

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 876

**STATISTIČKI POKAZATELJI I VIZUALIZACIJA  
USPJEŠNOSTI ISPITA U EDGARU**

Marija Kompar

Zagreb, lipanj 2023.

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 876

**STATISTIČKI POKAZATELJI I VIZUALIZACIJA  
USPJEŠNOSTI ISPITA U EDGARU**

Marija Kompar

Zagreb, lipanj 2023.

## **ZAVRŠNI ZADATAK br. 876**

Pristupnica: **Marija Kompar (0036533623)**  
Studij: Elektrotehnika i informacijska tehnologija i Računarstvo  
Modul: Računarstvo  
Mentor: prof. dr. sc. Igor Mekterović

Zadatak: **Statistički pokazatelji i vizualizacija uspješnosti ispita u Edgaru**

### Opis zadatka:

Edgar je informacijski sustav koji se koristi za ispitivanje studenata na Fakultetu elektrotehnike i računarstva. U okviru nekog predmeta, tipično se provodi niz većih ili manjih ispita (domaćih zadaća, projekata, itd.) tijekom semestra putem kojih studenti skupljaju bodove koji im ulaze u konačnu ocjenu. Posebno istaknuti su međuispit na polovici semestra i završni ispit na kraju semestra. Trenutno Edgar nudi osnovne statističke pokazatelje (prosječne vrijednost, medijan) te vizualizaciju razdiobe i kutijaste dijagrame (engl. boxplot) - ukupno i po grupama s predavanja. U okviru rada potrebno je proučiti i donijeti sažeti prikaz teorije koja se bavi ocjenom uspješnosti ispita i pitanja (Teorija odgovora na zadatke, engl. Item Response Theory). Isto tako, istražiti i opisati kako se u značajnijim sustavi za provođenje ispita (npr. Moodle) obrađuju rezultati ispita i donosi ocjena uspješnosti ispita, ocjena interne konzistencije ispita, kako se statistički obrađuju pitanja iz ispita itd. U praktičnom dijelu rada potrebno je primijeniti stečena znanja te napraviti sveobuhvatni izvještaj o odabranom ispitu u Edgaru koji će uključivati sve relevantne statistike i vizualizacije, te obraditi ispit u cjelini (prosječne vrijednost, medijan, devijacija, asimetrija, pouzdanost, pogreške, itd.) i na razini pojedinačnih pitanja koja su se pojavila u ispitu (osnovni statistički pokazatelji, indeks lakoće, diskriminacijski indeks, itd.). Donijeti ocjenu ostvarenog pristupa te smjernice za budući razvoj.

Rok za predaju rada: 9. lipnja 2023.



# Sadržaj

<b>Uvod</b>	1
<b>1 Opis problema</b>	3
1.1 Prikaz postojećeg izvještaja za nastavnike na aplikaciji Edgar	3
1.2 Vrste ispita i zadataka	4
1.3 Statistički pokazatelji	5
1.3.1 Cronbach alfa	5
1.3.2 Indeks lakoće	8
1.3.3 Diskriminacijski indeks	10
1.3.4 Diskriminativna učinkovitost	11
1.3.5 Vrijednost slučajnog pogotka	12
1.4 Teorija odgovora na zadatke	14
<b>2 Primjer statistike ispita na platformi Moodle</b>	18
<b>3 Model podataka sustava Edgar</b>	21
3.1 Relacijska shema	21
3.2 Opis korištenih tablica i atributa iz sustava Edgar	22
<b>4 Opis korištenih tehnologija</b>	26
4.1 PgAdmin	26
4.2 RStudio	26
<b>5 Oblikovanje i implementacija rješenja zadatka</b>	28
5.1 Povezivanje baze podataka s radnim okruženjem	28
5.2 Upute za pokretanje izvorne datoteke	28
5.3 Komponente izvještaja o uspjehu rezultata ispita	29
5.3.1 Histogram	29
5.3.2 Kutijasti dijagram (engl. <i>boxplot</i> )	31
5.3.3 Statistički pokazatelji za ispit	32
5.3.4 Grafički prikazi distribucije bodova pojedinih zadataka	35
5.3.5 Statistički pokazatelji za zadatak	36
<b>Zaključak</b>	39
<b>Literatura</b>	40

## Uvod

U današnje vrijeme, mnogi fakulteti u Hrvatskoj prilagodili su se trendovima u obrazovanju i koriste razna tehnološka rješenja za provođenje ispita na daljinu. Jedno od takvih rješenja je sustav Edgar, razvijen na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Pod fleksibilnim i nadziranim uvjetima, Edgar omogućuje ocjenjivanje raznih komponenti kolegija, poput međuispita i završnih ispita, ispitnih rokova, laboratorijskih vježbi, kratkih provjera znanja i sl., pritom podržavajući velike količine ispitnih zadataka s posebnim načinima vrednovanja.

Ispiti su nužan dio akademskog obrazovanja jer daju uvid u očekivanu usvojenost znanja i stečene kompetencije studenata. Definicija kvalitetnog ispita u obrazovnom sustavu može se iskazati kao sposobnost stavki u ispitu da na pošten i ispravan način mjere znanje studenta. Potrebno je uključiti različite vrste stavki koje pokrivaju mnoge aspekte znanja i vještina te ispituju teme relevantne za predmet. Pitanja trebaju biti jasna i precizno definirati ono što se traži od studenta. Bitan faktor je i postavljanje vremenskog okvira pisanja ispita, kao i pravedno ocjenjivanje odgovora studenata. Cilj svakog sastavljača ispita trebao bi biti precizno primjenjivanje navedenih faktora, ali to je često izrazito složen posao koji iziskuje statističku analizu, interpretaciju i formulaciju rezultata.

Iako u sklopu aplikacije Edgar postoji statistički izvještaj za nastavnike, on uključuje samo osnovne statističke pokazatelje i dijagrame. Za razliku od pojedinih platformi za provođenje ispita na daljinu (npr. Moodle), Edgar zasad ne nudi ocjenu uspješnosti ispita na temelju nekih drugih statističkih pokazatelja koji daju detaljniji osvrt na riješenost ispita i pojedinih zadataka na ispitu. Mnoge značajke uspješnosti ispita tumačene su u sklopu Teorije odgovora na zadatke (engl. *Item response theory*), obitelji matematičkih modela iz područja psihometrije, koja se koristi za određivanje kvalitete mjernih instrumenata poput provjera znanja.

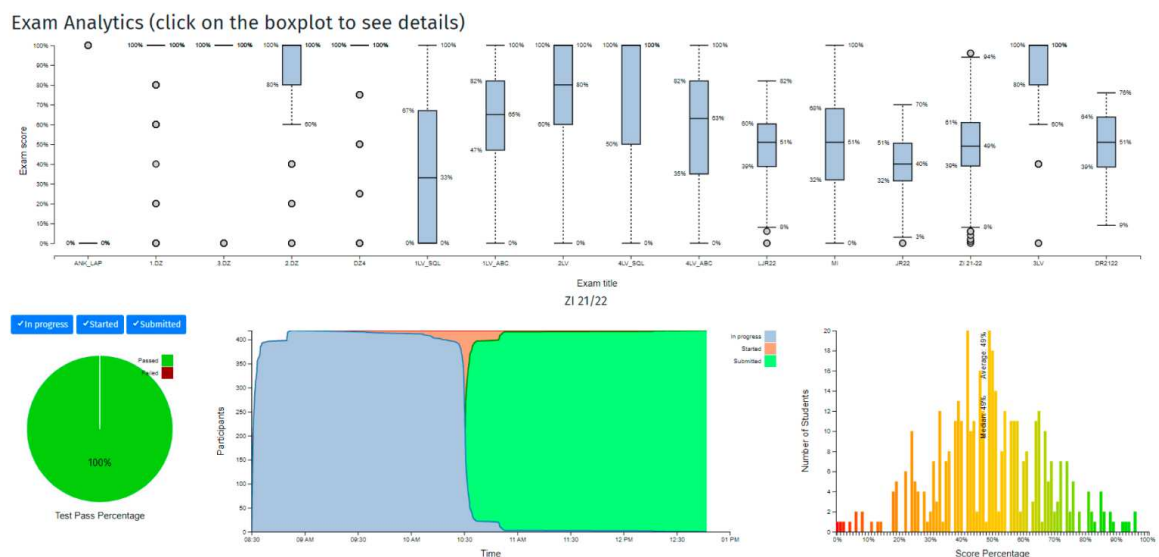
Cilj ovoga rada je definirati potrebne alate za proširivanje postojećeg statističkog izvještaja prikazanom na aplikaciji Edgar, objasniti povezanost Teorije odgovora na zadatke sa završnim radom te obaviti detaljan statistički izvještaj rezultata

ispita u Edgaru. Uz osnovne statističke pokazatelje, novi će izvještaj pobliže opisati rezultate pojedinih zadataka na ispitu koristeći pokazatelje poput indeksa lakoće, diskriminacijskog indeksa, diskriminativne učinkovitosti, vrijednosti slučajnog pogotka i koeficijenta unutarnje konzistentnosti među zadacima.

# 1 Opis problema

## 1.1 Prikaz postojećeg izvještaja za nastavnike na aplikaciji Edgar

Na aplikaciji Edgar postoji prikaz statistike za sve ispite koji su se održali kroz sustav tijekom akademske godine. Izvještaj uključuje tzv. kutijaste dijagrame (engl. *boxplot*) koji prikazuju distribuciju bodova svakoga ispita i smješteni su na istom koordinatnom sustavu. Klikom na bilo koji dijagram prikazuju se detalji o odabranom ispitu koji uključuju postotak prolaznosti ispita, vrijeme pisanja i predavanja ispita te obojeni histogram s razdiobom bodova koji uključuje i oznake za srednju vrijednost (aritmetičku sredinu) te medijan (središnji podatak u uzorku), kao što je prikazano na slici ispod.



Sl. 1.1.1 Prikaz statističkog izvještaja za nastavnike koji prikazuje sve ispite u akademskoj godini na određenom predmetu



## 1.2 Vrste ispita i zadataka

Baza ispita provedenih kroz sustav Edgar sadrži iznimno velik broj podataka iz kojih je moguće odrediti ocjenu djelotvornosti tih ispita kombiniranjem generiranih statističkih pokazatelja te interpretacijom dobivenih vrijednosti. Primjeri dosadašnjih već provedenih ispita vrlo su korisni u razumijevanju pravilnog sastavljanja novih, ali pomažu razumjeti i neke druge faktore poput kvalitete nastave, kvalitete profesora i prilagođenosti nastavnoga gradiva mogućnostima studenata.

Različiti ispiti sastoje se od mnogih vrsta pitanja, od kojih se svaka vrsta treba različito tretirati jer se neki statistički pokazatelji ne mogu računati za sve koje postoje i sve modele ocjenjivanja istih. Edgar podržava sljedeće vrste pitanja:

- „Classic (ABC) questions“ – pitanja višestrukog izbora. Sastoje se od pitanja i ponuđenih odgovora od kojih je jedan ili više njih točno, format pitanja pogodan je za izračunavanje svih statističkih pokazatelja.
- Pitanja s unosom koda – mogu biti pitanja za unos koda iz programskih jezika SQL (engl. *Structured Query Language*), Java, C, itd., najčešće donose više od jednog boda i nije moguće pogađanje i odabir točnog odgovora.
- „Free text“ – pitanja sa slobodnim unosom teksta, uglavnom podrazumijevaju tekstualne esejske odgovore ili vrlo velike unose programskog koda.
- „Homework question“ – pitanja iz domaćih zadaća višestrukog izbora.

Edgar je platforma za mnoge predmete koji iziskuju sposobnosti programiranja, zbog čega su mnoga pitanja unosa programskog koda („Evaluated SQL question“, „Evaluated C-Lang question“, itd.) ili tekstualnog opisa (engl. *free text questions*). Takva pitanja provjeravaju primjenu stečenog znanja na nove probleme, mjere sposobnost analitičkog razmišljanja i artikuliranja ideja. Budući da je područje primjene znanja i kreativnosti vrlo široko, poželjno je pitanja vrlo detaljno definirati. Popularno ih je razložiti u više dijelova za koje se definira

određeni udio bodova radi jednostavnijeg ocjenjivanja različitih aspekata ispitivanog područja. Ti zadaci uobičajeno vrijede više bodova od pitanja jednostavnijeg formata, kao npr. pitanja višestrukog izbora (engl. *multiple choice questions*).

