

VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
**MATEMATINĖS STATISTIKOS KATEDRA**

Magistro baigiamasis darbas

**Utilizavimo proceso laiko eilučių modelis**

**Time Series Model for Waste Utilization**

Olga Michailova

(parašas)

Vilnius, 2012

MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
**MATEMATINĖS STATISTIKOS KATEDRA**

Darbo vadovas:

Doc. Dr. Rimantas Eidukevičius (parašas)

Recenzentas:

Doc. Dr. Vytautas Kazakevičius (parašas)

Registracijos Nr.: .....

Darbo gynimo data: .....

# TURINYS

<b>ĮVADAS</b> .....	<b>4</b>
<b>1. TEORINĖ DALIS</b> .....	<b>6</b>
<b>1.1. Paprastos autoregresijos modelis</b> .....	<b>6</b>
1.1.1. Sąvokos ir žymėjimai.....	6
1.1.2. Statistinis įvadas.....	7
1.1.2.1 Prognozavimas.....	8
1.1.2.2 Prognozavimas su PROC ARIMA.....	8
1.1.2.3. Postūmio atgal B operatorius (Backshift).....	8
1.1.2.4. Yule-Walker lygtys kovariacijoms.....	9
<b>1.2. Slenkančio vidurkio modelis</b> .....	<b>11</b>
1.2.1. Sąvokos ir žymėjimai.....	11
<b>1.3. Modelio indentifikavimas</b> .....	<b>12</b>
1.3.1. Stacionarumas ir apverčiamumas.....	12
1.3.2. Laiko eilučių identifikavimas.....	13
1.3.2.1. Autokovariacijų funkcija $\gamma(j)$ .....	13
1.3.2.2. Autokoreliacijos (ACF).....	15
1.3.2.3. Dalinė autokoreliacija (PACF).....	15
1.3.2.4. Autokoreliacijos (ACF) įvertiniai.....	16
1.3.2.4. Dalinių autokoreliacijų (PACF) įvertiniai.....	16
1.3.2.5. Atvirkštinės autokoreliacijos (IACF).....	16
1.3.2.7. Atvirkštinių autokoreliacijų įvertiniai.....	16
1.3.3. $\chi^2$ testas paklaidoms.....	17
1.3.4. Modelio identifikavimo santrauka.....	17
<b>1.4. Tiesinė keletu kintamųjų regresija</b> .....	<b>18</b>
<b>2. PRAKTINĖ DALIS</b> .....	<b>21</b>
2.1. Statistiniai duomenys.....	21
2.2. Duomenų paruošimas analizei.....	21
<b>2.3 Prognozavimas</b> .....	<b>28</b>
<b>2.4 Regresinė analizė</b> .....	<b>32</b>
<b>IŠVADOS</b> .....	<b>37</b>
<b>SANTRAUKA</b> .....	<b>38</b>
<b>SUMMARY</b> .....	<b>39</b>
<b>PRIEDAI</b> .....	<b>41</b>

## IVADAS

Dabartiniu metu mūsų šalies ekonominė padėtis reikalauja daugelį įmonių taupyti ir optimizuoti savo išteklius. Neišimtis ir UAB „Rietavo veterinarinė sanitarija“. Tai vienintelė įmonė Lietuvoje, utilizuojanti gyvūninės kilmės atliekas.

Ji nusprendė tausoti savo žmogiškuosius ir energijos išteklius.

Iš viso galima išskirti du utilizavimo etapus. Kai baigiasi vienas, tada pradedamas kitas. Pirmojo proceso, kurį aš toliau vadinsiu tiesiog utilizavimo procesu, pabaiga – tai momentas, kai nusistovi svoris, jau nebekinta taip staigiai. Tada paleidžiamas kitas etapas- sterilizacija. Šiuo metu visa eiga prižiūri žmogus. Jis ir nustato kada baigiasi vienas ir prasideda kitas procesas. Tad momentas, kada procesas jau nusistovėjo nustatomas „iš akies“. Tai gana neoptimalu, tad reikia optimizuoti šį procesą. Jei jį pavyktų formalizuoti, galima būtų surasti sprendimo taisyklę, kuri leistų tą, kai įvyksta nusistovėjimas, nustatyti tiksliau išvengiant laiko ir finansinių sąnaudų.

Galimybė prognozuoti nusistovėjimą arba taip vadinama pasikeitimo tašką gali sumažinti pačio utilizavimo proceso savikainą.

Norint pagerinti šios įmonės darbą, reikia optimizuoti utilizavimo procesą. Tuo metu kai procesą valdo žmogus atsiranda nusistovėjimo aptikimo paklaidos.

Mano darbo tikslas padėti optimizuoti utilizavimo procesą, t.y. savalaikiai nustatyti kada pasibaigia utilizavimo procesas ir galima pradėti kitą. Tam man reikėjo ištirti utilizavimo procesus.

Savo darbe analizavau gyvūninės kilmės atliekų utilizavimo įmonės pateiktus duomenis. Analizavau svorio kitimą vykstanti utilizavimo metu.

Darbai su duomenimis bei analizei panaudojau duomenų analizės sistemos SAS paketą bei programas Microsoft Office Excel ir Word.

Darbą sudaro dvi dalys: teorinė ir praktinė. Praktinėje suformuluoju užduotis ir jas sprendžiu naudodamasi teorinėje dalyje išdėstyta medžiaga.

Darbo teorinėje dalyje aprašomi laiko eilučių modeliai ir tiesinė regresija.

Praktinė dalis suskirstyta į kelias dalis. Pirmiausia aprašomi analizuoti duomenys – iš kur jie buvo gauti, kaip buvo išrenkami bei susisteminami darbai. Tada pasinaudojus laiko eilučių modeliais, buvo aprašytas metodas galintis padėti rasti pasikeitimo tašką.

Darbo pabaigoje pateikiamos išvados, gautos atlikus analizes ir tyrimus, darbo santrauka lietuvių ir anglų kalbomis, naudotos literatūros sąrašas ir priedai.

#### PASTABA

Visi darbe naudojami papildomi priedai (sukurtos programos, visi duomenys, duomenų lentelės ir t.t.) yra įrašyti į kompaktą, kuris yra pridėtas prie šio darbo.

# 1. TEORINĖ DALIS

## 1.1. Paprastos autoregresijos modelis

### 1.1.1. Sąvokos ir žymėjimai

Tegu  $e_t \sim N(0, \sigma^2)$ ,  $t=0,1,2,\dots$  ir nekoreliuotos, jos dar vadinamos baltu triukšmu. Dažnai stebimi dydžiai  $Y_t$  (čia  $t$  laikas) gali būti prognozuojami naudojant jų ankstesnes reikšmes  $Y_{t-1}$ ,  $Y_{t-2}$ , ...

**Apibrėžimas.**  $Y_t - \mu = \rho(Y_{t-1} - \mu) + e_t$  vadinama autoregresijos procesu, žymima AR(1) (žr. [2]).

Tada

$$Y_{t-1} - \mu = \rho(Y_{t-2} - \mu) + e_{t-1} .$$

Paskutinę lygybę įrašome į pirmąją ir gauname:

$$Y_t - \mu = \rho^2 Y_{t-2} - \rho\mu + \rho e_{t-1} + e_t ,$$

pratęšę gauname

$$Y_t - \mu = \sum_{j=0}^{t-1} e_{t-j} \rho^j + \rho^t (Y_0 - \mu) .$$

Jei  $|\rho| < 1$ , tai ankstesnių eilučių įtaka dabarčiai yra silpna. Tarkime, kad

$$\text{var}(Y_{t-1}) = \frac{\sigma^2}{1 - \rho^2} ,$$

$$\text{Tada } \text{var}(Y_t) = \text{var}(\rho Y_{t-1} + e_t) = \frac{\sigma^2}{1 - \rho^2} .$$

Tokiu atveju dydžių  $Y_t$  dispersijos yra pastovios. Tai vienas iš sekos stacionarumo požymių.

### 1.1.2. Statistinis įvadas

Prateskime reiškinį formulėje  $Y_t - \mu = \sum_{j=0}^{t-1} e_{t-j} \rho^j + \rho^t (Y_0 - \mu)$  į praeitį iki begalybės:

$$Y_t = \sum_{j=0}^{\infty} \rho^j e_{t-j} + \mu.$$

Jei  $|\rho| < 1$ , tai kuo senesnis „šokas“ (t.y.  $e_{t-j}$ ), to jo įtaka mažesnė.

Lygybė  $Y_t - \mu = \sum_{j=0}^{t-1} e_{t-j} \rho^j + \rho^t (Y_0 - \mu)$  yra vadinama eilutės Wold'o pavidalu ar reprezentacija, kuri padeda rasti dydžių  $Y_t$  ir  $Y_{t-j}$  kovariaciją  $\text{cov}(Y_t, Y_{t-j})$  (žr. [2]).

Tegu  $j > 0$ , tada

$$\text{cov}(Y_t, Y_{t-j}) = \text{cov}\left(\sum_{k=0}^{\infty} \rho^k e_{t-k} + \mu, \sum_{k=0}^{\infty} \rho^k e_{t-j-k} + \mu\right) = \frac{\rho^{|j|} \sigma^2}{1 - \rho^2}.$$

Pažymėkime  $\gamma(j) = \text{cov}(Y_t, Y_{t-j})$ , o  $\gamma(0) = \frac{\sigma^2}{1 - \rho^2}$ , t.y. dydžių  $Y_t$  dispersija.

Gavome, kad kovariacija nepriklauso nuo  $t$ . tai yra kita sekos stacionarumo savybė.

Iš Wold'o reprezentacijos lengva matyti, kad  $EY_t = \mu$ . Tai yra dar viena sekos stacionarumo savybė (sekos elementų vidurkis nepriklauso nuo laiko).

Apskaičiuotos dispersijos ir kovariacijos padeda parinkti duomenims tinkamą modelį, pvz. bandant priderinti duomenims  $Y_t - \mu = \rho(Y_{t-1} - \mu) + e_t$  modelį, reikia rasti kovariacijų įverčius ir patikrinti ar jie yra reikiamo pavidalo, t.y. ar jie mažėja eksponentiškai.

$$\gamma(j) = \rho^{(j)} \text{var}(Y_t)$$

$$\frac{\gamma(j)}{\gamma(j-1)} = \frac{\rho^{(j)} \text{var}(Y_t)}{\rho^{(j-1)} \text{var}(Y_t)} = \frac{\rho^{(j)}}{\rho^{(j-1)}} = \rho.$$

Paimsime kiek sudėtingesnę  $Y_t$  priklausomybę nuo praeities:

$$Y_t - \mu = \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + e_t.$$

Šis modelis vadinamas antros eilės autoregresijos modeliu, žymimas AR(2). (žr. [2]).

### 1.1.2.1 Prognozavimas

Bendru atveju prognozė po L žingsnių atrodo šitaip:

$$\hat{Y}_{n+L} = \mu + \rho^L (Y_n - \mu)$$

su paklaida

$$Y_{n+L} - \hat{Y}_{n+L} = e_{n+L} + \rho e_{n+L-1} + \dots + \rho^{L-1} e_{n+1}.$$

Prognozavimo etapai:

1. Tiriama kovariacijų  $\gamma(j)$  įverčiai ir tikrinama ar jie mažėja eksponentiškai.
2. Jei taip, tai panaudojamas  $Y_t - \mu = \rho(Y_{t-1} - \mu) + e_t$  modelis ir randami vidurkio  $\mu$  ir koeficiento  $\rho$  įverčiai.
3. Tada apskaičiuojama prognozė  $\hat{Y}_{n+L} = \mu + \rho^L (Y_n - \mu)$  ir jos paklaidos dispersija  $\sigma^2(1 + \rho^2 + \rho^4 + \dots + \rho^{2L-2})$ . Paskutinėse lygybėse parametrai  $\mu$  ir  $\rho$  pakeičiami jų įverčiais (žr. [2]).

