Таким чином, застосування виведення на основі аналогії є ефективним як при наявності представницької бази прецедентів, так і при невеликій кількості відомих прецедентів.

Експериментальні дослідження показали необхідність застосування різних критеріїв існування аналогії між моделлю і прототипом, а також різних методів виведення залежно від особливостей множини прототипів. В подальших дослідженнях планується зосередити увагу на тому, які властивості множини прототипів і яким чином впливають на якість результатів розв'язання задачі прогнозування властивостей.

1. A.Aamodt, E.Plaza Case-based reasoning: foundational issues, methodological variations and System approaches, AI Communications 7(1), 1994, pp.39-59 2. Уемов А.И. Аналогия в практике научного исследования, М.:Наука, 1970. – 264с. 3. Кнут Д.Е. Искусство программирования для ЭВМ, М., – 1977, Т.2 4. Величко В.Ю., Москалькова Н.М. Розв'язування задачі прогнозування властивостей для структурно-атрибутивних моделей за допомогою виведення за аналогією// Искусственный интеллект – 1999 – №2. – С.378-385 5. Величко В.Ю., Москалькова Н.М. Подход к решению задачи комплектации на основе вывода по аналогии// Управляющие системы и машины, – 1999 – №4. С.62-65 6. Гладун В.П., Величко В.Ю., Киселева Н.Н., Москалькова Н.М. Вывод гипотез о составе и свойствах объектов на основе аналогии// Искусственный интеллект, – 2000 – №1. С.44-52

УДК 681.3

О.М. Верес
НУ «Львівська політехніка»,
кафедра "Інформаційні системи та мережі"

РЕПРЕЗЕНТАТИВНІСТЬ КРИТЕРІАЛЬНИХ ТЕСТОВИХ НОРМ СЛАБОСТРУКТУРОВАНОЇ ЗАДАЧІ УКЛАДАННЯ РОЗКЛАДУ

The principles of normalization of a scale of test dancing parties are represented. The operations on analysis of allocation of test dancing parties, construction of test norms and check them representative.

Подані принципи нормалізації показників оптимізаційних критеріїв, множина яких формується з використанням тест-опитування.

Після визначення множини запитань тесту виникає проблема психометрії тестів – питання про тестові норми [1]. Дві основні передумови традиційної західної тестології: питання застосування статистичних норм (квантилів розподілу балів) як діагностичної норми і питання зведення всіх емпіричних розподілів до нормальної моделі.

Статистична природа тестових шкал

Типічний вимірювальний тест у психодіагностиці – це послідовність коротких завдань або пунктів, що дає, в результаті виконання її досліджуваними, послідовність початкових значень, яка потім приводиться до однозначної кількісної інтерпретації.

Сумарний бал по тесту підраховується за допомогою ключа: ключ встановлює числове початкове значення по кожному пункту. Вихідне по окремому завданню піддається впливу не тільки з боку фактора, що вимірюється – здібності або риси досліджуваного, але й сторонніх шумових факторів, які є іррелевантними стосовно задачі вимірювання. Основний прийом, що дозволяє ліквідувати спотворюючий вплив факторів на результат (сумарний бал), є балансування цього впливу за допомогою повторення. При цьому фактично передбачається, що повторення забезпечує рандомізацію випадкового фактора, в результаті чого при додаванні початкових значень позитивні і негативні ефекти випадкових факторів взаємопоглинаються [2].

В оптимальному тесті набір і послідовність завдань організується таким чином, щоб збільшити частку постійної компоненти і зменшити частку випадкової у величині сумарного бала.

Для того, щоб оцінити ефективність, диференційну цінність всієї процедури вимірювання, необхідно співвіднести розміри помилки вимірювання з розмірами розсіву сумарних балів. В термінах статистики мова йде про порівняння "істинної" дисперсії розподілу сумарних балів з дисперсією "похибки". Тому аналіз розподілу необхідний не тільки при використанні статистичних норм, але й у випадку абсолютних і критеріальних норм. Як відомо, випадковий розподіл сумарних балів має зручну графічну ілюстрацію у вигляді кривих розподілу: гістограми і акумуляти. Очевидно, що диференційна спроможність тесту зводиться до нуля, якщо криві, що ілюструють "істинну" дисперсію і дисперсію "похибки", збігаються [3].