Mnoge statističke mjere korištene za robusniji izvještaj rezultata podrazumijevaju pitanja višestrukog izbora. Ona se sastoje od tekstualnog pitanja i minimalno dva ponuđena odgovora, od kojih su jedan ili više ispravni, a ostali se nazivaju tzv. „ometačima“ (engl. *Distractor*) [1]. Ovakav format ispitnih pitanja pogodan je za brzo i jednostavno ocjenjivanje, pogotovo za predmete s velikim brojem upisanih studenata [1]. Ipak, nedostatak ovakvih zadataka je da ispitanici mogu točno odgovoriti na temelju površnog pamćenja, bez traženog razumijevanja i sposobnosti za primjenu znanja, a mogu i pogoditi točan odgovor bez ikakvog znanja o ispitnom gradivu. Pri formuliranju adekvatnih pitanja, poželjno je korištenje značajnih i relevantnih „ometača“ koji otežavaju očiti izbor pitanja i potiču na razmišljanje i primjenu naučenih tema. Također, dobra praksa je izbjegavanje indicija na točan odgovor i obeshrabrivanje pogađanja kaznenim bodovima za odabir pogrešnog odgovora. Mnogi ispiti u sustavu Edgar sastoje se od pitanja višestrukog izbora. Prikladnost ispitnih pitanja vrlo je bitno svojstvo i može se iskazati određenim statističkim mjerama.

## **1.3 Statistički pokazatelji**

### **1.3.1 Cronbach alfa**

U kontekstu samog ispita, nova statistička mjera koja čini nadopunu postojećeg izvještaja je Cronbach alfa, koju je 1951. godine razvio Lee Cronbach kao mjeru pouzdanosti unutarnje konzistentnosti testa ili upitnika [3]. Predstavlja opseg u kojem sve stavke/pitanja unutar testa mjere isti koncept ili psihološki konstrukt, što u ovom slučaju podrazumijeva znanje studenta. Izračun se obavlja prema formuli , gdje „ $N$ “ predstavlja broj zadataka, „ $\bar{c}$ “ predstavlja srednju vrijednost kovarijance između zadataka, a „ $\bar{v}$ “ srednju vrijednost varijanci rezultata.

$$Cronbach\ alfa = \frac{N * \bar{c}}{\bar{v} + (N-1) * \bar{c}} \quad (1)$$

Za objašnjenje izračuna, uzmimo u obzir sljedeći primjer: upitnik s 4 pitanja na koja se odgovara Likert skalom s 5 ponuđenih odgovora, od kojih broj 1 označava „Vrlo nezadovoljno“ i 5 „Vrlo zadovoljno“ [7]. Upitnik sadrži sljedeća pitanja:

1. Na moj telefonski, e-mail ili pismeni upit odgovoreno je u razumnom vremenu.
2. Zadovoljan/na sam s vremenom pristizanja usluge.
3. Vrijeme u kojem sam čekao/la na uslugu bilo je razumno.
4. Zadovoljan/na sam s uslugom koju sam dobio/la.

Upitnik je proveden na 60 ispitanika i u tablici ispod prikazan je isječak rezultata iz uzorka.

Tablica 1.3.1.1 Isječak rezultata iz uzorka koji služi kao primjer za računanje Cronbach alfe

	Prvo pitanje	Drugo pitanje	Treće pitanje	Četvrto pitanje
Ispitanik 1	1	1	2	5
Ispitanik 2	4	3	4	2
Ispitanik 3	4	2	4	2
Ispitanik 4	4	5	5	2
...				
Ispitanik 57	5	5	4	2
Ispitanik 58	5	4	4	1
Ispitanik 59	4	5	3	1
Ispitanik 60	2	3	4	2

Za izračun srednje kovarijance potrebno je izračunati kovarijancu između pojedinih pitanja. Na primjer, formula za kovarijancu između pitanja 1 (označeno „a“) i pitanja 2 (označeno „b“) prikazana je ispod. „ $a_i$ “ odgovara rezultatu i-tog

ispitanika na zadatku a, dok „ $\mu_a$ “ odgovara srednjoj vrijednosti rezultata na zadatku a.

$$Cov(a, b) = \sum ((a_i - \mu_a) * (b_i - \mu_b)) / (N - 1) \quad (2)$$

Postoji 6 različitih parova pitanja između kojih se računa srednja vrijednost kovarijance, prikazana na formuli ispod za ovaj konkretan primjer.

$$\bar{c} = \frac{(Cov(1,2)+Cov(1,3)+ Cov(1,4)+Cov(2,3)+Cov(2,4)+Cov(3,4))}{6} \quad (3)$$

Srednja kovarijanca iznosi otprilike 0.2123, a srednja varijanca rezultata svih pitanja iznosi 1.4948. Dakle, za dani primjer Cronbach alfa je  $(4 * 0.2123) / (1.4948 + 3 * 0.2123)$ , što iznosi 0.7853, po mnogim standardima zadovoljavajući rezultat [7]. Ipak, taj se rezultat može poboljšati izbacivanjem pitanja čiji rezultati znatno odstupaju od ostalih iz uzorka. U ovom slučaju, kao što je vidljivo na isječku u tablici, može se primijetiti da kod mnogih ispitanika odgovor na četvrto pitanje nema baš smisla u kombiniranim četvorkama poput (4,5,3,1) kod ispitanika 59 i (1,1,2,5) kod ispitanika 1. Izbacivanjem četvrtoga pitanja dobiva se veća vrijednost alfe koja sada iznosi 0.92167 [7].

Ova se mjera često koristi u istraživanju i psihometriji za usporedbu pouzdanosti različitih verzija testa, kao i za identifikaciju slabih stavki ili stavki koje nisu dobro povezane s ostalima u testu. Poprima vrijednosti između 0 i 1, što je uglavnom prikazano u obliku postotka, a na njezinu vrijednost može utjecati broj pitanja, njihova međusobna povezanost i varijacije unutar skupova podataka. Na primjer, mala vrijednost alfe može biti rezultat nedovoljnog broja pitanja i velike heterogenosti stavki. Postoje različite preporuke o zadovoljavajućim iznosima, ali taj broj općenito spada u interval između 0.75 i 0.95 [3]. Postizanje željenih rezultata može biti izvedeno identificirajući pitanja koja pružaju male vrijednosti korelacije između rezultata na samom pitanju i ukupnog rezultata na testu, dakle pronalaženje nezadovoljavajućih koeficijenata diskriminacije koji se u ovom izvještaju računaju za svaki zadatak. Visoka vrijednost alfe ( $> 0.90$ ) može ukazati na redundancije u podacima i zahtijevati da se broj pitanja smanji. Također, valja napomenuti da se Cronbach alfa računa samo za pitanja višestrukog izbora s

uvjetom dovoljne varijabilnosti u rezultatima, što podrazumijeva da varijance rezultata zadataka poprimaju vrijednosti različite od nule.

### 1.3.2 Indeks lakoće

Indeks lakoće (engl. *facility index*) predstavlja udio ispitanika koji su točno odgovorili na pitanje. Može poprimiti vrijednosti između 0 i 1, a najčešće se prikazuje u postotku. Računa se formulom (4), gdje  $N_{correct}$  označava broj studenata koji su točno odgovorili, a  $N_{total}$  broj svih studenata koji su odgovarali na pitanje.

$$indeks\ lakoće = N_{correct}/N_{total} \quad (4)$$

Vrlo nizak rezultat ukazuje na teška i iznimno teška i apstraktna pitanja ili pitanja koja imaju grešku, zbunjujuća su ili zavaravajuća. Za prosječnog učenika odgovarajući indeks lakoće je 35%-65%, a vrlo visoki indeksi indiciraju da je pitanje temeljan koncept koji treba biti opće poznat, ili je ono loše formulirano tako da je točan odgovor očit [1]. Vrijedi naglasiti da, budući da mnogi ispiti imaju bazu s velikim brojem pitanja koja su nasumično dodijeljena pojedinim studentima, neki podaci s rezultatima nemaju dovoljno varijabilnosti da bi izmjereni indeks lakoće bio vjerodostojan i ukazao na prirodu pitanja. Primjer takve situacije je zadatak s identifikacijskim brojem 45788 iz Međuispita iz Naprednog razvoja programske potpore za web 2022/2023. Baza podataka na tom ispitu sadrži čak 123 zadatka, od kojih se samo 25 nasumično pojavilo na pojedinoj instanci testa, a navedeni zadatak pojavio se samo na 3 instance. Budući da je na svakoj od njih ostvaren točan odgovor, indeks lakoće za to pitanje iznosi 100% i upitno je koliko je taj podatak mjerodavan jer je dobiven na iznimno malom uzorku. Također, nizak indeks lakoće ne mora odbaciti uključivanje odgovarajućeg pitanja u ispit ako je ono dio podučavanoga gradiva koje se smatra neophodno za prolazak predmeta [1]. Indeks lakoće također se može iskazati za zadatke koji nemaju višestruki izbor. Većina je takvih zadataka u Edgaru otvorenog tipa, podrazumijevaju upis programskog koda ili tekstualni unos rješenja i boduju se parcijalni, zbog čega mogu poprimiti bilo koju ocjenu na skali od 0% i 100%. Za

takve zadatke moguće je intuitivno odrediti prag koji čini razliku između „točnih“ i „netočnih“ odgovora. Budući da u obrazovnom sustavu najčešći prag koji definira pad ili prolaz predmeta ili ispita 50%, takav se koristi u programskom dijelu završnoga rada. U bazi podataka postoje i pitanja višestrukog izbora koja imaju više točnih odgovora i parcijalno se vrednuju, ovisno o tome jesu li svi točni odgovori odabrani. Za takvu vrstu zadataka, prag kojim se definira prolaz i pad također je 50%. Izračun indeksa lakoće za različite vrste pitanja prikazan je na tri izmišljena uzorka s 18 ispitanika koji predstavljaju postotak bodova za pitanja višestrukog izbora s jednim točnim odgovorom, ista pitanja s više točnih odgovora i pitanja unosa teksta. Pretpostavimo da drugo pitanje ima dva točna odgovora, od kojih svaki donosi 0.5 bodova, a netočan odgovor se kažnjava negativnim bodovima od -0.25.

Tablica 1.3.2.1 Primjer izračuna indeksa lakoće za različite vrste pitanja

	Bodovi ostvareni na zadatku u obliku postotka	Indeks lakoće
Pitanje višestrukog izbora s jednim točnim odgovorom	1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0	61.11%
Pitanje višestrukog izbora s više točnih odgovora	-0.5, 1, 0.5, 0.5, 1, 1, 0.5, -0.5, 1, 1, 0.25, 0, -0.5, -0.5, 1, -0.5, 0, 0	44.44%
Pitanje slobodnog unosa teksta	0, 0.4, 1, 0.7, 0.3, 1, 0.9, 0.6, 0.2, 0.3, 0.5, 0.5, 1, 0.3, 0.8, 1, 0, 0	55.55%

Budući da sva tri primjera podrazumijevaju 18 ispitanika,  $N_{total} = 18$ . Za prvo pitanje  $N_{correct}$  predstavlja broj rezultata od 100% i ima vrijednost 11, dakle indeks lakoće za prvo pitanje je  $11/18 * 100\%$ , što iznosi 61.11%. Za drugo pitanje,  $N_{correct}$  iznosi 8, a indeks lakoće je  $8/18 * 100\%$ , što iznosi 44.44%. Treće pitanje ima 10 primjeraka iz uzorka s postotkom jednakim ili većim od 0.5, dakle njegov indeks lakoće iznosi 55.55%.

