

СТАТИСТИЧЕСКИ АНАЛИЗ ЗА ОЦЕНКА КАЧЕСТВОТО НА ИЗПИТЕН ТЕСТ

Райна Алашка

alraina@abv.bg

*ВТУ „Тодор Каблешков”, ул. „Гео Милев” 158, София 1574,
БЪЛГАРИЯ*

Резюме: Извършен е статистически анализ на реални изпитни данни. Пресметнати са различни измерителни показатели за качеството на отделните задачи и теста като цяло. Дадени са корелационните зависимости между тях и в съвкупност. Направен е дисперсионен анализ за влиянието на задачите, разпределени по основни теми. Правят се изводи за успеваемостта на кандидатите по темите и за качеството на теста за постигане на основните поставени цели.

Ключови думи: тест, изпит, корелационен анализ, дисперсионен анализ, надеждност, образование, статистика.

УВОД

В настоящата статия се прави статистически анализ, основан на класическата теория на тестовите за независима оценка качеството на приемен тест за ВТУ „Тодор Каблешков”, проведен на 28 юли 2009 година. Тестът се състои от 20 задачи с избираем отговор и 10 задачи със свободен отговор. Задачите са независими една от друга, т.е. информацията в една задача не подпомага отговора на друга задача. Към всяка задача с избираем отговор са дадени 4 възможности за отговор, от които точно една е правилният отговор. Максималният бал на теста е 120 точки. За правилен отговор на задачите с избираем отговор се получават – 3 точки, за грешен – 0 точки, за неотбелязан отговор – 1 точка. За задачите със свободен отговор се дават 6 точки при правилен отговор и 0 точки при грешен или неотбелязан отговор. Тестът е решаван от 650 кандидат-студента. Пресметнати са основни показатели за качеството на теста, като трудност на задачите, дискриминация на задачите, различни видове корелации между задачите и

надеждност на теста. Направен е дисперсионен анализ за зависимост на оценките по задачи от това към коя тема те принадлежат.

Условно задачите от теста са разделени в следните теми:

Таблица 1

№	Тема	Задача №	Брой задачи
1	Алгебрични действия. Действия със степени, логаритми и показатели. Прогресии.	1, 6, 7, 8, 9, 10, 14, 22, 24	9
2	Уравнения, неравенства, системи.	2, 3, 4, 5, 21	5
3	Планиметрия.	15, 16, 17, 18, 19, 23, 28	7
4	Стереометрия.	29, 30	2
5	Тригонометрия.	13, 25	2
6	Комбинаторика. Класическа вероятност. Граница на функция. Производна на функция.	11, 12, 20, 26, 17	5
Общо:			30

Целта на теста е да се определи група от кандидати, които имат добри математически познания и владеят работата с математически справочник и те да се разграничат от кандидатите с по-ниски действителни способности. Друга цел на теста е да се ограничи възможността от много на брой различни оценки.

2. АНАЛИЗ НА КАЧЕСТВАТА НА ЗАДАЧИТЕ ОТ ТЕСТА

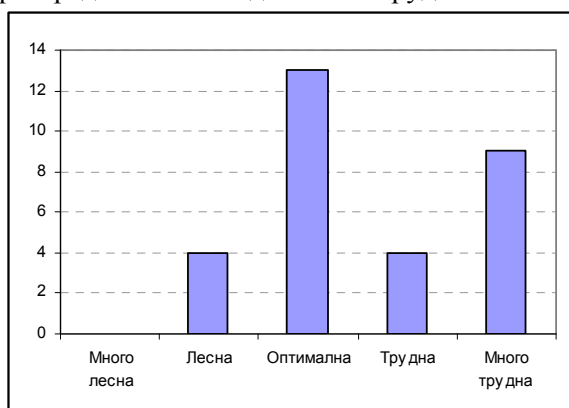
1. Трудност

Коефициентът на трудност е равен на процента на кандидатите, които са решили правилно задачата, изчислен от броя на всички кандидати, които са я решавали. Допустимите стойности на коефициента са между 0 и 100. Броят на възможните отговори на една задача оказва влияние при тълкуване на коефициента на трудност. В таблица 2 е дадено тълкуването на коефициента на трудност за тестови задачи с 4 алтернативни отговора. Оптималната трудност на задача с 4 алтернативни отговора е 62,5.