Роль упосередкованих еталонів в психометрії виконують самі тести: в тому розумінні, в якому складність задач можна розглядати як величину, прямо пропорційно сполучену зі здібністю (чим складніша задача, тим вищим повинен бути рівень здібностей, необхідних для її розв'язання). Ні складність, ні вагомість пунктів тесту неможливо виявити інакше, ніж за допомогою проведення тесту.

Отже, коли як єдиний еталон вимірювань психодіагностика розглядає сам тест, то як міра вимірюваної властивості виступає місцезнаходження бала на кривій розподілу. Використовується відсоткова шкала. Як універсальна міра, придатна для різних (за своїм якісним напрямом і кількістю пунктів) тестів, використовується "відсоткова міра". *Процентиль – відсоток досліджуваних з вибірки стандартизації, які отримали рівний або нижчий бал, ніж бал даного досліджуваного.* Таким чином, як джерело даної міри виступає нормативна вибірка (вибірка стандартизації), на якій побудовано нормативний розподіл тестових балів. На відсоткових шкалах грунтуються всі традиційні шкали, що використовуються в тестології (Т–очки *MMPI*, бали *IQ*, стени 16*PF* і інші).

Підкреслимо, що з точки зору теорії вимірювань відсоткові шкали належать до порядкових шкал. Для того, щоб будувати на базі таких шкал кількісний прогноз, необхідно покращити рівень вимірювання [4]. Перехід до шкал інтервалів проводять або на бази емпіричного розподілу, або на базі довільної моделі теоретичного розподілу.

Процентильна нормалізація шкали

Опишемо ряд процедур, які широко використовуються для штучної нормалізації.

1.Нормалізація пунктів. Ключ для даного пункту коректується на базі нормальної моделі. Якщо серед нормативної вибірки завдання виконали тільки 16% досліджуваних, то даному пункту на інтервальній шкалі "складності" (при умові апріорного прийняття нормальної моделі з параметрами M=0 і σ=1) відповідає значення +1. Якщо виконало 75% обстежуваних, то бал пункту на сигма–шкалі дорівнює –0,67. В результаті додавання балів по пунктах, відкоректованих нормалізацією, сумарні бали краще наближаються до нормального розподілу.

2.Нормалізація розподілу сумарних балів (або інтервальна нормалізація). В цьому випадку за таблицею нормального розподілу (нормального інтеграла) проводиться перехід від процентильної шкали до сигма–шкали: використовується функція, обернена інтегральній, – від ординати проводиться перехід до абсциси нормального розподілу.

Наведемо приклад інтервальної нормалізації в табл.1. Нехай рядок ***X*** містить початкові бали (не нормалізовані) по тесту, отримані простим підрахунком правильних відповідей. В рядку ***P*** – частоти, з якими зустрічаються початкові бали у вибірці з 62 обстежуваних. В рядку ***F*** – кумулятивні частоти: $\boldsymbol{F_i} = \sum_{j=1}^{i} \boldsymbol{P_{ji}}$. В стрічці $\boldsymbol{F}^*$ – кумулятивні бали: $\boldsymbol{F_i^*} = \boldsymbol{F_i} - \frac{1}{2}\boldsymbol{P_i}$. В рядку *PR* – процентильні ранги: $PR_i=F_i^*\cdot 100/n$. В рядку σ даються нормалізовані бали, які отримали із відповідних процентильних рангів по таблицях (див. табл.1).

Складність, з якою зустрічаються початківці при використанні інтервальної нормалізації, полягає в тому, що звичайні статистичні таблиці не пристосовані для психометрії: необхідно шукати значення процентильного рангу всередині таблиці, а відповідну сигма–оцінку – з краю. Для полегшення орієнтації наведемо фрагмент таблиці відповідностей *PR*, σ і стенів (табл.2).

