

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

Фізико-математичний факультет

Кафедра математичного аналізу та теорії ймовірностей

«На правах рукопису»
УДК _____

До захисту допущено:
Завідувач кафедри
_____ Олег КЛЕСОВ
«14» травня 2021 р.

Магістерська дисертація

з напрямку підготовки 111 «Математика»

на тему: Застосування MIRT і IRT моделей до аналізу тестів з вищої математики

Виконала:

студентка II курсу, групи ОМ-91мн

Лисенко Діана Ростиславівна
Науковий керівник:

кандидат фізико-математичних наук, доцент

Круглова Наталія Володимирівна
Рецензент:

кандидат фізико-математичних наук, доцент кафедри
математики та теоретичної радіофізики

Єфіменко Світлана Володимирівна. КНУ ім. Тараса

Шевченка

Засвідчую, що в цій дипломній роботі
немає запозичень із праць інших авторів
без відповідних посилань.
Студентка _____

Київ — 2021 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

«04» лютого 2021 р.

ЗАВДАННЯ
на магістерську дисертацію студентці

Лисенко Діані Ростиславівні

1. Тема дисертації: «Застосування MIRT і IRT моделей до аналізу тестів з вищої математики», науковий керівник дисертації: Круглова Наталія Володимирівна, кандидат фізико-математичних наук, доцент, затверджені наказом по університету від «26» березня 2021 р. №901-с.
2. Термін подання студентом роботи: 13 травня 2021 року.
3. Об'єкт дослідження: тестові МКР з вищої математики для студентів ФІОТ-у НТУУ «КПІ ім. Ігоря Сікорського».
4. Предмет дослідження: математичні моделі тестів.
5. Перелік завдань, які потрібно розробити: виконати аналіз літератури та методів дослідження тестів, застосувати одномірні та багатомірні моделі тестів до питань типу “вбудовані відповіді”, “числові”, “множинний вибір”, проаналізувати отримані результати, проаналізувати якість контрольних робіт, перевірити відповідність вибраних моделей даним тестування, скласти системи рівнянь для обчислення латентних параметрів багатовимірних моделей, виписати ітераційні формули для обчислення цих параметрів методом Ньютона-Рафсона.
6. Перелік ілюстративного матеріалу: графік функції успіху; графік функції виміру;

характеристичні криві рівнів підготовленості іспитників та складності завдань; види матриць первинних балів; дані, що пройшли первинну обробку для подальшого застосування політомічної моделі; гістограма частот сумарних балів іспитників; ІСС або характеристична крива складності та інформаційна крива завдань; графіки та поверхні очікуваних балів іспитників з різними рівнями підготовленості за завдання; інформаційна крива усіх тестових питань; ліній рівня, характеристичні поверхні складності питань та поверхні очікуваних балів іспитників для двовимірних та двофакторних моделей; ІС та ІСС криві для 1-PL, 2-PL, 3-PL та GPCM моделей.

7. Перелік публікацій:

- 1) Д.Р. Лисенко, Н.В. Круглова, О.О. Диховичний. Аналіз тестів типу “вбудовані відповіді” за допомогою моделей IRT // Тези доповідей X всеукраїнської наукової конференції молодих математиків. – 2021 р. – С. 106–107.
 - 2) Н.В. Круглова, О.О. Диховичний, Д.Р. Лисенко. Використання кластерного аналізу і IRT для поділу респондентів психологічного тестування на групи // Матеріали IX міжнародної науково-практичної конференції “Математика в сучасному технічному університеті”. – 2021 р. – С. 90–91.
 - 3) Н.В. Круглова, О.О. Диховичний, Д.Р. Лисенко. IRT і метод головних компонент як інструменти формування банку завдань дистанційного семестрового контролю // Матеріали IX міжнародної науково-практичної конференції “Математика в сучасному технічному університеті”. – 2021 р. – С. 302–304.
 - 4) N. Kruglova, O. Dykhovychnyi, D. Lysenko, N. Bogdanova. Simulation of distribution of the maximum of the Chentsov field on ‘steps’ // Adaptive Systems of Automatic Control. — 2020. — Vol. 2 No. 37 — p. 93–101.
 - 5) N. Kruglova, O. Dykhovychnyi, D. Lysenko. Application of IRT and MIRT models to analysis of analytical geometry tests// Adaptive Systems of Automatic Control. — 2021. — Vol. 1. (пройшла рецензування, готується до друку)
8. Дата видачі завдання: 04 лютого 2021 року.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Аналіз літератури по IRT та методів дослідження тестів.	04.02.2021- 21.02.2021	виконано
2	Застосування одномірних та багатомірних моделей тестів до питань типу “вбудовані відповіді” та “множинний вибір”.	22.02.2021- 28.02.2021	виконано
3	Аналіз отриманих результатів та якості моделей у відповідності до даних.	01.03.2021- 08.03.2021	виконано
4	Аналіз літератури по MIRT та методів дослідження тестів.	09.03.2021- 31.03.2021	виконано
5	Застосування багатомірних моделей тестів до питань типу “вбудовані відповіді” та “множинний вибір”.	01.04.2021- 18.04.2021	виконано
6	Аналіз отриманих результатів та якості моделей у відповідності до даних.	19.04.2021- 13.05.2021	виконано

Студентка
Науковий керівник дисертації

Діана ЛИСЕНКО
Наталія КРУГЛОВА

АНОТАЦІЯ

Магістерську роботу виконано на 64 аркушах, вона містить 1 додаток та перелік посилань на використані джерела з 15 найменувань. У роботі наведено 11 таблиць та 22 рисунки.

Метою даної дипломної роботи є дослідження багатовимірних моделей IRT та складення системи рівнянь для обчислення методом Ньютона-Рафсона латентних параметрів багатовимірних моделей, а також створення математичного та програмного забезпечення для комп'ютерної візуалізації результатів тестування.

У роботі було проведено моделювання та аналіз тестування за допомогою IRT та MIRT методів. А саме: обробка матриці первинних балів, обчислення коефіцієнту надійності тесту, застосування IRT та MIRT моделей, побудова графіків характеристичних функцій та поверхонь, здійснено перевірку моделей на адекватність, застосовано критерії вибору моделі.

У процесі виконання роботи було проведено аналіз вищезгаданих отриманих результатів.

Ключові слова: сучасна теорія тестів, багатовимірна теорія тестів, R, ІІС, характеристична крива важкості завдання, “вбудовані відповіді”, 1-PL, 2-PL, 3-PL, GPCM моделі.

ABSTRACT

The master's thesis is performed on 64 sheets, it contains 1 appendix and a list of references to the sources used with 15 items. The paper contains 11 tables and 22 figures.

The aim of this thesis is to study multidimensional IRT models and compile a system of equations for calculating the latent parameters of multidimensional models by the Newton-Rafson method, as well as to create mathematical and software for computer visualization of test results.

The modeling and analysis of testing results due to the methods of IRT and MIRT were performed in the work. Namely: processing of the matrix of primary scores, calculation of the Cronbach's coefficient, application of IRT and MIRT models, plotting of characteristic functions and surfaces, testing of models for adequacy, application of model selection criteria.

In the course of performance of work the analysis of the above-mentioned received results was carried out.

Key words: IRT, MIRT, R, IIC, ICC, "embedded answers", 1-PL, 2-PL, 3-PL, GPCM models.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТА ТЕРМІНІВ.....	8
ВСТУП.....	9
1 ПОСТАНОВКА ЗАДАЧІ.....	10
2 КОРОТКІ ІСТОРИЧНІ ВІДОМОСТІ.....	11
3 ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ТЕСТІВ ТА ДЕЯКІ ОДНОМІРНІ IRT МОДЕЛІ.....	14
3.1 Функція успіху та латентні параметри.....	14
3.2 Логістична модель Раша.....	15
3.3 Логістична модель Бірнбаума.....	18
3.4 Трьохпараметрична модель Бірнбаума.....	19
3.5 Модель Муракі та модель Андерсена.....	20
3.6 Види матриць первинних балів.....	20
4 БАГАТОМІРНІ IRT МОДЕЛІ.....	22
4.1 Компенсаторні та некомпенсаторні моделі.....	22
4.2 Компенсаторна 2-PL модель.....	22
4.3 Виведення системи рівнянь для знаходження латентних параметрів 2-PL моделі з двома вимірами.....	23
4.4 Багатовимірна модель Муракі.....	25
5 ПЕРЕВІРКА АДЕКВАТНОСТІ МОДЕЛЕЙ.....	27
6 КРИТЕРІЇ ВИБОРУ МОДЕЛІ.....	28
6.1 Критерій Акаїке (AIC).....	28
6.2 Баєсівський інформаційний критерій (BIC) або критерій Шварца (SC).....	29
7 ДОСЛІДЖЕННЯ ТА АНАЛІЗ ДАНИХ.....	30
7.1 Первинна обробка результатів тестування з аналітичної геометрії.....	30
7.2 Змішана модель або модель А.....	32
7.3 Двовимірна модель Муракі або модель В.....	39
7.4 Багатовимірна двофакторна модель Муракі або модель С.....	42
7.5 Порівняння моделей за AIC та BIC критеріями.....	46
7.6 Аналіз МКР з ТЙ.....	47
ВИСНОВКИ.....	57
ПЕРЕЛІК ПОСИЛАНЬ.....	59
Додаток А. Лістинг програми.....	61

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТА ТЕРМІНІВ

ВМ – вища математика;

ТЙ – теорія ймовірностей;

МКР – модульна контрольна робота;

КТТ – класична теорія тестів;

IRT – сучасна теорія тестів;

MIRT – багатомірна сучасна теорія тестів;

вбудовані відповіді – пропущені слова;

1-PL модель – логістична модель Раша;

2-PL модель – двопараметрична логістична модель Бірнбаума;

3-PL модель – трьохпараметрична логістична модель Бірнбаума;

GPCM – Generalized Partial Credit Model або модель Муракі;

PCM – Partial Credit Model або модель Андерсена;

ІІС – Item Information Curve;

ІСС – Item Characteristic Curve.

ВСТУП

Сьогодні актуальною проблемою є створення якісної дистанційної системи навчання. Окрім того перед викладачами постають нові задачі контролю та перевірки знань учнів або студентів. Оскільки кількість бажаючих отримати освіту збільшується, то викладачі змушені боротися за якість рівня знань, задля формування та підготовки майбутніх спеціалістів. Звичайно, що одному викладачу важко контролювати і надавати кожному своєму підопічному допомогу особисто. Мало того – звичайна перевірка знань за допомогою тестів, де треба обрати лише відповідь із за-даного переліку, теж не є ефективною. Оскільки розв’язок на такі завдання часто можна або вгадати, або швидко списати, більш того – тести закритого плану не містять у собі усі ключові моменти розв’язку певного завдання. Тому максимум, що може отримати студент в кінці написання такої роботи – або стрес і абсолютне незрозуміння того, що відбувалось і що від нього вимагалось під час тесту; або впевненість, що він списав, а значить так буде тривати весь його період навчання у кращому випадку; або студент вирішить завдання правильно, але витратить на розв’язок занадто багато часу, що може означати, що студент не вміє вирішувати задачу раціональним шляхом, а це повністю протирічить засадам отримання якісної освіти. Більш того – завдання не мають бути занадто легкими, або занадто важкими; та й загальні результати тестування можуть рішуче відрізнитися одне від одного, якщо одні і ті ж питання дати студентам, наприклад, з різних факультетів. Тож звичайні методи КТТ тут вже не є дієвими.

У цій роботі продемонстровано на одному із результатів опитування у доступній, безоплатній, відкритій платформі для створення персоналізованого навчального середовища – для студентів, та керування навчанням – для викладачів – Moodle переваги та можливості сучасної теорії тестів IRT та, досі не дослідженої вітчизняними науковцями, – теорії MIRT на прикладі питань типу “вбудовані відповіді” (пропущені слова). Це один із типів питань у системі Moodle, який є досить гнучким і може виступати заміном звичайних закритих тестів.

1 ПОСТАНОВКА ЗАДАЧІ

Метою даної магістерської роботи є створення програми, яка забезпечує комп'ютерну візуалізацію результатів тестування до питань типу “вбудовані відповіді” на платформі Moodle за допомогою методів КТТ, IRT та MIRT, коли кількість опитуваних в тестуванні більше 70, а кількість питань більше 8. Застосовані методи значно полегшують педагогічні, психологічні та соціологічні дослідження: вони виявляють дуже важкі та легкі питання у тесті; завдання, де відповідь можна вгадати; питання, які розв’язуються за допомогою спеціального програмного забезпечення; знаходять завдання, які мають низьку диференціюючу здатність; виявляють рівень знань опитуваних в суміжних областях. [1,2] Комбінація таких методів є більш ефективною, ніж звичайне застосування КТТ, яка залежить від вибірки тестування. Єдина складність полягає в тому, що для роботи програми необхідно завантажити початкові результати тестування з платформи Moodle або ж заповнити таблицю Excel вручну (у випадку обробки дихотомічних даних), що значно збільшує час досліджень.

При розв’язку задачі на дипломну роботу необхідно вирішити наступні завдання:

- 1) обробити матрицю первинних балів, завантаженої з платформи Moodle;
- 2) побудувати гістограми сумарних балів іспитників;
- 3) знайти коефіцієнт Кронбаха та виявити завдання з високою кореляцією;
- 4) застосувати різноманітні моделі IRT та MIRT до оброблених матриць;
- 5) порівняти моделі за допомогою AIC та BIC критеріїв;
- 6) провести графічний аналіз якості тестів з BM;
- 7) проаналізувати отримані результати, а також провести аналіз адекватності моделей за допомогою критерія χ^2 ;
- 8) побудувати систему для знаходження латентних параметрів двопараметричної 2-PL моделі, виписати необхідні ітераційні формули для обчислення невідомих параметрів.

2 КОРОТКІ ІСТОРИЧНІ ВІДОМОСТІ

Теорія тестів бере свій початок ще з середини 80-х рр. XIX ст., коли співробітники лабораторій, що займалися вивченням органів чуття в Німеччині: Вільгельм Вундт, Гюстав Фехнер, Ернст Вебер та їх колеги першими намагалися пояснити точність застосовуваних методів в психології з точки зору природничих наук.

Одночасно з ними вчені Великобританії працювали над індивідуальними відмінностями людини, яка згодом буде мати великий вплив на вимірювання розумових характеристик індивіда. Найбільший вклад зробив Френсіс Гальтон. Його ідеї мали величезний вплив на його відомого сучасника та двоюрідного брата – Чарльза Дарвіна. Гальтон у 1869 р., використовуючи результати іспиту з математики всіх студентів Кембриджського університету та вступного іспиту в Королівський військовий коледж, продемонстрував, що розподіл розумових здібностей можна описати математично за допомогою щільності нормального розподілу. Майже 20 років по тому він опублікував есе, в якому запропонував кореляційну процедуру для аналізу коваріації двох або більше характеристик.

Пізніше британський статистик Карл Пірсон вивів статистичну формулу для обчислення коефіцієнта кореляції, спираючись на результати Гальтона. Інші британські вчені взяли з них приклад. Найбільш помітний вклад зробив Чарльз Спірмен. Його теорія інтелекту базувалась на більш просунутих кореляційних процедурах, які він назвав факторним аналізом. Сьогодні кореляційні методи залишаються одними з найбільш відомих статистичних методів, які часто використовуються при валідації тестів.

Незважаючи на велике значення вкладів німецьких та британських психологів, саме досягнення двох французів, Альфреда Біне та Теофіла Симона (1905-1908), просунули тестування від рівня наукових вправ до рівня прикладних методів, які негайно зможуть знайти застосування в навчальному класі, клініці або на роботі. Обидва клінічних психологи займалися розробкою методу для виявлення учнів із

затримкою в розвитку. За результатами виконання своїх завдань Біне оцінював “ментальний вік” іспитників, який порівнювався з хронологічним, щоб потім учнів можна було розподілити. Тож Біне увів додатково поняття норми. Окрім того нинішня шкала Стенфорда-Біне мало відрізняється від тесту, який був створений і представлений ще в ті роки школярам Парижу. Тож фактично сучасні розробники тестів і до сьогодні використовують методи аналізу завдань та встановлення норм, які не дивлячись на значні модифікації, в принципі залишилися схожими на процедури, що використовувалися Біне ще на початку XX ст.

У США роботи Біне використовувалися масово для написання тестів визначення розумового розвитку, що було характерним для індустріальної революції. У 1917 р., коли США вступили у Першу світову війну, група із 6-ти психологів отримала фінансову підтримку від міністерства оборони у розмірі 700\$ і за 3 місяці розробила 5 форм групового екзамену, що був проведений зі всім військовим персоналом. Пізніше була додана група невербальних тестів для іспитників, які не знають англійської. Результати тестів використовувалися при розподілі новобранців, визначенні серед них кандидатів на навчання в офіцерських школах, відсіюванні людей з затримками у розвитку та навіть для діловодства у військових судах. Після цього психологи намагалися розширити сферу впливу тестів ще на звичайний люд, з ціллю розподілу в професійній та учбовій діяльності. [3]

Сьогодні, у зв'язку з активним впровадженням дистанційного навчання у закладах освіти, часто викладачами застосовуються тести, як основний метод збору та обробки результатів засвоєння певної теми або розділу студентами з деякого предмету. Проте по-справжньому гарно сконструйований тест, насправді, вимагає від укладача тесту розуміння як його скласти, які результати приблизно варто очікувати. Зокрема, якщо гістограми розподілу згрупованих або незгрупованих емпіричних даних не схожі на нормальний розподіл, то скоріше всього, тест був зконструйований невдало. Або якщо результати зміщені вправо або вліво відносно очікуваного середнього балу за тест, то це теж погано, бо означає, що завдання були або занадто легкі, або складні. [4]

Один із підходів до конструювання тесту, який призводить до більш повного

відображення того, як функціонують окремі завдання, отримав назву сучасної теорії тестів (IRT), або теорії латентної характеристики.

У теорії IRT розробник тесту припускає, що відповіді на завдання тесту можуть бути пояснені шляхом аналізу латентних характеристик, яких менше, ніж завдань тесту. Насправді більшість застосувань теорії впливає з того, що єдина латентна характеристика пояснює відповіді на завдання тесту. У “серці” цієї теорії лежить матмодель того, як іспитники з різними рівнями підготовки за латентними характеристиками повинні відповідати на завдання. [3]

На сьогодні класична теорія тестування (КТТ) не втратила своєї актуальності та одночасно використовується з більш сучасними теоріями тестів такими, як IRT (Item Response Theory) та MIRT (Multidimensional Item Response Theory). Сучасні ж теорії тестів швидко розвиваються та мають широке застосування не лише в психології, соціології, менеджменті, педагогіці, але й в політології та навіть у тестуванні програмного забезпечення. IRT та MIRT більш точно диференціюють іспитників, тому застосовуються при професійному відборі у великих структурах та військовому секторі. Ці теорії дозволяють покращувати та виявляти недоліки у нових психологічних методиках, визначати питання, що потребують коригування, адаптують відомі тестування до сучасного іспитника.

Тож згадаймо думку Іммануїла Канта, який жив в епоху розквіту класичної механіки, що в будь-якій науці рівно стільки науки, скільки в ній математики.

3 ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ТЕСТІВ ТА ДЕЯКІ ОДНОМІРНІ IRT МОДЕЛІ

3.1 Функція успіху та латентні параметри

Ймовірність правильного вирішення певного завдання іспитником на тестуванні — це функція принаймні двох змінних:

- 1) s — рівня підготовленості іспитника;
- 2) t — важкості завдання.

