УНИВЕРЗИТЕТ У БЕОГРАДУ

МАТЕМАТИЧКИ ФАКУЛТЕТ

ПРОЦЕНА ВАЛИДНОСТИ УСКЛАЂИВАЊА РАЗЛИЧИТИХ НАЧИНА ОЦЕЊИВАЊА МАТЕМАТИЧКИХ ТЕСТОВА

МАСТЕР РАД

Аутор

Лазар Јовановић

Ментор

проф. др Милан Божић

БЕОГРАД

2014.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

САДРЖАЈ

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Апстракт

Циљ овог истраживања био је досезање резултата који би показали да ли различити начини оцењивања математичких тестова утичу на валидност тестирања. У раду смо, применом различитих модела теорије ставског одговора (IRT - Item Response Theory), поредили успех испитаника на тесту на два различита начина оцењивања. За потребе овог истраживања, којим је обухваћено 84 испитаника првог разреда гимназије изабраних у узорак, направили смо четири теста у којима су дата иста питања, али у различитом редоследу њихових тежина. Испитали смо да ли постоји корелација између укупних вредности на тесту за различите начине оцењивања и да ли се и колико укупан успех разликује, као и значај тих разлика. Такође, одредили смо способност испитаника, као и тежине питања. Приказана су питања која су испитаници лако решавали, као и питања која нису могли лако да реше. Методом теорије ставског одговора одредили смо различите параметре питања (тежина питања, дискриминативност питања и погађање питања). Након што смо одредили параметре, показали смо да не постоје статистички значајне разлике међу параметрима питања у ова два различита начина оцењивања теста. Тестирали смо три различита модела IRT, користећи наведене параметре. У зависности од добијених резултата, закључили смо да је 3PL модел IRT најосетљивији на разлике у оцењивању јер нуди највише вредности параметара погађања, док су преостала два модела IRT мање осетљива на разлике у оцењивању.

Кључне речи: валидност теста, начин оцењивања, параметри питања, IRT модели.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Abstract

The purpose of the survey was reaching results that would show whether different ways of grading math tests affect the test validity. We compared the success of the examinees on the test, using different models of Item Response Theory in two different test estimation ways. The research involved 84 participants of the first year of high school, selected in the sample. For the purpose of this survey, we made four tests with the same questions, given in a different order, depending on their difficulty. We examined whether there is a correlation between the total values of the test by different estimation ways, if and how much total success differs and the significance of these differences also. In addition, we determined the ability of examinees, as well as questions difficulty. We presented the questions that examinees could easily solve, as well as the questions that could not be easily resolved. Using Item Response Theory Model we have set various question parameters (question difficulty, question discriminative and question guessing). When we set these parameters, we have shown that there are no statistically significant differences between the question parameters of these two different ways of the test estimation. We tested three different IRT model using these parameters. Depending on the results, we concluded that the 3PL IRT model is the most sensitive to differences in the estimation because it offers the highest parameter guessing value, while the other two IRT models are less sensitive to differences in the estimation.

Keywords: test validity, estimation way, question parameters, IRT models.

1 Теорија ставског одговора (IRT)

Теорија ставског одговора (IRT - Item Response Theory) је савремени начин статистичког тестирања, тренутно најпопуларнија област активног истраживања, која се, такође, понекад назива и латентном теоријом особина. За разлику од класичне теорије тестирања, теорија ставског одговора (у даљем тексту: IRT) захтева јаче претпоставке и има много интуитивнији приступ мерењу. У IRT тачност се дефинише на основу латентних особина које меримо тестом. IRT даје теоретско оправдање за многе ствари које меримо тестом, а за које га класична тест теорија не може дати.

IRT моделира вероватноћу одговора на питања у зависности од присуства латентне црте код испитаника. На основу склопа одговора испитаника на сва питања у тесту оцењује се мера присуства латентне особине код тог испитаника, односно додељује му се мера – број. Тај процес додељивања бројева назива се *скалирањем*.

IRT обухвата општи скуп линеарних модела и придружених статистичких процедура. Помоћу модела IRT директно израчунавамо вероватноћу да ће појединачан испитаник одговорити позитивно на неко питање, односно да ће заокружити неку од алтернатива. Такође, IRT модели прецизно дефинишу начин израчунавања грешака као разлику између добијених и израчунатих вероватноћа. Те разлике, скалиране и трансформисане на погодан начин, служе да се оцени да ли је модел погодан да подржи податке.

Модели IRT су математичке једначине које описују везу између основног нивоа испитаника о латентној особини и вероватноће одређеног одговора, користећи нелинеарну монотону функцију на основу које добијамо карактеристичну криву.

Математички, сви IRT модели представљају се функцијом вероватноће:

$$\boldsymbol{P_{ij}} = \ (\boldsymbol{\theta}; \boldsymbol{b}; \boldsymbol{a}; \boldsymbol{c}; \dots) = \boldsymbol{P}\left(\boldsymbol{X_{ij}} = \mathbf{1} \mid \boldsymbol{\theta_i},\ \boldsymbol{a_j},\ \boldsymbol{b_j}, \boldsymbol{c_j}, \dots\ \right)$$

Ова функција служи да се израчуна вероватноћа неког позитивног одговора испитаника на појединачном питању у функцији параметра питања (тежина питања, дискриминативност и погађање). Постоји велика разноликост IRT модела у погледу избора и броја параметара. Параметри које смо навели су основни параметри IRT модела и најчешће се срећу. ([11] Фајгељ, 2005)

1.1 Почеци развоја IRT

Иако многи мисле да је IRT савремена психометријска теорија, концепти и методологије развијени су пре више од 80 година. Развој IRT започиње 50-их и 60-их година прошлог века. Оснивачи и почетници били су Фридрих Лорд, математичар Георг Раш и социолог Пол Лазард. Кључни људи који су наставили развој и прогрес IRT су Бенџамин Дрејк и Дејвид Андрић. Широка примена овог начина тестирања започиње 70-их и 80-их година прошлог века. Најчешћа примена је у образовању, где се користи ради развоја и усавршавања тестирања. ([03] Baker, 2001)

Концептуални основ за развој IRT даје Л. Л. Терстон (1925. године) у свом раду под називом „Метод скалирања психолошких и образованих тестова“. Терстон је радио на развоју вишефакторске анализе варијансе. Његове колеге, Ричардсон (1943) и Фергусон (1943), радили су на даљем

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

усавршавању IRT, те су развили шиљкасти лук модел као средство за приказивање исправне пропорције за појединачна питања у функцији нормализованих резултата. ([04] Partchev, 2004)

Лавли је 1943. године описао могућност процене процедура максималне веродостојности за параметре, као и линеарну процену тих параметара. Фридрих Лорд (1942) уводи идеју о латентној особини. У својим радовима за развој IRT Георг Раш (1960) говори о потреби за стварањем статистичких модела који ће мерити објективност оцењивања људи и параметара одвојено, али уз могућност њиховог поређења. Раш је инспирисао Герхарда Фишера (1968) да продужи примену његовог модела. У развој IRT укључују се и Дејвид Андрић, Џефри Мастерс, Даглас Грахам и Марк Вилсон. ([06] Chong Ho Yu, 2008)

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

1.2 Претпоставке једнодимензионалности и локалне независности

Основна претпоставка IRT је претпоставка једнодимензионалности, али су са њеним каснијим развојем, неки модели постали вишедимензионални, па је стога и ова претпоставка морала бити модификована. Сва питања у тесту могу имати само један заједнички предмет мерења, али ако постоји више предмета мерења, сви морају бити обухваћени примењеним IRT моделом.

У пракси не постоји питање, а ни тест, који би мерили само једну особину. У одговарању на питања увек учествују мотивација, погађање, вештина да се ради брзо... Статистички речено, када локална независност захтева да корелације између питања буду нулте, односно када се варијанса која потиче од параметра способности елиминише, тада одговори на било који пар питања постају статистички независни.

Претпоставка локалне независности гласи да *вероватноћа одговарања на једно питање не зависи од одговора на друга питања.* Ако различита питања мере још неке особине које нису обухваћене у оквиру параметра способности, тада ће између одговора и питања постојати неки необјашњиви односи. Када је претпоставка једнодимензионалности задовољена, задовољена је и претпоставка локалне независности, а ако претпоставка једнодимензионалности није задовољена, неће бити ни локалне независности. ([11] Фајгељ, 2005)

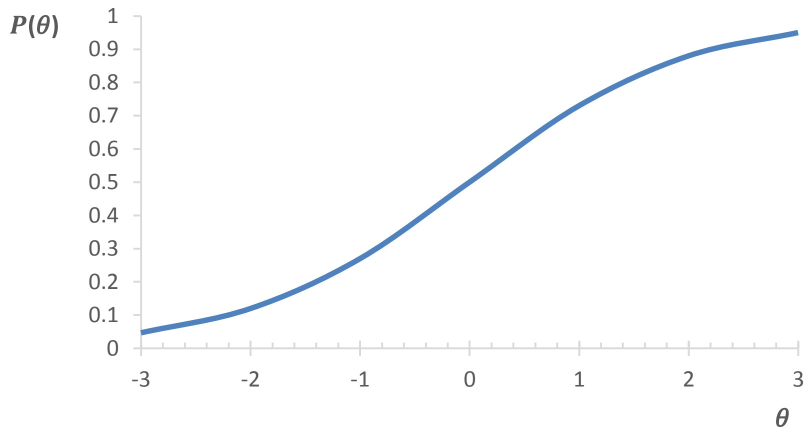
2 Карактеристична крива питања

Графички приказ IRT модела добија се исцртавањем функције вероватноће $P_{ij} = (\theta; b; a; c; \dots)$ и назива се *карактеристичном кривом питања*. Овај назив указује на то да се теорија ставског одговора базира на кривој одговора на питања.

Уколико неко жели да измери (оцени) количину латентне особине коју особа поседује, неопходно је да постоји скала мерења способности, уз помоћ које може да се дефинише обим мерења. Овакве скале способности имају средњу вредност нула и јединичне интервале који се крећу до позитивне и негативне бесконачности. Физички се може утврдити способност неке особе, међутим, предност ових скала је у мерењу специфичности особина способности. Такође, њима меримо и међусобни однос способности више особа. ([06] Chong Ho Yu, 2008)

Како се опсег скала способности креће од негативне до позитивне бесконачности, ради лакшег рачунања ми ћемо ограничити опсег рада. Уобичајен приступ који се узима за мерење способности јесте да се направи тест са што више питања. Свако од тих питања може мерити неки аспект одређене способности која нас интересује. Са техничке тачке гледишта, неопходно је да питања буду таква да испитанику пруже могућност да на њих одговори. Основна претпоставка је да сваки испитаник који одговара на питања у тесту поседује основна знања о области коју истражујемо.

Сваком испитанику можемо да придружимо неку нумеричку вредност на основу које га рангирамо на мерној скали, односно скали способности, и ту вредност обично означавамо са $\boldsymbol{\theta}$. На сваком нивоу способности постоји одређена вероватноћа да ће испитаник дати тачан одговор на дато питање и њу означавамо са $\boldsymbol{P(\theta)}$**.** Нивое способности означићемо са ***é***, а вероватноћу на том нивоу способности са ***P(é)***. Према томе карактеристичну криву неког питања можемо представити као глатку криву у облику латиничног слова S (**Графикон 1**). ([03] Baker, 2001)



Графикон 1: Карактеристична крива питања

Вероватноћа тачног одговора је близу нуле у најнижем нивоу способности, али се, како се крећемо ка вишим нивоима способности, она повећава и приближава јединици. Оваква глатка крива (S) описује однос вероватноће тачног одговора и способности испитаника. Стога ће свако питање у тесту имати своју карактеристичну криву. Може се рећи да је карактеристична крива питања регресија одговора питања на једну или више латентних варијабли које репрезентују психичку особину или особине које се мере. ([11] Фајгељ, 2005)

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Иако сви параметри утичу на облик карактеристичне криве питања, основне варијабле IRT модела су:

1. манифестно понашање, тј. одговор на питање – зависна варијабла
2. латентна црта која се мери – независна варијабла.

Ако питања имају различите нагибе, може доћи до укрштања њихових карактеристичних крива. Због тога је тешко рангирати испитанике и питања. Укрштање карактеристичних крива је логички погрешно, а до њега врло лако долази ако су питања различите дискриминативности. Ипак, оно се не одбацује увек у потпуности, већ се у неким моделима условно прихвата као нешто што пракса намеће.

Постоје и технички проблеми са оцењивањем параметра дискриминативности јер постојећи софтвери по правилу захтевају да се вредности параметра ограниче. У зависности од тих ограничења, оцене параметра биће мање или веће. ([12] Фајгељ, 2013)

2.1 Модели карактеристичне криве питања

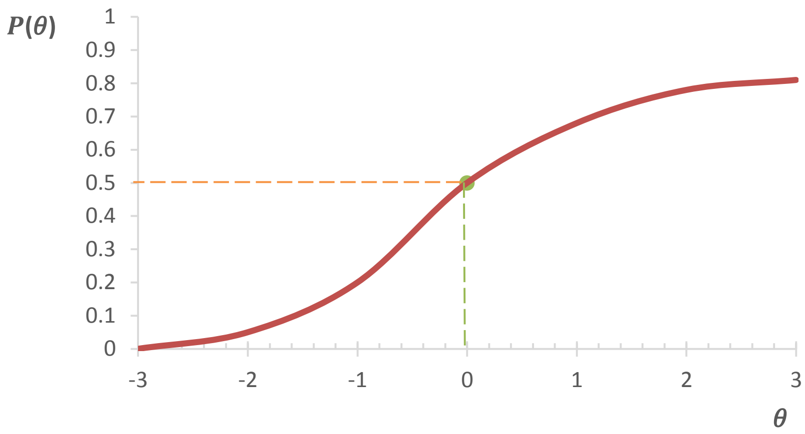
2.1.1. Једнопараметарски логистички модел (1PL)

Сваки модел IRT предвиђа вероватноћу да ће особа одговорити на одређено питање. Испитаници могу бити на различитим нивоима способности, па се стога питања могу разликовати по различитим аспектима, она која су тежа или она која су лакша. Ради лакшег праћења вероватноћу означавамо са $\boldsymbol{P_{ij}}$, тако да се индекс $\boldsymbol{i}$ односи на питање, а индекс $\boldsymbol{j}$ на особу. Са $\boldsymbol{P(\theta)}$ означићемо вероватноћу тачног одговора у зависности од функције способности $\boldsymbol{\theta}$. Међутим, P такође зависи од својстава питања. Питања из теста испитиваћемо помоћу модела IRT који имају један, два и три параметра, па ће тиме њихове вероватноће бити $\boldsymbol{P_{ij}(\theta_j, b_i)}$, $\boldsymbol{P_{ij}(\theta_j, b_i, a_i)}$ и $\boldsymbol{P_{ij}(\theta_j, b_i, a_i, c_i)}$, где су $\boldsymbol{a_i}$, $\boldsymbol{b_i}$, $\boldsymbol{c_i}$, параметри питања. Питања која имају више од две понуђене опције имаће такође параметре који су у вези са опцијама одговора. Најједноставнији модел IRT има само један параметар питања. Код једнопараметарске функције све њене вредности налазе се између 0 и 1. Једначина једнопараметарског логистичког модела је: ([03] Baker, 2001)

$$\boldsymbol{P_{ij}(\theta_i, b_i) = \frac{exp(\theta_j - b_i)}{1 + exp(\theta_j - b_i)} = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}}}$$

Параметар $\boldsymbol{b_i}$ назваћемо тежинским параметром. Најзанимљивији део ове формуле је $\boldsymbol{exp(\theta_j - b_i)}$ у бројиоцу (именилац је ту само да обезбеди да функција не постане мања од 0, а већа од 1). Ако се сконцентришемо на $\boldsymbol{exp(\theta_j - b_i)}$, примећујемо да једнопараметарски модел предвиђа да вероватноћа тачног одговора зависи од интеракције између индивидуалне способности $\boldsymbol{\theta_j}$ и параметра питања b_i.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 2: Заједничко представљање способности и тежине питања на истој оси

На **Графикону 2**, на истој оси представљени су заједно способност и тежина. Крећући се од координатног почетка надесно по скали нивоа особине $\boldsymbol{\theta}$, видимо да испитаници са ниском цртом имају малу вероватноћу тачног одговора. Они имају вероватноћу само случајно да погоде. У том делу криве велико повећање црте не доводи до повећања вероватноће тачног одговора. У средњој области криве мале промене способности доводе до драматичне промене у вероватноћи. Важна тачка на кривој је *тачка прелома* која пада на 50% вероватноће када се гледа ордината. Она се увек налази у тачки $\boldsymbol{b_i}$. Пошто се читава крива односи на једно питање, тачка прелома криве дефинише тежину питања – параметар тежине. Место карактеристичне криве на ординати одређује тежину питања: лево су лака, а десно тешка питања. Као што тежину питања дефинишемо преко нивоа црте код које постоји 50% тачних одговора, тако и неког испитаника можемо да дефинишемо преко тежине питања која он решава. На пример, испитаници који у око 88% случајева решавају питање тежине $-0{,}4$ биће за једну јединицу бољи од просечних испитаника. ([11] Фајгељ, 2005)

Код IRF (Item Response Function) у 1PL моделу функције које се односе на различита питања су паралелне и не секу се, а параметри различитих тежина само померају криву на лево или десно и облик остаје непромењен.

2.1.2. Двопараметарски логистички модел (2PL)

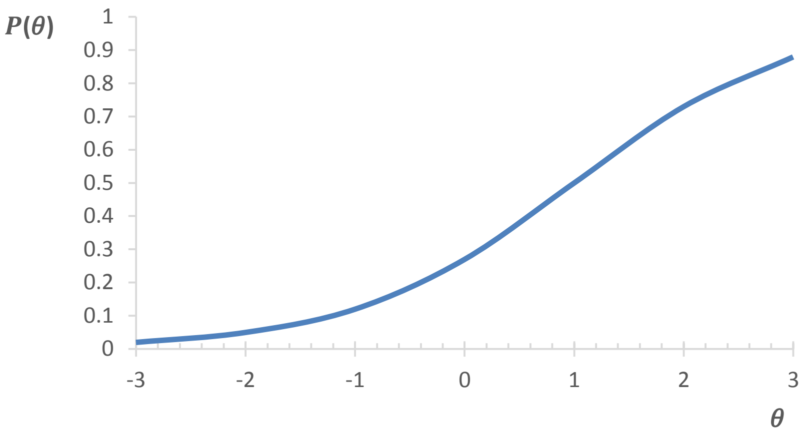
Двопараметарски логистички модел одређује вероватноћу давања тачног одговора у било којем тесту способности уз помоћ два параметра. Двопараметарски логистички модел се често назива *Birnbaum* моделом. За разлику од 1PL, вишепараметарски модели више су усмерени ка тачном скоровању – оцењивању особине, па се за њих обично каже да су модели података.

Једначина двопараметарског логистичког модела је: ([03] Baker, 2001)

$$\boldsymbol{P_{ij}(\theta_i, b_i, a_i) = \frac{exp[a_i(\theta_j - b_i)]}{1 + exp[a_i(\theta_j - b_i)]} = \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}}}$$

Разлика у односу на 1PL модел јесте у изразу $\boldsymbol{e^{[(\theta_j - b_i)]}}$, који је допуњен са $\boldsymbol{a_i}$ и сада је $\boldsymbol{e^{[a_i(\theta_j - b_i)]}}$. И овде ће $\boldsymbol{b_i}$ бити параметар тежине. Ако експонент $\boldsymbol{a_i(\theta_j - b_i)}$ представимо као $z_j = a_j(\theta_i - b_j) = a_j\theta_i - \delta_j$, где је $\delta_j = -a_j b_j$, добићемо регресиону једначину, где је $\boldsymbol{a_i}$ нагиб (регресиони коефицијент), а $\boldsymbol{\delta}_\mathrm{j}$ одсечак (регресиона константа).

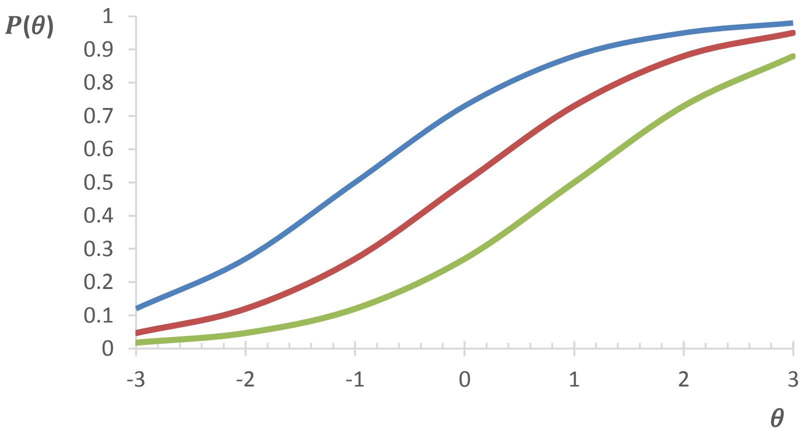
Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 3: Карактеристична крива 2PL модела

Захваљујући S облику криве (**Графикон 3**) нагиб криве се мења, као и функција на том нивоу способности, и може да достигне максималну вредност када је ниво способности једнак тежини питања. Због овога дискриминациони параметар не представља општи нагиб карактеристичне криве питања. У општем случају нагиб је $a/\mathbf{4}$ у $\boldsymbol{b} = \boldsymbol{\acute{e}}$. Теоријски опсег параметра је $-2.80 < a < +2.80$.

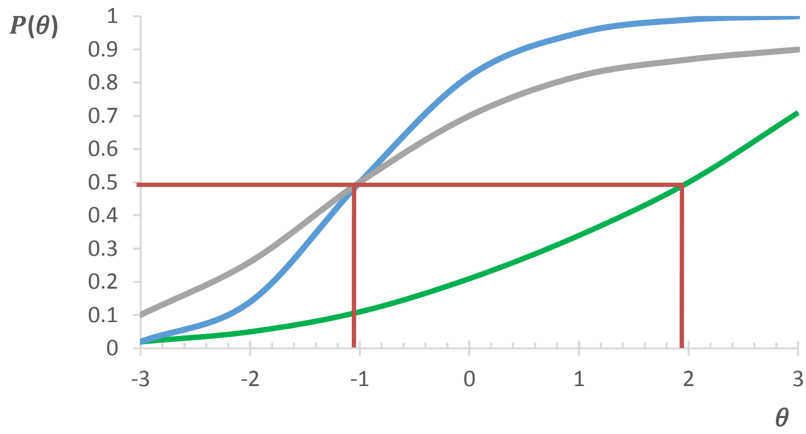
Дискриминација одражава стрмост карактеристичне криве питања, на основу које можемо видети колико испитаника има способност испод, а колико изнад просека питања. Дакле, стрмија крива одговара питању које боље дискриминише испитанике. Равнија крива даје да су разлике у тачним одговорима на нижим нивоима способности скоро исте као и на високим нивоима.



