

Ekonometrija

1 uždutis

Eglė Kaleckaitė

2011 m. spalio 9 d.

Mano nagrinėjama įmonių grupė yra trečia, o endogeninis kintamasis – val_Ax. Prieš pradėdama tolimesnį darbą su duomenimis, iš pradžių juos pertvarkiau formatu, kuris būtų tinkamas panelinių duomenų analizei su R, t.y. dabar duomenis sudaro 11 stulpelių:

```
nr, time, veikla, grupe, paj, dsk, val, atlyg, ter, nace1, nace2
```

Taigi atsirado papildomas stulpelis su data - *time*, kuriame skaičius po kablelio žymi ketvirtį, t.y. x.00 žymi pirmą x metų ketvirtį, x.25 – antrą, x.50 – trečią ir x.75 – ketvirtą.

1 Duomenų aprašymas

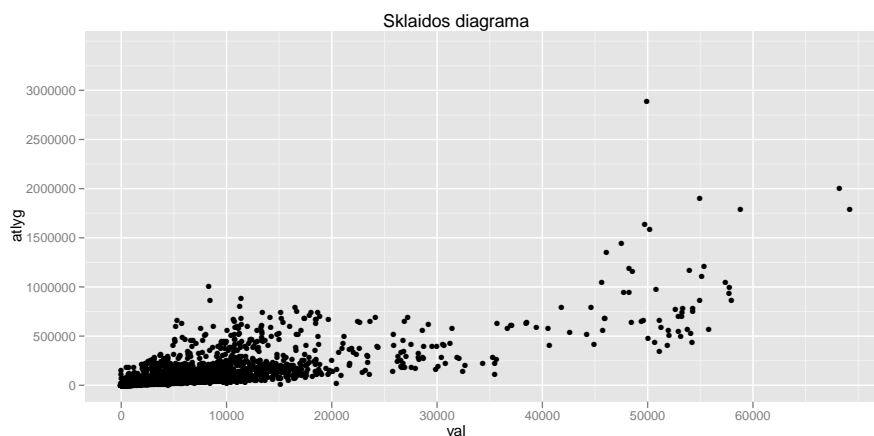
Žemiau patektoje lentelėje matyti kiekvienas duomenų stulpelis ir jo pagrindinės charakteristikos.

nr		time		veikla		grupe		paj	
Min.	: 530	Min.	:2005	Min.	:70	Min.	:3	Min.	: 0
1st Qu.:	4246	1st Qu.:	2006	1st Qu.:	70	1st Qu.:	3	1st Qu.:	20199
Median	: 6520	Median	:2007	Median	:70	Median	:3	Median	: 161270
Mean	: 6645	Mean	:2007	Mean	:70	Mean	:3	Mean	: 1013081
3rd Qu.:	9278	3rd Qu.:	2008	3rd Qu.:	70	3rd Qu.:	3	3rd Qu.:	551697
Max.	:11944	Max.	:2009	Max.	:70	Max.	:3	Max.	:68509767
								NA's	: 15460
dsk		val		atlyg		ter			
Min.	: 0.500	Min.	: 0.0	Min.	: 0	Min.	: 11.00		
1st Qu.:	2.000	1st Qu.:	711.8	1st Qu.:	2784	1st Qu.:	13.00		
Median	: 3.000	Median	: 1808.0	Median	: 7500	Median	: 19.00		
Mean	: 6.281	Mean	: 3849.4	Mean	: 29180	Mean	: 23.31		
3rd Qu.:	7.000	3rd Qu.:	4369.5	3rd Qu.:	22772	3rd Qu.:	21.00		
Max.	: 359.500	Max.	:69208.0	Max.	:3427509	Max.	: 91.00		
NA's	:4309.000	NA's	:16368.0	NA's	: 4309	NA's	:352.00		
nace1		nace2							
Min.	:452100	Min.	: 24000						
1st Qu.:	:702000	1st Qu.:	:682000						
Median	:702000	Median	:682000						
Mean	:703170	Mean	:665281						
3rd Qu.:	:702000	3rd Qu.:	:682000						

Max. :930500 Max. :960900
 NA's : 288 NA's : 288

Stulpeliai *veikla* ir *grupe* yra neįdomūs, nes jie yra konstantos. Taip pat galima pasakyti, jog duomenys korektiški ženklų prasme, t.y. nėra neigiamų reikšmių. Minimalios pajamų, valandų ir atlyginimo reikšmės yra 0. Lentelė taip pat parodo ir NA (praleistų) reikšmių kiekį kiekvienam rodikliui.

