



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA
MATEMATIKU I INFORMATIKU



Marija Pešić

Upotreba logističke regresije u modeliranju verovatnoće bankrota preduzeća

-MASTER RAD-

Mentor:

Prof.dr Zagorka Lozanov-Crvenković

Novi Sad, 2016.

Sadržaj

Uvod	4
1 Predviđanje bankrota	6
1.1 Definicija bankrota	6
1.2 Istorijski razvoj modela za predviđanje bankrota	7
1.2.1 Almtanov Z-skor model	9
1.2.2 Ohlsonov O-skor model	10
1.3 Izbor varijabli za model predviđanja bankrota srpskih preduzeća	11
1.3.1 Empirijski pristup	11
1.3.2 Teorijski pristup	13
1.4 Izbor statističkog modela za predviđanje bankrota srpskih preduzeća	15
2 Logistička regresija	17
2.1 Uvod	17
2.2 Populacioni model	19
2.2.1 Osnovni populacioni model	19
2.2.2 Modifikacija osnovnog populacionog modela	21
2.3 Sigmoid funkcija	23
2.4 Razlike između linearne i logističke regresije	23
2.5 Logistički regresioni model i logit transformacija	25
2.6 Postupak modeliranja logističkom regresijom	26
2.7 Logistička regresija “korak po korak”	28
2.8 Ocenjivanje značajnosti koeficijenata	30
2.8.1 Wald test	32
2.8.2 Score test	33
2.9 Slaganje logističkog regresionog modela sa podacima	34
2.9.1 Metod maksimalne verodostojnosti	35
2.10 Procena slaganja modela sa podacima (goodness-of-fit)	37
2.10.1 Pirsonova hi-kvadrat statistika i odstupanje	39
2.10.2 Hosmer-Lemeshow test	40
2.10.3 Tabele klasifikacije	41

2.11	Interpretacija logističkog regresionog modela	41
2.11.1	Interpretacija logističkog modela za dihotomnu nezavisnu promenljivu	42
2.11.2	Interpretacija logističkog modela za neprekidnu nezavisnu promenljivu	44
3	Primena logističke regresije u modeliranju bankrota	46
3.1	Veličina uzorka	46
3.2	Izbor početnih varijabli za model	47
3.3	Ispravnost podataka u uzorku	48
3.3.1	Provera ispravnosti podataka kategorijalnih promenljivih	48
3.3.2	Provera ispravnosti podataka neprekidnih promenljivih	49
3.4	Multikolinearnost varijabli	55
3.5	Postupak logističke regresije u SPSS-u	56
3.6	Rezultati	58
3.7	Konačan model	67
4	Zaključak	68
	Literatura	69

Uvod

Predviđanje bankrota je jedna od glavnih oblasti interesovanja u sferi finansijske analize i predstavlja deo analize procene kreditnog rizika koja se sve više razvija, kako u svetu, tako i u Srbiji. Preduzeća raznih delatnosti procenjuju kreditne sposobnosti svojih trgovinskih partnera, banke i druge finansijske institucije procenjuju kreditni rizik kako klijenata, tako i svog portfolija kako bi zadovoljili pravila i regulative upravljanja rizicima, a svaka kompanija mora da kontroliše novčane tokove tako da može da ispuni svoje obaveze, što je opet povezano sa verovatnoćom bankrota.

Najgori scenario koji se može dogoditi jednoj firmi kao prodavcu robe, jeste da njen dužnik ode u stečaj (da bankrotira). Samim tim, ukoliko je verovatnoća da kupac ode u stečaj u skorijoj budućnosti (narednih godinu dana) velika, ne bi uopšte trebalo dalje analizirati to preduzeće, već bi se umesto određivanja limita za saradnju na odloženo plaćanje, insistiralo isključivo na avansnim uplatama.

U skladu sa tim, ovaj rad će se baviti verovatnoćom bankrota kao dela procesa procene kreditnog rizika preduzeća. Glavni fokus će biti na predviđanju bankrota srpskih privatnih preduzeća.

Ima više motivacija za izbor ovakve teme rada. Glavna motivacija potiče iz toga što je ova tematika, iako veoma popularna u svetu, prilično neistražena na teritoriji Srbije. Drugi izvor motivacije proističe iz praktične primene procene kreditnog rizika, budući da je verovatnoća bankrota potencijalnih partnera jedan od ključnih parametara za određivanje kreditne ocene u poslovanju. I treći izvor motivacije je dosadašnje iskustvo u vođenju stečajnih postupaka u Srbiji, jer rezultati ukazuju da stečajni postupci u Srbiji u proseku traju 2,7 godina, dok je u Nemačkoj taj rok 1,2 godine, u zemljama OECD-a ¹ 1,7, u Hrvatskoj 3,1, u Bugarskoj i Rumuniji 3,3 godine. Što se tiče stepena namirenja poverilaca, analize i iskustva ukazuju da je namirenje u Srbiji 25,4% od ukupnih potraživanja, u zemljama OECD-a poverioci naplate 68,6% duga, u Nemačkoj 52,5%, u Hrvatskoj 30,5%, u Bugarskoj 32,1%, u Rumuniji 29,5% (vidi u [7]).

¹Organizacija za ekonomsku saradnju i razvoj (34 članice)

Zbog navedenih motivacija, formiranje modela koji predviđa bankrot preduzeća u Srbiji bi predstavljalo važan doprinos za srpsko poslovanje.

Istraživanje je sprovedeno koristeći statističku metodu - **logističku regresiju**, na osnovu finansijskih podataka iz završnih finansijskih računa srpskih preduzeća koja su otišla u stečaj, kao i njima sličnih preduzeća koja su “zdrava”, tj. nisu otišla u stečaj.

Cilj ovog rada je da odgovori na sledeća pitanja:

- Koje finansijske pokazatelje treba koristiti prilikom modeliranja verovatnoće bankrota srpskih preduzeća?
- Da li je logistička regresija odgovarajući model za predviđanje bankrota u Srbiji?
- Da li je dobijeni model zadovoljavajući?

Kako istraživanje ne bi bilo prevelikog obima, napravljeno je ograničenje u izboru analiziranih preduzeća, u smislu da se koncentriše isključivo na srpska privatna preduzeća. U tom kontekstu, privatno preduzeće je društvo sa ograničenom odgovornošću definisano Zakonom o privrednim društvima (vidi u [2]). Ova odluka je doneta na osnovu činjenice da se rizik od bankrota značajno razlikuje u javnom i privatnom sektoru, i mešanjem ovih sektora mogu doći do nevalidnog modela.

Rad je podeljen u četiri celine. Prvi deo rada odnosi se na teorijsku pozadinu pojma bankrota i izbor preliminarnih varijabli sa teorijskog aspekta za model predviđanja bankrota srpskih preduzeća. Drugi deo rada predstavlja teorijski osnov logističke regresije koja je izabrana kao naj-reprezentativniji statistički metod za modeliranje bankrota srpskih preduzeća. Treći deo rada se bavi samom primenom logističke regresije prilikom modeliranja bankrota srpskih preduzeća. I na kraju, u četvrtom delu rada, izložiće se rezultati ovakve primene logističke regresije, i doći će se do odgovora na pitanje da li je dobijeni model zadovoljavajući.

Veliku zahvalnost dugujem svojoj mentorki Zagorki Lozanov-Crvenković na pruženim savetima i smernicama tokom pisanja ovog rada. Takođe, zahvalila bih se članovima komisije Ljiljani Gajić i Ivani Štajner-Papuga na svom prenetom znanju i lepoj saradnji.

I na kraju, najveću zahvalnost dugujem svojoj porodici, a naročito mami, na nesebičnoj podršci tokom čitavog školovanja.

Marija Pešić

Glava 1

Predviđanje bankrota

1.1 Definicija bankrota

U Srbiji je bankrot regulisan Zakonom o stečajnom postupku.

Stečaj je zakonom utvrđen postupak koji se sprovodi nad imovinom dužnika radi namirenja poverilaca i onemogućavanja daljeg lošeg poslovanja privrednog subjekta (vidi u [1]). Ovaj postupak se sprovodi isključivo preko suda.

Cilj stečaja jeste najpovoljnije kolektivno namirenje stečajnih poverilaca ostvarivanjem najveće moguće vrednosti stečajnog dužnika, odnosno njegove imovine (vidi u [1]).

Razlozi otvaranja stečaja su:

- **Trajna nesposobnost plaćanja:** blokada računa preko 30 dana ili neplaćanje dospele obaveze 45 dana
- **Preteća nesposobnost plaćanja:** čini se verovatnim da stečajni dužnik neće izmirivati postojeće obaveze u roku
- **Prezaduženost:** imovina je manja od obaveza
- **Nepostupanje po usvojenom planu reorganizacije ili ako je plan reorganizacije izdejstvovan na prevaran ili nezakonit način.**

Iako u nekim istraživanjima koji se bave verovatnoćom bankrota, ove vrste stečaja čine posebne grupe ispitivanja, u ovom radu se one neće odvajati. Takva odluka je doneta na osnovu činjenice da iz ugla upravljanja kreditnim rizikom nema razlike koji je razlog otvaranja stečaja, dokle god je do njega došlo. Dakle, posmatraće se ona preduzeća za koja je otvoren stečajni postupak, bez obzira na razlog otvaranja istog.

Stečaj se, u smislu Zakona o stečaju, sprovodi **bankrotstvom** ili **reorganizacijom**.

- Pod bankrotstvom se podrazumeva namirenje poverilaca prodajom celokupne imovine stečajnog dužnika, odnosno stečajnog dužnika kao pravnog lica. Poverioci će izglasati bankrotstvo kada procene da firma nema mogućnosti da se izvuče reorganizacijom.
- Pod reorganizacijom se podrazumeva namirenje poverilaca prema usvojenom planu reorganizacije i to redefinisanjem dužničko-poverilačkih odnosa, statusnim promenama dužnika ili na drugi način koji je predviđen planom reorganizacije. Reorganizacija je drugim rečima mera za sprečavanje potpunog bankrota preduzeća.

Napomena: Stečaj i likvidacija označavaju različite postupke. Likvidacija je utvrđena Zakonom o privrednim društvima, i za razliku od stečaja, kada dužnik nema uslove da namiri poverioce, postupak likvidacije se pokreće onda kada dužnik ima dovoljno sredstava da izmiri svoje obaveze, nakon čega se likvidira. Pokretanje likvidacije ne sprečava određivanje i sprovođenje izvršenja, niti druge postupke koji se vode protiv društva, te je ovaj postupak bolji za poverioce.

Kako je ideja ovog rada da se analizira najgori scenario kojim može krenuti saradnja sa nekim preduzećem, analiziraće se preduzeća koja su ušla u stečajni postupak u smislu izglasanog *bankrotstva*. Radi jednostavnosti, u daljem radu, pojam bankrota i otvaranja stečajnog postupka će označavati istu stvar.

Strogo govoreći, bankrot je posledica procesa koji počinje finansijski i završava se zakonski, tako da je stvarnu tačku nastanka bankrota teško odrediti (vidi u [10]). U ovom radu, tačka nastanka bankrota podrazumevaće momenat kada je otvoren stečajni postupak i kada je izglasan bankrot preduzeća.

Otvaranjem stečajnog postupka, gase se postojeći računi dužnika kod banaka, a stečajni upravnik otvara novi račun pravnog lica koje u imenu sadrži reči “U STEČAJU”. Menice koje su date poveriocima pre stečaja, ne mogu se koristiti. Svi sudski i vansudski postupci za naplatu potraživanja se obustavljaju.

1.2 Istorijski razvoj modela za predviđanje bankrota

Najranija literatura o predviđanju bankrota datira još iz 1930-ih. Nakon Velike ekonomske krize, došlo je do praktične potrebe za predviđanjem bankrota, i od tog momenta, bankrot predstavlja značajnu temu za istraživanje naučnika.

Jedna od prvih publikacija o analizi racio pokazatelja kompanija koja su otišle u stečaj jeste bilten iz 1930. godine koji je objavio Bureau of Business Research. Oni su analizirali dvadeset i devet preduzeća koja su otišla u stečaj i njihova dvadeset i četiri različita racio pokazatelja. Utvrđeno je da osam pokazatelja imaju sličnosti kod preduzeća koja su otišla u stečaj i da pokazuju dobre indikatore za rast procenu zdravlja kompanije.

1932. godine, Paul J. Fitzpatrick objavljuje istraživanje u kojem se po prvi put porede finansijski pokazatelji preduzeća koja su otišla u stečaj sa finansijskim pokazateljima zdravih preduzeća. Njegova baza podataka obuhvata trinaest preduzeća koja su otišla u stečaj i trinaest onih koja su zdrava. Najznačajniji rezultat ovog istraživanja, osim što je ustanovljeno da kompanije koje su otišle u stečaj imaju pokazatelje ispod proseka, a zdrave kompanije natprosečne pokazatelje, se smatra to što su dva pokazatelja sa najvećim uticajem na verovatnoću bankrota: neto vrednost duga i neto vrednost dobitka, a pokazatelj tekuće likvidnosti¹ i pokazatelj ubrzane likvidnosti² nisu imali veliki uticaj.

Od 1930-ih do sredine 1960-ih godina, modeli predviđanja bankrota se uglavnom koncentrišu na pronalaženje razlika između zdravih i nezdravih preduzeća. Svoj doprinos istraživanju imaju i Winakor (1935), Merwin (1942) i Chudson (1945).

Međutim, razvoj modela za predviđanje bankrota zaista počinje člankom iz 1966. godine čiji je autor William Beaver. Polazna tačka njegovog istraživanja je bila činjenica da, iako su ranije studije poredile finansijske i racio pokazatelje zdravih i nezdravih preduzeća, nije postojao koristan model za poslovanje, odnosno nisu postojali empirijski dokazi za korisnosti korišćenih racio pokazatelja. Njegova baza podataka se sastojala od sedamdeset i devet zdravih i sedamdeset i devet nezdravih preduzeća u periodu od 1954. do 1964. godine. Kriterijum za izbor nezdravih preduzeća je bio taj da postoji bar jedan finansijski izveštaj godinu dana pre nego što je preduzeće otišlo u stečaj. Na tom kriterijumu će se bazirati i ovaj rad.

Analiza koju je Beaver koristio bila je prilično jednostavna. Izračunavši srednju vrednost svakog korišćenog finansijskog pokazatelja, poredio je sa njima pokazatelje pojedinačnih preduzeća. Rezultati ovakvog istraživanja su bili u skladu sa prethodnom literaturom.

Ograničenja ovakve analize su bila problematična za tumačenje rezultata. Naime, poređenje pokazatelja sa srednjom vrednošću može da nam pokaže razlike između racio pokazatelja zdravih i nezdravih preduzeća, ali na osnovu toga se ne može reći da li ovi pokazatelji imaju prediktivnu moć. Za tako nešto je potrebna informacija o veličini te razlike, a pogotovo o distribuciji.

¹Pokazatelj tekuće likvidnosti služi za merenje sposobnosti kompanije da servisira kratkoročne obaveze.

Formula: Obrtna imovina / Kratkoročne obaveze

²Pokazatelj ubrzane likvidnosti meri sposobnost kompanije da izmiri svoje kratkoročne obaveze onim delom svoje obrtne imovine koji se relativno brzo može konvertovati u gotov novac, bez prodaje zaliha.

Formula: (Obrtna imovina - Zalihe) / Kratkoročne obaveze

Rezultati istraživanja su ukazivali na to da je najuticajnija promenljiva na verovatnoću bankrota odnos poslovnih prihoda i ukupnog duga. Upotrebom ove promenljive, 92% kompanija je bilo ispravno klasifikovano. Drugi najznačajniji prediktor je bila neto marža, sa 91% tačnosti ([13]).

1968. godine je objavljeno prvo istraživanje koje je bazirano na višestrukoj diskriminacionoj analizi. Istraživanje je objavio **Edward Altman**, a model je dobio naziv **Z-skor**. Ovo je i dan-danas najpoznatiji model za predviđanje bankrota.

1.2.1 Almtanov Z-skor model

Kao što je rečeno, pre Altmanovog modela, finansijska analiza se bazirala samo na jednoj promenljivoj, tj. bila je univarijantna. Koristeći višestruku diskriminacionu analizu, Altman je minimizirao greške klasifikacije preduzeća.

Baza podataka koju je Altman koristio je sadržala šezdeset i šest preduzeća: trideset i tri preduzeća koja su otišla u stečaj, i trideset i tri zdrava preduzeća. On je najpre izabrao preduzeća koja su otišla u stečaj između 1946. i 1965. godine, a potom je izabrao zdrava preduzeća slične veličine (po aktivu).

Zbog velikog broja varijabli koje su empirijski dokazane kao značajne u predviđanju bankrota, Altman je uključio dvadeset i dva različita ratio pokazatelja, koja je dalje klasifikovao u pet kategorija: likvidnosti, profitabilnosti, moći povratka investicije, solventnosti i aktivnosti. Na kraju, iz svake od ovih kategorija, on je izabrao po jednu varijablu na osnovu ukupne prediktorne moći, a ne na osnovu značaja pojedinačne varijable.

Konačan Z-skor model glasi ([4]):

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.99X_5$$

gde je:

X_1 - Obrtni kapital / Ukupna aktiva

X_2 - Neraspoređena dobit / Ukupna aktiva

X_3 - Poslovni rezultat / Ukupna aktiva

X_4 - Tržišna vrednost kapitala / Ukupne obaveze

X_5 - Prihod od prodaje / Ukupna aktiva

Granice za promenljivu Z su sledeće:

- **Zona bankrotstva** $Z < 1.81$
- **Siva zona** $1.81 \leq Z < 2.99$
- **Bezbedna zona** $Z \geq 2.99$

Ovako definisan model je imao tačnost predviđanja od 95%.

Nakon Z-skor modela, 1977. godine, Altman, Haldeman i Harayanan su kreirali ZETA model koji je imao za cilj da unapredi Z-skor model. U poređenju sa Z-skor modelom, ZETA model je pokazivao preciznije ocene neuspešnih preduzeća od 2 do 5 godina pre bankrotstva, dok je tačnost za prvu godinu gotovo jednaka za oba modela. Uprkos povećanoj tačnosti, Z-skor model je zadržao status najpoznatijeg modela za predviđanje bankrota.

1.2.2 Ohlsonov O-skor model

1980. godine, **James Ohlson** uvodi model koji se bazira na logističkoj regresiji i koji mu dozvoljava upotrebu veće baze podataka. O logističkoj regresiji će biti više reči u narednom poglavlju.

U originalnom Ohlsonovom modelu bilo je uključeno devet varijabli ([5]):

$$O = -1.3 - 0.4X_1 + 6.0X_2 - 1.4X_3 + 0.8X_4 - 2.4X_5 - 1.8X_6 + 0.3X_7 - 1.7X_8 - 0.5X_9$$

gde je:

X_1 - Ukupna aktiva

X_2 - Ukupne obaveze / Ukupna aktiva

X_3 - Obrtni kapital / Ukupna aktiva

X_4 - Kratkoročne obaveze / Obrtna imovina

X_5 - Neto prihod / Ukupna aktiva

X_6 - Sredstva iz poslovanja / Ukupne obaveze

X_7 - 1 ako postoji neto gubitak u prethodne 2 godine, 0 inače

X_8 - 1 ako su ukupne obaveze veće od ukupne aktive, 0 inače

X_9 - (Trenutni neto prihod - Neto prihod u prethodnom periodu) / (|Trenutni neto prihod| + (|Neto prihod u prethodnom periodu|))

Ove varijable predstavljaju dva različita faktora: **finansijsko stanje** (X_1 - X_4) i **finansijsko poslovanje** (X_5 - X_9) kompanije.

Za pet varijabli je utvrđeno da su statistički značajne za predviđanje bankrota. Te promenljive su: X_1 , X_2 , X_3 , X_5 i X_6 .

Ovaj model je ocenjivao verovatnoću bankrota sa uspehom od 96,12%.

Tačnost predviđanja Altmanovog Z-skor i Ohlsonovog O-skor modela se tokom godina ispostavila prilično dobrom. Uprkos tome, naučnici konstantno pokušavaju da povećaju ovu moć predviđanja, te su primenjivane i neke nove tehnike: analiza preživljavanja, veštačka inteligencija, genetski algoritmi, neuronske mreže itd. Međutim, nijedan od njih nije značajno povećao tačnost predviđanja, već su samo podigli kompleksnost istraživanja (vidi u [3]). To je verovatno glavni razlog zašto su i dalje dve najzastupljenije metode za predviđanje bankrota višestruka diskriminaciona analiza i logistička regresija.

1.3 Izbor varijabli za model predviđanja bankrota srpskih preduzeća

Koje varijable, i koliko njih treba uključiti u model za predviđanje bankrota kako bi se postigao najveći stepen tačnosti, predstavlja diskutabilnu temu već duži vremenski period.

U ranijim studijama o predviđanju bankrota, nije uvek postojala pozadinska teorija koja bi objasnila koje finansijske pokazatelje i koji broj racio pokazatelja treba koristiti kako bi se bankrot najbolje predvideo ([3]). Metod izbora varijabli se oslanjao isključivo na empirijsko testiranje, te su rezultati u velikoj meri zavisili od korišćenih podataka, i samim tim nisu bili najtačniji.

