

ОЦІНКА НАДІЙНОСТІ КАТЕГОРІАЛЬНИХ ШКАЛ ПРИ АНАЛІЗІ ДАНИХ СОЦІОЛОГІЧНИХ ДОСЛІДЖЕНЬ

Національний аерокосмічний університет ім. М.Є. Жуковського «Харківський авіаційний інститут»

Анотація

В роботі розглянута оцінка надійності вимірювання категоріальних шкал. Описано методи перевірки первинного вимірювання на надійність: перевірка обґрунтованості, стійкості, точності вимірювання за класичними методами. Запропоновано використання сучасної теорії тестів (IRT) для політомічних питань на основі моделі Nominal Response Model (NRM). Приклади обчислення зроблено за допомогою пакетів відкритого інтегрованого середовища RStudio.

Ключові слова: надійність шкал, сучасна теорія тестів (IRT), категоріальні питання, RStudio, Nominal Response Model

Abstract

In this work the measurement reliability rating categorical scales are considered. Methods of measure the reliability, validity checking, stability, precision are proposed on base classical methods. We used the modern test theory (IRT) for polytomous items based on model Nominal Response Model (NRM). Examples of calculations are present in open integrated environment RStudio with using special packages.

Keywords: reliability of scales, item response theory (IRT), categorical item, RStudio, Nominal Response Model.

Вступ

Вимірювання властивостей об'єктів різної природи є одним зі складних завдань, тому що успіх конкретного дослідження і якість інформації завжди залежить від якості вимірювання. Необхідність вимірювання соціальних характеристик пояснюється як теоретичними, так і практичними завданнями, які виникають у соціально-економічних системах. Вимірювання якісних властивостей за допомогою кількісних параметрів є необхідною умовою будь-якого соціологічного дослідження [1].

В Україні багато організацій у своїй діяльності часто спираються на висновки соціологічних досліджень. Якщо соціологічні дані достовірні, задовольняють науковим критеріям обґрунтованості та надійності, отримані за допомогою апробованих методик, то вони можуть сприяти прийняттю оптимальних рішень. Але іноді отримана інформація спотворюється через помилки, допущені на етапі збору даних (некоректно формулювання питань у соціологічних анкетах, вибір невідповідної шкали оцінювання, порушення методики проведення дослідження і т.п.). Також недостовірні результати можуть виникнути і в процесі аналізу даних через неправильно обрані методи аналізу й некоректну інтерпретацію даних. Також слід відзначити, що проведення соціологічних досліджень завжди є складним і дорогим процесом, тому що потребує певних зусиль, часу й матеріальних витрат. Помилки, що допущені на початкових етапах при складанні анкети, можуть істотно вплинути на кінцевий результат. Тому проблема підвищення якості та ефективності соціологічних досліджень є актуальною проблемою.

Метою даної роботи є аналіз існуючих методів оцінки надійності питань соціологічних анкет, як інструментів вимірювання соціально-економічних процесів.

Результати дослідження

Серед існуючих методів, які дозволяють визначити якість інструментів соціологічного дослідження, найбільш розповсюдженим є метод оцінки надійності, запропонований в [2]. Цей метод базується на аналізі надійності шкали вимірювання з позицій статистичного аналізу і передбачає перевірку обґрунтованості, стійкості, точності і правильності вимірювання.

Обґрунтованість шкали – це впевненість в тому, що за допомогою цієї шкали цілеспрямовано ви-

мірюють саме ту властивість об'єкту дослідження, яка є головною метою аналізу, а не інше.

Щоб підвищити обґрунтованість вимірювання, використовують ряд технічних прийомів:

1. Логічні міркування на основі досвіду і здорового глузду. Наприклад, завдання опитування: визначити частоту комунікацій з друзями. В анкеті є питання «Скільки разів на тиждень Ви зустрічаєтеся з друзями?». Логічно, що це питання слід деталізувати і питати про кількість зустрічей в будні дні, у суботу та неділю. Деталізувати питання для визначення мети зустрічі. Таким чином, отримана інформація буде більш точна при розрахунку середньої частоти комунікацій.

2. Тест за «еталонною» групою. Сенс перевірки полягає у зіставленні даних, отриманих шляхом вимірювання за шкалою, з достовірними відомостями про об'єкт вимірювання.

3. Пошук незалежного критерію вимірювання, який би фіксував або вимірював ту ж саму властивість об'єкта з досить високим ступенем надійності (реєстрація приладом, дані незалежних джерел інформації і т.п.).