Kadangi kiti rodikliai yra beveik konstantos ir žymi tik priklausymą vienai ar kitai grupei, tai mus labiau domina *val* bei *atlyg* ir *dsk* priklausomybė. Tai galima pamatyti paveikslėliuose 1 ir 2. Paveikslėliuose galima pastebėti didelę duomenų koncentraciją ties mažesnėmis reikšmėmis. Taigi, galima sakyti, jog mažesnių įmonių yra ženkliai mažiau nei didelių. Taip pat galime pastebėti ryškią valandų ir darbuotojų skaičiaus tiesinę priklausomybę. Būtų keistą, jei būtų kitaip. Tačiau tarp atlyginimų ir valandų priklausomybė ne tokia aiški, lyg ir galima įžvelgti eksponentinę priklausomybę.

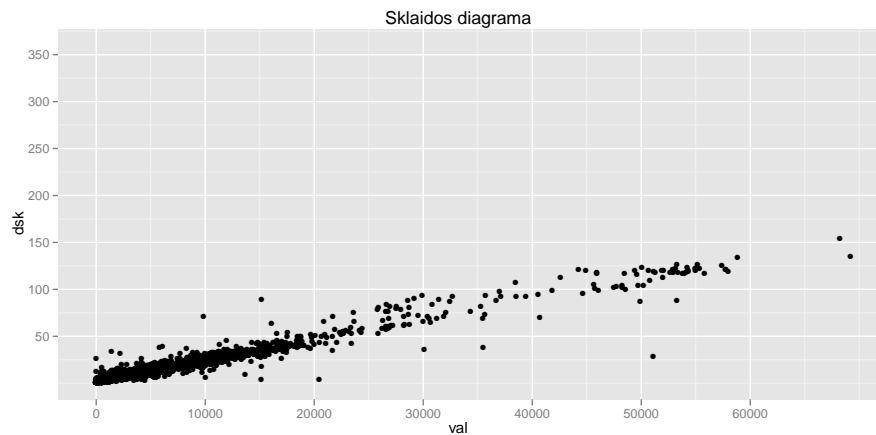


1 pav.: Kintamųjų *val* ir *atlyg* sklaidos diagrama

Iš viso turima 1430 įmonių priklausančių trečiai grupei, tačiau domina tik tos, kurių rodiklis *val* turi bent vieną reikšmę per visus metus. Kitu atveju, duomenys nesuteiks naudingos informacijos. Tokių įmonių yra 651.

Dabar pagrindinės duomenų charakteristikos atrodo taip:

nr	time	veikla	grupe	paj
Min. : 539	Min. :2005	Min. :70	Min. :3	Min. : 0
1st Qu.: 4819	1st Qu.:2006	1st Qu.:70	1st Qu.:3	1st Qu.: 45000
Median : 7187	Median :2007	Median :70	Median :3	Median : 198609
Mean : 6980	Mean :2007	Mean :70	Mean :3	Mean : 1110513
3rd Qu.: 9432	3rd Qu.:2008	3rd Qu.:70	3rd Qu.:3	3rd Qu.: 621630
Max. :11944	Max. :2009	Max. :70	Max. :3	Max. :68509767
				NA's : 5695
dsk	val	atlyg	ter	
Min. : 0.500	Min. : 0.0	Min. : 0	Min. : 11.00	
1st Qu.: 2.000	1st Qu.: 711.8	1st Qu.: 4500	1st Qu.: 13.00	
Median : 4.500	Median : 1808.0	Median : 13916	Median : 18.00	



