# Χρονικές σειρές

## $11^{ο}$ μάθημα: Προβλέψεις

## Εαρινό εξάμηνο 2018-2019
## Τμήμα Μαθηματικών ΑΠΘ


Διδάσκουσα: **Αγγελική Παπάνα**

Μεταδιδακτορική Ερευνήτρια

Πολυτεχνική σχολή, Α.Π.Θ. & Οικονομικό Τμήμα, Πανεπιστήμιο Μακεδονίας

http://users.auth.gr/~agpapana/

## Προβλέψεις

Ο κύριος σκοπός της εξειδίκευσης και εκτίμησης ενός μοντέλου ARIMA είναι η διενέργεια βραχυχρόνιων προβλέψεων. Με βάση το εκτιμώμενο μοντέλο και τις υπάρχουσες πληροφορίες μέχρι την χρονική περίοδο $\mathrm{T}$, γίνετε πρόβλεψη της τιμής της χρονοσειράς στις περιόδους $\mathrm{T}+1, \mathrm{T}+2, \ldots$

Στην περίπτωση του απλού γραμμικού υποδείγματος, η καλύτερη πρόβλεψη μιας μελλοντικής τιμής $\Upsilon_f$ της χρονοσειράς $\Upsilon$, για δεδομένη τιμή $X_f$ της ερμηνευτικής μεταβλητής $\mathrm{X}$ είναι η

$$\boxed{\widehat{\Upsilon}_f = \hat{\beta}_0 + \hat{\beta}_1 X_f}$$

η οποία στην ουσία είναι η $\mathrm{E}(\Upsilon_f)$.

Η διαφορά $\boxed{e = \Upsilon_f - \widehat{\Upsilon}_f}$ παριστάνει το **σφάλμα πρόβλεψης**.

## ΑΠΛΕΣ ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ

### Μέθοδος κινητού μέσου

Μια από τις απλούστερες μεθόδους πρόβλεψης είναι αυτή του κινητού μέσου όρου (moving average).

Η προβλεπόμενη τιμή στη χρονική περίοδο t προκύπτει ως ο μέσος όρος των $k$ προηγούμενων μετρήσεων:

$$\boxed{\mathrm{F_t} = \frac{\Upsilon_{\mathrm{t}-1} + \Upsilon_{\mathrm{t}-2} + \cdots + \Upsilon_{\mathrm{t}-k}}{k}}$$

Η μέθοδος αυτή εφαρμόζεται συχνά ως βοηθητική μέθοδος για να εξομαλύνει τις τυχαίες διακυμάνσεις (θόρυβος) και να αποκαλύψει τις τάσεις.

Ο απλός κινητός μέσος είναι πολύ χρήσιμος για να απομακρύνει την τυχαία μεταβλητότητα στην πρόβλεψη, όταν η ζήτηση δεν έχει τάση και εποχικότητα.

Είναι πολύ σημαντικό να επιλεγεί το κατάλληλο διάστημα για τον κινητό μέσο. Ὅσο μεγαλύτερο είναι το επιλεγμένο διάστημα τόσο περισσότερο εξομαλύνεται το τυχαίο στοιχείο στην πρόβλεψη.

Όταν όμως υπάρχει τάση στην ζήτηση, δηλαδή αυξάνεται ή μειώνεται σε συνάρτηση με το χρόνο, ο κινητός μέσος ενός μεγάλου διαστήματος θα εξομαλύνει και την τάση. Επομένως, ένα μικρότερο χρονικό διάστημα αν και θα παρουσιάζει μεγαλύτερη διακύμανση ακολουθεί με μεγαλύτερη ακρίβεια την τάση της ζήτησης.

# Π.χ. Μέθοδος κινητού μέσου για $k = 4$

| Εβδομάδα | Παρατηρήσεις Ζήτησης | Πρόβλεψη |
|---|---|---|
| 1 | $D_1$ | |
| 2 | $D_2$ | |
| 3 | $D_3$ | |
| 4 | $D_4$ | |
| 5 | $D_5$ | |
| 6 | $D_6$ | |
| 7 | ? | $F_7$ |
| 8 | $D_8$ | $F_8$ |
| 9 | $D_9$ | $F_9$ |
| 10 | $D_{10}$ | $F_{10}$ |
| 11 | $D_{11}$ | $F_{11}$ |
| 12 | $D_{12}$ | $F_{12}$ |

$$F_7 = \frac{D_3 + D_4 + D_5 + D_6}{4}$$

$$\mathrm{F}_8 = \frac{D_4 + D_5 + D_6 + D_7}{4}$$

$$\mathrm{F}_9 = \frac{D_5 + D_6 + D_7 + D_8}{4}$$

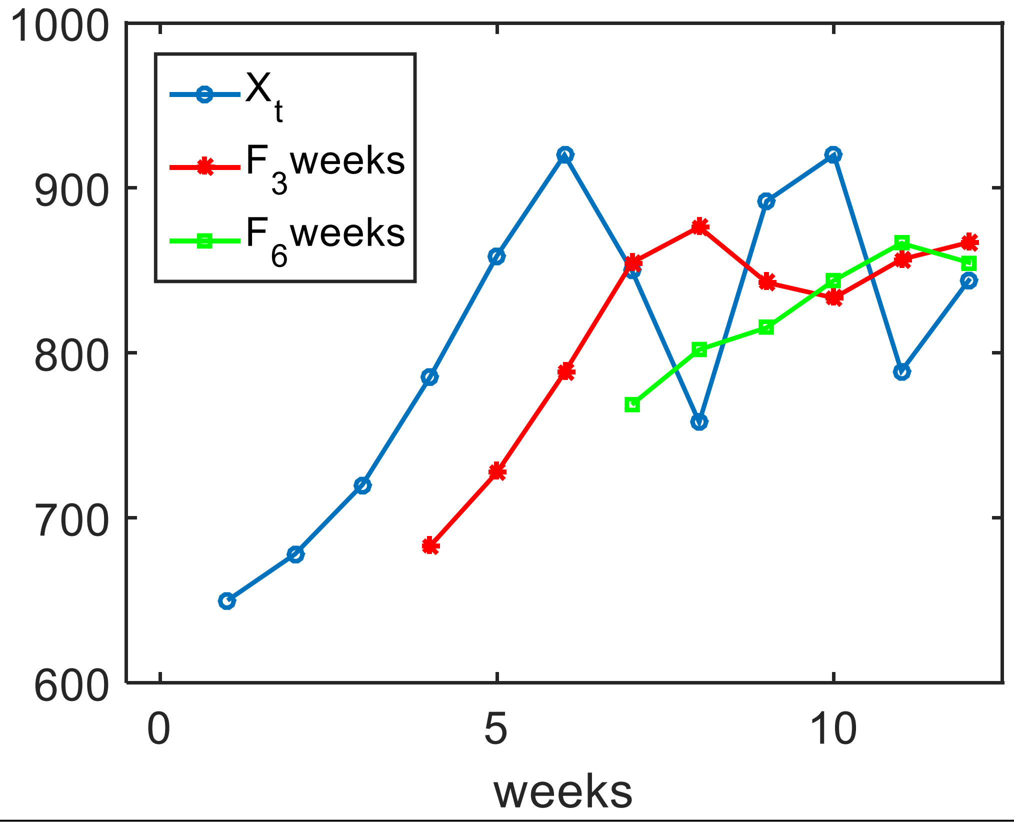
κ.ο.κ.

