

# Panelinių duomenų modelis.

Tomas Dzedulionis ir Dominykas Dzingelevičius

2020-12-14

## Modelis

**TIKSLAS** - Nustatyti ar nusikalstumo lygiui Lietuvos regionuose turį įtakos išsilavinimo, nedarbo lygis ir BVP vienam gyventojui. Jei įtaka yra, kokia ji.

**DUOMENYS** - 2005 - 2019 metai. 10 Lietuvos regionų.

## Duomenų importas ir susitvarkymas

```
Sys.setlocale("LC_ALL","Lithuanian")
```

```
## [1] "LC_COLLATE=Lithuanian_Lithuania.1257;LC_CTYPE=Lithuanian_Lithuania.1257;LC_MONETARY=Lithuanian_Lithuania.1257"
```

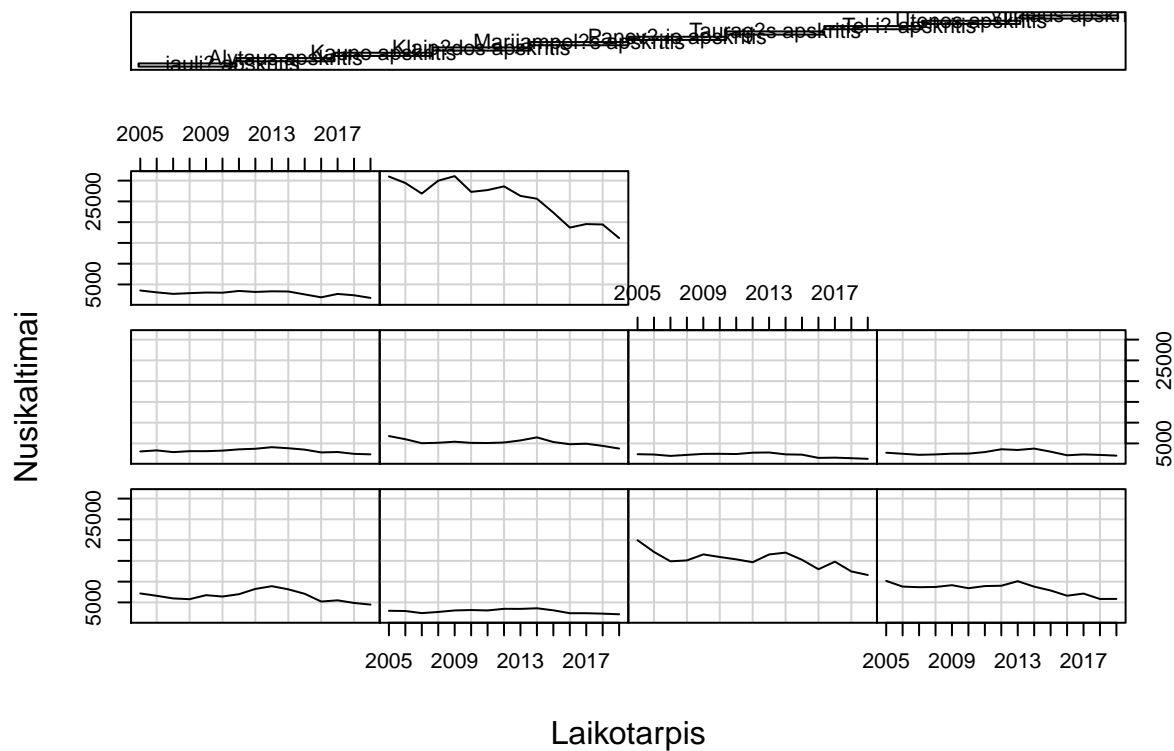
```
data <- read.delim("https://drive.google.com/uc?export=download&id=189Gq3W33BjoFkeDmlEcOCLjR4wAtC6-f", as.is=T)
data <- data%>% filter(Administracine.teritorija != "Lietuvos Respublika",
                      Administracine.teritorija != "Sostinė regionas",
                      Administracine.teritorija != "Vidurio ir vakarų Lietuvos regionas") %>%
  mutate(Vidutinis = Vidutinis/Viso.issilavinusiu* 100,
         Aukstas= Aukstas/Viso.issilavinusiu* 100,
         Zemas = Zemas/Viso.issilavinusiu* 100) %>%
  rename( "teritorija" =Administracine.teritorija)
data$Zemas <- round(data$Zemas, 1)
data$Aukstas <- round(data$Aukstas, 1)
data$Vidutinis <- round(data$Vidutinis, 1)

data.p <- pdata.frame(data,index=c("teritorija", "Laikotarpis"))
data.p[is.na(data.p)] <- 0
```

