



VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Imčių metodai

Projektinis darbas

Darbą atliko:

A.Užpelkis, R.Gulbinovič, M.Kamarauskas, M.Amšiejus

Duomenų mokslas II kursas, 2 grupė

Recenzentai:

T.Ramanauskas, I.Silkauskaitė, J.Sergejeva, A.Gervickaitė

Vilnius 2021

Turinys

IVADAS	3
DUOMENYS	3
TYRIMO ETAPAI	3
PASIRINKTI METODAI	3
SVORIŲ SKAIČIAVIMAS	4
PARAMETRŲ VERTINIMAS	4
REZULTATAI	4
IŠVADOS	7
R KODAS	8
Šaknies iš sukaupųjų dažnių metodas.....	8
Proporcingas skirstymas į sluoksnius	9
Elementų svorių nustatymas	9
Imties išrinkimas	10
Aprašomas imties planas	10
Masės, greičio ir galios vidurkių vertinimas	10
Parametrų vertinimas pagal kuro tipą.....	11

ĮVADAS

Automobiliai mūsų gyvenimuose tapo kasdienybe. Šiais laikais dažnas gali sau leisti būti šios transporto priemonės savininku. Kiekvienas žmogus savaip naudoja mašiną, o tai gali privesti prie įdomios informacijos, tendencijų. Dėl šios priežasties mes bandysime išvelgti kaip naudojamos šios transporto priemonės Lietuvoje. Tam išsikelsime kelis uždavinius:

1. Įvertinti pagrindinių automobilių charakteristikų (greitis, masė, galia) vidurkius;
2. Nustatyti vidutinę automobilio ridą pagal jo kuro tipą;
3. Įvertinti automobilių kiekį pagal kuro tipą.

DUOMENYS

Duomenys buvo paimti iš valstybinės įmonės „Regitra“. Duomenys yra atviri, juose žymima fizinių asmenų įregistruotų kelių transporto priemonių parko nuolat atnaujinama informacija. Mūsų duomenys paskutinį kartą atnaujinti buvo 2021 metų 04 mėnesio 02 dieną. Tyrime naudojome tik M1 tipo automobilius (iki 9 vietų su vairuotoju, mažiau nei 3,5t masė). Tyrimo kintamieji buvo: kėbulo tipas, nuosava masė, galia (Kw), degalai, maksimalus greitis, sėdimų vietų skaičius, rida, pirmoji registracijos data, savivaldybė.

Šaltinis: [Regitra](#)

TYRIMO ETAPAI

1. Sutvarkome duomenis, atmetame negalimas reikšmes;
2. Pasirenkame ėmimo būdą;
3. Išrenkame imtį pagal pasirinktą būdą;
4. Įvertiname norimus parametrus

PASIRINKTI METODAI

Imčiai išrinkti nusprendėme taikyti paprastąjį atsitiktinį sluoksningį ėmimą. Tai reiškia, kad mūsų turimą mašinų populiaciją teks skirstyti į sluoksnius. Nusprendėme, kad tikslingiausia į sluoksnius skirstyti pagal automobilio pirmąją registracijos datą (darėme išvadą, kad pirmoji registracijos data tiesiogiai siejasi su automobilio pagaminimo data). Kadangi registracijos data nėra

kategorinis kintamasis, pasirinkome taikyti šaknies iš sukaupųjų dažnių (Dalenius & Hodges) metodą. Šis metodas padeda rasti tokias sluoksnių ribas, kad įvertinio dispersija, esant optimaliam paskirstymui, būtų minimali. Gauname, kad automobiliai bus suskirstyti į 3 intervalus: nuo 1950 iki 2000, nuo 2000 iki 2010 ir nuo 2010 iki 2021. [Kodas](#)

Skirstant į sluoksnius reikėjo nustatyti jų dydžius. Tarėme, kad turime galimybę ištirti $n=5000$ mašinų bei mašinos tyrimo kaina yra vienoda. Pasirinkome mašinas skirstyti į sluoksnius proporcingai sluoksnių didumui populiacijoje, nes papildoma informacija nebuvo turima, o kaina buvo vienoda. Gavome, kad sluoksnių dydžiai atitinkamai metams yra 264, 3664 ir 1072. [Kodas](#)

SVORIŲ SKAIČIAVIMAS

Tam, kad galėtume naudoti paketą „Survey“ reikia suskaičiuoti kiekvieno elemento svorį:

$$w_i = \frac{N_h}{n_h},$$

kur w_i yra elemento (mašinos) svoris, N_h - mašinų skaičius h populiacijos sluoksnyje, n_h - mašinų skaičius h sluoksnių imtyje. Gauname, kad pirmajame sluoksnyje svoris bus 30.56818, antrame - 30.56632, trečiame - 30.55317. [Kodas](#)

PARAMETRŲ VERTINIMAS

Pasinaudojus „sampling“ paketą išrenkame atsitiktinę sluoksnių imtį. [Kodas](#).