U ovom radu, međutim, kako bi se došlo do što verodostojnijeg izbora varijabli za analizu verovatnoće bankrota, uzet je u obzir i empirijski i teorijski pristup.

1.3.1 Empirijski pristup

Varijable koje su izabrane empirijskim putem odnose se na najčešće korišćene varijable u dosadašnjim studijama na temu predviđanja bankrota.

Tokom godina, u mnogobrojne studije na ovu temu je uključeno na stotine različitih varijabli preuzetih iz finansijskih izveštaja i podataka sa tržišta.

Samo su 42 varijable korišćene u pet ili više radova, a 674 varijable su korišćene u jednom do dva rada na ovu temu (vidi u [6]). Dve najčešće korišćene varijable su odnos neto prihoda i ukupne aktive (poznatije kao **neto prinos na aktivu**) i odnos obrtnih sredstava i tekućih obaveza (poznatije kao **pokazatelj tekuće likvidnosti**).

- **Neto prinos na aktivu** prikazuje poslovnu efikasnost. Ovaj racio pokazatelj pokazuje koliko je jedinica neto dobitka ostvareno po jedinici angažovanih poslovnih sredstava.

- **Pokazatelj tekuće likvidnosti** meri sposobnost kompanije da servisira kratkoročne obaveze, odnosno sposobnost kompanije da ispuní obaveze kraće od jedne godine sa sredstvima koja dospevaju u jednogodišnjem periodu. Drugim rečima, ovaj ratio pokazatelj meri kratkoročno finansijsko stanje kompanije.

Ostale varijable koje su često korišćene u dosadašnjoj literaturi su: odnos obrtnog kapitala³ i ukupne aktive (**koeficijent obrta kapitala**), **odnos neraspoređenog dobitka i ukupne aktive**, **odnos poslovnih prihoda i ukupnog duga**⁴, odnos neto rezultata i poslovnih prihoda (**neto marža**), odnos ukupnog duga i kapitala (**ratio duga i kapitala**), odnos ukupnih obaveza i ukupne aktive (**koeficijent zaduženosti**) i odnos poslovnih prihoda i ukupne aktive (**koeficijent obrta ukupne imovine**) ([6]).

- **Koeficijent obrta kapitala** pokazuje koliko puta se kapital kompanije obrne u toku jedne godine, odnosno koliko uspešno kompanija koristi kapital vlasnika sa ciljem generisanja prihoda.

Kako se obrtni kapital definiše kao razlika između obrtnih sredstava i kratkoročnih finansijskih obaveza preduzeća, očigledno je da kompanija mora da balansira između efikasnosti i kratkoročnog finansijskog zdravlja. To znači da što je ovaj koeficijent veći, veća je i mogućnost kompanije da ispuní kratkoročne obaveze, ali što je više kapitala vezano za poslovanje, preduzeće posluje manje efikasno.

- **Odnos neraspoređenog dobitka i ukupne aktive** je ratio pokazatelj koji sadrži informacije o zrelosti kompanije, politike plaćanja dividendi kao i kumulativne profitabilnosti tokom vremena.

Preduzeća koja nemaju neraspoređeni dobitak, ili ga imaju u manjem iznosu se posmatraju kao mlada preduzeća koja tek treba da razviju svoj biznis. Iz tog razloga, interno generisani novčani tokovi nisu dovoljni da pokriju sva investiranja, te se ovakva preduzeća u velikoj meri oslanjaju na finansiranje iz tuđih izvora.

Sa druge strane, preduzeća koja imaju veće iznose neraspoređenog dobitka, smatraju se zrelim kompanijama koje mogu da pokriju potencijalne investicije i ne moraju da se oslanjaju na tuđe izvore.

Ovaj ratio pokazatelj se može posmatrati i kao sposobnost kompanije da plati dividende (vidi u [8]).

- **Odnos poslovnih prihoda i ukupnog duga** je uključen u listu promenljivih zbog empirijski dokazane moći predviđanja ([6]), a pokazuje koliki deo duga je preduzeće u stanju da pokrije sa trenutnim profitom.

³Pojam obrtni kapital se koristi za označavanje onog dela obrtnih sredstava koji je dugoročno finansiran. Predstavlja razliku između obrtnih sredstava (obrtne aktive) i kratkoročnih finansijskih obaveza preduzeća.

⁴Ukupan dug predstavlja zbir kratkoročnih i dugoročnih obaveza

- **Neto marža** pokazuje koliku procentualnu dobit ostvaruje kompanija u odnosu na svoje poslovne prihode. Neto marža prikazuje koliki deo poslovnih prihoda ostvarenog u poslovnoj godini kompanija zadržava u obliku neto dobiti, koju može koristiti za reinvestiranje ili za raspodelu vlasnicima.
- **Racio duga i kapitala** pokazuje izloženost preduzeća finansijskom riziku. Ovaj racio pokazuje do koje mere se preduzeće finansira iz sopstvenih izvora. Ukoliko je ovaj racio pokazatelj manji, veća je finansijska fleksibilnost preduzeća.
Prema Modigliani-Millerovoj teoremi iz 1958. godine⁵, vrednost kompanije ne zavisi od svog kapitala (vidi u [9]), ali pri donošenju kreditnih odluka, ovo jeste važno jer je kapital kompanije primarni izvor koji se koristi za finansiranje gubitaka.
Važno je napomenuti da sa stanovišta menadžmenta preduzeća nije dobro ni kada je racio duga i kapitala previše mali, jer u slučaju kada preduzeće obrće veći profit od kamate za dug, povećava se i iznos koji se vraća akcionarima. Dok sa stanovišta procene kreditnog rizika, ovaj racio ipak treba da bude što manji.
- **Koeficijent zaduženosti** takođe pokazuje izloženost preduzeća finansijskom riziku, ali za razliku od racia duga i kapitala, pokazuje do koje mere se preduzeće finansira iz tuđih izvora. Ukoliko je ovaj racio pokazatelj manji, veća je finansijska fleksibilnost preduzeća.
- **Koeficijent obrta ukupne imovine** pokazuje koliko se puta ukupna imovina kompanije obrne u toku jedne godine, odnosno koliko uspešno kompanija koristi svoju imovinu sa ciljem generisanja prihoda.

Na osnovu svega navedenog, može se primetiti da je finansijski podatak koji se najčešće koristi pri analizi verovatnoće bankrota **ukupna aktiva**.

Broj varijabli u modelima varira isto koliko i same varijable, a prosečan broj varijabli koje su se koristile u dosadašnjim modelima je devet.

1.3.2 Teorijski pristup

Kako bi se prevazišli nedostaci objašnjenja zašto neke promenljive predviđaju bankrot bolje od drugih, tokom godina su razvijeni razni teorijski okviri koji na logičan način objašnjavaju uspeh empirijskih modela za predviđanje bankrota.

⁵M-M Teorema: Posmatramo dva preduzeća koja su identična po svemu osim po svojoj finansijskoj strukturi. Neka se prvo preduzeće finansira isključivo iz sopstvenog kapitala, dok se drugo preduzeće finansira delimično iz sopstvenog kapitala, a delimično iz tuđih izvora, tj. duga. Tada je vrednost ova dva preduzeća ista.

Najjednostavniji teorijski okvir se odnosi na sagledavanje finansijskih pokazatelja kao indikatora zdravlja kompanije. **Likvidnost, profitabilnost i stabilnost** su tri glavne kategorije ovih pokazatelja i treba ih tumačiti u smislu: što je veća vrednost pokazatelja, kompanija je zdravija, dok manja vrednost pokazatelja ukazuje na nezdravo preduzeće sa većim rizikom od bankrota (vidi u [11]).

Na analizu finansijskih izveštaja može da se gleda kao na merenje pomenutih pokazatelja: likvidnosti, profitabilnosti, stabilnosti. Analiza finansijskih izveštaja nastoji da utvrdi prednosti i mane preduzeća, koje se dalje koriste u procesu donošenja odluka i procenu kreditnog rizika.

Pored ovih pokazatelja, postoji još jedan pokazatelj koji spada u najvažnije pokazatelje sa staništa kreditnog rizika: **solventost**. Solventnost je sposobnost preduzeća da u svakom trenutku svojim novčanim sredstvima podmiri svoje dospele novčane obaveze. Drugim rečima, solventnost je platežna sposobnost preduzeća.⁶ Kako je solventnost od izuzetnog značaja za kreditnu analizu, ona utiče u velikoj meri i na verovatnoću bankrota (vidi u [12]).

Solventnost zavisi od odnosa između kupovine i prodaje, tekućih obaveza i potraživanja, kao i iznosa gotovine i gotovinskih ekvivalenata.

Na osnovu toga, postoje tri važna racio pokazatelja koja se takođe uključuju u model, a koja utiču na modele predviđanja bankrota na sledeći način ([13]):

1. **Velicina obrtne imovine** - Što je veći iznos obrtne imovine, manja je verovatnoća bankrota.
2. **Neto priliv gotovine iz poslovnih aktivnosti**⁷ - Što je veći iznos priliva gotovine iz poslovnih aktivnosti, manja je verovatnoća bankrota preduzeća.
3. **Rashodi od kamata** - Što je manji iznos rashoda od kamata, manja je verovatnoća bankrota preduzeća.

U Tabeli 1 su predstavljene sve odabrane varijable za početni model predviđanja bankrota preduzeća kao i hipoteze o uticaju svake od njih na verovatnoću bankrota.

⁶Solventnost ne treba mešati sa pojmom likvidnosti, koji predstavlja sposobnost podmirivanja kratkoročnih dospelih obaveza.

⁷Poznatiji pod nazivom **potraživanja po osnovu prodaje**

Tabela 1

Varijabla	Uticaj na verovatnoću bankrota		Izbor varijable	
	↑ - povećava verovatnoću		Teorijski pristup (T)	
	↓ - smanjuje verovatnoću		Empirijski pristup (E)	
Neto prinos na aktivu	↓		E	
Pokazatelj tekuće likvidnosti	↓		E	
Koeficijent obrta kapitala	↓		E	
Neraspoređeni dobitak/ukupna aktiva	↓		E	
Poslovni prihodi/ukupni dug	↓		E	
Neto marža	↓		E	
Racio duga i kapitala	↑		E	
Koeficijent zaduženosti	↑		E	
Koeficijent obrta ukupne imovine	↓		E	
Logaritam vrednosti obrtne imovine	↓		T	
Priliv gotovine iz poslovnih aktivnosti	↓		T	
Logaritam vrednosti rashoda od kamata	↑		T	

1.4 Izbor statističkog modela za predviđanje bankrota srpskih preduzeća

Glavni cilj ovog rada jeste analiziranje međusobnih veza i uticaja finansijskih pokazatelja na verovatnoću da preduzeće ode u bankrot.

Linearna diskriminaciona analiza, kako u prostom, tako i u višestrukobnom obliku, bila je i ostala jedna od najčešće korišćenih statističkih tehnika za predviđanje ove verovatnoće.

Diskriminaciona analiza je zapravo klasifikovanje pojedinaca u grupe prema određenim kriterijumima. Osnovna ideja linearne diskriminacione analize jeste da se pronađe najbolja linearna kombinacija nezavisnih varijabli simultanom analizom međusobnog odnosa između velikog broja varijabli, kojom bi se jedinka klasifikovala u jednu od dve ili više alternativnih grupa. Kada se radi o predviđanju bankrota, grupe posmatranja su očigledno preduzeća koja su otišla u stečaj i ona koja nisu.

Najbolja linearna diskriminaciona funkcija je ona koja minimizira preklapanje vrednosti distribucije između grupa. Diskriminaciona analiza je adekvatna tehnika u slučajevima kada je zavisna varijabla kategorička (nominalna, deskriptivna), a nezavisne varijable numeričke.

Naravno, postoje i neke pretpostavke o prirodi varijabli korišćenih u diskriminacionoj funkciji koje moraju biti ispunjene. Jedna od njih je normalnost distribucije svake varijable. Druga pretpostavka podrazumeva da su kovarijansne matrice dve posmatrane grupe jednake. Ukoliko ova pretpostavka nije ispunjena, ne treba koristiti linearnu već kvadratnu formu.

Rezultat višestruke diskriminacione analize je vrednost koja se koristi za kategorizaciju zapažanja u više različitih grupa u zavisnosti od granične vrednosti. Međutim, ono što ne može da se dobije ovakvom analizom, jeste procena verovatnoće pripadnosti tim različitim grupama. Procena verovatnoće može da se izvede **korelacijom i regresijom**.

Korelaciona analiza je skup statističkih metoda kojima se istražuje jačina veze između posmatranih pojava.

Regresiona analiza se odnosi na skup statističkih metoda kojima se otkriva da li postoje veze između posmatranih varijabli, i od kolikog značaja su postojeće veze.

Cilj korelacione analize je da se ispita da li između varijacija posmatranih pojava postoji kvantitativno slaganje i, ako postoji, u kom stepenu, a da je pritom svejedno koja se od njih označava kao nezavisna, a koja kao zavisna promenljiva.

Kod regresione analize nužno je unapred identifikovati koja pojava će imati ulogu zavisne promenljive, a koja nezavisne promenljive. Cilj regresione analize je da se odredi onaj regresioni model koji najbolje opisuje vezu između pojava i da se na osnovu tog modela ocene i predvide vrednosti zavisne promenljive za odabrane vrednosti objašnjavajuće promenljive.

Regresiona analiza je jedan od najvažnijih i najčešće korišćenih statističkih metoda i ima vrlo široku primenu. Čitava jedna oblast ekonomske nauke, ekonometrija, velikim delom bavi se samo regresijom.

Prilikom istraživanja međusobnih veza dve promenljive primenjuju se metodi proste regresione i korelacione analize, a u slučaju posmatranja više promenljivih, metodi višestruke regresije i korelacije.

Iz svega navedenog, u ovom radu će se koristiti **logističku regresiju** kao statistički model za predviđanje bankrota.

Nakon Ohlsonovog modela za predviđanje bankrota koji je koristio logističku regresiju, ovaj metod je postao metod koji se najčešće koristi za ovakav tip istraživanja. Među njima su studije poput Aziz i dr. (1988), Theodossiou (1991), McGurr i DeVaney (1998), Kahya i Theodossiou (1999) itd.

Glava 2

Logistička regresija

2.1 Uvod

Regresiona analiza je jedan od najvažnijih i najčešće korišćenih statističkih metoda i ima veliku primenu u ekonomiji i ostalim društvenim naukama (kao u [26]).

Termin **regresija** znači vraćanje unazad, i prvi ju je upotrebio engleski naučnik Frensis Galton 1885. godine, prilikom istraživanja naslednih osobina. On je ispitivao vezu između visine roditelja i dece. Otkrio je da roditelji koji su znatno viši od proseka imaju decu nižu od njih, ali višu od proseka, i obrnuto. Visine su se “vraćale unazad”, odnosno, regresirale su prema proseku (vidi u [27]).

Danas se, međutim, reč regresija koristi u znatno širem značenju: da ukaže na statistički model kojim se izražava povezanost između zavisne varijable i odabranog skupa nezavisnih varijabli.

Regresija je sastavni deo svake analize podataka koja se bavi traženjem veze između zavisnih i nezavisnih promenljivih. Njen cilj je da se pronađe model koji je najbolje prilagođen podacima, koji sadrži samo one nezavisne promenljive koje imaju uticaj na ishod zavisne promenljive i koji dobro opisuje vezu između zavisne i nezavisnih promenljivih ([27]).

Najpoznatiji primer modeliranja jeste **linearni regresioni model**.

Linearni model, koji se koristi da se objasni slučajna promenljiva Y pomoću neslučajne promenljive x , obično se daje u obliku:

$$Y = \mu(x) + \varepsilon$$

Funkcija μ se naziva deterministički deo modela, a ε je slučajni deo modela i naziva se greška. Nezavisna promenljiva x je kontrolisana, vrednosti zavisne promenljive Y se mogu meriti, dok se vrednosti promenljive ε ne mogu meriti. Promenljiva ε sadrži u sebi sve ostale promenljive koje utiču na vrednost promenljive Y , ali nisu obuhvaćene posmatranjem ([14]).

Za linearni model, funkcija μ je linearna funkcija nepoznatih parametara. Najopštiji oblik ove funkcije zavisi od $p + 1$ nepoznatih parametara $\beta_0, \beta_1, \dots, \beta_p$ i može se zapisati kao ([14]):

$$\mu(x) = \beta_0 + \beta_1 q_1(x) + \beta_2 q_2(x) + \dots + \beta_p q_p(x)$$

gde su $q_1, q_2, \dots, q_p$ poznate funkcije.

Napomena: Linearnost se odnosi na način na koji se parametri i greške relacije pojavljuju u regresionoj jednačini, a ne na odnos među varijablama. Na primer, modeli:

$$y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 \cos x_i + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 \ln x_i + \varepsilon_i$$

su linearni jer su linearni u parametrima.

Klasične pretpostavke linearne regresije (vidi u [28]):

1. $E(\varepsilon_i) = 0$
2. $Var(\varepsilon_i) = \sigma^2$ (homoskedastičnost)
3. ε_i ima normalnu raspodelu
4. ε_i i ε_j su nezavisne za $i \neq j$
5. x_i su determinističke vrednosti i

$$\sum (x_i - \bar{x})^2 \neq 0,$$

$$\frac{\sum (x_i - \bar{x})^2}{n} < \infty, n \rightarrow \infty$$

Dakle, kod linearnog regresionog modela se za zavisnu promenljivu pretpostavlja da je neprekidna. Međutim, čest je slučaj da je rezultujuća promenljiva diskretna, obično binarna sa dve, a u ređim slučajevima i više mogućih vrednosti. Ovakva situacija se sreće i u ovom radu, gde zavisna promenljiva predstavlja prisustvo ili odsustvo značajne verovatnoće bankrota. Standardna metoda analize u ovakvoj situaciji je **logistički regresioni model**.

Razlika između linearne i logističke regresije se ogleda, kako u izboru parametara, tako i u pretpostavkama. Nezavisne varijable u logističkoj regresiji mogu biti kategorijalne, ili kombinacija kategorijalnih i neprekidnih, pri čemu ne postoje pretpostavke o raspodeli za ove varijable (vidi u [19]).

Klasične pretpostavke logističke regresije ([28]):

1. $E(\varepsilon) = n$, gde je n broj Bernulijevih eksperimenata¹
2. $Var(\varepsilon_i) = p$, gde je p verovatnoća “uspeha”
3. ε_i ima binomnu raspodelu
4. ε_i i ε_j su nezavisne za $i \neq j$
5. x_i su determinističke vrednosti i
$$\sum (x_i - \bar{x})^2 \neq 0,$$
$$\frac{\sum (x_i - \bar{x})^2}{n} < \infty, n \rightarrow \infty$$

Osim što se koristi za predviđanje zavisne promenljive na osnovu vrednosti nezavisnih promenljivih, logistička regresija se koristi i za rangiranje nezavisnih promenljivih po važnosti, kao i za procenu efekta njihove međusobne interakcije ([15]).

Zavisna promenljiva u ovom radu uzima samo dve vrednosti, tj. **dihotomna** je. Iz tog razloga se kodira sa 0 i 1, gde 0 označava “neuspeh” ili odsustvo neke karakteristike (preduzeće nije otišlo u stečaj), a 1 “uspeh” ili prisustvo neke karakteristike (preduzeće jeste otišlo u stečaj). Ako bi rezultirajuća promenljiva bila neprekidna a ne binarna, koristio bi se dijagram rasipanja² rezultata u odnosu na nezavisnu promenljivu.

2.2 Populacioni model

Engleski demograf i ekonomista, Tomas Maltus, 1789. godine objavio je rad “An Essay on the Principle of Population” u kojem je izložio svoj stav o problemu prenaseljenosti. On je objasnio da broj stanovnika naše planete raste brзом geometrijskom progresijom, dok količina proizvedenih resursa raste sporom aritmetičkom progresijom. Na taj način, u jednom momentu neće biti dovoljno resursa za postojeći broj stanovnika i moraće da dođe do nekog vida katastrofe: rat, siromaštvo, bolest (vidi u [16]). Ovo stanje je dobilo naziv demografska (Maltusova) katastrofa, a model koji ga prati **osnovni (Maltusov) populacioni model**.

2.2.1 Osnovni populacioni model

Maltusovo istraživanje se zasnivalo na pretpostavci da u nekom trenutku na Zemlji živi $p(0)$ ljudi, a da je u svakom narednom trenutku, populacija srazmerna populaciji u prethodnom tre-

¹Bernulijev eksperiment podrazumeva da postoje samo dva međusobno isključiva ishoda u svakom eksperimentu (preduzeće je otišlo u stečaj/preduzeće nije otišlo u stečaj), pri čemu je verovatnoća uspeha p ista pri svakom eksperimentu, i da su eksperimenti nezavisni.