Тоді

$$p = p(s, t) - \text{функція успіху}.$$

Перш ніж перейти до питання як вимірювати s, t кількісно, візьмемо до уваги наступні міркування. Нехай важкість завдання t_1 в k разів більше $t_2 \Rightarrow t_1 = kt_2$. Нехай також рівень підготовленості s_1 одного із іспитників в k разів більше рівня підготовленості $s_2 \Rightarrow s_1 = ks_2 \Rightarrow k = \frac{s_1}{s_2} = \frac{t_1}{t_2} \Rightarrow \frac{s_1}{t_1} = \frac{s_2}{t_2}$. Тоді можна зробити висновок, що ймовірність того, що перший учасник тестування, у якого підготовленість краща, відповість на складніше питання вірно, має співпадати з ймовірністю, що другий учасник, у якого рівень підготовленості нижчий, відповість правильно на менш складне — друге питання. Тобто поняття s, t між собою пов'язані, а функція успіху залежить насправді від відношення цих величин:

$$p = p(s, t) = p_1(\xi), \text{ де } \xi = \frac{s}{t}.$$

Змінні s та t називаються **латентними параметрами**, тобто такими, що їх не можна спостерігати, бо вони описують приховані характеристики іспитників та самих завдань тестування. Також точні аналітичні записи для функцій p та p_1 в останній формулі невідомі, проте можна говорити про їх апроксимації. Тоді природними є накладання таких умов на p та p_1 :

- i. функція p_1 гладка, тобто неперервна разом зі своєю похідною, та

монотонно зростаюча на всій області визначення $[0; +\infty)$, тому що збільшення частки $s:t$ має означати, що ймовірність правильного рішення завдання теж збільшилась;

- ii. $\lim_{\xi \rightarrow 0} p_1(\xi) = p_1(0) = 0$, тобто той, хто не готовий, не зможе відповісти правильно під час тестування;
- iii. $\lim_{\xi \rightarrow +\infty} p_1(\xi) = 1$, що гарантує гарну оцінку тому іспитникові, у якого рівень підготовленості в рази перевищує складність завдання;
- iv. $p_1(1) = 0.5$, тобто насправді максимальна невизначеність у передбаченні результату виникає, коли рівень підготовленості і складність завдання співпадають. [5]

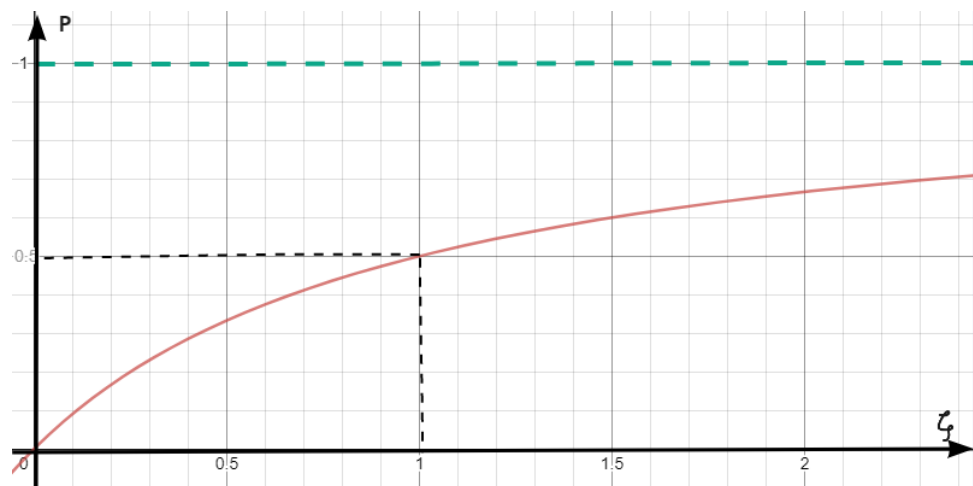


Рис. 3.1.1 — Приблизний графік функції успіху

3.2 Логістична модель Раша

Модель Раша — це найпростіша модель ймовірності успіху. Вона була запропонована датським математиком та є однопараметричною:

$$p(s, t) = \frac{s}{s+t} = \frac{s/t}{s/t+1} = \frac{\xi}{\xi+1}.$$

Окрім того ця модель підходить до рішення задачі про знаходження

ймовірності дістати чорну кулю із урни, коли в ній знаходяться s чорних куль та t — білих.

Нагадаємо, що аргументи функції успіху є латентними параметрами, тому їх виміряти ми не можемо, проте значення самої функції ми можемо виміряти з деякою точністю в реальному житті під час тестування. Тоді насправді цікава обернена функція $\xi = p_1^{-1}(p)$ до функції успіху. Ця обернена функція існує тому, що $p = p_1(\xi)$ монотонна. Можна тоді одразу зауважити, що оскільки $\xi = p_1^{-1}(p)$ обернена до функції успіху, то графік оберненої буде симетричним графіку на рис. 3.1.1 відносно бісектриси першого квадранта, тобто матиме вигляд:

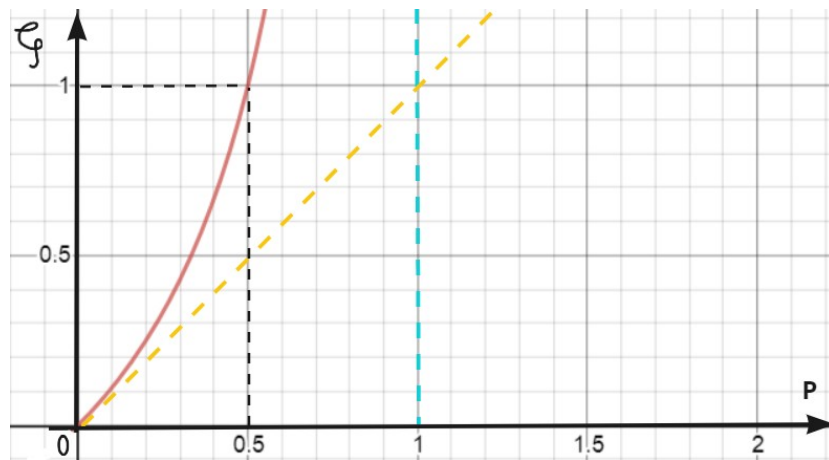


Рис. 3.2.1 — Приблизний графік функції виміру

Оскільки обернена функція до функції успіху виражає латентний параметр ξ через p , що доступна для тестування, то її називають **функцією вимірювання**.

Тобто функції вимірювання та успіху — взаємно обернені:

$$p = \frac{\xi}{\xi + 1} \Rightarrow (\xi + 1) \cdot p = \xi \Rightarrow \xi \cdot p + p = \xi \Rightarrow p = \xi \cdot (1 - p) \Rightarrow \frac{p}{1 - p} = \xi \Rightarrow \frac{p}{q} = \xi, \text{ де } q = 1 - p.$$

Тобто при багаторазових проведеннях тестів, можна оцінити ймовірність успіху та латентний параметр ξ . До того ж, у задачі з повторною вибіркою куль, коли треба порахувати відносну частоту появи чорної кулі (куля дістається наугад багато разів і повертається кожен раз назад в урну), користуються саме формулою

$$\frac{p}{q} = \xi, \text{ де } q = 1 - p.$$

На практиці аргументи $s \geq 0, t \geq 0$ зручно виражати в логарифмічному масштабі.

$$\text{Тоді нехай } \ln(s)=\theta, \ln(t)=r \Leftrightarrow s=e^{\theta}, t=e^r \Rightarrow p=\frac{e^{\theta}}{e^{\theta}+e^r}=\frac{e^{\theta}/e^{\theta}}{1+e^{r-\theta}}=\frac{1}{1+e^{r-\theta}}=\frac{1}{1+e^{-(\theta-r)}},$$

де $p=\frac{1}{1+\exp(-(\theta-r))}$, називається **логістичною моделлю Раша або 1-PL моделлю**,

оскільки параметри θ і r вимірюються за одною шкалою — логіт. Можна зробити висновок, що логістична модель Раша насправді залежить від одного параметра $(\theta-r)$, тобто вона однопараметрична.

Також $\lim_{(\theta-r) \rightarrow +\infty} p=1$, $\lim_{(\theta-r) \rightarrow -\infty} p=0$ та $p=0.5$, коли $\theta=r$.

Якщо зафіксувати $\theta=\theta_0$, тобто рівень підготовленості індивідуума, то можна помітити, що ймовірність p як функція від r повністю описує потенціальні можливості індивіда з деяким рівнем підготовленості θ_0 і тому називається **характеристичною функцією рівня підготовленості θ_0** . І навпаки, якщо зафіксувати $r=r_0$, а θ вважати будь-яким, то можна отримати **характеристичну функцію важкості r_0** . Графіки характеристичних функцій називаються відповідними **характеристичними кривими**:

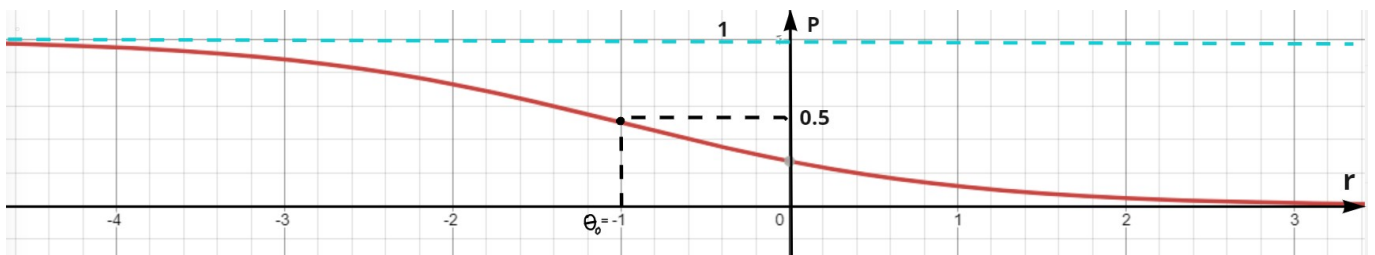


Рис. 3.2.2 — Характеристична крива підготовленості θ_0

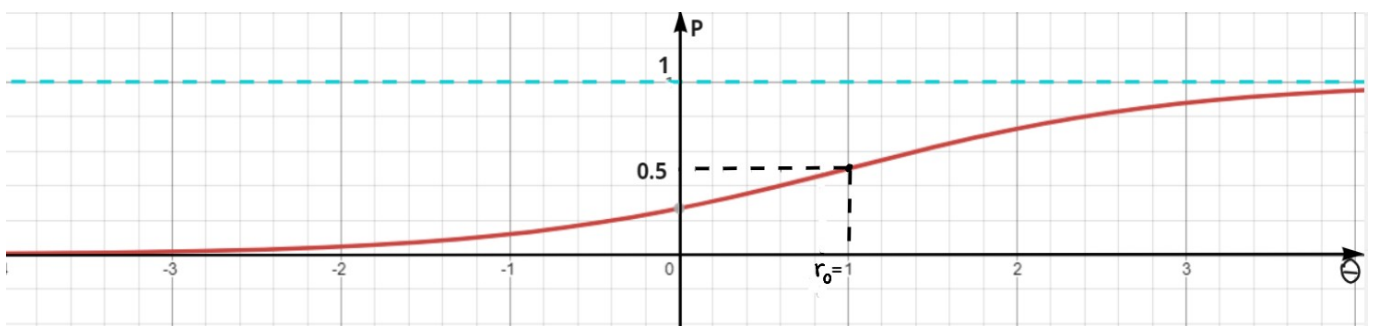


Рис. 3.2.3 — Характеристична крива рівня складності r_0

Можна помітити з рис. 3.2.2, що характеристична крива рівня підготовленості θ_0 — це характеристична крива рівня підготовки 0, яка зміщена без деформації вздовж осі абсцис на θ_0 одиниць. Це означає збільшення (якщо $\theta_0 > 0$) або

зменшення (якщо $\theta_0 < 0$) ймовірностей успішного виконання іспитників з $\theta = \theta_0$ завдань різної складності. Тобто характеристичні криві, що відповідають різним рівням підготовки, не перетинаються. Це хороша відмінність основної логістичної моделі Раша від інших функцій успіху, де перетин можливий.

3.3 Логістична модель Бірнбаума

Логістична модель Бірнбаума — це двопараметрична модель, яка є узагальненням моделі Раша. В ній додається новий параметр d — диференціююча або дискримінаційна здатність завдання. Тобто чим більше значення d , тим краще. Оскільки, коли d — мале, то характеристична крива важкості r_0 буде пологою, а це означає, що іспитники з хорошим і не дуже рівнем підготовки виконують дане завдання з приблизно однаковим успіхом. І навпаки, коли d — велике, то результати іспитників з різними рівнями підготовки будуть відрізнятися істотно. Ймовірність успіху вираховується за формулою:

$$p = (1 + \exp(-d[\theta - r]))^{-1}.$$

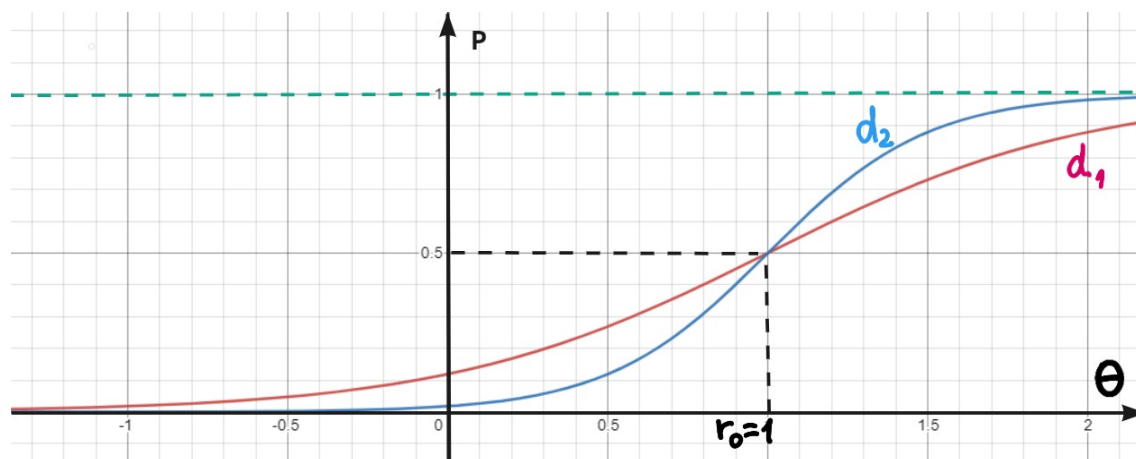


Рис. 3.3.1 — Характеристичні криві складності r_0 завдання з різними значеннями параметра d , де $0 < d_1 < d_2$

Бачимо, що при $d=1$ двопараметрична модель Бірнбаума співпадає з 1-PL моделлю Раша.

Якщо $d=1.7$, то тоді модель Бірнбаума добре апроксимує (з точністю до 0.01)

функцію розподілу ймовірностей нормального закону з нульовим матсподіванням та одиничною дисперсією:

$$N\left(\frac{\theta}{r}=r_0; d\right)=\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d(\theta-r_0)} \exp(-x^2/2) dx \Rightarrow |N\left(\frac{\theta}{r}=r_0; d\right)-p\left(\frac{\theta}{r}=r_0; d\right)|<0.01$$

$\forall \theta \in \mathbb{R}$, а $p\left(\frac{\theta}{r}=r_0; d\right)$ означає характеристичну функцію важкості r_0 в моделі

Бірнабаума з параметром $d=1.7$. Це дозволяє інтерпретувати, за необхідності, різні обчислення з логістичною моделлю з погляду детально вивченого нормального закону розподілу ймовірностей.

3.4 Трьохпараметрична модель Бірнабаума

Трьохпараметрична модель Бірнабаума або як вона ще відома – 3-PL модель має вигляд:

$$p\left(\frac{\theta}{r}=r_0; d; c\right)=c+(1-c) \cdot [1+\exp(-d(\theta-r))]^{-1}$$

та ґрунтується на формулі повної ймовірності двох подій. Одна із цих подій полягає у можливості вгадати правильну відповідь на якесь завдання з ймовірністю c , навіть, якщо $\theta \ll r_0$. [5]



Рис. 3.4.1 — Характеристична крива складності r_0 з дискримінаційною спроможністю $d=4 \cdot \tan \alpha$ та шансом вгадування c

3.5 Модель Муракі та модель Андерсена

Модель Муракі або GPCM (a generalized partial credit model) застосовується в політомічних завданнях, тобто в таких, де є декілька підпитань або підпунктів m . Ця модель описує ймовірність іспитником досягти в i -му завданні деякого рівня k :

$$p_{i,k} = \frac{\exp\{\sum_{j=0}^k d_i(\theta - r_{ij})\}}{\sum_{k=0}^m \exp\{\sum_{j=0}^k d_i(\theta - r_{ij})\}},$$

де d_i – дискримінаційна або диференціююча спроможність i -го завдання, θ – рівень підготовленості іспитника та j – індекс сумування. [2]

Якщо $d=1$, то GPCM перетворюється в **PCM (Partial Credit Model)** або інакша назва – **модель Андерсена**.

3.6 Види матриць первинних балів

Тут необхідно пояснити, що існує декілька схем оцінювання завдань: **дихотомічна** та **недихотомічна (або політомічна)** – одні з найпопулярніших схем, які використовуються в тестах. Дихотомічні змінні – це 0 або 1 балів, а політомічна схема оцінювання застосовується, коли може зараховуватись за відповідь не лише 0 або 1 бал, але й проміжні між 0 та 1, або більші за 1 значення. Наприклад, шкала Лайкерта (від “повністю згоден” до “повністю не згоден”) використовує якраз недихотомічну схему оцінювання. [3]

Політомічна матриця								Дихотомічна матриця							
	A	B	C	D	E	F	G		E	F	G	H	I	J	
1	1	2	3	4	5	6	7	1	5	6	7	8	9	10	
2	1	0	0	2	1	0	1	2	0	1	0	1	0	0	
3	1	0	3	3	0	4	3	3	1	1	0	0	0	1	
4	3	3	3	1	1	4	2	4	0	0	0	1	0	1	
								5	1	1	1	1	1	1	

Рис. 3.5.1 — Види матриць первинних балів

Таким чином, якщо

n — це кількість іспитників,

m — це кількість питань в тесті,

то для дихотомічної матриці первинних балів логістична модель Раша (1-PL), Бірнбаума (2-PL), трьохпараметрична логістична модель Бірнбаума (3-PL) набувають відповідно вигляду:

$$p_{ij} = \frac{1}{1 + \exp(-(\theta_i - r_j))}, \quad p_{ij} = \frac{1}{1 + \exp(-d_j(\theta_i - r_j))}, \quad p_{ij} = c_j + (1 - c_j) \frac{1}{1 + \exp(-d_j(\theta_i - r_j))},$$

де $i = \overline{1, n}$; $j = \overline{1, m}$; d_j — диференціююча здатність j -го завдання; θ_i — рівень підготовленості i -го іспитника; r_j — рівень складності j -го завдання; c_j — ймовірність вгадати відповідь у j -му завданні.

3.7 Коефіцієнт надійності тесту

Альфа Кронбаха або **коефіцієнт Кронбаха** — це коефіцієнт надійності тесту.

Він є найвідомішим та найвживанішим, проте спеціалістами рекомендуються інші як його альтернатива. Альфа Кронбаха визначається за наступною формулою:

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\sum_{i=1}^p \sigma_{x_i}^2}{\sigma_x^2} \right),$$

де p — це кількість завдань, значенням X_i позначають спостережуваний бал предмета i , також $X = X_1 + X_2 + \dots + X_p$, σ_x^2 — це дисперсія спостережуваних загальних результатів тесту, а $\sigma_{x_i}^2$ — це коваріація між X_i та X_j , де $i \neq j$. [6]

4 БАГАТОМІРНІ IRT МОДЕЛІ

4.1 Компенсаторні та некомпенсаторні моделі

Розглянемо кілька багатомірних моделей MIRT (Multidimensional Item Response Theory). На відміну від звичайних одновимірних IRT моделей, головною перевагою цих моделей є те, що рівень знань i -го іспитника виступає в ролі s -вимірного вектора, тобто: $(\theta_{i1}, \theta_{i2}, \dots, \theta_{is})$. Звичайно, що MIRT моделі можна вважати узагальненням IRT моделей на випадок даних, які є багатовимірними.

MIRT моделі поділяються на два типи: *компенсаторні* та *некомпенсаторні*.

Некомпенсаторні моделі є мультиплікативними. Їх використовують тоді, коли недостатній рівень знань з одного предмету ніяк не компенсувати гарними вміннями чи навичками з іншого предмету.

