

РУСЛАНА МОСКОТИНА,

PhD із соціології, фахівець 1 категорії лабораторії прикладних соціологічних досліджень факультету соціології Київського національного університету імені Тараса Шевченка (03022, Київ, просп. Глушкова, 4-д)

RUSLANA MOSKOTINA,

PhD in Sociology, 1st Category Specialist, Laboratory of Applied Sociological Research, Faculty of Sociology, Taras Shevchenko National University of Kyiv (4-d, Glushkova Avenue, Kyiv, 03022)

rmoskotina@ukr.net

<https://orcid.org/0000-0002-2195-3121>

Побудова індексу на підставі індикаторів методики GSR-5 із використанням моделі ґрадуйованих відповідей

Тестові методики, як правило, асоціюються із психологічними дослідженнями. Але вони також застосовуються у соціологічних дослідженнях. Тестом є набір пунктів (індикаторів), призначених для вимірювання латентної характеристики [Раєк, Cole, 2020: р. 1]. Латентними характеристиками можуть бути як психологічні (якщо йдеться про психологічні тести), так і соціальні (якщо йдеться про соціологічні тести) феномени. Серед ключових властивостей соціологічних тестів можна виокремити інтегральність, яка передбачає можливість об'єднання індикаторів тесту в один інтегральний показник [Дембицький, 2016: с. 143; Дембіцький, 2019: с. 36; Головаха, Панина, 1997: с. 5]. Інтегральний показник є числовим вираженням латентної характеристики. До інтегральних показників належать, в тому числі, індекси.

Серед теорій, які стосуються розроблення тестів, можна виокремити класичну теорію тестів (classical test theory) та сучасну теорію тестів (item response theory, або IRT). Одиницею аналізу у класичній теорії тестів виступає, як правило, сумарний бал відповідей респондента на запропонований набір пунктів [Раєк, Cole, 2020: р. 1]. Тобто маємо справу з адитивним індексом. Серед переваг адитивних індексів можна виокремити простоту обрахунку та інтерпретації їхніх значень. Однак вони мають і певні недоліки. Розраховуючи значення адитивного індексу, ми виходимо з того, що кожен пункт має однакову вагу. Проте логічно припустити, що так буває не завжди. Також передбачається, що відстані між варіантами відповідей є однаковими для всіх пар сусідніх альтернатив

Цитування: Москотіна, Р. (2022). Побудова індексу на підставі індикаторів методики GSR-5 із використанням моделі ґрадуйованих відповідей. *Соціологія: теорія, методи, маркетинг*, 3, 25–39, <https://doi.org/10.15407/sociology2022.03.025>.

кожного з індикаторів. Наприклад, якщо всі індикатори мають 3-бальну шкалу, відстань між варіантами відповідей 1 та 2 вважається такою ж, як відстань між альтернативами 2 і 3. Та якщо йдеться про порядкові шкали, ця умова може не виконуватись. Тож потрібен інструмент побудови індексів, що усував би вказані недоліки. Один з таких інструментів — модель градуйованих відповідей — пропонує сучасна теорія тестів.

Сучасна теорія тестів та модель градуйованих відповідей

Сучасна теорія тестів позначає низку математичних моделей зв'язку між латентною змінною та пунктами тесту — моделей IRT. Кожен респондент займає якусь позицію на шкалі латентної змінної, що визначає імовірність того, чи обере він той або інший варіант відповіді для певного пункту [Bean, Bowen, 2021]. Латентна змінна позначається як θ . Сучасна теорія тестів ґрунтується на низці припущень [Paek, Cole, 2020: p. 4; de Ayala, 2022: pp. 21–23]:

1. Слід визначити розмірність моделі IRT. Під розмірністю розуміють кількість латентних змінних, в які можна об'єднати пункти тесту. Моделі можуть бути одновимірними (тобто всі індикатори об'єднуються в одну латентну змінну) та багатовимірними.
2. Пункти тесту мають бути локально незалежними. Зокрема, локальна незалежність для одновимірних моделей передбачає, що однієї латентної змінної достатньо для пояснення всіх зв'язків між відповідями на пункти тесту.
3. Варто обрати функцію, що описує залежність між імовірністю відповіді респондента на той чи інший пункт тесту й латентною змінною. Її ще називають функцією відповідей на завдання (item response function). Вибір функції, своєю чергою, залежить від вибору моделі IRT.

Обираючи модель IRT, слід виходити з того, який тип шкали має кожен з пунктів тесту. Для дихотомічних змінних застосовуються модель Раша, 2-параметрична логістична модель, 3-параметрична логістична модель, 4-параметрична логістична модель. Для порядкових шкал досить часто використовується модель градуйованих відповідей (graded response model) [Samejima, 1997]. На ній зупинимося більш детально, адже в нашому випадку побудова індексу здійснюватиметься на підставі індикаторів, що мають порядкову шкалу. Модель градуйованих відповідей передбачає побудову функцій для кожного варіанту відповіді кожного з пунктів тесту — функцій відповідей для категорій (category response functions). Вони характеризують імовірність вибору певного варіанту відповіді на той чи інший пункт тесту залежно від значення латентної змінної. Спочатку розраховуються ймовірності вибору варіанту відповіді k і більше на кожен з пунктів тесту залежно від значення θ . Це можна подати у вигляді такої формули [Meade, Lautenschlager, 2004: p. 365; Bean, 2021]:

$$P_{ik}^*(\theta_s) = \frac{e^{a_i(\theta_s - b_{ik})}}{1 + e^{a_i(\theta_s - b_{ik})}}, \quad (1)$$

де $P_{ik}^*(\theta_s)$ — імовірність того, що респондент s з певним значенням θ обере варіант відповіді k і більше на пункт тесту i ; a — параметр дискримінації, дає змогу визначити, наскільки пункт тесту i диференціює респондентів з різними

значеннями θ (також може розглядатись як характеристика сили зв'язку між θ і пунктом тесту i); b — параметр розташування. Інтерпретується як значення θ , для якого респондент з імовірністю 0,5 обере варіант відповіді k і більше на пункт тесту i . Кількість параметрів b становить $n-1$, де n — кількість варіантів відповіді для пунктів тесту.