Графикон 4: Три питања са истим дискриминационим али различитим тежинским параметром

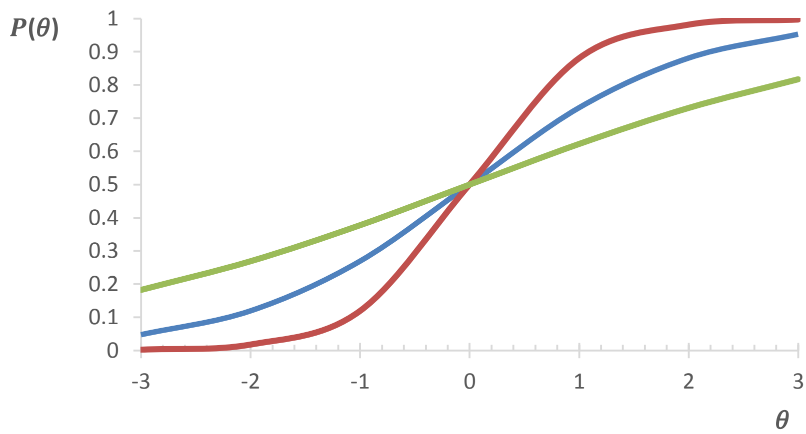
На **Графикону 4** дате су три карактеристичне криве са истим дискриминационим параметром али различитим тежинским параметром. Лева крива представља лако питање јер је вероватноћа тачног одговора већа од остале две криве за ниже нивое способности и приближава се јединици за високе нивое способности. Централна крива представља питање средње тежине јер је вероватноћа тачног одговора мала за ниже нивое, а у вишим нивоима се приближава јединици. Десна крива представља тешко питање. Вероватноћа је ниска за већину нивоа способности и једино расте када се достижу виши нивои. Још се може видети да је за више нивое способности вероватноћа тачног одговора 0.8, и то за најтежа питања.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 5: Криве различитих тежинских и дискриминационих параметара

На **Графикону 5** представљене су три криве, од којих две представљају питања која су исте тежине -1. Као и код 1PL модула, тежина се налази на нивоу способности где је вероватноћа 0.5. Плава крива је много стрмија од сиве криве из разлога што питање које представља плава крива има већи дискриминациони фактор од питања које представља сива крива. Што се тиче зелене криве, она има исти дискриминациони фактор као сива крива, али је померена у десно јер је параметар тежине овог питања већи. Може се приметити још да се плава и сива крива секу. Ово се не може десити у 1PL моделу. Разлог за то је и чињеница да је питање чија је сива крива теже за испитанике на ниском нивоу способности, док је питање чија је плава крива теже за испитанике на високим нивоима способности. ([03] Baker, 2001)

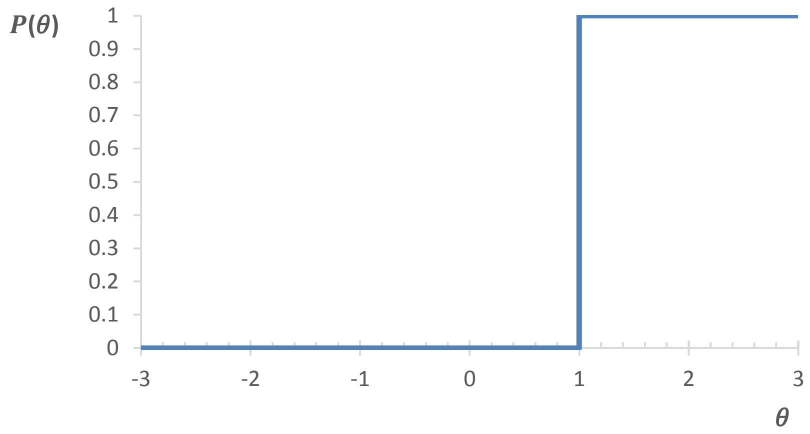


Графикон 6: Три питања са различитим дискриминационим а истим тежинским параметром

Графикон 6 представља три карактеристичне криве које имају исту тежину, али су различите у погледу дискриминације. Црвена крива има висок степен дискриминације јер је крива прилично стрма, а вероватноћа тачног одговора се повећавањем способности мења. Само на кратком растојању, лево од средине криве, вероватноћа тачног одговора је мања од 0.5, а десно, у односу на средину криве, вероватноћа је већа од 0.5. Плава крива представља питање са умереном дискриминацијом. Самим тим и нагиб криве ће бити мањи у односу на претходну криву, а вероватноћа тачног одговора ће, како се способност буде повећавала, имати драстично мање промене. Вероватноћа тачног одговора је близу нуле за нижу способност, а близу јединице за више нивое способности испитаника. Зелена крива

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

представља питање са најмањом дискриминацијом. Крива има врло мали нагиб, а вероватноћа тачног одговора се мења полако, како се пролази дуж скале способности. ([03] Baker, 2001)



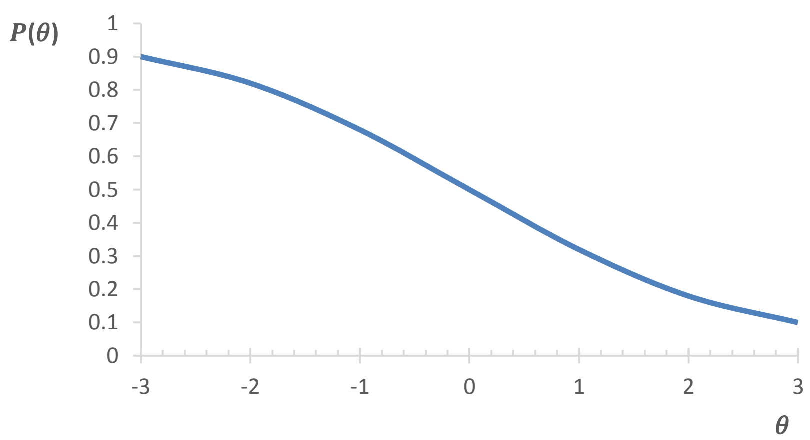
Графикон 7: Идеална дискриминација питања

На **Графикону 7** може се видети да је, у случају када је $\theta < 1$, вероватноћа давања тачног одговора 0, док је у случају када је $\theta > 1$ вероватноћа давања тачног одговора 1. Ово значи да питање савршено дискриминише испитанике чије су способности изнад и испод резултата способности. Оваква питања не праве разлику између испитаника са способностима изнад 1, као ни разлику између испитаника испод 1. ([03] Baker, 2001)

На основу свега овога, конструктор теста би требало да узме у разматрање питање које је драстично дискриминисано или оно које је мање дискриминисано, и да одлучи да ли је уопште такво питање пожељно за одређену сврху и да ли ће тим питањем измерити оно што жели.

2.1.3. Негативна дискриминација

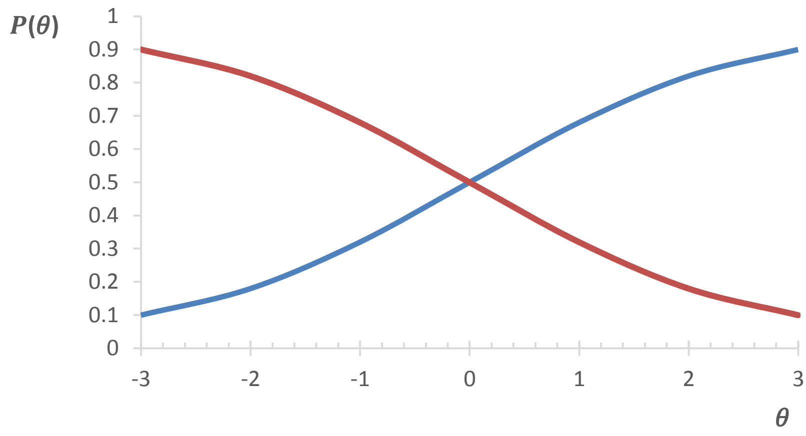
Код већине тестова, дискриминациони параметар има позитиван нагиб (тј. вероватноћа тачног одговора се повећава како се повећава ниво способности). Ипак, дискриминације могу бити и негативне и код таквих питања вероватноћа тачног одговора опада када способност расте од најнижег до највишег нивоа способности.



Графикон 8: Негативна дискриминација питања

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питања са негативном дискриминацијом (**Графикон 8**) јављају се на два начина. Прво, нетачан одговор на питање са два избора (нпр. под a и b) увек ће имати негативан дискриминациони параметар ако је тачан одговор имао позитивну вредност. Друго, понекад ће тачан одговор на нека питања дати негативан индекс дискриминације. Ово нам говори да нешто није у реду са питањем: или је лоше формулисано или постоји нека дезинформација која преовладава код испитаника високе способности. У оба случаја треба додатно обратити пажњу на питање. За већину тема које нас интересују у IRT, параметар дискриминације биће позитиван. ([05] Chong Ho Yu, 2013)



Графикон 9: Позитивна и негативна дискриминација питања

2.1.4. Тропараметарски логистички модел (3PL)

При решавању задатака једна од важних чињеница је да испитаници могу да добијају тачне одговоре погађањем. Вероватноћа, у овом случају, укључује мали број компонената што је последица погађања. Ниједан од претходна два модела не узима феномен погађања у разматрање. Бирнбаум је 1968. године модификовао двопараметарски логистички модел који укључује параметар који доприноси погађању. Такав модел је назван *тропараметарским логистичким моделом*, иако технички и није логистички модел. Једначина за тропараметарски логистички модел ([03] Baker, 2001) ([04] Partchev, 2004) је:

$$\boldsymbol{P_{ij}(\theta_i, b_i) = c + (1-c)\frac{e^{(a_i(\theta_j - b_i))}}{1 + e^{(a_i(\theta_j - b_i))}}}$$

Параметар $\boldsymbol{c}$ представља вероватноћу добијања тачног одговора погађањем. Вредност параметра c не варира у зависности од нивоа способности. Према томе, испитаници најниже и највише способности имају исту вероватноћу да погађањем одговоре тачно на питање. Он се креће у распону $0 < c < 1$, али се вредности изнад 0.35 не сматрају прихватљивим, па се узима $\boldsymbol{\acute{e}} \; < \; c \; < \; 0.35$. Ефекат употребе параметарског погађања **c** је да се дефиниција параметра тежине мења. У претходна два модела $\boldsymbol{b}$ је била тачка у којој је вероватноћа тачног одговора 0.5, али је сада доња граница карактеристичне криве питања вредност $\boldsymbol{c}$, а не 0. Резултат овога је да је параметар тежине тачка на скали способности у којој је:

$$\boldsymbol{P(\acute{e}) = c + (1-c) * (0.5) = {}^{(1+c)}/_{2}}$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Ова вероватноћа је на пола пута између вредности c и 1. Оно што се овде десило је да параметар c дефинише најнижу вредност вероватноће тачног одговора. Дакле, параметар тежине дефинише тачке на скали способности, где је вероватноћа тачног одговора на пола пута између ове границе и 1. Параметар дискриминације је пропорционалан нагибу карактеристичне криве у тачки $\acute{e} = b$. Код 3PL модела нагиб карактеристичне криве питања је заправо $\acute{e} = {(1-c)}/{4}$.

2.1.4.1 Погађање и 3PL модел

Погађање је параметар питања који зависи од различитих фактора (нпр. формулација питања). Неки сматрају да је погађање првенствено понашање лоших испитаника на тешким питањима. IRT третира погађање као експлицитно и егзактно израчунљиво мерно својство питања. Својства питања и испитаника су увек преплетена, тако да се у IRT параметри питања одређују на основу одговора испитаника.

3PL модел покушава да се прилагоди погађању увођењем трећег параметра c. Међутим, није баш уверљиво да се повеже понашање (природа) погађања са карактеристиком питања. Неки људи имају већу склоност ка погађању неких ствари за разлику од других. Са друге стране, постоје ситуације у којима је свако приморан да погађа. ([04] Partchev, 2004)

Чак и ако одлуку о давању неодговора испитаник жели донети рационално, он се мора ослањати на неке процене, а не на извесне факте. Као прво, испитаник мора да процењује своје знање. Ако је сасвим сигуран да не зна одговор, његова рационална одлука би требало да буде да ништа не заокружи. Ту процену није лако донети, барем из једног разлога. Наиме, познато је да је општа популација оптимистична у погледу властитог знања. Већина испитаника није склона да мисли да апсолутно не зна одговор. У том тренутку ваља донети једну другу процену. Она се односи на поређење степена властитог незнања са вероватноћом да ће одговор ипак бити тачан. Ова процена вероватно интерферира са народном мудрошћу „Ко не рескира, не профитира". ([12] Фајгељ, 2013)

Када се крене са погађањем, престаје се са применом комбиновања претходна два модела. Модел погађања предвиђа вероватноћу тачног одговора који је независан од способности, и он је једнак ${1}/{k_i}$, где је k_i број могућих одговора за питање i.

Све ово даје једначину ([04] Partchev, 2004):

$$\mathrm{P}(\theta_i, a_i, b_i, \pi_{ij}) = \pi_{ij}\frac{1}{k_i} + (1-\pi_{ij})\frac{e^{\left(a_i(\theta_j - b_i)\right)}}{1+e^{\left(a_i(\theta_j - b_i)\right)}}$$

Вредности π_{ij} могу се тумачити као вероватноћа да особа реагује по моделу погађања или 2PL моделу. Вредности су специфичне, јер као што је већ речено, нема свако исту склоност ка погађању, а склоност зависи од интеракције између способности и тежине питања, док погађање једино има смисла када је питање и сувише тешко у односу на способност.

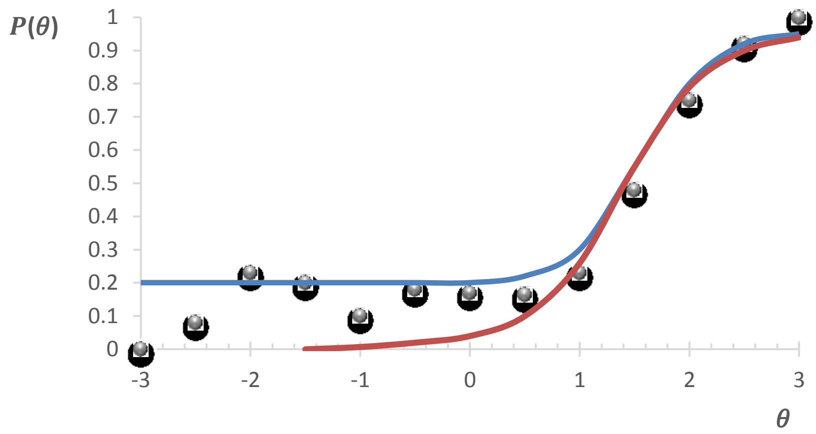
На овај начин смо завршили са нечим што је слично 3PL моделу, али доста компликованије, јер уместо k_i имамо π_{ij}, који зависе од комплексности особе (испитаника) и питања. У пракси све може бити још компликованије ако се погађање не догађа случајно. Зависно од нивоа способности, испитаник треба да има рационалну стратегију тако што ће искључити оне одговоре који имају најмању

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

могућност да су тачни, а онда да од преосталих изабере један одговор. Стопа успеха ће зависити од питања.

2.1.4.2 Погађање и оцена параметра

Ако узорак испитаника има нормалну стандардизовану расподелу способности, тј. средња вредност је 0, а стандардна девијација 1 и ако, рецимо, претпоставимо да питање има више понуђених формулација са, на пример, четири опције одговора, имаћемо доста могућности за посматрање давања одговора погађањем. Вероватноћа давања тачног одговора је $0.25 = {}^{1}/_{4}$. У 2PL моделу, чији су параметри $b = 0.84$ и $a = 1.84$, израчунава се вероватноћа тачног одговора (која је 0.04) и она се упоређује са 0.25, а затим се симулира одговор коришћењем веће од те две вредности. На овај начин испитаник прелази на погађање које обећава више вероватноће за давање тачног одговора од размишљања и тако добијамо и параметар $\boldsymbol{c}$ који ћемо користити у 3PL моделу.



Графикон 10: Карактеристичне криве питања 2PL и 3PL модела

На **Графикону 10** приказане су две криве, плава крива представља 3PL модел са параметрима $b = 0.94, a = 2.52$ и $c = 0.2$, а црвена крива представља 2PL модел. Види се да су горњи делови ових кривих доста близу један другом. Ако би се комбиновао горњи део 2PL модела са мањом вредношћу, тј. 0.04 уместо 0.25, могли бисмо, такође, да га поредимо са 3PL моделом, али би он сада имао различит параметар тежине. ([04] Partchev, 2004)

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

3 Функција вероватноће

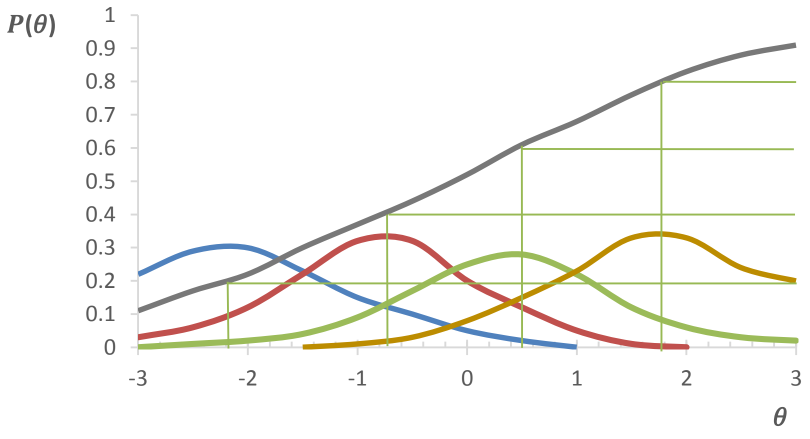
Питањима која се тестирају путем 1PL модела, којих има k, може се придружити једна од $k+1$ посматраних вредности $(0, 1 \ldots, k)$. Број могућих одговора на тесту је 2^k, тако да, на пример, за тест са пет питања постоје 32 могуће комбинације одговора. Сваки од њих има одређену вероватноћу одговарања. Сваки испитаник има своје моделе одговарања, који се међусобно искључују, и збир њихових вероватноћа биће 1. Ово важи за скуп података у целини, али и за сваки конкретни ниво способности.

Како израчунати вероватноћу да ће особа способности $\boldsymbol{\theta_j}$ одговарати на тесту по одређеном моделу, нпр. $(\top, \top, \bot, \top, \bot)$? Ми већ знамо како да израчунамо одвојено вероватноће сваког одговора у тесту, $\boldsymbol{P(\theta_j, b_1)}, \ldots, \boldsymbol{P(\theta_j, b_5)}$, али се поставља питање шта је њихова заједничка вероватноћа. IRT чини важну претпоставку локалне независности. То значи да су одговори дати на различита питања међусобно независни у односу на способност. Одговори могу бити у великој корелацији из разлога што су у обзир узете реакције испитаника различитих способности. Међутим, ако узмемо у обзир само лица која имају исту латентну способност, корелација између одговора требало би да нестане. Како су $\boldsymbol{P(\theta_j, b_1)}, \ldots, \boldsymbol{P(\theta_j, b_5)}$ функције од $\boldsymbol{\theta_j}$, можемо их помножити и добити вероватноћу за цео узорак. То следи из претпоставке условне независности, према којој су одговори дати на појединачна питања у тесту међусобно независни од $\boldsymbol{\theta_j}$.

Функцију $\boldsymbol{L(\theta) = \Pi_i \mathrm{P}_i(\theta, b_i)\theta^{u_i} * Q_i(\theta, b_i)^{1-u_i}}$, где је $\boldsymbol{u_i} \in (\boldsymbol{0}, \boldsymbol{1})$ називамо *функцијом вероватноће*. Ово је, у ствари, вероватноћа за сваки модел одговора, а збир свих ових функција је 1 за свако $\boldsymbol{\theta}$. IRT предвиђа да вероватноћа одговора на тесту даје праву (тачну) способност испитаника. ([04] Partchev, 2004)

3.1 Максимална веродостојност оцене способности

Оцену способности можемо добити на више начина, а један од начина је да изједначимо добијену (посматрану) вредност на тесту са очекиваном вредности. Други начин је да се приступи класичном начину добијања оцене способности, заснованом на принципу максималне веродостојности. Способност која има највећу вероватноћу у посматраном узорку постаће оцена способности. ([04] Partchev, 2004)

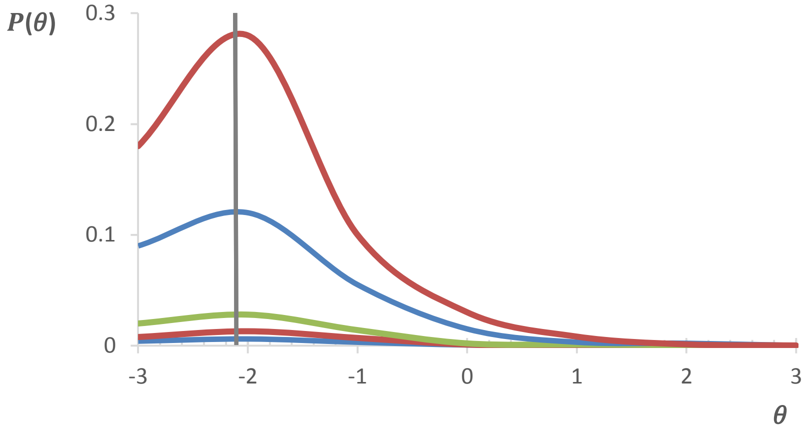


Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Графикон 11: Оцена максималне веродостојности

Графикон 11 приказује функцију вероватноће за одговоре: $(\top, \bot, \bot, \bot, \bot)$, $(\top, \top, \bot, \bot, \bot)$, $(\top, \top, \top, \top, \bot)$, али такође приказује и функцију информације за налажење оцене способности вредности $1-4$. Максимална веродостојност ће производити исте оцене као и изједначавање посматране са очекиваном способности.

Код 1PL модела оцена способности једино зависи од тога на колико питања је одговорено исправно, а не на која питања имамо тачан одговор. То не значи да је функција вероватноће инваријантна у односу на одговор, већ значи да функција вероватноће има исти број тачних одговора за сваки ниво способности.



Графикон 12: Функције вероватноће за пет питања

Графикон 12 приказује функције вероватноће за пет питања чија је укупна вредност 1. Свих пет функција дају исту оцену способности, чак иако нису исте функције. Лако је видети да су функције вероватноће различите, јер ако особа добије само једно питање, очекује се да оно буде лако, док би било изненађујуће да то питање буде тешко.

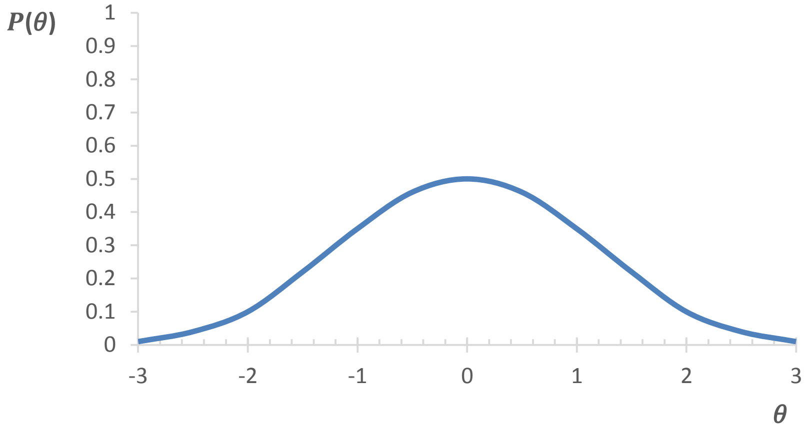
3.2 Оцењивање параметара питања

Ако претпоставимо да су параметри питања и способност познати, можемо лако нацртати функцију питања, функцију информације итд., а лако се може наћи и оцена способности. Налажење оцена параметара питања биће много једноставније уколико се зна тачна (права) способност испитаника. Међутим, постоји проблем истовременог оцењивања параметра питања и параметара особе (испитаника). За сада се може претпоставити да су способности познате, а пажња се усмерава на проблем налажења **S** криве на емпиријским подацима. ([03] Baker, 2001)

Теоретски, вредности способности могу да представљају било коју вредност између $-\infty$ и $+\infty$. Међутим, нису све вредности у том интервалу једнаке вероватноће.

