Софийски университет „Св. Кл. Охридски“,
Факултет по математика и информатика

Райна Милкова Алашка

ПРИЛОЖЕНИЕ НА ВЕРОЯТНОСТНИ МОДЕЛИ ЗА АНАЛИЗ НА РЕЗУЛТАТИ ОТ ИЗПИТИ И ТЕСТОВЕ

А В Т О Р Е Ф Е Р А Т

на дисертационен труд
за присъждане на образователна и научна степен „Доктор”

Област на висше образование:
1. Педагогически науки.
Професионално направление:
1.3. Педагогика на обучението по …
Научна специалност:
05.07.03. „Методика на обучението по математика и информатика“

Научен ръководител: проф. д-р Кирил Георгиев Банков
София, 2017 г.

СЪДЪРЖАНИЕ

Обща характеристика на дисертационния труд

Актуалност на темата

През последните години оценяването на придобитите знания и умения от ученици и студенти все по често се извършва чрез тестове. Тестовото изпитване има редица предимства, като: бърза проверка (все по-често и автоматизирана), бързо диагностициране на типичните грешки на изпитваните, оценяването не зависи от субективни фактори, проверява се по-голям обем учебно съдържание и др. За да бъде тестовото измерване обективен и надежден метод за оценяване, необходимо е то да се извършва с „качествен“ тест. Конструирането на един тест става по определени правила. Спазването на тези правила е задължително условие за получаване на „качествен“ тест с добри измерителни показатели. В българското образование се наблюдава масовото производство на тестове, голяма част от които не са съобразени с тези правила за конструиране на тестове. Съществува мнението, че писането на тест е лесна работа и може да се извършва от всеки. Не се разбира важността от анализа на резултатите от тестирането и съответните изводи.

В исторически план първата теория на тестовете, която възниква, е така наречената *класическата теория*. За начало на тази теория се водят работите на английския психолог Чарлз Спирмън през годините между 1904 и 1913 [15]. Класическата теория на тестовете се използва широко в практиката и е в основата за пресмятане на параметрите на тестовете и анализ на получените резултати. По-късно в развитието на педагогическите изследвания при разработване на изследователските модели и анализ на данни широко приложение намира и вариант на вероятностно моделиране – IRT (Item Response Theory). Тази теория включва различни параметрични вероятностни модели за определяне на връзката между изпълнението на дадена тестова задача от конкретно лице и неговата способност. Основен недостатък на класическата теория на тестовете е, че основните числови характеристики на задачите (трудност, дискриминация) зависят от измерваните лица и обратно, постиженията (способностите) на измерваните лица зависят от задачите, с които ги оценяваме. Тази взаимно свързаност не съществува при вероятностното моделиране. Основни предимства на последната теория са: (1) способността на изпитваните и трудността на задачите могат да се представят на една и съща скала; (2) способността на изпитваните не зависи от трудността на задачите, с които ги оценяваме; (3) трудността на задачите не зависи от способността на ученици, които оценяваме. Вероятностното моделиране (IRT) се наблюдава и при широко-мащабни международни изследвания, като TIMSS (Trends in Mathematics and Science Study) и PISA (Programme for International Student Assessment).

В настоящия дисертационен труд са дадени практически приложения на тези две теории за анализ на резултати от изпити и тестове. Направен е съпоставителен анализ между тях. Разгледани са техните предимства и недостатъци.

От друга страна, вероятностното моделиране, в частност моделите за разпознаване на риска, намира сериозно развитие от втората половина на 20 век. То се прилага предимно в сферата на икономиката, бизнеса, застраховането, медицината и други социално-икономически области за оценка на риска, като се използват различни вероятностни модели. Някои от публикациите по тази тематика са [34], [35], [58], [62], [65], [66], [74], [76], [81].

В настоящия дисертационен труд са приложени част от тези модели. Направени са техни модификации и многомерни обобщения.

Цел и задачи на дисертационния труд

Основна цел в дисертационния труд е прилагането на различни тестови теории, статистически методи и вероятностни модели, на основата на конкретни изследвания:

- за анализ на резултати от изпити и тестове;
- за оценка влиянието на различни фактори върху резултатите на изпитваните;
- за решаване на редица проблеми, които възникват в образованието.

За постигането на тази цел в дисертационния труд се решават следните *основни задачи:*

1. Проучване и анализ на използваните в образованието тестови теории;
2. Приложение на линейния вероятностен модел (ЛВМ), квадратичния вероятностен модел, Логит-модела и Пробит-модела в образованието.
3. Определяне на оптималната продължителност на занятие в подготвителния курс по математика за 3 и 4 клас на Първа частна математическа гимназия (ПЧМГ);
4. Определяне на оптималната дължина на теста за прием в ПЧМГ;
5. Разработване на домашни работи (задачи за самостоятелна подготовка), съобразени с възможностите на студентите и учебната програма по Висша математика 2 част (ВМ2) във ВТУ „Тодор Каблешков“;
6. Анкетно проучване за идентифициране нагласите на студентите към домашните работи;
7. Определяне на зависимостта на вероятността за успешно полагане на изпита по ВМ2 от количеството на решените примери от домашните задания;
8. Определяне на зависимостта на вероятността за успешно полагане на изпита по ВМ2 от количеството решени примери от зададените домашни работи и посещаемостта на учебните занятия (упражнения, лекции);
9. Изследване на вероятността за успех на изпита по статистика във ВТУ „Тодор Каблешков“, като функция на броя часове на самоподготовка;
10. Оценка на различните форми на обучение по математика във ВТУ „Тодор Каблешков“;
11. Анализ на изпит по ВМ2 по теми;
12. Пресмятане на основните числови характеристики и оценка качеството на изпитен (приемен) тест във ВТУ „Тодор Каблешков”;
13. Сравнителен анализ между реални изпитни тестове във ВТУ „Тодор Каблешков”;
14. Статистически анализ на тест по математика за ученици и оценка за влиянието на различни фактори;
15. Определянето на оптималната дължината на теста и продължителността за изпълнението му, чрез модел от тестове с различна дължина;
16. Сравнителен анализ между различните теории и модели за анализ на тест.

Място на емпиричните изследванията

Проведените изследванията със студенти са извършени във ВТУ „Тодор Каблешков”. Проведените изследвания с ученици от 4 клас са проведени в ПЧМГ и Частно Начално Училище „Питагор”. Проведените изследвания с ученици от 7 клас са проведени в ПЧМГ, Софийска Математическа Гимназия „Паисий Хилендарски“ (СМГ), 38 ОУ „Васил Априлов“ и 45 ОУ „Константин Величков“.

Обект на изследванията

1. Учениците от подготвителните курсове за 4 клас в ПЧМГ, учениците от Частно Начално Училище „Питагор” и кандидатите за 5 клас в ПЧМГ през учебната 2014-2015 година.
2. Студентите редовно и задочно обучение от I курс във ВТУ „Тодор Каблешков” през учебната 2014-2015 година.
3. Кандидат-студентите във ВТУ „Тодор Каблешков” през 2008 и 2009 година.
4. Ученици от седми клас от различни училища в град София.

Хипотеза на дисертационния труд

Познатите от социално-икономически области вероятностни модели за оценка на риска и техни модификации могат успешно да се прилагат за изследвания и анализи в образованието, както самостоятелно, така и в комбинация с класическата теория на тестовете и IRT-моделите.

Методи на изследванията

Методите, които се прилагат са: Линеен вероятностен модел (ЛВМ); Логит-модел; Пробит-модел; Квадратичен вероятностен модел,; Еднопараметричен „модел на Раш“, Двупараметричен „модел на Бирнбаум“; Класическата Теория на Тестовете (КТТ); Метод на най-малките квадрати; Регресионен анализ; Корелационен анализ; Дисперсионен анализ; Метод на Шефе (S-метод), Статистическа обработка на данни, получени от анкетни карти; Графично представяне на резултати.

Структура и съдържание на дисертационния труд

Настоящата дисертация съдържа въведение, три глави, заключение, литература и приложения. Основната част на дисертацията е от 240 страници, а изложението е придружено с фигури и таблици. Включени са 94 фигури и 83 таблици. Приложенията са 5. Списъкът на цитираната литературата включва 81 заглавия. Списъкът от публикации на автора по дисертацията съдържа 7 заглавия, четири от които са цитирани в две различни дисертации за присъждане на образователната и научна степен „Доктор“ в областта на педагогиката и образованието.

В дисертацията релациите (равенства, неравенства, формули) са означени в ляво на реда на релацията за всяка глава отделно с (П.Н), където П е номерът на параграфа в главата, а Н е номерът на съответната релация. Фигурите са отбелязани в средата под самата фигура с Фиг. Г.Ф, където Г е номерът на главата, а Ф е номерът на съответната фигура в главата. Таблиците са отбелязани над самата таблица с Таблица Г.Т, където Г е номерът на главата, а Т е номерът на съответната таблица в главата. Цитиранията са

[Л], където Л е номерът на съответната литература, понякога е споменато и името на автора на съответната литература.

В дисертацията се използват познания по различни математически дисциплини, като теория на апроксимациите, теория на вероятностите и статистика, математически анализ, линейна алгебра, аналитична геометрия и други.

Проверени са 10 изследователски хипотези и над 100 статически хипотези.

Първа глава „Вероятностни модели и възможности за използването им в образованието” се състои от увод и осем параграфа. В увода кратко е описана необходимостта от изследванията, съдържащи се в главата, дадени са кратки исторически бележки. Описани са известни вероятностни модели и се дава нов по-различен подход, с възможности за използването им в образованието. Дадени са теоретичните основи на тези модели и са направени техни подобрения и модификации за успешното им прилагане в образованието. Тези модели са използвани за различни приложения в образованието, както в тази глава, така и в следващата. Могат да се прилагат и за други незасегнати в дисертацията въпроси, не само в образованието, но и в други сфери на обществото.

В **първи параграф** е описан методът на най-малките квадрати (МНМК). Дадени са различни негови модификации – за приближаване на функции, за регресионен анализ, за корелационен анализ, за решаване на преопределени системи. Изведени са директно формули за различна размерност, които са приложени многократно в дисертацията. Дадено е определение за ***условието за хомоскедастицитет***, което означава да имаме еднакво разпределение на дисперсиите σ_i^2 на остатъците. Тогава е в сила $\sigma_1^2 = \sigma_2^2 = ... = \sigma_m^2$.

Във **втори параграф** е представен ***линейният вероятностен модел*** (ЛВМ) в две точки:

1. Еднофакторен модел и 2. Двуфакторен модел.

В разглежданията при ЛВМ зависимата променлива Y е дихотомна и има само две значения – 0 и 1, които са индикатор за наличие и отсъствие на някакво явление.

$$Y_i = \begin{cases} 1, & \text{при наличие на дадено явление,} \\ 0, & \text{при отсъствие на дадено явление.} \end{cases}$$

При еднофакторния ЛВМ е използван двукратно методът на най-малките квадрати (МНМК): първия път за предварително намиране на вероятностите и пресмятане на теглата и втория път за същинското пресмятане на вероятностите; теглата са пропорционални на единица върху средноквадратичното отклонение на остатъците и имат вида: $(\sqrt{W_i})^{-1} = (\sqrt{P_i(1-P_i)})^{-1}$, вж. [18], [19]], [34] .

Ще отбележим, че при еднофакторния модел след прилагането на МНМК се стига до решаването на система две на две, а при двуфакторния модел след прилагането на МНМК се стига до решаване на система три на три.

В **трети параграф** е направена модификация на линейния вероятностен модел, като функцията за приближение не е линейна, а полином от по-висока степен. Така получаваме полиномиален вероятностен модел (ПВМ). По-подробно е описан моделът, когато функцията е полином от втора степен. Този модел наричаме квадратичен вероятностен модел (КВМ). Методът е приложен в първите две изследвания, изложени в дисертацията (изследване 1 и изследване 2).

Целта на **Изследване 1** е определяне на оптималната продължителност на занятието в подготвителния курс по математика за 4 клас. *Място на изследването* е Първа Частна Математическа Гимназия (ПЧМГ). *Обект* на изследването са 44 ученици от 4 клас. *Предмет на изследването* е броят на допуснатите грешки в зависимост от изминалото време от началото на занятието. *Данните* са получени от предварително раздадените бланки за отговори. *Хипотеза за резултата от изследването*: Броят на допуснатите грешки зависи от продължителността на занятието. *Метод на теоретичното изследване* е квадратичният вероятностен модел.

Получените данни от изследването са поместени в таблица 1.1.

Таблица 1.1 Дял на грешните отговори с изменение на времето

Продължителност на обучението в минути	40	60	80	100	120	140	160	180
Дял (част) на грешно решените задачи	0,50	0,40	0,30	0,20	0,05	0,10	0,15	0,25

От емпиричните данни се забелязва, че грешките до един момент намаляват и след това започват да се увеличават. Това ни дава основание да приложим квадратичен вероятностен модел, който се описва с уравнението $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$, където X е времето от началото на занятието, а Y е частта на грешките за това време. Намерена е точката на минимум на параболата и е пресметната оптималната продължителност на занятието.

Резултати от изследването:

Модел	Очаквана стойност на вероятността	Точка на минимум на параболата
КВМ (без тегло)	$\hat{Y}_i = 0,98780 - 0,01320.X_i + 0,00005.X_i^2$	$X_0 = 132,39$, $\hat{Y}(X_0) = 0,114$
КВМ (с тегло)	$Y_i^* = 1,03250 - 0,01414.X_i + 0,00005.X_i^2$	$X_0^* = 131,46$, $Y^*(X_0^*) = 0,103$

Оптималната продължителност е 131,46 минути и ако учебните часове са по 45 минути, това прави 3 учебни часа, като преподавателят има малко повече от 3 минути за заключителни бележки. След това време курсистите са изморени или разконцентрирани и започват да правят повече грешки.

Извод на изследването: Оптималната продължителност на занятието е 3 учебни часа.

Практическо приложение на изследването: В резултат от проведеното изследване и направените изводи, продължителността на занятията в подготвителните курсове в ПЧМГ са по 3 учебни часа.

Реализирано е **Изследване 2**, *целта* на което е определяне на оптималната дължина (броя на задачите) в теста за прием в 5 клас на ПЧМГ. *Място на изследването* е ПЧМГ. *Обект на изследването* е извадка от 67 ученици от подготвителните курсове по математика за 4 клас, учениците от Частно Начално Училище „Питагор” и кандидати за 5 клас в ПЧМГ. *Предмет на изследването* е броят на учениците, успешно положили изпита, в зависимост от броя на задачите в теста. *Данните* са получени от резултатите на учениците от пробни и приемни изпити. *Хипотеза за резултата от изследването*: Броят на успешно положилите изпита зависи от дължината (броя на задачите) на теста. *Метод на теоретичното изследване* е квадратичният вероятностен модел.

Получените данни от изследването са поместени в таблица 1.5.

Таблица 1.5 Успеваемост на учениците според броя на задачите в теста.

Брой задачи в теста	10	15	20	25	30	35	40
Брой учениците, успешно положили изпита	13	19	31	35	34	28	20
Дял (част) на учениците, успешно положили изпита	0,19	0,28	0,46	0,52	0,51	0,42	0,30

От емпиричните данни се забелязва, че частта на изпитваните, успешно положили изпита, има вид на парабола. Определени са нейните коефициентите.

Приложен е квадратичен вероятностен модел с уравнение $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$, където X е броят на задачите в теста, а Y е частта на успешно положилите изпита при този брой задачи. Намерена е точката на максимум на параболата и е определен оптималният брой задачи в изпитната тема.

Резултати от изследването:

Модел	**Очаквана стойност на вероятността**	**Точка на максимум на параболата**
КВМ (без тегло)	$\widehat{Y}_i = -0,3700 + 0,0652.X_i - 0,0012.X_i^2$	$X_0 = 26,95$, $\widehat{Y}\ (X_0) = 0,508$
КВМ (с тегло)	$Y_i^* = -0,3297 + 0,0639.X_i - 0,0012.X_i^2$	$X_0 = 26,50$, $Y_i^*(X_0) = 0,517$

Извод на изследването: Оптималната дължина на теста е 27 задачи.

Практическо приложение на изследването: В резултат от проведеното изследване и направените изводи, от 2017 година ПЧМГ промени формата на приемния си изпит, като броят на задачите в него съответства на получените от изследването резултати.

