

Sveučilište Josipa Jurja Strossmayera u Osijeku

Odjel za matematiku

Diplomski studij – Financijska matematika i statistika

Istraživanje o putnicima s Titanica

Seminarski rad

Student: Ime Prezime

Kolegij: Statistika

Osijek, 2020.

Sadržaj

1. Uvod.....	3
1.1. RMS Titanic.....	3
1.2. Opis varijabli.....	3
2. Analiza podataka	4
2.1. Preživljavanje	4
2.1.1. Povezanost preživljavanja s brojem članova obitelji na brodu	5
2.2. Spol	6
2.2.1. Povezanost spola s preživljavanjem	7
2.3. Klasa	8
2.3.1. Povezanost klase s preživljavanjem	9
2.4. Dob.....	10
2.4.1. Povezanost dobi s preživljavanjem.....	10
3. Zaključak.....	12
4. Literatura.....	13

1. Uvod

U ovom seminarskom radu provest ćemo statističku analizu podataka iz baze „titanic3.csv“. Baza sadrži podatke o uzorku putnika s broda RMS Titanic. Sve statističke analize rađene su u programskom jeziku R (inačica 4.0.1.).

1.1. RMS Titanic

RMS Titanic bio je britanski preoceanski brod koji je potonuo u Atlantskom oceanu 15. travnja 1912. godine nakon sudara s ledenjakom za vrijeme svog prvog putovanja iz Southamptona u New York City. Od 2223 putnika i članova posade preživjelo ih je samo 706. Poginulo je ili ih se utopilo 1517, od toga 1360 muškaraca te 157 žena i djece.

Pretpostavke koje se žele potvrditi analizom podataka su:

- Proporcija stradalih je veća od proporcije preživjelih
- Osobe koje su putovali s nekim imale su veće šanse za preživljavanje
- Preživljavanje je ovisilo o spolu, tj. žene su imale veće šanse za preživljavanje
- Preživljavanje je ovisilo o dobi putnika, točnije mlađi putnici su imali veće šanse za preživljavanje (poznato je da su djeca imala prednost prilikom ukrcavanja u čamce za spašavanje)
- Preživljavanje je ovisilo o klasi kojom je osoba putovala (viša klasa je tipično bliža čamcima za spašavanje)

1.2. Opis varijabli

Baza sadrži 1309 jedinki i 14 varijabli. Svaka jedinka predstavlja jednog putnika broda Titanic. U Tablici 1. nalazi se opis i tip varijabli koje će biti korištene u analizi:

Tablica 1. Opis varijabli

Ime varijable	Opis varijable	Tip varijable
pclass	Klasa kojom je putnik putovao (1st, 2nd ili 3rd)	Kvalitativna ordinalna
survived	Je li putnik preživio ili ne (0 = Ne, 1 = Da)	Kvalitativna nominalna
sex	Spol (male=muškarac, female=žena)	Kvalitativna nominalna
age	Dob	Numerička neprekidna
sibsp	Broj braće, sestara ili supružnika na brodu	Numerička diskretna
parch	Broj roditelja ili djece na brodu	Numerička diskretna

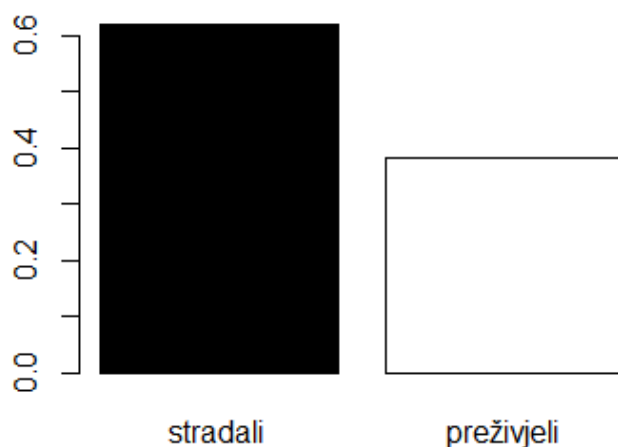
2. Analiza podataka

2.1. Preživljavanje

U Tablici 2. nalaze se frekvencije i relativne frekvencije varijable survived. Vidljivo je da je više stradalih nego preživjelih, a to je u skladu s poznatom činjenicom da nije bilo dovoljno čamaca za spašavanje na Titanicu. Na Slici 1. nalazi se stupčasti dijagram za istu varijablu.

Tablica 2. Tablica frekvencija i relativnih frekvencija za varijablu survived

	frekvencija	%
preživjeli	500	38,2
stradali	809	61,8



Slika 1. Stupčasti dijagram relativnih frekvencija varijable survived

Zanima nas je li proporcija preživjelih statistički značajno manja od proporcije stradalih. Korištenjem binomnog testa potvrđujemo našu slutnju ($p < 2.2e-16$),

odnosno možemo tvrditi da je proporcija preživjelih statistički značajno manja od proporcije stradalih.

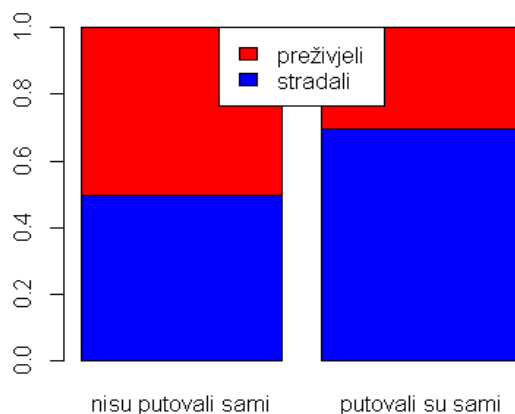
2.1.1. Povezanost preživljavanja s brojem članova obitelji na brodu

Želimo provjeriti je li činjenica da je osoba putovala sama ili s nekim članom obitelji povezana s mogućnošću preživljavanja potopa broda. U tablici 3. vidljive su frekvencije i relativne frekvencije broja putnika koji su putovali sam ili s nekim prema preživljavanju.

Tablica 3. Tablica frekvencija i relativnih frekvencija načina putovanja prema preživljavanju

	preživjeli	stradali	Ukupno u retku
Putnik je putovao sam	261	258	519
redak %	50,3%	49,7%	
Putnik je putovao s nekim	239	551	790
redak %	30,3%	69,7%	
Ukupno	500	809	1309

Proporcije putnika koji su putovali sami i onih koji nisu posebno među stradalima i preživjelima vidljive su na Slici 2.



Slika 2 Stupčasti dijagram relativnih frekvencija načina putovanja prema preživljavanju

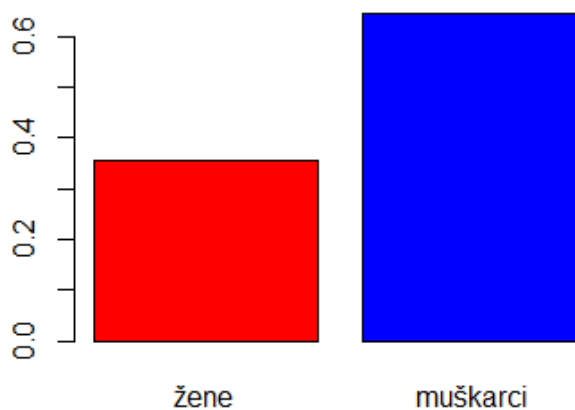
Sada nas zanima je li proporcija preživjelih veća među onima koji su putovali sami nego među onima koji su putovali s barem još jednim članom obitelji. Provođenjem testa o jednakosti proporcija za dva nezavisna uzorka potvrđujemo slutnju ($p=2.243e-13$), tj. možemo tvrditi da je proporcija preživjelih manja među putnicima koji su putovali sami nego među onima koji su putovali s barem još jednim članom obitelji.

2.2. Spol

Zanimljivo je da je više muškaraca putovalo Titanicom nego žena. U Tablici 4. nalaze se frekvencije i relativne frekvencije putnika po spolu.

Tablica 4. Tablica frekvencija i relativnih frekvencija

spol	frekvencija	%
muški	843	64,4
ženski	466	35,6



Slika 3. Stupčasti dijagram relativnih frekvencija putnika po spolu

Primjećujemo da je proporcija muškaraca nešto veća od proporcije žena u uzorku. Provjerimo je li ta razlika statistički značajna, tj. binomnim testom testiramo na razini značajnosti 0.05 je li proporcija muškaraca veća od 0.5. S obzirom da je p-vrijednost izrazito mala ($p < 2.2e-16$) potvrđujemo našu slutnju.

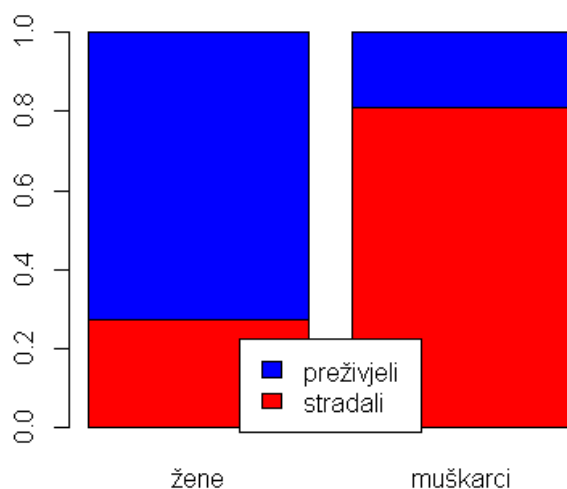
2.2.1. Povezanost spola s preživljavanjem

S obzirom na to da je poznato kako su prednost kod ukrcavanja na čamce za spašavanje imale žene i djeca, logično je za pretpostaviti kako preživljavanje ovisi o spolu putnika. Testiramo postoji li zavisnost između spola i preživljavanja χ^2 -testom o nezavisnosti i potvrđujemo da je preživljavanje ovisilo o spolu, odnosno da su žene imale prednost prilikom ukrcavanja na čamac za spašavanje ($p < 2.2e-16$).

Tablica 5. Tablica frekvencija i relativnih frekvencija spola prema preživljavanju

	preživjeli	umrli	Ukupno u retku
muškarci	161	682	790
redak %	19,1%	80,9%	
žene	339	127	519
redak %	72,7%	27,3%	
Ukupno	500	809	1309

Iz Tablice 5. i Slike 4. vidljivo je da je među muškarcima veća proporcija stradalih nego kod žena.



Slika 4. Stupčasti dijagram relativnih frekvencija spola prema preživljavanju

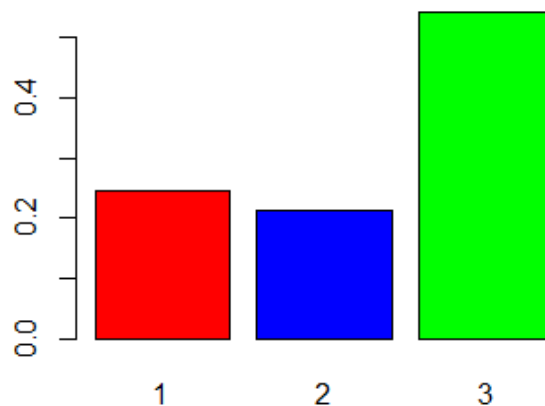
Konkretno, uočavamo da je među ženama proporcija preživjelih 72,7% , dok je kod muškaraca proporcija preživjelih 19,1%.

2.3. Klasa

U Tablici 6. nalaze se frekvencije i relativne frekvencije klase kojom je osoba putovala. Vidljivo je da je najviše putnika putovala u trećoj (najlošijoj) klasi. Pripadni stupčasti dijagram nalazi se na Slici 5.

Tablica 6. Tablica frekvencija i relativnih frekvencija klase kojom je osoba putovala

klasa	frekvencija	%
prva	323	24,68
druga	277	21,16
treća	709	54,16



Slika 5. Stupčasti dijagram relativnih frekvencija klase kojom je osoba putovala

S obzirom na to da je poznato kako je mnogo putnika s Titanica bilo poprilično siromašno i putovali su u SAD u potrazi za poslom, želimo provjeriti je li statistički značajno veća proporcija putnika koji su putovali trećom klasom (54,16%), u kojoj su uvjeti bili najlošiji, nego onih koji su putovali prvom ili drugom klasom (45,84%). Testirajmo je li ta razlika statistički značajna. Korištenjem binomnog testa potvrđujemo našu slutnju ($p=0.001411$), tj. možemo tvrditi da je veća proporcija putnika koji su putovali trećom klasom od proporcije onih koji su putovali prvom ili drugom klasom.

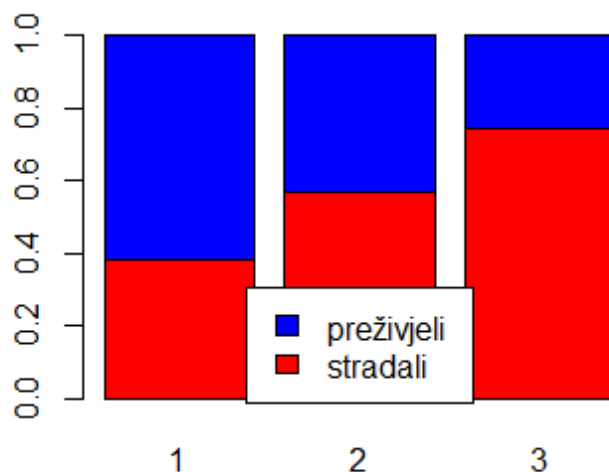
2.3.1. Povezanost klase s preživljavanjem

Poznato je kako su putnici prve klase zbog svog položaja na brodu imali najbolji pristup čamcima za spašavanje pa želimo provjeriti ovisi li preživljavanje o klasi kojom je putnik putovao. Testiramo slatnju χ^2 -testom o nezavisnosti i zaključujemo da je na razini značajnosti 0.05 preživljavanje ovisilo o klasi kojom je putnik putovao ($p < 2.2e-16$).

