

УДК 371.26(035)

ПРОГРАМНІ ЗАСОБИ ДЛЯ ОБҐРУНТУВАННЯ ЯКОСТІ ТЕСТОВИХ ЗАВДАНЬ

Фетісов В.С., Чернишова Е.О.

У статті розглядаються основи роботи із представником специфічного класу програмного забезпечення, що здійснює математико-статистичний аналіз завдань тесту на відповідність вимогам моделі Г.Раша – програмним продуктом WINSTEPS.

Ключові слова: тест, програмне забезпечення для обґрунтування якості тестових завдань, Г.Раш, WINSTEPS.

В статье рассматриваются основы работы с представителем специфического класса программного обеспечения для проведения математико-статистического анализа теста на соответствие требованиям модели Г.Раша – программным продуктом WINSTEPS.

Ключевые слова: тест, программное обеспечение для обоснования качества тестовых заданий, Г.Раш, WINSTEPS.

The article elucidates the foundations of work with an exemplar of a specific type of software for mathematical-statistical analysis of test items for their conformity to the requirements of the Rasch model – WINSTEPS software.

Key words: test, software for analysis of the quality of test items, G.Rasch, WINSTEPS.

Найважливіше значення для тестових завдань є обґрунтування їх якості. На Заході існує система знань і технологій, що набула назву *Rasch Measurement* (RM). Це можна перевести як “система об’єктивованих педагогічних вимірів, що побудована за теорією Г.Раша”. Вона будується на трьох припущеннях:

1. Рівень складності завдань і рівень підготовленості осіб, які тестуються, можна виміряти в одній шкалі, із загальною стандартною одиницею виміру.

2. За наявності такої шкали вірогідність правильної відповіді випробовуваного може стати залежною від різниці між рівнем його підготовленості і рівнем складності завдання тесту.

3. Результат протиставлення випробовуваного із завданнями тесту можна прогнозувати. Чим вище рівень підготовленості випробовуваного, тим вище має бути вірогідність його правильної відповіді на завдання фіксованого рівня складності.

Теорія Г.Раша дозволяє встановити зв’язок між рівнем знань осіб, які проходять тестування, і результатами виконання тестів, що дозволяє у свою чергу визначити рівень знань незалежно від складності тестових завдань. Для вимірювання складності завдань і рівня знань застосовується одиниця виміру, що набула назву *логіт*. Г.Раш увів дві міри: “логіт рівня знань” і “логіт рівня складності завдання”. Першу він визначив як натуральний логарифм відношення частки правильних відповідей особи, яка проходить тестування для усіх

завдань тесту, до частки неправильних відповідей, а другу – як натуральний логарифм іншого відношення – частки неправильних відповідей на завдання тесту до частки правильних відповідей на те саме завдання для всіх осіб, які проходять тестування. Єдина логарифмічна шкала дозволяє встановити необхідну відповідність між рівнем навчання і складністю завдання і, більше того, провести корекцію результатів тестування для тестів різної складності.

На моделі Г.Раша базується специфічний клас програмного забезпечення, що здійснює математико-статистичний аналіз завдань тесту на відповідність вимогам цієї моделі. Цей клас налічує велику кількість комп’ютерних програм, що розрізняються за різними критеріями: обсягами вхідної інформації, кількістю математико-статистичних методів, обсягами (кількістю) результатних таблиць і т.ін. Існують як безплатні (freeware), так і пропріетарні програми цього класу. Вони дозволяють швидко опрацювати великі обсяги даних, розрахувати потрібні статистики, провести апроксимацію фактичних даних за нормальним розподілом. У результаті користувач одержує інформацію про формальні властивості кожного тестового завдання і тесту в цілому, про рівень і структуру підготовленості осіб, які проходять тестування, про надійність і валідність результатів тестування, про міру відповідності результатів випробовуваних рівню складності завдань і т.ін.

Програми цього класу здійснюють також шкалювання рівнів підготовки осіб, які проходять тестування, і рівнів складності тестових завдань ("test calibration").

Наприклад, статистичний пакет *QUEST* призначено для аналізу як тестових завдань, так і питань соціологічних анкет. Обробка даних у ньому може проводитися як відповідно до класичної (статистичної) теорії педагогічних вимірів, так і за теорією RM. Програма *WINMIRA* має широкий набір методів обробки даних на основі Item Response Theory. Особливо слід відмітити повну інтеграцію цього засобу із статистичним пакетом SPSS. Програмний засіб *CONQUEST* надає змогу проведення не лише одновимірних, але і багатовимірних вимірів. Пакет *RASCAL* дозволяє шкалювати завдання і випробовуваних на основі одновимірної моделі Раша, якщо дані подані в дихотомічній шкалі ("0" або "1"). Але найбільш популярними і розповсюдженими програмними продуктами вважаються австралійська розробка *RUMM 2020* та американський програмний засіб *WINSTEPS*.

