



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Στοχαστική Ανάλυση Χρονοσειρών

Περιεχόμενο:

Διάστημα εμπιστοσύνης στο διάγραμμα ACF

Μοντέλα $ARIMA(p, d, q)$ και $SARIMA(p, d, q)(P, D, Q)_m$

Στατιστικοί Ελεγχοι Υποθέσεων



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Διάστημα Εμπιστοσύνης στο διάγραμμα ACF



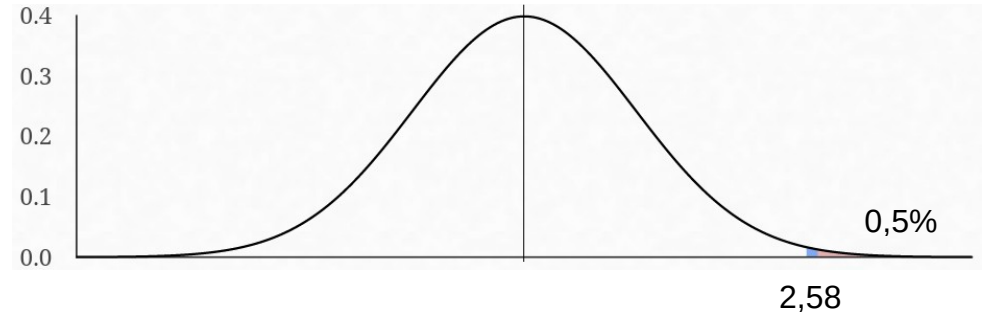
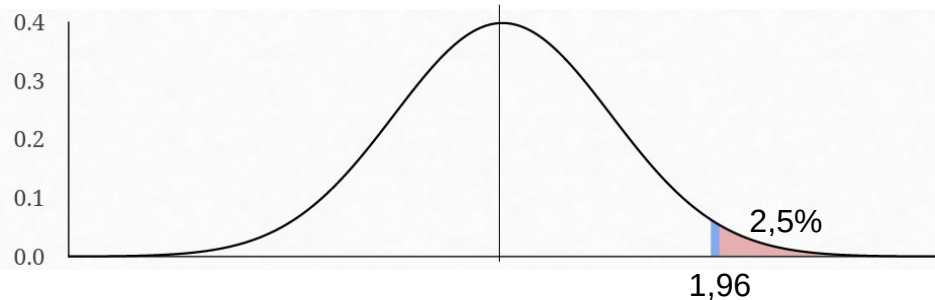
Διάστημα Εμπιστοσύνης στο διάγραμμα ACF

Εστω $Z \sim N(0, 1)$, $0 < \alpha < 1$, και z_α ο αριθμός για τον οποίο
 $P(Z > z_\alpha) = \alpha$ ή $\Phi(z_\alpha) = 1 - \alpha$.

Το z_α ονομάζεται α – ποσοστιαίο σημείο της κανονικής κατανομής. Ενδεικτικά:

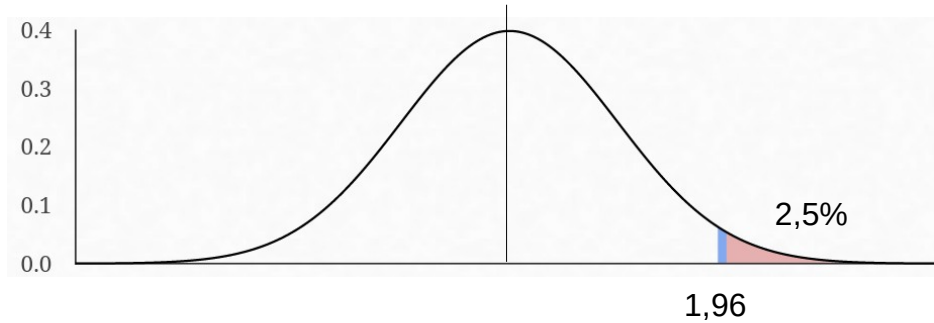
Αν $\alpha = 0,025$, τότε $z_{0,025}$ είναι ο αριθμός για τον οποίον $\Phi(z_{0,025}) = 0,975$ ή $z_{0,025} = 1,96$.

Αν $\alpha = 0,005$, τότε $z_{0,005}$ είναι ο αριθμός για τον οποίον $\Phi(z_{0,005}) = 0,995$ ή $z_{0,005} = 2,58$.

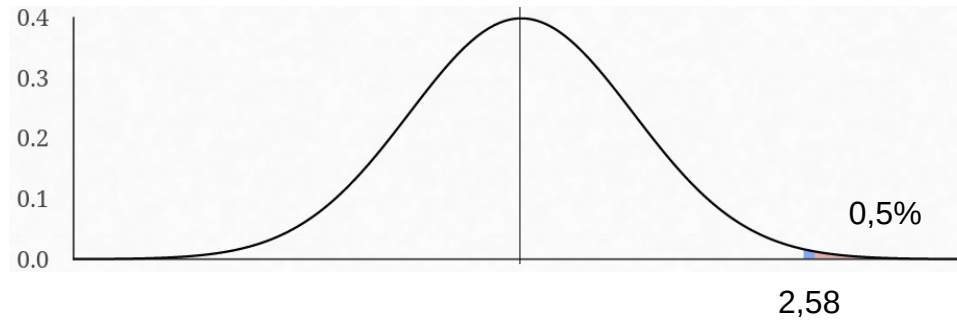




Διάστημα Εμπιστοσύνης στο διάγραμμα ACF



$$Z \sim N(0, 1) \Leftrightarrow P(-1.96 < Z < 1.96) = 0.95.$$



$$Z \sim N(0, 1) \Leftrightarrow P(-2.58 < Z < 2.58) = 0.99.$$



Διάστημα Εμπιστοσύνης στο διάγραμμα ACF

Έστω $X_t \sim \text{iid } N(0, \sigma^2)$ (White Noise). Αν r_k η αυτοσυσχέτιση υστέρησης k , αποδεικνύεται ότι:

- Ο r_k ακολουθεί κανονική κατανομή.