В **четвърти параграф** е разгледан ***логистичният вероятностен модел (Logit-модел)***. Представена е функцията на логистичното разпределение и са дадени редица нейни свойства. Функцията на разпределение $\varphi(x) = p(x)$, която задава кумулативната вероятност до точката x се представя с формулата: $\varphi(x) = \dfrac{e^x}{1+e^x} = \dfrac{1}{1+e^{-x}}$. От теорията на апроксимациите е известно, че: $\sup\limits_{x\in(-\infty;+\infty)} \left|\varphi(1,702.x) - F(x)\right| < 0,01$, където $F(x)$ е функцията на разпределение на стандартното нормално разпределение. Този факт обяснява защо в параметричните вероятностни модели се използва константата 1,7.

При логистичния вероятностен модел зависимата променлива (следствието) Y^*е ненаблюдаема, обикновено наречена скрита (латентна) променлива. Това, което се наблюдава е фиктивна променлива Y_i, чрез която се опитваме да опишем ненаблюдаемата променлива Y^*. За скритата (латентна) променлива Y^* фиктивната променлива Y_i има дихотомно проявление и се дефинира като:

$$Y_i = \begin{cases} 1, & \text{ако } Y^* \text{ показва наличие на изследваното явление;} \\ 0, & \text{ако } Y^* \text{ показва липса на изследваното явление.} \end{cases}$$

При логистичния вероятностен модел разпределението е логистично и има вида:

$$Y = \frac{1}{1+e^{-(\beta_0+\beta_1 X+\varepsilon)}}.$$

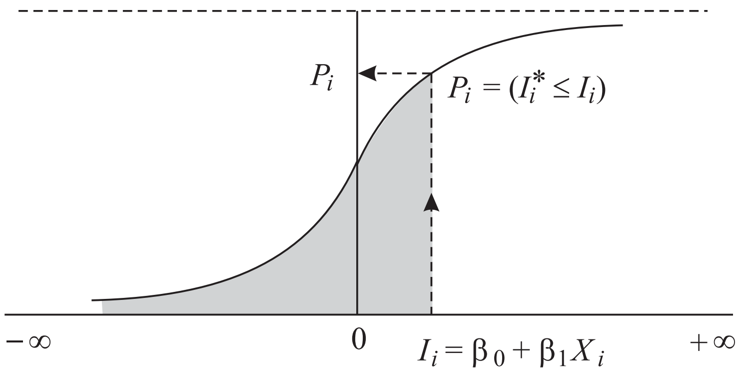
Шансовата пропорция при логистичните вероятностни модели се дефинира като:

$$\frac{P}{1-P} = \frac{1+e^{(\beta_0+\beta_1 X+\varepsilon)}}{1+e^{-(\beta_0+\beta_1 X+\varepsilon)}} = e^{(\beta_0+\beta_1 X+\varepsilon)}.$$

Логаритмуваме двете страни на равенство и получаваме $L = \ln\frac{P}{1-P} = \beta_0 + \beta_1 X + \varepsilon$, който в литературата се нарича *logit* (логит), затова този модел се нарича *Логит*-модел. Приближаваме логита по метода на най-малките квадрати (МНМК). Дадени са теглата, които имат вида: $\sqrt{W_i} = \sqrt{N_i \widehat{p_i}(1-\widehat{p_i})}$, вж. [18], [19]], [34].

Вероятностите да се сбъдне събитието могат да се пресметнат и без тегла, но тогава методът може да доведе до неточности.

В **пети параграф** е описан ***пробит моделът*** (Probit-модел), при който се използва стандартно нормално разпределение. За използването на пробит модела е конструиран ненаблюдаван индекс за полезност $I_i = \beta_0 + \beta_1 X_i$, наречен нормит, за който е в сила: $P_i = F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0+\beta_1 X_i} e^{-\frac{z^2}{2}} dz$, където $F(z)$ е функцията на стандартното нормално разпределение, т.е. $I \in N(0;1)$ (фиг.1.13.) Изходните променливи отново могат да се претеглят с тегло $\sqrt{W_i}$, където $\sqrt{W_i} = \sqrt{\frac{N_i (I_i)^2}{\widehat{p}_i (1-\widehat{p}_i)}}$, вж. [18], [19]], [34] и след това се прилага МНМК.



Фиг. 1.13 Пресмятане на вероятността P_i при зададен индекс за полезност I_i

В **шести параграф** е представен кратко ***тобит моделът*** (Tobit-модел) [34], [74], [81], който носи името на Нобеловия лауреат по икономика от 1981г. Джеймс Тобин. Този модел е модификация на вероятностния пробит модел. Той разглежда тъй наречените цензурирани извадки, което означава, че липсват данни за зависимата променлива на част от извадката. При него се прилага методът на максималното правдоподобие (ММП).

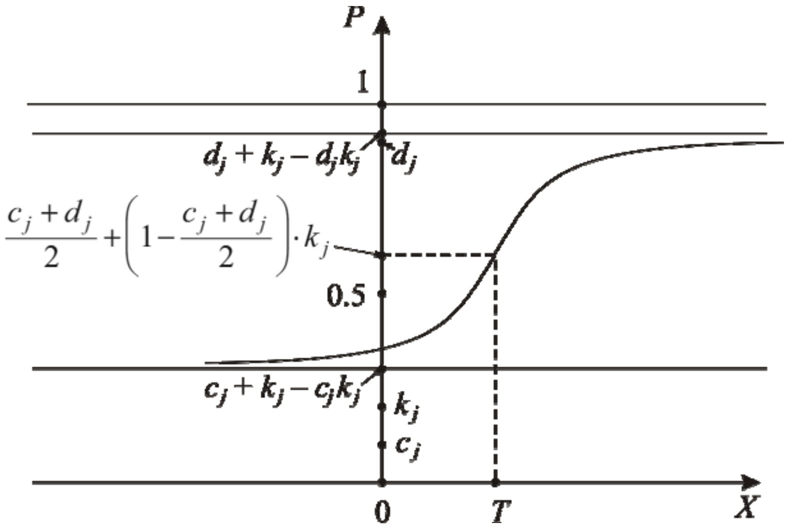
В **седми параграф** са представени четирите параметрични вероятностни модела, прилагани в образованието. Те използват основните параметри на задачите. Такива са параметрите на трудност, дискриминация, налучкване и мотивация. Споменава се за тяхното развитие в исторически план и за техните автори. Някои от публикациите по тази тематика са [12], [51], [52], [53], [54], [55], [59], [67], [69], [72], [79].

През последните години с развитието на информационните и комуникационни технологии нарастват случаите на компрометиране (изтичане на информация за изпита, получаване отвън на верните отговори и др.) на резултатите от изпити и тестове. По информационните медии се изнасят редица данни за компрометиране на държавните зрелостни изпити и изпитите за външно оценяване. Министерството на образованието и науката (МОН) предприема различни мерки, които да ограничат тези негативни явления.

Предложен от автора е нов ***петпараметричен модел***, който дава възможност да отчетем „лошата“ компетентност (компрометираност). При него, освен описаните в предходните вероятностни модели параметри, участва и нов пети параметърът k - „**компрометиране**“. Той показва преписването, подсказването, изтичането на информация за задачата и др. Характеристичната функция на всяка задача е от вида:

$$p(X) = k_j + (1-k_j)\left(c_j + (d_j - c_j)\cdot\frac{1}{1+e^{-a_j\left(X-T_j\right)}}\right).$$

Стойностите на параметъра k_j се изменят в интервала $[0,1]$, но при стойности по-големи от 0,5, те на практика не се прилагат, защото натежават във формулата. На Фиг. 1.23 е представена характеристична крива за петпараметричния модел.



Фиг. 1.23 Характеристична крива – 5 параметъра

Част от резултатите на този параграф са публикувани в [9].

В **осми параграф** са оценени параметрите трудност и дискриминация на заданието, като са сравнени еднопараметричният модел на Раш (вж. [69]) и двупараметричният модел на Бирнбаум (вж. [54]) с подходящи логистични модели, получени по МНМК. Резултатите показват, че при зададени способности на изпитваните, за да имаме най-малка средноквадратична грешка, трудността на заданието и дискриминацията трябва да се задават с конкретни формули, които са посочени. Това са препоръчителната трудност и препоръчителната дискриминация.

При **еднопараметричния модел на Раш** вероятността изпитван със способност *X* да се справи със задача с трудност *T* се дава с формулата: $P\left(X\right)=\dfrac{1}{1+e^{-(X-T)}}$, където *X* е способността на изпитвания, а *T* е трудността на задачата.

За определяне на оптималните параметри в модел на Раш конструираме логистичен вероятностен модел, като търсим L от вида $L_i = X_i - b_0$, където X_i е способността на група i. Тогава за вероятността имаме: $P=\dfrac{1}{1+e^{-(X-b_0)}}$.

Използваме МНМК за намиране на коефициента b_0. След преобразуване за коефициента b_0 получаваме:

$$b_0 = \overline{X} - \ln\left(\overline{\frac{\widehat{P}}{1-\widehat{P}}}\right)_G = \overline{X} - \ln\left(\overline{\widehat{P}}\right)_G + \ln\left(\overline{1-\widehat{P}}\right)_G ,$$

където използвахме означението $\overline{V_G} = \sqrt[m]{V_1.V_2....V_m}$ за средното геометрично на V_i

Сравняваме получения резултат при този логит-модел с еднопараметричния модел на Раш и получаваме, че те ще съвпадат, ако за трудността е изпълнено:

$$T = b_0 = \overline{X} - \ln\left(\overline{\widehat{P}}_G\right) + \ln\left(\overline{\left(1-\widehat{P}\right)}_G\right).$$

Директното прилагане на МНМК може да доведе до неточности, поради различния вид на отклоненията ε_i. Нарушено е условието за хомоскедастицитет (за равенство на стандартните отклонения при различни стойности на факторната променлива). За да решим този проблем използваме тегла, които имат вида: $\sqrt{W_i} = \sqrt{N_i \widehat{P}_i (1-\widehat{P}_i)}$. Прилагаме обобщения МНМК и за трудността получаваме:

$$T = b_0^* = \frac{\sum_{i=1}^{m} X_i \sqrt{W_i}}{\sum_{i=1}^{m} \sqrt{W_i}} - \frac{\sum_{i=1}^{m} L_i \sqrt{W_i}}{\sum_{i=1}^{m} \sqrt{W_i}}$$

При **двупараметричния вероятностен модел на Бирнбаум** вероятността изпитван със способност *X* да се справи със задача с трудност *T* и дискриминация *a* се дава с формулата: $P\left(X\right)=\dfrac{1}{1+e^{-1,7a(X-T)}}$.

За определяне на оптималните параметри в модел на Бирнбаум конструираме логистичен вероятностен модел, като търсим L от вида $L_i = b_1 X_i + b_0$, където X_i е способността на група i. Тогава за вероятността имаме: $P=\dfrac{1}{1+e^{-(b_1 X+b_0)}}$.

Сравняваме двата модела и получаваме, че за да са еднакви, коефициентът на дискриминация трябва да е $a = \frac{b_1}{1,7}$, а коефициентът на трудност да е $T = -\frac{b_0}{b_1}$.

Коефициентите b_0 и b_1 се получават по МНМК. Те са решения на системата:

$$\left| \begin{array}{lllll} b_0.m & + & b_1\sum_{i=1}^{m} X_i & = & \sum_{i=1}^{m} L_i \\ b_0\sum_{i=1}^{m} X_i & + & b_1\sum_{i=1}^{m} X_i^2 & = & \sum_{i=1}^{m} (X_i L_i) \end{array} \right.$$

Част от резултатите на този параграф са публикувани в [7].

Втора глава е озаглавена „**Приложение на вероятностните модели и S-метода за изследвания и измервания в образованието”**. Тя съдържа увод и шест параграфа.

В тази глава са дадени приложения в образованието на линейния вероятностен модел (ЛВМ), Логит-модела и Пробит-модела с една, две и повече променливи. Като конкретен пример са разгледани резултатите от изпита по Висша математика 2 част на студенти от ВТУ „Тодор Каблешков”. Върху тези резултати са приложени различни вероятностни модели за пресмятане на вероятността за успешно полагане на изпита в зависимост от посещаемостта на учебните занятия, изпълнението на домашните задания и други. Дадени са и статистически методи за оценка на различните методи, форми, стил и др. на преподаване от преподавателите, които оказват влияние върху получените резултати от студентите на изпита. Определена е връзката между отделните теми от изпита.

В **първи параграф** е разгледана учебната дисциплина Висша математика 2 част. Представени са анотацията на дисциплината, правилата за провеждане на изпита и начинът за формиране на оценка. Направен е анализ на резултати по дисциплината през последните години. Представени са проблемите при обучението по математика на студентите. Описана е важната роля на домашните работи (задачи за самостоятелна подготовка) за частичното преодоляване на тези проблеми. Разработени са домашните работи по Висша математика 2 част, които са съобразени както с учебния план по дисциплината, така и с липсата на основни знания от елементарната математика на първокурсниците. Описани са структурата и функцията на домашните работи. Съставени са общо 7 домашни работи (85 задачи, 850 примера), които обхващат целия учебен материал.

Извършено е **Изследване 3**, което има за *цел* да се идентифицират нагласите на студентите към домашните работи. *Предмет на изследването* е анкетно проучване, пряко насочено към оценка отношението на студентите към домашните работи в няколко аспекта: ефект за развитие на способностите; начин на задаване на домашните работи; необходимо условие за постигане успех на изпита по ВМ2; затруднения при решаване на домашните работи. *Място на изследването* е ВТУ „Тодор Каблешков”. *Обект на изследването* са 120 студенти от ВТУ „Тодор Каблешков”. Анкетираните студенти са разделени по равно в четири групи: първата от тях са студенти редовно обучение, за които домашната работа е задължителна; втората група са студенти редовно обучение, за които домашната работа не е задължителна; третата група са студенти задочно обучение, за които домашната работа е задължителна; четвъртата група са студенти задочно обучение, за които домашната работа не е задължителна. *Инструмент на изследването* е анкетна карта. *Методите на теоретичното изследване* са статистическа обработка на получените резултати и графично представяне на данните. *Хипотеза на изследването:* Тъй като домашните работи не влияят пряко на изпитната оценка (нямат дял при формирането на оценката), а са примерни задачи за самоподготовка за изпита, то голяма част от студентите ще имат положително отношение към тях. *Резултати от изследването:* 71% от анкетираните студенти определят положителен ефект на домашните работи за развитие на способностите си; интерес към домашните работи в обучението по ВМ 2 част проявяват 65% от студентите; 73% от анкетираните посочват като подходящ начина на задаване на домашната работа; 78% от студенти смятат, че задачите в домашните работи са подредени добре; анкетираните студенти по безспорен начин определят, че изпълнението на домашните работи е необходимо условие за успешното полагане на изпита (81% от тях). *Изводи от изследването:* По-голяма част от студенти определят, че домашните работи имат положителен ефект за развитие на способностите им и считат, че те са необходимо условие за постигане успех на изпита по ВМ 2 част. Одобряват начина на задаване на домашните работи.

Във **втори параграф** са дадени приложения на линейния вероятностен модел (ЛВМ), Логит-модела и Пробит-модела с една факторна променлива.

Направено е **Изследване 4**, *целта* на което е определяне на зависимостта на вероятността за успешно полагане на изпита от количеството на решените примери от домашните задания. *Място на изследването* е ВТУ „Тодор Каблешков”. *Обект на изследването* са 50 студента – редовно и задочно обучение от I курс. *Предмет на изследването* са резултатите от изпита по ВМ 2 част. *Данните* са получени от изпитните работи, изпитните протоколи, справки от преподавателите за изпълнение на домашните задания. *Хипотеза за резултата от изследването*: Вероятността за успех на изпита зависи от количеството на решените примери от домашните задания. *Методи на теоретичното изследване:* Приложени са 6 еднофакторни модела. Те са линейният вероятностен модел (ЛВМ), Логит-моделът и Пробит-моделът, всеки от тях без тегло и с тегло.

Получените данни от изследването са обобщени в таблица 2.7.