2 pav.: Kintamųjų *val* ir *dsk* sklaidos diagrama

Mean	: 8.167	Mean	: 3849.4	Mean	: 43096	Mean	: 21.69
3rd Qu.:	9.000	3rd Qu.:	4369.5	3rd Qu.:	40688	3rd Qu.:	21.00
Max.	: 154.500	Max.	: 69208.0	Max.	: 2883086	Max.	: 91.00
NA's	: 1705.000	NA's	: 5952.0	NA's	: 1705	NA's	: 212.00
nace1		nace2		t			
Min.	: 452100	Min.	: 24000	Min.	: 2005		
1st Qu.:	702000	1st Qu.:	682000	1st Qu.:	2006		
Median	: 702000	Median	: 682000	Median	: 2007		
Mean	: 703986	Mean	: 657128	Mean	: 2007		
3rd Qu.:	702000	3rd Qu.:	682000	3rd Qu.:	2008		
Max.	: 930500	Max.	: 960900	Max.	: 2009		
NA's	: 188	NA's	: 188				

Kad susidaryčiau aiškesnį vaizdą, kokie yra turimi duomenys, išsibrėžiau kiekvienos įmonės kiekvieno rodiklio grafikus. Jie patalpinti šio dokumento prisegtuose *pdf* formate, pavadinimu *all_ind.pdf*. Panašu, jog duomenys turi daug tuščių reikšmių ir trūkių. Iš pradžių pabandyčiau tai ignoruoti ir sudaryti panelinių duomenų modelį. Taip pat paruošiu kelis duomenų masyvus, t.y. skelsiu duomenis pagal tuščių reikšmių kiekį bei skaidysiu į grupes. Ir bandysiu pagerinti rezultatus.

2 Modeliavimas

Kaip jau minėta, pradžioje sudarysiu paprastą panelių duomenų modelį. Kadangi kintamieji *nace1* ir *nace2* žymi tą patį tik pagal skirtingus reikalavimus, pasiliksiu vieną iš jų. Tegul tai būna *nace1*. Sudarysiu *pooled* panelinių duomenų modelį (R paketas *plm*) duomenims iki 2008 metų:

```
val ~ t + dsk + atlyg + ter + nace1
```

Oneway (individual) effect Pooling Model

```
Call:
plm(formula = fm, data = model.data, effect = "individual", model = "pooling")
```

Unbalanced Panel: n=691, T=1-12, N=4702

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-23000	-342	51	396	38600

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	1.9181e+05	4.7360e+04	4.0500	5.204e-05 ***
t	-9.5991e+01	2.3602e+01	-4.0670	4.840e-05 ***
dsk	3.9346e+02	2.1132e+00	186.1888	< 2.2e-16 ***
atlyg	4.9017e-03	2.5771e-04	19.0203	< 2.2e-16 ***
ter	-1.1681e+00	1.2337e+00	-0.9469	0.3438
nace1	5.5020e-04	6.7892e-04	0.8104	0.4177

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 2.1153e+11

Residual Sum of Squares: 8591400000

R-Squared : 0.95938

Adj. R-Squared : 0.95816

F-statistic: 22184.7 on 5 and 4696 DF, p-value: < 2.22e-16

Du kintamieji yra nereikšmingi (*ter* ir *nace1*), tačiau jei jie nereikšmingi, jų dydžiai ir neturės reikšmingos įtakos endogeniniam kintamajam. Tokių kintamųjų pašalinti nėra prasmės, tuo labiau, kai nežinoma, ar gautas modelis geras. Gerų rezultatų nerodo ir paklaidų kvadratų sumos. Dar galima pastebėti, jog laiko įtaka yra neigiama. Būtų galima logaritmuoti *atlyg*, tačiau šis kintamasis turi nulinių reikšmių. Pabandžiau įtraukti teritorijos žymimuosius kintamuosius, tačiau rezultatai tik pablogėjo. Taip pat galima bandyti įtraukti sezoninius žymimuosius kintamuosius. Tai ir padarysiu. Įtraukiau tris žymimuosius kintamuosius *s2*, *s3* ir *s4*, kurie žymės antrą, trečią ir ketvirtą sezonus atitinkamai.

```
val ~ t + s2 + s3 + s4 + dsk + atlyg + ter + nace1
```