### 1.3.3 Diskriminacijski indeks

Veliki pokazatelj kvalitete pitanja s višestrukim odgovorima njegova je snaga diskriminiranja između studenata s više i onih s manje znanja. Diskriminacijski koeficijent ili indeks (engl. *discrimination index*) je mjera razlike vjerojatnosti da općenito uspješnija skupina studenata točno odgovori na pitanje naspram druge skupine. Uspješnija skupina podrazumijeva ispitanike s ukupnim rezultatom većim od 50%, a neuspješna one s manjim. Diskriminacijski indeks se računa prema formuli (5), gdje „ $high_{correct}$ “ označava broj ispitanika iz grupe s visokim ukupnim rezultatom koji su točno odgovorili na pitanje, „ $low_{correct}$ “ označava broj ispitanika iz grupe s niskim ukupnim rezultatom, a „ $N$ “ ukupni broj ispitanika iz obje grupe.

$$diskriminacijski\ indeks = \frac{high_{correct} - low_{correct}}{N} \quad (5)$$

Dakle, za računanje se prikupljaju podaci rezultata svih studenata na jednom ispitu i rezultati istih na pojedinom pitanju iz ispita, koji se označavaju binarno s 0 za netočan odgovor ili 1 za točan. U ovom je dijelu također potrebno definirati prag prolaza za „točan“ ili „netočan“ odgovor kod pitanja slobodnog unosa koji vrijede više od jednog boda, tako da je skup instanci podijeljen na one čiji je rezultat ispod i iznad 50%. Poželjno je da indeks poprimi što veću vrijednost, jer to ukazuje na dobro podudaranje između zadatka i cijeloga ispita te podupire pretpostavku da zadatak dobro diskriminira između boljih i lošijih studenata. Male, ali pozitivne vrijednosti indiciraju da zadatak ne pruža dobru diskriminaciju, a negativne vrijednosti su problematične i često nastaju zbog pogrešaka u samom pitanju. Zadaci višestrukog izbora uglavnom imaju manji diskriminacijski indeks jer je u njihovom slučaju veća vjerojatnost pogotka točnog pitanja bez znanja potrebnog za njegovo razumijevanje. Pretpostavimo izmišljeni ispit s 10 ispitanika i dvije vrste pitanja, višestruki izbor i slobodni unose teksta. Podaci su rezultata prikazani su u tablici.

Tablica 1.3.3.1 Primjer izračuna diskriminacijskog indeksa za dvije vrste pitanja

	Postotak bodova na pitanju višestrukog izbora	Postotak bodova na pitanju slobodnog unosa teksta	Ukupni postotak ostvaren na ispitu
Ispitanik 1	1	0.2	0.25
Ispitanik 2	1	0.8	0.7
Ispitanik 3	0	0.75	0.87
Ispitanik 4	1	0.5	0.34
Ispitanik 5	1	1	1
Ispitanik 6	1	0.9	0.92
Ispitanik 7	1	1	0.50
Ispitanik 8	1	0.35	0.44
Ispitanik 9	1	0.15	0.25
Ispitanik 10	1	1	0.66
Diskriminacijski indeks	10%	50%	

Za prvo pitanje,  $high_{correct}$  iznosi jer postoji 5 ispitanika koji su točno odgovorili na to pitanje, a svojim ukupnim rezultatom svrstavaju se u uspješniju skupinu. Za to pitanje  $low_{correct}$  ima vrijednost 4 jer postoji četvero ispitanika koji su točno odgovorili na njega, a pripadaju skupini s manjim uspjehom. Diskriminacijski indeks za ovaj zadatak ima vrijednost 10%, što je vrlo nizak i nezadovoljavajući rezultat. Za drugo pitanje,  $high_{correct}$  iznosi 6, a  $low_{correct}$  1, dakle diskriminacijski indeks je 50%, što je poprilično zadovoljavajuće i definitivno bolje od prvog zadatka.

### 1.3.4 Diskriminativna učinkovitost

Diskriminativna učinkovitost (engl. *discriminative efficiency*) mjera je slična diskriminacijskom indeksu, ali se odnosi na cjelokupnu sposobnost mjernog instrumenta da učinkovito diferencira između individualaca ili grupe s različitim



razinama mjerenoga svojstva. Iznimno je bitan faktor pri analiziranju uspješnosti ispita, posebno onih „glavnih“ koji čine velik udio završne ocjene jer bi takvi trebali biti izrazito namijenjeni da sadrže pitanja na kojima se na pravedan način ostvaruju bodovi. Može se računati kombiniranjem navedenih mjera, indeksa lakoće i diskriminacijskog indeksa, kao što je prikazano na formuli (6).

$$\text{diskriminativna učinkovitost} = \frac{\text{diskriminacijski indeks}}{\text{indeks lakoće}} \quad (6)$$

Izračun se može pokazati na izmišljenoj situaciji s četiri pitanja s njihovim indeksom lakoće i diskriminacijskim indeksom, kao što je prikazano u tablici ispod.

Tablica 1.3.4.1 Primjer izračuna diskriminativne učinkovitosti za četiri pitanja

	Diskriminacijski indeks	Indeks lakoće	Diskriminativna učinkovitost
Pitanje 1	75%	50%	150%
Pitanje 2	10%	95%	10.53%
Pitanje 3	45%	85%	52.94%
Pitanje 4	60%	60%	100%

Visoka učinkovitost tipično se dobiva za pitanja s niskim indeksom lakoće i visokom diskriminacijom, dok se male vrijednosti postižu za pitanja s iznimno visokim indeksom lakoće i niskom diskriminacijom. Dakle, uzimajući u obzir dobro poznavanje prirode faktora za izračunavanje, interpretiranje dobivenih rezultata vrlo je intuitivno. Vrijednosti između 30% i 50% pružaju prikladnu diskriminaciju, a one iznad 50% vrlo dobru. Vrijednosti ispod 15% vrlo su nezadovoljavajuće i tako se tretiraju u izvještaju.

### 1.3.5 Vrijednost slučajnog pogotka

Vrijednost slučajnog pogotka (engl. *random guess score*) je mjera rezultata koji može biti dobiven nasumičnim pogađanjem odgovora na pitanje, bez potrebnog znanja i razumijevanja. Podrazumijeva se da vrijedi samo za pitanja višestrukog

izbora gdje je moguće ostvariti takav rezultat odabirom jednog ili više ponuđenih odgovora. Za izračunavanje je bitno utvrditi broj opcija koje se nude kao rješenje zadatka i broj opcija koje zaista jesu točno rješenje te izračunati kolika je vjerojatnost slijepog odabira upravo te kombinacije opcija. Vrijednost slučajnog pogotka računa se prema formuli (7), gdje „ $N_{correct}$ ” označava broj točnih odgovora na pitanje, „ $N$ ” broj svih ponuđenih odgovora i „ $\max(score)$ ” maksimalan rezultat koji može biti ostvaren na zadatku.

$$vrijednost\ slučajnog\ pogotka = \frac{\max(score)}{N} * \frac{1}{\left(\frac{N}{N_{correct}}\right)} \quad (7)$$

Element pogađanja uvijek je prisutan kod zadataka višestrukog izbora, ali iskustvo pokazuje da se ova mjera može smanjiti povećanjem broja ukupnih i ispravnih opcija, kao i kažnjavanjem pogađanja negativnim bodovima. Vrijednost slučajnog pogotka iznimno je bitan faktor pri interpretaciji rezultata ispita jer se može koristiti pri usporedbi s ostvarenim rezultatom pojedinog studenta te dobivenom razlikom procijeniti razinu njegovog znanja i stručnosti. Različite vrijednosti slučajnog pogotka mogu se pokazati na izmišljenom primjeru u kojem prvo pitanje ima maksimalan rezultat 1 i 4 ponuđena odgovora od kojih je samo jedan točan. Drugo pitanje ima također maksimalan rezultat 1, ali 5 ponuđenih odgovora od kojih je samo jedan točan. Sljedeće pitanje ima maksimalan rezultat 2 i 5 ponuđenih odgovora od kojih su 2 točna i posljednje pitanje ima maksimalan rezultat 1 i 6 ponuđenih odgovora od kojih su 3 točna.

Tablica 1.3.5.1 Primjer izračuna vrijednosti slučajnog pogotka za četiri različita pitanja

	Maksimalan rezultat	Broj ponuđenih odgovora	Broj točnih odgovora	Vrijednost slučajnog pogotka
Pitanje 1	1	4	1	6.25%
Pitanje 2	1	5	1	4%
Pitanje 3	2	5	2	4%
Pitanje 4	1	6	3	0.83%

Iz primjera se može vidjeti da vrijednosti slučajnog pogotka opadaju porastom broja ponuđenih odgovora, a rastu porastom maksimalnog rezultata i broja točnih odgovora.

## 1.4 Teorija odgovora na zadatke

Teorija odgovora na zadatke obitelj je matematičkih metoda koja pruža okruženje za razumijevanje i analizu svojstava mjernih instrumenata, naročito u kontekstu pitanja s višestrukim izborom. Počiva na sljedećim pretpostavkama:

- Monotonost – pretpostavka da s porastom latentne varijable koja se mjeri (npr. znanje), raste i vjerojatnost točnoga odgovora na pitanje.
- Lokalna nezavisnost – pretpostavka da su odgovori dani na zasebne stavke u testu međusobno neovisni s obzirom na određenu razinu sposobnosti.
- Jednodimenzionalnost – pretpostavka da je svojstvo koje se mjeri dominantan uzročnik odgovora danih na stavke.
- Invarijanta – pretpostavka da se parametri stavke mogu procijeniti s bilo koje pozicije na krivulji.

Ako su pretpostavke zadovoljene, razlike u opaženim točnim odgovorima između ispitanika rezultat su varijacija u njihovom mjerenom latentnom svojstvu [6]. Teorija nudi različite modele koji uzimaju u obzir jedan ili više faktora za sofisticiran i obuhvatan pristup analizi podataka koja iznosi točne i informativne mjere sposobnosti i svojstava individualaca [2]. 1-parametarski (Rasch) model, uz vrijednost latentne varijable  $\theta$ , temelji se samo na težini pitanja, a 2-parametarski model na težini i diskriminacijskom indeksu. U radu je opisan samo 3-parametarski model, koji, uz težinu i diskriminaciju, uzima u obzir i vjerojatnost

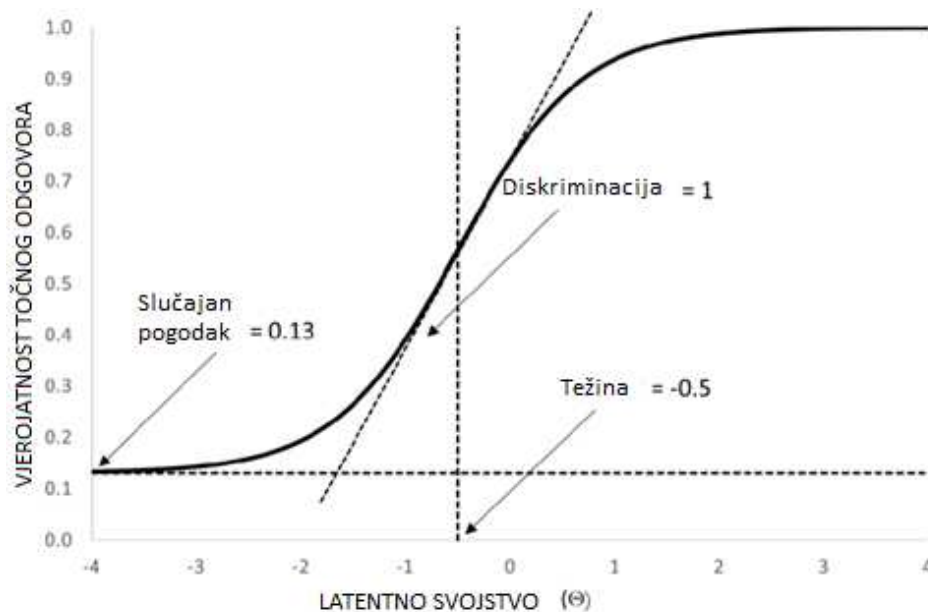
slučajnog pogotka. Formula za računanje vjerojatnosti točnoga odgovora u ovisnosti o razini znanja prikazana je ispod (8). Osim latentne varijable  $\theta$ , koja predstavlja razinu znanja, formula uključuje i konstante  $a$ ,  $b$  i  $c$ .

