в реальних даних, і матрицю, що показує кореляцію між факторами при контролі за іншими змінними. Для того, щоб визначити кількість факторів на основі критерію HULL, важливо порівняти власні числа спостережуваної матриці з власними числами матриці часткової кореляції.

Розраховують середнє власне число (AVE) як для емпіричної, так і для часткової кореляційної матриці, і потім порівнюють їх. Кількість факторів визначається, коли підраховується кількість власних чисел, які перевищують середнє власне число матриці часткової кореляції, у емпіричній матриці.

Даний критерій дає змогу визначити найбільш підходящу кількість факторів, що допомагає запобігти перевантаженню моделі та отримати більш точні результати факторного аналізу.

Критерій HULL має значну перевагу, оскільки він ґрунтується на порівнянні емпіричної кореляційної матриці з частковою кореляційною матрицею. Цей підхід враховує зв'язки між факторами і дозволяє визначити точну кількість незалежних факторів у моделі, тим самим підвищуючи ефективність.

Однак, важливо зазначити, що HULL - критерій не є ідеальним і може не завжди надавати точну кількість факторів. Тому слід рекомендувати використання його разом з іншими факторами та проведення додаткових досліджень, щоб забезпечити відповідність та надійність отриманих результатів від факторного аналізу.

Наведені нижче формули можуть виразити алгоритм, який використовується для критерію HULL в EFA:

1. По-перше, ми повинні обчислити кореляційну матрицю для змінних у вихідному наборі даних.

2. Далі кореляційна матриця перетворюється на фактори, в результаті чого утворюється матриця факторів. Власні числа цієї матриці факторів представлені як λ.

3. Для отримання довільних даних створюється набір даних з рівною кількістю змінних і спостережень. Випадкові дані можна генерувати як з рівномірного, так і з нормального розподілу.

4. Обробляючи довільні дані, обчислюється кореляційна матриця, яка потім розбивається на фактори, в результаті чого з'являється факторна матриця, отримана з випадкових даних. Випадкова факторна матриця представлена власними числами λ*.

5. Виконайте крок 4 кілька разів, наприклад, 100 разів, щоб обчислити середнє значення власних чисел випадкової матриці. M*λ - позначення, яке використовується для представлення середнього значення власних чисел випадково згенерованої матриці факторів.

6. Наступною порівнюється саме значення λ з відповідним середнім значенням випадкових значень Mλ. Якщо значення λ більше, ніж значення Mλ, то цей фактор вважається важливим, і на його місце обирається новий фактор. Кількість факторів збільшується, доки λ не стане меншим за M*λ.

7. Максимальна кількість факторів, яка вважається значущою, визначається шляхом порівняння λ з середнім значенням випадкових власних чисел M*λ.

ІІІ. Підходи до визначення параметрів моделі

1. ЕМ алгоритм (Expectation-Maximization)

Бок та Ейткін (1981) [5] представили методику для підвищення надійності оцінювання параметрів завдань у численних тестах. Їхній метод ґрунтується на алгоритмі максимізації (ЕМ). Цей алгоритм можна розділити на два ключові етапи.

На етапі обчислюються теоретичні розподіли відповідей на тестові запитання для кожного учасника, враховуючи поточні оцінки параметрів моделі. Потім ці теоретичні розподіли використовуються для визначення очікуваних значень кількості відповідей на кожне екзаменаційне запитання.

На етапі максимізації алгоритм використовує очікувані значення, отримані на попередньому кроці, для покращення параметрів моделі з метою оптимізації функції правдоподібності даних. Метод максимальної правдоподібності використовується для виконання цього кроку з максимальною точністю.

2. Алгоритм Метрополіса-Гастінгса-Роббінса-Монро (MH-RM).

Алгоритм MH-RM, також відомий як Metropolis-Hastings Robbins-Monroe [10]- це метод, який використовується для отримання більш надійних оцінок параметрів моделі в сценарії, що включає тестування з множинним вибором (MIRT), шляхом включення розподілу початкових значень.

MH-RM алгоритм поєднує в собі методи ітеративної оптимізації та Монте-Карло, що підвищує його ефективність. Він спирається на випадково вибраний набір параметрів, згенерований з попереднього розподілу, у поєднанні з оберненим перетворенням Монте-Карло, щоб підвищити точність оцінок параметрів моделі. Алгоритм максимізує ймовірність того, що оцінки параметрів досягнуть своїх оптимальних значень. Унікальною особливістю алгоритму є його здатність ефективно обробляти складні MIRT-моделі з великою кількістю параметрів.

Алгоритм MH-RM виявився корисним і надійним методом для отримання оцінок параметрів у моделі з попереднім розподілом, коли він використовується разом з MIRT. Він дозволяє отримати оцінки параметрів, які максимізують вірогідність спостережуваних даних, а також узгоджуються з оцінками, отриманими за допомогою інших методологій MIRT.

На відміну від традиційного методу EM, який розглядає параметр τ як набір "значущих" параметрів із заздалегідь визначеним розподілом і поєднує його з рівнянням правдоподібності, алгоритм MH-RM використовує стохастичні методи для заповнення відсутніх параметрів τ. Після того, як відсутні значення підставляються, вони вважаються встановленими даними. Потім ці дані беруться до уваги при коригуванні параметрів на рівні елементів за допомогою звичайних методів пошуку коренів, які використовують повну функцію логарифмічної правдоподібності. Цей метод підбору, хоча і не є точним або остаточним, зазвичай забезпечує більш прямий і практичний спосіб оцінки багатовимірних інтегралів, ніж метод числової квадратури. MH-RM - це нещодавня розробка, спрямована на усунення невідповідностей, які виникають при використанні стохастичних методів для заповнення τ-параметрів.

IV. Підходи до оцінки достатності моделі

Для вибору найбільш підходящої моделі використовуються спеціальні статистичні критерії, оскільки алгоритми EFA здатні визначати різні виміри моделі. Щоб оцінити, наскільки добре модель узгоджується зі спостережуваними даними в MIRT, ми використовуємо кілька критеріїв відповідності моделі, таких як RMSEA, TLI, AIC, BIC та CFI. Використання декількох критеріїв відповідності моделі може сприяти більш об'єктивній оцінці того, чи є модель MIRT адекватною.

