

Κεφάλαιο 22

Ανάλυση Χρονοσειρών

22.1 Ανάλυση Χρονοσειρών

Με τον όρο Χρονοσειρά εννοούμε μια σειρά από παρατηρήσεις που παίρνονται σε ορισμένες χρονικές στιγμές ή περιόδους που ισαπέχουν μεταξύ τους. Υπάρχουν ένα μεγάλο εύρος στατιστικών μεθόδων για την ανάλυση χρονοσειρών. Γενικά οι μέθοδοι αυτοί ανήκουν σε δύο κατηγορίες: αυτές που βασίζονται στη μελέτη συναρτήσεων που εξαρτώνται από τον χρόνο, και σε αυτές οι οποίες εξαρτώνται από τις συχνότητες και οι οποίες ερευνούν τις περιοδικές ιδιότητες που μπορεί να έχει η σειρά.

Τα τρία κυριότερα στοιχεία της ανάλυσης χρονοσειρών είναι η περιγραφή, η επεξήγηση και η πρόβλεψη των εξαρτημένων δεδομένων. Η περιγραφή επιτυγχάνεται με τη βοήθεια διαφόρων γραφημάτων, η επεξήγηση χρησιμοποιώντας κάποιες μορφής μοντέλα για να εξερευνηθούν οι μηχανισμοί δημιουργίας της χρονοσειράς, και η πρόβλεψη περιλαμβάνει τη χρησιμοποίηση ενός μοντέλου για να προβλεφθούν μελλοντικές τιμές της σειράς. Στο παράδειγμα το οποίο θα χρησιμοποιηθεί στο κεφάλαιο αυτό η ανάλυση θα επικεντρωθεί στο πεδίο του χρόνου.

Η **συνάρτηση αυτοδιακύμανσης** (autocovariance function) είναι το βασικό εργαλείο για την περιγραφή της σειριακής εξάρτησης μιας μονομεταβλητής, στάσιμης (χωρίς περιοδικότητα) χρονοσειράς και ορίζεται από

$$\gamma(k) = E(X_t - \mu)(X_{t+k} - \mu), k = 0, \pm 1, \pm 2, \dots$$

όπου X_t , $t = 0, \pm 1, \pm 2, \dots$ οι τιμές της σειράς, μ η μέση τιμή της και k η υστέρηση (lag) για την οποία υπολογίζεται η αυτοδιακύμανση. Η **συνάρτηση αυτοσυσχέτισης** (autocorrelation function) είναι η κανονικοποιημένη μορφή της συνάρτησης

αυτοδιακύμανσης και ορίζεται ως

$$\varrho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\gamma(k)}{\sigma_x^2}$$

όπου σ_x^2 η διακύμανση της χρονοσειράς.

Η συνάρτηση αυτοδιακύμανσης μπορεί να εκτιμηθεί από

$$\widehat{\gamma(k)} = \frac{1}{n} \sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x}).$$

Συνεπώς, η συνάρτηση αυτοσυσχέτισης μπορεί να εκτιμηθεί από

$$\widehat{\varrho(k)} = \frac{\widehat{\gamma(k)}}{\widehat{\gamma(0)}}$$

Για να διερευνηθούν τυχών μηχανισμοί δημιουργίας των δεδομένων, αλλά και να υποδειχθούν κατάλληλα μοντέλα, χρησιμοποιούνται οι γραφικές παραστάσεις της αυτοδιακύμανσης ή της αυτοσυσχέτισης συναρτήσει της υστέρησης k . Τέτοια μοντέλα ανάλυσης χρονοσειρών είναι της μορφής

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \epsilon_t$$

όπου ϵ_t το σφάλμα. Το πιο πάνω μοντέλο είναι γνωστό ως **διαδικασία αυτοπαλινδρόμησης βαθμού p** .