$$P(\text{točan odgovor} | \theta) = c + \frac{(1-c)}{1 + e^{-a(\theta-b)}} \quad (8)$$

Konstanta  $a$  predstavlja diskriminacijski parametar koji se računa na isti način kao i prethodno opisani diskriminacijski indeks. Veće vrijednosti su poželjne i ukazuju na to da pitanje dobro diskriminira između grupa s različitim razinama znanja. Konstanta  $b$  predstavlja mjeru težine pitanja. Ta težina poprima onu vrijednost latentne varijable  $\theta$  za koju ispitanik ima 50% vjerojatnosti za točan odgovor. Na primjer, pri vrlo niskoj težini, poput 0, ispitanik čija je razina znanja 0 ima 50% šanse točno odgovoriti na pitanje. U skladu s tim, niske vrijednosti ovog parametra ukazuju na jednostavna, a visoke na teška pitanja. Ova mjera vrlo je bitna jer dijeli skup prostora vjerojatnosti u dva jednaka dijela i čini prag kojim se razine znanja dijele u one s manjom i većom vjerojatnosti točnog odgovora. Konstanta  $c$  je parametar pseudo-slučajnog pogađanja koji je bitan samo za pitanja višestrukog izbora s dva ili više ponuđenih odgovora. Predstavlja vjerojatnost da ispitanik točno odgovori na pitanje pri vrlo niskim razinama znanja. Ovom se parametru tipično pridjeljuje vrijednost vjerojatnosti slučajnog pogotka. Ta vjerojatnost računa se prema formuli ispod (9), gdje „ $N$ “ predstavlja broj svih ponuđenih odgovora, a „ $N_{correct}$ “ broj točnih.

$$c = \frac{1}{\binom{N}{N_{correct}}} \quad (9)$$

Kao grafička reprezentacija veze između vjerojatnosti točnog odgovora i razine znanja, koristi se karakteristična krivulja zadatka. Može imati više izgleda za različite inačice modela, ali je uglavnom u S-obliku logističke funkcije [2]. Os  $x$  na karakterističnoj krivulji predstavlja vrijednosti latentne varijable  $\theta$ , a os  $y$  predstavlja vjerojatnost točnog odgovora na pitanje. Izgled krivulje prikazan je na slici ispod.



Slika 1.4.1: Karakteristična krivulja zadatka za 3-parametarski model

Na slici su označeni i parametri težine, diskriminacije i vjerojatnosti slučajnog pogotka. Mala vrijednost težine pitanja rezultira pomakom krivulje u lijevo, što označava da je za točan odgovor potrebna manja vrijednost  $\theta$  (znanja), a velika težina rezultira pomakom u desno, indicirajući da je veća razina znanja potrebna za točan odgovor. Za ovaj konkretan primjer, težina pitanja poprima vrijednost -0.5, zbog čega je krivulja pomaknuta u lijevo i može se zaključiti da je potrebno manje znanja za točan odgovor. Parametar diskriminacije određuje nagib krivulje; za visoku diskriminaciju nagib je strm, a za nisku je blag. Krivulja ima brži rast za ispitanike s većom razinom  $\theta$ , što ukazuje na to da oni imaju veću vjerojatnost točnoga odgovora, a brže opada za ispitanike s nižom razinom  $\theta$ . Vrijednost diskriminacije u primjeru na slici iznosi 1 i nagib krivulje je strm, što znači da pitanje poprilično dobro diskriminira između uspješnije i neuspješnije grupe ispitanika. Parametar vjerojatnosti slučajnog pogotka na krivulji je predstavljen donjom asimptotom krivulje, linijom kojoj se krivulja približava kad se vrijednost  $\theta$  snižava prema negativnoj beskonačnosti. Ovom se mjerom određuje vertikalni pomak krivulje u odnosu na os y. Za veću vjerojatnost slučajnog pogotka, cjelokupna vjerojatnost točnoga odgovora visoka je za cijeli model, a za nisku vjerojatnost slučajnog pogotka suprotno. Budući da u danom primjeru ovaj parametar iznosi 0.13, vjerojatnost točnoga odgovora ne može poprimiti manje vrijednosti od 0.13 čak ni za minimalne razine znanja. Za primjer sa slike,

uvrštavanjem parametara u formulu (8), vjerojatnost točnog odgovora za vrijednost  $\theta$  od 0 i parametre  $s$  krivulje iznosi 0.6715, dakle 67.15%. Vrijedi napomenuti da je metrika latentne varijable  $\theta$  donekle proizvoljna, ali je u referentnoj populaciji, koja je predstavljena kao metrička norma, procijenjena srednja vrijednost  $\theta$  nula sa standardnom devijacijom 1. Sukladno tome, vrijednosti  $\theta$  na grafu su centrirane oko nule. Za ispitanike koji su po svojim razinama znanja za jednu vrijednost standardne devijacije iznad srednje vrijednosti referentne populacije, njihova  $\theta$  po standardnoj praksi iznosi 1. Vrijednosti  $\theta$  teoretski mogu poprimati vrijednosti između negativne i pozitivne beskonačnosti, ali većina ispitanika ima  $\theta$  između -3 i 3.

Teorija odgovora na zadatke uspješno je primjenjiva u mnogo okruženja koji podrazumijevaju procjene, poput obrazovanja i psihologije. Također, može se koristiti za osmišljavanje skala i mjera uključivanjem stavki s visokom snagom diskriminiranja koje povećavaju preciznost mjernog alata i smanjuju nepraktično odgovaranje na duge upitnike [6].

## 2 Primjer statistike ispita na platformi Moodle

Moodle (*Modular Object-Oriented Dynamic Learning Environment*) je jedan od najpopularnijih sustava za upravljanje učenjem (engl. *LMS, Learning Management System*). Nudi iznimno fleksibilno i bogato okruženje za odvijanje online nastave zbog mogućnosti organizacije i podjele materijala za učenje, interakcije sa studentima i ocjenjivanja njihovog napretka provođenjem online ispita. Za svaki ispit proveden na Moodleu, uz dopuštenje profesora, studenti mogu vidjeti sveobuhvatan izvještaj testne statistike. Izvještaj se dijeli na prvi dio, koji uključuje sažetak cijelog ispita i drugi dio, koji opisuje statističke pokazatelje za svako ispitno pitanje. U prvom dijelu, uz naziv predmeta, naziv ispita i cjelokupni broj studenata koji su mu pristupili, prikazan je popis statističkih pokazatelja izračunatih za rezultate ispita. Budući da je na Moodleu nekim ispitima, poput kratkih provjera znanja, moguće pristupiti više puta, za svaki se pristup računa prosječna ocjena (srednja vrijednost) rezultata. Za kasnije pokušaje ona je tipično uvijek veća od prvoga. Umjesto srednje vrijednosti, prikazan je medijan, drugi pokazatelj koji opisuje sredinu vrijednosti među podacima. Za svaki ispit izračunata je standardna devijacija i mjere iskrivljenosti (engl. *skewness*) i izbočenosti (engl. *kurtosis*) za prvi pokušaj pristupa ispitu. Prvi pokušaj najvjernije prikazuje stvarno znanje i sposobnosti, dok na ostalima postoji mogućnost pamćenja ponovljenih pitanja. Prvi dio uključuje i Cronbach alfu kao mjeru unutarnje konzistentnosti zadataka na ispitu, kao i omjer pogreške (engl. *error ratio*) i standardnu pogrešku (engl. *standard error*). Omjer pogreške mjeri koliki udio varijacije u ocjenama dolazi od slučajne varijacije [8]. Budući da varijacija u ocjenama dolazi iz činjenice da neki studenti imaju više znanja od ostalih i određene količine slučajne varijacije, poželjno je da ova mjera poprima što manju vrijednost [8]. Standardna pogreška uzima se iz vrijednosti omjera pogreške i predstavlja mjeru koliko slučajne varijacije postoji u pojedinoj ocjeni na testu. Na primjer, ako standardna pogreška iznosi 10%, a student je ostvario rezultat od 60%, onda se stvarna sposobnost studenta nalazi u intervalu od 50% do 70% [8]. Svi navedeni pokazatelji prikazani su na slici ispod [9] preuzetoj sa službene stranice Moodleove dokumentacije.

## Quiz information

Download full report as

Comma separated values (.csv)

Download

Quiz name	Water Tester Quiz
Course name	Water!
Open the quiz	Friday, 1 February 2019, 1:50 PM
Number of complete graded first attempts	23
Total number of complete graded attempts	31
Average grade of first attempts	50.43%
Average grade of all attempts	55.48%
Average grade of last attempts	57.38%
Average grade of highest graded attempts	57.38%
Median grade (for first attempt)	40.00%
Standard deviation (for first attempt)	15.81%
Score distribution skewness (for first attempt)	0.5274
Score distribution kurtosis (for first attempt)	-0.2490
Coefficient of internal consistency (for first attempt)	12.66%
Error ratio (for first attempt)	93.46%
Standard error (for first attempt)	14.77%

### Sl. 2.1 Prikaz testne statistike i njenih komponenti na platformi Moodle

Drugi dio izvještaja ispitne statistike na Moodleu podrazumijeva analizu pojedinih zadataka s ispita. Svaki zadatak opisan je svojim rednim brojem pojavljivanja na ispitu, svojim imenom i brojem koliko je puta riješen, tj. koliko je studenata dalo odgovor na taj zadatak. Ovaj dio uključuje prethodno definirane mjere, koje podrazumijevaju indeks lakoće, diskriminacijski indeks, diskriminativnu učinkovitost i rezultat slučajnog pogotka. Navedene mjere poslužile su kao predložak i inspiracija za kreiranje statističkog izvještaja za Edgar. Također, analiza zadataka sadrži podatke o njihovoj standardnoj devijaciji te mjerama predviđene težine (engl. *intended weight*) i ostvarene težine (engl. *effective weight*). Predviđena je težina definirana pri kreiranju zadatka, npr., ako je za zadatak postavljeno da donosi 5 bodova od ukupnih 20 na ispitu, njegova predviđena težina je 25%. Ostvarena težina procjenjuje se iz rezultata i mjeri koliko je stvarni udio varijacije nastao zbog tog pitanja [8]. Poželjno je da ove dvije mjere



imaju slične ili jednake vrijednosti. Na slici ispod [10] prikazana je analiza zadataka u nekom ispitu u sustavu Moodle. Može se primijetiti da je naslov pitanja naglašen crvenom bojom za pitanja čija diskriminativna učinkovitost poprima vrijednosti manje od 15%.