**Παράδειγμα** Απλός κινητός μέσος

Στον παρακάτω πίνακα υπολογίζεται ο κινητός μέσος της ζήτησης ενός προϊόντος για 3 ($k = 3$) και 6 εβδομάδες ($k = 6$).

| Εβδομάδα | Ζήτηση | 3-εβδ. | 6-εβδ. |
|---|---|---|---|
| 1 | 650 | | |
| 2 | 678 | | |
| 3 | 720 | | |
| 4 | 785 | 682.67 | |
| 5 | 859 | 727.67 | |
| 6 | 920 | 788.00 | |
| 7 | 850 | 854.67 | 768.67 |
| 8 | 758 | 876.33 | 802.00 |
| 9 | 892 | 842.67 | 815.33 |
| 10 | 920 | 833.33 | 844.00 |
| 11 | 789 | 856.67 | 866.50 |
| 12 | 844 | 867.00 | 854.83 |

Στο γράφημα φαίνεται συγκριτικά η επίδραση που έχει στον κινητό μέσο το διαφορετικό χρονικό διάστημα στην πρόβλεψη.




Η ζήτηση εμφανίζει συνεχής αύξηση μέχρι την $6^{η}$ εβδομάδα την οποία ακολουθεί ο κινητός μέσος των 3 εβδομάδων. Ο κινητός μέσος των 6 εβδομάδων εξομαλύνει περισσότερο την ζήτηση από ότι εκείνος των 3 εβδομάδων. Σε περιπτώσεις μεγάλων σχετικά αυξομειώσεων, όπως συμβαίνει από την έκτη εβδομάδα και μετά οι δύο μέσοι εμφανίζουν μικρές διαφορές με το μέσο των 3 εβδομάδων να ακολουθεί με λίγο μεγαλύτερη ακρίβεια την πραγματική ζήτηση.

## Σταθμισμένος κινητός μέσος

Σε αντίθεση με το απλό κινητό μέσο που δίνει το ίδιο βάρος σε κάθε παρατήρηση, ο σταθμισμένος κινητός μέσος παρέχει την δυνατότητα στάθμισης κάθε ιστορικού στοιχείου με διαφορετικό βάρος.

Το άθροισμα όλων των σταθμίσεων (βάρη) θα πρέπει να ισούται με 1 $(w_1 + w_2 + \cdots + w_k = 1)$.

Ο κινητός μέσος με βάρη υπολογίζεται σύμφωνα με τον παρακάτω τύπο:

$$\boxed{\mathrm{F_t} = w_1 \Upsilon_{\mathrm{t}-1} + w_2 \Upsilon_{\mathrm{t}-2} + \cdots + w_k \Upsilon_{\mathrm{t}-k}}$$

$k$ = ο αριθμός των περιόδων που χρησιμοποιούνται στην πρόβλεψη

**Παράδειγμα** Σταθμισμένος κινητός μέσος

Έστω ότι ένα πολυκατάστημα θέλει να προβλέψει τις πωλήσεις που θα έχει την $4^{η}$ εβδομάδα και σύμφωνα με προηγούμενες προβλέψεις έχει βρει ότι η καλύτερη πρόβλεψη θα πρέπει να χρησιμοποιεί το 50% της πιο πρόσφατης περιόδου, το 30% των πωλήσεων πριν δύο περιόδους και το 20% πριν 3 περιόδους.

| Εβδομάδα | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Πωλήσεις | 650 | 678 | 720 | ; |
| Βάρη | 0.2 | 0.3 | 0.5 | |

Η πρόβλεψη για την τέταρτη εβδομάδα θα είναι:

$$\mathrm{F}_4 = 0.5 \times 720 + 0.3 \times 678 + 0.2 \times 650 = 693.4$$

**Επιλογή Βαρών**: Η εμπειρία καθώς και οι δοκιμές διαφόρων βαρών στο παρελθόν είναι πιο απλοί τρόποι για την επιλογή των καταλληλότερων στην πρόβλεψη. Γενικά, οι πιο πρόσφατες παρατηρήσεις είναι πιθανό είναι καλύτεροι δείκτες του μέλλοντος και επομένως πρέπει να έχουν μεγαλύτερη βαρύτητας στον υπολογισμό του μέσου. Όταν, για παράδειγμα, υπάρχει εποχικότητα τα βάρη θα πρέπει να επιλέγονται ανάλογα με τις τιμές που έχει η χρονοσειρά στους αντίστοιχους μήνες.
Ο σταθμισμένος κινητός μέσος θεωρείται πιο ακριβείς και αξιόπιστο μέσο πρόβλεψης από τον απλό κινητό μέσο όσο αφορά την επίδραση των προηγούμενων δεδομένων στην πρόβλεψη.

## Μέθοδος εκθετικής εξομάλυνσης

Ένα από τα σημαντικότερα μειονεκτήματα της μεθόδου του κινητού μέσου όρου είναι ότι προσδίδει το ίδιο βάρος σε όλες τις $k$ προηγούμενες παρατηρήσεις.

Είναι όμως λογικό η τρέχουσα παρατήρηση να επηρεάζεται σε μεγαλύτερο βαθμό από τις πιο πρόσφατες παρατηρήσεις.

Η μέθοδος της εκθετικής εξομάλυνσης (exponential smoothing) λαμβάνει υπόψη το γεγονός αυτό και προσδίδει μεγαλύτερο βάρος στις πρόσφατες ιστορικές παρατηρήσεις.

Έστω ότι οι ιστορικές παρατηρήσεις εκτείνονται μέχρι τη χρονική περίοδο $t$ και στόχος είναι η πρόβλεψη των επόμενων περιόδων.

Η εκθετική εξομάλυνση προσδίδει στην παρατήρηση $t$ βάρος ίσο με $w$, στη επόμενη $(t-1)$ βάρος ίσο με $w(1-w)$, στην επόμενη $(t-2)$ βάρος ίσο με $w(1-w)^2$ και γενικά στην $(t-k)$ παρατήρηση προσδίδει βάρος ίσο με $w(1-w)^k$.

Καθώς η τιμή του **συντελεστή εξομάλυνσης ($w$)** κυμαίνεται από 0 έως 1, η ακολουθία των βαρών συνεχώς φθίνει.