Oneway (individual) effect Pooling Model

Call:

```
plm(formula = fm.s, data = model.data1, effect = "individual",
     model = "pooling")
```

Unbalanced Panel: n=691, T=1-12, N=4702

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-22900.0	-337.0	59.6	397.0	38500.0

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	2.0652e+05	5.0142e+04	4.1187	3.876e-05	***
t	-1.0330e+02	2.4999e+01	-4.1322	3.656e-05	***
s2	-9.0789e+01	5.6563e+01	-1.6051	0.108540	
s3	-1.5269e+02	5.7615e+01	-2.6501	0.008073	**
s4	7.8624e+01	5.9353e+01	1.3247	0.185337	
dsk	3.9365e+02	2.1100e+00	186.5621	< 2.2e-16	***
atlyg	4.8759e-03	2.5732e-04	18.9486	< 2.2e-16	***
ter	-1.2040e+00	1.2315e+00	-0.9776	0.328312	
nace1	5.4479e-04	6.7962e-04	0.8016	0.422815	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 2.1153e+11

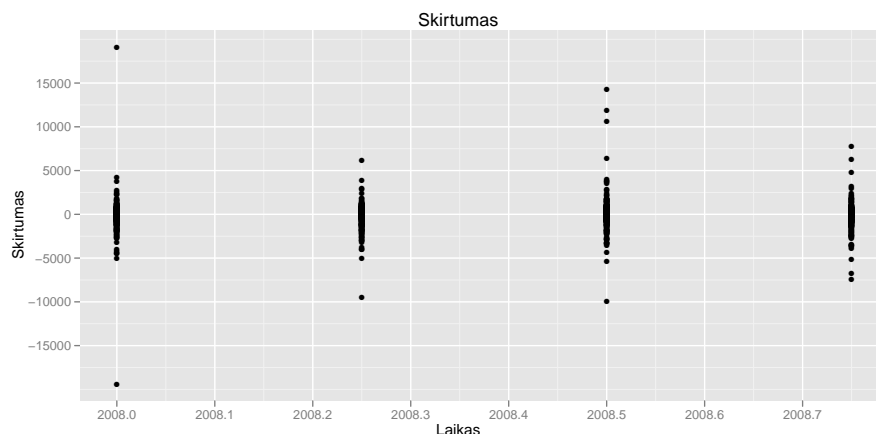
Residual Sum of Squares: 8554800000

R-Squared : 0.95956

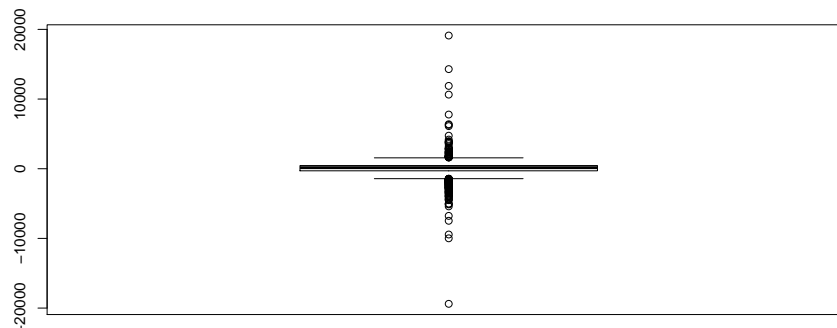
Adj. R-Squared : 0.95772

F-statistic: 13918.3 on 8 and 4693 DF, p-value: < 2.22e-16

Matome, kad ir *s3* yra reikšmingas. Nepašalinu ir likusių kintamųjų. Pabandyčiau padaryti šio modelio prognozę. Prognozių grafikai patalpinti šio dokumento prisegtuke pavadinimu *Valandos_prognoze_skirtumas_zym.pdf*. Iš paveikslėlių matyti, jog gan dažnai prognozė daug nenuklysta nuo realių duomenų (turint omenyje, jog dirbama su gan dideliais dydžiais). Tačiau yra ir nemažai atvejų, kai prognozė labai nutolsta nuo realybės. Skirtumai tarp valandų ir jų prognozės pavaizduoti paveikslėliuose 3 ir 4. Reikia pastebėti, jog nepavaizduotos tos reikšmės, kurioms nebuvo realių duomenų, nes palyginimui jos nereikšmingos. Tačiau taip pat svarbu paminėti, jog išprognozuotų reikšmių buvo daugiau nei realiųjų.



3 pav.: Taškais pavaizduotos visų įmonių skirtumai tarp valandų ir jų prognozės laike



4 pav.: Skirtumų tarp valandų ir jų prognozių boxplot