4. Використання думок експертів для відбору пунктів шкали. Так, при побудові шкал, які визначають частоту зустрічі з друзями для жінок і чоловіків, треба врахувати можливі зсуви, тому що результати можуть будуть істотно різними для цих груп.

5. Поєднання декількох показників для реєстрації одного певного властивості, або побудова індексу. Типи індексів вкрай різноманітні. Вони широко використовуються в психологічних тестах, в соціально-економічних дослідженнях. Суть індексного обґрунтованості в тому, що, відповідно до гіпотези, даній властивості знаходиться безліч його проявів, для кожного з яких формують окрему шкалу. Потім вимірювання за окремими шкалами або підсумовуються, або з них утворюють логічні конструкції.

Стійкість шкали – це здатність шкали давати однозначну інформацію при будь-якому повторному вимірі. Стійкість часто ототожнюють з надійністю процедури у цілому, однак це різні характеристики шкали. Існують різні способи визначення стійкості.

Перший спосіб – це повторне вимірювання. Один і той же об'єкт вимірюється двічі з двох-тритижневим часовим інтервалом і за допомогою однакової процедури вимірювання. Шкала вважається стійкою, якщо збігу між першими і другими серіями вимірювань будуть досить високими. Показник стійкості оцінюється як відношення кількості пар відповідей, які збіглися при першому та повторному опитуванні до загальної кількості запитань.

Стійкість шкали може визначатися за допомогою залучення експертів, які будуть відповідати на питання соціологічного дослідження і оцінювати один об'єкт із заданими характеристиками. Якщо результати опитування збігаються, то шкалу визначають стійкою.

Найбільш популярним методом визначення надійності є метод «розщеплення тесту навпіл». Цей метод також широко застосовується при оцінці якості педагогічних тестів. Даний метод заснований на припущенні паралельності двох половин тесту і передбачає поділ результатів тестування на дві частини: дані по непарних завдань тесту (X) і по парних (Y). Кореляція двох половин тестів зростає в міру зростання однорідності (гомогенності) тесту. Для визначення надійності шкал використовуються різні показники: Спирмена-Брауна, лямбда Гуттмана і коефіцієнта α -Кронбаха [3].

Точність (правильність) - достатня чутливість шкали, як оптимальна дрібність, що дає стійкий вимір без наявності систематичних помилок. Перевірити надійність шкали за даним критерієм можна за допомогою усунення явних помилок: виявлення працюючої частини шкали, випробування декількох шкал різної дрібності з розрахунком абсолютної помилки кожної з них і порівняння відносних помилок.

Досягнення стійких даних при максимальній дрібності метрики підвищує точність вимірювання.

Правильність і точність вимірювання можна підвищити шляхом розрахунку відносної помилки вимірювання. Відносна помилка дозволяє зіставляти правильність вимірів за двома і більше шкалами різної чутливості і таким шляхом вибрати оптимальний інструмент. Методика визначення дрібності наведена в [2].

Описані вище методи можуть застосовуватися як для дихотомічних шкал оцінювання, так і категоріальних. Методики є доволі складними і потребують спеціальних навичок у статистичних обчисленнях. Хоча деякі методи аналізу автоматизовані і є функціями таких пакетів для обробки статистичних даних, як SPSS та SAS. Але для аналізу надійності шкал та перевірки роботи окремих категорій питань анкет можна також використати модель сучасної теорії тестів (Item Response Theory - IRT) для політомічних завдань (Nominal Response Model - NRM) [4-5].

У моделі NRM не має заздалегідь визначеної залежності між порядком взаємовиключних катего-

рій та сумарними оцінками за усі відповіді на запитання, тобто не накладається умова, що вибір вищої категорії відповідає вищому рівню підготовки. Це дозволяє використовувати цю модель не тільки для тестів, але і для звичайних питань з категоріальною шкалою. По суті, модель NRM визначає популярність категорій запитання і ймовірність відповіді на запитання за формулою:

$$p_{lij}(\theta) = \frac{\exp(a_{jl}\theta_i + c_{jl})}{\sum_{k=1}^m \exp(a_{jk}\theta_i + c_{jk})}$$

де l – номер категорії ($l=1,2,\dots,m$), θ_i - рівень підготовки, a_{jl} - параметр дискримінації, c_{jl} - параметр складності.