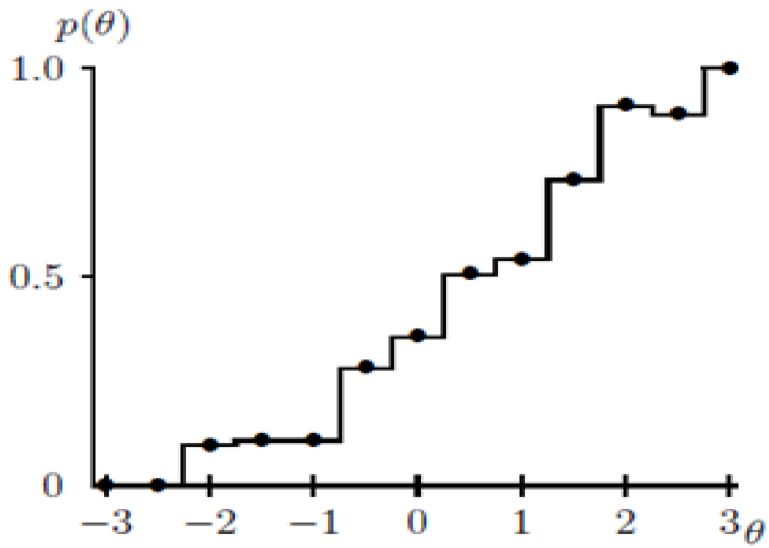
На **Графикону 13** приказана је стандардна, нормална расподела и, према овом моделу, већина способности је близу просека (тј. нуле), а вредности мање од -3 и веће од $+3$ биће веома ретке. Ово има важне последице по врсте одговора.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 13: Нормална расподела

Узимамо узорак од 1000 испитаника са нормалном дистрибуцијом способности и нека имамо питање средње тежине (рецимо $b = 0.5$). Можемо поделити опсег способности на велики број интервала, тако да се рачунају тачни одговори у сваком интервалу, као и пропорције тачних одговора. Резултати ће бити слично приказани као на **Графикону 14**. Тачке се налазе на средини сваког интервала. Стварне (праве) вредности параметара питања су непознате у тесту, па је један од задатака IRT да оцени ове параметре. Када се добију оцене параметара, оне ће обезбедити информације техничких особина питања у тесту. Параметри једног питања биће оцењени под претпоставком да су познати резултати (вредности) способности испитаника. У стварности, ови резултати нису познати, али је лакше објаснити како се добија оцена параметра питања ако је ова претпоставка испуњена. ([03] Baker, 2001)

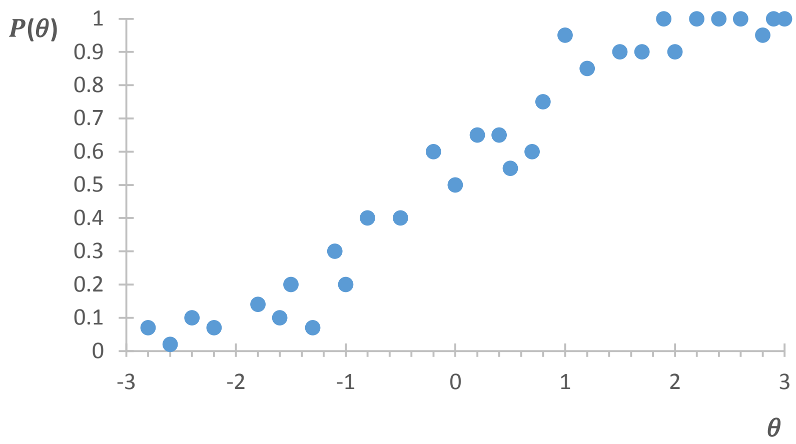


Графикон 14: Тачни одговори по интервалима ([03] Baker, 2001)

У случају типичног теста, од M испитаника који одговарају на N питања у тесту, резултати способности биће представљени преко различитих нивоа способности на скали способности у одређеном опсегу. У овом случају испитаници ће бити подељени у, рецимо $\mathcal{J}$, група на скали, тако да сви испитаници у оквиру једне групе имају исти ниво способности $\acute{e}_j$, и ту ће бити m_i испитаника у групи $\mathcal{J}$, где $j = 1, \ldots \mathcal{J}$. Унутар једне групе имаћемо резултат способности r_j, што је резултат свих испитаника који су одговорили тачно на питање. На нивоу способности $\acute{e}_j$ проценат тачног одговора биће $p(\acute{e}_j) = {}^{r_j}/_{m_j}$, те на овај начин добијамо још и оцену вероватноће тачног одговора на том нивоу. Сада се за

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

вредности r_j на скали способности могу добити $p(é_j)$ за сваки ниво способности $é_j$. Резултати добијених вредности (резултата) могу се представити и графички, као што је на **Графикону 15.**



Графикон 15: Оцена вероватноће по интервалима

Следећи корак је проналажење карактеристичне криве која одговара (апроксимативна је) пропорцији тачног одговора. Прво је неопходно изабрати логистички модел. Иако се може користити било који од три логистичка модела, овде ћемо употребити двопараметарски. Процедура која се користи за проналажење криве заснована је на процени максималне веродостојности. Према овом приступу почетне вредности параметра питања су $a = 1,\ b = 0$, а затим се израчунавају вредности $P(é_j)$ за сваки ниво способности. Посматрана вредност $p(é_j)$ и израчуната вредност $P(é_j)$ одређују се у свим групама способности. Након тога следе корекције оцена параметра питања које побољшавају карактеристичну криву, која дефинише везу између процењене вредности параметара и посматране пропорције тачног одговора. Процес корекције оцена наставља се све док се не добију толико мале оцене да је побољшање корекције могуће. Када се то постигне, поступак се прекида и тренутне вредности од b и a су оцене параметара питања.

Имајући у виду ове вредности, једначина карактеристичне криве питања користи се за израчунавање вероватноће тачног одговора $P(é_j)$ на сваком нивоу способности. Добијена крива је карактеристична крива која је апроксимативна подацима за то питање.

Важно разматрање у оквиру IRT је да ли модел карактеристичне криве одговара подацима који су дати за то питање. Мерење пропорција тачних одговора, које дају карактеристичну криву, ради се уз помоћ *Hi-kvadratnog indeksa*. Овај индекс се дефинише: ([03] Baker, 2001)

$$\chi^2 = \sum_{j=1}^{\mathcal{J}} m_j \frac{[p(\theta_j) - P(\theta_j)]^2}{P(\theta_j)Q(\theta_j)}$$

$\mathcal{J}$ - број група способности

$é_j$ - ниво способности групе j

m_j - број испитаника који су на нивоу $é_j$

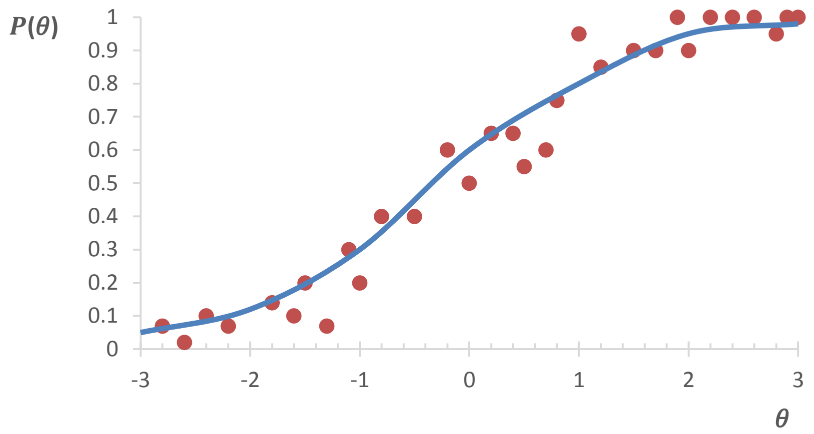
$p(é_j)$ је посматрана пропорција тачних одговора за групу j.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

$\boldsymbol{P(\acute{e}_j)}$ је вероватноћа тачног одговора за групу $\boldsymbol{j}$ израчуната на основу модела карактеристичне криве која користи оцене параметара.

$$\boldsymbol{Q(\theta_j) = 1 - P(\theta_j)}$$

Ако је вредност добијеног индекса већа од вредности критеријума, карактеристична крива питања, одређена вредностима оцена параметра питања, неће одговарати подацима. Ово може да изазове две ствари. Прво, може се добити погрешан модел карактеристичне криве. Друго, вредности посматраних пропорција тачног одговора не могу бити добро распоређене (разбацане), без обзира на модел.



Графикон 16: Проналажење карактеристичне криве која најбоље описује емпиријске податке

На пример, у случају неког питања, на **Графикону 16**, где имамо параметре питања $a = 1.27$ и $b = 0.39$, добијени индекс *Хи-квадрат (χ2) теста* је 28.88, а критеријум је 45.91 и ово се добро уклапа. Овде се види да двопараметарски логистички модел одговара пропорцијама тачног одговора.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Стварна процедура процене максималне веродостојности MLE (Maximum-Likelihood Estimation*)* прилично је сложен математички захтев и подразумева прорачун за свако питање у тесту. Из ових разлога није потребно ићи у детаље поступка, већ је довољно да се зна да процедура постоји, да подразумева доста рачунања и да карактеристична крива може да мери оно што нама треба.

3.3 Процедуре за оцене способности

Максимална веродостојност, према IRT, користи се за оцењивање способности испитаника. У случају оцењивања параметра питања имали смо итеративни процес. Процедура започиње познатом и примарном вредношћу способности испитаника и параметра питања, а затим се они користе за израчунавање вероватноће тачног одговора за свако питање. Побољшавање оцене способности добија се усклађивањем израчунатих вероватноћа са вектором (нула и јединица) одговора испитаника. Овај процес се понавља све док усклађивање постане довољно мало да је промена оцене способности занемарљива. Резултат овога је оцена параметра способности испитаника.

Једначина ове оцене биће: ([04] Partchev, 2004)

$$\boldsymbol{\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} a_i \left(u_i - P_i(\hat{\theta}_s)\right)}{\sum_{i=1}^{N} a_i{}^2 P_i(\hat{\theta}_s) * PQ_i(\hat{\theta}_s)}}$$

где је:

$\hat{e}_s$ – оцена способности у итерацији s

a_i - дискриминациони параметар, $i = 1, \dots, H$

u_i – одговор 1 – тачан, 0 – нетачан, $i = 0, 1$

$\mathrm{P}_i(\acute{e}_s)$ – вероватноћа тачних одговора за питање i, на нивоу способности $\acute{e}$ са итерацијом s

$Q_i(e_s) = 1 - \mathrm{P}_i(\acute{e}_s)$ – вероватноћа нетачног одговора за питање i, на нивоу способности $\acute{e}$ са итерацијом s.

Коначна вредност из ове једначине је оцена способности испитаника. Оцена способности биће у истој метрици као и нумеричка вредност параметра питања. Једна лепа особина ове једначине је да се може користи за сва три модела карактеристичне криве.

3.3.1. Оцена способности код 1PL модела

Оцењивање способности је вероватно најважнији део IRT. Можемо рећи да су оба приступа оцењивања способности, први изједначавање посматраног резултата на тесту са очекиваним резултатом и други - оцена максималне веродостојности способности, еквивалентни.

Свакој особи која учествује у тестирању помоћу 1PL модела, са нпр. пет питања, може се придружити вредност од 0 до 5. За резултате $1, 2, 3, 4$ може се наћи оцена способности у тачки где је очекивана вредност теста једнака посматраној вредности. Другим речима, оцене способности, добијене за посматране вредности, представљају оне вредности θ у којима функција теста узима вредности од 1 до 4. Функција одговора у 1PL моделу постаје 0 за особе чија је способност $-\infty$, односно 5 за лица чија способност иде у $+\infty$. Из тога следи да ће оцена способности за вредност 0 бити $-\infty$, а за савршен резултат (у нашем случају 5) биће $+\infty$. ([04] Partchev, 2004)

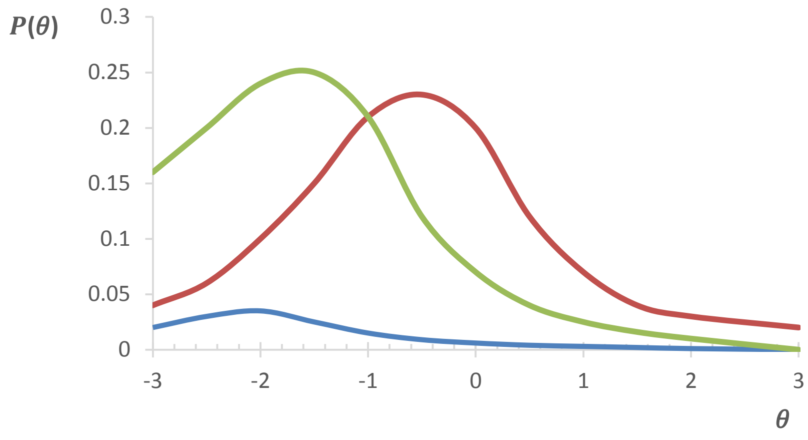
3.3.2. Оцена способности 2PL модела

Код 1PL модела максимална вероватноћа оцене способности може се на скали способности пронаћи као тачка у којој је очекивана вредност једнака посматраној вредности. Она такође зависи и од тога на колико питања је одговорено тачно, али не и на која питања је одговорено тачно, док код 2PL модела ово неће важити.

Из формуле 1PL модела $\sum_i P(\hat{\theta}, b_i) = \sum_i u_i$ где је $u_i \in (0, 1)$ вредност питања, код 2PL модела ово ће бити замењено са $\sum_i a_i P(\hat{\theta}, b_i, a_i) = \sum_i a_i u_i$, а уместо једноставне, добијамо пондерисане вредности, са параметром дискриминације a_i, као пондерима.

Све док се a_i разликују, што је обично случај, различити начини одговарања имају исту посматрану вредност (број тачних одговора), која неће довести до исте оцене способности. Сада неће бити важно само колико, већ и на која је питања одговорено тачно. ([04] Partchev, 2004)

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 17: Три различита начина одговарања

То се види на **Графикону 17**, где су приказана три начина одговарања: $(\top, \bot, \bot), (\bot, \top, \bot), (\bot, \bot, \top)$, који имају исту посматрану вредност 1. Зелена крива је веома равна, што одражава малу вероватноћу тачног одговора на тешка питања, која се јављају заједно са нетачним одговорима на два лака питања. Међутим, важно је приметити да су врхови (максимална вредност) кривих у различитим тачкама скале способности. Ово није био случај код 1PL модела. Принцип максималне веродостојности и даље важи, а може се наћи и максималне вероватноће оцене способности на месту где функција вероватноће, за посматрани начин одговарања, достиже свој максимум. Нпр. тест са пет питања имаће 32 начина одговора. Оцена максималне веродостојности за све нетачне одговоре је $-\infty$, а за све тачне је $+\infty$, док се за преосталих 30 начина може пронаћи максимизацијом вероватноће.

3.3.3. Оцена способности 3PL модела

Као и код претходна два модела, оцена способности код 3PL модела може се добити максимизирањем функције вероватноће. Функција вероватноће дефинисана је на исти начин као у претходна два модела, с тим што вероватноће $\boldsymbol{P}$ и $\boldsymbol{Q}$ подржавају 3PL модел. ([04] Partchev, 2004)

$$\boldsymbol{L(\theta) = \Pi_i P_i(\theta, b_i, a_i, c_i)^{u_i}\, Q_i(\theta, b_i, a_i, c_i)^{1-u_i}}, \text{ где је } \boldsymbol{u_i \in (0, 1)}$$

Максимална веродостојност оцене способности зависи од тога на колико је питања одговорено тачно, али и која су то питања.

3.3.4. Прецизност и грешка мерења

Прецизност је супротна од грешке. Међутим, и прецизност и грешка могу нам бити од велике користи за истраживање. Грешка мерења изражена је у истим јединицама као и само мерење, тако да је можемо поредити са оценом способности, или да је искористимо за интервал поверења оцене. Варијанса оцене способности може бити оцењена и као реципрочна вредност тест функције информације: ([03] Baker, 2001)

$$\boldsymbol{Var(\hat{\theta}) = \frac{1}{\mathrm{I}(\hat{\theta})}}$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Пошто је стандардна грешка мерења (SEM - *Standard error of measurement*) једнака квадратном корену варијансе, за 1PL модел имаћемо:

$$\boldsymbol{SEM(\theta)} = \sqrt{\boldsymbol{1}/_{\mathbf{I}(\boldsymbol{\theta}^{m})}} = \sqrt{\boldsymbol{1}/_{\boldsymbol{\Sigma}_i \boldsymbol{P}_i(\boldsymbol{\theta}, \boldsymbol{b}_i), \boldsymbol{Q}_i(\boldsymbol{\theta}, \boldsymbol{b}_i)}}$$

Формуле стандардних грешака 2PL и 3PL модела:

$$\boldsymbol{SEM(\theta)} = \sqrt{\boldsymbol{1}/_{\sum_i \boldsymbol{a}_i^2\, \boldsymbol{P}_i(\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{a}_i)\, \boldsymbol{Q}_i(\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{a}_i)}}$$

$$\boldsymbol{SEM(\theta)} = \sqrt{\boldsymbol{1} \Big/ \sum_i \boldsymbol{a}^2 \frac{\boldsymbol{Q(\theta)}}{\boldsymbol{P(\theta)}} \left[\frac{\boldsymbol{P(\theta)} - \boldsymbol{c}}{\boldsymbol{1} - \boldsymbol{c}}\right]^{\boldsymbol{2}}}$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

4 Функције и калибрација

4.1 Функција информације питања и теста

Када поседујемо неке информације то значи да имамо сазнања о одређеној теми или предмету. Статистичко значење информације приписује се Р. А. Фишеру, који је информације дефинисао као реципрочне вредности прецизности, са којом параметар може бити оцењен. Уколико постоји могућност да се оцени параметар са одређеном прецизношћу, тада би требало знати више информација о том параметру. Мера прецизности је варијанса оцењивача, који се означава са σ^2.

Количина информације означава се са I и ово нам даје формулу: $\boldsymbol{I} = \frac{\mathbf{1}}{\boldsymbol{\sigma}^{\mathbf{2}}}$

Када се подигне стандардна девијација на квадрат, добија се варијанса која је мера, тј. са којом можемо да оценимо способност на одређеном нивоу. Из формуле за количину информације имамо да је она, на датом нивоу, једнака реципрочној вредности варијансе. Уколико је количина информација велика, то значи да са великом прецизношћу можемо да оценимо способност испитаника на том нивоу, тј. ове оцене биће близу праве вредности. Ако је количина информација мала, значи да се способност не може оценити са великом прецизношћу. Користећи одговарајућу формулу, можемо добити количину информација за сваки ниво и те вредности могу бити смештене дуж мерне скале од негативне до позитивне бесконачности. Пошто је способност променљива, такође ће и информације бити променљиве. Функција информације говори нам како је сваки ниво способности оцењен, а такође је важно да функција не зависи од расподеле испитаника дуж скале способности. Идеална функција је хоризонтална линија (где су ови нивои оцењени са истом прецизношћу), међутим, то је тешко постићи. ([03] Baker, 2001)

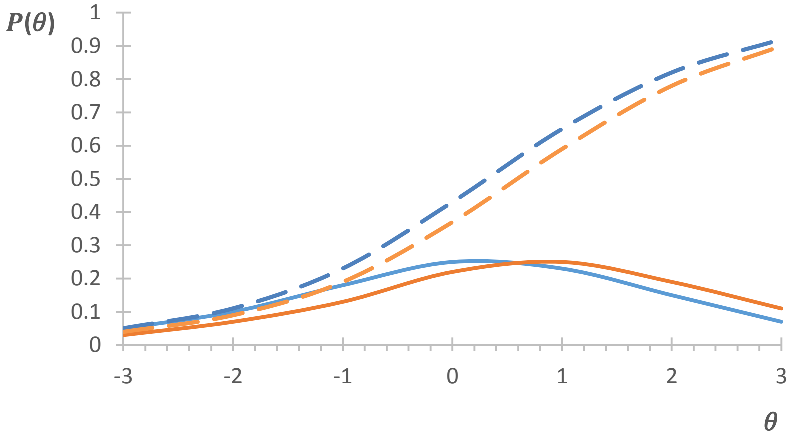
Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

4.1.1. Функција информације питања у 1PL моделу

Функција информације има истакнуту улогу у IRT – њоме можемо проценити способност испитаника. Овде се испитује једно питање и за то питање ћемо имати функцију информације. Свако питање у тесту даје неку информацију о способности испитаника, али количина информација зависиће од тога колика је тежина питања у односу на могућности испитаника. 1PL модел је једини случај који укључује информацију питања, док се код других модела врши комбинација са другим факторима. Функција информације 1PL модела ([04] Partchev, 2004) је:

$$\mathbf{I}_i(\boldsymbol{\theta}, \boldsymbol{b}_i) = \mathbf{P}_i(\boldsymbol{\theta}, \boldsymbol{b}_i)\boldsymbol{Q}_i(\boldsymbol{\theta}, \boldsymbol{b}_i)$$

На основу ове формуле и карактеристичне криве 1 PL модела добијамо **Графикон 18**:



Графикон 18: Функције информације и одговора за два питања 1PL модела

Из ове формуле се види да је максимална вредност ове функције 0.25 и она се јавља у тачки где је вероватноћа тачног и нетачног одговора једнака 0.5. Другим речима, свако питање у 1PL моделу има највећу вредност функције информације ако је тежина питања једнака са способношћу испитаника. Уколико је способност мања или већа од тежине питања, тада се функција информације смањује, тј. опада.

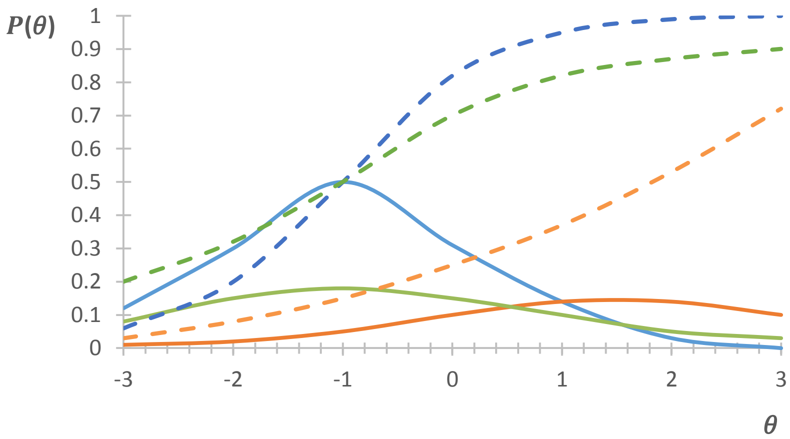
Закључак из свега овога јесте да су нам потребна питања различите тежине како бисмо мерењем способности код испитаника добили добре оцене.

4.1.2. Функција информације питања у 2PL моделу

Функција информације питања у 2PL моделу ([04] Partchev, 2004) је:

$$\mathbf{I}_i(\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{a}_i) = \boldsymbol{a}_i{}^2 \mathbf{P}_i(\boldsymbol{\theta}, \boldsymbol{b}_i) \boldsymbol{Q}_i(\boldsymbol{\theta}, \boldsymbol{b}_i)$$

Утицај дискриминативног параметра $\boldsymbol{a}_i$ је велики јер се појављује у облику квадрата. То значи да квадрат дискриминативног параметра испод $\mathbf{1}$ прилично смањује функцију информације, док је дискриминациони параметар изнад $\mathbf{1}$ значајно повећава.



Графикон 19: Функције информације и одговора за три питања 2PL модела

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

На **Графикону 19,** функције одговора означене су испрекидано и њихова боја усклађена је са бојом функције информације питања.

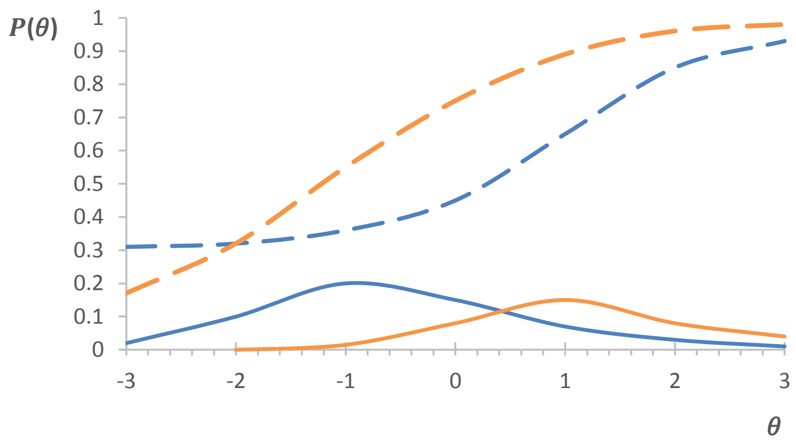
Код 1PL модела све функције информације питања имале су исти облик, исти максимум 0.25 и једноставно су се померале дуж осе способности, тако да свака функција информације питања има максимум у тачки где је способност једнака тежини питања.