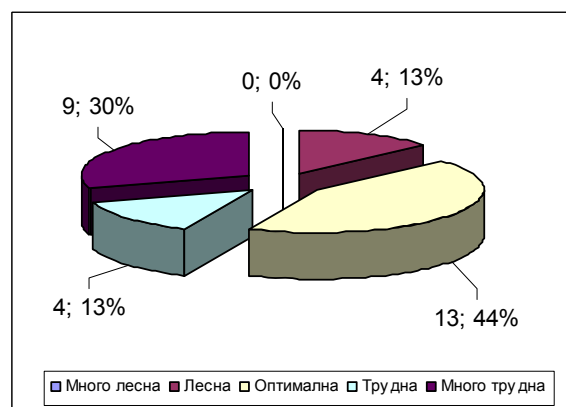
Таблица 2

Вид на задачата	Коефициент на трудност
Много лесна	86 – 100
Лесна	71 – 85
Оптимална	41 – 70
Трудна	26 – 40
Много трудна	0 – 25

На фиг.1 и фиг.2 са представени съответно хистограма на честотното разпределение и кръгова диаграма на процентното разпределение на задачите по трудност.



Фиг.1



Фиг.2

Тестът не съдържа много лесни задачи, тъй като прекалено лесните задачи дават една постоянна добавка към бала на всяко лице

2. Дискриминативна сила

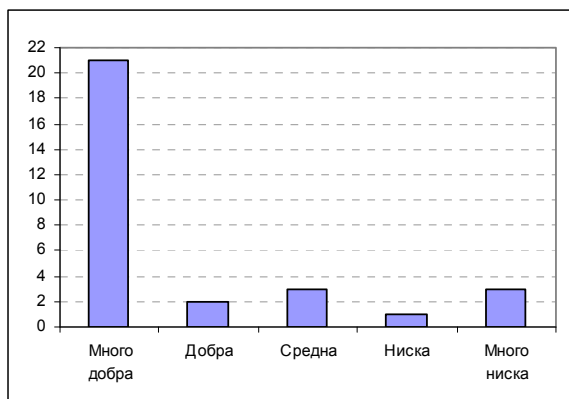
Дискриминацията на задачата е степента, в която тя разграничава кандидатите с високи постижения от тези с ниски постижения. При анализ на дискриминацията на задачите чрез коефициента на дискриминацията се определят две групи от кандидати – „силна група” и „слаба група”. Двете групи се определят, след като кандидатите се подредат по бал и се вземат първите 27% (първите 176 кандидата) от тях, получили най-високи резултати, и последните 27% (последните 176 кандидата) от тях, получили най-ниски резултати на теста.

Коефициентът на дискриминация на една тестова задача е равен на разликата от частта на вярно отговорилите кандидати от силната група и частта на вярно отговорилите кандидати от слабата група. Възможните стойности на коефициента са между – 1 и 1. Колкото коефициентът е по-близо до 1, толкова повече задачата прави по-добро отделяне на лицата, получаващи висок общ бал, от лицата получаващи нисък общ бал. Тълкуването на дискриминацията е дадено в таблица 3.

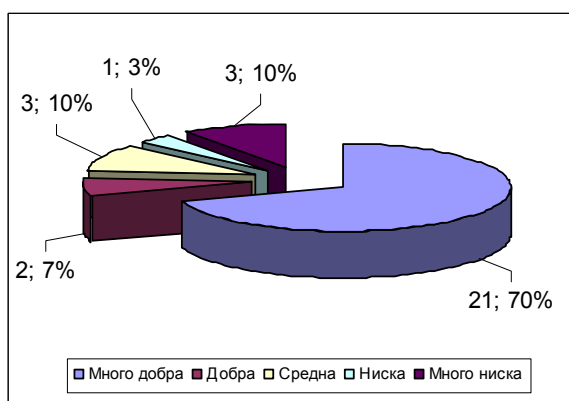
Таблица 3

Вид на задачата	Коефициент на дискриминация
Много добра	0,41 – 1,00
Добра	0,31 – 0,40
Средна	0,21 – 0,30
Ниска	0,11 – 0,20
Много ниска	< 0,10

По-голямата част от задачите в теста (70%) са с много добра дискриминация, т.е. те са отличен разграничител на двете групи. На фиг.3 и фиг.4 са представени съответно хистограма на честотното и кръгова диаграма на процентното разпределение на задачите по дискриминативна сила.