Bendras įspūdis nėra blogas, bet pabandyčiau prognozes pagerinti. Kadangi duomenis sudaro įvairaus ilgio valandų laiko eilutės, būtų galima tokias eilutes skirstyti į ilgesnes ir trumpesnes pagal kintamąjį *val*. Ilgesnėmis laikysiu eilutes, kurias sudaro bent 4 netuščios reikšmės, visa kita bus laikoma kaip trumpos laiko eilutės. Taip pat pastebėjau, jog valandas sudaro ir pastovios laiko eilutės (konstantos). Konstantas prognozuoti lengva, tam nereikia sudaryti regresijos, užtenka pratęsti tą pačią reikšmę. Tokių yra 9:

```
[1] 1368 1574 1735 7050 7537 9331 9836 10632 11429
```

Jau anksčiau sudarytas modelis, tik ilgesnėms duomenų laiko eilutėms:

Oneway (individual) effect Pooling Model

Call:

```
plm(formula = fm.s, data = long, effect = "individual", model = "pooling")
```

Unbalanced Panel: n=723, T=2-16, N=6153

Residuals :

	Min.	1st Qu.	Median	3rd Qu.	Max.
	-22700.0	-332.0	69.3	412.0	38600.0

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	1.8737e+05	3.1700e+04	5.9108	3.587e-09 ***
t	-9.3754e+01	1.5797e+01	-5.9351	3.098e-09 ***
s2	-7.6873e+01	4.9169e+01	-1.5634	0.11800
s3	-1.0511e+02	4.9633e+01	-2.1177	0.03424 *
s4	8.7291e+01	5.0463e+01	1.7298	0.08372 .
dsk	3.9628e+02	1.8675e+00	212.1952	< 2.2e-16 ***
atlyg	4.2060e-03	2.1167e-04	19.8709	< 2.2e-16 ***
ter	-9.8901e-01	1.0997e+00	-0.8993	0.36851

```
nace1          5.2527e-04  6.3577e-04   0.8262   0.40873
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Total Sum of Squares:    2.5884e+11
Residual Sum of Squares: 1.125e+10
R-Squared           : 0.95654
Adj. R-Squared      : 0.95514
F-statistic: 16903.1 on 8 and 6144 DF, p-value: < 2.22e-16
```

Nors įmonių skaičius mažesnis, paklaidų kvadratų sumos padidėjo.

Nors ir su ilgesnėmis laiko eilutėmis, duomenys vistiek turi labai daug tuščių reikšmių. Todėl pagalvojau, kad juos galima užpildyti pasitelkus R f-ją *na.spline*. Rezultatai pateikti žemiau. Šis būdas nepadėjo, todėl toliau darbą tęsiu su neužpildytais duomenimis.

