НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Фізико-математичний факультет

Кафедра математичного аналізу та теорії ймовірностей

«На правах рукопису»
УДК ____517.237_____

До захисту допущено:
Завідувач кафедри
_____ Олег КЛЕСОВ,
«___»__травня__2023 р.

Магістерська дисертація

на здобуття ступеня магістра

за освітньо-науковою програмою «Страхова та фінансова математика»

зі спеціальності 111 «Математика»

на тему: «Вибір розмірності моделі MIRT для аналізу тестів з вищої математики»

Виконав:
студент ІІ курсу, групи ОМ-11мн
Попрожук Марко Олегович _____

Керівник:
кандидат фізико-математичних наук,
доцент
Диховичний Олександр Олександрович _____

Рецензент:
доцент кафедри математики
та теоретичної радіофізики
КНУ ім. Тараса Шевченка
Єфіменко Світлана Володимирівна _____

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

Київ – 2023 року

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Фізико-математичний факультет

Кафедра математичного аналізу та теорії ймовірностей

Рівень вищої освіти – другий (магістерський)

Спеціальність – 111 «Математика»

Освітньо-наукова програма «Страхова та фінансова математика»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олег КЛЕСОВ

«08»_лютого_2023 р.

ЗАВДАННЯ

на магістерську дисертацію студенту

Попрожука Марко Олеговича

1. Тема дисертації «Вибір розмірності моделі MIRT для аналізу тестів з вищої математики», науковий керівник дисертації Диховичний Олександр Олександрович, кандидат фізико-математичних наук, доцент кафедри математичного аналізу та теорії ймовірності, затверджені наказом по університету від «27» березня 2023 р. №1337-с.
2. Термін подання студентом дисертації 19 травня 2023 року.

3. Об’єктом дослідження є моделі MIRT для статистичного аналізу тестів з вищої математики.

4. Предметом дослідження є вибір розмірності моделі MIRT.

5. Перелік завдань, які потрібно розробити:

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

2) Опанувати основні моделі та методи MIRT.

3) Дослідити методи EFA вибору розмірності моделі: PA, EKC, HULL.

4) Обрати алгоритми оцінювання латентних параметрів моделей.

5) Дослідити методи перевірки адекватності моделей MIRT.

6) Дослідити програмну реалізацію відібраних алгоритмів та методів у середовищі R.

7) На підставі відібраних алгоритмів і програм провести статистичний аналіз результатів контрольної роботи бакалаврів РТФ.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу: 23 слайдів.

7. Орієнтовний перелік публікацій:

1. Попрожук М. О., Круглова Н. В., Диховичний О. О. Вибір розмірності моделі MIRT для аналізу тестів з вищої математики // Тези XI Всеукраїнській науковій конференції молодих математиків, 11-13 травня 2023. Український державний університет імені Михайла Драгоманова.

2. Попрожук М. О., Круглова Н. В., Диховичний О. О. Застосування Wolfram Mathematica 13.2.1 для створення тестових завдань з вищої математики // Тези XI Всеукраїнській науковій конференції молодих математиків, 11-13 травня 2023. Український державний університет імені Михайла Драгоманова.

3. Kruglova N., Dykhovychnyi O., Poprozhuk M. Technologies for creating and analyzing tests in advanced mathematics // The Sixth Baltic-Nordic Conference on Survey Statistics. BaNoCoSS-2023. 21-25 August 2023. Helsinki, Finland. (подано на розгляд оргкомітету).

8. Дата видачі завдання 08 лютого 2023 року

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.	08.02.23-26.02.23	виконано
2.	Опанувати основні моделі та методи MIRT.	27.02.23-12.03.23	виконано
3.	Дослідити методи EFA вибору розмірності моделі: PA, EKC, HULL.	13.03.23-25.03.23	виконано
4.	Обрати алгоритми оцінювання латентних параметрів моделей.	26.03.23-14.04.23	виконано
5.	Дослідити методи перевірки адекватності моделей MIRT.	15.03.23-23.03.23	виконано
6.	Дослідити програмну реалізацію відібраних алгоритмів та методів у середовищі R.	24.04.23-04.05.23	виконано
7.	На підставі відібраних алгоритмів і програм провести статистичний аналіз результатів контрольної роботи бакалаврів РТФ.	05.05.23-14.05.23	виконано

Студент Марко Попрожук

Науковий керівник Олександр Диховичний

Реферат

Магістерська дисертація: 65 сторінки, 33 першоджерела, 23 слайдів презентації. Робота складається зі вступу, п’яти розділів, висновків та списку використаної літератури.

В дисертаційній роботі досліджується вибір розмірності моделі MIRT для аналізу тестів з вищої математики. Основною метою дисертаційного дослідження є вибір розмірності моделі MIRT для аналізу тестів з вищої математики. Об’єктом дослідження є моделі MIRT для аналізу тестів з вищої математики. Предметом дослідження є 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 контрольної роботи з вищої математики бакалаврів РТФ.

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

Abstract

Master's Thesis: 65 pages, 33 primary sources, 23 presentation slides. The work consists of an introduction, five chapters, conclusions, and a list of references.

The dissertation research focuses on selecting the dimensionality of the MIRT model for analyzing tests in higher mathematics. The main objective of the research is to determine the dimensionality of the MIRT model for analyzing tests in higher mathematics. The object of the study is MIRT models for analyzing tests in higher mathematics, while the subject of the study is the selection of the dimensionality 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 chapter presents a statistical analysis of the results of a mathematics test for bachelor students in the Faculty of Engineering.

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

Зміст

ВСТУП

Наразі, через агресію РФ проти України, дистанційна освіта для українських студентів була й залишається однією із основних форм навчання. Як наслідок, створення якісного контенту для перевірки знань студентів в таких умовах є надзвичайно важливою задачею. У цій царині основними напрямами є конструювання та статистичний аналіз якості створених тестів.

Команда викладачів з Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» забезпечує процес створення нових тестів з вищої математики та їх статистичного аналізу.

Основна увага при створенні нових тестів приділяється статистичному аналізу тестів. Мета такого аналізу - покращення якості тестів. На основі проведеного аналізу відбувається переформатування тесту, видалення або переформулювання занадто легких або занадто складних завдань.

Основу статистичного аналізу традиційно складають Clasical Test Theory (CTT) та Item Response Theory (IRT) [1]. Для аналізу тестів з вищої математики ми використовується більш розвинена теорія – Multidimensional Item Response Theory (MIRT) [2] , що дає можливість більш тонко вирізняти окремі риси та здібності іспитників. Центральним питанням застосування MIRT є питання підбору адекватної моделі для проведення аналізу і, зокрема, вибору розмірності моделі. Цій тематиці присвячено роботи [3, 4, 5].

В роботі [1] показано, що для дослідження якості тестів з ВМ доречно застосовувати модель MIRT, зокрема двовимірну компенсаторну модель 2PL. У цій роботі основу вибору склав Емпіричний критерій Кайзера. Але використання лише єдиного критерію може бути недостовірним. Тому у даній магістерській дисертації запропоновано обрати декілька алгоритмів EFA, провести оцінку параметрів і на підставі декількох критеріїв адекватності провести вибір найкращої моделі.

Загалом схема має вигляд:

I. Проведення розвідувального факторного аналізу (EFA). На цьому етапі застосовуються декілька алгоритмів: PA(Parallel Analysis), EKC(Empirical Kaiser criterion), HULL(Horn's Parallel Analysis with Minimum Average Partial).

II. Оцінка параметрів моделей (CFA). Пропонується компенсаторні 2PL моделі розмірностей, які обрано на першому етапі. Політомічна й дихотомічна. Дихотомічна будується на підставі «розщеплення» політомічних задач на дихотомічні.

III. Вибір «найкращої» моделі на підставі застосування критеріїв: AIC, BIC, RMSEA, CFI,TLI.

IV. Аналіз завдань для «найкращої» моделі.

I. Теоретичні основи статистичного аналізу тестів

1. Теоретичні основи IRT

Основу сучасного статистичного аналізу тестів складають Класична Теорія Тестів (КТТ) [1] та, так звана, Сучасна Теорія Тестів (IRT)[6]. Суть IRT полягає у встановленні певної залежності між імовірністю правильної відповіді на тестове завдання і латентними (прихованими) параметрами, які чисельно характеризують як завдання, так і іспитників.

Відмітимо переваги IRT порівняно з КТТ [7]

• Стійкість та інваріантність параметрів підготовки іспитників.

• Стійкість та інваріантність оцінок параметрів складності та диференціюючої здатності завдань.

• Побудова єдиних шкал для вимірювання всіх латентних параметрів.

Обробка результатів контролюючих заходів в умовах дистанційної освіти з вищої математики методами IRT проводилась роботах [4,8]

Логістичні моделі у IRT

Основу IRT складають, так звані, логістичні моделі. Найбільш загальною одновимірною моделлю є модель – 4-PL [9]. Пояснимо її.

Нехай тест містить N дихотомічних завдань, які пройшло M іспитників. Ймовірність правильної відповіді j-того іспитника на i-те завдання визначається формулою:

$$P\left(X_{ij}=1|\,\theta_j,a_i,b_i,u_i,v_i\right)=u_i+\frac{v_i-u_i}{1+e^{-a_i(\theta_j-b_i)}},$$

- θ_j - рівень підготовленості j-того іспитника;
- b_i - рівень складності завдання ;
- $a_i>0$ - диференціююча здатність завдання;

- u_i - нижній асимптотичний параметр характеристичної функції;
- v_i - верхній асимптотичний параметр характеристичної функції i-того завдання.

3-PL модель є частинним випадком 4-PL моделі при $v_i = 1$, при чому для цієї моделі параметр u_i - це ймовірність вгадати правильну відповідь в i-тому завданні.

$$P(X_{ij} = 1 | \theta_j, a_i, b_i, u_i) = u_i + (1 - u_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

2-PL модель характеризується двома параметрами завдань: a_i, b_i, тоді як $v_i = 1$і $u_i = 0$ для всіх завдань.

$$P(X_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}.$$

Для політомічних завдань використовується модель Муракі (GPCM) [10], в якій ймовірність досягнення і-м іспитником k-го рівня j-ого завдання визначається за формулою:

$$P(X_{\kappa ij} = \kappa | \theta_j, a_i, b_i) = \frac{e^{\sum_{l=0}^{k} a_i(\theta_j - b_{il})}}{\sum_{k=0}^{m_i} e^{\sum_{l=0}^{k} a_i(\theta_j - b_{il})}}, i = \overline{1, N}, \ j = \overline{1, S}, k = \overline{0, m_i},$$

де

- a_j - диференціююча здатність завдання завдання;
- $b_{jk}, j = \overline{1, S}, k = \overline{0, S_j}$ - параметри складності k-го рівня j-го завдання;
- $\theta_i, i = \overline{1, N}$ - підготовленість і-того іспитника;
- m_j - кількість рівнів завдання .

2.Теоретичні основи MIRT

Однією з основних передумов для використання IRT є те, що всі завдання тесту вимірюють одну і ту ж компетентність комплексу вмінь. У деяких випадках завдання, що складають тест, можуть вимірювати різні вміння або різні складові комплексу вмінь. Наприклад, математичний тест з алгебри може містити два типи завдань: розв'язування рівнянь та задач з текстом. Перший тип завдань вимагає від студентів вміння маніпулювати алгебраїчними символами, тоді як другий тип потребує спочатку прочитати та перекласти текст, а потім застосувати вміння маніпулювання символами. Залежно від обсягу тексту, можна уявити послідовність математичних завдань, які вимірюють кілька різних комплексів вмінь - від чистого маніпулювання алгебраїчними символами до читання та перекладу тексту. Трауб (1981)[11] стверджував, що якщо враховувати всі необхідні вміння для розв'язання всіх завдань більшості когнітивних тестів, то одновимірність, ймовірно, більше виняток, ніж правило. Для таких випадків розвинуто теорію відгуку на завдання з кількома вимірами (MIRT)[2], яка дозволяє визначати різні складові вмінь, що вимірюються відповідними завданнями.

IRT намагається описати взаємодію між характеристиками завдань та латентними здібностями іспитника особи за допомогою ймовірнісних моделей. Взаємодія між групою іспитників та одним завданням завжди може бути описана одновимірно, оскільки завдання може вимірювати лише один навичку або одну композитну групу навичок. Проте, коли розглядається взаємодія між групою іспитників та кількома завданнями, припущення про одновимірність потрібно розглядати дуже уважно. Якщо тест здатен вимірювати декілька різних навичок, але іспитники різняться лише за рівнем володіння однією цих навичок, то взаємодію можна описати одновимірно. Аналогічно, якщо тест завдань вимірює рівні лише однієї навички (або рівні однієї композитної групи кількох навичок), а іспитники відрізняються за рівнем володіння кожною з цих навичок, то взаємодію також можна описати

одновимірно. Однак, якщо тест здатен відрізняти рівні кількох навичок, а іспитники відрізняються за рівнем володіння більше, ніж одним з цих навичок, то взаємодію потрібно описувати багатовимірно.

Передбачати одновимірність не слід, але завжди треба перевіряти її наявність. Для дослідження одновимірності відповідей існує кілька теоретично обґрунтованих підходів, які можуть застосовуватися практиками. Серед досліджень в цій галузі варто згадати нелінійні факторні аналізи Макдональда [12] та підхід умовних асоціацій Голланда та Розенбаума [13]. Іншою технікою є процедура статистичного тестування на відсутність суттєвої одновимірності, запропонована Нандакумаром та Стаутом [14]. Однак ці підходи дають змогу статистично перевірити, чи є певний підтест елементів одновимірним у порівнянні з іншими елементами тесту, але не дають можливості визначити реальну вимірність тесту.

MIRT Models

MIRT (Multidimensional Item Response Theory) - це статистична теорія, яка використовується для оцінювання кількох латентних змінних (або факторів) відповідно до відповідей на набір тестових завдань. У MIRT підході, кожне завдання має різні параметри, які описують його властивості, такі як складність завдання, дискримінативність, та можливість відповіді "не знаю".

MIRT відрізняється від класичної IRT моделі, так як дозволяє моделювати взаємодію між латентними змінними та тестовими завданнями. У той час як IRT модель моделює відповіді як залежність між відповідями та однією латентною змінною, MIRT дозволяє моделювати відповіді як залежність між відповідями та кількома латентними змінними. Це дозволяє більш детально досліджувати взаємозв'язок між кожним тестовим завданням та кожною латентною змінною, що може бути корисним у різних областях, наприклад, у педагогіці, психології, та інших сферах.

В MIRT для опису здібностей та параметрів питань використовуються багатовимірні параметри. Це означає, що кожна здібність (latent trait) та кожен параметр питання (item parameter) описується вектором чисел, де кожне число відповідає конкретному аспекту здібності або параметру питання.