TLI критерій

Критерій TLI (Tucker-Lewis Index) [21] використовується в CFA як міра для оцінки відповідності моделі факторного аналізу порівняно з вихідною моделлю нульової кореляції. Він обчислює різницю між матрицею спостережуваних даних і матрицею, створеною за допомогою моделі нульової кореляції, і порівнює її з різницею між матрицею спостережуваних даних і матрицею, створеною за допомогою досліджуваної моделі факторного аналізу, з метою визначення відстані між ними. Чим більше значення TLI наближається до 1, тим сильніше модель узгоджується з даними.

Для того, щоб визначити TLI, важливо встановити зв'язок між підігнаною моделлю та базовою моделлю, яка встановлює випадкові зв'язки між змінними.

Діапазон значень TLI становить від 0 до 1, причому вищі значення вказують на сильнішу відповідність моделі. До аналізу оцінки адекватності моделі за допомогою індексу TLI можна підійти наступним чином. Якщо значення TLI близьке до 1, це свідчить про задовільну відповідність моделі. Якщо значення TLI нижче 0,9, доцільно внести зміни до моделі, щоб підвищити її точність.

TLI можна описати як частку залишкової коваріації між спостережуваними змінними в моделі, порівняно із залишковою коваріацією в базовій моделі, що допомагає оцінити достовірність.

Статистика TLI критерію можна представити так:

$$TLI = \frac{\frac{Y^2}{n} - dg}{\frac{Y^2}{n} - dg + \frac{Y_0^2}{n_0} - dg_0},$$

де:

• Y^2- статистика χ^2 - квадрат тесту для перевірки адекватності моделі;

• dg - ступені свободи моделі (кількість) ;

• n - параметри моделі (кількість);

• Y_0^2- статистика x_p-квадрат тесту для базової моделі;

• dg_0 - для базової моделі;

• k_0 - в базовій моделі.

CFI критерій

Використання критерію порівняльної придатності (Comparative Fit Index, CFI) [21] в MIRT слугує вирішальним показником при оцінці придатності вимірювальної моделі. Він оцінює рівень узгодженості між фактичними спостережуваними даними та прогнозованими даними, отриманими за допомогою моделі.

Чим ближче до 1 значення CFI, тим кращою є відповідність моделі. Загалом, значення CFI, що перевищує 0,95, зазвичай вважається надійною ознакою того, що модель добре підігнана.

CFI вимірює збільшення відповідності моделі, порівнюючи її з базовою моделлю, яка не враховує жодних зв'язків між змінними. Високе значення CFI свідчить про те, що модель відповідає даним більш ефективно, ніж нульова модель.

Індекс відповідності моделі CFI в MIRT визначається математично шляхом порівняння досліджуваної моделі з нульовою моделлю, яка не враховує взаємозв'язки між змінними. Мета полягає в тому, щоб впевнено оцінити відповідність моделі.

Щоб обчислити CFI потрібно врахувати декілька величин:

1. χ^2 значення для моделі, яку ми досліджуємо.
2. χ^2 значення для нульової моделі.

CFI статистику можемо визначити так:

$$CFI = \frac{(Y_0^2 - Y_{модель}^2)}{Y_0^2},$$

де:

- $Y_{\text{модель}}^2$- χ^2 значення для моделі, яку ми досліджуємо.
- Y_0^2- χ^2 значення для нульової моделі.

RMSEA критерій

У MIRT для оцінки адекватності моделі факторного аналізу, що включає змінні різних типів, використовується показник RMSEA (середньоквадратична помилка апроксимації) [22. Він враховує точність того, наскільки добре модель відповідає зібраним даним, зокрема, визнаючи і враховуючи розбіжності (варіації) в даних, які не враховані моделлю.

Розрахунок індексу RMSEA ґрунтується на поділі даних на різні елементи, які складаються з апроксимації спостережуваними даними (модель) та елементів, які безпосередньо не спостерігаються (неспостережувані). Чим точніше модель відповідає даним, тим нижчий індекс RMSEA. Зазвичай, якщо значення RMSEA нижче 0,05, вважається, що модель добре відповідає даним, тоді як значення в діапазоні від 0,05 до 0,08 свідчить про помірну відповідність.

Цей показник широко використовується для оцінки адекватності моделі в MIRT. Він дозволяє оцінити якість підгонки моделі, враховуючи розбіжність між даними та апроксимацією моделі.

Індекс перевірки адекватності RMSEA моделі в MIRT розраховується шляхом порівняння непоясненої дисперсії даних із загальною дисперсією даних.

Статистику RMSEA можна визначити наступною формулою:

$$RMSEA = \sqrt{\frac{Y^2 - dg}{K(dg - 1)}}$$

де

- Y^2 - χ^2 статистика, що використовується для оцінки розбіжності між моделлю та даними,
- dg - рівні гнучкості, які вказують на складність моделі,
- K - розмір вибірки.

AIC критерій

Індекс AIC (Інформаційний критерій Акайке) [23] використовується для оцінки адекватності моделі. Цей індекс ґрунтується на концепції балансування між точністю моделі та рівнем її складності.

Зазвичай модель, яка має найнижче значення AIC, вважається найоптимальнішою, оскільки вона пропонує найкращий баланс між точністю та складністю моделі.

Слід підкреслити, що для точного використання індексу AIC в MIRT дуже важливо порівнювати моделі, які використовують однаковий набір даних та оцінені параметри. Це пов'язано з тим, що значення AIC може відрізнятися залежно від цих факторів.

Розрахунок статистики AIC здійснюється математично за наступною формулою:

$$AIC = -2\,ln(w) + 2n$$

Де

- $ln(w)$ - ступінь відповідності моделі даним показує логарифм функції максимальної правдоподібності.
- n - загальна кількість параметрів моделі відображає загальну кількість параметрів, які оцінюються в моделі.

ВІС критерій

ВІС (Байєсівський інформаційний критерій) [24] - статистична міра в MIRT, яка оцінює, наскільки добре модель відповідає даним. Цей підхід ґрунтується на концепціях байєсівського підходу і враховує точність узгодження моделі з даними та складність моделі.

Тестову статистику для оцінки адекватності моделі ВІС можна визначити так:

$$AIC = -2\ln(w) + n\,ln(K)$$

де:

- $ln(w)$ - Логарифм функції максимальної правдоподібності вказує на рівень точності, з якою модель відповідає даним;
- n - кількість параметрів;
- K - кількість спостережень (обсяг вибірки).

Як правило, менше значення ВІС свідчить про те, що модель краще відповідає даним.

