ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ | DEMOCRITUS UNIVERSITY OF THRACE

Στοχαστική Ανάλυση Χρονοσειρών

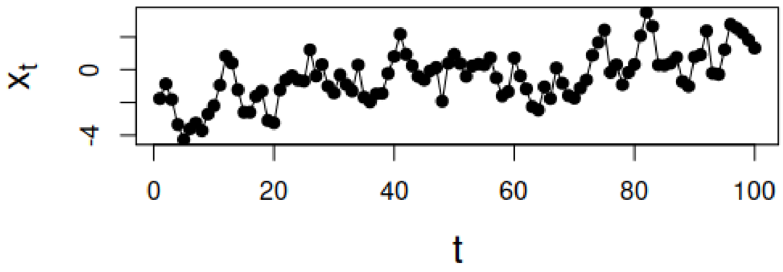
Περιεχόμενο:

Δραστηριότητες εξοικείωσης με το διάγραμμα ACF

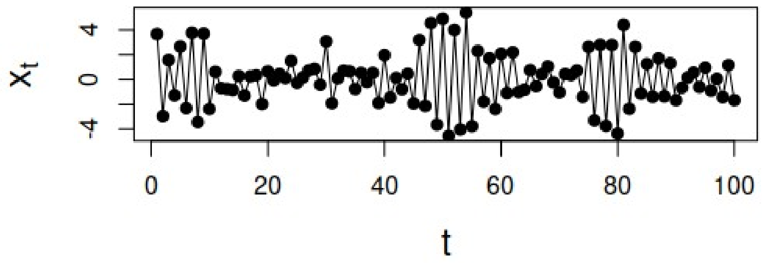
Μέθοδος Ελαχίστων Τετραγώνων (Least Squares)

Επαμεινώνδας Διαμαντόπουλος
Πέμπτη 18 Απριλίου 2024

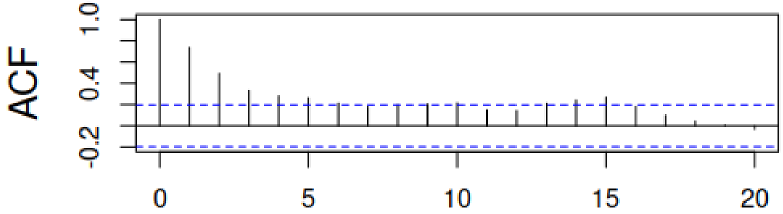
Στοχαστική Ανάλυση Χρονοσειρών (Νέο)
https://eclass.duth.gr/courses/1021383/