Зазвичай в MIRT використовують двовимірні моделі, де здібність описується двома параметрами (наприклад, здібність до математики та мовленнєвого розвитку), а кожен параметр питання - двома параметрами, що відповідають за його складність та дискримінативність.

Багатовимірні параметри дозволяють більш точно моделювати взаємодію між здібностями та параметрами питань, що підвищує якість оцінювання здібностей студентів або учнів. Однак, використання багатовимірних параметрів ускладнює процес оцінювання та аналізу даних.

Для опису даних відповідей на дихотомічні елементи застосовують дві основні моделі MIRT: компенсаторну та некомпенсаторну. Ймовірність правильної відповіді на елемент і може бути визначена з використанням компенсаторної двопараметричної логістичної моделі з m вимірами (MC2PL), запропонованої Reckase в 1985 році. [2].

Рівень підготовленості іспитника у багатовимірній моделі характеризується лінійною комбінацією компонент s-вимірного вектору $\overline{\theta_j} = (\theta_{j1}, \theta_{j2}, ..., \theta_{js})$. Нехай тест містить N завдань, які пройшло M іспитників

$$P\left(X_{ij} = 1 | \bar{\theta}_j, \bar{a}_i, d_i\right) = \frac{e^{\bar{a}_i \bar{\theta}_j' + d_i}}{1 + e^{\bar{a}_i \bar{\theta}_j' + d_i}} \quad i = \overline{1, N}, j = \overline{1, M}.$$

Можна розкласти експоненту e у цій моделі, щоб продемонструвати, як взаємодіють компоненти векторів a та θ.

$$\bar{a}_i \bar{\theta}_j' + d_i = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \cdots + a_{is}\theta_{js} + d_i = \sum_{l=1}^{s} a_{il}\theta_{jl} + d_i$$

У цій моделі експонента є лінійною функцією елементів вектора $\bar{\theta}$, де параметр d є терміном перетину, а елементи вектора a є параметрами нахилу. Вираз у показнику визначає лінію в просторі з m вимірами. Це має цікаву властивість: якщо показник встановлений на деяке постійне значення k, то всі вектори $\bar{\theta}$, які задовольняють виразу $k = \bar{a}_I \bar{\theta}_j' + d_i$, лежать на одній прямій лінії і дають однакову ймовірність правильної відповіді для моделі.

Можна сформулювати вираз для ймовірності правильної відповіді, використовуючи некомпенсаторну двопараметричну логістичну модель (MNC2-PL) [15], як:

$$P\left(X_{ij} = 1 | \bar{\theta}_j, \bar{a}_i, \bar{b}_i\right) = \prod_{l=1}^{s} \frac{e^{\bar{a}_{il}(\bar{\theta}_{jl} - \bar{b}_{il})}}{1 + e^{\bar{a}_{il}(\bar{\theta}_{jl} - \bar{b}_{il})}}$$

У цьому виразі X_{ij} - це оцінка (0,1) для питання i, $\bar{a}_{il} = (a_{i1}, a_{i2}, ..., a_{is})$ - вектор параметрів дискримінації питання, $\bar{b}_{il} = (b_{i1}, b_{i2}, ..., b_{is})$ - вектор параметрів складності для питання i, а $\bar{\theta}_{jl} = (\theta_{j1}, \theta_{j2}, ..., \theta_{js})$ - вектор параметрів здібності. Ця модель містить параметри дискримінації та складності для кожної з вимірів. Важливо зазначити, що цю модель можна подати як добуток двох або більше одновимірних моделей (двопараметрична логістична модель [2-PL] IRT). Зокрема, у випадку з двома вимірами некомпенсаторну модель можна переписати як:

$$P\left(X_{ij} = 1 | \bar{\theta}_j, \bar{a}_i, \bar{b}_i\right) = \frac{e^{a_{i1}(\theta_{j1} - b_{i1})}}{1 + e^{a_{i1}(\theta_{j1} - b_{i1})}} \frac{e^{a_{i2}(\theta_{j2} - b_{i2})}}{1 + e^{a_{i2}(\theta_{j2} - b_{i2})}}$$

Множинна природа компонентів моделі забороняє іспитнику компенсувати низький рівень здібностей на одній вимірі шляхом високого рівня на іншій вимірі. Один з випадків, коли компенсаторна модель була б доречною, - це тест, в якому запитання вимірюють дві здібності: читання і тематичні знання. Розгляньте тест з читання, в якому здобувачі повинні читати

та відповідати на запитання про пасажі про бейсбол. Здобувачі, які добре знають бейсбол, можуть компенсувати низький рівень читання завдяки їх розумінню теми. Так само, здобувачі, які відмінно володіють читанням, можуть компенсувати відсутність знань про бейсбол завдяки своїм навичкам читання.

Припустимо, що в американській школі старшої школи учні вивчають французьку як другу мову. Припустимо, що тест передбачає, що іспитники повинні спрягати французькі дієслова, як вказано у вказівках, написаних англійською мовою. Такий тест фактично вимірює дві навички: розуміння словникового запасу, пов'язаного зі спряженням дієслів, та знання французької мови. Якщо учень має добру знайомість зі спряженням дієслів, але мало знає французької мови, компенсація не відбудеться. Аналогічно, учні, які незнайомі з термінами, такими як "future perfect" або "past participle", не зможуть компенсувати це більшою володінням французькою мовою.

Модель M3-PL є розширенням моделі M2-PL і передбачає можливість наявності ненульової нижньої асимптоти в моделі. Ця модель є багатовимірним розширенням трьохпараметричної логістичної моделі UIRT, описаної у попередній розділі про IRT. Математичний вираз для моделі M3-PL наведено у формулі даного розділу (1), яка використовує ті самі символи, що і в попередніх визначеннях.

$$P\left(X_{ij}=1|\bar{\theta}_j,\bar{a}_i,u_i,d_i\right)=u_i+(1-u_i)\frac{e^{\bar{a}_i\bar{\theta}_j^{'}+d_i}}{1+e^{\bar{a}_i\bar{\theta}_j^{'}+d_i}} \qquad (1)$$

Модель M3-PL була розроблена з метою пояснення спостережуваних емпіричних даних, таких як ті, що наведені в роботі Лорда (1980)[16], де показано, що люди з низькою здібністю мають ненульову ймовірність правильної відповіді на багатовимірні завдання. Оскільки процес вибору правильної відповіді для осіб з низькими здібностями не здається пов'язаним з показниками, які оцінюються завданням тесту, модель містить один

параметр нижньої асимптоти, або параметр псевдо-вгадування сі, який визначає ймовірність правильної відповіді для осіб з дуже низькими значеннями θ.

Для політомічних, у яких передбачено $m_i, \overline{i=1,N}$ підзавдань використовуємо модель GPCM (S=2):

$$P(X_{kij}=k \mid \theta_{j1},\theta_{j2},a_{k1},a_{k2},d_k)=\frac{e^{a_{k1}\theta_{j1}+a_{k2}\theta_{j2}+d_{k-1}}}{\sum_{l=1}^{m_i} e^{a_{l1}\theta_{j1}+a_{l2}\theta_{j2}+d_{l-1}}},$$

де $i = \overline{1,\ N}$; $j = \overline{1,\ M}$; k = $\overline{1,\ m_i}$;

Імовірності правильної відповіді в MIRT зображають за допомогою характеристичних поверхонь завдань, які зображено на рисунках 1 і 2. На рисунку 1- компенсаторна модель з параметрами a_1 = 1,6, a_2 = 0,6, $d = 0{,}0$; на рисунку 2-некомпенсаторна a_1 = 1,6, a_2 = 0,6, $d_1 = 0{,}0$, d_2 = 0,0.

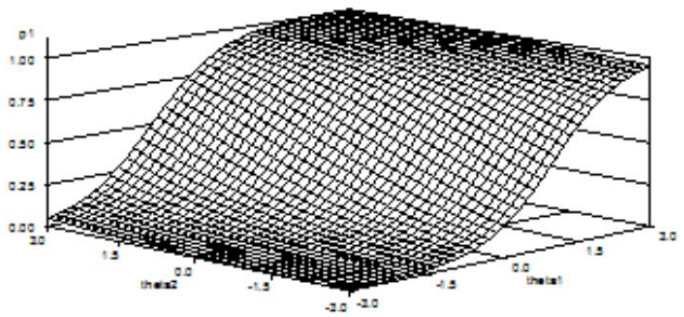


Рис. 1.

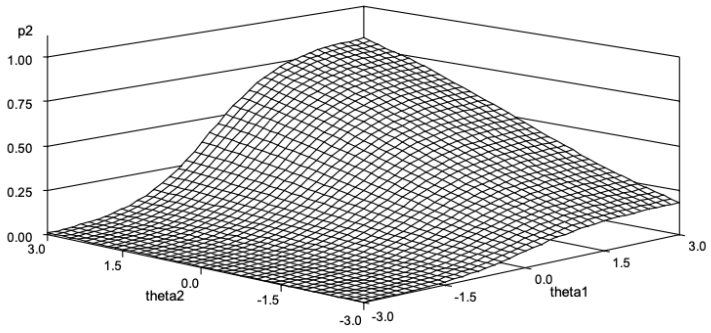


Рис. 2.

Для багатовимірних моделей складність i-го завдання - це відстань від точки $\bar{\theta}$ до точки, яка відповідає найкрутішому нахилу поверхні. Її знак вказує на розташування точки відносно початку координат, а значення визначається формулою:

$$\hat{B}_i = \frac{-d_i}{\sqrt{\sum_{k=1}^{s} a_{ik}^2}}$$

Багатовимірна диференціююча здатністю - це нахил характеристичної поверхні у точці найбільшої крутизни схилу у напрямі точки $\bar{\theta}$, її значення визначається формулою:

$$\hat{A}_i = \sqrt{\sum_{k=1}^{s} a_{ik}^2}$$

II. Математичні методи визначення розмірності моделей MIRT

1. Визначення розмірності моделі на підставі алгоритмів EFA

Основними методами визначення латентних факторів в MIRT є Дослідницький факторний аналіз (Exploratory Factor Analysis) (EFA) та Підтверджувальний факторний аналіз Confirmatory Factor Analysis (CFA).[1]

EFA в MIRT - це метод факторного аналізу, який використовується для виявлення латентних факторів у відповідях на багатовимірні тести або запитання, які вимірюють різні аспекти здібностей чи характеристик.

EFA є широко використовуваним статистичним методом для дослідження латентної структури багатьох спостережуваних змінних, особливо якщо немає чіткої теоретичної підстави для конкретної моделі. EFA визначає приховану структуру за допомогою підходу, що ґрунтується на даних, припускаючи наявність загальної факторної моделі. У цій моделі кожна спостережувана змінна концептуалізується як зважена сума множини (можливо корельованих) факторних змінних та єдиного унікального фактора. Загальні фактори пояснюють коваріації між спостережуваними змінними та є факторами теоретичного інтересу. Унікальні фактори, з іншого боку, виключно пояснюють дисперсії окремих спостережуваних змінних, що вважається вимірювальною помилкою щодо загальних факторів.

Визначення кількості латентних факторів є ключовою проблемою в EFA[17]. Якщо кількість факторів недооцінюється або переоцінюється, це може негативно вплинути на якість EFA. Недооцінення кількості факторів призводить до помилок у всіх завантаженнях факторів, незалежно від їх ваги в правильно визначеній моделі. Крім того, недооцінення погіршує факторні оцінки порівняно з оцінками у правильно визначеній моделі. На відміну від цього, переоцінення кількості факторів, як правило, призводить до менших зміщень у факторних оцінках. Проте переоцінення кількості факторів може призвести до розщеплення факторів, коли змінні з розділяються на кілька факторів після обертання. Крім того, переоцінення кількості факторів

призводить до менш парсимонійних моделей, які включають конструкти з малою або без пояснювальної вартості, і збільшує вірогідність виникнення Хейвуд-випадків, таких як оцінки від'ємної дисперсії.

Загальна факторна модель (для огляду, див. Jöreskog, 2007)[18] передбачає наявність S латентних загальних факторів $\theta_1,...,\theta_S$, які пояснюють варіацію в N спостережуваних (і стандартизованих) випадкових змінних $x_1,...,x_N$. Кожна окрема спостережувана змінна x_i вважається лінійною комбінацією факторів $\theta_1,...,\theta_S$ та унікального фактору ε_i, подібно до лінійної регресії.

$$x_i = \alpha_{i1}\theta_1 + \alpha_{i2}\theta_2 + ... + \alpha_{iS}\theta_S + \varepsilon_i, 1 \le i \le N$$

В моделі загального фактору, фактори ε_i не корелюють з будь-якими іншими факторами $\theta_1,...,\theta_S$ і всі $\varepsilon_{i'}$ для $i \ne i'$. Крім того, коефіцієнт α_{ij} представляє навантаження і-го спостережуваного показника на фактор j. Отже, метою є знаходження загальних латентних факторів, менших за кількість спостережуваних змінних, які пояснюють коваріації між спостережуваними змінними $x_1,...,x_N$ таким чином, що $x_1,...,x_N$ стають некорельованими при умові наявності латентних факторів $\theta_1,...,\theta_S$.

Існує кілька способів визначення кількості факторів у EFA, більшість з яких ґрунтуються на власних значеннях, які відображають дисперсію, що пояснюється кожним загальним фактором. До них належать такі відомі методи, як Empirical Kaiser Criterion, Hull Method та Parallel Analysis (PA)[17]. Так як загальна факторна модель є частковим випадком моделювання структурних рівнянь, показники адекватності моделі також часто використовуються для перевірки правильності вибраної кількості факторів.

CFA [1] є стандартним методом для оцінки підтверджувальної моделі. В той час як факторний аналіз (FA) моделює залежність між спостережуваними змінними та складовими факторами, CFA моделює залежність між

спостережуваними змінними та латентними факторами. В контексті багаторівневих ієрархічних моделей MIRT CFA використовується для моделювання взаємодії між різними рівнями вимірювання, такими як підшкали, і кількома латентними факторами.

У MIRT кожен індикатор вимірювання моделюється як функція двох або більше латентних факторів, що відображають різні аспекти здібностей чи характеристик. CFA в MIRT дозволяє включати кореляції між латентними факторами, що використовуються для опису різних вимірювань, а також різні вагові коефіцієнти для різних індикаторів, що відображають їх різні вклади у латентний фактор.

Для оцінювання параметрів MIRT можуть використовуватися методи максимальної правдоподібності або методи байєсівської оцінки [9]. Метод максимальної правдоподібності зазвичай використовується в більшості досліджень, але метод байєсівської оцінки може бути корисним, коли потрібно зайняти позицію щодо невизначеності параметрів, або коли дослідження має обмежену вибірку.