Знаючи імовірність вибору варіанта відповіді k і більше на кожен з пунктів тесту, можна розрахувати імовірність вибору варіанту відповіді k залежно від значення латентної змінної. Це можна виразити у вигляді такої формули:

$$p_{ik}^*(\theta_s) = P_{ik}^*(\theta_s) - P_{ik+1}^*(\theta_s), \quad (2)$$

де $p_{ik}^*(\theta_s)$ — імовірність того, що респондент s з певним значенням θ обере варіант відповіді k на пункт тесту i .

Графіки функцій мають вигляд кривих, побудованих для кожного варіанту відповіді кожного з пунктів тесту — кривих відповідей для категорій (category response curves). Наприклад, якщо індикатори мають 7-бальну порядкову шкалу, для кожного з них будується графік, на який нанесено 7 кривих. За допомоги їх можна візуально відобразити імовірність вибору варіанта відповіді k на той чи інший пункт тесту для кожного значення θ .

У сучасній теорії тестів використовують поняття «інформація». Інформацію можна визначити як кількісний показник знань про латентну змінну, отриману на підставі балу пункту тесту чи загального балу за шкалою [Nugent, 2017]. Кількість інформації для пункту тесту залежить від значень параметра дискримінації a та параметрів розташування b . Чим більше значення a , тим точніше цей пункт дає змогу оцінити значення латентної змінної. Також чим більша відстань між значеннями параметрів b , тим точніше пункт тексту дає змогу оцінити значення латентної змінної [Bean, 2021]. Визначити, наскільки точно кожен з пунктів дає змогу оцінити значення θ , можна графічно, за допомоги інформаційних кривих (item information curves). Чим більшою є площа під кривою, тим точніше цей пункт тесту дає змогу оцінити значення θ .

Модель градуйованих відповідей дає змогу згенерувати значення θ ; ця змінна і буде шуканим індексом. θ набуває як додатних, так і від'ємних значень, нерідко має стандартний нормальний розподіл із середнім значенням 0 та стандартним відхиленням 1. Це може викликати незручності в інтерпретації змінної. Модель градуйованих відповідей дає змогу співвіднести значення змінної θ з очікуваними значеннями адитивного індексу¹. Графічно це можна зобразити у вигляді характеристичної кривої для тесту (test characteristic curve). Крім того, можна зберегти значення змінної θ у шкалі значень адитивного індексу, що полегшує її подальшу інтерпретацію. В цьому полягає перевага моделі градуйованих відповідей порівняно з моделями конфірматорного факторного аналізу. Останні теж можуть бути використані для обрахунку значень індексів, що враховують різну вагу індикаторів та відстані між варіантами відповідей для кожного з них, але не пропонують інструментів їх перетворення в іншу шкалу.

1 Якщо точніше, то модель градуйованих відповідей співвідносить значення θ з істинними балами (true scores). Істинний бал — це сумарний бал відповідей респондента на індикатори тесту. Саме таким чином розраховується адитивний індекс.

Методика GSR-5: побудова індексу

Схарактеризувавши модель градуйованих відповідей, перейдемо безпосередньо до побудови індексу для індикаторів методики GSR-5¹, розробленої С. Дембіцьким [Дембіцький, 2022]. Кожен з них має 3-бальну порядкову шкалу² (див. табл. 1). Методика вимірює загальні установки щодо держави загального добробуту. Вони характеризують фонове ставлення населення до своєї країни (а саме до України). З одного боку, це ставлення зумовлене особливостями функціонування державних інститутів, подіями, що мають місце в Україні. З іншого боку, воно визначається стереотипними оцінками цих інститутів і подій. Індикатори методики GSR-5 стосуються різних часових вимірів. Пункти В4 і В2 характеризують уявлення респондентів про минуле та майбутнє української держави. Решта індикаторів стосуються уявлень про теперішній стан справ у країні. Ми спиралися на дані опитування Соціологічної групи «Рейтинг», проведеного у травні 2022 року. Застосовувався метод САТІ, вибіркова сукупність становила 2000 респондентів, що репрезентували населення України віком від 18 років і старше³.

Таблиця 1

Індикатори GSR-5

Індикатори	Варіанти відповідей
В1. Як Ви оцінюєте ефективність української держави на цьому етапі?	1. Центральні органи влади майже не справляються зі своїми обов'язками 2. Частково справляються 3. Справляються майже зі всіма своїми обов'язками
В2. Що ви думаєте про майбутнє України?	1. Найімовірніше ситуація буде погіршуватися 2. Ситуація не буде погіршуватися, але навряд чи покращиться 3. Найімовірніше ситуація буде покращуватися
В3. Якими, на Ваш погляд, є умови життя в Україні для більшості населення?	1. Загалом погані 2. Загалом задовільні 3. Загалом гарні
В4. Що Ви скажете про досягнення та невдачі України, починаючи з 1991 року і до сьогоднішнього дня?	1. Переважали невдачі 2. Невдачі і досягнення компенсували одне одного 3. Переважали досягнення
В5. Якою мірою Ви задоволені подіями, які відбуваються в нашій країні зараз?	1. Радше не задоволені 2. Наскільки задоволені, настільки і не задоволені 3. Радше задоволені

1 Від словосполучення «generalized state reality». Як можна здогадатися з назви, методика включає 5 індикаторів.

2 Кожен з пунктів включає варіант «Важко відповісти», а отже, вони мають номінальну шкалу. Цю альтернативу було перекодовано у пропущені значення, таким чином, змінні стали порядковими.