Έτσι, κάθε χρονική στιγμή $t$ υπολογίζεται ο όρος $\mathrm{F}_t$, ο οποίος ονομάζεται **επίπεδο (level)** της χρονοσειράς ως:

$$\boxed{\mathrm{F}_t = w\Upsilon_{t-1} + w(1-w)\Upsilon_{t-2} + \cdots + w(1-w)^k\Upsilon_{t-k-1}}$$

Εναλλακτικά, μπορεί να χρησιμοποιηθεί η ακόλουθη αναδρομική σχέση:

$$\boxed{\mathrm{F}_t = wY_\mathrm{t} + (1 - \mathrm{w})\mathrm{F}_{t-1}}$$

Σύμφωνα με τη μέθοδο εκθετικής εξομάλυνσης, το επίπεδο $\mathrm{F}_t$ αποτελεί την πρόβλεψη για όλες τις χρονικές στιγμές μετά την $t$:

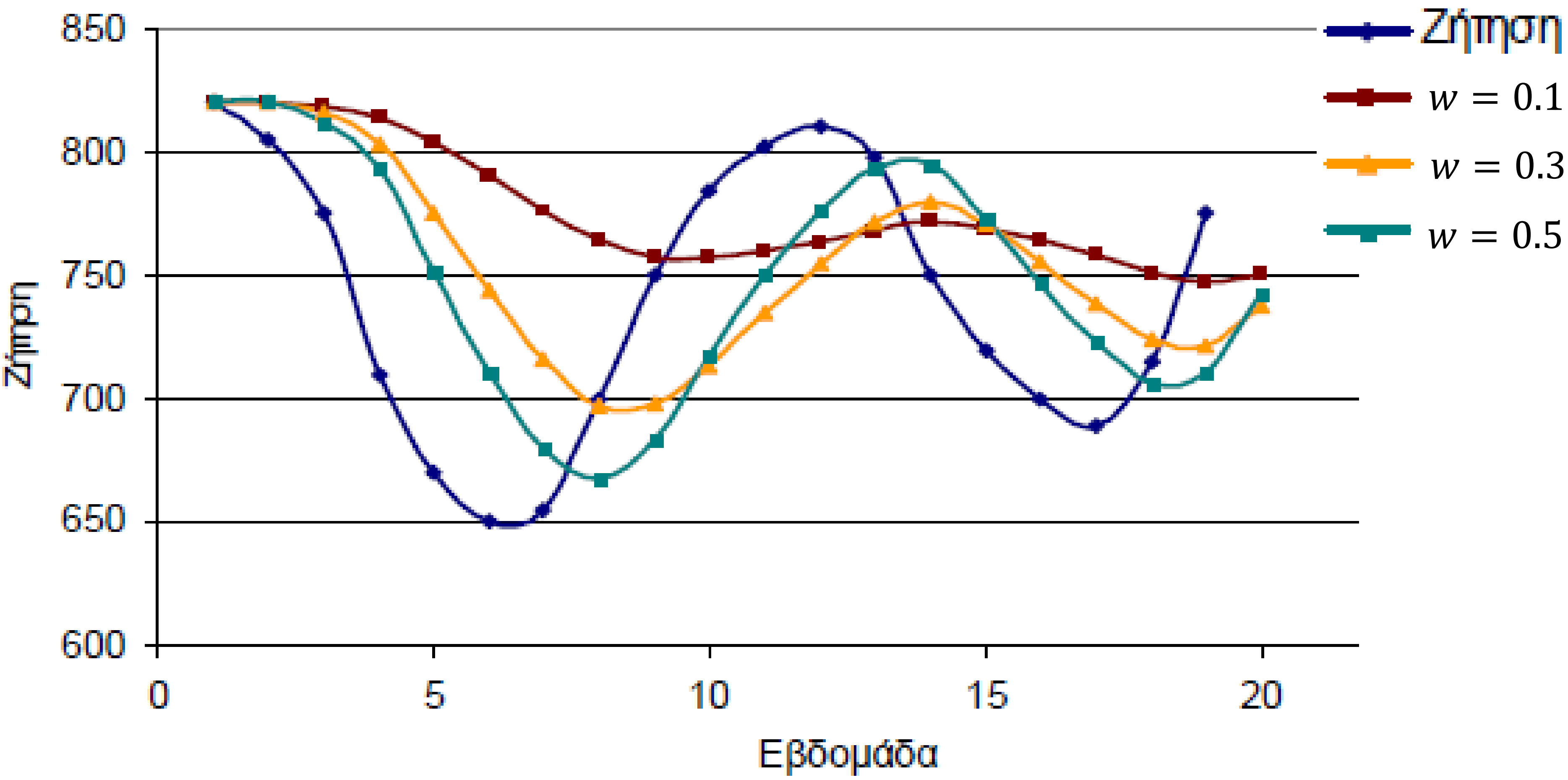
$$\boxed{\widehat{Y}_{t+k} = \mathrm{F}_t, k = 1,2, \ldots.}$$

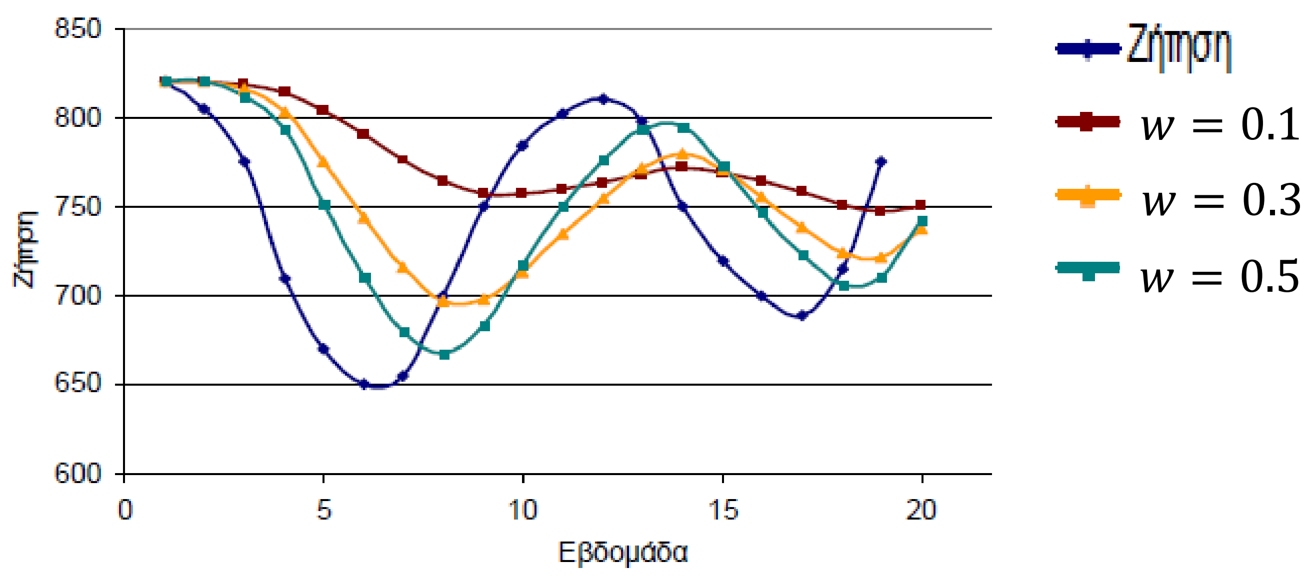
**Παράδειγμα** Εκθετική εξομάλυνση

Στον επόμενο πίνακα υπάρχει η ζήτηση $D_t$ μιας εταιρίας για 19 εβδομάδες. Να υπολογιστούν οι προβλέψεις για τις περιόδους 2 έως 20 με $w$ =0.10, $w$=0.30 και $w$ = 0.5.