Oneway (individual) effect Pooling Model

Call:

```
plm(formula = fm.s, data = long.na, effect = "individual", model = "pooling")
```

Balanced Panel: n=723, T=12, N=8676

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-23000	-701	-239	268	94400

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	-2.1065e+04	8.7935e+04	-0.2395	0.8107
t	1.0869e+01	4.3833e+01	0.2480	0.8042
s2	-2.2560e+01	1.0172e+02	-0.2218	0.8245
s3	-3.5925e+01	1.0366e+02	-0.3466	0.7289
s4	1.0037e+02	1.0649e+02	0.9426	0.3459
dsk	3.7582e+02	4.2400e+00	88.6369	<2e-16 ***
atlyg	6.4654e-03	4.6738e-04	13.8333	<2e-16 ***
ter	-3.2631e+00	2.2504e+00	-1.4500	0.1471
nace1	-1.0060e-03	1.3448e-03	-0.7481	0.4544

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Total Sum of Squares:    3.2084e+11
Residual Sum of Squares: 9.6131e+10
R-Squared           : 0.70037
Adj. R-Squared      : 0.69965
F-statistic: 2532.35 on 8 and 8667 DF, p-value: < 2.22e-16
```

Taip pat tą patį modelį pritaikiau ir trumpoms laiko eilutėms:

Oneway (individual) effect Pooling Model

```
Call:
plm(formula = fm.s, data = short, effect = "individual", model = "pooling")
```

Unbalanced Panel: n=45, T=1-7, N=216

```
Residuals :
      Min. 1st Qu.  Median 3rd Qu.    Max.
-4090.0  -487.0   -40.8   317.0 13400.0
```

```
Coefficients :
      Estimate Std. Error t-value Pr(>|t|)
(Intercept)  5.4537e+04  2.3085e+05  0.2362  0.81348
t            -2.6654e+01  1.1493e+02 -0.2319  0.81684
s2           4.1659e+01  3.5548e+02  0.1172  0.90682
s3           1.7546e+02  3.5575e+02  0.4932  0.62239
s4          -3.8530e+01  3.6090e+02 -0.1068  0.91508
dsk          3.8532e+02  2.3039e+01 16.7249 < 2e-16 ***
atlyg       3.8330e-03  2.9308e-03  1.3079  0.19237
ter         1.4844e+01  7.2532e+00  2.0465  0.04197 *
nace1      -2.5119e-03  2.9458e-03 -0.8527  0.39481
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Total Sum of Squares:    2809100000
Residual Sum of Squares: 414530000
R-Squared      : 0.85244
Adj. R-Squared : 0.81692
F-statistic: 149.472 on 8 and 207 DF, p-value: < 2.22e-16
```

Rezultatai taip pat nieko gero nežada. Bet ir nėra, ko norėti, kai laiko eilutės trumpesnės negu 4.

Dar pabandyti pagerinti ilgesniųjų laiko eilučių prognozes. Skaidysiu duomenis į keles grupes pagal įmonių dydį, t.y. *dsk*. Tai darysiu pagal histogramą 5. Pirmąją grupę sudarys įmonės su mažesniu nei 5 darbuotojų skaičiumi, antrąją – 5:10 darbuotojų, trečiąją – 10:20, ketvirtąją – 20:40 ir penktąją įmonės su 40 ir daugiau darbuotojų.

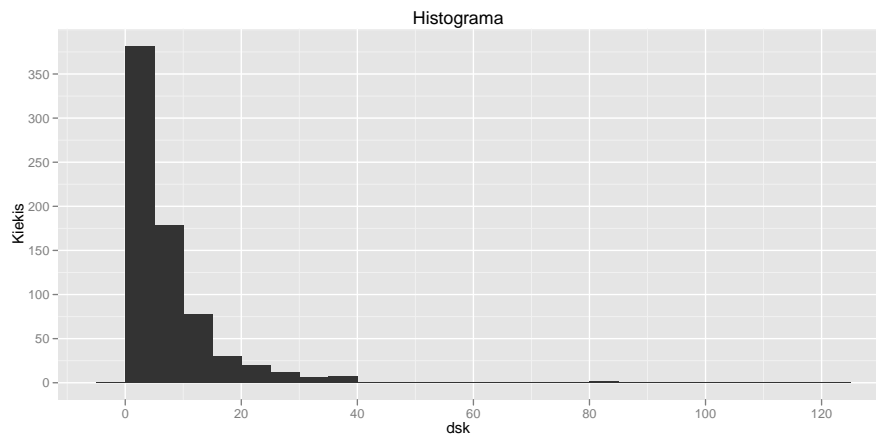
```
$gr1
$gr1$Summary
Oneway (individual) effect Pooling Model
```