WINSTEPS

Програма *WINSTEPS* (США) є однією з сімейства програм, призначених для аналізу результатів тестування, побудовану на технології за теорією Г.Раша, що були розроблені під керівництвом J.M.Linacre. Програма є комерційною, але є її безкоштовна версія, що називається *MINISTEP*. Вона дозволяє використовувати усі можливості *WINSTEPS*, але має обмеження на кількість питань у тесті (25) і кількість осіб (75). Скопіювати *MINISTEP* можна із сайту розробника за адресою <http://www.winsteps.com/index>.

Програма має англійський інтерфейс. Її опануванню сприятиме наявність потужної і докладної довідки, яку також можна скопіювати із сайту розробника, але вона, звичайно, також подана англійською мовою.

Робота з даними

Основний модуль має текстовий інтерфейс, дещо незвичайний для тих, хто не мав із ним справу. Але для роботи з даними є спеціальний модуль *Windata*, який має саме графічний інтерфейс. Звернутися до нього можна двома шляхами. По-перше, це можна зробити з меню кнопки <Пуск>, шляхом вибору у групі "Winsteps" пункту "Windata". За другим варіантом це можна зробити з основного модуля. Для цього після його завантаження у головному меню достатньо вибрати пункт **Data Setup**.

За будь-яким варіантом з'явиться вікно "*Ministep Control File Set-Up*" ("Установки файлу керування даними *Ministep*"), з якого можна здійснювати операції з файлами первинних даних у звичайному графічному вигляді. Крім того, і відображення первинних даних подається у комфортному вигляді.

Для завантаження існуючого файлу з даними слід виконати команду **File ► Read Control (and data file)** (зчитати файл керування і файл з даними), після чого з'явиться стандартне вікно для завантаження даних. Зазначимо, що розробник передбачив кілька варіантів імпорту даних, починаючи з текстового файлу і закінчуючи імпортом із відомих статистичних пакетів (SAS, SPSS тощо).

Вхідні дані

Починаючи працювати з програмою, користувачам доведеться розібратися з концепцією формування вхідних даних, за якою вони поділяються на

Other specifications in control file:

```

; This is file "example0.txt" - ";" starts a comment
@GENDER = $S1W1 ; KID gender in column 1 of person label
PSUBTOT = @GENDER ; subtotal by KIDs genders

```

Рис 1. Вікно *MINISTEP* з початковими даними

дві частини: безпосередньо дані і правила (інструкції) керування даними (Control File). Вони можуть бути об'єднані в одному файлі (у цьому разі первинні дані розташовуються після інструкцій керування, тобто наприкінці файлу) або розташовані у двох окремих файлах. Розробник рекомендує об'єднувати їх в одному файлі, якщо обсяг первинних даних є невеликим. Тому розробник включив до інсталяції системи достатньо велику кількість різноманітних прикладів. Він рекомендує перед створенням файлу (матриці) з початковими даними знайти схожий приклад і скористатися ним.

Нижче наведений приклад первинних даних, розташованих в одному (текстовому) файлі:

```
...
TITLE = 'LIKING FOR SCIENCE (Wright & Masters p.18)'
NI = 25 ; 25 items
ITEM1 = 1 ; responses start in column 1 of the data
NAME1 = 28 ; person-label starts in column 28 of the data
ITEM = ACT ; items are called "activities"
PERSON= KID ; persons are called "kids"
...
END NAMES; this follows the item names: - the data follow:
1211102012222021122021020 M Rossner, Marc Daniel
22222222222222222222222222222222 M Rossner, Lawrence F.
2211011012222122122121111 M Rossner, Toby G.
1010010122122111122021111 M Rossner, Michael T.
```

Вікно "Ministep Control File Set-Up"

Матриця первинних даних

Первинні дані подаються у вигляді рядків з даними. Кожний рядок містить інформацію для однієї людини: її **ідентифікатор**, наприклад, прізвище і послідовність **відповідей** на питання тесту, анкети тощо (надалі називатиме їх для кращого розуміння також **питання**). Якщо відповідь відсутня, то на цьому місці ставиться код, що застосовується для помилкових даних. Ідентифікатором не обов'язково має бути тільки один якийсь реквізит: їх може бути і кілька, наприклад, стать, прізвище, ім'я. Під "відповіддю" розуміють значення даних, що може числовим значенням або назвою категорії, значенням підрахунку або порядковим номером з множинного вибору.

Початок ідентифікатора і кожної відповіді в окремому рядку розташовані у чітко визначених позиціях. Таке подання даних нагадує матрицю, де кожний рядок містить відомості для однієї людини, через що до первинних даних застосовують також назву "матриця первинних даних".