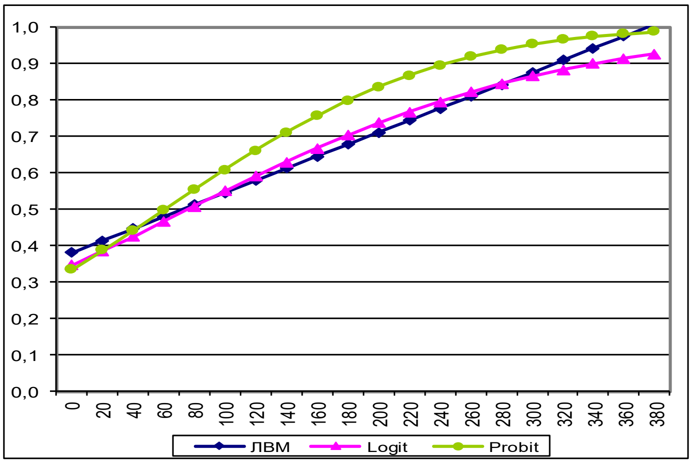
Таблица 2.7 Брой студенти, издържали изпита, в зависимост от решените примери

Брой решени примери	0	80	100	130	160	180	200	260	300
Брой студенти	10	5	2	2	3	8	10	5	5
Брой студенти издържали изпита	4	1	1	1	2	6	9	4	4

Резултати от изследването: При всички, разгледани модели има зависимост на стойността на вероятността за успешно полагане на изпита от броя на решените примери от домашните задания (задачите за самоподготовка).

Сравнението между трите групи вероятностни модели е извършено в четири направления.

Първо, при трите модела следствието е представено на слабата скала с две разновидности. За предоставянето на тези две разновидности се използват фиктивни означения $Y_i = \begin{cases} 1, & \text{студентът е издържал изпита} \\ 0, & \text{студентът не е издържал изпита} \end{cases}$. При третиране на следствието съществува принципна разлика. При ЛВМ се анализира дихотомната променлива, такава, каквато е. При логит и пробит моделите се приема, че съществува латентна променлива, за която се наблюдава дихотомно проявление. Разликата между логит и пробит модела е свързана с кумулативното честотно разпределение на остатъците. При логит модела то е логистично, а при пробит модела следва нормалния закон на разпределение. Както се вижда от Фиг. 2.19, двата модела са напълно сравними. Получените криви за стойността на вероятността за успешно полагане на изпита в зависимост от броя на решените примери от домашните задания са достатъчно близки една до друга. Основната разлика е, че кривата на пробит модела е по-стръмна и по-плътно се доближава до граничните стойности, докато логит моделът има леко повдигната опашка.



Фиг. 2.19 Сравнителна графика на ЛВМ, логит и пробит модела

Второ, получените оценки на параметрите на регресионното уравнение при трите вероятностни модела не са директно сравними. Причината за това е, че при трите модела разсейването на оценките е различно (колона 5 на таблица 2.10).

Трето, тълкуването на регресионните коефициенти при трите вероятностни модела е различно. При ЛВМ регресионният коефициент b_1^* измерва директно промяната във вероятността P_i студентът да издържи изпита по ВМ2. Той ни показва с колко средно ще се промени вероятността при промяна на броя решени примери с единица. Промяната във вероятността остава константа за всички значения на *X*. При логит моделите регресионният коефициент b_1^* измерва промяната в логаритъма на шанса *L* за единица промяна на *X*. Тук промяната във вероятността P_i студентът да издържи изпита по ВМ2 се свързва с регресионния коефициент b_1^* по-сложно, а именно $\Delta P_i = b_1^* P_i (1 - P_i)$. При пробит моделите обвързаността между вероятността и регресионния коефициент е най-сложно. Промяната във вероятността P_i се свързва с регресионния коефициент b_1^* с промяната на израза $b_1^* F(I_i)$.

Таблица 2.10 Основни характеристики на ЛВМ, логит и пробит модела

Модел	Оценки на параметрите		Функция за вероятността	Дисперсия на остатъците
	b_0^*	b_1^*		
ЛВМ	0,3804	0,0017	$P(X_i) = b_0^* + b_1^* X_i$	$\sigma^2 = P_i(1-P_i)$
Логит модел	-0,6342	0,0083	$P(X_i) = \dfrac{1}{1+e^{-(b_0^*+b_1^* X_i)}}$	$\sigma^2 = \dfrac{1}{N_i P_i(1-P_i)}$
Пробит модел	-0,4270	0,0070	$P(X_i) = F(b_0^* + b_1^* X_i) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{b_0^*+b_1^* X_i} e^{-\frac{z^2}{2}} dz$	$\sigma^2 = \dfrac{P_i(1-P_i)}{N_i (I_i)^2}$

Четвърто, като разгледаме ЛВМ виждаме, че коефициентът b_0^* е положителен, което показва, че разгледаният фактор не е единственият фактор, от който зависи успеваемостта на студентите на изпита. Същото показват и графиките на логит и пробит моделите.

Извод на изследването: Стойността на вероятността зависи от броя на решените примери, но това не е единственият фактор, който влияе за успешното полагане на изпита.

Практическо приложение на изследването: След изследването използването на домашните работи се прилага във всички групи и от всички преподаватели.

В **трети параграф** са дадени приложения на линейния вероятностен модел (ЛВМ), Логит-модела и Пробит-модела с две факторни променливи.

Направено е **Изследване 5**, *целта* на което е определяне на зависимостта на вероятността за успешно полагане на изпита от количеството решени примери от зададените домашни работи и посещаемостта на учебните занятия (упражнения, лекции). *Място на изследването* е ВТУ „Тодор Каблешков”. *Обект на изследването* са 50 студента – редовно и задочно обучение от I курс. *Предмет на изследването* са резултатите от изпита по ВМ 2 част. *Данните* са получени от справки от преподавателите за изпълнение на домашните задания и присъствени списъци, изпитните работи и изпитните протоколи. *Хипотеза за резултата от изследването*: Вероятността за успех на изпита зависи от количеството на решените примери от домашните задания и посещаемостта на учебните занятия. *Методи на теоретичното изследване:* Приложени са 6 двуфакторни модела. Те са линейният вероятностен модел (ЛВМ), Логит-моделът и Пробит-моделът, всеки от тях без тегло и с тегло.

Обобщените данни от изследването са дадени в таблица 2.12.

Таблица 2.12 Брой студенти, издържали изпита, в зависимост от решените примери и посещаемостта на учебните занятия

Брой примери	0	0	100	200	200	200	300	300
Брой занятия	20	30	10	10	20	30	20	30
Брой студенти	4	6	7	3	5	15	6	4
Брой издържали изпита	1	3	2	1	3	14	5	3

Резултати от изследването: При всички разгледани модели има зависимост на стойността на вероятността за успешно полагане на изпита от броя на решените примери от домашните задания (задачите за самоподготовка) и посещаемостта на учебните занятия (упражнения, лекции). В следващата таблица са поместени основните характеристики на двуфакторните модели, получени при изследването.

Модел	**Регресионно уравнение и очаквана стойност на вероятността**	**Критична права**
ЛВМ	$Y_i^* = -0,2168 + 0,0019.X_{1,i} + 0,0251.X_{2,i}$ $P_i = Y_i^*$	$P_i = 0,5$ $0,0019.X_{1,i} + 0,0251.X_{2,i} = 0,7168$
Логит модел	$\hat{L}_i^* = -2,7309 + 0,0069.X_{1,i} + 0,0951.X_{2,i}$ $\hat{P}_i^* = \dfrac{1}{1+e^{-\hat{L}_i^*}}$	$P_i = 0,5 \Rightarrow L_i = 0$ $0,0069.X_{1,i} + 0,0951.X_{2,i} = 2,7309$
Пробит модел	$\hat{I}_i^* = -1,9893 + 0,0035.X_{1,i} + 0,0923.X_{2,i}$. $\hat{P}_i^* = F(\hat{I}_i^*)$	$P_i = 0,5 \Rightarrow I_i = 0$ $0,0035.X_{1,i} + 0,0923.X_{2,i} = 1,9893$

Изводи и възможни приложения от изследването.

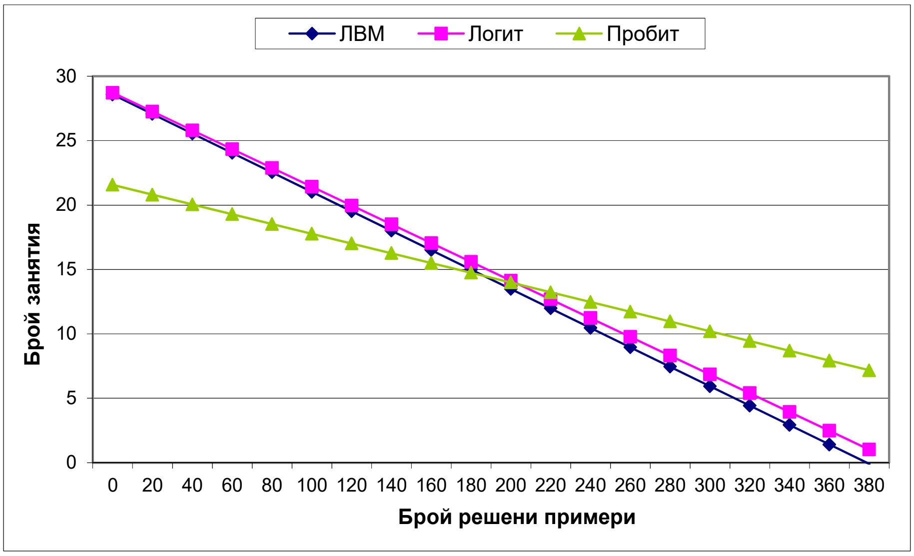
Първо, като разгледаме ЛВМ виждаме, че коефициентът b_0^* е отрицателен. Това показва, че разгледаните два фактора изчерпват основните фактори, от които зависи успеваемостта на студентите на изпита. Без известен брой посещения на учебните занятия или частично изпълнение на домашните работи (самоподготовка) е почти невъзможно да се положи успешно изпитът. Ако студент, който не е посещавал нито едно учебно занятие и не е изпълнявал нито една домашна работа, си вземе изпита, то това се дължи на преписване, подсказване, предварителни компетентности или на случайни причини. Това не е така при еднофакторния модел, разгледан в предишния параграф.

Второ, коефициентите пред факторните променливи в ЛВМ показват, че решаването на 10 примера повече повишава вероятността, изразена в проценти, с 1,9 %. Посещението на всяко занятие води до повишаване на вероятността с 2,5%. Можем да направим извода, че посещението на едно занятие е еквивалентно на решаването на 14 примера.

Трето, при трите модела коефициентите пред факторните променливи са положителни. От това и от вида на функциите на вероятността се вижда, че те са растящи функции спрямо всяка от тези променливи.

Четвърто, на Фиг. 2.23 са представени критичните прави в равнината X_1OX_2 за всеки от моделите. Ако една точка с координати (X_1^0, X_2^0) лежи на правата, то вероятността, според съответния модел е равна на 0,5, ако е под правата, вероятността е по-малка от 0,5, а ако е над правата, вероятността според съответния модел е над 0,5.

Ако обобщим трите модела и имаме точка с координати (X_1^0, X_2^0) от първи квадрант на равнината X_1OX_2, която лежи и под трите прави на моделите, то тогава вероятността студентът да си вземе изпита е по-малка от 0,5.



Фиг. 2.23 Критични прави за ЛВМ, логит и пробит модела

За точка с координати (X_1^1, X_2^1) от първи квадрант на равнината X_1OX_2, която лежи и над трите прави на моделите, вероятността студентът да си вземе изпита е по-голяма от 0,5.

Част от резултатите на този параграф са публикувани в [8].

В **четвърти параграф** е направено многомерно обобщение във векторен вид на параметричните вероятностни модели, както и на логистичния и пробит модела. Обосновано е защо се налага това обобщение. Определена е критичната права (равнина, хиперравнина), където вероятността е равна на 0,5. Посочен е най-краткият път до нея. По-подробно е описан **векторният вид на двуфакторния логит** модел.

Нека L е оценката за логита при двуфакторния модел от вида $L = b_0 + b_1X_1 + b_2X_2$.

Когато той е равен на нула ($L = 0$), вероятността за успех е равна на 0,5 ($P = 0,5$).

Пресмятаме градиента му (нормалния вектор $\vec{n}$ към правата, която се задава с уравнението $0 = b_0 + b_1X_1 + b_2X_2$ в равнината X_1OX_2):

$$grad(L) = grad(b_0 + b_1.X_1 + b_2.X_2) = \frac{\partial L}{\partial X_1}.\vec{i} + \frac{\partial L}{\partial X_2}.\vec{j} = b_1\vec{i} + b_2\vec{j} = \vec{n}(b_1, b_2).$$

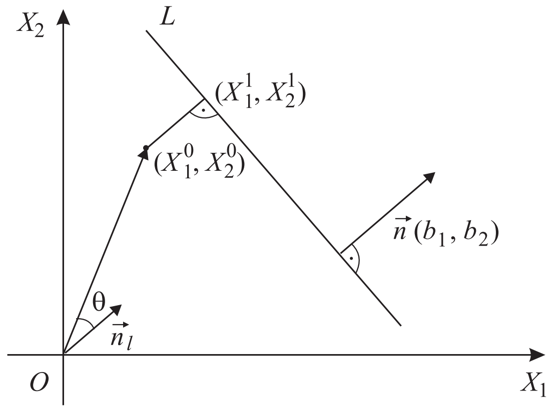
Тогава може да запишем логита във векторен вид: $L = b_0 + \vec{n}.\vec{X}$.

Нека $\overrightarrow{X^0} = \left(X_1^0, X_2^0\right)$ ($\overrightarrow{X^0} = X_1^0.\vec{i} + X_2^0.\vec{j}$) е векторът, характерен за студент с неговите стойности на факторните променливи X_1, X_2, за които вероятността да си вземе изпита е по-малка от 0,5. Тогава е в сила: $(grad(L).\overrightarrow{X^0}) = s < -b_0$.

Ако студентът иска да достигне критичната права или да я надмине (да повиши вероятността си за успех до и над 0,5) най-бързо, трябва да се движи по вектора на градиента. Трябва да увеличи показателите си и по двата фактора и да достигне поне до състояние $\overrightarrow{X^1}\left(X_1^1, X_2^1\right)$, такова че $(grad(L).\overrightarrow{X^1}) = -b_0$ (Фиг. 2.24).

За вектора $\overrightarrow{X^1}$ е изпълнено: $\overrightarrow{X^1} = \overrightarrow{X^0} + r.\underbrace{grad(L)}_{\vec{n}}$, където $r = \frac{(-b_0) - s}{|\vec{n}|^2} = \frac{(-b_0) - s}{b_1^2 + b_2^2}$.

Тъй като $s=\left(grad(L).\overrightarrow{X^0}\right)=\left(\vec{n}.\overrightarrow{X^0}\right)=\left|\vec{n}\right|.\left|\overrightarrow{X^0}\right|.\cos\theta$, способността зависи от два компонента - от дължината на вектора $\overrightarrow{X^0}\left(X_1^0,X_2^0\right)$ и от това, какъв е ъгълът θ между него и вектора $\vec{n}=grad(L)$.



Фиг. 2.24 Векторен вид на двуфакторния логит модел

При различни вектори $\overrightarrow{X_i^0}\left(X_{1i}^0,X_{2i}^0\right)$ с еднаква дължина $\left|\overrightarrow{X^0}\right|$, за способността ще имаме най-голяма стойност при този вектор, който е колинеарен на вектора $\vec{n}=grad(L)$ (ъгълът, заключен между него и $\vec{n}$, е равен на нула).

Логита записваме във вида:

$$L=b_0+\vec{n}.\overrightarrow{X}=\left|\vec{n}\right|\left(\frac{b_0}{\left|\vec{n}\right|}+\frac{\vec{n}}{\left|\vec{n}\right|}\cdot\overrightarrow{X}\right)=\left|\vec{n}\right|\left(\frac{b_0}{\left|\vec{n}\right|}+\overrightarrow{n_l}.\overrightarrow{X}\right)=\left|\vec{n}\right|\left(\overrightarrow{n_l}.\overrightarrow{X}-\frac{(-b_0)}{\left|\vec{n}\right|}\right),$$

където: трудността е $\dfrac{\left(-b_0\right)}{\left|\vec{n}\right|}$; способността е $\overrightarrow{n_l}.\overrightarrow{X}$; дискриминацията е $\left|\vec{n}\right|=\sqrt{b_1^2+b_2^2}$.