M2 критерій

Алгоритм M2 [10] оцінює дві функції похибок, а саме: для базової моделі та для альтернативної моделі, для того, щоб зробити порівняння. У цьому конкретному сценарії під базовою моделлю мається на увазі початкова модель, яка була використана для розрахунку параметрів. Альтернативна модель може стосуватися моделі з іншими параметрами або іншої моделі, яка краще пояснює тестові дані порівняно з початковою моделлю.

Для того, щоб оцінити дві моделі, алгоритм M2 використовує статистику, яка визначається шляхом аналізу розбіжностей між їхніми функціями похибок, а також кількості ступенів свободи, які має кожна з них. Якщо значення тестової статистики перевищує критичне значення, то альтернативна модель є кращою за базову.

Нижче наведено статистику, яка використовується в алгоритмі перевірк и адекватності M2 в рамках MIRT:

$$I = -2\sum_{p=1}^{P}\sum_{n=1}^{N_p}\sum_{q=1}^{K_p} l_{pnq}\, log\frac{j_{pnq}}{\hat{j}_{pnq}} + \sum_{p=1}^{P}\sum_{n=1}^{N_p} log\,\frac{\hat{j}_{pnq}}{1-\hat{j}_{pnq}}$$

Де

l_{pnq}- ваговий коефіцієнт для відповіді n учасника q на питання p;

j_{pnq}- теоретична ймовірність відповіді n учасника q на питання p згідно з альтернативною моделлю;

$\hat{j}_{pnq}$- теоретична ймовірність відповіді n учасника q на питання p згідно з базовою моделлю;

p - загальна кількість питань;

N_p- загальна кількість варіантів відповіді на запитання p;

K_p- загальна кількість учасників, що відповіли на запитання p;

V. Генерація матриць відповідей та їх аналіз

Зазвичай, в реальних тестуваннях важко отримати репрезентативні матриці відповідей іспитників. Це залежить від багатьох факторів, серед яких:

1) проблема списування під час проходження тесту;
2) викладачі працюють з обмеженою кількістю студентів (група, потік).
3) тестування не може проводитися більше 2 годин.

1. Для коректного використання MIRT моделей тест має включати не менше 30 питань, а кількість іспитників має бути не менше 500. Розробникам тестів потрібно вміти правильно підбирати модель, що описує дані тестування, оскільки коректний аналіз тестування неможливо провести, не визначивши розмірність моделі і її вигляд [6]. Ось чому наявність бази репрезентативних матриць відповідей іспитників є дуже важливою.

2. В роботі розглядались двовимірна і трьохвимірна компенсаторна 2PL модель, оскільки в [7] показано, що такі моделі найчастіше можна використати для опису результатів педагогічних тестувань.

Було згенеровано по 1000 матриць для двовимірної компенсаторної 2PL моделі для 500 іспитників з кількістю питань – 30 і для трьохвимірної. Рівні підготовленості по кожній компетенції були стандартними гауссівськими величинами, компоненти векторів диференціюючої здатності і складностей завдань – рівномірно розподілені. Всі згенеровані вектори параметрів питань фіксувались і використовувалися для кожної згенерованої матриці. Тобто, на вхід функції генерації матриці відповідей 1000 разів подавались ті самі параметри.

На рис. 1 і 2 елементи згенерованих матриць відповідей для двовимірної компенсаторної 2PL моделі і трьохвимірної відповідно. З прикладами цих матриць можна ознайомитись в електронному додатку до магістерської роботи.

1	Item_1	Item_2	Item_3	Item_4	Item_5	Item_6	Item_7
2	1	1	0	1	1	0	1
3	1	1	1	0	0	0	1
4	1	0	0	0	1	0	1
5	0	1	0	0	0	1	0
6	1	0	0	0	0	1	0
7	0	0	0	0	0	0	0
8	1	0	1	1	1	0	1
9	1	1	0	0	1	0	1
10	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1
12	0	0	0	1	0	1	0
13	1	0	1	1	1	0	0
14	1	1	0	1	1	0	1
15	0	0	1	0	1	1	0
16	1	1	1	0	1	0	0

Рис. 1

1	Item_1	Item_2	Item_3	Item_4	Item_5	Item_6
2	1	0	0	0	0	0
3	0	0	1	0	0	1
4	0	0	0	0	0	0
5	1	1	1	1	1	1
6	0	0	0	0	0	0
7	1	1	1	1	1	1
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	1	1	1	1
11	1	1	1	1	1	0
12	1	1	1	1	1	1
13	0	0	0	0	1	1
14	1	1	1	1	1	1
15	0	0	0	0	0	0

Рис. 2

До всіх матриць відповідей застосовувалась наступна схема аналізу результатів тестування:

1) Визначення розмірності моделі за допомогою емпіричного критерію Кайзера, критерію Хала і паралельного аналізу.
2) Обчислювались оцінки параметрів моделей. Вираховувалось середнє для кожного параметру.

3) За допомогою M2, RMSEA, SRMSR,TLI,CFI критеріїв перевірено, що модель адекватна даним тестування.

В результаті первинного аналізу матриць первинних балів було виявлено, що всі критерії показали однакову розмірність моделі, яка співпадала з розмірністю моделі, що подавалась на функцію моделювання матриці.

В табл. 1 і 2 результати застосування критерію Кайзера для двовимірної та трьохвимірної моделі відповідно.

Табл. 1

№ питання	Власне число
1	11.973
2	2.887
3	0.988
4	0.968
5	0.843
6	0.805
7	0.751
8	0.733
9	0.695
10	0.662
11	0.640
12	0.600
13	0.577
14	0.548
15	0.540
16	0.511
17	0.496
18	0.485
19	0.465
20	0.454
21	0.414
22	0.400
23	0.384
24	0.360
25	0.348
26	0.339
27	0.317
28	0.305
29	0.262
30	0.249

Табл. 2

№ питання	Власне число
1	15.628
2	2.588
3	1.934
4	0.822
5	0.637
6	0.621
7	0.57
8	0.503
9	0.492
10	0.456
11	0.443
12	0.433
13	0.396
14	0.377
15	0.359
16	0.348
17	0.344
18	0.329
19	0.313
20	0.301
21	0.288
22	0.249
23	0.243
24	0.233
25	0.223
26	0203
27	0.194
28	0.174
29	0.173
30	0.139

В таблицях 3 і 4 результати паралельного аналізу для цих моделей.

Двовимірна.