| Εβδ. | Ζήτηση | α=0.1 | α=0.3 | α=0.5 | Εβδ. | Ζήτηση | α=0.1 | α=0.3 | α=0.5 |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 820 | 820.00 | 820.00 | 820.00 | **11** | 802 | 759.70 | 734.82 | 750.47 |
| **2** | 805 | 820.00 | 820.00 | 820.00 | **12** | 810 | 763.93 | 754.97 | 776.23 |
| **3** | 775 | 818.50 | 815.50 | 812.50 | **13** | 798 | 768.53 | 771.48 | 793.12 |
| **4** | 710 | 814.15 | 803.35 | 793.75 | **14** | 750 | 771.48 | 779.44 | 795.56 |
| **5** | 670 | 803.74 | 775.35 | 751.88 | **15** | 720 | 769.33 | 770.61 | 772.78 |
| **6** | 650 | 790.36 | 743.74 | 710.94 | **16** | 700 | 764.40 | 755.42 | 746.39 |
| **7** | 655 | 776.33 | 715.62 | 680.47 | **17** | 689 | 757.96 | 738.80 | 723.19 |
| **8** | 700 | 764.19 | 697.43 | 667.73 | **18** | 715 | 751.06 | 723.86 | 706.10 |
| **9** | 750 | 757.77 | 698.20 | 683.87 | **19** | 775 | 747.46 | 721.20 | 710.55 |
| **10** | 784 | 757.00 | 713.74 | 716.93 | **20** | | 750.21 | 737.34 | 742.77 |

Στο παρακάτω γράφημα βλέπουμε την επίδραση της σταθεράς εξομάλυνσης $w$ στην πρόβλεψη.






Όλες οι προβλέψεις υπολείπονται της ζήτησης όταν αυξάνεται ή μειώνεται ενώ σε περίπτωση αλλαγής κατεύθυνσης καθυστερούν κατά μία χρονική περίοδο.

Όσο μεγαλύτερη είναι η τιμή της σταθεράς εξομάλυνσης τόσο πιο πιστά ακολουθεί η πρόβλεψη την καμπύλη της πραγματικής ζήτησης.

Η βελτίωση της προβλεψιμότητας του μοντέλου μπορεί να επιτευχθεί με τους εξής τρόπους:

**1.** Προσθήκη ενός νέου παράγοντα τάσης για την διόρθωση της καθυστέρησης σε περίπτωση ανοδικής ή καθοδικής πορείας της ζήτησης.

**2.** Προσαρμογή της σταθεράς εξομάλυνσης κατά περιόδους, που ονομάζεται και προσαρμογή πρόβλεψης (adaptive forecasting).

## Μέθοδος ελαχίστων τετραγώνων

Η μέθοδος των ελαχίστων τετραγώνων εφαρμόζει μία γραμμή στα δεδομένα που ελαχιστοποιεί το άθροισμα των τετραγώνων της κάθετης απόστασης μεταξύ των πραγματικών δεδομένων και των αντίστοιχων σημείων της ευθείας.

Η ευθεία ελαχίστων τετραγώνων / ευθεία παλινδρόμησης είναι:

$$\boxed{\hat{y} = \alpha X + b}$$

όπου

$$b = \frac{\sum y_i \mathrm{T}_i - \mathrm{T}\bar{\mathrm{T}}\bar{y}}{\sum \mathrm{T}_i^2 - \mathrm{T}\bar{\mathrm{T}}^2} = \frac{\sum y_i \mathrm{T}_i - T\bar{\mathrm{T}}\bar{y}}{\sum \mathrm{T}_i^2 - T\bar{\mathrm{T}}^2}$$

$$\alpha = \bar{y} - b\bar{X}$$

$\bar{\mathrm{T}}$ = ο μέσος του χρόνου των $\mathrm{T}_i = 1, \ldots, \mathrm{T}$