²Dijagram rasipanja (Scatter dijagram) je grafički prikaz međuzavisnosti između promenljivih X i Y

nutku, odnosno $p(1) = \lambda p(0)$, $p(2) = \lambda p(1)$, itd. gde je λ parametar koji opisuje neto priraštaj stanovništva (rast stanovništva prati geometrijsku progresiju), i može se dobiti iz postojećih podataka.

Ako se sa γ označi stopa nataliteta, odnosno konstantna brzina rađanja u jedinici vremena po jedinki, a sa δ stopa mortaliteta, odnosno konstantna brzina umiranja u jedinici vremena po jedinki, tada važi da je konstantan priraštaj ([15]):

$$\lambda = \gamma - \delta$$

Ako je sa $p(t)$ označena veličina populacije u trenutku t , onda je ona posle nekog vremenskog intervala Δt jednaka:

$$p(t + \Delta t) = p(t) + \lambda p(t) \Delta t$$

Drugim rečima, rast je srazmeran postojećoj populaciji i vremenu.

Sada dolazimo do problema za običnu diferencijalnu jednačinu:

$$p'(t) = \lambda p(t), \tag{2.1}$$

$$p(0) = p_0$$

Rešavanjem ove diferencijalne jednačine dobijamo izraz koji nam govori da je priraštaj u jedinici vremena proporcionalan veličini populacije:

$$\frac{dp(t)}{p(t)} = \lambda dt$$

Daljim sređivanjem ovog izraza dobijamo:

$$\ln|p(t)| = \lambda t + C$$

$$p(t) = e^{\lambda t} e^C$$

$$A := e^C$$

$$p(t) = A e^{\lambda t}$$

Dakle, rast populacije je eksponencijalan.

Konstantu A određujemo jer znamo početnu populaciju $p(0) = p_0$. Kako važi da je:

$$p(0) = p_0 = A e^0 = A$$

rešenje jednačine 2.1 je ([16]):

$$p(t) = p_0 e^{\lambda t}$$

Na ovaj način se dolazi do **osnovnog (Maltusovog) populacionog modela**.

Ovakvo rešenje posle nekog vremena dovodi do nerealnih procena i priraštaj bi na neki način trebalo ograničiti.

2.2.2 Modifikacija osnovnog populacionog modela

Najveća mana osnovnog (Maltusovog) modela jeste neograničen rast stanovništva. Iz tog razloga, Pierre-Francois Verhulst, belgijski matematičar, zaključio je da je potrebno napraviti model koji ograničava rast do neke maksimalne fiksne vrednosti koja je karakteristična za sistem koji se posmatra, odnosno do nekog maksimalnog kapaciteta sredine K . Ograničeni resursi usporavaju rast populacije, te populacija teži ka graničnom zasićenju (vidi u [17]).

Druga mana jeste to što linearne brzine rađanja i umiranja nisu konstante, već su date sa:

$$\begin{aligned}\gamma(t) &= \gamma_0 - \gamma_1 p(t) \\ \delta(t) &= \delta_0 + \delta_1 p(t) \\ \gamma_0 &> \delta_0 > 0, \gamma_1, \delta_1 > 0\end{aligned}$$

U Verhulstovom logističkom modelu maksimalni priraštaj označen je sa a , gde je ([15]):

$$a = \gamma_0 - \delta_0$$

tako da važi da je prirodni priraštaj

$$\lambda(t) = (\gamma_0 - \delta_0) - (\gamma_1 + \delta_1)p(t) = a - bp(t)$$

gde je $b = \gamma_1 + \delta_1$

Populaciona jednačina sada dobija oblik:

$$\begin{aligned}p'(t) &= \lambda(t)p(t) = ap(t) - bp^2(t) = a\left(1 - \frac{b}{a}p(t)\right)p(t) = a\left(1 - \frac{1}{K}p(t)\right)p(t) \\ a &> b > 0, p(0) &= p_0\end{aligned}$$

Prema Verhulstu, populacija u početku raste eksponencijalno sa stopom rasta a , ali se taj rast smanjuje kako se populacija približava maksimalnom kapacitetu sistema ([15]): $K = \frac{a}{b}$

Ovakvo ponašanje se modelira sledećom jednačinom:

$$\begin{aligned}\frac{dp(t)}{dt} &= ap(t)\left(1 - \frac{p(t)}{K}\right) \\ p(0) &= p_0\end{aligned}$$

Ova jednačina se naziva **logistička jednačina**.

Drugim rečima, u slučaju kada je populacija mala u odnosu na kapacitet K , izraz u zagradi je približno jednak jedinici, i populacija se ponaša prema Maltusovom populacionom modelu. Kada se populacija približi maksimalnom kapacitetu, izraz u zagradi teži nuli, tj. dolazi do otpora populacije prema daljem rastu u trenutku t (vidi u [15]).

Rešavajući logističku jednačinu:

$$\begin{aligned}\frac{dp(t)}{dt} &= ap(t)\left(1 - \frac{p(t)}{K}\right) \\ \int \frac{1}{p\left(1 - \frac{p}{K}\right)} dp &= at + C \\ \int \left(\frac{1}{p} + \frac{1}{K-p}\right) dp &= at + C \\ \ln|p| - \ln|K-p| &= at + C \\ \ln\left|\frac{K-p}{p}\right| &= -at - C \\ \left|\frac{K-p}{p}\right| &= e^{-at}e^{-C} \\ \frac{K}{p} - 1 &= Ae^{-at}\end{aligned}$$

dolazi se do opšteg rešenja koje se naziva **logistička funkcija**.

Kada se uvrsti i početni uslov, dobija se:

$$\begin{aligned}p(t) &= \frac{K}{1 + Ae^{-at}} \\ p(0) = p_0 &= \frac{K}{1 + A} \\ A &= \frac{K - p_0}{p_0}\end{aligned}$$

Rešenje logističke jednačine može da se prikaže i u ekvivalentnom obliku:

$$\begin{aligned}p(t) &= \frac{ap_0}{bp_0 + (a - bp_0)e^{-at}} \\ p(t) &= \frac{Kp_0}{p_0 + (K - p_0)e^{-at}}\end{aligned}$$

Kriva $p(t)$ ima S-oblik i naziva se **logistička kriva**.

Postoje različiti oblici logističke funkcije, a jedan od specijalnih slučajeva je **sigmoid funkcija** ili **sigmoid kriva**.

2.3 Sigmoid funkcija

Sigmoid funkcija, poznata i pod nazivom **standardna logistička funkcija**, data je sa ([15]):

$$P(t) = \frac{1}{1 + e^{-t}}$$

Promenljiva t se definiše kao $\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, gde su $\beta_i, i = 1 \dots k$ regresioni koeficijenti koji opisuju veličinu doprinosa odgovarajućeg rizičnog faktora x_i . Veličina P predstavlja verovatnoću da se neki događaj desi pod uticajem nezavisnih rizičnih faktora (vidi u [18]). Što su regresioni koeficijenti veći, nezavisne promenljive x_i povećavaju verovatnoću pozitivnog ishoda, i obrnuto.

Standardna sigmoid funkcija se dobija kao rešenje nelinearne diferencijalne jednačine prvog reda ([15]):

$$\begin{aligned}\frac{dP}{dt} &= P(1 - P) \\ P(0) &= \frac{1}{2}\end{aligned}$$

Ona je dakle strogo rastuća funkcija koja se može prikazati i u obliku:

$$\varphi(\nu) = \frac{1}{1 + e^{-a\nu}}$$

gde je parametar a parametar nagiba sigmoidne funkcije.

2.4 Razlike između linearne i logističke regresije

Kao što je napomenuto na početku ovog poglavlja, osnovna razlika između linearnog i logističkog regresionog modela jeste to što se kod linearnog modela pretpostavlja da je zavisna promenljiva neprekidna, dok je kod logističkog modela ona diskretna.

Druga razlika se tiče prirode veze između zavisne i nezavisne promenljive. Ova veza se najbolje može predstaviti preko očekivane vrednosti zavisne promenljive za datu vrednost nezavisne, $E(Y|x)$, gde Y označava zavisnu promenljivu, a x nezavisnu.

U linearnoj regresiji, ova vrednost se može izraziti kao linearna jednakost po x , tj.

$$E(Y|x) = \beta_0 + \beta_1 x$$

Ovaj izraz podrazumeva da $E(Y|x)$ može da uzme bilo koju vrednost sve dok se x kreće u intervalu od $-\infty$ do $+\infty$ (vidi u [29]).

U logističkoj regresiji, kako je zavisna promenljiva dihotomna, i uzima vrednosti 0 i 1, može se pretpostaviti da uzima vrednost 1 sa verovatnoćom π , a vrednost 0 sa verovatnoćom $1 - \pi$, tj.

$$Y : \begin{pmatrix} 0 & 1 \\ 1 - \pi & \pi \end{pmatrix}$$

Analogno, slučajna promenljiva $Y|x$ će takođe uzimati 0 i 1 sa verovatnoćama $1 - \pi(x)$ i $\pi(x)$, tj.

$$Y|x : \begin{pmatrix} 0 & 1 \\ 1 - \pi(x) & \pi(x) \end{pmatrix}$$

Pošto je najinteresantnija očekivana vrednost Y za dato x , ona može da se izračuna na sledeći način:

$$E(Y|x) = 0 \cdot (1 - \pi(x)) + 1 \cdot \pi(x) = \pi(x)$$

Iz tog razloga, ubuduće će se koristiti oznaka $\pi(x)$ za prikazivanje uslovne sredine od Y za dato x kada se koristi logistička raspodela ([19]).

Sa dihotomnim podacima, za uslovnu sredinu mora da važi $0 \leq E(Y|x) \leq 1$

I treća važna razlika između linearnog i logističkog regresionog modela tiče se uslovne raspodele za rezultujuću promenljivu.

U linearnom regresionom modelu, pretpostavlja se da se opservacija rezultujuće promenljive može izraziti sa:

$$Y = E(Y|x) + \varepsilon$$

gde je ε greška i predstavlja jedno odstupanje registrovane vrednosti od uslovne sredine. Najopštija pretpostavka je da ε ima normalnu raspodelu sa sredinom nula i nekom varijansom koja je konstantna za sve nivoe nezavisne promenljive. To znači da će u linearnom regresionom modelu uslovna raspodela rezultujuće promenljive za dato x biti normalna sa sredinom $E(Y) = E(Y|x)$ i varijansom $Var(Y)$ koja je konstantna.

Međutim, ovo ne važi u slučaju kada je rezultujuća promenljiva dihotomna. U ovoj situaciji, vrednost rezultujuće promenljive za dato x se može izraziti kao ([15]):

$$Y = \pi(x) + \varepsilon$$

gde veličina ε može uzeti jednu od dve moguće vrednosti. Za $y = 1$, važi da je $\varepsilon = 1 - \pi(x)$ sa verovatnoćom $\pi(x)$, dok je za $y = 0$, $\varepsilon = -\pi(x)$ sa verovatnoćom $1 - \pi(x)$. Dakle, ε ima raspodelu sa sredinom $E(\varepsilon) = 0$ i varijansom $Var(\varepsilon) = \pi(x)(1 - \pi(x))$.

$$\varepsilon : \begin{pmatrix} -\pi(x) & 1 - \pi(x) \\ 1 - \pi(x) & \pi(x) \end{pmatrix}$$

Kako $\pi(x)$ predstavlja verovatnoću, potrebno je da funkcija kojom se ova vrednost modelira bude neprekidna, monotona i da uzima vrednost između 0 i 1.

2.5 Logistički regresioni model i logit transformacija

Kao što je rečeno u dosadašnjem delu rada, standardna logistička funkcija je data sa:

$$P(t) = \frac{1}{1 + e^{-t}}$$

Logistički regresioni model ima sledeći oblik:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}, \beta_0, \beta_1 \in \mathbb{R}, \beta_1 \neq 0$$

Analogno, za slučaj kada postoji više nezavisnih promenljivih, ovaj model ima sledeći oblik ([15]):

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

Kako ova funkcija nije linearna po parametrima $\beta_i, i = 0 \dots p$, može se linearizovati odgovarajućom logističkom transformacijom. Tada važi:

$$1 - \pi(x) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

Odakle sledi da je:

$$\frac{\pi(x)}{1 - \pi(x)} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}$$

Odnosno:

$$\ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = g(x)$$

Ova jednakost se naziva **logist transformacija** i ona sada jeste linearna po komponentama $\beta_i, i = 0 \dots p$. Vrednost π pripada intervalu $[0,1]$, dok se vrednost logita kreće od $-\infty$ do $+\infty$, kao i da može biti neprekidna (vidi u [30]), pa je iz tog razloga logistička funkcija najprikladniji izbor za traženu funkciju koja će modelirati vrednost $\pi(x)$.

2.6 Postupak modeliranja logističkom regresijom

Nakon što se empirijskim i teorijskim pristupom došlo do početnog izbora varijabli (u mom slučaju do izbora od 11 varijabli koje su prikazane u tabeli zajedno sa hipotezama o uticaju varijabli na verovatnoću bankrota), potrebno je matematičkim putem izabrati one promenljive koje će ući u logistički regresioni model, proveriti tačnost dobijenog modela i interpretirati ga na odgovarajući način. Postoji nekoliko koraka za to:

1. Provera ispravnosti podataka u uzorku
2. Univarijabilna analiza svake promenljive
3. Analiza multikolinearnosti promenljivih
4. Multivarijabilna analiza svake promenljive
5. Ocenjivanje značajnosti svake promenljive koja je ušla u model
6. Procena slaganja dobijenog modela sa podacima
7. Interpretacija dobijenog modela

Provera ispravnosti podataka u uzorku: Čest je slučaj da prikupljeni podaci u uzorku sadrže i takozvane **netipične tačke** (outliers), odnosno nestandardne, odstupajuće vrednosti koje mogu loše da utiču na krajnji ishod modela u smislu izvođenja pogrešnih zaključaka. One predstavljaju opservacije koje izrazito odstupaju od celokupne distribucije podataka (vidi u [31]). Netipične tačke mogu da se otkriju tako što se podaci poređaju u varijacioni niz, a zatim se izračunaju aritmetičke sredine promenljivih bez 5% gornjih i 5% donjih slučajeva. Upoređuje se ta aritmetička sredina sa pravom aritmetičkom sredinom određenog obeležja. Ako se te dve aritmetičke sredine mnogo razlikuju, najverovatnije su 5% gornjih i 5% donjih slučajeva netipične tačke (vidi u [32]). Dalja provera otkrivanja netipičnih tačaka vrši se pomoću crtanja dijagrama ili histograma. U slučaju postojanja netipičnih vrednosti, moguće ih je ukloniti ili zameniti. U ovom radu, netipične vrednosti, ukoliko postoje, biće zamenjene verodostojnijim podacima.

Univarijabilna analiza svake promenljive: Svako građenje statističkog modela teži ka modelu sa što manjim brojem promenljivih koji će dobro objasniti podatke. Minimiziranje broja promenljivih u modelu se vrši radi dobijanja numerički stabilnijeg modela, kao i modela koji će se lakše generalizovati. Prilikom uključivanja više promenljivih u model, ocene standardne greške postaju veće, i model više zavisi od registrovanih podataka ([15]). Za kategorijalne promenljive sa nekoliko celobrojnih vrednosti analizira se tabela kontigencije³

³Neparametarski test nezavisnosti koji se primenjuje isključivo za kategorijski tip podataka

koja sadrži ishode 0 i 1, i ima k nivoa nezavisne promenljive. Ocenjuje se pojedinačni odnos šansi⁴ (zajedno sa granicama poverenja) za promenljive koje pokazuju umeren nivo povezanosti koristeći jedan od nivoa kao referentnu kategoriju ([15]).

Za neprekidne promenljive, kao što je slučaj u ovom radu, analiza bi trebala da obuhvati ([15]): fitovanje univarijabilnog modela radi dobijanja ocenjenog koeficijenta, standardne greške, testa količnika verodostojnosti za značajnost koeficijenta kao i univarijabilne Wald statistike. Nešto više o Wald statistici će biti napisano u daljem delu rada.

Osnovni problem kod univarijabilnog pristupa je što on zanemaruje mogućnost da skup promenljivih, od kojih je svaka slabo povezana sa rezultatom, može postati važan prediktor rezultata kada se posmatraju zajedno (kao u [15]).

Analiza multikolinearnosti promenljivih: U svakoj statističkoj analizi je veoma važno proveriti međukorelacije nezavisnih promenljivih. Multikolinearnost pokazuje kolika je međuzavisnost između nezavisnih varijabli. Što je veća multikolinearnost, to je veći problem upotrebe varijabli kao pokazatelja relativnog uticaja na zavisnu varijablu i sam model je manje tačan. Dakle, prilikom svakog statističkog modeliranja, neophodno je izvršiti analizu multikolinearnosti promenljivih, i ukoliko postoji, isključiti jednu od jako međukoreliranih promenljivih iz modela.

Multivarijabilna analiza svake promenljive: Bilo koja promenljiva koja je isključena iz modela nakon univarijabilnog testa može biti kandidat za multivarijabilni model zajedno sa svim ostalim promenljivama za koje se već zna da su značajne. Kako bi se izbegao problem isključivanja promenljive koja nije jako povezana sa rezultatom a može biti značajan prediktor u kombinaciji sa ostalim promenljivama, može da se izabere nivo značajnosti dovoljno velik kako bi se dopustilo i tim promenljivama uključivanje u model (vidi u [15]). Tehnika koja se najčešće koristi je metoda *korak po korak* u kojoj se promenljive koje su izabrane uključuju odnosno isključuju iz modela u nizu koraka zasnovanih na odgovarajućim statističkim kriterijumima. O ovoj metodi će biti navedeno više detalja kasnije.

Ocenjivanje značajnosti svake promenljive koja je ušla u model: Kako bi se ocenila značajnost svake promenljive koja je za sada ušla u model, najpre je potrebno za svaku od njih ispitati Wald ili Score statistiku. Nešto više o ovim statistikama će biti napisano u daljem delu rada. Promenljive koje ne doprinose modelu potrebno je eliminisati. Nakon toga, novi model se upoređuje sa starim, koristeći *test količnika verodostojnosti*. Nešto više o hi-kvadrat testu količnika verodostojnosti će biti rečeno u kasnijem delu rada. Krajnji rezultat treba naravno da bude da su sve promenljive koje su uključene u model značajne, a one koje su isključene iz njega statistički neznačajne.

⁴Mera povezanosti nezavisne promenljive sa ishodom od interesa

Procena slaganja dobijenog modela sa podacima: Nakon što se došlo do modela koji sadrži one promenljive koje su značajne, potrebno je proceniti i slaganje ocenjenog logističkog regresionog modela sa podacima. Drugim rečima, potrebno je proveriti koliko dobro ovaj model opisuje nezavisnu promenljivu. Ovaj test se naziva *goodness of fit* i o njemu će biti reči u daljem delu rada. Postoji nekoliko mogućnosti za procenu ovakvog slaganja, te će se u nastavku ovog rada više pažnje posvetiti i Pearsonovoj hi-kvadrat statistici, tabelama klasifikacije, kao Hosmer-Lemeshow testu.

Interpretacija dobijenog modela: Pošto su sprovedene sve potrebne analize i došlo se do modela koji je dobro prilagođen podacima i sadrži statistički značajne promenljive, ostaje da se donesu zaključci o tome šta ocenjeni koeficijenti govore o pitanjima zbog kojih je započeto istraživanje. Interpretacija logističkog regresionog modela zavisi od toga da li je nezavisna promenljiva kategorijalna ili neprekidna, te će se i na ovu temu detaljnije pisati u kasnijem delu rada.

2.7 Logistička regresija “korak po korak”

Izbor promenljivih metodom “korak po korak” (Stepwise Logistic Regression) se često koristi prilikom modeliranja regresijom. Upotreba ovakvog postupka obezbeđuje brz i efektivan način za proveru i kontrolu velikog broja promenljivih, kao i fitovanja više logističkih regresionih jednačina istovremeno ([15]).

Metod “korak po korak” baziran je na statističkom algoritmu koji proverava značajnost promenljivih, pa ih na osnovu utvrđenog pravila odlučivanja uključuje ili isključuje iz modela.

Značajnost promenljive se definiše pomoću statističke značajnosti njenog koeficijenta. U linearnoj regresiji, za postupak “korak po korak” koristi se **F test**⁵, jer postoji pretpostavka da greške imaju normalnu raspodelu. U logističkoj regresiji pak, pretpostavka je da greške imaju binomnu raspodelu, pa je značajnost ocenjena putem **hi- kvadrat testa** količnika verodostojnosti (o ovom testu će biti više reči kasnije). Prilikom bilo koje vrste ovakve procedure, najvažnija promenljiva, u statističkom smislu te reči, je ona koja prouzrokuje najveću promenu u logaritmu verodostojnosti (vidi u [15]).

Odlučujući aspekt u korišćenju logističke regresije “korak po korak” jeste izbor nivoa za procenu važnosti promenljive. On određuje koliko promenljivih će biti uključeno u konačan model. Najčešće se koristi izbor vrednosti α u rang od 0.05 do 0.25. Ovo naravno zavisi od prirode samog istraživanja.