Код 2PL модела, функција информације питања, ипак, свој максимум остварује у тежини питања. Међутим, њихови облици и вредности максимума зависиће од параметра дискриминације. Када је дискриминација висока (тј. функција је стрма), питање даје више информација о способности, а информације су сконцентрисане око питања (тачке) тежине. Питања ниске (мале) дискриминативности параметра, дају мање информација, а информације су разбијене дуж осе способности.

4.1.3. Функција информације питања у 3PL моделу

Ова функција мало је компликованија код 3PL модела, за разлику од 1PL и 2PL, а дефинисана је: ([04] Partchev, 2004)

$$\mathbf{I}_i(\boldsymbol{\theta}, \boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{c}_i) = \boldsymbol{a}^2 \frac{\boldsymbol{Q}(\boldsymbol{\theta})}{\mathbf{P}(\boldsymbol{\theta})} \left(\frac{\mathbf{P}(\boldsymbol{\theta}) - \boldsymbol{c}}{\mathbf{1} - \boldsymbol{c}} \right)^{\mathbf{2}}$$



Графикон 20: Функције информације и одговора за два питања 3PL модела

Графикон 20 приказује испрекидану IRF (*Item Response Function*) и IIF (*Item Information Function*) која је неиспрекидаna за два питања 3PL модела. Питање са плавом кривом линијом има $a = 1,\ b = -1,\ c = 0.1$, док питање са наранџастом линијом има $a = 1,\ b = 1,\ c = 0.3$. Параметар b помера функцију информације питања лево или десно, али не утиче на њен облик. Ова два питања имају исто $a = 1$, а различито c, па се може закључити да веће c доводи до општег смањења информације питања и врх IIF није више у тачки $\theta = b$.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

4.1.4. Функција информације теста

Како се тестом оцењује способност испитаника, на основу њега такође можемо добити количину информација за сваки ниво способности. Тест је скуп питања, па је информација теста на датом нивоу способности збир информација питања на том нивоу. Стога се функција информације теста дефинише као: ([04] Partchev, 2004)

$$\mathbf{I}(\boldsymbol{q}) = \sum_{i=1}^{N} \mathbf{I}_i(\boldsymbol{q}),$$

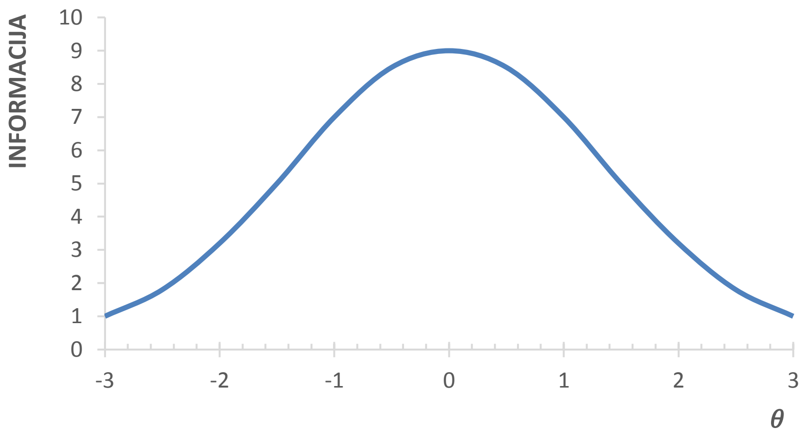
где је:

$\mathbf{I}(\boldsymbol{\acute{e}})$ количина информација теста на нивоу способности $\boldsymbol{\acute{e}}$

$\mathbf{I}_i(\boldsymbol{\acute{e}})$ количина информација теста за питање i на нивоу $\boldsymbol{\acute{e}}$

$\boldsymbol{N}$ број питања у тесту.

Општа функција информације теста биће много већа него за једну функцију информације питања. Према томе, тест мери способност прецизније него што то раде питања појединачно. Важна карактеристика дефиниције информације теста јесте да што више питања имамо у тесту, то имамо већу количину информација. Стога, тестови са већим бројем питања мериће способност испитаника са већом прецизношћу него тестови са мање питања. Количина информације теста може се представити **Графиконом 21** функције информације теста.



Графикон 21: Функција информације теста

У овом примеру максимална вредност функције информације теста је умерена, количина информација се смањује постепено када се ниво способности разликује од оне која одговара максимуму. Дакле, способност је оцењена са неком прецизношћу која је близу центра скале способности. Међутим, како се ниво способности приближава крају скале, количина информације теста значајно опада.

Функција информације теста у основи говори колико добро ради тест за оцењивање способности преко свих вредности способности. С друге стране, идеална функција информације теста често може бити дуж хоризонталне осе, што можда није најбоље за специфичну намену. Нпр. ако бисмо желели да направимо тест за доделу стипендија, овај метод (*ideal*) не би био оптималан. У овој ситуацији желимо да измеримо способност са значајном прецизношћу да бисмо раздвојили оне

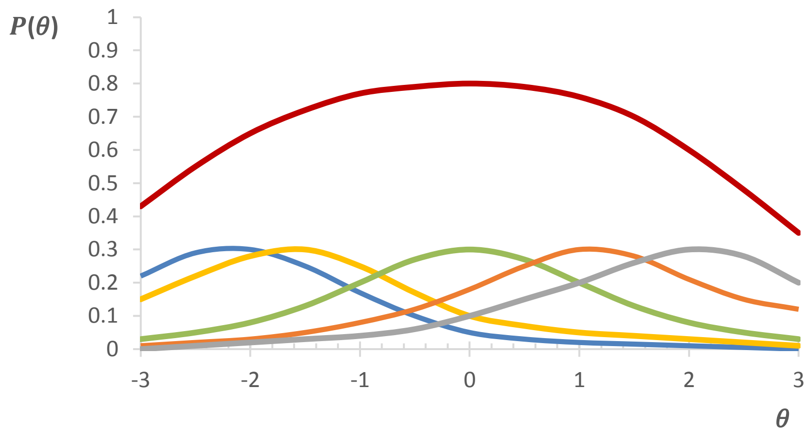
Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

који ће добити ову стипендију од оних који неће. Најбоља функција информације теста у овом случају има врх у граничној вредности. Иако се функција информације може добити за свако питање у тесту, она се ретко користи. Количина информација од једног питања је мала, па се оцена способности испитаника не одређује на основу једног питања. На основу свега овога закључујемо да су количина информације теста на сваком нивоу способности и функција информације теста од примарног интереса. Како се информације теста добијају сумирањем информација питања на датом нивоу способности, количина информација представљаће тај ниво способности. Математичка дефиниција количине информације питања зависиће од конкретног модела. ([04] Partchev, 2004)

4.1.5. Функција информације теста у 1PL моделу

У тесту, функција одговора је сума функција свих одговора на дата питања. Поставља се питање каква је функција информације теста у односу на функције информација питања. Функција информације теста у 1PL моделу је: ([04] Partchev, 2004)

$$\mathbf{I}_j(\boldsymbol{\theta}_j) = \sum_i \mathbf{I}_{ij}(\boldsymbol{\theta}_j, \boldsymbol{b}_i)$$



Графикон 22: Функција информације теста (црвена линија) и пет функција информације питања 1PL модела

Графикон 22 приказује функцију информације теста за пет питања, а заједно са овом функцијом приказано је и свих пет функција информације питања. Све ове функције приказане су на једној скали (графикону). Може се видети да тест у целини даје много више информација него питања појединачно, а уз то и количину информација на ширем опсегу способности.

Најважнија особина функције информације теста је да предвиђа прецизност на основу које можемо мерити вредности латентне способности. Још увек не можемо посматрати способност посебно, јер још не постоје начини како то измерити, али са сигурношћу знамо шта можемо постићи са тачношћу мерења у било ком нивоу способности.

4.1.6. Функција информације теста у 2PL моделу

Функција информације теста 2PL модела дефинише се на исти начин као и код 1PL модела. Она представља суму функција информације питања над питањима у тесту.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Функције информације питања различите су у односу на оне из 1PL модела, па је формула за функцију информације теста 2PL модела: ([04] Partchev, 2004)

$$I_j(\theta_j) = \sum_i I_{ij}(\theta_j, b_i, a_i) = \sum_i a_i{}^2 P(\theta, b_i, a_i)\, Q(\theta, b_i, a_i)$$

Из овога се види да функције информације питања зависе од дискриминативног параметра a_i. Облик функције информације теста заобљен је и непредвидив, нарочито у тестовима са мало питања. У пракси би требало да имамо функцију која је висока и глатка на одређеном опсегу способности, али би најбоље било имати криву која је стрма. Ово се може постићи са већим бројем питања која имају велику дискриминацију параметара и равномерно распоређен опсег способности. Питања са веома ниском дискриминацијом параметара обично се одбацују из практичне употребе.

4.1.7. Функција информације теста у 3PL моделу

Функција информације теста у 3PL моделу дефинисана је као сума функција информација питања над питањима у тесту. Оно што се разликује између ова три модела су функције информације питања, тако да је функција информације теста за 3PL модел: ([04] Partchev, 2004)

$$I_j(\theta_j) = \sum_i I_{ij}(\theta_j, b_i, a_i, c_i) = \sum_i a_i{}^2\, \frac{Q(\theta)}{P(\theta)} \left(\frac{P(\theta) - c}{1 - c}\right)^2$$

Функција информације теста као и код 2PL модела зависи од параметара дискриминације a_i, али у 3PL моделу је додатни утицај параметара „погађања“ c_i. Што су већи c_i, функција информације теста опада и мења свој максимум у односу на b_i. Као резултат свега овога, облик функције информације теста може бити компликован код 3PL модела. У практичној примени циљ је да добијемо функцију информације теста која је растућа и глатка на одређеном интервалу.

4.2 Калибрација

За дидактичке потребе, способност је приказана у виду метричких вредности на скали способности. Ове метричке вредности могу узети вредност нула, као централну вредност, и да се од нуле крећу ка позитивној, односно негативној бесконачности у јединичним интервалима. Нумеричке вредности параметара питања, као и параметара способности испитаника, исказане су такође метричким вредностима. ([03] Baker, 2001)

Када се крене са израдом теста, требало би да знамо шта којим питањем желимо да измеримо и да ли су питања која смо задали разумљива и решива међу испитаницима мале, средње и високе способности. Међутим, на почетку не можемо имати представу о свему, јер уколико се даје тест групи испитаника, нећемо унапред знати колико латентних особина испитаници поседују. Као резултат овога, поставља се главни задатак да се утврде вредности параметара питања и способности испитаника. У IRT овај задатак зове се *тест калибрација*, и она ће обезбедити оквир за тумачење резултата испитаника. Ово можемо постићи тако што ћемо групи од M испитаника оценити тест од N питања. Затим се примењују метрички поступци на податке одговора питања, са циљем формирања скале способности која на јединствен начин одређује комбинацију питања теста. На овај начин и оцене параметара питања и оцене способности испитаника биће изражене метрички.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

4.2.8. Процес калибрисања

За калибрацију се користи итеративни поступак, који укључује две фазе оцене максималне веродостојности. У првој фази оцењују се параметри N питања из теста, а у другој фази оцењује се способност M испитаника. Процес итерације врши се до тренутка док се не добије скуп оцена параметара. Тада можемо рећи да је тест калибрисан и да смо дефинисали скалу способности.

У оквиру прве фазе, оцењивање способности сваког испитаника третира се као права метрика, која је изражена у функцији латентне особине. Тада се параметри сваког питања у тесту оцењују путем максималне веродостојности, о којој је било речи у једном од претходних поглавља. Ради се питање по питање, јер је претпоставка да су питања независна. ([03] Baker, 2001)

У другој фази претпостављамо да је оцена параметра питања, добијена у првој фази, заправо вредност параметра питања. Затим се оцењује способност сваког испитаника коришћењем максималне веродостојности. Још се претпоставља да је способност испитаника независна у односу на друге способности, па ће се оцене добијати истовремено за сваког испитаника.

Овај поступак се наставља све док се неком повољном конвергенцијом критеријум не испуни. Ефекат свега овога је да се параметри од N питања у тесту и нивои способности за M испитаника оцењују истовремено.

4.2.9. Метрички проблем

Једна од мана калибрације је могућност да не добијемо потребне показатеље на скали способности на јединствен начин. ([03] Baker, 2001)

Може се десити да добијемо много различитих вредности, које раде подједнако добро. У техничком смислу, метрика је јединствен начин којим се долази до линеарне трансформације. Свака од две фазе, у оквиру једне итерације, изводи се коришћењем нешто другачије метрике на скали способности. Општи итеративни процес конвергира, као што и метрика, на скали способности, конвергира ка одређеној вредности и мерној јединици. Главна карактеристика овог процеса је да резултат мерења на скали способности зависи од одређеног скупа елемената, који чине тест. На овај начин можемо добити податке који зависе од комбинације способности испитаника и питања теста.

Способност да се пронађу питања и испитаници дуж заједничке скале способности је доста моћна карактеристика IRT. Ова функција омогућава да се интерпретира резултат калибрације у јединственом оквиру, који ће обезбедити смислене вредности параметара питања.

4.2.10. Позиционирање оцена на скали способности

Ако се погледа формула 1PL модела, из ње можемо да издвојимо $\boldsymbol{e}^{(\theta_i - b_j)}$, и да кажемо да је ово одређено до на константу. На пример, ако бисмо сваком нивоу способности $\boldsymbol{\theta_i}$ додали исти број, онда бисмо се могли вратити у формулу и сваком параметру тежине $\boldsymbol{b_i}$ додати исти тај број, тј. имаћемо $\boldsymbol{e}^{\left(\theta_i + B - (b_j + B)\right)} = \boldsymbol{e}^{(\theta_i - b_j)}$.

На исти начин, дискриминациони фактор $\boldsymbol{a_j}$ код 2PL и 3PL модела може бити помножен са константом, па ћемо имати:

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

$$e^{\left(\frac{a_j}{A}(A\theta_i - Ab_j)\right)} = e^{(\theta_i - b_j)}$$

Уопштено, нека је $\boldsymbol{\theta_i^* = A + B\theta}$. За 3PL модел можемо добити линеарну трансформацију коришћењем:

$$\boldsymbol{a_i^* = \frac{a_j}{A}, b_j^* = Ab_j + B, c_j^* = c_j}$$

Последица овога јесте да ћемо морати да поставимо произвољна ограничења способности θ_i**,** обично да је средња вредност 0, а варијанса 1, или ћемо, алтернативно, поставити ограничења на оцене параметара питања како бисмо идентификовали наш модел. ([03] Baker, 2001)

Претпоставимо да желимо да оценимо параметре питања за два теста и да у њима учествују различити испитаници, при том испитаници за оба теста имају различите способности. Уколико дозволимо програму да стандардизује способност, оцена способности за способније групе ће пасти и самим тим ће их пратити и тежине питања. С друге стране, за мање способније групе оцене ће бити знатно веће, а тиме и тежине питања. Дакле, параметри питања два теста неће бити на истом нивоу на скали. Проблем неће нестати ако бисмо стандардизовали параметре питања уместо оцена способности, а уз то, не може се тврдити да ова два теста имају исту просечну тежину. Неопходно је постављање параметара добијених од ова два различито калибрисана узорка испитаника на истој скали. Начини којима ћемо ово постићи зависе од практичне ситуације.

На пример, можемо да радимо калибрацију на два узорка одвојено, а затим да се користе оцене заједничких питања, тако да ће се оцене питања наћи на истој скали.

Ово се постиже линеарном трансформацијом, чији се параметри могу добити из односа: ([03] Baker, 2001)

$$\boldsymbol{A} = \frac{\sigma(b_j^*)}{\sigma(b_j)} = \frac{\mu(a_j)}{\mu(a_j^*)} \text{ и } \boldsymbol{B = \mu(b_j^*) - \mathrm{A}\mu(b_j)}$$

где је $\boldsymbol{\mu}$ средња вредност, а $\boldsymbol{\sigma}$ је стандардна девијација. Грчка слова значе да су ови односи на нивоу модела. У зависности од тога како смо добили A, имамо однос $\boldsymbol{\mu}/\boldsymbol{\sigma}$ или $\boldsymbol{\mu}/\boldsymbol{\mu}$. ([03] Baker, 2001)

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

5 Метод

5.1 Циљ

Основни циљ овог истраживања био је да се преко модела IRT упореде два начина оцењивања (бодовања) питања на тесту. Такође, применом дефиниција и формула модела IRT тестирали смо адекватност ова три модела за упоређивање различитих начина оцењивања.

5.2 Статистичке јединице

Јединице нашег истраживања биле су ученици средње школе - Гимназије у Врњачкој Бањи, који су изабрани у узорак. У истраживању је учествовало 84 ученика, од тога 60,71 % девојчица и 39,29 % дечака. Ученици који тог дана нису били у школи нису учествовали у тестирању, а током касније обраде података су искључени, тј. посматрали смо их као посебну врсту неодговора и такве јединице смо искључили из обраде. Треба нагласити да ни ученици као ни њихови професори нису знали ко је све од ученика изабран у узорак, тачније професори нису знали која су тачно одељења изабрана све до тренутка уласка у учионицу.

5.3 Начин избора узорка

Узорак који смо користили у нашем истраживању је пригодан узорак. Биран је тако што је на територији једне општине изабрана једна школа, а затим су у оквиру те школе бирана одељења. Од четири одељења изабрали смо три, тако да су обухваћени сви ученици тих одељења.

Избору сваког узорка претходило је и прављење оквира за избор узорка. Због ограничених могућности, оквир смо свели на територију општине Врњачка Бања и ученике у тој општини, а био је дефинисан циљном популацијом. Укључени су сви 15-огодишњи ученици са територије поменуте општине, али и ученици оних школа које су касније могле бити искључене због критеријума искључивања. Оквир за избор узорка укључивао је идентификационе податке попут: јединствене нумеричке идентификације ученика и кодиране информације о школи (на пример врста школе). Такође, треба водити рачуна о броју јединица које су у узорку, као и о броју јединица које искључујемо, тј. да ли уопште постоје јединице које искључујемо, и њих треба свести на минимум. Искључења која су обухваћена и нашим истраживањем односе се на: интелектуално ометене ученике (они који имају менталне или емоционалне неспособности) и на функционално ометене ученике, тј. оне који су трајно физички онемогућени да би могли валидно бити оцењени.

Нашим узорком није било неопходно обезбедити репрезентативност за популацију ученика, јер смо тестирали начин оцењивања питања и моделе IRT, а не саме ученике.

5.4 Проблем и хипотезе

Основно истраживачко питање које смо поставили у овом раду односило се на: тестирање валидности два начина оцењивања теста, и то преко корелација са успехом у школи и са успехом са пријемног испита. При његовом разрађивању, водили смо рачуна и о томе да ли је на испитане појаве и

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

њихове везе утицала оцена из математике на крају осмог разреда, или, евентуално, број поена на пријемном испиту из тог предмета.

Ослањајући се на претходно изнета теоријска разматрања, формулисали смо неколико истраживачких хипотеза. ([12] Фајгељ, 2013) Наша прва претпоставка била је да ће бољи успех ученика из школе дати и боље резултате на тесту. Затим, друга претпоставка била је да ће уситњен начин оцењивања (други начин оцењивања) дати бољи резултат на тесту, а трећа да ће се у просеку, у другом начину оцењивања, параметри питања значајно разликовати.

5.5 Типови података (варијабле)

Резултате тестирања објединили смо у две базе (за први и за други начин оцењивања), направљене у *Excel* табели ***(Прилог 2).*** Оне су садржале податке за укупно 84 испитаника и за сваког испитаника постојале су вредности обележја. Сви одговори у бази били су кодирани на начин који ће испод бити наведен.

Стандардним статистичким моделима било је јако тешко или немогуће анализирати категоријске податке јер, се за њихову анализу примењују методе оптималног скалирања, развијене и имплементиране у оквиру SPSS (*Statistical Package for Social Sciences*) статистичког софтвера. Поред овог софтвера постоје и други који нам, такође, дају жељене резултате. Један од таквих је SAS (*Statistical Analysis Software*), који има широку примену у статистичким истраживањима. Обрада података које смо прикупили тестирањем извршена је уз помоћ SAS софтвера ***(Прилог 3).*** За обраду података користили смо уобичајене функције и процедуре које се налазе у склопу SAS софтвера.

Категоричке номиналне променљиве су:

- Успех у решавању задатака, кодиран на следећи начин:

I начин	II начин
-1 = нетачно	-1 = нетачно
0 = није одговорено	0 = није одговорено
1 = тачно	0.25 = од 4 понуђена одговора тачан је само један
	0.5 = тачна је половина понуђених одговора
	0.75 = од 4 понуђена одговора тачно је три
	1 = тачно

- Пол: представља пол испитаника и кодиран је на следећи начин:

$$0 = \text{мушки}$$

$$1 = \text{женски}$$

- Параметри IRT модела:

$$a = \text{параметар дискриминације}$$

$$b = \text{параметар тежине}$$

$$c = \text{параметар погађања}$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Нумеричке номиналне променљиве су:

- Оцена на крају осмог разреда из математике
- Број поена на пријемном испиту из математике.

5.6 Инструмент (тест)

Податке смо прикупили методом теста над петнаестогодишњим ученицима гимназије, изабраних у узорак, који су на дан тестирања били у школи. Врста теста коју смо користили приликом истраживања је тест знања. Садржај питања са теста, као што је претходно речено, био је шаренолик, јер су питања која смо задали различитог нивоа тежине. Ти нивои су:

1. Ниво репродуктивног знања
   - Примењује одређена научна знања у једноставним, познатим ситуацијама
   - Директно резонује на основу датих података у једноставној ситуацији, изводи дословне интерпретације
   - Познаје основне термине

   Питања првог нивоа била су: троугао, избор, степенице, тест, размена, извоз ***(Прилог 1)***.
2. Ниво интеграције
   - Бира и повезује објашњења из различитих научних дисциплина и примењује их на различите животне ситуације
   - Бира битна знања и примењује једноставне истраживачке моделе
   - Користи и интерпретира научне концепте из различитих дисциплина

   Питања другог нивоа била су: коцкице, макета, полице, фарма ***(Прилог 1)***.
3. Ниво евалуативног знања
   - Препознаје научне компоненте у различитим сложеним животним ситуацијама и њих примењује на научне концепте и знања
   - Пореди, евалуира и бира одговарајуће научне податке
   - Пореди, проверава и потврђује хипотезе
   - Даје објашњења и аргументе засноване на подацима

Питања трећег нивоа била су: најбољи ауто, интернет дописивање, стабла јабуке, столар ***(Прилог 1)***.

Питања отвореног типа била су троугао, коцкице и столар, док су остала затвореног типа. Питања столар, извоз, фарма и стабла јабуке имала су слику на основу које су испитаници могли доћи до тачног одговора.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Основни образац био је тест са четрнаест задатака који обухватају двадесет питања. Тест ***(Прилог 1)*** задали смо на четири различита начина, тј. имали смо четири групе различитог распореда тежина питања, тако да су у првом тесту питања ишла од најлакшег ка најтежем, у другом обрнуто, док су у трећем и четвртом тесту питања комбинована. Важна чињеница била је да сви тестови (све четири групе) имају потпуно иста питања. Тиме смо желели да избегнемо преписивање и међусобна упоређивања решења задатака, али и да погледамо постигнућа испитаника када питања нису у типичном редоследу. Решавање теста било је ограничено на двадесет минута, што је представљало други начин којим смо желели да спречимо међусобна преписивања.