$\bar{y}$ = ο μέσος της πραγματικής τιμής των $y_i$

$\mathrm{T}$ = ο αριθμός των παρατηρήσεων

**Παράδειγμα**

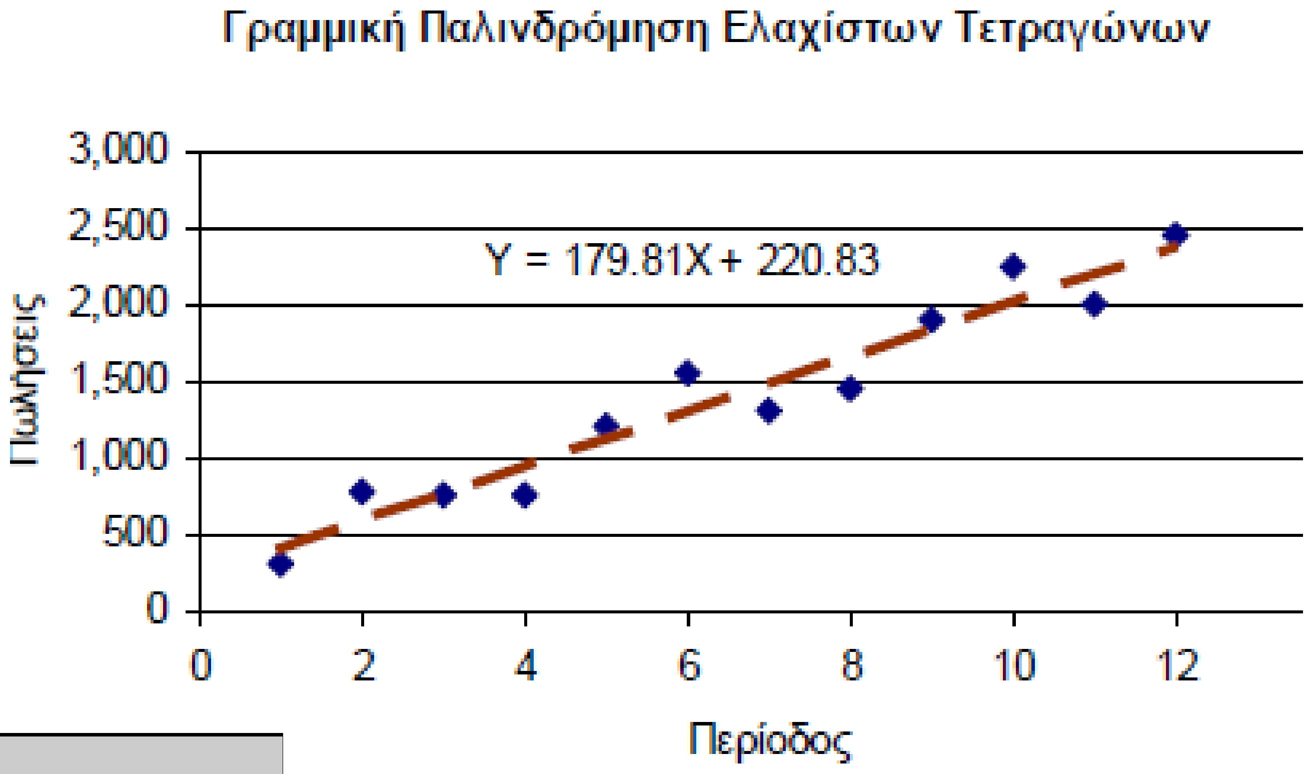
Μέθοδος ελαχίστων τετραγώνων

Δίνονται οι πωλήσεις μιας εταιρείας από 12 περιόδους. Να γίνει πρόβλεψη των πωλήσεων με την μέθοδο ελαχίστων τετραγώνων.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Περίοδος (Xi) | Πωλήσεις (yi) | Xi*yi | Xi2 | Y | (yi-YI)2 |
| | 1 | 300 | 300 | 1 | 400.641026 | 10128.62 |
| | 2 | 775 | 1550 | 4 | 580.448718 | 37850.2 |
| | 3 | 750 | 2250 | 9 | 760.25641 | 105.194 |
| | 4 | 750 | 3000 | 16 | 940.064103 | 36124.36 |
| | 5 | 1,200 | 6000 | 25 | 1119.87179 | 6420.529 |
| | 6 | 1,550 | 9300 | 36 | 1299.67949 | 62660.36 |
| | 7 | 1,300 | 9100 | 49 | 1479.48718 | 32215.65 |
| | 8 | 1,450 | 11600 | 64 | 1659.29487 | 43804.34 |
| | 9 | 1,900 | 17100 | 81 | 1839.10256 | 3708.498 |
| | 10 | 2,250 | 22500 | 100 | 2018.91026 | 53402.47 |
| | 11 | 2,000 | 22000 | 121 | 2198.71795 | 39488.82 |
| | 12 | 2,450 | 29400 | 144 | 2378.52564 | 5108.584 |
| Άθροισμα | 78 | 16675 | 134100 | 650 | | 331017.6 |
| Μέσος | 6.5 | 1389.58 | | | | |

$$b = \frac{\sum y_i X_i - T\bar{X}\bar{y}}{\sum X_i^2 - T\bar{X}^2} =$$
$$= \frac{\sum y_i X_i - T\bar{X}\bar{y}}{\sum X_i^2 - T\bar{X}^2} = 179.81$$

$$\alpha = \bar{y} - b\bar{X} =$$
$$= 1389.6 - 220.83 \times 6.5 =$$
$$= 220.83$$




Οι προβλέψεις για τις περιόδους 13 ως 16:

| Περίοδος | Πρόβλεψη | |
|---|---|---|
| 13 | 220.83 + 179.81 * 13= | 2558.333 |
| 14 | 220.83 + 179.81 * 14= | 2738.141 |
| 15 | 220.83 + 179.81 * 15= | 2917.949 |
| 16 | 220.83 + 179.81 * 16= | 3097.756 |

Τυπική απόκλιση σφάλματος της εκτίμησης της πρόβλεψης:

$$s = \sqrt{\frac{\sum_{t=1}^{M} \varepsilon_t{}^2}{T-1}} = 173.47$$

## Ανάλυση χρονοσειράς (decomposition of a time series)

Ως χρονοσειρά ορίζεται ένα σύνολο δεδομένων, με χρονολογική σειρά, που μπορεί να έχουν τάση, εποχικότητα / κυκλικότητα, αυτοσυσχέτιση και τυχαιότητα.

Με τον όρο ανάλυση χρονοσειράς εννοεί κανείς τον διαχωρισμό μιας χρονοσειράς στα επιμέρους στοιχεία της. Στην πράξη, είναι πιο απλό να βρεθεί η τάση και η εποχικότητα μιας χρονοσειράς αλλά πιο δύσκολο να αναγνωριστούν οι κύκλοι, η αυτοσυσχέτιση και τα τυχαία στοιχεία.

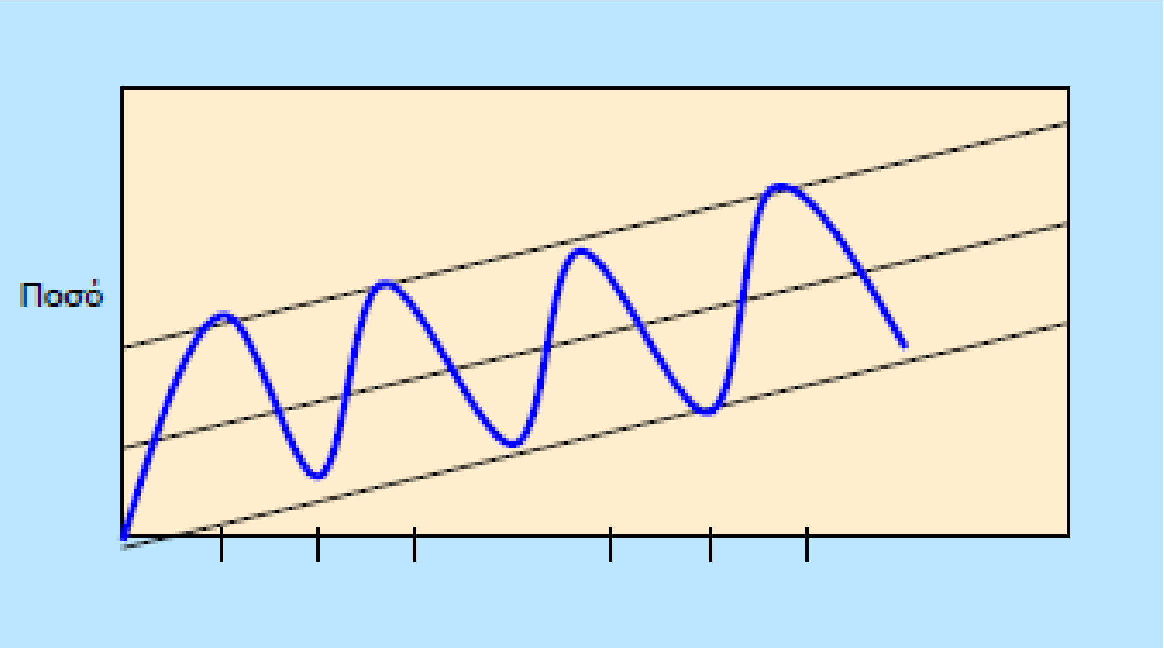
Όταν η ζήτηση περιλαμβάνει και στοιχεία τάσης και εποχικότητας, το ερώτημα είναι πώς τα στοιχεία αυτά σχετίζονται μεταξύ τους.

Υπάρχουν δύο είδη εποχικής διακύμανσης, η **προσθετική** και η **πολλαπλασιαστική**.

