

Vilniaus Universitetas
Matematikos ir informatikos fakultetas
Ekonometrinės analizės katedra

Kursinis darbas

Įskaitinių eismo įvykių Lietuvos keliuose analizė

Atliko: Linas Šyvis ir Kornelijus Samsonas

Vadovas: prof. Vydas Čekanavičius

VILNIUS 2016

Turinys

Tvarkome duomenis	3
Įvykių sk. 1000-iui gyventojų didžiausiuose miestuose grafikas.	7
Tikrinam įvykių priklausomybę nuo amžiaus	9
Forecastinam autoįvykius	9
Kaltininku procentas miestuose (Vyrų/Moterys)	9
koreliacijos	9
Literatūra	10

Reikalingi užduotims paketai.

```
library(fpp)
library(dplyr)
```

Nuskaitome duomenis.

```
rawdata <- read.csv2("2003-2015.csv", header = T)
agreguoti <- read.csv2("agreguoti2009-2015.csv", header = T)
gyventojai <- read.csv2("gyventojai.csv", header = T, skip = 1)
kor <- read.csv2("koreliacijos.csv", header = T)
```

Apžvelgiame duomenis

```
sapply(c(rawdata, agreguoti, gyventojai), head)
```

```
##      Vieta Data Laikas Dalyviu_skaicius Zuvusiu_skaicius Suzeistu_skaicius
## [1,]      1  20   1442                1                0                1
## [2,]      1  22   1442                2                0                2
## [3,]      1  26   3113                2                0                2
## [4,]      1  57  5797                2                0                1
## [5,]      1  75   754                2                0                1
## [6,]      1  76  5459                2                0                1
##      Dalyvavusiu_transporto_priemoniu_skaicius
## [1,]
## [2,]
## [3,]
## [4,]
## [5,]
## [6,]
##      Apgadintu_transporto_priemoniu_skaicius Kelio_reiksme Dangos_rusis
## [1,]
## [2,]
## [3,]
## [4,]
## [5,]
## [6,]
##      Dangos_bukle Paros_metas Meteorologines_salygos Kelio_kreive
## [1,]
## [2,]
## [3,]
## [4,]
## [5,]
## [6,]
##      Atitvarai Eismo_ivykio_dalyvio_kategorija
## [1,]
## [2,]
## [3,]
## [4,]
## [5,]
## [6,]
##      KET_pazeidimas_del_kurio_ivyko_eismo_ivykis Lytis Pilietybe
## [1,]
```

```
## [2,] 32 3 19
## [3,] 32 3 19
## [4,] 46 3 19
## [5,] 32 3 19
## [6,] 46 3 19
## Gimimo_data Neigalumas Saugos_dirzas Oro_pagalviu_suveikimas
## [1,] 6250 0 3 NA
## [2,] 4286 0 NA NA
## [3,] 8136 0 NA NA
## [4,] 2163 0 NA NA
## [5,] 14506 0 NA NA
## [6,] 4206 0 NA NA
## Vairuotojo_kvalifikacija Stazas Blaivumas Kaltas Metai Data Vieta
## [1,] 1 12 3 1 2009 1 28
## [2,] 1 32 3 1 2009 1 2
## [3,] 1 22 5 1 2009 1 3
## [4,] NA NA 5 1 2009 1 4
## [5,] 1 4 3 1 2009 1 8
## [6,] NA NA 5 1 2009 1 27
## Visi_ivykiai Del_neblaiviu_vairuotoju_kaltes Suzeistieji Zuvusieji
## [1,] 262 17 303 28
## [2,] 13 0 12 3
## [3,] 7 0 8 0
## [4,] 1 0 0 1
## [5,] 1 0 2 0
## [6,] 3 0 1 2
## Vieta Data Visi Vyrai Moterys
## [1,] 3 2003 3431497 1601098 1830399
## [2,] 3 2004 3398929 1584018 1814911
## [3,] 3 2005 3355220 1562264 1792956
## [4,] 3 2006 3289835 1528463 1761372
## [5,] 3 2007 3249983 1506794 1743189
## [6,] 3 2008 3212605 1487419 1725186
```

Tvarkome duomenis

Paliekame tik 5 didžiuosius miestus

```
miestai <- c("Vilniaus m. sav.", "Kauno m. sav.", "Klaipėdos m. sav.", "Siauliu m. sav.",
             "Panevezio m. sav.")
data <- subset(rawdata, Vieta %in% miestai)
```

Turime stulpelius “Gimimo data” ir “Stazas”. Jų reikšmės tikėtina stipriai koreliuoja, patikrinkime. Pertvarkysime stulpelio “Gimimo data” reikšmes taip, kad matytume eismo įvykio dalyvio amžių ir patikrinsime amžiaus koreliaciją su “Stažo” reikšmėmis.

