

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

ODJEL ZA MATEMATIKU

Sveučilišni preddiplomski studij Matematika i računarstvo

# **TESTIRANJE NAD PODACIMA O PROMETNIM NEZGODAMA**

Seminarski rad

Kolegij: Statistički praktikum

Student: David Dujmović

Osijek, 2022.

## Sadržaj

1.Uvod.....	1
1.1. Opis varijabli .....	2
2.Analiza podataka .....	3
2.1. Proporcija spola .....	3
2.2. Prosječna dob.....	4
2.3. Godina prometne nezgode .....	5
2.4. Poveznica između nošenja pojasa i starosti.....	6
2.5. Udio frontalnih sudara među spolovima .....	7
2.6. Proporcija ozljeda među spolovima .....	8
2.7. Spol i preživjeli .....	10
2.8. Varijable dead i airbag .....	11
3. Zaključak.....	14

# 1.Uvod

**Prometna nezgoda** događa se kada se vozilo sudari s drugim vozilom, pješakom, životinjama ili drugim nepomičnim preprekama. Prometni sudari često rezultiraju ozljedama, invaliditetom, smrću i imovinskom štetom kao i financijskim troškovima kako društva tako i pojedinaca koji su uključeni. Brojne činjenice doprinose riziku od sudara, uključujući dizajn vozila, brzinu rada, dizajn ceste, cestovno okruženje, vozačke sposobnosti, oštećenja zbog alkohola ili droga i ponašanje, posebno ometajući vožnju, prekoračenje brzine. Brojne tjelesne ozljede mogu biti posljedica trauma sile uzrokovanih nezgodom, u rasponu od modrica i udaraca do katastrofalnih fizičkih ozljeda ili smrti. Nakon nezgode može doći do dugotrajne psihološke traume. Osobe koje su sudjelovale u prometnoj nezgodi često imaju strah ponovno voziti, tj. aktivno sudjelovati u prometu. U nekim slučajevima psihološka trauma može utjecati na život pojedinca, može uzrokovati poteškoće do posla, pohađanja škole ili obavljanja obiteljskih obaveza. Vodeći se ovim važnim činjenicama, želimo provesti istraživanje koje će nam pokazati dosta korisnih stvari na koje trebamo pripaziti prilikom vožnje.

## 1.1. Opis varijabli

Baza „nassCDS“<sup>1</sup> se sastoji od 26217 jedinki i 15 varijabli. Svaka jedinka predstavlja jednog putnika, na temelju čega vršimo istraživanje. U bazi se nalaze sljedeće varijable.

- dvcats - procjena udarne brzine vozila (kvalitativna ordinalna varijabla)
- weight - procjena težine, iako neizvjesne točnosti (numerička neprekidna varijabla)
- dead - označava je li ispitanik preživio ili ne (kvalitativna nominalna varijabla)
- airbag - sa zračnim jastukom ili bez zračnog jastuka (kvalitativna nominalna varijabla)
- seatbelt - sa pojasom ili bez pojasa (kvalitativna nominalna varijabla)
- frontal - frontalni ili ne frontalni udar {1-frontalni, 0-ne frontalni} (kvalitativna nominalna varijabla)
- sex - spol putnika {f-ženski spol, m-muški spol} (kvalitativna nominalna varijabla)
- ageofocc - starost osobe u godinama (numerička neprekidna varijabla)
- yearacc - godina kad se dogodila nesreća {1997-2002} (kvalitativna ordinalna varijabla)
- yearveh - godina modela vozila koji je sudjelovao u nesreći (kvalitativna ordinalna varijabla)
- abcat - jel se aktivirao jedan ili više zračnih jastuka (kvalitativna nominalna varijabla)
- occrole - vozač ili nevozač (kvalitativna nominalna varijabla)
- deploy - (0-ako zračni jastuk nije bio dostupan ili se nije aktivirao, 1-ako se aktivirao jedan ili više zračnih jastuka) (kvalitativna nominalna varijabla)
- injSeverity - (0- nema ozljeda, 1-moguća ozljeda, 2-nema invalidnosti, 3-invalidnost, 4-poginuo, 5-nepoznato, 6-prije smrti) (kvalitativna ordinalna varijabla)
- caseid - jedinstveni identifikacijski broj vozila (numerička neprekidna varijabla)

---

<sup>1</sup><https://vincentarelbundock.github.io/Rdatasets/datasets.html?fbclid=IwAR0ZG1DcSQdIDbT7NC9eNyyQpoupos9o6aIwfKiDMwO3ndZti0D9>

## 2. Analiza podataka

### 2.1. Proporcija spola

Prema podacima koji se nalaze u varijabli sex, mi imamo 53% (13969) muškaraca i 47% (12248) žena. Je li proporcija muškaraca statistički značajno veća od proporcije žena, na razini značajnosti 0.05?

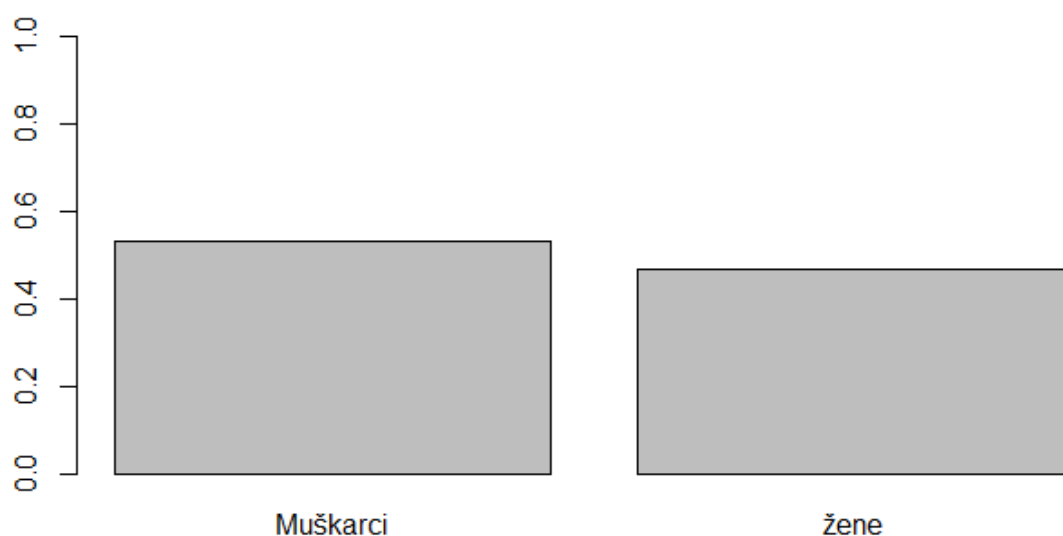
Ako sa  $\theta$  označimo proporciju muškaraca u ukupnoj populaciji onda imamo sljedeće hipoteze:

$$H_0: \theta = 0.5$$

$$H_1: \theta > 0.5$$

Korištenjem binomnog testa dobivamo sljedeću p-vrijednost =  $2.2e-16 < 0.05$ , pa na razini značajnosti 0.05 odbacujemo hipotezu  $H_0$  i prihvaćamo hipotezu  $H_1$ . To jest, možemo tvrditi da je proporcija muškaraca statistički veća od proporcije žena u ovom istraživanju.

Grafički je to prikazano na sljedeći način:



Slika 1. Proporcija muškaraca i žena u populaciji.

## 2.2. Prosječna dob

Prema dostupnim podacima sa stranice "Median age of the U.S.population" prosječna dob stanovnika koji žive u SAD-u 2000. godine je bila 35,3.

Provjerimo je li očekivana dob osobe koja je sudjelovala u prometnoj manja od očekivane dobi stanovnika SAD-a. Koristimo t-test, jer se radi o velikom uzorku.

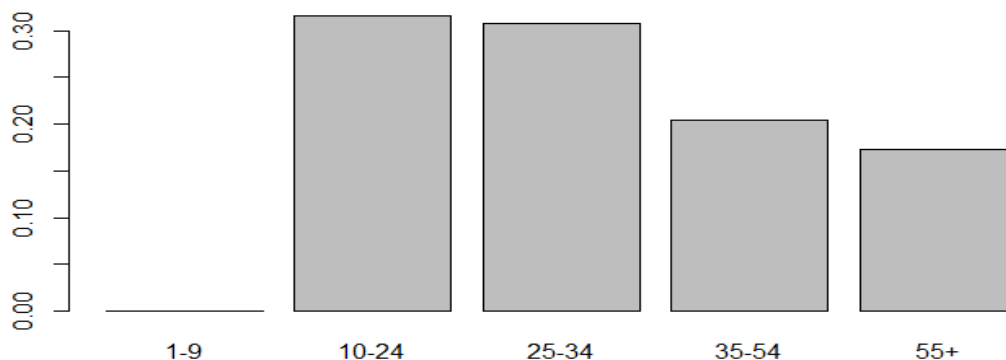
Hipoteze:

$$H_0: \mu = 35.3$$

$$H_1: \mu < 35.3$$

Možemo doći do sljedećeg zaključka:

p-vrijednost = 1 > 0.05 pa na razini značajnosti 0.05 ne odbacujemo  $H_0$  i prihvaćamo hipotezu  $H_1$ , tj. možemo tvrditi da je očekivana dob osobe koja je sudjelovala u prometnoj nesreći statistički manja od očekivane dobi stanovnika SAD-a. Iz grafičkog prikaza možemo primjetiti da je najmlađa osoba imala tek 11 godina dok je najstarija osoba imala preko 90 godina. Osobe koje najčešće sudjeluju u prometnoj nezgodi su prilično mlađe od 36 godina, kako se starost osoba povećava tako se smanjuje broj sudjelovanih osoba u nezgodi. Ovaj prikaz možemo smatrati opadajućim od 10-te godine pa na dalje.



Slika 2. Dobni prikaz osoba koje su sudjelovale u prometnoj nezgodi.

### 2.3. Godina prometne nezgode

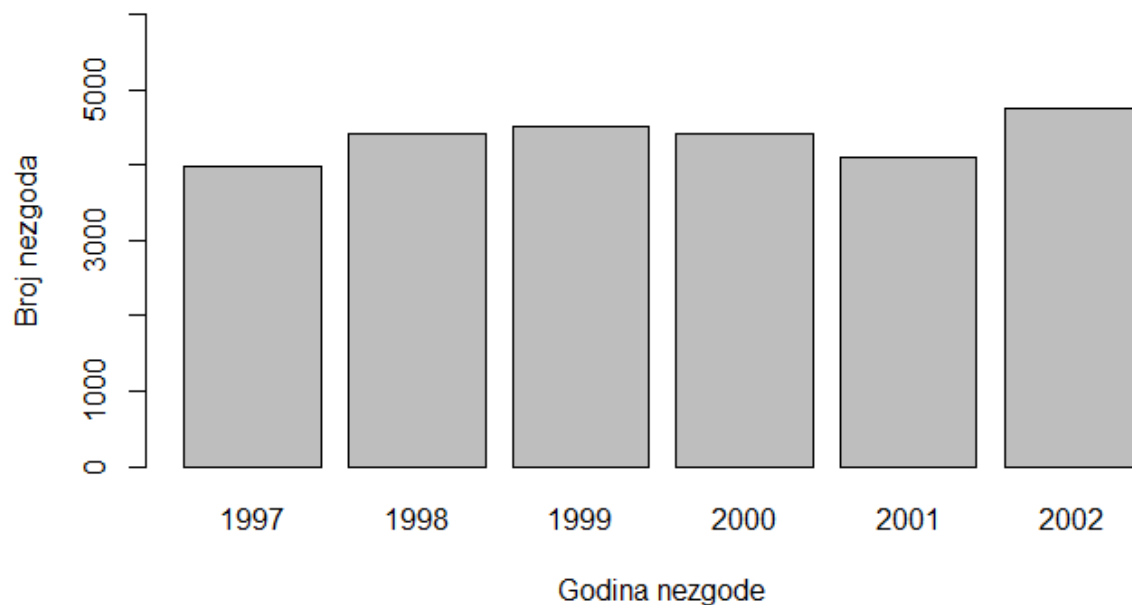
Sljedeća tablica prikazuje godinu prometne nezgode sa pripadnim brojem nezgoda te godine.

<i><b>1997.</b></i>	<i><b>1998.</b></i>	<i><b>1999.</b></i>	<i><b>2000.</b></i>	<i><b>2001.</b></i>	<i><b>2002.</b></i>
3975	4427	4516	4420	4115	4764

Iz tablice možemo izčitati da se 2002. godine dogodilo najviše nezgoda i to 4764. Na razini značajnosti 0.05, želimo provjeriti postoji li prednost koje godine će se dogoditi prometna nezgoda ili je to sve plod slučajnosti.

Korištenjem  $\chi^2$  testa dobivamo p-vrijednost  $< 2.2e-16 < 0.05$  pa na razini značajnosti 0.05 odbacujemo  $H_0$  i prihvaćamo  $H_1$ , tj. možemo tvrditi da postoji prednost godine kad će se dogoditi prometna nezgoda.

Iz grafičkog prikaza možemo uočiti da ne postoji godina u kojoj je došlo do drastičnog pada ili rasta nezgoda, ali broj nezgoda ovisi i o mnogo drugih faktora koje ćemo poslije komentirati.



Slika 3. Godina prometne nezgode sa pripadajućim brojem nezgoda.

## 2.4. Poveznica između nošenja pojasa i starosti

Promatramo kvalitativnu i numeričku varijablu, te želimo utvrditi koja skupina ljudi najviše koristi sigurnosni pojas prilikom vožnje, tj. hoćemo zaljučiti da li je istinita činjenica kad kažemo da mladi ljudi puno više prave prekršaje prilikom vožnje od iskusnijih vozača. Istražitelji uzroka prometnih nesreća naglašavaju važnost vezanja prilikom vožnje, jer naglašavaju da je puno veća vjerojatnost da osoba preživi udar kad je vezana pojasom. Uobičajena navika putnika je da samo vozač bude vezan, dok ostale osobe u vozilu ne koriste pojas. Kako starije osobe smatramo odgovornijima, pretpostaviti ćemo da je očekivana starost osoba koje su vezane veća od očekivane starosti onih osoba koje nisu vezane pojasom.

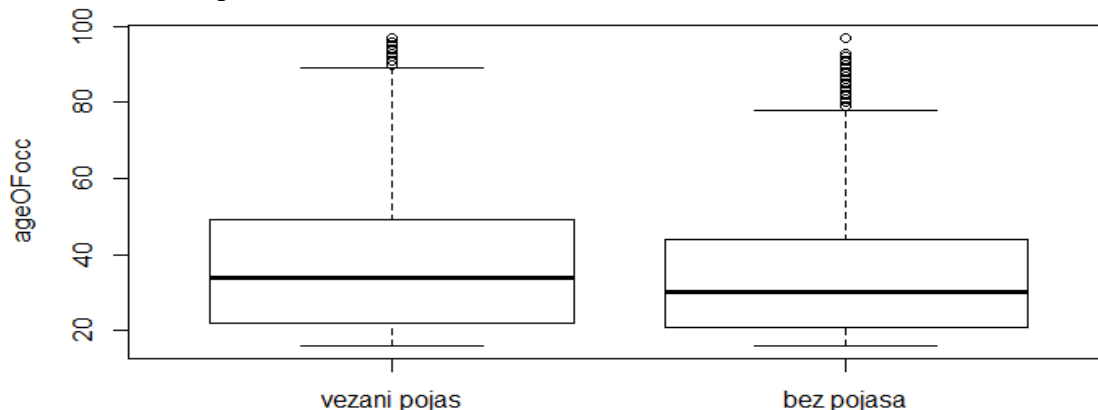
Hipoteze:

HO: Očekivana starost osoba koje su vezane pojasom je jednaka očekivanoj starosti osoba koje ne koriste pojas.

H1: Očekivana starost osoba koje koriste pojas veća je od očekivane starosti osoba koje ne koriste pojas.

Zaključak testa:

Korištenjem t-testa dobivam p-vrijednost  $< 2.2e-16 < 0.05$  pa na razini značajnosti 0.05 odbacujemo HO i prihvaćamo H1, tj. možemo tvrditi da je očekivana starost osoba koje koriste sigurnosti pojas za vezanje veća od očekivane starosti osoba koje i dalje pojas ne koriste. Grafički prikaz:



Slika 4. Očekivana starost osoba koje koriste ili ne koriste sigurnosni



## 2.5. Udio frontalnih sudara među spolovima

Promatramo dvije kvalitativne varijable, varijabla sex se sastoji od muškog i ženskog spola, te varijabla frontal od kategorije da je sudar bio frontalni i kategorije da sudar nije bio frontalni. Oko 60% žena je imalo frontalni sudar i oko 70% muškaraca također. Željeli bismo provjeriti je li proporcija žena koje su imale frontalni udar manji od proporcije muškaraca koji su imali frontalni sudar. Možemo primjetiti da je broj muškaraca veći od broja žena u populaciji, te provodimo odgovarajući test.

Tablica relativnih frekvencija odgovarajućih varijabli:

Frontal	Sex	
	Žene	Muškarci
Frontalni sudar	0.3882266	0.3290142
Ne frontalni sudar	0.6117734	0.6709858

Testiramo je li proporcija žena koje su imale frontalni udar manji od proporcije muškaraca koji su imali frontalni udar, tj. da li vrijedi naša pretpostavka.

Hipoteze:

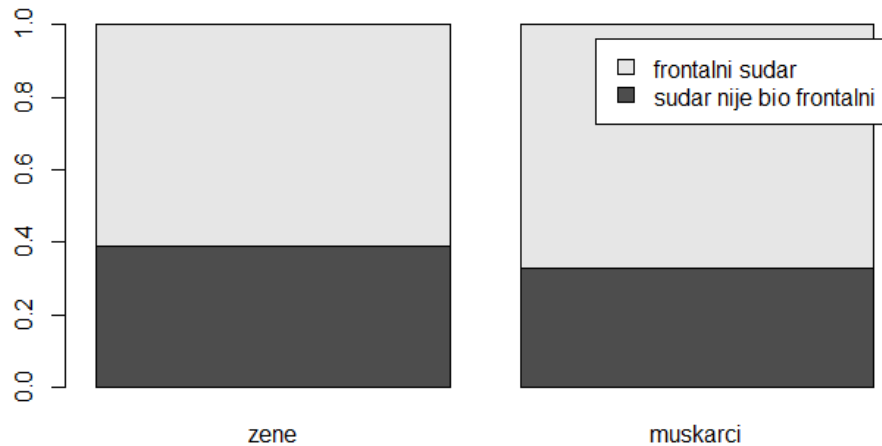
HO: Proporcija frontalnih sudara među ženama jednaka je proporciji frontalnih sudara među muškarcima

H1: Proporcija frontalnih sudara među ženama manja je od proporcije frontalnih sudara među muškarcima

Zaključak testa:

Korištenjem testa o proporciji dobivamo p-vrijednost  $< 2.2e-16 < 0.05$  pa na razini značajnosti 0.05 možemo tvrditi da je proporcija onih koji su imali frontalni sudar veća kod muškaraca nego kod žena. Dakle, muškarci su skloniji većim pogreškama, pa samim time i opasniji u prometu, kako za svoje zdravlje i sigurnost tako i za ostale osobe.

Grafički prikaz žena i muškaraca koji su imali frontalni udar:



Slika 5. Proporcija muškaraca i žena kod frontalnog udara.

## 2.6. Proporcija ozljeda među spolovima

Na temelju prethodnog provođenja testova, zaključili smo da su muškarci u prosjeku mlađe osobe, te sklonije frontalnom udaru. Želimo provjeriti da li muškarci imaju veće ozljede prilikom nesreće. Htjeli bismo provjeriti je li proporcija žena koje nisu imale ozljeda manja od proporcije muškaraca koji nisu imali ozljeda. Kada gledamo tablicu frekvencija, mogli smo također promatrati situaciju kada su osobe preminule i sve ostale slučajeve. Pošto je broj preminulih osoba mali u odnosu na cijelu populaciju, ali i dalje dovoljno velik za tu kategoriju, ne bi trebalo biti velikih odstupanja. Zato upravo promatramo ovo podjelu, jer nam grafički prikaz slikovito objašnjava situaciju, i nekako nam je lakše shvatiti i opisati tu pojavu. Posljedice prometne nezgode mogu na osobama ostaviti stalne ožiljke, te nezgodu u općem smislu možemo shvatiti nekim dijelom kao slučajan događaj, te ju tako onda trebamo i shvaćati.

Tablica relativnih frekvencija odgovarajućih varijabli:

injSeverity	Spol	
	žene	muškarci
Nema ozljeda	0.3512712	0.583686
Ostale posljedice	0.6487288	0.416314

Provjeravamo da li je proporcija žena koje nisu imale ozljede manja od proporcije muškaraca koji nisu imali ozljeda. Kako smatramo da je kod žena udarna brzina manja nego kod muškaraca, pretpostavljamo da su onda i ozljede puno blaže. Iz tablice relativnih frekvencija uočavamo velike pomake među spolovima.

Hipoteze:

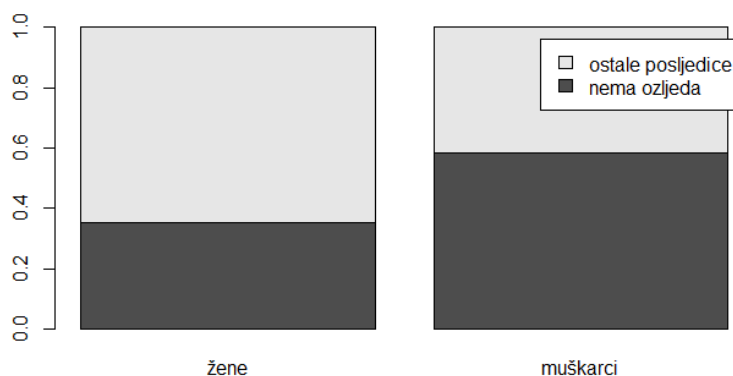
H<sub>0</sub>: Proporcija ozljeda među ženama je jednaka proporciji ozljeda među muškarcima

H<sub>1</sub>: Proporcija ozljeda među ženama je manja od proporcije ozljeda među muškarcima

Zaključak testa:

Primjenom prop-testa dobivamo sljedeće zaključke; p-vrijednost  $< 2.2e-16 < 0.05$  pa na razini značajnosti 0.05 odbacujemo H<sub>0</sub> i prihvaćamo H<sub>1</sub>, tj. možemo tvrditi da je proporcija onih koji nisu imali ozljeda manja kod žena nego kod muškaraca.

Grafički prikaz muškaraca i žena koji nisu imali ozljeda prilikom prometne nezgode:



Slika 6. Proporcija muškaraca i žena koji nisu imali ozljeda i onih koji su imali određene posljedice.

## 2.7. Spol i preživjeli

Promatramo dostupne podatke o kvalitativnim varijablama, te želimo nešto zaključiti o spolu osoba koje su preživjele prometnu nezgodu. Uspoređujemo preživjele osobe jer je njihov udio u populaciji puno veći od broja umrlih, što je u skladu sa našom pretpostavkom. U ovoj varijabli “dead” imamo samo dvije podjele, jer kod našeg trenutnog zaključivanja nas ne zanimaju pojedinosti, nego samo broj preživjelih osoba koje promatramo. Na temelju prethodnih testova mogli bi pretpostaviti da je proporcija žena koje su preživjele prometnu nezgodu veća od proporcije muškaraca koji su preživjeli prometnu nezgodu.

Tablica relativnih frekvencija odgovarajućih varijabli:

dead	sex	
	žene	muškarci
alive	0.96211626	0.94874365
dead	0.03788374	0.05125635

Očekivamo puno veći udjeli kod preživjelih osoba, pa ćemo sada pobliže promatrati taj podskup. Iz tablice frekvencija možemo uočiti da je udio preživjelih žena malo veći od udjela preživjelih muškaraca, ali to treba statistički provjeriti i pokazati da li je to statistički značajna razlika na razini značajnosti 0.05. Preko hipoteza želimo odrediti je li proporcija žena koje su preživjele prometnu nezgodu veća od proporcije muškaraca koji su preživjeli.

Hipoteze:

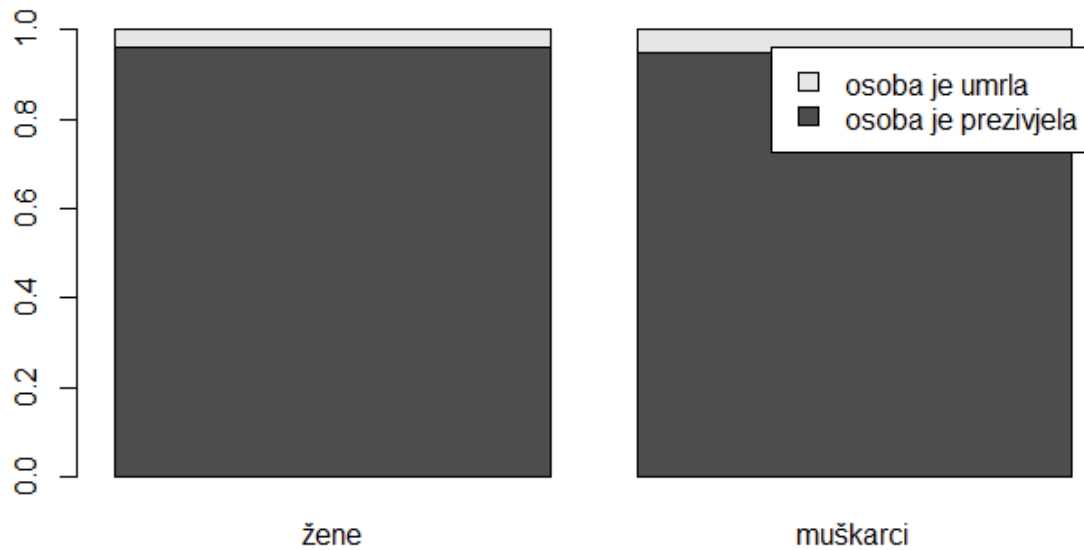
HO: Proporcija žena koje su preživjele prometnu nezgodu je jednak proporciji muškaraca koji su preživjeli

H1: Proporcija žena koje su preživjele prometnu nezgodu je veća od proporcije muškaraca koji su preživjeli

Zaključak testa:

Primjenom prop-testa dolazimo do sljedećih saznanja,  $p\text{-vrijednost} = 1.105e-07 < 0.05$  pa na razini značajnosti 0.05 odbacujemo  $H_0$  i prihvaćamo  $H_1$ , možemo tvrditi da je proporcija preživjelih veća kod žena nego kod muškaraca.

Grafički prikaz žena i muškaraca koji su preživjeli(umrli) prometnu nezgodu:



Slika 7. Proporcija žena i muškaraca koji su preživjeli ili umrli tijekom prometne nezgode.

## 2.8. Varijable dead i airbag

Osobe koje su preživjele u prometnoj nezgodi, preživjeli su na temelju slučajnosti, sreće, sigurnosnog pojasa ili na temeljenu činjenice da su se aktivirali zračni jastuci. Možemo na temelju podataka iz varijable “abcat” zaključiti koji zračni jastuk se aktivirao, bilo to kod vozača ili kod ostalog putnika u vozilu. Naš ljudski instikt govori da noviji automobili imaju više zračnih jastuka, a samim time kvalitetnije izvedbe, te veće šanse sa preživjeti. Pretpostavimo da je vjerojatnost da će osoba preživjeti, ako se zračni jastuk aktivirao, veća od vjerojatnosti da će osoba umrijeti sa aktiviranim zračnim jastukom.

Tablica relativnih frekvencija odgovarajućih varijabli:

airbag	dead	
	alive	dead
yes	0.5554979	0.4330508
no	0.4445021	0.5669492

Na temelju tablice možemo uočiti da postoji razlika između umrlih i preživjelih osoba za koje se aktivirao zračni jastuk. Provjerimo na razini značajnosti 0.05 da li je proporcija osoba za koje se aktivirao zračni jastuk među preživjelim većā od proporcije osoba za koje se aktivirao zračni jastuk među umrlima.

Hipoteze:

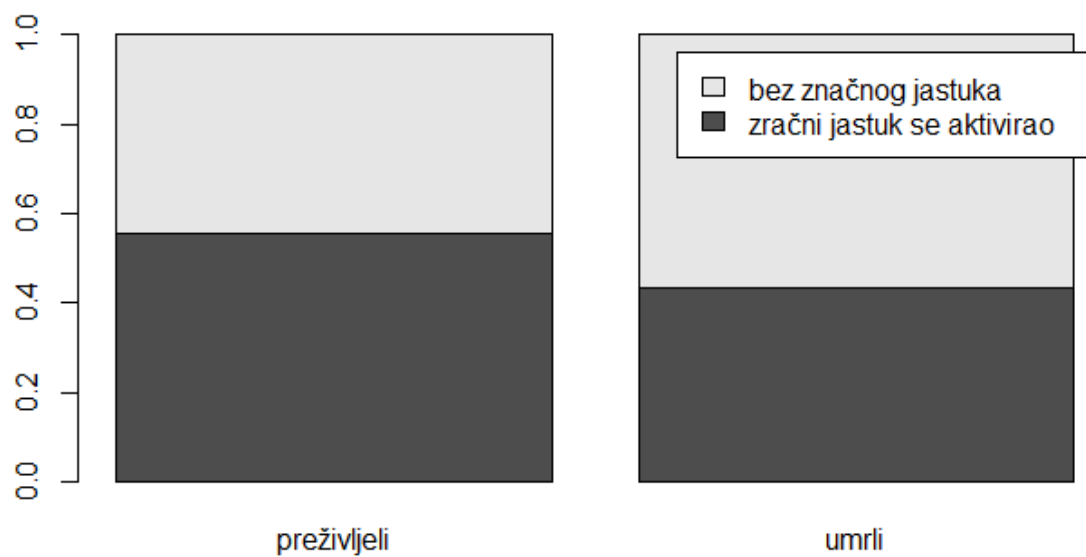
HO: Proporcija osoba za koje se aktivirao zračni jastuk među preživjelim je jednaka proporciji među umrlima.

H1: Proporcija osoba za koje se aktivirao zračni jastuk među preživjelim je većā od proporcije među umrlima.

Zaključak testa:

Dobivamo sljedeće podatke, p-vrijednost  $< 2.2e-16 < 0.05$  pa na razini značajnosti 0.05 odbacujemo HO i prihvaćam H1, možemo tvrditi da je proporcija osoba za koje se aktivirao zračni jastuk većā kod preživjelih nego kod umrlih.

Grafički prikaz:



Slika 8. Proporcija preživjelih i umrlih osoba kod kojih se aktivirao/nije aktivirao zračni jastuk.

### 3. Zaključak

Ovo istraživanje je provedeno na dostupnoj bazi podataka “nassCDS” koja sadrži 15 varijabli. Tijekom ovog rada smo proveli dobar dio testova i opisali što više različitih varijabli, ali nismo baš sve usporedbe vršili. Na temelju jednog uzorka kojeg smo istražili želimo zaključke donijeti na cjelokupnoj populaciji. Podaci predstavljaju sve činjenice vezane za prometne nezgode u SAD-u od 1997. godine do 2002. godine. Proporcija muškaraca u uzorku je statistički veća od proporcije žena, dolazimo do zaključka da su muškarci opasniji sudionici u prometu, suprotno našoj pretpostavci. Možemo primjetiti da je najmlađa osoba imala tek 11 godina dok je najstarija osoba imala preko 90 godina. Najveći broj prometnih nezgoda se dogodio 2002. godine i to 4764 slučaja. Kako je broj prometnih nesreća plod neke slučajnosti tako postoje i neki čimbenici koji indirektno utječu na to. Kada gledamo starost osobe koje koriste pojas dolazimo do zaključka, da očekivana starost osoba koje koriste sigurnosti pojas za vezanje, veća od očekivane starosti osoba koje i dalje pojas ne koriste. Muškarci su skloniji frontalnom sudaru nego žene, što je suprotno od naše pretpostavke. Promatranjem osoba koje nisu imali ozljeda prilikom prometne nezgode uočavamo da muškarci puno bolje prolaze nego žene. Ali kad gledamo osobe koje su preminule u prometnoj nezgodi, to su pretežno osobe muškog spola. I na kraju dolazimo do zaključka da dobra većina osoba kod kojih se zračni jastuk aktivira, preživljavaju prometnu nezgodu, što je u skladu sa pretpostavkom.