3 Детальнішу інформацію про опитування групи «Рейтинг» та його результати можна переглянути за посиланням: https://ratinggroup.ua/research/ukraine/dvenadcatyy_obschenacionalnyy_opros_dinamika_ocenki_obraza_gosudarstva_18-19_maya_2022.html

Вважається, що «чим більшою мірою, на думку дослідника, результати вимірювання зумовлені різними дискурсами і/або навколишньою соціальною дійсністю респондента, тим вагомішою є необхідність конструювання індексу» [Дембіцький, 2019: с. 269]. Вище зазначалося, що загальні установки щодо держави загального добробуту залежать від особливостей функціонування державних інститутів, від подій, що відбуваються в країні. Стереотипні уявлення респондентів про роботу державних інститутів та щодо подій в Україні, зі свого боку, теж чинять вплив. Однак, по-перше, ці стереотипні уявлення сформовані зовнішніми чинниками. Зокрема, тим, як у тих чи інших ЗМІ подається інформація про ситуацію в країні. По-друге, респонденти можуть переглянути власні стереотипні уявлення у відповідь на зміни, що відбуваються в країні. З цього робимо висновок, що результати вимірювання все ж досить суттєво визначаються навколишньою дійсністю респондента. Це свідчить на користь того, що ми можемо будувати індекс для індикаторів методики GSR-5.

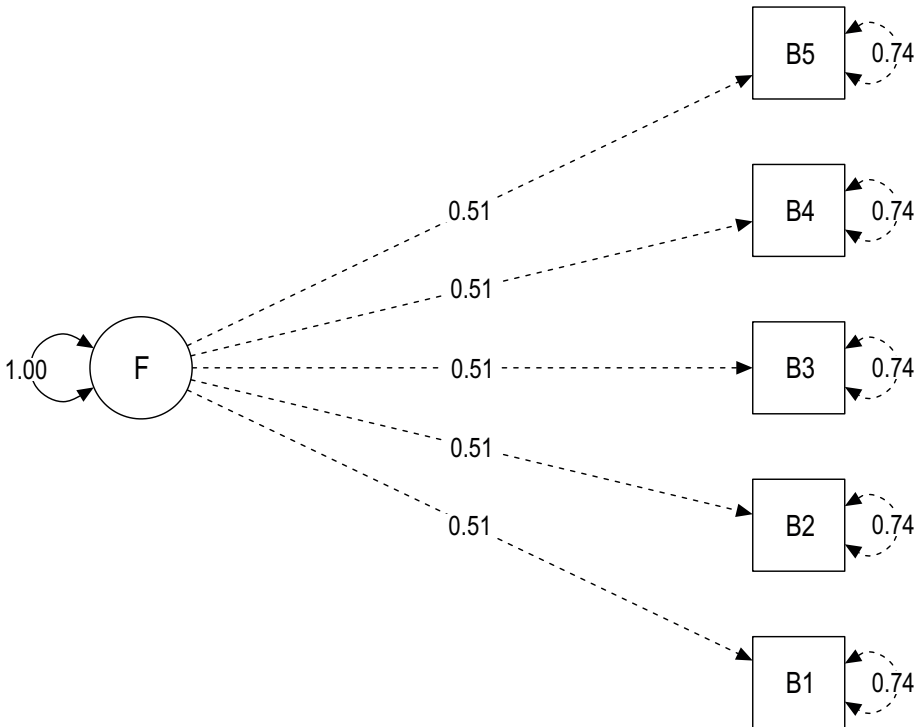


Рис. 1. Модель конфірмаційного факторного аналізу (факторні навантаження та дисперсії стандартизовано)

Спершу визначимо, чи має сенс побудова адитивного індексу для пунктів методики. Слід переконатися, що індикатори мають однакову вагу. З цією метою було проведено конфірмаційний факторний аналіз (КФА) у програмному середовищі R (пакет «lavaan»). Позаяк дані виміряні у порядковій шкалі, було застосовано метод DWLS, адже він забезпечує більш точні оцінки параметрів

[Mindrila, 2010: p. 61]. Модель КФА передбачала виокремлення однієї латентної змінної, до якої входили всі індикатори методики GSR-5, що мали однакові факторні навантаження (див. рис. 1). Аналіз відповідності моделі даним оцінювався за показниками CFI, TLI, RMSEA, SRMR¹. Якщо значення CFI та TLI перевищують 0,95, а значення RMSEA та SRMR є меншими за 0,06 та 0,08 відповідно, це свідчить про хорошу відповідність моделі даним [West et al., 2012].

Як виявилось, модель КФА не демонструє прийнятної відповідності даним: значення CFI становить 0,838, значення TLI — 0,820, а RMSEA та SRMR — 0,091 та 0,092 відповідно. Можна припустити, що індикатори мають різну вагу, а отже, виникає потреба в альтернативній моделі. Тож побудовано одновимірну модель градуїзованих відповідей (пакет «mirt» в R).

Щоб перевірити відповідність моделі градуїзованих відповідей (моделі GRM) даним, застосовувалася статистика C_2 . Її розроблено для порядкових шкал; її перевагою є також те, що вона не накладає обмежень на кількість індикаторів залежно від кількості варіантів відповіді на кожен з них² [Paek, Cole, 2020; Cai, Monroe, 2014]. Для статистики розраховується рівень значимості. Якщо він перевищує 0,05, модель відповідає даним. У нашому випадку рівень значимості для C_2 перевищує 0,05, що свідчить про відповідність моделі GRM даним.

Відповідність моделі даним також оцінювалася за показниками CFI, TLI, RMSEA, SRMSR. Якщо CFI та TLI більші або дорівнюють 0,95, а RMSEA та SRMSR не перевищують 0,05, це свідчить про хорошу відповідність моделі даним [Noe-Grijalva et al., 2022]. Значення CFI та TLI становлять 0,999, RMSEA — 0,007, SRMSR — 0,015. Такі значення свідчать про хорошу відповідність моделі GRM даним. Так як побудована нами модель є одновимірною і демонструє гарну відповідність даним, це свідчить на користь того, що однієї латентної змінної достатньо для пояснення зв'язків між відповідями на пункти методики GSR-5.