Stulpelį “Gimimo data” pertvarkome į “Amžius”.

```
Amzius <- as.numeric(format(Sys.Date(), format = "%Y")) - as.numeric(format(as.Date(data$Gimimo_data,
                                         format = "%Y"), format = "%Y"))
data <- data[, -22]
data <- cbind(data, Amzius)
head(data$Amzius)
```

```
## [1] 36 45 73 33 39 40
```

Dabar turime eismo įvykių dalyvių amžių, kuro koreliaciją su stažu galime patikrinti.

Patikrinsime koreliaciją.

```
sum(is.na(data$Stazas))
```

```
## [1] 8184
```

```
sum(is.na(data$Amzius))
```

```
## [1] 1316
```

```
cor(data$Amzius, data$Stazas, use = "pairwise.complete.obs")
```

```
## [1] 0.7957286
```

Matome, kad koreliacijos koeficientas gana didelis, todėl galėsime pašalinti amžiaus arba stažo stulpelį.

```
table(data$Blaivumas)
```

```
##
## Apsvaiges nuo narkotiniu, psichotropiniu ar kitu psichika veikianciu medziagu
##                                     14
##                               Atsisake buti patikrintas
##                                     12
##                               Blaivus
##                               21615
##                               Neapsvaiges
##                                     0
##                               Neblaivus
##                               2218
```

Matome, kad stulpelį “Blaivumas” sudaro 6 skirtingos reikšmės. Sutrauksime jas visas į “Blaivus” ir “Neblaivus”.

```
data[data[, "Blaivumas"] %in% c("Apsvaiges nuo narkotiniu, psichotropiniu ar kitu psichika veikianciu m
    "Atsisake buti patikrintas"), "Blaivumas"] <- "Neblaivus"
data$Blaivumas <- factor(data$Blaivumas)
table(data$Blaivumas)
```

```
##
##   Blaivus Neblaivus
##   21615      2244
```

Šalinsime nereikalingas reikšmes.

```
sapply(lapply(data, is.na), sum)
```

```
##                               Vieta
##                               0
##                               Data
##                               0
##                               Laikas
##                               0
##                               Dalyviu_skaicius
##                               0
##                               Zuvusiu_skaicius
##                               0
##                               Suzeistu_skaicius
##                               0
##   Dalyvavusiu_transporto_priemoniu_skaicius
##                               0
##   Apgadintu_transporto_priemoniu_skaicius
##                               1019
##                               Kelio_reiksme
##                               23690
##                               Dangos_rusis
##                               0
##                               Dangos_bukle
##                               0
##                               Paros_metas
##                               0
##                               Meteorologines_salygos
##                               0
##                               Kelio_kreive
##                               0
##                               Atitvarai
##                               0
##   Eismo_ivykio_dalyvio_kategorija
##                               56
## KET_pazeidimas_del_kurio_ivyko_eismo_ivykis
##                               0
##                               Lytis
##                               300
##                               Pilietybe
##                               1038
##                               Gimimo_data
##                               1316
##                               Neigalumas
##                               1482
##   Oro_pagalviu_suveikimas
##                               18134
##   Vairuotojo_kvalifikacija
##                               6893
##                               Stazas
##                               8184
##                               Blaivumas
##                               265
##                               Kaltas
```

```
##                                0
##                                Amzius
##                                1316
```