⁵F test proverava da li su varijanse dve različite populacije jednake

Postoje dva tipa metode “korak po korak”, jedan se odnosi na unos varijabli u početni model jedna po jedna, tzv. **izbor unapred** (“Forward selection”), a drugi na slučaj kada su unete sve varijable u početni model, a zatim se one koje najmanje doprinose modelu jedna po jedna uklanjaju, tzv. **eliminacija unazad** (“Backward deletion”).

Opis algoritma za izbor unapred:

Neka u modelu postoji ukupno p mogućih nezavisnih promenljivih, gde je za svaku utvrđeno da li je značajna prilikom analiziranja izlazne promenljive.

Najpre se fituje model koji sadrži samo odsečak, β_0 , izračunavanjem njegovog logaritma verodostojnosti L_0 . Nakon toga, svaki od p mogućih univarijabilnih logističkih regresionih modela se fituje, i upoređuje se njihovi odgovarajući logaritmi verodostojnosti. Bira se ona promenljiva koja ima najmanju p-vrednost testa količnika verodostojnosti, x_1 , i ako je p-vrednost manja od α uključujemo ovu promenljivu u model.

Nakon toga se fituje logistički regresioni model koji sadrži izabranu promenljivu x_1 , i određuje se da li je bilo koja od $p - 1$ preostalih promenljivih značajna dok je promenljiva x_1 u modelu. Ponovo se bira promenljiva sa najmanjom p-vrednosti testa količnika verodostojnosti, x_2 . Ukoliko je p-vrednost za test količnika verodostojnosti manja od α uključuje se i ova promenljiva u model.

Proces se nastavlja na ovaj način, dokle god postoje nove promenljive za analizu. Važno je napomenuti da se može desiti da, kada u model uđe, na primer, promenljiva x_2 , promenljiva x_1 više nije značajna za model. Dakle, ne bi bilo loše uključiti i proveru za tzv. eliminaciju unazad.

Opis algoritma za eliminaciju unazad: Ponovo se pretpostavlja da u modelu postoji ukupno p mogućih nezavisnih promenljivih, gde je za svaku utvrđeno da li je značajna prilikom analiziranja izlazne promenljive.

U ovom slučaju, najpre se fituje potpun model koji sadrži svih p mogućih nezavisnih promenljivih.

Sada se računa logaritam verodostojnosti kada je promenljiva x_1 izbačena iz modela. Računa se količnik verodostojnosti ovakvog i punog modela, kao i p-vrednost ove statistike. Odlučuje se da li ovako izabranu promenljivu treba eliminisati tako što se upoređuje odgovarajuća p-vrednost sa unapred izabranim nivoom α_1 . Ako nije poželjno da se izbací previše promenljivih, koristi se veća vrednost α_1 , i obrnuto.

Kao i kod izbora unapred, proces eliminacije se nastavlja dokle god postoje promenljive koje bi trebalo eliminisati iz modela.

Metod “korak po korak” se završava nakon što sve promenljive u modelu imaju p-vrednosti za eliminisanje koje su manje od α a promenljive koje nisu uključene u model imaju p-vrednost za unos koja je veća od α_1 . Model sada sadrži one promenljive koje su značajne u odnosu na kriterijum za α i α_1 . One mogu ili ne moraju biti promenljive koje su prikazane u konačnom modelu. U slučaju da se izabrane vrednosti za α i α_1 slažu sa našim stavom o statističkoj značajnosti, tada model na kraju poslednjeg koraka može sadržati značajne promenljive. Međutim, ako je korišćena vrednost za α i α_1 koja je manje stroga, tada je potrebno birati promenljive za konačni model iz tabele koja prikazuje rezultate “korak po korak” procedure ([15]).

Napomena: p-vrednosti izračunate u proceduri izbora “korak po korak” nisu p-vrednosti u tradicionalnom kontekstu testiranja hipoteza. Umesto toga, one imaju značenje indikatora relativnog značaja među promenljivama.

Izbor promenljivih metodom “korak po korak” je upravo onaj izbor promenljivih koji će se koristiti u ovom radu radi modeliranja verovatnoće bankrota preduzeća.

2.8 Ocenjivanje značajnosti koeficijenata

Dalje analiziranje fitovanog modela se odnosi na ocenjivanje značajnosti promenljivih koje su ostale u modelu. Ovo obično uključuje formulisanje i testiranje statističkih hipoteza za određivanje da li su nezavisne promenljive u modelu “značajno” povezane sa rezultujućom promenljivom.

Ono što je u ovom momentu važno odrediti jeste da li model koji sadrži neku promenljivu, bolje opisuje rezultujuću promenljivu nego model koji je ne sadrži.

Ako su predviđene vrednosti zavisne promenljive na osnovu modela koji sadrži jednu određenu nezavisnu promenljivu tačnije od vrednosti koje su predviđene na osnovu modela koji ne sadrži tu promenljivu, tada se kaže da je promenljiva u modelu značajna.

U linearnoj regresiji, testiranje značajnosti koeficijenta vrši se korišćenjem **analize varijanse**. Totalna suma kvadrata odstupanja registrovanih vrednosti od njihove sredine sastoji se od *rezidualne sume kvadrata* i *regresione sume kvadrata* (vidi u [14]).

Rezidualna suma kvadrata je suma kvadrata odstupanja registrovanih vrednosti oko regresione prave:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Regresiona suma kvadrata je suma kvadrata odstupanja predviđenih vrednosti na osnovu regresionog metoda od sredine zavisne promenljive:

$$SSR = \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) - \left(\sum_{i=1}^n (y_i - \bar{y}_i)^2 \right)$$

Kada je reč o logističkoj regresiji, poređenje registrovane i predviđene vrednosti dobijene iz modela koji sadrži nezavisnu promenljivu i modela koji je ne sadrži, bazira se na logaritmu **funkcije verodostojnosti**.

Registrovana vrednost zavisne promenljive se odnosi na vrednost koja se dobija iz zasićenog modela⁶ (vidi u [33]).

Funkcija verodostojnosti ima sledeći oblik:

$$D = -2 \ln \left(\frac{lf}{lz} \right) \quad (2.2)$$

gde je lf verodostojnost fitovanog modela, a lz verodostojnost zasićenog modela

Izraz $\frac{lf}{lz}$ se naziva **količnik verodostojnosti**. Koristi se $-2 \ln$ kako bi se dobila veličina čija je raspodela poznata, tako da se ova statistika može koristiti za testiranje hipoteza. Ovaj test je poznat pod nazivom **test količnika verodostojnosti**.

Kasnije, uvođenjem pojma metoda maksimalne verodostojnosti⁷ i fukcije verodostojnosti u okviru ovog metoda, izraz 2.2 postaje:

$$D = -2 \sum_{i=1}^n \left(y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right)$$

gde je $\hat{\pi}_i = \hat{\pi}(x_i)$.

Statistika D, u ovoj jednakosti, naziva se **odstupanje** (vidi u [19]), a poznata je i kao **logaritamski test statistika količnika verodostojnosti**.

⁶Zasićen model je onaj model koji sadrži onoliko parametara koliko ima podataka

⁷Više o ovom metodu u poglavlju 2.9.1

Kako bi se procenili značajnost nezavisne promenljive, upoređuje se vrednost D za model koji sadrži nezavisnu promenljivu i za model koji je ne sadrži. Promena u D koja nastaje zbog uključivanja nezavisne promenljive u model je data sa:

$$G = D(mb) - D(ms)$$

gde je mb model bez nezavisne promenljive a ms model sa nezavisnom promenljivom.

Pošto obe vrednosti D imaju isti imenilac (verodostojnost zasićenog modela), G se može izraziti i kao:

$$G = -2\ln\left(\frac{D(vmb)}{D(vms)}\right)$$

gde je vmb verodostojnost modela bez nezavisne promenljive a vms verodostojnost modela sa nezavisnom promenljivom.

- Univarijabilni logistički regresioni model

Kada promenljiva nije u modelu, ocena maksimalne verodostojnosti od β_0 je $\ln\left(\frac{\sum y_i}{\sum(1-y_i)}\right)$, a predviđena vrednost, $\frac{\sum y_i}{n}$, je konstantna. U tom slučaju vrednost G je:

$$G = -2\ln\left[\frac{\left(\frac{\sum y_i}{n}\right)^{\sum y_i} \left(\frac{\sum(1-y_i)}{n}\right)^{\sum(1-y_i)}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}}\right]$$

Ukoliko je tačna hipoteza da je $\beta_1 = 0$, statistika G ima hi-kvadrat raspodelu sa jednim stepenom slobode⁸.

- Multivarijabilni logistički regresioni model

Test količnika verodostojnosti za ukupnu značajnost p koeficijenata za nezavisne promenljive je izveden na isti način kao i u univarijabilnom slučaju. Jedina razlika je u tome što su fitovane vrednosti za model $\hat{\pi}_i$ bazirane na vektoru $\hat{\beta}$ koji sadrži $p + 1$ parametar. Tada G ima hi-kvadrat raspodelu sa p stepeni slobode pod nultom hipotezom da je svih p koeficijenata nagiba za kovarijate u modelu jednako 0 ([15]).

Testiranje hipoteza o vektoru parametra β se može vršiti i korišćenjem dva statistički ekvivalentna testa: **Wald testa** i **Score testa**⁹.

2.8.1 Wald test

Testira se hipoteza $H_0(\beta_1 = 0)$ da test statistika ima standardnu normalnu raspodelu.

⁸Raspodelu definisanu gustinom $f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$, ($x > 0$) nazivamo hi-kvadrat raspodelom sa n stepeni slobode.

⁹Takođe poznat i pod imenom **Lagrange multiplier test**

Wald test za logistički regresioni model je dobijen kao ([20]):

$$W = \frac{\hat{\beta}_1}{\hat{\delta}(\hat{\beta}_1)}$$

gde je $\hat{\delta}(\hat{\beta}_1)$ ocena standardne devijacije ocene $\hat{\beta}_1$.

Ako je H matrica drugih izvoda funkcije $\ln L(\beta_0, \beta_1)$ definisana sa

$$H = [h_{ij}]_{2 \times 2}, i, j \in \{0, 1\}$$

$$h_{ij} = \frac{\partial^2 \ln L(\beta_0, \beta_1)}{\partial \beta_i \partial \beta_j}, i, j \in \{0, 1\}$$

tada se lako dobija ocena za disperziju ocene $\hat{\beta}_1$:

$$\hat{\delta}^2(\hat{\beta}_1) = (| - h_{ij} |_{\beta_1 = \hat{\beta}_1})^{-1}$$

Waldov test može biti i jednostran i dvostran. U slučaju dvostranosti, donošenje odluke o prihvatanju/odbacivanju H_0 sa zadatim pragom značajnosti α vrši se na sledeći način ([20]):

$$|W| \leq c(1 - \frac{\alpha}{2}) \Rightarrow \text{prihvatamo } H_0$$

$$|W| > c(1 - \frac{\alpha}{2}) \Rightarrow \text{odbacujemo } H_0$$

gde je c granica kritične oblasti.

Test statistika količnika verodostojnosti i Wald statistika daju približno iste vrednosti kad su u pitanju veliki uzorci, pa ako je neka studija dovoljno obimna nije bitno koja statistika se koristi. Međutim, ako su uzorci manjeg obima, statistike mogu značajno da se razlikuju i pokazano je da je test statistika količnika verodostojnosti u ovakvim situacijama ipak tačnija (vidi u [15]).

Wald test često ima nedostatak da se ne odbacuje nulta hipoteza iako su koeficijenti značajni, tako da je preporučljivije koristiti test količnika verodostojnosti.

2.8.2 Score test

Testira se hipoteza $H_0(\beta_1 = 0)$ da test statistika ima hi-kvadrat raspodelu.

Score test za logistički regresioni model je dobijen kao ([20]):

$$S = \frac{u(\hat{\beta}_1)^2}{I(\hat{\beta}_1)}$$

gde je $u(\hat{\beta}_1)$ izvod logaritma funkcije verodostojnosti u tački $\beta_1 = \hat{\beta}_1$:

$$u(\hat{\beta}_1) = \frac{\partial \ln L(\beta_0, \beta_1)}{\partial \beta_1} \Big|_{\beta_1 = \hat{\beta}_1}$$

a $I(\hat{\beta}_1)$ predstavlja Fišerov informacioni kriterijum u istoj tački $\beta_1 = \hat{\beta}_1$:

$$I(\hat{\beta}_1) = E \left(- \frac{\partial^2 \ln L(\beta_0, \beta_1)}{\partial \beta_1^2} \right) \Big|_{\beta_1 = \hat{\beta}_1}$$

Test statistika količnika verodostojnosti, Wald statistika i Score statistika su asimptotski ekvivalentni testovi hipoteza (vidi u [20]). Drugim rečima, rastom obima uzorka, vrednosti ovih test statistika će biti sve sličnije.

Prednost Score testa u odnosu na test količnika verodostojnosti i Wald test je što se može koristiti za traženje izostavljenih varijabli kada je njihov broj veliki (vidi u [21]). Sa druge strane, pokazano je da su na uzorcima manjeg obima tačnije test statistike količnika verodostojnosti i Wald test.

Kada se radi o linearnim modelima, veza između prezentovanih modela je sledeća ([22]):

$$\text{Wald test} \geq \text{Test količnika verodostojnosti} \geq \text{Score test}$$

Za modele koji nisu linearni, često će se pre koristiti test količnika verodostojnosti, ali ovo naravno zavisi i od konkretnog slučaja.

2.9 Slaganje logističkog regresionog modela sa podacima

Neka postoji uzorak od n nezavisnih registrovanih vrednosti parova (x_i, y_i) , $i = 1, 2, \dots, n$ gde y_i označava vrednost rezultujuće dihotomne promenljive, a x_i označava registrovanu vrednost nezavisne promenljive za i -ti subjekat.

Pretpostavimo takođe da je rezultujuća promenljiva kodirana sa 0 ili 1, gde 0 predstavlja odsustvo, a 1 prisustvo neke karakteristike, kao što u ovom radu 0 predstavlja zdravo preduzeće koje nije ušlo u stečaj, a 1 ulazak preduzeća u stečaj.

“Fitovanje” logističkog regresionog modela u jednakosti $\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ zahteva da ocenimo vrednosti za nepoznate parametre β_0 i β_1 (vidi u [14]).

Kao i do sada, da bi se bolje objasnio pojam “fitovanja modela”, posmatra se najjednostavniji logistički regresioni model koji sadrži samo jednu nezavisnu promenljivu, a analogna priča važi i za više nezavisnih promenljivih.

U linearnoj regresiji, najčešći metod za ocenjivanje regresionih parametara je **metod najmanjih kvadrata**.

U ovom metodu, biraju se one vrednosti β_0 i β_1 koje minimiziraju sumu kvadrata odstupanja registrovane vrednosti za Y , od predviđene vrednosti koja je dobijena iz modela (kao u [14]). Ocene koje daje metod najmanjih kvadrata imaju mnogo poželjnih statističkih svojstava kao što su konzistentnost¹⁰, asimptotska efikasnost¹¹ i asimptotska nepristrasnost¹².

Međutim, kada se metod najmanjih kvadrata primeni na model sa dihotomnim ishodom, ocene više nemaju te iste osobine.

Kod logističke regresije, za ocenjivanje regresionih koeficijenata, koristi se **metod maksimalne verodostojnosti**. Ovaj metod daje vrednosti za $\beta_i, i = 0 \dots p$ koje maksimiziraju verovatnoću dobijanja registrovanog skupa podataka ([15]). Drugim rečima, utvrđuju se verovatnoće registrovanih podataka za različite kombinacije vrednosti regresionih koeficijenata.

2.9.1 Metod maksimalne verodostojnosti

Metod maksimalne verodostojnosti daje vrednosti za nepoznate parametre koji maksimiziraju verovatnoću dobijanja registrovanog skupa podataka (vidi u [14]). Pokazalo se da ovaj metod daje ocene koje su asimptotski efikasnije od ocena dobijenih na bilo koji drugi način (vidi u [35]).

Da bi se opisao metod maksimalne verodostojnosti, potrebno je najpre definisati **funkciju verodostojnosti** koja izražava verovatnoću registrovanih podataka u funkciji nepoznatih parametara. To je funkcija regresionih koeficijenata koja predstavlja verovatnoću koja kombinuje doprinose svih subjekata u istraživanju.

Neka je zavisna promenljiva

$$Y : \begin{pmatrix} 0 & 1 \\ 1 - \pi & \pi \end{pmatrix}$$

tada izraz

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

za proizvoljnu vrednost $\beta = (\beta_0, \beta_1)$, daje uslovnu verovatnoću $P\{Y = 1|x\} = \pi(x)$ i $P\{Y = 0|x\} = 1 - \pi(x)$.

¹⁰ $\lim_{n \rightarrow \infty} MSE(\hat{\beta}) = 0$, gde je $MSE(\hat{\beta}) = E((\hat{\beta} - \beta)^2)$

¹¹ Asimptotska efikasnost podrazumeva da ocena koeficijenta ima osobinu konzistentnosti, najmanju asimptotsku varijansu i asimptotsku distribuciju sa konačnom sredinom i varijansom

¹² $\lim_{n \rightarrow \infty} E(\hat{\beta}) = \beta$

Za one parove (x_i, y_i) gde je $y_i = 1$ doprinos funkciji verodostojnosti je $\pi(x_i)$, a za one parove (x_i, y_i) gde je $y_i = 0$ doprinos funkciji verodostojnosti je (vidi u [15]): $1 - \pi(x_i)$, gde je $\pi(x_i)$ vrednost funkcije $\pi(x)$ za konkretno x_i .

Dakle, za par (x_i, y_i) doprinos funkciji verodostojnosti je dat sledećim izrazom:

$$\pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

Pošto se radi pod pretpostavkom da su registrovane vrednosti nezavisne, funkcija verodostojnosti je dobijena kao proizvod gornjeg izraza, tj.

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (2.3)$$

Verodostojnost se može predstaviti i kao:

$$l(\beta) = \prod_{i=1}^n \left(\frac{\pi(x_i)}{(1 - \pi(x_i))} \right)^{y_i} (1 - \pi(x_i))$$

gde se izraz

$$\frac{\pi(x_i)}{1 - \pi(x_i)}$$

naziva šansa za $P\{Y = 1|x_i\}$ i jednak je:

$$\frac{\pi(x_i)}{1 - \pi(x_i)} = e^{e^{\beta_0 + \beta_1 x_{1i}}}$$

odnosno verodostojnost predstavlja funkciju registrovanih vrednosti zavisne i nezavisnih promenljivih i nepoznatih parametara (vidi u [36]).

Radi jednostavnosti, koristi se logaritam ove funkcije, odnosno logaritam verodostojnosti:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^n (y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)))$$

Koristeći informacije koje imamo o izgledu logističkog regresionog modela, $\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$, može se izračunati:

$$\begin{aligned} \ln\left(\frac{\pi(x_i)}{(1 - \pi(x_i))}\right) &= \beta_0 + \beta_1 x_i \\ \ln(1 - \pi(x_i)) &= -\ln(1 + e^{\beta_0 + \beta_1 x}) \end{aligned}$$

pa se logaritam verodostojnosti može zapisati na drugačiji način:

$$L(\beta) = \sum_{i=1}^n (y_i(\beta_0 + \beta_1 x) - \ln(1 + e^{\beta_0 + \beta_1 x}))$$

Ocene parametara se traže tako da maksimiziraju funkciju verodostojnosti. Da bi se našlo $\beta = (\beta_0, \beta_1)$ koje maksimizira funkciju $L(\beta)$, diferencira se $L(\beta)$ u odnosu na β_0 i β_1 , dobijene jednačine se izjednačavaju sa nulom pa se rešava odgovarajući sistem:

$$0 = \frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(y_i - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \right) = \sum_{i=1}^n (y_i - \pi(x_i))$$

$$0 = \frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(y_i x_i - x_i \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \right) = \sum_{i=1}^n (x_i (y_i - \pi(x_i)))$$

Posledica prve jednakosti je da važi

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \pi(x_i)$$

odnosno, suma registrovanih vrednosti za y je jednaka sumi očekivanih vrednosti na osnovu modela.

Ove jednačine su nelinearne po β_0 i β_1 , i sistem se ne može rešiti analitički, ali postoje numerički algoritmi koji omogućavaju pronalaženje rešenja, tj. sistem se rešava nekim od iterativnih postupaka.

Vrednost koja se dobija kao rešenje ovih iteracija naziva se **ocena maksimalne verodostojnosti** i označava se sa $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$

Da bi rešenja sistema bile ocene parametara β_0 i β_1 po metodi maksimalne verodostojnosti, matrica drugih parcijalnih izvoda po β_0 i β_1 mora biti negativno definitna¹³.

2.10 Procena slaganja modela sa podacima (goodness-of-fit)

Sada se može pretpostaviti da model sadrži sve promenljive koje treba da su u modelu, tj. da su sve promenljive u modelu značajne, i da su unete u korektnom funkcionalnom obliku. Sledeće što je potrebno razmotriti su metode koje procenjuju slaganje ocenjenog logističkog regresionog modela sa podacima.

Potrebno je doći do odgovora na pitanje koliko efikasno dobijeni model opisuje rezultujuću promenljivu, tzv. **goodness-of-fit**.