22.2 Παράδειγμα

Τα δεδομένα που θα χρησιμοποιηθούν παρουσιάζουν τον αριθμό των ηλιακών κηλίδων από το 1771 ως το 1870. Η εντολή `ts` χρησιμοποιείται για να δημιουργήσει ένα αντικείμενο χρονοσειράς.

```
> sun.ts<-ts(round(sunspot.year[71:171]),start=1771,end=1870)
```

```
> sun.ts
```

```
Time Series:
```

```
Start = 1771
```

```
End = 1870
```

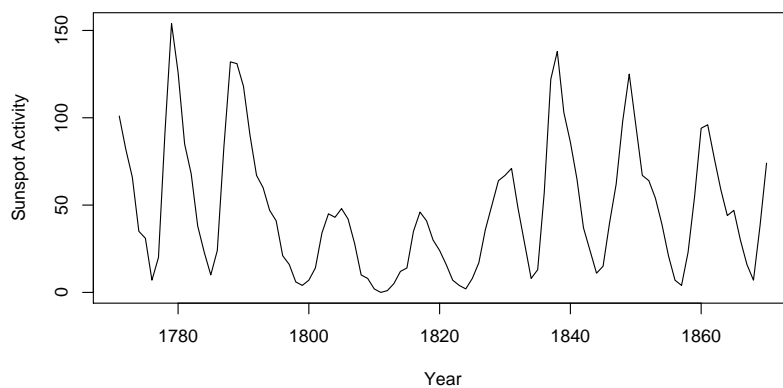
```
Frequency = 1
```

```
[1] 101 82 66 35 31 7 20 92 154 126 85 68 38 23 10 24 83 132  
[19] 131 118 90 67 60 47 41 21 16 6 4 7 14 34 45 43 48 42  
[37] 28 10 8 2 0 1 5 12 14 35 46 41 30 24 16 7 4 2
```

```
[55]  8  17  36  50  64  67  71  48  28  8  13  57 122 138 103  86  65  37
[73] 24  11  15  40  62  98 125  96  67  64  54  39  21  7  4  23  55  94
[91] 96  77  59  44  47  30  16  7  38  74
```

Το πρώτο βήμα στην ανάλυση χρονοσειρών είναι να γίνει η περιγραφή της χρονοσειράς χρησιμοποιώντας το γράφημα των τιμών της συναρτήσεως του χρόνου. Το γράφημα αυτό αποκαλύπτει τα κυριότερα στοιχεία της σειράς όπως είναι η τάση και η περιοδικότητα. Το γράφημα κατασκευάζεται με την εντολή `ts.plot` και παρουσιάζεται στο Σχήμα 22.1.

```
> ts.plot(sun.ts,xlab="Year",ylab="Sunspot Activity")
```



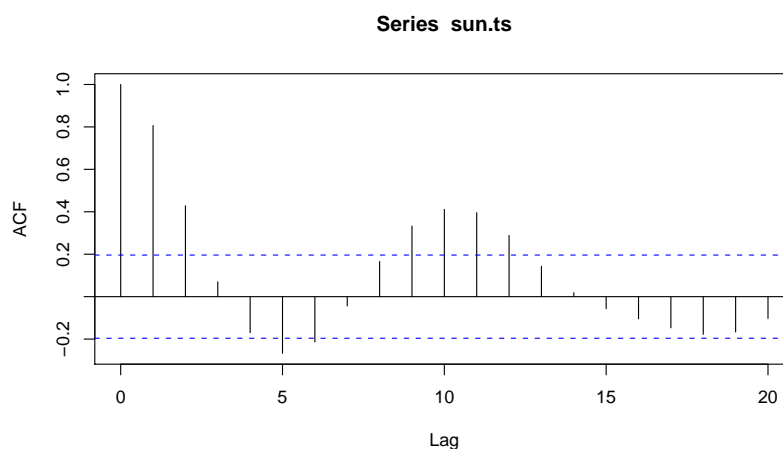
Σχήμα 22.1: Γράφημα της χρονοσειράς του αριθμού των ηλιακών κηλίδων από το 1771 ως το 1870.