Kai kurie stulpeliai turi NA reikšmių, reikia jas pašalinti.

```
salinamieji <- c("Lytis", "Apgadintu_transporto_priemoniu_skaicius", "Pilietybe",
  "Neigalumas", "Blaivumas", "Amzius")
ndata <- data[complete.cases(data[salinamieji]), ]
sapply(lapply(ndata, is.na), sum)
```

```
##                                Vieta
##                                0
##                                Data
##                                0
##                                Laikas
##                                0
##                                Dalyviu_skaicius
##                                0
##                                Zuvusiu_skaicius
##                                0
##                                Suzeistu_skaicius
##                                0
##    Dalyvavusiu_transporto_priemoniu_skaicius
##                                0
##    Apgadintu_transporto_priemoniu_skaicius
##                                0
##                                Kelio_reiksme
##                                20823
##                                Dangos_rusis
##                                0
##                                Dangos_bukle
##                                0
##                                Paros_metas
##                                0
##                                Meteorologines_salygos
##                                0
##                                Kelio_kreive
##                                0
##                                Atitvarai
##                                0
##    Eismo_ivykio_dalyvio_kategorija
##                                0
##    KET_pazeidimas_del_kurio_ivyko_eismo_ivykis
##                                0
##                                Lytis
##                                0
##                                Pilietybe
##                                0
##                                Gimimo_data
##                                0
##                                Neigalumas
##                                0
```

```
##          Oro_pagalviu_suveikimas
##                      15707
##          Vairuotojo_kvalifikacija
##                      5080
##                      Stazas
##                      6275
##          Blaivumas
##                      0
##                      Kaltas
##                      0
##          Amzius
##                      0
```

Pašalinome NA reikšmes (stulpeliuose, kuriuose NA > 5000, NA palikome, kad neprarastume didžiosios dalies duomenų).

5 didžiausius Lietuvos miestus lyginsime pagal įvykius 1000-iui gyventojų.

Sukuriame rodiklį 1000 gyventojų.

```
tukst <- gyventojai[,3]/1000
gyventojai <- cbind(gyventojai, tukst)
```

Įvykių sk. 1000-iui gyventojų didžiausiuose miestuose grafikas.

****Skaičiuojame mėnesinį įvykių skaičių.**

```
by_month <- group_by(data, format(as.Date(data$Data, format = "%Y-%m-%d"), "%Y-%m"),
  Vieta)
agreguoti <- summarise(by_month, sum(Kaltas), sum(Dalyviu_skaicius), sum(Suzeistu_skaicius),
  sum(Zuvusiu_skaicius))
colnames(agreguoti) <- list("Data", "Vieta", "Ivykiai", "Dalyviai", "Suzeisti",
  "Zuve")
```

****Išsaugome didžiųjų miestų įvykių skaičių ir gyventojų laiko eilutes.**

```
ivykiusk <- function(i)
  ts(agreguoti[agreguoti$Vieta == i, "Ivykiai"], start=c(2003,1), frequency = 12)

gyvsk <- function(i)
  ts(rep(gyventojai[gyventojai$Vieta == i, "Visi"], each = 12), start = c(2003,1), frequency = 12)

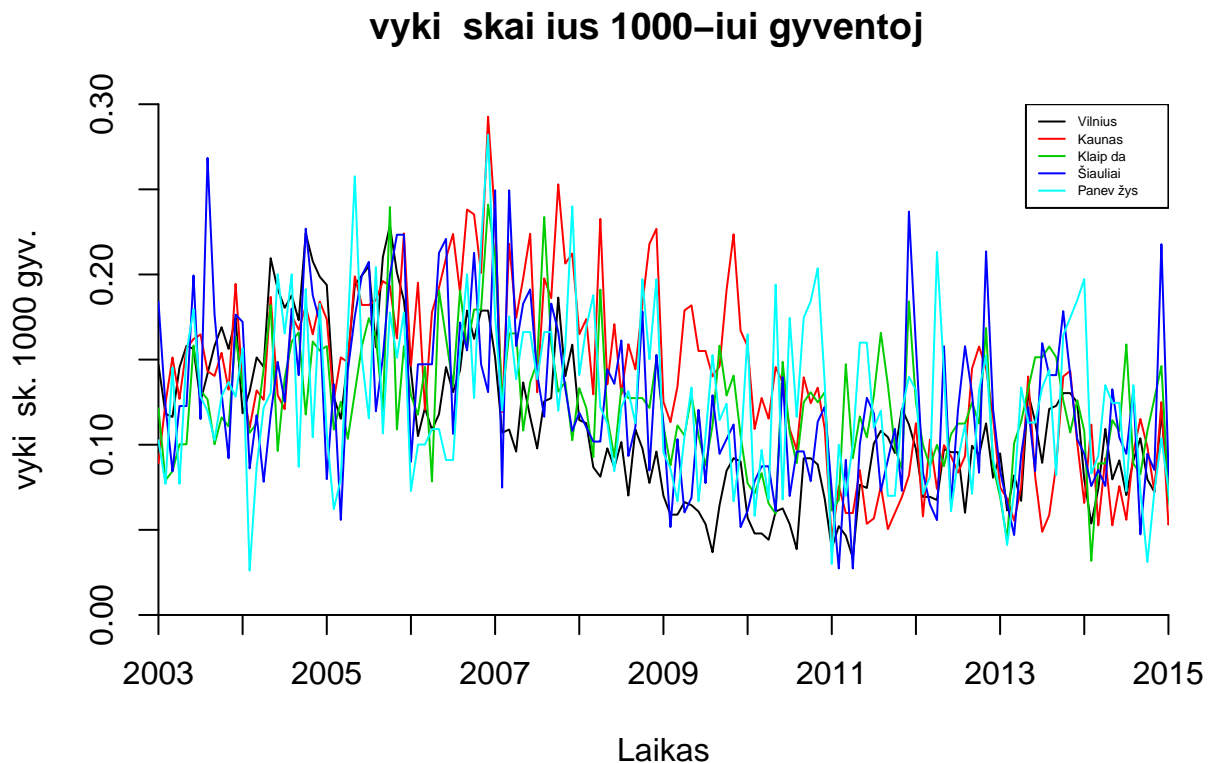
ivykiai <- sapply(miestai, ivykiusk)
names(ivykiai)<-sprintf(miestai)
gyven <- sapply(miestai, gyvsk)
names(gyven)<-sprintf(miestai)

tukst <- function(i)
  ivykiai[,i] * 1000 / gyven[,i]

tukstgyv <- ts(sapply(miestai, tukst), start = c(2003,1), frequency = 12)
names(tukstgyv)<-sprintf(miestai)
```