Таблиця 1

X	3	4	5	6	7	8	9	10	
P	2	18	13	8	10	6	4	1	n=62
F	2	20	33	41	51	57	61	62	
F^*	1	11	26.5	37	46	54	59	61.5	∑=100
PR	1.6	17.7	42.7	59.7	74.2	87.1	95.2	99.2	M=0
σ	–2.1	–0.9	–0.2	0.2	0.6	1.1	1.7	2.4	σ=1

Таблиця 2

PR	99	95	90	85	80	75	70	65	50	55	
σ	2,33	1,64	1,28	1,04	0,84	0,68	0,52	0,39	0,25	0,13	
стен	10	10	9	8	8	7	6,5	6,5	6	6	
PR	50	45	40	35	30	25	20	15	10	5	1
σ	0,0	-0,13	-0,25	-0,39	-0,52	-0,68	-0,84	-1,04	-1,28	-1,64	-2,33
стен	5,5	5	5	4,5	4	4	3	3	2	1	1

В результаті використання процедури нормалізації дослідник отримує для нормативної вибірки таблицю переведення початковий балів в нормалізовані бали.

Оскільки нормальний розподіл описується лише двома параметрами – середнім М (міра положення) і середнім квадратичним (або стандартним) відхиленням σ (мірою розсіювання), то діагностичні норми у випадку нормалізованих шкал описуються в одиницях відхилення від середнього по виборці.

Перехід до нормального розподілу створює дуже зручні умови для кількісних операцій з діагностичною шкалою: як і зі шкалою інтервалів, з нею можна виконувати операції лінійного перетворення (множення і додавання), можна описувати діагностичні норми в компактному вигляді (в одиницях відхилень), можна використовувати лінійний коефіцієнт кореляції Пірсона, критерії для перевірки статистичних гіпотез, побудовані при використанні нормального розподілу, тобто весь аппарат традиційної “гауссівської” статистики (основаної на гауссівському нормальному розподілі).

У традиційній психометрії нормальний розподіл виступає як інструментальне поняття, що полегшує оперування данними.

Підрахунок параметрів і оцінка типу розподілу

Для опису розподілу вибірок, як правило, використовуються відомі параметри:
Середнє арифметичне:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{m}\sum p_j y_j \quad , \qquad (1)$$

де x_i – бал *i*–го обстежуваного;

y_j – значення *j*–го по порядку зростання бала;

p_j – частота *j*–го, що зустріли, бала;

n – кількість обстежуваних у вибірці (об'єм);

m – кількість градацій шкали (кількість балів).

Середнє квадратичне (стандартне) відхилення:

$$S = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} \approx \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n-1}} \quad , \qquad (2)$$

де $\sum x^2$ – сума квадратів тестових балів для n обстежуваних.

Асиметрія:

$$As = \frac{1}{S^3}(0 - 3C^2 x + 2x^3) \quad , \qquad (3)$$

де $\bar{x}$ – середнє арифметичне;

S – стандартне відхилення;

θ – середнє кубічне: $\theta = \sqrt[3]{\frac{1}{n}\sum x^3}$;

C – середнє квадратичне: $C = \sqrt{\frac{1}{n}\sum x^2}$.

Ексцес:

$$Ex = \frac{1}{S^4}(Q^4 - 4\theta^3\bar{x} + 6C^2\bar{x}^2 - 3\bar{x}^4) - 3 \quad , \qquad (4)$$

де Q – середнє значення четвертої степені: $Q = \sqrt[4]{\frac{1}{n}\sum x^4}$.

Стандартна помилка середнього арифметичного (математичне сподівання) оцінюється за формулою:

$$S_m = \frac{S}{\sqrt{n}} \quad . \qquad (5)$$

На основі помилки математичного сподівання будуються довірчі інтервали: $(\bar{x} - 2S_m, \bar{x} + 2S_m)$.