Також слід визначити, наскільки пункти методики відповідають моделі GRM. У разі невідповідності індикатора моделі ми, за потреби, можемо його вилучити. Для оцінювання відповідності пунктів застосовувалася статистика $S\text{-}\chi^2$, на її основі розраховується показник RMSEA. Якщо його значення для індикатора не перевищує 0,06, він добре відповідає побудованій моделі [Bean, 2021]. Значення RMSEA менші, ніж 0,06, для кожного з індикаторів методики, а отже, всі вони добре відповідають побудованій моделі (див. табл. 2).

1 Значення CFI (Comparative Fit Index) знаходяться в діапазоні від 0 до 1. TLI (Tucker-Lewis Index) набуває значень від 0 до 1, але подекуди вони можуть виходити за вказані межі. Значення RMSEA (Root Mean Square Error of Approximation) та SRM(S)R (Standardized Root Mean Square Residual) перебувають в межах від 0 до 1.

2 На відміну від статистики M_2^* , також розробленої для порядкових змінних (хоча загалом вона може бути застосована і для дихотомічних змінних). Вона накладає обмеження на кількість індикаторів залежно від кількості варіантів відповіді. Наприклад, для розрахунку статистики потрібно щонайменше 6 дихотомічних змінних, а якщо індикатори мають 5-бальну шкалу, їх потрібно не менш як 12 [Paek, Cole, 2020; Cai, Hansen, 2013]. Нагадаємо, що ми маємо справу з 5 індикаторами, які мають 3-бальну шкалу; якщо керуватися вказаними критеріями, ми не маємо змоги розрахувати статистику M_2^* .

Таблиця 2

Відповідність індикаторів методики GSR-5 побудованій моделі

Індикатори	RMSEA на підставі $S-\chi^2$
V1. Як Ви оцінюєте ефективність української держави на цьому етапі?	0,030
V2. Що ви думаєте про майбутнє України?	0,013
V3. Якими, на Ваш погляд, є умови життя в Україні для більшості населення?	0,000
V4. Що Ви скажете про досягнення та невдачі України, починаючи з 1991 року і до сьогоднішнього дня?	0,019
V5. Якою мірою Ви задоволені подіями, які відбуваються в нашій країні зараз?	0,031

Пересвідчившись, що модель демонструє хорошу відповідність даним, перейдемо до аналізу значень її параметрів — параметра дискримінації a і параметрів розташування b (див. табл. 3).

Таблиця 3

Параметри моделі GRM

Індикатори	a	b_1	b_2
V1. Як Ви оцінюєте ефективність української держави на цьому етапі?	1,636	-2,617	-0,170
V2. Що ви думаєте про майбутнє України?	1,571	-2,212	-1,113
V3. Якими, на Ваш погляд, є умови життя в Україні для більшості населення?	1,327	-0,962	2,225
V4. Що Ви скажете про досягнення та невдачі України, починаючи з 1991 року і до сьогоднішнього дня?	0,562	-1,499	2,703
V5. Якою мірою Ви задоволені подіями, які відбуваються в нашій країні зараз?	0,474	0,299	3,429

Як уже зазначалося, параметр a характеризує силу зв'язку між пунктами та латентною змінною θ . Індикатор V1 демонструє найбільшу силу зв'язку з латентною змінною, а V5 — найменшу. Різні значення параметру дискримінації вказують на те, індикатори мають неоднакову вагу. Модель GRM дає змогу цю обставину взяти до уваги при розрахунку значень латентної змінної.

Також маємо параметри b : їхня кількість становить $n - 1$, де n — кількість варіантів відповіді, що мають індикатори. В нашому випадку індикатори мають по 3 варіанти відповідей, тож у нас два параметри b . Параметр b інтерпретується як значення θ , для якого респондент з імовірністю 0,5 обере варіант відповіді k і більше для того чи іншого пункту. Тобто якщо b_1 становить $-2,617$, це означає, що для $\theta = -2,617$ респондент з імовірністю 0,5 обере варіанти відповіді «Частково справляються» або «Справляються майже зі всіма своїми обов'язками» для індикатора V1. А коли b_2 дорівнює $-0,170$, це означає, що для вказаного значення θ респондент з імовірністю 0,5 обере варіант «Справляються майже зі всіма своїми обов'язками» для індикатора V1. Значення параметрів b також показують, які ділянки шкали θ охоплюють варіанти відповіді того чи іншого

індикатора. Так, індикатори B1 і B2 характеризуються найбільшою силою зв'язку з латентною змінною, але краще «працюють» для лівої (від'ємної) частини шкали θ . B3 і B4 покривають ліву і праву (додатну) частини шкали змінної θ . А індикатор B5 краще «працює» для правої частини шкали θ .

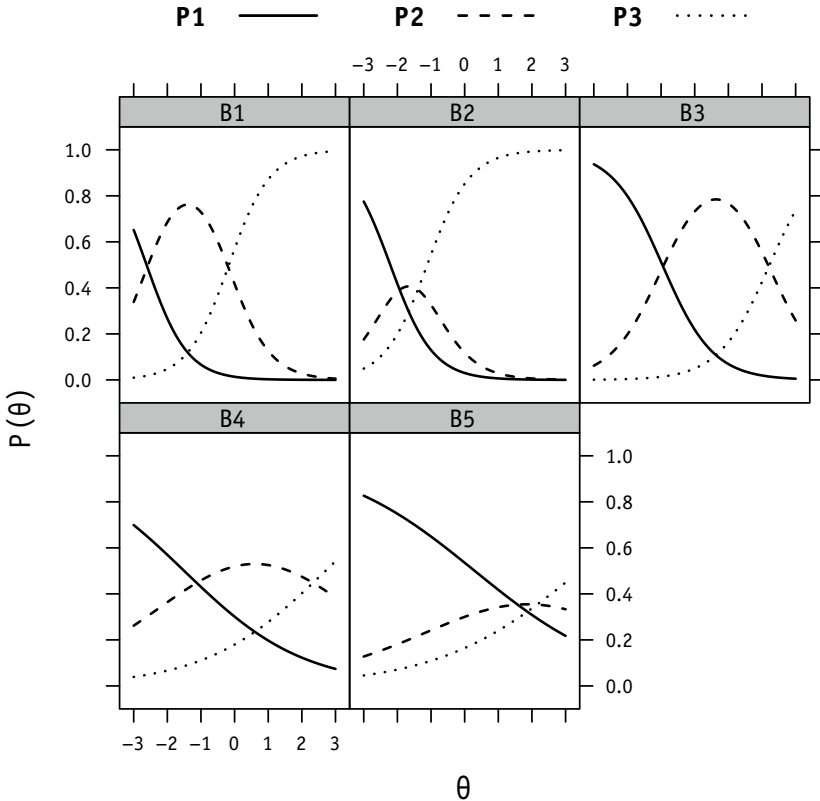


Рис. 2. Криві відповідей для категорій (модель GRM)