Приложение на изложения метод е направено в **Изследване 6**, *целта* на което е определяне на вероятността за успешно полагане на изпита по статистика в зависимост от броя на часовете за практическа и теоретична самоподготовка. *Обект на изследването* са 75 студента – редовно и задочно обучение от I курс във ВТУ „Тодор Каблешков“. *Предмет на изследването* са резултатите от изпита по статистика. *Данните* са получени от попълнените от студентите въпросници, изпитните работи и изпитните протоколи. *Хипотеза за резултата от изследването*: Вероятността за успех на изпита по статистика зависи от броя на часовете за практическа и теоретична самоподготовка. *Методи на теоретичното изследване:* Двуфакторен логит модел и неговия векторен вид.

Резултати от изследването:

Модел	**Регресионно уравнение и очаквана стойност на вероятността**	**Критична права**
Логит модел	$\widehat{L}_i^*=-1,8327+0,0687.X_{1,i}+0,0568.X_{2,i}$ $\widehat{P}_i^*=\dfrac{1}{1+e^{-\widehat{L}_i^*}}$	$P_i=0,5\Rightarrow L_i=0$ $0,0687.X_{1,i}+0,0568.X_{2,i}-1,8327=0$

Да разгледаме група студенти, за които вероятността да си вземат изпита е точно 0,5. За да намерим колко най-малко часове самоподготовка от двата вида са били необходими на един студент от тази група, трябва да намерим пресечната точка на правата $0,0687.X_{1,i} + 0,0568.X_{2,i} - 1,8327 = 0$ и правата, която е перпендикулярна на нея и минава през точката (0; 0). Координатите на тази точка са $\left(X_1^*; X_2^*\right) = (16;13)$. Тя показва най-бързия начин ($\left|\overrightarrow{X^*}\right|$ е най-малко) по който един студент може да стигне до желаната вероятност 0,5.

Извод: Оптималното време, необходимо за самоподготовка на един студент за да постигне успех на изпита (с вероятност 0,5), е 16 часа за практическа и 13 часа за теоретична самоподготовка.

Част от резултатите на този параграф са публикувани в [8].

В **пети параграф** е разгледан методът на Шефе (Schefe) [21], [44], означен кратко като S-метод, за множествени сравнения. Този метод е част от post-hoc тестовете за множествени сравнения. Въведени са различни видове контрасти.

Методът е приложен в **Изследване 7**. *Цел на изследването* е установяване наличието или отсъствието на зависимост между резултатите от изпита по ВМ 2 и формата на обучение. *Място на изследването* е ВТУ „Тодор Каблешков”. *Обект на изследването* са 79 студента – редовно и задочно обучение от I курс. *Предмет на изследването* са резултатите от изпита по ВМ 2 част. *Данните* са получени от изпитните работи и изпитните протоколи. *Хипотеза за резултата от изследването*: Формата на обучение оказва влияние на резултатите от изпита по ВМ 2 част. *Методи на теоретичното изследване:* Дисперсионен анализ, Метод на Шефе.

Резултати от изследването:

Таблица 2.20 Основни характеристики по групи

Форма	Група	Брой	Сума	Средно	Дисперсия
Форма 1	Задочно обучение със задължителна домашна работа	21	1232	58,67	2097,83
Форма 2	Редовно обучение със задължителна домашна работа	29	1877	64,72	1801,35
Форма 3	Редовно обучение	16	606	37,88	855,98
Форма 4	Задочно обучение	13	192	14,77	492,53

Таблица 2.21 Таблично представяне на резултати от дисперсионния анализ

Източник на дивиация	Сума от квадратите на отклоненията	Степени на свобода	Оценки на дисперсията	Емпирична стойност	Теоретична стойност
Между-групова	26329,1	3	8776,36	$F_e = 5,92$	$F(0,05,3,75) = = F_T = 2,73$
Вътрешно-групова	111144,5	75	1481,93	$F_e > F_T \Rightarrow$ Приемаме H_1.	
Обща	137473,6	78			

Изводът е, че ефектът от обучението по четирите различни форми е различен. Не е ясно обаче кои именно форми или групи от форми се различават и дали има форми, които са с еднакъв ефект.

Сега ще проверим дали обучението по форма 2 (редовно обучение със задължителна домашна работа) дава резултати, които се различават от резултатите на другите три форми на обучение. Записваме нулевата хипотеза във вида:

$$H_0 : (+1)\mu_1 + (-3)\mu_2 + (+1)\mu_3 + (+1)\mu_4 = 0 .$$

В този сложен контраст всички коефициенти са различни от нула и всички средни участват в изчисленията. В таблица 2.24 са дадени необходимите данни и изчисления за този контраст.

Таблица 2.24 Изчисляване на емпиричната характеристика при сложен контраст

Група (форма на обучение)	1	2	3	4
Средно в групата $\overline{y}_j$	58,67	64,72	37,88	14,77
Брой в групата n_j	21	29	16	13
Коефициент	+1	-3	+1	+1
Емпирична характеристика	$\overline{F}_e = \dfrac{(58,67 - 3.64,72 + 37,88 + 14,77)^2}{1481,93.\left(\dfrac{1}{21} + \dfrac{9}{29} + \dfrac{1}{16} + \dfrac{1}{13}\right)} = 9,32$			

Сравняваме съответната емпирични характеристики с теоретичната. Тъй като $9,32 = \overline{F}_e > \overline{F}_T = 8,18$, нулевата хипотеза се отхвърля и се приема алтернативната хипотеза. Можем да направим *извод*, че обучението по форма 2 дава среден резултат, който е различен от средния резултат на комбинацията от другите три форми.

В **шести параграф** е направен анализ на изпита по Висша математика 2 част по теми. Отчетени са резултатите от статистическото изследване с S-метода за връзката между отделните теми от изпита. Използван е и корелационен анализ [10], [28] за връзката между отделните теми от конкретен изпит. Взети са предвид резултатите от ЛВМ, за да се оцени защо някои теми са по-предпочитани от студентите пред други. Дадени са възможни обяснения за това. Направен е анализ за трудността на всяка задача от изпита. Оценена е надеждността на изпита като цяло и са направени препоръки за неговото подобряване.

Трета глава е озаглавена „**Статистически и вероятностни методи за сравняване и оценка на кандидат-студентски тестове и оценка на реален изпитен тест за ученици**”.

В първите три параграфа на тази глава е приложена Класическата Теория на Тестовете (КТТ) за пресмятане на параметрите на тестовете и анализ на получените резултати [4], [5], [6], [13], [15], [22], [33], а в останалите параграфи е приложена теорията от вероятностното моделиране, известна още като IRT (Item Response Theory) и теорията на логит моделите.

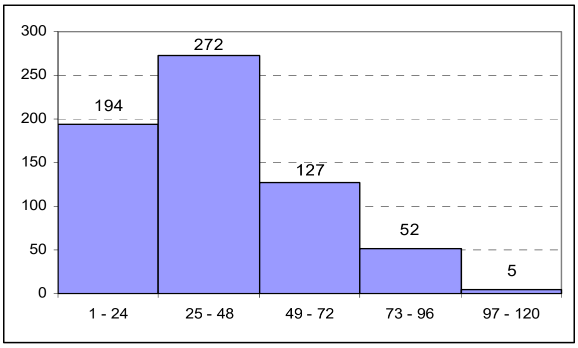
В първи параграф е извършено **Изследване 8**, *целта* на което е да се направи независима оценка на качеството на приемен тест за ВТУ „Тодор Каблешков”, проведен на 28 юли 2009 година. *Хипотеза за резултата от изследването*: Приемният тест за ВТУ „Тодор Каблешков”, проведен на 28 юли 2009 година има добри измерителни показатели и изпълнява целите, за които е съставен. *Задачи на изследването:* Представяне на основни емпирични статистики на теста; Анализ на качествата на задачите от теста; Надеждност на теста и стандартна грешка на измерването. *Място на изследването* е ВТУ „Тодор Каблешков”. *Данните* са получени от изпитните работи на 650 кандидат-студенти. *Метод на теоретичното изследване* е статистически анализ на данни, основан на класическата теория на тестовете.

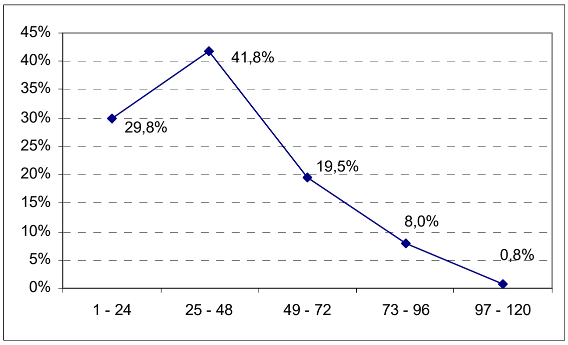
Част от резултатите от изследването са публикувани в [4] и [6].

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. 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 по 6 точки при правилен отговор и по 0 точки при грешен или неотбелязан отговор. Максималният бал на теста е 120 точки.

Резултати от изследването:

Баловете на кандидат-студентите, изразени в точки, са групирани в 5 групи. На Фиг. 3.1 е представена хистограма на абсолютните честоти, а на Фиг. 3.3 - полигонът на относителното честотно разпределение на кандидатите.



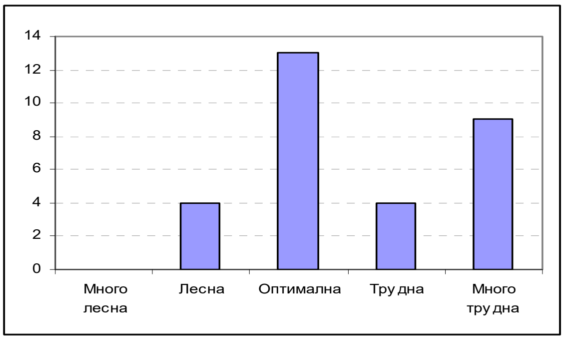


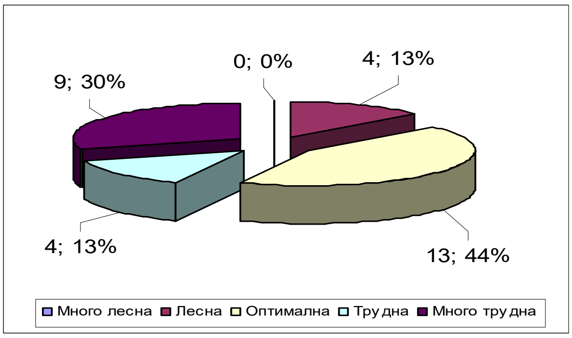
Фиг. 3.1 Абсолютните честоти **Фиг.3.3** Относително разпределение

Най-много кандидати (41,8%) попадат във втората група (25 – 48), която съответства на среден.

Анализът на задачите от теста е необходим за откриване на евентуални дефекти в измерителните качества на теста, дължащи се на лошо функциониращи задачи, които не измерват в необходимата степен постиженията, които мери целия тест и/или са с неподходяща трудност.

На Фиг. 3.5 и Фиг. 3.6 са представени съответно честотното разпределение и кръгова диаграма на процентното разпределение на задачите от теста по трудност.

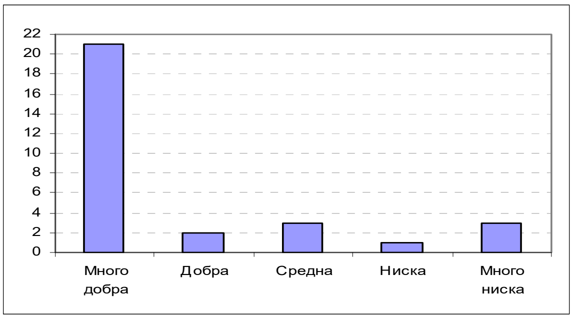


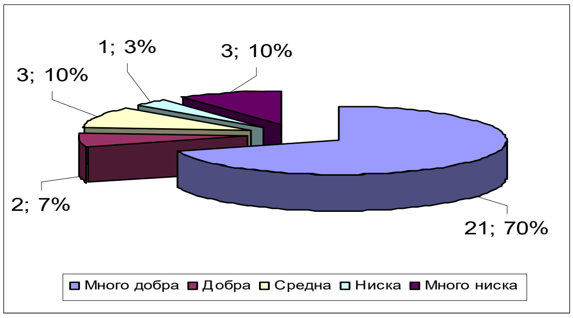


Фиг. 3.5 Брой задачи по трудност **Фиг. 3.6** Структура на задачите по трудност

Тестът не съдържа много лесни задачи, тъй като прекалено лесните задачи дават една постоянна добавка към бала на всяко лице. По-големият брой трудни задачи води до изкривяване на разпределението на бала на ляво. Такъв тест работи за отделянето на лица с отлично разбиране на материала от останалите.

Дискриминацията на задачата е степента, в която тя разграничава кандидатите с високи постижения от тези с ниски постижения. На Фиг. 3.7 и Фиг. 3.8 са представени съответно хистограма и кръгова диаграма на процентното разпределение на задачите по дискриминативна сила.

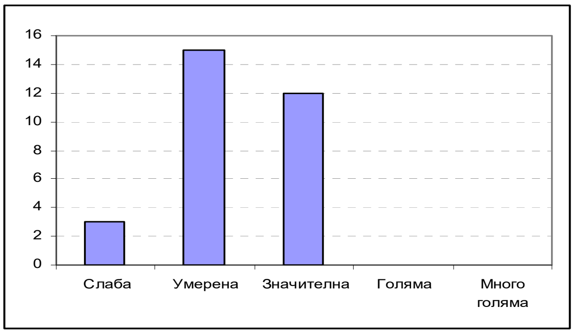


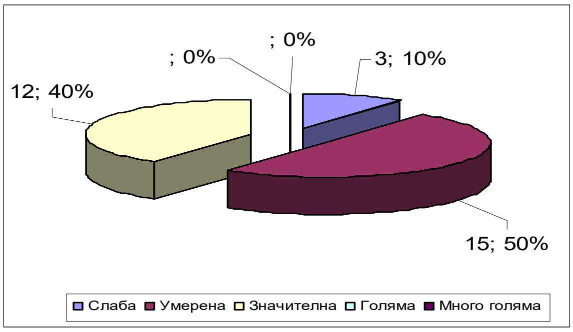


Фиг. 3.7 Брой задачи по дискриминация **Фиг. 3.8** Структура на задачите по дискриминация

При равни други условия, колкото по-големи са коефициентите на дискриминация на задачите, толкова по надежден е теста. Разпределението на баловете се разпростира по-широко, когато средната дискриминация нараства. Надеждността на теста може да се увеличи чрез използване на средно трудни задачи с висока дискриминация.

Корелацията дава друг начин за оценяване на дискриминацията на тестовите задачи. Тя определя връзката на задачата с останалите задачи. На Фиг. 3.7 и Фиг. 3.8 са представени съответно хистограма и кръгова диаграма на процентното разпределение на задачите по корелационната им зависимост с общия бал.





Фиг. 3.9 Брой задачи по корелационната им зависимост с общия бал

Фиг. 3. 10 Структура на задачите по корелационната им зависимост с общия бал

Изводи:

Голяма част от задачите, разгледани 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 намали броят на отличните оценки. Липсват задачи с много лесна трудност, което показва слабата подготовка на кандидатите като цяло. Спорен момент е даването на една точка за неотбелязан отговор за първите 20 задачи. Това дава един постоянен бал за кандидатите, които евентуално изпитват трудност при решаването на съответната задача. От друга страна това ограничава случайното попълване на отговорите (налучкването на верния отговор).

Повечето задачи в теста са с много добър и добър коефициент на дискриминация (70%), което дава възможност да се разграничат индивидуалните способностите 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 със значителна степен на корелация. Липсват задачи с голяма и много голяма корелация. Това показва сравнителната независимост на задачите, което води до по-доброто измерване от теста. Прави впечатление слабата корелация на задачи с номер 9, 13 и 25. Същите имат и много нисък коефициент на дискриминация и съответно труден и много труден коефициент на трудност. Това прави задачите неподходящи за изпитваните кандидати и не дава принос за разграничителната способност на теста, а само ограничава максималния брой точки. Забелязва се, че за решаването на тези задачи е необходимо да се направят повече от три предварителни преобразования. Това показва, че трябва да се избягва даването на такива задачи в теста. Корелациите между всеки две задачи са сравнително ниски, което показва, че тестът е направен така, че да покрива висок диапазон от учебния материал по математика за средния курс. Дистракторите на задачите са сравнително добре подбрани, с малки изключения. Такива дистрактори са например: за задача № 6 – 4 отговор, за задача № 8 – 3 отговор, за задача № 14 – 2 отговор. Те са събрали малък процент отговори.