Neka su registrovane uzoračke vrednosti rezultujuće promenljive prikazane u vektorskom obliku sa $y' = (y_1, y_2 \dots y_n)$. Fitovane vrednosti se označavaju sa $\hat{y}$, gde je $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$.

¹³Simetrična matrica $H \in M_n(\mathbb{R})$ je negativno definitna ako je $\forall x \in \mathbb{R}^n \setminus \{0\} (Hx|x) < 0$

Za model se kaže da je prilagođen podacima ako važi ([15]):

1. mere rastojanja između y i $\hat{y}$ su male
2. doprinos svakog para, $(y_i, \hat{y}_i)$, $i = 1, 2, 3, \dots, n$ ovim merama je nesistematski i mali u odnosu na grešku modela.

Iz tog razloga, kompletno procenjivanje fitovanog modela obuhvata izračunavanje mera rastojanja između y i $\hat{y}$, i ispitivanje pojedinačnih komponenti tih mera (vidi u [15]).

Osnovne mere za goodness-of-fit predstavljaju opšti pokazatelj koliko dobro se model slaže sa podacima, ali ne govori o tome da li je dati model bolji od nekog drugog modela (vidi u [37]). Iz tog razloga, često se koriste strategije koje porede više različitih modela a ne sam goodness-of-fit.

Kada je procenjeno koliko se dobro model slaže sa podacima, potrebno je pozabaviti se brojem kovarijatnih obrazaca. Kovarijatni obrazac predstavlja opservacije sa istim vrednostima za sve nezavisne promenljive ([37]).

Goodness-of-fit je procenjen preko grupisanja fitovanih vrednosti određenih pomoću kovarijati u modelu, a ne preko ukupnog skupa kovarijanti. Ukoliko fitovan model sadrži p nezavisnih promenljivih, $\hat{x} = (x_1, x_2, \dots, x_p)$, a J označava broj različitih vrednosti za registrovano x , za sve subjekte koji imaju istu vrednost za x , važi $J < n$ (kao u [19]).

Broj subjekata za koje je $x = x_j$ se označava sa m_j , $j = 1, 2, 3, \dots, J$, odnosno važi da je $\sum m_j = n$. Neka je sa y_j označen broj pozitivnih odgovora, $y = 1$, među m_j subjekata za koje važi $x = x_j$ i neka važi da je $\sum y_j = n_1$.

Raspodela za goodness-of-fit statistiku se dobija tako što se pusti da n bude dovoljno veliko. Ako se i broj kovarijatnih obrazaca povećava, tada svaka vrednost m_j teži da bude što manja. Za raspodele dobijene pod pretpostavkom da n postaje veliko se kaže da su **n-asimptotski** ([15]).

Ukoliko se fiksira broj grupa na neku vrednost J , i povećava obim uzorka n , onda će se povećavati i broj elemenata u svakoj grupi tj. ako se fiksira $J < n$ i pusti se da n bude dovoljno veliko, tada svaka vrednost m_j takođe teži da postane što veća. Za raspodele dobijene pod pretpostavkom da svako m_j postaje veliko, kaže se da su **m-asimptotske** ([15]).

Slučaj koji se najčešće javlja u praksi je $J \approx n$ i on predstavlja najveći izazov u razvijanju raspodela goodness-of-fit statistike ([19]).

2.10.1 Pirsonova hi-kvadrat statistika i odstupanje

U linearnoj regresiji osnovne mere za procenu slaganja modela sa podacima čine **funkcije reziduala** koje su definisane kao razlika između posmatrane i fitovane (procenjene) vrednosti.

U logističkoj regresiji, sa druge strane, postoji nekoliko mogućnosti za procenu razlike između ove dve vrednosti. Fitovane vrednosti u logističkoj regresiji su izračunate za svaki kovarijantni obrazac i zavise od ocenjene verovatnoće za taj kovarijantni obrazac. Fitovana vrednost za j -ti kovarijantni obrazac se obeležava sa $\hat{y}_j$ i važi:

$$\hat{y}_j = m_j \hat{\pi}_j = m_j \frac{e^{\hat{g}(x_j)}}{1 + e^{\hat{g}(x_j)}}$$

gde je $\hat{g}(x_j)$ ocenjen logit.

Za određen kovarijantni obrazac, **Pirsonov rezidual** se definiše kao:

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}$$

Statistika koja se bazira na ovim rezidualima je **Pirsonova hi-kvadrat statistika**:

$$\chi^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2$$

Rezidual odstupanja se definiše na sledeći način:

$$d(y_j, \hat{\pi}_j) = \pm \left(2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{m_j - y_j}{m_j (1 - \hat{\pi}_j)} \right) \right] \right)^{1/2}$$

Statistika koja se bazira na ovim rezidualima odstupanja naziva se **odstupanje**:

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2$$

Obe ove statistike, χ^2 i D , imaju hi-kvadrat raspodelu sa $J - (p + 1)$ stepeni slobode.

U slučaju kada je $J \approx n$, problematika je malo komplikovanija jer je raspodela n -asimptotska. Jedan način da se izbegnu moguće smetnje sa raspodelama za χ^2 i D , kada je $J \approx n$ je grupisanje podataka na takav način da se koristi m -asimptotska raspodela (vidi u [15]).

Prednosti ovih statistika su, između ostalog, što se nalaze u skoro svim softverskim statističkim paketima. Još jedna dobra karakteristika je i ta da se statistike relativno lako računaju korišćenjem elementarnih kalkulacija kao i njihova odgovarajuća p -vrednost. Često se dešava da ove dve statistike imaju različite vrednosti, ako su te razlike jako velike smatramo da χ^2 aproksimacija ovih raspodela nije odgovarajuća (vidi u [19]).

2.10.2 Hosmer-Lemeshow test

Kako bi se izbegla problematična upotreba odstupanja prilikom spomenutog slučaja kada je $J \approx n$, David Hosmer i Stanley Lemeshow su predložili grupisanje bazirano na vrednostima ocenjenih verovatnoća. Njihov statistički test, pod nazivom Hosmer–Lemeshow test je veoma široko rasprostranjen, čak se tvrdi da je to SPSS-ov najpouzdaniji test kvaliteta predikcije modela.

Ova statistika zahteva da model ima najmanje tri kovarijatna obrazca, a najbolje rezultate pokazuje kada je broj kovarijatnih obrazaca blizu n (vidi u [39]).

Najpre pretpostavimo da je $J = n$.

U tom slučaju imamo n kolona koje odgovaraju vrednostima ocenjenih verovatnoća, sa prvom kolonom kojoj odgovara najmanja vrednost, i sa n -tom kolonom kojoj odgovara najveća vrednost ([15]). Predložena su dva postupka grupisanja i to formiranjem tabele zasnovane na percentilima ocenjenih verovatnoća i fiksiranim vrednostima ocenjenih verovatnoća.

Bez obzira na to koji postupak grupisanja je u pitanju, Hosmer–Lemeshow statistika $\hat{C}$ je dobijena računanjem Pirsonove hi-kvadrat statistike iz tabele sa observiranim i ocenjenim očekivanim frekvencijama.

Statistika $\hat{C}$ se definiše na sledeći način ([40]):

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

gde je n_k ukupan broj subjekata u k -toj grupi, o_k broj jedinica među c_k kovarijatnih obrazaca:

$$o_k = \sum_{j=1}^{c_k} y_j$$

a $\bar{\pi}_k$ prosečna ocenjena verovatnoća:

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n_k}$$

Hosmer i Lemeshow su pokazali da je za $J = n$, kao i za $J \approx n$ raspodela $\hat{C}$ statistike dobro aproksimirana sa χ^2 raspodelom (vidi u [15]).

Kako raspodela $\hat{C}$ statistike zavisi od m -asimptotske raspodele, prikladnost p -vrednosti zavisi od validnosti pretpostavki da su ocenjene očekivane frekvencije velike. Smatra se da se model dobro slaže sa podacima ako je p -vrednost odgovarajuće χ^2 statistike veća od 0,05 (kao u [15]).

2.10.3 Tabele klasifikacije

Ukoliko je potrebno da se sprovede sažimanje rezultata fitovanog logističkog regresionog modela, to se može uraditi pomoću tabele klasifikacije. **Tabela klasifikacije** je rezultat ukrštanja rezultujuće promenljive sa dihotomnom promenljivom čije su vrednosti izvedene iz ocenjenih logističkih verovatnoća (vidi u [14]).

Da bi se kreirala tabela klasifikacije 2×2 predviđenih vrednosti iz modela za rezultujuću promenljivu nasuprot tačnoj vrednosti rezultujuće promenljive, mora se najpre definisati **nivo odlučivanja** c sa kojim se poredi svaka ocenjena verovatnoća. Ukoliko je $\hat{\pi}_i > c$, važi da je $\hat{y} = 1$, odnosno ukoliko je $\hat{\pi}_i < c$, važi da je $\hat{y} = 0$ ([15]).

Dva važna pojma kada se govori o tabelama klasifikacija su **senzitivnost** i **specifičnost**.

Senzitivnost testa je mera tačnosti testa koja se odnosi na populaciju kod koje postoji prisustvo osobine (vidi u [38]). Drugim rečima, senzitivnost je verovatnoća da je predviđena vrednost zavisne promenljive 1, ukoliko je zavisna promenljiva primila vrednost 1 tj. $P(\hat{y} = 1|y = 1)$.

Specifičnost testa je mera tačnosti testa koja se odnosi na populaciju kod koje ne postoji prisustvo osobine (vidi u [38]). Drugim rečima, specifičnost je verovatnoća da je predviđena vrednost zavisne promenljive 0, ako je njena stvarna vrednost 0 tj. $P(\hat{y} = 0|y = 0)$.

Klasifikacija je osetljiva na relativnu veličinu dve komponentne grupe i uvek favorizuje klasifikaciju u veće grupe ([15]), što ne zavisi od prilagođenosti modela podacima. Mere izvedene iz tabele klasifikacije 2×2 (kao što su senzitivnost i specifičnost) dosta zavise od raspodele verovatnoća u uzorku.

Modeli na bazi mera izvedenih iz tabele klasifikacije 2×2 se ne mogu meriti jer se ove mere ne mogu posmatrati nezavisno od raspodele verovatnoća u uzorcima na kojima su bazirani. Model koji je procenjen u dve populacije, korišćenjem mera senzitivnosti ili specifičnosti može da da vrlo različite utiske o njegovom učinku. Iz tog razloga, tabela klasifikacije je najprikladnija kada je klasifikacija postavljena kao cilj analize, inače bi trebala da bude samo dopuna strožijim metodama procene slaganja modela sa podacima (vidi u [15]).

2.11 Interpretacija logističkog regresionog modela

Sada može da se kaže da je logistički regresioni model prilagođen podacima (fitovan je), a da su promenljive koje ostaju u modelu značajne, tj. da su odgovarajući regresioni koeficijenti različiti od nule.

Zaključak koji treba da se donese na osnovu ocenjenih koeficijenata i njihove značajnosti, jeste šta nam ocenjeni koeficijenti govore o pitanjima zbog kojih je i započeto istraživanje.

Prilikom izvođenja ovih zaključaka, tj. interpretacije samog modela, određuje se funkcionalna veza između zavisne i nezavisne promenljive i definišu se odgovarajuće jedinice promene za nezavisnu promenljivu.

Prvo je potrebno odrediti koja će funkcija zavisne promenljive dati linearnu funkciju po nezavisnim promenljivama. U linearnom regresionom modelu, to je *funkcija identiteta*¹⁴ jer je zavisna promenljiva, po definiciji, linearna po parametrima (vidi u [14]).

Funkcionalnu vezu između zavisne i nezavisne promenljive u logističkom regresionom modelu daje logit funkcija, tj.:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

Zbog jednostavnosti, u nastavku će se raditi samo sa jednom nezavisnom promenljivom.

Ocenjeni koeficijenti za nezavisne promenljive predstavljaju nagib funkcije zavisne promenljive po jedinici promene za nezavisnu promenljivu ([14]).

U linearnom regresionom modelu koeficijent nagiba β_1 iznosi:

$$\beta_1 = y(x + 1) - y(x)$$

za $y(x) = \beta_0 + \beta_1 x$.

U logističkom regresionom modelu sa druge strane, koeficijent nagiba β_1 predstavlja promenu u logitu po jedinici promene nezavisne promenljive, tj. :

$$\beta_1 = g(x + 1) - g(x)$$

Interpretacija fitovanog logističkog regresionog modela zavisi od toga da li je nezavisna promenljiva **dihotomna** ili **neprekidna**.

2.11.1 Interpretacija logističkog modela za dihotomnu nezavisnu promenljivu

Najpre se posmatra slučaj kada je nezavisna promenljiva dihotomna, odnosno kada može uzeti dve vrednosti, 0 i 1, kao što je slučaj i sa ovim modelom.

¹⁴Funkcija identiteta f na skupu M je definisana kao funkcija sa domenom i kodomenom M koja zadovoljava $f(x) = x$, tj. $f : M \rightarrow M$, za sve elemente $x \in M$.

Pošto koeficijent β_1 predstavlja stopu promene zavisne promenljive po jedinici promene nezavisne promenljive, važi da je:

$$g(1) - g(0) = (\beta_0 + \beta_1) - \beta_0 = \beta_1$$

Da bi dobijeni rezultat mogao da se interpretira, uvodi se novi pojam, **odnos šansi** (odds ratio), koji predstavlja meru povezanosti nezavisne promenljive sa ishodom od interesa ([15]).

Šansa označava odnos verovatnoća da se određeni događaj desi i verovatnoće da se taj isti događaj ne desi. Kako je nezavisna promenljiva kodirana sa 0 i 1, posebno se računaju šanse za oba slučaja (vidi u [34]).

Šansa da je zavisna promenljiva uzela vrednost 1, kada nezavisna promenljiva uzme vrednost takođe 1 je:

$$Odds = \frac{P(Y = 1|X = 1)}{P(Y = 0|X = 1)} = \frac{\pi(1)}{1 - \pi(1)}$$

Šansa da je zavisna promenljiva uzela vrednost 1, kada nezavisna promenljiva uzme vrednost 0 je:

$$Odds = \frac{P(Y = 1|X = 0)}{P(Y = 0|X = 0)} = \frac{\pi(0)}{1 - \pi(0)}$$

Odnos šansi (unakrsni odnos šansi), u oznaci OR , je sada definisan kao odnos ove dve šanse, tj.

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}}$$

Moguće vrednosti logističke verovatnoće se predstavljaju sledećom tablicom:

Y	$x = 1$	$x = 0$
$y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
$y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Ukupno:	1	1

Iz ove tablice se vidi zašto se odnos šansi OR naziva i unakrsni odnos šansi, jer se očigledno dobija kao odnos unakrsnog proizvoda elemenata na glavnoj dijagonali i elemenata na sporednoj dijagonali predstavljene tabele.

Zamenom izraza iz tabele u izraz za OR dobijamo:

$$OR = \frac{\left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}\right) / \left(\frac{1}{1 + e^{\beta_0 + \beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1 + e^{\beta_0}}\right) / \left(\frac{1}{1 + e^{\beta_0}}\right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Ocenjen odnos šansi se dobija kada β_1 zamenimo sa $\hat{\beta}_1$, tj.:

$$\hat{OR} = e^{\hat{\beta}_1}$$

Ova jednostavna veza između koeficijenta β i odnosa šansi je jedan od osnovnih razloga zašto se logistička regresija pokazala kao moćan analitički alat (vidi u [15]).

Sada uvodimo pojam **relativnog rizika**, u oznaci RR . Relativni rizik predstavlja odnos verovatnoća uspeha u okviru dve grupe.

U ovom slučaju to se može prikazati na sledeći način:

$$RR = \frac{P(Y = 1|X = 1)}{P(Y = 1|X = 0)} = \frac{\pi(1)}{\pi(0)}$$

Izraz za odnos šansi se sada može zapisati:

$$OR = RR \frac{1 - \pi(0)}{1 - \pi(1)}$$

Odnos šansi često aproksimira relativni rizik. To će se očigledno desiti kada $\frac{1 - \pi(0)}{1 - \pi(1)} \rightarrow 1$, odnosno kada su verovatnoće neuspeha u obe grupe približno jednake. To se dešava u slučajevima kada je verovatnoća $\pi(x)$ dovoljno mala, bez obzira na to da li je $x = 1$ ili $x = 0$ ([15]).

2.11.2 Interpretacija logističkog modela za neprekidnu nezavisnu promenljivu

Sada se posmatra logistički regresioni model koji sadrži neprekidnu nezavisnu promenljivu. Pretpostavimo da je logit linearan po toj promenljivoj, odnosno:

$$g(x) = \beta_0 + \beta_1 x$$

U ovom slučaju, za razliku od slučaja kada je nezavisna promenljiva diskretna, promena od jedne jedinice mere nezavisne promenljive najčešće nije naročito interesantna ([15]). Dakle, da bi se obezbedila pravila interpretacija modela, smatraćemo da se desila promena od c jedinica. Tada je promena u logitu sledeća:

$$g(x + c) - g(x) = \beta_0 + \beta_1(x + c) - \beta_0 + \beta_1 x = \beta_1 c$$

Sada odnos šansi dobija sledeći oblik:

$$OR = e^{c\beta_1}$$

a njegova ocena je:

$$\hat{OR} = e^{c\hat{\beta}_1}$$

U ovom slučaju važi da c može da uzme bilo koju vrednost, ali se mora voditi računa o tome da se na jasan način pokaže kako se rizik da je ishod prisutan menja sa promenom nezavisne promenljive (vidi u [19]).

Glava 3

Primena logističke regresije u modeliranju bankrota

Kompletno statističko istraživanje se sprovodi u IBM SPSS Modelaru. SPSS ima čitavu porodicu tehnika za logističku regresiju koje služe za istraživanje prediktivne moći skupova ili blokova promenljivih i omogućavaju zadavanje načina unosa promenljivih. Postepena tehnika uključivanja promenljivih u model “korak po korak” (stepwise logistic regression) omogućava zadavanje velike grupe mogućih prediktora odakle SPSS sam bira podskup koji ima najveću moć predviđanja.

Procedura koja će se upotrebiti u ovom radu se naziva *Binary Logistic* i ona se koristi u slučaju kada je zavisna promenljiva dihotomna. Kada zavisna promenljiva ima više od dve moguće vrednosti, upotrebljava se skup procedura *Multinomial Logistic*.

3.1 Veličina uzorka

Kao i u većini ostalih statističkih tehnika, mora se najpre razmotriti veličina i priroda uzorka ukoliko će se koristiti logistička regresija. Analiza postaje problematična ukoliko je uzorak mali, a treba da se uključi veliki broj prediktora. Takav problem se naročito primećuje kod kategorijskih prediktora sa ograničenim brojem slučajeva u svakoj kategoriji (vidi u [23]).

Nedostatak malih uzoraka je u tome što je moguća identifikacija samo najvećih statističkih razlika. Sa druge strane, velike veličine uzorka povećavaju statističku snagu tako što će bilo koja razlika, bilo praktično relevantna ili ne, biti razmatrana kao statistički značajna. Iz ovih razloga, najbolje je da veličina uzorka bude negde između ekstrema.

Za ovaj rad izabran je uzorak od 100 preduzeća od čega polovinu čine preduzeća koja su otišla u stečaj, a drugu polovinu čine preduzeća koja nisu otišla u stečaj. Bilo bi poželjno da je posmatrani uzorak veći, ali je došlo do ograničenja u smislu nepotpunih finansijskih podataka

za preduzeća koja su ušla u stečaj. Svakako bi bilo dobro da se u budućnosti napravi korekcija ovog modela sa većom veličinom uzorka.

3.2 Izbor početnih varijabli za model

Kako se ovaj rad bavi isključivo srpskim preduzećima, sve varijable koje se koriste za analizu se izračunavaju na osnovu finansijskih izveštaja koji se zasnivaju na Zakonu o računovodstvu i reviziji ("Službeni glasnik RS", br. 46/06 i 111/09).

Po uzoru na dosadašnje modele za predviđanje bankrota, za razvoj modela se koriste podaci iz završnih finansijskih izveštaja godine koja prethodi otvaranju postupka. Pošto pretpostavka o dostupnosti finansijskih izveštaja na samom kraju svake godine nije tačna (finansijski izveštaji moraju da prođu proces revizije pre nego što postanu javno dostupni), posmatraće se ona preduzeća koja su ušla u stečajni postupak zaključno sa 31.12.2015. godine.

Početne varijable za model su one koje su dobijene empirijskim i teorijskim pristupom u prethodnom delu ovog rada:

Varijabla	Uticaj na verovatnoću bankrota	Izbor varijable
	↑ - povećava verovatnoću	Teorijski pristup (T)
	↓ - smanjuje verovatnoću	Empirijski pristup (E)
Neto prinos na aktivu	↓	E
Pokazatelj tekuće likvidnosti	↓	E
Koeficijent obrta kapitala	↓	E
Neraspoređeni dobitak/ukupna aktiva	↓	E
Poslovni prihodi/ukupni dug	↓	E
Neto marža	↓	E
Racio duga i kapitala	↑	E
Koeficijent zaduženosti	↑	E
Logaritam vrednosti obrtne imovine	↓	T
Priliv gotovine iz poslovnih aktivnosti	↓	T
Logaritam vrednosti rashoda od kamata	↑	T

Sve izabrane varijable su neprekidnog tipa.