- $$\text{Var}(r_k) = \frac{n - k}{n(n + 2)}$$

Οριακά, καθώς $k \rightarrow \infty$: $r_k \sim N(0, 1/n) \Leftrightarrow \sqrt{n}r_k \sim N(0, 1)$.

$$P(-1.96 < \sqrt{n}r_k < 1.96) = 0.95 \Leftrightarrow P\left(\frac{-1.96}{\sqrt{n}} < r_k < \frac{1.96}{\sqrt{n}}\right) = 0.95.$$

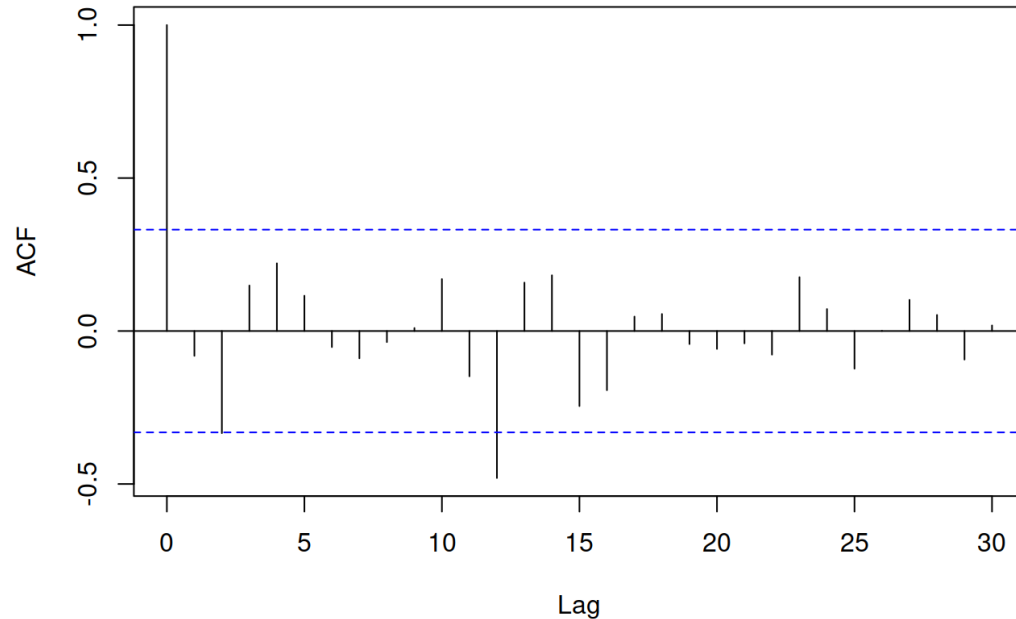
Σημείωση: Μία απόδειξη είναι διαθέσιμη εδώ: <https://stats.stackexchange.com/a/611874/27608>. Πρωτότυπο άρθρο:

Box, G. E. P., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, 65(332), 1509–1526.

<https://doi.org/10.1080/01621459.1970.10481180>



Διάστημα Εμπιστοσύνης στο διάγραμμα ACF



Συμπεραίνουμε ότι, το 95% διάστημα εμπιστοσύνης υπολογίζεται από τον τύπο

$$95 \% \text{ C.I.} = \left(-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}} \right)$$

Όταν μία αυτοσυσχέτιση είναι εκτός του 95% δ.ε., υπάρχει ένδειξη πως η χρονοσειρά δεν είναι i.i.d. $N(0, \sigma^2)$.

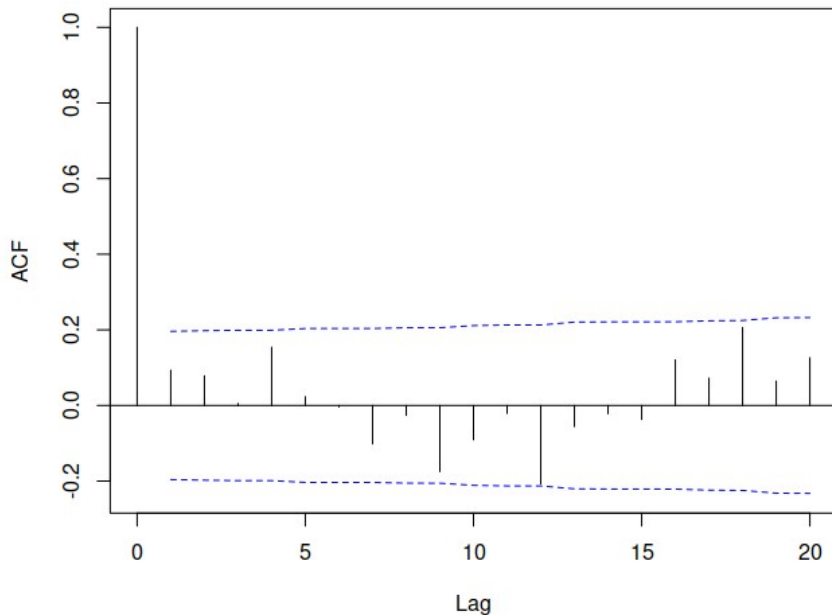


Διάστημα Εμπιστοσύνης στο διάγραμμα ACF

Αν αντί της υπόθεσης $X_t \sim \text{iid } N(0, \sigma^2)$, ξεκινήσουμε από την υπόθεση $X_t \sim \text{MA}(q)$ για κάποιο $q \in \mathbb{N}$, τότε αποδεικνύεται πως το αντίστοιχο 95% διάστημα εμπιστοσύνης για την αυτοσυσχέτιση r_k , $k = 1, 2, \dots$, έχει όρια:

$$\pm 1.96 \sqrt{\frac{1}{n} \left(1 + 2 \sum_{i=1}^k x_i^2 \right)}.$$

Στην περίπτωση αυτή, το εύρος αυξάνει με το k .