Метод максимальної правдоподібності (maximum likelihood, ML)[9] є одним з найбільш популярних методів оцінювання параметрів в CFA. Його суть полягає у тому, що за допомогою статистичної процедури максимізується ймовірність того, що модель, яку ми побудували, відображає спостережувані дані.

У CFA з методом максимальної правдоподібності, ми спочатку будуємо модель, яка включає загальні фактори та їхні спостережувані показники. За допомогою статистичної процедури максимальної правдоподібності, ми шукаємо такі значення параметрів моделі (наприклад, загальні фактори та їхні коефіцієнти навантаження), які максимізують ймовірність того, що наша модель описує спостережувані дані.

Іншими словами, метод максимальної правдоподібності дозволяє нам знайти такі значення параметрів моделі, які найбільш вірогідно описують наші спостереження. При цьому, ми можемо використовувати різні критерії для оцінки адекватності моделі, такі як індекси збігу, які дозволяють оцінити якість підгонки моделі до спостережень.

Ми будемо застосовувати алгоритми EM та NH-RM.

Метод максимальної правдоподібності є дуже потужним і популярним методом в CFA, але він також може бути чутливим до ненормальності даних та великих викидів. У таких випадках можуть бути використані альтернативні методи, такі як метод мінімуму квадратів (least squares estimation) [9].

Методи байєсівської оцінки [9] можуть бути застосовані в CFA для оцінки параметрів моделі та інформативних попередніх розподілів цих параметрів. Замість традиційних методів оцінки, таких як Метод Максимальної Правдоподібності (ML) або Метод Моментів (MM), байєсівські методи використовують байєсівське правило для оновлення оцінок параметрів моделі з урахуванням нової інформації.

Алгоритми визначення розмірності моделі у MIRT

Більшість критеріїв для визначення кількості факторів в EFA базуються на власних значеннях [19]. Для кращого розуміння зв'язку між власними значеннями та моделлю загального фактору, останню можна виразити у матричній формі. Для $\bar{X} = (x_1,...,x_N)', \bar{\theta} = (\theta_1,...,\theta_S)', \bar{\varepsilon} = (\varepsilon_1,...,\varepsilon_N)'$ та матриці завантажень розмірності $N \times S$ позначеної як Λ,

$$\bar{X} = \Lambda\bar{\theta} + \bar{\varepsilon} \quad (1)$$

Можна виразити матрицю кореляції проявних змінних X як

$$B = E(\bar{X}\bar{X}') \quad (2)$$

де E – математичне сподівання, оскільки спостережувані змінні стандартизовані. З рівняння 1 випливає, що

$$\bar{X}\bar{X}' = (\Lambda\bar{\theta} + \bar{\varepsilon})(\Lambda\bar{\theta} + \bar{\varepsilon})' \quad (3)$$

$$= \Lambda\bar{\theta}\bar{\theta}'\Lambda' + \Lambda\bar{\theta}\bar{\varepsilon}' + \bar{\varepsilon}\bar{\theta}'\Lambda' + \bar{\varepsilon}\bar{\varepsilon}' \quad (4)$$

Ми використовуємо позначення $\Phi = E(\bar{\theta}\bar{\theta}')$ для матриці кореляцій між загальними факторами $\bar{\theta}$ та $\Delta = E(\bar{\varepsilon}\bar{\varepsilon}')$ для матриці коваріацій унікальних факторів $\bar{\varepsilon}$. Оскільки $\bar{\theta}$ та $\bar{\varepsilon}$ є незалежними, модель виражає матрицю кореляцій у такий спосіб:

$$B = \Lambda\Phi\Lambda' + \Delta \quad (5)$$

Матриця Δ в рівнянні 5 є діагональною, оскільки модель загального фактору передбачає незалежність всіх унікальних факторів ε_i, $\varepsilon_{i'}$, для $i \neq i'$. Елементи δ_i матриці Δ називаються факторами унікальності і представляють частину дисперсії виявленої змінної x_i, яка не залежить від латентних факторів і є частиною дисперсії x_i, яка може бути пояснена латентними факторами. Модель загального фактора оцінює матрицю Λ з точністю до обертання таким чином, що

$$\tilde{B}_C \approx \Lambda\Lambda' \quad (6)$$

В рівнянні 6 матриця $\tilde{B}_C$ отримується після заміни діагональних елементів B на унікальності. Один із розв'язків методу найменших квадратів для рівняння 6 визначає завантаження Λ, що пропорційні так званим власним векторам $\tilde{B}_C$ [18]. Загалом, власними векторами називаються вектори, для яких:

$$Av = \lambda v, v \neq 0 \quad (7)$$

Це спільне твердження стосується квадратних матриць A розмірності NxN, векторів $\bar{v}$ довжини N і власних чисел λ. Для симетричних, позитивно-невід'ємних матриць, як от коваріаційні матриці або матриці R_C, існує рівно N (не обов'язково різних) невід'ємних власних чисел. Найголовніше, j-те найбільше власне число матриці $\tilde{B}_C$ відповідає дисперсії, поясненій j-м фактором в загальній факторній моделі.

PCA часто використовують як альтернативу EFA для аналізу факторів на основі загальної факторної моделі. Проте, PCA є методом зменшення кількості даних, який не враховує унікальну дисперсію. Якщо метою є виявлення латентної структури, що відображає коваріації між змінними з випадковою похибкою, що є більш реалістичним в психологічних дослідженнях, зазвичай використовують EFA. Основна різниця полягає у тому, що власні числа PCA обчислюються на основі кореляційної матриці B, а не B_C. Проте, гіпотетична загальна факторна модель населення повністю визначає кореляційну матрицю B і пов'язані власні числа B. Деякі дослідження показують, що EFA та PCA можуть давати порівняні результати на практиці.

Незалежно від використаного критерію для визначення кількості збережених факторів, існують умови, які можуть спростити або ускладнити виявлення правильної кількості факторів. Зокрема, коли насиченість фактору низька (наприклад, через низькі факторні навантаження, обмежену кількість показників на фактор або високі міжфакторні кореляції), правильну кількість факторів стає складніше визначити, тому що власні числа, пов'язані зі справжніми факторами, числово наближені до решти власних[19].

Оскільки власні числа можна вивести з моделі загального фактору (див. рівняння 5 та 6), ми можемо аналітично визначити ефекти різних факторних моделей на очікувані власні числа. Для гіпотетичної моделі популяційного загального фактору з S факторами, N показниками на фактор, стандартизованими факторними навантаженнями l та міжфакторною

кореляцією ρ, власні числа можна вивести з матриці кореляцій B, яка випливає з цієї моделі.

$$\lambda_1 = 1 + (N-1)l^2 + (S-1)N\rho l^2$$

$$\lambda_2 = \ldots = \lambda_S = 1 + (N-1)l^2 - N\rho l^2 \quad (8)$$

$$\lambda_{S+1} = \ldots = \lambda_{NS} = 1 - l^2$$

[19]. Подібним чином, ми можемо вивести власні числа B_C, як

$$\lambda_1 = Nl^2 + (S-1)N\rho l^2$$

$$\lambda_2 = \ldots = \lambda_S = Nl^2 - N\rho l^2 \quad (9)$$

$$\lambda_{S+1} = \ldots = \lambda_{NS} = 0$$

З використанням вибірки отримані власні числа будуть відрізнятися від власних чисел генеральної популяції, які були визначені в рівняннях 8 і 9. Брекен та ван Ассен[19] показали, що дисперсія вибірки впливає на власні числа у трьох аспектах. По-перше, власні числа $\lambda_1,\ldots,\lambda_{[S/2]}$ (перша половина власних чисел, що пов'язані зі справжніми факторами) збільшуються через використання випадкових кореляцій в матриці кореляцій вибірки при виділенні факторів. По-друге, власні числа $\lambda_{[S/2]},\ldots,\lambda_S$ (друга половина власних чисел, що пов'язані зі справжніми факторами) зменшуються, оскільки перші фактори пояснюють більшу частину варіації. По-третє, перша половина залишкових власних чисел $(\lambda_{S+1},\ldots,\lambda_{[S+(N-S)/2]})$ знову збільшується. Аналогічно власним числам $\lambda_1,\ldots,\lambda_{[S/2]}$, додаткові фактори використовують залишкові випадкові кореляції, які не були враховані попередніми факторами. В результаті цього виникає більш неоднозначний набір власних чисел, оскільки різниця між λ_S і λ_{S+1} зменшується.

Загалом, це доводить, що будь-який метод екстракції, який використовує вибіркові власні числа, буде ефективним в разі наявності великої кількості індикаторів з високими навантаженнями та слабких кореляцій між факторами. У випадку, коли індикатори мають низькі навантаження, кількість індикаторів на фактор невелика та кореляції між факторами є сильними, виникають серйозні труднощі, особливо при обмеженому обсязі вибірки.

Методи визначення кількості факторів.

EFA зазвичай використовується тоді, коли немає чіткої теоретичної основи для визначення кількості латентних факторів, що лежать в основі спостережуваних змінних. В цьому розділі ми здійснимо короткий огляд традиційних та нових методів визначення кількості факторів, які слід зберегти в EFA.

2. Паралельний аналіз

У 1965 році Horn запропонував метод паралельного аналізу (PA)[20], який використовується для визначення кількості факторів, що потрібно зберегти, і ґрунтується на генерації випадкових змінних. PA порівнює власні значення кореляційної матриці, яку слід аналізувати, з власними значеннями, отриманими з некорельованих нормальних змінних. Для отримання "очікуваних" власних значень використовується процес моделювання Монте-Карло, який повторює вихідні дані за обсягом вибірки та кількістю змінних. При використанні PA у факторному аналізі процедура є практично однаковою, за винятком того, що діагональ кореляційної матриці замінюється на квадрати множинних кореляцій, які є першим кроком у наближенні загальних змінних в EFA. Раніше для визначення значущості фактору використовувалось порівняння його власного значення з середнім значенням власних значень, отриманих з некорельованих даних. В даний час рекомендується використовувати власне значення, яке відповідає певному перцентілю, такому як 95-й відсоток розподілу власних значень, отриманих з випадкових даних [21].

Різні дослідження підтверджують відповідність методу PA визначенню кількості факторів [22]. Дослідження Цвіка та Велісера (1986)[23] показали, що PA є найточнішим серед розглянутих методів, показуючи меншу варіабельність та чутливість до різних факторів. Глорфельд (1995) [21] погоджується з цією оцінкою і стверджує, що при огляді результатів функціонування різних методів мало є причин вибирати інший метод, крім PA. Декілька академічних журналів, таких як "Освітня та психологічна вимірювання" [24], підтримують цю позицію. Загалом, можна сказати, що відносно широко погоджуються з тим, що PA є найкращою доступною альтернативою для вирішення проблеми визначення кількості факторів у EFA та PCA.

Алгоритм Parallel Analysis для факторного аналізу може бути сформульований математично наступним чином:

1.Отримати вибірку даних з N спостереженнями та S змінними (або факторами).

2.Побудувати матрицю кореляцій B з NxN розміром на основі вибірки даних.

3.Згенерувати k нормально розподілених вибірок даних з N спостереженнями та S змінними. Кожну з цих вибірок далі обробити, як з оригінальну вибіркою, щоб отримати k матриць кореляцій, B_1, B_2, ..., B_k.

4.Для кожної матриці кореляцій B_i обчислити її власні значення та зберегти тільки перші h найбільших власних значень, де h - кількість факторів, яку ми хочемо порівняти з фактичною кількістю факторів в оригінальних даних.

5.У середньому з к вибірок даних згенерувати кількість критичних власних значень, які можна використати як порівняльні значення для визначення кількості факторів в оригінальних даних.

6.Порівняти перші h власних значень оригінальної матриці кореляцій B з відповідними критичними значеннями, отриманими в попередньому кроці. Якщо хоча б одне власне значення більше відповідного критичного значення, то прийняти, що кількість факторів, що зумовлюють оригінальні дані, більша за h.

Отже, алгоритм PA полягає в порівнянні перших h власних значень оригінальної матриці кореляцій з відповідними критичними значеннями, отриманими на основі випадково згенерованих даних, що мають ту ж саму кореляційну матрицю.

3. Емпіричний критерій Кайзера

Емпіричний критерій Кайзера (EKC)[19] враховує випадкові варіації власних значень у критерії Кайзера. На рівні популяції він ідентичний критерію Кайзера та виділяє всі фактори з відповідними власними значеннями кореляційної матриці, що більші за одиницю. Однак, на рівні вибірки критерій враховує розподіл власних значень для нормально розподілених даних. При нульовій моделі розподіл власних значень асимптотично підкоряється розподілу Марченко-Пастура [25], де верхня межа цього розподілу є посиланням на перше власне значення λ, яке визначається

$$\lambda_1 = (1+\sqrt{\frac{S}{N}})^2$$

Для набору даних з N спостережень та S елементів, наступні власні значення поправляються за допомогою відсотка поясненої дисперсії, що виражається у власних значеннях попередніх факторів. Посилання на j-те власне значення визначається як

$$\lambda_j = \max(\frac{S-\sum_{i-0}^{j-1}\lambda_j}{S-j+1}[1+\sqrt{\frac{S}{N}}]^2,1)$$

Отже, при вищих попередніх власних значеннях посилання на власне значення знижується, тому що відношення невиправданої дисперсії буде

меншим. Згідно з оригінальним критерієм Кайзера, посилання на власне значення не може бути менше одиниці.

Для всіх $1 \leq j \leq S$ та m загальних факторів, визначаються умови, за яких ЕКС може правильно визначити кількість факторів. Для ортогональних факторів, умови є менш складними і полягають у високих значеннях коефіцієнта альфа Кронбаха a_j та кількості спостережень N, а також у короткості шкал та низькій кореляції між факторами. У разі корельованих факторів, умови є більш складними, але з високою ймовірністю будуть виконані, якщо а та N високі, шкали коротші та кореляції факторів низькі. Дослідження Braeken та van Assen (2017)[19] підтверджують, що ЕКС показує високі показники правильності (> 0,90), якщо виконуються відповідні умови, і низькі (< 0,50), якщо умови не виконуються. Зокрема, ЕКС показала кращі результати, ніж традиційний РА з 95-м відсотковим персентилем як критерієм, коли фактори корельовані та вимірюються лише кількома елементами з дуже високими завантаженнями. Дослідження також підтвердили, що ЕКС дає порівнянні результати з відновленим РА та CD у симуляційних дослідженнях з високою кількістю факторів та декількома спостережуваними змінними.