Parametrus vertinti naudojame „survey“ paketą, kuriam reikalingas ėmimo plano sudarymas. [Kodas](#).

Turint imtį ir ėmimo planą, galima vertinti norimus parametrus:

- vidurkius (maksimalaus greičio, masės, galios). [Kodas](#).
- pagal kuro tipą įvertinti automobilių kiekį ir ridos vidurkį. [Kodas](#).

REZULTATAI

Įvertintas maksimalaus greičio vidurkis:

Vidurkis: 195.57189 km/h. 95% pasikliautinis intervalas: (194.9884; 196.1554) km/h. Galime laikyti, kad automobilių populiacijos maksimalaus greičio vidurkis bus intervale (194; 197) km/h.

Ivertintas galios vidurkis:

Vidurkis: 101.94765 kW. 95% pasikliautinis intervalas: (101.0597; 102.8356) kW. Galime laikyti, kad automobilių populiacijos galios vidurkis bus intervale (101; 103) kW.

Ivertintas masės vidurkis:

Vidurkis: 1595.5480 kg. 95% pasikliautinis intervalas: (1587.977 1603.119) kg. Galime laikyti, kad automobilių populiacijos masės vidurkis bus intervale (1587, 1604) kg.

Ivertintas automobilių kiekis, pagal kėbulo tipą:

KEB_PAVADINIMAS	sk	se
Daugiatikslis	22462.62600	741.09848
Hečbekas	7243.31263	450.92697
Kabrioletas	244.50427	84.96588
Kemperiai	91.68581	52.05686
Kupė	733.46021	146.83638
Sedanas	7548.59066	458.17257
Universalas	39942.74456	881.42501

Gauname, kad populiacijoje populiariausi universalai, antroje vietoje daugiatikslis kėbulo tipo automobiliai. Mažiausiai populiarius yra kemperiai ir kabrioletai.

Ivertintas automobilių kiekis, pagal kuro tipą:

DEGALAI	sk	se
Benzinas	27355.0558	806.95643
Benzinas / Dujos	5562.9591	396.41902
Benzinas / Elektra	1466.7758	206.16385
Benzinas / Elektra / Dujos	275.0311	90.05458
Benzinas / Etanolis	152.8185	67.19313

Benzinas / Gamtinės_dujos	122.2390	60.09304
Dyzelinas	117730.3548	883.87579
Dyzelinas / Elektra	152.7659	67.06604

Gauname, kad populiariausi automobiliai populiacijai yra dyzeliniai. Benzininių yra netgi apie 5 kartus mažiau. Dėl įdomumo palyginkime su turėta populiacija:

--	Benzinas	Benzinas / Dujos	Benzinas / Elektra	Benzinas / Elektra / Dujos
3	27524	5225	1845	238
Benzinas / Etanolis		Benzinas / Etanolis / Dujos		Benzinas / Gamtinės_dujos
110		16		65
Dujos	Dyzelinas	Dyzelinas / Dujos	Dyzelinas / Elektra	Elektra
16	117609	1	78	71
Gamtinės_dujos				
16				

Išties populiacijoje didžiąją dalį sudaro dyzeliniai varikliai, (skirtumas tarp mūsų sumos įverčio ir populiacijos tik 121).

Ivertintas ridos vidurkis:

Vidurkis: 218371.0 km, 95% pasikliautinis intervalas: (208440.3; 228301.7) km.

Ivertintas automobilių ridos vidurkis, pagal kuro tipą:

DEGALAI	RIDA	se
Benzinas	186526.7	16326.124
Benzinas / Dujos	209426.6	5053.499
Benzinas / Elektra	116520.3	13782.120
Benzinas / Elektra / Dujos	223865.9	18201.897
Benzinas / Etanolis	193094.6	12259.445
Benzinas / Gamtinės_dujos	181989.8	25613.121
Dyzelinas	227573.7	5369.942

Dyzelinas / Elektra 176620.2 43918.716

Automobilių, varomų dyzelinu, benzinu/elektra/dujomis, benzinu/dujomis, vidutinės ridos yra didžiausios, viršijančios 200 000 km.