## Grafikai

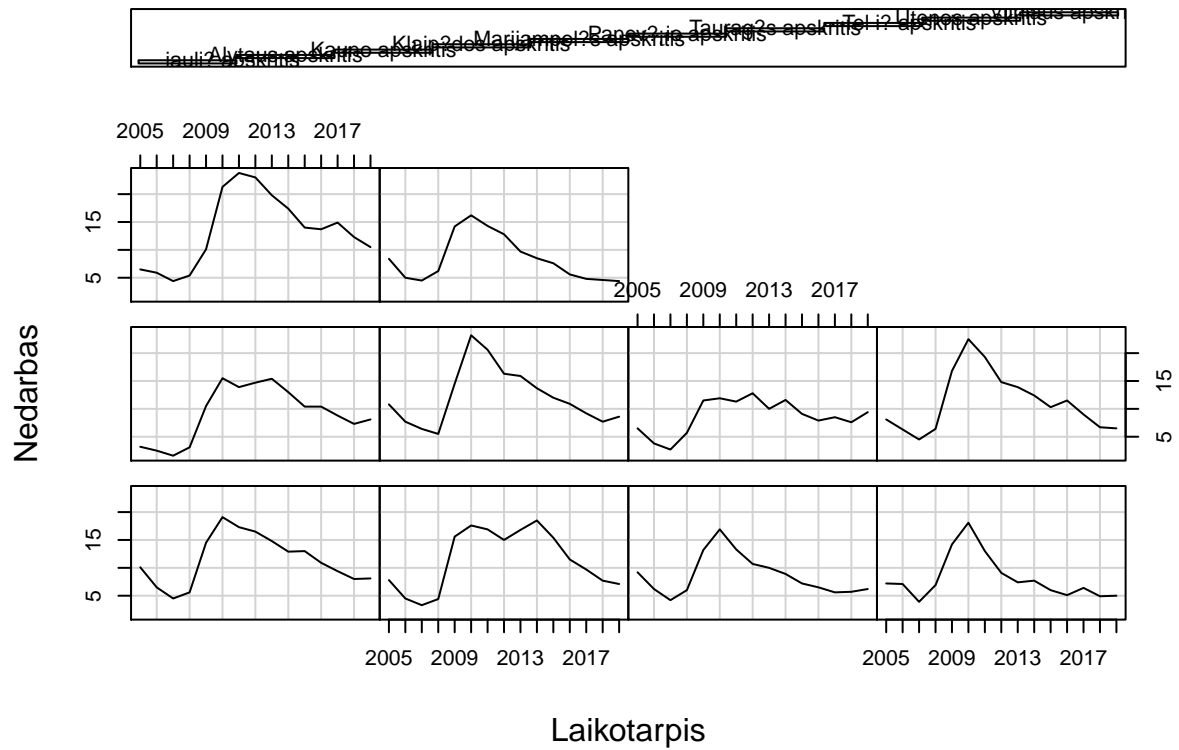
```
coplot(Nusikaltimai ~ Laikotarpis|teritorija, type="l", data=data.p)
```