Ακόμα ένα παράδειγμα: Διάστημα Εμπιστοσύνης για μέση τιμή

Αν $n > 30$ και οι τ.μ. X_i , $i = 1, 2, \dots, n$, είναι ισόνομες και ανεξάρτητες, το Κεντρικό Οριακό Θεώρημα μας εξασφαλίζει ότι: $X_1 + X_2 + \dots + X_n \sim N(n \cdot \mu, n \cdot \sigma^2)$, ή ισοδύναμα ότι:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N(\mu, SE^2) \Leftrightarrow \frac{\bar{X} - \mu}{SE} \sim N(0, 1).$$

Από την τελευταία σχέση παίρνουμε:

$$P\left(-1,96 < \frac{\bar{X} - \mu}{SE} < 1,96\right) = 0,95 \Leftrightarrow P(\bar{X} - SE \cdot 1,96 < \mu < \bar{X} + SE \cdot 1,96) = 0,95.$$

Σημείωση:

Στην πράξη η τυπική απόκλιση σ του πληθυσμού δεν είναι γνωστή και προσεγγίζεται από την δειγματική τυπική απόκλιση s . Ωστόσο, στην περίπτωση αυτή, το σφάλμα μέτρησης μειώνει την ακρίβεια των υπολογισμών και αυτό πρέπει να ανακλαστεί στον τρόπο υπολογισμού του διαστήματος. Για το λόγο αυτό χρησιμοποιείται η κατανομή του Student $t(n - 1)$ στη θέση της κανονικής. Πιο συγκεκριμένα, στην περίπτωση αυτή, το α - διάστημα εμπιστοσύνης υπολογίζεται ως:

$$(\bar{X} - SE \cdot t_{n-1, \alpha/2}, \bar{X} + SE \cdot t_{n-1, \alpha/2})$$



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Στατιστικοί Ελεγχοι Υποθέσεων



Στατιστικοί Ελεγχοι Υποθέσεων

Στην Επαγωγική Στατιστική προσπαθούμε να ελέγξουμε μία υπόθεση που αφορά τον πληθυσμό από ένα δείγμα. Αυτό, το πετυχαίνουμε σε 4 βήματα:

- (α) Προσδιορίζουμε μία στατιστική υπόθεση H_0 που εκφράζει ένα γεγονός για κάποιο στατιστικό μέγεθος (π.χ. X, Y ανεξάρτητες, $\mu = 10$, $\sigma_1 = \sigma_2$, $X_t \sim \text{iid } N(0, 1)$, κλπ)
- (β) Υπολογίζουμε ένα στατιστικό Σ_Δ στο οποίο αντανακλάται η διαφοροποίηση του δείγματος από την υπόθεση H_0 και για το οποίο γνωρίζουμε την θεωρητική κατανομή των τιμών του, δεδομένης της H_0 . (π.χ. χ^2 , $t = (x - 10)/SE$, $F = \sigma_1^2 / \sigma_2^2$, Q_m , κλπ)
- (γ) Υπολογίζουμε την πιθανότητα p το στατιστικό να πάρει περισσότερο ακραία τιμή από την τιμή που υπολογίστηκε. Συνήθως το όριο απόρριψης είναι το 0,05.
- (δ) Αν $p < 0,05$ τότε λέμε ότι η H_0 απορρίπτεται και αποδεχόμαστε τη συμπληρωματική της πρόταση (H_1).
- (ε) Ακολουθεί η καταγραφή του αποτελέσματος.



Στατιστικοί Ελεγχοι Υποθέσεων

Ενδεικτικά παραδείγματα

Δοκιμασία ομοιογένειας χ^2

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k - 1).$$

Δοκιμασία one sample t – test

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(n - 1)$$

Δοκιμασία Box-Pierce⁽¹⁾ $Q(m) = n(n + 2) \sum_{j=1}^m \frac{r_j^2}{n - j} \sim \chi^2(m)$ ή $\chi^2(m - p - q - 1)$

Δοκιμασία ανεξαρτησίας χ^2

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2(k - 1)(m - 1).$$

Δοκιμασία independent samples t – test

$$\frac{\bar{X}_1 - \bar{X}_2}{SE_p} \sim t(n_1 + n_2 - 2)$$

(1) Σχετική αναφορά: <https://online.stat.psu.edu/stat510/lesson/3/3.2>



Στατιστικοί Ελεγχοι Υποθέσεων

Καταγραφή μίας στατιστικής δοκιμασίας

Αναφέρουμε την τιμή του στατιστικού, τους βαθμούς ελευθερίας και την τιμή της στατιστικής σημαντικότητας p .

Ακολουθεί η ερμηνεία του αποτελέσματος.



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Σύνοψη Προηγούμενου Μαθήματος



Σύνοψη Προηγούμενου Μαθήματος

AR(1): $X_t = \delta + \varphi_1 \cdot X_{t-1} + w_t$.

- $\rho_h = \varphi_1^h, h \geq 0$.

AR(2): $X_t = \delta + \varphi_1 \cdot X_{t-1} + \varphi_2 \cdot X_{t-2} + w_t$.

- $\rho_1 = \varphi_1 / (1 - \varphi_2)$.
- $\rho_2 = \varphi_1^2 / (1 - \varphi_2) + \varphi_2$.
- $\rho_3 = (\varphi_1^3 + \varphi_1 \cdot \varphi_2) / (1 - \varphi_2) + \varphi_1 \cdot \varphi_2$.

ARMA(1, 1): $X_t = \varphi X_{t-1} + w_t - \theta w_{t-1}$.

- $\rho_1 = (\varphi + \theta)(1 + \varphi\theta) / (1 + 2\varphi\theta + \theta^2)$.
- $\rho_h = \varphi^{h-1} \rho_1, h \geq 2$.

MA(1): $X_t = \mu + w_t + \theta_1 w_{t-1}$.

