

Автор:

Тупигін Катерина Василівна,
студент 51 ОБмз групи

Науковий керівник:

Кухар Людмила Олександрівна,
старший викладач кафедри комп'ютерної інженерії

СТАТИСТИЧНІ МЕТОДИ ВИЗНАЧЕННЯ ВАЛІДНОСТІ ТА НАДІЙНОСТІ ТЕСТУ

Анотація. Доповідь присвячена основним критеріям якості методу вимірювання рівня знань, найважливішими з яких є валідність та надійність. Об'єктом дослідження є вимірювання рівня знань учнів, студентів. Предметом – методи їх вимірювання. Під час дослідження були використані такі методи, як: абстрагування, аналіз і синтез тощо. Результатами дослідження є групування методів, які дають змогу оцінити тест, підкреслюючи його основні недоліки.

Ключові слова: тест, тестування, валідність, об'єктивність, надійність, точність.

Вступ. Вибір того чи іншого методу вимірювання рівня знань ґрунтується на критеріях, які визначають якість засобу вимірювання. Найважливішими критеріями якості є валідність, об'єктивність, надійність і точність [4, 109].

Постановка задачі. Доцільним буде розглянути кожен із них.

Мета роботи. Метою дослідження є розгляд даних критеріїв, щоб наголосити на тому, які саме критерії дають змогу оцінити тест, виявити його недоліки та переваги.

Основна частина. Фундаментальним є критерій *валідності* [4, 109]. Поняття валідності та валідизації інструментів вимірювання є одним чи не з найскладніших для розуміння та застосування в теорії освітніх вимірювань [2, 67].

Валідність – комплексна характеристика тесту, що включає відомості про область досліджуваних явищ і репрезентативність діагностичної процедури яка для неї використовується (Л.Ф. Бурлачук). Інакше кажучи, тест називається валідним, якщо він вимірює те, для вимірювання чого він призначений (П. Клайн). А. Анастасі вважає валідність тесту як поняття, яке вказує нам, що тест вимірює і наскільки гарно він це робить.

Поняття валідність використовується для загальної характеристики тесту в аспекті відповідності здобутих результатів меті та вимогам оцінювання [4, 109].

Деякі дослідники (Л. Крокер, Дж. Алгіна), вважають, що існує три основних типи валідності:

- *Змістова валідність* для ситуацій, в яких користувач тесту хоче з тестової оцінки тестованого отримати висновки для більшої області завдань, подібних до тих, які безпосередньо використовуються в тесті.
- *Валідність, пов'язана з критерієм*, для ситуацій, в яких користувач тесту хоче з тестової оцінки тестованого отримати висновок відносно виконання індивідуумом діяльності за деякої поведінковою зміною практичного значення.
- *Конструктна валідність* для ситуацій, в яких «ні критерій, ні генеральна сукупність змісту не вважаються повністю адекватними для визначення якості, призначеного для вимірювання...» (Cronbach and Meehl, 1955), але користувач тесту хоче отримати висновок з тестової оцінки результатів виконання тесту, які можуть бути згруповані за ознакою вимірювання певного психологічного конструкту.

Автори посібника [1, 14] пропонують аналізувати якість тестів і тестових завдань на основі визначення їх головних характеристик і параметрів, коли загальне поняття валідності щодо процесу вимірювання і оцінювання рівня знань диференціюється за функціональною ознакою таким чином: валідність методу (валідність змісту, відповідності, прогнозу); валідність тесту (валідність тестових завдань, процедури тестування, процедури оцінювання).

Валідність змісту і відповідності може бути кількісно визначено через *коефіцієнт валідності*. *Коефіцієнт валідності* дорівнює коефіцієнту кореляції між результатами, одержаними різними методами за однакових умов, і показує, наскільки збігаються результати вимірювань [1, 14].

Ще одним критерієм якості методу вимірювання є *об'єктивність*. Він вказує, наскільки мінімізований вплив суб'єктивних факторів. Критерії валідності й об'єктивності фактично є первинними, оскільки порушення одного з них призводить до порушення критеріїв надійності та точності, які, незважаючи на їх важливість, виступають як похідні [1, 15].

