

СТАТИСТИЧНІ ПОКАЗНИКИ ВГАДУВАННЯ ЗАКРИТИХ ТЕСТІВ

При тестуванні найбільше розповсюдження мають тести на основі завдань закритого типу. У них кожне завдання має жорсткий перелік відповідей на твердження, сформульоване у змістовній частині завдання. Можливість вгадати відповідь з запропонованих варіантів при розгляді окремого завдання може привести до набору значної кількості слушних відповідей при розгляді усього тесту навіть тоді, коли тестуємий не має ніякої уяви з питань, які розглядаються.

При відсутності знань і оголошенні результатів після завершення тестування відповідь на будь-яке завдання буде незалежною, тобто не буде залежати від інших. Результат «аналізу» (опиту) дворівневий – чи вгадав, чи ні. При таких умовах ймовірність появи m слушних відповідей при N незалежних випробуваннях (опитах), у кожному з яких ймовірність вгадування дорівнює p , описується біноміальним законом

$$D_{N,p} = C_N^m p^m q^{N-m}, \quad (1)$$

де C_N^m – сполучення з N елементів за m ;

$$m = 0, 1, 2 \dots N;$$

$$q = 1 - p - \text{ймовірність неправильної відповіді при розгляді окремого завдання.}$$

Ймовірність вгадування при розгляді окремого завдання закритого типу дорівнює

$$\delta = \frac{n_+}{i}, \quad (2)$$

де n_+ – кількість слушних відповідей у завданні.

За запропонованою нижче методикою можна також аналізувати тести на співвідношення та на послідовність. У першому – необхідно встановити зв'язок між двома наборами понять, а у другому – вказати правильну послідовність виконання якихось взаємодій. У них відсутні варіанти відповідей, тому їх формально не відносять до закритих тестів. Однак, у зв'язку з тим, що у відповідях не можливо вийти за набір наданих понять, то можна вважати, що вони являють собою замаскований варіант закритих тестів. Ймовірність слушного угадування одного завдання у таких тестах дорівнює:

$$p = \frac{n_+}{n!}. \quad (3)$$

На рисунку 1 надані розподіли при 15 завданнях у тесті (білеті) при різних кількостях варіантів відповідей, одна з яких вірна ($n_+ = 1$). При великих n (та незначних N) помітна асиметрія і ексцес розподілу. Причому при великих n має місце значна ймовірність нульової кількості вгадувань, що і відповідає нульовим знанням тестуємого. Однак максимальна ймовірність все ж таки зміщена до більшого числа m .

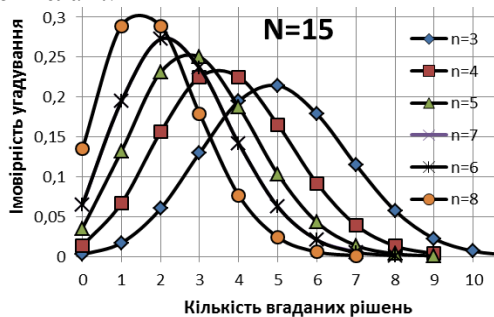


Рисунок 1

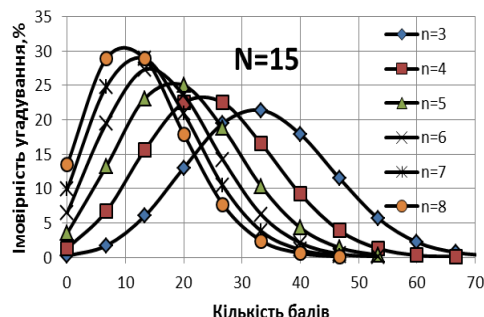


Рисунок 2

Показники біноміального розподілу визначаються добутком ($N * p$). Середнє значення за осями абсцис та ординат

$$m_{cp} = N * p, \quad (3)$$

а середнє квадратичне відхилення

$$\sigma_m = \sqrt{N * p * q}. \quad (4)$$

При зростанні ($N * p$) розподіл наближається до нормального. Критерієм можливості застосування усіх надбань нормального закону є ($N * p$) > 10.