Фиг.3



Фиг.4

3. Корелация между бала на дадена задача и общия бал

Корелацията дава друг начин за оценяване на дискриминацията на тестовите задачи. Тя определя връзката на задачата с останалите задачи. Корелацията не се изчислява от определените силна и слаба група, а се извежда от средния бал на лицата, които са отговорили правилно на задачата, и от средния бал на всички кандидати.

Коефициентът на корелация на Пирсън за i -тата задача се пресмята по формулата:

$$(1) \quad R_i = \frac{\sum (X_{ij} - \bar{X}_i)(Y_j - \bar{Y})}{\sqrt{\sum (X_{ij} - \bar{X}_i)^2 \cdot \sum (Y_j - \bar{Y})^2}}$$

Коефициентът на корелация е по-информативна характеристика от

коефициента на дискриминация. В неговото изчисляване участват всичките тестови балове, докато коефициентът на дискриминация се основава само на част от получените балове.

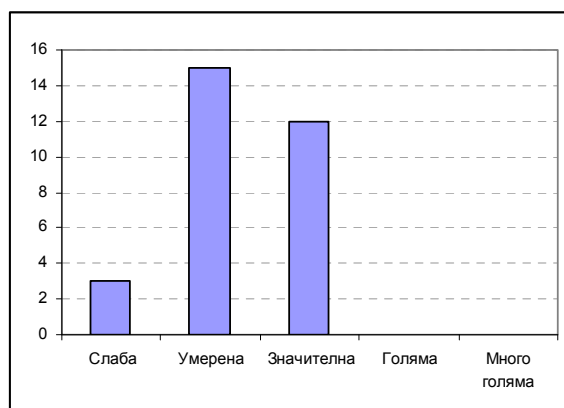
Допустимите стойности на коефициента на корелация са между -1 и 1. Висока положителна корелация показва, че тези кандидати, които са отговорили правилно на задачата, имат по-висок от средния общ бал. В таблица 4 е дадено тълкуването на коефициента на корелация на Пирсън.

Таблица 4

Корелационна зависимост	Коефициент на корелация (R)
Слаба	$0 < R_i \leq 0,3$
Умерена	$0,3 < R_i \leq 0,5$
Значителна	$0,5 < R_i \leq 0,7$
Голяма	$0,7 < R_i \leq 0,9$
Много голяма	$0,9 < R_i \leq 1$

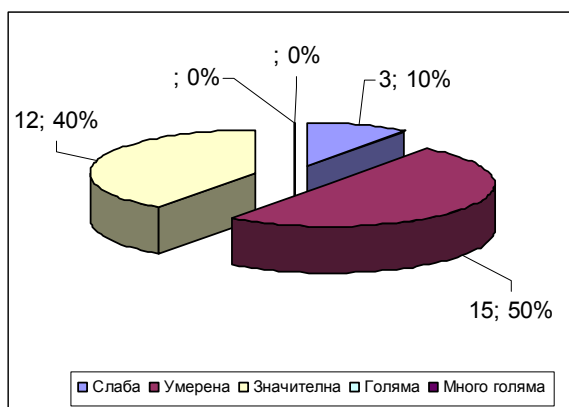
Отрицателната корелация на задачата с общия бал е силен показател за несъвместимост на задачата с теста. Ако има такава задача в теста, тя трябва да се изключи от него или да се подобри (промени). В теста няма задачи с отрицателна корелация.

На фиг.5 е представена с хистограма на честотното разпределение на задачите по корелационната им зависимост с общия бал.



Фиг.5

На фиг.6 са представена кръгова диаграма на процентното разпределение на задачите по корелационната им зависимост с общия бал.



Фиг.6

4. Корелация на дадена задача с всички останали задачи от теста

Пресмята се коефициента на корелация между всеки две задачи по формулата на Пирсън:

$$(2) \quad R_{ij} = \frac{\sum (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{\sqrt{\sum (X_{ik} - \bar{X}_i)^2 \cdot \sum (X_{jk} - \bar{X}_j)^2}}$$

Коефициентите на корелация между всеки две задачи се систематизират в „корелационна матрица“.

Връзката на всяка задача с останалите може да си види от най-високата, най-ниската и средната корелация на задачата с останалите. По-големият брой отрицателни корелации показва, че задачата мери нещо по-различно от това, което мерят другите задачи. В теста няма задачи с отрицателна корелации помежду си, т.е. няма несъвместими задачи.