Given : teritorija



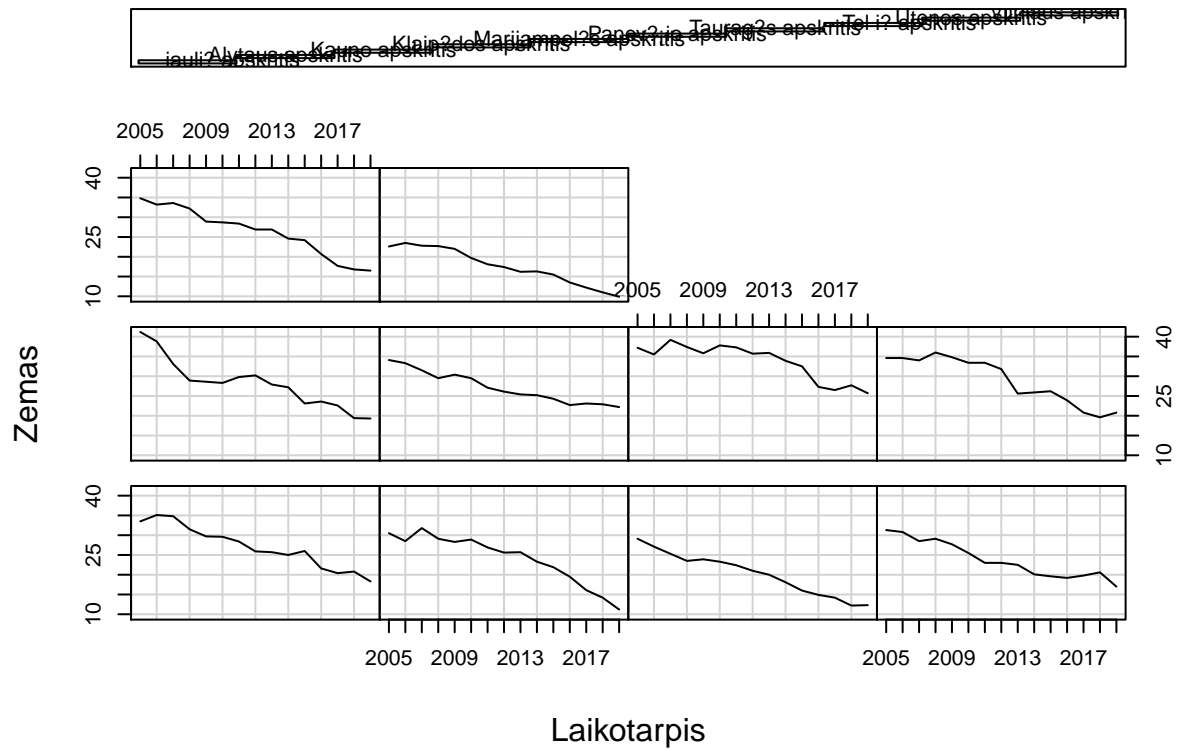
```
coplot(Nedarbas ~ Laikotarpis|teritorija, type="l", data=data.p)
```

Given : teritorija



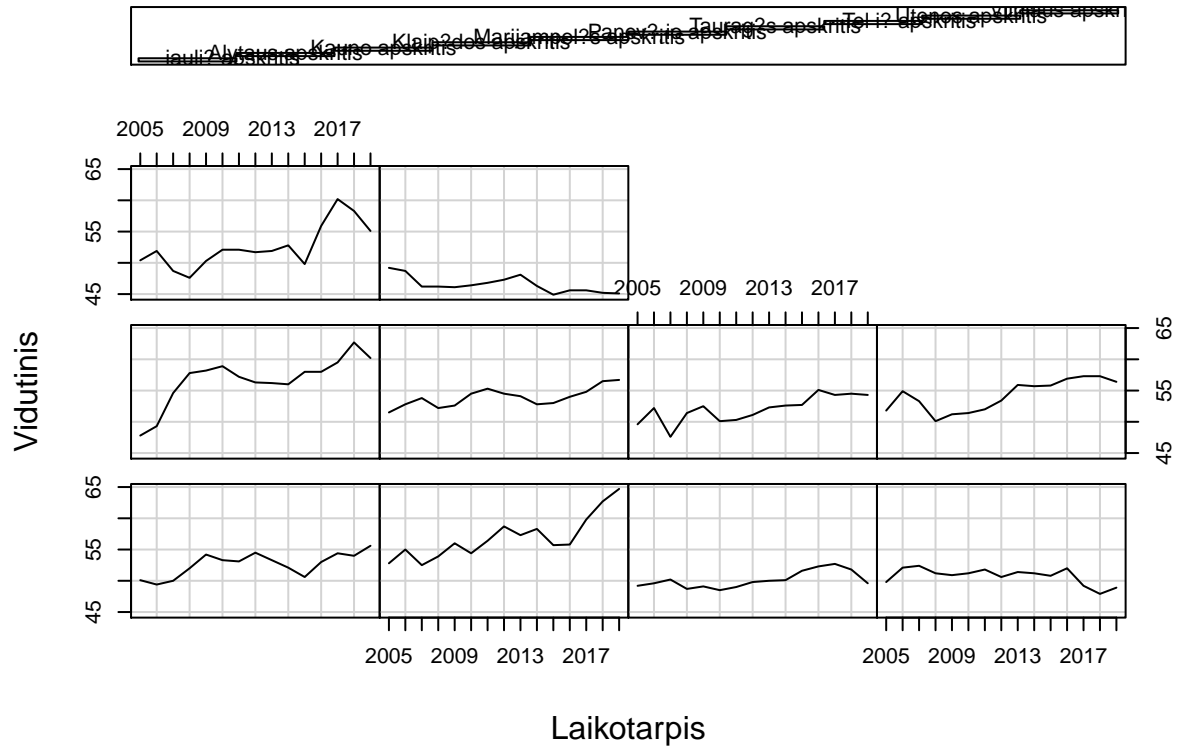
```
coplot(Zemas ~ Laikotarpis|teritorija, type="l", data=data.p)
```