У даній роботі розглядатимуться лише компенсаторні моделі. Вони є адитивними, тобто припускають можливість компенсувати слабкі знання іспитника з однієї області за рахунок високого рівня знань та умінь в іншій області. [7]

4.2 Компенсаторна 2-PL модель

Для цієї моделі ймовірність того, що i -ий іспитник правильно відповість у j -му дихотомічному завданні дорівнює:

$$p_{ij} = \frac{\exp\left(\sum_{l=1}^s d_{jl}\theta_{il} + r_j\right)}{1 + \exp\left(\sum_{l=1}^s d_{jl}\theta_{il} + r_j\right)},$$

де n — це кількість іспитників, m — це кількість питань в тесті, $i = \overline{1, n}$;

$j=\overline{1,m}$; $l=\overline{0,s}$; s – вимірність вектора підготовленостей; d_{jl} – диференціюючі здатності j -го завдання; θ_{il} – рівні підготовленості i -го іспитника; r_j – рівень складності j -го завдання.

Окрім того помітимо, що багатовимірна складність завдань обчислюється за наступною формулою:

$$R_j = \frac{-r_j}{\sqrt{\sum_{q=1}^s d_{jq}^2}}, j=\overline{1,m},$$

а багатовимірна диференціююча спроможність завдань вираховується за формулою:

$$D_j = \sqrt{\sum_{q=1}^s d_{jq}^2}, j=\overline{1,m}. \quad [7]$$

4.3 Виведення системи рівнянь для знаходження латентних параметрів 2-PL моделі з двома вимірами $\theta=(\theta_1, \theta_2)$

Нехай дано: $p_{ij} = \frac{\exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}{1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}$, де $i=\overline{1,n}$; $j=\overline{1,m}$; n — це кількість іспитників; m — це кількість питань в тесті; d_{j1}, d_{j2} – диференціюючі здатності j -го завдання; θ_{i1}, θ_{i2} – рівні обізнаності i -го іспитника; r_j – рівень складності j -го завдання.

Введемо позначення: $\varphi_{ij} = d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j$, тоді $p_{ij} = \frac{\exp(\varphi_{ij})}{1 + \exp(\varphi_{ij})}$. Позначимо також через $q_{ij} = 1 - p_{ij} = \frac{1}{1 + \exp(\varphi_{ij})}$.

Введемо також функцію правдоподібності. Вона буде мати наступний вигляд:

$$L = \prod_{i=1}^n \prod_{j=1}^m p_{ij}^{y_{ij}} q_{ij}^{1-y_{ij}}, y_{ij} \in \{0; 1\}. \text{ Прологарифмуємо її, щоб отримати функцію}$$

логарифмічної правдоподібності:

$$\Lambda = \ln(L) = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} \ln p_{ij} + (1 - y_{ij}) \ln(q_{ij})) = \sum_{i=1}^n \sum_{j=1}^m [y_{ij} (\ln e^{\varphi_{ij}} - \ln(1 + e^{\varphi_{ij}})) - (1 - y_{ij}) \ln(1 + e^{\varphi_{ij}})] =$$

$$= \sum_{i=1}^n \sum_{j=1}^m [y_{ij} \varphi_{ij} - y_{ij} \ln(1 + e^{\varphi_{ij}}) - \ln(1 + e^{\varphi_{ij}}) + y_{ij} \ln(1 + e^{\varphi_{ij}})] = \sum_{i=1}^n \sum_{j=1}^m [y_{ij} \varphi_{ij} - \ln(1 + e^{\varphi_{ij}})], y_{ij} \in \{0; 1\}.$$

Отже, функція логарифмічної правдоподібності набуває наступного вигляду:

$$\Lambda = \sum_{i=1}^n \sum_{j=1}^m [y_{ij}(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j) - \ln(1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j))], y_{ij} \in \{0; 1\}.$$

Для знаходження невідомих параметрів, скористаємося методом максимальної вірогідності. Оскільки функція логарифмічної правдоподібності диференційовна, то необхідно, щоб виконувались умови екстремума, тобто градієнт має дорівнювати нулю. Для цього беремо перші частинні похідні та прирівнюємо їх до нуля, щоб знайти максимум:

$$\frac{\partial \Lambda}{\partial d_{j1}} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij}\theta_{i1} - \frac{\theta_{i1} \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}{1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}) = \sum_{i=1}^n \sum_{j=1}^m \theta_{i1}(y_{ij} - p_{ij}) = 0;$$

$$\frac{\partial \Lambda}{\partial d_{j2}} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij}\theta_{i2} - \frac{\theta_{i2} \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}{1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}) = \sum_{i=1}^n \sum_{j=1}^m \theta_{i2}(y_{ij} - p_{ij}) = 0;$$

$$\frac{\partial \Lambda}{\partial r_j} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \frac{\exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}{1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}) = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - p_{ij}) = 0;$$

$$\frac{\partial \Lambda}{\partial \theta_{i1}} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij}d_{j1} - \frac{d_{j1} \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}{1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}) = \sum_{i=1}^n \sum_{j=1}^m d_{j1}(y_{ij} - p_{ij}) = 0;$$

$$\frac{\partial \Lambda}{\partial \theta_{i2}} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij}d_{j2} - \frac{d_{j2} \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}{1 + \exp(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)}) = \sum_{i=1}^n \sum_{j=1}^m d_{j2}(y_{ij} - p_{ij}) = 0, \text{ де } i = \overline{1, n}, j = \overline{1, m}.$$

Потім шукаємо другі частинні похідні. Від'ємна визначеність матриці других частинних похідних (гессіана) — це достатня умова існування максимуму. Наведемо тільки подвійні частинні похідні, оскільки вони будуть використані далі при розв'язанні системи:

$$\frac{\partial^2 \Lambda}{\partial d_{j1}^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-\theta_{i1}^2 \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j))}{(1 + \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)))^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-\theta_{i1}^2 \exp(-\varphi_{ij})}{(1 + \exp(-\varphi_{ij}))^2} < 0;$$

$$\frac{\partial^2 \Lambda}{\partial d_{j2}^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-\theta_{i2}^2 \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j))}{(1 + \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)))^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-\theta_{i2}^2 \exp(-\varphi_{ij})}{(1 + \exp(-\varphi_{ij}))^2} < 0;$$

$$\frac{\partial^2 \Lambda}{\partial r_j^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-\exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j))}{(1 + \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)))^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-\exp(-\varphi_{ij})}{(1 + \exp(-\varphi_{ij}))^2} < 0;$$

$$\frac{\partial^2 \Lambda}{\partial \theta_{i1}^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-d_{j1}^2 \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j))}{(1 + \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)))^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-d_{j1}^2 \exp(-\varphi_{ij})}{(1 + \exp(-\varphi_{ij}))^2} < 0;$$

$$\frac{\partial^2 \Lambda}{\partial \theta_{i2}^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-d_{j2}^2 \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j))}{(1 + \exp(-(d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + r_j)))^2} = \sum_{i=1}^n \sum_{j=1}^m \frac{-d_{j2}^2 \exp(-\varphi_{ij})}{(1 + \exp(-\varphi_{ij}))^2} < 0; \text{ де } i=\overline{1,n}, j=\overline{1,m}.$$

Система рівнянь, утворена з перших частинних похідних логарифмічної функції правдоподібності, розв'язується наближено за допомогою ітераційних формул алгоритму Ньютона-Рафсона:

$$\theta_{i1}^{(k+1)} = \theta_{i1}^{(k)} + \frac{\frac{\partial \Lambda}{\partial \theta_{i1}}}{\frac{\partial^2 \Lambda}{\partial \theta_{i1}^2}} = \theta_{i1}^{(k)} - \frac{\sum_{i=1}^n \sum_{j=1}^m d_{j1}(y_{ij} - p_{ij})}{\sum_{i=1}^n \sum_{j=1}^m d_{j1}^2 p_{ij} q_{ij}}, \quad \theta_{i2}^{(k+1)} = \theta_{i2}^{(k)} + \frac{\frac{\partial \Lambda}{\partial \theta_{i2}}}{\frac{\partial^2 \Lambda}{\partial \theta_{i2}^2}} = \theta_{i2}^{(k)} - \frac{\sum_{i=1}^n \sum_{j=1}^m d_{j2}(y_{ij} - p_{ij})}{\sum_{i=1}^n \sum_{j=1}^m d_{j2}^2 p_{ij} q_{ij}}, \text{ де}$$

$i=\overline{1,n}, j=\overline{1,m}$. Помітимо, що останні дві формули можна об'єднати в одну:

$$\theta_{is}^{(k+1)} = \theta_{is}^{(k)} + \frac{\frac{\partial \Lambda}{\partial \theta_{is}}}{\frac{\partial^2 \Lambda}{\partial \theta_{is}^2}} = \theta_{is}^{(k)} - \frac{\sum_{i=1}^n \sum_{j=1}^m d_{js}(y_{ij} - p_{ij})}{\sum_{i=1}^n \sum_{j=1}^m d_{js}^2 p_{ij} q_{ij}}, \text{ де } i=\overline{1,n}, j=\overline{1,m}, s \in \{1; 2\}.$$

Тоді інші формули набудуть вигляду:

$$d_{js}^{(k+1)} = d_{js}^{(k)} + \frac{\sum_{i=1}^n \sum_{j=1}^m \theta_{is}(y_{ij} - p_{ij})}{\sum_{i=1}^n \sum_{j=1}^m \theta_{is}^2 p_{ij} q_{ij}}, \quad r_j^{(k+1)} = r_j^{(k)} + \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - p_{ij})}{\sum_{i=1}^n \sum_{j=1}^m p_{ij} q_{ij}}, \text{ де } i=\overline{1,n}, j=\overline{1,m}, s \in \{1; 2\}.$$

Звичайно, що тоді виникає питання вибору початкових значень для ітераційних формул. У даному випадку $r_j^{(0)} = \ln \frac{Q_j}{P_j}$, де $\frac{Q_j}{P_j}$ — це відношення кількості правильних відповідей до неправильних на j -те питання; $\theta_{is}^{(0)} = \ln \frac{q_i}{p_i}$, де $\frac{q_i}{p_i}$ — це відношення кількості правильних відповідей до неправильних i -го іспитника; а $d_{js}^{(0)} = 0$.

У силу великого об'єму обчислень для реалізації цих ітераційних формул та пошуку невідомих параметрів — у магістерській дисертації використовувалося спеціальне програмне забезпечення та спеціальні пакети R.

4.4 Багатовимірна модель Муракі

Для цієї моделі ймовірність того, що i -ий іспитник правильно відповість у j -му завданні на k підпитань дорівнює:

$$p_{ijk} = \frac{\exp(k \sum_{l=1}^s d_{jl} \theta_{il} + \sum_{q=0}^k r_{jq})}{\sum_{v=0}^m \exp(v \sum_{l=1}^s d_{jl} \theta_{il} + \sum_{q=0}^v r_{jq})},$$

де n — це кількість іспитників, m_j — це кількість рівнів в завданні, $i=\overline{1,n}$; $j=\overline{1,m}$; $k=\overline{0,m_j}$; $l=\overline{1,s}$; s — вимірність вектора підготовленостей; m_j — кількість рівнів у завданні; d_j — диференціююча здатність j -го завдання; θ_i — рівні підготовленості i -го іспитника; r_{jq} — параметри складності j -го завдання q -го рівня.

Окрім того помітимо, що багатовимірна складність k -го кроку в j -му завданні обчислюється за наступною формулою:

$$R_{jk} = \frac{-r_{jk}}{\sqrt{\sum_{q=1}^s d_{jq}^2}}, j_k=\overline{1,m}, k=\overline{1,m_j}. \quad [8]$$

а багатовимірна диференціююча спроможність завдань вираховується за формулою:

$$D_{jk} = \sqrt{\sum_{q=1}^s d_{jq}^2}, j_k=\overline{1,m}. \quad [8]$$

5 ПЕРЕВІРКА АДЕКВАТНОСТІ МОДЕЛЕЙ

Перевірка адекватності моделей буде здійснюватися за допомогою аналога критерія χ^2 . Для одновимірних моделей IRT використовуватиметься функція Goodness-of-Fit, а для багатовимірних моделей MIRT – M2. Обидва ці методи ґрунтуються на статистиці Пірсона:

$$\sum_{r=1}^{2^p} \frac{\{O(r) - E(r)\}^2}{E(r)},$$

де p — це кількість питань в тесті, $O(r)$ — це спостережувана частота відповідей на питання в тесті, а $E(r)$ — це очікувана частота відповідно.

Рівень статистичної значущості в усіх перевірках моделей буде покладатися $\alpha=0,05$. Надалі порівнюватимуться отримані результати саме з таким рівнем значущості. Якщо результат буде більшим, ніж 0.05, то будемо робити висновок, що модель адекватна. [7]

6 КРИТЕРІЇ ВИБОРУ МОДЕЛІ

Існує багато критеріїв вибору моделі. Розглянемо деякі з них.

6.1 Критерій Акаїке (AIC)

AIC (Akaike's information criterion) – це один із критеріїв вибору моделі. Критерій японського математика Акаїке, як його ще інакше називають. Коли тести складаються з елементів із множинним вибором та вбудованою (тобто запропонованою) відповіддю, дослідники стикаються з питанням, яка комбінація моделей теорії IRT буде відповідним чином представляти дані, зібрані з цих тестів змішаного формату. Імовірно, найбільш визнаним інформаційним критерієм є критерій Акаїке (AIC), який обчислюється наступним чином:

$$AIC = d + 2p,$$

де d – статистика відхилення для даної моделі або інакше $d = -2\Lambda = -2\ln(L)$, а p – кількість параметрів, оцінена в даній моделі. При порівнянні двох конкуруючих моделей, модель із найнижчим значенням AIC буде обрана як модель, що демонструє кращу відповідність даним. Однією з характеристик AIC є те, що вона, як правило, схильна вибирати більш високопараметризовані моделі. Крім того цей критерій є асимптотично ефективним, тобто вибирає модель, яка найкраще наближує правильну/справжню модель, коли правильна/істинна модель насправді не входить до набору моделей порівняння. Хоча AIC є ефективним критерієм вибору моделі, він не є послідовним. Послідовні критерії вибору дозволяють вибрати правильну/істинну модель з імовірностями, близькими до або однаковими до 1, коли правильна/істинна модель насправді є серед набору моделей порівняння.

6.2 Баєсівський інформаційний критерій (BIC) або критерій Шварца (SC)

BIC (Bayesian information criterion) – це розширений критерій AIC завдяки Шварцу, який враховує не лише складність моделі, але й обсяг вибірки:

$$BIC = d + p \cdot \ln(N),$$

де N – це розмір вибірки.

Цей критерій застосовується в політомічних моделях IRT (коли варіантів відповідей більше, ніж два).

Ці критерії найкраще застосовувати до одно- та двопараметричних моделей, тому що уже в трьопараметричних вони виявляються не такими ефективними. [9-12]

7 ДОСЛІДЖЕННЯ ТА АНАЛІЗ ДАНИХ

7.1 Первинна обробка результатів тестування з аналітичної геометрії

Об'єктом дослідження є контрольна робота з аналітичної геометрії, яка проводилась в осінньому семестрі 2020 року на ФІОТ-і НТУУ “КПІ ім. Ігоря Сікорського” серед 105 студентів. Її аналіз проводився за допомогою змішаної моделі: 2, 3, 4, 7, 10 завдання були “числові”, а інші — “вбудовані відповіді”. Тестові питання типу “вбудовані відповіді” являються найбільш гнучкими завданнями, в яких можна поєднувати питання типу “числове”, “коротка відповідь” або “множинний вибір”, “на відповідність”. Це дозволяє перевіряти проміжні етапи розв’язання задачі, краще диференціювати знання студентів, визначати помилки у розв’язанні, ускладнювати процес використання допоміжних ресурсів.

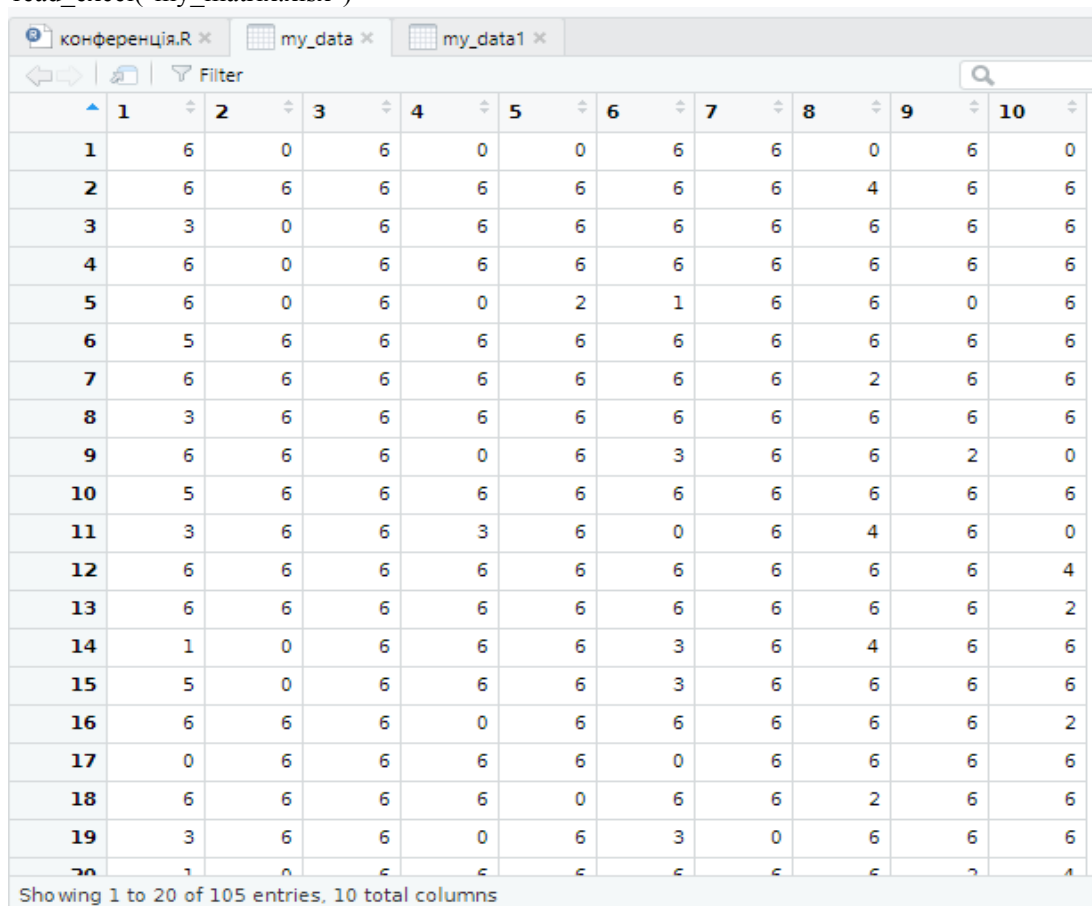
Завдання на множинний вибір можна досліджувати дихотомічними моделями, решту — політомічними. Завдання під №1 та №2 містили 4 питання, а завдання під №5, №8 і №9 мали 3 питання. Необхідно також зазначити, що завдання №1-3 та №7 були створені для перевірки вмінь студентами використовувати властивості векторів; а завдання №4-6 та №8-10 перевіряли знання прямих і кривих на площині, а також знання площин у просторі. [13]

Інструментом для досліджень виступила мова програмування R. Вона є безкоштовною та поширюється завдяки ліцензії GNU General Public License. Ця мова містить багато бібліотек, які можна додатково завантажити та використовувати у роботі за необхідності, наприклад, коли треба виконати статистичні обчислення, проаналізувати та зобразити дані у графічному вигляді. Це дуже зручно. Окрім того, існують спеціально створені бібліотеки для IRT та MIRT моделей. Усі функції та пакети можна знайти на сайті Comprehensive R Archive Network (CRAN). [14]

Етап математично-статистичної обробки матриці балів, що була завантажена із платформи Moodle, полягає в тому, щоб прибрати всі зайві стовпці, наприклад,

“затрачений час” або “загальний бал за тест”; зайві рядки, наприклад, “середній бал”, результати, де студенти отримали 0 балів за всі завдання, або максимальний бал. [1, 2, 5] Після того необхідно виконати операцію шкалування: 0 балів залишити нулем, наступному балу призначити значення 1, і т.д., тобто розглядаємо в подальшому натуральні числа. Це робиться для того, щоб можна було застосовувати дихотомічні або політомічні моделі. Нижче наведено частину програми, яка це все виконує та результати:

```
while (i<=10) {k <- 1
while (k<=105) {if (!any(my_vector==my_matrix[k,i])) {my_vector <- c(my_vector, my_matrix[k,i])}
k <- k+1}
i <- i+1}
my_vector <- sort(my_vector)
for (i in 1:105) {for (j in 1:10) {my_matrix[i,j] <- match(my_matrix[i,j], my_vector) - 1}}
print(my_matrix)
library('openxlsx')
write.xlsx(my_matrix, "my_matrix.xlsx")
my_data <- read_excel("my_matrix.xlsx")
```