Root	Real Data	Mean	Percentil
1	11.973	1.487	1.555
2	2.887	1.416	1.459
3	0.988	1.368	1.410
4	0.968	1.324	1.368
5	0.843	1.289	1.323
6	0.805	1.252	1.278
7	0.751	1.218	1.248

8	0.733	1.186	1.217
9	0.695	1.157	1.182
10	0.662	1.129	1.153
11	0.640	1.100	1.122
12	0.600	1.074	1.101
13	0.577	1.047	1.069
14	0.548	1.018	1.040
15	0.540	0.994	1.015
16	0.511	0.969	0.989
17	0.496	0.943	0.963
18	0.485	0.918	0.942
19	0.465	0.892	0.911
20	0.454	0.871	0.895
21	0.414	0.847	0.866
22	0.400	0.823	0.841
23	0.384	0.801	0.823
24	0.360	0.777	0.800
25	0.348	0.751	0.776
26	0.339	0.727	0.747
27	0.317	0.700	0.723
28	0.305	0.674	0.698
29	0.262	0.644	0.674
30	0.249	0.604	0.640

Трьохвимірна:

Root	Real Data	Mean	Percentile
1	15.628	1.474	1.528
2	2.588	1.412	1.459
3	1.934	1.365	1.406
4	0.822	1.324	1.369
5	0.637	1.287	1.320
6	0.621	1.252	1.279
7	0.570	1.216	1.248
8	0.503	1.187	1.214
9	0.492	1.158	1.180
10	0.456	1.129	1.158
11	0.443	1.101	1.127
12	0.433	1.072	1.095
13	0.396	1.047	1.072
14	0.377	1.020	1.044
15	0.359	0.995	1.020
16	0.348	0.971	0.995
17	0.334	0.946	0.971
18	0.329	0.923	0.945
19	0.313	0.898	0.918
20	0.301	0.873	0.897
21	0.288	0.849	0.872
22	0.249	0.826	0.848
23	0.243	0.800	0.823
24	0.233	0.776	0.796
25	0.223	0.753	0.779
26	0.203	0.726	0.751
27	0.194	0.700	0.724
28	0.173	0.673	0.701
29	0.172	0.643	0.671
30	0.139	0.602	0.634

В таблиці 5 наведено розмірність, яку було визначено кожним критерієм.

Таблиця 5

Критерій	Двовимірна 2PL	Трьохвимірна 2PL
Hull	2	3
EMPKC	2	3
RAWPAR	2	3

Як ми бачимо, усі критерії показали однакову розмірність моделі. Ця розмірність співпадає з тими даними, що подавались на функцію генерації матриць первинних балів. Це є суттєвою відмінністю від застосування критеріїв до реальних даних. Зазвичай, ці критерії можуть давати різні значення, і тоді вибираються результати більш сильного критерію.

Після визначення розмірності моделі потрібно обчислити оцінки параметрів моделі. Для цього застосовувалась функція mirt (пакет mirt) з вбудованим алгоритмом розв'язання систем нелінійних рівнянь щодо параметрів моделі.

В таблицях 6 і 7 наведено параметри, що подавались на функцію генерації, і обчислені середні оцінки параметрів.

Табл.6. Двовимірна модель.

a1	a2	d	estimation a1	estimation a2	estimation d
1,70286902	0,05813399	0,63254928	1,82	0,101	0,786
1,87584171	1,53699354	0,36210218	1,632	1,144	0,973
0,08961821	1,1071419	-0,19233926	0,321	1,105	-0,206
0,40858678	1,15163701	-0,51160677	0,532	1,299	-0,698
1,63824562	1,2914201	-0,04308555	1,873	1,312	-0,18
0,04701143	0,90426956	-0,29216498	0,772	0,998	-0,409
1,96965476	0,4417091	-0,61117834	1,009	0,52	-0,72
1,32424795	0,51840285	0,05497988	-1,747	0,764	0,0475
0,90988289	0,44195934	0,0184148	1,121	0,504	0,091

1,52438427	0,95170131	-0,37538908	1,605	1,05	-0,378
1,40002736	0,01201929	0,72917825	1,41	0,03	0,728
1,20935775	1,46416934	0,25378728	1,204	1,475	0,249
1,36907939	0,90288275	0,87925389	1,378	0,901	0,876
1,52454501	0,2609561	0,73247974	1,534	0,259	0,742
0,10931179	1,20196253	0,0960362	0,108	1,21	0,101
1,35042444	1,2500337	-0,33595432	1,349	1,261	-0,334
0,69691451	0,38665958	0,77075303	0,698	0,385	0,778
1,58936233	1,85373701	-0,87562965	1,578	1,861	-0,874
1,55918295	0,44200466	0,0517683	1,557	0,442	0,051
0,88057724	1,60407211	0,50549217	0,8883	1,61	0,49
0,24022501	0,03996966	-0,0223098	0,239	0,04	-0,019
1,44800104	1,05660808	-0,32067434	1,4448	1,057	-0,3221
1,05523493	1,10222301	0,03813861	1,0567	1,108	0,041
0,45779395	1,55419499	0,55067471	0,459	1,5687	0,562
1,70105383	0,70143101	0,92443248	1,71	0,711	0,934
0,76691876	0,88192222	-0,47069573	0,756	0,872	-0,469
1,13949271	0,57131104	0,64566088	1,14	0,561	0,645
1,71912057	1,90310843	0,2941818	1,729	1,931	0,285
0,18529245	1,81512255	-0,5837016	0,179	1,825	-0,579
1,74203375	1,07277157	-0,76799187	1,73	1,081	-0,779

Табл. 7. Трьохвимірна модель

a1	a2	a3	d	estimation a1	estimation a2	estimation a3	estimation d
1,2045872	1,61188219	2,73688784	-0,99040811	1.192	1.600	2.725	-1.003
0,7007952	1,10699501	1,26241735	0,80575707	0.688	1.095	1.250	0.793
1,32958136	1,33714294	0,43900633	-0,10176466	1.317	1.325	0.427	-0.114
0,67257334	2,25136656	2,89186728	-0,4350184	0.660	2.239	2.880	-0.447
2,84134139	0,70076336	2,50998535	0,82582778	2.829	0.688	2.498	0.813
2,92101292	2,14473894	0,88243379	0,08862704	2.909	2.132	0.870	0.076
1,92155126	0,03199817	0,92205764	-0,93898271	1.909	0.020	0.910	-0.951
2,89866416	2,53159438	1,50445155	0,47558945	2.886	2.519	1.492	0.463
1,19170833	0,86330921	0,42925785	0,21682225	1.179	0.851	0.417	0.204
2,50868389	2,6631393	1,60335547	0,8678327	2.496	2.651	1.590	0.855
2,66857222	0,22898139	0,68800848	-0,93105831	2.656	0.217	0.676	-0.943
0,34155629	0,74573406	2,9410938	-0,25614309	0.329	0.733	2.929	-0.268
2,27395963	2,76077279	2,7137862	0,16481286	2.262	2.748	2.701	0.152
0,94625732	2,07323939	2,6876928	0,3079817	0.934	2.061	2.675	0.296
2,87774218	1,23406827	0,24242866	-0,71883815	2.865	1.222	0.230	-0.731
1,97714528	1,07726002	1,30844774	0,1163921	1.965	1.065	1.296	0.104