### 1.1.2.2 Prognozavimas su PROC ARIMA

Prognozavimui naudojama SAS procedūra PROC ARIMA. Sintaksė (žr. [3]):

```
PROC ARIMA DATA=lentelė;  
IDENTIFY VAR=Y CENTER;  
ESTIMATE P=1 NOCONSTANT;  
FORECAST LEAD=5;  
RUN;
```

### 1.1.2.3. Postūmio atgal B operatorius (Backshift)

Pažymėkime  $B(Y_t) = Y_{t-1}$ . Tada

$$B^2(Y_t) = B(B(Y_t)) = B(Y_{t-1}) = Y_{t-2},$$

$$B^3(Y_t) = Y_{t-3}.$$

Pasiremiamą lygybę  $(1-x)^{-1} = 1+x+x^2+x^3+\dots$ , kai  $|x| < 1$ .

$Y_t$  užrašoma kaip:

$$Y_t = \sum_{j=0}^{\infty} w_j e_{t-j}.$$



### 1.1.2.4. Yule-Walker lygtys kovariacijoms

Apskaičiuojamos kovariacijos  $\gamma(j)$  modeliui  $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + e_t$ .

Dydį  $Y_t$  dauginame iš  $Y_{t-j}$  ir skaičiuojame šios sandaugos vidurkį  $E(Y_t Y_{t-j})$ .

Yule-Walker lygtys duotam procesui:

- $\gamma(1) = \alpha_1 \gamma(0) + \alpha_2 \gamma(-1)$
- $\gamma(2) = \alpha_1 \gamma(1) + \alpha_2 \gamma(0)$

Pažymėkime, kad  $\gamma(-j) = \gamma(j)$  ir kad autokoreliacijos  $\rho(j) = \frac{\gamma(j)}{\gamma(0)}$  nepriklauso nuo dispersijos  $\sigma^2$ .

Jei nagrinėjamas pirmos eilės autoregresijos procesas AR(1), tai

- $\gamma(1) = \alpha_1 \gamma(0)$
- $\rho(1) = \frac{\gamma(1)}{\gamma(0)}$ .

Bendru atveju antros eilės autoregresijos procesui :

- $\gamma(0) = \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \sigma^2$
- $\gamma(j) = \alpha_1 \gamma(j-1) + \alpha_2 \gamma(j-2)$ , kai  $j > 0$ .

Jei paklaidų  $e_t$  dispersija yra žinoma, tai Yule-Walker lygtis galima išspręsti ir rasti autokovariacijas  $\gamma(j)$ .

Kovariacijų  $\gamma(j)$  įverčius galima rasti su procedūra PROC ARIMA.

Nagrinėsime procesą:

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \alpha_2 (Y_{t-2} - \mu) + e_t.$$

Procesą  $Y_t$  užrašome kaip triukšmų  $e_t$  tiesinę kombinaciją:

$$Y_t - \mu = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}.$$

Lygtį perrašome pasinaudoję postūmio atgal operatoriumi B:

$$Y_t - \mu = e_t + w_1 B(e_t) + w_2 B^2(e_t) = (1 + w_1 B + w_2 B^2) e_t,$$

$$(1 + w_1 B + w_2 B^2)(Y_t - \mu) = e_t.$$

Kairiuosiuose skliaustuose raidę B pakeičiame algebriniu kintamuoju M ir gauname taip vadinamąją charakteringą lygtį:

$$1 + w_1 M + w_2 M^2 = 0.$$

Šaknų modulio nelygybė 1 yra proceso stacionarumo sąlyga garantuojanti senų „šokų“ silpną įtaką dabarčiai (žr. [2]).

**Apibrėžimas.** Bendro autoregresijos p-tosios eilės proceso pavidalas:

$$Y_t - \mu = \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + \dots + \alpha_p(Y_{t-p} - \mu) + e_t.$$

Procesas žymimas AR(p) (žr. [2]).

Perrašome jį pasinaudodami operatoriumi B:

$$Y_t - \mu - \alpha_1(B - \mu) - \alpha_2(B^2 - \mu) - \dots - \alpha_p(B^p - \mu) = e_t,$$

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)(Y_t - \mu) = e_t,$$

$$Y_t - \mu = (1 + w_1 B + w_2 B^2 + \dots)e_t.$$

Paskutinėje lygybėje svoriai  $w_j$  artėja prie nulio, jei charakteringosios lygties  $1 - \sum_{i=1}^p \alpha_i M^i = 0$  šaknų moduliai didesni už 1 (žr. [2]).

Yulo-Walkerio lygtys gaunamos taip:

lygybė 
$$Y_t - \mu = \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + \dots + \alpha_p(Y_{t-p} - \mu) + e_t$$

dauginama iš  $(Y_{t-j} - \mu)$ ,  $j=0,1,\dots$ , ir apskaičiuojami sandaugų vidurkiai. Jei žinomi kovariacijų  $\gamma(j)$  įverčiai, tai randami koeficientų  $\alpha_j$  įverčiai, o kovariacijų  $\gamma(j)$  pavidalas padės parinkti duomenims tinkamą modelį (žr. [2]).

## 1.2. Slenkančio vidurkio modelis

ARMA modelis- autoregresijos ir slenkančio vidurkio procesas.

### 1.2.1. Sąvokos ir žymėjimai

Apibrėžimas. Pirmos eilės slenkančio vidurkio modelis žymimas MA(1). Jo pavidalas:

$$Y_t = \mu + e_t - \beta e_{t-1},$$

Kur  $e_t$  - vadinamas baltas triukšmas,

$$Ee_t = 0,$$

$$\text{var}(e_t) = Ee_t^2 = \sigma^2,$$

$\text{cov}(e_t, e_{t+j}) = 0$ ,  $j \neq 0$ ,  $t$  – sveiki skaičiai.

- $EY_t = E(\mu + e_t - \beta e_{t-1}) = \mu$
- $\text{cov}(Y_t, Y_{t-1}) = \text{cov}(\mu + e_t - \beta e_{t-1}, \mu + e_{t-1} - \beta e_{t-2}) = -\beta\sigma^2$
- $\text{cov}(Y_t, Y_{t-2}) = 0$
- $\text{cov}(Y_t, Y_{t-3}) = 0$
- $\text{cor}(Y_t, Y_t) = \frac{\gamma(0)}{\gamma(0)} = 1$
- $\text{cor}(Y_t, Y_{t-1}) = \frac{\gamma(1)}{\gamma(0)} = -\frac{\beta}{1 + \beta^2}$
- $\gamma(j) = \rho(j) = 0$ , kai  $j > q$ .

Apibrėžimas. Bendras slenkančio vidurkio proceso pavidalas yra toks:

$$Y_t = \mu + e_t - \sum_{i=1}^q \beta_i e_{t-i}.$$

Jis žymimas MA(q) (žr. [2]).

Perrašome modelį pasinaudodami postūmio operatoriumi:

$$Y_t = \mu + (1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q) e_t.$$

Mišriojo autoregresijos ir slenkančio vidurkio modelio ARMA(p,q) pavidalas:

$$Y_t - \mu - \alpha_1(Y_{t-1} - \mu) - \dots - \alpha_p(Y_{t-p} - \mu) = e_t - \beta_1 e_{t-1} - \dots - \beta_q e_{t-q}.$$

Užrašome su postūmio operatoriumi:

$$(1 - \alpha_1 B - \dots - \alpha_p B^p)(Y_t - \mu) = (1 - \beta_1 B - \dots - \beta_q B^q) e_t.$$

### 1.3. Modelio indentifikavimas

#### 1.3.1. Stacionarumas ir apverčiamumas

Nagrinėkime ARMA(p,q) modelį

$$(1 - \sum_{i=1}^p \alpha_i B^i)(Y_t - \mu) = (1 - \sum_{i=1}^q \beta_i B^i) e_t.$$

**Apibrėžimas.** Modelis stacionarus, jei charakteringojo polinomo  $1 - \sum_{i=1}^p \alpha_i M^i$  visų šaknų moduliai didesni už 1 (taip stacionarumą apibrėžia knygos autorius žr. [2] psl. 55).

Procesas yra stacionarus, jeigu jo vidurkis ir dispersija yra pastovieji dydžiai, t.y. nepriklauso nuo postūmio laike.

Stacionarumas užtikrina, kad senos „šokų“ reikšmės turi mažą poveikį dabarčiai. Todėl kelių pradinių paklaidų reikšmių pakeitimas nuliais turi mažos įtakos pakankamai ilgų eilučių tyrimui.

**Apibrėžimas.** Eilutė apverčiama, jei lygties  $1 - \sum_{i=1}^q \beta_i M^i = 0$  visų šaknų moduliai didesni už 1.

Dėl apverčiamumo, procesą  $Y_t$  galima užrašyti kaip sumą šoko  $e_t$  ir svertinės ankstesnių  $Y$  sumos.

### 1.3.2. Laiko eilučių identifikavimas

Modelio pavidalui nustatyti panaudojama SAS procedūra PROC ARIMA.

Randami šių funkcijų įverčiai:

1. autokorelacių funkcija ACF(autocorrelation function)
2. atvirkštinė autokorelacių funkcija IACF(inverse autocorrelation function)
3. dalinių autokorelacių funkcija PACF(partial autocorrelation function).

Šios funkcijos bus aprašytos toliau.

Procedūros PROC ARIMA sakinyje IDENTIFY kontroliuoja šių funkcijų įverčių radimą ir atitinkamų grafikų spausdinimą.

```
PROC ARIMA DATA=lentele;  
IDENTIFY VAR=Y;  
RUN;
```

#### 1.3.2.1. Autokovariacijų funkcija $\gamma(j)$

Stacionarioms laiko eilutėms  $\gamma(j) = \text{cov}(Y_t, Y_{t-j}) \quad \forall t$ .

- AR(1) modeliui:  $Y_t = \rho Y_{t-1} + e_t$ ,

$$\gamma(j) = \frac{\rho^{|j|} \sigma^2}{1 - \rho^2},$$

$$\sigma^2 = \text{var}(e_t).$$

- AR(2) modeliui:  $Y_t - \alpha_1 Y_{t-1} - \alpha_2 Y_{t-2} = e_t$ ,

su pradinėmis reikšmėmis  $\gamma(0)$ ,  $\gamma(1)$ , o vėliau tenkina lygtį

$$\gamma(j) = \alpha_1 \gamma(j-1) + \alpha_2 \gamma(j-2).$$

Kovariacijos gali svyruoti ir svyravimo periodas priklauso nuo koeficientų  $\alpha_1$ ,  $\alpha_2$ . Pradinės reikšmės randamos iš Yule- Walkerio lygčių.

- Bendroms AR(p) eilutėms:  $Y_t - \sum_{i=1}^p \alpha_i Y_{t-i} = e_t$  pradinės reikšmės  $\gamma(0), \gamma(1), \dots, \gamma(p-1)$ ,

o  $\gamma(j) = \sum_{i=1}^p \alpha_i \gamma(j-i)$ , čia  $j=p, p+1, \dots$ .

Stacionariu atveju  $|\gamma(j)| \leq c\lambda^j$ ,  $\lambda \in (0,1)$ . Kovariacijos gali svyruoti, bet vistiek turi artėti prie 0.

- MA(1) modeliui:  $Y_t - \mu = e_t - \beta e_{t-1}$ ,

$$\gamma(0) = \sigma^2(1 + \beta^2),$$

$$\gamma(1) = \gamma(-1) = \text{cov}(Y_t, Y_{t-1}) = -\beta\sigma^2,$$

$$\gamma(2) = \gamma(3) = \dots = 0, \text{ t.y. } \gamma(j) = 0, \text{ kai } |j| > 1.$$

- MA(q) modeliui:  $Y_t - \mu = e_t - \sum_{i=1}^q \beta_i e_{t-i}$ ,

$$\gamma(0), \gamma(1), \dots, \gamma(q), \gamma(q+1) = \dots = 0.$$

- ARMA(p,q) modeliui:

$$\left(1 - \sum_{i=1}^p \alpha_i B^i\right)(Y_t - \mu) = \left(1 - \sum_{i=1}^q \beta_i B^i\right)e_t,$$

pažymime  $r = \max(p-1, q)$ .