Η εποχική διακύμανση υποθέτει ότι το ποσό της εποχικότητας είναι σταθερό ανεξαρτήτως της τάσης ή του μέσου ποσού.

$$\boxed{\textbf{Πρόβλεψη = Τάση + Εποχικότητα}}$$

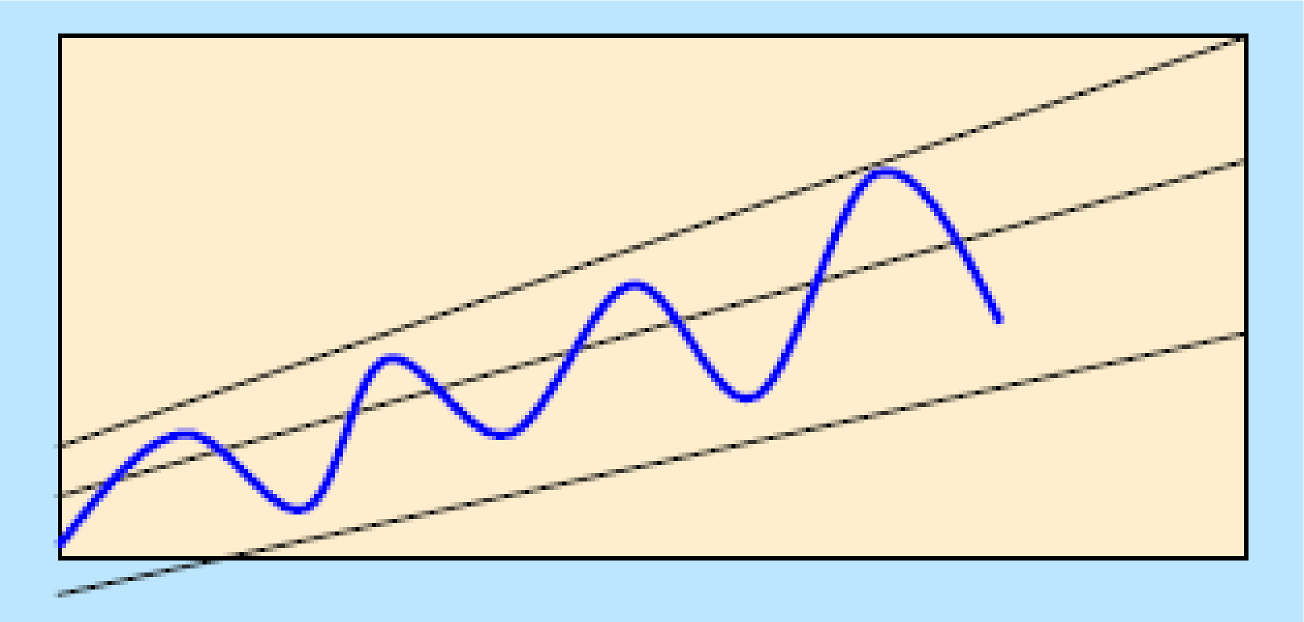
Παράδειγμα αυξανόμενης τάσης με σταθερά ποσά εποχικότητας (προσθετική εποχική μεταβλητότητα).



Στην περίπτωση της **πολλαπλασιαστικής εποχικής μεταβλητότητας**, η τάση πολλαπλασιάζεται με τους παράγοντες εποχικότητας.

**Πρόβλεψη = Τάση * Παράγοντας Εποχικότητας**

Παράδειγμα αυξανόμενης εποχικής διακύμανσης καθώς αυξάνεται η τάση.



Με την ανάλυση μιας χρονοσειράς βρίσκονται τα βασικά στοιχεία της σειράς όπως η τάση και η εποχικότητα / κυκλικότητα.

Υπολογίζονται δείκτες για την εποχικότητα / κυκλικότητα.

Στη συνέχεια η διαδικασία πρόβλεψης βρίσκει την τάση και την προσαρμόζει με βάση τους δείκτες εποχικότητας και κυκλικότητας, οι οποίοι έχουν καθοριστεί από την ανάλυση της χρονοσειράς.

Πιο αναλυτικά, η διαδικασία είναι η ακόλουθη:

**Ανάλυση της χρονοσειράς στα στοιχεία της**

α) Εύρεση του στοιχείου εποχικότητας

β) Αφαίρεση του στοιχείου εποχικότητας (deseasonalize) από τη ζήτηση

γ) Εύρεση του στοιχείου τάσης

**Πρόβλεψη των μελλοντικών τιμών του κάθε στοιχείου**

α) Πρόβλεψη του στοιχείου τάσης στο μέλλον

β) Πολλαπλασιασμός του στοιχείου τάσης με το στοιχείο εποχικότητας

**Παράδειγμα** Ανάλυση Χρονοσειράς

Έστω ότι τα τελευταία χρόνια, μια εταιρεία πούλησε κατά μέσο όρο 1000 κομμάτια ενός συγκεκριμένου προϊόντος ετησίως. Κατά μέσο όρο, 200 κομμάτια πουλήθηκαν την άνοιξη, 350 το καλοκαίρι, 300 το φθινόπωρο και 150 το χειμώνα. Ο εποχικός παράγοντας είναι ο λόγος του ποσού που πουλήθηκε κάθε εποχή δια το μέσο όρος πωλήσεων της εποχής.

Μέσος όρος πωλήσεων 4 εποχών:

$$\frac{1000}{4} = 250$$

| | Πωλήσεις | Μέσος Όρος | Εποχικός Παράγοντας | |
|---|---|---|---|---|
| **Άνοιξη** | 200 | 250 | 200/250= | 0,8 |
| **Καλοκαίρι** | 350 | 250 | 350/250= | 1,4 |
| **Φθινόπωρο** | 300 | 250 | 300/250= | 1,2 |
| **Χειμώνας** | 150 | 250 | 150/250= | 0,6 |
| **Άθροισμα** | 1000 | | | |

Με βάση τα στοιχεία αυτά, αν η αναμενόμενη ζήτηση για τον επόμενο χρόνο είναι 1100 κομμάτια, προβλέπουμε ότι η ζήτηση κάθε εποχή για τον επόμενο χρόνο θα είναι ως εξής:

Μέσος όρος πωλήσεων 4 εποχών:

$$\frac{1100}{4} = 275$$

**Πρόβλεψη = Τάση * Παράγοντας Εποχικότητας**

| | Πωλήσεις | Μέσος Όρος | Εποχικός Παράγοντας | Πρόβλεψη |
|---|---|---|---|---|
| Άνοιξη | | 275 | 0,8 | 220 |
| Φθινόπωρο | | 275 | 1,4 | 385 |
| Καλοκαίρι | | 275 | 1,2 | 330 |
| Χειμώνας | | 275 | 0,6 | 165 |
| Άθροισμα | 1100 | | | |

Πρόβλεψη = $275 * 0.8 = 220$

Πρόβλεψη = $275 * 1.4 = 385$

Πρόβλεψη = $275 * 1.2 = 330$

Πρόβλεψη = $275 * 0.6 = 165$

Ο εποχικός παράγοντας είναι δυνατόν να ανανεώνεται περιοδικά καθώς θα γίνονται διαθέσιμα καινούρια στοιχεία.

## Μέθοδος αυτοπαλινδρόμησης

Ένα από τα χαρακτηριστικά των τυχαίων χρονοσειρών είναι ότι οι παρατηρήσεις είναι πιθανολογικά ανεξάρτητες. Πολλές χρονοσειρές δεν ικανοποιούν το κριτήριο αυτό καθώς κάθε παρατήρηση εξαρτάται από τις προηγούμενές της. Η ιδιότητα αυτή ονομάζεται **αυτοσυσχέτιση**.

Για παράδειγμα, στη συχνότερη περίπτωση αυτοσυσχέτισης, τη θετική αυτοσυσχέτιση, μεγάλες τιμές τείνουν να ακολουθούνται από μεγάλες τιμές και το αντίστροφο.