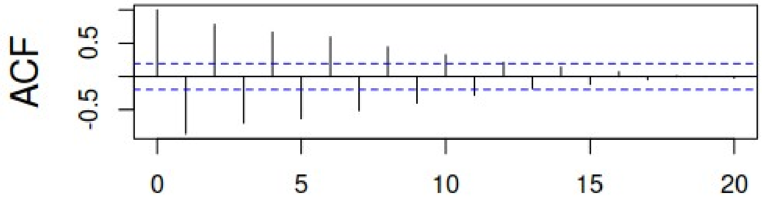


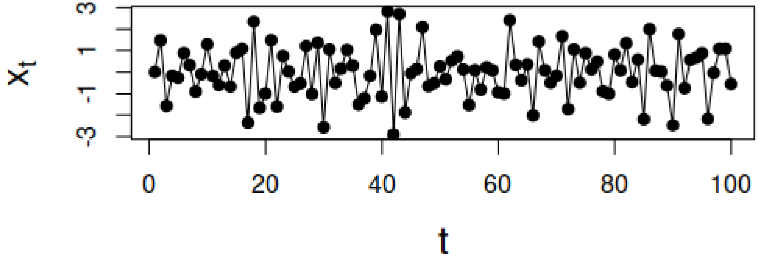
Διάγραμμα 1



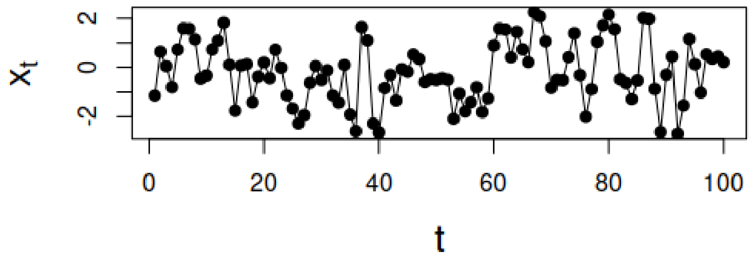
Διάγραμμα 2



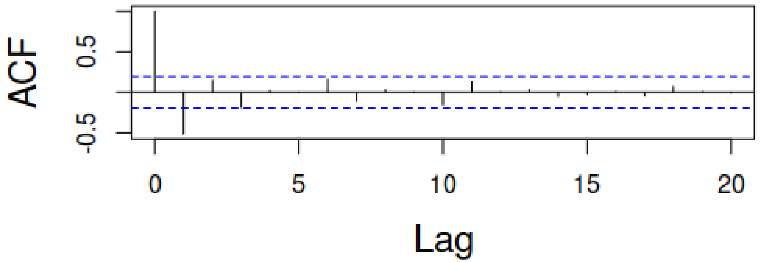


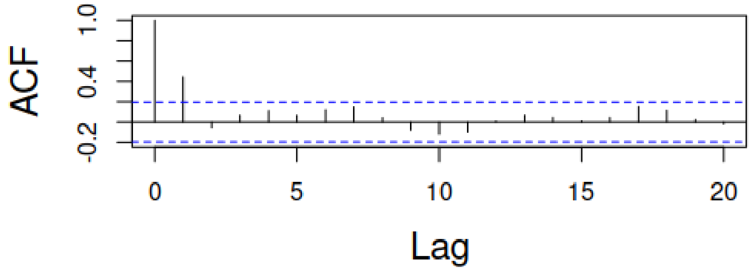


Διάγραμμα 3



Διάγραμμα 4





MA(1), θ < 0

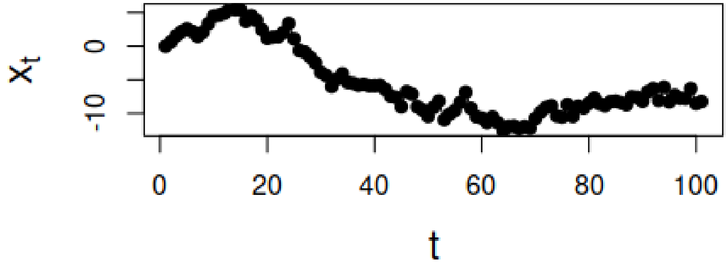
MA(1), θ > 0

AR(1), φ < 0

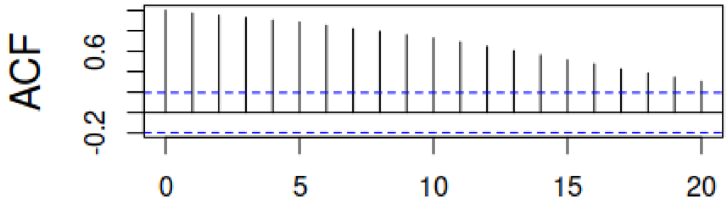
AR(1), φ > 0

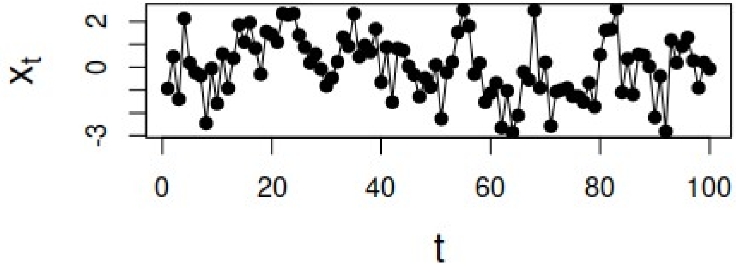
Δραστηριότητα

Αντιστοιχίστε τα 4 χρονοδιαγράμματα με τα αντίστοιχα μοντέλα.

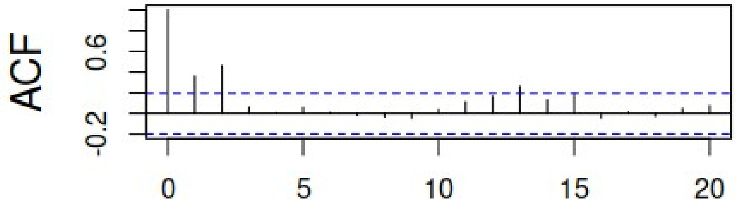


Διάγραμμα 1





Διάγραμμα 2

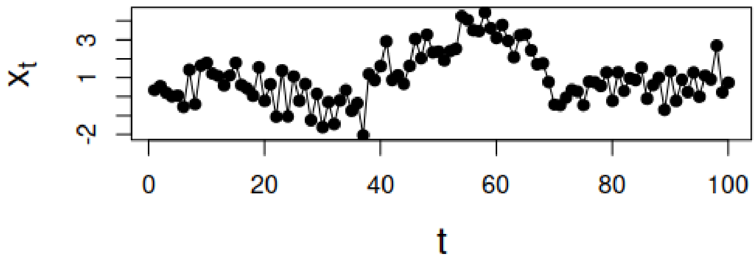


MA(2)

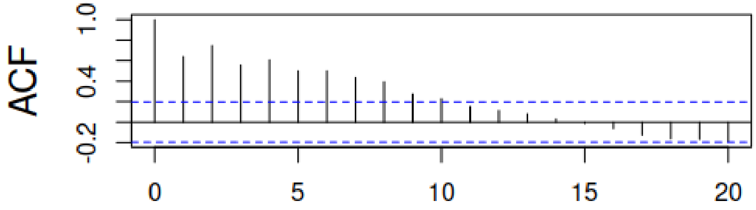
ARIMA(0,1,0)

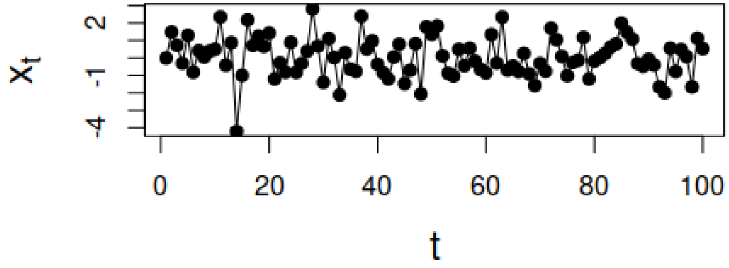
AR(2)