Редагувати відповіді дуже просто: слід встановити курсор у клітинки з потрібною відповіддю і ввести нове значення.

Правила (специфікації) керування (первинними даними)

Правила визначають різноманітні установки для первинних даних, розрахунків, аналізів і т.ін. Вони кодуються за допомогою певних команд, які, у свою чергу, мають атрибути. Наприклад, команди, що відображаються на наведеному вище рисунку, призначені для керування саме первинними даними.

- TITLE. Назва тесту (анкети).

- PERSON = A data row is a. Це довідкова інформація про те, дані для яких категорій осіб містяться у рядку. Наприклад, це можуть бути діти (KID), жінки (FEMALE) і т.ін.

- NAME1 = First personal label column. Це числове значення, що показує, з якої позиції починаються ідентифікаційні дані.

- NAMELEN = Personal label length. Числове значення, що визначає довжину (загальну кількість позицій) ідентифікаційних даних.

- Number of data rows. Кількість рядків (осіб).

- Number of data columns. Загальна кількість позицій у рядку.

- ITEM1 = First item column. Числове значення, що вказує, з якої позиції починаються відповіді.

- NI = Number of items. Загальна кількість питань (відповідей).

- XWIDE = columns per response. Кількість позицій, що відводиться на одну відповідь (1 або 2 позиції).

- CODES = Valid codes. Значення містить множину допустимих кодів для кодування відповідей, які складають суцільний рядок без проміжків.

Разом із тим система надає можливість користувачу застосовувати велику кількість команд, що мають найрізноманітніше призначення. Наприклад, перемикач UIMEAN (UPMEAN), що розташований у правій частині вікна "Ministep Control File Set-Up", визначає для яких об'єктів будуть розраховані середні значення: відповідно для відповідей або осіб. USCALE визначає кількість одиниць для одного логіту, а UDECIM – кількість значень дробової частини для показників у звітах.

Інструменти

Крім безпосередньо даних і правил специфікації керування ними, вікно містить інструменти для роботи з даними.

Інструмент «**Item Labels**» ("Назви відповідей"). Натиснувши цю кнопку, користувач одержує можливість надати назву кожній відповіді або змінити її.

Інструмент «**Remove excess rows and columns**» ("Вилучити рядки і стовпчики") дозволяє вилучати інформацію з матриці, після чого слід застосувати інструмент «**Refresh Data Display**» ("Оновити дані") для оновлення матриці даних.

Перевірка первинних даних на відповідність їх можнині кодів допустимих кодів здійснюється за допомогою інструмента «**Scan data for codes**». ("Перевірити дані на (допустимі) коди"). Після перевірки знайдені неправильні значення кодів у матриці будуть виділені іншим кольором.

Інструмент «**Category labels CLFILE**» ("Назви текстових меток") призначено для введення, зміни і відображення текстових назв, що відповідають числовим значенням відповідей.

Аналіз даних

Після введення даних і контролю їх вірогідності переходять до їх аналізу, який здійснюється у головному модулі.

Після його завантаження програма запросить завантажити файл керування даними. Це можна зробити шляхом введення його повного імені з клавіатури, а можна натиснути <Enter> для виклику діалогового вікна вибору файлів.

Control file name? (e.g., exam1.txt). Press Enter for Dialog Box:

Після завантаження файлу керування даними, система запропонує ввести ім'я для вихідного

файла з результатами аналізу. Але він не є обов'язковим, тому можна створити тимчасовий файл, натиснувши для цього <Enter>.

Report output file name (or press Enter for temporary file, Ctrl+O for Dialog Box):

Якщо первинні дані розміщені у двох файлах, то слід ввести ще додатково один файл після запрошення:

Extra specifications (if any). Press Enter to analyze: Нарешті натискаємо <Enter> для початку аналізу. Після цього система виводить відомості про вхідні дані і таблицю збіжності (Convergence Table).



Початкове, грубе оцінювання збіжності дають показники у рядку "Calculating Fit Statistics". Стандартизовані залишки (Standardized Residuals)

CONVERGENCE TABLE

PROX ITERATION	ACTIVE COUNT			EXTREME 5 RANGE		MAX LOGIT CHANGE	
	KID	ACT	CATS	KID	ACT	MEASURES	STRUCTURE
1	75	25	3	3.78	3.20	3.8918	.0740
2	74	25	3	4.59	3.71	.8258	-.6158
3	74	25	3	4.83	3.92	.2511	-.1074