Истраживање је подразумевало да ученици пре почетка решавања теста дају неке од својих основних идентификационих података. Идентификационе податке чинили су: одељење, пол, оцена из математике на крају осмог разреда и број поена са пријемног испита из математике.

5.7 Поступак спровођења истраживања

Захтеви које смо поштовали за спровођење истраживања обухватају четрдесетдводневни период, и то од 01.03. до 31.08. датума у текућој години, и он се назива *периодом тестирања*. Према овим захтевима, тестирање није било препоручљиво током првих шест недеља школске године, јер се сматра да су ученикове способности мање на почетку школске године, него на крају претходне школске године, тако да је тестирање било предвиђено за април, а све то је било у договору са директором и предметним наставницима. На основу захтева, истраживање је обухватало популацију коју чине ученици старости од 15 година и 3 месеца до оних са 16 година и 2 месеца. Варијација до месец дана у овом узрасту је дозвољена.

Како бисмо предупредили осипање узорка испитивање је извршено у једном наврату и трајало је један школски час по одељењу. Испитивања су протекла без икаквих проблема, а испитаници су тестове попуњавали савесно. Из анализа нису били искључени испитаници са изразито ниским резултатима.

Учешће у истраживању било је добровољно, а испитаницима је гарантована анонимност и поверљивост индивидуалних података.

5.8 Обавеза заштите података прикупљених током истраживања

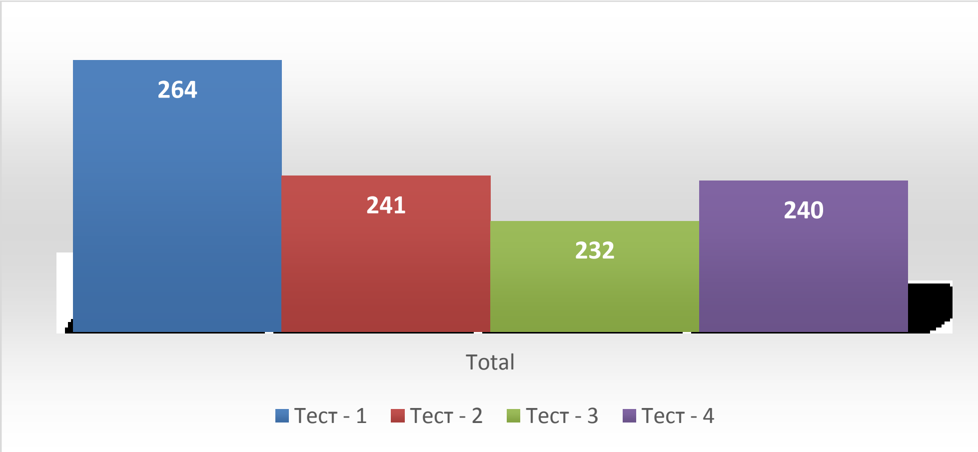
Према Закону о званичној статистици („Службени гласник РС“, број 104/09), подаци прикупљени током истраживања користиће се искључиво у статистичке сврхе. Прикупљени подаци представљају строгу тајну и свака злоупотреба прикупљених података је строго кажњива. Стога ће се прикупљени подаци објављивати као групни, никако појединачно ([10] „Службени Гласник Републике Србије“ број 104/09).

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

6 Резултати и дискусија

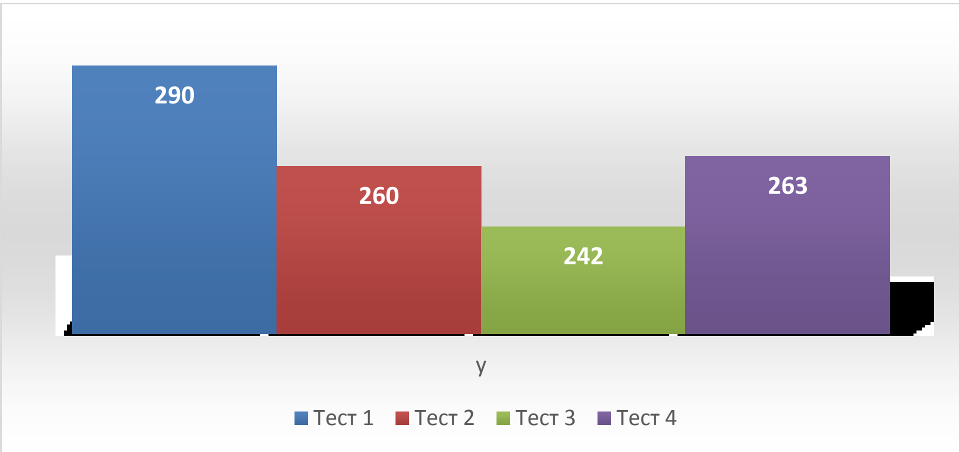
6.1 Помоћни подаци

Најпре је било важно да упоредимо четири различита теста како бисмо утврдили да ли је распоред задатака имао ефекат на укупан скор испитаника. Приликом одређивања укупног скора за први начин оцењивања (**Графикон 23**) за сва четири теста добили смо мале разлике. Вредности укупног скора (тотали) представљали су нам почетну тачку за даљу обраду података. Овим смо, такође, желели да видимо да ли смо коректно задали тестове и да ли њихове резултате можемо користити у даљој обради, а у супротном бисмо имали проблем. Одређивањем вредности укупног скора испитаника за други начин оцењивања теста добили смо такође мале разлике (вредности су: $290, 260, 242, 263$) - **Графикон 24**.



Графикон 23: Укупан скор испитаника за први начин оцењивања теста

Затим смо се питали да ли постоје разлике укупних скорова између тестова првог и другог начина оцењивања. Да бисмо ово урадили користили смо АНОВА (*Analysis Of Variance*) анализу варијансе. Код првог начина оцењивања, закључили смо да не постоје статистички значајне разлике укупних скорова између тестова $(F = 0.13, p > 0.05)$ - ***Прилог 4***. Такође, и код другог начина оцењивања дошли смо до истог закључка $(F = 0.08, p > 0.05)$ - ***Прилог 5***.

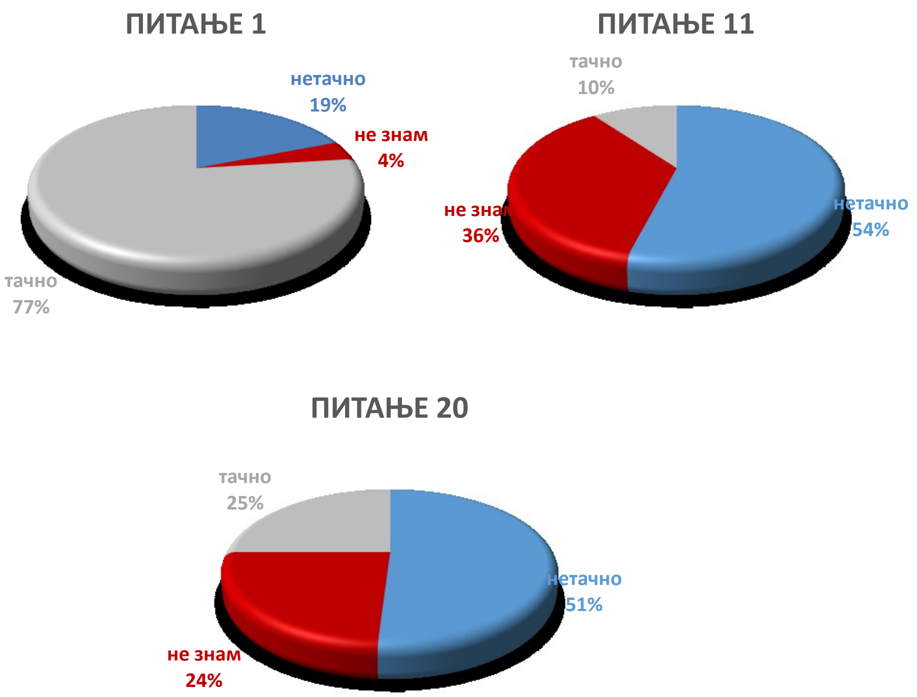


Графикон 24: Укупан скор испитаника за други начин оцењивања теста

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Како бисмо могли да докажемо да ове разлике између укупних скорова два начина оцењивања нису велике, потражили смо корелацију између њих. Добијени коефицијент корелације (***Прилог 6***) износи 0.95 (значајно на нивоу 0.05), што представља врло јаку корелацију. Тиме смо показали да можемо узети било који од ових скорова.

Након што смо израчунали и упоредили укупне скорове два начина оцењивања, посматрали смо постигнућа испитаника по питањима. Стопа одговора на питања била је изузетно висока, око 95 %, што нам је само указивало на то да су се испитаници потрудили да одговоре скоро на сва питања. Међутим, стопа неодговора би се сигурно повећавала са повећањем тежине питања у тесту. Питања су била разнолика, било је оних са вишеструким избором, одговора као и дихотомних питања. Оно што нас је још занимало била је стопа тачних, нетачних одговора, као и број оних који нису дали никакав одговор. Издвојили смо три питања (**Графикон 25**) - оно које је било најлакше (***питање 1***), оно које је било средње тежине (***питање 11***), као и најтеже питање (***питање 20***). Класификацију тежина питања преузели смо из базе задатака PISA (*Programme for International Student Assessment*) OECD (*Organisation for Economic Co-operation and Development*).



Графикон 25: Расподела одговора на три питања различите тежине

Још један корак ка утврђивању валидности различитих начина оцењивања био је да потражимо корелацију оцене на крају осмог разреда из математике са успехом на тесту, као и корелацију поена са пријемног испита из математике и успеха на тесту. Као резултат добили смо да је корелација мала у оба претходна случаја и то:

- за први начин оцењивања корелација оцене и успеха је 0.41 (значајно на нивоу 0.05), а корелација пријемног испита и успеха је 0.34 (значајно на нивоу 0.05) - ***Прилог 7***;
- за други начин оцењивања корелација оцене и успеха је 0.45 (значајно на нивоу 0.05), а корелација пријемног испита и успеха је 0.36 (значајно на нивоу 0.05) - ***Прилог 8***.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

На основу ових резултата могли смо да закључимо да оба начина оцењивања готово исто корелирају са оценом на крају осмог разреда из математике и са пријемним испитом из математике, па су, самим тим, подједнако валидни.

6.2 Пробна тежина питања и способности

У циљу одређивања параметара питања, које ћемо касније користити да добијемо IRT моделе, било је неопходно да на почетку покушамо да одредимо неке пробне вредности тих параметара. Део тачног одговора сваке особе назвали смо *пробна ученичка способност* (TSP – *Tentative Student Proficiency*) и *пробна тежина питања* (TID – *Tentative Item Difficulty*). За добијање пробних вредности параметра у разматрање смо узели својство питања, као и тестирање способности испитаника.

Како бисмо могли да почнемо са одређивањем пробних параметара, требало је на почетку за свако питање одредити вектор који садржи низ од: -1 – нетачан, 0 – није одговорио, 1 – тачан, за први начин оцењивања, као и вектор који садржи низ од: -1, 0, 0.25, 0.5, 0.75, 1, за други начин оцењивања. На овај начин добили смо две различите базе података које су биле спремне за даљу обраду.

Након прављења база, прво смо одредили пробну способност и пробну тежину питања за сваког испитаника, користећи први начин оцењивања. Ради прегледности формирали смо табелу од првих десет испитаника једног одељења из прве базе, у којој смо одредили способност испитаника и тежине питања, што се може видети у **Табели 1**:

rbr	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	TSP
1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1	1	0	0.75
2	1	1	0	1	1	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0.6
3	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0.8
4	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0.7
5	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	0	0	0.75
6	1	1	1	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0.4
7	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	0	1	0	0	0.65
8	0	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0.4
9	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	0	0	0.75
10	1	1	0	0	1	0	0	1	1	1	0	1	1	1	1	1	0	1	0	0	0.6
TID	0.22	0.27	0.27	0.25	0.19	0.41	0.46	0.21	0.28	0.42	0.88	0.14	0.38	0.4	0.59	0.28	0.75	0.36	0.75	0.74	

Табела 1: Однос пробних параметара способности и тежине питања 1

На основу обрађених података запазили смо да је испитаник са ***р. бр. 1*** одговорио тачно на петнаест питања, што значи да се његова способност може изразити нумерички као 75%. Код испитаника ***р. бр. 2*** видимо 12 тачних одговора, те његову способност можемо нумерички изразити као 60%, док испитаника ***р. бр. 3*** има 16 тачних одговора, што је 80%, итд. Такође, из нашег примера примећује се да нисмо могли да говоримо о способности само на основу броја тачних одговора, него је требало да рачунамо и својства питања. У овој табели нема испитаника који су имали исте резултате по врстама. Из овога следи да не можемо дати трајни закључак да су испитаници имали исти ниво способности, јер испитаници са редним бројевима 1, 5 и 9 нису на иста питања дали тачне, односно погрешне одговоре.

Следећи корак био је да се израчунају пробне процене својства питања. Након рачунања имали смо да је, за испитанике у овом одељењу из узорка било најтеже ***питање 11***, па затим ***питање 19*** и ***питање 20***. Ово је значило да 88% испитаника није било у могућности да одговори тачно на ***питање***

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

11. Другим речима, питање је толико тешко да 88% испитаника није могло да одговори на њега. Из ове табеле такође смо могли видети да је за ***питање 2*** и ***питање 3*** исти тежински ниво, али су се она разликовала јер је на ***питање 3*** испитаник, који је високо способан, одговорио погрешно, док је на ***питање 2*** одговорио тачно. Постојала је могућност да је питање било неразумљиво или да је збунило боље испитанике.

rbr	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	TSP
1	1	1	0	1	1	1	1	1	1	0.5	0.5	1	1	1	1	1	0	1	1	0.75	0.84
2	1	1	0.75	1	1	0	0	1	1	1	0.5	1	1	1	1	1	1	0	0	0.5	0.74
3	1	0	0.75	1	1	1	1	1	1	1	0.5	1	1	1	1	1	1	1	1	0.5	0.89
4	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0.25	0	0	0.75	0.75
5	1	1	1	1	1	1	1	1	1	1	0.5	1	0	1	1	1	0	1	0	0.75	0.81
6	1	1	1	1	0	1	0	1	0	0.75	0	1	0	0	0	1	0.5	0	0	0	0.46
7	1	1	1	1	1	1	1	1	1	0.5	0	1	0	0	1	1	0.25	1	0	0.25	0.7
8	0	1	1	1	1	0	0	0	0	0.25	0.5	1	1	1	1	1	0.5	0	0	0.5	0.54
9	1	1	1	0	1	1	1	1	1	1	0.5	1	1	1	0	1	0	1	0	0.75	0.76
10	1	1	0	1	1	0	0	1	1	1	0.5	1	1	1	1	1	0	1	0.5	0.5	0.73
TID	0.13	0.23	0.3	0.1	0.1	0.26	0.48	0.06	0.23	0.23	0.62	0.18	0.23	0.26	0.29	0.1	0.48	0.29	0.58	0.52	

Табела 2: Однос пробних параметара способности и тежине питања 2

Користећи други вектор, из друге базе, одредили смо нове пробне вредности параметара. Могли смо да приметимо (**Табела2**) да су се вредности сваког ученика повећале на индивидуалном нивоу, а тежина питања се смањила, тј. за разлику од првог начина оцењивања повећала се могућност да већи број испитаника одговори тачно на одређено питање. На пример, код испитаника ***р. бр. 8*** способност се повећала и сада је изнад просека од 50%. На основу ових првих пробних резултата за други начин оцењивања, могли смо претпоставити да ће и остали рeзтултати који следе имати веће вредности.

Када смо добили пробне вредности параметара способности и питања, могли смо, како за први, тако и за други начин оцењивања, предвидети вероватноћу делимичног одговарања на питања, и то функцијом:

$$P = {}^{1}\!/_{1 + e^{(-(sposobnost - težina))}}$$

Применом ове формуле за први начин оцењивања добили смо **Табелу 3**:

rbr	ver1	ver2	ver3	ver4	ver5	ver6	ver7	ver8	ver9	ver10	ver11	ver12	ver13	ver14	ver15	ver16	ver17	ver18	ver19	ver20	TSP
1	0.65	0.63	0.58	0.66	0.66	0.61	0.57	0.67	0.64	0.6	0.45	0.67	0.6	0.62	0.61	0.66	0.52	0.61	0.53	0.48	0.75
2	0.62	0.59	0.54	0.62	0.62	0.58	0.53	0.63	0.6	0.56	0.42	0.64	0.56	0.58	0.57	0.62	0.49	0.57	0.5	0.44	0.6
3	0.66	0.64	0.59	0.67	0.67	0.62	0.58	0.68	0.65	0.61	0.47	0.68	0.61	0.63	0.62	0.67	0.54	0.62	0.55	0.49	0.8
4	0.64	0.62	0.57	0.65	0.65	0.6	0.55	0.65	0.62	0.59	0.44	0.66	0.59	0.61	0.59	0.65	0.51	0.59	0.52	0.47	0.7
5	0.65	0.63	0.58	0.66	0.66	0.61	0.57	0.67	0.64	0.6	0.45	0.67	0.6	0.62	0.61	0.66	0.52	0.61	0.53	0.48	0.75
6	0.57	0.54	0.5	0.57	0.57	0.53	0.48	0.58	0.55	0.51	0.37	0.59	0.51	0.53	0.52	0.57	0.44	0.52	0.45	0.39	0.4
7	0.63	0.6	0.56	0.63	0.63	0.59	0.54	0.64	0.61	0.57	0.43	0.65	0.57	0.6	0.58	0.63	0.5	0.58	0.51	0.45	0.65
8	0.57	0.54	0.5	0.57	0.57	0.53	0.48	0.58	0.55	0.51	0.37	0.59	0.51	0.53	0.52	0.57	0.44	0.52	0.45	0.39	0.4
9	0.65	0.63	0.58	0.66	0.66	0.61	0.57	0.67	0.64	0.6	0.45	0.67	0.6	0.62	0.61	0.66	0.52	0.61	0.53	0.48	0.75
10	0.62	0.59	0.54	0.62	0.62	0.58	0.53	0.63	0.6	0.56	0.42	0.64	0.56	0.58	0.57	0.62	0.49	0.57	0.5	0.44	0.6
TID	0.22	0.27	0.27	0.25	0.19	0.41	0.46	0.21	0.28	0.42	0.88	0.14	0.38	0.4	0.59	0.28	0.75	0.36	0.75	0.74	

Табела 3: Вероватноћа тачног одговора 1

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Из табеле могли смо видети да је вероватноћа да испитаник ***р. бр. 3*** на ***питање 18*** одговори тачно 0.62. Пошто је испитаник са ***р. бр. 3*** имао пробну способност 0.8, а пробна тежина ***питања 18*** је била 0.36, закључили смо да овај испитаник има вишу способност у односу на питање. Уколико се десида су, као код испитаника ***р. бр. 9***, пробна способност и тежина питања једнаки, тада је постојало 50% вероватноће да тај испитаник на ***питање 19*** одговори тачно. Још једна могућност ове формуле била је да предвиди како су испитаници прошли кроз сва питања. Тако на пример, приметили смо да се вероватноћа тачног одговора за испитаника ***р. бр. 8*** кретала од 0.37 до 0.57, што је значило да је већину питања урадио слабије од половине.

Када смо претходну формулу вероватноће применили на други начин оцењивања, дошли смо до нових вредности вероватноће за свако питање, односно до **Табеле 4**:

rbr	ver1	ver2	ver3	ver4	ver5	ver6	ver7	ver8	ver9	ver10	ver11	ver12	ver13	ver14	ver15	ver16	ver17	ver18	ver19	ver20	TSP
1	0.67	0.65	0.63	0.68	0.68	0.64	0.59	0.69	0.65	0.65	0.55	0.6	0.65	0.64	0.63	0.68	0.59	0.63	0.56	0.58	0.84
2	0.65	0.62	0.61	0.65	0.65	0.62	0.56	0.66	0.62	0.62	0.53	0.65	0.62	0.62	0.61	0.65	0.56	0.61	0.54	0.55	0.74
3	0.68	0.66	0.64	0.69	0.69	0.65	0.6	0.7	0.66	0.66	0.57	0.66	0.66	0.65	0.65	0.69	0.6	0.65	0.58	0.59	0.89
4	0.65	0.63	0.61	0.66	0.66	0.62	0.57	0.67	0.63	0.63	0.53	0.57	0.63	0.62	0.61	0.66	0.57	0.61	0.54	0.56	0.75
5	0.66	0.64	0.62	0.67	0.67	0.63	0.58	0.68	0.64	0.64	0.55	0.58	0.64	0.63	0.63	0.67	0.58	0.63	0.56	0.57	0.81
6	0.58	0.56	0.54	0.59	0.59	0.55	0.5	0.6	0.56	0.56	0.46	0.5	0.56	0.55	0.54	0.59	0.5	0.54	0.47	0.49	0.46
7	0.64	0.62	0.6	0.65	0.65	0.61	0.55	0.65	0.62	0.62	0.52	0.61	0.62	0.61	0.6	0.65	0.55	0.6	0.53	0.54	0.7
8	0.6	0.58	0.56	0.61	0.61	0.57	0.51	0.62	0.58	0.58	0.48	0.57	0.58	0.57	0.56	0.61	0.51	0.56	0.49	0.5	0.54
9	0.65	0.63	0.61	0.66	0.66	0.62	0.57	0.67	0.63	0.63	0.53	0.66	0.63	0.62	0.62	0.66	0.57	0.62	0.54	0.56	0.76
10	0.65	0.62	0.61	0.65	0.65	0.62	0.56	0.66	0.62	0.62	0.53	0.65	0.62	0.62	0.61	0.65	0.56	0.61	0.54	0.55	0.73
TID	0.13	0.23	0.3	0.1	0.1	0.26	0.48	0.06	0.23	0.23	0.62	0.18	0.23	0.26	0.29	0.1	0.48	0.29	0.58	0.52	

Табела 4: Вероватноћа тачног одговора 2

Овде смо приметили да су вероватноће тачног одговора за свако питање за сваког испитаника повећане, чиме се повећава могућност давања тачног одговора. Ово је било значајно за испитаника са ***р. бр. 8*** јер се оваквим начином оцењивања теста повећала његова могућност давања тачног одговора на сва питања, и то је сада у интервалу од 0.49 до 0.62.

Ово је додатно дало на значају другом начину оцењивања јер се, поред повећања пробне вероватноће делимичног одговора повећала и пробна способност, док се пробна тежина питања смањила, али и даље све на индивидуалном нивоу.

6.2.1. Скала способности и тежине питања

После свих израчунавања параметара питања као и одређивања способности испитаника, постављало се питање да ли се подаци о питањима и способностима испитаника могу објединити на једној заједничкој скали и како.

Међутим, IRT је имао одговор и на ово питање. Још једна позитивна страна IRT јесте постојање могућности да обједини нешто што је на први поглед немогуће објединити. То смо постигли конвертовањем мера тих вредности у такозвану $logit$ скалу.