Įvertintas automobilių ridos vidurkis, pagal sėdimų vietų skaičių:

SEDIMU_VIETU_SK	RIDA	se
2	173437.1	16456.286
3	180931.0	0.000
4	171552.3	6779.284
5	218929.4	6124.000
6	245228.2	20791.078
7	220969.8	4050.334
8	220513.6	12736.708
9	225707.1	8992.314

Automobiliai, turintys daugiau nei 4 sėdimas vietas (su vairuotoju) vidutiniškai nuvažiuoja daugiau nei 200 000 km. Ypatingai didelis 6 vietų automobilių nuvažiuotų kilometrų skaičius, tačiau jo standartinis nuokrypis taip pat žymiai didesnis nei kitų automobilių, todėl drąsiai tvirtinti negalime, kad šešiaviečiai automobiliai nuvažiuoja daugiausiai.

IŠVADOS

Patikrinus kiekvieną įvertinį su variacijos koeficiento įvertiniu, gauname, kad nei vienas nesiekia 5%. Taip pat suskaičiavus plano efektą, visur gaunama, kad $pe(\hat{\theta}) < 1$, vadinasi mūsų paprastoji sluoksninė imtis išties gavosi efektyvesnė nei paprastoji atsitiktinė imtis. Iš gautų rezultatų matome, kad tirta populiacija mėgsta sąlyginai negalingas mašinas (apie 102Kw), bet gana greitas – maksimalus greitis apie 196 km/h. Populiariausias kėbulo tipas populiacijoje yra universalas ir daugiatikslis (praktiškai automobiliai su daug erdvės), mažiausiai populiarūs – kemperiai (brangūs, Lietuvoje dar nepopuliarūs) ir kabrioletai (nelabai tinkami lietuviškam orui, per daug egzotiniai). Daugiausia populiacijoje būta dyzelinių variklių (potencialiai dėl pigaus kuro), paskui benzininių. Hibridų/elektrinių automobilių populiacijoje yra tik nedidelė dalis. Vidutiniškai

rida buvo įvertinta 218000 km. Taip pat matome, kad su pigesniais kuro šaltiniais (dyzelinu, elektra/benzinu, dujomis) yra nuvažiuojami didesni atstumai. Tie, kas nusiperka taupesnius automobilius, išnaudoja tą faktą ir nuvažiuoja ilgesnius atstumus. Taip pat tie, kurie perka talpesnius automobilius (5 ir daugiau sėdimų vietų) keliauja daugiau. Taigi galime daryti išvadą, kad populiacijoje tirti asmenys labiau renkasi praktiškus, taupesnius automobilius, su kuriais galima pigiau nuvažiuoti didesnius atstumus bei sutalpinti daugiau bagažo.

R KODAS

Šaknies iš sukaupųjų dažnių metodas

```
#Populiacijos dydis
#1
#Norime nustatyti sluoksniu dydžius.
#Tam naudosime saknies iš sukaupųjų dažnių metoda (Dalenius & Hodges).
#Pirma skirstome mūsų visą amžių į intervalus po 10 (metų) ir skaičiuojame
#pasitaikiusių elementų kieki juose:
dazniai<-c()
for(i in seq(1950, 2020, 10)){
  daznis<-nrow(subset(masinos, Metai>=i & Metai<(i+10)))
  dazniai<-c(dazniai, daznis)
}
#randame dažnių saknis
sak_dazniai<-sqrt(dazniai)
#randame sukaupųjų dažnių saknis
suk_dazniai<-cumsum(sak_dazniai)
#mūsų norimos ribos:
riba<-suk_dazniai[length(suk_dazniai)]/3;riba

210.5569

riba*2

421.1137

suk_dazniai

1.000000    3.000000    6.464102   13.744212  103.186931  437.843471  618.328927
631.670591
```



```
#gauname, kad imant kas 10 metu ir norint suskirstyti i tris panasaus dydzio
intervalus, gauname, kad automobiliai iki
#2000 metu pakliuna i pirma intervala, iki 2010 - antra, o visi velesni - 3
masinos$sluoksniai<-ifelse(masinos$Metai<2000, 1, ifelse(masinos$Metai<2010, 2,
3))
```