Імовірність вибору варіантів відповіді кожного з пунктів методики GSR-5 можна зобразити у вигляді кривих відповідей для категорій (див. рис. 2). Як бачимо, перший варіант відповіді (P1) для індикаторів B1–B4 покриває, більшою мірою, від'ємну частину шкали θ . Для індикатора B5 він охоплює фактично всю шкалу латентної змінної. Третій варіант відповіді (P3) переважно покриває додатну частину шкали θ . Для пункту B2 він охоплює значну ділянку шкали латентної змінної. Другий варіант відповіді (P2) для індикаторів B1 і B2 покриває переважно від'ємну частину шкали. Причому для пункту B2 цей варіант відповіді охоплює порівняно невелику ділянку шкали θ . Для індикатора B3 другий варіант відповіді охоплює і додатні, і від'ємні значення шкали латентної змінної. Те саме можна сказати і про індикатори B4 й B5. Ми бачимо, що деякі варіанти відповідей для певних пунктів можуть охоплювати значну ділянку шкали латентної змінної, а деякі — лише її невелику частину. Так, перший і другий варіант відповіді індикатора B2 займають майже всю ділянку шкали змінної θ . Тож

можна поставити під сумнів те, що для кожного з пунктів методики GSR-5 відстані між парами сусідніх альтернатив сприймаються респондентами однаково. Модель GRM при побудові латентної змінної враховує відстані між варіантами відповідей кожного з індикаторів.

Аналіз кривих відповідей для категорій допомагає в подальшій інтерпретації значень латентної змінної. θ може набувати як додатних, так і від'ємних значень. Перший варіант відповіді для індикаторів методики характеризує негативні уявлення про Україну. І цей же варіант відповіді охоплює, як правило, ліву частину шкали θ . Тобто від'ємні значення латентної змінної характеризують негативні установки щодо держави загального добробуту. Третій варіант відповіді покриває переважно праву частину шкали θ і характеризує позитивні уявлення про Україну. Тобто додатні значення θ свідчать про позитивні установки щодо держави загального добробуту.

Побудуємо інформаційні криві для індикаторів методики (див. рис. 3). Чим більша площа під кривою, тим точніше пункт методики дає змогу оцінити значення латентної змінної. Найбільшою площею під кривою характеризуються пункти B1 і B2, а найменшою — індикатори B4 та B5. Причому B1 та B2 точніше оцінюють значення від'ємної частини шкали латентної змінної. Особливо добре це помітно для пункту B2. А індикатор B3, хоч і характеризується помітно меншою площею під кривою, «працює» на всій ділянці шкали θ .

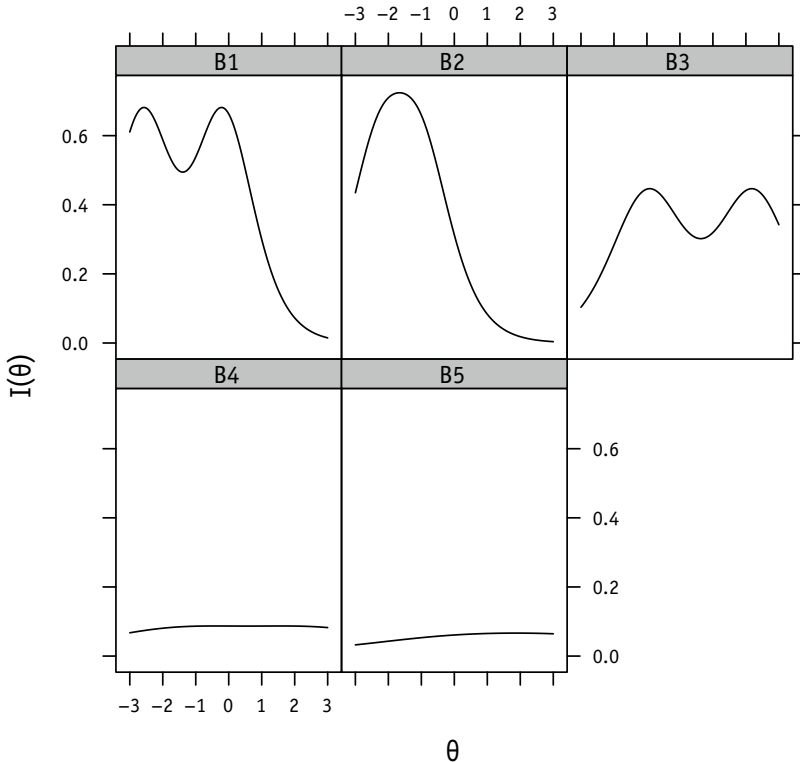


Рис. 3. Інформаційні криві (модель GRM)

Нарешті побудуємо характеристичну криву для тесту, яка дасть змогу співвіднести значення змінної θ та адитивного індексу (див. рис. 4). Значення останнього перебуває в межах від 5 до 15. Чим більшим є значення індексу, тим більш позитивними є загальні установки щодо держави загального добробуту. За допомоги цієї кривої ми можемо оцінити, яке значення адитивного індексу відповідає тому чи іншому значенню θ . Так, якщо θ дорівнює -3 , значення адитивного індексу становить 6,25. А якщо θ становить 3, значення адитивного індексу дорівнює 13,42.