WN

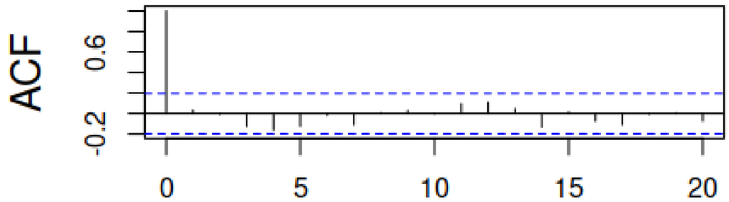


Διάγραμμα 3



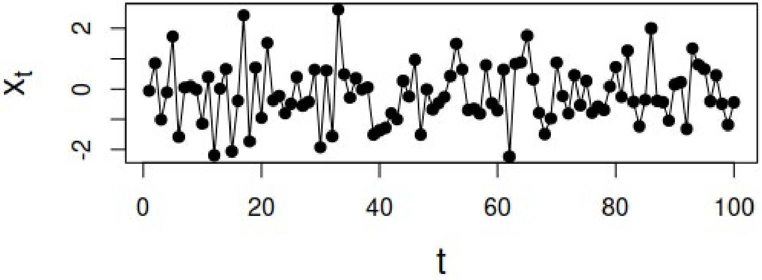


Διάγραμμα 4

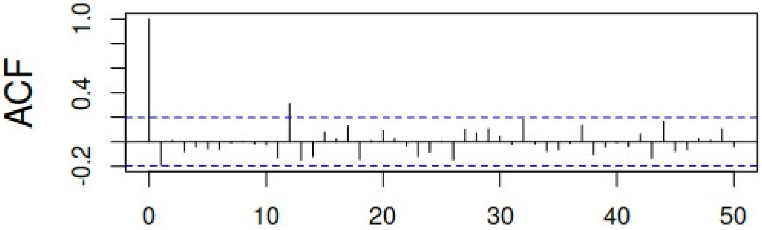


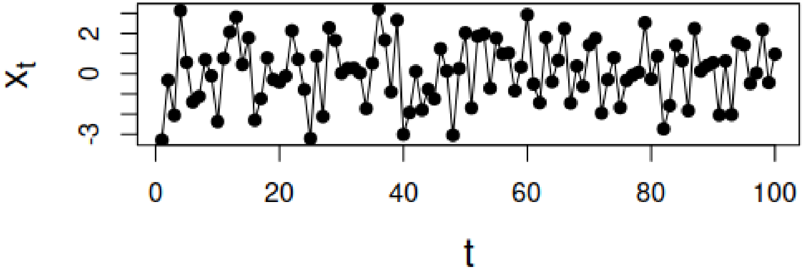
Δραστηριότητα

Αντιστοιχίστε τα 4 χρονοδιαγράμματα με τα αντίστοιχα μοντέλα.