Given : teritorija



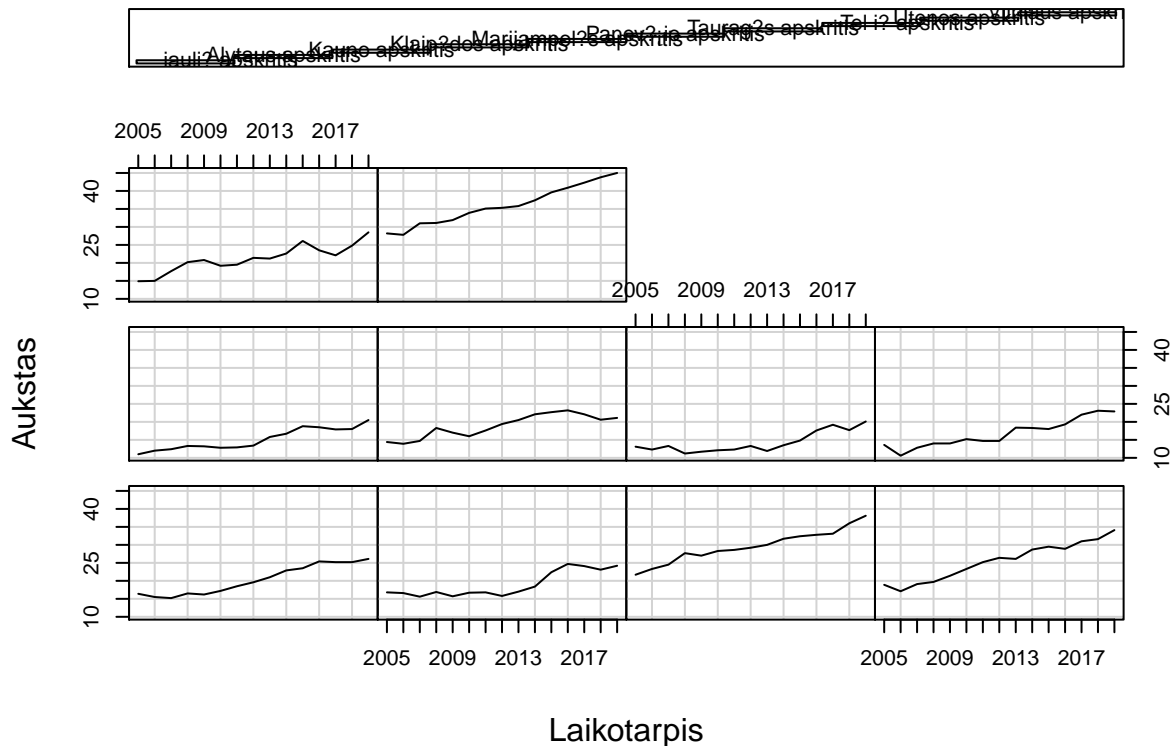
```
coplot(Vidutinis ~ Laikotarpis|teritorija, type="l", data=data.p)
```

Given : teritorija



```
coplot(Aukstas ~ Laikotarpis | teritorija, type="l", data=data.p)
```

Given : teritorija



## Stacionarumo tikrinimas [1]

Matome, kad duomenys, išskyrus nedarbą, nėra stacionarūs, kadangi  $p$  reikšmės  $> 0.05$ .

```
adf.test(data.p$Nedarbas)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Nedarbas
## Dickey-Fuller = -5.7376, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
adf.test(data.p$Nusikaltimai)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Nusikaltimai
## Dickey-Fuller = -2.3303, Lag order = 5, p-value = 0.4388
## alternative hypothesis: stationary
adf.test(data.p$Zemas)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Zemas
## Dickey-Fuller = -3.007, Lag order = 5, p-value = 0.1569
```

```
## alternative hypothesis: stationary
adf.test(data.p$Aukstas)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Aukstas
## Dickey-Fuller = -1.5032, Lag order = 5, p-value = 0.7834
## alternative hypothesis: stationary
adf.test(data.p$Vidutinis)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Vidutinis
## Dickey-Fuller = -3.2202, Lag order = 5, p-value = 0.08721
## alternative hypothesis: stationary
adf.test(data.p$BVP)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$BVP
## Dickey-Fuller = -1.6356, Lag order = 5, p-value = 0.7282
## alternative hypothesis: stationary
```