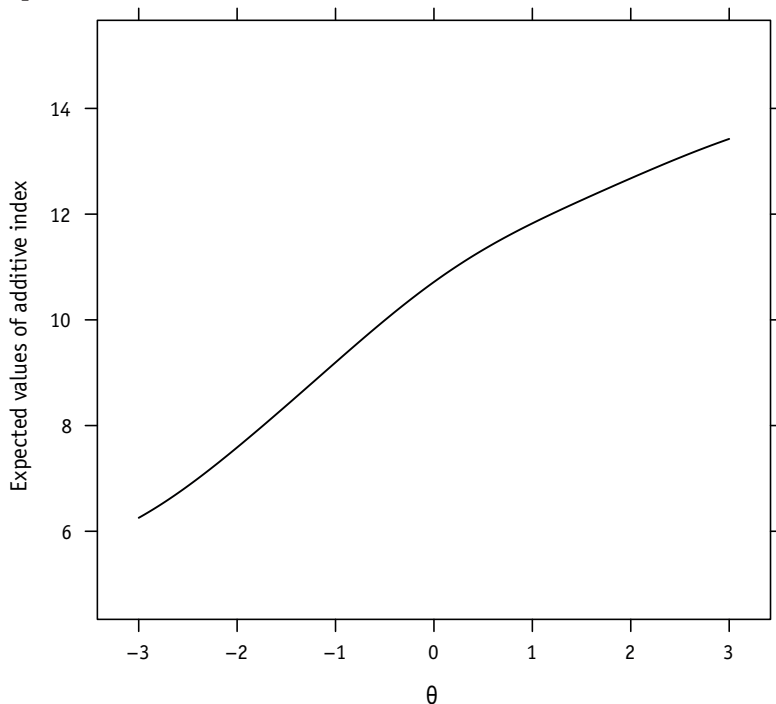


Рис. 4. Характеристична крива для тесту (модель GRM)

Ми маємо змогу зберегти θ у шкалі значень адитивного індексу. Причому, як видно з характеристичної кривої, перетворення значень латентної змінної θ у шкалу значень адитивного індексу є нелінійним.

Новостворений індекс характеризується кращою диференціовальною здатністю (його шкала має більшу кількість градацій), ніж адитивний індекс. Для обох індексів — адитивного та розрахованого за допомоги моделі GRM (індекс GRM) — було побудовано боксплоти (див. рис. 5).

Бачимо, що адитивний індекс має більший розмах та більший міжквартильний розмах. Значення першого квартилю для адитивного індексу менше, а третього квартилю більше, ніж для розрахованого за допомоги моделі GRM. Мінімальне значення для адитивного індексу менше, а максимальне є більшим порівняно з індексом GRM. При цьому медіани для них візуально не дуже від-

різняються. Для адитивного індексу медіанне значення становить 11,00, а для індексу GRM — 10,74.

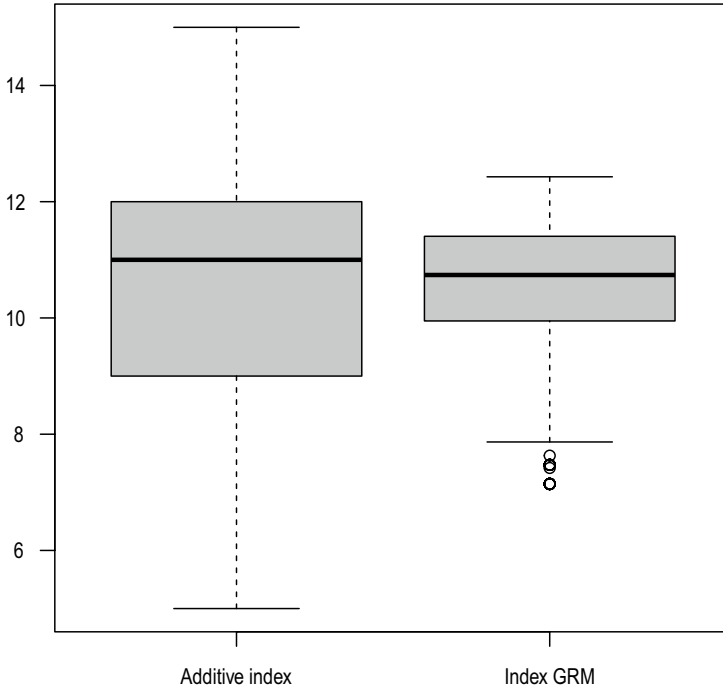


Рис. 5. Порівняння значень індексів (адитивного та розрахованого за допомогою моделі GRM)

Висновки

Ми вели мову про побудову індексу для індикаторів методики GSR-5 за допомогою моделі ґрадуєваних відповідей. Найпростіший спосіб його конструювання — знайти сумарний бал відповідей респондента на пункти методики. Адитивний індекс передбачає, що індикатори мають однакову вагу. Однак побудувавши однофакторну модель КФА з однаковими факторними навантаженнями для пунктів, ми з'ясували, що вона не демонструє прийнятної відповідності даним. Відтак розрахунок адитивного індексу не є доцільним. Тож було застосовано одновимірну модель ґрадуєваних відповідей, яка продемонструвала хорошу відповідність даним. З одного боку, такий спосіб розрахунку індексу є більш трудомістким, ніж обрахунок адитивного індексу. З іншого боку, індекс, побудований із використанням моделі ґрадуєваних відповідей, має кілька переваг. По-перше, він враховує те, що індикатори мають різну вагу та роблять неоднаковий внесок в індекс. По-друге, враховано відстані між варіантами відповідей, адже вони не обов'язково є однаковими. Завдяки цьому індекс має вищу диференціувальну здатність порівняно з адитивним. Крім того, модель ґраду-

йованих відповідей з огляду на полегшення інтерпретації дає змогу подати бали θ у шкалі значень адитивного індексу. У цьому полягає її перевага порівняно з моделями конфірмаційного факторного аналізу, які теж можуть бути використані для побудови індексів, але не дають змоги здійснювати перетворення латентних змінних з однієї шкали в іншу.

