

ВГАДУВАННЯ ПРИ ТЕСТУВАННІ (СТАТИСТИЧНИЙ АСПЕКТ)

*Ймовірність вгадування відповідей при тестуванні за допомогою закритих тестів описується біноміальним законом. Розглянуті пов'язані з цим властивості простих закритих однотипних тестів з вибором однієї відповіді при показниках і характеристиках тесту, які найчастіше використовуються на практиці. У результаті перерахування кількості вгаданих завдань у 100-бальні шкалу проаналізований їх вплив на рівень оцінки за національною шкалою та шкалою **ESTD**. Показана велика вірогідність отримання задовільної оцінки при повній відсутності знань та вірогідність набирання значної кількості «вгаданих» балів, що дозволяє отримати позитивну оцінку за рахунок отримання невисоких додаткових балів в результаті рішення завдань на підставі знань. Пропонується вводити корекцію набраних балів.*

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

Ключові слова: тестування; закриті тести; ймовірність вгадування; результати тестування; оцінки.

Постановка проблеми. Аналіз досліджень і публікацій. Тестування як метод оцінки рівня знань, отриманих в результаті навчання, в останній час набуло значного розповсюдження як у середній, так і у вищій школі. Згідно з теорією тестування, розробка якісних тестів триває декілька років [1–8]. У цей час необхідно здійснити не одну тисячу попередніх випробувань, обробок, аналізів результатів тощо, у результаті яких встановлюються не лише такі статистичні параметри, як середнє значення, середньоквадратичне відхилення, але й коефіцієнти надійності, валідності, індекс диференціовальної здатності та ін. Це зазначено навіть у рекомендаціях Міністерства освіти України [9]. Різноманітність спеціальностей призводить до значного різноманіття матеріалу, що викладається, навіть у таких загальних дисциплінах, як фізика, хімія, математика (вища математика). Їх наповнення повинно бути адаптованим до потреб спеціальності та спеціалізації. Ще більше різнопланового матеріалу у спеціальних дисциплінах, які повинні відобразити також особливості регіонів тощо. Це робить неможливим розробку загальнодержавних тестів для окремих дисциплін, які викладаються у вищій школі. На наш погляд, лише у деяких випадках можлива розробка науково обґрунтованих тестів сукупною групою у складі фахівців текстологів та викладачів з предметних дисциплін. Однак доступність і масовість інформаційних технологій робить навіть такі тести недієздатними після декількох випробувань.

Застосування більш простих тестів дозволяє отримати не лише кінцеву (екзамен, залік, перевірка залишкових знань тощо), але і проміжну (блокові та модульні контрольні) інформацію про хід та якість навчання. На практиці подібне тестування також стимулюється різноманітними контролюючими структурами, які вимагають його проведення при різних періодичних та вибіркових перевірках. Розробити таку кількість та різноманіття тестів на основі використання теорії тестування просто неможливе. Це призводить до того, що більшість тестів, які використовуються на практиці, формуються на основі завдань закритого типу, часто без врахування деяких негативних властивостей.

Метою роботи є розгляд однієї з таких властивостей, що пов'язана з можливістю (ймовірністю) вгадування вірної відповіді при повній відсутності знань з питання, що розглядається.

Викладення основного матеріалу. Уточнимо деякі положення. Під тестом будемо розуміти сукупність з декількох окремих завдань, які об'єднані в одному документі, наприклад, у білеті. Кількість завдань у будь-якому білеті однакова і дорівнює N . Кожне завдання є завданням закритого типу, тобто має жорсткий набір відповідей на твердження, сформульоване у змістовній частині завдання. Будемо аналізувати тести, в яких всі завдання мають *одну* вірну відповідь. Кількість наданих відповідей (n) в усіх завданнях тесту однакова. Позначимо такі тести простішими однотипними.

Будемо вважати, що людина, яка тестується (учень, студент), не обтяжена знаннями з жодного із запитань, розглянутих у будь-якому завданні тесту, а просто вказує на одну з наданих відповідей. Таку «обробку» одного завдання, згідно з термінологією статистики, будемо вважати одним опитом. Результат опиту дворівневий – вгадав чи ні. Ймовірність такого вгадування при розгляді окремого завдання дорівнює:

$$\delta = \frac{1}{i}. \quad (1)$$

Ймовірність вгадування вірної відповіді в окремому завданні призводить до ймовірності вгадування більшої їх кількості при розгляді усього тесту, де тестуемий зробить N спроб (опитів) вгадування. Спроби незалежні, бо результат оголошується після завершення тестування.