	1	2	3	4	5	6	7	8	9	10
1	6	0	6	0	0	6	6	0	6	0
2	6	6	6	6	6	6	6	4	6	6
3	3	0	6	6	6	6	6	6	6	6
4	6	0	6	6	6	6	6	6	6	6
5	6	0	6	0	2	1	6	6	0	6
6	5	6	6	6	6	6	6	6	6	6
7	6	6	6	6	6	6	6	2	6	6
8	3	6	6	6	6	6	6	6	6	6
9	6	6	6	0	6	3	6	6	2	0
10	5	6	6	6	6	6	6	6	6	6
11	3	6	6	3	6	0	6	4	6	0
12	6	6	6	6	6	6	6	6	6	4
13	6	6	6	6	6	6	6	6	6	2
14	1	0	6	6	6	3	6	4	6	6
15	5	0	6	6	6	3	6	6	6	6
16	6	6	6	0	6	6	6	6	6	2
17	0	6	6	6	6	0	6	6	6	6
18	6	6	6	6	0	6	6	2	6	6
19	3	6	6	0	6	3	0	6	6	6
20	1	0	6	6	6	6	6	6	2	4

Рис. 7.1.1 — Дані, що пройшли первинну обробку для

подальшого застосування політомічної моделі

Первинний статистичний аналіз матриці балів іспитників включає в себе

побудову гістограми балів. Наведемо частину програми для побудови гістограми:

```
suma <- c()
for (i in 1:105) {SUM <- 0
  for (j in 1:10){SUM <- SUM + my_matrix[i,j]}
  suma <- c(suma,SUM)}
hist(suma,col=10,main="Histogram of Person Parameters")
```

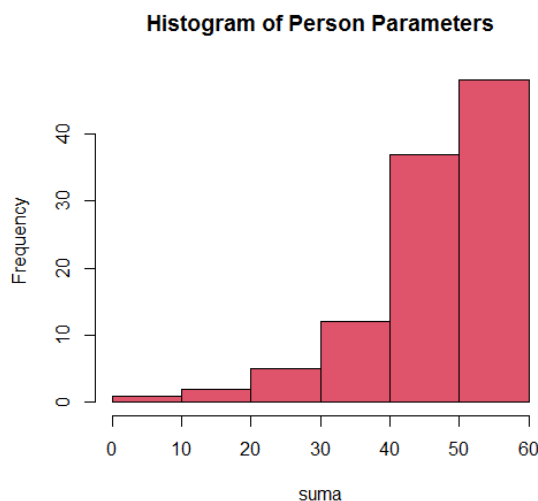


Рис. 7.1.2 — Гістограма частот сумарних балів іспитників

Гістограма сумарних балів іспитників має зміщення вправо – це пояснюється високим рівнем математичних знань іспитників даної спеціальності (бал ЗНО по математиці вище 185 балів). [15]

Потім виконувались обчислення коефіцієнту Кронбаха. Його значення вийшло рівним 0.704. Це означає, що тест складений вдало і він має високу надійність.

Далі до перетвореної матриці балів будуть застосовуватися три моделі.

7.2 Змішана модель або модель А

Ця модель являється комбінацією двох: одновимірної 2-PL моделі та багатовимірної моделі Муракі. Варто зазначити, що інколи таку модель називають мішаною або змішаною. Насправді досить використати модель Муракі (GPCM) до усіх питань, оскільки вона є узагальненням 2-PL моделі. І обов'язково відповіді на питання під номерами 2-4, 7 та 10 необхідно перетворити в нулі або одинички, щоб спостерігались графіки результатів дихотомічних питань. Нижче наведено

результати програми для побудови інформаційної кривої третього питання тесту:

```
my_data1 <- read_excel('my_matrix_dich.xlsx')
# install.packages('mirt')
#install.packages('latticeExtra')
library(mirt)
library(latticeExtra)
#polytomous items
A_model <- mirt(my_data1, 1, SE=TRUE)
itemplot(A_model, 3, type = 'infotrace')
itemplot(A_model, 3, type = 'score')
```

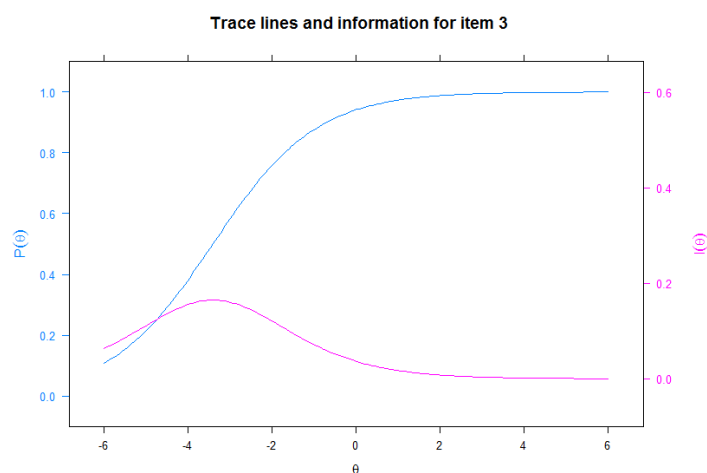


Рис. 7.2.1 — ICC або характеристична крива складності (позначено синім) та інформаційна крива завдання №3 (позначено рожевим)

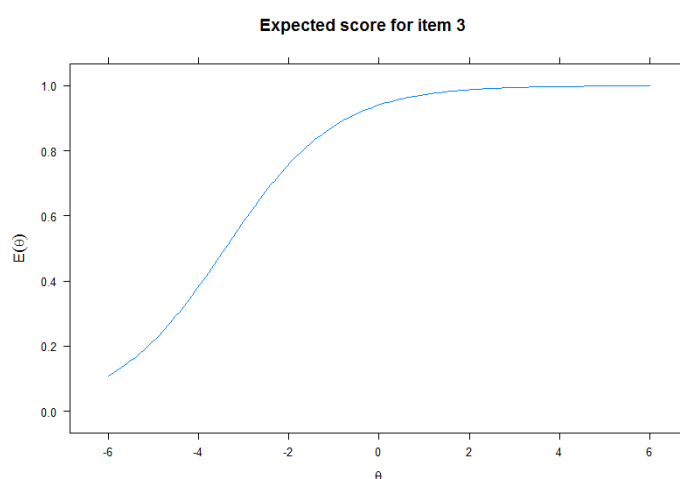


Рис. 7.2.2 — Очікувані оцінки за питання №3
іспитників з різними рівнями підготовленості

На рис. 7.2.1 зображено блакитним ІІС або Item Information Curve – це характеристична крива рівня складності завдань, що показує ймовірність правильного рішення завдання студентами з різними рівнями підготовленості при

умові, що складність завдань зростає. Хоча теоретичний діапазон аргументів ІСС знаходиться на проміжку $(-\infty; +\infty)$, проте на практиці зазвичай зустрічається скінченний відрізок $[-6; +6]$.

Рожевим кольором зображено інформаційну функцію завдання №3 для 2-PL моделі. Інформаційна функція розраховується так само, як і у випадку 1-PL моделі — як добуток ймовірності правильної відповіді на ймовірність неправильної. Але в ній виникає деяка відмінність від інформаційної функції 1-PL моделі — це множник у вигляді квадрату диференціюючої здатності завдання. Наслідком є те, що чим більше значення диференціюючої здатності — тим більшу інформацію в собі несе питання. Оскільки диференціюючий параметр може змінюватися залежно від завдання, то графіки функцій інформативності завдань теж можуть виглядати по-різному. Аналітичний запис функцій інформативності завдання для 2-PL моделі:

$$I_j(\theta_i, d_j, r_j) = d_j^2 p_{ij} q_{ij}.$$

На рис. 7.2.2 зображено очікувані бали питання №3 іспитників з різними рівнями підготовленості. Окрім цього можна побудувати графіки для усіх питань.

```
i <- 1
while (i <= 10) {
  print(itemplot(A_model, i))
  print(itemplot(A_model, i, type = 'infotrace'))
  i <- i+1}
```

Результати наведені в наступній таблиці.

Таблиця 7.2.1

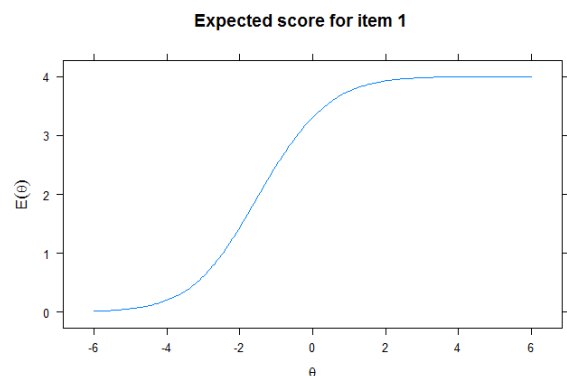
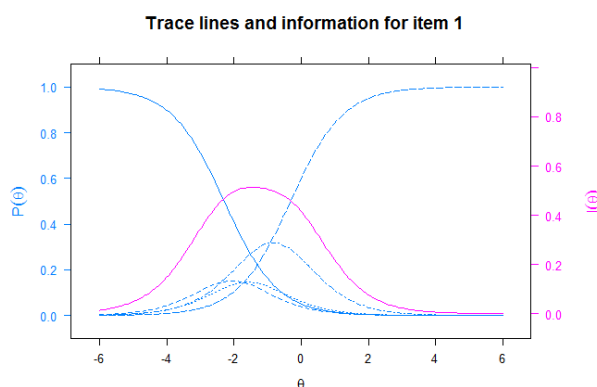
Характеристичні криві рівня складності завдання,
інформаційні криві та графіки очікуваних
балів іспитників з кожного питання тесту