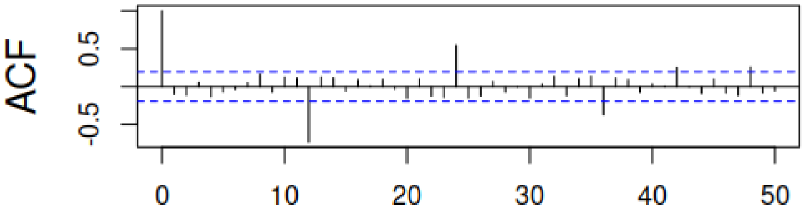


Διάγραμμα 1



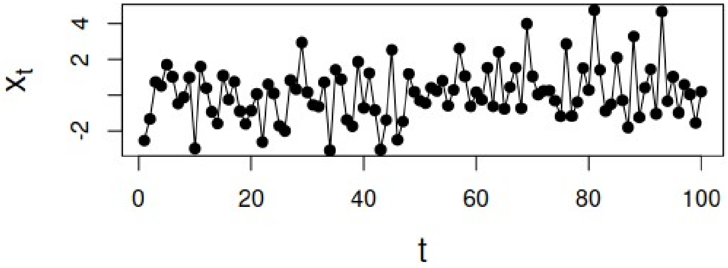


Διάγραμμα 2

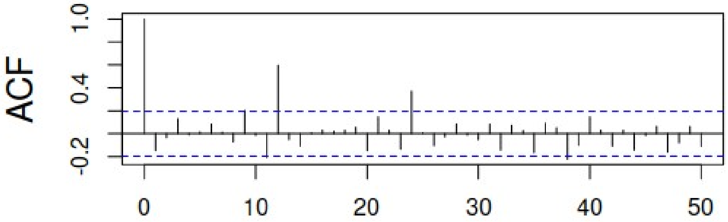


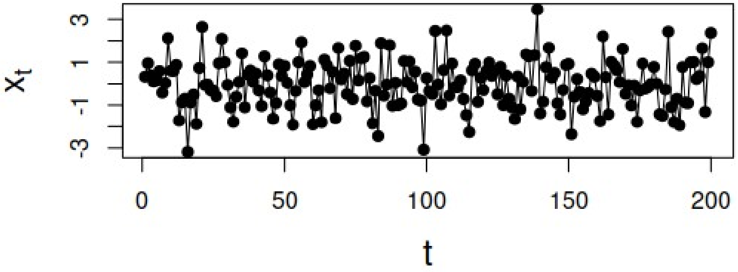
SARIMA$(0,0,0)(1,0,0)_{12}$

SARIMA$(0,0,0)(0,0,1)_{12}$

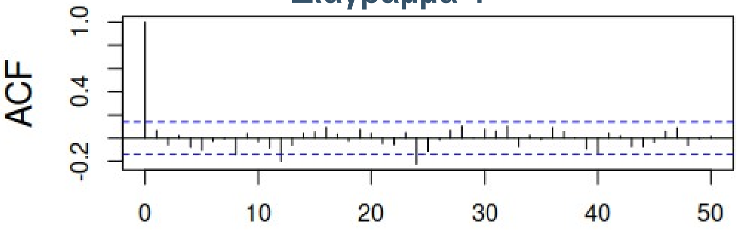


Διάγραμμα 3



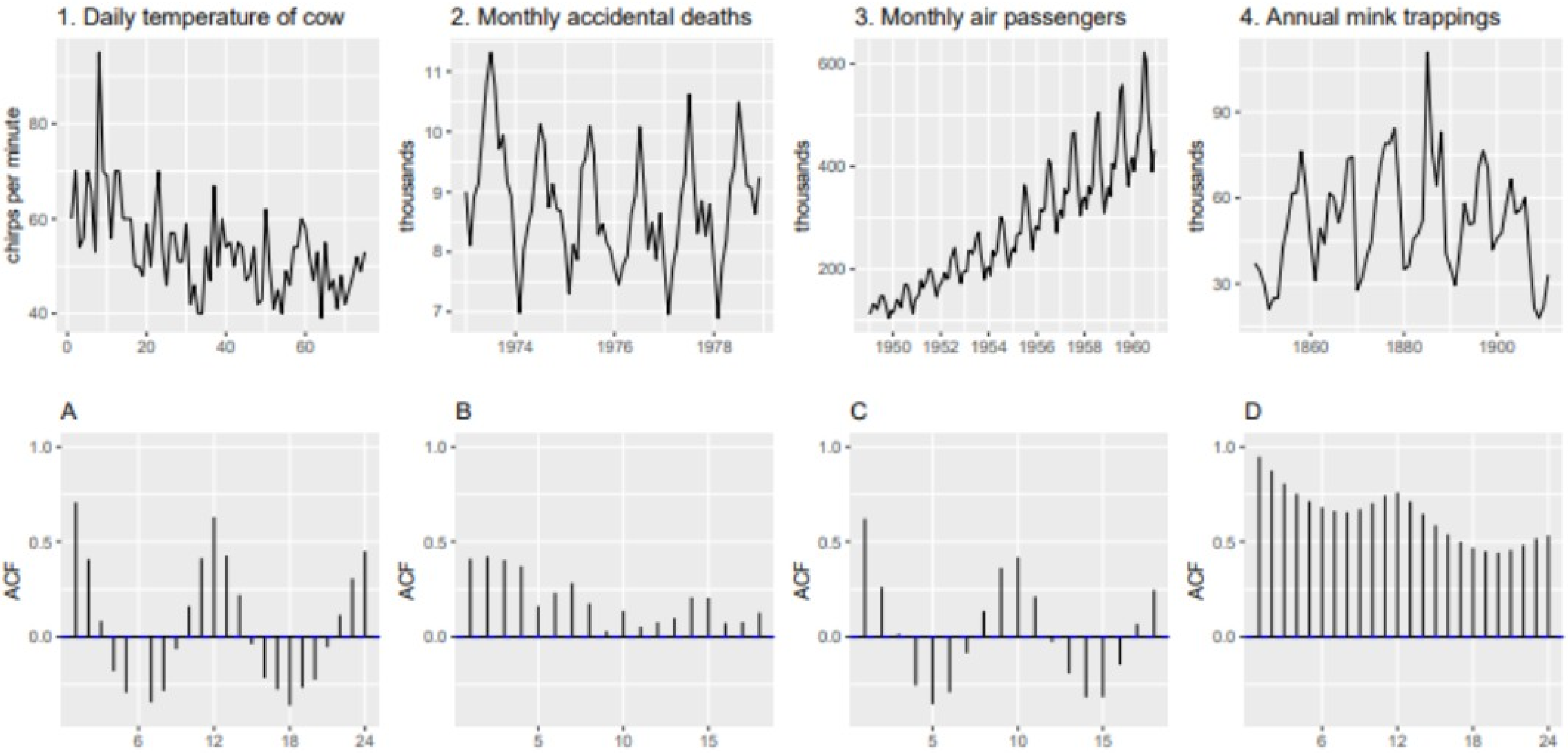


Διάγραμμα 4



Δραστηριότητα

Αντιστοιχίστε τα 4 χρονοδιαγράμματα με τα αντίστοιχα μοντέλα.


1. Daily temperature of cow
chirps per minute
2. Monthly accidental deaths
thousands
3. Monthly air passengers
thousands
4. Annual mink trappings
thousands
A
ACF
B
ACF
C
ACF
D
ACF

Προσομοίωση ARIMA με τη γλώσσα R

WN ts.sim <- arima.sim(list(order = c(0,0,0)), n = 100)

AR(2) ts.sim <- arima.sim(list(order = c(0,0,0), ar = c(0.4, 0.5)), n = 100)

AR(1) ts.sim <- arima.sim(list(order = c(0,0,0), ar = 0.8), n = 100)

MA(1) ts.sim <- arima.sim(list(order = c(0,0,0), ma = 0.7), n = 100)

MA(2) ts.sim <- arima.sim(list(order = c(0,0,0), ma = c(0.4, 0.5)), n = 100)

ARIMA(0,1,0) ts.sim <- arima.sim(list(order = c(0,1,0)), n = 100)

```
source("https://utopia.duth.gr/epdiaman/files/kedivim/MyRFunctions.R")
my.plot.ts.acf(ts.sim)
```

Προσομοίωση SARIMA με τη γλώσσα R

library(astsa)

SARIMA(0,0,0)(1,0,0)$_{12}$ ts.sim = sarima.sim(D=0, sar = 0.6, S=12, n=100)

SARIMA(0,0,0)(0,0,1)$_{12}$ ts.sim = sarima.sim(D=0, sma = 0.6, S=12, n=100)

SARIMA(0,0,0)(0,0,2)$_{12}$ ts.sim = sarima.sim(D=0, sma = c(-0.45, -0.45), S=12, n=200)

source("https://utopia.duth.gr/epdiaman/files/kedivim/MyRFunctions.R")
my.plot.ts.acf(ts.sim)

Μέθοδοι Υπολογισμού των Συντελεστών

Μέθοδοι Υπολογισμού των Συντελεστών

Για τον υπολογισμό των συντελεστών ενός μοντέλου που προσαρμόζεται σε μία χρονοσειρά, υπάρχουν πολλές μέθοδοι. Οι κυριότερες είναι οι εξής:

- Μέθοδος ελαχίστων τετραγώνων (Least Squares - LS).

- Εκτιμητής μέγιστης πιθανοφάνειας (Maximum Likehood Estimator – MLE).

Μέθοδος των ελαχίστων τετραγώνων

Υπολογισμός συντελεστών 1/3

Έστω (x_i, y_i), i = 1, 2, …, n τα n ζεύγη παρατηρήσεων και έστω

$$\hat{y} = \hat{\alpha} + \hat{b}x$$

η εξίσωση της ευθείας που αναζητούμε. Τα σφάλματα της εκτίμησης για κάθε ένα x_i, i = 1, 2, …, n, είναι:

$$\text{resid}_i(\hat{\alpha}, \hat{b}) = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{b}x_i$$

Αναζητούμε την εξίσωση ευθείας που ελαχιστοποιεί το άθροισμα τετραγώνων αυτών των σφαλμάτων.

$$\hat{\alpha}, \hat{b} = ;, \text{ ώστε } S(\hat{\alpha}, \hat{b}) = \sum_{i=1}^{n} \text{resid}_i^2(\hat{\alpha}, \hat{b}) = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{b}x_i)^2 = \text{Minimum}$$

Για τον υπολογισμό των α, b, αρκεί να βρούμε τις κρίσιμες τιμές: $\nabla S(\hat{\alpha}, \hat{b}) = 0 \Leftrightarrow \dfrac{\partial S}{\partial \hat{\alpha}} = 0$ και $\dfrac{\partial S}{\partial \hat{b}} = 0$

Είναι:

$$\frac{\partial S}{\partial \hat{\alpha}} = \sum_{i=1}^{n} -2(y_i - \hat{\alpha} - \hat{b}x_i) = -2\left[\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\alpha} - \hat{b}\sum_{i=1}^{n} x_i\right] = -2\left[n\bar{y} - n\hat{\alpha} - \hat{b}n\bar{x}\right]$$

Και: $\dfrac{\partial S}{\partial \hat{\alpha}} = 0 \Leftrightarrow \bar{y} - \hat{\alpha} - \hat{b}\bar{x} = 0$ ή $\hat{\alpha} = \bar{y} - \hat{b}\bar{x}$

Σημείωση

Από την τελευταία σχέση προκύπτει ειδικότερα ότι το σημείο $(\bar{x}, \bar{y})$ ανήκει στην ευθεία ελαχίστων τετραγώνων.