Kako bi se došlo do vrednosti nabrojanih varijabli, bilo je potrebno preuzeti sledeće finansijske pokazatelje iz završnih finansijskih izveštaja:

- Ukupna aktiva
- Poslovni prihodi
- Neto rezultat
- Obrtna imovina
- Neraspoređeni dobitak
- Kratkoročne obaveze
- Dugoročne obaveze
- Kapital
- Obrtna imovina
- Potraživanja po osnovu prodaje
- Finansijski rashod

Kao što je ranije rečeno, izabrano je 100 preduzeća na osnovu kojih će se sprovoditi analiza. Polovina ovih ispitanika čine preduzeća koja su otišla u stečaj zaključno sa 31.12.2015.godine. Za ovu grupu, podaci su preuzeti iz završnih finansijskih izveštaja godine koja prethodi otvaranju postupka. Druga polovina ispitanika čine ona preduzeća koja nisu otišla u stečaj, a izabrana su uparivanjem sa prvom grupom ispitanika na osnovu veličine ukupne aktive i poslovnih prihoda. Za ovu grupu ispitanika, podaci su preuzeti iz istih završnih finansijskih izveštaja kao za njihovo upareno preduzeće iz prve grupe.

3.3 Ispravnost podataka u uzorku

Pre nego što se započne sama analiza podataka, potrebno je da se uklone podaci koji mogu da poremete ispravnost krajnjeg rezultata. Uvek postoji mogućnost da prilikom unosa podataka u program dođe do greške istraživača. Neke analize su izuzetno osetljive na netipične tačke (*outliers*), odnosno na ekstremne vrednosti koje su van opsega mogućih vrednosti te promenljive (vidi u [23]). Sa druge strane, moguće je da postoje jake veze između nezavisnih promenljivih što takođe može loše uticati na rezultat.

3.3.1 Provera ispravnosti podataka kategorijalnih promenljivih

Za kategorijalne promenljive, koristi se funkcija *Descriptive Statistics/Frequencies* kako bi se videlo da li su svi podaci unutar opsega mogućih vrednosti, kao i da li neki podaci eventualno nedostaju.

U ovom slučaju, nema izabranih kategorijalnih promenljivih za model.

3.3.2 Provera ispravnosti podataka neprekidnih promenljivih

Za neprekidne promenljive, koristi se funkcija *Descriptive Statistics/Descriptives*, na osnovu koje se dobijaju informacije o srednjoj vrednosti, standardnoj devijaciji, minimalnoj/maksimalnoj vrednosti promenljive, itd.

U ovom slučaju, postoji 11 početnih promenljivih i sve su neprekidnog tipa. Na osnovu vrednosti minimuma i maksimuma iz dobijenih rezultata, zaključuje se da svi podaci imaju smisla, tj. kreću se u svojim mogućim opsezima.

Međutim, za promenljive **Pokazatelj tekuće likvidnosti, Koeficijent obrta kapitala, Odnos poslovnih prihoda i ukupnog duga i Racio duga i kapitala** primećuje se da iznos srednje vrednosti nije u očekivanim intervalima, te se za ove promenljive proverava postojanje netipičnih tačaka (*outliers*).

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Neto prinos na aktivu	100	-9,77	,65	-,2242	1,26829
Pokazatelj tekuće likvidnosti	100	,00	2616,00	34,9250	262,07138
Koeficijent obrta kapitala	100	,00	13273,00	465,2220	2018,37376
Nerasporedjeni dobitak/ukupna aktiva	100	,00	1,04	,2625	,33120
Poslovni prihodi/ukupan dug	100	,00	115,26	3,7942	12,65346
Neto marža	100	-60,90	42,36	-1,9615	10,15910
Racio duga i kapitala	100	,00	158330,00	3977,0640	22489,52684
Koeficijent zaduženosti	100	,00	15,56	,9438	1,78014
Logaritam vrednosti obrtne imovine	100	4,66	9,62	7,6933	,87372
Priliv gotovine po osnovu poslovnih aktivnosti	100	,00	1059382000,00	78630780,0000	184793517,92370
Logaritam vrednosti rashoda od kamata	100	-1,00	9,20	4,7305	3,12113
Valid N (listwise)	100				

Otkrivanje netipičnih tačaka može da se izvrši, između ostalog, crtanjem histograma, dijagrama i drugih grafičkih prikaza. Za svaku netipičnu tačku koja može značajno da utiče na krajnji rezultat, naći će se adekvatna zamena u smislu uključivanja u model novog preduzeća koje ispunjava potrebne uslove, kako bi analiza bila što uspešnija i verodostojnija.

Koristiće se funkcija *Descriptive Statistics/Explore*.

Descriptives

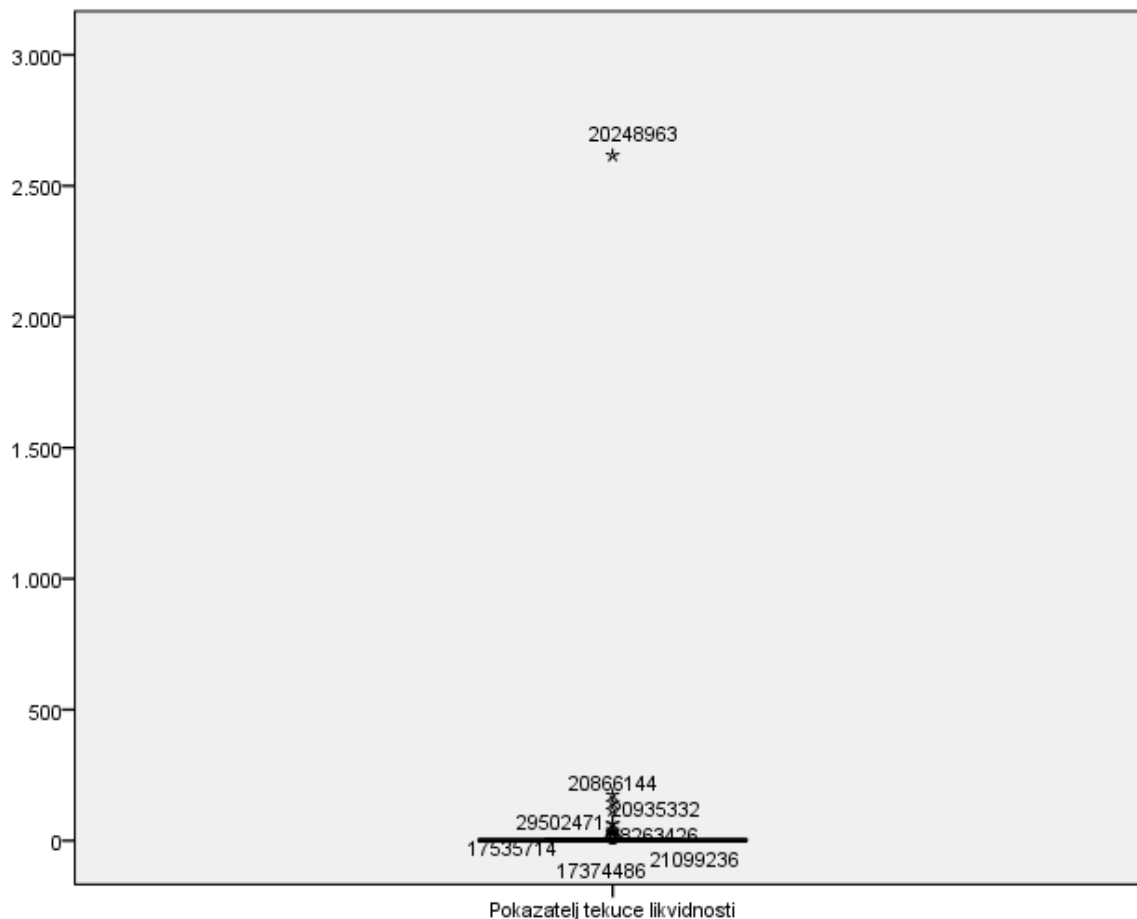
			Statistic	Std. Error
Pokazatelj tekuce likvidnosti	Mean		34,9250	26,20714
	95% Confidence Interval for Mean	Lower Bound	-17,0756	
		Upper Bound	86,9256	
	5% Trimmed Mean		4,1933	
	Median		1,1500	
	Variance		68681,408	
	Std. Deviation		262,07138	
	Minimum		,00	
	Maximum		2616,00	
	Range		2616,00	
	Interquartile Range		3,78	
	Skewness		9,847	,241
	Kurtosis		97,883	,478
Koeficijent obrta kapitala	Mean		465,2220	201,83738
	95% Confidence Interval for Mean	Lower Bound	64,7329	
		Upper Bound	865,7111	
	5% Trimmed Mean		71,0800	
	Median		,2000	
	Variance		4073832,627	
	Std. Deviation		2018,37376	
	Minimum		,00	
	Maximum		13273,00	
	Range		13273,00	
	Interquartile Range		4,23	
	Skewness		5,541	,241
	Kurtosis		32,338	,478
Poslovni prihodi/ukupan dug	Mean		3,7942	1,26535
	95% Confidence Interval for Mean	Lower Bound	1,2835	
		Upper Bound	6,3049	
	5% Trimmed Mean		1,8439	
	Median		,2550	
	Variance		160,110	
	Std. Deviation		12,65346	
	Minimum		,00	
	Maximum		115,26	
	Range		115,26	
	Interquartile Range		3,31	
	Skewness		7,392	,241
	Kurtosis		62,517	,478
Racio duga i kapitala	Mean		3977,0640	2248,95268
	95% Confidence Interval for Mean	Lower Bound	-485,3460	
		Upper Bound	8439,4740	
	5% Trimmed Mean		156,5200	
	Median		1,1000	
	Variance		505778817,435	
	Std. Deviation		22489,52684	
	Minimum		,00	
	Maximum		158330,00	
	Range		158330,00	
	Interquartile Range		16,80	
	Skewness		6,576	,241
	Kurtosis		43,353	,478

Informacije u rezultujućoj tabeli **Descriptives** pokazuju koliki problem predstavljaju slučajevi sa netipičnim tačkama. Pojam **5 % Trimmed Mean** je broj koji se dobija zanemarivanjem gornjih i donjih 5 % slučajeva i bez njih se izračunava nova srednja vrednost (kao u [23]). Upoređivanjem prvobitne i nove srednje vrednosti, dolazi se do zaključka da li ekstremne vrednosti mnogo utiču na srednju vrednost ili ne. Posmatrajući srednju vrednost izračunatu bez gornjih i donjih 5% slučajeva (*Trimmed Mean*) i “pravu” srednju vrednost, primećuje se da vrednosti za sve četiri izdvojene varijable nisu naročito bliske te bi trebalo ukloniti one vrednosti koje su problematične za dalju analizu.

Lociranje podataka koji treba da se uklone iz modela, odnosno koji treba da se zamene drugim vrednostima koje će ući u uobičajen interval, najlakše je izvršiti crtanjem i analiziranjem dijagrama ili drugih tipova grafičkih prikaza.

Pokazatelj tekuće likvidnosti

Kada je reč o varijabli **Pokazatelj tekuće likvidnosti**, na osnovu dobijenog dijagrama, vidi se da je podatak koji je potrebno zameniti boljim onaj koji se odnosi na preduzeće sa matičnim brojem 20248963.



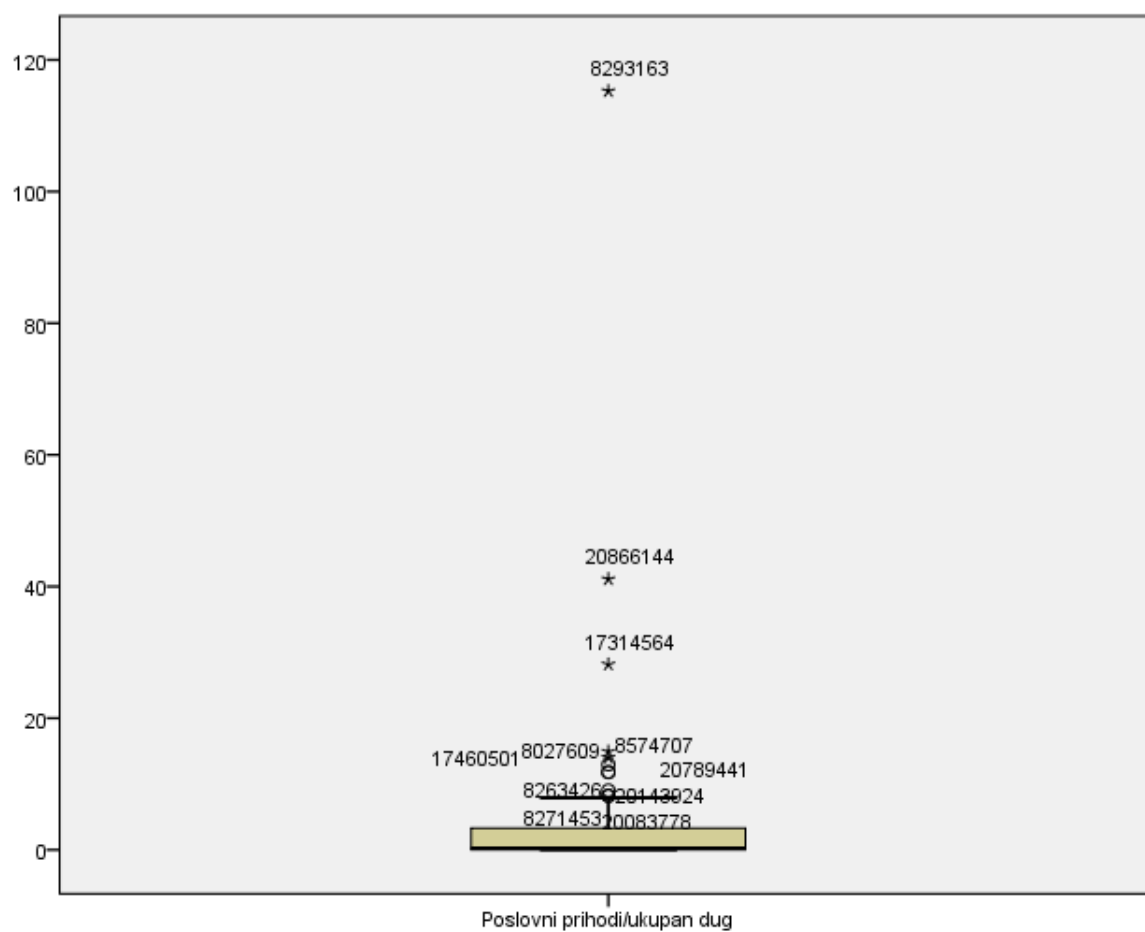
Koeficijent obrta kapitala

Analogno prethodnoj analizi, kada se radi o varijabli **Koeficijent obrta kapitala**, podaci koje je potrebno zameniti boljim su oni koji se odnose na preduzeća sa matičnim brojevima 20545658 i 06589553.



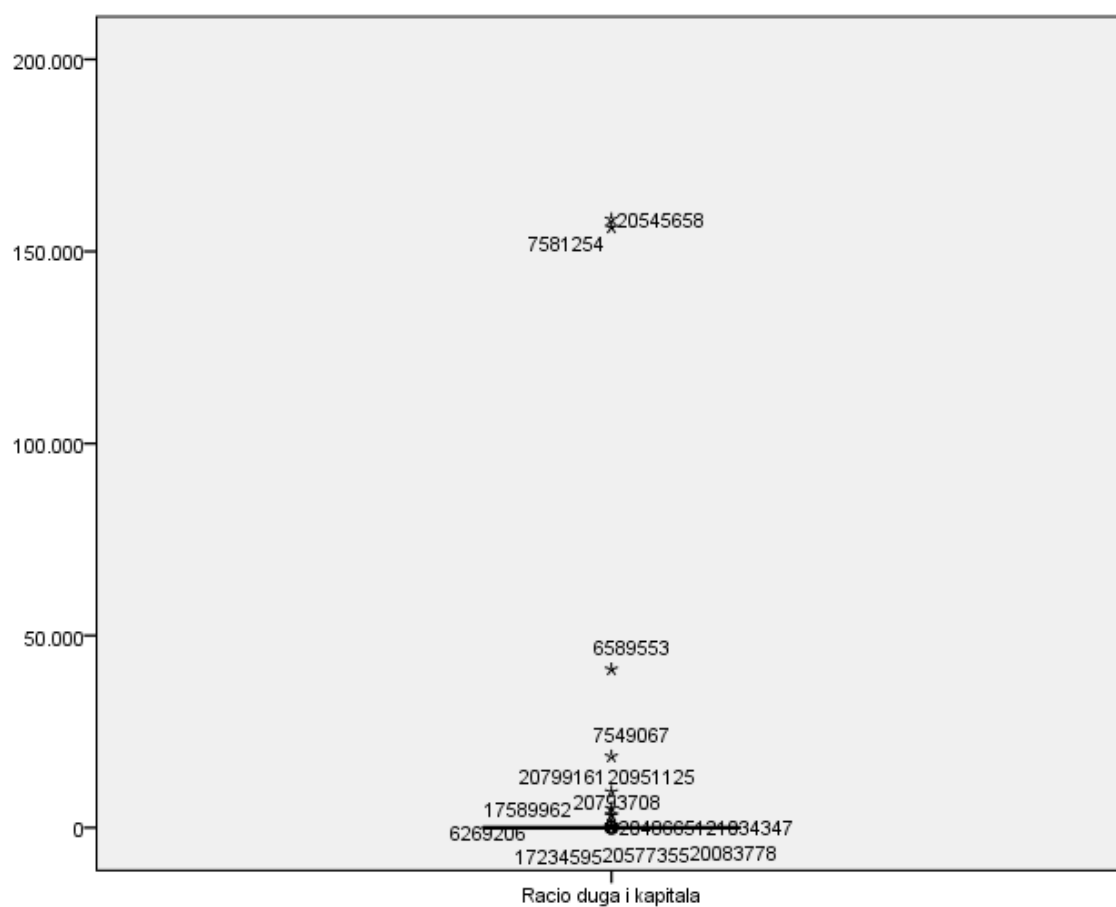
Odnos poslovnih prihoda i ukupnog duga

Istom analizom, za varijablu **Odnos poslovnih prihoda i ukupnog duga** potrebno je zameniti podatak koji se odnosi na preduzeće sa matičnim brojem 08293163.



Racio duga i kapitala

I na kraju, za varijablu **Racio duga i kapitala** zaključuje se da je potrebno zameniti podatke koji se odnose na preduzeća sa matičnim brojevima 20545658 i 07581254.



Nakon što su ekstremne vrednosti (netipične tačke) identifikovane i zamenjene boljim podacima, još jednom je pokrenuta funkcija *Descriptive Statistics/Descriptives* na osnovu koje se primećuje da se iznos očekivane srednje vrednosti popravila u odnosu na prethodnu analizu.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Neto prinos na aktivu	100	-9,77	,65	-,2106	1,26233
Pokazatelj tekuće likvidnosti	100	,00	388,60	12,6430	46,38366
Koeficijent obrta kapitala	100	,00	5198,00	168,5110	759,57605
Neraspoređeni dobitak/ukupna aktiva	100	,00	11,00	,3595	1,12414
Poslovni prihodi/ukupni dug	100	,00	41,11	2,6414	5,77719
Neto marža	100	-60,90	42,36	-1,9535	10,16091
Ratio duga i kapitala	100	,00	18538,00	595,9430	2396,57687
Koeficijent zaduženosti	100	,00	15,56	,9259	1,77568
Logaritam vrednosti obrtne imovine	100	4,66	9,62	7,6988	,86516
Priliv gotovine po osnovu poslovnih aktivnosti	100	,00	1059382000,00	78684320,0000	184771227,95472
Logaritam vrednosti rashoda od kamata	100	-1,00	9,20	4,6924	3,12488
Valid N (listwise)	100				

Takođe, na dijagramima koji su iscrtani funkcijom *Descriptive Statistics/Explore*, nema više toliko ekstremnih iskanja podataka do sada.

Prikupljeni podaci su sada spremni za dalju analizu.

3.4 Multikolinearnost varijabli

Prilikom svake statističke analize, potrebno je proveriti da li postoje jake međukorelacije prediktorskih (nezavisnih) promenljivih. Idealan scenario je da su prediktorske promenljive jako povezane sa zavisnom promenljivom, ali ne i međusobno.

Logistička regresija ne počiva na pretpostavkama o raspodeli rezultata merenja prediktorskih promenljivih, ali je osetljiva na visoke korelacije između prediktorskih promenljivih (multikolinearnost). Multikolinearnost postoji kada su nezavisne promenljive jako korelirane ($r \geq 0,9$) (kao u [23]).

Multikolinearnost se testira korišćenjem funkcije *Statistics/Collinearity diagnostics*.