Braeken та van Assen (2017)[19] показали, що точність ЕКС була високою тільки за умови, коли всі точності більше 0.93, і нижчою, якщо ці умови є протилежними (менше 0.83). Однак, теоретичні умови, які гарантують високу ефективність ЕКС та інших критеріїв екстракції, потребують інформації, яка доступна дослідникам тільки в разі припущення конкретної факторної структури. Це може бути неприйнятно в контексті EFA, оскільки зазвичай вона застосовується для уникнення припущень щодо підлягаючої факторної структури.

Алгоритм Емпіричного критерію Кайзера можна описати наступним чином:

1.Обчислити кореляційну матрицю між спостереженнями.

2.Обчислити власні числа цієї матриці.

3.Відсортувати власні числа в порядку спадання.

4.Визначити кількість власних чисел, більших за 1 (тобто власних чисел, які більше, ніж середнє значення власних чисел).

5.Ця кількість і є рекомендованою кількістю факторів, яку слід використовувати в аналізі.

4. Hull Method

Критерій HULL (англ. "Horn's Parallel Analysis with Minimum Average Partial" або "Horn's Parallel Analysis with Minimum Average Partial Correlation Matrix") - це один з критеріїв визначення кількості факторів у факторному аналізі з використанням методу головних компонент (PCA) або методу максимальної правдоподібності (MLE).

Критерій HULL базується на порівнянні емпіричного кореляційного міжфакторного матриці з матрицею часткових кореляцій між факторами. Він вимагає обчислення двох матриць: емпіричної кореляційної міжфакторної матриці та матриці часткових кореляцій. Для визначення кількості факторів за критерієм HULL необхідно порівняти власні значення емпіричної матриці з власними значеннями матриці часткових кореляцій.

Для цього необхідно обчислити середнє власне значення (MAVE) для емпіричної та часткової кореляційних матриць та порівняти їх. Кількість факторів визначається як кількість власних значень емпіричної матриці, що більше за відповідне середнє власне значення часткової кореляційної матриці.

Критерій HULL дозволяє визначити оптимальну кількість факторів, що може допомогти уникнути перенавантаження моделі та забезпечити більш точні результати факторного аналізу.

Важливою перевагою критерію HULL є те, що він базується на порівнянні емпіричної кореляційної матриці з матрицею часткових кореляцій,

що враховує взаємозв'язки між факторами та дозволяє визначити кількість дійсно незалежних факторів у моделі.

Однак важливо зазначити, що критерій HULL не є універсальним і не завжди може дати правильну кількість факторів. Тому рекомендується використовувати його разом з іншими критеріями та здійснювати додаткову перевірку на адекватність та робастність результатів факторного аналізу.

Алгоритм критерію HULL в EFA можна виразити наступними формулами:

1.Спочатку необхідно обчислити кореляційну матрицю між змінними у вихідних даних.

2.Далі, проводиться розклад кореляційної матриці на фактори, що дає факторну матрицю. Власні значення цієї факторної матриці позначаються як λ.

3.Для генерації випадкових даних, створюється набір даних з аналогічною кількістю змінних та спостережень. Випадкові дані можуть бути згенеровані з рівномірного або нормального розподілу.

4.З випадкових даних обчислюється кореляційна матриця та розкладається на фактори, отримуючи випадкову факторну матрицю. Власні значення випадкової факторної матриці позначаються як λ*.

5.Повторюється пункт 4 багато разів, наприклад, 100 разів, щоб отримати середнє значення власних значень випадкової факторної матриці. Середнє значення власних значень випадкової факторної матриці позначається як M*λ.

6.Далі порівнюється власне значення λ з відповідним середнім значенням випадкових власних значень Mλ. Якщо λ більше за Mλ, то фактор вважається значущим, і на його місце вибирається новий фактор. Кількість факторів збільшується, поки λ не стане меншим за M*λ.

7.Остаточна кількість значущих факторів визначається як максимальна кількість факторів, при якій λ більше за відповідне середнє значення випадкових власних значень M*λ.

Отже, алгоритм критерію HULL в EFA використовує розклад кореляційної матриці на фактори та порівнює власні значення цієї факторної матриці з власними значеннями факторної матриці, отриманої з випадкових даних. Цей процес дозволяє визначити кількість значущих факторів для подальшого аналізу.

III. Методи оцінки параметрів моделі

1. Expectation-Maximization(EM) алгоритм

Bock та Aitkin (1981)[26] запропонували метод оцінки параметрів елементів великої кількості тестів, що ґрунтується на алгоритмі Expectation-Maximization (EM), розробленому Dempster, Laird та Rubin (1977)[27].Алгоритм складається з двох основних кроків - крок очікування (Expectation) та крок максимізації (Maximization).

На кроці очікування, алгоритм обчислює теоретичні розподіли відповідей на тестові питання для кожного респондента на основі поточних оцінок параметрів моделі. Ці теоретичні розподіли потім використовуються для обчислення очікуваних значень кількості відповідей респондентів на кожне тестове питання.

На кроці максимізації, алгоритм використовує отримані на попередньому кроці очікувані значення для оновлення параметрів моделі, що максимізують функцію правдоподібності даних. Цей крок виконується за допомогою методу максимальної правдоподібності.

Кроки очікування та максимізації повторюються до тих пір, поки не буде досягнуто заданої точності або не буде досягнуто максимальної кількості ітерацій.

Оскільки алгоритм є досить складним громіздким, і особливо у політомічному випадку, обмежимось випадком, коли всі дані є дихотомічними.

Наведемо приклад відповідної системи для випадку компенсаторної 2-PL моделі. Ці формули отримано у магістерській дисертації [28].

У цьому випадку S=2,

$$p_{ij} = P(X_{ij} = 1 | \theta_{j1}, \theta_{j2}, a_{i1}, a_{i2}, d_i) = \frac{e^{a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + d_i}}{1 + e^{a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + d_i}}$$

, де $i = \overline{1, N}$; $j = \overline{1, M}$;

M – це кількість іспитників;

N – це кількість питань в тесті;

a_{i1}, a_{i2} – диференціюючі здатності ; i-го завдання; θ_{j1}, θ_{j2} – рівні підготовленості j-го іспитника; d_i – рівень складності i-го завдання.

Введемо позначення: $\Delta_{ij} = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + d_i$, тоді $p_{ij} = \frac{e^{\Delta_{ij}}}{1+e^{\Delta_{ij}}}$. Позначимо також через $q_{ij} = 1 - p_{ij} = \frac{1}{1+e^{\Delta_{ij}}}$.

$$\mathcal{L} = \prod_{i=1}^{N}\prod_{j=1}^{M} p_{ij}^{x_{ij}} \ (1-p_{ij})^{(1-x_{ij})}, x_{ij} \in \{0;1\}.$$

$$\mathcal{L}^* = ln\mathcal{L} = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij} \ln p_{ij} + \left(1 - x_{ij}\right) \ln q_{ij.}$$

$$\mathcal{L}^* = \sum_{i=1}^{N}\sum_{j=1}^{M}\left[x_{ij}\left(\Delta_{ij} - \ln\left(1 + e^{\Delta_{ij}}\right)\right) - \left(1 - x_{ij}\right)\ln\left(1 + e^{\Delta_{ij}}\right)\right] =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_{ij}\Delta_{ij} - \ln\left(1 + e^{\Delta_{ij}}\right)\right).$$

$$\begin{cases} \frac{\partial \mathcal{L}^*}{\partial a_{is}} = \sum_{i=1}^{N}\left(x_{ij}\theta_{js} - \frac{e^{\Delta_{ij}}}{1+e^{\Delta_{ij}}}\theta_{js}\right) = \sum_{i=1}^{N}\left(x_{ij} - p_{ij}\right)\theta_{js} = 0; s = 1;2, \\ \frac{\partial \mathcal{L}^*}{\partial \theta_{js}} = \sum_{j=1}^{M}\left(x_{ij} - p_{ij}\right)a_{is} = 0; s = 1;2, \\ \frac{\partial \mathcal{L}^*}{\partial d_i} = \sum_{i=1}^{N}\left(x_{ij} - p_{ij}\right) = 0\,. \end{cases}$$

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

MH-RM (Metropolis-Hastings Robbins-Monro) [9]- це алгоритм, що використовується для отримання оцінок параметрів моделі з розподілом апріорних значень в контексті теорії тестування на основі відповідей з множинним вибором (MIRT).

Алгоритм MH-RM поєднує методи Монте-Карло та ітераційної оптимізації. Він базується на згенерованому випадковому виборі параметрів з апріорного розподілу та зворотньому перетворенні Монте-Карло для оновлення оцінок параметрів моделі. Алгоритм забезпечує збіжність оцінок

параметрів до їх оптимальних значень з максимальною ймовірністю. Особливістю алгоритму є те, що він може працювати з досить складними моделями MIRT, які містять багато параметрів.

Алгоритм MH-RM є ефективним інструментом для отримання оцінок параметрів моделі з розподілом апріорних значень в контексті MIRT. Він дозволяє отримати оцінки параметрів, що максимізують правдоподібність спостережуваних даних та збігаються з оцінками, отриманими з інших методів MIRT.

Алгоритм MH-RM використовує стохастичні методи для заповнення відсутніх параметрів θ, що відрізняє його від традиційного підходу EM, де параметр θ розглядається як набір "важливих" параметрів з відомим розподілом і їх інтегрують з рівнянням правдоподібності. Після заповнення відсутніх значень, вони розглядаються як відомі, і на основі цих даних оновлюються параметри на рівні елементів за допомогою звичайних методів пошуку коренів, які використовують повну функцію логарифма правдоподібності. Цей метод заповнення не є точним або визначеним, але зазвичай дозволяє більш просте та зручне оцінювання багатовимірних інтегралів, ніж їх чисельна квадратура. MH-RM - це більш нова спроба контролювати неточності, що виникають при використанні стохастичних методів заповнення параметрів θ.

Показано, що при використанні стохастично заповнених моделей повних даних та правильному обліку похибки можна розрахувати оцінки максимальної правдоподібності та стандартні помилки параметрів. Для цього спочатку обчислюють початкові значення параметрів θ за допомогою вибірки Метрополіса-Гастінгса. Далі, за допомогою вектора градієнта та матриці Гессе, які розраховуються на основі повних даних, оновлюються початкові параметри за допомогою коректування Ньютона-Рафсона. Цей процес повторюється кілька разів для прогрівання моделі. Після цього оцінки параметрів використовуються для розрахунку приблизних початкових значень

для алгоритму MH-RM. Нарешті, останній набір оновлень параметрів здійснюється за допомогою методу, керованого алгоритмом розв'язування кореня Роббінса-Монро, який повільно збігається, коли константа зменшення приростає до нуля. Цей підхід дозволяє належним чином врахувати похибки, що виникають внаслідок вибірки Метрополіса-Гастінгса, при максимізації оцінок параметрів. Алгоритм MH-RM можна використовувати для отримання оцінок як експлораторних, так і підтверджувальних моделей відповідей на питання.

IV. Методи перевірки адекватності моделі

Оскільки алгоритми EFA можуть визначати різні розмірності моделі, то проводиться вибір найбільш адекватної моделі на підставі спеціальних статистичних критеріїв. Ми використовуємо декілька критеріїв оцінки придатності моделі, які використовуються в MIRT для оцінки якості відповідності моделі спостережуваним даним, а саме, RMSEA, TLI, AIC, BIC, CFI. Використання декількох критеріїв оцінки придатності моделі може допомогти зробити більш об'єктивну оцінку адекватності моделі MIRT.

Критерій TLI

Критерій TLI(Tucker-Lewis Index) [29] в CFA - це критерій, який використовується для оцінки адекватності моделі факторного аналізу відносно базової моделі нульової кореляції. Він вимірює відстань між матрицею спостережень та матрицею, що побудована на основі моделі нульової кореляції, у порівнянні з відстанню між матрицею спостережень та матрицею, що побудована на основі розглядуваної моделі факторного аналізу. Чим ближче значення TLI до 1, тим краще модель відповідає даним.

Для розрахунку TLI потрібно визначити відношення між підігнаною моделлю та базовою моделлю, яка встановлює випадкові зв'язки між змінними. Значення TLI може варіюватися від 0 до 1, де більш високі значення вказують на кращу відповідність моделі. проаналізувати результати оцінки адекватності моделі, використовуючи індекс TLI можна наступним чином. Якщо значення TLI близьке до 1, це вказує на добру відповідність моделі. Якщо значення TLI менше 0,9, то модель слід переглянути.

Індекс перевірки адекватності моделі TLI визначається як відношення залишкової коваріації між спостережуваними змінними в моделі та залишкової коваріації в базовій моделі.

Статистика критерію TLI виглядає наступним чином:

$$TLI = \frac{\frac{\chi^2}{k} - df}{\frac{\chi^2}{k} - df + \frac{\chi_0^2}{k_0} - df_0},$$

де:

- χ^2 - статистика χ^2-квадрат тесту для перевірки адекватності моделі;
- df - кількість ступенів свободи для моделі;
- k - кількість параметрів моделі;
- χ_0^2 - статистика x_i-квадрат тесту для базової моделі;
- df_0 - кількість ступенів свободи для базової моделі;
- k_0 - кількість параметрів в базовій моделі.

Критерій CFI

Критерій перевірки адекватності моделі CFI (Comparative Fit Index)[29] в MIRT є одним із показників, що використовується для оцінки адекватності моделі вимірювання. Він вимірює ступінь відповідності між спостережуваними даними і передбаченими даними, отриманими з моделі.

Значення CFI знаходиться в діапазоні від 0 до 1, де значення ближче до 1 вказує на кращу адекватність моделі. Зазвичай значення CFI більше 0.95 вважається показником доброї адекватності моделі.

CFI порівнює покращення у придатності моделі з базовою моделлю, де нульова модель повністю не враховує залежності між змінними. Високе значення CFI вказує на те, що модель показує кращу підгонку даних, ніж нульова модель.

Індекс перевірки адекватності моделі CFI в MIRT математично визначається шляхом порівняння моделі, яку ви досліджуєте, з нульовою моделлю, яка не враховує залежності між змінними.

Для обчислення CFI потрібно враховувати дві величини:

1. Значення χ^2 (chi-square) для моделі, яку ви досліджуєте.
2. Значення χ^2 для нульової моделі.

Статистика CFI визначається наступним чином:

$$CFI = \frac{(\chi_0^2 - \chi_{модель}^2)}{\chi_0^2},$$

де:

- $\chi_{модель}^2$ - значення χ^2 для моделі, яку ви досліджуєте.
- χ_0^2 - значення χ^2 для нульової моделі.