Υπολογισμός συντελεστών 2/3

Περαιτέρω, από την σχέση $\dfrac{\partial S}{\partial \hat{b}} = 0$ παίρνουμε:

$$\sum_{i=1}^{n} x_i (y_i - \hat{\alpha} - \hat{b} x_i) = 0 \Leftrightarrow \sum_{i=1}^{n} x_i (y_i - \bar{y} + \hat{b}\bar{x} - \hat{b} x_i) = 0 \Leftrightarrow \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \hat{b} \sum_{i=1}^{n} x_i (x_i - \bar{x}) = 0$$

$$\mathbf{\hat{b}} = \frac{\sum_{i=1}^{n} \mathbf{x_i} (\mathbf{y_i} - \mathbf{\bar{y}})}{\sum_{i=1}^{n} \mathbf{x_i} (\mathbf{x_i} - \mathbf{\bar{x}})}$$

Υπολογισμός συντελεστών 3/3

Εύκολα, αποδεικνύεται ότι

$$\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})=\sum_{i=1}^{n}x_i(y_i-\bar{y})-\sum_{i=1}^{n}\bar{x}\,y_i+n\,\bar{x}\,\bar{y}=\sum_{i=1}^{n}x_i(y_i-\bar{y})$$

$$\sum_{i=1}^{n}(x_i-\bar{x})\cdot(x_i-\bar{x})=\sum_{i=1}^{n}x_i(x_i-\bar{x})-\sum_{i=1}^{n}\bar{x}\,x_i+n\,\bar{x}\,\bar{x}=\sum_{i=1}^{n}x_i(x_i-\bar{x})$$

Από τις τελευταίες δύο ισότητες, συνάγεται ότι μία ισοδύναμη έκφραση των συντελεστών είναι:

$$\hat{\alpha}=\bar{y}-\hat{b}\,\bar{x}\ ,\ \ \hat{b}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{s_{XY}^2}{s_X^2}=r_{XY}\frac{s_Y}{s_X}$$

Σημαντικά αθροίσματα τετραγώνων

Αποδεικνύεται ότι: $$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

Βασικά βήματα απόδειξης:

$$(y_i-\bar{y})^2=(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2=(\hat{y}_i-\bar{y})^2+(y_i-\hat{y}_i)^2+2(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)$$

$$\hat{y}_i-\bar{y}=\hat{b}(x_i-\bar{x}),\ \text{ από }\ \bar{y}=\hat{\alpha}+\hat{b}\,\bar{x}\ \text{ και }\ \hat{y}_i=\hat{\alpha}+\hat{b}\,x_i$$

$$y_i-\hat{y}_i=(y_i-\bar{y})+(\bar{y}-\hat{y}_i)=(y_i-\bar{y})-\hat{b}(x_i-\bar{x})=(y_i-\bar{y})-\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}(x_i-\bar{x})$$

$$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\hat{b}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\hat{y}_i)=\hat{b}\sum_{i=1}^{n}(x_i-\bar{x})\left[(y_i-\bar{y})-\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}(x_i-\bar{x})\right]=0$$

Σημείωση: Η απόδειξη είναι διαθέσιμη εδώ: https://en.wikipedia.org/wiki/Explained_sum_of_squares

$$\sum_{i=1}^{n}\left(\mathbf{y_i}-\bar{\mathbf{y}}\right)^2=\sum_{i=1}^{n}\left(\hat{\mathbf{y}}_i-\bar{\mathbf{y}}\right)^2+\sum_{i=1}^{n}\left(\mathbf{y_i}-\hat{\mathbf{y}}_i\right)^2$$

Στα πλαίσια της γραμμικής παλινδρόμησης τα παραπάνω αθροίσματα τετραγώνων έχουν ιδιαίτερη σημασία και ονομασία.

α) $\text{TSS}=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2$: συνολικό άθροισμα τετραγώνων (total sum of squares)

Το συνολικό άθροισμα τετραγώνων (TSS) εκφράζει τη συνολική μεταβλητότητα της εξαρτημένης μεταβλητής.

β) $\text{ESS}=\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2$: επεξηγημένο άθροισμα τετραγώνων (explained sum of squares)

Το επεξηγημένο άθροισμα τετραγώνων (ESS) είναι το μέρος της μεταβλητότητας της εξαρτημένης μεταβλητής που εξηγείται από τις επεξηγηματικές μεταβλητές του γραμμικού μοντέλου.

γ) $\text{RSS}=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$: άθροισμα τετραγώνων υπολοίπων (residual sum of squares)

Το άθροισμα τετραγώνων υπολοίπων (RSS) εκφράζει το μέρος της μεταβλητότητας της εξαρτημένης μεταβλητής που δεν εξηγείται από τις επεξηγηματικές μεταβλητές του γραμμικού μοντέλου.

Η σχέση $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ γράφεται:

$$TSS = ESS + RSS$$

Διαιρώντας, με το συνολικό άθροισμα τετραγώνων (TSS) παίρνουμε:

$$\frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Συντελεστής Προσδιορισμού (R^2) 1/2

Το πόσο καλά προσαρμόζεται το γραμμικό μοντέλο στα δεδομένα ποσοτικοποιείται και από το συντελεστή προσδιορισμού (coeffficient of determination) ο οποίος ορίζεται ως:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Αν το μοντέλο προβλέπει χωρίς σφάλμα όλες τις παρατηρήσεις, τότε

$$y_i - \hat{y}_i = 0 \Leftrightarrow R^2 = 1 = 100\%.$$

Αν τα σφάλματα προβλέψεων του μοντέλου είναι όσο η διαφορά από την μέση τιμή (baseline model), τότε

$$y_i - \hat{y}_i = y_i - \bar{y} \Leftrightarrow R^2 = 0 = 0\%.$$

Αν τα σφάλματα προβλέψεων ξεπερνούν την απόσταση των παρατηρήσεων από τη μέση τιμή, τότε $R^2 < 0$.