2,79418826	0,79460147	1,23548572	0,44670728	2.782	0.782	1.223	0.434
0,4395304	2,19491465	0,19662127	0,96752529	0.427	2.183	0.184	0.955
1,70040062	1,87366286	0,85704987	-0,21913522	1.688	1.861	0.845	-0.231
1,09444735	1,41623738	2,11947413	0,05673674	1.082	1.404	2.107	0.044
0,81966147	2,92454891	0,30715609	-0,1539111	0.807	2.912	0.295	-0.166
2,2195575	2,70624268	2,50991174	-0,22856864	2.207	2.694	2.498	-0.241
1,51996496	0,73030084	0,1122247	0,91076935	1.508	0.718	0.100	0.898
2,00418111	0,45098488	0,6133381	0,53006666	1.992	0.439	0.601	0.518
2,13707318	2,03904536	1,11536391	0,95099479	2.125	2.027	1.103	0.939
1,88874483	2,06071859	0,49341871	-0,01906154	1.876	2.048	0.481	-0.031
2,97496582	2,74444516	2,88510333	0,64106217	2.963	2.731	2.873	0.629
0,0730616	2,28655578	1,7925407	0,57647553	0.061	2.274	1.780	0.564
2,00725314	0,51468235	2,0102285	-0,5105304	1.996	0.502	1.998	-0.523
2,62263262	1,63045982	1,34741781	-0,62132331	2.610	1.618	1.334	-0.634

З останніх двох таблиць видно, що середні оцінки параметрів моделі відрізняються від тих значень, що подавались на функцію генерації. Це можна пояснити декількома факторами:

1) Точність оцінок залежить від об'ємів вибірок. Потрібно збільшити кількість питань і іспитників у тестуванні.

2) При моделюванні у матрицях первинних балів виникають рядки і стовпчики, де кількість 1 не належать інтервалу 5-95%. Тобто, є іспитники, які відповіли менше, ніж на 5% усіх питань, або більше, ніж на 95% відсотків. Або є питання, на які відповіло більше 95% іспитників, та питання, на які відповіло менше 5% іспитників. Такі рядки і стовпчики потрібно видаляти для коректної роботи ітераційного процесу розв'язання нелінійних систем щодо параметрів моделей. Оскільки ці дані порушують збіжність ітераційного процесу.

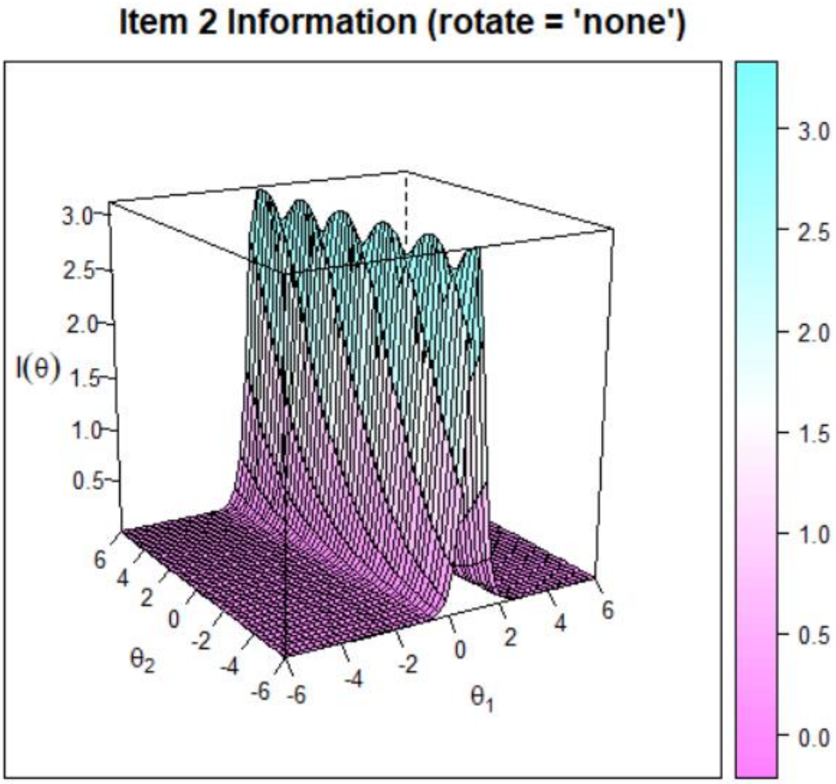
Після обчислення параметрів моделей потрібно перевірити адекватність моделі даним. В таблиці 8 наведено дані, отримані декількома критеріями.