Для отримання оцінки результати тестування зазвичай переводять у 100 бальну шкалу з наступним перетворенням у національні оцінки та оцінки **ESTD**. Перетворення часто проводять за критеріями, наданими у таблиці. В ній градації обірвані на оцінці «задовільно», бо аналізується можливість отримання позитивної оцінки при повній відсутності знань.

| Бали | Національна оцінка | Оцінка ESTD |
|-------------|--------------------|--------------------|
| < 25 | Не задовільно | F |
| 25 ... < 50 | | FX |
| ≥ 50 | Задовільно | E |

Для простого однорідного тесту (усі завдання білету мають одну правильну відповідь при однаковій кількості варіантів) природно однаково оцінювати кожну правильну відповідь. Тому при 100 бальній шкалі за одну слушну відповідь буде нараховане $K_o = 100/N$ балів. Якщо на заставі перерахунку змінити

масштаб горизонтальної осі, а вертикальну визначити у процентах, то розподіли рисунку 1 будуть мати вигляд, наданий на рисунку 2.

Аналізуючи з точки зору критеріїв таблиці розподіли з різними показниками можна зробити такі висновки:

- при деяких параметрах має місце значна вірогідність отримання позитивної оцінки;
- з іншої сторони спостерігається суттєва можливість набрати бали діапазону **F**, що, як раз, і відповідає істинним знанням тестуемого (студент, який отримав оцінки цього діапазону, повинен пройти повторний курс навчання);
- під час тестування можна набрати значну кількість балів за рахунок угадування, що дозволяє отримати позитивну оцінку при незначних додаткових знаннях, які виявляються при розгляді окремих завдань.

Результат аналізу вказує на те, що статистичні властивості тестів покращуються з ростом кількості завдань і кількості варіантів відповідей. Вплив цих показників різний і може бути встановлений шляхом математичного аналізу виразу (1), що, мабуть, має теоретичний інтерес. На практиці необхідні уяви можна отримати у результаті розрахунку розподілів у діапазоні параметрів, які використовуються.

Найгірші властивості мають тести з кількістю відповідей $n = 3$. Навіть при тесті з 50 завданнями більше одного проценту тестуємих у середньому можуть отримати задовільну оцінку. Ще більш незадовільними властивостями характеризуються тести з двома варіантами відповідей. Вірогідність отримання позитивної оцінки у таких тестах коливається біля 50%.

Незадовільні властивості тестів з малим числом завдань. Наприклад, при 5 завданнях не менш одного процента мають шанс отримання позитивної оцінки навіть при 7..8 варіантами відповідей. А запропонувати таку кількість рівноцінних відповідей у кожному завданні важко. Однак на практиці тести з малим числом завдань не виняткові.

Враховуючи складність розробки тестів зі значною кількістю завдань та кількістю варіантів відповідей найбільш оптимальними, на наш погляд, є тести з $N = 15...30$ і $n = 4, 5$. Для кінцевих випробувань (заліки, екзамени) бажано підвищити кількість завдань до приблизно 50.

Ще одним недоліком закритих тестів є можливість набрати значну кількість балів за рахунок вгадування. Зазвичай їх буде недостатньо для отримання задовільної оцінки. Однак додаткову кількість балів тестуємих може набрати за рахунок знань учбового матеріалу тільки за декількома завданнями. Цих додаткових балів буде значно менше, ніж у тому випадку, коли тест не буде мати визначеного недоліку. Тому результати реального тестування, де має місце як вгадування, так і «знання», повинні бути скоректовані. Однією з можливих формул подібної корекції результатів тестування відповідно до 100 бальної системи оцінювання може бути вираз:

$$X_{кор} = X - \frac{100 - X}{n}, \quad (6)$$

X – кількість балів, отриманих за результатами тестування.

На практиці при тестуванні використовують змішані тести, окремі завдання яких мають різну кількість відповідей чи більш ніж одну правильну відповідь. У результаті машинного моделювання було показано, що середнє значення і середнє квадратичне відхилення визначаються зваженою оцінкою, у якій вагові коефіцієнти залежать від відносного числа окремих складових змішанного тесту.