Συντελεστής Προσδιορισμού (R^2) 2/2

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Ο συντελεστής R^2 συνήθως παρουσιάζεται ως ποσοστό με μεγαλύτερες τιμές να καταδεικνύουν καλύτερη προσαρμογή στα δεδομένα. Τα όρια αποδοχής της αποτελεσματικότητας του μοντέλου ποικίλουν ανάλογα με την επιστημονική περιοχή.

Στην απλή γραμμική παλινδρόμηση αποδεικνύεται ότι ο συντελεστής προσδιορισμού είναι το τετράγωνο του συντελεστή συσχέτισης Pearson μεταξύ των x_i και των y_i.

$$R^2 = r^2$$

Ειδικότερα: Στην απλή γραμμική παλινδρόμηση είναι αδύνατον για το R^2 να πάρει αρνητικές τιμές.

Σημείωση
Μία απόδειξη μπορεί να βρεθεί εδώ:
https://math.stackexchange.com/questions/129909/correlation-coefficient-and-determination-coefficient

Τυπικό Σφάλμα Συντελεστών

$$\hat{\alpha} = \bar{y} - \hat{b}\bar{x}\ ,\ \ \hat{b} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{XY}^2}{s_X^2} = r_{XY}\frac{s_Y}{s_X}$$

Αποδεικνύεται ότι το τυπικό σφάλμα των παραπάνω υπολογισμών δίνεται από τους τύπους

$$SE(\hat{\alpha}) = \sqrt{\frac{1}{n(n-2)}\sum_{i=1}^{n}x_i^2\ \sum_{i=1}^{n}(y_i - \hat{y})^2 \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{n-1}{n(n-2)}\sum_{i=1}^{n}x_i^2 \cdot \frac{RSS}{s_x^2}}$$

$$SE(\hat{b}) = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y})^2 \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{n-1}{n-2} \cdot \frac{RSS}{s_x^2}}$$

Τυπικό Σφάλμα Συντελεστών

Η γνώση του τυπικού σφάλματος του γραμμικού μοντέλου, μας επιτρέπει:

(α) Να ελέγχουμε αν απορρίπτεται ή όχι η υπόθεση H_0: b = 0 (αντ. η H_0: α = 0) υπολογίζοντας το στατιστικό

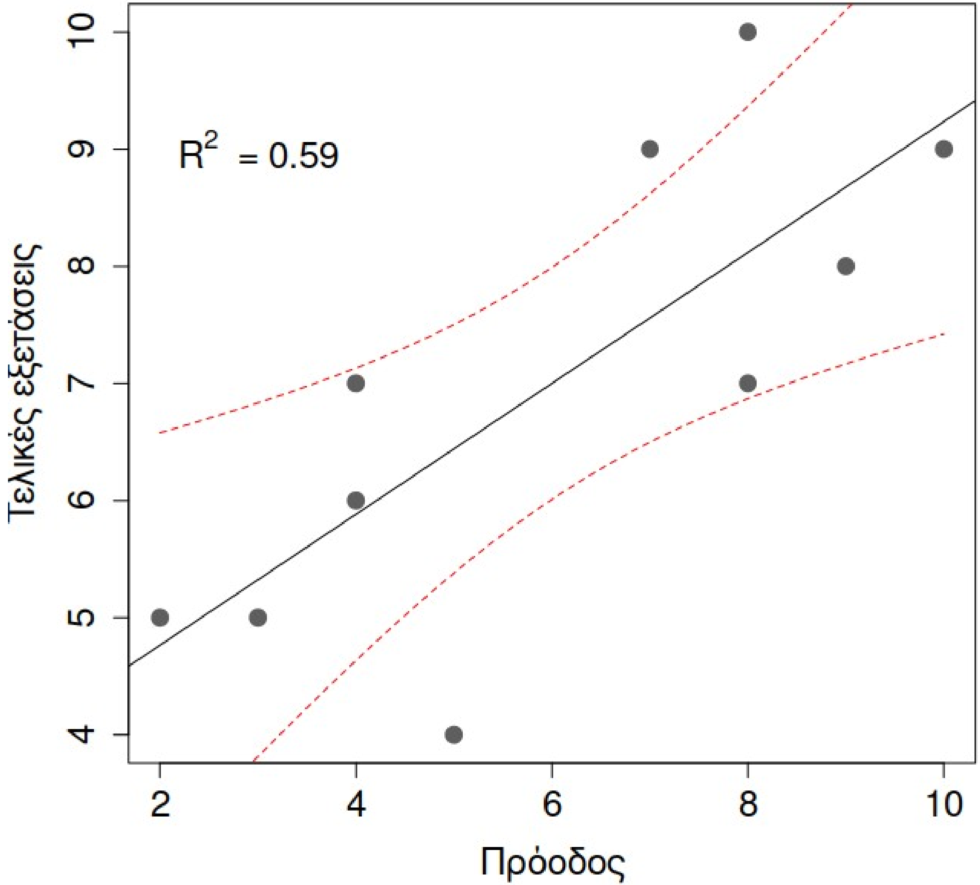
$$H_0: b=0: \ \ t = \frac{\hat{b}}{SE(\hat{b})} \sim t_{n-2}, \ \ p = P(|t| > |t_0|)$$

(β) Να υπολογίζουμε το 95% διάστημα εμπιστοσύνης για τους συντελεστές του μοντέλου

$$\left(\hat{b} - t_{n-2;0.025} SE(\hat{b}), \ \hat{b} - t_{n-2;0.025} SE(\hat{b})\right)$$

Σημείωση:
Μία απόδειξη για το γεγονός ότι οι συντελεστές ακολουθούν την κατανομή t μπορεί να βρεθεί εδώ:
https://stats.stackexchange.com/questions/117406/proof-that-the-coefficients-in-an-ols-model-follow-a-t-distribution-with-n-k-d

```
fit = lm(exams ~ test)
newx = seq(min(test), max(test), length.out=100)
preds = predict(fit, newdata = data.frame(test=newx), interval = 'confidence')
plot(exams ~ test, xlab = "Πρόοδος", pch = 19 ,col = "grey37" , ylab = "Τελικές εξετάσεις", cex = 1.5, cex.lab = 1.5, cex.axis = 1.5)
rsquare = round(summary(fit)$r.squared, 3)
text(3, 9, bquote ("R"^{2} ~ " ="~.(rsquare)), cex = 1.5)
polygon(c(rev(newx), newx), c(rev(preds[ ,3]), preds[ ,2]), border = NA)
abline(fit)
lines(newx, preds[ ,3], lty = 'dashed', col = 'red')
lines(newx, preds[ ,2], lty = 'dashed', col = 'red')
```



Υπολογισμός συντελεστών – Γενική Περίπτωση 1/3

Η γενική περίπτωση μπορεί να περιέχει πολλές επεξηγηματικές μεταβλητές.