Brėžiame grafiką.

```
plot(0, type = "n", xlim = c(2003,2015), ylim = c(0,0.3), yaxs="i", xaxs="i", main = "Įvykių skaičius 1000 gyventojui",  
par(xaxp=c(2003, 2015, 12), yaxp=c(0, 0.3, 6))  
axis(1)  
axis(2)  
for(i in 1:5)  
  lines(tukstgyv[,i], col = i, lwd = 1)  
legend("topright", col = c(1:5), lty = 1, lwd = 1, legend = c("Vilnius", "Kaunas", "Klaipėda", "Šiauliai", "Panevėžys"))
```



[Reikalingi komentarai]

Palyginsime proporcijas

Matome, kad $p\text{-value} < 0.05$, vadinasi H_0 atmetam. Proporcijos statistiškai reikšmingai skiriasi. Patikrinkime tarp kurių miestų proporcijos yra statistiškai reikšmingai skirtingos.

Proporcijos statistiškai reikšmingai skiriasi tarp Šiauliu m. sav. ir Panevėžio m. sav. ir tarp Vilniaus m. sav. ir Panevėžio m. sav. Patikrinkime, ar Šiauliuose/Vilniuje įvykių dėl neblaivių vairuotojų kaltės yra statistiškai reikšmingai daugiau negu Panevėžyje.

Paaiškėjo, kad Šiauliuose/Vilniuje įvykių dėl neblaivių vairuotojų kaltės yra statistiškai reikšmingai daugiau negu Panevėžyje.

Tikrinam įvykių priklausomybę nuo amžiaus

Forecastinam autoįvykius

Tikrinsime hipotezę ar Vilniaus mieste eismo įvykiu kaltininkai dažniau yra Vyrų nei moterų.

$H_0 = V \text{ sk. } =$

Kaltininkų procentas miestuose (Vyrų/Moterys)

Kaltininkų procentas Vilniaus mieste.

Kaltininkų procentas Kauno mieste.

Kaltininkų procentas Šiaulių mieste.

Kaltininkų procentas Klaipėdos mieste.

Kaltininkų procentas Panevėžio mieste.

Tikrinsime hipotezę ar Vilniaus mieste eismo įvykiu kaltininkai dažniau yra Vyrų nei moterų.

koreliacijos

Palyginti zuvusių/sužeistų mėnesinius vidurkius tarp laikotarpio ir vietos. Ar skiriasi įvykių skaičius naktį ir dieną?

Literatūra

- <https://www.epolicija.lt/atviri-duomenys>
- <http://osp.stat.gov.lt>
- <http://www.lakd.lt>
- <http://www.lkpt.lt>