Како бисмо дошли до $logit$ скале, неопходно је било дефинисати помоћне вредности које су нас довеле до тражене скале. Три битне вредности које су нам биле потребне јесу: $količnik\ šansi, verovatnoća, logit$. Тако је:

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

$$\boldsymbol{količnik\ šansi} = \boldsymbol{Q/P}$$

где је $\boldsymbol{Q}$ представљало број нетачних одговора, а $\boldsymbol{P}$ број тачних одговора за дато питање. Ради провере имали смо променљиву $verovatnoća$ која је представљала вероватноћу давања тачног одговора, па смо могли $količnik\ šansi$ да дефинишемо и као $(\boldsymbol{1}-\boldsymbol{P})/\boldsymbol{P}$.

$$\boldsymbol{logit} = \boldsymbol{log}(\boldsymbol{količnik\ šansi})$$

Када смо добијене податке за први и други начин оцењивања заменили у претходно дефинисане формуле, добили смо податке за питања која су приказана у **Табели 5** и **Табели 6**:

odd	0.29	0.37	0.37	0.33	0.23	0.71	0.86	0.27	0.4	0.75	8.33	0.16	0.61	0.68	1.47	0.4	3.2	0.58	3.2	3
prob	0.77	0.72	0.72	0.75	0.81	0.58	0.53	0.78	0.71	0.57	0.1	0.85	0.61	0.59	0.4	0.71	0.23	0.63	0.23	0.25
logit	-0.53	-0.42	-0.42	-0.47	-0.62	-0.14	-0.06	-0.56	-0.39	-0.12	0.92	-0.77	-0.21	-0.16	0.16	-0.39	0.50	-0.23	0.50	0.47

Табела 5: logit за питања (први начин оцењивања)

odd	0.29	0.38	0.23	0.29	0.22	0.62	0.87	0.27	0.42	0.38	2.14	0.17	0.53	0.65	1.40	0.40	1.80	0.52	2.43	1.24
prob	0.77	0.73	0.82	0.77	0.82	0.62	0.54	0.79	0.70	0.72	0.32	0.86	0.65	0.61	0.42	0.71	0.36	0.66	0.29	0.45
logit	-0.53	-0.42	-0.65	-0.53	-0.66	-0.21	-0.06	-0.56	-0.37	-0.42	0.33	-0.78	-0.28	-0.19	0.15	-0.40	0.26	-0.28	0.39	0.09

Табела 6: logit за питања (други начин оцењивања)

Када је $logit$ негативан, тј., када су вредности испод нуле, тада смо то питање сматрали лаким, док смо питање које има позитиван $logit$ сматрали тешким питањем. Уколико би се десило да имамо исти број тачних и нетачних одговора, за то питање $količnik\ šansi$ био би једнак 1, а $logit$ једнак нули, па бисмо тиме имали питање средње тежине.

У IRT не постоје вредности $logit$ скале за екстремно тешка и за екстремно лака питања. Таквим питањима сматрали смо она код којих $logit$ не постоји при рачунању $log(količnik\ šansi)$ али, уколико су постојали такви екстремни случајеви у IRT, њих смо искључивали.

Такође, на основу ових дефиниција добили смо и податке за способност, за оба начина оцењивања, које смо приказли у **Табели 7** и **Табели 8:**

odd	3.00	1.50	4.00	2.33	3.00	0.67	1.86	0.67	3.00	1.50	3.00	1.50	1.22	1.22	1.50	5.67	5.67	3.00	2.33	0.82
prob	0.75	0.60	0.80	0.70	0.75	0.40	0.65	0.40	0.75	0.60	0.75	0.60	0.55	0.55	0.60	0.85	0.85	0.75	0.70	0.45
logit	0.48	0.18	0.60	0.37	0.48	-0.18	0.27	-0.18	0.48	0.18	0.48	0.18	0.09	0.09	0.18	0.75	0.75	0.48	0.37	-0.09

Табела 7: logit за способност испитаника (први начин оцењивања)

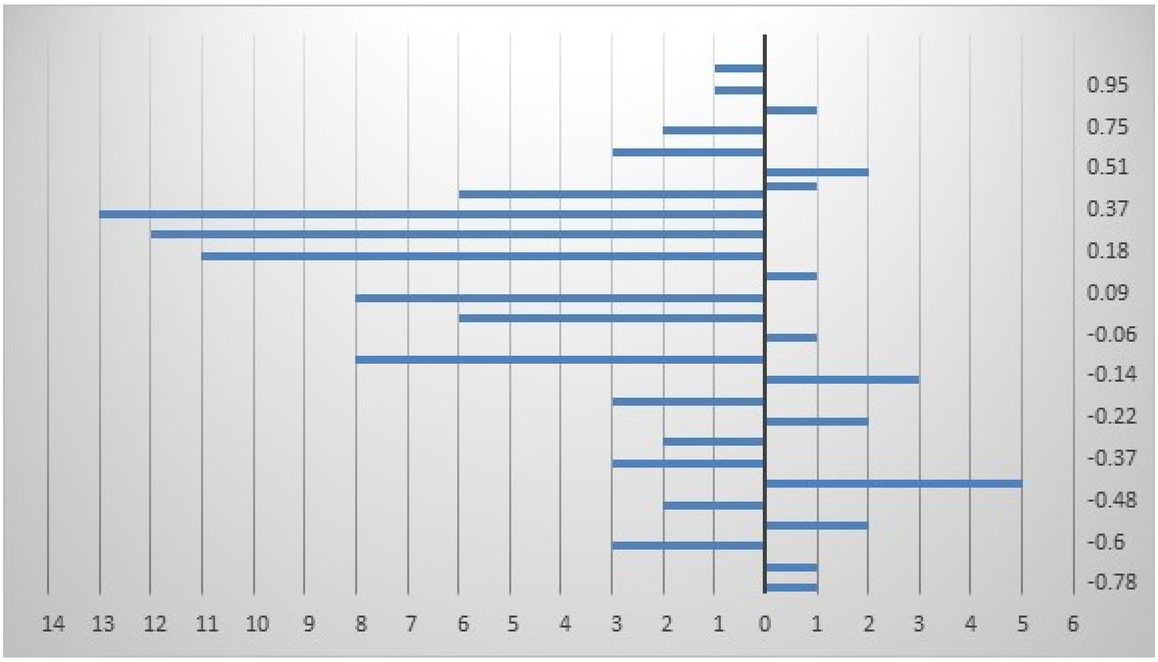
odd	0.51	1.76	1.67	3.21	2.20	3.00	1.29	0.45	1.76	1.05	1.00	0.51	0.33	0.31	0.90	0.82	0.86	0.21	0.78	1.22
prob	0.66	0.36	0.38	0.24	0.31	0.25	0.44	0.69	0.36	0.49	0.50	0.66	0.75	0.76	0.53	0.55	0.54	0.83	0.56	0.45
logit	-0.29	0.25	0.22	0.51	0.34	0.48	0.11	-0.34	0.25	0.02	0.00	-0.29	-0.48	-0.51	-0.04	-0.09	-0.07	-0.67	-0.11	0.09

Табела 8: logit за способност испитаника (други начин оцењивања)

За одређивање $logit$ вредности код способности испитаника имали смо да је $\boldsymbol{količnik\ šansi} = \boldsymbol{P/Q}$ или $\boldsymbol{P}/(\boldsymbol{1}-\boldsymbol{P})$. Позитиван $logit$ нам је говорио да је испитаник на високом нивоу способности, а негативан $logit$ да се испитаник налазио на нижим нивоима способности.

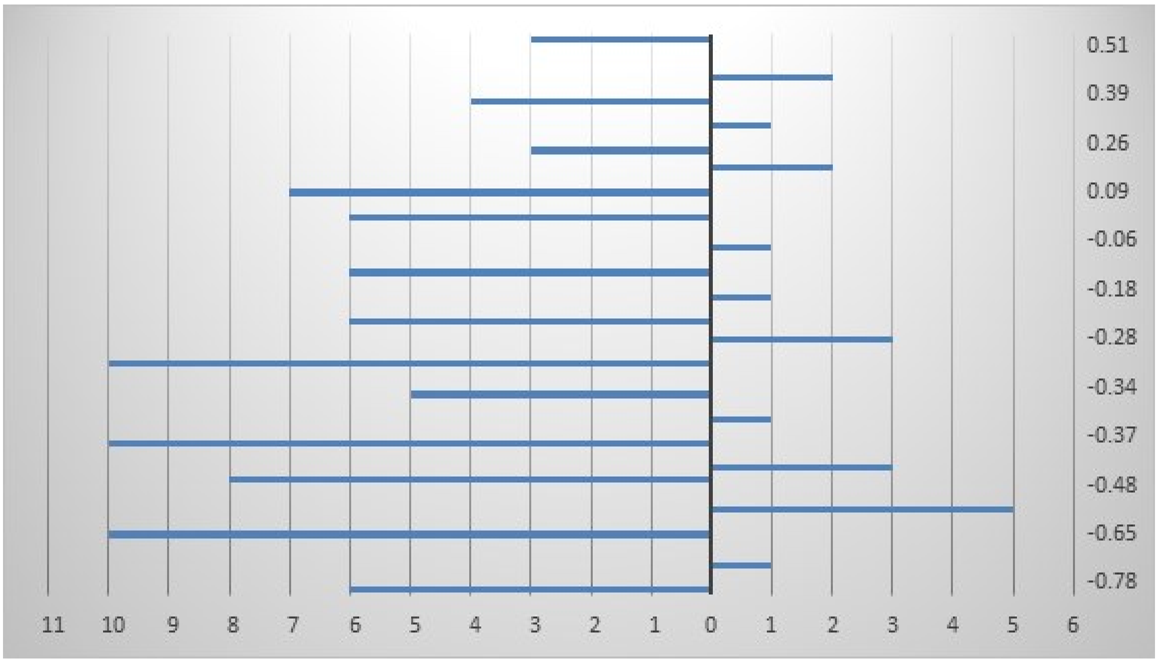
Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Из ових табела добили смо тражену $logit$ скалу за оба начина оцењивања, **Графикон 26** и **Графикон 27:**



Графикон 26: $logit$ скала за први начин оцењивања

Графикон 26 и **графикон 27** обично се називају *мапама испитаник-питање*. На тим мапама имамо вредности параметара питања и вредности за способност испитаника. Када смо подвукли линију у нули, добили смо средњу вредност и за способност испитаника и за тежину питања. На тај начин, испитаници који су били испод линије прошли су само половину питања. Испитаници који су били на горњој десној страни су бољи и способнији од питања на доњој левој страни.



Графикон 27: $logit$ скала за други начин оцењивања

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

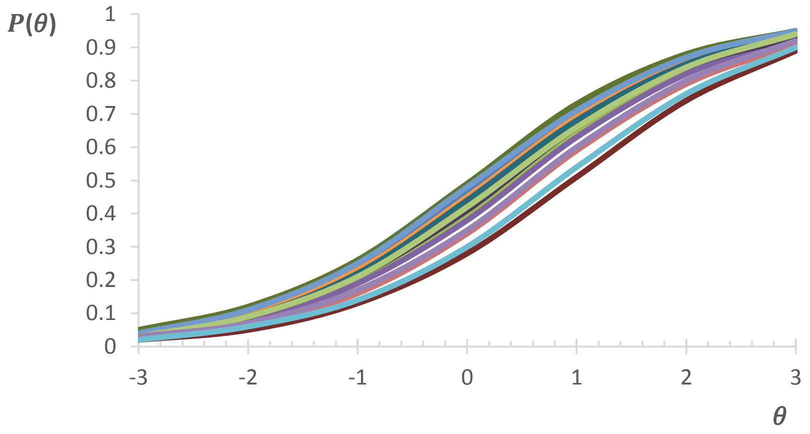
6.3 Поређење три логистичка модела

Када смо кроз неколико итерација, да ли кроз пробне вредности или емпиријским поређењем података, добили параметре питања који нам одговарају, могли смо да приступимо тражењу IRT модела у коме ће наши подаци дати најбоље резултате.

Као што смо видели из њихових дефиниција функција, модели IRT разликују се по броју параметара. Тиме ће се и резултати који се добију обрадом истих података разликовати јер се користе различити модели. Када се раде спецификације података, неопходно је да се изабере строго један модел, који ће најбоље представљати посматране податке. Пошто нисмо били сигурни који модел да изаберемо, посматрали смо сва три модела упоредо и тумачили тако добијене резултате.

Анализирањем података и рачунањем свих параметара, добили смо вредности или разлике неких вредности које су мале или доста мале да се готово могу занемарити и, уместо да за све нивое способности користимо различите вредности параметра, користићемо једну вредност која ће бити јединствена за све нивое.

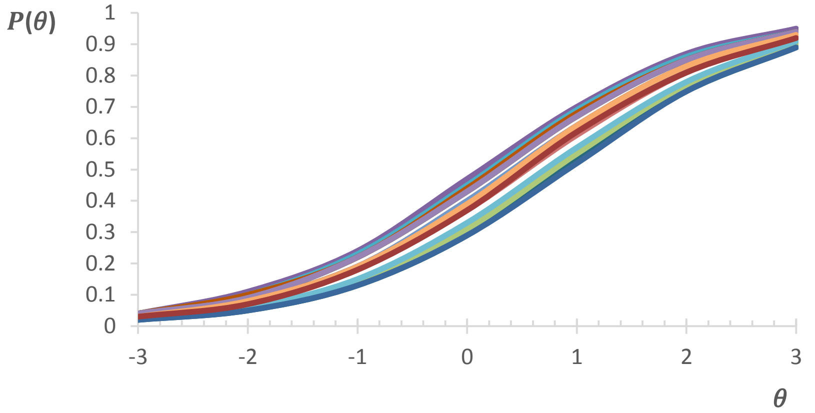
Користећи први начин оцењивања теста израчунали смо тежине свих питања са теста и одредили смо нивое способности. На основу тога применили смо једначину за једнопараметарски логистички модел, чиме смо добили карактеристичне криве 1PL модела за свако питање (**Графикон 28**):



Графикон 28: Карактеристичне криве питања 1PL модела – први начин оцењивања

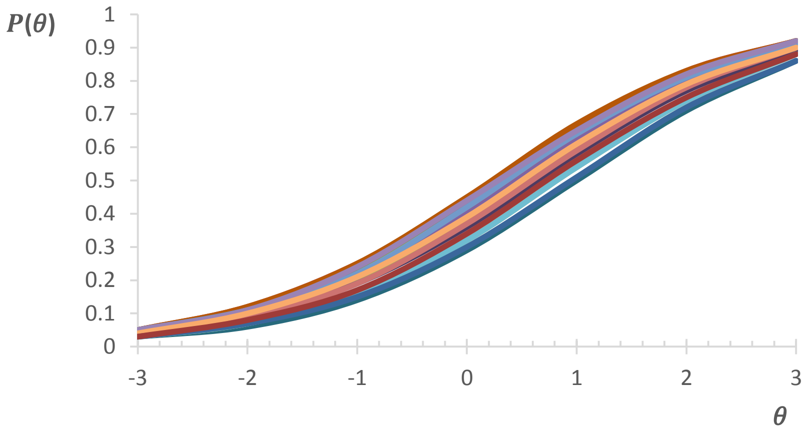
Са овог графикона закључили смо да како расте ниво способности, повећава се и вероватноћа давања тачног одговора. Исто се могло рећи и за карактеристичне криве 1PL модела за други начин оцењивања **(Графикон 29)**. Чак су и вредности кривих у оба оцењивања биле готово исте. Такође, урадили смо и статистичку значајност разлика параметара тежине за оба начина оцењивања. Вредност статистика била је: $\boldsymbol{F = 1.68}$ $(\boldsymbol{p > 0.05})$. Тиме смо закључили да не постоје статистички значајне разлике између параметара тежине ***(Прилог 9).***

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 29: Карактеристичне криве питања 1PL модела – други начин оцењивања

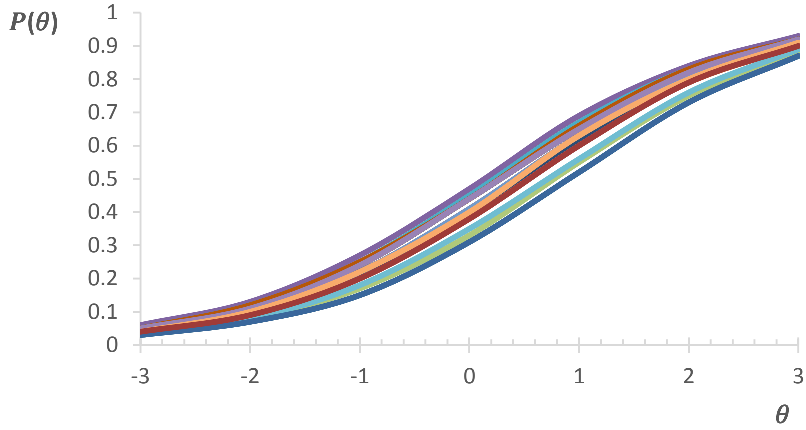
Када смо добили тежинске параметре приступили смо рачунању дискриминационог параметра и то према његовој дефиницији. На тај начин смо добили све параметре за израчунавање функције карактеристичне криве 2PL модела. Израчунавањем дискриминационих параметара добили смо да се они не разликују драстично за ова два начина оцењивања. Анализирањем питања може се рећи да имамо умерену дискриминацију. Оправдање за узимање исте вредности параметра дискриминације од 0.9 проналазимо у закључку да не постоје статистички значајне разлике између дискриминационих параметара ова два начина оцењивања $(\boldsymbol{F = 2.18, p > 0.05})$ - ***Прилог* 10**. Сходно свему наведеном, добили смо графикон **Графикон 30** и **Графикон 31:**



Графикон 30: Карактеристичне криве питања 2PL модела – први начин оцењивања

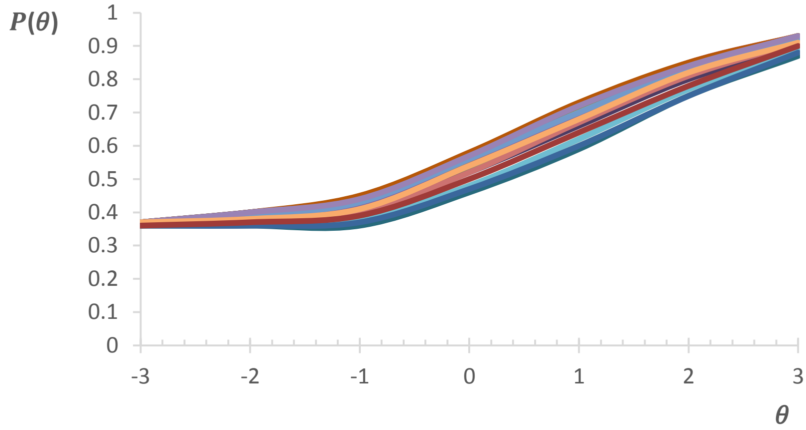
Са **Графикона 30** и **Графикона 31** запазили смо да су све криве, као и код 1PL модела, паралелне и да није дошло до међусобног пресецања кривих јер су сви дискриминациони параметри позитивни. Када смо поредили вредности ових кривих са вредностима кривих 1PL модела, уочили смо да се оне не разликују много.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 31: Карактеристичне криве питања 2PL модела – други начин оцењивања

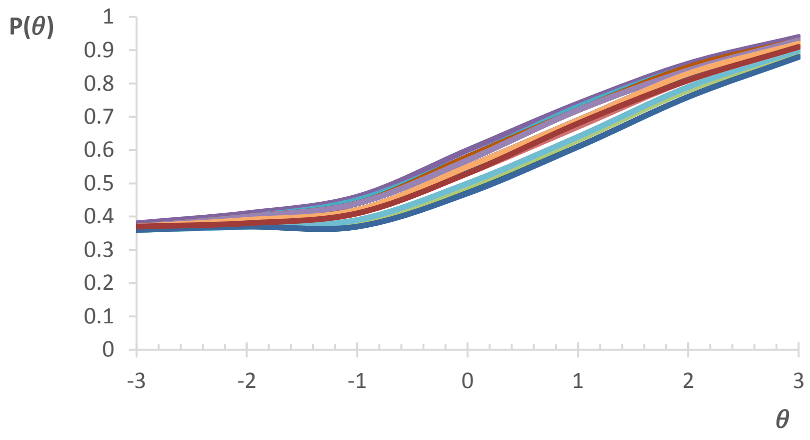
Остало нам је да, поред ова два параметра питања, пронађемо и трећи параметар како бисмо могли да израчунамо вредности функција 3PL модела и тиме добијемо карактеристичне криве питања овог модела. Применом дефиниције за налажење параметра погађања c, добили смо различите вредности у распону од 0.1 до 0.6 за сваки ниво способности код оба начина оцењивања. Као и код претходна два параметра, и овде смо тестирали зависне узорке. Као резултат тога постоји статистички значајна разлика параметара погађања: $(\boldsymbol{F = 53.98, p < 0.05})$ - ***Прилог 11***.



Графикон 32: Карактеристичне криве питања 3PL модела – први начин оцењивања

Како смо добили резултат да су разлике значајне, то је значило, да је код једног начина оцењивања могућност погађања повећана. Други начина оцењивања остављао је већу могућност да се нешто погоди. Разлог за то произилазио је из чињенице да је испитаник код првог начина оцењивања, да би добио поен, морао све да погоди, док би код другог начина оцењивања, уколико би урадио само мало, добио бар део поена, што је представљало већу вероватноћу него да погоди све. Такође, погађање није баш добро за нас, јер не знамо да ли су се испитаници двоумили који одговор да дају, или одговор нису уопште ни знали него су ишли „на срећу“. У ствари, одлуку о давању неодговора испитаници су доносили на основу процена које се односе на поређење степена властитог незнања са вероватноћом да ће одговор ипак бити тачан.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs



Графикон 33: Карактеристичне криве питања 3PL модела – други начин оцењивања

Упоредивши ова два графикона (**Графикон 32** и **Графикон 33**) са графиконима 1PL i 2PL модела, видели смо да су њихове вредности карактеристичних кривих веће у односу на вредности кривих ова два модела.

На основу свега претходно изнетог, могли смо да закључимо, да су прва два модела IRT неподесна, тј. мање су осетљива јер не виде никакву разлику између два начина оцењивања. Једино је трећи модел IRT дао неку разлику, а то је разлика у случајном погађању, па је самим тим он био најосетљивији на разлике у оцењивању јер је нудио највише вредности параметара.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

7 Закључак

Свођењем свих резултата добијених приликом обраде података, уочавамо да се не могу баш извући једноставни закључци. Међутим, за неке ствари ипак можемо да потврдимо одређена наша тврђења. Један од првих закључака био је да не постоје статистички значајне разлике укупних скорова између тестова код првог начина оцењивања. Такође, исти закључак извели смо и за други начина оцењивања теста. Затим смо утврдили да постоји врло висока корелација између успеха ученика на тесту за два различита начина оцењивања. Дакле, могли смо да закључимо да редослед задатака није утицао на решавање теста у глобалу.

Закључили смо да оба начина оцењивања готово исто корелирају са оценом на крају осмог разреда из математике и са пријемним испитом из математике, чиме смо само поткрепљивали тврдњу да су оба начина оцењивања подједнако валидна.

Мерењем значајности параметара тежине, као и значајности параметара дискриминативности, утврдили смо да не постоје статистички значајне разлике на овим параметрима између два начина оцењивања. Такође, и за параметре погађања мерили смо значајност разлика и утврдили да разлике постоје између параметара погађања првог и другог начина оцењивања. Тиме смо дошли до закључка да је трећи модел IRT најосетљивији на разлике у оцењивању, док су прва два модела IRT мање осетљива и да не дају разлику између два начина оцењивања.

На основу свега претходног могли смо да закључимо да који год од ова два начина оцењивања применимо, успех испитаника биће готово исти и на исти начин ће бити повезан са спољашњим критеријумима попут оцене или успеха са пријемног. Такође, ако нешто разликује два начина оцењивања то је вероватноћа случајног погађања која се повећава при оцењивању са делимичним поенима.

Дакле, оцењивање делимичним поенима даје подједнако добру процену као и $0-1$ оцењивање, али је сложеније, тражи више времена, а повећава могућност случајног погађања. Према томе, за овај вид бодовања теста можемо рећи да је подједнако добар или лошији од бинарног бодовања (решио – није решио задатак).

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

ПОПИС ГРАФИКОНА И ТАБЕЛА

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Литература

[01] David S. Moore, W. I. (2006). *Statistics concepts and controversies.* The Ohio State University USA.

[02] Lohr, S. L. (2010). *Sampling: Design and Analysis.* Arizona State University USA.

[03] Baker, F. B. (2001). *The Basics of Item Response Theory.* University of Wisconsin USA.

[04] Partchev, I. (2004). *A visual guide to item response theory.* Friedrich-Schiller-Universitat Jena.