Задача визначення ймовірності вгадування вірних відповідей при розгляді тесту збігається з класичною задачею математичної статистики [12]: визначення ймовірності вгадування однієї з n граней грального кубика при N незалежних опитах (розіграшах).

Ймовірність появи m вірних відповідей за N незалежних опитів (випробувань), у кожному з яких ймовірність вгадування дорівнює p , описується біноміальним законом:

$$D_{N,p} = C_N^m p^m q^{N-m}, \quad (2)$$

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

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

$q = 1 - p$ – ймовірність невірної відповіді при розгляді окремого завдання.

На рисунку 1 наведені діаграми ймовірності вгадування m відповідей у тесті з десяти завдань, у кожному з яких надано п'ять варіантів відповідей. Відображення у формі діаграми застосовано тому, що біноміальний розподіл уривчастий, дискретний. Однак у наступних рисунках, де будуть наведені відразу декілька залежностей при різних параметрах розподілу, дискретні точки, які як раз відповідають дискретним значенням біноміального закону, будуть з'єднані плавною кривою. На наш погляд, це дозволяє зручніше розглядати наведені на рисунках результати. Побудова графіків та розрахунок ймовірності біноміального розподілу виконано програмою *Microsoft Excel* пакету *Microsoft Office*.

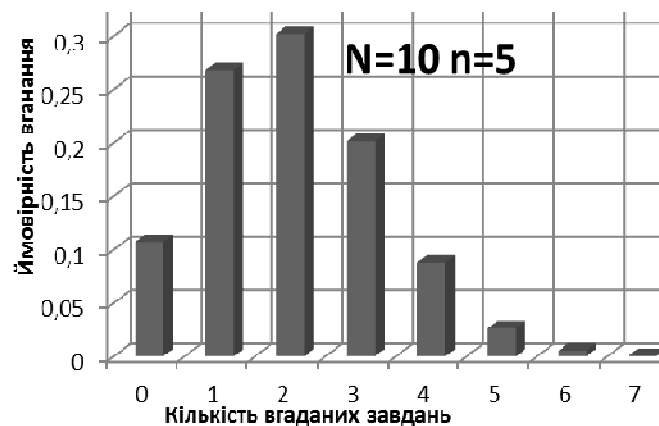


Рис. 1

Біноміальний розподіл асиметричний, крім випадку $p = 0,5$, який в подальшому не розглядається. Причому, при $p < 0,5$ асиметрія позитивна, при $p > 0,5$ – негативна.

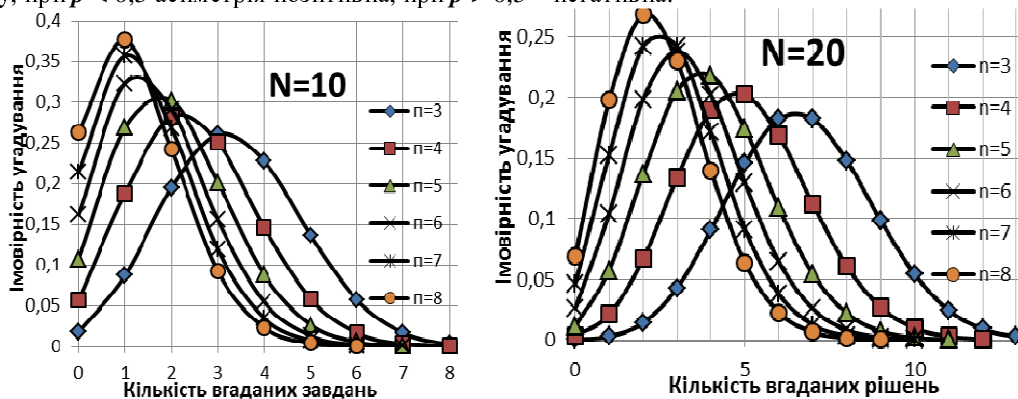


Рис. 2

На рисунку 2 надані розподіли при загальній кількості завдань у білеті 10 та 20 при різній кількості варіантів відповідей. При великих n (та незначних N) помітна асиметрія та ексцес розподілу. Причому при великих n має місце значна ймовірність відсутності вгадувань, що відповідає нульовим знанням тестованого. Однак максимальна ймовірність все ж таки зміщена до більшого числа m .