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 наблюдавания и действителния бал. Високата надеждност води до ограничаване грешката при измерване на действителния бал. Коефициентът на надеждност по Спирмън–Браун (0,878) и долната граница на надеждността (коефициентът Алфа на Кронбах – 0,874) са високи и много близки един до друг, което показва висока надеждност на теста. Коефициентът на корелация между задачите с четен и нечетен номер е висок – 0,975, което показва правилно разпределение на задачите в теста. Същото се вижда и от високата корелацията между задачите с избираем и свободен отговор (0,869).

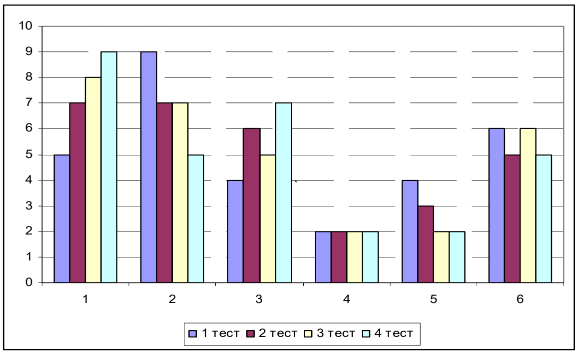
Извод на изследването: Приемният тест за ВТУ „Тодор Каблешков”, проведен на 28 юли 2009 година има добри измерителни показатели и изпълнява целите, за които е съставен.

Във **втори параграф** е разгледано **Изследване 9,** *целта* на което е да се направи сравнителен анализ на 4 различни теста за прием във ВТУ „Тодор Каблешков”. *Хипотези за резултата от изследването*: 1. Не са налице съществени различия между основните характеристики на задачите в различните тестове. 2. Различните тестове измерват едно и също количество знание от кандидатите, като отделят еднакви групи от кандидати за прием във ВТУ „Тодор Каблешков”. *Задачи на изследването:* Сравнителен анализ на основни емпирични статистики на тестовете; Сравнителен анализ на качествата на задачите от тестовете; *Място на изследването* е ВТУ „Тодор Каблешков”. *Данните* са получени от всички 2046 изпитни работи на кандидат-студентите. *Методи на теоретичното изследване:* сравнителен анализ, основан на класическата теория на тестовете, корелационен анализ, дисперсионен анализ.

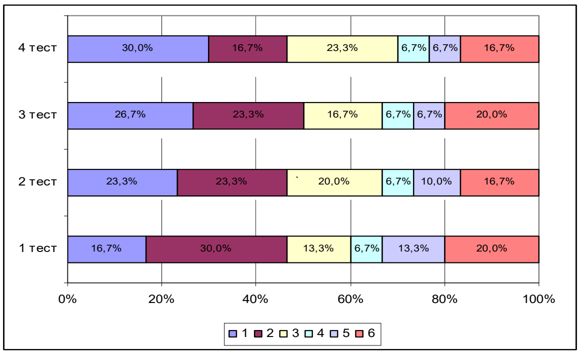
Резултати от изследването:

Задачите от тестовете са разпределени на 6 основни теми: Тема 1 – Алгебрични действия. Степенна, логаритмична и показателна функция. Прогресии; Тема 2 – Уравнения, неравенства, системи; Тема 3 – Планиметрия; Тема 4 – Стереометрия; Тема 5 – Тригонометрия; Тема 6 – Комбинаторика. Класическа вероятност. Граница на функция. Производна на функция.

На Фиг. 3.11 и Фиг. 3.12 е показана структурата на тестовете по броя на задачите от различните теми за четирите теста.

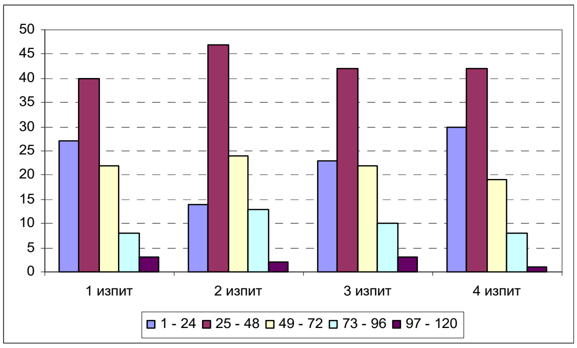


Фиг. 3.11 Брой задачи по теми

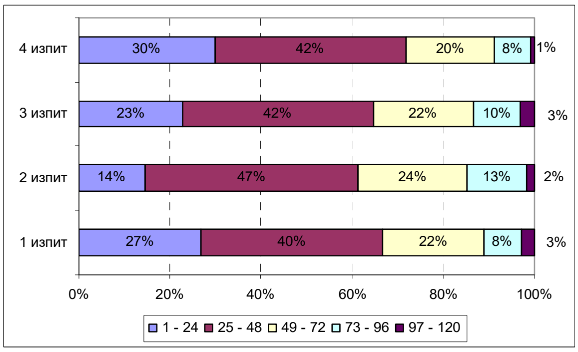


Фиг. 3.12 Структура на теста по теми

Баловете на кандидат-студентите за всеки от тестовете са групирани в пет групи с еднаква ширина на интервала. На Фиг. 3.14 са построени сравнителни хистограми на абсолютните честоти, приведени към 100 кандидата по групи за всеки от тестовете. На Фиг. 3.15 е направено сравнение между тестовете, чрез линейни диаграми за процентното съотношение на кандидатите по групи .

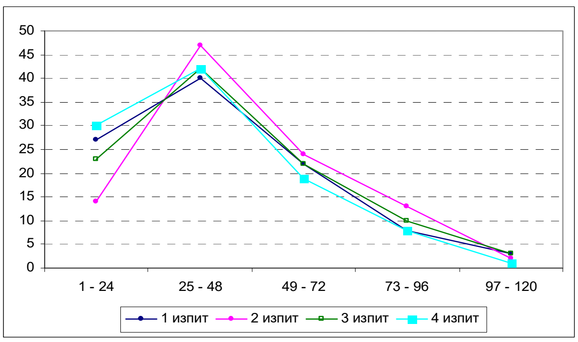


Фиг. 3.14 Абсолютните честоти

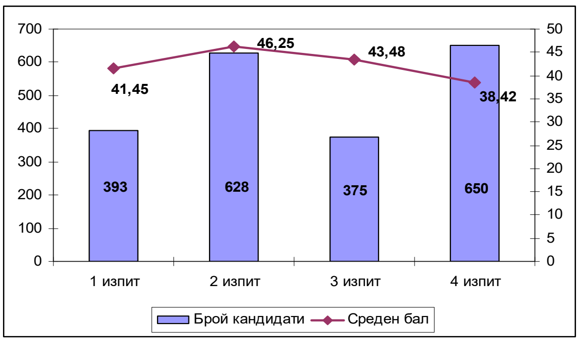


Фиг. 3.15 Процентното съотношение

На Фиг. 3.16 са съпоставени полигоните за различните тестове по групи. На Фиг. 3.17 са представени хистограма според броя кандидати, положили изпитите и полигонът за средния получен бал от тях.



Фиг. 3.16 Сравнение на полигоните на различните тестове по групи



Фиг. 3.17 Брой кандидати и среден бал по тестове

Стойностите на средното, медианата и модата за четирите теста са близо и между тях съществуват неравенствата $M_o < M_e < \bar{X}$, което означава, че разпределенията им са дясно изтеглени асиметрични.

Сравнителен анализ на качествата на задачите от тестовете показва, че задачите от четвърта тема „Стереометрия" са много трудни за кандидатите във всички години (тестове). Причината за това е, че в задължителните програми по математика в средния курс се изучава по-малко стереометрия. За повишаване разграничителната способност на теста, задачите от тема „Стереометрия" трябва да са с по-малка трудност. Пета тема „Тригонометрия" и шеста тема „Комбинаторика. Класическа вероятност. Граница на функция. Производна на функция" също се оказват трудни. Това може да се обясни с факта, че част от задачите от тези теми са извадени от задължителните програми в средния курс и се изучават само в профилираните паралелки. От тези наблюдения могат да се изведат следните изводи: Първо, в теста трябва да се сведат до минимум задачите, които не се изучават в задължителните програми. Второ, програмите по висша математика трябва да се съобразят с тази липса и да я запълнят, а не да се предполага, че този материал е изучаван в средния курс.

Извод на изследването: Проведените корелационен и дисперсионен анализ показват, че различните тестове измерват едно и също количество знание от кандидатите, като отделят еднакви групи от кандидати за прием във ВТУ „Тодор Каблешков". Тестовете са подходящо съставени и осъществяват първоначалните си цели.

Част от резултатите на този параграф са публикувани в [27].

В последните четири параграфа е изложено **Изследване 10.**

Цел на изследването е да се приложат различните тестови теории и логистични модели върху тест със структурата на националния изпит-тест за прием след завършен 7 клас в държавните и в общинските училища и да се направи съпоставителен анализ между тях. За целта на изследването е използван нормативният подход, както при изготвянето на теста, така и при интерпретацията на получените резултати.

Задачи на изследването:

1. Да се създаде подходящ инструментариум (тест);
2. Да се направи анализ на теста и интерпретация на получените резултати, основани на класическата теория на тестовете;
3. Да се направи анализ на теста и интерпретация на получените резултати, чрез използване на еднопараметричния модел на Раш;
4. Чрез използване на квадратичен метод да се определи оптималната дължина на теста и на оптимално време за изпълнението му;
5. Да се направи анализ на теста и интерпретация на получените резултати, чрез използване на двупараметричния модел на Бирнбаум;
6. Да се направи сравнителен анализ между еднопараметричния модел на Раш и двупараметричния модел на Бирнбаум;
7. Да се оценят съответно параметрите на трудност на задачите в еднопараметричния модел на Раш и параметрите на трудност и дискриминация на

задачите в двупараметричния модел на Бирнбаум, чрез приложение на логит–моделите, които са разгледани в параграф 8 на Първа глава;

8. Да се направи сравнителен анализ на едноименните модели, получени по двете методики за определяне на параметрите им;

9. Да се направи съпоставителен анализ на едноименните параметри на задачите, получени при Класическата теория на тестовете, Теорията на вероятностното моделиране и Теорията на логит моделите.

Инструментариум на изследването:

1. Тест по математика, съобразен с Държавните образователни изисквания за учебното съдържание и изпитната програма по математика за 7 клас. Тестът съдържа 40 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 правилен отговор. Продължителността на теста е 180 минути.

2. Анкетна карта, съдържаща 10 основни въпроса с различни възможности за отговор. Засегнати са въпроси за допълнителните занимания с математика, за отношението на родителите и учениците към математиката, за честотата на даваните домашни, контролни и изпитвания по математика и други. Времето за попълване на анкетната карта е приблизително 15 минути.

Методиката на изготвяне на инструментариума е съобразена с утвърдената практика в образованието.

Място на изследването: Изследването е проведено успоредно в ПЧМГ, СМГ„Паисий Хилендарски“, 38 ОУ „Васил Априлов“ и 45 ОУ „Константин Величков“.

Обект на изследването: В изследването участват общо 302 ученици от седми клас от различни училища в град София. 249 от тях изучават математика само в часовете по задължителна подготовка (ЗП) – 4 часа седмично. 53 ученици (от математически паралелки) изучават математика и под формата на: задължително избираема подготовка (ЗИП) – 2 часа седмично; свободно избираема подготовка (СИП) – 3 часа седмично.

Период на изследването: Изследването е проведено две седмици преди националния изпит-тест за прием след завършен 7 клас в държавните и в общинските училища.

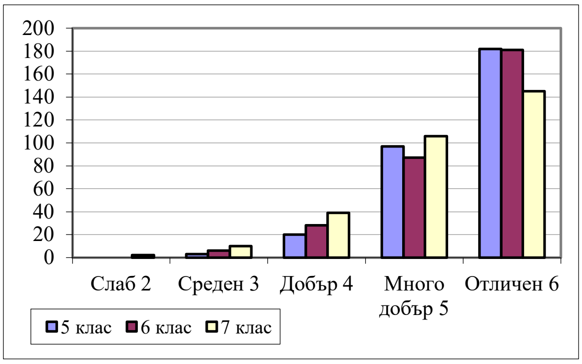
Хипотези за резултата от изследването: Всяка от тестовите теории може да се прилага самостоятелно, но за по-добър анализ и съдържателни резултати е по-добре да се използват успоредно няколко теории.

Методи на теоретичното изследване: Класическата Теория на Тестовете (КТТ); Еднопараметричен „модел на Раш“; Квадратичният вероятностен модел; Двупараметричен „модел на Бирнбаум“; Логит-модели; Сравнителен анализ; Корелационен анализ; Дисперсионен анализ.

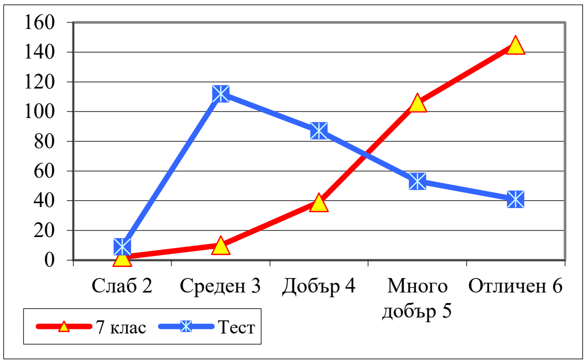
В **трети параграф** е направен анализ на качествата на задачите от теста, основан на Класическата Теория на Тестовете. Пресметнати са коефициентите на трудност на задачите, дискриминация на задачите, различни видове корелации между задачите. Оценени са надеждността на теста и стандартна грешка на измерването. Пресметнати са различни коефициенти за оценка на надеждността на теста: корелация между задачите, разделяне на теста на две, Алфа на Кронабах. Извършен е анализ на получените резултати. Направен е сравнителен анализ по отделни групи – пол, тип училище. Оценено е влиянието на различните фактори, като брой изучавани часове, брой допълнителни часове и самостоятелно занимание. Направен е корелационен анализ за зависимост на получените резултати и оценки по математика на учениците в предишни периоди за различните училища. Направено е заключение за обективността на тези оценки. В изследването е анализирана необходимостта от независимо външно оценяване.

Изводи от направените анализи: Изпитният тест е с добри измерителни качества и висока степен на надеждност, задачите са добре корелирани помежду си и са с оптимална трудност и много добра дискриминация.

На Фиг. 3.24 и Фиг. 3.25 са представени сравнителна хистограма на оценките по години и сравнителен полигон на оценките от теста и годишната оценка от 7 клас.



Фиг. 3.24 Сравнение на хистограмите



Фиг. 3.25 Сравнение на полигоните

Сравнението на получените резултати на учениците по видове паралелки е дадено в таблица 3.24.

Таблица 3.24 Резултати на учениците по видове паралелки

	Брой ученици	Среден брой решени задачи	Среден бал	Максимален бал	Минимален бал
Математически паралелки	53	40	76	95	38
Обикновени паралелки	249	24	43	89	14
Общо	302	27	49	95	14

Полигонът на резултатите от теста е изтеглен вдясно, за разлика от полигона от оценките в 7-ми клас, който е изтеглен в ляво. Това показва, че учениците са надценени в училище и има необходимост от външно оценяване. В анкетата над 92% от учениците посочват, че са правилно оценени или подценени, т.е. те нямат реална представа за знанията си.

В обикновените паралелки учениците изучават математика само в часовете по задължителна подготовка (ЗП) – 4 часа седмично. В математически паралелки учениците изучават математика 9 часа седмично под различни форми (ЗП – 4 часа, ЗИП – 2 часа, СИП – 3 часа). Сравнителната таблица за резултатите на учениците от математическите паралелки и тези на учениците от обикновените паралелки показват по-добър резултат на първите, т.е. налице е положителен ефект на профилираната подготовка по математика върху знанията и уменията на учениците. Това се дължи на подбора на учениците в тях, на средата, начина на преподаване и броя часове по математика.