[05] Chong Ho Yu, P. (2013). *A Simple Guide to the Item Response Theory (IRT) and Rasch Modeling.* Arizona State University USA.

[06] Chong Ho Yu, A. J.-P. (2008). *A Non-Technical Approach for Illustrating Item Response Theory.* Arizona State University USA.

[07] Трнавац Н., Ђ. Ј. (1998). *Педагогија.* Београд: Научна књига.

[08] OECD. (2009). *Take the Test Sample Questions from OECD's PISA Assessments.* Paris.

[09] OECD. (2008). *PISA 2006. Technical Report. OECD.* Paris.

[10] „Службени Гласник Републике Србије" број 104/09.

[11] Фајгељ, С. (2005). *Психометрија, метод и теорија психолошког мерења.* Београд: Центар за примењену психологију.

[12] Фајгељ, С. Ј. (2013). *Погађање у когнитивном тестирању.* Нови Сад: Филозфски факултет.

[13] Вујић, С. Б. (2010). *Однос између похађања предшколског образовања и школског успеха успеха ученика и ученица и могућности унапређења предшколског образовања у Србији.* Психолошка истраживања, 14, 105-140.

[14] Тењовић, Л. (2005). *Статистика у психологији.* Београд: Центар за примењену психологију.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

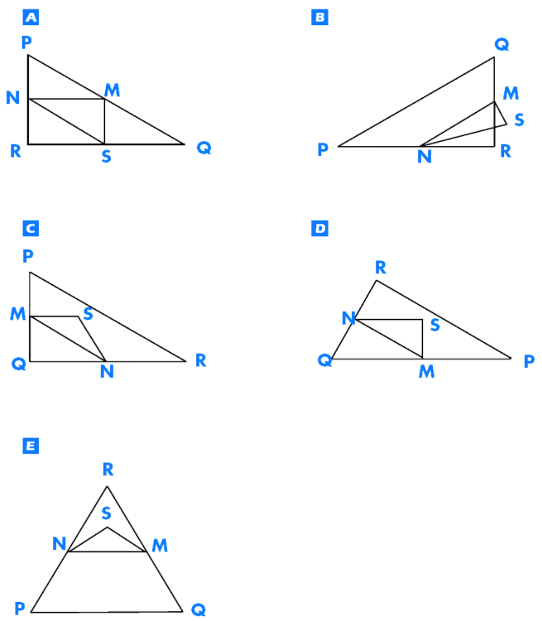
Прилог 1

([08] OECD, 2009)

Тест знања из математике 1

Школа:
Одељење:
Пол: М Ж
Оцена из математике на крају 8. разреда:
Број поена са пријемног испита:

Питање 1. Троугао



Питање:
Којој фигури на слици одговара следећи опис:

Троугао PQR има прав угао код темена R, страница RQ је мања од странице PR. M је средиште од PQ, а N је средиште од QR. S је тачка унутар троугла. Страница MN је већа од странице MS.

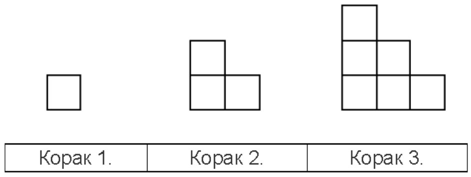
Одговор:

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 2. Макета

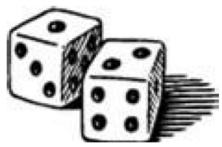
Роберт прави макету степеница користећи квадрате, и то у три корака, на следећи начин:



Питање: Колико квадрата треба да користи за четврти корак?

Одговор: .

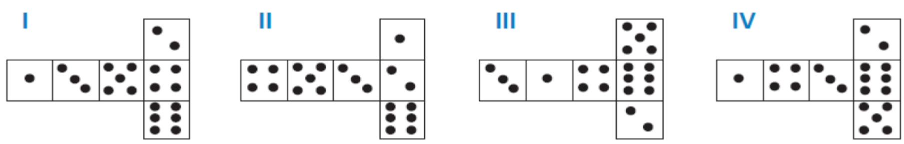
Питање 3. Коцкице



Са десне стране дате су две коцкице.
Када се сабере број тачака на супротним странама коцкице увек се добије седам.

Коцкице можете направити једноставно сечењем, лепљењем и савијањем картона.
На слици испод дате су четири исечене коцкице.

Питање:
Који од наведених облика може бити пресавијен у коцкицу, тако да збир на супротним странама буде седам ?



Облик	Који облик испуњава захтев?
I	Да. / Не.
II	Да. / Не.
III	Да. / Не.
IV	Да. / Не.

Питање 4. Избор

У пицерији можете наручити пицу са два основна састојка: сир и парадајиз.
Такође у овом ресторану можете направити и своју пицу са екстра додацима.
Екстра додаци које можете изабрати су: маслине, шунка, печурке, печеница.

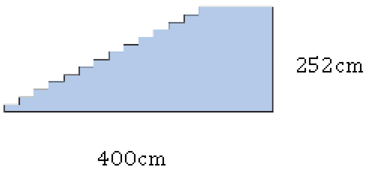
Марко жели да наручи пицу са два различита екстра додатка.

Питање:
На колико различитих начина Марко може да направи своју пицу?

Одговор:

Питање 5. Степенице

На слици су дате степенице које имају 14 степеника. Укупна висина степеница је 252cm.



Питање:
Колика је висина сваког од 14 степеника?

Одговор:

Питање 6. Тест

У једној основној школи наставница даје тестове који носе 100 поена. Филип је на прва четири теста имао просек 60 поена по тесту. На петом тесту имао је 80 поена.

Питање:
Који је просек поена Филип имао после свих пет тестова?

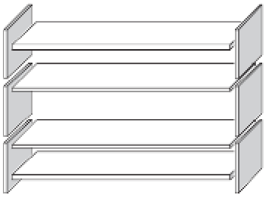
Одговор:

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 7. Полице

За склапање једне полице за књиге неоходне су следеће компоненте:

4 дуже плоче
6 краћих плоча
12 краћих шрафова
2 дужа шрафа
14 матица



Питање:
Колико полица може да се направи од: 26 дужих плоча, 33 краће плоче, 200 краћих шрафова, 20 дужих шрафова и 510 матица?

Одговор:

Питање 8. Размена

Меи-Линг из Сингапура се спрема да иде на размену студената у Јужну Африку. Пре пута она је морала да замени сингапурске доларе (сгд) у јужноафрички ранд (зар).

Меи-Линг је сазнала да је девизни курс између долара и ранда:

1сгд = 4,2 зар

Меи-Линг је заменила 3000 сингапурских-долара у јужноафричке ранде.

Питање:
Колико је јужноафричких ранда Меи-Линг добила?

Одговор:

После три месеца Меи-Линг се враћа у Сингапур али јој је остало 3900 зар. Треба да их промени у сингапурске доларе при чему је сада однос:

1 сгд = 4.0 зар

Питање:
Колико сингапурских долара Меи-Линг може да добије?

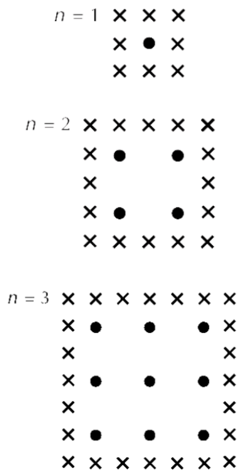
Одговор:

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 9. Стабла јабуке

Земљорадник је засадио стабла јабуке у облику квадрата. У циљу заштите дрвећа од ветра, око њих је поставио стабала четинара. Све ово је претстављено дијаграмом:

✕ = стабло четинара
● = стабло јабуке



Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање:
Попуните табелу:

n	Број стабала јабука	Број стабала четинара
1.	1	
2.		
3.		
4.		
5.		

Постоје две формуле које можете да користите за израчунавање броја стабала јабуке и број стабала четинара, тако да је:

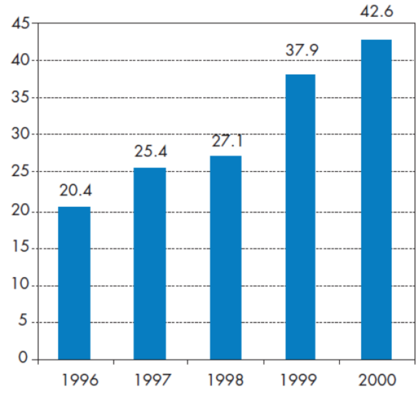
Број стабала јабуке = n^2
Број стабала четинара = 8*n , где је **n** број редова стабала јабуке.

Питање:
Пронаћи вредност n, за коју је број стабала јабуке једнак броју стабала четинара?

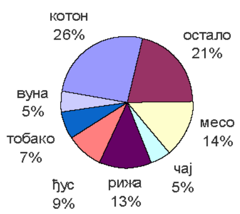
Одговор:

Питање 10. Извоз

На графицима су приказани подаци о извозу са Новог Зеланда, у милионима долара, у периоду од 1996. до 2000. и у 2000-ој.



Промет извоза Новог Зеланда у 2000.



Питање:
Колика је укупна вредност (у милионима долара) извоза са Новог Зеланда у 1998-ој?

Одговор:

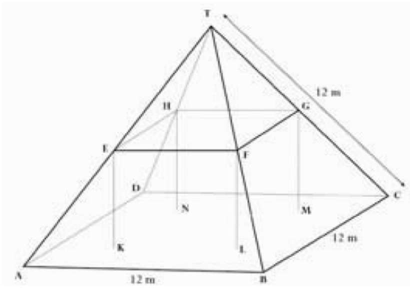
Питање:
Колика је вредност извезеног ђуса са Новог Зеланда у 2000-ој?

a. 1,8 милиона
b. 2.3 милиона
c. 2.4 милиона
d. 3,4 милиона
e. 3,8 милиона

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 11. Фарма

Дата је слика једне фарме, чији је кров облика пирамиде. Испод те слике је математички модел крова фарме, правилна четворострана пирамида, са унетим мерама.



Поткровље је база правилне четворостране пирамиде. Греде које држе кров су ивице коцке EFGHKLMN.
E је средиште од AT, F је средиште од BT, G је средиште од CT а H је средиште од DT.
Све ивице пирамиде имају дужину 12 метара.

Питање:
Израчунати површину поткровља ABCD.
Површина поткровља ABCD = ______ m^2.

Питање:
Израчунати дужину EF, једну од ивица коцке.
Дужина EF = ______ m.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 12. Најбољи ауто

Ауто магазин сваке године бира аутомобил године. У табели је дато пет аутомобила са оценама њихових карактеристика, који су по одабиру овог магазина најбољи:

Аутомобил	Безбедност (S)	Потрошња горива (F)	Изглед спољашњости (E)	Опрема (T)
Ca	3	1	2	3
M2	2	2	2	2
Sp	3	1	3	2
N1	1	3	3	3
KK	3	2	3	2

Оцене се тумаче на следећи начин:
3 бода = Одличан
2 бода = Добар
1 бод = Лош

За израчунавање укупне оцене аутомобила, ауто магазин користи следећу формулу:

Укупан резултат = (3*S) + F + E + T

Питање:
Израчунати укупну оцену за аутомобил „Ca“?
Одговор:

Произвођач аутомобила „Ca“ мисли да је правило за укупан скор био неправедан.

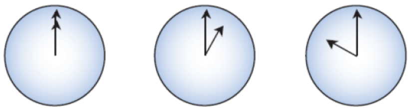
Питање:
Запишите правило за израчунавање укупног скора, тако да „Ca“ буде најбољи ауто. Ваше правило треба да обухвати све четири променљиве, тако да се слободан простор у једначини испод правилно попуни позитивним бројевима.

Укупан скор = ... *S + ... * F + ... * E + ...* T

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 13. Интернет дописивање

Марк (Сиднеј, Аустралија) и Хас (Берлин, Немачка) често се дописују путем интернета. Како би могли да се дописују морају се конектовати у исто време.
Да би пронашао одговарајуће време за дописивање Марк је упоређивао следећа времена:



Гринич : 00:00 поноћ | Берлин: 1:00 јутро | Сиднеј: 10:00 јутро

Питање:
Које је време у Берлину ако је у Сиднеју 7:00 после подне?

Одговор:

Марк и Ханс нису доступни на мрежи између 9:00 ујутру и 4:30 после подне по њиховом локалном времену, када су у школи. Такође нису доступни и између 11:00 после подне и 7:00 ујутру по њиховом локалном времену јер спавају.

Питање: Када је најбоље време за дописивање између Марка и Ханса?

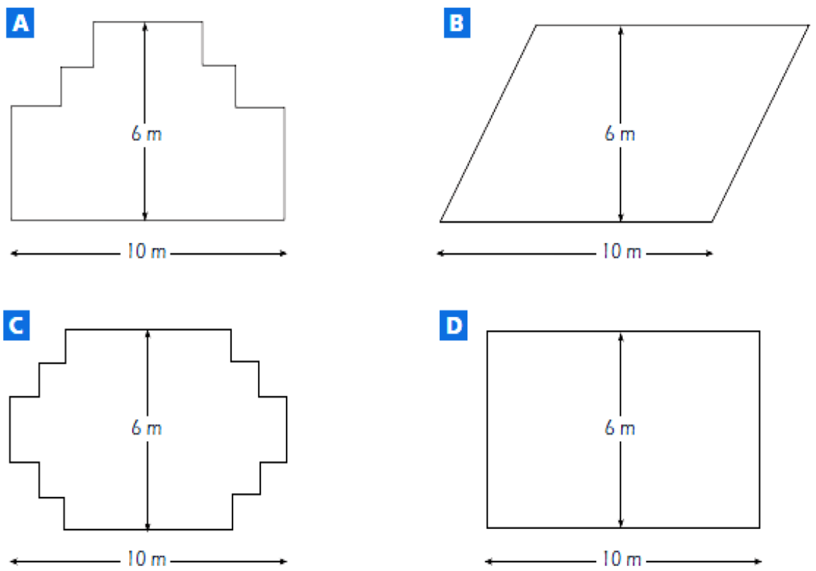
У табели написати локална времена:

Место	Време
Сиднеј	
Берлин	

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Питање 14. Столар

Столар има 32 метра дрвне грађе и жели да огради башту са цвећем. Разматра следеће облике башта:



Питање:
Који од понуђених облика башта столар може да огради са 32 метра дрвне грађе?

Попунити табелу:

Облици баште	Која башта са цвећем може бити ограђена? (може / не може)
Облик А	
Облик B	
Облик C	
Облик D	

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 2

br_testa	ime_testa	skola	odeljenje	redni broj ucenika	pol	ocena_iz_8_razreda	broj_poena	trougao	maketa	kockice	izbor	stepenice	test	police	razmena		stabla_jabuke		izvoz		farma		najbolji_auto		internet_dopisivanje		stolar
															a	b	a	b	a	b	a	b	a	b	a	b	
1	Test1	Gimnazija	I/3	1	0	5	18	1	1	-1	1	1	1	1	1	1	-1	-1	1	1	1	1	1	0	1	1	-1
1	Test1	Gimnazija	I/3	2	0	5	20	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	1	1	1	0	-1	-1	-1
1	Test1	Gimnazija	I/3	3	1	5	20	1	-1	-1	1	1	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	-1
1	Test1	Gimnazija	I/3	4	1	5	19	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	-1	-1	0	-1
1	Test1	Gimnazija	I/3	5	1	4	13.5	1	1	1	1	1	1	1	1	1	1	-1	1	-1	1	1	1	0	1	0	-1
1	Test1	Gimnazija	I/3	6	1	5	19.5	1	1	1	1	0	1	0	1	-1	-1	0	1	-1	0	0	1	-1	-1	0	-1
1	Test1	Gimnazija	I/3	7	1	5	15	1	1	1	1	1	1	1	1	1	-1	-1	1	0	-1	1	1	-1	1	-1	-1
1	Test1	Gimnazija	I/3	8	0	4	16	-1	1	1	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	-1	-1	-1
1	Test1	Gimnazija	I/3	9	0	5	17	1	1	1	1	1	1	1	1	1	1	-1	1	1	1	0	1	0	1	0	-1
2	Test2	Gimnazija	I/3	10	1	5	18	1	1	-1	-1	1	-1	-1	1	1	1	-1	1	1	1	1	1	-1	1	-1	-1
2	Test2	Gimnazija	I/3	11	0	4	16	1	1	-1	1	1	1	1	1	1	-1	-1	1	1	1	1	1	0	1	1	-1
2	Test2	Gimnazija	I/3	12	1	5	18	1	1	-1	1	1	1	1	1	1	-1	-1	1	1	0	0	1	1	-1	0	-1
2	Test2	Gimnazija	I/3	13	1	5	19	1	-1	1	1	1	0	-1	1	1	-1	-1	1	-1	1	0	1	1	1	0	-1
2	Test2	Gimnazija	I/3	14	1	4	19	1	1	-1	1	1	1	0	1	1	1	1	1	-1	-1	0	1	0	-1	0	-1
2	Test2	Gimnazija	I/3	15	1	5	19.5	-1	1	1	-1	1	-1	1	1	1	1	-1	1	1	1	1	0	1	-1	0	0
2	Test2	Gimnazija	I/3	16	0	5	18.5	1	1	-1	1	1	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	-1
2	Test2	Gimnazija	I/3	17	0	5	16	1	-1	1	1	1	-1	1	1	1	1	-1	1	1	1	1	1	1	1	1	1
3	Test3	Gimnazija	I/3	18	1	5	20	1	1	1	1	1	1	-1	1	1	1	-1	1	-1	1	1	1	0	1	1	-1
3	Test3	Gimnazija	I/3	19	0	5	20	1	1	-1	1	1	1	-1	1	1	1	-1	1	1	1	1	1	-1	1	-1	-1
3	Test3	Gimnazija	I/3	20	0	5	20	-1	1	1	1	1	-1	-1	1	-1	1	-1	1	0	-1	0	1	0	1	0	-1
3	Test3	Gimnazija	I/3	21	0	5	17	1	1	1	1	1	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	-1
3	Test3	Gimnazija	I/3	22	1	5	18	1	1	0	1	-1	1	0	1	1	1	-1	1	1	0	0	1	0	1	1	1
3	Test3	Gimnazija	I/3	23	1	5	18	1	1	1	1	1	1	1	1	-1	-1	-1	1	0	1	1	1	0	1	1	-1
3	Test3	Gimnazija	I/3	24	1	5	17.5	1	-1	-1	1	1	-1	-1	1	1	1	-1	1	1	1	1	1	1	1	-1	1
4	Test4	Gimnazija	I/3	25	0	4	17	1	1	1	1	1	1	0	0	0	1	-1	1	-1	1	0	-1	1	0	0	0
4	Test4	Gimnazija	I/3	26	0	5	20	1	1	1	1	1	0	1	1	0	1	-1	1	0	1	1	1	0	1	1	-1
4	Test4	Gimnazija	I/3	27	1	5	18.5	1	-1	-1	1	1	1	-1	1	1	1	-1	1	1	1	1	1	-1	1	-1	-1
4	Test4	Gimnazija	I/3	28	0	5	20	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
4	Test4	Gimnazija	I/3	29	1	4	15	-1	-1	1	1	0	1	1	1	1	-1	-1	1	1	-1	0	1	0	-1	1	0
4	Test4	Gimnazija	I/3	30	1	5	19	1	1	1	1	1	1	-1	1	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1
4	Test4	Gimnazija	I/3	31	1	5	20	1	-1	-1	1	1	1	-1	1	1	1	-1	1	1	1	1	1	1	1	-1	1
1	Test1	Gimnazija	I/1	32	1	3	11	1	-1	1	1	1	1	1	1	1	-1	-1	1	1	1	0	1	0	1	-1	-1
1	Test1	Gimnazija	I/1	33	0	3	16	-1	1	1	1	1	-1	0	0	0	-1	0	0	0	0	0	0	0	-1	0	-1
1	Test1	Gimnazija	I/1	34	0	4	18	1	-1	1	0	1	-1	1	0	0	1	0	1	1	0	0	0	0	1	-1	1
1	Test1	Gimnazija	I/1	35	1	3	15	1	1	1	1	1	1	1	1	1	-1	-1	1	1	0	0	1	1	1	0	0
1	Test1	Gimnazija	I/1	36	1	5	19	-1	1	-1	1	1	1	-1	1	1	-1	0	1	1	1	0	1	1	-1	0	-1
1	Test1	Gimnazija	I/1	37	1	5	19	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	0	1	0
2	Test2	Gimnazija	I/1	38	1	5	20	1	-1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	-1	-1	-1	0
2	Test2	Gimnazija	I/1	39	1	4	18	-1	1	-1	1	1	1	-1	1	1	-1	0	1	1	0	0	1	-1	0	-1	-1
2	Test2	Gimnazija	I/1	40	1	2	12	-1	1	1	1	1	0	0	0	0	-1	0	0	-1	0	0	0	0	-1	-1	-1
2	Test2	Gimnazija	I/1	41	1	5	18	-1	-1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	-1	1	-1	1
2	Test2	Gimnazija	I/1	42	0	3	16	1	1	-1	0	0	0	0	1	1	0	0	1	1	1	0	1	-1	-1	0	-1
2	Test2	Gimnazija	I/1	43	1	4	16.5	1	1	1	0	1	0	1	1	1	1	0	1	1	1	0	1	0	-1	-1	0
3	Test3	Gimnazija	I/1	44	1	5	20	1	-1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	-1	-1
3	Test3	Gimnazija	I/1	45	1	5	17	1	-1	-1	1	0	-1	1	1	1	-1	0	1	1	0	0	0	0	1	0	1
3	Test3	Gimnazija	I/1	46	0	5	17.5	1	1	1	1	1	0	1	1	0	0	0	1	0	0	0	1	-1	1	0	-1
3	Test3	Gimnazija	I/1	47	0	3	13.5	1	1	-1	1	1	1	1	1	1	1	-1	1	1	0	0	1	-1	-1	-1	1
3	Test3	Gimnazija	I/1	48	1	5	18.5	1	0	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	-1	0
3	Test3	Gimnazija	I/1	49	1	4	18	1	1	0	1	1	1	-1	0	0	-1	-1	0	0	1	1	1	-1	1	0	0
4	Test4	Gimnazija	I/1	50	0	5	17.5	1	-1	-1	-1	-1	0	-1	-1	1	1	0	1	0	-1	0	1	-1	1	-1	1
4	Test4	Gimnazija	I/1	51	1	4	17	-1	1	1	-1	1	-1	1	1	1	-1	0	1	-1	1	1	1	1	-1	-1	1
4	Test4	Gimnazija	I/1	52	1	5	15	1	1	1	1	0	1	1	1	1	-1	0	1	-1	1	1	1	-1	-1	0	-1
4	Test4	Gimnazija	I/1	53	0	2	15	1	1	1	1	1	-1	-1	1	1	1	-1	1	1	0	0	1	1	1	-1	0
4	Test4	Gimnazija	I/1	54	0	4	16	1	-1	-1	1	1	1	-1	1	1	-1	0	1	1	1	-1	1	-1	1	-1	0
4	Test4	Gimnazija	I/1	55	1	5	19	1	-1	1	1	1	1	1	1	1	-1	-1	1	1	1	1	1	-1	-1	0	0
1	Test1	Gimnazija	I/2	56	0	4	17	1	1	1	-1	-1	1	1	-1	0	1	1	1	1	1	1	0	0	0	0	1
1	Test1	Gimnazija	I/2	57	0	5	18	1	1	0	0	0	-1	-1	0	0	1	-1	-1	-1	-1	0	1	1	1	-1	0
1	Test1	Gimnazija	I/2	58	1	5	20	1	1	1	1	1	1	1	1	1	-1	-1	1	1	1	1	0	0	1	-1	-1
1	Test1	Gimnazija	I/2	59	0	3	13	-1	-1	1	-1	1	0	0	0	0	1	1	1	-1	1	-1	-1	0	1	0	-1
1	Test1	Gimnazija	I/2	60	1	4	16.5	1	1	1	1	1	-1	-1	1	1	1	-1	0	0	1	-1	1	-1	1	1	1
1	Test1	Gimnazija	I/2	61	1	4	15	0	0	1	0	1	1	1	1	1	0	0	1	1	0	0	0	0	1	-1	1
1	Test1	Gimnazija	I/2	62	1	5	19	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	1	-1	1	-1	1
1	Test1	Gimnazija	I/2	63	0	2	12	1	1	1	-1	-1	-1	0	0	0	1	0	1	-1	0	0	1	0	-1	-1	-1
2	Test2	Gimnazija	I/2	64	1	4	17	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	-1	1	-1	0
2	Test2	Gimnazija	I/2	65	1	5	19.5	1	1	1	0	1	0	0	1	1	1	-1	1	1	1	-1	0	0	1	1	1
2	Test2	Gimnazija	I/2	66	0	3	14	0	0	0	1	1	1	-1	1	1	1	-1	1	1	-1	-1	1	0	0	0	-1
2	Test2	Gimnazija	I/2	67	0	5	19	1	1	1	1	1	-1	-1	1	1	1	1	1	0	0	0	1	1	1	1	-1
2	Test2	Gimnazija	I/2	68	0	5	18	-1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	-1	1	1	1	1
2	Test2	Gimnazija	I/2	69	1	4	16	1	1	1	0	1	1	1	0	0	1	-1	1	1	1	-1	1	-1	1	-1	1
2	Test2	Gimnazija	I/2	70	0	2	12	1	-1	1	-1	1	0	0	1	-1	-1	-1	1	-1	0	0	1	-1	-1	-1	-1
3	Test3	Gimnazija	I/2	71	1	5	19	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	-1	-1	1
3	Test3	Gimnazija	I/2	72	1	3	14.5	-1	-1	1	1	1	0	0	1	1	1	-1	1	1	1	1	-1	-1	0	0	0
3	Test3	Gimnazija	I/2	73	1	4	17	1	1	1	-1	1	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1
3	Test3	Gimnazija	I/2	74	1	4	18	1	1	1	1	1	0	1	1	1	1	-1	1	1	-1	-1	1	-1	-1	-1	-1
3	Test3	Gimnazija	I/2	75	1	4	18	0	0	1	-1	-1	1	1	1	1	1	1	-1	-1	0	0	0	0	1	1	1
3	Test3	Gimnazija	I/2	76	1	3	15	1	1	1	1	1	1	-1	-1	-1	1	-1	1	1	0	0	1	0	1	0	0
3	Test3	Gimnazija	I/2	77	0	3	14	1	1	1	0	0	0	0	1	1	1	-1	1	1	1	-1	-1	-1	0	0	0
4	Test4	Gimnazija	I/2	78	0	4	15.5	-1	1	1	1	1	-1	1	1	1	0	0	1	1	0	0	1	-1	1	-1	-1
4	Test4	Gimnazija	I/2	79	1	5	20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