## Duomenų stacionarizavimas.

Diferencijuojame laiko eilutes, tam, kad gauti pokyčius. Nusikaltimų duomenis logaritmuojame ir diferencijuojame, taip gaudami procentinius augimo pokyčius.

```
data.p$Nusikaltimai <- log(data.p$Nusikaltimai)
data.p$Nusikaltimai <- diff(data.p$Nusikaltimai)
data.p$Zemas <- diff(data.p$Zemas)
data.p$Vidutinis <- diff(data.p$Vidutinis)
data.p$Aukstas <- diff(data.p$Aukstas)
data.p$BVP <- diff(data.p$BVP)
data.p[is.na(data.p)] <- 0
```

## Stacionarumo tikrinimas [2]

Matome, jog duomenys vienetinės šaknies nebeturi - tapo stacionarus. p reikšmės  $< 0.05$ .

```
adf.test(data.p$Nedarbas)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Nedarbas
## Dickey-Fuller = -5.7376, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
adf.test(data.p$Nusikaltimai)
##
## Augmented Dickey-Fuller Test
```

```
##
## data: data.p$Nusikaltimai
## Dickey-Fuller = -5.4403, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
adf.test(data.p$Zemas)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Zemas
## Dickey-Fuller = -6.9576, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
adf.test(data.p$Aukstas)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Aukstas
## Dickey-Fuller = -4.9841, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
adf.test(data.p$Vidutinis)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$Vidutinis
## Dickey-Fuller = -7.8634, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
adf.test(data.p$BVP)
##
## Augmented Dickey-Fuller Test
##
## data: data.p$BVP
## Dickey-Fuller = -5.4848, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

## Pastovių konstantų modelis.

Sudarėme pastovių konstantų modelį su išraiška:

$$\log(Nusikaltimai) = \alpha + \beta_1 Nedarbas + \beta_2 Zemas + \beta_3 Vidutinis + \beta_4 Aukstas + \beta_5 BVP$$

Matome, kad BVP komponentė nėra statistiškai reikšminga, teks jos atsisakyti.

```
ols<- lm(Nusikaltimai ~ Nedarbas + Zemas + Vidutinis +Aukstas + BVP, data=data.p)
summary(ols)
##
## Call:
## lm(formula = Nusikaltimai ~ Nedarbas + Zemas + Vidutinis + Aukstas +
##     BVP, data = data.p)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35047 -0.07278  0.01540  0.06959  0.30421
##
```



