УДК 37.091

АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТОВОГО КОНТРОЛЮ НА ОСНОВІ КЛАСИЧНОЇ ТЕОРІЇ

Макарчук Олег, Резінов Артем

Науковий керівник: канд.-ф.-м. наук, доцент Макарчук О.П.

Центральноукраїнський державний університет імені Володимира Винниченка, м. Кропивницький, Україна

Анотація. Робота присвячена проблемі обробки результатів тестового контролю на основі засад класичної теорії. Акцент здійснюється на аналіз завдань, що мають дихотомічну структуру. Розглянуто такі характеристики як кореляційні показники, валідність тесту, надійність тесту, дискримінативність тесту.

Ключові слова: педагогічний тест, дихотомічні завдання, валідність тесту, надійність тесту, дискримінативність тесту, кореляція.

Analysis of test control results based on classical theory

O Makarchuk, A Rezinov

Scientific supervisor: Candidate of Physics and Mathematics Science Makarchuk O.P.

Volodymyr Vynnychenko Ukrainian State University, Kropyvnystky, Ukraine

Annotation. The work is devoted to the problem of processing test control results based on the principles of classical theory. Emphasis is placed on the analysis of tasks that have a dichotomous structure. Such characteristics as correlation indicators, test validity, test reliability, and test discrimination were considered.

Key words: pedagogical test, dichotomous tasks, test validity, test reliability, test discrimination, correlation.

1. Постановка проблеми.

В останні роки особливо у сфері освіти широкого розповсюдження та розвитку набули нові системи контролю та оцінки навчальних досягнень, які орієнтуються на широке використання педагогічних тестових форм.

Актуальність тестових методів контролю засвоєння та якості знань, в більшій мірі визначається їхніми можливостями технологічного характеру, що передбачає можливість отримання об'єктивної інформації про якість знань та підготовки суб'єктів навчання.

Сучасний період розвитку освіти, зокрема вищої технічної, характеризується всебічним поширенням тестового підходу до контролю знань і потребує застосування сучасних статистичних математичних методів до аналізу якості як загалом тестів, так і окремих тестових завдань. Безумовно статистичний підхід, який обраховує кількісні характеристики успіхів є на сьогоднішній час найпродуктивнішим в розумінні визначення рівня якості засвоєних знань.

Об'єктом дослідження даної роботи є математичні методи оцінки якості засвоєних знань при тестовому контролі.

Предмет дослідження статистичні показники оцінки якості знань при тестовому контролі.

Мета дослідження: здійснити оцінку якості засвоєних знань при тестовому контролі на основі методів класичної теорії.

Результати дослідження є актуальними та можуть бути використані при викладанні спец курсу з статистики та освітніх технологій.

2. Статистичні показники оцінки якості засвоєних знань.

Відповідний тестовий комплекс включав в себе 10 тестових питань, які мають закриту форму тобто існує лише одна правильна відповідь яка оцінюється в 1 бал, в протилежному випадку неправильна відповідь оцінюється в 0 балів. В тестуванні взяло участь десять учнів. Питання тестового контролю відповідали підготовці до ЗНО (НМТ) з математики.

Первинні кількісно-статистичні показники представлені в таблиці 1.

Таблиця 1. Кількісно-статистичні показники результати апробації тестових завдань.

Учень\завдання	1	2	3	4	5	6	7	8	9	10	X_i
1	0	1	1	0	0	1	1	0	0	0	4
2	1	1	1	1	1	0	1	0	1	0	7
3	1	0	1	0	1	1	0	1	0	1	6
4	0	0	1	0	0	1	1	1	1	0	5
5	0	1	0	1	1	1	1	1	1	1	8
6	0	1	0	0	0	1	1	1	0	0	4
7	1	0	1	0	1	0	1	0	1	0	5
8	1	1	1	1	0	1	1	1	0	0	7
9	1	0	0	1	1	1	1	0	0	0	5
10	0	0	1	0	0	1	1	1	1	0	5
R_j	5	5	7	4	5	8	9	6	5	2	
W_j	5	5	3	6	5	2	1	4	5	8	
p_j	0,5	0,5	0,7	0,4	0,5	0,8	0,9	0,6	0,5	0,2	
q_j	0,5	0,5	0,3	0,6	0,5	0,2	0,1	0,4	0,5	0,8	

Відомо, що мода це таке значення певної вибірки, яке має найбільше частотне значення. У нашому випадку значення моди дорівнює: $M_0 = 5$.

Необхідно також відзначити, що медіаною являється таке значення, яке ділить упорядкований вибірковий набір статистичних так, що одна половина статистичних значень являється меншою за медіану, а відповідно інша більшою. У нашому випадку $M_e = 5$.

Маємо наступне значення середнього вибіркового моменту: $\bar{X} = 5{,}6$.

В нашому випадку ми маємо таке значення середнього квадратичного відхилення: $S_X^2 = 1{,}64$.

Асиметрія та ексцес дорівнює відповідно $A = 0{,}492$ та $E = -0{,}973$.