redni broj ucenika	pol	ocena_iz_8_razreda	broj_poena	trougao	maketa	kockice	izbor	stepenice	test	police	razmena		stabla_jabuke		izvoz		farma		najbolji_auto		internet_dopisivanje		stolar
											a	b	a	b	a	b	a	b	a	b	a	b	
1	0	5	18	1	1	-1	1	1	1	1	1	1	0.5	0.5	1	1	1	1	1	0	1	1	0.75
2	0	5	20	1	1	0.75	1	1	-1	-1	1	1	1	0.5	1	1	1	1	1	1	-1	-1	0.5
3	1	5	20	1	-1	0.75	1	1	1	1	1	1	1	0.5	1	1	1	1	1	1	1	1	0.5
4	1	5	19	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0.25	-1	0	0.75
5	1	4	13.5	1	1	1	1	1	1	1	1	1	1	0.5	1	-1	1	1	1	0	1	0	0.75
6	1	5	19.5	1	1	1	1	0	1	0	1	-1	0.75	0	1	-1	0	0	1	0.5	-1	0	-1
7	1	5	15	1	1	1	1	1	1	1	1	1	0.5	-1	1	0	-1	1	1	0.25	1	-1	0.25
8	0	4	16	-1	1	1	1	1	-1	-1	-1	-1	0.25	0.5	1	1	1	1	1	0.5	-1	-1	0.5
9	0	5	17	1	1	1	-1	1	1	1	1	1	1	0.5	1	1	1	0	1	0	1	0	0.75
10	1	5	18	1	1	-1	1	1	-1	-1	1	1	1	0.5	1	1	1	1	1	-1	1	0.5	0.5
11	0	4	16	1	1	0.5	-1	1	1	1	1	1	0.75	0.5	1	1	1	1	1	0	1	1	0.75
12	1	5	18	1	1	0.25	1	1	1	1	1	1	0.5	0.5	1	1	0	0	1	1	-1	0	0.75
13	1	5	19	1	-1	1	1	1	0	-1	1	1	0.5	0.75	1	-1	1	0	1	1	1	0	0.25
14	1	4	19	1	1	0.5	1	1	1	0	1	1	1	1	1	1	-1	0	1	0	-1	0	0.5
15	1	5	19.5	-1	1	1	-1	1	0.5	1	1	1	1	0.75	1	1	1	1	1	1	0.75	0	0
16	0	5	18.5	1	1	0.5	1	1	1	1	1	1	1	0.75	1	1	1	1	0	1	1	1	0.25
17	0	5	16	1	-1	1	1	1	0.5	1	1	1	1	0.25	1	1	1	1	1	1	1	1	1
18	1	5	20	1	1	1	1	1	1	-1	1	1	1	0.5	1	-1	1	1	1	0	1	1	0.25
19	0	5	20	1	1	0.5	1	1	1	-1	1	1	1	0.75	1	1	1	1	1	0.25	1	0.5	0.25
20	0	5	20	-1	1	1	1	1	-1	-1	1	-1	1	0.5	1	0	-1		1	0	1	0	0.75
21	0	5	17	1	1	1	1	1	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	0.5
22	1	5	18	1	1	0	1	-1	1	0	1	1	1	0.75	1	1	0	0	1	0	1	1	1
23	1	5	18	1	1	1	1	1	1	1	1	-1	0.5	-1	1	0	1	1	1	0	1	1	0.5
24	1	5	17.5	1	-1	-1	1	1	-1	-1	1	1	1	0.25	1	1	1	1	1	1	1	-1	1
25	0	4	17	1	1	1	1	1	1	0	0	0	1	0.25	1	1	1	0	-1	1	0	0	0
26	0	5	20	1	1	1	1	1	0	1	1	0	1	0.75	1	1	1	1	1	0	1	1	-1
27	1	5	18.5	1	-1	0.75	1	1	1	-1	1	1	1	0.5	1	1	1	1	1	0.25	1	-1	0.5
28	0	5	20	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
29	1	4	15	-1	-1	1	1	0	1	1	1	0	0.75	-1	1	1	-1	0	1	0	0.25	1	0
30	1	5	19	1	1	1	1	1	1	-1	1	1	-1	0.75	1	1	-1	-1	-1	-1	-1	-1	0.75
31	1	5	20	1	-1	0.75	1	1	1	-1	1	1	1	0.25	1	1	1	1	1	1	1	-1	1
32	1	3	11	1	-1	1	1	1	1	1	1	1	0.75	0.5	1	1	1	0	1	0	1	0.5	0.25
33	0	3	16	-1	1	1	1	1	-1	0	0	0	0.5	0	0	0	0	0	0	0	-1	0	0.5
34	0	4	18	1	-1	1	0	1	-1	1	0	0	1	0	1	1	0	0	0	0	1	-1	1
35	1	3	15	1	1	1	1	1	1	1	1	1	1	0.25	1	1	0	0	1	1	1	0	0
36	1	5	19	-1	1	0.75	1	1	1	-1	1	1	0.75	0	1	1	1	0	1	1	0.25	0	0.25
37	1	5	19	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	1	1	0
38	1	5	20	1	-1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	-1	-1	0.5	0
39	1	4	18	-1	1	0.25	1	1	1	-1	1	1	-1	0	1	1	0	0	1	-1	0	-1	0.25
40	1	2	12	-1	1	1	1	1	0	0	0	0	0.25	0	0	-1	0	0	0	0	-1	-1	0.5
41	1	5	18	-1	-1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0.75	1	-1	1
42	0	3	16	1	1	0.25	0	0	0	0	1	1	0	0	1	1	1	0	1	0.25	-1	0	0.25
43	1	4	16.5	1	1	1	0	1	0	1	1	1	1	0	1	1	1	0	1	0	-1	-1	0
44	1	5	20	1	-1	1	1	1	0.5	1	0	0	0	0	0	0	0	0	0	0	1	-1	0.25
45	1	5	17	1	-1	-1	1	0	0.5	1	1	1	0.5	0	0	0	1	1	0	0	1	0	1
46	0	5	17.5	1	1	1	1	1	0	1	1	0	0	0	1	0	0	0	1	-1	1	0	-1
47	0	3	13.5	1	1	-1	1	1	1	1	1	1	1	0.75	1	1	0	0	1	-1	-1	-1	1
48	1	5	18.5	1	0	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	-1	0
49	1	4	18	1	1	0	1	1	1	-1	0	0	-1	0.75	0	0	1	1	1	-1	1	0	0
50	0	5	17.5	1	-1	0.5	1	-1	0	-1	-1	1	1	0	1	0	-1	0	1	0.25	1	-1	1
51	1	4	17	-1	1	1	1	1	-1	1	1	1	0.75	0	1	-1	1	1	1	1	-1	-1	1
52	1	5	15	1	1	1	1	0	1	1	1	1	-1	0	1	-1	1	1	1	0.25	-1	0	0.5
53	0	2	15	1	1	1	1	1	-1	-1	1	1	1	0.25	1	1	0	0	1	1	1	-1	0
54	0	4	16	1	-1	0.5	1	1	1	-1	1	1	1	0	1	1	1	-1	1	0.75	1	0.5	0
55	1	5	19	1	-1	1	1	1	1	1	1	1	0.75	0.75	1	1	1	1	1	0.75	-1	0	0
56	0	4	17	1	1	1	-1	-1	1	1	-1	0	1	1	1	1	1	1	0	0	0	0	1
57	0	5	18	1	1	0	0	0	-1	-1	0	0	1	0.75	-1	-1	-1	0	1	1	1	-1	0
58	1	5	20	1	1	1	1	1	1	1	1	1	0.75	0.25	1	1	1	1	0	0	1	-1	-1
59	0	3	13	-1	-1	1	-1	1	0	0	0	0	1	1	1	-1	1	-1	-1	0	1	0	0.25
60	1	4	16.5	1	1	1	1	1	-1	-1	1	1	1	-1	0	0	1	-1	1	0.25	1	1	1
61	1	4	15	0	0	1	0	1	1	1	1	1	0	0	1	1	0	0	0	0	1	-1	1
62	1	5	19	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	1	0.75	1	0.5	1
63	0	2	12	1	1	1	-1	-1	0.5	0	0	0	1	0	1	-1	0	0	1	0	-1	-1	0.75
64	1	4	17	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	-1	1	-1	0
65	1	5	19.5	1	1	1	0	1	0	0	1	1	1	-1	1	1	1	-1	0	0	1	1	1

Прилог 3

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Део кода програма

```
options linesize=256 pagesize=3276 nodate nocenter nolabel ;
%global  CC  excel;
%let CC=c;
libname a "&cc:\lazar\master\dat\mysaslib";
%let excel=&CC.:\lazar\master\dat\mysaslib\;
options mstored sasmstore=clan;
libname clan "&cc:\asasapp\clan97_339";

proc import out= work.baza
            datafile= "C:\Lazar\master\dat\Baza1.xls"
            DBMS=EXCEL2000 REPLACE;
     *SHEET="Sheet1$";
     GETNAMES=YES;
run;

data a.baza1;
set baza;
run;

data baza;
set a.baza1;
run;

data baza1;
set baza;
array p(20) p1-p20;
do i=1 to 20;
      if p(i) = -1 then
          p(i)=0;
end;
drop i;
run;

data suma;
set baza1;
suma=p1+p2+p3+p4+p5+p6+p7+p8+p9+p10+p11+p12+p13+p14+p15+p16+p17+p18+p19+p20;
run;

proc summary data=suma;* missing nway;
var suma;
class brt;
output out=brt_suma sum (suma)= ;
run;

data sposobnost;
set suma;
sposobnost=sum/20;
run;

data suma_netacnih;
set suma;
array p (20) p1-p20;
do i=1 to 20;
      if p(i) = 1 then
          p(i)=0;
      else if p(i) = 0 then
              p(i)=1;
end;
drop i;
run;

proc summary data=suma_netacnih;* missing nway;
var p1--p20;
class od;
*id rbr;
*class rbr;
output out=wtab_suma_netacnih sum (p1--p20)= snp1 snp2 snp3 snp4 snp5 snp6 snp7 snp8 snp9 snp10
snp11 snp12 snp13 snp14 snp15 snp16 snp17 snp18 snp19 snp20;
run;


data tezina_pitanja;
set wtab suma netacnih ;
array snp (20) snp1-snp20;
array t (20) t1-t20;
do i=1 to 20;
if snp(i) ne 0 then
t(i)= snp(i)/ freq ;
t(i)=round(t(i),0.01);
end;
drop i;
run;
```

```
proc sort data=sposobnost;
by od;
run;

proc sort data=tezina_pitanja1;
by od;
run;

data baza2;
merge sposobnost(in=a) tezina_pitanja1(in=b);
by od;
if a and b;
run;

data probna_verovatnoca;
set baza2;
array ver(20) ver1-ver20;
array t(20) t1-t20;
do i=1 to 20;
      if t(i) ne 0 then
ver(i)=1/(1+exp(-(sposobnost-t(i))));
ver(i)=round(ver(i),0.01);
       end;
drop i;
run;

data baza3;
set tezina pitanja1;
q1=-3; q2=-2; q3=-1; q4=0; q5=1; q6=2; q7=3;
run;

data diskriminacija;
set baza3;

v1=(exp(2*(q1-0))/(1+exp(2*(q1-0))));
v2=(exp(2*(q2-0))/(1+exp(2*(q2-0))));
v3=(exp(2*(q3-0))/(1+exp(2*(q3-0))));
v4=(exp(2*(q4-0))/(1+exp(2*(q4-0))));
v5=(exp(2*(q5-0))/(1+exp(2*(q5-0))));
v6=(exp(2*(q6-0))/(1+exp(2*(q6-0))));
v7=(exp(2*(q7-0))/(1+exp(2*(q7-0))));

v8=(exp(1*(q1-0))/(1+exp(1*(q1-0))));
v9=(exp(1*(q2-0))/(1+exp(1*(q2-0))));
v10=(exp(1*(q3-0))/(1+exp(1*(q3-0))));
v11=(exp(1*(q4-0))/(1+exp(1*(q4-0))));
v12=(exp(1*(q5-0))/(1+exp(1*(q5-0))));
v13=(exp(1*(q6-0))/(1+exp(1*(q6-0))));
v14=(exp(1*(q7-0))/(1+exp(1*(q7-0))));

v15=(exp(0.5*(q1-0))/(1+exp(0.5*(q1-0))));
v16=(exp(0.5*(q2-0))/(1+exp(0.5*(q2-0))));
v17=(exp(0.5*(q3-0))/(1+exp(0.5*(q3-0))));
v18=(exp(0.5*(q4-0))/(1+exp(0.5*(q4-0))));
v19=(exp(0.5*(q5-0))/(1+exp(0.5*(q5-0))));
v20=(exp(0.5*(q6-0))/(1+exp(0.5*(q6-0))));
v21=(exp(0.5*(q7-0))/(1+exp(0.5*(q7-0))));


v1=round(v1,0.0001);v2=round(v2,0.0001);v3=round(v3,0.0001);v4=round(v4,0.0001);v=
round(v5,0.0001);v6=round(v6,0.0001);v7=round(v7,0.0001);
v8=round(v8,0.0001);v9=round(v9,0.0001);v10=round(v10,0.0001);v11=round(v11,0.0001);v12=round(v12,0.
0001);v13=round(v13,0.0001);v14=round(v14,0.0001);
v15=round(v15,0.0001);v16=round(v16,0.0001);v17=round(v17,0.0001);v18=round(v18,0.0001);v19=round(v1
9,0.0001);v20=round(v20,0.0001);v21=round(v21,0.0001);
run;
```

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 4

Nitrogen Content of Red Clover Plants
The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
imet	5	Test1 Test2 Test3 Test4 ime_testa

Number of Observations Read	86
Number of Observations Used	84

Dependent Variable: ukupno

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	9.584161	3.194720	0.13	0.9442
Error	80	2022.558696	25.281984		
Corrected Total	83	2032.142857			

R-Square	Coeff Var	Root MSE	ukupno Mean
0.004716	69.69670	5.028119	7.214286

Source	DF	Anova SS	Mean Square	F Value	Pr > F
imet	3	9.58416149	3.19472050	0.13	0.9442

Alpha	0.05
Error Degrees of Freedom	80
Error Mean Square	25.28198
Critical Value of Studentized Range	3.71070

Tukey's Studentized Range (HSD) Test for ukupno

Note: *This test controls the Type I experimentwise error rate.*

Comparisons significant at the 0.05 level are indicated by ***.				
Imet Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
Test3 - Test4	0.4500	-3.7220	4.6220	
Test3 - Test2	0.7500	-3.3721	4.8721	
Test3 - Test1	0.8804	-3.1533	4.9141	
Test4 - Test3	-0.4500	-4.6220	3.7220	
Test4 - Test2	0.3000	-3.8221	4.4221	
Test4 - Test1	0.4304	-3.6033	4.4641	
Test2 - Test3	-0.7500	-4.8721	3.3721	
Test2 - Test4	-0.3000	-4.4221	3.8221	
Test2 - Test1	0.1304	-3.8516	4.1124	
Test1 - Test3	-0.8804	-4.9141	3.1533	
Test1 - Test4	-0.4304	-4.4641	3.6033	
Test1 - Test2	-0.1304	-4.1124	3.8516	

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 5

Nitrogen Content of Red Clover Plants
The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
imet	5	Test1 Test2 Test3 Test4 imet

Number of Observations Read	86
Number of Observations Used	84

Dependent Variable: ukupno

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5.139363	1.713121	0.08	0.9692
Error	80	1655.029535	20.687869		
Corrected Total	83	1660.168899			

R-Square	Coeff Var	Root MSE	ukupno Mean
0.003096	45.79741	4.548392	9.931548

Source	DF	Anova SS	Mean Square	F Value	Pr > F
imet	3	5.13936335	1.71312112	0.08	0.9692

Tukey's Studentized Range (HSD) Test for ukupno

Note: *This test controls the Type I experimentwise error rate.*

Alpha	0.05
Error Degrees of Freedom	80
Error Mean Square	20.68787
Critical Value of Studentized Range	3.71070

Comparisons significant at the 0.05 level are indicated by ***.				
Imet Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
Test4 - Test1	0.3918	-3.2570	4.0407	
Test4 - Test2	0.5637	-3.1651	4.2925	
Test4 - Test3	0.6625	-3.1115	4.4365	
Test1 - Test4	-0.3918	-4.0407	3.2570	
Test1 - Test2	0.1718	-3.4302	3.7739	
Test1 - Test3	0.2707	-3.3782	3.9195	
Test2 - Test4	-0.5637	-4.2925	3.1651	
Test2 - Test1	-0.1718	-3.7739	3.4302	
Test2 - Test3	0.0988	-3.6300	3.8276	
Test3 - Test4	-0.6625	-4.4365	3.1115	
Test3 - Test1	-0.2707	-3.9195	3.3782	
Test3 - Test2	-0.0988	-3.8276	3.6300	

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 6

Korelacija između totala using PROC CORR
The CORR Procedure

2 Variables:	t nt

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
t	4	244.25000	13.76893	977.00000	232.00000	264.00000
nt	4	262.00000	22.28041	1048	235.25000	289.75000

Pearson Correlation Coefficients, N = 4 Prob > |r| under H0: Rho=0		
	t	nt
t	1.00000	0.95428 0.0457
nt	0.95428 0.0457	1.00000

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 7

Korelacija ocene i uspeha
The CORR Procedure

3 Variables:	oc8	brpprij	ukupno_poena

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
oc8	84	4.27381	0.90980	359.00000	2.00000	5.00000
brpprij	84	17.19048	2.24132	1444	11.00000	20.00000
ukupno_poena	85	14.25882	65.13341	1212	-5.00000	606.00000

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations			
	oc8	**brpprij**	**ukupno_poena**
oc8	1.00000 84	0.81016 <.0001 84	0.40967 0.0001 84
brpprij	0.81016 <.0001 84	1.00000 84	0.34174 0.0015 84
ukupno_poena	0.40967 0.0001 84	0.34174 0.0015 84	1.00000 85

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 8

Korelacija ocene i uspeha
The CORR Procedure

3 Variables:	oc8 brpprij ukupno

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
oc8	84	4.27381	0.90980	359.00000	2.00000	5.00000
brpprij	84	17.19048	2.24132	1444	11.00000	20.00000
ukupno	84	9.93155	4.47236	834.25000	0.50000	19.00000

Pearson Correlation Coefficients, N = 84 Prob > |r| under H0: Rho=0			
	oc8	**brpprij**	**ukupno**
oc8	1.00000	0.81016 <.0001	0.45547 <.0001
brpprij	0.81016 <.0001	1.00000	0.36310 0.0007
ukupno	0.45547 <.0001	0.36310 0.0007	1.00000

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 9

Značajnost parametra težine
The TTEST Procedure (t - test nezavisnih uzoraka)

Variable: t (težina)

br2	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	20	0.4125	0.2171	0.0485	0.1400	0.8800
2	20	0.2835	0.1674	0.0374	0.0600	0.6200
Diff (1-2)		0.1290	0.1938	0.0613		

br2	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		0.4125	0.3109	0.5141	0.2171	0.1651	0.3171
2		0.2835	0.2052	0.3618	0.1674	0.1273	0.2445
Diff (1-2)	Pooled	0.1290	0.00491	0.2531	0.1938	0.1584	0.2498
Diff (1-2)	Satterthwaite	0.1290	0.00464	0.2534			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	2.10	0.0420
Satterthwaite	Unequal	35.694	2.10	0.0424

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	1.68	0.2662

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 10

Značajnost parametra diskriminacije
The TTEST Procedure (t - test nezavisnih uzoraka)

Variable: d (diskriminativnost)

br2	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	20	0.8990	0.00641	0.00143	0.8900	0.9100
2	20	0.8950	0.00946	0.00212	0.8700	0.9100
Diff (1-2)		0.00400	0.00808	0.00255		

br2	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		0.8990	0.8960	0.9020	0.00641	0.00487	0.00936
2		0.8950	0.8906	0.8994	0.00946	0.00719	0.0138
Diff (1-2)	Pooled	0.00400	-0.00117	0.00917	0.00808	0.00660	0.0104
Diff (1-2)	Satterthwaite	0.00400	-0.00120	0.00920			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	1.57	0.1257
Satterthwaite	Unequal	33.403	1.57	0.1268

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	2.18	0.0978

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Прилог 11

Značajnost parametra pogadjanja
The TTEST Procedure (t - test nezavisnih uzoraka)

Variable: c (pogađanje)

br1	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	7	0.2229	0.0934	0.0353	0.1000	0.3400
2	7	0.4600	0.6862	0.2594	0.0400	2.0000
Diff (1-2)		-0.2371	0.4897	0.2618		

br1	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		0.2229	0.1365	0.3092	0.0934		
2		0.4600	-0.1747	1.0947	0.6862		
Diff (1-2)	Pooled	-0.2371	-0.8075	0.3332	0.4897		
Diff (1-2)	Satterthwaite	-0.2371	-0.8722	0.3979			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	12	-0.91	0.3828
Satterthwaite	Unequal	6.2222	-0.91	0.3987

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6	6	53.98	0.0001

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs