

Seminarska naloga 2

Računalniška podpora statistike

Alen Kahteran

22. 11. 2020

Uvod

Pred kratkim sem si hotel odgovoriti na vprašanje ali obstaja kakšna razlika med moškimi in ženskami, ki povzročijo nesrečo z osebnim avtomobilom v različnih okoliščinah.

Podatke o vseh (od leta 1995 dalje) prometnih nesrečah v Sloveniji, lahko dobimo na spletni strani policije¹. Odločil sem se, da raziščem podatke za leto 2020, saj je možno, da je kje opazen tudi vpliv COVID-19. Seveda za ta namen bi bila potrebna podrobnejša analiza, ter primerjava s prejšnjimi leti, vendar je v vsakem primeru najprej potrebno analizirati le letošnje podatke. Trenutni podatki vsebujejo prometne nesreče le do konca avgusta 2020.

Naše podatke bomo primarno delili na moške in ženske, za tem pa še na posamezne skupine, glede na določeno spremenljivko. Okoliščine so lahko različnih oblik. Večina je povezanih s samo prometno nesrečo (gostota prometa, lokacija, vremenske razmere, itd.), nekaj pa jih je z voznikom (starost, vozniški staž, itd.). Najprej je bilo potrebno podatke korektno pripraviti za obdelavo, kar si lahko pogledamo v naslednjem poglavju.

Čiščenje podatkov

Preden se kakorkoli dotaknemo podatkov, je potrebno vedeti kakšne podatke sploh imamo. Tu nam je policija poleg podatkov, pripravila tudi opise (in formate) vseh spremenljivk. Žal ti niso točno povedali kateri opis pripada kateri spremenljivki, vendar je že samo ime spremenljivke povedalo večino informacij. Tako da v tabeli 1 si lahko pogledamo imena spremenljivk, njihove opise ter ali so bile uporabljene ali ne.

Za večino spremenljivk je bil podan format. Ponekod, kjer je bil podatek ločen na dva stolpca, kot npr. DatumPN in UraPN sem ustrezno pretvoril v datum z uro. Podobno sem storil pri vozniskem stažu, kjer sem vse pretvoril v mesece.

Spremenljivke ki sem jih obdržal so bile ZaporednaStevilkaPN, DatumPN, VNaselju, VzrokNesrece, VremenskeOkoliscine, StanjePrometa, StanjeVozisca, VrstaVozisca, Starost, Spol, VozniskiStazVMesecih in VrednostAlkotesta. Povzročitelj, VrstaUdelezenca in Drzavljanstvo sem uporabil le toliko, da smo dobili pravi vzorec naših začetnih podatkov (povzročitelje nesreč, ki so vozili osebni avtomobil in imajo slovensko državljanstvo).

UraPN in VozniskiStazVLetih sta bila ustrezno dodana drugim spremenljivkam.

Ostale spremenljivke so ali preveč razdrobljeni podatki, ali neuporabni saj se osredotočamo na slovensko populacijo, ali pa vezani na posledice nesreče. Tako si pogledjmo še preostale spremenljivke v naslednjem poglavju.

Pregled podatkov

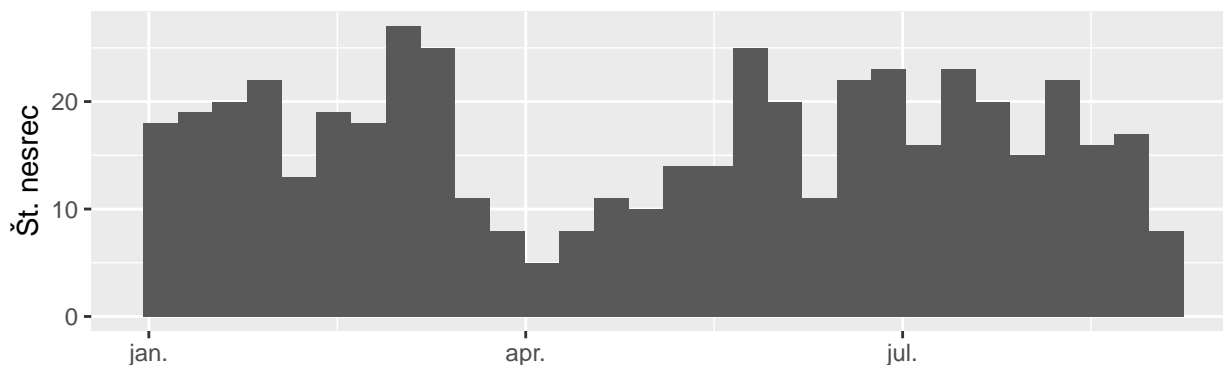
Najprej si pogledjmo na sliki 1 porazdeljenost nesreč v času. V normalnih okoliščinah bi pričakovali, da so nesreče čez celo leto enakomerno porazdeljene (dogodki so časovno neodvisni). V našem časovnem obdobju

¹<https://www.policija.si/o-slovenski-policiji/statistika/prometna-varnost>

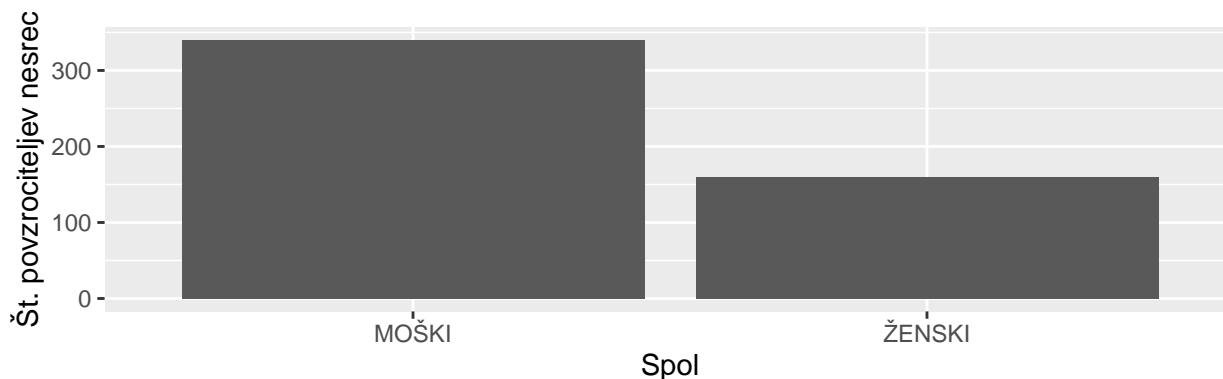
Tabela 1: Opisi spremenljivk

Ime spremenljivke	Opis spremenljivke (in morebiten format)	Uporaba
ZaporednaStevilkaPN	številka za štetje in ločevanje posamezne prometne nesreče	da
KlasifikacijaNesrece	klasifikacija nesreče glede na posledice (Izračuna se avtomatično glede na najhujšo posledico pri udeležencih v prometni nesreči)	ne
UpravnaEnotaStoritve	upravna enota, na območju katere se je zgodila prometna nesreča	ne
DatumPN	datum nesreče (format: dd.mm.llll)	da
UraPN	ura nesreče (format: hh)	ne
VNaselju	indikator ali se je nesreča zgodila v naselju (D) ali izven (N)	da
Lokacija	lokacija nesreče	ne
VrstaCesteNaselja	vrsta ceste ali naselja na kateri je prišlo do nesreče	ne
SifraCesteNaselja	oznaka ceste ali šifra naselja kjer je prišlo do nesreče	ne
TekstCesteNaselja	tekst ceste ali naselja, kjer je prišlo do nesreče	ne
SifraOdsekaUlice	oznaka odseka ceste ali šifra ulice, kjer je prišlo do nesreče	ne
TekstOdsekaUlice	tekst odseka ali ulice, kjer je prišlo do nesreče	ne
StacionazaDogodka	točna stacionaža ali hišna številka, kjer je prišlo do nesreče	ne
OpisKraja	opis prizorišča nesreče	ne
VzrokNesrece	glavni vzrok nesreče	da
TipNesrece	tip nesreče	ne
VremenskeOkoliscine	vremenske okoliščine v času nesreče	da
StanjePrometa	stanje prometa v času nesreče	da
StanjeVozisca	stanje vozišča v času nesreče	da
VrstaVozisca	stanje površine vozišča v času nesreče	da
GeoKoordinataX	Geo Koordinata X (Gauß-Krüger-jev koordinatni sistem)	ne
GeoKoordinataY	Geo Koordinata Y (Gauß-Krüger-jev koordinatni sistem)	ne
ZaporednaStevilkaOsebeVPN	številka za štetje in ločevanje oseb, udeleženih v prometnih nesrečah	ne
Povzročitelj	kot kaj nastopa oseba v prometni nesreči	da
Starost	starost osebe (LL)	da
Spol	spol	da
UEStalnegaPrebivalisca	upravna enota stalnega prebivališča	ne
Drzavljanstvo	državljanstvo osebe	da
PoskodbaUdelezenca	poškodba osebe	ne
VrstaUdelezenca	vrsta udeleženca v prometu	da
UporabaVarnostnegaPasu	ali je oseba uporabljala varnostni pas ali čelado (polje se interpretira v odvisnosti od vrste udeleženca) (Da/Ne)	ne
VozniskiStazVLetih	vozniški staž osebe za kategorijo, ki jo potrebuje glede na vrsto udeleženca v prometu (LL)	ne
VozniskiStazVMesecih	vozniški staž osebe za kategorijo, ki jo potrebuje glede na vrsto udeleženca v prometu (MM)	da
VrednostAlkotesta	vrednost alkotesta za osebo, če je bil opravljen (n.nn)	da
VrednostStrokovnegaPregleda	vrednost strokovnega pregleda za osebo, če je bil odrejen in so rezultati že znani (n.nn)	ne

(od 1. 1. 2020 do 31. 8. 2020) vidimo da temu ni tako, saj se je v sredini marca 2020 začelo ustavljanje javnega življenja zaradi COVID-19. Ker je takrat večina ljudi bila doma, in ne v avtu, je v našem časovnem obdobju pričakovano, da bo v obdobju, ko je bilo ustavljeno javno življenje, manj nesreč. Kljub temu to ne bi smelo vplivati na rezultate, saj se ne osredotočamo na to kdaj so se zgodili, temveč kdo jih je storil.



Slika 1: Porazdelitev nesreč v času



Slika 2: Porazdelitev spola povzročiteljev nesreč.

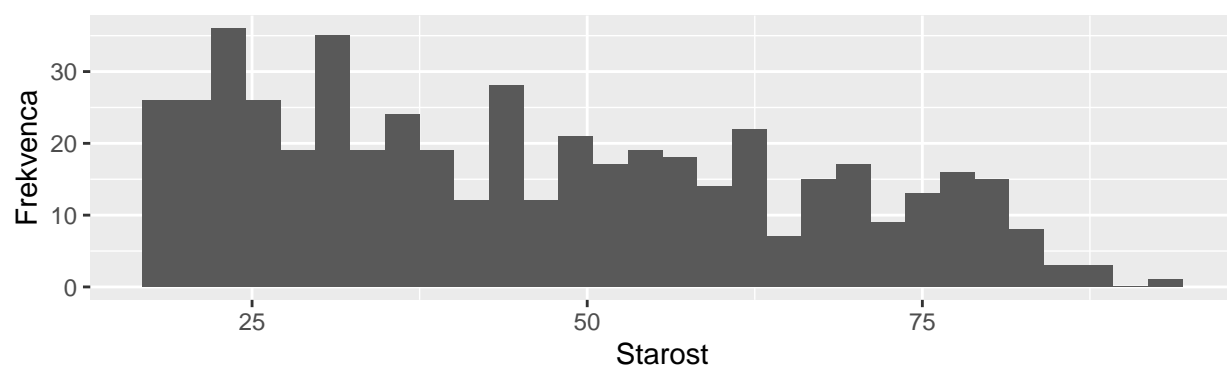
Zato si na sliki 2 najprej pogledjmo porazdelitev našega vzorca po spolu. Videti je da je moških približno dvakrat toliko kot žensk (moških - 340, žensk - 160). Pogledjmo si naš vzorec še za ostale spremenljivke

Starost

Porazdelitev starosti lahko vidimo na sliki 3. Videti je da št. povzročiteljev nesreč pada s starostjo. Če primerjamo našo porazdelitev s starostno porazdelitvijo Slovenije², je videti da ta ne pada s starostjo. To bi mogoče lahko pripisali dvem stvarim, ali temu da so mlajši vozniki osebnih avtomobilov bolj pogosti da povzročijo nesrečo (zaradi voznih izkušenj), ali pa temu, da starejši ljudje manj vozijo avte.

Poglejmo si še opisne statistike Starosti za oba spola, za moške in za ženske v tabeli 2 in še grafični prikaz opisnih statistik starosti za moške in ženske s škatlo z ročaji na sliki 4. Iz grafa okvirjev z ročaji, je videti da sta si porazdelitvi precej podobni. Videti je da so porazdelitve asimetrične (Zgornji del je daljši), in je zato mediana smiselna mera središčnosti, medtem ko je interkvartilni razmik smiselna mera razpršenosti. Žal iz teh fotografij ne moremo nič sklepati. Poleg tega je iz tabele razvidno, da naš vzorec vsebuje celoten spekter starosti, od mladih, ki so komaj opravili izpit, do starejših (75+).

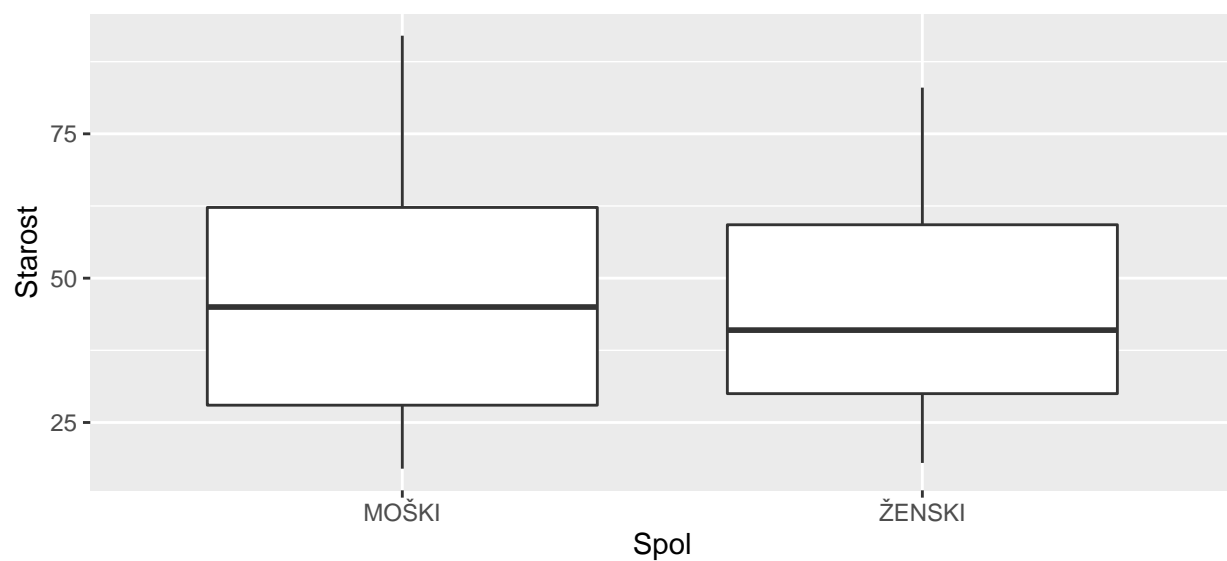
²<https://pxweb.stat.si/SiStatData/pxweb/sl/Data/-/05C5002S.px/>



Slika 3: Porazdelitev starosti

Tabela 2: Opisne statistike starosti

	Skupno	Moški	Ženske
Min.	17.0	17.0	18.0
1st Qu.	28.0	28.0	30.0
Median	44.0	45.0	41.0
Mean	45.6	46.1	44.5
3rd Qu.	61.0	62.2	59.2
Max.	92.0	92.0	83.0
IQR	33.0	34.2	29.2

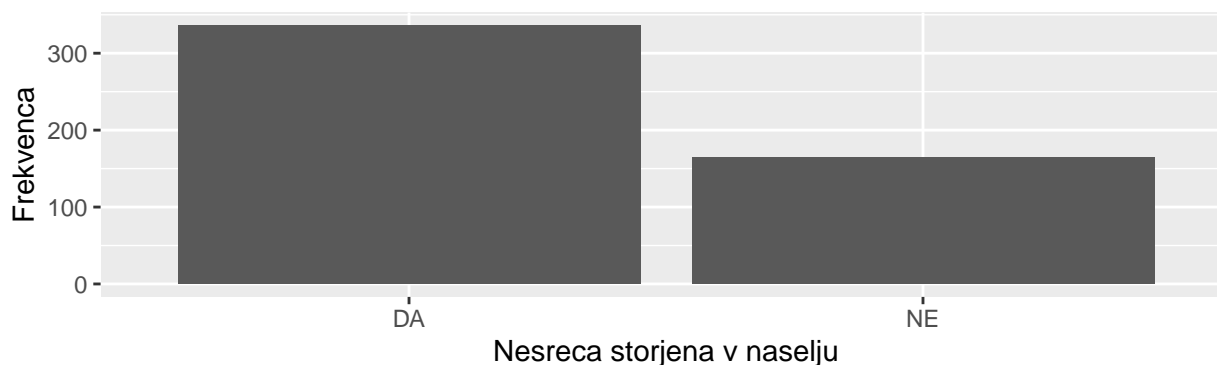


Slika 4: Škatle z brki za starost po spolu.

V naselju

Vidimo da se v našem vzorcu zgodi več nesreč v naselju (V naselju - 336, Izven naselja - 164). To sem tudi pričakoval, saj mislim da se največ vožnje z avtomobilom zgodi ravno v naselju. Podoben razmislejš je da tudi zato, ker je tam največ ljudi in avtov.

V tabeli 3 si oglejmo kontingenčno tabelo opazovanih frekvenc za spol povzročiteljev nesreč, ter ali je bila nesreča v naselju, ali ne. Po občutku bi rekel da večjih odstopanj od pričakovanj (če gledamo vsote) ni. To bomo v naslednjem poglavju preverili s testom χ^2 .

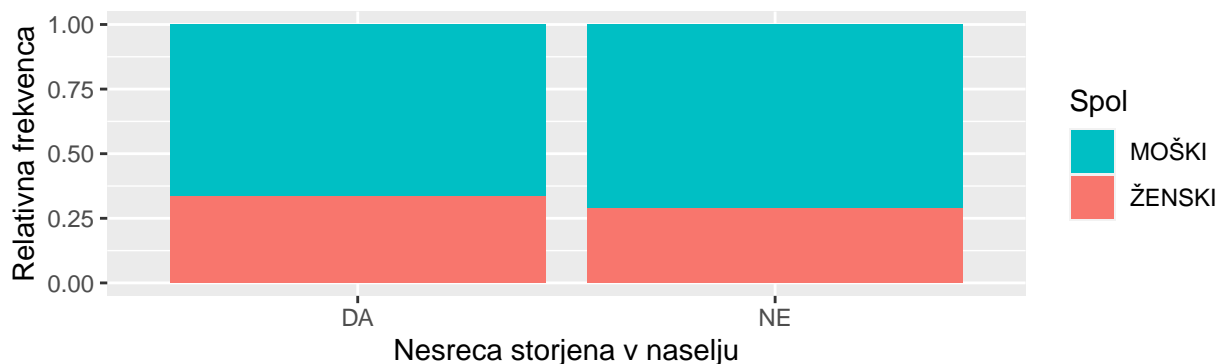


Slika 5: Frekvenca ali je bila nesreča storjena v naselju ali ne

Tabela 3: Kontingenčna tabela za spol in ali je bila nesreča storjena v naselju

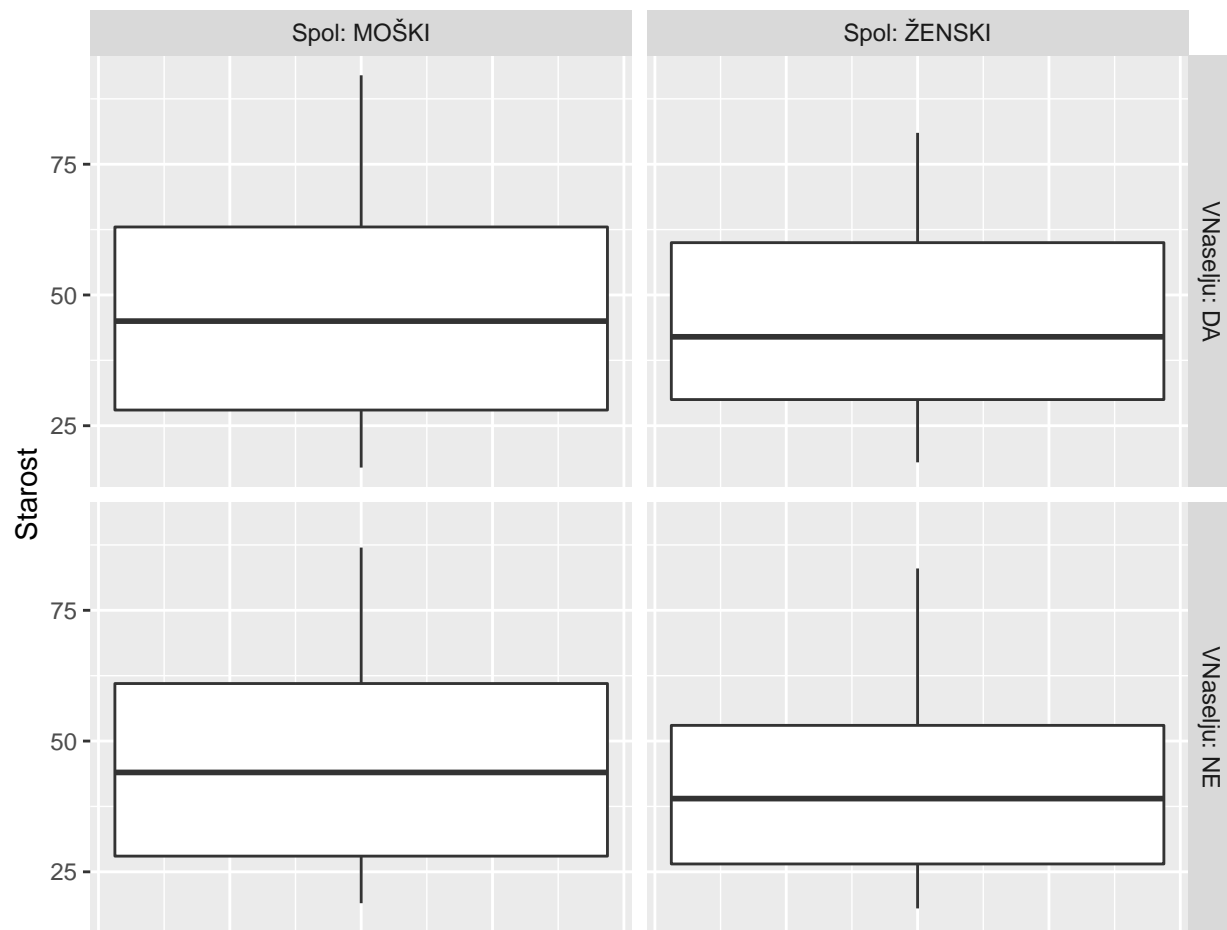
	MOŠKI	ŽENSKI	VSOTA
DA	223	113	336
NE	117	47	164
VSOTA	340	160	500

Podobno lahko sklepamo iz slike 6, saj vidimo da nekih velikih razlik v relativnih frekvencah ni.



Slika 6: Relativen prikaz, ali je bila nesreča storjena v naselju ali ne, po spolu.

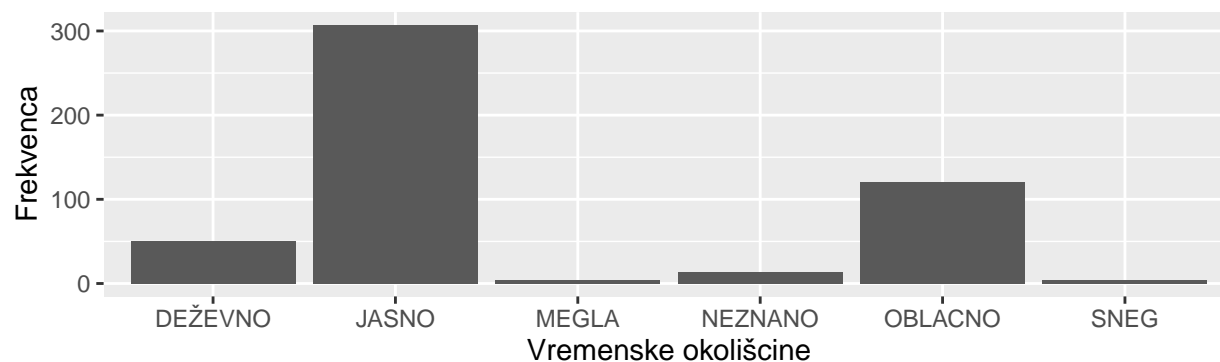
Na sliki 7 je prikazana porazdelitev starosti znotraj vsake izmed skupin. Če primerjamo s sliko 4, večjih odstopanj v starosti ni. Ker so skupine dovolj velike ($n \geq 30$), bomo lahko s t -testom preverili če se povprečje starosti bistveno razlikuje med spoloma.



Slika 7: Relativen prikaz, ali je bila nesreča storjena v naselju ali ne, po spolu.

VremenskeOkoliscine

Na sliki 8 so prikazane absolutne frekvence nesreč v različnih vremenskih okoliščinah. Vidimo da se največ nesreč zgodi pri vremenu JASNO



Slika 8: Frekvence vremenskih okoliščin

Tabela 4: Kontingenčna tabela za spol in vremenske okoliščine

	MOŠKI	ŽENSKI	VSOTA
DEŽEVNO	36	14	50
JASNO	206	101	307
MEGLA	3	1	4
NEZNANO	7	7	14
OBLAČNO	86	35	121
SNEG	2	2	4
VSOTA	340	160	500

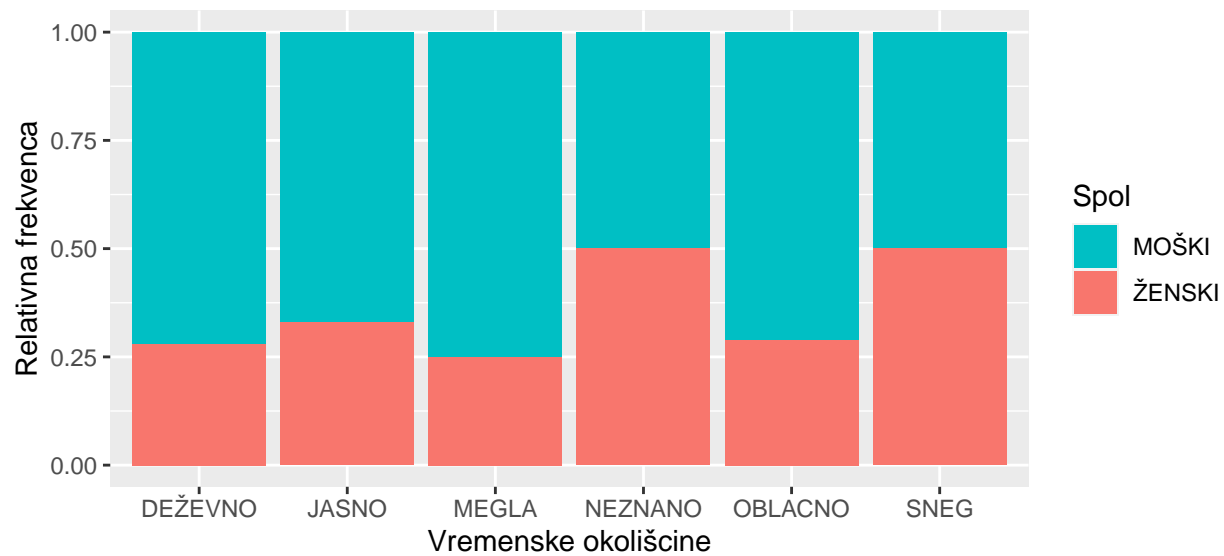
StanjePrometa

Tabela 5: Kontingenčna tabela za spol in stanje prometa

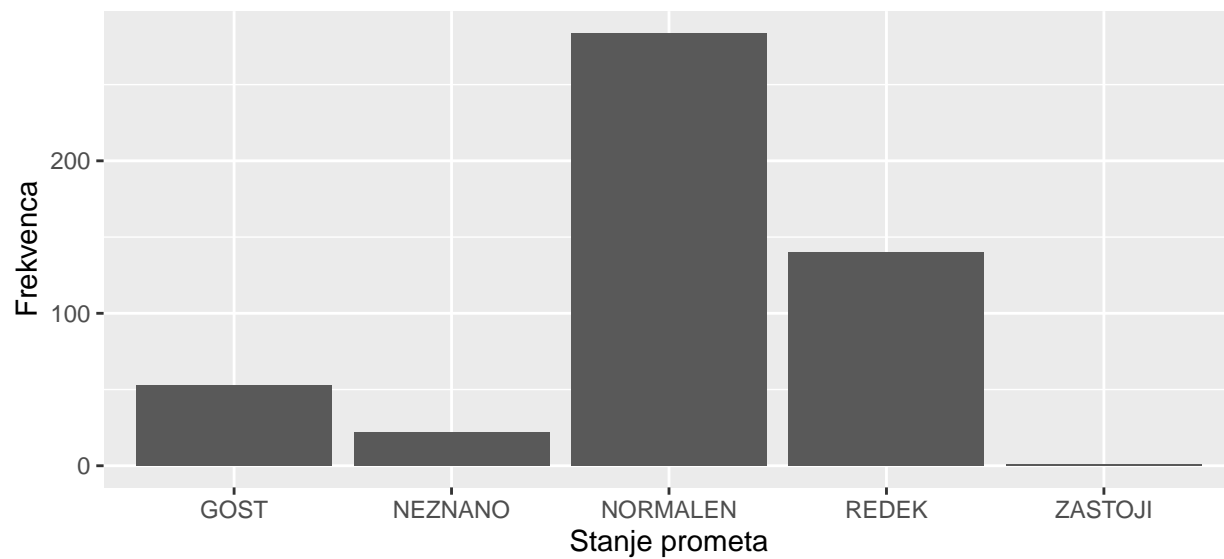
	MOŠKI	ŽENSKI	VSOTA
GOST	39	14	53
NEZNANO	10	12	22
NORMALEN	189	95	284
REDEK	101	39	140
ZASTOJI	1	0	1
VSOTA	340	160	500

StanjeVozisca

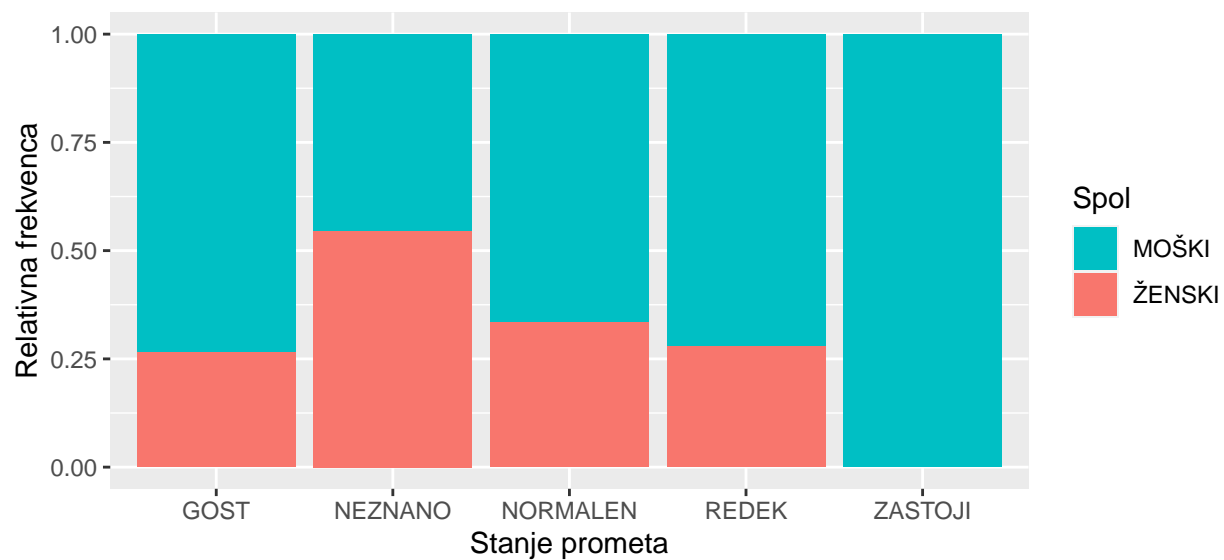
VrstaVozisca



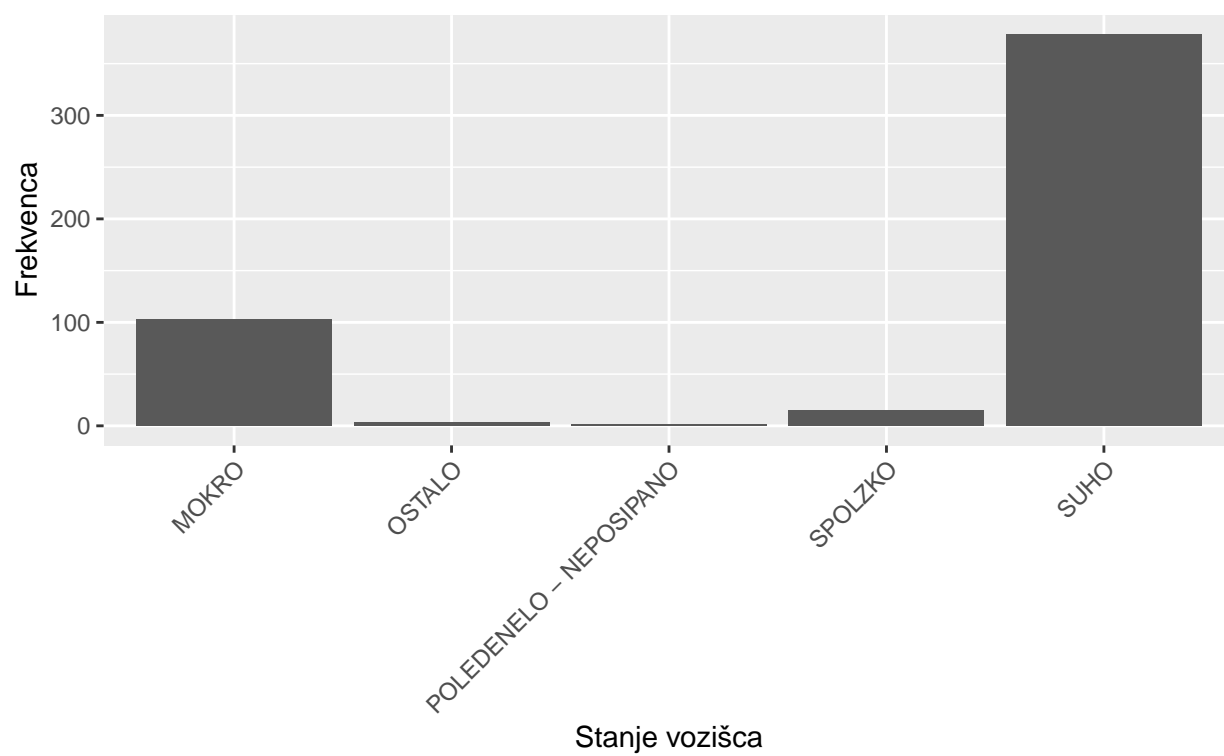
Slika 9: Relativen prikaz vremenskih razmer po spolu.