Randomos pradinės reikšmės  $\gamma(0), \gamma(1), \dots, \gamma(r)$  ir  $\gamma(j) = \sum_{i=1}^p \alpha_i \gamma(j-i)$ ,  $j > r$ .

- Balto triukšmo sekai:

$$\gamma(0) = \sigma^2,$$

$$\gamma(1) = \dots = 0,$$

$$\gamma(j) = 0, \quad j > 0.$$

### 1.3.2.2. Autokoreliacijos (ACF)

Modelio parinkimui svarbu ne tiek kovariacijų  $\gamma(j)$  dydis, kiek jų bendras pavidalas (struktūra). Todėl dažnai tiriamos autokoreliacijos  $\rho(j) = \frac{\gamma(j)}{\gamma(0)}$ .

$$\rho(0) = 1,$$

$$\rho(-j) = \rho(j).$$

### 1.3.2.3. Dalinė autokoreliacija (PACF)

Dalinės autokoreliacijos apskaičiuojamos atlikus šiuos žingsnius:

**1 žingsnis.** Apskaičiuojame stebėjimų  $Y_t$  regresiją pagal  $Y_{t-1}$ . Gautas koeficientas prie  $Y_{t-1}$  žymimas  $\hat{\pi}_1$ .

**2 žingsnis.** Stebėjimams  $Y_t$  pritaikome tiesinės regresijos modelį pagal  $Y_{t-1}$  ir  $Y_{t-2}$ . Gautas koeficientas prie  $Y_{t-2}$  žymimas  $\hat{\pi}_2$ .

...

**j žingsnis.** Stebėjimams  $Y_t$  pritaikome tiesinės regresijos modelį pagal  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-j}$ . Gautas koeficientas prie  $Y_{t-j}$  žymimas  $\hat{\pi}_j$ .

Šie koeficientai  $\hat{\pi}_i, i=1, \dots, j$  ir yra dalinių autokoreliacijų įverčiai.

Procesui AR(p) dalinės autokoreliacijos  $\pi_j = 0$ , kai  $j > p$ .

Teorinės dalinės autokoreliacijos randamos sprendžiant šias tiesines sistemas:

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(j-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(j-2) \\ \dots & \dots & \dots & \dots \\ \gamma(j-1) & \gamma(j-2) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_j \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \dots \\ \gamma(j) \end{pmatrix}$$

Fiksuotam  $j$  išsprendus šią sistemą randama  $j$ -toji dalinė autokoreliacija  $\pi_j = b_j$ . Dalinės autokoreliacijos naudingos identifikuojant autoregresijos procesus. Slenkančio vidurkio ir ARMA procesams dalinės autokoreliacijos nelygios 0.

### 1.3.2.4. Autokoreliacijos (ACF) įvertiniai

Autokovariacijos įvertinį pažymime  $c(j)$ . Jis lygus

$$c(j) = \frac{1}{n} \sum_{t=1}^{n-j} (Y_t - \bar{Y})(Y_{t+j} - \bar{Y})$$

Autokoreliacijos įvertinį (ACF) pažymime  $r(j)$ . Jis lygus

$$r(j) = \frac{c(j)}{c(0)}$$

Autokoreliacijos standartinės paklaidos radimas:

**1 žingsnis.** Autokoreliacijai  $r(j)$  priskiriama variacija  $\frac{1}{n} \sum_{i=-j+1}^{j-1} r^2(i)$ .

**2 žingsnis.** Standartinė paklaida yra paskutiniosios išraiškos kvadratinė šaknis.

**3 žingsnis.** Atitinkamos dispersijos įvertis tuo atveju, kai teisinga:  $\gamma(i) = 0, i \geq j$  ir  $\gamma(i) \neq 0, i < j$ .

### 1.3.2.4. Dalinių autokoreliacijų (PACF) įvertiniai

Dalinės autokoreliacijos (PACF) randamos sprendžiant sistemas, kurios apibrėžiamos autokovariacijomis  $\gamma(j)$ . Šios  $\gamma(j)$  pakeičiamos įvertiniais  $c(j)$ , sistemos išsprendžiamos ir taip randami dalinių autokoreliacijų įvertiniai.

### 1.3.2.5. Atvirkštinės autokoreliacijos (IACF)

Turime procesą ARMA(p,q). Jo atvirkštinės autokoreliacijos (IACF) yra proceso ARMA(q,p) autokoreliacijos (ACF).

### 1.3.2.7. Atvirkštinių autokoreliacijų įvertiniai

Tarkime, kad duomenys gauti stebint procesą AR(i). Prideriname jį duomenims, randame koeficientų įvertinius. Toliau užrašome „atvirkštinį“ modelį MA(i) ir randame jo autokovariacijų įvertinius.



Bendru atveju nei modelio pavidalas, nei jo eilė nėra žinomi. Bet bendras ARMA procesas gali būti užrašytas kaip begalinės autoregresijos procesas  $AR(\infty)$ . Todėl ARMA procesus galime aproksimuoti pakankamai didelės eilės autoregresiniais procesais.

### 1.3.3. $\chi^2$ testas paklaidoms

Nulinė hipotezė teigia, kad stebima eilutė yra baltas triukšmas; alternatyva- eilutė nėra baltas triukšmas.

$H_0$  : eilutė yra baltas triukšmas;

$H_1$ : eilutė nėra baltas triukšmas.

Naudojama Ljungo modifikuota Box-Pierce Q statistika :

$$n(n+2) \sum_{j=1}^k \frac{r^2(j)}{n-j} \sim \chi_k^2,$$

kur k teigiamas sveikas skaičius.

Identifikacijos ir įvertinimo etapuose ši statistika naudojama triukšmų „baltumui“ nustatyti. Kad modelis būtų tinkamas būtina, kad paklaidos atitiktų balto triukšmo savybes, tuo pačiu, kad statistikos reikšmė būtų maža, priešingu atveju modelis netinkamas.

### 1.3.4. Modelio identifikavimo santrauka

	MA(q)	AR(p)	ARMA(p,q)	Baltas triukšmas
ACF	D(q)	T	T	0
PACF	T	D(p)	T	0
IACF	T	D(p)	T	0

Čia T - eksponentinis mažėjimas; D(q)- charakteristika lygi 0, kai  $j > q$ .

ACF- autokoreliacijos,

PACF- atvirkštinė autokoreliacija,

IACF-dalinė autokoreliacija.

## 1.4. Tiesinė keletu kintamųjų regresija

Tiesinė regresija yra vienas iš bendros tiesinių modelių teorijos taikymo pavyzdžių. Šiuo atveju eksperimento plano matricos elementai gali būti bet kokie realieji skaičiai.

Tarkime, kad norime prognozuoti a. d.  $Y$  remdamiesi  $m$  kovariantėmis  $X_1, \dots, X_m$ . Pažymėkime  $\mathbf{X} = (X_0, X_1, \dots, X_m)^T$  kovariančių vektorių, papildytą  $X_0 \equiv 1$ . Tarkime, kad fiksuojant kovariantės  $\mathbf{X}$  reikšmes  $\mathbf{X}^{(i)} = \mathbf{x}^{(i)} = (x_{0i}, x_{1i}, \dots, x_{mi})^T$  gauti nepriklausomi a. d.  $Y$  stebėjimai  $Y_i, i = 1, \dots, n$ , t.y.  $\mathbf{x}^{(i)}$  yra neatsitiktiniai arba yra nepriklausomų vienodai pasiskirsčiusių atsitiktinių vektorių  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$  realizacijos.

Tarkime, kad stebėjimus galime užrašyti:

$$Y_i = \mu(\mathbf{x}^{(i)}) + e_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} + e_i;$$

Čia  $e_i$  nepriklausomi vienodai pasiskirstę a. d., turintys normalųjį skirstinį su nuliniu vidurkiu ir dispersija  $\sigma^2$ ;  $\beta_0, \beta_1, \dots, \beta_m$  - nežinomi parametrai (koeficientai).

Tiesinė keletu kintamųjų regresija yra vienas iš tiesinių modelių teorijos taikymo pavyzdžių. Pažymėję  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)^T$ ,  $\mathbf{e} = (e_1, \dots, e_n)^T$  ir

$$A = \begin{pmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{mn} \end{pmatrix}$$

Galime užrašyti  $\mathbf{Y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e}$ ,  $\mathbf{E}(\mathbf{Y}) = \mathbf{A}\boldsymbol{\beta}$ ,  $\mathbf{V}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ . Parametro  $\boldsymbol{\beta}$  įvertinys randamas mažiausių kvadratų metodu:

$$\hat{\boldsymbol{\beta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}; \mathbf{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}.$$

Regresijos  $\mu(x) = E(Y|x)$  įvertinys:

$$\hat{\mu}(x) = \hat{\beta}^T x = C_m.$$

Nepaslinktas dispersijos  $\sigma^2$  įvertinys:

$$\hat{\sigma}^2 = s^2 = \frac{SS_E}{n-m-1}.$$

$$SS_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

$$\hat{Y}_i = \hat{\mu}(x^{(i)}) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{im}.$$

Atsitiktiniai dydžiai  $\hat{Y}_i$  ir  $Y_i$  vadinami atitinkamai prognozuotomis ir stebėtomis priklausomo kintamojo  $Y$  reikšmėmis, o a. d.  $\hat{e}_i = Y_i - \hat{Y}_i$  vadinami liekamosiomis paklaidomis.

$$SS_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

$$SS_R = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2,$$

$$SS_T = SS_R + SS_E = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Parametro  $\beta_i$  pasiklovimo intervalas, esant pasiklovimo lygmeniui  $Q = 1 - \alpha$ , yra

$$\hat{\beta}_i \pm \sqrt{c_{ii}} st_{\alpha/2}(n-m-1);$$

Čia  $c_{ii}$  yra matricos  $(A^T A)^{-1}$  elementai.

Parametro  $\mu(x)$  pasiklovimo intervalas, esant pasiklovimo lygmeniui  $Q = 1 - \alpha$ :

$$\hat{\mu}(x) \pm b(x) st_{\alpha/2}(n-m-1);$$

$$\text{čia } b^2(x) = x^T (A^T A)^{-1} x.$$

Tarkime, kad reikia patikrinti hipotezę:

$$H_{1,\dots,m} : \beta_1 = \dots = \beta_m = 0,$$

t.y. regresijos modelis netinka. Kriterijaus statistika

$$F_{1,\dots,m} = \frac{MS_R}{MS_E}$$

Turi Fišerio skirstinį su  $m$  ir  $n-m$ -ą laisvės laipsniais. Hipotezė  $H_{1,\dots,m}$  atmetama esant reikšmingumo lygmeniui  $\alpha$ , jei  $F_{1,\dots,m} > F_{\alpha}(m, n-m-1)$  (žr. [1]).

Tarkime, kad reikia patikrinti hipotezę:

$H_j : \beta_j = 0$ , t. y. Kovariantė  $x_j$  nėra reikšminga priklausomo kintamojo prognozei.

Kriterijaus statistika

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{c_{jj}}}$$

Turi Stjudento skirstinį su  $n-m$ -ą laisvės laipsniais.

Modelio tinkamumą parodo determinacijos koeficientas (R-square):

$$R^2 = \frac{SS_R}{SS_t}$$

Kuo determinacijos koeficiento reikšmė didesnė, tuo geriau  $Y$  reikšmes aprašo į regresijos modelį įtrauktos kovariantės. Dydis  $R_{Y \bullet x_1 \dots x_m} = \sqrt{R^2}$  vadinamas empiriniu dauginiu koreliacijos koeficientu (žr. [1]).