Για την ανάλυση αυτόσυσχετιζόμενων χρονοσειρών χρησιμοποιούνται οι τιμές υστέρησης (lag values), οι οποίες προκύπτουν με μετατόπιση των τιμών της χρονοσειράς κατά ένα αριθμό θέσεων προς τα εμπρός.

Ένα μέτρο της αυτοσυσχέτισης είναι ο συντελεστής αυτοσυσχέτισης, ο οποίος για περίοδο υστέρησης $k$ και για μια χρονοσειρά με παρατηρήσεις από $t = 1$ έως $t = \mathrm{T}$ ορίζεται ως:

$$\boxed{r_k = \frac{(Y_{k+1} - \bar{Y})(Y_1 - \bar{Y}) + \cdots + (Y_T - \bar{Y})(Y_{T-k} - \bar{Y})}{(Y_1 - \bar{Y})^2 + (Y_2 - \bar{Y})^2 + \cdots + (Y_T - \bar{Y})^2}}$$

Ένας πρακτικός κανόνας είναι να θεωρείται ένας συντελεστής αυτοσυσχέτισης σημαντικός όταν η απόλυτη τιμή του είναι τουλάχιστον διπλάσια από το τυπικό σφάλμα.

Στις περιπτώσεις που κάποιος (ή κάποιοι) συντελεστής αυτοσυσχέτισης είναι σημαντικός μπορεί να χρησιμοποιηθεί ένα μοντέλο αυτόπαλινδρόμησης.

## ΠΡΟΒΛΕΨΗ ΜΕ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

### Προβλέψεις με μοντέλα AR(1)

Έστω το μοντέλο $\boxed{\Upsilon_t = \delta + \alpha_1 \Upsilon_{t-1} + \varepsilon_t}$

Για $\mathrm{T} = \mathrm{T} + 1$: $\boxed{\Upsilon_{\mathrm{T}+1} = \delta + \alpha_1 \Upsilon_{\mathrm{T}} + \varepsilon_{\mathrm{T}+1}}$

Αν οι παράμετροι $\delta$ και $\alpha_1$ είναι γνωστές, τότε με βάση τις πληροφορίες μέχρι την περίοδο $\mathrm{T}$, **μια πρόβλεψη για την περίοδο $\mathbf{T+1}$ είναι η υπό συνθήκη προσδοκώμενη τιμή της $\boldsymbol{\Upsilon_{\mathrm{T}+1}}$**.

Δηλαδή αν $\widehat{\Upsilon}_{\mathrm{T}+1}$ παριστάνει την πρόβλεψη, τότε:

$$\widehat{\Upsilon}_{\mathrm{T}+1} = \mathrm{E}(\Upsilon_{\mathrm{T}+1} | \pi\lambda\eta\rho o\varphi o\rho\iota\varepsilon\varsigma\ \mu\varepsilon\chi\rho\iota\ \tau\eta\nu\ \pi\varepsilon\rho\iota o\delta o\ \mathrm{T})$$

ή $$\widehat{\Upsilon}_{\mathrm{T}+1} = \mathrm{E}(\Upsilon_{\mathrm{T}+1} | \Upsilon_{\mathrm{T}}, \Upsilon_{\mathrm{T}-1}, \ldots, \Upsilon_1)$$

ή $$\widehat{\Upsilon}_{\mathrm{T}+1} = \mathrm{E}_{\mathrm{T}}(\Upsilon_{\mathrm{T}+1})$$

Γενικά $\boxed{\widehat{\Upsilon}_{T+h} = E_T(\Upsilon_{T+h})}$

είναι μια πρόβλεψη $h$ περιόδους μπροστά με βάση τις πληροφορίες μέχρι την περίοδο Τ.

Η παραπάνω πρόβλεψη **ελαχιστοποιεί** τον **μέσο του τετραγώνου του σφάλματος (Mean Square Error, MSE)**, δηλαδή ελαχιστοποιεί την σχέση:

$$\boxed{E(\widehat{\Upsilon}_{T+h} - \Upsilon_{T+h})^2}$$

Η πρόβλεψη αυτή θεωρείται ***άριστη (optimal forecast).***

Από την σχέση $\boxed{\Upsilon_{T+1} = \delta + \alpha_1 \Upsilon_T + \varepsilon_{T+1}}$

Έχουμε

$$E_T(\Upsilon_{T+1}) = E_T(\delta + \alpha_1 \Upsilon_T + \varepsilon_{T+1})$$

$$\boxed{E_T(\Upsilon_{T+1}) = \delta + \alpha_1 E_T(\Upsilon_T) + E_T(\varepsilon_{T+1})}$$

- Μέχρι την περίοδο $\boxed{\mathrm{t} = \mathrm{T}}$, κάθε όρος με δείκτη $\mathrm{t} \leq \mathrm{T}$ είναι γνωστός και
  - $\mathrm{E_T(\Upsilon_T) = \Upsilon_T}$ και
  - $\mathrm{E_T(\varepsilon_T) = \varepsilon_T}$.
- Για $\boxed{\mathrm{t} > \mathrm{T}}$ όμως είναι:
  - $\mathrm{E_T(\varepsilon_t) = E(\varepsilon_t) = 0}$.

Με βάση τα παραπάνω, η σχέση $\boxed{\mathrm{E_T(\Upsilon_{T+1}) = \delta + \alpha_1 E_T(\Upsilon_T) + E_T(\varepsilon_{T+1})}}$ γίνεται:

$$\boxed{\mathrm{E_T(\Upsilon_{T+1}) = \delta + \alpha_1 \Upsilon_T}}$$

Και για το σφάλμα πρόβλεψης ισχύει: $\hat{\varepsilon}_{\mathrm{T+1}} = \Upsilon_{\mathrm{T+1}} - \widehat{\Upsilon}_{\mathrm{T+1}}$, όπου

- $\hat{\varepsilon}_{\mathrm{T+1}} = \varepsilon_{\mathrm{T+1}}$ και
- $\mathrm{Var}(\hat{\varepsilon}_{\mathrm{T+1}}) = \mathrm{Var}(\varepsilon_{\mathrm{T+1}}) = \sigma^2$

- Για πρόβλεψη στην περίοδο $\boxed{\mathrm{t} = \mathrm{T} + 2}$, έχουμε:
  - $\widehat{\Upsilon}_{\mathrm{T}+2} = \delta + \alpha_1 \widehat{\Upsilon}_{\mathrm{T}+1}$
  - $\hat{\varepsilon}_{\mathrm{T}+2} = \varepsilon_{\mathrm{T}+2} + \alpha_1 \hat{\varepsilon}_{\mathrm{T}+1} = \varepsilon_{\mathrm{T}+2} + \alpha_1 \varepsilon_{\mathrm{T}+1}$
  - $\mathrm{Var}(\hat{\varepsilon}_{\mathrm{T}+2}) = \sigma^2(1 + \alpha_1^2)$