Критерій RMSEA

Індекс перевірки адекватності моделі RMSEA (Root Mean Square Error of Approximation) [30] в MIRT використовується для оцінки відповідності моделі факторного аналізу з мішаними змінними. Він враховує якість підгонки моделі до спостережуваних даних, зокрема, враховується помилка апроксимації (розброс даних, який не пояснюється моделлю).

Індекс RMSEA обчислюється на основі розбивки даних на кілька компонентів, зокрема на апроксимовані даними (модель) і непроявленими складовими (нев'язка). Чим менше значення індексу RMSEA, тим краще модель відповідає даним. Зазвичай приймається, що значення RMSEA менше 0,05 вказує на добру відповідність моделі даним, а значення між 0,05 і 0,08 вказує на помірну відповідність.

Цей індекс є одним з найпоширеніших індексів перевірки адекватності моделі в MIRT, оскільки він забезпечує оцінку якості підгонки моделі з урахуванням нев'язки апроксимації даних.

Індекс перевірки адекватності моделі RMSEA в MIRT визначається на основі співвідношення між розбросом даних, який не пояснюється моделлю, і загальним розбросом даних.

Статистика RMSEA визначається наступним чином:

$$RMSEA = \sqrt{\frac{\chi^2 - df}{N * (df - 1)}}$$

де

- χ^2 - статистика χ^2-квадрат, яка використовується для оцінки відхилення моделі від даних,
- df - ступені свободи, що відображають складність моделі,
- N - розмір вибірки.

Значення RMSEA менше 0,05 зазвичай вказує на добру відповідність моделі даним, а значення між 0,05 і 0,08 вказує на помірну відповідність. Чим менше значення RMSEA, тим краще модель підходить до даних.

Критерій AIC

Індекс AIC (Akaike's Information Criterion) [31] є мірою, яка використовується для оцінки адекватності моделі в контексті MIRT. Цей індекс базується на принципі компромісу між точністю моделі та її складністю.

Індекс AIC використовується для порівняння різних моделей MIRT і вибору найбільш адекватної моделі. Зазвичай, модель з найменшим значенням AIC вважається кращою, оскільки вона надає найкращий компроміс між точністю і складністю моделі.

Важливо зауважити, що для правильного використання індексу AIC в MIRT необхідно порівнювати моделі з однаковим набором даних і однаковими параметрами оцінювання, оскільки значення AIC може варіюватися в залежності від цих факторів.

Статистика AIC визначається математично за допомогою наступної формули:

$$AIC = -2 * \log(l) + 2 * k$$

Де

- $\log(l)$-логарифм максимальної функції правдоподібності відображає якість підгонки моделі до даних;
- k-кількість параметрів моделі показує загальну кількість параметрів, що оцінюються в рамках моделі.

Критерій BIC

Критерій перевірки адекватності моделі BIC[32] (Bayesian Information Criterion) в MIRT є статистичним показником, що використовується для оцінки якості підгонки моделі до даних. Він базується на ідей байєсівського підходу і враховує як точність підгонки моделі до даних, так і складність моделі.

Статистика перевірки адекватності моделі BIC визначається наступним чином:

$$AIC = -2 * \log(l) + k * \ln(N)$$

де:

- $\log(l)$-логарифм максимальної функції правдоподібності відображає якість підгонки моделі до даних;
- k - кількість параметрів моделі;
- N - обсяг вибірки (кількість спостережень).

Зазвичай менше значення BIC вказує на кращу адекватність моделі до даних.

Критерій M2

Алгоритм M2[9] полягає в порівнянні значень двох функцій витрат: функції витрат для базової моделі та для альтернативної моделі. Базова модель в цьому випадку є початковою моделлю, яку було застосовано для оцінювання параметрів. Альтернативна модель може бути моделлю з іншими параметрами або іншою моделлю, яка описує дані тесту краще за базову модель.

Для порівняння двох моделей алгоритм M2 використовує статистику, яка розраховується за допомогою різниці між двома функціями витрат та їх кількістю ступенів свободи. Якщо значення статистики тесту перевищує критичне значення з таблиці розподілу Хі-квадрат з відповідними ступенями свободи та певним рівнем значущості, то можна стверджувати, що альтернативна модель є кращою за базову модель.

При використанні алгоритму M2 у теорії тестів MIRT, альтернативна модель може бути моделлю з різними параметрами для одного з факторів (наприклад, з різним числом факторів або різними матрицями навантажень), або може бути іншою моделлю IRT з різними параметрами.

У алгоритмі перевірки адекватності M2, який використовується у теорії тестів MIRT, використовується статистика тесту, яка називається "розподіленим тестом логарифмічної шансовості" (distributed likelihood ratio test). Ця статистика дозволяє порівняти дві моделі та визначити, яка з них краще пояснює дані.

Розподілений тест логарифмічної шансовості базується на різниці логарифмічних шансовостей між двома моделями. Для порівняння базової моделі (наприклад, моделі з одним чинником) з альтернативною моделлю (наприклад, моделлю з двома чинниками) розраховується логарифмічна шансовість для кожної з цих моделей. Різниця між цими логарифмічними шансовостями обчислюється і використовується для порівняння адекватності моделей.

Функція витрат, яка використовується в алгоритмі перевірки адекватності M2 у теорії тестів MIRT, має наступний вигляд:

$$Q = -2\sum_{j=1}^{J}\sum_{k=1}^{K_j}\sum_{i=1}^{N_j}\omega_{jki}\log\frac{p_{jki}}{\hat{p}_{jki}} + \sum_{j=1}^{J}\sum_{k=1}^{K_j}\log\frac{\hat{p}_{jki}}{1-\hat{p}_{jki}}$$

Де

ω_{jki} - ваговий коефіцієнт для відповіді k на питання j від учасника i;

p_{jki} - теоритична ймовірність відповіді k на питання j від учасника i згідно з альтернативною моделлю;

$\hat{p}_{jki}$ - теоритична ймовірність відповіді k на питання j від учасника i згідно з базовою моделлю;

j-кількість питань;

K_j - кількість варіантів відповіді на питання j;

N_j - кількість учасників, які відповіли на питання j;

V. Статистичний аналіз результатів КР

1.Розроблена у магістерській дисертації методика була застосована до аналізу контрольної роботи з дисципліни «математичний аналіз-1» для бакалаврів 1-го курсу на РТФ. Об'єм вибірки склав – 79 студентів. Кількість завдань-20.

2. Було сформовано два набори даних. Перший – безпосередньо результати тестування(змішаний дихотомічні і політомічні), другий – повністю дихотомічний, сформований шляхом розщеплення політомічних питань типу «вбудовані відповіді» на дихотомічні.

3. У якості засобу обробки даних було обрано статистичного програмування R [33]. Зокрема, пакети EFA.dimension та MIRT.

4. Було проведено EFA за допомогою алгоритмів PA, EKC, HULL, які реалізовано у функції DIMTEST пакету EFA.dimension. Результати наведено у таблиці 1.

Таблиця 1. *Результати EFA*

	Політомія	Дихотомія
PA	2	3
EKC	2	2
HULL	1	1

5. Для всіх запропонованих розмірностей було проведено оцінку параметрів за допомогою функції mirt пакету MIRT. Результати оцінювання наведено у таблицях (2-6).

Таблиця 2. Р*езультати оцінювання GPCM моделі S=1*

№ пит	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

анн я																		
a1	0,283	0,847	4,927	1,071	5,286	1,811	1,13	0,61	1,047	1,083	0,572	0,557	0,603	0,443	0,382	0,548	0,295	0,485
ak0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ak1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ak2	2			2	2			2	2	2	2	2	2	2	2	2	2	2
ak3								3	3	3	3	3		3	3	3	3	3
ak4								4		4	4			4	4	4	4	4
ak5											5			5	5		5	5
ak6														6	6		6	6
ak7															7		7	
ak8															8			
ak9															9			
d0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d1	2,799	1,992	9,25	0,126	8,674	2,221	3,09	1,57	2,804	2,288	1,33	1,338	1,426	1,219	0,967	0,707	-0,12	0,082
d2	3,844			3,664	15,089			1,638	5,338	4,377	0,811	0,516	2,99	1,871	2,005	-0,122	0,113	-0,095
d3								0,898	4,129	5,633	3,52	-0,11		3,411	1,432	0,535	1,389	0,234
d4								0,29		4,4	3,782			3,164	2,24	1,61	1,216	0,383
d5											2,767			4,029	2,452		1,37	1,329
d6														4,65	1,107		2,709	-0,397

d7															2,16		2,416	
d8															2,714			
d9															3,055			

Таблиця 3. *Результати оцінювання GPCM моделі S=2*

№ питання	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a1	-0,178	-0,482	-5,659	-0,937	-7,157	-1,5355	-0,665	-0,545	-1,124	-0,97	0,53	-0,615	-0,608	-0,392	-0,356	-1,667	-0,42	-0,606
a2	0,298	1,097	8,956	1,227	1,78	0,845	1,263	0,312	1,939	0,402	0,188	-0,079	0,079	0,201	0,075	-0,705	-0,103	0
ak0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ak1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ak2	2			2	2			2	2	2	2	2	2	2	2	2	2	2
ak3								3	3	3	3	3		3	3	3	3	3
ak4								4		4	4			4	4	4	4	4
ak5											5			5	5		5	5
ak6														6	6		6	6
ak7															7		7	
ak8															8			
ak9															9			
d0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

d1	2,854	2,204	19,813	0,927	11,992	2,188	3,392	1,6	5,616	2,28	1,328	1,388	1,438	1,248	0,941	2,112	0,086	0,224
d2	3,909			4,699	20,693			1,673	9,25	4,344	0,804	0,569	3,011	1,913	1,958	2,07	0,497	0,159
d3								0,93	7,435	5,589	3,511	-0,116		3,456	1,373	2,987	1,917	0,562
d4								0,311		4,385	3,777			3,208	2,175	3,563	1,851	0,728
d5											2,772			4,072	2,386		2,063	1,621
d6														4,695	1,045		3,399	-0,236
d7															2,105		3,032	
d8															2,67			
d9															3,028			

Таблиця 4. *Результати оцінювання 2PL моделі S=1*

№ питання	a1	d
1	0,528	1,382
2	1,432	2,81
3	1,848	3,931

4	1,414	2,795
5	1,673	3,339
6	25,703	39,786
7	0,531	1,671
8	-0,007	-0,568
9	3,268	-4,616
10	0,866	-2,18
11	1,018	2,266
12	1,285	-1,867
13	0,813	1,475
14	2,315	-2,775
15	1,279	2,221
16	0,617	1,276
17	0,396	1,35
18	0,161	2,385
19	0,639	-0,526
20	-0,021	-0,857
21	0,102	1,746
22	0,659	1,562
23	0,454	-1,815
24	-0,113	-1,444
25	-1,587	4,101

Таблиця 5. *Результати оцінювання 2PL моделі S=2*

№ питання	a1	a2	d
1	-0,5	-0,204	1,418
2	-1,687	0,989	3,317
3	-81,783	-69,506	161,462

4	-65,231	-56,493	115,092
5	-1,91	0,992	3,89
6	-69,555	13,351	113,028
7	-0,618	0,597	1,799
8	0,149	-0,476	-0,579
9	-3,829	-3,654	-6,446
10	-16,774	-119,129	-119,702
11	-1,031	-0,256	2,353
12	-0,879	-0,659	-1,688
13	-0,554	-0,535	1,512
14	-2,274	-0,525	-2,704
15	-1,976	1,702	3,306
16	-0,845	0,65	1,443
17	-0,145	-0,51	1,418
18	21,604	-75,938	115,571
19	-0,508	-0,203	-0,481
20	0,082	-0,362	-0,867
21	0,027	-0,23	1,774
22	-0,591	-0,572	1,68
23	-0,694	0,711	-2,078
24	0,012	0,028	-1,443
25	2,88	0	5,708

Таблиця 6. *Результати оцінювання 2PL моделі S=3*

№ питання	a1	a2	a3	d
1	-0,656	-0,016	-0,36	1,413
2	-1,836	1,514	0,562	3,494
3	-46,229	-33,045	-3,303	86,717

4	-76,829	-84,556	-21,219	134,799
5	-2,132	1,371	0,336	3,972
6	-60,026	9,75	29,185	96,525
7	-0,741	0,499	0,582	1,814
8	0,349	-1,533	0,883	-0,889
9	-60,524	-73,353	-52,913	-128,653
10	-0,118	-54,623	-36,501	-70,496
11	-1,652	-1,112	1,172	3,097
12	-1,078	-0,638	-0,054	-1,94
13	-0,499	-0,901	0,372	1,524
14	-12,19	-0,01	-1,741	-11,984
15	-52,591	16,717	66,14	78,177
16	-0,75	0,619	0,387	1,351
17	-0,375	-0,22	-0,477	1,402
18	10,857	-55,62	-44,224	99,478
19	-0,643	0,016	-0,434	-0,573
20	0,104	-0,452	-0,033	-0,892
21	0,077	15,318	-37,773	44,647
22	-0,828	0,262	-2,049	2,397
23	-26,208	66,345	-53,194	-84,635
24	-0,026	0,173	0	-1,448
25	1,61	0	0	4,186

6. Після було проведено перевірку адекватності усіх моделей за допомогою критерієв : AIC, BIC, RMSEA, CFI,TLI. Результати наведено у таблиці 7.