Можем да твърдим, че за успешното представяне на изпита са необходими повече часове по математика под различна форма.

Сравнението на резултатите по пол и направеният t-тест не дават съществени различия между половете.

Част от резултатите на този параграф са публикувани в [5].

В **четвърти параграф** е приложен ***еднопараметричният модел на Раш***, който е част от вероятностното моделиране или IRT (Item Response Theory).

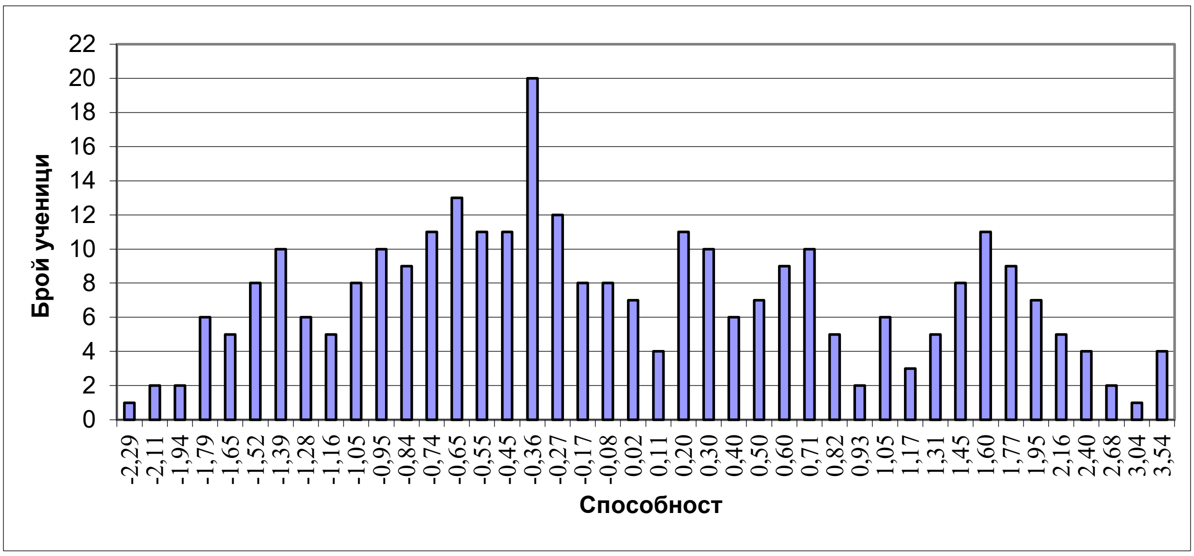
Оценени са способностите на учениците и трудността на задачите в теста.

В таблица 3.25 са дадени средните статистически величини и средноквадратичното отклонение, съответно на оценките на способността на учениците, стандартната грешка, първичния и действителния бал.

Таблица 3.25 Основни статистики на способността на изпитваните

	Оценка на способностите на изпитваните	**Стандартна грешка на оценката на способността**	**Първичен бал**	**Действителен бал**
Средна стойност	0,05	0,38	26,57	24,58
Средноквадратично отклонение	1,17	0,08	10,27	13,10
Максимална стойност	3,54	0,84	48	49,65
Минимална стойност	-2,29	0,33	7	3,50

На Фиг. 3.28 е представена хистограма на честотното разпределение на учениците по способност.



Фиг. 3.28 Разпределение на учениците по способност

В изследването са участвали ученици с равнище на способност от – 2,29 до 3,54 логита. Средната способност на ученици е 0,05. Модата (стойността с най-голяма честота) на способността е – 0,36. Медиана (средната по положение) на разпределението е равна на – 0,17.

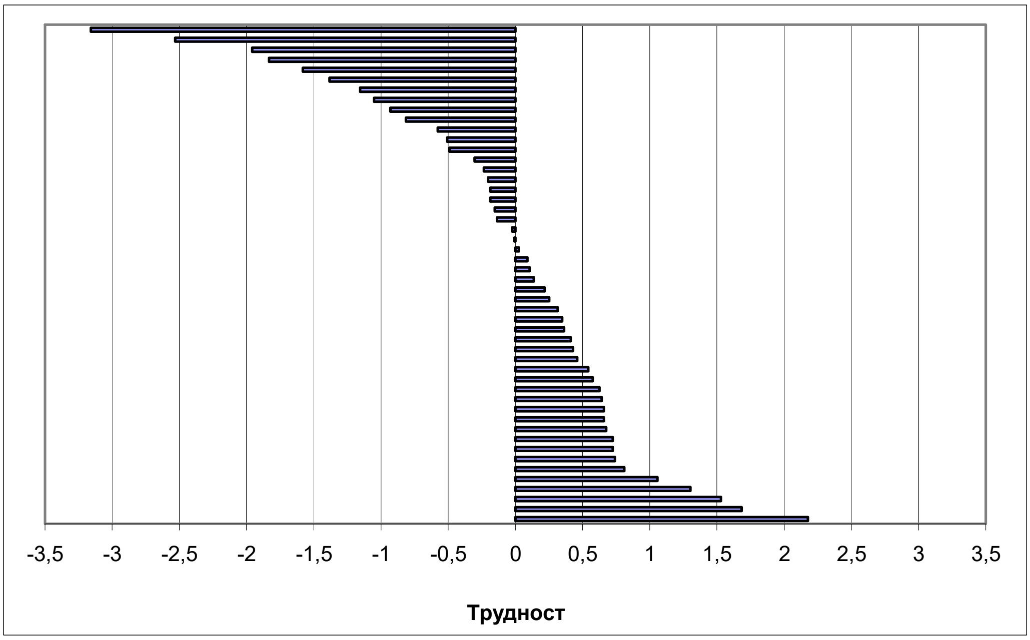
В таблица 3.26 са дадени средните статистически величини и средноквадратичното отклонение съответно на процента решени задачи, оценките на трудността на задачите в теста, стандартната грешка.

Задачите в теста са с равнище на трудност от – 3,16 до 2,18 логита. Средната трудност на задачите е – 0,02.

Таблица 3.26 Основни статистики на трудността на задачите

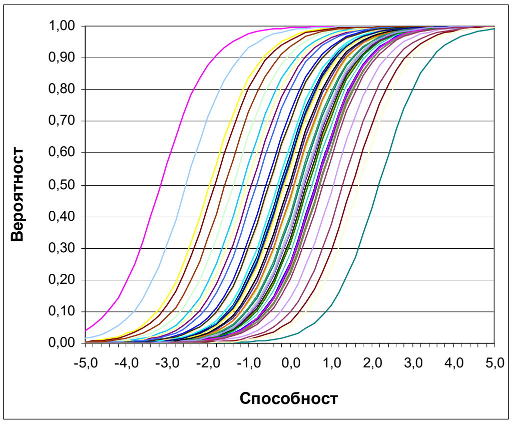
	Процент решени задачи	**Оценка на трудността на задачите**	**Стандартна грешка на оценката на трудността**
Средна стойност	53,15%	-0,02	0,15
Средноквадратично отклонение	17,44%	1,03	0,03
Максимална стойност	94,04%	2,18	0,29
Минимална стойност	16,23%	-3,16	0,14

На Фиг. 3.30 е представено разпределение на задачите на теста по трудност.

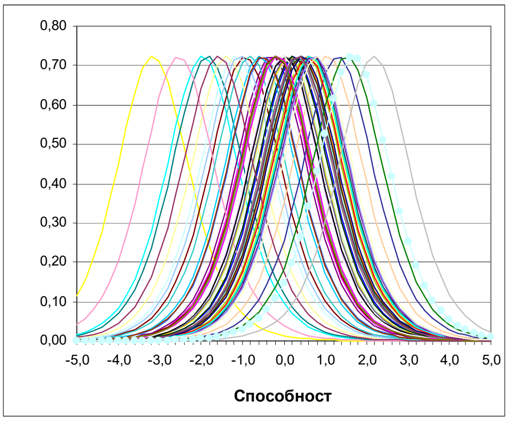


Фиг. 3.30 Разпределение на задачите по трудност

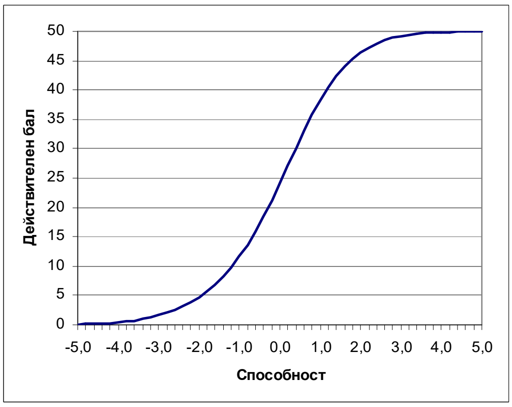
Пресметнати са вероятностите за правилно решаване на задачите от теста. Показани са графиките на характеристичната и информационната функция, както на целия тест, така и по задачи.



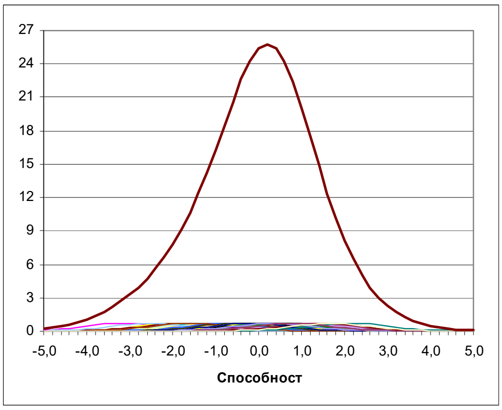
Фиг. 3.31 Характеристични криви на задачите в теста



Фиг. 3.32 Информационни криви на задачите в теста

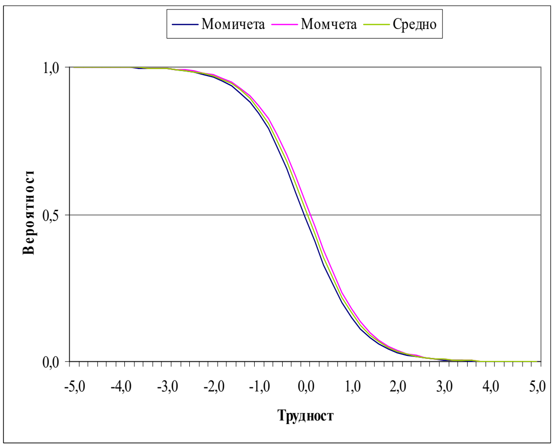


Фиг. 3.29 Характеристична крива на теста

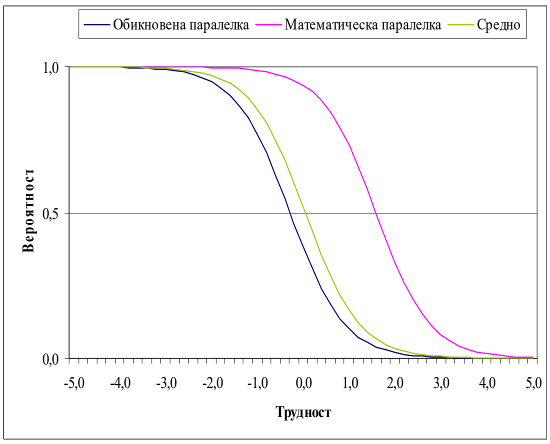


Фиг. 3.33 Информационна крива на теста

Представени са графиките на характеристичната функция, както на средното за всички ученици, така и за учениците по пол и по вид паралелка.



Фиг. 3.34 Характеристични криви на учениците по пол



Фиг. 3.35 Характеристични криви на учениците по вид паралелка

Определена е **оптималната дължина на теста** (броят на задачите, включени в теста).

Тест, който не е с оптимална дължина понижава своите измерителни качества. Например, при тест с малка дължина и лесни задачи, учениците с малки способности и учениците с големи способности, ще имат един и същ действителен бал, т.е. тестът ще има слаба разграничителна сила. Ако тестът е с голяма дължина, учениците с малки способности няма да се справят с по-трудните задачи поради слаба подготвеност, а учениците с големи способности – поради липса на достатъчно време. За определяне на оптималната дължина на теста е използван квадратичен метод с изходно уравнение $\sigma = \beta_0 + \beta_1 k + \beta_2 k^2 + \varepsilon$, където k е дължината на теста, а σ - стандартните отклонения на стойността на вероятността за правилен отговор на една задача при тази дължина на теста. За оптимална дължина на теста е избрана тази дължина, при която се достига максимум на стандартното отклонение на стойността на вероятността. В тази точка (на максимум) разграничителната сила на теста е най-голяма.

Резултати:

Етап	**Очаквана стойност**	**Точка на максимум на параболата**
Без тегло	$\widehat{\sigma}_i = 0,0459 + 0,0120.k_i - 0,0002.k_i^2$	$k_0 = 37,75$, $\widehat{\sigma}(k_0) = 0,273$
С тегло	$\sigma_i^* = 0,0347 + 0,0131.k_i - 0,0002.k_i^2$	$k_0^* = 36,78$, $\sigma_i^*(k_0^*) = 0,275$

При първия етап (без тегло) оптималната дължина на теста е 37,75 или закръглено до цяло число 38, тогава стандартното отклонение ще е най-голямо 0,273. При втория етап (с тегло) оптималната дължина на теста е 36,78 или закръглено до цяло число 37, тогава стандартното отклонение ще е най-голямо 0,275.

Дължината на един тест и времето за неговото изпълнение са тясно свързани. Изходният тест се състои от 50 задачи и време за изпълнение 180 минути. Средното време за решаване на една задача е 3 минути и 36 секунди. Ако умножим оптималната дължина на теста по средното време, необходимо за решаване на една задача, ще получим: 37 x (3 минути и 36 секунди) ≈ 133 минути.

Извод: Оптималният тест ще има дължина 35 задачи и време за изпълнение 120 минути.

В **пети параграф** е приложен ***двупараметричният модел на Бирнбаум***.

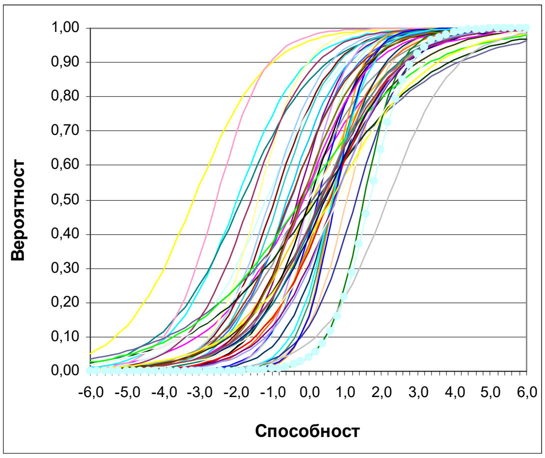
Оценена е дискриминацията на задачите в теста. В таблица 3.35 са дадени средните статистически величини и средноквадратичното отклонение, съответно на оценките на дискриминация на задачите в теста.

Таблица 3.35 Основни статистики на дискриминацията на задачите

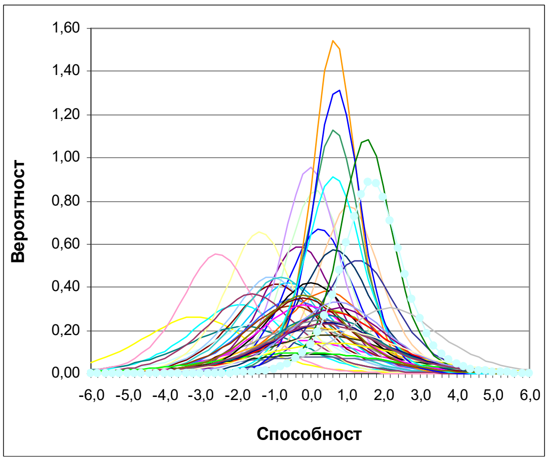
	Точково-бисериален коефициент	**Бисериален коефициент**	**Оценка на дискриминацията на задачите**
Средна стойност	0,438	0,575	0,747
Средноквадратично отклонение	0,102	0,126	0,264
Максимална стойност	0,652	0,826	1,463
Минимална стойност	0,244	0,306	0,322

Дискриминацията на задачите в теста се изменя от 0,322 до 1,463. Средната дискриминация на задачите е 0,747.