Koreguotas determinacijos koeficientas (Adj R-Sq)

$$R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-m-1}$$

Parodo modelio tinkamumą. Atsižvelgiama į imties didumą ir nepriklausomų kintamųjų skaičių.

Dalinės (angl. partial) koreliacijos koeficientai parodo  $Y$  priklausomybę nuo vieno kintamojo, kai eliminuojama likusių kintamųjų įtaka. Tiesinę  $Y$  ir  $X_j$  priklausomybę įvertina koreliacijos koeficientas, tačiau jis gali būti didelis vien todėl, kad abu kintamieji  $Y$  ir  $X_j$  priklauso nuo

kokių nors kitų kintamųjų. Kintamieji, į kurių įtaką skaičiuojant dalinę koreliaciją atsižvelgta, vadinami kontroliuojamais kintamaisiais. Kontroliuojamų kintamųjų skaičius vadinamas dalinės koreliacijos eile. Daugialypėje regresijoje dažniausiai skaičiuojamos (m-1)-osios eilės dalinės koreliacijos (žr. [1]).

## 2. PRAKTINĖ DALIS

### 2.1. Statistiniai duomenys

Duomenys buvo gauti iš įmonės utilizuojančios gyvūninės kilmės atliekas. Duomenų faile buvo pateikti duomenys gauti vykstant utilizacijos procesui. Kaip atrodo duomenys galima pamatyti prieduose (Priedas 1).

Utilizacijos proceso metu katile yra fiksuojami šie rodikliai: data, laikas, svoris, garų slėgis, drėgmė ir temperatūra pirmame ir antrame katile. Taip pat duomenų faile kiekvienam stebėjimui yra priskirtas unikalus numeris- ID. Duomenys yra fiksuojami apytiksliai kas 10 sekundžių. Duomenys yra fiksuojami vienu metu dviejuose katiluose.

Pats procesas atrodo šitaip: į katilą pakraunamos mėsos atliekos, jos apdorojamos ten termiškai, po to baigus apdorojimui yra pradedama sterilizacija, ji duomenyse atsispindi slėgio ir temperatūros padidėjimu. Dėl sterilizacijos įvyksta laikinas svorio rodiklio padidėjimas.

### 2.2. Duomenų paruošimas analizei

Pradėjus dirbti su duomenimis išsiaiškinau, kad duomenų faile yra kelios laiko atkarpos kur duomenys buvo fiksuojami nenutrūkstamai. Atskiriau tas duomenų atkarpas į 6 laiko intervalus, kur rodikliai buvo fiksuojami nenutrūkstamai, rezultatus pateikiau 2.1 lentelėje.

2.1 lentelė.

Nr.	ID		Data		Laikas	
	nuo	iki	nuo	iki	nuo	iki
1	326164	331849	2011.06.15	2011.06.15	00:00:20	15:50:33

2	331850	337535	2011.06.15	2011.06.15	00:00:20	15:50:33
3	337536	345578	2011.06.17	2011.06.18	00:00:05	00:00:27
4	345579	353770	2011.06.19	2011.06.19	00:00:18	22:52:09
5	353771	361962	2011.06.20	2011.06.20	00:00:17	22:48:08
6	361963	370154	2011.06.18	2011.06.18	00:00:27	23:34:05

Čia: Nr - nenutrūkstamo laiko intervalo numeris; ID - duomenų faile esantis unikalus numeris ir nuo kurios ir iki kurios jo reikšmės vyksta nenutrūkstamas rodiklių fiksavimas.

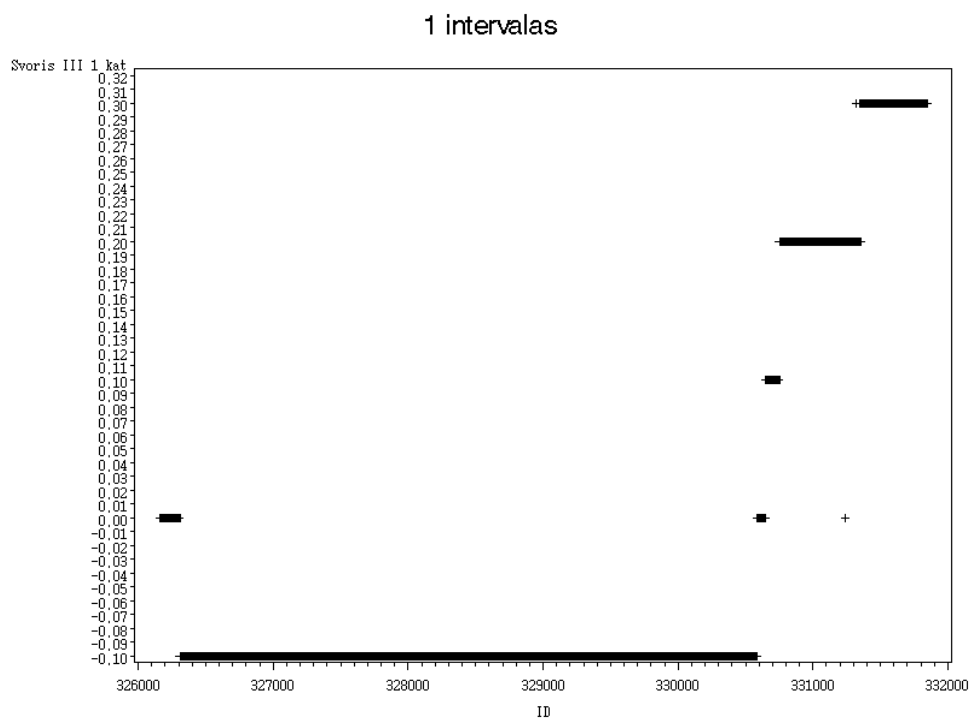
**Pastabos.** Intervalų numeriai yra priskirti tokia tvarka, kokia duomenų faile yra pateikti duomenys. Pastebėkime, kad du pirmi intervalai sutampa. Taip pat 6 intervalas gali būti 3 tęsinys.

Toliau nagrinėjau kiekvieną intervalą atskirai. Tam nusprendžiau nubraižyti kiekviename intervale ir kiekviename katile vykstanti svorio kitimo grafiką atskirai.

Tam pradžioje duomenis iš Microsoft Excel rinkmenos „duomenys.xls“ importavau į duomenų analizės sistemos SAS paketą. Jame duomenis atskyriau į intervalus, kur duomenys buvo fiksuojami nenutrūkstamai. Nubraižiau kiekvieno intervalo ir kiekviename katile vykstantį svorio kitimo grafiką.

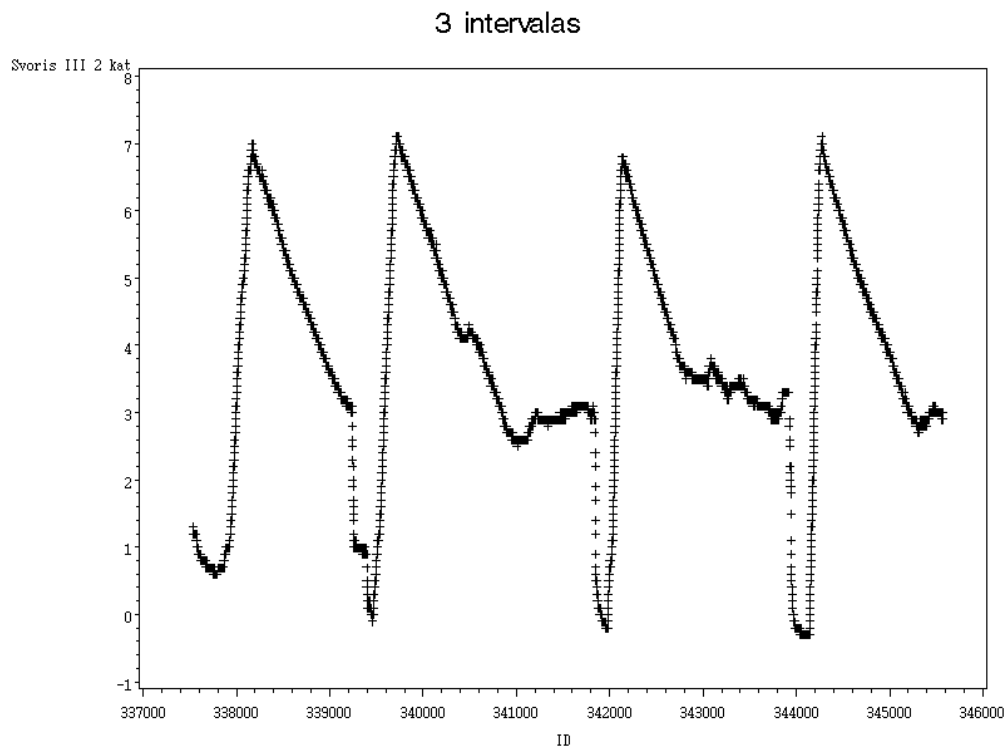
1 ir 2 intervalų svorio kitimo grafikuose utilizavimo procesų nematau, 1 intervalo svorio kitimas pavaizduotas 2.1 grafike.

## 2.1 grafikas. 1 intervalo svorio kitimas.

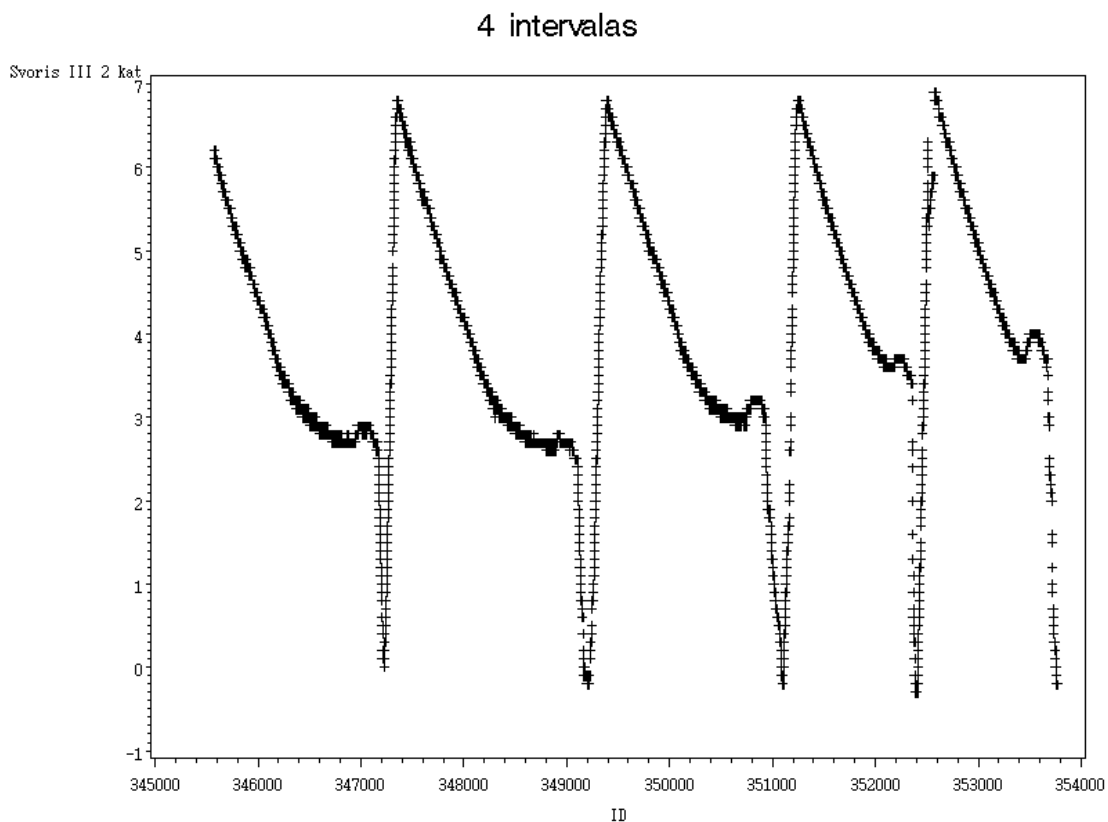


5 intervalas taip pat mažai informatyvus. O 3, 4 ir 6 intervaluose galime išskirti atskirus utilizavimo procesus. Grafikai 2.2, 2.3 ir 2.4.