```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.090493   0.026966  -3.356  0.00101 **
## Nedarbas     0.006615   0.002019   3.277  0.00132 **
## Zemas       -0.287613   0.125805  -2.286  0.02370 *
## Vidutinis    -0.289576   0.125839  -2.301  0.02282 *
## Aukstas      -0.305662   0.126252  -2.421  0.01672 *
## BVP          0.004040   0.006845   0.590  0.55599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1101 on 144 degrees of freedom
## Multiple R-squared:  0.1369, Adjusted R-squared:  0.1069
## F-statistic: 4.568 on 5 and 144 DF,  p-value: 0.0006756
```

Sudarome naują modelį:

$$\log(\text{Nusikaltimai}) = \alpha + \beta_1 \text{Nedarbas} + \beta_2 \text{Zemas} + \beta_3 \text{Vidutinis} + \beta_4 \text{Aukstas}$$

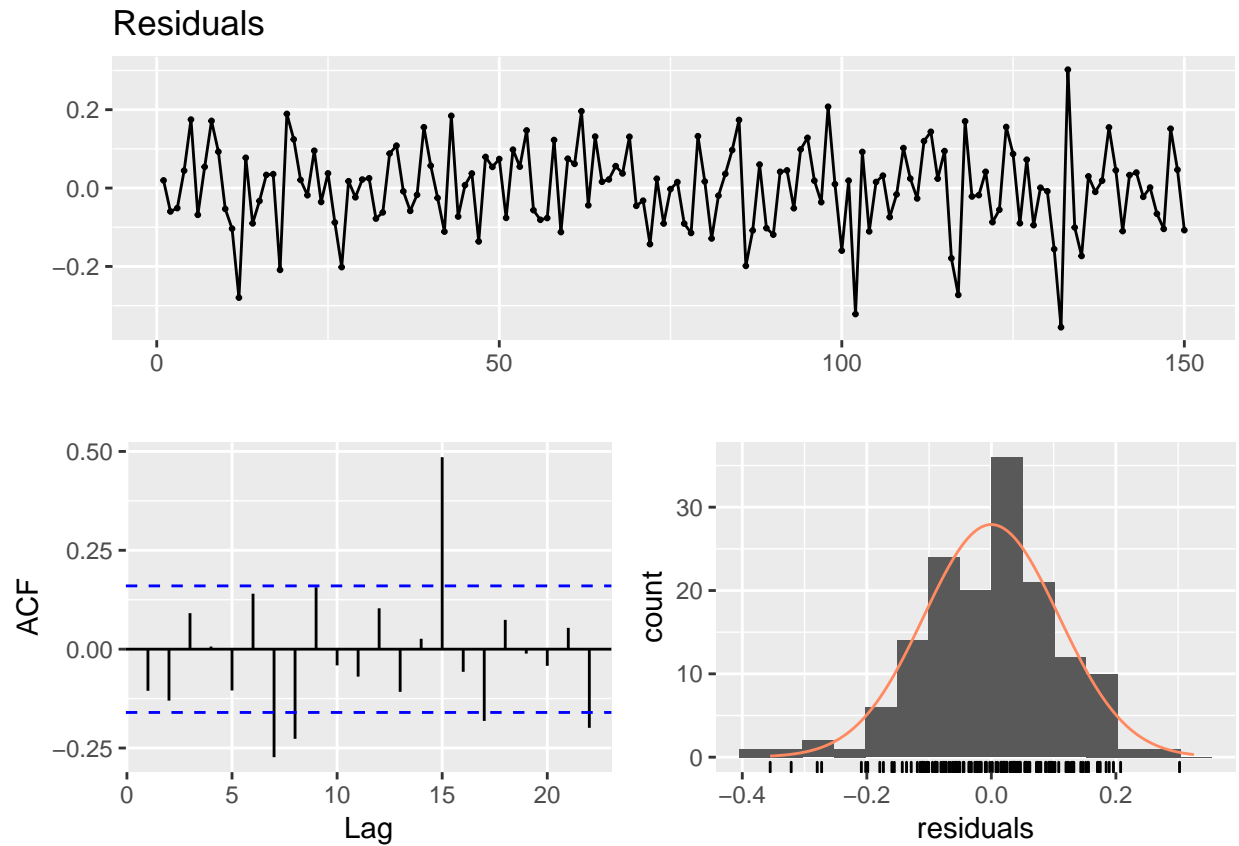
```
ols<- lm(Nusikaltimai ~ Nedarbas + Zemas + Vidutinis + Aukstas, data=data.p)
summary(ols)
##
## Call:
## lm(formula = Nusikaltimai ~ Nedarbas + Zemas + Vidutinis + Aukstas,
##     data = data.p)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35520 -0.07382  0.01574  0.06970  0.30218
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.081297   0.021959  -3.702  0.000303 ***
## Nedarbas     0.006114   0.001827   3.346  0.001044 **
## Zemas       -0.285099   0.125450  -2.273  0.024520 *
## Vidutinis    -0.287000   0.125481  -2.287  0.023631 *
## Aukstas      -0.303929   0.125934  -2.413  0.017052 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1099 on 145 degrees of freedom
## Multiple R-squared:  0.1348, Adjusted R-squared:  0.1109
## F-statistic: 5.648 on 4 and 145 DF,  p-value: 0.0002972
```

Matome, jog visi kintamieji yra statistiškai reikšmingi. Turime patikrinti ar nėra autokoreliacijos:

```

durbinWatsonTest(ols)
## lag Autocorrelation D-W Statistic p-value
## 1 -0.1053024 2.203789 0.27
## Alternative hypothesis: rho != 0
checkresiduals(ols$residuals)

```



Matome, jog Durbin-Watson testas nerodo reikšmingos autokoreliacijos ( $p > 0.05$ , neatmetam  $H_0$ ). Paklaidų grafike stebime autokoreliaciją 7-tame lage, tačiau toks lagas yra per daug tolimas, kad būtų reikšmingas, kadangi duomenys yra metiniai.

#### Galutinis modelis

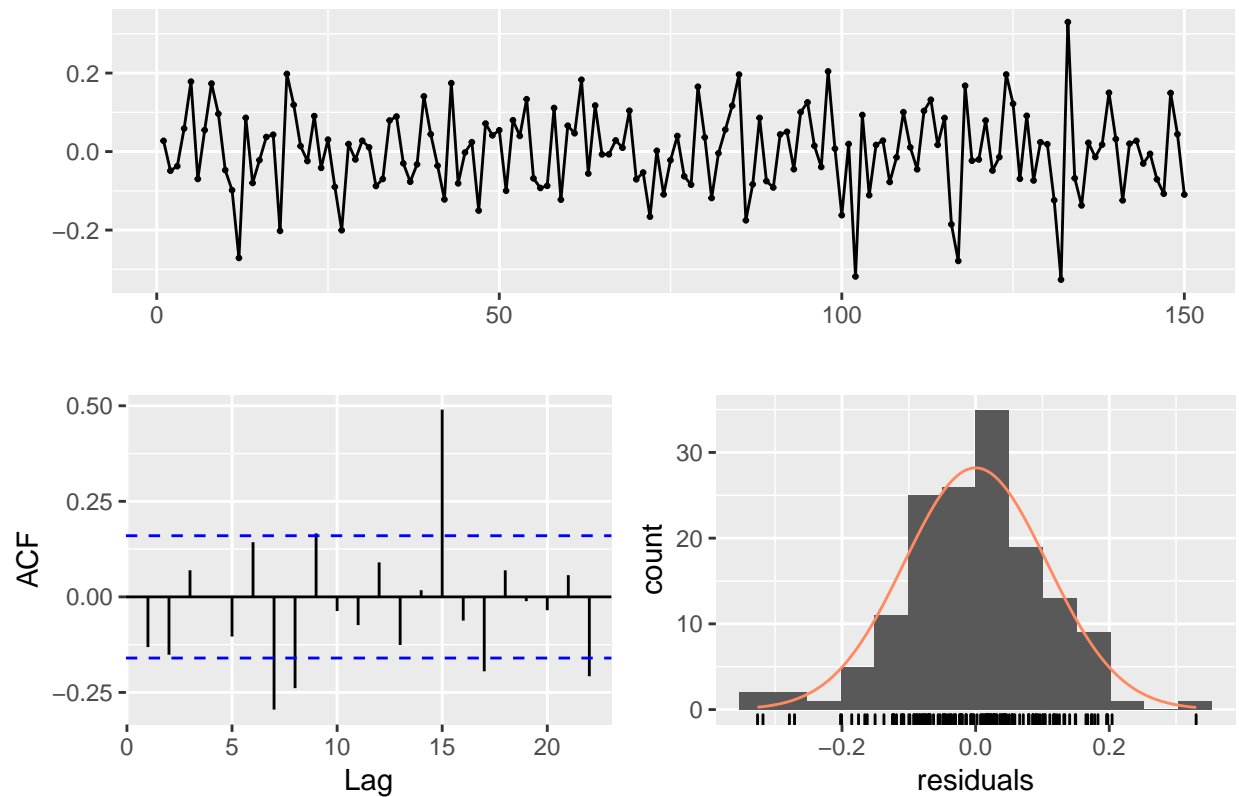
$$\log(Nusikaltimai) = -0.081297 + 0.006114 * Nedarbas + (-0.285099) * Zemas + (-0.287) * Vidutinis + (-0.303929) * Aukstas$$

## Fiksuotų efektų modelis

$$\log(\text{Nusikaltimai}) = \alpha_1 + \beta_1 \text{Nedarbas} + \alpha_2 + \beta_2 \text{Zemas} + \alpha_3 + \beta_3 \text{Vidutinis} + \alpha_4 + \beta_4 \text{Aukstas}$$

```
fixedeff <- plm(data.p$Nusikaltimai~data.p$Nedarbas+data.p$Zemas+data.p$Vidutinis+data.p$Aukstas,data=d,
summary(fixedeff)
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = data.p$Nusikaltimai ~ data.p$Nedarbas + data.p$Zemas +
##      data.p$Vidutinis + data.p$Aukstas, data = data.p, model = "within")
##
## Balanced Panel: n = 10, T = 15, N = 150
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.326371 -0.070495  0.010199  0.070025  0.330015
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## data.p$Nedarbas  0.0072178  0.0020025   3.6044 0.000438 ***
## data.p$Zemas    -0.2878980  0.1282695  -2.2445 0.026417 *
## data.p$Vidutinis -0.2899920  0.1282800  -2.2606 0.025371 *
## data.p$Aukstas   -0.3075217  0.1287369  -2.3888 0.018278 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2.009
## Residual Sum of Squares: 1.7106
## R-Squared:    0.14851
## Adj. R-Squared: 0.067113
## F-statistic: 5.9298 on 4 and 136 DF, p-value: 0.00019844
fixef(fixedeff) ## Fiksuotu efektu konstantos kiekvienai apskriciai
##      \212iauli? apskritis      Alytaus apskritis      Kauno apskritis
##      -0.100268                -0.093427                -0.077155
##      Klaip?dos apskritis Marijampol?s apskritis      Panev?\236io apskritis
##      -0.075913                -0.069659                -0.117646
##      Taurag?s apskritis      Tel\232i? apskritis      Utenos apskritis
##      -0.090594                -0.086370                -0.126142
##      Vilniaus apskritis
##      -0.083001
checkresiduals(fixedeff$residuals)
```

## Residuals



Visos komponentės statistiškai reikšmingos, tikrinkime, kuris modelis - pastovių konstantų ar fiksuotų efektų - yra geresnis.

Kadangi  $p > 0.05$ ,  $H_0$  neatmetame. Pastovios konstantos modelis yra geresnis.