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. У [2] критерієм можливості користування усіма надбаннями нормального закону є $(N * p) > 10$. На наш погляд, для отримання якісних показників нормальним розподілом можна користуватися і за менш жорстких умов.

Тестування найчастіше використовують з метою отримання оцінки, для чого результати необхідно представити у бальній шкалі, за яку зазвичай використовують 100-бальну. В подальшому її перетворюють у національні оцінки та оцінки ESTD. Перетворення часто проводять за критеріями, наданими у таблиці 1, в якій градації оцінок обірані на оцінці «задовільно», бо аналізується можливість отримання позитивної чи негативної оцінки при повній відсутності знань.

Таблиця 1

Бал	Національна оцінка	Оцінка ESTD
<25	Не задовільно	<i>F</i>
25...<50		<i>FX</i>
≥50	Задовільно	<i>E</i>

Для однорідного тесту природно *однаково* оцінювати кожен вірну відповідь. Тому за 100-бальною шкалою за одну вірну відповідь буде нараховане $K_6 = 100/N$ балів. Якщо на заставі перерахунку змінити масштаб горизонтальної осі, а вертикальну визначити у процентах, то розподіли рисунку 2 будуть мати вигляд, наданий на рисунку 3.

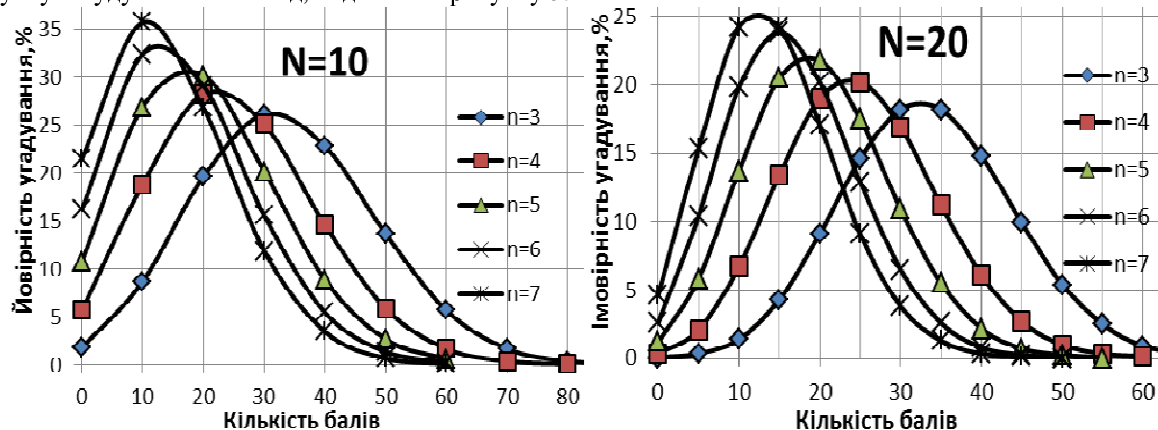


Рис. 3

Аналізуючи, з точки зору критеріїв таблиці 1, ці та подібні розподіли з іншими показниками, можна зробити такі висновки:

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

Результати аналізу розподілів з різними параметрами, з точки зору перших двох критеріїв, зведені у таблицю 2. У ній наведені ймовірності (у процентах з округленням):

- отримання задовільної оцінки – P_{50} (50 балів та вище);
- потрапляння у зону F (S_F).

Визначення значення P_{50} проводилося шляхом складання усіх ймовірностей при отриманні 50 та більшій кількості балів [12]. При визначенні S_F підсумовувались ймовірності у діапазоні від 0 до 25 балів.