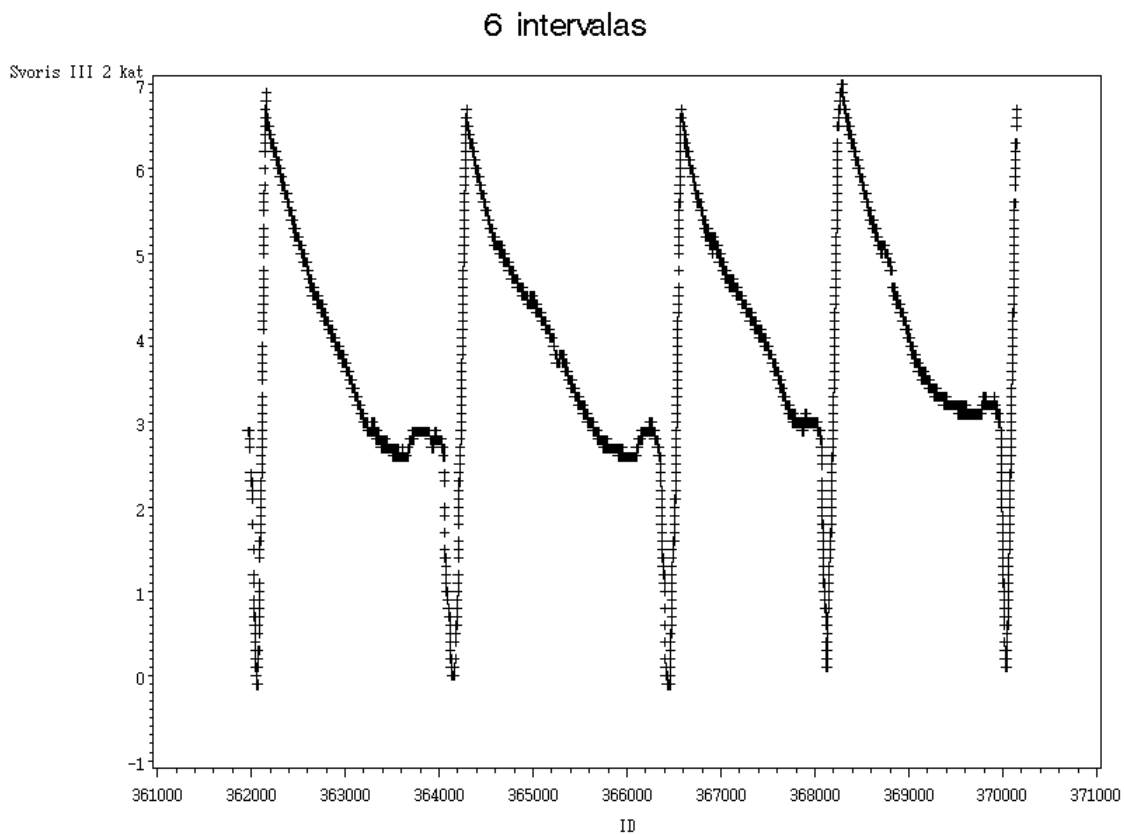
## 2.2 grafikas. 3 intervalo 2 katilo svorio kitimo grafikas.



2.3 grafikas. 4 intervalo 2 katilo svorio kitimo grafikas.



2.4 grafikas. 6 intervalo 2 katilo svorio kitimo grafikas.



Toliau nagrinėjau 3, 4 ir 6 intervalus. Juose atskiriau atskirus utilizavimo procesus.



2.2 lentelėje pateikiau 1 katilė vykstančių procesų pradžią ir pabaigą nurodžius jų ID numerius.

**2.2 lentelė. 1 katilė vykstantys procesai.**

Procesas	ID	
	nuo	iki
1	338620	339863
2	340353	340868
3	343548	344180
4	345286	345578
5	346066	347180
6	347729	348642
7	349083	350242
8	351947	352577
9	361963	362700
10	363384	364411
11	365140	366502
12	367084	368369
13	368823	370154

**Pastabos.** Proceso pradžia laikau, kai katilas yra pilnai pakrautas, t.y. kai yra maksimalus svoris, o proceso pabaiga - kai yra paleidžiama sterilizacija. 4, 8, 9 ir 13 procesus toliau nenagrinėjau, nes trūksta dalies jų duomenų.

Taip pat padariau ir su 2 katilo duomenimis.

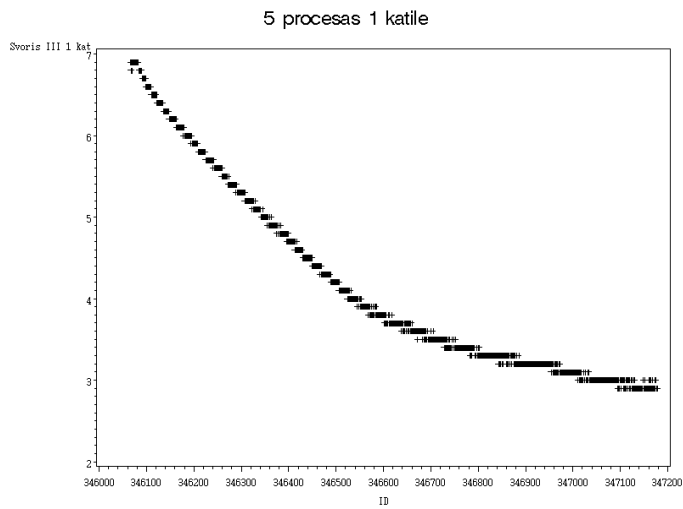
**2.3 lentelė. 2 katilė vykstantys procesai.**

Procesas	ID	
	nuo	iki
1	338169	339160
2	339756	340478
3	342178	343049
4	344291	345389
5	345579	346940
6	347351	348892
7	349387	350740
8	351277	352150
9	352586	353452
10	362158	363655
11	364290	366090
12	366575	367868
13	368293	369765

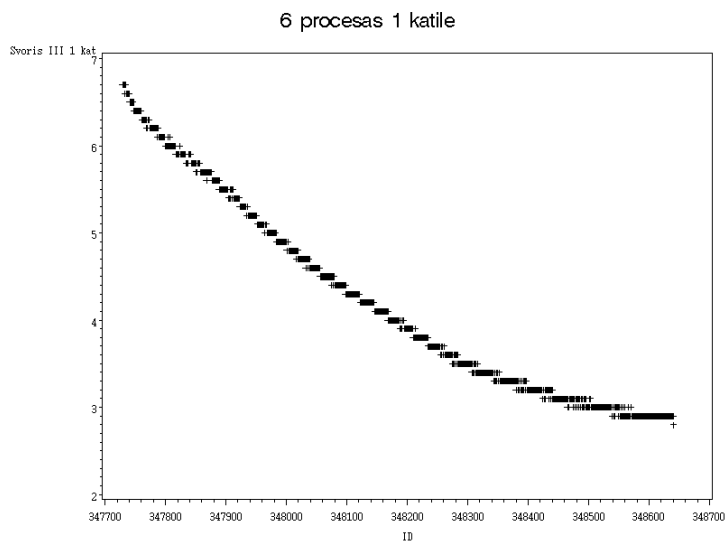
**Pastaba.** 5 procesą toliau nenagrinėsiu, nes trūksta dalies duomenų.

Kiekvieną procesą nubraižiau atskirai. Keli iš jų pavaizduoti 2.5, 2.6, 2.7, 2.8 grafikuose.

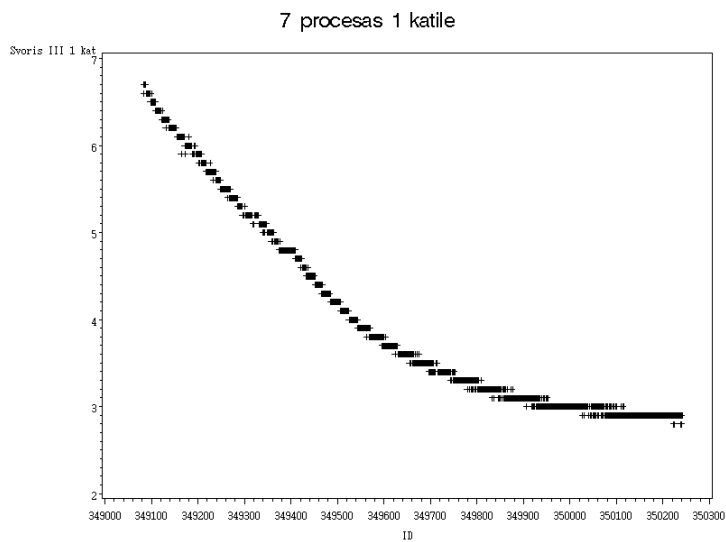
**2.5 grafikas. 5 procesas vykstantys 1 katile.**



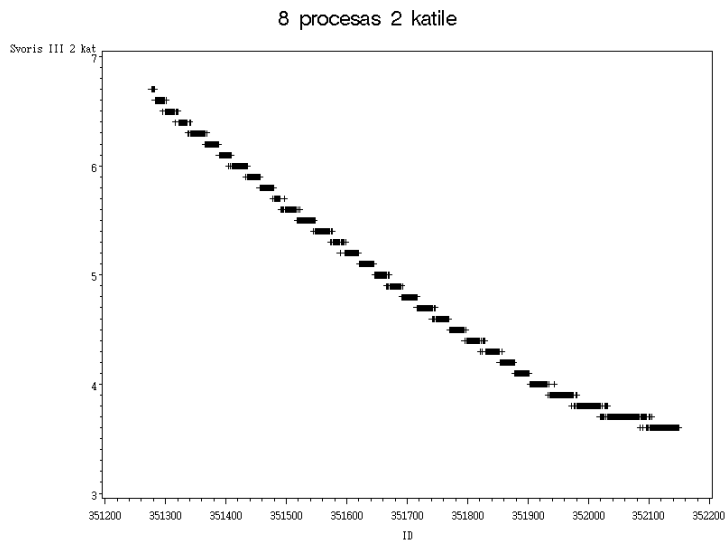
**2.6 grafikas. 6 procesas vykstantys 1 katile.**