№ пи-
тання

1

ІСС та інформаційна крива завдання

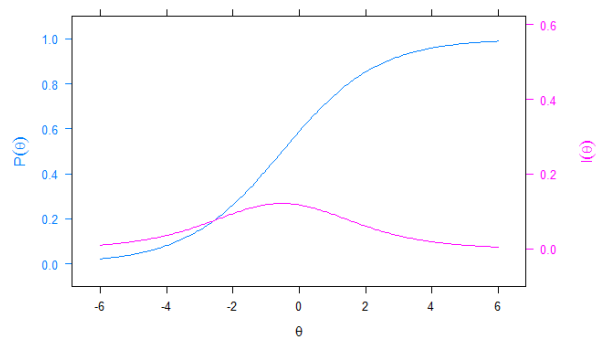
Очікувані бали іспитників



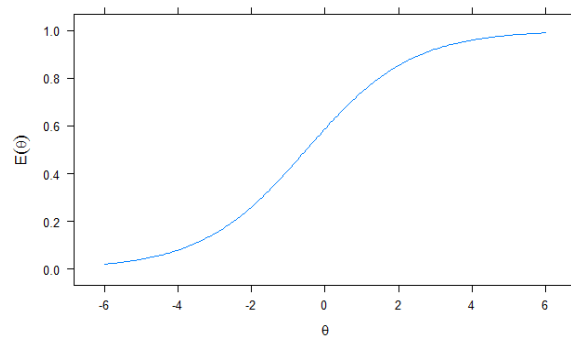
Продовження табл. 7.2.1

2

Trace lines and information for item 2

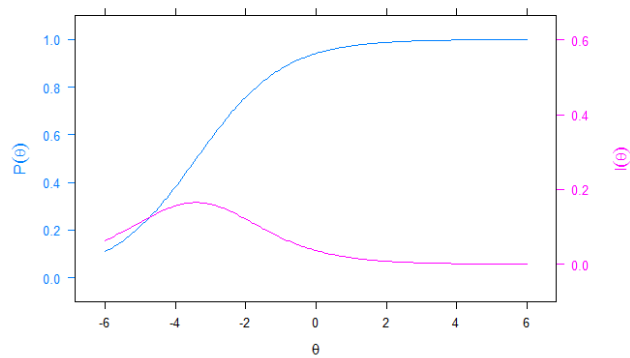


Expected score for item 2

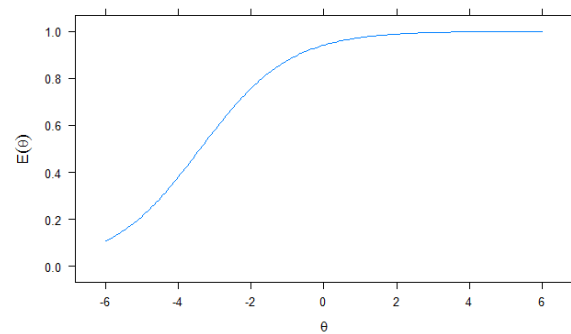


3

Trace lines and information for item 3

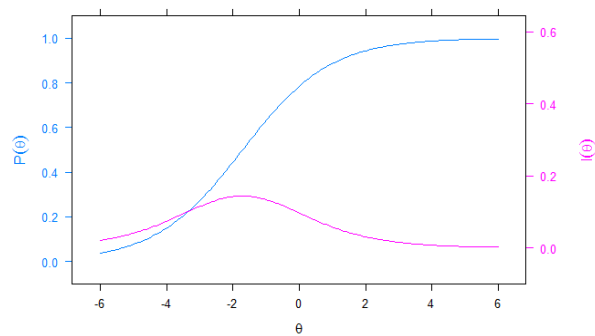


Expected score for item 3

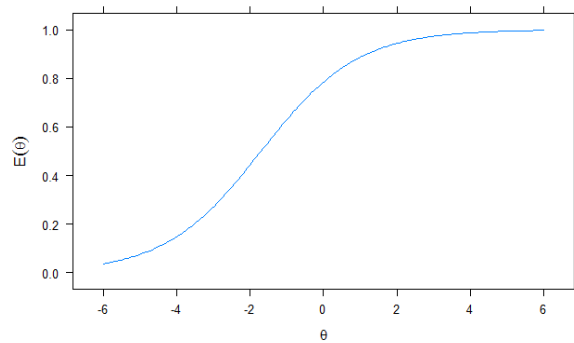


4

Trace lines and information for item 4

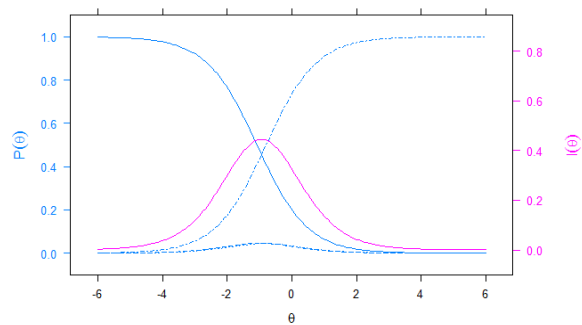


Expected score for item 4

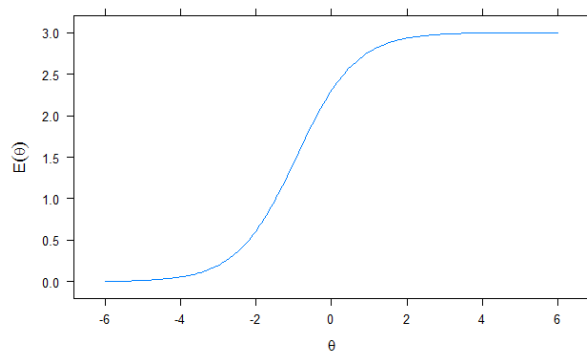


5

Trace lines and information for item 5



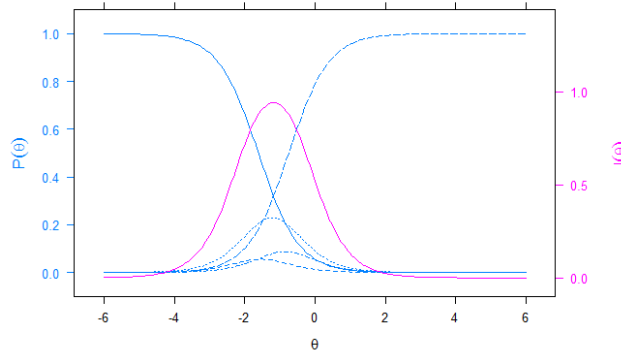
Expected score for item 5



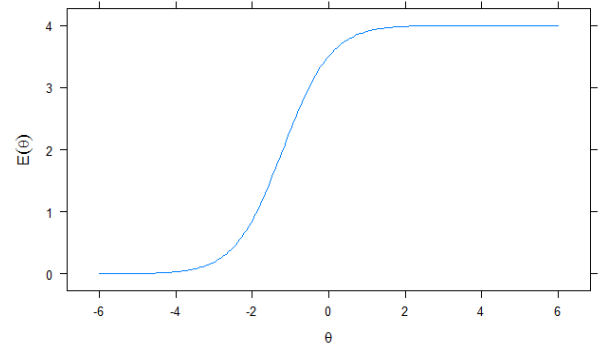
Продовження табл. 7.2.1

6

Trace lines and information for item 6

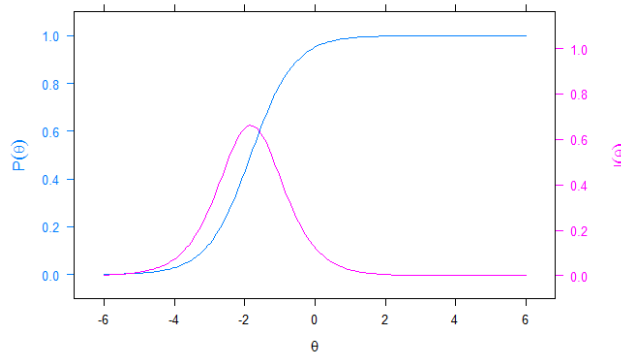


Expected score for item 6

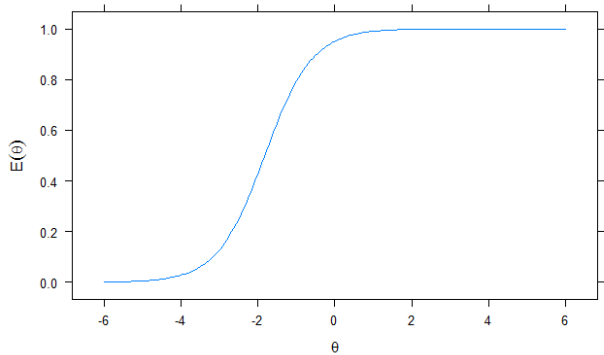


7

Trace lines and information for item 7

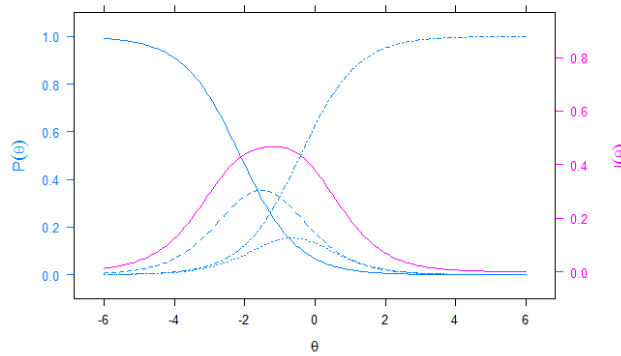


Expected score for item 7

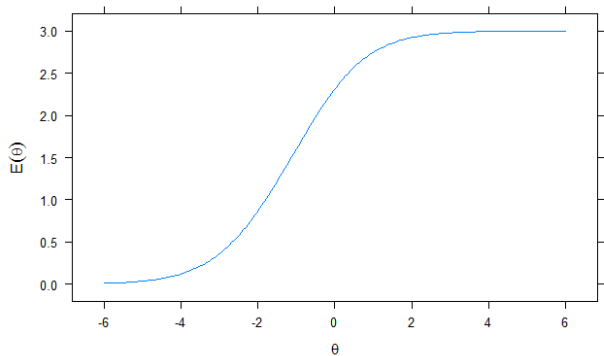


8

Trace lines and information for item 8

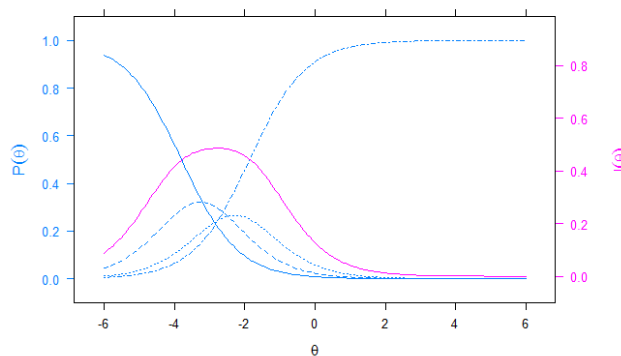


Expected score for item 8

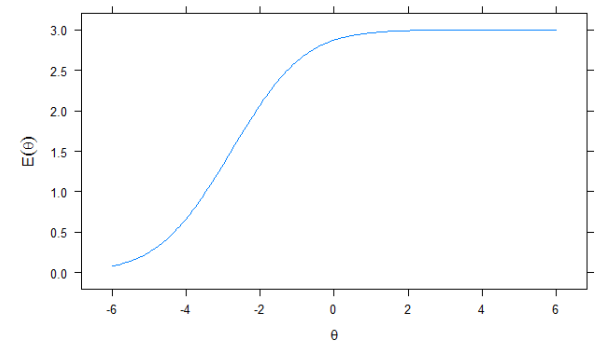


9

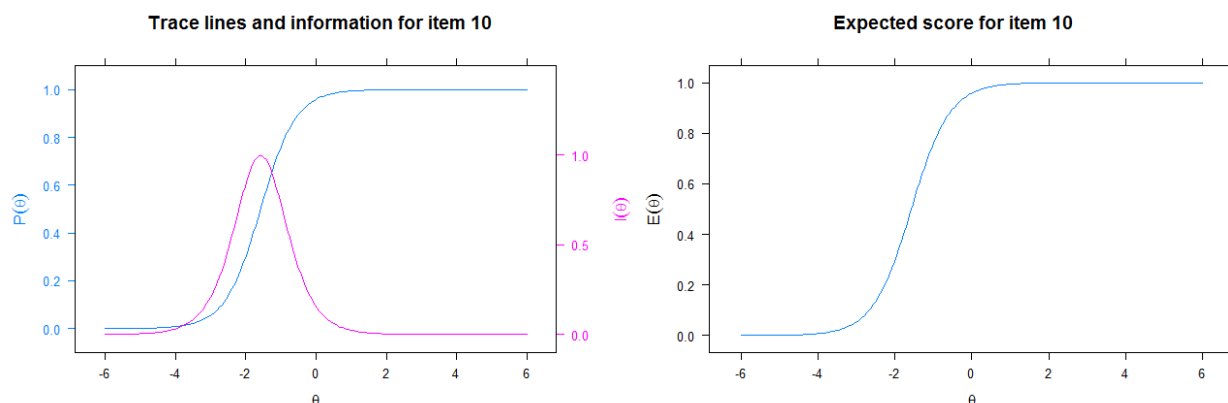
Trace lines and information for item 9



Expected score for item 9



10



```
plot(A_model,type="trace")
```

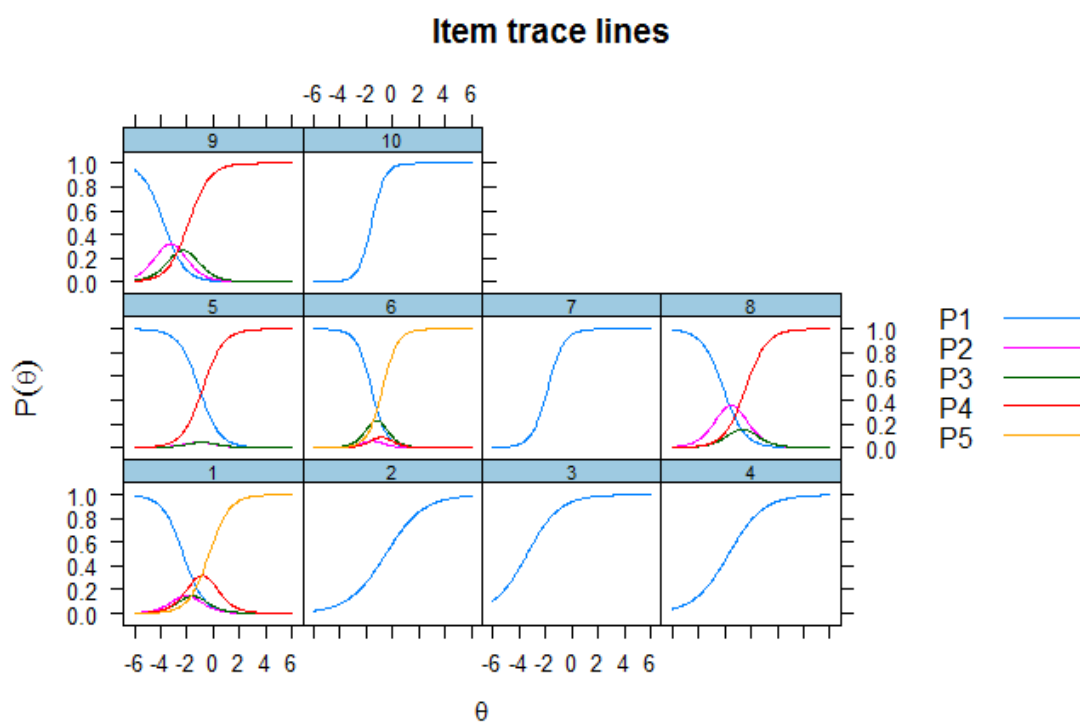


Рис. 7.2.3 — Характеристичні криві рівня складності завдань

По графікам, що наведені в таблиці 7.2.1 зліва та з рис. 7.2.3, чітко видно, що завдання №2-4, 7, 10 дихотомічні, а всі інші — політомічні. Можна зробити висновок, що у цілому дихотомічні завдання складено правильно. У них закладена помірна складність та додатна диференціююча здатність. Також питання під №3 та №7 являються найпростішими. Окрім цього перше завдання є нескладним, а підпитання в ньому поводять себе відносно однаково. Третє підпитання в завданні

№5 та 4 підпитання в завданні №6 мають найнижчу складність. Перші питання п'ятої та шостої завдань являються найскладнішими. [13]

Також можна порівняти результати, які були отримані графічно, з аналітичними. Для цього використаємо метод для виведення коефіцієнтів моделі:

`coef(A_model)`

Таблиця 7.2.2

Коефіцієнти GPCM					
	d	r_1	r_2	r_3	r_4
1	0.448	0.372	-1.316	-2.876	-2.056
2	0.561	-0.621			
3	0.916	-3.096			
4	1.539	-0.703			
5	0.403	4.931	-0.188	-7.779	
6	0.457	3.192	-4.286	1.498	-5.796
7	1.504	-1.913			
8	0.655	-1.631	0.381	-2.397	
9	0.757	-2.431	-1.706	-3.271	
10	1.253	-1.346			

Видно, що характеристичні криві складності завдань із табл. 7.2.1 та рис. 7.2.3 відповідають оціненим параметрам із табл. 7.2.2. Зокрема, якщо поглянути на п'яте та шосте питання, то видно, що деякі підзавдання «просідають» в силу високої складності. Окрім цього можна також побудувати об'єднану інформаційну криву для всього тесту: `plot(A_model, type = 'info')`

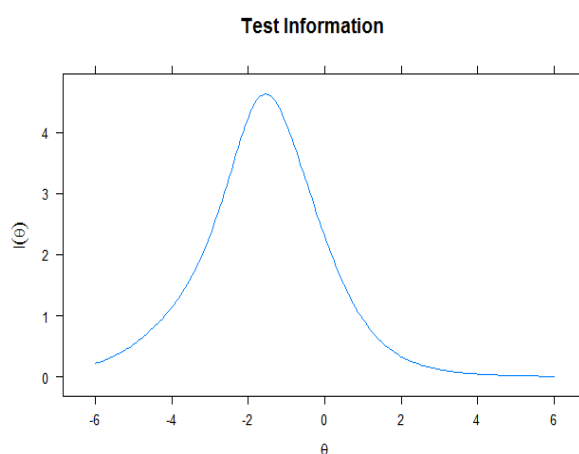


Рис. 7.2.4 — Інформаційна крива всіх тестових питань

Із рис. 7.2.4 видно, що графік дзвоноподібний. Отже, тест складено вдало. Проте максимум кривої зміщений у ліву сторону. Звідси можна зробити висновок,

що тест буде найбільш інформативний для слабо підготовлених студентів.

Останній крок — це перевірка адекватності моделі. Для цього використаємо аналог критерія χ^2 — функцію Goodness of Fit (G2):

```
extract.mirt(A_model, 'G2')
extract.mirt(A_model, 'p')
```

Оскільки $p\text{-value}=0.138>0.05$, то модель адекватна.

7.3 Двовимірна модель Муракі або модель В

Це багатовимірна модель, що містить дві ознаки. Для аналізу завдань №2-4, 7, 10 використовуватимемо багатовимірну 2-PL модель Бірнбаума. Надалі вважається, що $\theta=(\theta_{i1}, \theta_{i2})$ — це рівні підготовленості i -го іспитника. Зокрема, знання векторної алгебри — це компетентність θ_{i1} , а знання прямих, кривих, площин (застосування векторів) — це компетентність θ_{i2} .

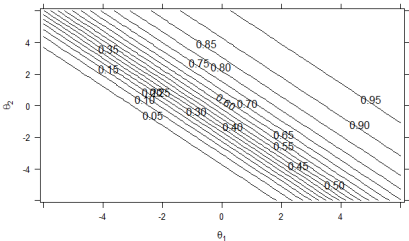
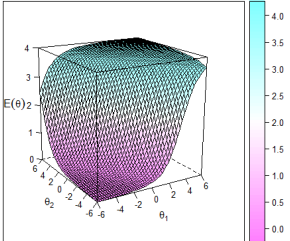
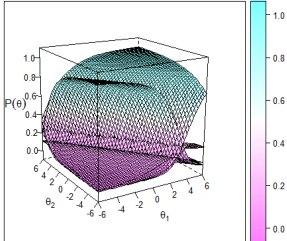
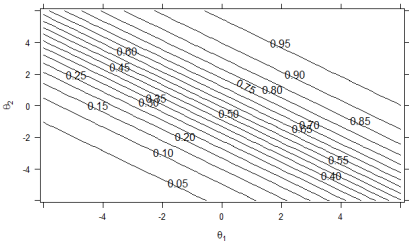
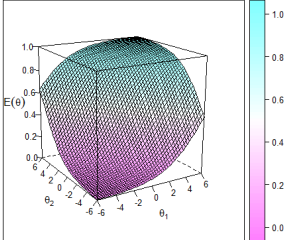
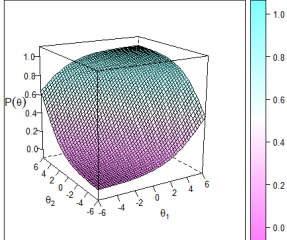
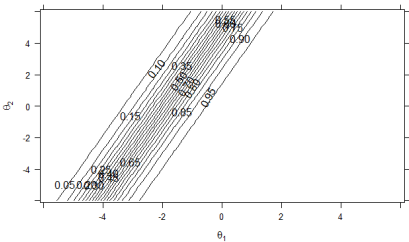
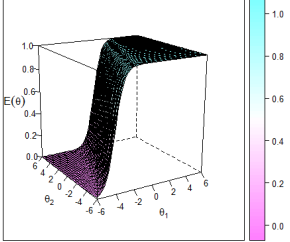
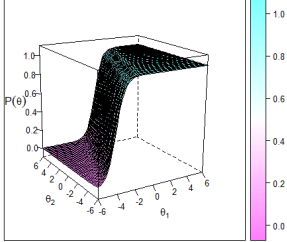
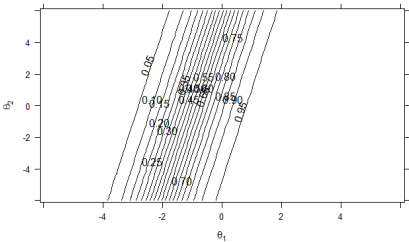
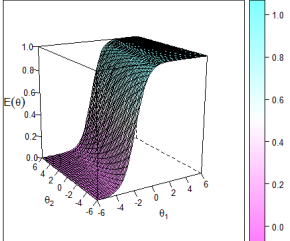
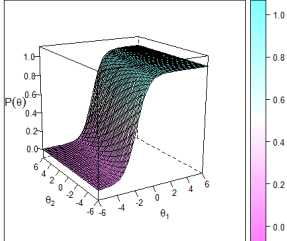
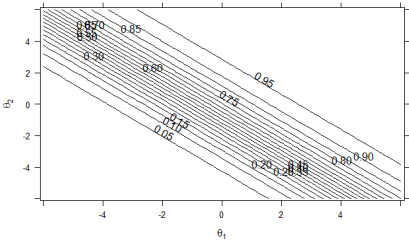
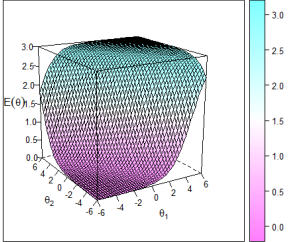
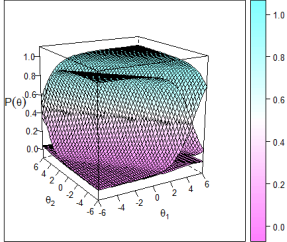
```
B_model <- mirt(my_data1, 2, SE=TRUE)
vector_of_types <- c("gpcm", rep("2PL", 3), rep("gpcm", 2), "2PL", rep("gpcm", 2), "2PL")
bmod <- mirt(my_data1, pol_model, itemtype = vector_of_types, method = 'MHRM', verbose=FALSE)
i <- 1
while (i<=10) {
  print(itemplot(bmod, type='tracecontour', i))
  print(itemplot(bmod, type='score', i))
  print(itemplot(bmod, i))
  i <- i+1
}
```

Результати можна подивитися в табл. 7.3.1 на наступній сторінці. Аналізуючи лінії рівня, можна зробити висновок, що для завдань під №3 і №4 ансамблі контурних прямих мають додатній коефіцієнт нахилу. На противагу — в інших завданнях кут нахилу ансамблю ліній рівня має від'ємний кутовий коефіцієнт. Ці результати можна трактувати наступним чином: незнання векторів не можна компенсувати знанням прямих, площин або поверхонь в МКР.

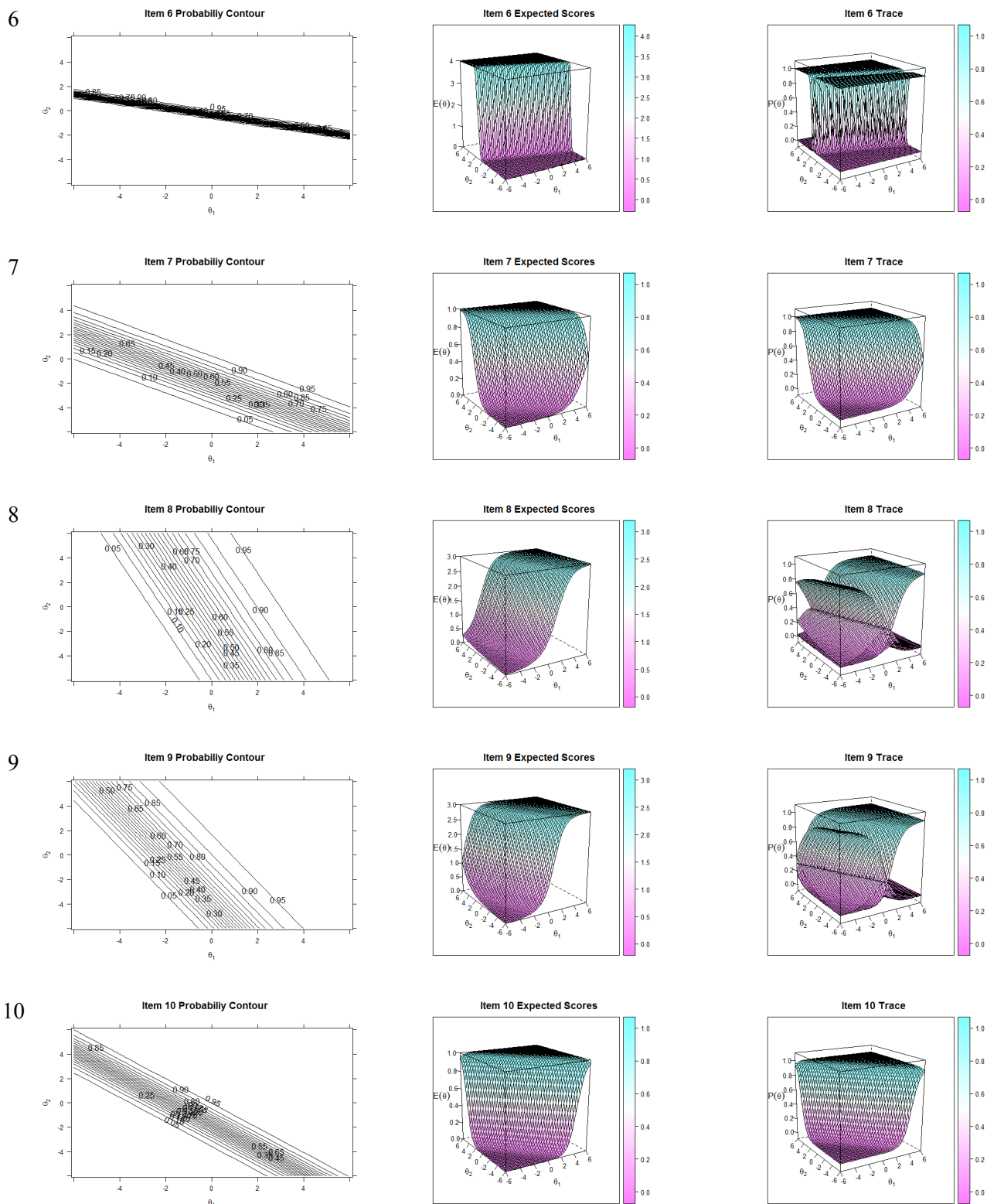
Також, наприклад, завдання №6 має високу диференціюючу здатність, що пояснюється високою щільністю ансамблю контурних прямих.

Таблиця 7.3.1

Лінії рівня, характеристичні поверхні складності питань
та поверхні очікуваних оцінок іспитників для моделі В

№ пит.	Контурні прямі або лінії рівня для завдання	Поверхня очікуваних оцінок іспитників	Характеристична поверхня складності завдання
1	Item 1 Probability Contour 	Item 1 Expected Scores 	Item 1 Trace 
2	Item 2 Probability Contour 	Item 2 Expected Scores 	Item 2 Trace 
3	Item 3 Probability Contour 	Item 3 Expected Scores 	Item 3 Trace 
4	Item 4 Probability Contour 	Item 4 Expected Scores 	Item 4 Trace 
5	Item 5 Probability Contour 	Item 5 Expected Scores 	Item 5 Trace 

Закінчення табл. 7.3.1



Із наведеного вище аналізу та графіків табл. 7.3.1 можна зробити висновок, що ймовірно, використання двовимірної компенсаторної моделі для аналізу тесту з аналітичної геометрії на ФІОТ-і є недоречним. Компроміс — спробувати

використати некомпенсаторну модель.