**Quiz structure analysis**

Download table data as a comma separated values text file

Q#	Question name	Attempts	Facility index	Standard deviation	Random guess score	Intended weight	Effective weight	Discrimination index	Discriminative efficiency
1	El Nino	23	43.48%	50.69%	25.00%	20.00%	34.44%	1.96%	3.16%
2	Match the river	23	95.65%	20.85%	33.33%	20.00%	17.67%	16.93%	100.00%
3	What type of water?	23	21.74%	42.17%	0.00%	20.00%	33.26%	28.51%	42.50%
4	Every 20 seconds !!!	23	91.30%	28.81%	50.00%	20.00%	14.63%	-15.77%	-35.29%
5	Rich and poor	23	0.00%	0.00%	0.00%	20.00%	0.00%		

Sl. 2.2 Prikaz podataka i statističkih pokazatelja za pojedine zadatke na ispitu u sustavu Moodle

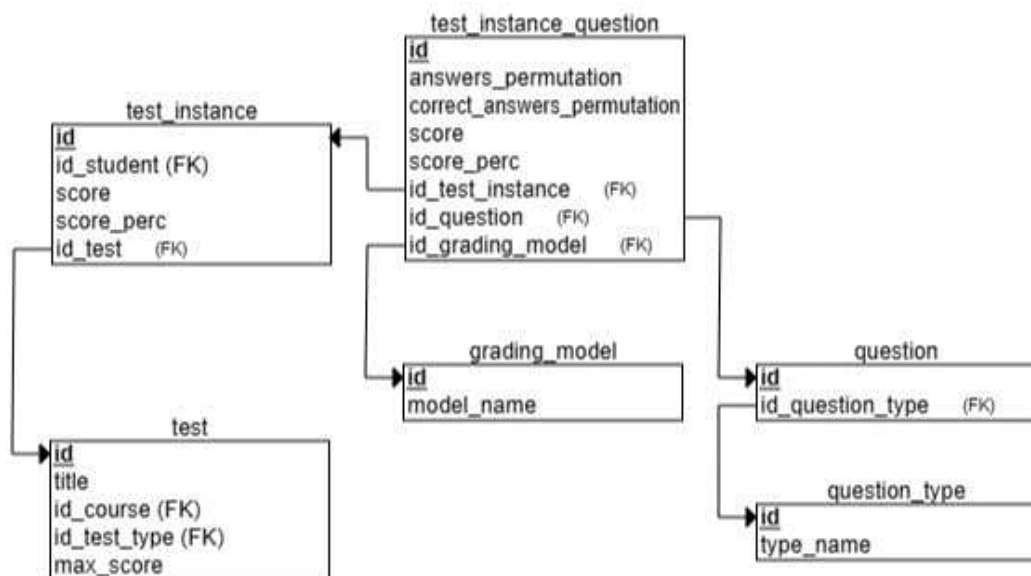
Klikom na bilo koje ime zadatka prikazuje se tablica s istim podacima koji su navedeni u gornjoj tablici, ali posebno izdvojeno za pojedini zadatak.

U sustavu Moodle se na analizi testne statistike mogu prepoznati uzorci koji narušavaju pretpostavku da studenti koji ostvare visoku ocjenu na testu imaju veću vjerojatnost za odgovoriti točno na pojedini zadatak [8]. Fokus je na analizi rezultata svakog pitanja i identifikaciji onih koji ne spadaju u referentne okvire definirane za njihove pokazatelje. Takva se pitanja reformiraju, zamjenjuju s drugim pitanjima koji bolje diskriminiraju između uspješnih i neuspješnih rezultata ili izbacuju radi povećanja unutarnje konzistentnosti.

## 3 Model podataka sustava Edgar

### 3.1 Relacijska shema

Relacijska shema je skica koja opisuje organizaciju podataka u relacijskoj bazi. Prikazuje tablice, attribute tablica i međusobni odnos između njih. Sadrži oznake o primarnim i stranim ključevima unutar tablica, kao i povezanosti pojedinih tablica koje pomoću ključeva ovise jedna o drugoj. Podaci iz prikazane pojednostavljene sheme na slici ispod koriste se kod izračuna statističkih značajki.



Sl. 3.1: Pojednostavljena relacijska shema dijela baze podataka sustava Edgar koja se odnosi na provjere znanja i uspjehe na provjerama

### 3.2 Opis i prikaz korištenih tablica i atributa iz sustava Edgar

Baza podataka sustava Edgar relacijska je baza podataka, što znači da su podaci spremjeni u tablice strukturirane u obliku redova i stupaca. Redovi tablica označavaju instance entiteta koji tablica opisuje, a stupci označavaju atribute, tj. svojstva tih entiteta. Tablice su međusobno logički povezane njihovim primarnim i stranim ključevima. U bazi su spremljene 133 tablice, od kojih je vrlo malen udio korišten pri rješavanju završnog zadatka. Također, nisu svi atributi koji se nalaze u tablici potrebni za programski dio zadataka pa su prikazani oni najosnovniji. Skup najbitnijih korištenih tablica sadrži:

- „test“ – tablica koja opisuje pojedine ispite, identifikacijski broj ispita ključan je atribut u rješavanju problema jer se cijeli programski dio izvodi na osnovu njegove vrijednosti

Tablica 3.2.1 Opis atributa za tablicu „test“ iz baze podataka sustava Edgar

Atribut	Tip podataka	Opis
id	integer	identifikacijski broj ispita, cjelobrojni tip podataka koji je primarni ključ tablice
title	varchar (255)	ime ispita
id_course	integer	identifikacijski broj predmeta, strani ključ na tablicu „course“
id_test_type	integer	identifikacijski broj tipa ispita, strani ključ na tablicu „test_type“
max_score	numeric (5,2)	maksimalan rezultat koji se može postići za pojedini ispit

- „test\_instance“ – instanca jednoga ispita, predstavlja ispit pojedinog studenta i sadrži informacije o identitetu studenta i ostvarenom rezultatu za taj ispit

Tablica 3.2.2 Opis atributa za tablicu „test\_instance“ iz baze podataka sustava Egdar

Atribut	Tip podataka	Opis
id	integer	identifikacijski broj instance, primarni ključ
id_test	integer	identifikacijski broj ispita, strani ključ na tablicu „test“
id_student	integer	identifikacijski broj studenta, strani ključ na tablicu „student“
score	numeric (10,3)	ukupni ostvareni rezultat na jednoj instanci ispita
score_perc	numeric (10,3)	postotak ukupnog ostvarenog rezultata na instanci ispita

- „question\_type“ – daje informaciju o tipu pitanja u ispitu

Tablica 3.2.3 Opis atributa za tablicu „question\_type“ iz baze podataka sustava Edgar

Atribut	Tip podataka	Opis
id	integer	identifikacijski broj tipa pitanja, primarni ključ tablice
type_name	varchar (50)	ime tipa pitanja

- „test\_instance\_question“ – pitanje koje se pojavilo na određenoj instanci ispita, sadrži informacije o ispitu na kojemu se zadatak pojavio, ostvarenom rezultatu pojedinog studenta za taj zadatak, oznaku o modelu ocjenjivanja i oznaku o točnim odgovorima za pitanja višestrukog izbora

Tablica 3.2.4 Opis atributa za tablicu „test\_instance\_question“ iz baze podataka sustava Edgar

Atribut	Tip podataka	Opis
id	integer	identifikacijski broj pitanja na instanci ispita, primarni ključ tablice
id_test_instance	integer	identifikacijski broj instance ispita, strani ključ na tablicu „test_instance“
id_question	integer	identifikacijski broj pitanja, strani ključ na tablicu „question“
id_grading_model	integer	identifikacijski broj modela ocjenjivanja, strani ključ na tablicu „grading_model“
answers_permutation	integer[]	permutacija ponuđenih odgovora za zadatke višestrukog izbora
correct_answers_permutation	integer[]	permutacija točnih odgovora za zadatke višestrukog izbora
score	numeric (10,3)	rezultat ostvaren na zadatku
score_perc	numeric (10,3)	postotak rezultata ostvarenog na zadatku

- „question“ – sadrži podatke o pitanju na ispitu

Tablica 3.2.5 Opis atributa za tablicu „question“ iz baze podataka sustava Edgar

Atribut	Tip podataka	Opis
id	integer	identifikacijski broj pitanja, primarni ključ tablice
id_question_type	integer	identifikacijski broj tipa pitanja, strani ključ na tablicu „question_type“

- „grading\_model“ – model ocjenjivanja zadatka, daje informaciju koliko bodova i negativnih bodova donosi pojedini zadatak

Tablica 3.2.6 Opis atributa za tablicu „grading\_model“ iz baze podataka sustava Edgar

Atribut	Tip podataka	Opis
id	integer	identifikacijski broj modela, primarni ključ tablice
model_name	varchar (50)	ime modela, u formatu x/y/z, gdje x predstavlja broj bodova za točan odgovor, y za neodgovoreno i z za netočan

## 4 Opis korištenih tehnologija

Tehnologije korištene pri izradi završnog rada su PgAdmin za lokalnu pohranu baze podataka i analizu strukture tablica unutar nje i RStudio za analizu samih podataka, manipulaciju podacima i izračunavanje potrebnih statističkih pokazatelja.

### 4.1 PgAdmin

PgAdmin besplatan je alat za upravljanje i administraciju PostgreSQL baza podataka [5]. Pruža pregledno grafičko korisničko sučelje (engl. *Graphical User Interface, GUI*) koje omogućuje korisnicima vizualan uvid u strukturu baze podataka te olakšava navigaciju, izvršavanje upita i administrativne zadatke [5]. Omogućuje kreiranje novih baza podataka, kao i njihov pregled i uređivanje. Upiti (engl. *query*), koji podrazumijevaju naredbe i instrukcije za izvršavanje operacija nad podacima, pišu se u programskom jeziku SQL (*Structured Query Language*). Moguće je obaviti pretragu potrebnih podataka, unos novih u definirane tablice te brisanje i ažuriranje postojećih [5]. Korisnik piše upite unutar „query tool“ alata te pregledava rezultate u jednostavnom korisničkom sučelju. Također, podržani su razni administrativni zadaci kao što su upravljanje korisnicima i ulogama, sigurnosno upravljanje i postavljanje ovlasti nad bazom podataka te praćenje performansi. PgAdmin je popularan alat među PostgreSQL korisnicima zbog svojih moćnih funkcionalnosti, jednostavnosti korištenja i podrške za različite operativne sustave. Olakšava administraciju baza podataka i pomaže korisnicima da učinkovito upravljaju svojim podacima.

### 4.2 RStudio

RStudio je integrirano razvojno okružje (engl. *Integrated Development Environment, IDE*) specifično dizajnirano za programski jezik R [4]. Pruža korisnicima alate i funkcionalnosti za razvoj, izvršavanje i analizu statističkih programa. Njegove bogate tehničke karakteristike iznimno olakšavaju rad s jezikom R, uključujući automatsko izvršavanje koda, prikazivanje dokumentacije

i interaktivno razrješavanje pogrešaka u kodu. Korisničko sučelje omogućuje preglednu navigaciju između programskog dijela dokumenta, konzole za vanjski ispis, preglednika podataka, grafičkih prikaza i drugih prozora. Velika prednost ovog alata elegantna je instalacija mnogih paketa i biblioteka, koja podrazumijeva jednostavno preuzimanje s interneta i uključivanje unutar projekta [4]. Paketi su iznimno robusni i sadrže mnoge lako dostupne i dobro objašnjene funkcije ključne za provođenje statističke analize podataka. Umjesto vanjskih paketa, sam RStudio ima mnoge ugrađene alate za učitavanje, manipulaciju i vizualizaciju podataka. Zbog svih navedenih pogodnosti, u vrlo je širokoj upotrebi među statističarima i analitičarima podataka.

U kontekstu samog završnog rada, RStudio je poslužio kao iznimno koristan alat zbog mogućnosti povezivanja na PostgreSQL bazu podataka, unaprijed ugrađenih funkcija za izračunavanje mnogih statističkih pokazatelja i raznolikih alata za uređivanje grafova potrebnih za prikaz razdiobe podataka.



## 5 Oblikovanje i implementacija rješenja zadatka

### 5.1 Povezivanje baze podataka s radnim okruženjem

Rješenje zadatka osmišljeno je u obliku R skripte koja se pokreće u alatu RStudio i generira pdf dokument s detaljnim izvještajem o statistici rezultata na ispitu. Za početak rada prvo je bilo potrebno povezati programski dio s bazom podataka, što je ostvareno korištenjem paketa „RPostgreSQL“, pomoću kojeg se stvara objekt baze podataka unutar razvojnog okruženja RStudio. Na ovaj se način nad tim objektom mogu jednostavno provoditi upiti u jeziku SQL i pridruživati tražene vrijednosti raznim varijablama i objektima. Za definiranje nužnih podataka za razna izračunavanja i ostvarivanje ciljeva, bilo je neophodno detaljno proučiti već integriranu bazu.