**2. 7 grafikas. 7 procesas vykstantys 1 katile.**



## 2.8 grafikas. 8 procesas vykstantys 2 katile.

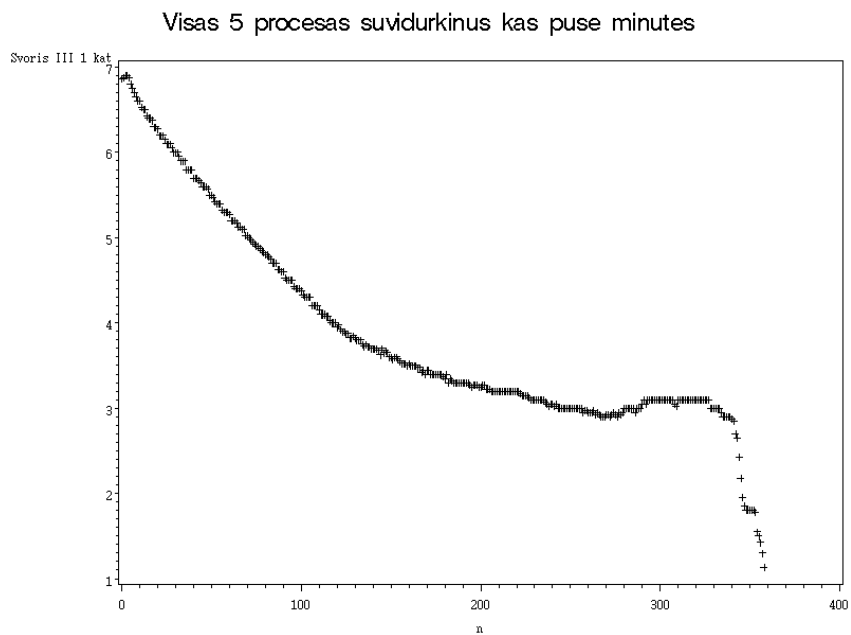


## 2.3 Prognozavimas

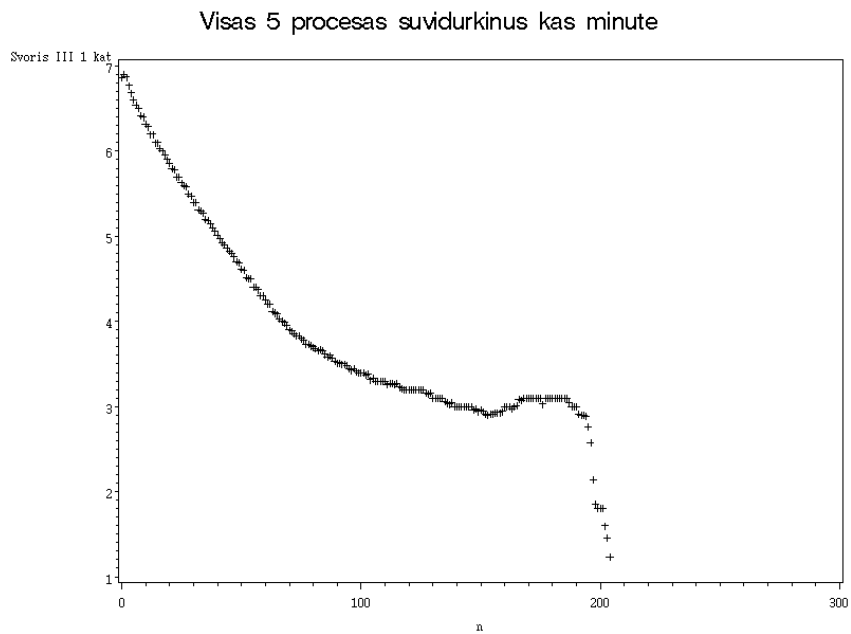
Kad galima būtų analizuoti utilizavimo proceso duomenis, aš juos pertvarkiau. Suskaičiavau vidurkius kas 4 stebėjimus (4 stebėjimai sudaro pusę minutės), tai nėra tikras kas pusės minutės suvidurkinimas, bet toliau aš naudosiu šį terminą. Taip pat kas 7 stebėjimus (sudaro minutę), kas 10 stebėjimų (pusantros minutės) ir kas 13 stebėjimų (dvi minutės) paskaičiavau vidurkius. Atsižvelgus į gautus rezultatus nusprendžiau su kuriais duomenimis toliau atliksiu analizę.

Toliau pateikiu gautus rezultatus 1 katile vykusio 5 proceso pavyzdžiu. Gavau šiuos grafikus.

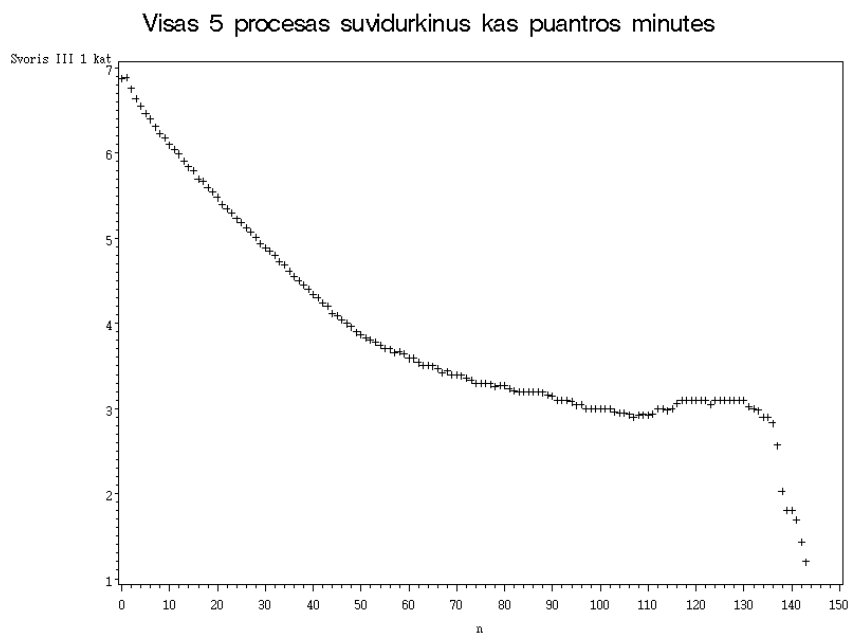
**2.9. grafikas. Visas 5 procesas nuo pakrovimo į 1 katilą iki iškrovimo, suvidurkinus kas pusę minutės.**



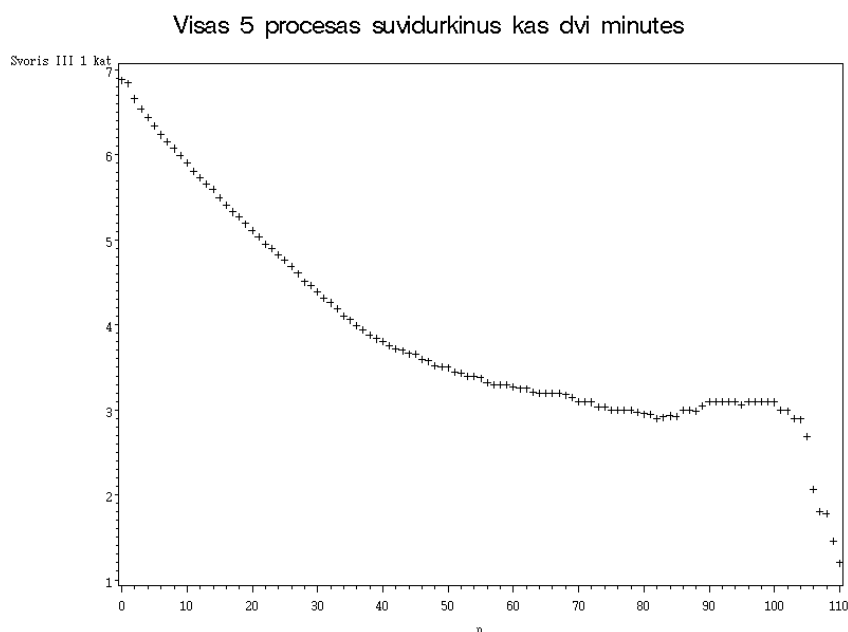
**2.10. grafikas. Visas 5 procesas nuo pakrovimo į 1 katilą iki iškrovimo, suvidurkinus kas minutę.**



**2.11. grafikas. Visas 5 procesas nuo pakrovimo į 1 katilą iki iškrovimo, suvidurkinus kas pusantros minutės.**



**2.12. grafikas. Visas 5 procesas nuo pakrovimo į 1 katilą iki iškrovimo, suvidurkinus kas dvi minutes.**



Toliau atliksiu prognozavimą duomenims suvidurkintiems kas pusę minutės ir bandau jiems pritaikyti savo utilizavimo proceso pabaigos ieškojimo metodą. Ieškome utilizavimo proceso pasikeitimo taško (change point).

Pasiūliau itin paprasta net naivu metodą pasikeitimo taškui rasti.

Metodo mintis tokia: paimeame duomenų dalį ir jiems sudarome prognozę. Po to žiūrime ar realūs duomenys patenka į pasikliautinąjį prognozės intervalą. Jei taip, tai padidiname duomenų

kiekį link utilizavimo proceso pabaigos. Mes siekiame sužinoti kiek arti reikia prieit prie utilizavimo proceso pabaigos, kad realūs duomenys nepatektų į prognozės intervalą ir ar realių duomenų nepatekimas į prognozės intervalą leidžia nustatyti pasikeitimo tašką.

Tam paėmiau duomenų dalį ir prognozavau į priekį. Vėliau paėmiau daugiau duomenų vis artėdama prie utilizavimo proceso pabaigos.

Pateiksiu skaičiavimus atliktus su 5 proceso vykusio 1 katile duomenimis.

Paleidu procedūrą PROC ARIMA:

```
proc arima data=minutev;  
  identify var=vid center NLAG=15;  
run;  
quit;
```

Gaunu, kad pradiniems duomenims autokoreliacijų funkcija labai lėtai mažėja prie 0 (Output langą rasite 2 priede). Todėl tolesnę analizę atlikau pirmiems skirtumams  $Y_t - Y_{t-1}$ .

```
proc arima data=minutev;  
  identify var=vid(1) NLAG=15;  
run;  
quit;
```

Taip pat taikiau pirmos eilės slenkančio vidurkio modelį.

Procedūra ARIMA apskaičiuoja 95% prognozių pasiliautinius intervalus.

Paėmiau dalį duomenų pasibaigiančių už 50 stebėjimų iki pasikeitimo taško ir atlikau jiems prognozę tap pat dalį duomenų pasibaigiančių už 10 stebėjimų nuo pasikeitimo taško prognozė.

Programos kodas ir rezultatai prieduose ( Priedas 3).

Pastebiu, kad pasikeitimo taškas šiuo atveju yra 268 stebėjimas ir jo reikšmė yra 2,9.

Gavau, kad kiek arti bebūtume prie pasikeitimo taško, t.y. kada yra paleidžiama sterilizacija, vis vien realūs duomenys niekada nevišyja prognozės intervalo rėžių.

Taigi gavau, kad šis metodas neduoda mums reikiamų rezultatų, t.y. juo naudodamiesi mes negalime rasti utilizavimo proceso pabaigos. Tai gali būti dėl to, kad sterilizacijos metu matavimo prietaiso užfiksuotas svorio pasikeitimas nėra pakankamai didelis.

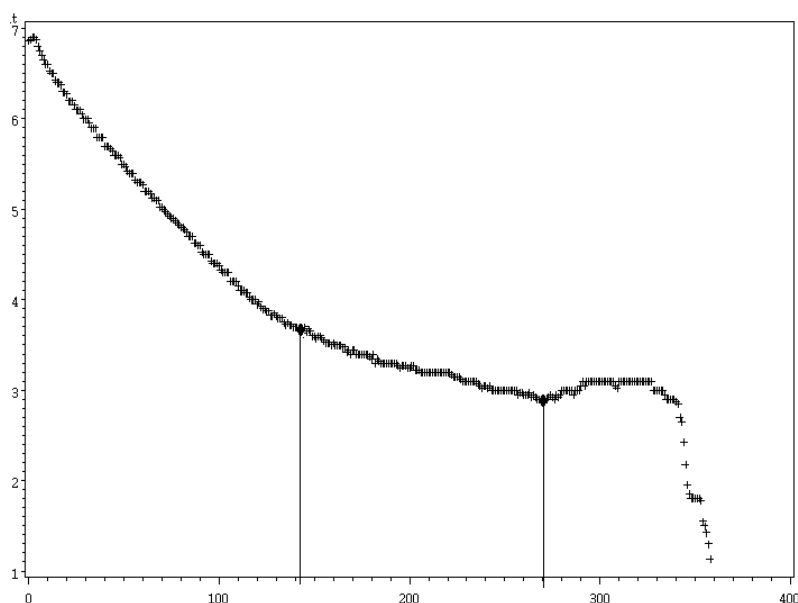
## 2.4 Regresinė analizė

Toliau utilizavimo procesams taikiau tiesinės regresijos modelį. Tam kiekvienam procesui atskirai radau du pasikeitimo taškus (change point), t.y. taškus kur proceso elgsena pasikeičia.