Таблиця 2. Кореляційна матриця результатів апробації тесту.

r_{XY}	1	2	3	4	5	6	7	8	9	10
1	1	-0,2	0,218218	0,408248	0,6	-0,5	-0,33333	-0,40825	-0,2	-1,38778E-17
2	-0,2	1	-0,21822	0,408248	-0,2	0	0,333333	-4,5E-17	-0,2	-4,16334E-17
3	0	0	1	-0,35635	-0,21822	-0,32733	-0,21822	-0,08909	0,21821789	-0,21821789
4	0,5	-0,5	-0,10206	1	0,408248	-0,10206	0,272166	-0,16667	0	0,102062073
5	0,6	-0,2	-0,40825	0,5	1	-0,5	-0,33333	-0,40825	0,2	0,5
6	-0,21822	0,218218	-0,53452	-0,21822	-0,21822	1	-0,16667	0,612372	-0,5	0,25
7	-0,33333	0,333333	0,408248	0,166667	-0,33333	-0,21822	1	-0,27217	0,333333333	-0,666666667
8	-0,65465	0,218218	-0,08909	-0,76376	-0,65465	0,52381	-0,21822	1	0	0,40824829
9	-0,2	-0,2	0,408248	-1,4E-17	0,2	-0,65465	0,333333	-0,21822	1	1,38778E-17
10	-1,4E-17	-1,4E-17	-0,61237	-0,25	0,5	0,327327	-0,66667	0,327327	1,38778E-17	1
$\sum r$	0,493795	0,669769	0,070204	0,894834	1,083826	-0,45112	0,002396	0,377066	0,851551224	1,375425806

Обрахуємо дихотомічний коефіцієнт кореляції за формулою [2, с.102]:

$$\varphi_{mk} = \frac{p_{mk} - p_m p_k}{\sqrt{p_m q_m p_k q_k}},$$

Таблиця 3. Кореляційна матриця, побудована з коефіцієнтів кореляції для дихотомічних даних.

φ_{mk}	1	2	3	4	5	6	7	8	9	10
1	1	-0,21	0	-0,52	0,16	-0,21	-0,33	-0,65	-0,34	-0,887
2	-0,21	1	-0,52	0,408248	-0,2	0	0,33	--,006	-0,2	-0,009
3	0	0	1	-0,35635	-0,21	-0,32	-0,21	-0,08	0,218	0,33
4	-0,52	-0,45	-0,10	1	0,40	-0,10	0,27	-0, 67	0	0,786
5	0,16	-0,2	-0,405	0,5	1	-0,5	-0,33	-0, 825	0,2	0,5
6	-0,21	0,21	-0,53	-0,21822	-0,21	1	-0,16	0, 372	-0,5	0,25
7	-0, 33	0, 933	0,40	0,16	-0,33	-0,21	1	-0,27	0,33	-0,666
8	-0, 65	0,28	-0,08	-0,76	0,65	0,52	-0,21	1	0	0, 4829
9	-0,52	-0,2	0,40	-0,78	0,2	-0,65	0,33	-0,212	1	-0,234
10	-0,34	-0,023	-0, 37	-0,25	0,5	0,327	-0,66	0, 327	-0,887	1
$\sum \varphi_{mk}$	0,43	0, 769	0,703	0,48	1,08	-0,45	0,0043	0,706	--0,85	1, 425

Обрахуємо бісеріальні та точково-бісеріальні коефіцієнти використовуючи методологію [2, с.105].

Таблиця 4. Бісеріальні та точково-бісеріальні коефіцієнти.

j	1	2	3	4	5	6	7	8	9	10
$(r_{pbis})_j$	0,19	0,49	0,71	0,88	0,68	0,81	0,55	0,801	0,07	0,06
$(r_{bis})_j$	0,17	0,38	0,82	0,86	0,7	0,83	0,59	0,91	0,17	0,23

Трудність для j-го завдання обчислюється за формулою:

$$p_j = \frac{R_j}{N},$$

де p_j рівне долі правильних відповідей на відповідне j-те завдання, R_j дорівнює кількості учасників тестування, які правильно дали відповідь на j-те завдання.

З одного боку класична теорія аналізу результатів апробації тестовго контролю трактує те, що трудність зростає з ростом кількості учнів. Однак з іншого боку, таке твердження суперечить дещо здоровому глузду.

Є наступний варіант виправлення цього негаразду. Тут потрібно говорити, про коректну збалансованість тесту.

Таким чином, для коректно збалансованого тесту, необхідне існування декількох важких завдань з величиною p_j достатньо близькою до 0, також необхідно, щоб деяка кількість завдань була простою з величиною p_j достатньо близькою до значення 1. Було б не погано, щоб інші завдання мали значення p_j в діапазоні від 0,3 до 0,7. Зрозуміло, що класичним фактором коректності було, б наближення результатів до моделі нормального розподілу.