- $\rho_1 = \theta / (1 + \theta^2)$ και $\rho_h = 0, h \geq 2$.

MA(2): $X_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}$.

- $\rho_1 = (\theta_1 + \theta_1 \cdot \theta_2) / (1 + \theta_1^2 + \theta_2^2)$,
- $\rho_2 = \theta_2 / (1 + \theta_1^2 + \theta_2^2)$,
- $\rho_h = 0, h \geq 3$.

Σημείωση: w_t i.i.d. και $w_t \sim N(0, \sigma^2)$

Γενική εξίσωση

AR(p): $\Phi(B)X_t = \delta + w_t$: Αν όλες οι ρίζες του Φ βρίσκονται έξω από τον μοναδιαίο δίσκο, τότε η AR(p) είναι στάσιμη.

MA(q): $X_t = \mu + \Theta(B)w_t$: Αν όλες οι ρίζες του Θ , βρίσκονται έξω από το μοναδιαίο δίσκο τότε η MA(q) είναι αντιστρέψιμο.

ARMA(p, q): $\Phi(B)X_t = \delta + \Theta(B)w_t$

ARIMA(p, d, q): $\Phi(B)\Delta^d X_t = \delta + \Theta(B)w_t, \Delta = I - B$.



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

ARIMA με περιοδικότητα (ή SARIMA)



Μοντέλα SARIMA(p, d, q)(P, D, Q)_m

Στην περίπτωση όπου υπάρχει περιοδικότητα στη χρονοσειρά τότε αυτή εξ' ορισμού δεν είναι στάσιμη καθώς η μέση τιμή δεν είναι διαχρονικά σταθερή (λαμβάνει διαφορετική τιμή σε διαφορετικά χρονικά διαστήματα).

Η αναγνώριση της περιοδικότητας μπορεί να γίνει:

(α) Από το χρονοδιάγραμμα των τιμών με απλή παρατήρηση.

(β) Από το διάγραμμα αυτοσυσχετίσεων ACF: Θετική αυτοσυσχέτιση σε σταθερό πλήθος διαφορών δείχνει περιοδικότητα αντίστοιχης περιόδου.

Στην περίπτωση όπου υπάρχει περιοδικότητα στη χρονοσειρά τότε αυτή εξ' ορισμού δεν είναι στάσιμη καθώς η μέση τιμή δεν είναι διαχρονικά σταθερή.

Η πρώτη ενέργεια στην περίπτωση αυτή είναι ο εντοπισμός της εξίσωσης που θα εξαλείψει από τα δεδομένα την περιοδικότητα, ώστε στη συνέχεια να γίνει εφαρμογή ενός ARIMA μοντέλου στην χρονοσειρά που θα απομείνει ως υπόλοιπο μετά την εφαρμογή της εποχικής εξίσωσης.



Μοντέλα SARIMA(p, d, q)(P, D, Q)_m

Ένας τρόπος για την εξάλειψη της περιοδικότητας από τα δεδομένα, είναι η προσαρμογή ενός νέου ARIMA μοντέλου για την εξήγηση της εποχικής συνιστώσας της χρονοσειράς.

Όταν λαμβάνουμε εποχιακές διαφορές σε μια χρονοσειρά, ουσιαστικά αφαιρούμε την τιμή ενός σημείου δεδομένων από το αντίστοιχο σημείο δεδομένων της προηγούμενης περιόδου. Με τον τρόπο αυτόν, αφαιρούμε την εποχικότητα, αναδεικνύοντας τη γενικότερη τάση της σειράς.

Για την περιγραφή της περιοδικότητας μπορεί να χρησιμοποιηθούν αυτοπαλινδρομικοί όροι (πλήθος P), διαφορές (πλήθος D) και κινούμενοι μέσοι όροι (πλήθος Q). Ένα τέτοιο μοντέλο το συμβολίζουμε

$$\text{SARIMA}(p, d, q)(P, D, Q)_m,$$

όπου m η περίοδος της χρονοσειράς.



Μοντέλα SARIMA(p, d, q)(P, D, Q)_m

Εξίσωση SARIMA(p, d, q)(P, D, Q)_m.

Συμβολίζουμε με

- $\varphi(\omega) = 1 - \varphi_1\omega - \varphi_2\omega^2 - \dots - \varphi_p\omega^p$, το πολυώνυμο της AR(p).
- $\Phi(\omega) = 1 - \Phi_1\omega - \Phi_2\omega^2 - \dots - \Phi_P\omega^P$, το πολυώνυμο της Seasonal AR(P).
- $\theta(\omega) = 1 - \theta_1\omega - \theta_2\omega^2 - \dots - \theta_q\omega^q$, το πολυώνυμο της MA(q).
- $\Theta(\omega) = 1 - \Theta_1\omega - \Theta_2\omega^2 - \dots - \Theta_Q\omega^Q$, το πολυώνυμο της Seasonal MA(Q).

Τότε:

$$\text{Εξίσωση ARMA}(p, q)(P, Q)_m: \quad \varphi(B)\Phi(B^m)X_t = \theta(B)\Theta(B^m)w_t.$$

$$\text{Εξίσωση ARIMA}(p, d, q)(P, D, Q)_m: \quad \varphi(B)\Phi(B^m)\Delta^d \Delta_m^D X_t = \theta(B)\Theta(B^m)w_t.$$

$$\Delta X_t = X_t - X_{t-1}, \quad \Delta^d X_t = \Delta \cdot \Delta \cdot \Delta \cdot \dots \cdot \Delta X_t, \quad \Delta_m X_t = X_t - X_{t-m}, \quad \Delta_m^D X_t = \Delta_m \cdot \Delta_m \cdot \Delta_m \cdot \dots \cdot \Delta_m X_t.$$



SARIMA(0, 0, 0)(0, 1, 0)₁₂

Παράδειγμα: Ένα SARIMA(0, 0, 0)(0, 1, 0)₁₂ μοντέλο, χωρίς σταθερά, έχει εξίσωση της μορφής