Το πιο πάνω γράφημα δείχνει ότι υπάρχει μια περιοδικότητα στα δεδομένα αλλά δεν φαίνεται να υπάρχει κάποια τάση. Στη συνέχεια υπολογίζονται οι αυτοσυσχετίσεις και κατασκευάζεται το γράφημά τους συναρτήσεως της υστέρησης (Σχήμα 22.2). Η συνάρτηση στην R για κατασκευή του γραφήματος της συνάρτησης αυτοσυσχετίσεως είναι η `acf`.

```
> acf(sun.ts,lag.max=20,type="correlation")
```

Το όρισμα `lag.max` ορίζει τον μέγιστο αριθμό της υστέρησης για την οποία θα γίνει η γραφική παράσταση της συνάρτησης αυτοσυσχετίσεως, ενώ το όρισμα `type`

αν θα χρησιμοποιηθεί η αυτοδιακύμανση ή η αυτοσυσχέτιση. Εξ ορισμού το `type` είναι η αυτοσυσχέτιση και για αυτό στο παράδειγμά μας δεν ήταν αναγκαίο να οριστεί. Στο γράφημα οι δύο οριζόντιες διακεκομμένες γραμμές δείχνουν το 95% διάστημα εμπιστοσύνης για τον έλεγχο $H_0 : \rho = 0$. Είναι φανερό ότι η αυτοσυσχέτιση έχει μια ημιτονοειδή μορφή και αυτό υποδεικνύει ότι την περιοδικότητα που παρουσιάστηκε και στο γράφημα του Σχήματος 22.1. Επίσης, φαίνεται ότι αυτή η περιοδικότητα είναι ίση με 10 με 11 χρόνια, ένα στοιχείο που είναι γενικά γνωστό για την δραστηριότητα μιας ηλιακής κηλίδας.



Σχήμα 22.2: Συνάρτηση Αυτοσυσχέτισης του αριθμού των ηλιακών κηλίδων από το 1771 ως το 1870.

Πριν να γίνει προσπάθεια μοντελοποίησης της χρονοσειράς, πρέπει να γίνει προσπάθεια να εξαληφθεί η περιοδικότητα. Ένας απλός τρόπος να επιτευχθεί αυτό είναι να θεωρηθούν οι κατάλληλες διαφορές της σειράς. Συνεπώς, υποθέτοντας περιοδικότητα 11 χρόνων, μπορούν να θεωρηθούν οι διαφορές μεταξύ των σημείων ανά 11 χρόνια. Δηλαδή, ορίζουμε μια νέα χρονοσειρά

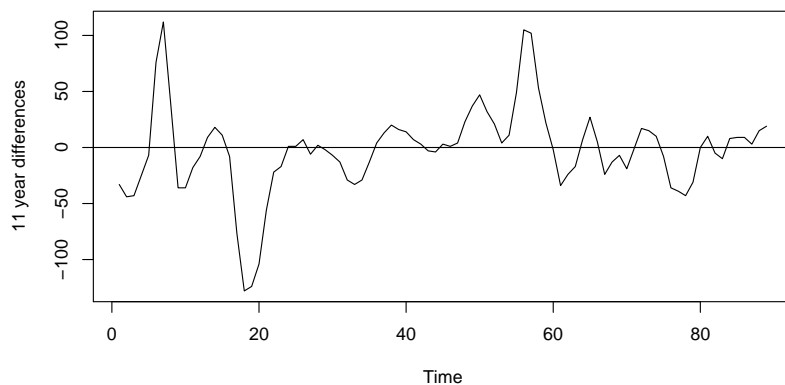
$$y_t = x_t - x_{t-11}.$$

Στην R αυτό γίνεται όπως πιο κάτω:

```
> newsun<-sun.ts[12:100]-sun.ts[1:89]
> ts.plot(newsun,xlab="Time",ylab="11 year differences")
> abline(h=0)
> acf(newsun,lag.max=20)
```

Το γράφημα της νέας χρονοσειράς (Σχήμα 22.3), από την 11-χρονη διαφορά των τιμών της αρχικής, ελάχιστο στοιχείο περιοδικότητας. Επίσης, η συνάρτηση αυτοσυσχέτισης στο Σχήμα 22.4 δείχνει ότι η ημιτονοειδής μορφή δεν είναι σε τόσο μεγάλο βαθμό όπως πριν, αλλά η αυτοσυσχέτιση περιορίστηκε σε υστέρηση ίση με 2.