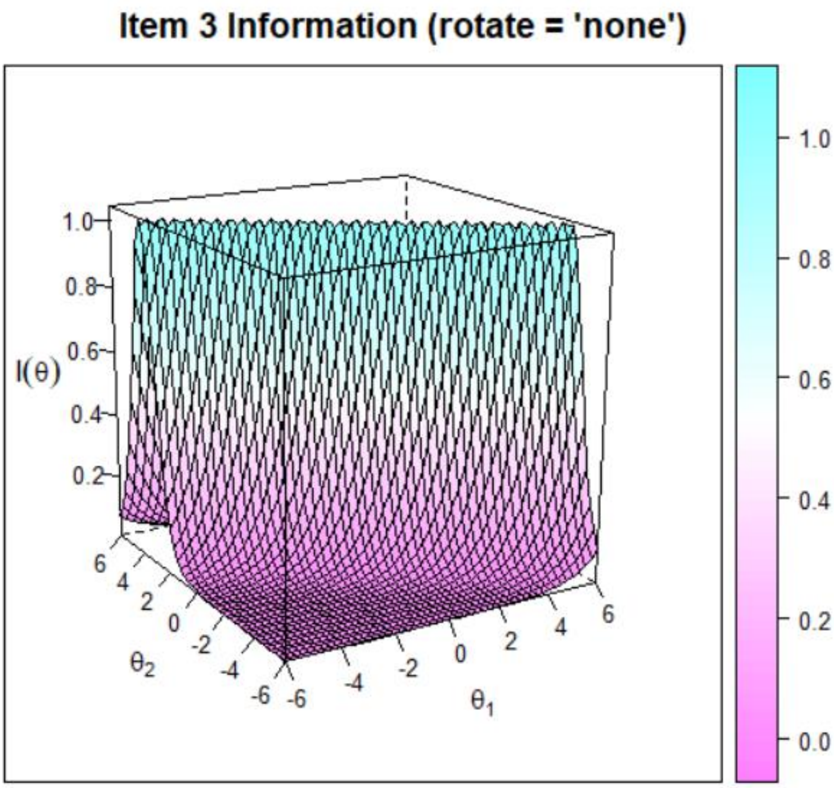
Модель\Критерій	M2	RMSEA	SRMSR	TLI	CFI
2PL2	389.0567	0.999596	0.008342045	0.01844536	0.0256336
2PL3	371.9355	0.01174036	0.03197654	0.9994733	0.9995787

З останньої таблиці видно, що моделі адекватні даним.

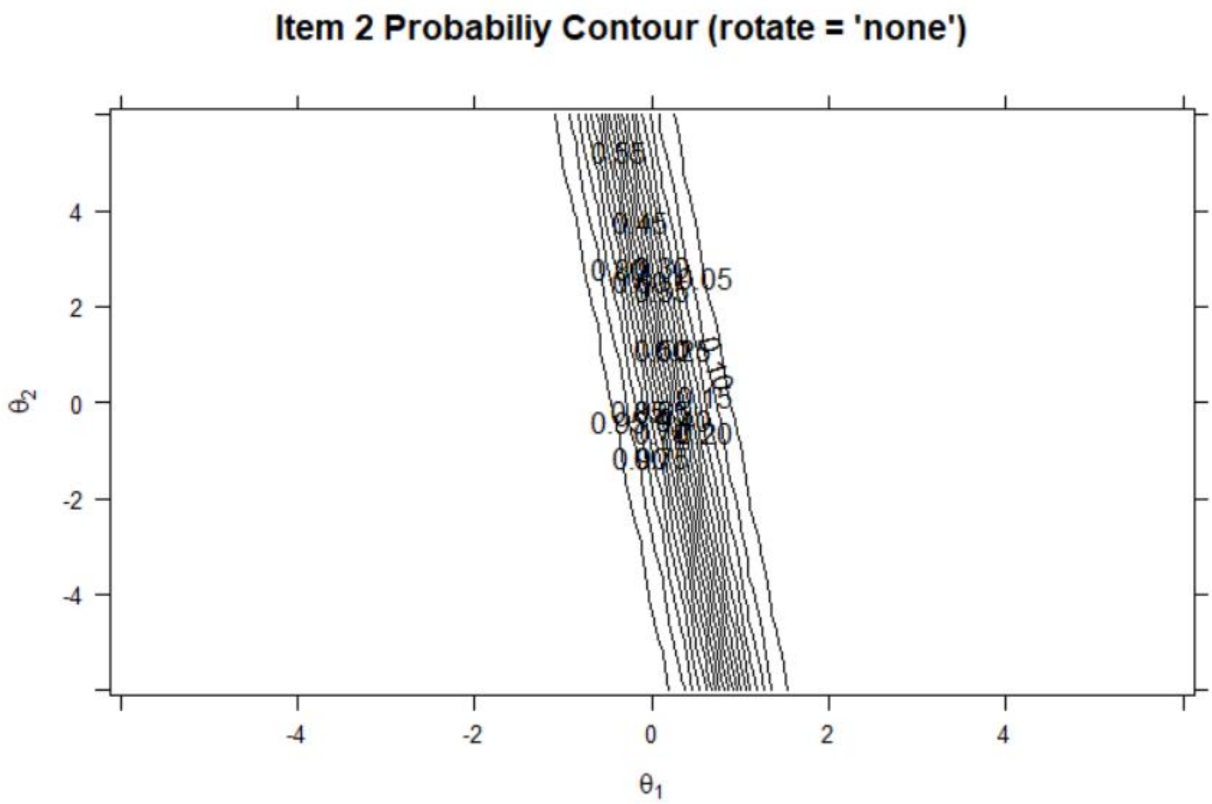
Наведемо фрагмент графічної ілюстрації MIRT аналізу згенерованої матриці первинних балів для компенсаторної двовимірної моделі.