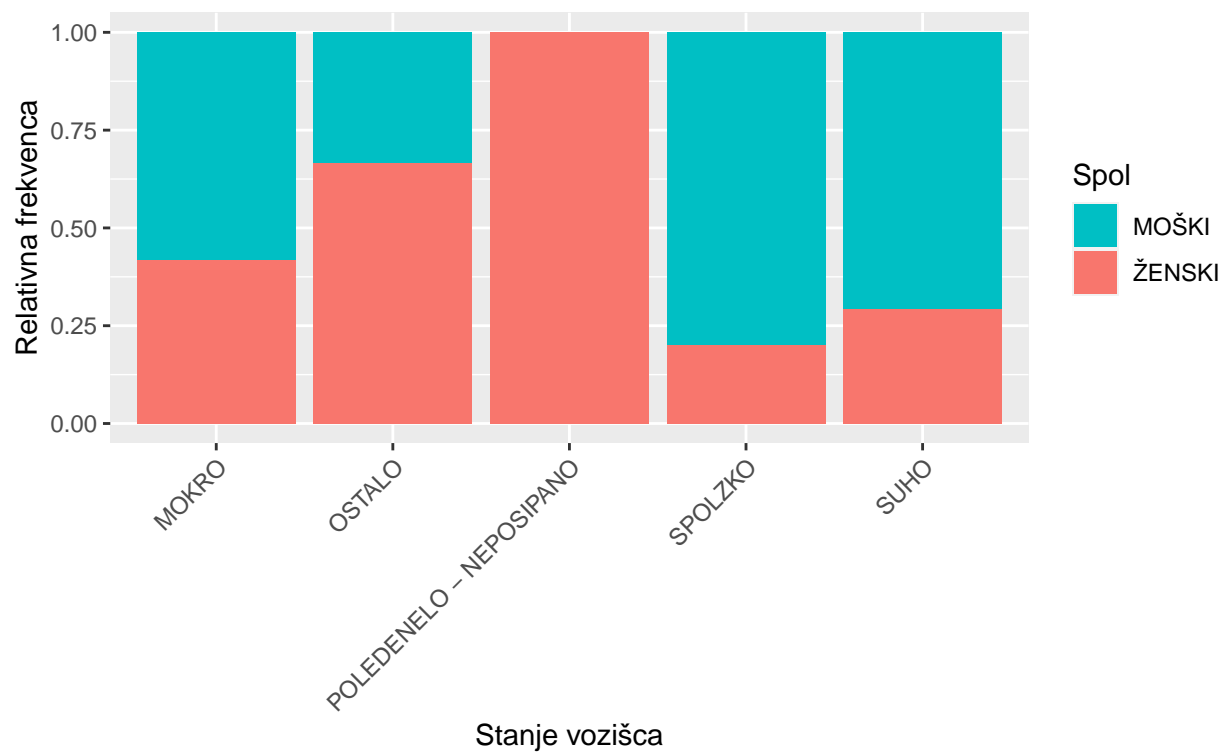
Slika 10: Frekvence stanj prometa



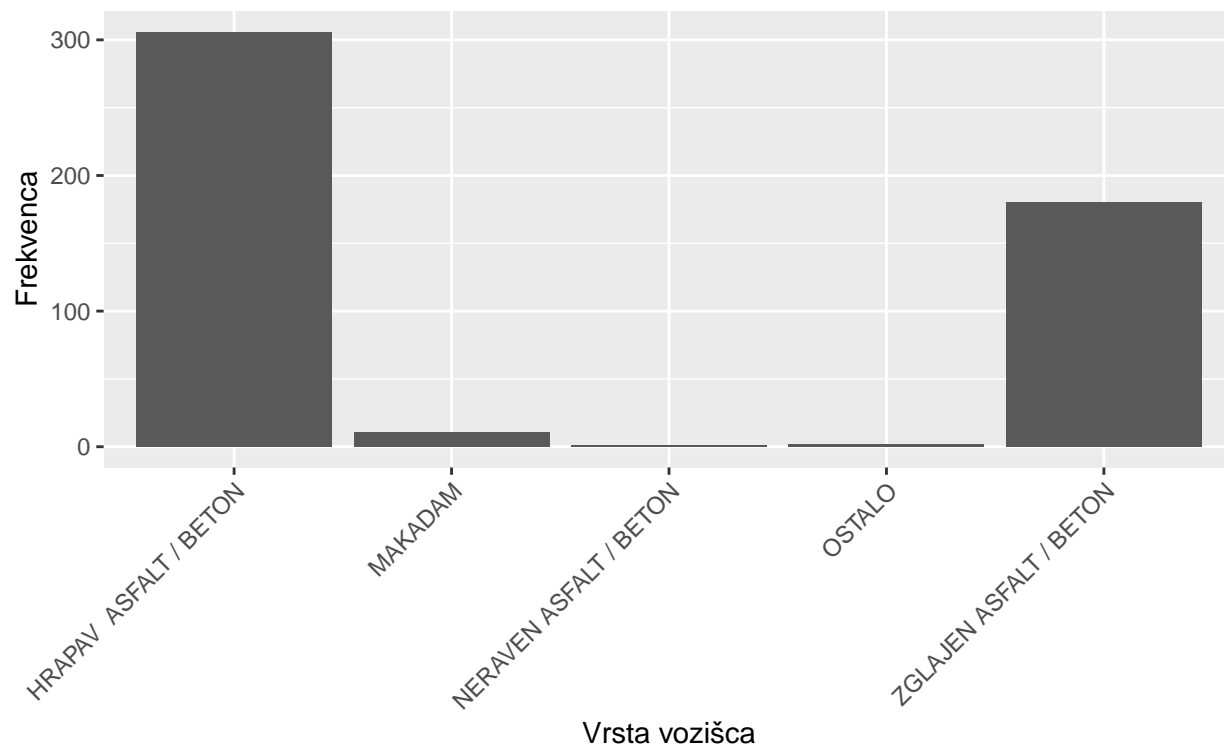
Slika 11: Relativen prikaz stanj prometa po spolu.



Slika 12: Frekvence stanj vozišča



Slika 13: Relativen prikaz stanj vozišča po spolu.



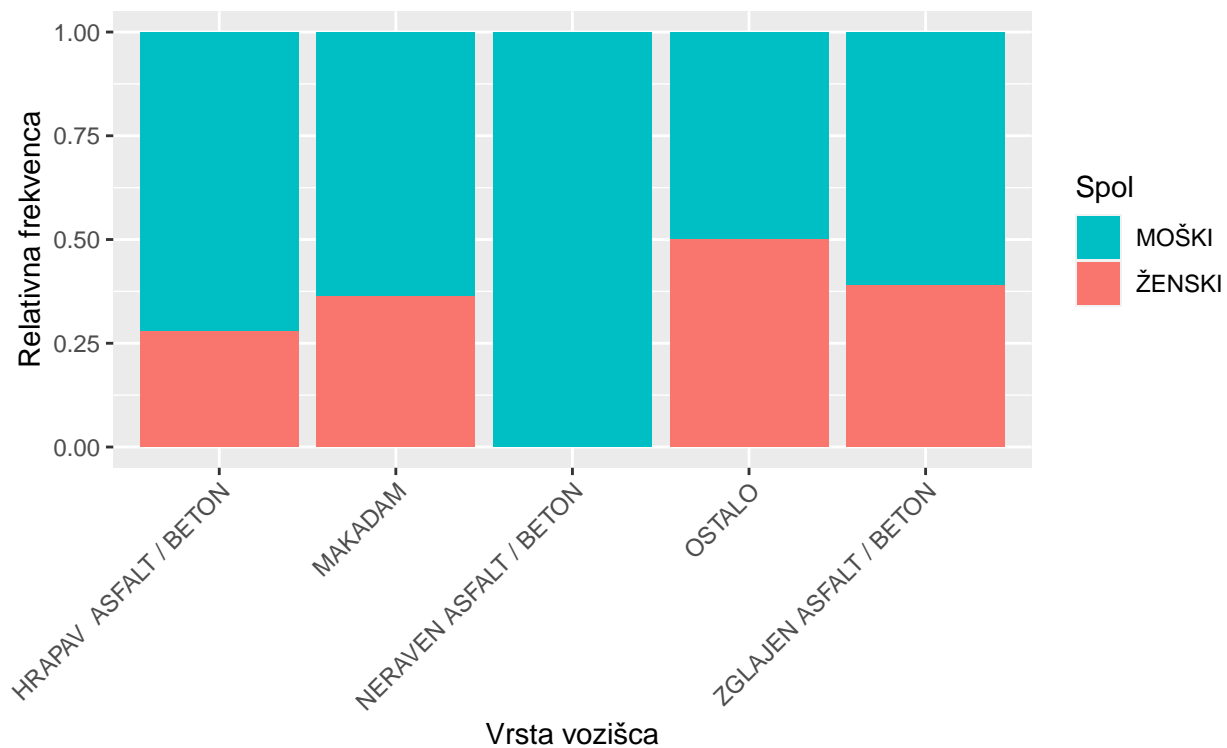
Slika 14: Frekvence vrst vozišča

Tabela 6: Kontingenčna tabela za spol in stanje vozišča

	MOŠKI	ŽENSKI	VSOTA
MOKRO	60	43	103
OSTALO	1	2	3
POLEDENELO - NEPOSIPANO	0	1	1
SPOLZKO	12	3	15
SUHO	267	111	378
VSOTA	340	160	500

Tabela 7: Kontingenčna tabela za spol in vrsto vozišča

	MOŠKI	ŽENSKI	VSOTA
HRAPAV ASFALT / BETON	221	85	306
MAKADAM	7	4	11
NERAVEN ASFALT / BETON	1	0	1
OSTALO	1	1	2
ZGLAJEN ASFALT / BETON	110	70	180
VSOTA	340	160	500



Slika 15: Relativen prikaz vrst vozišča po spolu.