Пресметнати са вероятностите за правилно решаване на задачите от теста. Показани са графиките на характеристичните и информационните функции на задачите.



Фиг. 3.39 Характеристични криви на задачите в теста



Фиг. 3.40 Информационни криви на задачите в теста

По-стръмните характеристични криви съответстват на задачите с по-голяма дискриминация. Максимумът на информационната функция на задачата се получава в точката, която е равна на трудността и стойността му е по-голяма при задачите с по-голяма дискриминация, дори да е еднаква трудността им. Следователно по-високите криви съответстват на задачи с по-голяма дискриминация.

Извършено е сравнение на вероятностните модели на Раш и Бирнбаум.

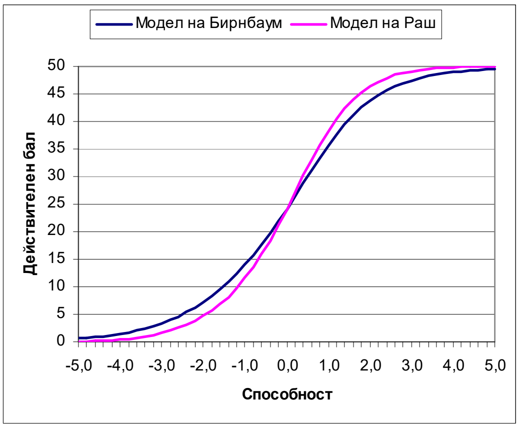
В таблица 3.37 е показано сравнение между първичния бал и действителния бал, пресметнат съответно по моделите на Раш и Бирнбаум.

Таблица 3.37 Сравнение между първичния бал и действителния бал на теста

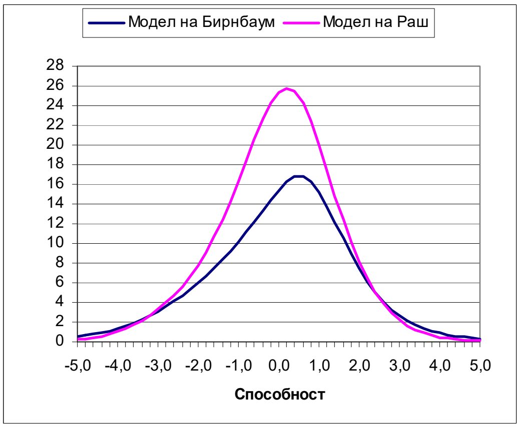
	Оценка на способностите на изпитваните	**Първичен бал**	**Действителен бал - Раш**	**Действителен бал - Бирнбаум**
Средна стойност	0,05	26,57	24,58	24,68
Средноквадратично отклонение	1,17	10,27	13,10	11,12
Максимална стойност	3,54	48	49,65	48,45
Минимална стойност	-2,29	7	3,50	5,86

Забелязва се, че моделът на Бирнбаум по-добре се приближава до първичните данни относно средноквадратичното отклонение и при граничните стойности на способностите.

На Фиг. 3.41 и Фиг. 3.42 са сравнени съответно характеристичните и информационните криви на теста, получени при двата модела.



Фиг. 3.41 Сравнение на характеристични криви на теста от двата модела



Фиг. 3.42 Сравнение на информационните криви на теста от двата модела

Моделът на Раш дава възможност да се оцени трудността на задачите, независимо от изпитваните и да се оцени способността на изпитваните независимо от набора на задачите. Наличието само на един параметър опростява изчисленията.

Предимствата на модела на Бирнбаум са, че той дава възможност да се диференцират задачите по трудност, а също така дава възможност да се определят и евентуално да се заменят тези от тях, които не могат достатъчно добре да разграничат изпитваните според способностите им.

Всеки от тези два модела има и свои недостатъци. Например при модела на Раш това са твърдите (непроменливите) условия на свойствата на използваните задачи. При модела на Бирнбаум е възможно при един и същи брой верни отговори да се получи различен тестови бал.

В **шести параграф** са приложени ***логит–моделите***, които са разгледани в параграф 8 на първа глава, за да се оценят съответно параметрите на трудност на задачите в еднопараметричния модел на Раш и параметрите на трудност и дискриминация на задачите в двупараметричния модел на Бирнбаум.

Сравнени са резултатите на едноименните модели, получени по двете методики (начини) за определяне на параметрите им. I начин – методиката описана в параграф 4, II начин – чрез използване на логит–модели.

Еднопараметричен модел на Раш

Таблица 3.37 Сравнение между първичния бал и действителния бал на теста

	Оценка на способностите на изпитваните	**Първичен бал**	**Действителен бал – Раш (I начин)**	**Действителен бал – Раш (II начин)**
Средна стойност	0,05	26,57	24,58	26,31
Средноквадратично отклонение	1,17	10,27	13,10	13,23
Максимална стойност	3,54	48	49,65	49,76
Минимална стойност	-2,29	7	3,50	3,49

Можем да заключим, че при втория вариант има по-голяма дискриминация около средната трудност на теста и този вариант дава по-близки средни стойности до средните стойности на първичния бал.

Двупараметричен модел на Бирнбаум

Таблица 3.44 Сравнение между първичния бал и действителния бал на теста

	Оценка на способностите на изпитваните	Първичен бал	Действителен бал – Бирнбаум (I вариант)	Действителен бал – Бирнбаум (II вариант)
Средна стойност	0,05	26,57	24,68	26,24
Средноквадратично отклонение	1,17	10,27	11,12	7,97
Максимална стойност	3,54	48	48,45	45,07
Минимална стойност	-2,29	7	5,86	10,81

От таблица 3.44 се вижда се, че при втория вариант е по-малко средноквадратичното отклонение. Този модел дава по-близки средни стойности до средните стойности на първичния бал.

В **седми параграф** е направен съпоставителен анализ на едноименните параметри на задачите, получени при Класическата теория на тестовете, Теорията на вероятностното моделиране и Теорията на логит моделите.

В таблица 3.45 са дадени основните статистики на трудността на задачите в теста, пресметнати по различните теории.

Таблица 3.45 Сравнение на коефициентите на трудност на задачите

	Коефициент на трудност - КТТ (стандартизирани стойности)	Коефициент на трудност (IRT)	Коефициент на трудност Раш - Логит	Коефициент на трудност Бирнбаум - Логит
Средна стойност	0,00	-0,02	-0,16	-0,17
Средноквадратично отклонение	1,00	1,03	0,92	1,05
Максимална стойност	2,12	2,18	1,91	2,54
Минимална стойност	-2,34	-3,16	-2,82	-2,78

Между оценките се наблюдават изключително високи коефициенти на корелация (от 0,975 до 0,99). Можем да заключим, че между оценките на трудност, получени по трите теории се наблюдава висока степен на съгласуваност и в този смисъл са взаимнозаменяеми.

В таблица 3.47 са дадени средните величини и средноквадратичното отклонение на коефициентите на корелация и дискриминация на задачите в теста, пресметнати по различните теории.

Таблица 3.47 Сравнение на коефициентите на дискриминация на задачите

	Коефициент на дискри-минация (КТТ)	**Точково-бисериален коефициент (Коефициент на Пирсън)**	**Бисериален коефициент**	**Коефициент на дискри-минация (IRT)**	**Коефициент на дискри-минация Логит**
Средна стойност	0,51	0,44	0,58	0,75	0,41
Средноквадратично отклонение	0,15	0,10	0,13	0,26	0,13
Максимална стойност	0,82	0,65	0,83	1,46	0,69
Минимална стойност	0,15	0,24	0,31	0,32	0,16

Измерителите на дискриминацията са по-малко устойчиви от тези на трудността. Стойностите по IRT са по-високи от тези на КТТ, които от своя страна са по-високи от точково–биселарния коефициент на корелация. Стойностите на дискриминация пресметнати по Логит–модела са близки до бисериалния коефициент на корелация.

Можем да заключим, че между оценките на дискриминация, получени по трите теории, се наблюдава определена степен на съгласуваност, макар и не така добре изразена, както при оценките на трудността.

Обобщение на резултатите от трета глава

В сравнение с класическата теория IRT моделите имат своите предимства и ограничения. Главните ***предимства*** на IRT моделите са: Оценките на параметрите на задачите, определени чрез IRT моделите, не зависят от избора (способността) на оценяваните лица; Оценката на способностите на оценяваните лица не зависи от равнището на трудност на теста като цяло и равнището на трудност на отделните тестови задачи; Посредством IRT моделите лесно се установява съответствие между равнището на трудност на задачите и равнището на способност на изпитваните, което повишава качеството на измерването; Равнището на способност на оценяваните лица и равнището на трудност на задачите се представят на една и съща скала; IRT моделите позволяват да се намират локални коефициенти на надеждност за тестовите резултати на оценявани лица със зададено равнище на способност, въз основа на информационната функция. ***Ограниченията*** на IRT моделите са: За тях са необходими по-строги предположения, отколкото при използване на класическата теория; IRT моделите изискват добра математическа подготовка, поради което тази теория по-слабо се разбира от част на педагогическата колегия; Необходима е по-голяма извадка на оценяваните лица за обосноваване на оценките на параметрите.

За определяне на надеждността и валидността на един тест, значителна роля има класическата теория на тестовете. Тя е полезна и за решаване на следните задачи от педагогическите изследвания и измервания: разработка на спецификациите и съдържателната рамка на теста, създаване на банка от тестови задачи и тяхната експертна оценка, апробация и анализ на тестовите задачи, конструиране на окончателен вариант на теста и др.

Съчетаването на двете тестови теории води до повече съдържателни резултати, отколкото тяхното самостоятелно използване.

Логит–моделите, които въведохме и разгледахме в параграф 8 на Първа глава и използвахме в параграф 6 на настоящата глава, имат своите предимства. При тях по-лесно и едновременно се оценяват съответно параметрите на трудност на задачите в еднопараметричния модел на Раш и параметрите на трудност и дискриминация на задачите в двупараметричния модел на Бирнбаум. Коефициентите на корелация между параметрите, получените при Логит–моделите и съответстващите им параметри от IRT моделите са много високи, което е едно доказателство за тяхна приложимост за изследвания и анализи в образованието.

ЗАКЛЮЧЕНИЕ

Представен е математическият апарат, необходим за изчислителната страна на статистическите методи и вероятностните модели, използвани в дисертацията. За всеки от моделите подробно са описани етапите и стъпките за неговата реализация. Многомерните модели са представени и във векторен вид. За всеки от тях са дефинирани съответно понятията: критична права, критична равнина или критична хиперравнина.

Всички изчисления и диаграми са реализирани чрез програмата за електронни таблици Microsoft Excel. Графиките са реализирани чрез обектно-ориентираната програма CorelDraw. Голяма част от изчисленията биха могли да се осъществят и със специализирани пакети за статистическа обработка като Statistica и SPSS, както и със специализираните програми за анализ на тестови задачи като Xcalibre4.1.8 (Item Parameter Estimation Software), RUMM (Rasch Unidimensional Measurement Model) и др.

В изпълнение на поставените в дисертацията задачи са проведени 10 изследвания с различни групи ученици и студенти. Всяко от изследванията е представено с неговите основни етапи и компоненти: от избора на обекта и предмета, през целта и задачите, формулиране на хипотеза на конкретното изследване, изработване на инструментариум, осигуряване на теоретичен метод, избор на технология за анализ и тълкуване на получените резултати.

Направените от всяко изследване изводи дават основание да се приеме верността на хипотеза на дисертационния труд: „Познатите от социално-икономически области вероятностни модели за оценка на риска и техни модификации могат успешно да се прилагат за изследвания и анализи в образованието както самостоятелно, така и в комбинация с класическата теория на тестовете и IRT-моделите“.

АВТОРСКА СПРАВКА

Основни приноси на дисертацията са:

1. Реализиране на идеята за използването в образованието на описаните в първа глава: Линеен вероятностен модел; Логистичен вероятностен (Logit) модел; Пробит (Probit) модел;
2. Въведен е квадратичен (полиномиален) вероятностен модел и са дадени негови приложения в образованието за определяне на оптималната продължителност на занятието и оптималната дължина на теста;
3. Предложен е петпараметричен вероятностен модел и са дадени примери за приложението му в образованието;

4. Определянето на параметрите на задачите чрез сравняване на еднопараметричния и двупараметрия модел със съответни логистични модели.
5. Прилагането на вероятностните модели от Първа глава във висшето образование и сравнителния анализ между тях;
6. Прилагането на двуфакторните вероятностни модели в образованието, векторният им вид и определянето на критичната права за вероятността;
7. Многомерното обобщение на логистичния и пробит вероятностни модели, определянето на критичната хиперравнина и сравняване с модела на Раш и модела на Бирнбаум;
8. Използването на Класическата теория на тестовете (КТТ) за оценка на кандидат-студентски изпити във ВТУ „Тодор Каблешков“, включително и сравнителен анализ между няколко изпита, както и изводите и препоръките, които се правят за нивото на подготовка по математика в средния курс на обучение и за броя часове, отделени за някои теми;
9. Приложението на различните тестови теории и логистични модели за анализ на тест по математика за ученици от 7 клас и интерпретацията на получените резултати;
10. Определянето на оптималната дължина на теста и продължителността за изпълнението му, чрез модел от тестове с различна дължина, като се отчита големината на отклонението;
11. Сравнителният анализ между модела на Раш и модела на Бирнбаум с едноименните им модели, когато параметрите са определени чрез съответни логистични модели;
12. Извършеният съпоставителен анализ на едноименните параметри на задачите, получени при Класическата теория на тестовете, Теорията на вероятностното моделиране и Логит моделите от първа глава.

Практическите приложения на дисертационния труд, които вече са осъществени или са в процес на реализация са:

1. Определяне продължителността на занятията в подготвителните курсове в ПЧМГ. От 2014 година, във връзка с проведеното изследване, продължителността на занятията са по 3 учебни часа.
2. Определяне оптималната дължина (броят на задачите) в теста за прием в 5 клас на ПЧМГ. В резултат от проведеното изследване и направените изводи, от 2017 година ПЧМГ промени формата на приемния си изпит, като броят на задачите в него съответства на получените от изследването резултати.
3. В резултат и от направения във Втора глава експеримент и анализ на изпита по Висша математика 2 част, от учебната 2016 – 2017 година, бяха извършени промени в съдържанието на учебните програми, във формата на изпитните теми и в начина на формиране на крайната оценка по предметите Висша математика 1 част, Висша математика 2 част, Висша математика 3 част, Висша математика 4 част и Статистика.
4. От направения в Трета глава сравнителен анализ на резултатите от кандидат-студентските изпити по математика за ВТУ „Тодор Каблешков” и направените изводи, през 2016 година от приемния изпит по математика бяха премахнати задачите (темите), които се изучават само в профилираните паралелки. За да се

компенсират някои от тези теми и да се даде равен шанс на всички студенти, от учебната 2016 – 2017 година, в програмата по Висша математика 1 част е въведен нов модул „Увод във висшата математика”. Извършена беше актуализация и на програмата на кандидат-студентските курсове по математика.

5. Разработените от автора задачи за самостоятелна подготовка (домашни работи) по Висша математика 2 част се използват успешно от всички преподаватели по математика във ВТУ „Тодор Каблешков“. Поради мобилността на преподаватели от ВТУ „Тодор Каблешков“ към други университети, авторът е получил отзиви за тяхното успешно приложение и в други университети.

Други възможни практически приложения на дисертационния труд са:

Дадени са препоръки към студентите от ВТУ „Тодор Каблешков” за успешно полагане на изпита по Висша математика 2 част. На основа на тези препоръки, те могат да изградят оптимална стратегия за подготовката си за полагане на изпита.

Дисертационният труд може да бъде полезен за преподавателите и изследователите в областта на образованието и сродни области. Доказателство за това е, че част от публикациите по дисертацията са вече цитирани в две успешно защитени дисертации за присъждане на образователната и научна степен „Доктор“ в областта на педагогиката и образованието.

Дисертационният труд може да се използва като ръководство за анализ на резултати от изпити и тестове.