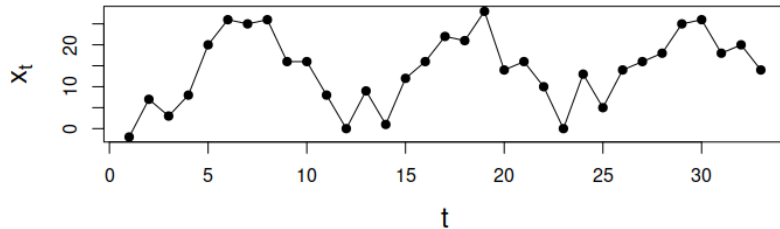
$$\Delta^{12}X_t = w_t \text{ ή } (1 - B^{12})X_t = w_t \text{ ή } X_t - X_{t-12} = w_t.$$

temp = c(-2, 7, 3, 8, 20, 26, 25, 26, 16, 16, 8, 0, 9, 1, 12, 16, 22, 21, 28, 14, 16, 10, 0, 13, 5, 14, 16, 18, 25, 26, 18, 20, 14)

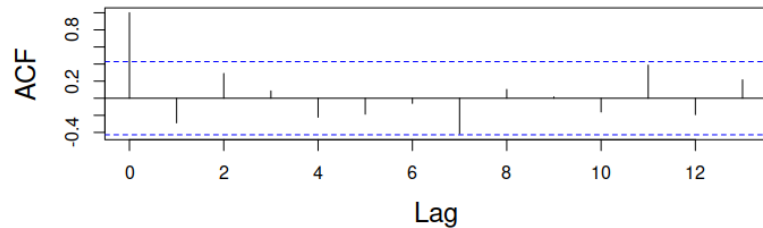
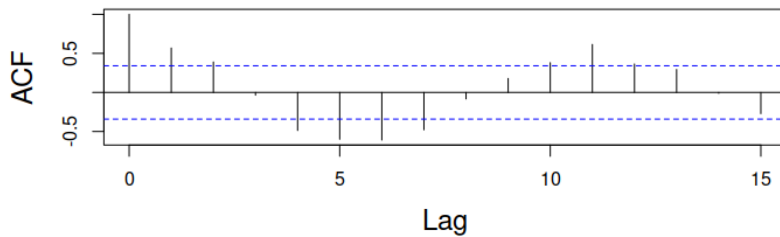
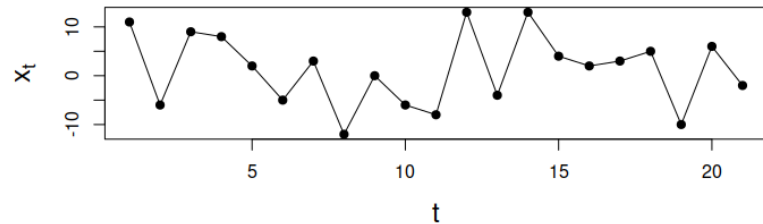
Οι διαφορές των παραπάνω ανά 12 είναι οι εξής:

$X_{13} - X_1 = 11, X_{14} - X_2 = -6, \dots, 9, 8, 2, -5, 3, -12, 0, -6, -8, 13, -4, 13, 4, 2, 3, 5, -10, 6, -2.$

Χρονοδιάγραμμα



Χρονοδιάγραμμα





SARIMA(0, 0, 0)(0, 1, 1)₁₂

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
DEMOCRITUS UNIVERSITY OF THRACE

Παράδειγμα: Ένα SARIMA(0, 0, 0)(0, 1, 1)₁₂ μοντέλο, χωρίς σταθερά, έχει εξίσωση της μορφής

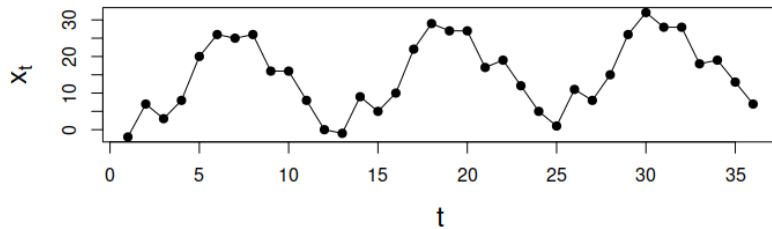
$$\Delta^{12}X_t = w_t \text{ ή } (1 - B^{12})X_t = (I + \Theta B^{12})w_t \text{ ή } X_t - X_{t-12} = w_t + \Theta w_{t-12}.$$

temp2 = c(-2, 7, 3, 8, 20, 26, 25, 26, 16, 16, 8, 0, -1, 9, 5, 10, 22, 29, 27, 27, 17, 19, 12, 5, 1, 11, 8, 15, 26, 32, 28, 28, 18, 19, 13, 7)

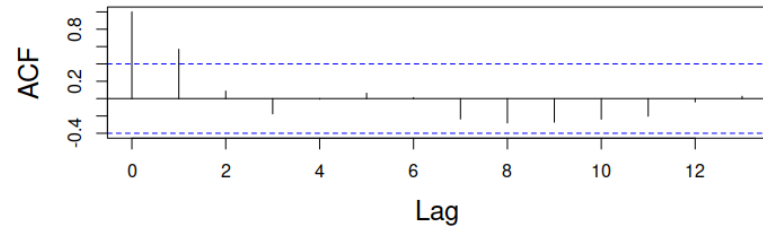
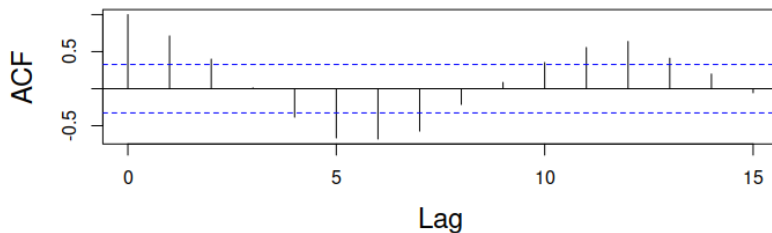
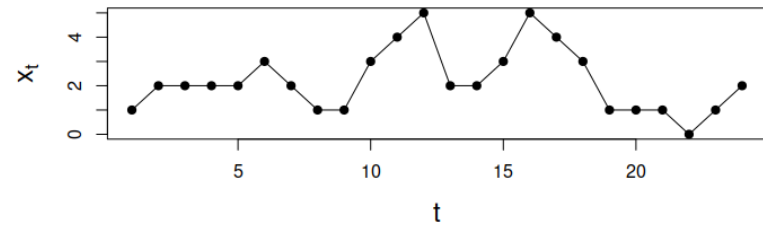
Οι διαφορές των παραπάνω ανά 12 είναι οι εξής:

$X_{13} - X_1 = 1$, $X_{14} - X_2 = 2, \dots$ 2, 2, 2, 3, 2, 1, 1, 3, 4, 5, 2, 2, 3, 5, 4, 3, 1, 1, 1, 0, 1, 2

Χρονοδιάγραμμα



Χρονοδιάγραμμα





SARIMA(1, 0, 0)(0, 1, 1)₁₂

Παράδειγμα

Ένα SARIMA(1, 0, 0)(0, 1, 1)₁₂ μοντέλο, χωρίς σταθερά, έχει εξίσωση της μορφής

$$(1 - B^{12})X_t = \varphi(1 - B^{12})X_{t-1} + (1 + \Theta B^{12})w_t$$

ή

$$X_t - X_{t-12} = \varphi X_{t-1} - \varphi X_{t-13} + w_t + \Theta w_{t-12},$$



SARIMA(0, 2, 1)(0, 0, 1)₁₂

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
DEMOCRITUS UNIVERSITY OF THRACE

Παράδειγμα

Ένα SARIMA(0, 2, 1)(0, 0, 1)₁₂ μοντέλο, χωρίς σταθερά, έχει εξίσωση της μορφής

$$(1 - B)^2 X_t = (1 + \theta B)(1 + \Theta B^{12}) w_t,$$

ή

$$X_t - 2X_{t-1} + X_{t-2} = w_t + \Theta w_{t-12} + \theta w_{t-1} + \theta \Theta w_{t-13}.$$

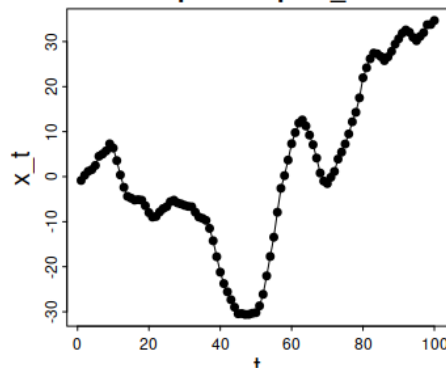


SARIMA(0, 2, 1)(0, 0, 1)₁₂

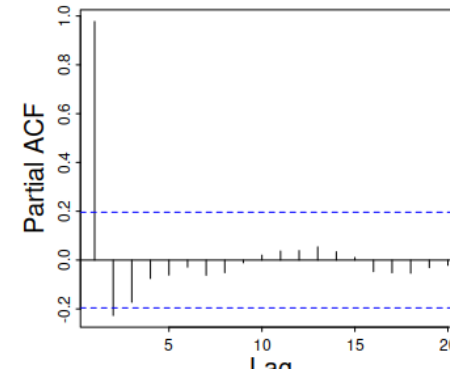
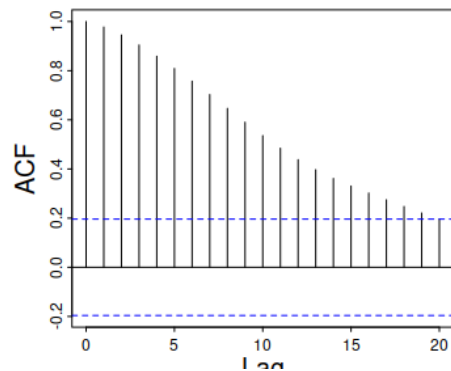
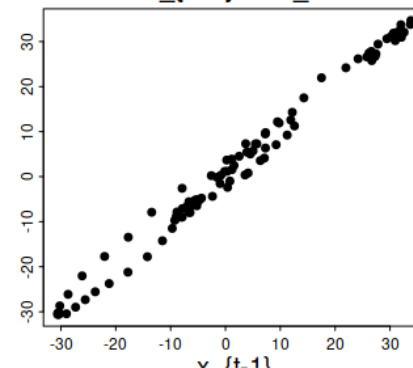
Παράδειγμα

$$X_t - 2X_{t-1} + X_{t-2} = \varepsilon_t - 0.2\varepsilon_{t-12} + 0.1\varepsilon_{t-1} - 0.02\varepsilon_{t-13}$$

Χρονοσειρά x_t



$x_{\{n-1\}}$ vs x_n



```
library(astsa)  
x = sarima.sim(ar = 0, d=2, ma=0.1, sar = 0, D=0, sma=-0.2, S=12, n=100)  
my.plot.ts.simple(x)
```



SARIMA(2, 1, 0)(0, 2, 2)₁₂

Παράδειγμα

Ένα ARIMA(2, 1, 0)(0, 2, 2)₁₂ μοντέλο, χωρίς σταθερά, έχει εξίσωση της μορφής

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})^2 Y_t = (1 + \Theta_1 B^{12} + \Theta_2 B^{24}) e_t$$