Асиметрія і ексцес нормального розподілу повинні дорівнювати нулю. Якщо вони суттєво відхиляються від нуля (хоча б один з двох параметрів), то це означає анормальність отриманого емпіричного розподілу.

Перевірку значимості асиметрії можна провести на основі загальної нерівності Чебишова:

$$|As| \le \sqrt{\frac{S_a}{1-p}} \quad , \qquad (6)$$

де S_a – дисперсія емпіричної оцінки асиметрії:

$$S_a = \frac{6(n-1)}{(n+1)(n+3)} \quad , \qquad (7)$$

p – рівень значимості або ймовірність помилки першого роду: помилки в тому, що буде зроблено висновок про незначимість асиметрії у присутності значимої асиметрії (у формулу підставляють стандартні p=0,05 або p=0,01 і перевіряють виконання нерівності).

Подібним чином оцінюють значимість ексцесу:

$$|Es| \le \sqrt{\frac{S_e}{1-p}} \quad , \qquad (8)$$

де S_e – емпірична дисперсія оцінки ексцесу, визначається за формулою:

$$S_e = \frac{24(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \quad , \qquad (9)$$

Гіпотези про відсутність асиметрії і ексцесу приймаються з ймовірністю помилки P (нехтовно мала), якщо виконуються нерівності (6) і (8).

Простіший метод перевірки нормальності емпіричного розподілу базується на універсальному критерії Колмогорова. Для кожного тестового бала y_j (для кожного інтервалу рівнозначності при дискретизації безперервної хронометричної шкали) обчислюється величина D_j – модуль відхилення емпіричної і теоретичної інтегральної функції розподілу:

$$D_j = \left|F(y_j) - U(z_j)\right| , \qquad (10)$$

де F – емпірична інтегральна функція (значення кумуляти в даній точці y_j); U – теоретична інтегральна , взята з таблиць. Серед D_j знаходять максимальне значення D_{max} і величину $\lambda_e = D_{max}\sqrt{n}$ порівнюють з табличним значенням λ_t критерієм Колмогорова.

Якщо перевірка співвідношення емпіричного і нормального розподілу дає позитивні результати, то це означає, що отриманий розподіл можна розглядати як стійкий – репрезентативне стосовно до генеральної сукупності і, отже, на його основі можна визначити репрезентативні тестові норми. Якщо перевірка не виявить нормальності на необхідному рівні, то це означає, що або вибірка мала і нерепрезентативна до популяції, або вимірювана властивість і структура тесту (спосіб підрахунку) взагалі не дають нормального розподілу.

В принципі аж ніяк не обов’язково всі нормативні розподіли зводити до пуассонормальних. Можна також успішно користуватися добре розробленими моделями гамма–розподілу, пуассонівського розподілу і т.п. Критерій Колмогорова дозволяє оцінити близькість емпіричного розподілу до довільного теоретичного розподілу.

Наявність значимої додатної асиметрії свідчить про те, що в системі факторів, які детермінують значення вимірюваного показника, переважають фактори, що діють в одному напряму – в бік збільшення значень.

Стандартизація шкали

У психометриці слід розрізняти дві форми стандартизації. Під стандартизацією тесту розуміють передусім стандартизацію самої процедури проведення інструкцій, бланків, способу реєстрації, умов і т.п. Без стандартизації тесту неможливо отримати нормальний розподіл тестових балів і, очевидно, тестових норм.

Під стандартизацією шкали розуміють лінійне перетворення масштабу нормальної (або штучно нормалізованої) шкали. В загальному випадку формула стандартизації має такий вигляд:

$$z_i = \sigma \frac{x_i - \overline{X}}{S} + M \quad , \qquad (11)$$

де x_i – вихідний бал за "сирою" шкалою, для якої доведена нормальність розподілу;

X – середнє арифметичне по "сирому" розподілу;

S – "сире" стандартне відхилення;

M – математичне сподівання за вибраною стандартною шкалою;

σ – стандартне відхилення за стандартною шкалою.