Можна порівняти отримані результати графічно з аналітичними:

coef(bmod)

Табл. 7.3.2

Коефіцієнти дихотомічних завдань двовимірної моделі

	d_{j1}	d_{j2}	r_j	D_j	R_j
2	-0.499	0.33	0.35	0.60	-0.59
3	-1.559	-0.71	3.528	1.71	-5.90
4	-2.515	-0.064	1.426	2.52	-2.38
7	-1.043	1.249	2.978	1.63	-4.98
10	-1.777	0	1.969	1.78	-3.29

Табл. 7.3.3

Коефіцієнти політомічних завдань двовимірної моделі

	d_{j1}	d_{j2}	r_{j1}	r_{j2}	r_{j3}	r_{j4}	D_j	R_{j1}	R_{j2}	R_{j3}	R_{j4}
1	-0.318	0.322	-0.164	0.43	1.722	2.643	0.45	0.36	-0.95	3.81	-5.84
5	-0.268	0.277	-2.002	-1.927	1.206		0.39	5.19	5	-3.13	
6	-1.092	2.48	1.6	5.821	6.718	9.599	2.71	-0.59	-2.15	-2.48	-3.54
8	-0.611	0.126	1.035	0.769	2.343		0.62	-1.66	-1.23	-3.76	
9	-0.677	0.319	1.83	3.123	5.684		0.75	-2.45	-4.17	-7.59	

Якщо проаналізувати табл. 7.3.2 і табл. 7.3.3, то можна побачити, що всі завдання є помірно складними, а, в цілому, складність тесту є невисокою. Найпростішим знову є завдання під №3 та №7.

Адекватність моделі оцінювалась знову за допомогою аналога критерія χ^2 :
M2(bmod)

Оскільки $p\text{-value}=0.727>0.05$, то модель адекватна.

7.4 Багатовимірна двофакторна модель Муракі або модель С

Це багатовимірна модель, в якій частині завдань надається одна ознака, тобто один фактор, а іншій частині завдань — інша ознака, тобто другий фактор. Зокрема для МКР з аналітичної геометрії було сформовано всього дві ознаки, по яким питанням під №1-3 та №7 відповідали знання з теми “вектори”, а іншим питанням — “прямі, площини, поверхні”.

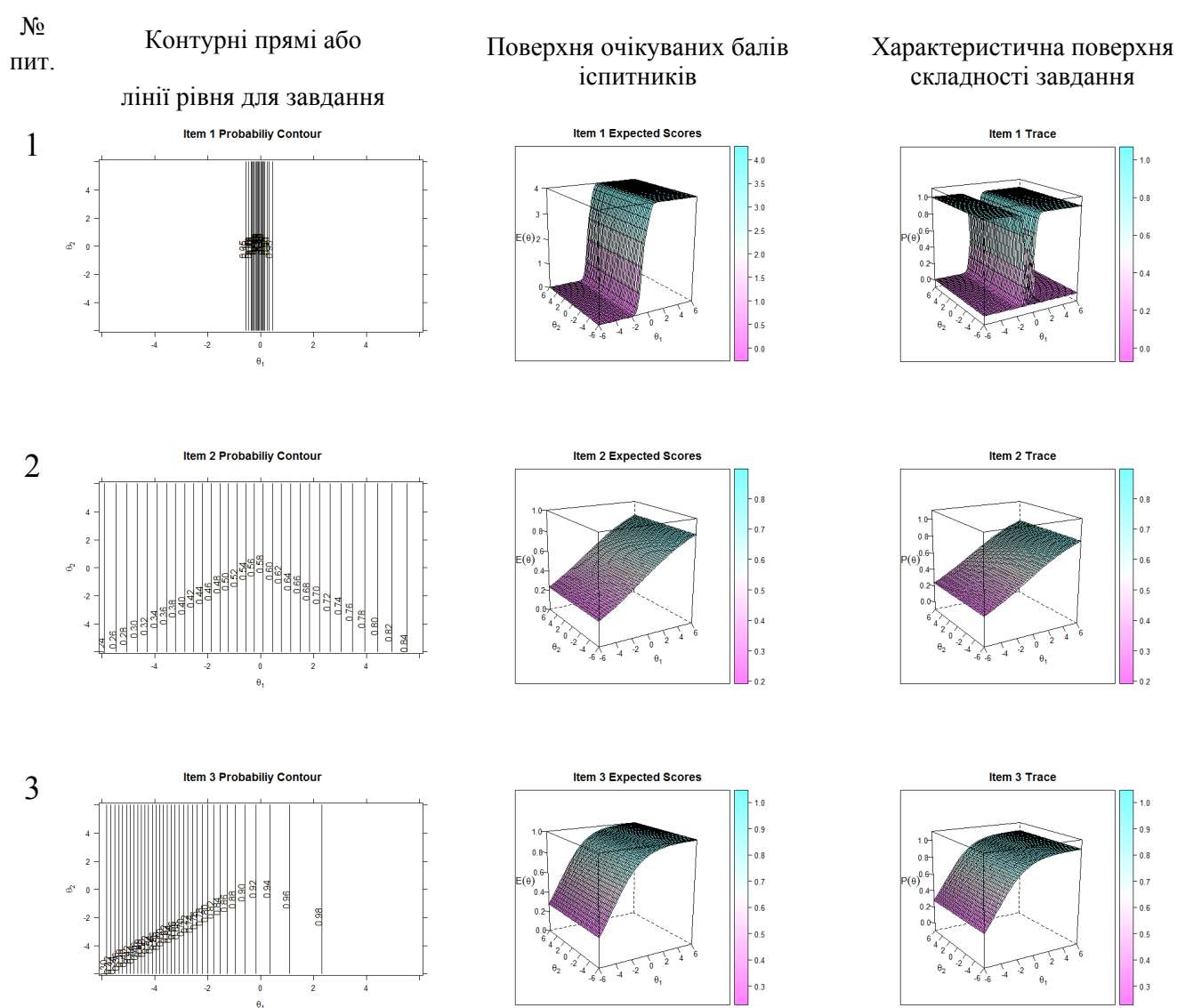
```

library(mirt)
C_model <- mirt.model('F1 = 1-3,7
                     F2=4-6,8-10')
vector_of_types <- c(rep("gpcm",10))
cmmod <- mirt(my_data, C_model, itemtype = vector_of_types, method = 'MHRM', verbose=FALSE)
i <- 1
while (i<=10) {
  print(itemplot(cmmod,type='tracecontour',i))
  print(itemplot(cmmod,type='score',i))
  print(itemplot(cmmod,i))
  i <- i+1
}

```

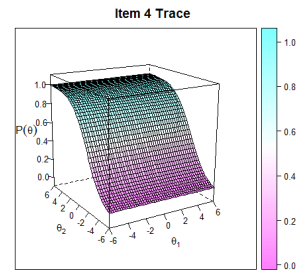
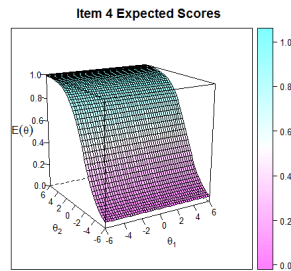
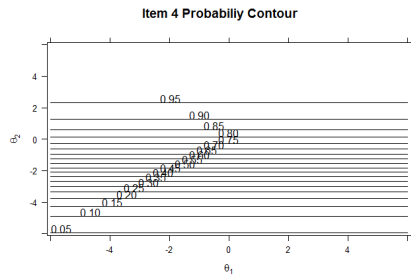
Таблиця 7.4.1

Лінії рівня, характеристичні поверхні складності питань
та поверхні очікуваних балів іспитників для моделі С

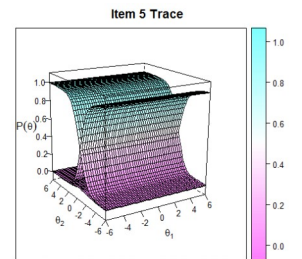
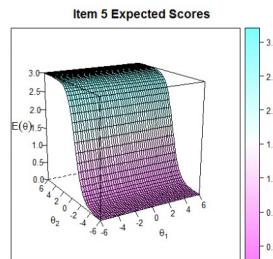
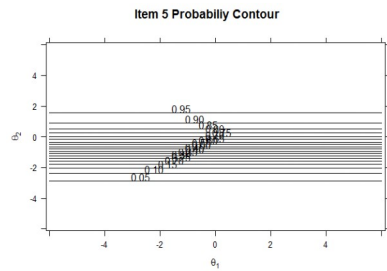


Продовження табл. 7.4.1

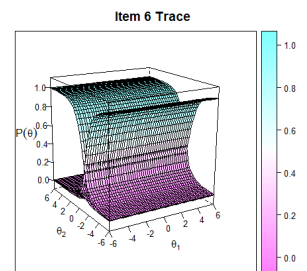
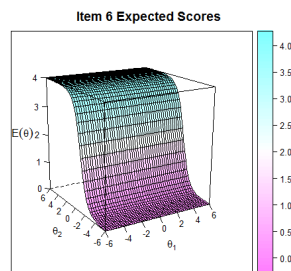
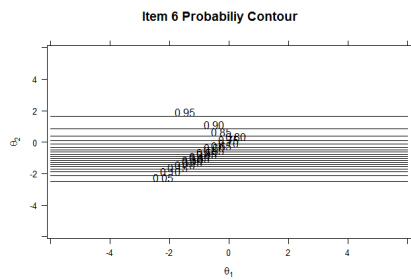
4



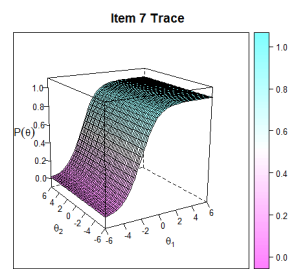
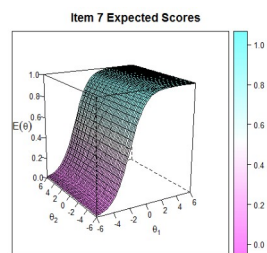
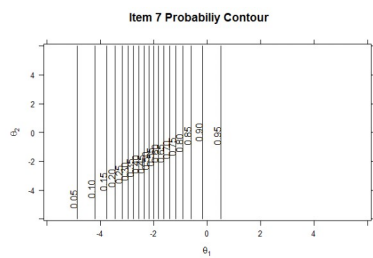
5



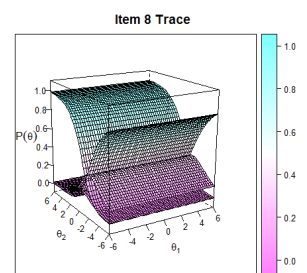
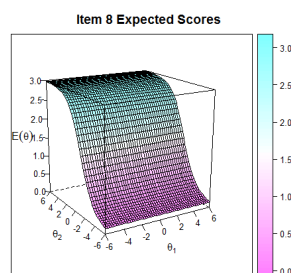
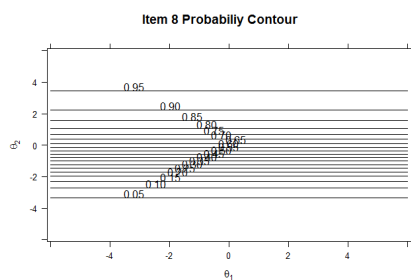
6



7

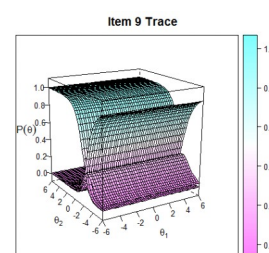
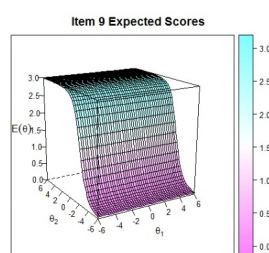
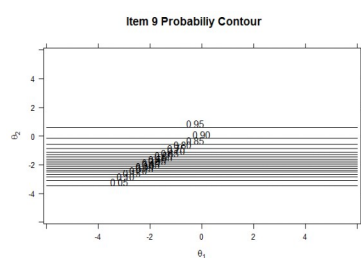


8

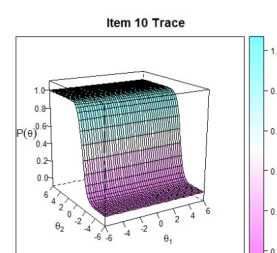
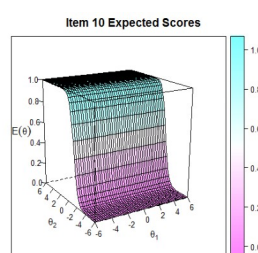
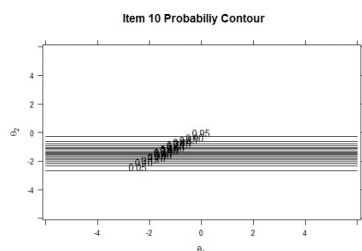


Закінчення табл. 7.4.1

9



10



Ансамблі контурів ймовірностей набувають вигляду вертикальних або горизонтальних прямих в силу того, що кутовий коефіцієнт 0 або 1, бо одним завданням відповідала одна компетентність, а іншим — друга ознака.

Із табл. 7.4.1 видно, що контури досить рівномірні, окрім питань №1, 5, 6, 9, 10. А високу щільність цих графіків можна пояснити високим рівнем диференціюючої здатності питань, що відображається на відповідних характеристичних поверхнях. Окрім того, питання №2 краще доробити або переформулювати.

Отримані графічним способом результати знову можна порівняти з аналітичними наступним чином:

coef(cmod)

Таблиця 7.4.2

Коефіцієнти дихотомічних завдань двофакторної моделі

	d_1	d_2	r	R
2	0.235	0	0.288	-1.22553
3	0.589	0	2.518	-4.27504
4	0	1.778	1.172	-0.65917
7	1.119	0	2.34	-2.09115
10	0	1.771	1.967	-1.11067

Таблиця 7.4.3

Коефіцієнти поліномічних завдань двофакторної моделі

	d_1	d_2	r_1	r_2	r_3	r_4	R_1	R_2	R_3	R_4
1	0	9.035	9.583	16.974	22.49	22.678	-1.06	-1.88	-2.49	-2.51
5	0	0.336	-2.044	-1.986	1.156		0.23	0.22	-0.13	
6	0.37	0	-1.584	0.303	-0.41	2.249	0.18	-0.03	0.05	
8	0	0.595	0.987	0.713	2.294		-0.11	-0.08	-0.25	
9	0	0.902	2.04	3.525	6.149		-0.23	-0.39	-0.68	

Якщо провести аналіз табл. 7.4.2 та табл. 7.4.3, то можна спостерігати наступну картину: так само, як і в попередніх результатах, найпростішими виявились питання під №3 та №7. Диференціююча здатність виявилась невід’ємною. Хоча завдання відносно легкі. Окрім того, перші підпункти в питаннях №5 та №6 — складні.

7.5 Порівняння моделей за AIC та BIC критеріями

Відповідність моделей оцінювалась за допомогою двох інформаційних критеріїв: AIC та BIC.

Таблиця 7.5.1

Порівняння моделей за AIC та BIC критеріями

Критерій	Модель А	Модель В	Модель С
Кількість ітерацій	18	78	1142
AIC	1441.768	1449.803	1477.354
BIC	1513.425	1558.616	1562.281

Як видно з табл. 7.5.1, кращою виявилась модель А. [13]

7.6 Аналіз МКР з ТЙ

Схожим чином до МКР з аналітичної геометрії, аналізувалась МКР з ТЙ, яка проводилась онлайн серед студентів II курсу ФІОТ КПІ ім. Ігоря Сікорського восени 2020 року. Всього контрольна складалась із десяти питань, в роботі далі досліджуються лише вісім із них. А саме — “вбудовані відповіді”. Всі інші завдання були на множинний вибір. Акцент був зроблений саме на вбудованих відповідях, оскільки це найменш досліджений тип питань. А в Україні вони взагалі мало вивчені за допомогою методів саме IRT.

Спершу було побудовано гістограму сумарних балів іспитників:

```
setwd('C:/Users/DianaUser/Downloads/Telegram Desktop')
library("readxl")
my_data <- read_excel('conf.xlsx')
my_matrix <- data.matrix(my_data, rownames.force = NA)
i <- 1
my_vector <- c()
while (i <= 10) {
  k <- 1
  while (k <= 60) {
    if (!any(my_vector == my_matrix[k,i])) {
      my_vector <- c(my_vector, my_matrix[k,i])
    }
    k <- k+1
  }
  i <- i+1
}
my_vector <- sort(my_vector)
for (i in 1:60) {
  for (j in 1:10) {
    my_matrix[i,j] <- match(my_matrix[i,j], my_vector) - 1
  }
}
suma <- c()
for (i in 1:60) {SUM <- 0
  for (j in 1:10){
    SUM <- SUM + my_matrix[i,j]
  }
  suma <- c(suma, SUM)}
hist(suma, main="Histogram of Person Parameters")
```

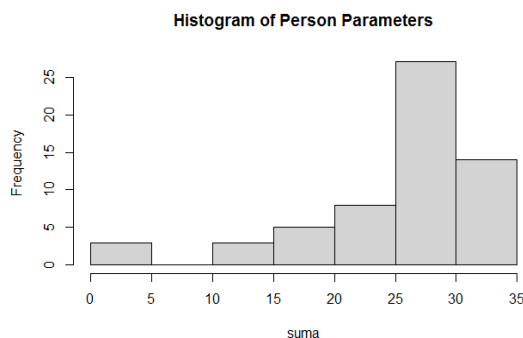


Рис. 7.6.1 — Гістограма сумарних балів іспитників

На рис. 7.6.1 гістограма зміщена вправо — це можна пояснити тим, що у іспитників даної спеціальності високий рівень знань в математиці. Зокрема, вступний бал ЗНО на цей факультет у 2019 році був 185 балів та вище.

Далі обчислювався коефіцієнт Кронбаха для дихотомічної та політомічної матриць балів:

```
library('openxlsx')
write.xlsx(my_matrix, "my_matrix.xlsx")
my_data <- read_excel("my_matrix.xlsx")
library(ltm)
descript(my_data)
```

Табл. 7.6.1

Коефіцієнти Кронбаха для різних видів матриць первинних балів

	Коефіцієнт або альфа Кронбаха
Для дихотомії	0.9267
Для політомії	0.8439

Такі високі значення коефіцієнта Кронбаха свідчать про вдало складений тест. Окрім цього можна за допомогою `descript(my_data)` знайти значення латентних параметрів для кожного із завдань та виявити завдання з високою кореляцією. Серед них виявились найбільш корельованими завдання №2 та №7, оскільки коефіцієнт кореляції вийшов в політомічній матриці рівним 0.991, що близько до одиниці.

Потім було припасовано модель Муракі до матриці первинних балів, яка була завантажена із платформи Moodle.

Третій — останній етап дослідження полягав у застосуванні 1-PL, 2-PL та 3-PL моделей до дихотомічної матриці первинних балів.

Було проведено аналіз адекватності усіх моделей даним тестування за допомогою аналога критерія χ^2 , розраховано латентні параметри моделей та побудовані ІС та ІСС криві. Тут варто пояснити, що графік інформаційності завдання показує, яку кількість інформації надає кожне питання тесту. Якщо побудувати спільний графік ІС для усіх завдань, то найкращим є те завдання, яке набуває найбільшого значення.