Модель будується за принципом багатовимірної моделі Раша для педагогічних тестів [5], яка визначає залежність між рівнем підготовки особи, яка відповідає вирішує тестове завдання та складністю цього завдання. Але у разі NRM накладається додаткові обмеження на параметри: $a_{j0}\theta_i + c_{j0} = 0$. Це дозволяє використовувати цю модель для номінальних категоріальних шкал, а також для впорядкованих відповідей на шкалі Лайкерта.

Для перевірки роботи даної моделі було проаналізовано результати одного з соціологічних досліджень, де респондентам пропонувалося 9 запитань стосовно їх часу відпочинку. Кожне запитання припускало відповідь за 5-категоріальною шкалою. Було отримано результати анкетування за 132 респондентами, які визначили ступінь відношення до різних видів відпочинку.

За допомогою пакету відкритого інтегрованого середовища RStudio й пакету mirt [6-7] було проведено розрахунки та побудовано графіки характеристичних кривих, які відображають ймовірність відповідей на запитання з кожної категорії. На рис. 1 наведено загальний вид кривих за кожним запитанням x11-x33.

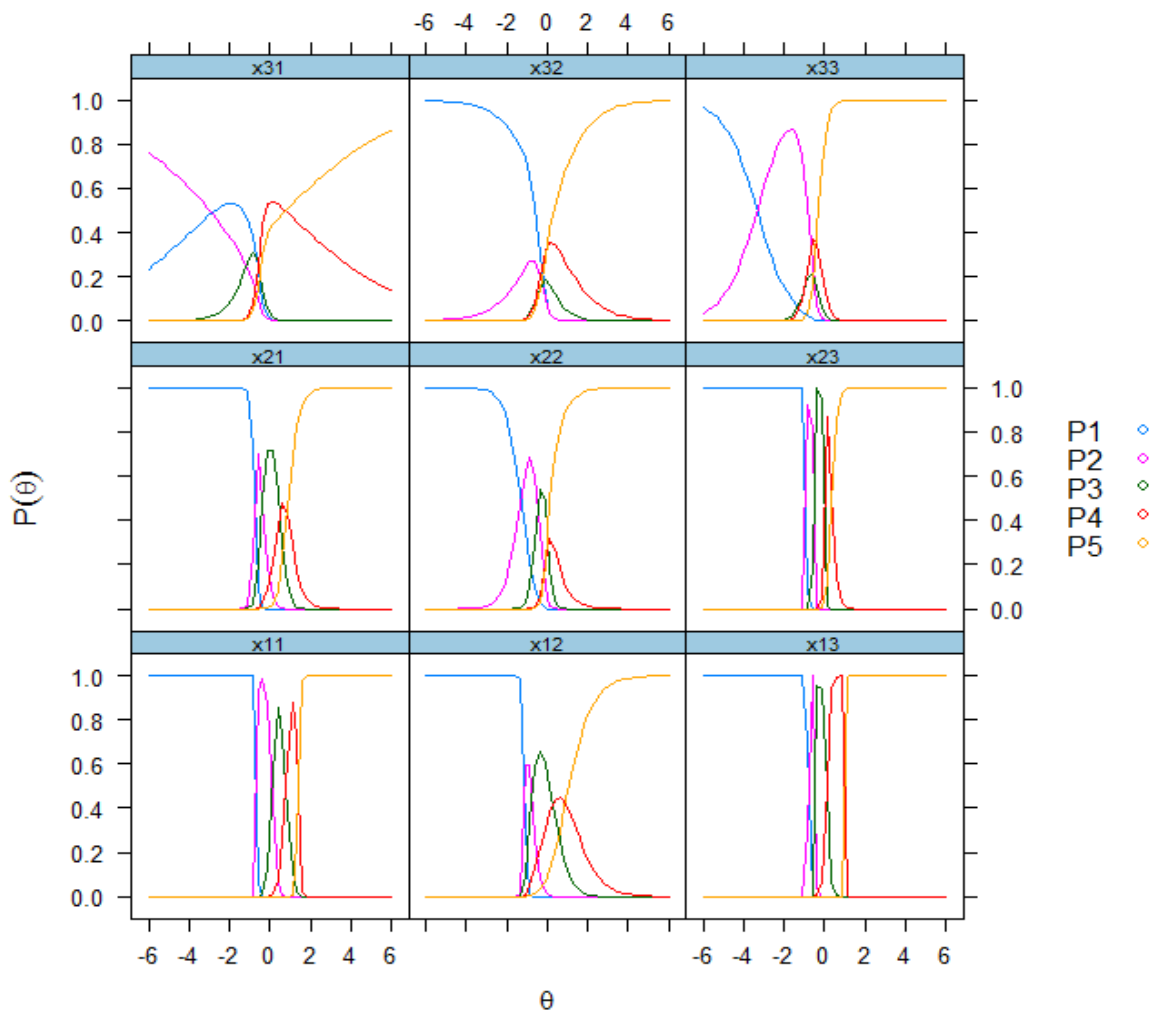


Рис. 1. Характеристичні криві за моделлю NRM для категоріальних запитань анкети

Як інтерпретувати ці графіки? За віссю абсцис – рівень підготовленості учасників, який для шкал запитань анкети дорівнює 0. За віссю ординат – ймовірність відповіді на запитання, яка дозволяє визначити, з якою ймовірністю буде обрана та чи інша категорія шкали. Строго монотонними будуть лише криві, які відповідають найменш та найбільш популярним категоріям запитання. Криві, які відповідають питанням x11, x13 та x23 мають категорії, які працюють однаково, тобто ймовірність вибору цих категорій респондентами однакова. Для запитань x12 та x22 популярність категорій різна, але вони усі «працюють». Проблеми можуть виникнути із запитаннями x31-x33, оскільки на деякі категорії респонденти відповідають частіше, ніж на інші, а деякі – обираються не часто. Тому можна розглянути можливість скорочення кількості категорій для даних запитань з метою отримання більш точних показників. Але серед категорій запитань практично не має тих, які «не працюють», тому можна стверджувати, що питання були сформульовані доволі коректно.