Таблиця 7. *Результати перевірки адекватності моделей*

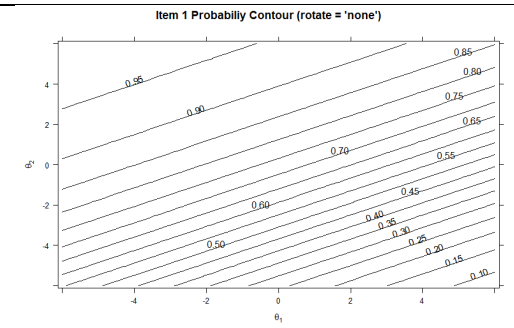
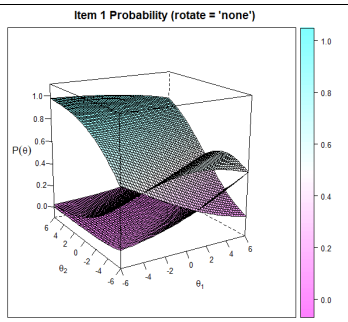
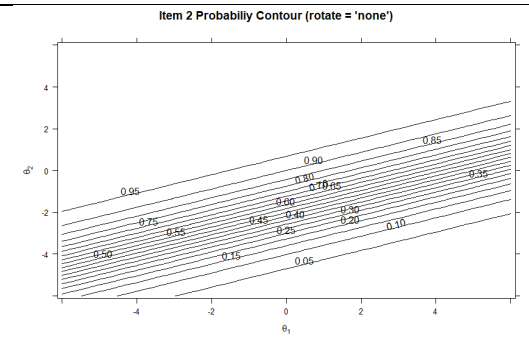
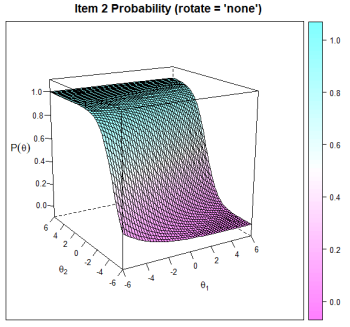
	AIC	BIC	M2	RMSEA	CFI	TLI

GPCM1	829,8795	899,4252	118,1731	0,04008931	1	1,093184
GPCM2	832,3751	834,7619	97,79429	0,02838533	1	1,128016
2PL1	1043,206	1135,713	287,3868	0,06635958	0,8804574	0,8695899
2PL2	1045,727	1122,638	265,8541	0,05472845	0,9815373	0,9924695
2PL3	1043,782	1223,247	290,4729	0,06386327	0,9578431	0,9742875

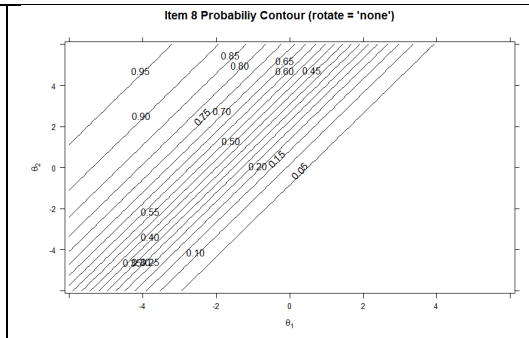
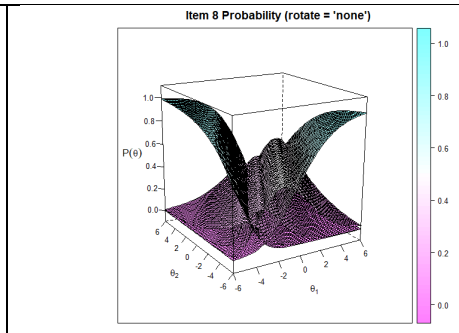
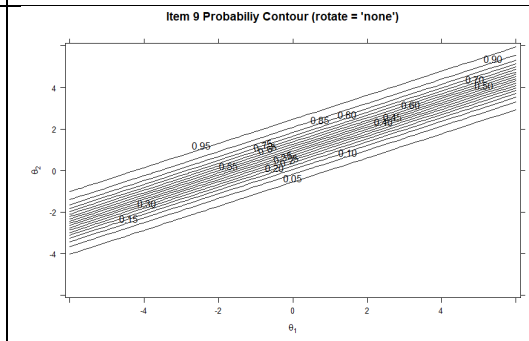
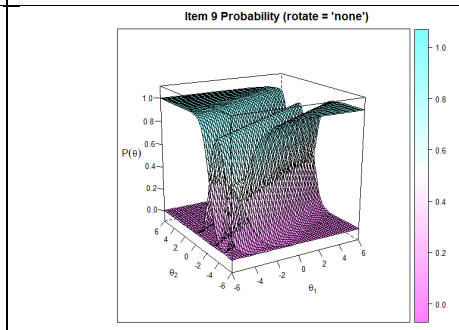
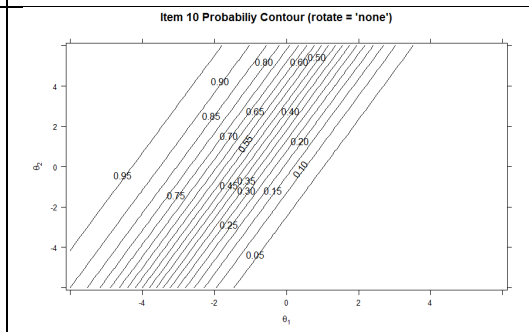
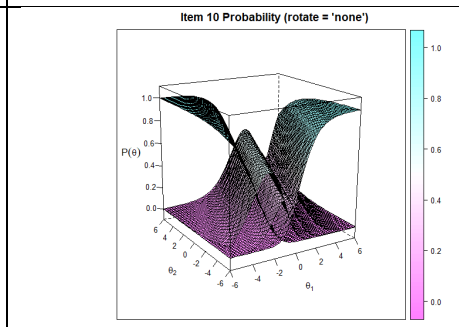
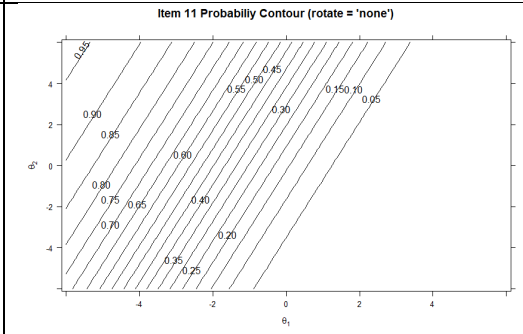
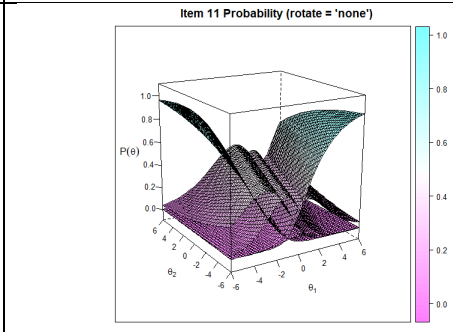
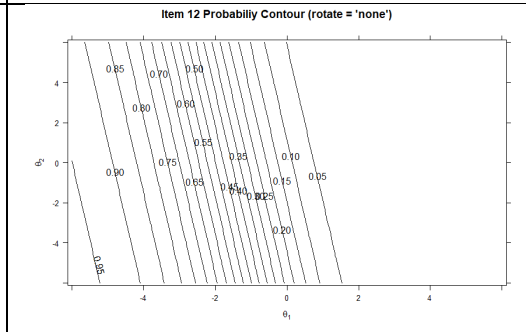
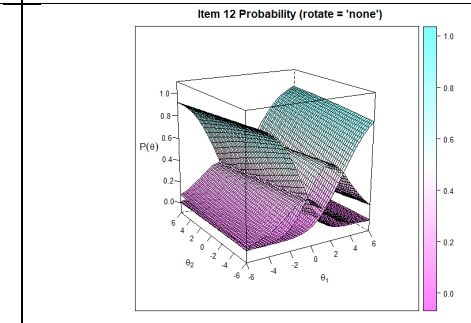
Як видно з таблиці, для політомічних даних найкращою є модель розмірності 2. Для дихотомічних даних також – розмірності 2.

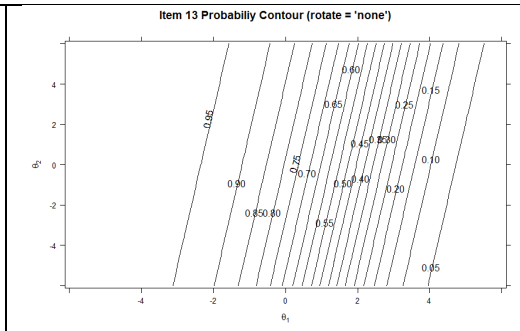
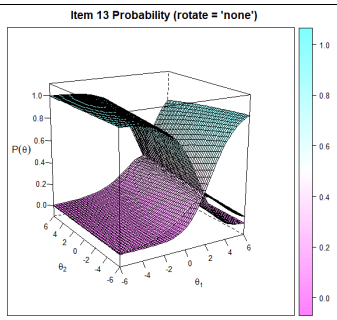
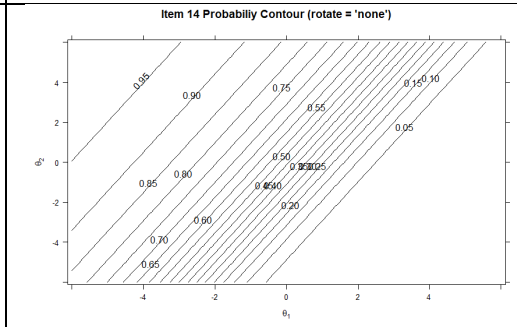
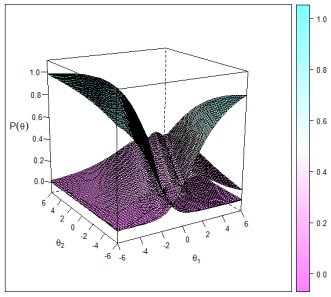
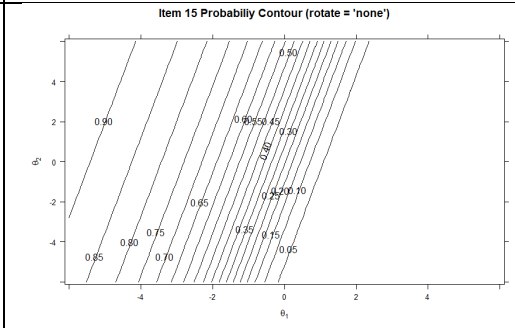
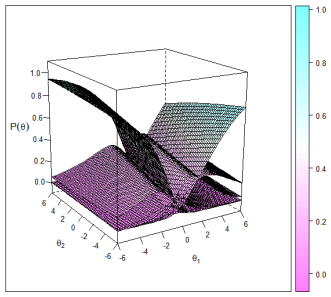
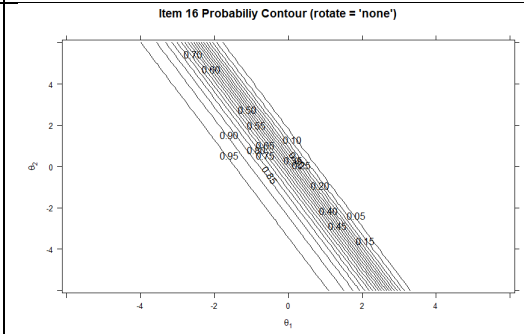
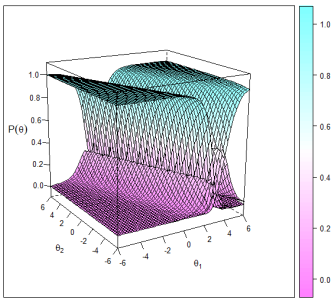
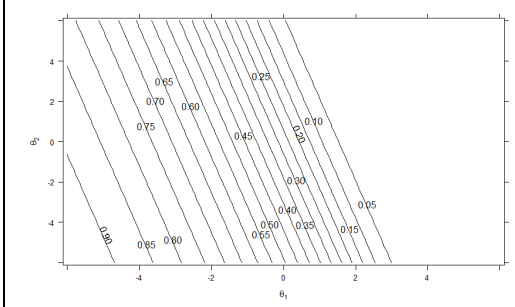
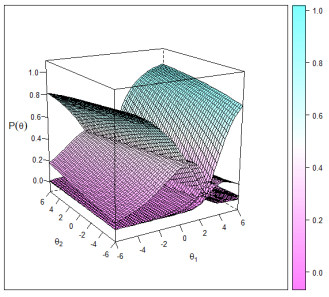
7. Для обраних моделей наведено оцінені параметри у таблицях і відповідні графіки. Таблиці 8-9

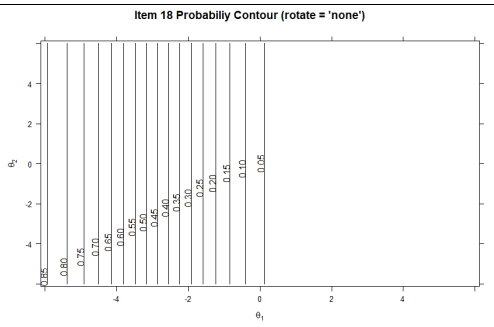
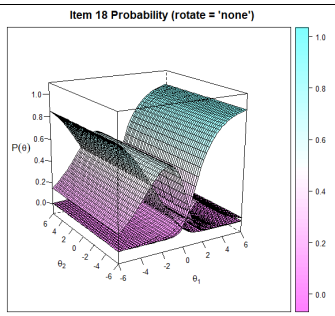
Таблиця 8. *Характеристичні поверхні та контурні рівні для завданнь політомічної моделі*

№ питання	Контурні рівні	Характеристичні поверхні
1		
2		

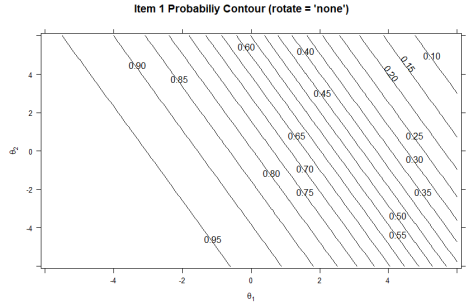
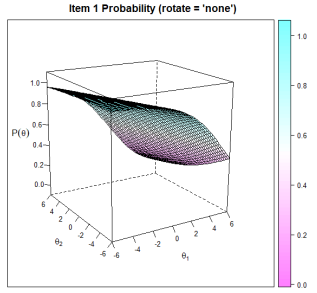
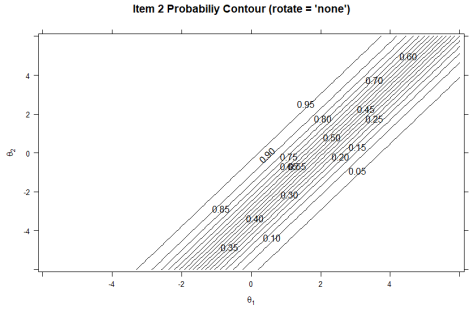
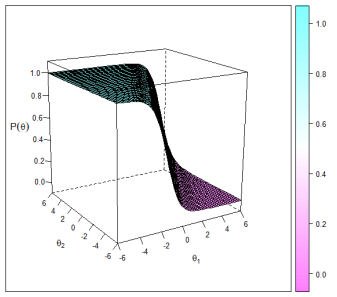
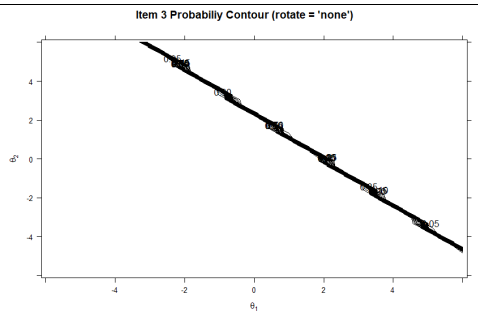
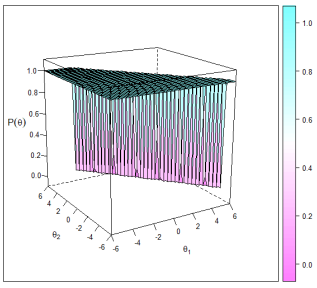
3	Item 3 Probabiliy Contour (rotate = 'none')	Item 3 Probability (rotate = 'none')
4	Item 4 Probabiliy Contour (rotate = 'none')	Item 4 Probability (rotate = 'none')
5	Item 5 Probabiliy Contour (rotate = 'none')	Item 5 Probability (rotate = 'none')
6	Item 6 Probabiliy Contour (rotate = 'none')	Item 6 Probability (rotate = 'none')
7	Item 7 Probabiliy Contour (rotate = 'none')	Item 7 Probability (rotate = 'none')