Συνεχίζοντας, μπορεί να εφαρμοστεί στη νέα χρονοσειρά των 11-χρονων διαφορών ένα μοντέλο αυτοπαλινδρόμησης και αυτό στην R επιτυγχάνεται με τη συνάρτηση `ar.yw`. Σημαντικό στοιχείο της συνάρτησης αυτής είναι ότι μπορεί να εκτιμηθεί η τάξη του μοντέλου που χρειάζεται η σειρά με την επιλογή της τιμής της υστέρησης k η οποία ελαχιστοποιεί το κριτήριο AIC.



Σχήμα 22.3: Γράφημα της 11-χρονης διαφοράς του αριθμού των ηλιακών κηλίδων από το 1771 ως το 1870.

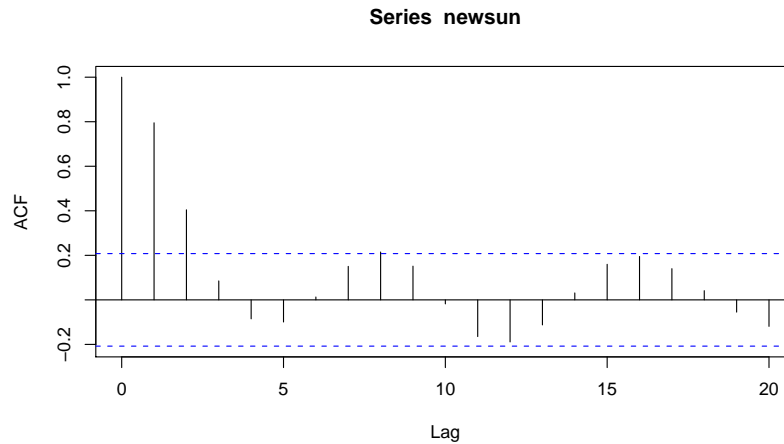
```
> sun.ar<-ar.yw(newsun)
> sun.ar
```

```
Call: ar.yw.default(x = newsun)
```

```
Coefficients:
```

```
      1      2      3      4      5
1.4985 -1.1434  0.6369 -0.4126  0.1961
```

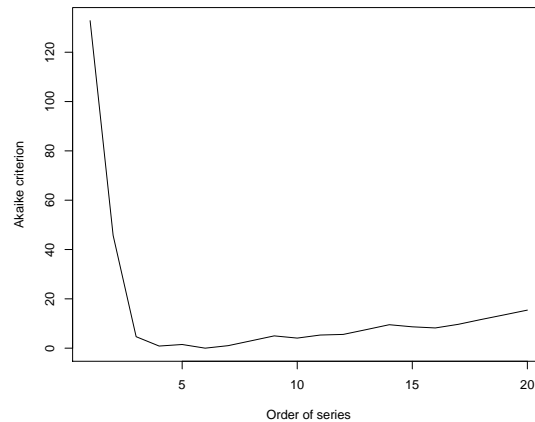
```
Order selected 5  sigma^2 estimated as  313.4
```



Σχήμα 22.4: Συνάρτηση Αυτοσυσχέτισης της 11-χρονης διαφοράς του αριθμού των ηλιακών κηλίδων από το 1771 ως το 1870.

Το αποτέλεσμα του μοντέλου παρουσιάζει τους εκτιμώμενους συντελεστές του όπως και την τάξη του. Επίσης δίνεται και η εκτίμηση της διασποράς του σφάλματος του μοντέλου. Συνεπώς, η τάξη του μοντέλου αυτοπολινδρόμησης που εφαρμόζει καλύτερα τα δεδομένα εκτιμήθηκε να είναι ίση με 5. Μπορεί να κατασκευαστεί το γράφημα του AIC όπως πιο κάτω και παρουσιάζεται στο Σχήμα 22.5. Στο γράφημα φαίνεται ότι το κριτήριο AIC ελαχιστοποιείται στην τιμή 6, αλλά αυτό εξηγείται από το γεγονός ότι το γράφημα ξεκινά από το σημείο 1, το οποίο βασικά αντιστοιχεί στην τιμή του AIC για το μοντέλο τάξης 0.

```
> ts.plot(sun.ar$aic,xlab="Order of series",ylab="Akaike criterion")
```