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - 2B^{12} + B^{24}) Y_t = (1 + \Theta_1 B^{12} + \Theta_2 B^{24}) e_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - B + \phi_1 B^2 + \phi_2 B^3)(1 - 2B^{12} + B^{24}) Y_t \\ = (1 + \Theta_1 B^{12} + \Theta_2 B^{24}) e_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - B + \phi_1 B^2 + \phi_2 B^3 - 2B^{12} + 2\phi_1 B^{13} + 2\phi_2 B^{14} + 2B^{13} \\ - 2\phi_1 B^{14} - 2\phi_2 B^{15} + B^{24} - \phi_1 B^{25} - \phi_2 B^{26} - B^{25} + \phi_1 B^{26} + \phi_2 B^{27}) Y_t \\ = (1 + \Theta_1 B^{12} + \Theta_2 B^{24}) e_t$$

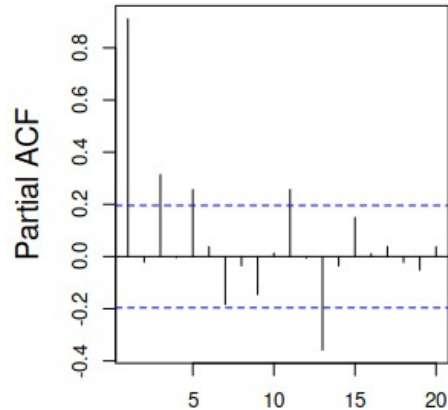
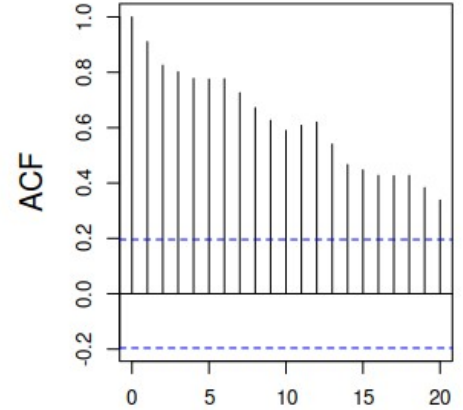
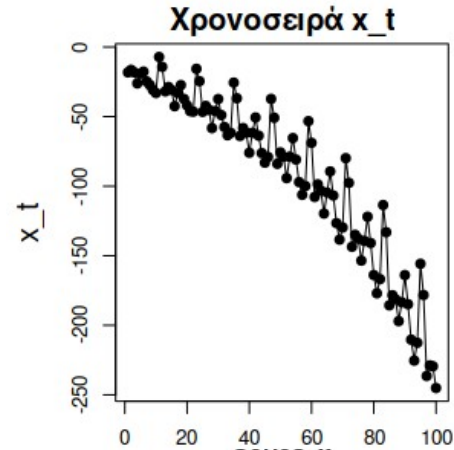
$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - Y_{t-1} + \phi_1 Y_{t-2} + \phi_2 Y_{t-3} - 2Y_{t-12} + 2\phi_1 Y_{t-13} + 2\phi_2 Y_{t-14} \\ + 2Y_{t-13} - 2\phi_1 Y_{t-14} - 2\phi_2 Y_{t-15} + Y_{t-24} - \phi_1 Y_{t-25} - \phi_2 Y_{t-26} - Y_{t-25} + \phi_1 Y_{t-26} \\ + \phi_2 Y_{t-27} = e_t + \Theta_1 e_{t-12} + \Theta_2 e_{t-24}$$



SARIMA(2, 1, 0)(0, 2, 2)₁₂

ΔΗΜΟΚΡΕΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
DEMOCRITUS UNIVERSITY OF THRACE

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - Y_{t-1} + \phi_1 Y_{t-2} + \phi_2 Y_{t-3} - 2Y_{t-12} + 2\phi_1 Y_{t-13} + 2\phi_2 Y_{t-14} + 2Y_{t-13} - 2\phi_1 Y_{t-14} - 2\phi_2 Y_{t-15} + Y_{t-24} - \phi_1 Y_{t-25} - \phi_2 Y_{t-26} - Y_{t-25} + \phi_1 Y_{t-26} + \phi_2 Y_{t-27} = e_t + \Theta_1 e_{t-12} + \Theta_2 e_{t-24}$$



library(astsa)

x = sarima.sim(ar = c(0.1, -0.3), d=1, ma=0, sar = 0, D=2, sma=c(-0.2, 0.1), S=12, n=100)

my.plot.ts.simple(x)



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Επιλέγοντας ένα μοντέλο



Βήματα ανάλυσης χρονοσειράς

1ο Στάδιο: Ταυτοποίηση.

Σε αυτό γίνεται μια προσπάθεια να αναγνωριστούν τα πιθανά υποδείγματα που ταιριάζουν με τα δεδομένα. Η αναγνώριση γίνεται μέσα από τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.

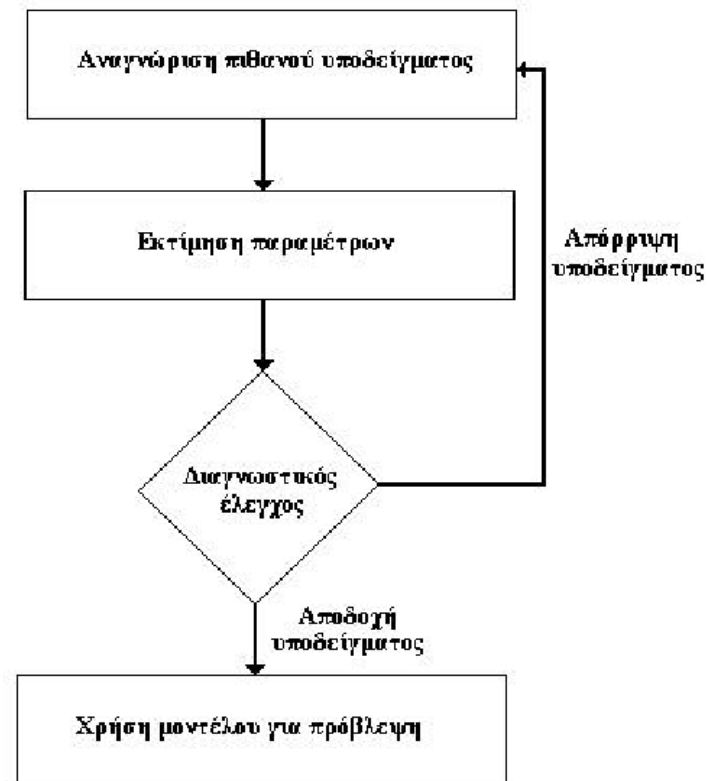
2ο Στάδιο: Εκτίμηση και δοκιμή.