Među putnicima prve klase je ujedno i najveći udio preživjelih (61.9%). Detaljniji uvid se može dobiti u Tablici 7., dok je ilustracija dostupna na Slici 6.

Tablica 7. Tablica frekvencija i relativnih frekvencija preživljavanja putnika s obzirom na klasu u kojoj su putovali

	preživjeli	umrli	Ukupno u retku
klasa 1	200	123	323
redak %	61,9%	38,1%	
klasa 2	119	158	277
redak %	43%	57%	
klasa 3	181	528	709
Redak %	25,5%	74,5%	
Ukupno	500	809	1309



Slika 6. Stupčasti dijagram relativnih frekvencija preživljavanje prema klasi putovanja

Primjećujemo razlike u proporcijama preživjelih među klasama i stoga želimo provjeriti je li statistički značajno veća proporcija preživjelih među putnicima koji su putovali prvom klasom nego među onima koji su putovali drugom ili trećom klasom.

Pokazuje se da je razlika statistički značajna ($p < 2.2e-16$), tj. proporcija preživjelih među putnicima koji su putovali prvom klasom je statistički značajno veća od proporcije onih koji su putovali drugom ili trećom klasom.

2.4. Dob

Deskriptivna statistika dobi putnika nalazi se u Tablici 7. Prosječna dob putnika Titanica iznosi 29.88 uz standardnu devijaciju 14.41. Zanimljivo je da je i najstariji putnik osoba stara 80 godina.

Tablica 8. Deskriptivna statistika dobi putnika Titanica

minimum	donji kvartil	medijan	gornji kvartil	maksimum	prosjek	st. ev.
0.1667	21	28	39	80	29.88	14.41

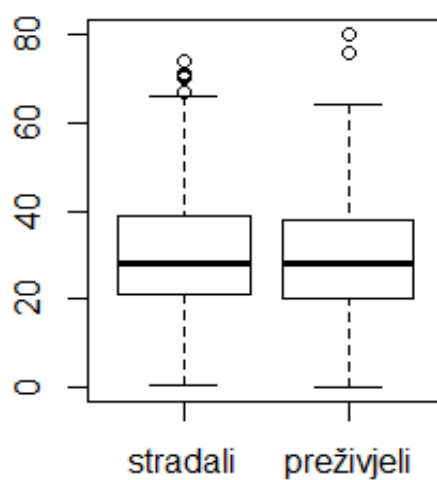
2.4.1. Povezanost dobi s preživljavanjem

Logično je za pretpostaviti kako je očekivana dob preživjelih manja od očekivane dobi stradalih jer je poznato kako su prednost pri ukrcavanju u čamce za spašavanje imala djeca. U Tablici 8. dostupna je deskriptivna statistika dobi putnika prema preživljenju.

Tablica 9. Deskriptivna statistika dobi putnika prema preživljavanju

	minimum	donji kvartil	medijan	gornji kvartil	maksimum	prosjek	st. dev.
preživjeli	0.33	21	28.5	41	80	30.97	14.9
stradali	0.17	21	28	36.5	74	29.07	14

Unatoč našoj intuiciji preživjeli su prosječno stariji (30.97) u odnosu na stradale (29.07). Grafička usporedba dobi prema preživljenju dostupna je na Slici 7.



Slika 7. Kutijasti dijagram dobi putnika prema preživljavanju

Suprotno našoj pretpostavci provjerimo je očekivana dob preživjelih statistički značajno veća od očekivane dobi stradalih. Na razini značajnosti 0.05 možemo potvrditi našu slutnju (Welchov t-test, $p = 0.0183$).

3. Zaključak

Na osnovu podataka iz baze „titanic3.csv“, na razini značajnosti 0.05, donijeli smo sljedeće zaključke o putnicima s broda RMS Titanic:

- proporcija preživjelih je manja od proporcije stradalih
- proporcija preživjelih manja je među putnicima koji su putovali sami nego među onima koji su putovali s barem jednim članom obitelji
- proporcija muškaraca veća je od proporcije žena
- među preživjelim putnicima veća je proporcija žena nego muškaraca
- veća je proporcija putnika koji su putovali trećom klasom od proporcije onih koji su putovali prvom ili drugom klasom
- veća je proporcija preživjelih među putnicima prve klase nego među putnicima druge i treće klase
- očekivana dob preživjelih veća je od očekivane dobi stradalih unatoč suprotnoj pretpostavci.

4. Literatura

- <https://www.gagolewski.com/resources/data/titanic3.csv>
- https://hr.wikipedia.org/wiki/RMS_Titanic
- <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.html>