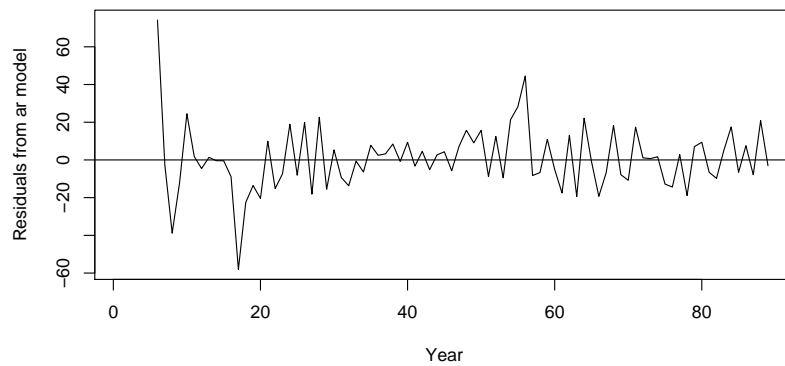


Σχήμα 22.5: Γράφημα του κριτηρίου AIC από την εφαρμογή μοντέλων αυτοπαλινδρόμησης στα δεδομένα.

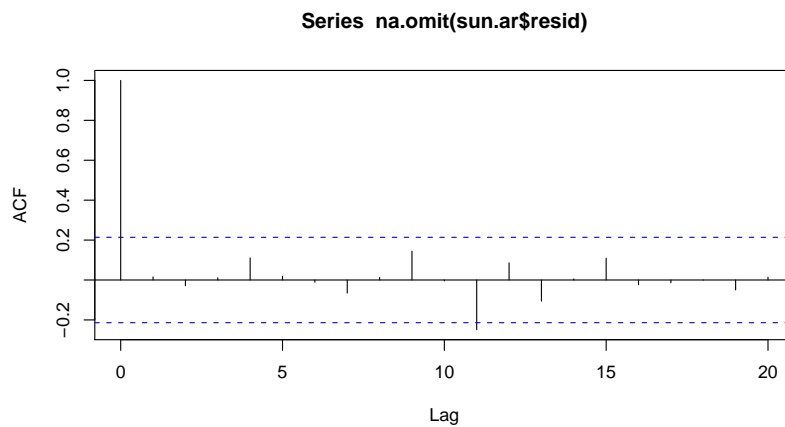
Τα γράφημα των υπολοίπων κατασκευάζονται με τις εντολές

```
> ts.plot(sun.ar$resid,xlab="Year",ylab="Residuals from ar model")  
> abline(h=0)  
> acf(na.omit(sun.ar$resid),lag.max=20)
```

Στην περίπτωση που το μοντέλο εφαρμόζει καλά στα δεδομένα, τότε τα υπόλοιπα θα είχαν τη μορφή τυχαίας σειράς στο Σχήμα 22.6, δηλαδή θα είχαν τη μορφή μιας χρονοσειράς λευκού θορύβου. Είναι φανερό πως τα υπόλοιπα δεν παρουσιάζουν μια προφανή δομή. Το κορελλόγραμμα των υπολοίπων βοηθάει επίσης στο να ελεγχθεί η τυχαιότητά τους. Σε μια σειρά λευκού θορύβου δεν παρουσιάζεται καμία ημιτονοειδής μορφή αλλά και η αυτοσυσχέτιση για όλες τις υστερήσεις είναι μεταξύ των ορίων του 95% διαστήματος εμπιστοσύνης. Αυτό δείχνει και το γράφημα της συνάρτησης αυτοσυσχέτισης στο Σχήμα 22.7.



Σχήμα 22.6: Γράφημα των υπολοίπων του μοντέλου αυτοπαλινδρόμησης.



Σχήμα 22.7: Συνάρτηση Αυτοσυσχέτισης των υπολοίπων του μοντέλου αυτοπαλινδρόμησης.