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	,108	,336		,322	,748		
Neto prinos na aktivu	-,010	,028	-,026	-,363	,717	,665	1,504
Pokazatelj tekuće likvidnosti	,000	,001	-,012	-,107	,915	,268	3,735
Koeficijent obrta kapitala	5,860E-006	,000	,009	,099	,922	,419	2,387
Neraspoređeni dobitak/ukupna aktiva	-,025	,049	-,056	-,513	,610	,281	3,559
Poslovni prihodi/ukupan dug	-,004	,006	-,045	-,634	,528	,684	1,462
Neto marža	-,001	,003	-,020	-,330	,743	,878	1,139
Racio duga i kapitala	1,358E-005	,000	,065	,733	,465	,432	2,313
Koeficijent zaduženosti	,010	,023	,034	,411	,682	,505	1,982
Logaritam vrednosti obrtne imovine	,000	,046	-,001	-,008	,993	,526	1,900
Priliv gotovine po osnovu poslovnih aktivnosti	2,172E-010	,000	,080	1,129	,262	,673	1,485
Logaritam vrednosti rashoda od kamata	-,014	,011	-,090	-1,270	,208	,671	1,489

a. Dependent Variable: Stečaj (1-da, 0-ne)

Veoma male vrednosti pokazatelja **Tolerance** (manje od 0,1), ukazuju na to da promenljiva ima visoke korelacije sa ostalim promenljivama u modelu. U tom slučaju se jedna od jako međukoreliranih promenljivih uklanja iz modela. U ovom slučaju, pak, nema ovako malih vrednosti pokazatelja **Tolerance**, pa u se ovom momentu neće izbacivati nijedna promenljivu iz početnog modela.

3.5 Postupak logističke regresije u SPSS-u

Nakon što su provereni osnovni uslovi i karakteristike promenljivih koji treba da budu zadovoljeni kako bi se sprovela logistička regresiju, prelazi se na samo modeliranje.

Kao što je i ranije rečeno, logistička regresija služi za ocenu koliko dobro skup prediktorskih promenljivih predviđa/objašnjava kategorijsku zavisnu promenljivu. Ovom statističkom metodom se dobija pokazatelj relativne važnosti svake nezavisne promenljive ili interakcije između njih, zbir pokazatelja tačnosti klasifikacije slučajeva na osnovu modela kao i ocena kvaliteta predviđanja rezultata (vidi u [23]).

S obzirom na to da su netipične tačke uklonjene zamenom problematičnih podataka onim koji su verodostojniji, kao i na to da nema visokih korelacija između promenljivih, sve varijable (ukupno 11) koje su prvobitno određene za model, uključene su u analizu.

Izabrana metoda za modeliranje željenog modela je logistička regresija. Ova metoda se u SPSS-u pokreće funkcijom *Analyze/Regression/Binary Logistic*. Za zavisnu varijablu bira se ona pod nazivom “stečaj” koja prima vrednost 1 ukoliko je preduzeće ušlo u stečaj, i vrednost 0 ukoliko nije. Za nezavisne varijable se bira već odabrani skup od 11 neprekidnih promenljivih.

Metod koji će se koristiti za izbor promenljivih jeste već spominjani metod “korak po korak” (Stepwise regression). SPSS Modelar nudi dve vrste metoda “korak po korak” - unapred i unazad.

Metod “korak po korak” unapred postepeno dodaje nezavisne promenljive u osnovni model koji ne sadrži nijednu nezavisnu promenljivu, već sadrži samo konstantu. **Metod “korak po korak” unazad**, sa druge strane, postepeno uklanja nezavisne promenljive iz potpunog modela koji sadrži sve nezavisne promenljive uključene u model.

U okviru obe ove metode, “korak po korak” unapred i “korak po korak” unazad, postoje tri različita tipa metoda u SPSS Modelaru: uslovni metod (**Conditional**), metod količnika verodostojnosti (**Likelihood Ratio**) i Wald metod (**Wald**).

Naredne definicije su preuzete iz [25]:

- Izbor unapred (Conditional) - Postupak za izbor varijabli za model zasnovan na značajnosti score statistike, dok se varijable ne uključuju u model na osnovu uslovne procene parametara
- Izbor unapred (Likelihood Ratio) - Postupak za izbor varijabli za model zasnovan na značajnosti score statistike, dok se varijable ne uključuju u model na osnovu testa količnika verodostojnosti
- Izbor unapred (Wald) - Postupak za izbor varijabli za model zasnovan na značajnosti score statistike, dok se varijable ne uključuju u model na osnovu Waldove statistike
- Eliminacija unazad (Conditional) - Postupak za uklanjanje varijabli iz modela zasnovan na uslovnim procenama parametara
- Eliminacija unazad (Likelihood Ratio) - Postupak za uklanjanje varijabli iz modela zasnovan na testu količnika verodostojnosti
- Eliminacija unazad (Wald) - Postupak za uklanjanje varijabli iz modela zasnovan na Waldovoj statistici

U ovom primeru, koristi se metoda “korak po korak” unazad - metod količnika verodostojnosti, u SPSS Modelaru označeno kao *Method Backward: LR*.

Kao što je navedeno u prvom delu rada, odlučujući aspekt u korišćenju logističke regresije “korak po korak” jeste izbor α nivoa za procenu važnosti promenljive. On određuje koliko promenljivih će biti uključeno u konačan model. U ovom primeru će se koristiti izbor $\alpha = 0.05$.

Što se tiče verovatnoće za unos promenljivih u model metodom “korak po korak”, kao i verovatnoće za eliminaciju promenljivih iz modela istom metodom, biraju se vrednosti 0,05 i 0,10 redom.

Nakon unetih podataka o željenim opcijama, može da se pokrene samo modeliranje u SPSS Modelaru.

3.6 Rezultati

Pokretanjem funkcije *Analyze/Regression/Binary Logistic*, kao i u većini SPSS-ovih analiza, dobija se pregršt rezultata.

U tabeli **Case Processing Summary** dobijaju se pojedinosti o veličini uzorka. Posmatrani uzorak se sastoji od 100 preduzeća od kojih je polovina ušlo u stečaj, a druga polovina nije. Preduzeća koja su ušla u stečaj dobijaju vrednost zavisne varijable 1, a ona koja nisu ušla u stečaj dobijaju vrednost zavisne varijable 0, kao što je prikazano u narednoj rezultujućoj tabeli, **Dependent Variable Encoding**.

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	100	100,0
	Missing Cases	0	,0
	Total	100	100,0
Unselected Cases		0	,0
Total		100	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
preduzece nije uslo u stečaj	0
preduzece je uslo u stečaj	1

U odeljku **Block 0** dobijaju se rezultati analize bez ijedne nezavisne promenljive koja treba da uđe u model. U tabeli **Classification Table** se vidi da je “ispravno” klasifikovanih slučajeva bilo 50%. Cilj je da se tačnost ovog predviđanja poboljša nakon unosa nezavisnih promenljivih u model.

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			Stecaj (1-da, 0-ne)		Percentage Correct
			preduzece nije uslo u stecaj	preduzece je uslo u stecaj	
Step 0	Stecaj (1-da, 0-ne)	preduzece nije uslo u stecaj	0	50	,0
		preduzece je uslo u stecaj	0	50	100,0
Overall Percentage					50,0

a. Constant is included in the model.

b. The cut value is ,500

Dalje, u odeljku **Block 1: Method = Backward Stepwise (Likelihood Ratio)** se dobijaju rezultati ispitivanja modela sa uključenim skupom prediktorskih promenljivih. U tabeli **Omnibus Tests of Model Coefficients** se vide zbirni pokazatelji performansi modela. Taj test se zove *goodness of fit* i pokazuje koliko dobro model predviđa rezultate, kao što je bilo reči u prethodnom delu rada. Naravno, ovaj skup rezultata treba da bude što značajniji, odnosno veličina *Sig.* bi trebalo da bude manja od 0,05. U ovom slučaju, značajnost iznosi 0,000 tj. $p > 0,0005$ iz čega se zaključuje da ovaj model bolje predviđa podatke od početnog modela prikazanog u odeljku **Block 0**. Pokazatelj hi-kvadrat testa u krajnjem modelu iznosi 76,277 sa 3 stepena slobode.

Block 1: Method = Backward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	83,108	11	,000
	Block	83,108	11	,000
	Model	83,108	11	,000
Step 2 ^a	Step	-,080	1	,777
	Block	83,027	10	,000
	Model	83,027	10	,000
Step 3 ^a	Step	-,153	1	,695
	Block	82,874	9	,000
	Model	82,874	9	,000
Step 4 ^a	Step	-,169	1	,681
	Block	82,705	8	,000
	Model	82,705	8	,000
Step 5 ^a	Step	-,718	1	,397
	Block	81,987	7	,000
	Model	81,987	7	,000
Step 6 ^a	Step	-,719	1	,396
	Block	81,268	6	,000
	Model	81,268	6	,000
Step 7 ^a	Step	-,916	1	,338
	Block	80,352	5	,000
	Model	80,352	5	,000
Step 8 ^a	Step	-1,985	1	,159
	Block	78,367	4	,000
	Model	78,367	4	,000
Step 9 ^a	Step	-2,090	1	,148
	Block	76,277	3	,000
	Model	76,277	3	,000

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

U tabeli **Model Summary** se vide vrednosti *Cox & Snell R Square* i *Nagelkerke R Square* koje pokazuju koliki deo zavisne promenljive model objašnjava. U ovom slučaju, za krajnje dobijeni model, te vrednosti iznose 0,534 i 0,711. Drugim rečima, skup promenljivih koje čine dobijeni model objašnjava između 53,4% i 71,1% varijanse.

Rezultati prikazani u tabeli **Hosmer and Lemeshow Test** podržavaju tvrdnju da je model dobar. Tvrdi se da je to SPSS-ov najpouzdaniji test kvaliteta predikcije modela (vidi u [23]). Prema ovom testu, model je podržan ukoliko vrednost *Sig.* iznosi više od 0,05. Pokazatelj hi-kvadrat testa za Hosmer-Lemeshow test iznosi 0,513 uz značajnost 1 i 8 stepeni slobode.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	55,522 ^a	,564	,753
2	55,602 ^a	,564	,752
3	55,756 ^a	,563	,751
4	55,925 ^b	,563	,750
5	56,642 ^b	,560	,746
6	57,362 ^b	,556	,742
7	58,278 ^b	,552	,736
8	60,262 ^b	,543	,724
9	62,353 ^b	,534	,711

a. Estimation terminated at iteration number 15 because parameter estimates changed by less than ,001.

b. Estimation terminated at iteration number 14 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	2,706	8	,951
2	6,119	8	,634
3	6,116	8	,634
4	6,829	8	,555
5	5,945	8	,653
6	7,269	8	,508
7	2,877	8	,942
8	,396	8	1,000
9	,513	8	1,000

U tabeli **Classification Table** su prikazani pokazatelji koliko dobro model predviđa kategoriju (preduzeće je otišlo u stečaj/preduzeće nije otišlo u stečaj) za svaki ispitivani slučaj, tj. svaki pojedinačni korak u regresiji. Na osnovu rezultata, krajnji model ispravno klasifikuje 85% svih slučajeva, što je neuporedivo bolje od početnih 50% navedenih u **Block 0**.

U prethodnom delu rada je navedeno da su dva važna pojma kada se govori o tabelama klasifikacija **senzitivnost** i **specifičnost**. Senzitivnost modela je procentualni udeo grupe sa

ispitivanim obeležjem koji je model tačno prepoznao. U ovom slučaju, model je tačno klasifikovao 84% preduzeća koja su otišla u stečaj. Specifičnost modela je procentualni udeo grupe koja nema ispitivano obeležje i koji je model tačno prepoznao. U posmatranom primeru, specifičnost iznosi 86% preduzeća koja nisu otišla u stečaj.

Classification Table^a

Observed			Predicted		
			Stečaj (1-da, 0-ne)		Percentage Correct
			preduzece nije uslo u stečaj	preduzece je uslo u stečaj	
Step 1	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	46	4	92,0
		preduzece je uslo u stečaj	6	44	88,0
	Overall Percentage				90,0
Step 2	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	46	4	92,0
		preduzece je uslo u stečaj	6	44	88,0
	Overall Percentage				90,0
Step 3	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	47	3	94,0
		preduzece je uslo u stečaj	7	43	86,0
	Overall Percentage				90,0
Step 4	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	46	4	92,0
		preduzece je uslo u stečaj	6	44	88,0
	Overall Percentage				90,0
Step 5	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	44	6	88,0
		preduzece je uslo u stečaj	7	43	86,0
	Overall Percentage				87,0
Step 6	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	44	6	88,0
		preduzece je uslo u stečaj	7	43	86,0
	Overall Percentage				87,0
Step 7	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	44	6	88,0
		preduzece je uslo u stečaj	6	44	88,0
	Overall Percentage				88,0
Step 8	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	43	7	86,0
		preduzece je uslo u stečaj	8	42	84,0
	Overall Percentage				85,0
Step 9	Stečaj (1-da, 0-ne)	preduzece nije uslo u stečaj	43	7	86,0
		preduzece je uslo u stečaj	8	42	84,0
	Overall Percentage				85,0

a. The cut value is ,500

Tabela **Variables in the Equation** je najvažnija tabela u celokupnoj analizi logističkom regresijom jer daje krajnji izgled traženog modela. Ona nas informiše o doprinosu ili važnosti svake prediktorske promenljive. U prethodnom delu rada je detaljnije opisan postupak verifikacije značajnosti svake promenljive koja je ušla u model i naglašeno je da je potrebno za svaku od njih ispitati Wald statistiku.

Obratiće se najpre pažnja na kolonu *Sig.* U krajnjem modelu se pojavljuju samo one vrednosti koje su manje od 0,05. To znači da su odabrane promenljive koje značajno doprinose prediktivnim mogućnostima modela. Takođe, može se primetiti da se u svakom koraku isključuju one promenljive koje imaju najveću vrednost *Sig.*

Koeficijenti B zapravo predstavljaju koeficijente koji ulaze u krajnju jednačinu za izračunavanje verovatnoće da analizirani slučaj spada u određenu kategoriju (preduzeće otišlo u stečaj).

U posmatranom primeru, glavni faktori koji utiču na to da li će preduzeće otići u stečaj ili ne jesu: **neto prinos na aktivu, pokazatelj tekuće likvidnosti i racio duga i kapitala**. Ostale promenljive nisu značajno doprinele modelu.

Obratiće se pažnja na smer veze. Negativne vrednosti koeficijenta B pokazuju da povećanje vrednosti nezavisne promenljive ima za posledicu smanjenje verovatnoće da preduzeće ode u stečaj, a pozitivne vrednosti koeficijenta B imaju za posledicu povećanje verovatnoće da preduzeće ode u stečaj. U ovom primeru, racio duga i kapitala ima pozitivan koeficijent B, $B = 0,14$, što nam govori da za preduzeća koja imaju veći racio duga i kapitala postoji veća verovatnoća da preduzeće uđe u stečaj. Za druge dve promenljive koje su ušle u model, neto prinos na aktivu i pokazatelj tekuće likvidnosti, koeficijenti B su negativni, redom $B = -10,956$ i $B = -0,461$. To znači da za preduzeća koja imaju veći neto prinos na aktivu i pokazatelj tekuće likvidnosti, manje je verovatno da preduzeće ode u stečaj.

Smerovi veze dobijeni u konačnom modelu se poklapaju sa početnim pretpostavkama o uticaju promenljivih na verovatnoću bankrota, koje su donete na osnovu osnovnih ekonomskih zakona, te se zaključuje da je ovaj model dobar i sa empirijskog aspekta:

Varijabla	Pretpostavka o uticaju na verovatnoću bankrota	
	↑ - povećava verovatnoću	B koeficijent
	↓ - smanjuje verovatnoću	
Neto prinos na aktivu	↓	-
Pokazatelj tekuće likvidnosti	↓	-
Racio duga i kapitala	↑	+

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Neto_prinos_na_aktivu	-14,804	7,823	3,581	1	,058	,000	,000	1,697
	Pokazatelj_tekuce_likvidnosti	-,578	,298	3,759	1	,053	,561	,313	1,006
	Koeficijent_obrta_kapitala	,046	,044	1,060	1	,303	1,047	,960	1,141
	Odnos_nerasporedjenog_dobitka_i_ukupne_aktivne	-,541	1,324	,167	1	,683	,582	,043	7,791
	Odnos_poslovnih_prihoda_i_ukupnog_duga	-,137	,165	,693	1	,405	,872	,632	1,204
	Neto_marza	,024	,057	,180	1	,672	1,024	,916	1,146
	Racio_duga_i_kapitala	,058	,071	,660	1	,417	1,059	,922	1,217
	Logaritam_vrednosti_obrtne_imovine	,504	,553	,830	1	,362	1,655	,560	4,895
	Priliv_gotovine_po_osnovu_poslovnih_aktivnosti	,000	,000	,078	1	,780	1,000	1,000	1,000
	Koeficijent_zaduzenosti	,440	,415	1,122	1	,289	1,553	,688	3,504
	Logaritam_vrednosti_rashoda_od_kamata	-,311	,164	3,595	1	,058	,733	,532	1,011
	Constant	-1,801	3,713	,235	1	,628	,165		
	Step 2 ^a								
Step 2 ^a	Neto_prinos_na_aktivu	-14,940	7,869	3,605	1	,058	,000	,000	1,620
	Pokazatelj_tekuce_likvidnosti	-,596	,293	4,132	1	,042	,551	,310	,979
	Koeficijent_obrta_kapitala	,045	,044	1,041	1	,308	1,046	,959	1,140
	Odnos_nerasporedjenog_dobitka_i_ukupne_aktivne	-,516	1,320	,153	1	,696	,597	,045	7,937
	Odnos_poslovnih_prihoda_i_ukupnog_duga	-,127	,158	,640	1	,424	,881	,646	1,202
	Neto_marza	,024	,057	,174	1	,677	1,024	,916	1,145
	Racio_duga_i_kapitala	,058	,070	,674	1	,412	1,059	,923	1,215
	Logaritam_vrednosti_obrtne_imovine	,586	,468	1,570	1	,210	1,798	,718	4,499
	Koeficijent_zaduzenosti	,433	,415	1,088	1	,297	1,542	,684	3,478
	Logaritam_vrednosti_rashoda_od_kamata	-,311	,165	3,562	1	,059	,733	,531	1,012
	Constant	-2,362	3,111	,576	1	,448	,094		
	Step 3 ^a								
	Neto_prinos_na_aktivu	-15,345	7,843	3,828	1	,050	,000	,000	1,028
Step 3 ^a	Pokazatelj_tekuce_likvidnosti	-,625	,293	4,534	1	,033	,535	,301	,952
	Koeficijent_obrta_kapitala	,046	,045	1,086	1	,297	1,047	,960	1,143
	Odnos_poslovnih_prihoda_i_ukupnog_duga	-,121	,159	,572	1	,449	,886	,649	1,211
	Neto_marza	,025	,058	,182	1	,670	1,025	,915	1,149
	Racio_duga_i_kapitala	,060	,071	,707	1	,401	1,062	,923	1,222
	Logaritam_vrednosti_obrtne_imovine	,620	,458	1,828	1	,176	1,859	,757	4,564

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 4 ^a	Koeficijent_zaduzenosti	,454	,413	1,207	1	,272	1,574	,701	3,535
	Logaritam_vrednosti_rashoda_od_kamata	-,319	,165	3,739	1	,053	,727	,526	1,004
	Constant	-2,678	2,978	,808	1	,369	,069		
	Neto_prinos_na_aktivu	-13,926	6,896	4,078	1	,043	,000	,000	,664
	Pokazatelj_tekuce_likvidnosti	-,617	,294	4,391	1	,036	,540	,303	,961
	Koeficijent_obrta_kapitala	,045	,045	1,025	1	,311	1,046	,958	1,142
	Odnos_poslovnih_prihoda_i_ukupnog_duga	-,124	,160	,597	1	,440	,884	,646	1,209
	Racio_duga_i_kapitala	,060	,076	,622	1	,430	1,062	,915	1,233
	Logaritam_vrednosti_obrtne_imovine	,613	,457	1,800	1	,180	1,846	,754	4,521
	Koeficijent_zaduzenosti	,430	,409	1,106	1	,293	1,538	,690	3,428
Step 5 ^a	Logaritam_vrednosti_rashoda_od_kamata	-,327	,165	3,922	1	,048	,721	,522	,997
	Constant	-2,607	2,955	,778	1	,378	,074		
	Neto_prinos_na_aktivu	-14,740	6,717	4,816	1	,028	,000	,000	,207
	Pokazatelj_tekuce_likvidnosti	-,650	,296	4,825	1	,028	,522	,292	,932
	Koeficijent_obrta_kapitala	,037	,044	,712	1	,399	1,038	,952	1,132
	Racio_duga_i_kapitala	,066	,080	,693	1	,405	1,068	,914	1,249
	Logaritam_vrednosti_obrtne_imovine	,639	,450	2,016	1	,156	1,894	,784	4,575
	Koeficijent_zaduzenosti	,470	,407	1,332	1	,248	1,600	,720	3,553
	Logaritam_vrednosti_rashoda_od_kamata	-,305	,156	3,829	1	,050	,737	,543	1,000
	Constant	-3,034	2,898	1,096	1	,295	,048		
Step 6 ^a	Neto_prinos_na_aktivu	-12,706	5,987	4,504	1	,034	,000	,000	,378
	Pokazatelj_tekuce_likvidnosti	-,572	,262	4,757	1	,029	,565	,338	,944
	Racio_duga_i_kapitala	,116	,077	2,276	1	,131	1,123	,966	1,305
	Logaritam_vrednosti_obrtne_imovine	,589	,440	1,793	1	,181	1,801	,761	4,263
	Koeficijent_zaduzenosti	,395	,397	,985	1	,321	1,484	,681	3,233
	Logaritam_vrednosti_rashoda_od_kamata	-,276	,148	3,469	1	,063	,759	,568	1,015
	Constant	-2,899	2,868	1,022	1	,312	,055		
	Neto_prinos_na_aktivu	-11,254	5,385	4,368	1	,037	,000	,000	,497
	Pokazatelj_tekuce_likvidnosti	-,625	,260	5,786	1	,016	,535	,322	,891
	Racio_duga_i_kapitala	,134	,077	3,025	1	,082	1,144	,983	1,330
Step 7 ^a	Logaritam_vrednosti_obrtne_imovine	,607	,438	1,924	1	,165	1,835	,778	4,326

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 8 ^a	Logaritam_vrednosti_rashoda_od_kamata	-,273	,148	3,406	1	,065	,761	,570	1,017
	Constant	-2,782	2,836	,962	1	,327	,062		
	Neto_prinos_na_aktivu	-10,611	5,086	4,353	1	,037	,000	,000	,526
	Pokazatelj_tekuce_likvidnosti	-,522	,224	5,446	1	,020	,594	,383	,920
	Racio_duga_i_kapitala	,163	,076	4,619	1	,032	1,177	1,014	1,365
	Logaritam_vrednosti_rashoda_od_kamata	-,162	,115	1,972	1	,160	,851	,679	1,066
Step 9 ^a	Constant	1,050	,793	1,754	1	,185	2,857		
	Neto_prinos_na_aktivu	-10,956	5,051	4,704	1	,030	,000	,000	,348
	Pokazatelj_tekuce_likvidnosti	-,461	,205	5,068	1	,024	,630	,422	,942
	Racio_duga_i_kapitala	,140	,069	4,138	1	,042	1,151	1,005	1,317
	Constant	,205	,479	,184	1	,668	1,228		

a. Variable(s) entered on step 1: Neto_prinos_na_aktivu, Pokazatelj_tekuce_likvidnosti, Koeficijent_obrta_kapitala, Odnos_nerasporedjenog_dobitka_i_ukupne_aktive, Odnos_poslovnih_prihoda_i_ukupnog_duga, Neto_marza, Racio_duga_i_kapitala, Logaritam_vrednosti_obrtne_imovine, Priliv_gotovine_po_osnovu_poslovnih_aktivnosti, Koeficijent_zaduzenosti, Logaritam_vrednosti_rashoda_od_kamata.