Proporcingas skirstymas į sluoksnius

```
#2
#Tarkime galime imti 5000 masinu imti (n=5000). Reikia rasti n_h (h=1,2,3), t.y.
kiekvieno sluoksnio imties dydi
#Sluoksniu imtis skirstysime proporcingai pagal sluoksnio diduma:
```

```
n<-5000
```

```
N<-nrow(masinos)
```

```
152818
```

```
N1<-nrow(subset(masinos, masinos$sluoksniai == 1))
```

```
n1<-round(n*N1/N)
```

```
264
```

```
N2<-nrow(subset(masinos, masinos$sluoksniai == 2))
```

```
n2<-round(n*N2/N)
```

```
3664
```

```
N3<-nrow(subset(masinos, masinos$sluoksniai == 3))
```

```
n3<-round(n*N3/N)
```

```
1072
```

```
n1+n2+n3
```

```
5000
```

Elementų svorių nustatymas

```
#3
```

```
#Randame emimo svorius
```

```

masinos$svoris <- ifelse(masinos$sluoksniai == 1, N1/n1,
ifelse(masinos$sluoksniai == 2, N2/n2, N3/n3))
#Planui reikes populiacijos sluoksnio dydzio, taigi priskiriame:
masinos$N_h <- ifelse(masinos$sluoksniai == 1, N1, ifelse(masinos$sluoksniai ==
2, N2, N3))

```

Imties išrinkimas

```

#4
#Renkame paprastąją atsitiktinę sluoksninę imtį
#Pirma surusiuojame elementus pagal sluoksni
sorted_masinos <- masinos[order(masinos$sluoksniai),]
#Randame elemnto patekimo i imtį tikimybę
sorted_masinos$pik <- inclusionprobastrata(sorted_masinos$sluoksniai,
c(n1,n2,n3))
#Atrenkame pagalbinę imtį, kuri nurodo duomenų indeksus
pag_imtis_duom <-sampling::strata(sorted_masinos, stratanames=c("sluoksniai"),
c(n1,n2,n3), method="srswor",
sorted_masinos$pik, description = TRUE )
#konvertuojame pagalbinę imtį i duomenis
imtis_duom <-getdata(sorted_masinos, pag_imtis_duom)
imtis_duom$sk<-1

```

Aprašomas imties planas

```

#5
#Sukuriame imties planą
plan_sluks_masinos<-svydesign(ids = ~1, strata = ~sluoksniai, weights =
~svoris, data = imtis_duom, fpc = ~pik)

```

Masės, greičio ir galios vidurkių vertinimas

```

#6
#Ieskome parametru įvertinių
#Galios vidurkis
galios_vid<-svymean(~GALIA, plan_sluks_masinos, deff=T);galios_vid
confint(galios_vid)#pasikliovimo intervalas (95%)
cv(galios_vid)#variacijos koeficiento įvertinys
#maksimalaus greičio vidurkis
greic_vid<-svymean(~MAKS_GREITIS, plan_sluks_masinos, deff=T);greic_vid
confint(greic_vid)
cv(greic_vid)
#vidutine automobiliu mase

```

```

mase<-svymean(~NUOSAVA_MASE, plan_sluks_masinos, deff = T);mase
confint(mase)
#vidutine automobiliu rida
rida<-svymean(~RIDA, plan_sluks_masinos, deff=T);rida
confint(rida)
cv(rida)
rida_kuras<-svyby(~RIDA, ~DEGALAI, plan_sluks_masinos, svymean);rida_kuras
rida_vietos<-
svyby(~RIDA, ~SEDIMU_VIETU_SK, plan_sluks_masinos, svymean);rida_vietos

```

Parametrų vertinimas pagal kuro tipą

```

#masinu skaicius
svyttotal(~sk, plan_sluks_masinos)
svyby(~sk, ~MARKE, plan_sluks_masinos, svyttotal)

#masinos pagal degalus
svyby(~sk, ~DEGALAI, plan_sluks_masinos, svyttotal)
#palyginimui galime paskaiciuoti realius populiacijos duomenis
table(masinos$DEGALAI)

#vertiname automobiliu skaiciu pagal kebuki tipa
sk_kebulas<-svyby(~sk, ~KEB_PAVADINIMAS, plan_sluks_masinos, svyttotal);sk_kebulas

```