З використанням пакету *mirt*, а також інших можна провести і більш детальний аналіз. Розрахувати, наприклад, коефіцієнт надійності α -Кронбаха для всієї анкети, визначити похибки вимірювання та багато іншого.

Висновки

Таким чином, у роботі проведено аналіз існуючих методів оцінки надійності питань соціологічних анкет. Показано, що окрім існуючих класичних методів аналізу, які дозволяють визначити такі характеристики, як надійність, стійкість, точність шкал вимірювання, можна використовувати і сучасні підходи і методи – моделі теорії IRT. Розрахунки рекомендується робити за допомогою програмного середовища R для статистичної обробки даних, який є відкритим ресурсом з великою кількістю функцій.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Шкалирование при сборе и анализе социологической информации [Текст]: моногр. / Клигер С.А., Косолапов М.С., Толстова Ю.Н. - М.: Наука, 1978. – 112 с.
2. Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности [Текст]: Университетский учебник / В.А. Ядов. – 3-е изд., испр. – Москва: Омега-Л, 2007. – 567 с.
3. Мазорчук М.С. Обоснование выбора методов измерения надежности педагогических тестов / М.С. Мазорчук, В.С. Добряк, Е.О. Соколова, А.А. Сухобрус. - Радиоелектронні і комп'ютерні системи. – 2014.- №4(68). – С. 136 - 142.
4. Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
5. Моделі та методи сучасної теорії тестів: [навчально-методичний посібник] / Т.В. Лісова. – Ніжин: Видавець ПП Лисенко М.М., 2012. - 112 с.
6. CRAN (2016). The Comprehensive R Archive Network. Available at: <https://cran.rproject.org/>
7. R. Philip Chalmers *mirt*: A Multidimensional Item Response Theory Package for the R: Environment Journal of Statistical Software, May 2012, Volume 48, Issue 6. <http://www.jstatsoft.org/>.

Мазорчук Марія Сергіївна – канд. техн. наук, доцент, доцент кафедри інформатики Національного аерокосмічного університету ім. М.Є. Жуковського «ХАІ», Харків, Україна.

Добряк Вікторія Сергіївна - канд. техн. наук, доцент кафедри інформатики Національного аерокосмічного університету ім. М.Є. Жуковського «ХАІ», Харків, Україна.

Mazorchuk Mariia – PhD, Assistant Professor of Informatics Department, National Aerospace University “Kharkiv Aviation Institute”, e-mail: mazorchuk.mary@gmail.com.

Dobriak Viktoriia – PhD, Assistant Professor of Informatics Department, National Aerospace University “Kharkiv Aviation Institute” e-mail: viktoriya--13@mail.ru.