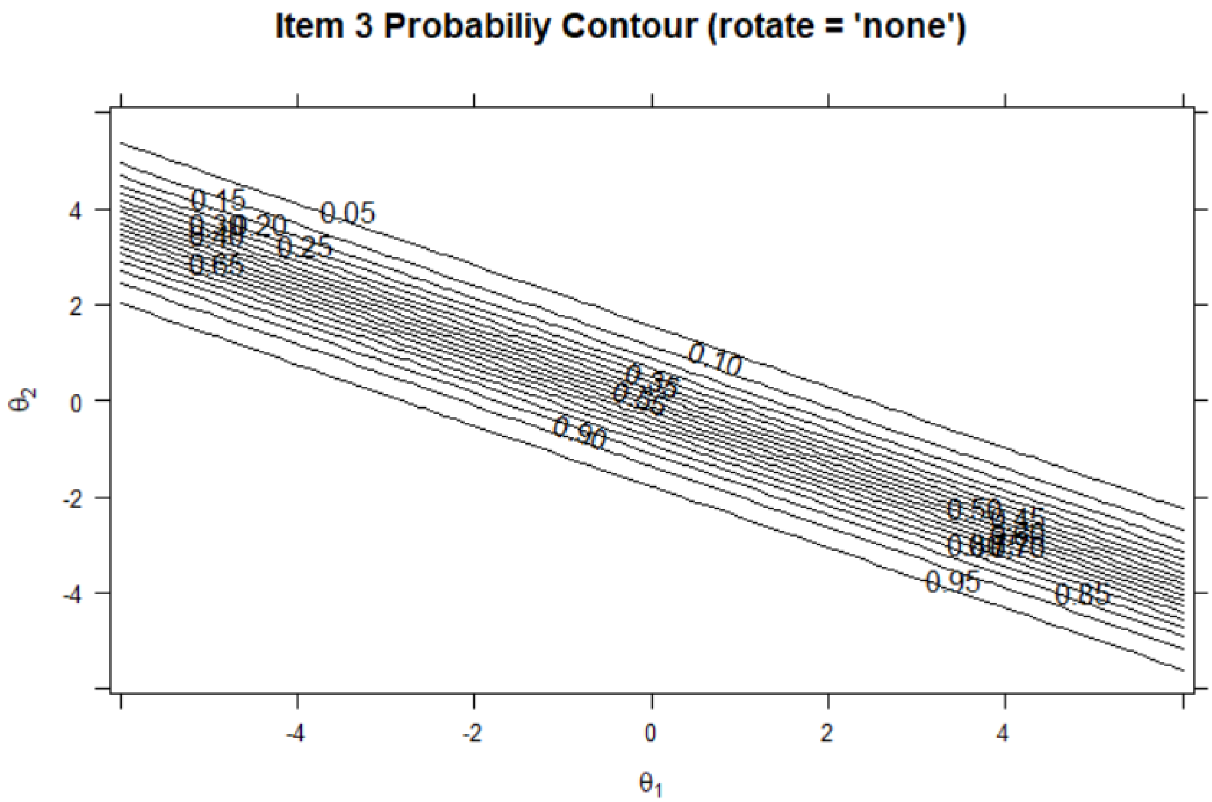
Інформаційні поверхні двох питань



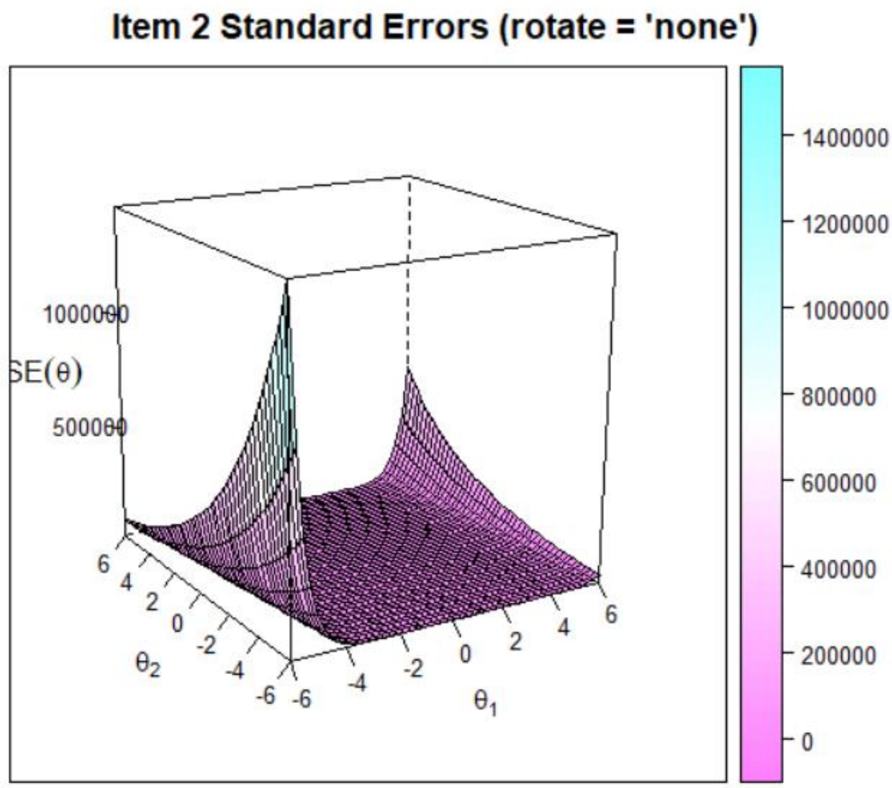

Item 3 Information (rotate = 'none')
I(θ)
θ1
θ2
1.0
0.8
0.6
0.4
0.2
0.0

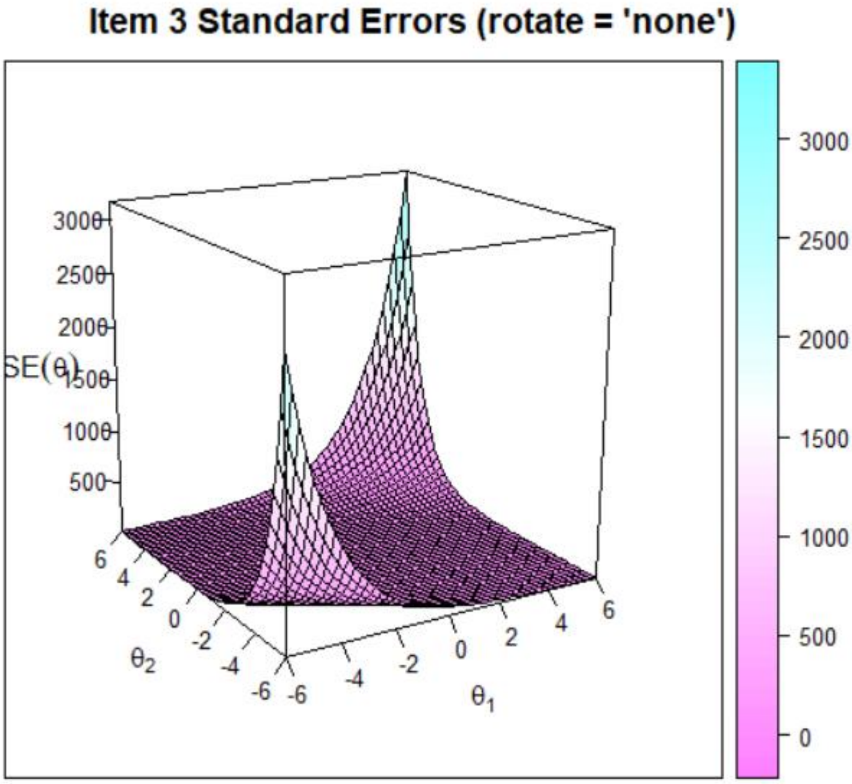
Лінії рівня


Item 2 Probabiliy Contour (rotate = 'none')
θ1
θ2
0.55
0.45
0.05
0.20
-4
-2
0
2
4



Поверхня похибок





Як видно із рисунків, питання 2 краще диференціює іспитників, ніж питання 3, оскільки лінії рівня розміщені більш щільно. Проте у 3 питанні краще проявляються обидві компетентності, тоді як у 2 питанні майже не проявляється перша компетентність. З інформаційних поверхонь і поверхонь похибок видно, що обидва питання добре характеризують іспитників, що мають стандартні рівні вираженості кожної компетенції. Тоді як при екстремальних значеннях рівнів підготовленостей спостерігається збільшення похибки.