5. Ефективност на въпросите

Качеството „ефективност“ на въпрос е свързано преди всичко с анализа на дистракторите му. Неподходящо подобрени алтернативи често водят до лоша дискриминация или до нежелана степен на трудност. Няма определен процент от отговори, който да се изисква за всяка алтернатива, но обикновено дистракторите, събрали под 5% от отговорите, не са за предпочитане. В изследвания тест алтернативите са подобрени добре, като са използвани най-често срещаните грешки между изпитваните.

6. Забележка:

За повече информация за всички данни и числовите им характеристики виж [4].

3. НАДЕЖНОСТ НА ТЕСТА

Надеждността на теста е мярка за точността на получените балове. Тя изразява вероятността за получаване на един и същ бал от един кандидат във времето или различни балове от различни кандидати по едно и също време. Надеждността е свързана с квадрата на корелацията между наблюдения и действителния бал. Високата надеждност води до ограничаване грешката при измерване на действителния бал.

1. Корелация между задачите

Като използваме корелационната матрица на задачите от теста (виж [4]) ще пресметнем надеждността на теста.

Средната корелация на всички задачи е:

$$(3) \quad \rho = \frac{1}{435} \sum_{i \neq j} R_{ij} = 0,193$$

(R_{ij} е корелацията между i и j задача).

Надеждността оценяваме по формулата на Спирмън – Браун:

$$(4) \quad r_{3030} = \frac{30 \cdot \rho}{1 + (30 - 1) \cdot \rho} = 0,878.$$

2. Разделяне на теста на две

Ще оценим надеждността на теста чрез корелацията между два подтеста. За целта разделяме тестовите задачи на две части – четни и нечетни номера на задачите, след което изчисляваме корелацията на баловете от две половинки.

Коефициента на корелация на Пирсън между двете половини на теста $R_{XY} = 0,951$.

Тъй като полученият коефициент на корелация отразява надеждност въз основа на половинките на теста, а не въз основа на неговата цялост, ще използваме „коригиращата“ формулата на Спирмън – Браун:

$$(5) \quad r_{22} = \frac{2 \cdot R_{XY}}{1 + R_{XY}} = 0,975.$$

3. Корелация между задачите с избираем и свободен отговор.

Коефициентът на корелация на Пирсън е $R_{XY} = 0,769$.

Използваме „коригиращата“ формулата на Спирмън – Браун за оценка на надеждността:

$$(6) \quad r_{22} = \frac{2.R_{XY}}{1 + R_{XY}} = 0,869.$$

4. Алфа на Кронбах

Коефициентът α не определя надеждността на теста, а е само една долна граница за нея. Приема стойности от 0 до 1. Колкото задачите са по надеждни и мерят един и същ действителен бал, толкова α е по-близо до 1. Той се пресмята по формулата:

$$(7) \quad \alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_X^2} \right),$$

където:

k – броят на задачите;

σ_i^2 – дисперсията на i – тата задача;

σ_X^2 – дисперсията на общия бал.

В нашия случай $\alpha = 0,874$.

5. Стандартна грешка на измерването

Стандартната грешка е показател за надеждност в смисъла на оценка за това, колко често могат да се очакват грешки в определен размер при многократното му използване.

Пресмятането на стандартната грешка SE при измерването се извършва по формулата:

$$(8) \quad SE = \sigma_X \sqrt{1 - R},$$

където:

σ_X – средноквадратично отклонение на

общия бал;

R – коефициент на надеждност по формулата на Спирмън – Браун.

Имаме: $SE = 7,61$.

Доверителният интервал за реалната оценка на всеки кандидат се определя въз основа на получената от него оценка X_i , стандартната грешка SE и избраната вероятност за грешка α .

Интерпретирайки α чрез процентно изразена „достоверност“, ако получената оценка от даден кандидат е равна на X_i , то:

♦ При 68% сигурност реалната оценка на този кандидат попада в интервала

$$(X_i - SE; X_i + SE)$$

♦ При 95% сигурност реалната оценка на този кандидат попада в интервала

$$(X_i - 2SE; X_i + 2SE)$$

♦ Почти с пълна сигурност реалната оценка на този кандидат попада в интервала

$$(X_i - 3SE; X_i + 3SE)$$

Ако доверителните интервали на двама кандидата се застъпват, може да се приеме със съответната вероятност за грешка, че действителните постижения (реалните оценки) на тези кандидати не се различават съществено и разликата се дължи на грешки при измерването чрез теста.