Якщо попередньо проводилася штучна нормалізація інтервалів шкали, то формула спрощується:

$$z'_j = \sigma \cdot z_j + M \quad . \qquad (12)$$

Перевірка стійкості розподілу

Загальна логіка перевірки стійкості розподілу базується на індуктивному міркуванні: якщо "половинний" (отриманий на половині вибірки) розподіл добре моделює конфігурацію цілого розподілу, то можна припустити, що і цілий розподіл буде також добре моделювати розподіл генеральної сукупності.

Найпростіший його варіант може бути зведеним до отримання таблиць переведення "сирих" очків в нормалізовану шкалу за даними всієї вибірки, потім застосування цих таблиць для кожного обстежуваного з половини вибірки: якщо розподіл нормалізованих балів з половини вибірки добре наближається до нормального, то це означає, що задані таблиці нормалізації тестових норм визначені стійко. Близкість до нормального розподілу перевіряється за допомогою критерію Колмогорова.

Статистично більш коректний метод перевірки однорідності двох розподілів, отриманих при поділі вибірки на рівні частини, знову ж таки пов'язаний із застосуванням критерію Колмогорова. Для цього з табличним значенням порівнюється величина [5]:

$$K_e = max\left| F_{j1} - F_{j2} \right| \cdot \sqrt{n/4} \quad , \qquad (13)$$

де F_{j1} – кумулятивная відносна частота для j–го інтервалу шкали для першої половини вибірки;

F_{j2} – та сама частота для другої половини;

n – чисельність повної вибірки;

K_e – емпіричне значення статистики Колмогорова.

Отже, апріорна передумова нормальності розподілу тестових балів базується скоріше на принципах операційної зручності, ніж на теоретичній необхідності. Психометрично коректні процедури отримання стійких тестових норм можливі за допомогою спеціальних методів непараметричної статистики (критерій "хі–квадрат" і т.п.) для розподілу довільної форми.

Репрезентативність критеріальних тестів

В тестах за критерієм як реальний еталон застосовується критерій, заради якого створюється тест – цільовий критерій. Особливе значення такий підхід має в тих сферах, де високі результати можуть дати вузькоспеціалізовані діагностичні методики, націлені на дуже конкретні і вузькі критерії. Така ситуація є в навчанні: тестування скероване на отримання інформації про рівень засвоєння відповідних знань, здібностей і навичок, повинно точно відображати рівень засвоєння цих навичок і тим самим давати надійний прогноз ефективності конкретної професійної діяльності, яка потребує засвоєння цих навичок [6].

Розглянемо операційну систему шкалування, що використовується при створенні критеріального тесту. Нехай маємо деякий критерій C_i, для прогнозування якого дослідник створює тест X. Для простоти приймемо C як дихотомічну змінну з двома значеннями – 1 і 0. $C_i = 1$ означає, що i–й суб'єкт досяг критерію (попав у "високу" групу за критерієм), $C_i = 0$ означає, що i–й суб'єкт не досяг критерію (попав у "низьку" групу). Психодіагност застосовує на нормативній вибірці тест X, і в результаті кожен індивід отримає тестовий бал X_i. Після того як для кожного індивіда з вибірки стає відомим значення C, психодіагност групує індивідів у порядку зростання бала X і для кожної поділки вихідної шкали сирих тестових балів підраховує емпіричну ймовірність P попадання у "високу" групу за критерієм. Тепер можна сформулювати основну вимогу до критеріального тесту: лінія регресії повинна бути монотонною функцією C від X. Іншими словами, ні для одного більш високого значення X ймовірність P не повинна бути меншою, ніж для якого–небудь менш високого значення X. Якщо ця умова виконується, то відкривається можливість для критеріального шкалування сирих балів X.

Операції з аналізу розподілу тестових балів, побудови тестових норм і перевірки їх репрезентативності

Коротко опишемо дії, які послідовно повинен провести дослідник при побудові тестових норм:

Сформувати вибірку стандартизації (випадкову або відібрану за яким небудь параметром) з тієї популяції, на якій передбачається застосовувати тест. Провести на кожному обстежуваному з вибірки тест в стислі терміни (щоб ліквідувати іррелевантний розсів, викликаний зовнішніми подіями, які відбулися за час дослідження).