Списък на публикациите по темата на дисертацията и участия на конференции

Авторът на дисертационния труд Райна Милкова Алашка има следните научни публикации по дисертацията:

1. Алашка Р.М. Статистически анализ за оценка качеството на изпитен тест. Механика, транспорт, комуникации, *научно списание*, брой 3, София, 2009, стр.IX-19 –IX-24
2. Алашка Р.М. Статистически анализ на тест по математика за ученици и оценка за влиянието на различни фактори. Механика, транспорт, комуникации, *научно списание*, брой 3, София, 2011, стр.IX-29 –IX-33
3. Алашка Р.М. Приложение на логистични вероятностни модели в образованието за оптимално определяне на параметрите в моделите на Раш и на Бирнбаум. Механика, транспорт, комуникации, *научно списание*, том14, брой 3/2, 2016
4. Алашка Р.М. Двуфакторни вероятностни модели в образованието. Механика, транспорт, комуникации, *научно списание*, том14, брой 3/2, 2016
5. Алашка Р.М., Михалев Д.Й. Основни числови характеристики на реален изпитен тест. Механика, транспорт, комуникации, *научно списание*, брой 3, София, 2009, стр.IX-25 –IX-30
6. Михалев Д.Й., Алашка Р.М. Сравнителен анализ на реални изпитни тестове. Механика, транспорт, комуникации, *научно списание*, брой 3, София, 2009, стр.IX-31 – IX-35
7. Алашка Р.М., Михалев Д.Й. Параметрични вероятностни модели. Механика, транспорт, комуникации, *научно списание*, , том14, брой 3/2, 2016

Първите четири публикации са самостоятелни, а другите публикации са в съавторство с доц. д-р Драго Михалев.

Публикациите са докладвани на **международни научни конференции**: „ТРАНСПОРТ 2009” , „ТРАНСПОРТ 2011” и „Механика, транспорт, комуникации 2016”.

Освен тези публикации авторът на дисертационния труд има, в съавторство с доц. д-р Драго Михалев, издаден учебник и справочник по близка тематика, съответно със заглавия:

8. Михалев Д.Й., Алашка Р.М. Теория на вероятностите и статистика. София, 2012, 312 стр.;
9. Алашка Р.М., Михалев Д.Й. Формули и таблици по вероятности и статистика. ВТУ „Тодор Каблешков”, София, 2014, 106 стр.

Авторът на дисертационния труд има дългогодишен опит в разработването и съставянето на тестове както за подготовка, така и за пробни и приемни изпити за ученици. Автор и съавтор е на учебни помагала, част от които са:

10. Паскалева З., Алашка М., Алашка Р. – Национално външно оценяване математика – 4 клас, Учебно помагало, Издателство Архимед, 2010
11. Паскалева З., Алашка М., Алашка Р. – Национално външно оценяване математика – 5 клас, Учебно помагало, Издателство Архимед, 2010
12. Паскалева З., Алашка М., Алашка Р. – Национално външно оценяване математика – 6 клас, Учебно помагало, Издателство Архимед, 2010
13. Паскалева З., Алашка М., Алашка Р.– Национално външно оценяване и прием след 7 клас (нов формат на изпита) – математика, Част 1, Учебно помагало, Издателство Архимед, 2012
14. Паскалева З., Алашка М., Алашка Р.– Национално външно оценяване и прием след 7 клас (нов формат на изпита) – математика, Част 2, Учебно помагало, Издателство Архимед, 2012
15. Паскалева З., Алашка М., Алашка Р.– Задачи по формата PISA с решения, Учебно помагало, Издателство Архимед, 2012
16. Алашка Р.– Тестове за прием в ПЧМГ – 4 клас, Учебно помагало, Издателство ПСМГ, 2014
17. Алашка Р.– Тестове за прием в ПЧМГ – 3. и 4. клас, Учебно помагало, Издателство ПСМГ, 2017

Декларация за оригиналност

Декларирам, че настоящият дисертационен труд **„Приложение на вероятностни модели за анализ на резултати от изпити и тестове”** е изцяло мой авторски продукт и в неговото разработване не са ползвани чужди публикации и разработки в нарушение на авторските им права.

ЛИТЕРАТУРА

1. *Аванесов В.С. Композиция тестовых заданий. 2 изд., испр. и доп., - Москва, Адепт, 1998, 217 стр.*
2. *Аванесов В.С. Форма тестовых заданий. - Москва, изд. Центр тестирования, 2005, 156 стр.*
3. *Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. - Москва, 1989, 167 стр.*
4. Алашка Р.М. Статистически анализ за оценка качеството на изпитен тест. Механика транспорт комуникации, *научно списание*, брой 3, София, 2009, стр.IX-19 –IX-24
5. Алашка Р.М. Статистически анализ на тест по математика за ученици и оценка за влиянието на различни фактори. Механика транспорт комуникации, *научно списание*, брой 3, София, 2011, стр.IX-29 –IX-33
6. Алашка Р.М., Михалев Д.Й. Основни числови характеристики на реален изпитен тест. Механика транспорт комуникации, *научно списание*, брой 3, София, 2009, стр.IX-25 –IX-30
7. Алашка Р.М. Приложение на логистични вероятностни модели в образованието за оптимално определяне на параметрите в моделите на Раш и на Бирнбаум. Механика транспорт комуникации, *научно списание*, том14, брой 3/2, 2016
8. Алашка Р.М. Двуфакторни вероятностни модели в образованието. Механика транспорт комуникации, *научно списание*, том14, брой 3/2, 2016
9. Алашка Р.М., Михалев Д.Й. Параметрични вероятностни модели. Механика транспорт комуникации, *научно списание*, , том14, брой 3/2, 2016
10. Алашка Р.М., Михалев Д.Й. Формули и таблици по вероятности и статистика. ВТУ „Тодор Каблешков", София, 2014, 106 стр.
11. *Анастази А., Урбина С. Психологическое тестирование.-Спр. пособ.: Питер, 2006. -688 стр.*
12. Банков К.Г. Широко мащабни оценъчно-диагностични педагогически изследвания – Хабилитационен труд за академична длъжност професор, София, 2012
13. Банков К.Г. Методи за оценяване на постиженията на ученици. Съвременната педагогическа реалност – резултати и очаквани постижения. Сборник с материали от научно-практическата конференция, 29-30.10.2009
14. *Банков К.Г. Измерване и оценяване на постиженията на учениците.* В. „Азбуки"*, брой 48, 2009*
15. Банков К.Г. Увод в тестологията, Издателство „Изкуство", София, 2012, 140 стр.
16. *Банков К.Г. Качествена интерпретация на резултатите от тестове за постижения. Математика и математическо образование – тридесет и първа пролетна конференция на СМБ, 2002*
17. *Великова Евг., Григорова Д., Славчова-Божкова М. Анализ и влияние на приема с матура в кандидат-студентската кампания 2009 във факултета по математика и информатика на СУ „Климент Охридски". Математика и математическо образование – тридесет и девета пролетна конференция на СМБ, 2010.*
18. Въндев Д. Л. Записки по приложна статистика 1, СУ, София, 2003.
19. Въндев Д. Л. Записки по приложна статистика 2, СУ, София, 2003.
20. *Гатев К., Гатева Н. Самоучител по статистика. София, УНСС, Издателство „Стопанство", 2000, 232 стр.*
21. Калинов Кр. Статистически методи в поведенческите и социални науки. 2 преработено и допълнено издание, Нов Български Университет, София, 2010, 569 стр.
22. Ким В.С. Тестирование учебных достижений. УГПИ, Уссурийск, 2007, 214 стр.
23. *Ким В.С. Анализ результатов тестирования в процесе Rasch measurement // Педагогические измерения, №4, 2005, стр 39-45.*
24. *Ким В.С. Развивающая функция тестовых заданий. // Педагогические измерения, №1, 2007, стр 77-84.*
25. *Львовский Е.Н. Статистические методы построения эмпирическх формул. Москва, Высшая школа, 1988, 240 стр.*

26. *Магранова Ю.В. Теория тестированиния как основа оценивания уровня знаний в современной системе образования. стр. 186-210.*
27. Михалев Д.Й., Алашка Р.М. Сравнителен анализ на реални изпитни тестове. Механика транспорт комуникации, *научно списание*, брой 3, София, 2009, стр.IX-31 –IX-35
28. Михалев Д.Й., Алашка Р.М. Теория на вероятностите и статистика. София, 2012
29. *Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. Москва, 2000.-168 стр.*
30. *Нейман Ю.М., Хлебников В.А. Педагогическое тестирование как измерение. Москва, Центр тестирования МО РФ, 2002.-67 стр.*
31. *Переверзев В.Ю. Технология разработки тестовых заданий. Справ. Руководство, Москва, Е-Медиа, 2005. – 265 стр.*
32. *Роберт Ван Криген, Стивен Баккер. Подготовка и проведение экзаменов. Руководство для организации и разработки ценрализованных экзаменов. – Амхем, Нидерланды, 1995.*
33. Стоименова Е.А. Измерителни качества на тестове. Нов Български Университет, София, 2000, 176 стр.
34. Съйкова Ив. Д., Стойкова-Къналиева Адр. Ст., Съйкова Св. Ст.. Статистическо изследване на зависимости. Унив. изд. „Стопанство", София, 2002, 453 стр.
35. Съйкова Ив. Д., Тодорова С. Статистическо изследване (постановка, методи и анализ на резултатите). НБУ, София, 2000.
36. *Тодорова С. Статистика в икономиката и бизнеса (методи, решения и изпитни тестове). НБУ, София, 2004.*
37. *Успенский А.Б., Федоров В.В. Вычислительные аспекты метода наименьших квадратов при анализе и планировании регресионных экспериментов. Москва, изд. МГУ, 1976.*
38. *Феськов Н.С., Якобчук А.П. Математические методы интерпретации результатов нормативно-ориентированвого тестирования. „Адукацыя і выхаванне", №3.-2007.*
39. *Химмельблау Д. Анализ процессов статистическими методами. М., Мир, 1973.*
40. *Хлебников В. А. Теоретические основы обьективного измерения учебных достижений учащихся. Москва: Федеральное государственое учреждение- Федеральный центр тестирования, 2005, 127 стр.*
41. *Челышкова М.Б. Организация контроля учебной деятельности студентов в условиях педагогического сотрудничества: Дис., Киев, 1990.*
42. *Челышкова М.Б. Разработка педагогических тестов на основе современных математических моделей. Москва, МИСИС, 1995, 195 стр.*
43. *Челышкова М.Б. Теория и практика конструирования педагогических тестов. Учебное пособие.- Москва, Логос, 2002, 432 стр.*
44. Шеффе Д. Дисперсионный анализ. Москва, Физматгиз, 1963.
45. *Ширяев А.Н. Статистический последовательный анализ. Оптимальные правила остановки. Москва, Наука, 1976.*
46. *Щиголев Б.М. Математическая обработка наблюдений. Москваа, Наука, 1969.*
47. *Яснопольский С.Л. Построение эмпирических формул и подбор их параметров методом наименьших квадратов и методом средних. Москва, изд. МИСиС, 1972.*
48. *Abedi J. The interrater/test reliability system (ITRS). Multivariate Behavioral Research, 1996, 31, 4, p 409-417.*
49. *Adedoyin O., Nenty H., & Chilisa B. Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. Educational Research and Review, 2008, vol.3 (2), p 83-93.*
50. *Andrich, D., Sheridan, B., Lyne, A & Luo, G. RUMM: A windows-based item analysis program employing Rasch unidimensional measurement models (Parth: Murdoch University), 2000.*
51. Baker F. B. The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, 2001, 2-nd ed.
52. Barton M.A., Lord F.M.. An upperasymptote for the three-parameter logistic item-response model. Princeton,N.J.: Educational Testing Service, 1981.
53. Bechger T., Maris G., Verstralen h. & Beguin A. Using classical test theory in combination with item response theory. Applied Psychological Measurement, 2003, 27(5), p 319-334.

54. Birnbaum A., Some Latent Trait Models and Their Use in Inferring an Examinee’s Ability. In F.M. Lord and M.R. Novick. Statistical Theories of Mental Test Scores. Readinf Mass.: Addison-Wesley, 1968. Ch. 17-20. –p 397-479.
55. Birnbaum M.H., editor. Measurement, Judgement and Decision Making. Academic Press, 1997.
56. *Bolt D.M., Cohen A.S. & Wollack J.A. A mixture item response model for multiple-choice data. Journal of Educational and Behavioral Statistics, 2001, 26(4), p 381-409.*
57. *Camilli G., Shepard L.A. Methods for Identifyinc Biased Test Items, Sage,1994.*
58. Carson R.T., Yixiao Sun. The Tobit model with a non-zero threshold. Royal Economic Society , Econometrics Journal, volume 10, 2007, pp 488-502.
59. Crocker L. Algina J. Introduction to Classical and Modern Test Theory. Harcourt Brace, 1986.
60. *Cronbach L. Essentials of psychological testing. NY: Harper & Row, 1977.*
61. *Everitt B.S. Making Sense of Statistics in Psychology. 1996.*
62. Fisher W. The Standard Model in the History of Natural Sciences. Econometrics and the Social Sciences. Journal of Physics: Conference Series, 2011, p 238.
63. *Hambleton R.K. Aplication of Item Response Theory. Vancouver, Ed. Res. Inst. B. C., 1983.*
64. *Holland P., Hoskens M. Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. Psychometrika, 2003, Vol.68, 1, p 123-149.*
65. Iman, Ronald L. et al. Modern Business Statistics, N. Y., 1989.
66. Johnston J. Econometric Methods. N.Y., 1963.
67. Keeves J.P. (Ed.). Educational Reserch, Methodology and Measurement: An International Handbook. Oxford, Pergamon Press, 1988.
68. *MacCann R., Stanley G. The Use of Rasch Modeling to Improve Standard Settings. Practical Assessment, Research & Evaluation, 2006, 11*
69. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, 1960, Danisch Institute of Educational Research. (Expanded edition, Chicago, 1980, The University of Chicago Press).
70. *Rasch G. On objectivity and specifity of the probabilistic basis for testing. In: Rasch Lectures. In honor of Georg Rasch’s 100 years birthday on Semtember, 2001.*
71. *Roid G., Haladyna T. A Technology for Test-Item Writing. Academic Press, 1981.*
72. Sazberger T. Does the Rasch Model Convert an Ordinal Scale into an Interval Scale? Rasch Measurement Transactions SIG American Educational Research Association, 2010, 24(2), p 1273-1275.
73. *Suen H.K. Principles of Test Theories. Erlbaum, 1990.*
74. Tobin J. Estimation on relationships for limited dependent variables. Econometrica 26, 1958, p 24-36.
75. *Traub R.E. Reliability for the Social Sciences: Theory and Appplications. Sage, 1984.*
76. Tukey J. Methodology and the Statistician’s Responsibility, J. Am. Stat. Ass., 1979, vol.74.
77. *Vandev D.L., Neykov N.M. About Regresion Estimators with Hight Breakdown Point, Statistics, 1998, p 111-129.*
78. *Wright B.D., Stone M.H. Best test design, Chicago, : Rasch Measurement. Mesa Press, 1979, 220 p.*
79. Wright B.D., IRT in the 1990s: Which Models Work Best? // Rasch Measurement Transactions, 1992, 6:1, p 196-200.
80. *Zeidner M., Most R. (editors). Psychological Testing. An Inside view. Consulting Psychologists Press, 1992.*
81. Zuehlke T. Estimation of a Tobit model with unknown censoring threshold. Applied Economics 35, 2003, p 1163-1169.

Забележка:

В литературата са включени всички заглавия, които са цитирани в дисертацията. Тези от тях, които не са цитирани в автореферата са с наклонен шрифт. Броят на цитираните заглавия в автореферата е 36.

БЛАГОДАРНОСТИ

Искрено благодаря на:

научния ми ръководител проф. Кирил Банков, който провокира в мен интерес към тестологията и ми даде много ценни съвети;

доц. Драго Михалев за добрите идеи, за полезните съвети, за отделеното време и оказаната помощ;

инж. Марияна Влъчкова, директор на Първа Частна Математическа Гимназия, която осигури условията за експериментиране в ПЧМГ и за голямото доверие, което ми гласува;

проф. Огнян Касабов за ценните препоръки и оказаната помощ;

проф. Даниела Тодорова, ректор на ВТУ „Тодор Каблешков“, за оказаната финансова подкрепа.

Благодарна съм и на колегите ми във ВТУ „Тодор Каблешков“ за проявеното разбиране, както и на всички ученици и студенти, участвали в експериментите.

Благодаря на семейството ми и на приятелите ми, които се лишиха от вниманието и помощта ми през последните години.