JMLE ITERATION	MAX SCORE RESIDUAL*	MAX LOGIT CHANGE	LEAST CONVERGED			CATEGORY STRUCTURE	
			KID	ACT	CAT	RESIDUAL	CHANGE
1	2.84	-.1955	60	22*	2	21.44	.0076
2	.71	-.0335	53	15*	0	-5.89	.0144
3	-.43	.0297	53	5*	1	3.47	.0101
4	.32	.0237	18	11*	1	2.71	.0079
5	.24	.0184	18	11*	0	-2.09	.0060
6	.19	.0141	18	11*	0	-1.63	.0045
7	.14	.0108	18	11*	0	-1.25	.0035
8	.11	.0082	18	11*	0	-.96	.0026
9	.08	.0062	18	11*	0	-.73	.0020
10	.06	.0048	18	11*	0	-.56	.0015

Standardized Residuals N(0,1) Mean: .06 S.D.: 1.04

в моделі Раша моделюються за нормальним розподілом. Тому суттєві відхилення від значення "0" для середньої величини (Mean) і значення "1" для стандартного відхилення (S.D., Standard Deviation) сигналізують про те, що первинні дані не відповідають моделі Раша, за якою вони мають відповідати саме нормальному розподілу.

У стовпчиках першої таблиці розташовані значення, що призначені для початкового оцінювання даних, розрахованих за нормальним розподілом:

1. PROX ITERATION. Порядковий номер ітерації.
2. ACTIVE COUNT. Інформація про загальну кількість спостережень (осіб), кількість питань (відповідей), рейтингову шкалу категорій.
3. EXTREME 5 RANGE PERSONS. Оцінювання поточної ітерації розмаху між середнім значенням для п'яти осіб, які мають максимальні результати, і середнім значенням для п'яти осіб, які мають мінімальні результати.
4. MAX LOGIT CHANGE. Максимальна зміна логіту для кожної особи або питання. Очікується, що з кожною ітерацією це значення буде зменшуватися до максимально припустимого значення збіжності (це значення є аргументом команди LCONV). Саме на це значення "орієнтується" система під час ітераційного процесу: він продовжується доки логіт не буде мати значення менше за максимально припустиме.

Друга таблиця містить максимальні оцінки ймовірності призначені для точного оцінювання даних:

1. JMLE ITERATION. Порядковий номер ітерації.
2. MAX SCORE RESIDUAL. Максимальна різниця між фактичними і теоретичними значеннями для кожної особи або питання. Очікується, що з кожною ітерацією це значення буде зменшуватися до максимально припустимого значення збіжності.
3. MAX LOGIT CHANGE. Максимальна зміна логіту для кожної особи або питання. Очікується, що з кожною ітерацією це значення буде зменшуватися до максимально припустимого значення збіжності (це значення є аргументом команди LCONV).
4. LEAST CONVERGED. Значення найменшої збіжності. Це порядковий номер особи, питання та категорії, для якого спостерігається мінімальне відхилення від збіжності.
5. CATEGORY RESIDUAL. Залишок для категорії. Максимальна різниця між фактичними і теоретичними значеннями для будь-якої категорії. Очікується, що з кожною ітерацією це значення буде зменшуватися.
6. STRUCTURE CHANGE. Зміни структури. Максимальна зміна логіту для будь-якого шкалювання структури. Це значення має довідковий характер і не використовується для оцінювання збіжності. Очікується, що з кожною ітерацією це значення буде зменшуватися.

Дослідження вимірювання і контроль якості тестових завдань

Дослідження здійснюється за допомогою великої кількості результатних таблиць, доступ до яких надає команда головного меню **Output Tables**. Результатні таблиці можна умовно поділити на кілька категорій: перші є загальними, другі здійснюють аналіз відносно відповідей (АСТ) і треті – відносно категорій осіб (у нашому прикладі – KID, тобто дітей).

Дослідження доцільно починати з таблиці, яка містить низку підсумкових показників. Для цього

слід виконати команду **Output TablesSummary statistics**.

Таблиця вимірювань відповідей рангує відповіді, дозволяючи швидко і просто визначити найбільш популярні і непопулярні відповіді. Наприклад, з наведеного нижче фрагменту таблиці, бачимо, що найулюбленішими заняттями серед дітей, за даними тестування, є виїзд на пікнік ("Go on picnic"), походи до зоопарку та музею, а найменш улюбленим – збирання пляшок та банок ("Find bottles and cans"). Під ранговою таблицею знаходиться таблиця з деталізацією вимірювань кожної відповіді.

Література

1. Офіційний сайт RUMM Laboratory [Електронний ресурс]. – Режим доступу: <http://www.rummlab.com/>. – Назва з екрану.
2. Офіційний сайт WINSTEPS [Електронний ресурс]. – Режим доступу: <http://www.winsteps.com/index>. – Назва з екрану.