```
Call:
plm(formula = fm.s, data = datt, effect = "individual", model = "pooling")
```

Unbalanced Panel: n=376, T=2-16, N=2626

```
Residuals :
      Min. 1st Qu.  Median 3rd Qu.    Max.
-1740.0  -211.0   -20.5   208.0  3670.0
```

```
Coefficients :
```

5 pav.: Ilgesniųjų laiko eilučių darbuotojų skaičiaus histograma

```

              Estimate Std. Error t-value Pr(>|t|)
(Intercept) 9.3500e+04 1.3626e+04  6.8619 8.450e-12 ***
t            -4.6940e+01 6.7891e+00 -6.9141 5.894e-12 ***
s2           5.5413e+00 2.1305e+01  0.2601  0.79481
s3           1.6915e+01 2.1506e+01  0.7865  0.43164
s4           2.2854e+01 2.1874e+01  1.0448  0.29620
dsk          3.3360e+02 5.7455e+00 58.0625 < 2.2e-16 ***
atlyg        7.3894e-03 5.1654e-04 14.3056 < 2.2e-16 ***
ter          4.7941e-01 4.5031e-01  1.0646  0.28715
nace1        6.7825e-04 3.2355e-04  2.0963  0.03615 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1160700000
Residual Sum of Squares: 382100000
R-Squared      : 0.6708
    Adj. R-Squared : 0.6685
F-statistic: 666.575 on 8 and 2617 DF, p-value: < 2.22e-16

$gr2
$gr2$Summary
Oneway (individual) effect Pooling Model

Call:
plm(formula = fm.s, data = datt, effect = "individual", model = "pooling")

Unbalanced Panel: n=184, T=3-16, N=1591

Residuals :
    Min. 1st Qu.  Median 3rd Qu.    Max.

```

-3650.0 -456.0 22.2 414.0 19000.0

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	1.1930e+05	4.1044e+04	2.9067	0.003703 **
t	-5.8900e+01	2.0447e+01	-2.8807	0.004022 **
s2	-7.3866e+01	6.2538e+01	-1.1811	0.237724
s3	-1.0345e+02	6.3342e+01	-1.6332	0.102626
s4	-4.7733e+01	6.4393e+01	-0.7413	0.458640
dsk	3.4963e+02	8.7209e+00	40.0907	< 2.2e-16 ***
atlyg	4.2306e-03	5.4317e-04	7.7887	1.216e-14 ***
ter	-1.3116e-01	1.5076e+00	-0.0870	0.930684
nacel	-1.7381e-03	6.9145e-04	-2.5137	0.012046 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 2.979e+09

Residual Sum of Squares: 1.209e+09

R-Squared : 0.59414

Adj. R-Squared : 0.59078

F-statistic: 289.489 on 8 and 1582 DF, p-value: < 2.22e-16

\$gr3

\$gr3\$Summary

Oneway (individual) effect Pooling Model

Call:

plm(formula = fm.s, data = datt, effect = "individual", model = "pooling")

Unbalanced Panel: n=109, T=4-16, N=1144

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-11000	-585	115	661	10800

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	3.9987e+05	7.1323e+04	5.6065	2.591e-08 ***
t	-2.0069e+02	3.5538e+01	-5.6473	2.059e-08 ***
s2	-2.3887e+01	1.0599e+02	-0.2254	0.8217325
s3	-9.3326e+01	1.0690e+02	-0.8730	0.3828228
s4	1.1523e+02	1.0876e+02	1.0595	0.2896002
dsk	3.6515e+02	8.3592e+00	43.6826	< 2.2e-16 ***
atlyg	3.6259e-03	5.3138e-04	6.8236	1.442e-11 ***
ter	-7.8089e+00	2.1531e+00	-3.6269	0.0002996 ***
nacel	4.0885e-03	1.3426e-03	3.0453	0.0023779 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 6.241e+09
 Residual Sum of Squares: 1811500000
 R-Squared : 0.70975
 Adj. R-Squared : 0.70416
 F-statistic: 346.924 on 8 and 1135 DF, p-value: < 2.22e-16

\$gr4
 \$gr4\$Summary
 Oneway (individual) effect Pooling Model

Call:
 plm(formula = fm.s, data = datt, effect = "individual", model = "pooling")

Unbalanced Panel: n=44, T=9-16, N=640

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-11000.0	-918.0	83.4	948.0	38400.0

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	1.7173e+05	1.8129e+05	0.9473	0.34387
t	-8.7030e+01	9.0339e+01	-0.9634	0.33573
s2	-2.4933e+02	2.6864e+02	-0.9281	0.35370
s3	-5.0674e+02	2.7132e+02	-1.8677	0.06227 .
s4	5.4017e+02	2.7506e+02	1.9638	0.04999 *
dsk	4.2566e+02	1.0577e+01	40.2453	< 2e-16 ***
atlyg	1.5370e-03	7.2438e-04	2.1218	0.03424 *
ter	1.4305e+01	8.6257e+00	1.6584	0.09773 .
nacel	2.9507e-03	3.2173e-03	0.9171	0.35942

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 1.5105e+10
 Residual Sum of Squares: 3580600000
 R-Squared : 0.76295
 Adj. R-Squared : 0.75223
 F-statistic: 253.866 on 8 and 631 DF, p-value: < 2.22e-16

\$gr5
 \$gr5\$Summary
 Oneway (individual) effect Pooling Model

Call:
 plm(formula = fm.s, data = datt, effect = "individual", model = "pooling")

Unbalanced Panel: n=10, T=12-16, N=152