### 5.2 Upute za pokretanje izvorne datoteke

Programski dio rješenja bilo je potrebno implementirati tako da podržava svaku vrstu ispita koja se nalazi unutar baze. Generirani pdf dokument koncipiran je kao izvještaj na razini jednog ispita i zadataka na tom ispitu. Ulazni parametar koji se zadaje programu je identifikacijski broj ispita koji se pridružuje varijabli „id\_test“. Dakle, potrebno je odrediti za koji ispit se želi dobiti izvještaj, pronaći njegov identifikacijski broj u bazi i zatim ga ručno unijeti kao sadržaj varijable, kao što je prikazano na slici ispod.

```
{r message=FALSE, warning=FALSE, include=FALSE}  
id_test <- 13846
```

Sl.5.2.1 Primjer pridruživanja vrijednosti varijabli „id\_test“ u izvornoj datoteci „ispiti.Rmd“

Svi isječci koda koji se nalaze u izvornoj datoteci imaju definirana svojstva za njihovo izvršavanje i vidljivost pri generiranju dokumenta, zbog čega nije potrebno i jedan dio programa ručno pokretati. Stvaranjem dokumenta svi se

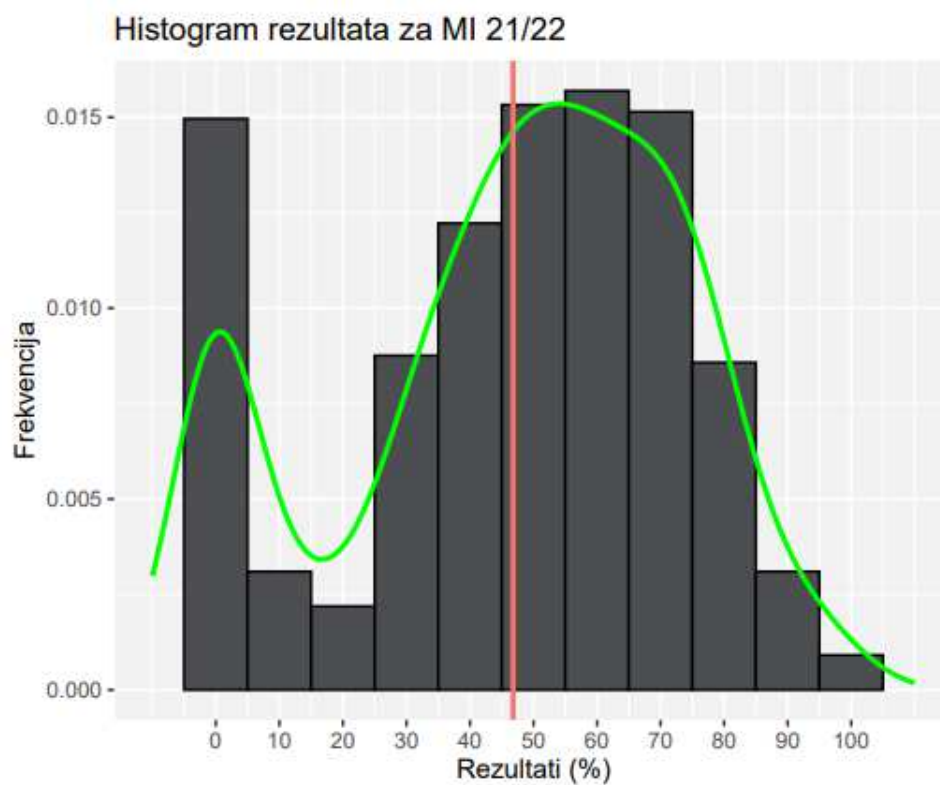
dijelovi automatski izvrše i prikazuju točne vrijednosti za svaki ispit. Dokument se može generirati na više načina i u različitim formatima, ali je najpreglednija opcija dobivanje pdf datoteke koja će biti spremljena u trenutni direktorij/kazalo. U sklopu grafičkog korisničkog sučelja alata RStudio postoji opcija „Knit“, koju je potrebno odabrati i zatim „Knit to PDF“. Za nekoliko trenutaka, generirat će se datoteka, odmah otvoriti u novom prozoru i spremiti na lokalno računalo.

## **5.3 Komponente izvještaja o uspjehu rezultata ispita**

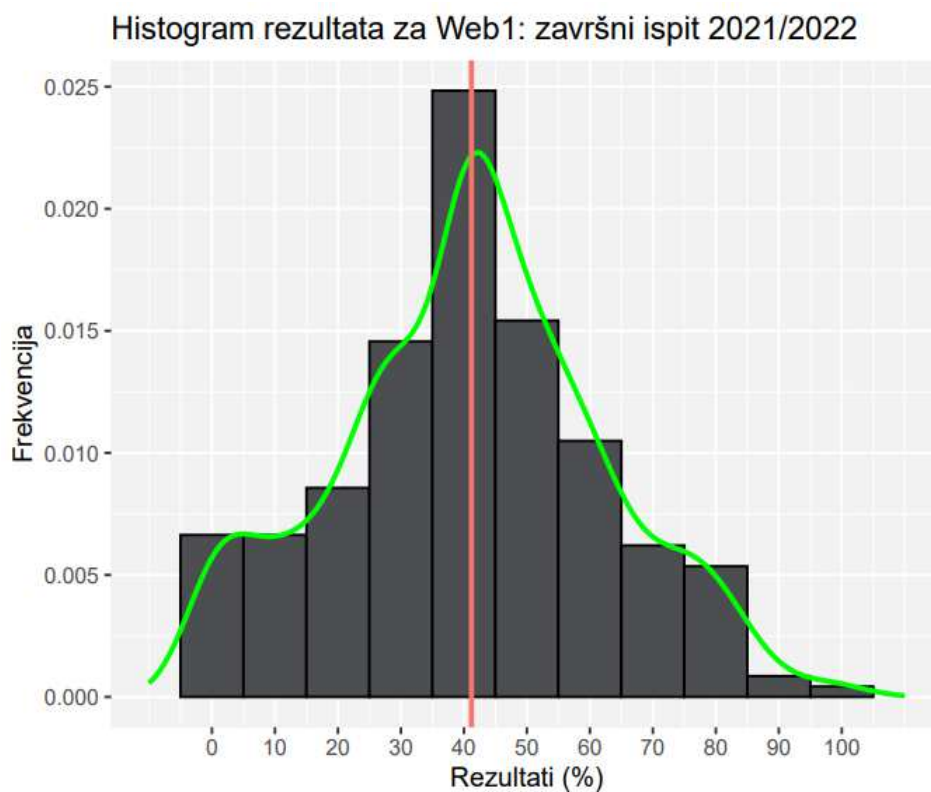
Izvještaj sadrži grafičke prikaze i izračunate statističke pokazatelje za cijeli ispit i pojedini zadatak na ispitu. Sve stavke sažeto su opisane radi smanjenja potrebe za traženjem njihovih definicija i značenja, te razumijevanja dobivenih rezultata. Izvještaj je podijeljen na dio koji se odnosi za cijeli ispit i zatim dio koji se odnosi na pojedine zadatke.

### **5.3.1 Histogram**

Prvo je prikazan histogram, graf distribucije numeričkih podataka na kojemu svaki stupac predstavlja broj studenata koji su ostvarili određen postotak na ispitu. Stupci predstavljaju kategorije postotaka podijeljenih u intervalu od 0 do 100, a instance testa grupirane su unutar tih kategorija prema ostvarenom rezultatu. Po izgledu histograma da se naslutiti priroda distribucije vrijednosti i može se usporediti s normalnom ili Gaussovom distribucijom zvonolikog oblika, koja se najčešće koristi kao predložak za zadovoljavajuću distribuciju bodova na ispitu. Na grafu je zelenom linijom označena gustoća vjerojatnosti koja predstavlja vjerojatnosti da se vrijednost podataka (rezultat na ispitu) nalazi unutar određenog raspona (širina stupca koja predstavlja interval od 10%). Ta se vjerojatnost iščitava kao površina ispod krivulje na području pojedinog intervala. Uglavnom prati vrhove stupaca i kontrastom boje jasnije ukazuje na izgled razdiobe. Crvena linija na grafu predstavlja srednju vrijednost, aritmetičku sredinu svih ostvarenih bodova.



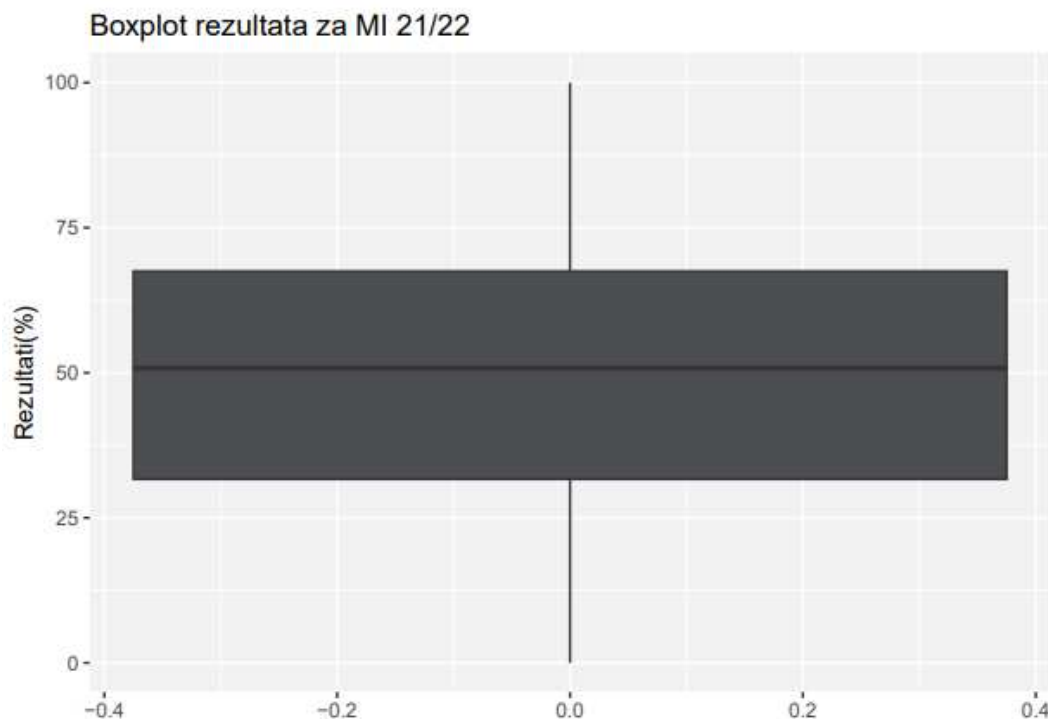
Sl. 5.3.1.1 Histogram za Međuispit iz predmeta Baze podataka 2021/2022



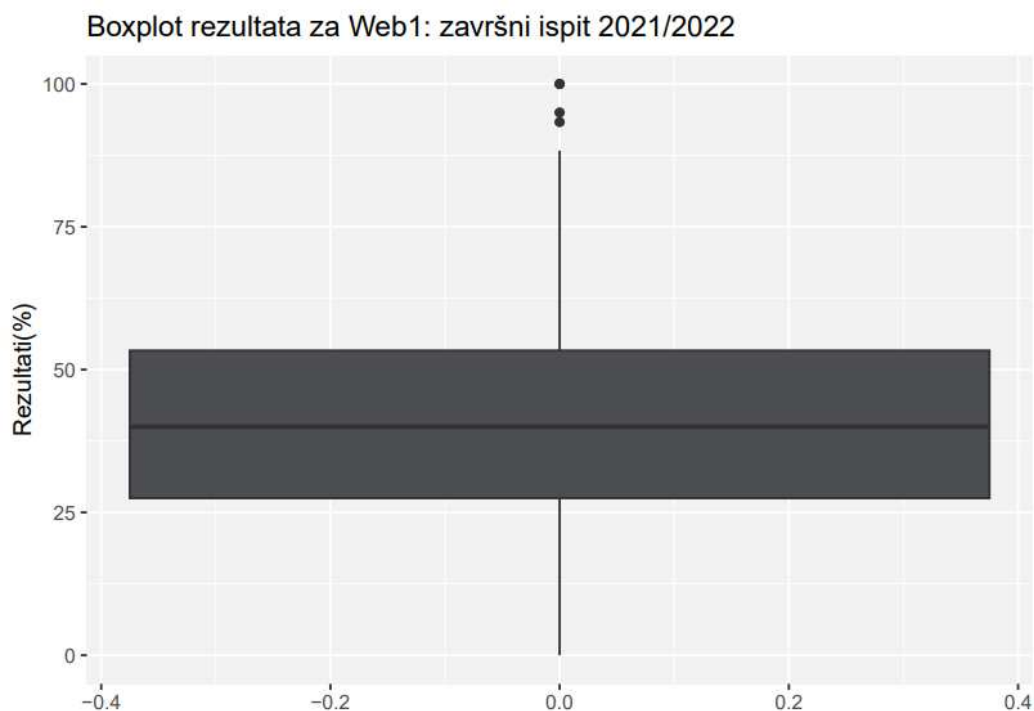
Sl. 5.3.1.2 Histogram za Završni ispit iz predmeta Razvoj programske potpore za web 2021/2022