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \varepsilon_i \quad (i = 1, \cdots, n).$$

Ορίζουμε:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{2n} & \cdots & x_{kn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Με το νέο συμβολισμό, η αρχική εξίσωση γράφεται:

$$y = X\beta + \varepsilon.$$

Με β συμβολίζεται το διάνυσμα των άγνωστων παραμέτρων και με ε το διάνυσμα των σφαλμάτων που αντιστοιχούν σε κάθε μία παρατήρηση.

Υπολογισμός συντελεστών – Γενική Περίπτωση 2/3

Αν το διάνυσμα b είναι η εκτίμηση του β, τότε μπορούμε να γράψουμε:

$$y = Xb + e.$$

Στην εξίσωση αυτή, e δηλώνει τις παρατηρήσιμες διαφορές, δηλαδή τα σφάλματα εκτίμησης του γραμμικού μοντέλου. Ειδικότερα:

$$e = y - Xb.$$

Στο σημείο αυτό ορίζουμε τη συνάρτηση:

$$\begin{aligned} S(b) &= \sum e_i^2 = e'e = (y - Xb)'(y - Xb) \\ &= y'y - y'Xb - b'X'y + b'X'Xb. \end{aligned}$$

Υπολογισμός συντελεστών – Γενική Περίπτωση 3/3

Με λίγη άλγεβρα πινάκων, βρίσκουμε ότι:

$$\frac{\partial S}{\partial b} = -2X'y + 2X'Xb.$$

Η ελάχιστη τιμή ως προς b, λαμβάνεται όταν η παράγωγος είναι ίση με 0. Συμπεραίνουμε, ότι, ο εκτιμητής ελαχίστων τετραγώνων ικανοποιεί την εξίσωση:

$$X'Xb = X'y.$$

Λύνοντας ως προς b βρίσκουμε:

$$b = (X'X)^{-1}X'y$$

Η μόνη προϋπόθεση υλοποίησης της παραπάνω διαδικασίας είναι ο πίνακας X’X να είναι αντιστρέψιμος ή ισοδύναμα το πλήθος παρατηρήσεων να είναι μεγαλύτερο ή ίσο από το πλήθος των αγνώστων συντελεστών του μοντέλου.

Εφαρμογή ελαχίστων τετραγώνων σε μοντέλα της μορφής X_t = M(t, θ)

Η μέθοδος αυτή είναι δυνατόν να εφαρμοστεί σε μοντέλα της μορφής:

$$X_t = M(t, \boldsymbol{\theta}) + \varepsilon_t$$

όπου:

- M(t, θ) είναι μία γνωστή συνάρτηση που εξαρτάται από το t,
- **θ** είναι το διάνυσμα των αγνώστων παραμέτρων και
- ε_t iid ~ N(0, σ^2), το τυχαίο σφάλμα της πρόβλεψης.

Ελαχιστοποιώντας το άθροισμα τετραγώνων των ε_t:

$$\sum_{t=1}^{n}[X_t - M(t, \boldsymbol{\theta})]^2$$

είναι δυνατόν να υπολογίσουμε εκτιμητές ελαχίστων τετραγώνων για τις παραμέτρους του διανύσματος **θ**.

Μέθοδος των ελαχίστων τετραγώνων στις χρονοσειρές

Παράδειγμα 1

Στον πίνακα παρουσιάζεται η παραγωγή ζαχαρότευτλου (σε τόνους) σε μία περιοχή ανά έτος. Να γίνει γραμμικό μοντέλο πρόβλεψης της εξέλιξης της παραγωγής στο χρόνο.

Έτος t	2010	2011	2012	2013	2014	2015	2016
Παραγωγή	40	45	46	42	47	50	46

Τυπολόγιο: $\hat{y} = \hat{\alpha} + \hat{b}\,t$ $\qquad \hat{b} = \dfrac{\sum_{i=1}^{\nu}(t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^{\nu}(t_i - \bar{t})^2} \qquad \hat{\alpha} = \bar{y} - \hat{b}\,\bar{t}$

Μέθοδος των ελαχίστων τετραγώνων στις χρονοσειρές

Παράδειγμα 1

Στον πίνακα παρουσιάζεται η παραγωγή ζαχαρότευτλου (σε τόνους) σε μία περιοχή ανά έτος. Να γίνει γραμμικό μοντέλο πρόβλεψης της εξέλιξης της παραγωγής στο χρόνο.

Έτος t	2010	2011	2012	2013	2014	2015	2016
Παραγωγή	40	45	46	42	47	50	46

Τυπολόγιο: $\hat{y} = \hat{\alpha} + \hat{b}\,t$ $\quad \hat{b} = \dfrac{\sum_{i=1}^{n}(t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^{n}(t_i - \bar{t})^2}$ $\quad \hat{\alpha} = \bar{y} - \hat{b}\,\bar{t}$

Αν οι υπολογισμοί γίνονται με το χέρι, τότε είναι πρακτικό το έτος t να αντικατασταθεί με [Έτος – Μέση τιμή] = {-3, -2, -1, 0, 1, 2, 3}, καθώς τότε η μέση τιμή του t είναι 0 και οι τύποι απλοποιούνται:

$$\hat{b} = \sum_{i=1}^{n} t_i y_i \Big/ \sum_{i=1}^{n} t_i^2 \qquad \hat{\alpha} = \bar{y}$$

Μέθοδος των ελαχίστων τετραγώνων στις χρονοσειρές

Παράδειγμα 1

Έτος	Παραγωγή (y)	t = Έτος – 2013	ty	t^2
2010	40	-3	-120	9
2011	45	-2	-90	4
2012	46	-1	-46	1
2013	42	0	0	0
2014	47	1	47	1
2015	50	2	100	4
2016	46	3	138	9
Άθροισμα	316	0	29	28

$$\hat{b} = \sum_{i=1}^{n} t_i y_i \Big/ \sum_{i=1}^{n} t_i^2 = 29/28 = 1{,}036$$

$$\hat{\alpha} = \bar{y} = 316/7 = 45{,}143$$

Μέθοδος των ελαχίστων τετραγώνων

Παράδειγμα 1

$$\hat{b} = \sum_{i=1}^{n} t_i y_i \Big/ \sum_{i=1}^{n} t_i^2 = 29/28 = 1{,}036 \qquad \hat{\alpha} = \bar{y} = 316/7 = 45{,}143$$

Εξίσωση πρόβλεψης:

Παραγωγή = 1,036 · (Έτος – 2013) + 45,143.