```

Residuals :
    Min. 1st Qu.  Median 3rd Qu.    Max.
-22900   -1850     346    1890   20500

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  8.9031e+05  9.3180e+05  0.9555  0.340949
t            -5.3493e+02  4.0479e+02 -1.3215  0.188444
s2           -1.4006e+03  1.1034e+03 -1.2693  0.206381
s3           -8.4022e+02  1.1528e+03 -0.7289  0.467267
s4            2.7847e+02  1.1649e+03  0.2390  0.811418
dsk           4.1116e+02  1.5958e+01 25.7653 < 2.2e-16 ***
atlyg         6.0684e-03  1.2279e-03  4.9420 2.134e-06 ***
ter          -2.1521e+02  7.7158e+01 -2.7892  0.006004 **
nace1         2.6250e-01  8.3571e-01  0.3141  0.753896
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Total Sum of Squares:    3.9656e+10
Residual Sum of Squares: 3273400000
R-Squared      : 0.91746
Adj. R-Squared : 0.86313
F-statistic: 198.676 on 8 and 143 DF, p-value: < 2.22e-16

```

Padariusi prognozes ir suskaičiavusi jų skirtumą nuo realių darbo valandų, gavau, jog skaidymas į grupes pagerino absoliutinių skirtumų sumą, t.y:

```
> sum(abs(errors.long$Skirtumas), na.rm = TRUE)
```

```
[1] 1093791
```

```
> sum(abs(errors.gr$Skirtumas), na.rm = TRUE)
```

```
[1] 993769.5
```

Dar bandžiau padaryti paprastą tiesinę regresiją kiekvienai įmonei atskirai. Tačiau geresnių rezultatų gauti nepavyko. Galutines prognozes, bei skirtumus nuo realių valandų galima pamatyti šio dokumento prisegtuکه pavadinimu *Prognozes.xls*.

3 Išvados

Labai gerų rezultatų gauti nepavyko. To priežastys gali būti:

- blogai sudarytas modelis
- per didelė įmonių įvairovė
- pasirinkti ne tie metodai
- skaidyta ne teisingai ar į perdideles grupes

Tačiau lyginant procentinį paklaidų dydį su realiais duomenimis, prognozė pakankamai gerai atitinka realybę.