Модель градуїзованих відповідей має справу зі змінними з порядковою шкалою. А індикатори методики GSR-5 включають варіант відповіді «Важко відповісти», що робить їх номінальними. Після перекодування альтернативи «Важко відповісти» у пропущені значення (як це було зроблено в нашому випадку) змінні стають порядковими. Однак ми втрачаємо певну кількість респондентів, котрі теж надали відповіді на ті чи інші пункти методики GSR-5. Тож наступним кроком може стати побудова інтегрального показника, який уможливить урахування відповідей тих респондентів, котрі обрали варіант «Важко відповісти».

Джерела

Головаха, Е., Панина, Н. (1997). *Интегральный индекс социального самочувствия (ИИСС): конструирование и применение социологического теста в массовых опросах*. Киев: Ин-т социологии НАН Украины.

Дембицький, С. (2016). Социологические тесты: сущность и валидизация. *Социология: теория, методы, маркетинг*, 3, 140–155.

Дембицький, С. (2019). *Розробка соціологічних тестів: методологія і практики її застосування*. Київ: Інститут соціології НАН України.

Дембицький, С. (2022). *Загальні установки щодо держави загального добробуту: концепт та особливості вимірювання*. [Препринт наданий автором.]

Bean, G.J., Bowen, N.K. (2021). Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*, 18 (6), 597–618.

Bean, J. (2021). *Using R for Social Work Research*. Retrieved from: https://bookdown.org/bean_jerry/using_r_for_social_work_research/

Cai, L., Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. *British Journal of Mathematical and Statistical Psychology*, 66 (2), 245–276. DOI: <https://doi.org/10.1111/j.2044-8317.2012.02050.x>

Cai, L., Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data (CRESST Report 839)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

de Ayala, R.J. (2022). *The theory and practice of item response theory*. S.l.: Guilford Press.

Meade, A.W., Lautenschlager, G.J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, 7 (4), 361–388. DOI: <https://doi.org/10.1177/1094428104268027>

Mindrila, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society*, 1 (1), 60–66.

Noe-Grijalva, M., Polo-Ambrocio, A., Gómez-Bedia, K., Caycho-Rodríguez, T. (2022). Spanish Translation and Validation of the COVID Stress Scales in Peru. *Frontiers in psychology*, 13, 840302. DOI: <https://doi.org/10.3389/fpsyg.2022.840302>

Nugent, W.R. (2017). Understanding DIF and DTF: Description, methods, and implications for social work research. *Journal of the Society for Social Work and Research*, 8 (2), 305–334. DOI: <https://doi.org/10.1086/691525>

Paek, I., Cole, K. (2020). *Using R for item response theory model applications*. S.l.: Routledge.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. Retrieved from: <https://www.R-project.org/>

Samejima, F. (1997). Graded response model. In: W.J. Van der Linden, R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York: Springer.

West, S.G., Taylor, A.B., Wu, W. (2012). Model fit and model selection in structural equation modeling. In: R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). S.I.: Guilford Press.

Отримано 10.08.2022

References

Bean, G.J., Bowen, N.K. (2021). Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*, 18 (6), 597–618.

Bean, J. (2021). *Using R for Social Work Research*. Retrieved from: https://bookdown.org/bean_jerry/using_r_for_social_work_research/

Cai, L., Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. *British Journal of Mathematical and Statistical Psychology*, 66 (2), 245–276. DOI: <https://doi.org/10.1111/j.2044-8317.2012.02050.x>

Cai, L., Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data (CRESSST Report 839)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESSST).

de Ayala, R.J. (2022). *The theory and practice of item response theory*. S.I.: Guilford Press.

Dembitskiy, S. (2016). Sociological tests: the essence and validation. [In Russian]. *Sociology: theory, methods, marketing*, 3, 140–155. [=Дембицкий 2016].

Dembitskiy, S. (2019). *Development of sociological tests: methodology and practice of its application*. [In Ukrainian]. Kyiv: Institute of Sociology, National Academy of Sciences of Ukraine. [=Дембицький 2019].

Dembitskiy, S. (2022). *General attitudes towards the general welfare state: concept and measurement*. [In Ukrainian]. Preprint provided by the author. [=Дембицький 2022].

Golovakha, E., Panina, N. (1997). *Integral index of social well-being (IISS): Design and application of a sociological test in mass surveys*. [In Russian]. Kyiv: Institute of Sociology, National Academy of Sciences of Ukraine. [=Головаха, Панина 1997].

Meade, A.W., Lautenschlager, G.J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods*, 7 (4), 361–388. DOI: <https://doi.org/10.1177/1094428104268027>

Mindrila, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society*, 1 (1), 60–66.

Noe-Grijalva, M., Polo-Ambrocio, A., Gómez-Bedia, K., Caycho-Rodríguez, T. (2022). Spanish Translation and Validation of the COVID Stress Scales in Peru. *Frontiers in psychology*, 13, 840302. DOI: <https://doi.org/10.3389/fpsyg.2022.840302>

Nugent, W.R. (2017). Understanding DIF and DTF: Description, methods, and implications for social work research. *Journal of the Society for Social Work and Research*, 8 (2), 305–334. DOI: <https://doi.org/10.1086/691525>

Paek, I., Cole, K. (2020). *Using R for item response theory model applications*. S.I.: Routledge.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. Retrieved from: <https://www.R-project.org/>

Samejima, F. (1997). Graded response model. In: W.J. Van der Linden, R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York: Springer.

West, S.G., Taylor, A.B., Wu, W. (2012). Model fit and model selection in structural equation modeling. In: R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). S.I.: Guilford Press.