Συμπεραίνουμε ότι:

Για κάθε έτος που περνάει αναμένουμε αυξημένη παραγωγή κατά 1,036 τόνους.

Μέθοδος των ελαχίστων τετραγώνων

Παράδειγμα 1

Υλοποίηση στο Excel / Calc

Συναρτήσεις SLOPE(y, x) και INTERCEPT(y, x) (ακολουθεί επίδειξη)

Υλοποίηση στη γλώσσα R

```
years = c(-3, -2, -1, 0, 1, 2, 3)
production = c(40, 45, 46, 42, 47, 50, 46)
lm(production ~ years)
```

(ακολουθεί επίδειξη)

Εφαρμογή ελαχίστων τετραγώνων σε AR(p) μοντέλα

Η μέθοδος αυτή είναι δυνατόν να εφαρμοστεί σε μοντέλα AR(p): $x_n = \sum_{k=1}^{p} a_k x_{n-k} + \epsilon_n$
όπου $\varepsilon_n \sim N(0, \sigma^2)$.
Αν $x_1, x_2, \ldots, x_N$, είναι οι τιμές της χρονοσειράς που έχουμε παρατηρήσει, τότε μπορούμε να βρούμε τους συντελεστές, λύνοντας το σύστημα των M = N + 1 – p εξισώσεων:

$$\begin{bmatrix} x_0 & x_1 & \cdots & x_{p-1} \\ x_1 & x_2 & \cdots & x_p \\ x_2 & x_3 & \cdots & x_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-p} & x_{N-p+1} & \cdots & x_{N-1} \end{bmatrix} \begin{bmatrix} a_p \\ a_{p-1} \\ \vdots \\ a_1 \end{bmatrix} = \begin{bmatrix} x_p \\ x_{p+1} \\ x_{p+2} \\ \vdots \\ x_N \end{bmatrix} \Rightarrow \begin{bmatrix} \mathbf{x}_{N-p,M} & \mathbf{x}_{N-p+1,M} & \cdots & \mathbf{x}_{N-1,M} \end{bmatrix} \mathbf{a} = \mathbf{x}_{N,M}$$

$$\Downarrow$$

$$\mathbf{X}_{N,M}\mathbf{a} = \mathbf{x}_{N,M}$$

Ο εκτιμητής ελαχίστων τετραγώνων είναι:

$$\hat{\mathbf{a}} = \left(\mathbf{X}_{N,M}^T \mathbf{X}_{N,M}\right)^{-1} \mathbf{X}_{N,M}^T \mathbf{x}_{N,M}$$

Εφαρμογή ελαχίστων τετραγώνων σε AR(p) μοντέλα

Παράδειγμα AR(1) $X_t = \phi\, X_{t-1} + \mu + Z_t$

Αρκεί να ελαχιστοποιήσουμε το άθροισμα τετραγωνικών σφαλμάτων:

$$E = \frac{1}{2\,n} \sum_{t=2}^{n} (X_t - \phi\, X_{t-1} - \mu)^2$$

Υπολογίζουμε τις μερικές παραγώγους:

$$\frac{\partial E}{\partial \mu} = \frac{1}{n} \sum_{t=2}^{n} (\phi\, X_{t-1} + \mu - X_t) = \phi\, \langle X_{t-1} \rangle + \mu - \langle X_t \rangle$$

$$\frac{\partial E}{\partial \phi} = \frac{1}{n} \sum_{t=2}^{n} (\phi\, X_{t-1} + \mu - X_t)\, X_{t-1} = \phi\, \langle X_{t-1}^2 \rangle + \mu \langle X_{t-1} \rangle - \langle X_t\, X_{t-1} \rangle$$

Εφαρμογή ελαχίστων τετραγώνων σε AR(p) μοντέλα

Παράδειγμα AR(1)

Για να βρούμε τα ακρότατα, αρκεί να εξισώσουμε τις δύο εξισώσεις με το 0:

$$\frac{\partial E}{\partial \mu} = \frac{1}{n}\sum_{t=2}^{n}(\phi\, X_{t-1} + \mu - X_t) = \phi\,\langle X_{t-1}\rangle + \mu - \langle X_t\rangle = 0$$

$$\frac{\partial E}{\partial \phi} = \frac{1}{n}\sum_{t=2}^{n}(\phi\, X_{t-1} + \mu - X_t)\, X_{t-1} = \phi\,\langle X_{t-1}^2\rangle + \mu\langle X_{t-1}\rangle - \langle X_t\, X_{t-1}\rangle = 0$$

Παίρνουμε το σύστημα: $$\begin{pmatrix} 1 & \langle X_{t-1}\rangle \\ \langle X_{t-1}\rangle & \langle X_{t-1}^2\rangle \end{pmatrix}\begin{pmatrix} \mu \\ \phi \end{pmatrix} = \begin{pmatrix} \langle X_t\rangle \\ \langle X_{t-1}\, X_t\rangle \end{pmatrix}.$$

Η λύση του είναι: $$\begin{pmatrix} \mu \\ \phi \end{pmatrix} = \frac{1}{var(X_{t-1})}\begin{pmatrix} cov(X_t X_{t-1}, X_{t-1}) - cov(X_t, X_{t-1}^2) \\ cov(X_{t-1},\ X_t) \end{pmatrix}.$$

Η συνέχεια στο…

https://utopia.duth.gr/epdiaman/files/dpth/LinearRegressionForTimeSeries.html