### 5.3.2 Kutijasti dijagram (engl. *boxplot*)

Nakon histograma, opisan je i prikazan tzv. kutijasti dijagram (engl. *boxplot*), graf distribucije numeričkih podataka u obliku kutije koji prikazuje distribuciju rezultata po kvartilima. Kvartili su statističke mjere koje se koriste za podjelu podataka u četiri jednaka dijela, dakle 25% ukupnih podataka. Prvi kvartil je vrijednost ispod koje se nalazi 25% najnižih rezultata iz uzorka i grafički je predstavljen kao donji rub pravokutnika. Linija u sredini označava medijan ili drugi kvartil, središnji podatak ispod kojeg je 50% rezultata iz uzorka. Nadalje, treći kvartil poprima vrijednost podatka ispod kojega se nalazi 75% uzorka. Izgled dijagrama također daje informaciju o simetričnosti ili asimetričnosti podataka; ako je pravokutnik centriran, gornji i donji rub pravokutnika su jednako udaljeni od srednje linije i izdanci (linije s donje i gornje strane) su jednake duljine s obje strane, ukazuje na simetričnu razdiobu. Moguće je i dobiti ekstremne vrijednosti (engl. *outliers*) u obliku krugova koji se protežu van granica izdanaka. Te vrijednosti u velikim razmjerima odstupaju od ostatka uzorka i utječu na simetričnost.



Sl. 5.3.2.1 Boxplot za Međuispit iz predmeta Baze podataka 2021/2022

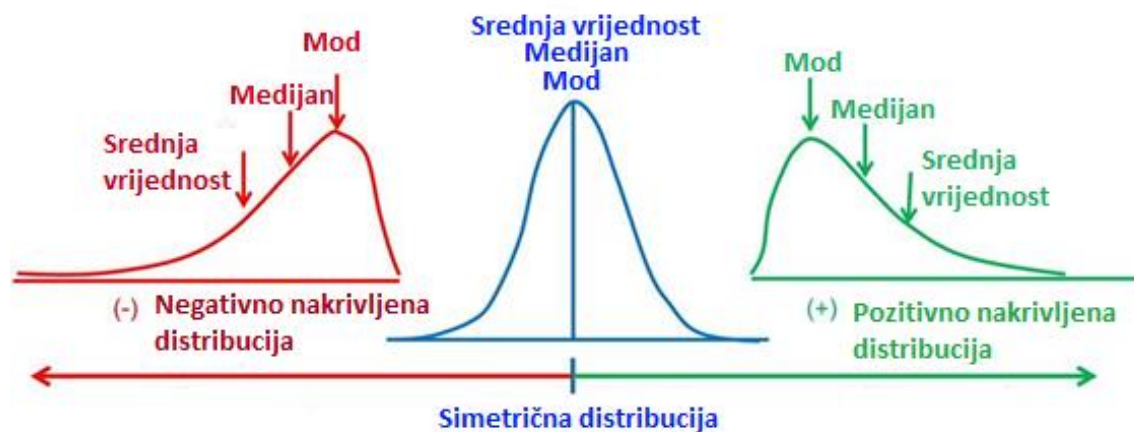


Sl. 5.3.2.2 Boxplot za Završni ispit iz predmeta Razvoj programske potpore za web 2021/2022.  
prisutnost ekstremnih vrijednosti

### 5.3.3 Statistički pokazatelji za ispit

Nakon dijagrama, unutar tablice su navedeni su statistički pokazatelji za ispit. Uključuju vrijednosti već implementirane u dosadašnjem izvještaju na Edgaru, što podrazumijeva srednju vrijednost, medijan, standardnu devijaciju, minimalan i maksimalan rezultat. Također je uključena standardna pogreška koja predstavlja mjeru nepreciznosti procjene srednje vrijednosti. Jedna od novosti na izvještaju je tzv. iskrivljenost (engl. *skewness*), koja mjeri asimetriju distribucije rezultata, tj. stupanj u kojemu je ona nakrivljena prema lijevo i desno. Pozitivne vrijednosti tumače se u obliku teškog repa s desne strane, što ukazuje na situaciju u kojoj je većina vrijednosti koncentrirana prema nižem dijelu, a ekstremne vrijednosti (engl. *outliers*) povlače srednju vrijednost prema većim rezultatima. Negativne vrijednosti imaju suprotan učinak, rep je težak s lijeve strane, a niske stršeće vrijednosti povlače srednju vrijednost s koncentrirane velike na manju. Vizualna reprezentacija mjere prikazana je na slici ispod. Crveno obojenim grafom prikazana je negativno nakrivljena distribucija, zeleno obojenim

pozitivno nakrivljena te u sredini, radi usporedbe, plavo obojena simetrična distribucija. Na slici se vidi odnos srednje vrijednosti, moda i medijana za sve distribucije. Mod je statistička mjera koja predstavlja određenu vrijednost koja se najčešće pojavljuje u uzorku. Koristan je za određivanje tipične ili reprezentativne vrijednosti.

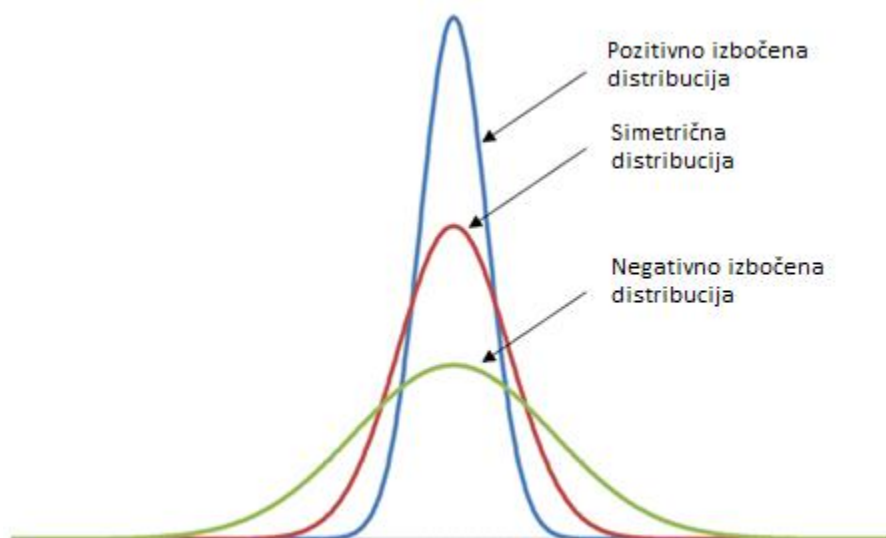


Sl. 5.3.3.1 Usporedba distribucija s negativnom i pozitivnom iskrivljenosti u odnosu na simetričnu distribuciju

U idealnim uvjetima, srednja vrijednost, mod i medijan imaju isti iznos, ali to često nije slučaj. Za negativno nakrivljenu distribuciju, srednja vrijednost je najmanja, medijan je veći, a mod ima najveću vrijednost, a za pozitivno nakrivljenu raspored vrijednosti je obrnut.

Sljedeća mjera je izbočenost (engl. *kurtosis*), koja mjeri vertikalna odstupanja od normalne distribucije te daje informaciju o prisutnosti ekstremnih vrijednosti. Negativan rezultat ukazuje na nizak i ravan vrh i lake repove distribucije (krajevi linije koji se protežu prema ekstremnim vrijednostima s lijeve i desne strane). Laki rep odnosi se na situaciju kada rep distribucije opada sporije od predviđenog ili očekivanog. To znači da ekstremne vrijednosti u repu distribucije nisu tako rijetke kao što se očekuje prema teorijskom modelu. S druge strane, pozitivan rezultat ukazuje na veliku izbočenost vrha i teške repove, koji opadaju brže od očekivanoga. Slika ispod prikazuje odnos pozitivne izbočenosti, negativne izbočenosti te normalne distribucije s vrijednosti izbočenosti 0. Pozitivno

izbočena distribucija prikazana je plavom linijom, negativno izbočena zelenom, a crvena linija predstavlja normalnu distribuciju umjerenih dimenzija.



Sl. 5.3.3.2 Odnos negativne, pozitivne i normalne izbočenosti razdioba

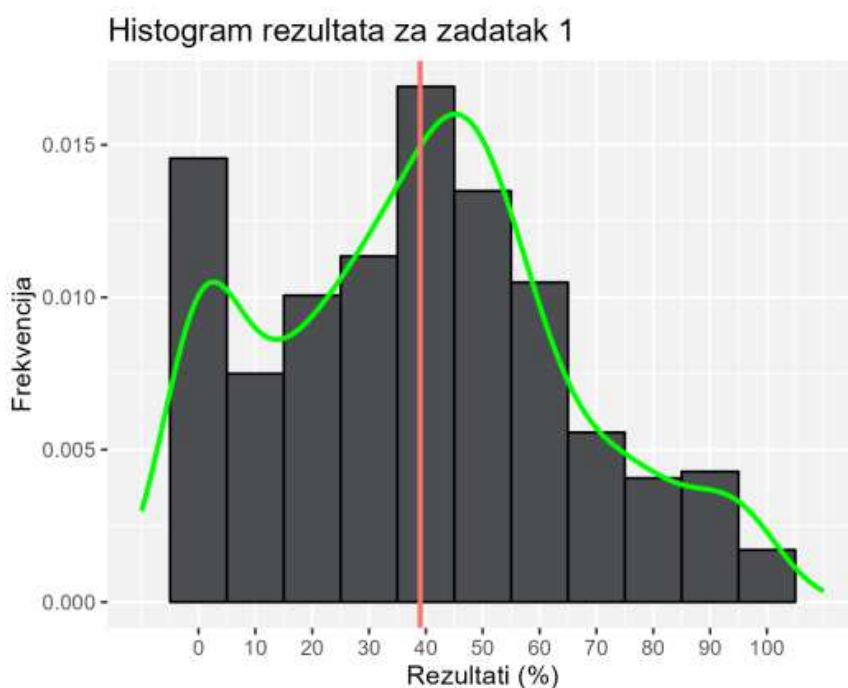
Posljednja vrijednost tablice je prethodno opisana Cronbach alfa. Vrijednosti ispod 50% (0.5) u tablici su označene crvenom, a veće vrijednosti zelenom bojom. Kao što je prije spomenuto, ovu mjeru moguće je izračunati samo za ispit s pitanjima višestrukog izbora. Također, za pojedine testove koji su toga tipa, Cronbach alfa nije prikazana jer ponekad u podacima nema dovoljno varijabilnosti za njezin izračun (ako su neki pitanja u svim slučajevima točno ili netočno riješena).