Pirmą pasikeitimo tašką, kurį žymiu cp1, radau vizualiai, antrą (cp2) radau pagal slėgio padidėjimą. Antras taškas reiškia sterilizacijos pradžią.

Pavyzdys 2.13 grafike.

**2.13 grafikas. Pasikeitimo taškai.**



Atrinkau 15 procesų tinkamų tolesniai analizei. Rinkausi tuos procesus, kurie turi tik du pasikeitimo taškus. Tai procesai, kur svoris kurį laiką kinta gan staigiai, o vėliau svoris kinta mažesniu greičiu arba nusistovi, kas galėtų reikšti, kad sterilizacija galėjo prasidėti anksčiau. Nenagrinėju taip pat procesų, kurie turi kelis svorio padidėjimus. Priežastis kodėl įvyksta keli svorio padidėjimai nežinoma. Buvo atrinkti iš 1 katilė vykusių šie procesai: 3, 5, 6, 10, o 2 katilo – 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13.

Tariau, kad regresinė kreivė aprašoma tokiu būdu:

$$X_t = a_0 + a_1 t + \alpha X_{t-1} + e_t.$$

Atlikau su SAS procedūra PROC REG skaičiavimus, gavau  $a_0$ ,  $a_1$  ir  $\alpha$  įverčius. SAS programa ir Output langas yra prieduose (Priedas 4).



Atlikau tokius skaičiavimus atrinktiems utilizavimo procesams, t.y. tiems kurie elgiasi panašiai. Procesus sunumeravau iš naujo. Rezultatai 2.4 lentelėje.

2.4 lentelė.  $a_0$ ,  $a_1$  ir  $\alpha$  įverčiai ir  $\hat{cp}1$  ir  $\hat{cp}2$  prognozuotos reikšmės.

procesas	a0	a1	alfa	cp1	cp2	$\hat{cp}1$	$\hat{cp}2$
1	5,89602	-0,01995	0,09388	98	162	102,33	255,35
2	4,98151	-0,01705	0,05742	140	268	1256,7	270,38
3	2,55	-0,0081	0,0674	150	230	159,4	304,78
4	3,442445	-0,0078	0,2385	160	271	159,53	292,89
5	5,63072	-0,1466	0,10941	165	290	178,55	325
6	4,9808	-0,01463	0,22232	120	180	120,02	268,37
7	5,09237	-0,0139	0,24421	110	257	129	268,19
8	2,55091	-0,00581	0,4594	240	381	168,34	305,4
9	3,4056	-0,00769	0,45801	190	338	154,3	292,4
10	3,31787	-0,00844	0,44813	160	218	153,87	293,79
11	5,15478	-0,01297	0,20338	120	211	128,29	266,86
12	6,11407	-0,01569	0,03722	138	371	101,09	250,48
13	4,40973	-0,01323	0,20706	79	451	132,38	277,54
14	2,78057	-0,00811	0,24412	88	301	163,17	302,4
15	4,81502	-0,1262	0,2779	190	371	172,03	326,15

Čia cp1 yra pirmas pasikeitimo taškas, o cp2- antras.

Tada pritaikau regresiją jau gautiems rezultatams. Mano tikslas sužinoti ar pasikeitimo taškas tiesiškai priklauso nuo gautų koeficientų įverčių ir kaip. Tikriname tokį sąryšį:

$$cp1 = \gamma_0 + \gamma_1 a_0 + \gamma_2 a_1 + \gamma_3 \alpha + e_t$$

Procedūros kodą ir rezultatų Output langą galima rasti prieduose (Priedas 5).

Hipotezė, kad visi koeficientai statistiškai nereikšmingi (regresijos modelis netinka), neatmetama (p - reikšmė lygi 0,202).

Determinacijos koeficiento reikšmė 0,3315, t. y. regresijos modelis paaikškina 33,15% duomenų sklaidos. Koreguotojo determinacijos koeficiento reikšmė 0,1492. Jis parodo modelio netinkamumą. Taip gavau, kad  $a_0$ ,  $a_1$  ir  $\alpha$  statistiškai nereikšmingi.

Vadinasi šis modelis netinka šiems mūsų duomenims. Taip galėjo atsitikti dėl to, kad taikiau matematinį modelį savo pasirinktam pasikeitimo taškui.

Pamėginau pritaikyti antram pasikeitimo taškui modelį:

$$cp2 = \gamma_0 + \gamma_1 a_0 + \gamma_2 a_1 + \gamma_3 \alpha + e_t.$$

Rezultatai Outpute prieduose (Priedas 6).

Čia determinacijos koeficiento reikšmė 0,0819, t. y. regresijos modelis paaikškina 8,19% duomenų sklaidos. Hipotezė, kad visi koeficientai statistiškai nereikšmingi, neatmetama (p - reikšmė lygi 0,8058). Šis modelis taip pat netinka. Taip pat  $a_0$ ,  $a_1$  ir  $\alpha$  statistiškai nereikšmingi.

Tarkime, kad regresinė kreivė aprašoma tokiu būdu:

$$X_t = a_0 + a_1 t + a_2 t^2 + \alpha X_{t-1} + e_t.$$

Atlikau su SAS procedūra PROC REG skaičiavimus, gavau  $a_0$ ,  $a_1$ ,  $a_2$  ir  $\alpha$  įverčius. Rezultatai 2.5 lentelėje.

2.5 lentelė.  $a_0$ ,  $a_1$ ,  $a_2$  ir  $\alpha$  įverčiai ir  $\hat{cp}1$  ir  $\hat{cp}2$  prognozuotos reikšmės.

procesas	a0	a1	a2	alfa	cp1	cp2	$\hat{cp}1$	$\hat{cp}2$
1	5,91458	-0,0203	0,00000335	0,09388	98	162	132,58	201,9
2	6,40041	-0,02525	0,00002566	0,05742	140	268	125,99	220,51
3	6,31872	-0,03114	0,0000735	0,0674	150	230	93,56	259,75
4	4,00987	0,01021	-0,00009761	0,2385	160	271	137,09	293,73
5	6,70319	-0,02287	0,00002756	0,10941	165	290	154,8	323,77
6	5,09922	-0,0156	0,00000518	0,22232	120	180	145,57	255,46
7	5,14223	-0,01427	0,00000223	0,24421	110	257	157,29	283,76

8	3,72006	-0,01054	0,00000897	0,4594	240	381	186,51	301,72
9	3,68439	-0,00941	0,0000054	0,45801	190	338	184,54	306,91
10	3,66952	-0,01024	0,00000663	0,44813	160	218	180,33	290,01
11	5,46885	-0,01493	0,00000924	0,20338	120	211	144,62	314,2
12	6,34286	-0,01535	-0,00000945	0,03722	138	371	119,59	278,4
13	5,23176	-0,01773	0,00003192	0,20706	79	451	120,44	310,67
14	5,00587	-0,01835	0,00005402	0,24412	88	301	106,93	359,87
15	4,91572	-0,01336	0,00000547	0,2779	190	371	158,15	299,31

Tikriname sąryšį

$$cp1 = \gamma_0 + \gamma_1 a_0 + \gamma_2 a_1 + \gamma_3 a_2 + \gamma_4 \alpha + e_t.$$

Gauti rezultatai prieduose (Priedas 7).

Determinacijos koeficiento reikšmė 0,4123, t. y. regresijos modelis paaiškina 41,23% duomenų sklaidos. Koreguotojo determinacijos koeficiento reikšmė 0,1772. Jis parodo modelio netinkamumą. Hipotezė, kad visi koeficientai statistiškai nereikšmingi, neatmetama (p - reikšmė lygi 0,2147). Šis modelis taip pat netinka.

O modeliui  $cp2 = \gamma_0 + \gamma_1 a_0 + \gamma_2 a_1 + \gamma_3 a_2 + \gamma_4 \alpha + e_t$ , hipotezė, kad visi koeficientai statistiškai nereikšmingi, taip pat neatmetama (p - reikšmė lygi 0,5716). Determinacijos koeficiento reikšmė 0,2343, t. y. regresijos modelis paaiškina 23,43% duomenų sklaidos. Koreguotojo determinacijos koeficiento reikšmė -0,072. Jis parodo modelio netinkamumą. Taip gavau, kad  $\gamma_0$ ,  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$  ir  $\gamma_4$  statistiškai nereikšmingi (Priedas 8).

**2.6 lentelė.**

Modelis	Pasikeitimo taškas	p - reikšmė	$R^2$	$R_a^2$
1	cp1	0,202	0,3315	0,1492
	cp2	0,8058	0,0819	-0,1684
	cp1	0,2147	0,4123	0,1772

2	cp2	0,5716	0,2343	-0,072
---	-----	--------	--------	--------

Čia 1 modelis yra  $X_t = a_0 + a_1t + \alpha X_{t-1} + e_t$ , o 2-as  $X_t = a_0 + a_1t + a_2t^2 + \alpha X_{t-1} + e_t$ ,  $R^2$  yra determinacijos koeficientas,  $R_a^2$  - koreguotas determinacijos koeficientas.

Išsiaiškinau, kad negalime taikyti tiesinės regresijos modelio pasikeitimo taškams. Pirmiesiems pasikeitimo taškams tai gali būt dėl to, kad juos aš radau vizualiai.

## IŠVADOS

Atlikus gautų duomenų sisteminimą bei analizę buvo:

- Paruošti analizei gyvūninės kilmės atliekų utilizavimo proceso duomenys.
- Suformuluotas ir tikrintas pasikeitimo taško nustatymo metodas.
- Gauta, kad šis modelis šiems duomenims netinkamas.
- Panaudojus tiesinę regresiją, rasti nežinomų parametrų įverčiai ir jais paremtas tiesinio sąryšio tarp pasikeitimo taško ir gautų įverčių ieškojimas.
- Nustatyta, kad nėra tiesinio sąryšio tarp pasikeitimo taškų ir gautų įverčių.

Pastabos: tolesniam utilizavimo proceso tyrimui reikia daugiau duomenų, kad galima būtų išskirti visus proceso niuansus. Taip pat reikia daugiau informacijos apie fizikinius ir cheminius procesus vykstančius katile. Spėju, kad proceso elgesys labai priklauso nuo pakrautų atliekų charakteristikų.

## Utilizavimo proceso laiko eilučių modelis

### SANTRAUKA

Šiame darbe buvo atlikta gyvūninės kilmės atliekų utilizavimo proceso analizė.

Pagrindinis uždavinys- rasti būdą prognozuoti utilizavimo proceso pabaigą ir tuo sumažinti energijos suvartojimą.

Naudojasi laiko eilučių prognozavimo modeliu. Aprašiau savo metodą pasikeitimo taškui rasti. Taip pat buvo panaudota tiesinė regresija.

Galimybė prognozuoti pasikeitimo tašką leistų žymiai sumažinti utilizavimo proceso savikainą.

**Raktiniai žodžiai** : utilizavimo procesas, pasikeitimo taškas, laiko eilutės, tiesinė regresija.

## **Time Series Model for Waste Utilization**

### **SUMMARY**

I this work, an analysis of animal waste utilization process was performed.

The main task was to find a way to predict the end of the desiccation process, because possibility to predict this end point may reduce energy consumption.

I used the time series forecasting model and proposed method for the change point detection. Linear regression was also used for this task.

**Key words :** Waste Utilization, change point, time series, linear regression

## LITERATŪRA

1. R. Levulienė. Statistikos taikymai naudojant SAS. Vilnius, Vilniaus universiteto leidykla, 2009
2. Brocklebank, John C., and David A. Dickey. SAS® for Forecasting Time Series, 2003
3. SAS Online Help



# PRIEDAI

## Priedas 2

### Output langas:

The ARIMA Procedure

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
15	-0.01960										. .											