Received 10.08.2022

РУСЛАНА МОСКОТИНА

Побудова індексу на підставі індикаторів методики GSR-5 із використанням моделі ґрадуйованих відповідей

Статтю присвячено побудові індексу для індикаторів методики GSR-5. Методика дає змогу вимірювати загальні установки щодо держави загального добробуту. Найпростішим способом є побудова адитивного індексу: розрахунок сумарного балу відповідей респондента на пункти методики. Такий індекс хоч і є простим в інтерпретації, має певні недоліки. По-перше, він передбачає, що всі індикатори мають однакову вагу. Проте логічно припустити, що це не завжди так. По-друге, вважається, що відстані між варіантами відповідей є однаковими для всіх пар сусідніх альтернатив кожного з індикаторів. Однак якщо маємо справу з порядковими змінними, ця умова може не виконуватись. Таким чином, виникає потреба в інструменті, який дасть змогу сконструювати індекс, котрий усував би вказані недоліки. Одним з таких інструментів може стати модель ґрадуйованих відповідей; вона призначена для роботи з порядковими змінними. Передусім було з'ясовано, чи доцільною є побудова адитивного індексу для індикаторів методики GSR-5. Побудувавши однофакторну модель конфірматорного факторного аналізу з однаковими факторними навантаженнями для пунктів, було з'ясовано, що вона не демонструє прийнятної відповідності даним. Звідси випливає, що розрахунок адитивного індексу для індикаторів GSR-5 недоцільний. Тому постає потреба в альтернативній моделі. Оскільки пункти методики GSR-5 мають порядкову шкалу, для них було побудовано одновимірну модель ґрадуйованих відповідей (модель GRM). Вона продемонструвала хорошу відповідність даним. Модель GRM дала можливість сконструювати такий індекс, який враховує різну вагу пунктів та відстані між варіантами відповідей для кожного з індикаторів. Він характеризується кращою диференціальною здатністю, ніж адитивний. Крім того, модель ґрадуйованих відповідей з метою полегшення інтерпретації дає змогу зберегти значення латентної змінної у шкалі значень адитивного індексу. В цьому полягає перевага моделі ґрадуйованих відповідей, порівняно з моделями конфірматорного факторного аналізу. Останні теж можуть бути застосовані як інструменти для побудови адитивних індексів, однак не передбачають перетворення латентних змінних з однієї шкали в іншу.

Ключові слова: модель ґрадуйованих відповідей, GSR-5, латентна змінна, конфірматорний факторний аналіз, адитивний індекс

РУСЛАНА МОСКОТИНА

Построение индекса на основе индикаторов методики GSR-5 с использованием модели градуированных ответов

Статья посвящена построению индекса для индикаторов методики GSR-5. Методика позволяет измерять общие установки по отношению к государству всеобщего благосостояния. Простейшим способом является построение аддитивного индекса: расчет суммарного балла ответов респондента на пункты методики. Такой индекс является простым в интерпретации, но имеет определенные недостатки. Во-первых, он предполагает, что все индикаторы имеют одинаковый вес. Но логично предположить, что так бывает не всегда. Во-вторых, считается, что расстояния между вариантами ответов одинаковы для всех пар соседних альтернатив каждого индикатора. Однако если мы имеем дело с порядковыми переменными,

это условие может не выполняться. Таким образом, возникает потребность в инструменте, который позволил бы сконструировать индекс, устраняющий указанные недостатки. Одним из таких инструментов является модель градуированных ответов; она предназначена для работы с порядковыми переменными. Прежде всего, была определена целесообразность построения аддитивного индекса для индикаторов методики GSR-5. После построения однофакторной модели конфирматорного факторного анализа с одинаковыми факторными нагрузками для пунктов было установлено, что она не демонстрирует приемлемого соответствия данным. Отсюда следует, что расчет аддитивного индекса для индикаторов GSR-5 нецелесообразен. Потому возникла потребность в альтернативной модели. Так как пункты методики GSR-5 имеют порядковую шкалу, для них была построена одномерная модель градуированных ответов (модель GRM). Она продемонстрировала хорошее соответствие данным. Модель GRM позволила сконструировать индекс, учитывающий разный вес пунктов и расстояния между вариантами ответов для каждого индикатора. Он характеризуется лучшей дифференцирующей способностью, чем аддитивный. Кроме того, модель градуированных ответов с целью облегчения интерпретации позволяет сохранить значение латентной переменной в шкале значений аддитивного индекса. В этом состоит преимущество модели градуированных ответов по сравнению с моделями конфирматорного факторного анализа. Последние тоже могут быть применены в качестве инструментов построения аддитивных индексов, однако не предполагают преобразования латентных переменных из одной шкалы в другую.

Ключевые слова: модель градуированных ответов, GSR-5, латентная переменная, конфирматорный факторный анализ, аддитивный индекс

RUSLANA MOSKOTINA

Construction of the index based on the GSR-5 items using the graded response model

This article is about building of the index for GSR-5 items (questions). They measure general attitudes towards the general welfare state. The simplest way is to calculate an additive index (a total score of the respondent's answers to GSR-5 items). Such an index is easy to interpret but it has some limitations. Firstly, it assumes that all the questions have the same weight. But it is logical to suppose that this is not always the case. Secondly, it is expected that the distances between the neighboring answer options are the same for each item. However, if we are dealing with variables that are measured on an ordinal scale this condition may not be fulfilled. Therefore, we need an appropriate tool allows us to construct an index that overcomes the limitations are mentioned above. One such a tool is a graded response model; it is designed to work with variables that are measured on an ordinal scale. First of all, it is found out is there appropriate to construct an additive index for GSR-5 items. After building the single factor CFA model (confirmatory factor analysis model) with the same factor loadings for each question it turned out that this model does not show an acceptable fit to the data. Thus, the calculation of the additive index for GSR-5 items is not appropriate. Therefore, there is a need for an alternative model. Since GSR-5 items are measured on an ordinal scale a unidimensional graded response model (GRM model) is applied. It shows a good fit to the data. With the GRM model it is possible to build the index which takes into account different weights of the questions and distances between the answer options for each item. The index is constructed with the graded response model shows more variability than the additive index. In addition the graded response model (in order to facilitate interpretation) allows us to represent values of the latent variable as the additive index values. This is the advantage of the graded response model compared to confirmatory factor analysis models. The latter can also be used as tools for constructing additive indices but they do not provide the transformation of latent variables from the one scale to another.

Keywords: graded response model, GSR-5, latent variable, confirmatory factor analysis, additive index