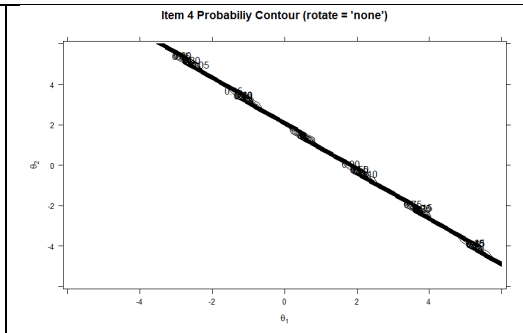
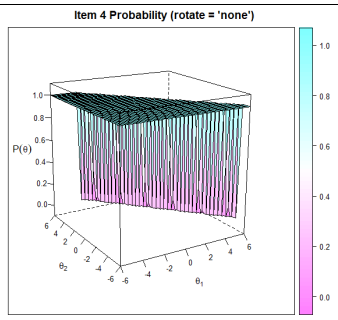
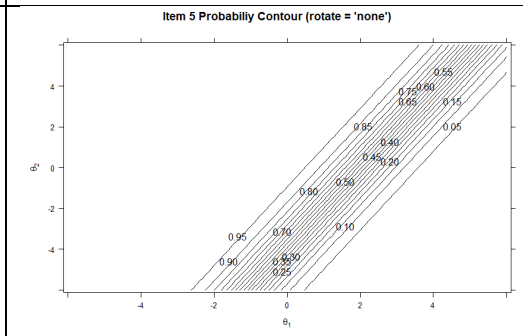
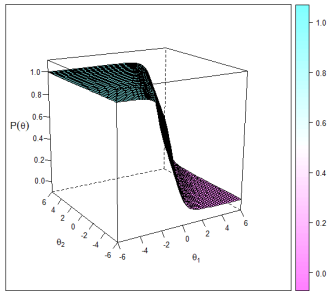
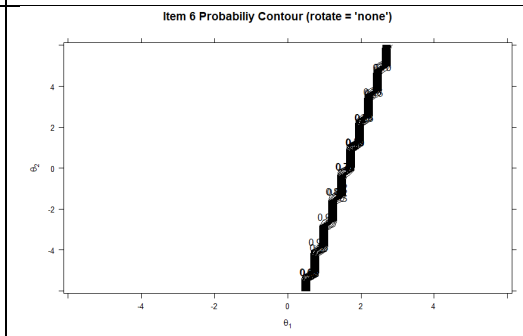
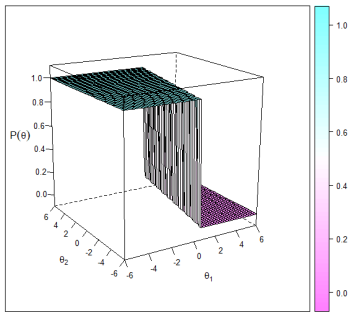
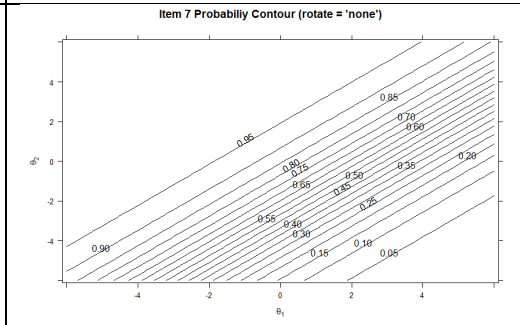
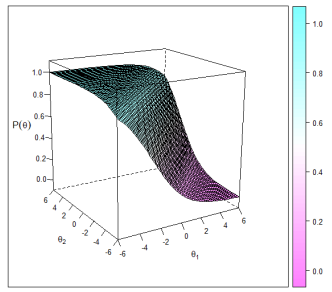
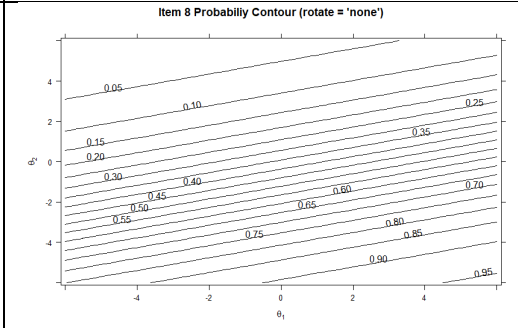
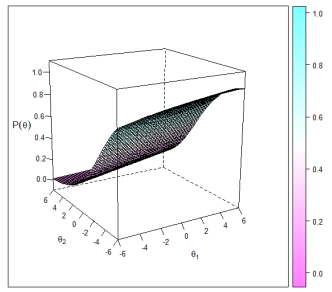
8	 Item 8 Probabiliy Contour (rotate = 'none')	 Item 8 Probability (rotate = 'none')
9	 Item 9 Probabiliy Contour (rotate = 'none')	 Item 9 Probability (rotate = 'none')
10	 Item 10 Probabiliy Contour (rotate = 'none')	 Item 10 Probability (rotate = 'none')
11	 Item 11 Probabiliy Contour (rotate = 'none')	 Item 11 Probability (rotate = 'none')
12	 Item 12 Probabiliy Contour (rotate = 'none')	 Item 12 Probability (rotate = 'none')

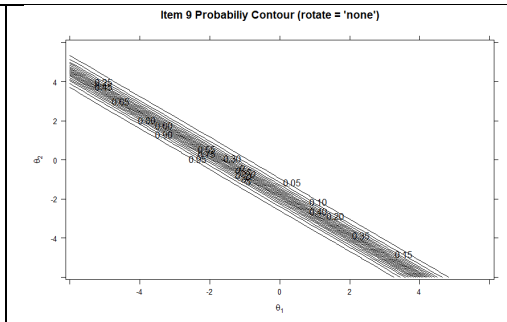
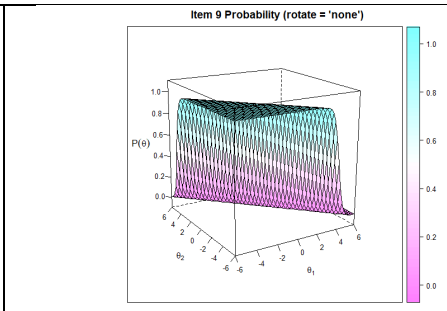
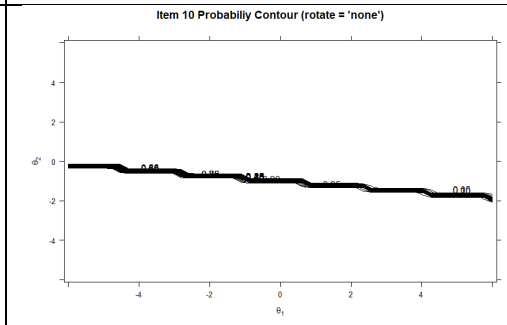
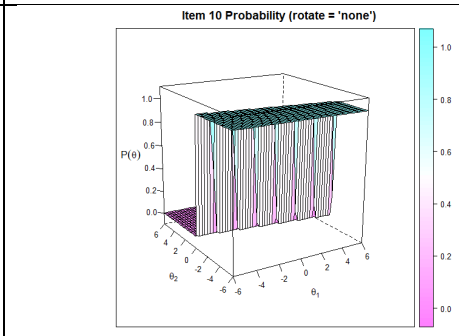
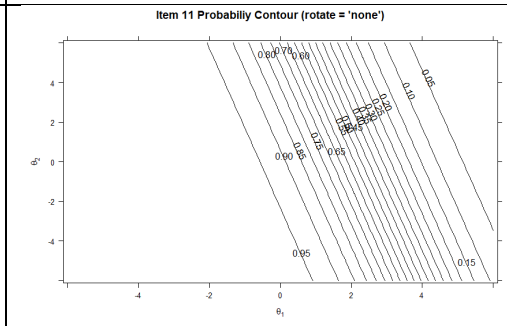
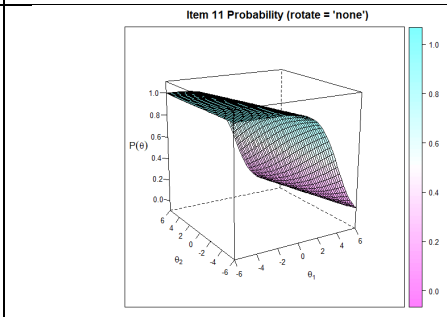
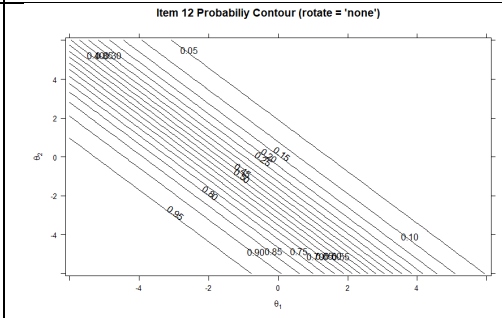
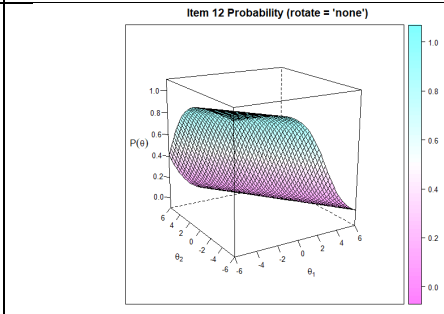
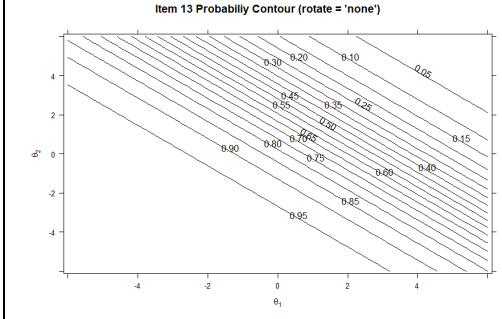
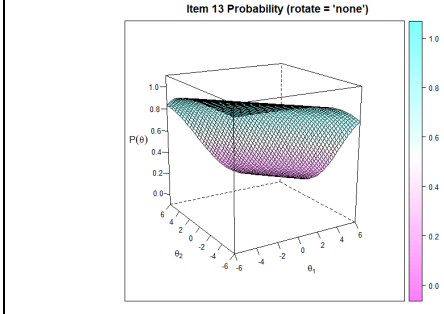
13	 Item 13 Probabiliy Contour (rotate = 'none')	 Item 13 Probability (rotate = 'none')
14	 Item 14 Probabiliy Contour (rotate = 'none')	 Item 14 Probability (rotate = 'none')
15	 Item 15 Probabiliy Contour (rotate = 'none')	 Item 15 Probability (rotate = 'none')
16	 Item 16 Probabiliy Contour (rotate = 'none')	 Item 16 Probability (rotate = 'none')
17	 Item 17 Probabiliy Contour (rotate = 'none')	 Item 17 Probability (rotate = 'none')

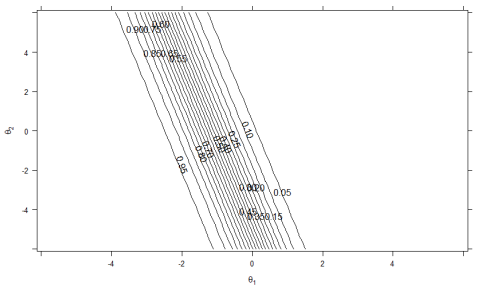
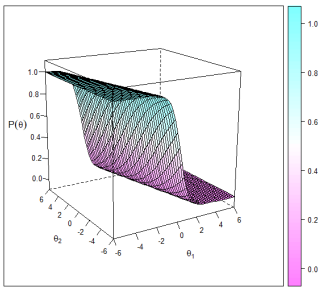
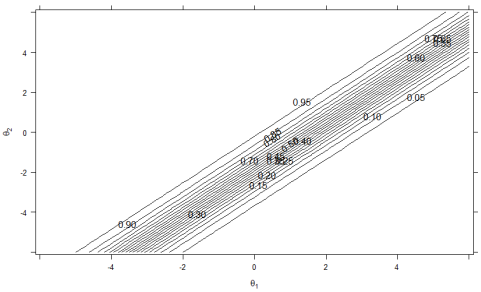
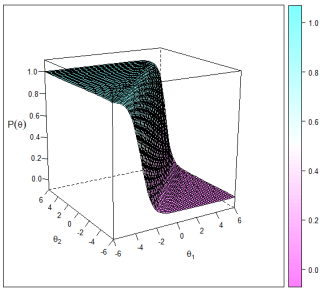
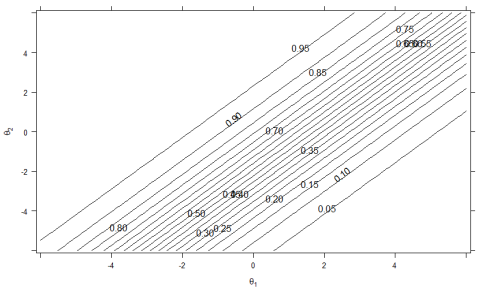
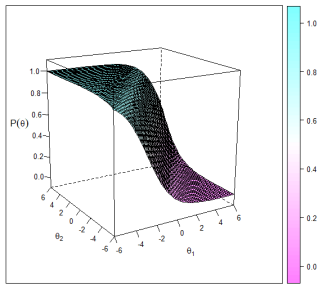
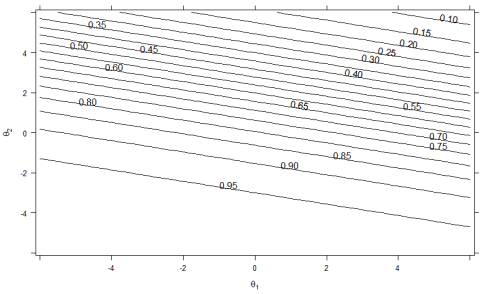
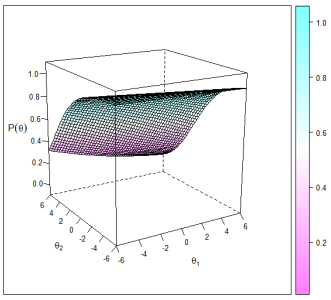
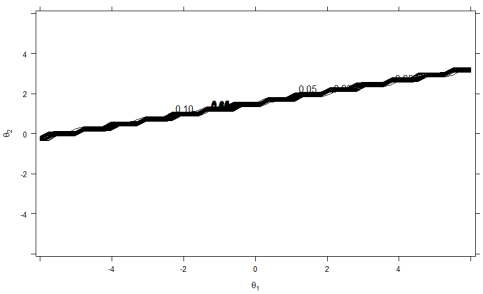
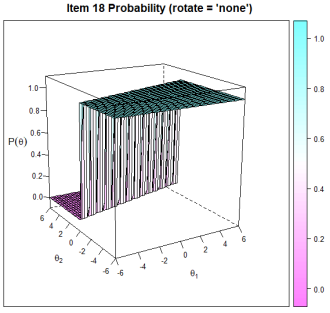
18		