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	0.98350										. *****											
2	0.00318										. .											
3	-0.00522										. .											
4	-0.00882										. .											
5	-0.00802										. .											
6	0.03421										. *											
7	0.00092										. .											
8	-0.00621										. .											
9	-0.00809										. .											
10	-0.00910										. .											
11	-0.01847										. .											
12	0.00288										. .											
13	0.00022										. .											
14	0.01820										. .											
15	0.03852										. *											

Autocorrelation Check for White Noise

To	Chi-Square	DF	Pr >	ChiSq	-----Autocorrelations-----																	
6	1953.81	6	<.0001	0.983	0.967	0.951	0.935	0.919	0.905													
12	3581.98	12	<.0001	0.890	0.876	0.861	0.847	0.832	0.818													

## Priedas 3

```
/* 50 stebejimu iki change point */  
data dalis10;
```

```

set minutev;
if n<218;
run;
proc arima data= dalis8;
  identify var=vid(1) noprint;
  estimate Q=1 noconstant noprint;
  forecast lead=70;
run;
quit;

/* 10 stebejimu iki change point */
data dalis11;
  set minutev;
  if n<258;
run;
proc arima data= dalis9;
  identify var=vid(1) noprint;
  estimate Q=1 noconstant noprint;
  forecast lead=30;
run;

```

## Output langas:

```

The ARIMA Procedure

Model for variable vid

Period(s) of Differencing  1

No mean term in this model.

Moving Average Factors

Factor 1: 1 + 0.0059 B**(1)

Forecasts for variable vid

Obs   Forecast   Std Error   95% Confidence Limits
219   3.2000     0.2069     2.7945     3.6055
220   3.2000     0.2935     2.6248     3.7752
221   3.2000     0.3598     2.4948     3.9052
222   3.2000     0.4156     2.3853     4.0147
223   3.2000     0.4648     2.2889     4.1111
224   3.2000     0.5093     2.2018     4.1982
225   3.2000     0.5502     2.1216     4.2784
226   3.2000     0.5882     2.0471     4.3529
227   3.2000     0.6240     1.9770     4.4230
228   3.2000     0.6578     1.9108     4.4892
229   3.2000     0.6899     1.8478     4.5522
230   3.2000     0.7206     1.7876     4.6124

```

231	3.2000	0.7501	1.7299	4.6701
232	3.2000	0.7784	1.6743	4.7257
233	3.2000	0.8058	1.6207	4.7793
234	3.2000	0.8322	1.5689	4.8311
235	3.2000	0.8578	1.5187	4.8813
236	3.2000	0.8827	1.4699	4.9301
237	3.2000	0.9069	1.4225	4.9775
238	3.2000	0.9305	1.3763	5.0237
239	3.2000	0.9535	1.3312	5.0688
240	3.2000	0.9759	1.2872	5.1128
241	3.2000	0.9979	1.2442	5.1558
242	3.2000	1.0194	1.2021	5.1979
243	3.2000	1.0404	1.1609	5.2391
244	3.2000	1.0610	1.1205	5.2795
245	3.2000	1.0812	1.0808	5.3192
246	3.2000	1.1011	1.0419	5.3581
247	3.2000	1.1206	1.0037	5.3963
248	3.2000	1.1397	0.9662	5.4338
249	3.2000	1.1586	0.9292	5.4708
250	3.2000	1.1771	0.8929	5.5071
251	3.2000	1.1954	0.8571	5.5429

03:27 Sunday, May 30, 2012

The ARIMA Procedure

Forecasts for variable vid

Obs	Forecast	Std Error	95% Confidence Limits	
252	3.2000	1.2134	0.8218	5.5782
253	3.2000	1.2311	0.7871	5.6129
254	3.2000	1.2486	0.7529	5.6471
255	3.2000	1.2658	0.7191	5.6809
256	3.2000	1.2828	0.6858	5.7142
257	3.2000	1.2996	0.6529	5.7471
258	3.2000	1.3161	0.6205	5.7795
259	3.2000	1.3325	0.5884	5.8116
260	3.2000	1.3486	0.5567	5.8433
261	3.2000	1.3646	0.5254	5.8746
262	3.2000	1.3804	0.4945	5.9055
263	3.2000	1.3960	0.4639	5.9361
264	3.2000	1.4114	0.4337	5.9663
265	3.2000	1.4267	0.4038	5.9962
266	3.2000	1.4418	0.3742	6.0258
267	3.2000	1.4567	0.3449	6.0551
268	3.2000	1.4715	0.3159	6.0841
269	3.2000	1.4862	0.2872	6.1128
270	3.2000	1.5007	0.2588	6.1412
271	3.2000	1.5150	0.2306	6.1694
272	3.2000	1.5292	0.2027	6.1973

273	3.2000	1.5433	0.1751	6.2249
274	3.2000	1.5573	0.1477	6.2523
275	3.2000	1.5712	0.1206	6.2794
276	3.2000	1.5849	0.0937	6.3063
277	3.2000	1.5985	0.0670	6.3330
278	3.2000	1.6120	0.0406	6.3594
279	3.2000	1.6254	0.0143	6.3857
280	3.2000	1.6386	-0.0117	6.4117
281	3.2000	1.6518	-0.0375	6.4375
282	3.2000	1.6649	-0.0631	6.4631
283	3.2000	1.6778	-0.0885	6.4885
284	3.2000	1.6907	-0.1137	6.5137
285	3.2000	1.7034	-0.1387	6.5387
286	3.2000	1.7161	-0.1635	6.5635
287	3.2000	1.7287	-0.1882	6.5882
288	3.2000	1.7412	-0.2126	6.6126

## Output langas:

The ARIMA Procedure

Model for variable vid

Period(s) of Differencing 1

No mean term in this model.

Moving Average Factors

Factor 1:  $1 + 0.00589 B^{**}(1)$

Forecasts for variable vid

Obs	Forecast	Std Error	95% Confidence Limits	
259	3.0000	0.1901	2.6273	3.3727
260	3.0000	0.2697	2.4714	3.5286
261	3.0000	0.3306	2.3520	3.6480
262	3.0000	0.3820	2.2513	3.7487
263	3.0000	0.4272	2.1627	3.8373
264	3.0000	0.4680	2.0826	3.9174
265	3.0000	0.5056	2.0090	3.9910
266	3.0000	0.5406	1.9405	4.0595
267	3.0000	0.5734	1.8761	4.1239
268	3.0000	0.6045	1.8152	4.1848
269	3.0000	0.6340	1.7573	4.2427

270	3.0000	0.6622	1.7020	4.2980
271	3.0000	0.6893	1.6490	4.3510
272	3.0000	0.7154	1.5979	4.4021
273	3.0000	0.7405	1.5487	4.4513
274	3.0000	0.7648	1.5010	4.4990
275	3.0000	0.7883	1.4549	4.5451
276	3.0000	0.8112	1.4101	4.5899
277	3.0000	0.8335	1.3665	4.6335
278	3.0000	0.8551	1.3240	4.6760
279	3.0000	0.8762	1.2826	4.7174
280	3.0000	0.8969	1.2421	4.7579
281	3.0000	0.9170	1.2026	4.7974
282	3.0000	0.9368	1.1639	4.8361
283	3.0000	0.9561	1.1261	4.8739
284	3.0000	0.9751	1.0889	4.9111
285	3.0000	0.9936	1.0525	4.9475
286	3.0000	1.0119	1.0168	4.9832
287	3.0000	1.0298	0.9816	5.0184
288	3.0000	1.0474	0.9471	5.0529

## Priedas 4

### Output langas:

The REG Procedure

Model: MODEL1

Dependent Variable: vid Svoris III 1 kat

#### Analysis of Variance

Source	DF	Sum of	Mean	F Value	Pr > F
		Squares	Square		
Model	2	14.80355	7.40177	15679.7	<.0001
Error	67	0.03163	0.00047206		
Corrected Total	69	14.83518			

Root MSE      0.02173    R-Square    0.9979  
 Dependent Mean    4.95714    Adj R-Sq    0.9978  
 Coeff Var        0.43830

#### Parameter Estimates

Variable	Label	Parameter		Standard		
		DF	Estimate	Error	t Value	Pr >  t
Intercept	Intercept	1	4.98151	0.78706	6.33	<.0001
x1		1	-0.01705	0.00269	-6.33	<.0001
vidm1		1	0.25012	0.11779	2.12	0.0374

## Priedas 5

### Procedūros kods:

```
proc reg data=koef;
  model cp1= a0 a1 alfa;
run;
quit;
```

### Output langas:

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: cp1 cp1

#### Analysis of Variance

Source	DF	Sum of		F Value	Pr > F
		Squares	Mean Square		
Model	3	8707.08555	2902.36185	1.82	0.2020
Error	11	17557	1596.11950		
Corrected Total	14	26264			

Root MSE      39.95146    R-Square    0.3315  
 Dependent Mean    143.20000    Adj R-Sq    0.1492  
 Coeff Var      27.89907

#### Parameter Estimates

Variable	Label	Parameter		Standard		
		DF	Estimate	Error	t Value	Pr >  t
Intercept	Intercept	1	-162.44995	721.32130	-0.23	0.8259
a0	a0	1	37.57849	109.02153	0.34	0.7368
a1	a1	1	-338.41324	342.87777	-0.99	0.3449

alfa alfa 1 380.45940 729.97996 0.52 0.6126

## Priedas 6

### Output langas:

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: cp2 cp2

#### Analysis of Variance

Source	DF	Sum of	Mean	F Value	Pr > F
		Squares	Square		
Model	3	7785.95492	2595.31831	0.33	0.8058
Error	11	87239	7930.85258		
Corrected Total	14	95025			

Root MSE 89.05533 R-Square 0.0819  
 Dependent Mean 286.66667 Adj R-Sq -0.1684  
 Coeff Var 31.06581

#### Parameter Estimates

Variable	Label	Parameter DF	Standard		t Value	Pr >  t
			Estimate	Error		
Intercept	Intercept	1	277.91008	1607.88873	0.17	0.8659
a0	a0	1	-6.53467	243.01861	-0.03	0.9790
a1	a1	1	-475.51864	764.30477	-0.62	0.5465
alfa	alfa	1	67.61466	1627.18966	0.04	0.9676

## Priedas 7

### Output langas:

The REG Procedure  
 Model: MODEL1  
 Dependent Variable: cp1 cp1

Analysis of Variance

Source	DF	Sum of	Mean	F Value	Pr > F
		Squares	Square		
Model	4	10829	2707.12582	1.75	0.2147
Error	10	15436	1543.58967		
Corrected Total	14	26264			

Root MSE      39.28854   R-Square   0.4123  
 Dependent Mean   143.20000   Adj R-Sq   0.1772  
 Coeff Var      27.43613

Parameter Estimates

Variable	Label	Parameter DF	Parameter	Standard	t Value	Pr >  t
			Estimate	Error		
Intercept	Intercept	1	-226.90734	285.08678	-0.80	0.4446
x1		1	41.83295	50.05357	0.84	0.4228
x2		1	-3366.69438	5566.83138	-0.60	0.5588
x3		1	-1126537	1028702	-1.10	0.2991
alfa	alfa	1	505.92416	285.69776	1.77	0.1070

## Priedas 8

The REG Procedure

Model: MODEL1

Dependent Variable: cp2 cp2

Analysis of Variance

Source	DF	Sum of	Mean	F Value	Pr > F
		Squares	Square		
Model	4	22263	5565.70025	0.76	0.5716
Error	10	72763	7276.25323		
Corrected Total	14	95025			

Root MSE      85.30096   R-Square   0.2343  
 Dependent Mean   286.66667   Adj R-Sq   -0.0720  
 Coeff Var      29.75615

Parameter Estimates



Variable	Label	Parameter		Standard		
		DF	Estimate	Error	t Value	Pr >  t
Intercept	Intercept	1	-108.54769	618.96351	-0.18	0.8643
x1		1	108.87667	108.67334	1.00	0.3400
x2		1	18708	12086	1.55	0.1527
x3		1	3230684	2233456	1.45	0.1786
alfa	alfa	1	377.52947	620.29004	0.61	0.5563