ID	Srednja vrijednost	Medijan	Standardna devijacija	Minimalan rezultat	Maksimalan rezultat	Standardna pogreška	Skewness	Kurtosis	Cronbach alfa
13,539	8.174	8.188	3.884	-1.25	18.75	0.868	0.163	0.099	21.36%

Sl. 5.3.3.3 Tablica statističkih pokazatelja za Međuispit iz Naprednog razvoja programske potpore za web 2022/2023

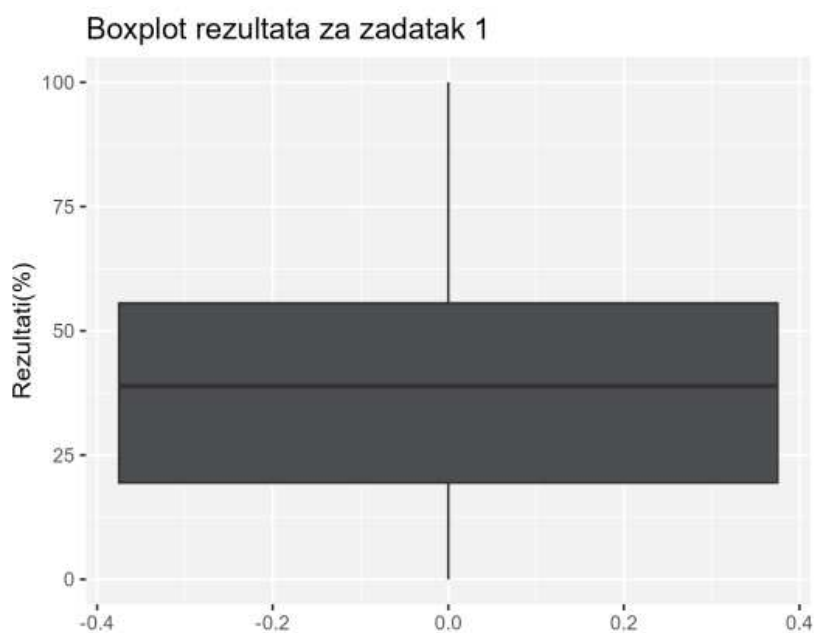
### 5.3.4 Grafički prikazi distribucije bodova pojedinih zadataka

Za pojedine ispite, moguć je prikaz histograma i kutijastih dijagrama ostvarenih bodova. Taj dio izvještaja ne prikazuje se za ispite koji sadrže prevelik broj zadataka od kojih se samo malen dio nasumično pojavljuje u određenim instancama ispita i nema dovoljno varijabilnosti u rezultatima da bi distribucija bila jasno vidljiva na dijagramima.



Sl. 5.3.4.1 Histogram rezultata za prvi zadatak za završni ispit iz predmeta Razvoj programske potpore za web 2021/2022





Sl. 5.3.4.2 Boxplot rezultata za prvi zadatak za završni ispit iz predmeta Razvoj programske potpore za web 2021/2022

### 5.3.5 Statistički pokazatelji za zadatak

Sljedeći dio izvještaja uključuje tablicu sa statističkim pokazateljima za sve zadatke iz ispita, kao i opis i tumačenje pojavljivanih vrijednosti i boja u tablici. Uz identifikacijske brojeve, priložen je tip pitanja koji čitatelju daje naslutiti razlog nekih dobivenih vrijednosti jer, kao što je prije spomenuto, pokazatelji se mogu znatno razlikovati za različite vrste pitanja. Tablica također uključuje srednju vrijednost bodova ostvarenih na pojedinom zadatku. Sadrži i indeks lakoće, čiji je optimalan raspon između 30% i 80%, zbog čega su ćelije s vrijednostima izvan tog intervala kodirane bojama. Zeleno obojene ćelije označavaju prevelik indeks, a crveno obojene premalen. Sljedeći stupci tablice definiraju diskriminacijski indeks i zatim diskriminativnu učinkovitost zadataka. Vrijednosti za diskriminativnu učinkovitost mogu varirati između iznimno velikih negativnih i pozitivnih brojeva, a, uzimajući u obzir parametre tablice o kojima on ovisi, vidljiv je isti uzorak za većinu zadataka; male vrijednosti za visok indeks lakoće i nisku diskriminaciju, a visok indeks za nižu lakoću i veću diskriminaciju. Budući da nezadovoljavajuća učinkovitost poprima vrijednosti ispod 15%, za takav zadatak cijeli je redak obojen crvenom bojom kako bi se naglasile ekstremne vrijednosti. Posljednji statistički pokazatelj je vrijednost slučajnog pogotka. Te

vrijednosti iznose 0% za zadatke slobodnog unosa teksta jer na njima nema mogućnosti nasumičnog pogađanja odgovora. Na slici ispod prikazan je isječak iz tablice generiran za zadatke iz jednog međuispita iz predmeta Napredni razvoj programske potpore za web 2022/2023. Radi sveobuhvatnosti prikaza, odabran je zanimljiv skup podataka iz sredine tablice na kojemu se mogu vidjeti vrijednosti i boje koje su se pojavile zbog njih. Također, u tom dijelu prikazane su dvije vrste zadataka s različitim rezultatima.

br.	Tip pitanja	ID	Srednja vrijednost	Indeks lakoće	Diskr. indeks	Diskr. učinkovitost	Slučajan pogodak
33	Free text	45043	0.259	25.93%	58.54%	225.76%	0%
34	Free text	45044	0.648	57.41%	34.71%	60.46%	0%
35	Free text	45045	0.356	35.59%	61.25%	172.1%	0%
36	Free text	45046	0.143	14.29%	33.08%	231.49%	0%
37	Classic (ABC) question	45048	0.424	59.26%	27.57%	46.52%	16%
38	Classic (ABC) question	45049	0.371	48.39%	43.69%	90.29%	20%
39	Classic (ABC) question	45050	0.452	52.38%	5.55%	10.6%	20%
40	Classic (ABC) question	45051	0.103	24.14%	37.95%	157.21%	20%
41	Classic (ABC) question	45052	0.905	91.67%	27.24%	29.72%	20%

Sl. 5.3.5.1 Tablica statističkih pokazatelja za zadatke iz međuispita iz predmeta Napredni razvoj programske potpore za web 2022/2023

br.	Tip pitanja	ID	Srednja vrijednost	Indeks lakoće	Diskr. indeks	Diskr. učinkovitost	Slučajan pogodak
1	Free text	46399	7.019	0.64%	82.13%	12832.81%	
2	Free text	46400	1.177	4.71%	50.97%	1082.17%	
3	Free text	46401	2.573	19.49%	57.5%	295.02%	
4	Free text	46402	1.592	21.63%	32%	147.94%	

Sl. 5.3.5.2 Tablica statističkih pokazatelja za zadatke iz završnog ispita iz predmeta Razvoj programske potpore za web 2022/2023

Dobiveni rezultati mogu se interpretirati u skladu s definiranim očekivanim vrijednostima pokazatelja za različite vrste zadataka. Vidljivo je da kombinacija niskog indeksa lakoće i visoke diskriminacije u većini primjera rezultira visokom diskriminacijskom učinkovitosti i obrnuto. Također, gledajući diskriminativnu učinkovitost, može se zamijetiti da pitanja slobodnog unosa imaju izrazito veliku učinkovitost u odnosu na pitanja višestrukog izbora i manji indeks lakoće, na što definitivno utječe nemogućnost nasumičnog pogađanja točnog odgovora. Čak i pri usporedbi samo unutar grupe pitanja višestrukog izbora, uglavnom pitanja s većom vrijednosti slučajnog pogotka imaju manji indeks lakoće i veću diskriminaciju.

Na kraju svakog izvještaja, prikazana je tablica sa srednjom vrijednosti pokazatelja za sve zadatke. Ako je prosječan indeks lakoće manji od 30%, ćelija je obojena crvenom bojom, a, ako je veći od 80%, obojena je zelenom. Također, za prosječnu diskriminativnu učinkovitost, ćelije s vrijednostima ispod 15% prikazuju se crvenom bojom. Većinom ove srednje vrijednosti spadaju u zadovoljavajući interval, ali postoji situacija kada su ekstremne, kao npr. u slučaju na slici ispod. Za ovaj ispit može se zaključiti da su svi zadaci u prosjeku bili iznimno teški.

ID	Prosjek indeksa lakoće	Prosjek diskriminacijskog indeksa	Prosjek diskriminativne učinkovitosti	Prosjek slučajnog pogotka
13,846	11.62%	55.65%	3589.48%	

Sl. 5.3.5.3 Tablica srednjih vrijednosti statističkih pokazatelja za Web1: završni ispit 2021/2022

## Zaključak

Zahvaljujući detaljnim podacima o rezultatima ispita spremljenih u bazu podataka sustava Edgar, moguće je statističkim metodama doći do različitih pokazatelja koji će pospješiti sastavljanje kvalitetnih ispita adekvatnih za nastavno gradivo i mogućnosti studenta. U kontekstu prilagođenosti ispita temama u sklopu predmeta, potrebno je provjeriti kako su riješeni zadaci koji pokrivaju te teme. Primjerice, ako imaju vrlo nizak indeks lakoće, potrebno je utvrditi razlog loše riješenosti, što može biti loše razumijevanje gradiva, neadekvatni materijali za učenje i nekompetentnost nastave. U sklopu nastave, trebao bi biti cilj omogućiti svim studentima pravedne temelje na kojima mogu izgraditi svoje znanje i sposobnosti. Rješavanje problema koji mogu nastati može se provesti na probnom uzorku na kojemu se prema rezultatima definira razina težine pojedinih vrsta pitanja koji se dalje ponovno formuliraju da bi ta razina ostvarila optimalne vrijednosti. Također, bitno je skrenuti pozornost na snagu diskriminiranja za različita pitanja. Kao što je prije viđeno, ta mjera uglavnom postiže zadovoljavajuće vrijednosti, pogotovo za pitanja slobodnog unosa teksta. Ipak, postoji nekoliko primjera zadataka koji ne ispunjavaju određene standarde. Problem pitanja višestrukog izbora sveprisutna je mogućnost za dobro rješavanje ispita bez ikakvog potrebnog znanja. Niski indeksi diskriminacije za takva pitanja mogu nalagati da je to pitanje bilo vrlo očito i lako za pogoditi, ili da se ponovilo s ispita iz prethodnih godina pa je unaprijed poznato. Dobra praksa kombiniranja prednosti ovih dviju vrsta pitanja je višestruki izbor zbog jednostavnosti ocjenjivanja i mogućnosti računanja Cronbach alfe, ali s jako detaljnim, dugim odgovorima koji iziskuju dobro razumijevanje i razlikovanje točnoga i netočnoga. Za budući rad vezano za temu predloženo je odrediti razine znanja i sposobnosti za grupe ispitanika na kojima se mogu izračunati parametri za modele Teorije odgovora na zadatke radi još detaljnijeg uvida u statistiku ispita. Prijedlog je, također, integrirati ovaj izvještaj unutar same aplikacije Edgar da bi bio dostupan svima koji žele, uz osobne rezultate, vidjeti sveobuhvatne rezultate za pisani ispit.

## Literatura

- [1] DiBattista, D., Kurzawa, L., *Examination of the Quality of Multiple-choice Items on Classroom Tests*. The Canadian Journal for the Scholarship of Teaching and Learning: Vol. 2, 2,4 (2011)
- [2] Tjabolo, S.A., Otaya, L.G., *Quality of School Exam Tests Based on Item Response Theory*, Universal Journal of Educational Research, 7,10 (2019)
- [3] Tavakol, M., Dennick, R., *Making sense of Cronbach's alpha*, *International Journal of Medical Education*, 2 (2011), str. 53-55.
- [4] RStudio. Wikipedija: Slobodna enciklopedija. Poveznica: [RStudio - Wikipedia](#); pristupljeno 3. lipnja 2023.
- [5] PgAdmin Documentation. Poveznica: [pgAdmin 4 — pgAdmin 4 6.21 documentation](#), pristupljeno: 3. lipnja 2023.
- [6] Columbia University Irving Medical Center, Item Response Theory. Poveznica: [Item Response Theory | Columbia University Mailman School of Public Health](#), pristupljeno: 8. lipnja 2023.
- [7] Statistics by Jim, Cronbach's Alpha: Definition, Calculations & Example. Poveznica: [Cronbach's Alpha: Definition, Calculations & Example - Statistics By Jim](#), pristupljeno: 8. lipnja 2023.
- [8] Moodle Quiz statistics report. Poveznica: [Quiz statistics report - MoodleDocs](#), pristupljeno: 10. lipnja 2023.
- [9] Moodle Quiz statistics report. Quiz information. Poveznica: [Quiz statistics report - MoodleDocs](#), pristupljeno: 10. lipnja 2023.
- [10] Moodle Quiz statistics report.Quiz structure analysis. Poveznica: [Quiz statistics report - MoodleDocs](#), pristupljeno: 10. lipnja 2023.