Στο στάδιο αυτό γίνεται η εκτίμηση των παραμέτρων του υποδείματος που αναγνωρίστηκε. Στη συνέχεια γίνεται η δοκιμή, ο έλεγχος προσαρμογής και ικανότητας πρόβλεψης του υποδείματος. Σε περίπτωση μη αποδοχής του υποδείματος, επαναλαμβάνεται η διαδικασία για να επιλεγεί ένα πιο κατάλληλο υπόδειγμα.

3ο Στάδιο: Χρήση.

Σε περίπτωση που το υπόδειγμα γίνει αποδεκτό ακολουθεί το τρίτο στάδιο, δηλαδή η χρήση του μοντέλου για πρόβλεψη.

Βήματα ανάλυσης για την κατασκευή μοντέλων ARIMA





Κριτήριο AIC (Akaike information criterion)

Το κριτήριο πληροφοριών Akaike (AIC) (Hirotugu Akaike(1927 – 2009) είναι ένας εκτιμητής του σφάλματος πρόβλεψης ενός μοντέλου. Συνεπώς, εκφράζεται με αυτό η σχετική ποιότητα πρόβλεψης ενός στατιστικού μοντέλου για ένα σύνολο δεδομένων.

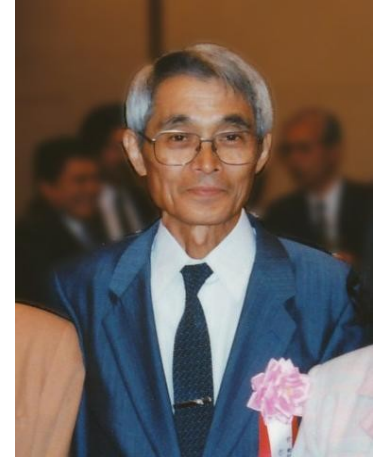
Το AIC ορίζεται ως το αντίθετο της μέγιστης τιμής του λογάριθμου της πιθανοφάνειας L του μοντέλου, προσαρμοσμένο ως προς τον αριθμό των παραμέτρων d στο μοντέλο,

$$AIC = 2 * d - 2 * \log\text{likelihood}.$$

Μεγαλύτερη τιμή πιθανοφάνειας, αντιστοιχεί σε μικρότερη τιμή του AIC.

Άρα, υπολογίζοντας το AIC για όλα τα πιθανά μοντέλα, επιλέγεται αυτό που έχει τη μικρότερη τιμή σε συνδυασμό με την απλότητά του ορισμού του.

Στην R αναφέρεται και η 2ης τάξης τιμή AICc η οποία συγκλίνει στην AIC όταν το δείγμα $n \rightarrow \infty$.



Hirotugu Akaike(1927 – 2009)

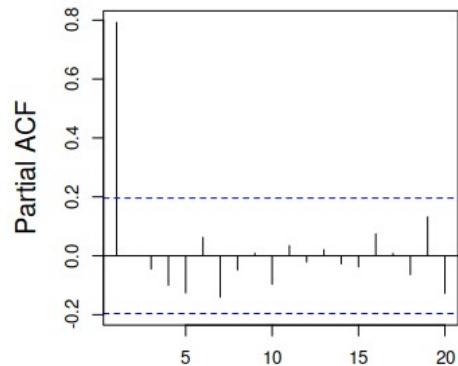
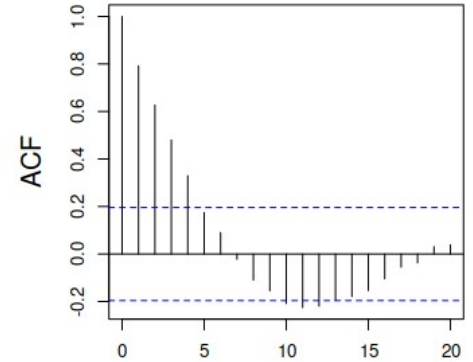
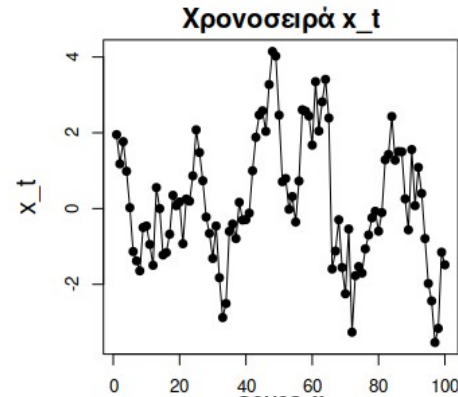


Κριτήριο AIC (Akaike information criterion)

```
x <- arima.sim(list(order = c(1,0,1), ar = 0.7, ma = -0.1), n = 100)  
my.plot.ts.simple(x)
```

```
model1 = arima(x,order=c(1,0,0))  
model2 = arima(x,order=c(2,0,0))  
model3 = arima(x,order=c(1,0,1))  
model4 = arima(x,order=c(0,0,2))
```

```
print(model1$aic)  
print(model2$aic)  
print(model3$aic)  
print(model4$aic)
```



```
> print(model1$aic)  
[1] 289.7061
```

```
> print(model2$aic)  
[1] 291.7021
```

```
> print(model3$aic)  
[1] 291.7022
```

```
> print(model4$aic)  
[1] 310.8712
```



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Η συνέχεια στο ...

<https://utopia.duth.gr/epdiaman/files/dpth/TimeSeriesActivity.html>