ІС для 1-PL моделі буде мати наступний вигляд:

```
library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
```



```
library(ltm)
fit <- rasch(my_data1)
plot(fit, type = "IIC", zrange = c(-4, 4))
```

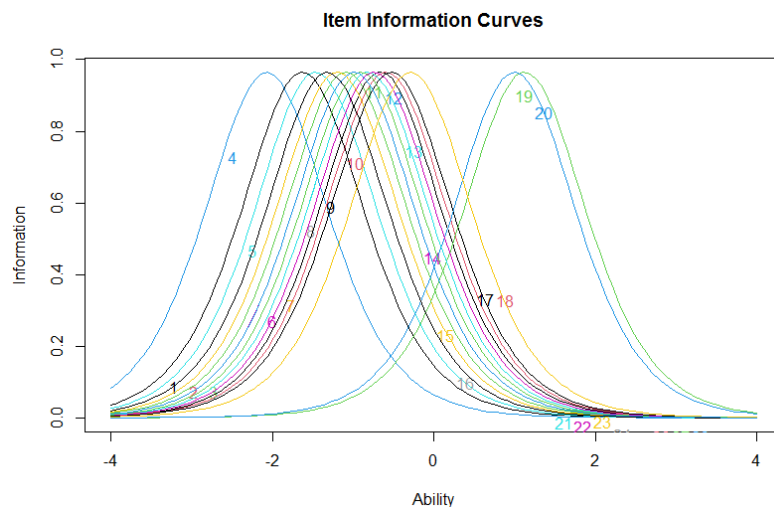


Рис. 7.6.1 — ІС для 1-PL моделі

ІС для 2-PL моделі буде мати наступний вигляд:

```
library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)
res_2pl <- ltm(my_data1 ~ z1)
plot(res_2pl, type = "IIC", zrange = c(-1, 0.5), legend = TRUE, cx = "topleft", cex=0.6)
```

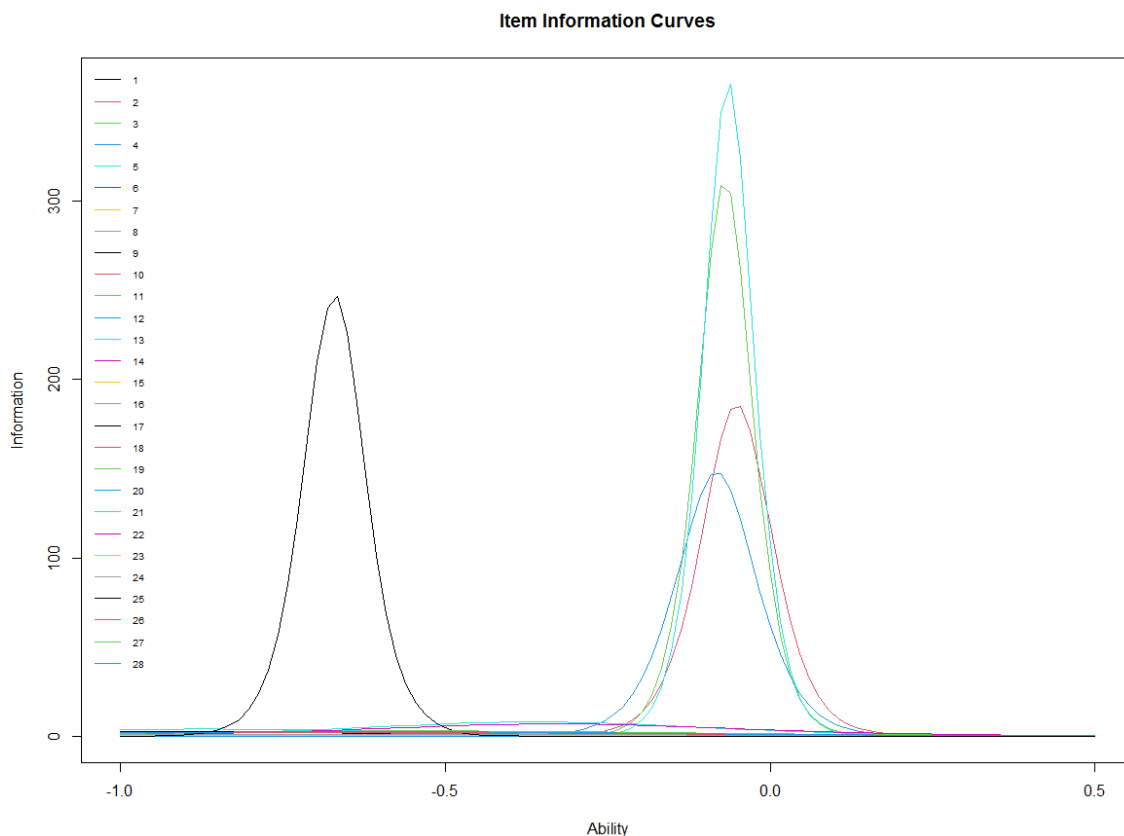


Рис. 7.6.2 — ІС для 2-PL моделі

Далі будувалась ІС для 3-PL моделі:

```
library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)
res_3pl <- tpm(my_data1 ~ z1)
plot(res_2pl, type = "IIC", zrange = c(-1.5, 0.5), legend = TRUE, cx = "topleft", cex=0.6)
```

Item Information Curves

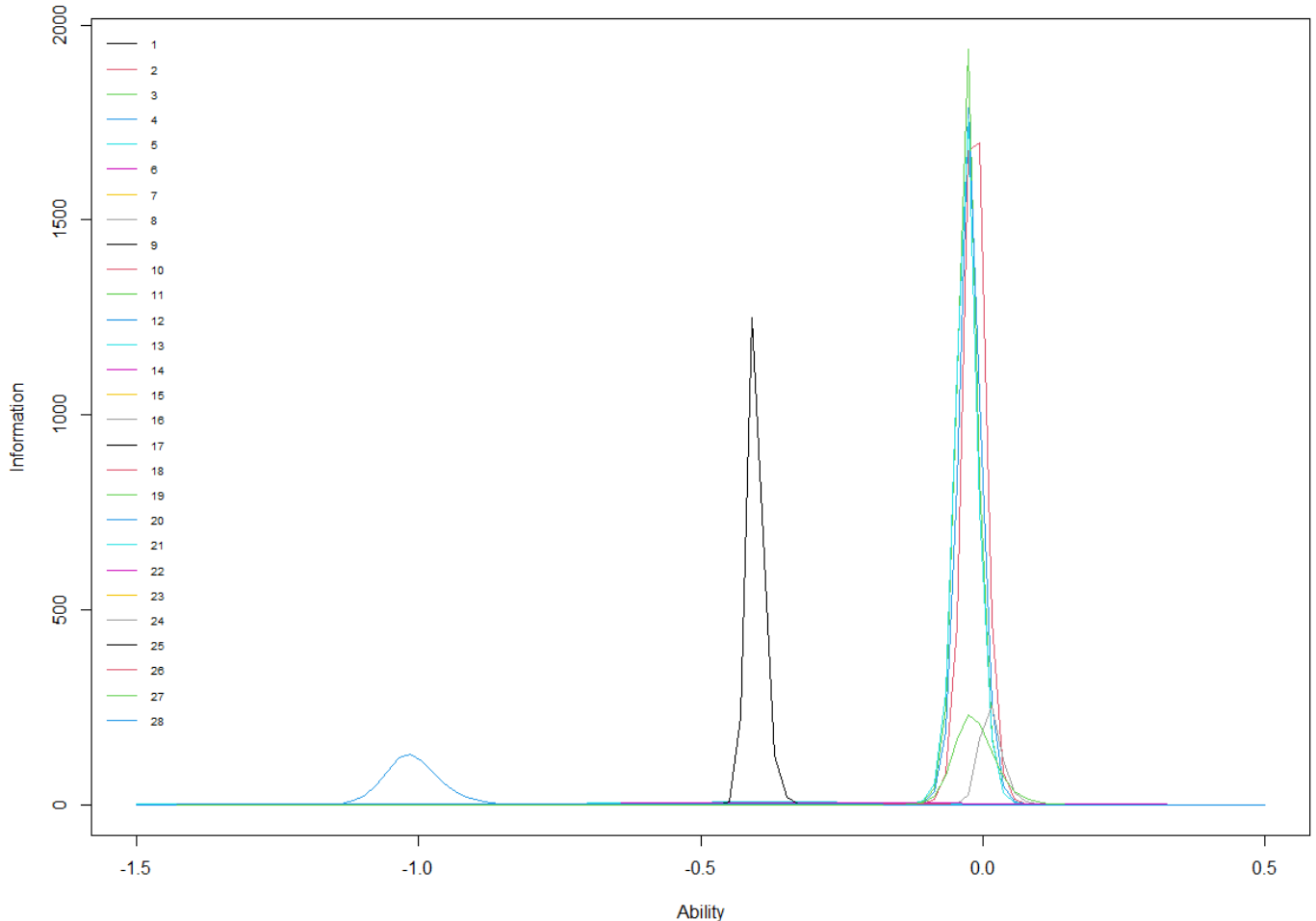


Рис. 7.6.3 — ІС для 3-PL моделі

А для безпосередньо завантаженої матриці первинних балів із платформи Moodle будувалась ІС моделі Муракі:

```
library("readxl")
my_data <- read_excel("conf.xlsx")
library(ltm)
Muraki_result <- gpcm(my_data)
plot(Muraki_result, type = "IIC", zrange = c(-4, 4), legend = TRUE, cx = "topleft", cex=1)
```

Графік міститиме всього 10 кривих, кожна з яких відповідатиме за окреме питання в тесті. Наприклад, з рис. 7.6.4 видно, що найінформативнішими будуть

питання 6 і 10.

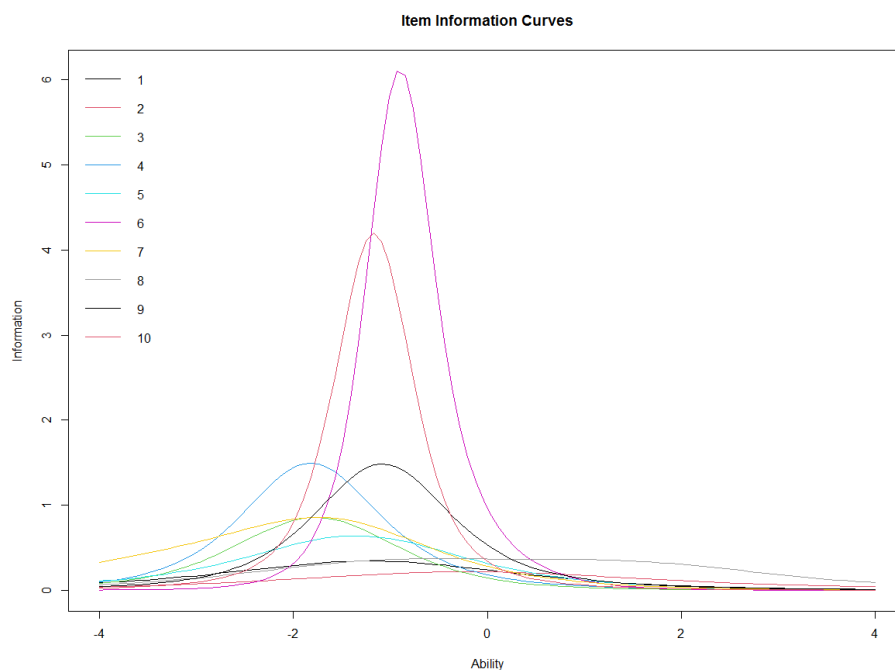


Рис. 7.6.4 — ІС для моделі Муракі

Також будувалися криві ICC або характеристичні криві рівня важкості завдань. ІС для дихотомічних завдань моделі Раша:

```
library("readxl")
my_data1<- read_excel("conf_dich.xlsx")
library(eRm)
res <- RM(my_data1)
plotjointICC(res, item.subset = c(1:28), legend = FALSE, lwd = 2, cex = 1.5)
```

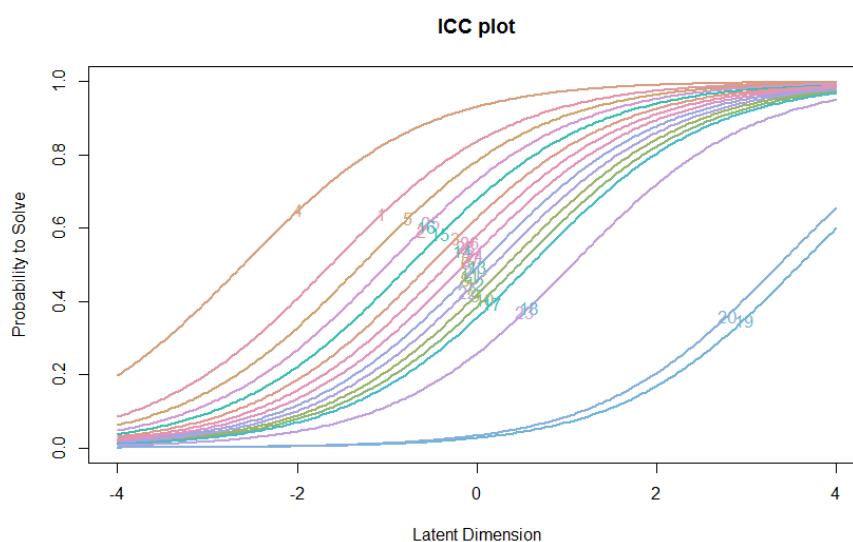


Рис. 7.6.5 — ICC для 1-PL моделі

Із рис. 7.6.5 видно, що 20 та 19 питання просідають дуже сильно. Тобто відрізняються своєю надмірною важкістю. Ці питання треба сформулювати інакше.

Далі будувалась ICC крива для 2-PL моделі:

```
library("readxl")
my_data1<- read_excel("conf_dich.xlsx")
library(ltm)
res_2pl <- ltm(my_data1 ~ z1)
plot(res_2pl)
```

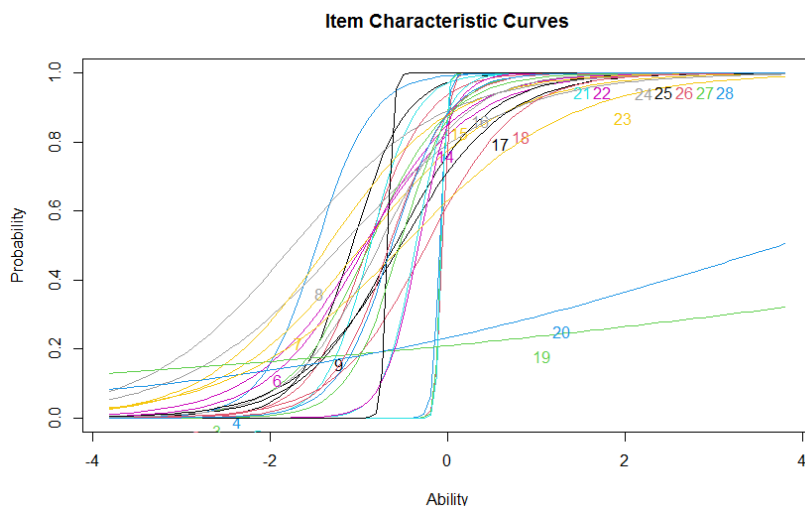


Рис. 7.6.6 — ICC для 2-PL моделі

Тут знову видно, що 19 та 20 завдання сильно просідають. Висновок той самий. Далі будуюмо ICC криву для 3-PL моделі:

```
library("readxl")
my_data1<- read_excel("conf_dich.xlsx")
library(ltm)
res_3pl <- tpm(my_data1)
plot(res_3pl)
```

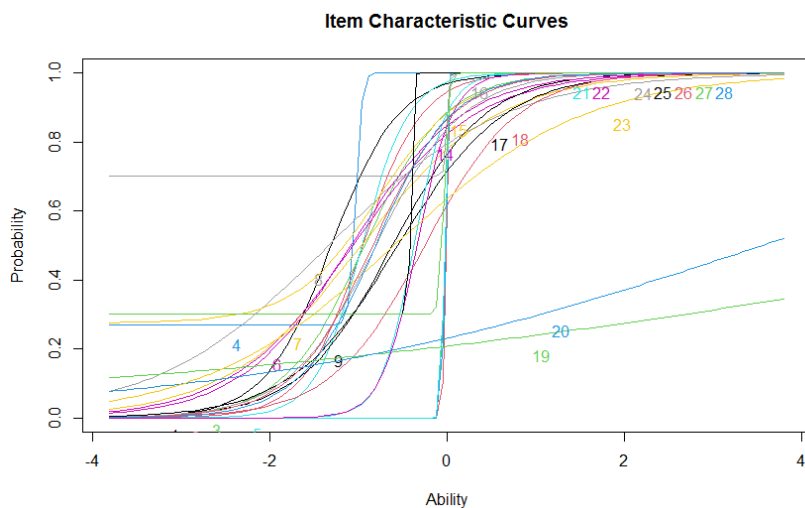


Рис. 7.6.7 — ICC для 3-PL моделі

Результати побудови ICC для політомічних завдань — моделі Муракі були зібрані для зручності в таблицю на рис. 7.6.8.

```
library('mirt')
Muraki_model <- mirt(my_data, 1, SE=TRUE)
plot(Muraki_model, type="trace")
```

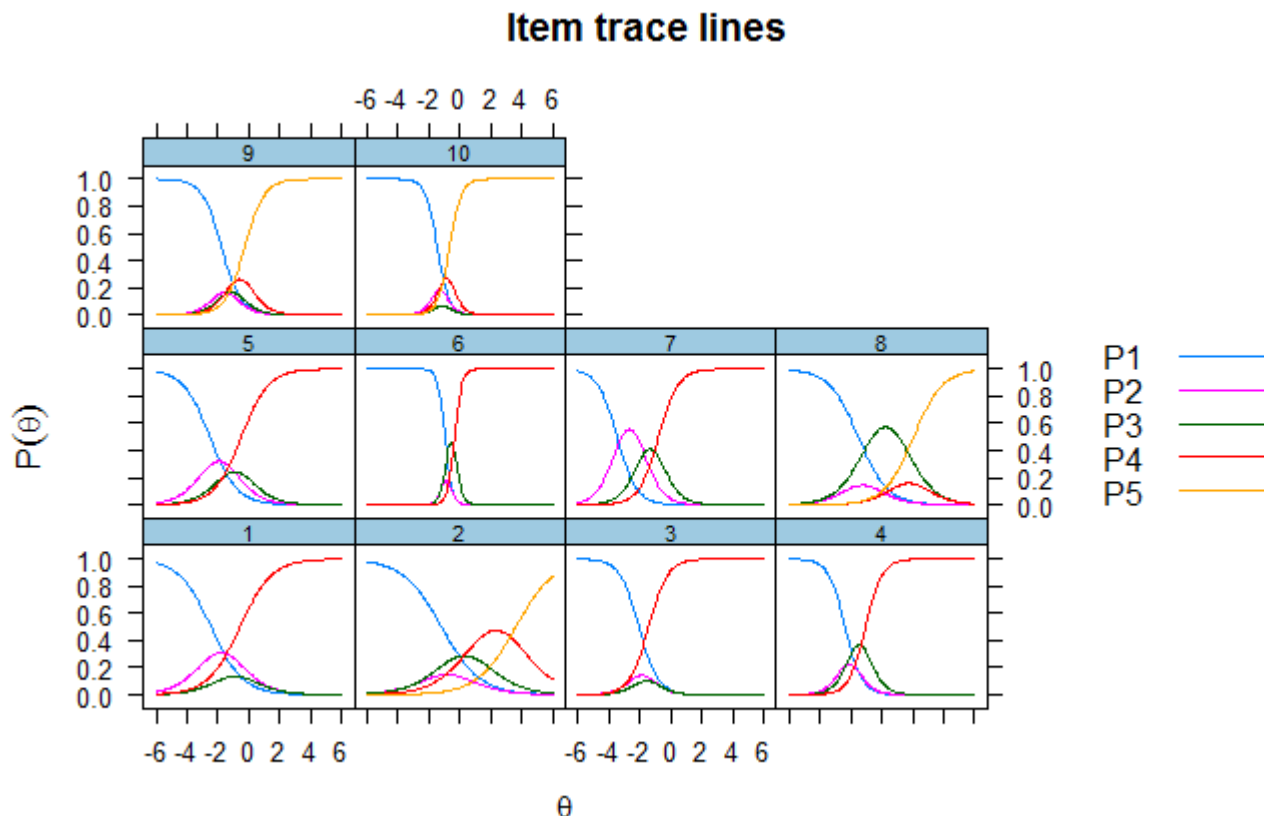


Рис. 7.6.8 — ICC політомічних завдань для моделі Муракі

Проведений IRT аналіз дав змогу зробити висновки, що найкраще описує дані тестування саме 1-PL модель. Проте в цьому методі є недолік — необхідно вручну формувати матрицю дихотомії, тобто перезаписувати значення в кожній комірці самостійно, а це значно збільшує час обробки даних тестування. Тому для зручності та економії часу найкраще застосовувати саме модель Муракі. [15]

Для цікавості можна подивитися на результати двовимірної MIRT моделі, що наведені в наступній таблиці та порівняти їх з IRT моделями:

```
two_dimensional_Muraki_model <- mirt(my_data, 2, SE=TRUE)
M2(two_dimensional_Muraki_model)
coef(two_dimensional_Muraki_model)
i <- 1
while (i <= 10) {
  print(itemplot(two_dimensional_Muraki_model, type='tracecontour', i))
  i <- i + 1
}
```

```
print(itemplot(two_dimensional_Muraki_model,type='score',i))
print(itemplot(two_dimensional_Muraki_model,i))
i <- i+1}
```

Таблиця 7.6.2

Лінії рівня, характеристичні поверхні складності питань
та поверхні очікуваних балів іспитників для двовимірної моделі

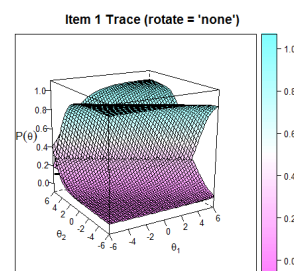
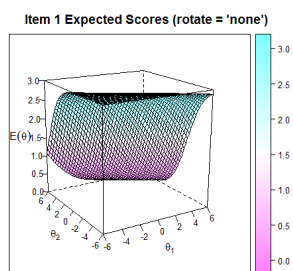
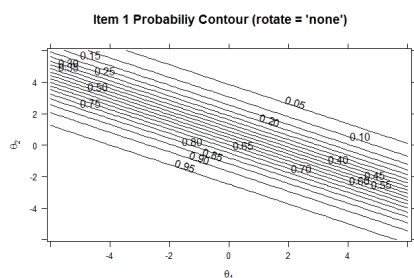
№
пит.