Поняття *надійності*, разом із поняттям валідності, є фундаментальною характеристикою тесту, без якої тестування не може вважатися вимірюванням. Разом з тим, порівняно з валідністю, надійність є більш технічною характеристикою, яка стосується насамперед проблеми точності вимірювання [2, 84].

Надійність методу вимірювання – це міра стійкості результатів, що впливає на точність, з якою можна виміряти ту чи іншу конкретну ознаку. Перевірка надійності методу стосується насамперед відновлення результатів при повторних вимірюваннях. [4, 113]

Ступінь надійності методу визначається за допомогою *коефіцієнта надійності* (R), який дорівнює коефіцієнту кореляції між результатами, отриманими однаковим методом за однакових умов, і показує, наскільки збігаються результати вимірювань [1,15].

Коефіцієнт надійності змінюється від 0 до 1 і розраховується з використанням певних методик. Найчастіше для його розрахунку використовують рівняння Спірмена-Брауна (*формула для коригування надійності Спірмена-Брауна*), що дає змогу визначити коефіцієнт надійності тесту після застосування методу внутрішньої послідовності. Тест поділяють навпіл, розраховують надійність, наприклад, парних завдань. Знаючи надійність половини тесту, за формулою Спірмена-Брауна можна визначити надійність усього тесту.

Формула має вигляд:

$$r = \frac{2r_{xy}}{1 + r_{xy}}, \quad (1)$$

де r_{xy} – коефіцієнт кореляції між двома половинами тесту. Коефіцієнт, обчислений за формулою Спірмена-Брауна показує, наскільки підвищується надійність тесту, якщо збільшити його обсяг.

Коефіцієнт α -Кронбаха вказує на надійність за внутрішньою узгодженістю вимірюваних характеристик.

$$\alpha = \frac{k}{k-1} \left(1 - \sum_{i=1}^k \frac{\sigma_i^2}{\sigma^2} \right), \quad (2)$$

де σ_i^2 – дисперсії для k окремих позицій,

σ^2 – дисперсія суми всіх позицій (порівнюється розсіювання кожного елемента із загальним розсіюванням всієї шкали [4, 117].

Якщо розсіювання результатів тесту менше, ніж розсіювання результатів для кожного окремого питання, то кожне окреме питання спрямоване на дослідження однієї й тієї ж ознаки, властивості або явища. А отже, сукупність запитань тесту дають результат, який можна вважати істинним.

Коефіцієнт надійності тісно пов'язаний з помилкою (похибкою) вимірювання, яка визначається *точністю* методу. *Точність методу* визначає мінімальну або систематичну похибку, з якою можна провести вимірювання даним методом [1, 16].

Як ми вже зазначали, надійність тесту стосується точності вимірювання. Натомість, валідність пов'язана з самою природою характеристик, які вимірюються. Невалідний тест може виявитися цілком надійним [2, 91].

Висновки. Таким чином, найважливішими критеріями якості тесту є валідність, об'єктивність, надійність і точність. Саме вони дають змогу оцінити якість тесту, вказуючи на основні його сильні та слабкі сторони.

Список використаних джерел

1. Булах І. Є./ Створюємо якісний тест: Навч. посіб./ І.Є. Булах, М.Р. Мруга — К.: Майстер клас, — 2006 — 160 с.
2. Ковальчук Ю.О. Теорія освітніх вимірювань. – Ніжин: Видавець ПП Лисенко М.М., 2012. – 200 с.
3. Крокер Л. Введение в классическую и современную теорию тестов: учебник/ Л. Крокер, Дж. Алгина; пер. с англ. Н.Н. Найденовой, В.Н. Симкина, М.Б. Чельшковой; под общ. ред. В.И. Звонникова, М.Б. Чельшковой. – М:Логос,2012.- 668с.
4. Кухар Л.О. / Конструювання тестів. Курс лекцій: навч. посіб. /Л.О. Кухар, В.П. Сергієнко. – Луцьк, 2010. – 182 с.