Згрупувати початкові бали з урахуванням вибраного інтервалу квантування (інтервал рівнозначності). Інтервал визначається величиною W / m, де $W = x_{max} - x_{min}$ – розмах; m – кількість інтервалів рівнозначності (градацій шкали).

Побудувати розподіл частот тестових балів (для заданих інтервалів рівнозначності) у вигляді таблиці і у вигляді відповідних графіків гістограми і кумуляти.

Розрахувати середнє і стандартне відхилення, а також асиметрію і ексцес. Перевірити гіпотези про значення асиметрії і ексцесу. Порівняти результати перевірки з візуальним аналізом кривих розподілу.

Перевірити нормальність одного з розподілів при допомозі критерія Колмогорова (при $n < 200$ за допомогою більш потужних критеріїв) або провести процентильну нормалізацію з переведенням до стандартної шкали, а також лінійну стандартизацію та порівняти їх результати (з точністю до цілих значень стандартних очків).

Якщо збігу не буде – нормальність відкидається, тоді перевірити стійкість розподілу шляхом розщеплення вибірки на дві випадкові половини. При збігові нормалізованих балів для половини і для цілої вибірки вважати нормалізовану шкалу стійкою.

Перевірити однорідність розподілу стосовно множини варіантів заданої популяційного ознаки (стать, професія та ін.) за допомогою критерію Колмогорова. Побудувати у спільних координатах графіки гістограми і кумуляти для повної і випадкової вибірок. При значних відмінностях розбити вибірку на різнорідні підвибірки.

Побудувати таблиці процентильних і нормалізованих тестових норм (для кожного інтервалу рівнозначності "сирого" бала). При наявності різнорідних підвибірок для кожної підвибірки повинна бути своя таблиця.

Визначити критичні точки (верхню і нижню) для довірчих інтервалів (на рівні $\boldsymbol{P} <$ 0,01) з урахуванням стандартної помилки при знаходженні середнього значення.

Обговорити конфігурацію отриманих розподілів з урахуванням механізму, яким передбачається розв'язувати той чи інший тест.

У випадку негативних результатів – відсутності стійких норм для шкали з заданим числом градацій (з заданою точністю прогнозу критеріальної діяльності) – провести обстеження більш широкої вибірки або відмовитися від плану використання даного тесту.

Як вже відзначалося, загальне розсіяння (дисперсію) результатів проведених вимірювань можна представити як результат додавання двох джерел різноманітності: самої властивості і нестабільності процедури вимірювання, що обумовлює наявність помилки вимірювання. Це представлення відображено у формулі, яка описує надійність тесту у вигляді відношення істинної дисперсії до дисперсії емпірично зареєстрованих балів:

$$\alpha = \frac{\boldsymbol{S}_{\boldsymbol{T}}^2}{\boldsymbol{S}_{\boldsymbol{x}}^2} \quad .$$

Висновок

Отже, для формування множини значень показників критеріїв оптимізації необхідно провести нормалізацію тестових балів, побудувати тестові норми і перевірити їх репрезентативність. Наведена математична модель є основою для побудови алгоритмів та програм нормалізації тестових балів і формування початкових масивів показників.

1. Общая психодиагностика: Под ред. А.А.Бодалева, В.В.Столина. М., 1987. -300с. 2. Готтсданкер Р. Основы психологического эксперимента. М., 1982. -С.74–75 3. Кимбл Г. Как правильно пользоваться статистикой. М., 1982.–С.55–70. 4. Клигер С.А. и др. Шкалирование при сборе и анализе социологической информации. М., 1978.-С.75–80. 5. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. М., 1982. -278с. 6. Гуревич К.М., Лубовский В.И. Предисловие // Анастази А. Психологическое тестирование. М., 1982.-С.5–14.