Još jedan koristan podatak u tabeli **Variables in the Equation** jeste kolona *Exp(B)*. Ove vrednosti predstavljaju odnose šansi (*odds ratios*) za svaku nezavisnu promenljivu. Odnos šansi je promena verovatnoće pripadanja jednoj kategoriji ishoda kada se vrednost određenog prediktora poveća za jednu jedinicu (vidi u [24]). Za svaki odnos šansi u koloni *Exp(B)* prikazan je i 95-procentni interval poverenja, *95 % C.I. for EXP(B)*, u smislu donje i gornje granice intervala. To predstavlja opseg za koji sa sigurnošću od 95% može da se tvrdi da obuhvata stvarnu vrednost odnosa šansi. Interval poverenja se menja u zavisnosti od veličine uzorka. Mali uzorak ima za posledicu širok interval poverenja, dok se povećanjem uzorka ovaj interval smanjuje. Kako je uzorak relativno mali, može da se primeti da je i interval poverenja prilično širok.

Poslednja tabela rezultata, **Casewise List**, govori o slučajevima u uzorku koje model ne predviđa dobro. Prikazani su oni slučajevi sa *ZResid* većim od 2. Iz informacija koje se dobijaju u tabeli, vidi se da je za slučaj pod rednim brojem 18 bilo predviđeno da će ući u stečaj (Predicted Group = 1), a zapravo nije otišlo u stečaj (Observed = 0).

U situacijama kada postoji veći broj ovakvih slučajeva, stvara se potreba za uključivanjem novih varijabli u model. U ovom primeru, samo jedno preduzeće (pod rednim brojem 18) iskače iz opsega predviđanja modela i ne uklapa se u opšti obrazac, što znači da je sam model relativno dobar i stabilan. U praksi je praktično nemoguće da se dobije idealno poklapanje između očekivanih i stvarnih vrednosti za veći broj posmatranih slučajeva.

Casewise List^b

Case	Selected Status ^a	Observed	Predicted	Predicted Group	Temporary Variable	
		Stecaj (1-da, 0-ne)			Resid	ZResid
18	S	0**	,900	1	-,900	-3,008

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2,000 are listed.

3.7 Konačan model

U tabeli **Variables in the Equation** se vidi koliko svaka promenljiva doprinosi modelu. Korišćena je metoda izbora “korak po korak” unazad, tako da se iz početnog modela u koji su ubačene sve posmatrane promenljive u svakom koraku izbacuje ona promenljiva koja ima najmanju značajnost. U poslednjem koraku su ostale tri promenljive, i one čine konačan model: **neto prinos na aktivu, pokazatelj tekuće likvidnosti i ratio duga i kapitala**.

Njihove odgovarajuće vrednosti B pokazuju koliko je svaka od njih značajna. On pokazuje njihov jedinstven doprinos, nakon što se statistički uklone uticaji preklapanja sa drugim promenljivama (vidi u [23]). Po redu važnosti, to su: **neto prinos na aktivu** ($B = -10,956$), **pokazatelj tekuće likvidnosti** ($B = -0,461$) i **ratio duga i kapitala** ($B = 0,14$).

Sada se zaključuje da jednačina za predviđanje verovatnoće bankrota srpskih privatnih preduzeća dobija sledeći oblik:

$$\log\left(\frac{p}{1-p}\right) = 0,205 - 10,956X_1 - 0,461X_2 + 0,14X_3$$

gde je:

X_1 - **neto prinos na aktivu**

X_2 - **pokazatelj tekuće likvidnosti**

X_3 - **ratio duga i kapitala**

Drugim zapisom, verovatnoća da srpsko privatno preduzeće ode u stečaj iznosi:

$$p = \frac{e^{0,205-10,956X_1-0,461X_2+0,14X_3}}{1 + e^{0,205-10,956X_1-0,461X_2+0,14X_3}}$$

gde je:

X_1 - **neto prinos na aktivu**

X_2 - **pokazatelj tekuće likvidnosti**

X_3 - **ratio duga i kapitala**

Glava 4

Zaključak

Ovaj rad se bavio predviđanjem verovatnoće bankrota srpskih privatnih preduzeća, upotrebom logističke regresije.

Cilj rada je, između ostalog, bio da odgovori na sledeća pitanja:

- Koje finansijske pokazatelje treba koristiti prilikom modeliranja verovatnoće bankrota srpskih preduzeća?
- Da li je logistička regresija odgovarajući model za predviđanje bankrota u Srbiji?
- Da li je dobijeni model zadovoljavajuć?

Na osnovu modela dobijenog upotrebom logističke regresije, glavni faktori koji utiču na verovatnoću bankrota srpskih privatnih preduzeća, po redu važnosti, su: neto prinos na aktivu, pokazatelj tekuće likvidnosti i racio duga i kapitala.

Neto prinos na aktivu meri poslovnu efikasnost preduzeća, pokazatelj tekuće likvidnosti nam prikazuje sposobnost preduzeća da servisira svoje kratkoročne obaveze, dok racio duga i kapitala predstavlja izloženost preduzeća finansijskom riziku. Sve tri promenljive upadaju u kategoriju racio pokazatelja i računaju se na osnovu sledećih finansijskih podataka: neto rezultat, ukupna aktiva, obrtna imovina, kratkoročne obaveze, ukupan dug i kapital.

Neto prinos na aktivu spada u kategoriju **pokazatelja profitabilnosti**, pokazatelj tekuće likvidnosti u kategoriju **pokazatelja likvidnosti** a racio duga i kapitala u kategoriju **pokazatelja solventnosti**.

Iz svega navedenog, zaključuje se da model obuhvata širok opseg finansijskih pokazatelja i njihovih kategorija, i da je logistička regresija pravi model za predviđanje bankrota srpskih privatnih preduzeća.

Uzorak na osnovu kojeg je praviljen model se sastoji od 100 preduzeća. Polovinu uzorka čine preduzeća koja su ušla u stečaj zaključno sa 31.12.2015.godine, a drugu polovinu preduzeća koja nisu ušla u stečaj, uparivanjem prema veličini ukupne aktive i poslovnih prihoda. Korišćeni su finansijski izveštaji godine koja je prethodila ulasku preduzeća u stečaj. Utvrđeno je da postoje značajne razlike između preduzeća koja su otišla u stečaj i onih koja su nastavila zdravo da posluju, odnosno model može uspešno da se formira.

Na osnovu dobijenih rezultata, krajnji model ispravno klasifikuje 85% svih slučajeva, odnosno model je tačno klasifikovao 84% preduzeća koja su otišla u stečaj i 86% preduzeća koja nisu otišla u stečaj. Iz tog razloga, sa zadovoljstvom mogu da odgovorim na poslednje pitanje, i da zaključim da je dobijeni model zadovoljavajuć.

Bibliografija

- [1] Zakon o stečaju ("Sl. glasnik RS", br. 104/2009, 99/2011 - dr.zakon, 71/2012 - odluka US i 3/2014)
- [2] Zakon o privrednim društvima ("Sl. glasnik RS", br. 125/2004)
- [3] Riku Saastamoinen, *Predicting bankruptcy of Finnish limited liability companies from historical financial statements*, 2015.
- [4] Altman, Edward I., *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, 1968.
- [5] James A. Ohlson., *Financial Ratios and the Probabilistic Prediction of Bankruptcy*, 1980.
- [6] Bellovary, J., Giacomino, D., Akers, M., *A Review of bankruptcy Prediction Studies: 1930 to Present*, 2007.
- [7] [https://economy.rs/Ekonomija i privreda](https://economy.rs/Ekonomija%20i%20privreda)
- [8] DeAngelo, H., DeAngelo, L.E., Stulz, R., *Dividend policy and the earned/contributed capital ix: a test of the life - cycle theory*, 2006.
- [9] Modigliani, F., Miller, M., *The Cost of Capital, Corporation Finance and the Theory of Investment*, 1958.
- [10] Karels, G.V., Prakash, A.J., *Multivariate Normality and Forecasting of Business Failure*, 1987.
- [11] Lim, T.C., *Bankruptcy Prediction: Theoretical Framework Proposal*, 2012.
- [12] Walter, J.E., *Determination of Technical Solvency. The Journal of Business* 30, 30-43, 1957.
- [13] Beaver, W., *Financial ratios as predictors of failure*, 1966.
- [14] Zagorka Lozanov-Crvenković, *Statistika*, 2012.
- [15] Gorica Gvozdić, *Primenjena logistička regresija, master rad*, 2011.

- [16] Thomas Robert Malthus, *An Essay on the Principle of Population*, 1798.
- [17] Pierre-Francois Verhulst, *Instructions on the law that the population follows in its growth*, 1838.
- [18] Jun Han, Claudio Morag, *The influence of the sigmoid function parameters on the speed of backpropagation learning*, 1995.
- [19] Hosmer W. David, Lemeshow Stanley, *Applied Logistic Regression-second edition*, 2000.
- [20] Engle, Robert F., *Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics*, 1983.
- [21] Fox J., *Applied regression analysis, linear models, and related methods*, 1997.
- [22] Johnston J., DiNardo J., *Econometric Methods Fourth Edition*, 1997.
- [23] Julie Pallant, *SPSS, Priručnik za preživljavanje*, 2011.
- [24] Barbara G. Tabachnick, Linda S. Fidell, *Using Multivariate Statistics*, 2007.
- [25] *IBM SPSS Guide*
- [26] Dragan Manasijević *Statistička analiza u SPSS programu*, 2011.
- [27] Ekonomski fakultet u Kragujevcu, *Osnovi statistike*, udžbenik - odabrana poglavlja
- [28] Frank E. Harrell Jr., *Regression Modeling Strategies*, 2001.
- [29] Pohar Maja, Blas Mateja, Turk Sandra, *Comparison Logistic Regression Models and Linear Discriminant Analysis*, 2004.
- [30] Chatterjee Samprit, Hadi S. Ali, *Regression Analysis by Example*, 2006.
- [31] Nemanja Vaci, *Poređenje metoda ocenjivanja parametara na podacima iz psiholingvističkih eksperimenata*, master rad, 2013.
- [32] Snežana Todorović, *Neke primene statistike u psihijatriji*, master rad, 2013.
- [33] Scott A. Czepiel, *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation*
- [34] Goldin Rebecca, *Odds Ratios*, 2007.
- [35] Andersen, Erling B., *Asymptotic Properties of Conditional Maximum Likelihood Estimators*, 1970.
- [36] Stevenson Mark, *An Introduction to Logistic Regression*, 2008.

- [37] Hallett C. David, *Goodness of Fit Test in Logistic Regression*, 1999.
- [38] Goran Trajković, *Performanse dijagnostičkih testova*
- [39] Menard Scott, *Applied Logistic Regression Analysis*, 2001.
- [40] Stevenson Mark, *An Introduction to Logistic Regression*, 2008.

Biografija



Marija Pešić je rođena 26.01.1991. godine u Novom Sadu. Osnovu školu “Đorđe Natošević” završila je 2005. godine. Iste godine upisala je prirodno-matematički smer Gimnazije “Isidora Sekulić” u Novom Sadu, koju je završila 2009. godine. Po završetku gimnazije upisuje Prirodno-matematički fakultet u Novom Sadu, smer primenjena matematika, modul - matematika finansija. Osnovne studije završava 2012. godine nakon čega upisuje master studije na istom fakultetu. Položila je sve ispite predviđene nastavnim planom i programom master studija 2014. godine.

2014. godine, u okviru TEMPUS projekta “Visuality and Mathematics“, provela je mesec dana u gradu Jyväskylä, u Finskoj.

U januaru 2015.godine izabrana je za projekat “Mladi lideri” kompanije Delta Holding u Beogradu, nakon kojeg se zaposlila u kompaniji Delta Agrar kao referent za procenu rizika poslovanja.

Od jula 2016.godine, zaposlena je u Raiffeisen banci u Beogradu kao analitičar podataka u okviru sektora za procenu rizika.

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:

RBR

Identifikacioni broj:

IBR

Tip dokumentacije: Monografska dokumentacija

TD

Tip zapisa: Tekstualni štampani materijal

TZ

Vrsta rada: Master rad

VR

Autor: Marija Pešić

AU

Mentor: prof. dr Zagorka Lozanov-Crvenković

MN

Naslov rada: Upotreba logističke regresije u modeliranju verovatnoće bankrota preduzeća

NR

Jezik publikacije: srpski (latinica)

JP

Jezik izvoda: s / e

JI

Zemlja publikovanja: Srbija

ZP

Uže geografsko područje: Vojvodina

UGP

Godina: 2016.

GO

Izdavač: Autorski reprint

IZ

Mesto i adresa: Novi Sad, Departman za matematiku i informatiku, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, Trg Dositeja Obradovića 4

MA

Fizički opis rada: (4, 79, 40, 4, 15, 0, 0)

(broj poglavlja/strana/lit. citata/tabela/slika/grafika/priloga)

FO

Naučna oblast: Matematika

NO

Naučna disciplina: Statističko modeliranje

ND

Predmetna odrednica / Ključne reči: statistika, logistička regresija, korak po korak analiza, Wald test, Score test, bankrot, Altman Z-score, Ohlson O-score

PO

UDK

Čuva se: Biblioteka Departmana za matematiku i informatiku Prirodno-matematičkog fakulteta Univerziteta u Novom Sadu

ČU

Važna napomena:

VN

Izvod: Ovaj rad se bavi verovatnoćom bankrota kao dela procesa procene kreditnog rizika. Glavni fokus je na predviđanju bankrota preduzeća. Istraživanje je sprovedeno koristeći statističku metodu - logističku regresiju, na osnovu istorijskih podataka iz završnih finansijskih računa srpskih preduzeća koja su otišla u stečaj, kao i njima sličnih preduzeća koja su „zdrava“, tj. nisu otišla u stečaj. Prvi deo rada odnosi se na teorijsku pozadinu pojma bankrota i na izbor preliminarnih varijabli sa teorijskog i empirijskog aspekta za model predviđanja bankrota srpskih preduzeća. Drugi deo rada predstavlja teorijski osnov logističke regresije koja je izabrana kao najreprezentativniji statistički metod za modeliranje bankrota srpskih preduzeća. U ovom delu rada se predstavlja matematička pozadina za izbor varijabli, uveden je pojam logističke regresije „korak po korak“ koja proverava značajnost promenljivih pa ih na osnovu utvrđenog pravila odlučivanja uključuje ili isključuje iz modela, opisano je ocenjivanje značajnosti promenljivih u modelu kao i testiranje slaganja logističkog regresionog metoda sa podacima. U trećem delu rada opisuje se sama primena logističke regresije prilikom modeliranja bankrota srpskih preduzeća. Korišćen je IBM SPSS Modelar, a u okviru njega procedura Binary Logistic. U ovom delu rada se prate svi koraci koji su navedeni u drugom poglavlju kao standardna procedura za sprovođenje logističke regresije. I na kraju, u četvrtom delu rada, izloženi su rezultati ovakve primene logističke regresije, i dat je odgovor na pitanje da li je dobijeni model zadovoljavajući.

IZ

Datum prihvatanja teme od strane NN veća: 14.9.2016.

DP

Datum odbrane: oktobar 2016.

DO

Članovi komisije:

KO

Predsednik: dr Ljiljana Gajić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, predsednik

Član: dr Zagorka Lozanov-Crvenković, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, mentor

Član: dr Ivana Štajner-Papuga, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, član

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
KEY WORD DOCUMENTATION

Accession number:

ANO

Identification number:

INO

Document type: Monograph documentation

DT

Type of record: Textual printed material

TR

Contents Code: Master's thesis

CC

Author: Marija Pešić

AU

Mentor: Zagorka Lozanov-Crvenković, Ph.D.

MN

Title: Application of the logistic regression for modelling bankrupt probability of companies

TI

Language of text: Serbian

LT

Language of abstract: en / s

LA

Country of publication: Serbia

CP

Locality of publication: Vojvodina

LP

Publication year: 2016.

PY

Publisher: Author's reprint

PU

Publ. place: Novi Sad, Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 4

PP

Physical description: (4, 79, 40, 4, 15, 0, 0)
(chapters/pages/literature/tables/pictures/graphics/appendices)

PD

Scientific field: Mathematics

SF

Scientific discipline: Statistical Modelling

SD

Subject / Key words: statistics, logistic regression, stepwise analysis, Wald test, Score test, bankruptcy, Altman Z-score, Ohlson O-score

SKW**UC:**

Holding data: Library of the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad

HD

Note:

N

Abstract: This thesis deals with bankrupt probability as a part of the process of credit risk assessment. The main focus is on predicting the bankruptcy of companies. The survey was conducted using a statistical method - logistic regression, based on historical data from annual financial accounts for Serbian companies that have bankrupted, and for those that are "healthy". The first part of the thesis refers to the theoretical background of the bankruptcy concept, as well as the preliminary selection of variables with the theoretical and empirical aspects of bankruptcy prediction model for Serbian companies. The second part represents the theoretical base of the logistic regression which was chosen as the most representative statistical method for modeling the bankruptcy of Serbian companies. In this part, mathematical background for the choice of variables is presented, stepwise logistic regression is introduced, the assessment of the significance of variables in the model is described, and testing the concurrence of the logistic regression method with the data is introduced. In the third part of thesis, the application of logistic regression for modeling the bankrupt probability of Serbian companies is described. IBM SPSS Modeler was used, and within it, Binary Logistic procedure. In this part of the thesis, all the steps that were listed in the second section were followed as a standard procedure for the implementation of logistic regression. Finally, in the fourth part of the thesis, the results of the application of logistic regression were presented, and the answer of the question whether the resulting model is satisfactory is given.

AB

Accepted by the Scientific Board on: 14.9.2016.

ASB

Defended: October 2016.

DE

Thesis defend board:

DB

President: Ljiljana Gajić, Ph.D., full professor, Faculty of Sciences, University of Novi Sad, president

Member: Zagorka Lozanov-Crvenković, Ph.D., full professor, Faculty of Sciences, University of Novi Sad, supervisor

Member: Ivana Štajner-Papuga, Ph.D., full professor, Faculty of Sciences, University of Novi Sad, member