Таблиця 2

N	$N =$	3	4	5	6	7	8
5	$S_{25}, \%$	46,2	63,3	73,8	80,4	84,9	88
	$P_{50}, \%$	21,5	10,5	6,1	3,8	2,6	1,9
7	$S_{25}, \%$	26,4	44,6	57,8	67,1	73,7	78,6
	$P_{50}, \%$	17,6	7	3,5	1,8	1,1	0,7
10	$S_{25}, \%$	30,1	52,7	67,9	77,7	84	88,2
	$P_{50}, \%$	21,7	8,1	3,4	1,8	0,8	
12	$S_{25}, \%$	18,3	39,2	56	68	76,2	82
	$P_{50}, \%$	18	5,8	2,1	1		
15	$S_{25}, \%$	20,9	46,1	64,8	76,9	84,5	89,2
	$P_{50}, \%$	8,8	1,7	0,5			
20	$S_{25}, \%$	15,1	41,5	63	76,8	85,4	90,5
	$P_{50}, \%$	9,2	1,4	0,3			
25	$S_{25}, \%$	22,1	56,2	78,1	89	94,4	96,9
	$P_{50}, \%$	4,1	0,4				
30	$S_{25}, \%$	16,7	51,7	76,1	88,6	94,5	97,3
	$P_{50}, \%$	4,5	0,4				
50	$S_{25}, \%$	10,3	51	81,3	93,7	97,8	99,3
	$P_{50}, \%$	1,2					

Результат, який маємо, аналізуючи таблицю, вказує на те, що статистичні властивості тестів покращуються з ростом кількості завдань і варіантів відповідей. Вплив цих показників різний і може бути встановлений шляхом математичного аналізу виразу (2), що, мабуть, має теоретичний інтерес. На практиці необхідні уяви можна отримати у результаті розрахунку розподілів у діапазоні параметрів, які використовуються.

Найгірші властивості з розглянутих мають тести з кількістю відповідей $n = 3$. Навіть при тесті з 50 завданнями більше одного проценту тестованих 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 %. Тому у даній роботі вони не розглядаються.

Незадовільні властивості тестів з малою кількістю завдань. Наприклад, при 5 завданнях не менш одного процента мають шанс отримання позитивної оцінки навіть при 7–8 варіантах відповідей. А запропонувати таку кількість рівноцінних відповідей у кожному завданні важко. Однак на практиці тести з малою кількістю завдань не виняткові. Зазвичай їх поява обумовлена, на наш погляд, двома причинами. По-перше, недостатньою проробкою навчального матеріалу розробником тесту. По-друге, недостатніми термінами на розробку тесту, що найбільш характерно при раптових перевірках.

Враховуючи складність розробки тестів зі значною кількістю завдань та 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 50.

Як було вказано вище, 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 тестований може набрати за рахунок знань навчального матеріалу лише за декількома завданнями. Цих додаткових балів буде значно менше, ніж у тому випадку, коли тест не матиме визначеного недоліку чи тест буде сформований на основі завдань з малою ймовірністю вгадування окремого завдання.

З урахуванням коефіцієнта перерахунку кількості вірно вгаданих завдань у бали, тестований у середньому набирає наступну кількість балів:

$$K_{cp \bar{a}} = m_{cp} * \frac{100}{N} = N * p * \frac{100}{N} = \frac{100}{N}, \quad (5)$$

яка, як видно, визначається лише кількістю варіантів відповідей, наданих у завданнях. Відповідні значення зведені до таблиці 3.

Таблиця 3

<i>n</i>	3	4	5	6	7	8
$K_{cp \bar{b}}$	33,3	25	20	16,7	14,3	12,5

Тобто якість тесту зростає зі збільшенням кількості варіантів відповідей. Так при трьох варіантах відповіді можна отримати задовільний бал (табл. 1) і лише при набраних 17 балів «за рахунок» знань. При семи варіантах кількість необхідних балів, отриманих за рахунок знань, зростає приблизно до 35. При відсутності недоліку закритих тестів необхідно було б набрати («за рахунок» знань) не менше 50 балів. Тому результати реального тестування, де має місце як вгадування, так і «знання», повинні бути скоректовані. У [4] розглядається застосування подібної корекції. Відповідно до 100-бальної системи оцінювання формула запропонованої там корекції має вигляд:

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

X – кількість балів, отриманих за результатами тестування.

При використанні (6) при низьких значеннях X результат може стати від'ємним. Тому корекція повинна починатися при значеннях $X > 100/(n + 1)$. До цього $X_{кор} = 0$. При незначних X використання формули практично ліквідує бали, набрані за рахунок вгадування. З ростом набраних балів вплив корекції швидко зменшується.

Крім закритих, достатнє розповсюдження мають тести, які умовно можна назвати тестами на співвідношення та на послідовність. У першому випадку необхідно встановити зв'язок між двома наборами понять, у другому – вказати вірну послідовність виконання якихось дій. Відсутні варіанти відповідей, тому їх формально не можна віднести до закритих тестів. Однак, у зв'язку з тим, що у відповідях не можна вийти за набір наданих понять, то вважатимемо, що вони являють собою замаскований варіант закритих тестів. Це дозволяє застосовувати для аналізу їх статистичних властивостей наведену вище методику.