ВИСНОВКИ

Змодельовано вибірки матриць первинних балів для двовимірної та трьохвимірної компенсаторної MIRT моделей із фіксованими параметрами тестових завдань.

Для визначення розмірностей моделей використано критерії EFA: HULL, PA, EKC.

Обчислено оцінки параметрів моделей за допомогою алгоритмів EM та MH-RM .

Перевірено адекватність моделей MIRT за допомогою M2, RMSEA, CFI, TLI.

В якості засобу реалізації відповідних алгоритмів обрано мову статистичного програмування R.

В результаті первинного аналізу матриць первинних балів було виявлено, що всі критерії показали однакову розмірність моделі, яка співпадала з розмірністю моделі, що подавалась на функцію моделювання матриці.

Змодельовані вибірки ідеально підходять для визначення розмірності моделі. Всі критерії дали однакову розмірність моделі, яка співпадала з вхідною.

При обчисленні параметрів моделей було виявлено, що оцінки параметрів моделей відрізняються від вхідних параметрів. Це може бути пов'язано з декількома причинами.

1) Якщо вибірка була невеликою, то отримані параметри можуть бути менш точними або мають велику дисперсію. Збільшення розміру вибірки поліпшує точність оцінок.

2) Вибрана модель має неоднозначність у розв'язанні нелінійних систем для параметрів моделі, особливо, якщо у вас невелика кількість тестових завдань або недостатньо даних для ідентифікації параметрів. Це може призводити до того, що програма знаходить локальний мінімум чи найкраще приближення, а не точний розв'язок.

3) Однією з можливих причин є те, що функція MIRT, яка обчислює параметри моделі використовує початкові значення параметрів, які можуть бути випадковими або неякісно підібраними.

Для вирішення цих питань рекомендується:

1)Збільшити розмір вибірки, якщо це можливо.

2)Використовувати обрані параметри моделі як початкові значення параметрів при використанні функції MIRT.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Braeken, J., & van Assen, M. A. An empirical Kaiser criterion. Psychological Methods, 22, 2017. С. 450–466. http://dx.doi.org/10.1037/met0000074
2. Horn, J. L. A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 1965. С. 179–185. http://dx.doi.org/10.1007/BF02289447
3. Zwick, W. R., & Velicer, W. F. Comparison of five rules for determining the number of components to retain. Psychological Bulletin,99, 1986. С. 432–442. http://dx.doi.org/10.1037/0033-2909.99.3.432
4. Glorfeld, L. W. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Edu cational and Psychological Measurement, 55, 1995. С. 377–393. http://dx.doi.org/10.1177/0013164495055003002
5. Bock RD, Aitkin M. "Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm." Psychometrika, 46(4), 1981. С. 443–459.
6. Ackerman T. An NCME instructional module on using multidimensional item response theory to evaluate educational and psychological tests. / T. Ackerman, M. Gierl, C. Walker. // Educational Measurement: Issues and Practice. – 2003. – №22. – С. 37–53.
7. Hartig J. Multidimensional IRT models for the assessment of competencies. / J. Hartig, J. Hohler. // Studies in Educational Evaluation. – 2009. – №35. – С. 57–63.
8. Crocker L., Algina J. Introduction to Classical and Modern Test Theory.—Belmont,CA:Wadsworth, 2006.—527 С.

9. Hambleton R.K., Swaminathan H., Rogers H. J. Fundamentals of Item Response Theory. – Newbury Park, CA: Sage, 1991. – 175 C.
10. R. Philip Chalmers. Mirt: A Multidimensional Item Response Theory Package for the R Environment
11. Traub, R. E., and Wolfe, R. G. Latent trait theories and assessment of educational achievement. In D. C Berliner (Ed.). Review of research in education 9, Washington, D.C.: American Educational Research Association, 1981.
12. Reckase M. D. Multidimensional item response theory.—New York:Springer, 2009.—354 C.
13. McDonald, R. P. The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 1981. C. 100-117.
14. Paul W. Holland; Paul R. Rosenbaum. Conditional Association and Unidimensionality in Monotone Latent Variable Models. / The Annals of Statistics, Vol. 14, No. 4. Dec., 1986. C. 1523-1543.
15. Stout, W. F. A nonparametric approach to assessing latent trait unidimensionality. Psychometrika, 52, 1987. C. 589-617.
16. Sympson, J. B. A model for testing multidimensional items. In D. J. Weiss (Ed.), Proceedings of the 1977 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978. C. 82-98.
17. Auerswald, M., & Moshagen, M. How to Determine the Number of Factors to Retain in Exploratory Factor Analysis: A Comparison of Extraction Methods Under Realistic Conditions. Psychological Methods. Advance online publication, 2019, January 21

http://dx.doi.org/10.1037/met0000200

18. Jöreskog, K. G. Factor analysis and its extensions. In R. Cudeck & R. C. MacCallum (Eds.), Factor analysis at 100: Historical developments and future directions. Mahwah, NJ: Erlbaum, 2007. C. 47–77

19. Humphreys, L. G., & Montanelli, R. G. An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 10, 1975. C. 193–205. http://dx.doi.org/10.1207/s15327906mbr1002_5

20. Marcˇenko, V. A., & Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. Mathematics of the USSR-Sbornik, 1, 1967. C. 457–483. http://dx.doi.org/10.1070/SM1967v001n04ABEH001994

21. Yu, C. Y. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Unpublished doctoral dissertation, University of California, Los Angeles, 2002. https://escholarship.org/content/qt7k49h7pm/qt7k49h7pm.pdf

22. Maydeu-Olivares, A., & Joe, H. Limited information goodness-of-fit testing in multidimensional contingency tables. Psychometrika, 71(4), 2006.C. 713-732.

23. Muthén, B., & Muthén, L. How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 9(4), 2002. C. 599-620.

24. Muthén, B., & Muthén, L. K. Mplus user's guide. Los Angeles, CA: Muthén & Muthén, 2012.