6. Дисперсионен анализ

Задачите от теста са разделени по теми в 6 групи (таблица 1). Ще установим дали типа на дадена задача оказва влияние върху общия бал от теста.

Нулевата хипотеза гласи, че всички средни на задачите са равни, независимо от темата към която принадлежи задачата, т.е. средният бал на задача от различните теми е еднакъв.

Алтернативната хипотеза е, че резултатът за задача от различните теми би бил различен.

Определяме равнище на значимост $\alpha = 0,01$.

Пресмятаме:

Междугрупова дисперсия:

$$(9) \quad \sigma_M^2 = \frac{\sum_{j=1}^k (\bar{X}_j - \bar{X}_0)^2 \cdot n_j}{k - 1} \approx 414388.$$

Вътрешногрупова дисперсия:

$$(10) \quad \sigma_B^2 = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2}{n - k} \approx 76643.$$

Изчислява се отношението:

$$(11) \quad F = \frac{\sigma_M^2}{\sigma_B^2} = 5,407.$$

По таблицата на Стюдънт се установява F_T . При 5 и 24 степени на свобода и при $P(t) = 0,99$, $F_T = 3,9$.

Съпоставят се емпиричната с табличната стойност за F . Имаме $F > F_T$, т.е. с риск за грешка 1% може да се твърди, че типа на задачата оказва съществено влияние върху бала от теста.

4. ИЗВОДИ ЗА КАЧЕСТВОТО НА ТЕСТА

Голяма част на задачите в теста са с оптимална трудност (44%). Същевременно

30% от задачите в теста са се оказали много трудни за кандидатите. Това е оправдано с една от поставените цели – да се ограничи броят на отличните оценки. Липсват задачи с много лесна трудност, което може да говори за недобрата подготовка на кандидатите като цяло. Повечето задачи са с много добър и добър коефициент на дискриминация (70%), което дава възможност да се разграничат индивидуалните способности на отделните кандидати. Само 3 задачи са с много нисък коефициент на дискриминация. Половината задачи са с умерен коефициент на корелация, а 40% от задачите са с значителна степен на корелация. Липсват задачи с голяма и много голяма корелация. Това говори за сравнителната независимост на задачите, което води до по-доброто измерване от теста. Само 3 задачи са със слаба корелация. Средната корелация на всички задачи е сравнително ниска, което говори, че теста е направен така, че да покрива висок диапазон от учебния материал по математика за средния курс. Коефициентът на надеждност по Спирмън-Браун и долната граница на надеждността (коефициентът Алфа на

Кронбах) са високи и много близки един до друг, което говори за висока надеждност на теста. Коефициентът на корелация между задачите с четен и нечетен номер е висок, това говори за правилното разпределение на задачите в теста. Същото се вижда и от високата корелацията между задачите с избираем и свободен отговор.

ЛИТЕРАТУРА:

- [1] Калоянов Т., Статистика, Издателство „Тракия-М”, София, 2004 г.
- [2] Стоименова Е., Измерителни качества на тестове, София, 2000 г.
- [3] Улучев Р., Михалев Д., Приложна математика, ВТУ „Тодор Каблешков”, София, 2008 г.
- [4] Алашка Р., Михалев Д., Основни числови характеристики на реален изпитен тест, София, 2009 г.
- [5] Конкурсен тест по математика – 28 юли 2009 година, София, 2009 година - http://vtu.bg/i_admission_b.html .

STATISTICAL ANALYSIS FOR QUALITY ASSESSMENT OF EXAMINATION TEST

Rayna Alashka
alraina@abv.bg

HST “Todor Kableshkov”, 158 Geo Milev Street, 1574 Sofia
BULGARIA

Key words: *test, examination, correlation analysis, dispersion analysis, reliability, education, statistics.*

Abstract: Statistical analysis of actual test data is given. Different indicators measuring for the quality of individual tasks and tests as a whole are calculated. The correlation relationships between them and the population are given. Dispersion analysis of the impact of the tasks allocated to the basic subjects is made. Conclusions about achievement of candidates on the basic subjects and for the quality of the test to achieve of the main targets are made.