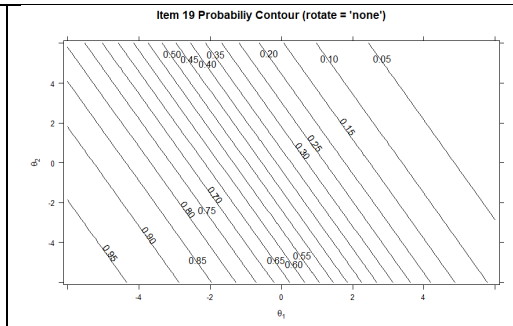
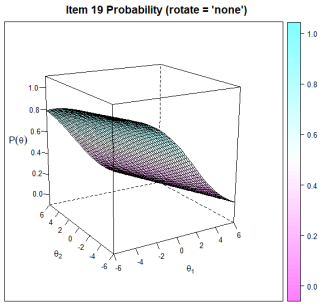
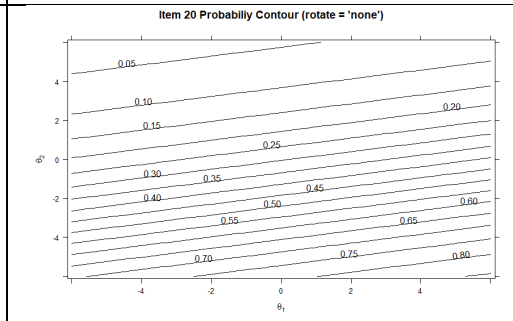
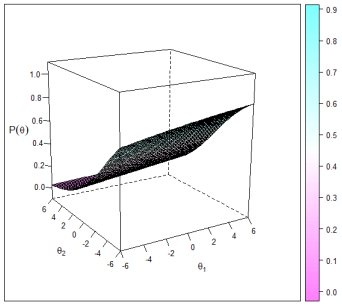
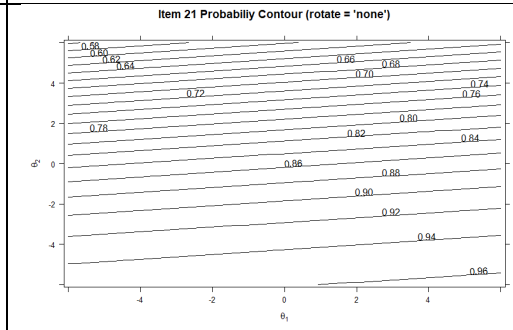
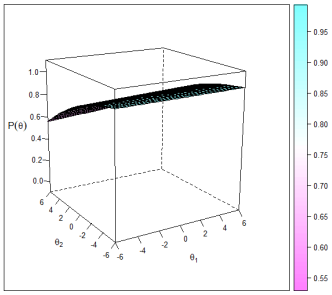
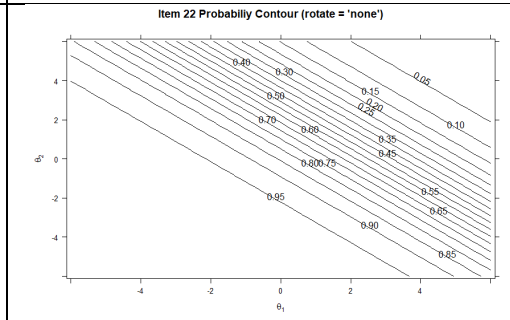
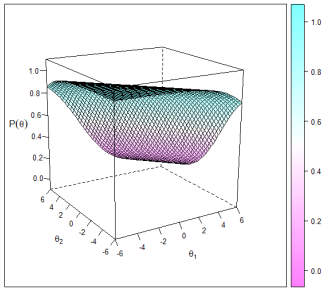
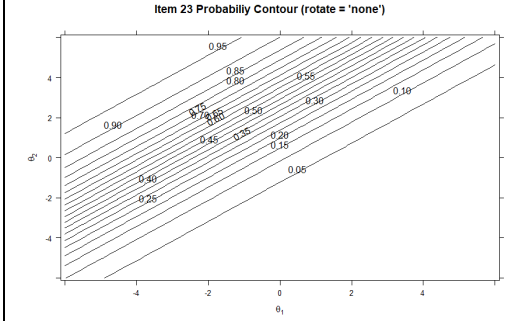
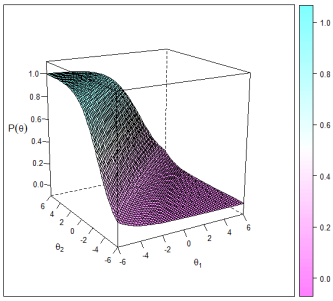
Таблиця 9. *Характеристичні поверхні та контурні рівні для завдання дихотомічної моделі*

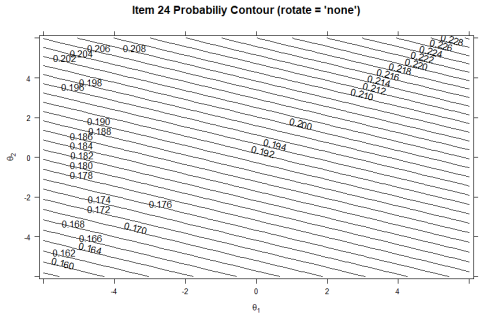
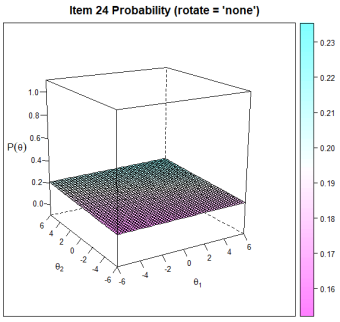
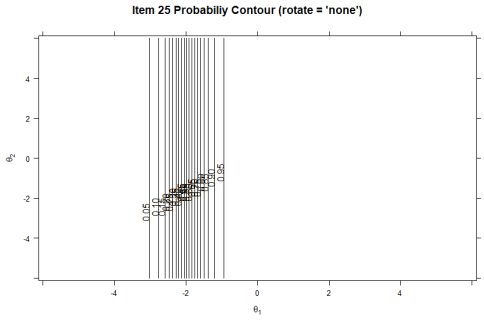
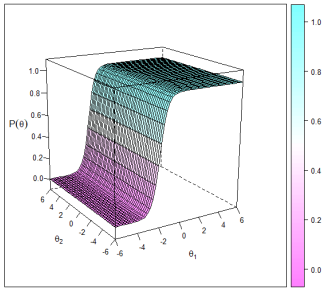
№ питання	Контурні рівні	Характеристичні поверхні
1		
2		
3		

4	 Item 4 Probabiliy Contour (rotate = 'none')	 Item 4 Probability (rotate = 'none')
5	 Item 5 Probabiliy Contour (rotate = 'none')	 Item 5 Probability (rotate = 'none')
6	 Item 6 Probabiliy Contour (rotate = 'none')	 Item 6 Probability (rotate = 'none')
7	 Item 7 Probabiliy Contour (rotate = 'none')	 Item 7 Probability (rotate = 'none')
8	 Item 8 Probabiliy Contour (rotate = 'none')	 Item 8 Probability (rotate = 'none')

9	 Item 9 Probabiliy Contour (rotate = 'none')	 Item 9 Probability (rotate = 'none')
10	 Item 10 Probabiliy Contour (rotate = 'none')	 Item 10 Probability (rotate = 'none')
11	 Item 11 Probabiliy Contour (rotate = 'none')	 Item 11 Probability (rotate = 'none')
12	 Item 12 Probabiliy Contour (rotate = 'none')	 Item 12 Probability (rotate = 'none')
13	 Item 13 Probabiliy Contour (rotate = 'none')	 Item 13 Probability (rotate = 'none')

14	Item 14 Probabiliy Contour (rotate = 'none') 	Item 14 Probability (rotate = 'none') 
15	Item 15 Probabiliy Contour (rotate = 'none') 	Item 15 Probability (rotate = 'none') 
16	Item 16 Probabiliy Contour (rotate = 'none') 	Item 16 Probability (rotate = 'none') 
17	Item 17 Probabiliy Contour (rotate = 'none') 	Item 17 Probability (rotate = 'none') 
18	Item 18 Probabiliy Contour (rotate = 'none') 	Item 18 Probability (rotate = 'none') 

19	 Item 19 Probabiliy Contour (rotate = 'none')	 Item 19 Probability (rotate = 'none')
20	 Item 20 Probabiliy Contour (rotate = 'none')	 Item 20 Probability (rotate = 'none')
21	 Item 21 Probabiliy Contour (rotate = 'none')	 Item 21 Probability (rotate = 'none')
22	 Item 22 Probabiliy Contour (rotate = 'none')	 Item 22 Probability (rotate = 'none')
23	 Item 23 Probabiliy Contour (rotate = 'none')	 Item 23 Probability (rotate = 'none')

24	 Item 24 Probabiliy Contour (rotate = 'none')	 Item 24 Probability (rotate = 'none')
25	 Item 25 Probabiliy Contour (rotate = 'none')	 Item 25 Probability (rotate = 'none')

ВИСНОВКИ

У даній магістерській дисертації розроблено методику вибору найкращої розмірності моделі MIRT для аналізу тестів з вищої математики.

Основні наукові та практичні результати роботи полягають у наступному:

1. Досліджено основні одновимірні (IRT) та багатовимірні (MIRT) моделі, які можуть бути застосовані для аналізу тестів з вищої математики.
2. Досліджено методи EFA, такі як PA, EKC, HULL, які дозволяють зробити первинний вибір розмірності моделі.
3. Обрано алгоритми EM та MH-RM для оцінки параметрів моделей.
4. Відібрано критерії перевірки адекватності моделей IRT та MIRT, такі як: AIC, BIC, RMSEA, CFI,TLI.
5. В якості засобу реалізації відповідних алгоритмів обрано мову статистичного програмування R.
6. Розроблену методику застосовано для аналізу контрольної роботи з вищої математики бакалаврів першого курсу РТФ.
7. На підставі проведеного EFA було обрано можливі вимірності моделей: для політомічної -1,2; для дихотомічної -1,2,3.
8. Проведено оцінювання латентних параметрів моделей та перевірка побудованих моделей на адекватність даним.
9. Показано, що для політомічних і дихотомічних даних «найкращою» модель розмірності 2.
10. Для обраних моделей проведено аналіз властивості завдань.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Crocker L., Algina J. Introduction to Classical and Modern Test Theory.—Belmont,CA:Wadsworth, 2006.—527 С.
2. Reckase M. D. Multidimensional item response theory.—New York:Springer, 2009.—354 С.
3. Mahmood Ul H., Frank M. Discrimination with unidimensional and multidimensional item response theory models for educational dataCommunications in Statistics // Simulation and Computation. —— 2019.—С. 1–21.
4. Круглова Н. В., Диховичний О. О. Добір математичної моделі для аналізу тестових завдань типу «вбудовані відповіді» з математичних дисциплін// Інформаційні технології і засоби навчання. 2022. —87 № 1. —— С. 166–184.
5. Kruglova N., Dykhovychnyi O. Choosing MIRT Model for Analysis of Quality of Pedagogical and Psychological Tests // 2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC). 2022.——С. 1–4.
6. Hambleton R.K., Swaminathan H., Rogers H. J. Fundamentals of Item Response Theory. – Newbury Park, CA: Sage, 1991. –175 С.
7. Моделі та методи сучасної теорії тестів: [навчально-методичний посібник] / Т.В. Лісова. – Ніжин: Видавець ПП Лисенко М.М., 2012. 112 С.
8. Диховичний. О. О. Комплексна методика аналізу якості тестів з вищої математики / О.О.Диховичний, А.Ф.Дудко // Науковий часопис НПУ імені М.П. Драгоманова. Серія №2. Комп’ютерно-орієнтовані системи навчання: Зб. наук. праць / Редрада. – К.: НПУ імені М.П. Драгоманова, 2015. – № 15 (22). – С. 140-144.
9. R. Philip Chalmers. Mirt: A Multidimensional Item Response Theory Package for the R Environment

10. E. Muraki, and B. Carlson, Full-information factor analysis for polytomous item responses. Applied Psychological Measurement. Vol. 19, N1, 1995. C. 73–90.

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. McDonald, R. P. The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 1981. C. 100-117.

13. 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.

14. Stout, W. F. A nonparametric approach to assessing latent trait unidimensionality. Psychometrika, 52, 1987. C. 589-617.

15. 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.

16. Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, N.J.: Lawrence Erlbaum, 1980.

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. Braeken, J., & van Assen, M. A. An empirical Kaiser criterion. Psychological Methods, 22, 2017. C. 450–466.

http://dx.doi.org/10.1037/met0000074

20. Horn, J. L. A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 1965. C. 179–185.

http://dx.doi.org/10.1007/BF02289447

21. 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. C. 377–393.

http://dx.doi.org/10.1177/0013164495055003002

22. 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

23. Zwick, W. R., & Velicer, W. F. Comparison of five rules for determining the number of components to retain. Psychological Bulletin,99, 1986. C. 432–442.

http://dx.doi.org/10.1037/0033-2909.99.3.432

24. Thompson, B. and Daniel, L.G. Factor Analytic Evidence for the Construct Validity of Scores: A Historical Overview and Some Guidelines. Educational and Psychological Measurement, 56, .1996. C. 197-208.

http://dx.doi.org/10.1177/0013164496056002001

25. 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

26. Bock RD, Aitkin M. "Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm." Psychometrika, 46(4), 1981. C. 443–459.

27. Dempster AP, Laird NM, Rubin DB. "Maximum Likelihood From Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society B, 39(1), 1977. C. 1–38.

28.https://ela.kpi.ua/bitstream/123456789/48225/1/Kononenko_magistr.pdf

29. 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

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

31. 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.

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

33. https://cran.r-project.org/