```
pFtest(fixedeff,ols) ## Neatmetam  $H_0$ , todėl pastovios konstantos geresnis.
##
## F test for individual effects
##
## data: data.p$Nusikaltimai ~ data.p$Nedarbas + data.p$Zemas + data.p$Vidutinis + ...
## F = 0.35556, df1 = 9, df2 = 136, p-value = 0.9538
## alternative hypothesis: significant effects
```

## Atsitiktinių efektų modelis

$$\log(\text{Nusikaltimai}) = \alpha + \beta_1 \text{Nedarbas} + \beta_2 \text{Zemas} + \beta_3 \text{Vidutinis} + \beta_4 \text{Aukstas} + (u + v)$$

```

randomeff <- plm(data.p$Nusikaltimai~data.p$Nedarbas+data.p$Zemas+data.p$Vidutinis+data.p$Aukstas,data=
summary(randomeff)
## Oneway (individual) effect Random Effect Model
## (Swamy-Arora's transformation)
##
## Call:
## plm(formula = data.p$Nusikaltimai ~ data.p$Nedarbas + data.p$Zemas +
## data.p$Vidutinis + data.p$Aukstas, data = data.p, model = "random")
##
## Balanced Panel: n = 10, T = 15, N = 150
##
## Effects:
##               var std.dev share
## idiosyncratic 0.01258 0.11215    1
## individual    0.00000 0.00000    0
## theta: 0
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -0.355205 -0.073824  0.015738  0.069695  0.302181
##
## Coefficients:
##               Estimate Std. Error z-value Pr(>|z|)
## (Intercept)   -0.0812967  0.0219594 -3.7021 0.0002138 ***
## data.p$Nedarbas  0.0061138  0.0018272  3.3460 0.0008198 ***
## data.p$Zemas   -0.2850991  0.1254499 -2.2726 0.0230495 *
## data.p$Vidutinis -0.2870002  0.1254807 -2.2872 0.0221839 *
## data.p$Aukstas  -0.3039294  0.1259337 -2.4134 0.0158041 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2.0237
## Residual Sum of Squares: 1.7509
## R-Squared:    0.13481
## Adj. R-Squared: 0.11094
## Chisq: 22.5925 on 4 DF, p-value: 0.00015271
stargazer(randomeff,type='text')
##
## =====
##               Dependent variable:
## -----
##               Nusikaltimai
## -----
## Nedarbas           0.006***
##                   (0.002)
##
## Zemas             -0.285**
##                   (0.125)
##

```

```
## Vidutinis          -0.287**
##                   (0.125)
##
## Aukstas           -0.304**
##                   (0.126)
##
## Constant          -0.081***
##                   (0.022)
##
## -----
## Observations      150
## R2                 0.135
## Adjusted R2        0.111
## F Statistic        22.592***
## =====
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

Visas komponentes ir vėl gauname statistiškai reikšmingas. Pasinaudodami Hausman testu tikriname, kuris modelis - atsitiktinių efektų ar fiksuotų efektų - yra geresnis. Gauname  $p > 0.05$ , todėl  $H_0$  neatmetame. Tai reiškia, kad tinkamesnis yra atsitiktinių efektų modelis.

```
phtest(fixedeff,randomeff) ##Neatmetam h0, todėl atsitiktiniu dydžiu geresnis.
##
## Hausman Test
##
## data: data.p$Nusikaltimai ~ data.p$Nedarbas + data.p$Zemas + data.p$Vidutinis + ...
## chisq = 1.9182, df = 4, p-value = 0.7508
## alternative hypothesis: one model is inconsistent
```

Palyginus atsitiktinių efektų determinacijos koeficientą (0.1348) ir pastovios konstantos modelio determinacijos koeficientą (0.1348) matome, kad jie yra vienodi. Tai reiškia, kad abu modeliai yra geresni už fiksuotų efektų modelį ir abu paaiškina vienodą dalį duomenų pokyčių.