Розглянемо дисперсійну характеристику для кожного із завдань тестового комплексу, яка обраховується за наступною формулою

$$\sigma_j = p_j q_j$$

де відповідно $q_j = 1 - p_j$.

Таблиця 5. Трудність та дисперсія тестових завдань.

j	1	2	3	4	5	6	7	8	9	10
p_j	85	37	44	24	36	23	77	92	17	25
σ_j	0,11	0,23	0,54	0,22	0,44	0,09	0,27	0,002	0,3	0,24

Проаналізуємо ще одне важливе поняття класичної тестової теорії поняття дискримінативності завдань. Дискримінативністю є здатність завдань тестової сукупності поділяти учнів на гірших та кращих. Є цілий ряд характеристик, які використовуються для аналізу поняття дискримінативності. Одним з них є показник розпізнавальної здатності.

Показник розпізнавальної здатності вираховується за співвідношенням:

$$D_j = (p_1)_j - (p_0)_j$$

де D_j – індекс дискримінативності для j-го завдання тесту; $(p_1)_j$ – доля учнів, що правильно виконали j-те завдання у підмножині випробовуваних (підгрупі з 27% кращих учнів) за результатами виконання тесту; $(p_0)_j$ – частка учнів, які правильно виконали j-те завдання у підмножині випробовуваних (у підгрупі з 27% гірших учнів за результатами виконання тесту).

Таблиця 6. Показники розпізнавальної здатності завдань.

j	1	2	3	4	5	6	7	8	9	10
$(p_1)_j$	1	0,8	1	1	0,7	0,8	1	0,6	0,9	0,8

$(p_0)_j$	0,8	0,3	0,2	0,3	0	0,3	0,4	0	0,4	0,5
D_j	0,2	0,5	0,8	0,7	0,7	0,5	0,6	0,6	0,5	0,3

В результаті обробки результатів досліджень можливо зробити наступні висновки:

1) доволі низька дисперсія індивідуальних балів учасників тестування свідчить про доволі слабку диференціацію учасників в групі. Значення асиметрії $0{,}492$ є додатним, тому тест можна вважати достатньо важким. Значення ексцесу $-0{,}973$ є від'ємним, тобто крива розподілу є унімодальною та плосковершинною.

2) в кореляційній матриці, які містять значенння кореляцій наявні як від'ємні сумарні значення, які є небажаними (наприклад для завдання з номером 6), так і значення доволі близькі до одиниці, яке свідчить про, те що деякі завдання для даної групи необхідно рекомендовано виклячити (такими є завдання з номерами 8 та 4), найбільш близьким до нуля є значення кореляції яке відповідає завданню з номером 3, це свідчить про деяку незначну корельованість цього завдання з всім тестом та можливість його можливого виключення.

3) в кореляційній матриці кореляцій існують як від'ємні сумарні значення, що є небажаними (наприклад для завдання з номером 9), так і наближені одиничні значення, які свідчить про, те що певні завдання для даної групи потрібно виклячити (завдання з номерами 8 та 5), найбільш близьке значення нуль сумарної кореляції відповідає завданню з номером три, що підтверджує про незначну корельованість цього завдання з усім тестом і також про можливість його виключення з тесту.

4) значення точково-бісеріальної кореляції тестових завдань свідчать про невисоку корельованість завдань з номерами 1 та 9, їх структуру та зміст необхідно проаналізувати. Завдання з номерами 2, 1, 9, 10 мають кореляцію у

прийнятному інтервалі. Завдання з номерами 3,4,8 мають високий рівень кореляції, тому їх зміст та форму слід проаналізувати. Ці завдання слід залишати тільки у критерійно-орієнтованих також вузько спеціальних тестах.

5) максимальний внесок у загальну дисперсію тесту робить завдання з номером 5, а також завдання з номерами 5, 3, 7. Такі завдання знаходяться у центральній частині відповідного ряду, вони є самими вдалими для нормативно-орієнтованих тестів. Дисперсія крайніх завдань 1 та 2, а також завдання з номерами 9 та 10 є невисокою. Такі завдання рекомендується включати у невеликій кількості у збалансований тест.

6) маємо завдання з номерами 3, 4, 5 які володіють найкращою роздільною здатністю; завдання з номерами 2, 6, 8, 9 функціонують доволі задовільно; завдання з номером 10 потребує деякої корекції, завдання 1 необхідно вилучити з тесту або ж перепроектувати.

Список літератури

1. Аванесов В. С. Вопросы методологии педагогических измерений. Педагогические измерения. – 2005. –№ 1. – С. 3-27.
2. Вимірювання в освіті: Підручник / За редакцією О.В. Авраменко.– Кіровоград: Лисенко В.Ф., 2011. – 360 с.
3. Вихрущ В.О. Методологія та методика наукового дослідження. – Тернопіль: Тайп, 2005. – 224 с.
4. Agresti, A. (1996). An introduction to categorical data analysis. New York: John Wiley & Sons.
5. Hougaard, P. (2000). Analysis of multivariate survival data. New York: Springer.