Контурні прямі або
лінії рівня для завдання

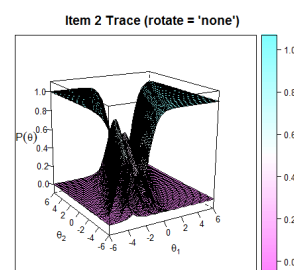
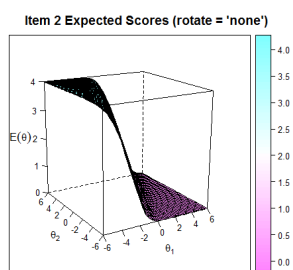
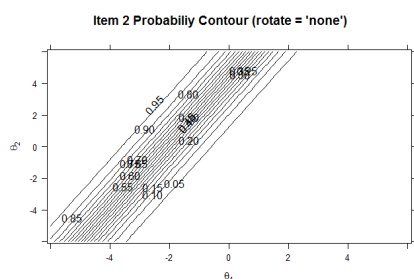
Поверхня очікуваних балів
іспитників

Характеристична поверхня
складності завдання

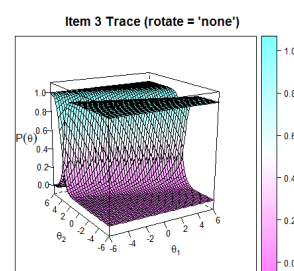
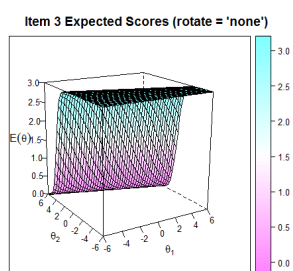
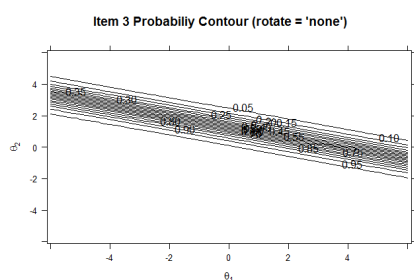
1



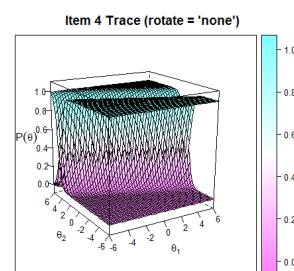
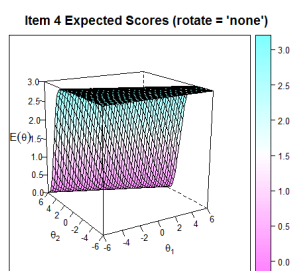
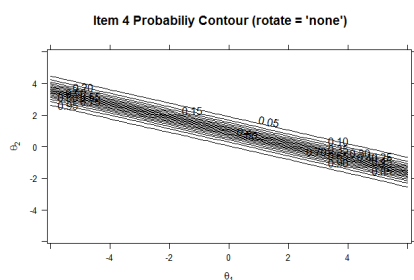
2



3

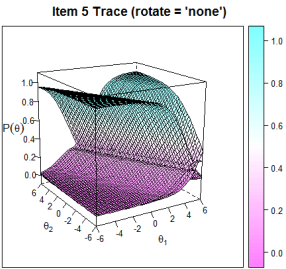
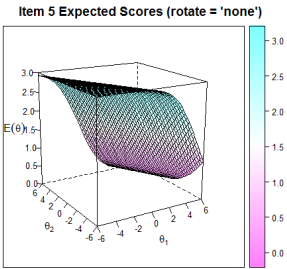
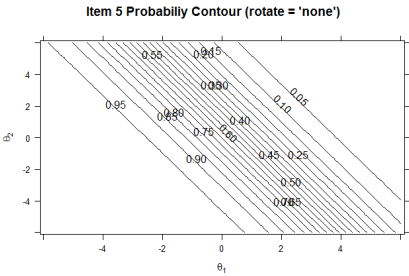


4

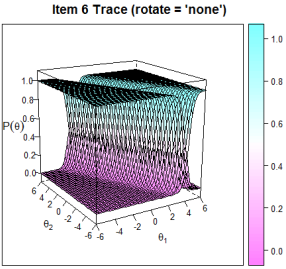
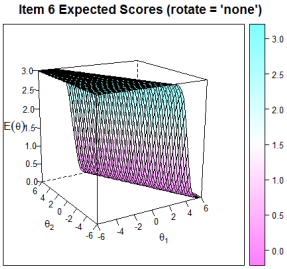
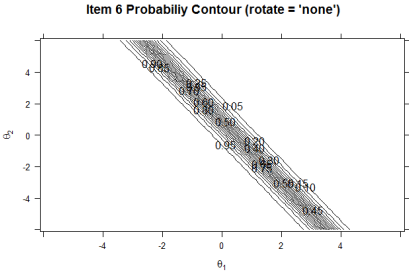


Продовження табл. 7.6.2

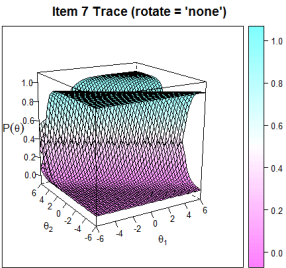
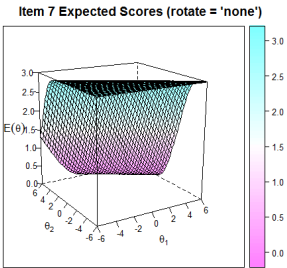
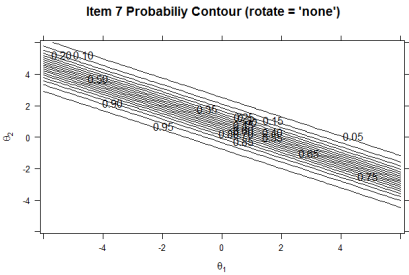
5



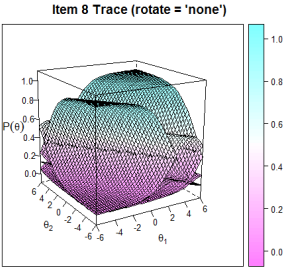
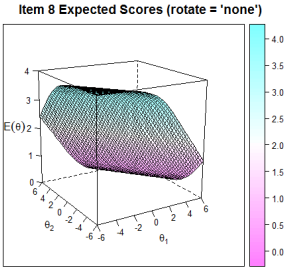
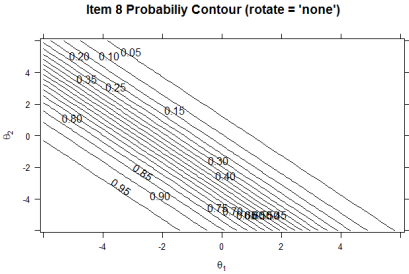
6



7

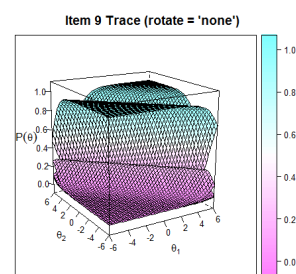
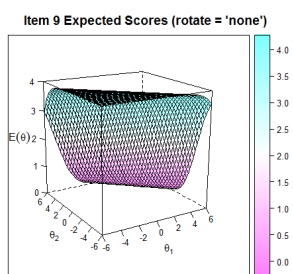
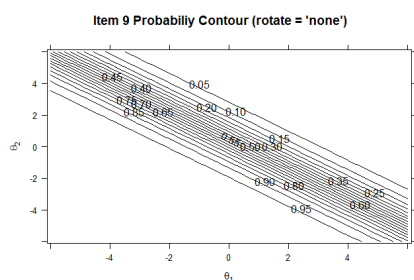


8

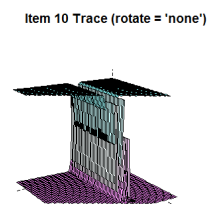
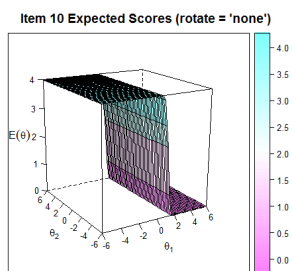
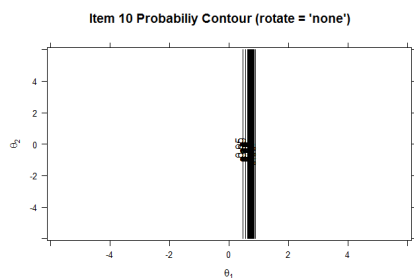


Закінчення табл. 7.6.2

9



10



У табл. 7.6.2 для другого завдання більше підійде компенсаторна модель за рахунок додатнього кутового коефіцієнту в ансамблі контурів ймовірностей, а в десятому чітко прослідковується домінація однієї компетентності, бо всі контури розташовані вертикально, окрім того, десяте завдання має високу диференціюючу здатність — це видно із щільно розташованих одне до одного контурів ймовірностей. Завдання №3, 4, 6 та 7 теж мають високу диференціюючу здатність.

ВИСНОВКИ

У даній роботі було розглянуто математичні методи аналізу тестів, що ґрунтуються на одновимірних та багатовимірних IRT моделях, та їх застосування до питань типу “вбудовані відповіді” на платформі Moodle. Такі методи надають змогу проводити більш детальний аналіз якості тесту, зокрема, визначати проблемні завдання, які необхідно доробити або сконструювати заново. Тобто знайти такі питання, що мають або занадто високу, або занадто низьку складність; завдання, що розв’язуються за допомогою спеціального програмного забезпечення або спеціальних засобів та питання до яких відповідь можна легко вгадати.

Моделі MIRT є узагальненням одномірних моделей. Для них вводяться додаткові латентні параметри, що відповідають за загальну обізнаність та певні компетентності іспитника. Тобто перевіряються не лише знання з якогось предмету, який викладається, але й уміння, які були набуті раніше, наприклад, під час навчання в університеті.

У МКР з аналітичної геометрії до дихотомічних даних було застосовано 2-PL моделі Бірнбаума, а до політомічних — модель Муракі. Перевірка адекватності моделей здійснювалась за допомогою аналогів критерія χ^2 . Було зроблено висновок, що кожна з моделей адекватно описує дані тестування.

У МКР з ТЙ до дихотомічних даних було застосовано моделі Раша, дво- та три-параметричні моделі Бірнбаума, а до політомічних — модель Муракі. Зроблено висновок, що модель Раша описує дані краще, ніж інші, але для цього необхідно створювати дихотомічну матрицю, що затратно по часу й зусиллям. Тому цілком прийнятним є застосування моделі Муракі. Потім отримані результати за допомогою IRT методів, порівнювались із результатами отриманими MIRT методами. Завдяки MIRT методам можна більш детально розглянути і дослідити конструкцію кожного питання та завдання — це надзвичайно допомагає у формуванні якісного банку питань для наступних тестувань. Проте існує ще досить

велика недосліджена область — це моделювання та підбір некомпенсаторних моделей. Тому питання застосування та аналіз питань за допомогою некомпенсаторних моделей досі вважається відкритим.

ПЕРЕЛІК ПОСИЛАНЬ

1. Круглова Н.В. Особенности застосування математичних моделей тестів в умовах дистанційного контролю / Н.В. Круглова, О.О. Диховичний, І.В. Алексеева // Математические машины и системы. — 2020. — №2. — С.105-116.
2. Карданова Е.Ю. О применимости политомической модели Г. Раша к тестовым заданиям различных форм, оцениваемым политомически // Вопросы тестирования в образовании. — 2006. — № 4. — С. 44-56.
3. Крокер Л. Введение в классическую и современную теорию тестов / Л. Крокер, Дж. Алгина — Москва: Логос, 2010. — 668 с.
4. Челышкова М.Б. Теория и практика конструирования педагогических тестов / М.Б. Челышкова — Москва: Логос, 2002. — 432 с.
5. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов // Нейман Ю.М., Хлебников В.А. — Москва: Центр тестирования МО РФ, 2000 — 168 с.
6. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
7. Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
8. Yao, L., & Schwarz R. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed format tests. *Applied Psychological Measurement*. 30, 469—492.
9. Akaike, Hirotugu. A new look at the statistical model identification (неопр.) // *IEEE Transactions on Automatic Control*. — 1974. — Т. 19, № 6. — С. 716—723.
10. Grasa, Antonio Aznar (1989) *Econometric Model Selection: A New Approach*,

Kluwer.

11. Greene, William H. (1997) *Econometric Analysis*, 3rd edition, Prentice-Hall.
12. Sakamoto, Y., Ishiguro, M., and Kitagawa G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.
13. N. Kruglova, O. Dykhovychnyi, D. Lysenko. Application of IRT and MIRT models to analysis of analytical geometry tests// *Adaptive Systems of Automatic Control*. — 2021. — Vol. 1. (пройшла рецензування, готується до друку)
14. Michael J. Crawley (2007). *The R Book*. John Wiley & Sons. ISBN 978-0-470-51024-7.
15. Д.Р. Лисенко, Н.В. Круглова, О.О. Диховичний. Аналіз тестів типу “вбудовані відповіді” за допомогою моделей IRT // *Тези доповідей X всеукраїнської наукової конференції молодих математиків*. – 2021 р. – С. 106–107.

Додаток А. Лістинг програми

```

setwd('C:/Users/DianaUser/Downloads/Telegram Desktop')
# To know your current working directory, type the function getwd() in R console
getwd()

# install.packages("readxl")
# Loading
library("readxl")

#my_data <- read_excel(file.choose())
# If you use the R code above in RStudio, you will be asked to choose a file.

# xlsx files
my_data <- read_excel("4_diploma.xlsx") #conf.xlsx
my_matrix <- data.matrix(my_data, rownames.force = NA)
# first column
#my_matrix[,1]

i <- 1
my_vector <- c()
while (i<=10) {

  k <- 1
  while (k<=105) {
    if (!any(my_vector==my_matrix[k,i])) {
      my_vector <- c(my_vector, my_matrix[k,i])
      k <- k+1
    }
    i <- i+1
  }

my_vector <- sort(my_vector)
print(my_vector)

for (i in 1:105) {
  for (j in 1:10) {
    my_matrix[i,j] <- match(my_matrix[i,j], my_vector) - 1}}

# вивести 1-ий рядок:
# my_matrix[1,]

suma <- c()
for (i in 1:105) {
  SUM <- 0
  for (j in 1:10){
    SUM <- SUM + my_matrix[i,j]
  }
  suma <- c(suma,SUM)
}
print(suma)
hist(suma,col=10,main="Histogram of Person Parameters")

#install.packages('openxlsx')
library('openxlsx')
write.xlsx(my_matrix, "my_matrix.xlsx")

```

```

my_data <- read_excel("my_matrix.xlsx")
print(my_matrix)

#модель A
my_data1 <- read_excel('my_matrix_dich.xlsx')
# install.packages('mirt')
#install.packages('latticeExtra')
library(mirt)
library(latticeExtra)
#polytomous items
A_model <- mirt(my_data1, 1, SE=TRUE)
itemplot(A_model, 3, type = 'infotrace')
itemplot(A_model, 3, type = 'score')

i <- 1
while (i<=10) {
  print(itemplot(A_model, i))
  print(itemplot(A_model, i, type = 'infotrace'))
  i <- i+1}

plot(A_model,type="trace")
coef(A_model)

extract.mirt(A_model, 'G2')
extract.mirt(A_model, 'p')
plot(A_model, type = 'info')

# модель B
B_model <- mirt(my_data1, 2, SE=TRUE)
vector_of_types <- c("gpcm",rep("2PL",3),rep("gpcm",2),'2PL', rep("gpcm",2),"2PL")
bmod <- mirt(my_data1, B_model, itemtype = vector_of_types, method = 'MHRM', verbose=FALSE)
i <- 1
while (i<=10) {
  print(itemplot(bmod,type='tracecontour',i))
  print(itemplot(bmod,type='score',i))
  print(itemplot(bmod,i))
  i <- i+1}

C_model <- mirt.model('F1 = 1-3,7
                     F2=4-6,8-10')
vector_of_types <- c(rep("gpcm",10))
cmmod <- mirt(my_data, C_model, itemtype = vector_of_types, method = 'MHRM', verbose=FALSE)
i <- 1
while (i<=10) {
  print(itemplot(cmmod,type='tracecontour',i))
  print(itemplot(cmmod,type='score',i))
  print(itemplot(cmmod,i))
  i <- i+1
}

AIC(A_model, B_model, C_model)
BIC(A_model, B_model, C_model)

```

```

# МКР по ТЙ!

setwd('C:/Users/DianaUser/Downloads/Telegram Desktop')
# To know your current working directory, type the function getwd() in R console
getwd()

# install.packages("readxl")
# Loading
library("readxl")

my_data <- read_excel('conf.xlsx')
my_matrix <- data.matrix(my_data, rownames.force = NA)

i <- 1
my_vector <- c()
while (i<=10) {
  k <- 1
  while (k<=60) {
    if (!any(my_vector==my_matrix[k,i])) {
      my_vector <- c(my_vector, my_matrix[k,i])
      k <- k+1}
    i <- i+1}

my_vector <- sort(my_vector)
print(my_vector)

for (i in 1:60) {
  for (j in 1:10) {
    my_matrix[i,j] <- match(my_matrix[i,j], my_vector) - 1}}

print(my_matrix)

suma <- c()
for (i in 1:60) {
  SUM <- 0
  for (j in 1:10){
    SUM <- SUM + my_matrix[i,j] }
  suma <- c(suma,SUM)}
print(suma)
hist(suma, main="Histogram of Person Parameters")
library('openxlsx')
write.xlsx(my_matrix, "my_matrix.xlsx")
my_data <- read_excel("my_matrix.xlsx")
library(ltm)
descript(my_data)
library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)
fit <- rasch(my_data1)
plot(fit, type = "IIC", zrange = c(-4, 4))
library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)
res_2pl <- ltm(my_data1 ~ z1)
plot(res_2pl, type = "IIC", zrange = c(-1, 0.5), legend = TRUE, cx = "topleft", cex=0.6)

library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)

```

```

res_3pl <- tpm(my_data1 ~ z1)
plot(res_2pl, type = "IIC", zrange = c(-1.5, 0.5), legend = TRUE, cx = "topleft", cex=0.6)
library("readxl")
my_data <- read_excel("conf.xlsx")
library(ltm)
Muraki_result <- gpcm(my_data)
plot(Muraki_result, type = "IIC", zrange = c(-4, 4), legend = TRUE, cx = "topleft", cex=1)
library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(eRm)
res <- RM(my_data1)
plotjointICC(res, item.subset = c(1:28), legend = FALSE, lwd = 2, cex = 1.5)

library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)
res_2pl <- ltm(my_data1 ~ z1)
plot(res_2pl)

library("readxl")
my_data1 <- read_excel("conf_dich.xlsx")
library(ltm)
res_3pl <- tpm(my_data1)
plot(res_3pl)

library('mirt')
Muraki_model <- mirt(my_data, 1, SE=TRUE)
plot(Muraki_model,type="trace")
two_dimensional_Muraki_model <- mirt(my_data, 2, SE=TRUE)
M2(two_dimensional_Muraki_model)
coef(two_dimensional_Muraki_model)
i <- 1
while (i<=10) {
  print(itemplot(two_dimensional_Muraki_model,type='tracecontour',i))
  print(itemplot(two_dimensional_Muraki_model,type='score',i))
  print(itemplot(two_dimensional_Muraki_model,i))
  i <- i+1}

```