Якщо у обох тестах на співвідношення і на послідовність вказано, що вірним може бути лише одне співвідношення чи послідовність, а інші, наприклад, навіть частково вірні, віднесені до помилкових, то вони принципово не відрізняються від розглянутих вище. З точки зору статистичних властивостей, відмінність буде лише у значенні ймовірності угадування вірного рішення одного завдання p .

У таблиці 4 наведений приклад розрахунку ймовірності встановлення вірного співвідношення між двома наборами понять, котрі умовно позначені літерами **A**; **B**, **B** (перший набір) і **a**; **b**; **v** (другий набір).

Таблиця 4

Перший набір понять	Можливі сполучення другого набору понять					
A	a	a	b	b	v	v
B	b	v	a	v	a	b
B	v	b	v	a	b	a

Лише перше сполучення другого стовпчика буде визначене як вірне. Загальна кількість усіх можливих сполучень другого набору дорівнює кількості можливих переставлень з n елементів, тобто $n!$ (n факторіал). Тому ймовірність вірного угадування одного завдання дорівнює:

$$p = \frac{1}{n!}. \quad (7)$$

Цією ж формулою визначається ймовірність вірного угадування послідовності виконання дій. Порівняння з (1) вказує на меншу ймовірність угадування окремих завдань тестів на співвідношення і на послідовність, крім випадку $n = 2$. У останньому разі, як і для закритого тесту, ймовірність дорівнює 0,5.

При інших значеннях вона стрімко зменшується, що відповідає більш якісним показникам цих тестів, порівняно з простими закритими.

Висновки. Можливість вгадати відповідь з запропонованих варіантів при розгляді окремого завдання закритого типу призводить до набору значної кількості вірних відповідей при розгляді усього тесту навіть тоді, коли тестований не має жодної уяви щодо питань, які розглядаються. Відповідь на будь-яке завдання буде незалежною, тобто не буде залежати від інших. За таких умов ймовірність вгадування описується біноміальним законом, показниками якого будуть кількість завдань у тесті і кількість варіантів відповіді у кожному завданні.

У роботі проаналізовано ймовірності вгадування у діапазоні параметрів, які часто використовуються у практичному тестуванні. Аналіз виконаний на прикладі простих однотипних тестів (усі завдання мають однакову кількість варіантів відповідей, одна з яких вірна).

Запропонований варіант перерахування кількості вгаданих відповідей у 100-бальну шкалу з наступним переходом (за одним з можливих варіантів трансформування) до оцінок за національною шкалою і шкалою **ESTD**, для визначення яких, зазвичай, проводиться тестування. Проаналізований вплив параметрів тесту на:

- вірогідність потрапляння оцінки у групу **F** (за **ESTD**), яка вказує на необхідність повторного навчання, що відповідає рівню знань тестованого;
- вірогідність отримання задовільної оцінки при повній відсутності знань;
- отримання значної кількості балів, що дозволяє отримати позитивну оцінку за рахунок отримання незначних додаткових балів в результаті рішення завдань на основі знань.

Для зменшення впливу вгадування пропонується коректувати результати реального тестування.

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 50.

Додатково розрахована ймовірність вгадування вірної відповіді для завдань на співвідношення та на послідовність. Це дозволяє використати наведену вище методику для розрахунку і аналізу статистичних властивостей тестів на основі таких завдань. Враховуючи швидке зменшення ймовірності вгадування вірної відповіді для окремого завдання зі збільшенням кількості понять, що використовуються у завданнях, ці типи тестів мають безперечні переваги перед тестами з завданнями закритого типу.

Список використаної літератури:

1. *Аванесов В.С.* Композиция тестовых заданий / *В.С. Аванесов.* – 2-ое изд. – М. : Центр тестирования, 2002. – 240 с.
2. *Аванесов В.С.* Истоки и основные понятия математической теории измерений (Item Response Theory). Статья вторая / *В.С. Аванесов* // Педагогические измерения. – № 3. – 2007. – С. 3–36.
3. *Аванесов В.С.* Проблема объективности педагогических измерений / *В.С. Аванесов* // Педагогические измерения. – № 3. – 2008. – С. 3–40
4. *Чельшикова М.Б.* Теория и практика конструирования педагогических тестов : учеб. пособ. / *М.Б. Чельшикова.* – М. : Логос, 2002. – 432 с
5. *Анастаси А.* Психологическое тестирование / *А.Анастаси.* – Кн. 1. – М. : Педагогика, 1982. – 320 с
6. *Ким В.С.* Анализ результатов тестирования в Rasch Measurement / *В.С. Ким* // Педагогические измерения. – № 4. – 2005. – С. 39–45.
7. *Деменченко О.Г.* Математические основы Rasch Measurement / *О.Г. Деменченко* // Педагогические измерения. – № 1. – 2010.
8. Технология оценивания тестов в зависимости от типа и уровня сложности тестовых заданий на основе интегрированной модели / *М.Бондаренко, В.Семенец, Н.Белоус и др.* // International Book Series «Information Science and Computing». – 2009 [Електронний ресурс]. – Режим доступу : http://www.foibg.com/ibs_isc/ibs-12/ibs-12-p07.pdf.
9. Методичні рекомендації з розроблення складових галузевих стандартів вищої освіти. – К. : Інститут інноваційних технологій і змісту освіти МОН України, 2013. – 89 с.
10. *Міхайленко В.В.* Теорія ймовірностей, математична статистика та випадкові функції. Курс лекцій : навч. посібник / *В.В. Міхайленко.* – Житомир : ЖІТІ, 2003. – 292 с.
11. *Барановський В.В.* Теорія ймовірностей та математична статистика / *В.В. Барановський, Н.В. Барановська, О.К. Лопатін.* – 5-те вид. – К. : Центр навч. літ-ри, 2010 – 424 с.
12. *Калинина В.Н.* Математическая статистика / *В.Н. Калинина, В.Ф. Панкин.* – 2-ое изд. – М. : Высш. шк., 1998. – 336 с.

КУПКІН Євген Савелійович – кандидат технічних наук, доцент кафедри КСУ та А Житомирського державного технологічного університету.

Наукові інтереси:

– моделювання систем управління.

Стаття надійшла до редакції 07.11.2014

Купкін Є.С. Вгадування при тестуванні (статистичний аспект)
Купкин Е.С. Угадывание при тестировании (статистический аспект)
Kupkin E.S. Guessing when testing (statistical aspect)

УДК 303.714

Угадывание при тестировании (статистический аспект) / Е.С. Купкин

Вероятность угадывания ответов при тестировании с помощью закрытых тестов описывается биномиальным законом. Рассмотрены связанные с этим свойства простых закрытых однопольных тестов с выбором одного ответа, при показателях и характеристиках теста, которые часто используются на практике. Результаты угадывания переведены в 100-балльную шкалу, с последующим преобразованием в оценки по национальной шкале и шкале ESTD. Показана большая вероятность получения удовлетворительной оценки при полном отсутствии знаний и набора значительного количества "угаданных" баллов. Последнее позволяет получить положительную оценку за счет незначительных дополнительных баллов в результате решения заданий на основании знаний. Предлагается проводить коррекцию результатов тестирования.

Учитывая сложность разработки тестов со значительным количеством заданий и количеством вариантов ответов наиболее оптимальными, на наш взгляд, являются тесты с 15...30 заданиями и 4, 5 вариантами ответов. Для конечных испытаний (зачеты, экзамены) желательно повысить количество заданий до приблизительно 50.

Ключевые слова: тестирование; закрытые тесты; вероятность угадывания; результаты тестирования; оценки.

УДК 303.714

Guessing when testing (statistical aspect) / E.S. Kupkin

The probability of guessing the answers when tested using the closed test described binomial law. Considered related properties of simple open simple tests with a choice of one answer, in terms of characteristics test most commonly used in practice. The results of guessing are translated in 100 ball scale, with subsequent transformation to the estimations on a national scale and scale of ESTD. Shown large probability of receipt of satisfactory estimation at complete absence of knowledge and probability of set of far "of the guessed" points. The last allows getting a positive estimation due to insignificant additional points as a result of decision of tasks on the basis of knowledge. It is proposed to carry out the correction of test results.

Taking into account complication of development of tests with the far of tasks and amount of variants of answers most optimal, in our view, there are tests from 15 by 30 tasks and 4 by 5 variants of answers. For eventual tests (tests, examinations) it is desirable to promote the amount of tasks to approximately 50.

Keywords: testing; closed tests; probability of guessing; testing results; estimations.