- Γενικά για πρόβλεψη h περιόδους μπροστά, δηλαδή στην περίοδο $\boxed{\mathrm{t} = \mathrm{T} + \mathrm{h}}$, έχουμε:
  - $\widehat{\Upsilon}_{\mathrm{T}+\mathrm{h}} = \delta + \alpha_1 \widehat{\Upsilon}_{\mathrm{T}+\mathrm{h}-1}$
  - $\mathrm{Var}(\hat{\varepsilon}_{\mathrm{T}+\mathrm{h}}) = \sigma^2(1 + \alpha_1^2 + \alpha_1^4 + \cdots + \alpha_1^{2(\mathrm{h}-1)})$

- Η διακύμανση του σφάλματος πρόβλεψης αυξάνει μη γραμμικά καθώς μεγαλώνει η περίοδος πρόβλεψης.
- Η πρόβλεψη συγκλίνει προς το μέσο:

$$\lim_{h\to\infty} \widehat{\Upsilon}_{\mathrm{T+h}} = \frac{\delta}{1-\alpha_1} = \mu$$

## Προβλέψεις με μοντέλα MA(1)

Έστω το μοντέλο $\boxed{\Upsilon_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}}$

- Η πρόβλεψη την περίοδο $\mathrm{T} + 1$ είναι:

$$\widehat{\Upsilon}_{\mathrm{T}+1} = \mathrm{E}_{\mathrm{T}}(\mu + \varepsilon_{\mathrm{T}+1} + \theta_1 \varepsilon_{\mathrm{T}}) \quad \text{ή}$$

$$\widehat{\Upsilon}_{\mathrm{T}+1} = \mu + \theta_1 \varepsilon_{\mathrm{T}}$$

  - Το σφάλμα πρόβλεψης είναι:

$$\hat{\varepsilon}_{\mathrm{T}+1} = \Upsilon_{\mathrm{T}+1} - \widehat{\Upsilon}_{\mathrm{T}+1} = \varepsilon_{\mathrm{T}+1}$$

  - Η διακύμανση του σφάλματος πρόβλεψης είναι:

$$\mathrm{Var}(\hat{\varepsilon}_{\mathrm{T}+1}) = \mathrm{Var}(\varepsilon_{\mathrm{T}+1}) = \sigma^2$$

Είναι
$E_T(\varepsilon_{T+2}) = E_T(\varepsilon_{T+1}) = 0$

- Η πρόβλεψη την περίοδο $T+2$ είναι:

$\widehat{Y}_{T+2} = E_T(Y_{T+2}) = E_T(\mu + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1})$ ή

$\widehat{Y}_{T+2} = \mu$

- Το σφάλμα πρόβλεψης είναι:

$\hat{\varepsilon}_{T+2} = Y_{T+2} - \widehat{Y}_{T+2} = \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1}$

- Η διακύμανση του σφάλματος πρόβλεψης είναι:

$Var(\hat{\varepsilon}_{T+2}) = Var(\varepsilon_{T+2} + \theta_1 \varepsilon_{T+1}) = \sigma^2(1 + \theta_1^2)$

- Γενικά, η πρόβλεψη την περίοδο $T+h$ με $h > 1$ είναι:

$\widehat{Y}_{T+h} = \mu$ και

$Var(\hat{\varepsilon}_{T+h}) = \sigma^2(1 + \theta_1^2)$

Από την σχέση

$\widehat{Y}_{T+h} = \mu, \text{για } h > 1$

είναι φανερό ότι ένα υπόδειγμα MA(1) είναι κατάλληλο μόνο για πρόβλεψη μια περίοδο μπροστά, αφού για $h > 1$ η πρόβλεψη είναι πάντα ο μέσος όρος.

Η πρόβλεψη για μια περίοδο μπροστά: $\widehat{Y}_{T+1} = \mu + \theta_1 \varepsilon_T$ εξαρτάται από την τιμή του τυχαίου όρου ε στην περίοδο $T$.

Επειδή η τιμή αυτή είναι άγνωστη, στην πράξη αντικαθίσταται από την εκτίμηση της, η οποία προκύπτει από τις προηγούμενες τιμές.

- Η πρόβλεψη της $\Upsilon$ για ένα βήμα εμπρός, δηλ. την περίοδο $t = 1$ είναι:

$\Upsilon_1 = \mu + \varepsilon_1 + \theta_1\varepsilon_0$

Η πρόβλεψη είναι $\widehat{\Upsilon}_1 = E_0(\Upsilon_1) = \mu + \theta_1\varepsilon_0$ $\quad\quad (E_0(\varepsilon_1) = 0)$

Η τιμή $\varepsilon_0$ αναφέρεται στην περίοδο πριν από το δείγμα, και συνήθως δεν είναι γνωστή, οπότε αντικαθίσταται από την προσδοκώμενη τιμή της που είναι μηδέν.

Επομένως:

- $\widehat{\Upsilon}_1 = \mu$
- $\hat{\varepsilon}_1 = \Upsilon_1 - \widehat{\Upsilon}_1 = \Upsilon - \mu$

- Η πρόβλεψη της $\Upsilon$ για την περίοδο $\mathrm{t} = 2$ είναι:

$\Upsilon_2 = \mu + \theta_1 \varepsilon_1 = \mu + \theta_1 \hat{\varepsilon}_1$ και

$\hat{\varepsilon}_2 = \Upsilon_2 - \widehat{\Upsilon}_2$

κ.ο.κ

Συνεχίζοντας με τον ίδιο τρόπο, εκτιμάμε όλα τα $\varepsilon$, δηλαδή τα $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$,…, $\hat{\varepsilon}_{\mathrm{T}}$.

Έτσι έχουμε την εκτίμηση της $\varepsilon_{\mathrm{T}}$ για να μπορέσουμε να βρούμε την πρόβλεψη στην περίοδο $\mathrm{T} + 1$, κ.ο.κ.

## Προβλέψεις με μοντέλα ARMA

Έστω το μοντέλο **ARMA(1,1)**: $\boxed{\Upsilon_t = \delta + \alpha_1 \Upsilon_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}}$

- Η άριστη πρόβλεψη την περίοδο $t = T + 1$ είναι:

$$\widehat{\Upsilon}_{T+1} = E_T(\delta + \alpha_1 \Upsilon_T + \varepsilon_{T+1} + \theta_1 \varepsilon_T) \quad \text{ή}$$
$$\widehat{\Upsilon}_{T+1} = \delta + \alpha_1 \Upsilon_T + \theta_1 \varepsilon_T$$

- Το σφάλμα πρόβλεψης είναι:

$$\hat{\varepsilon}_{T+1} = \Upsilon_{T+1} - \widehat{\Upsilon}_{T+1} = \varepsilon_{T+1}$$

- Η διακύμανση του σφάλματος πρόβλεψης είναι:

$$\mathrm{Var}(\hat{\varepsilon}_{T+1}) = \mathrm{Var}(\varepsilon_{T+1}) = \sigma^2$$

Με τον ίδιο τρόπο μπορούμε να κάνουμε προβλέψεις για 2,3 και γενικά $h$ περιόδους μπροστά.

## Προβλέψεις με μοντέλα ARIMA

Έστω το αυτοπαλίνδρομο ολοκληρωμένο μοντέλο πρώτης τάξης ARIMA(1,1,0):

$\boxed{w_t = \delta + \alpha_1 w_{t-1} + \varepsilon_t}$ όπου $\boxed{w_t = \Delta\Upsilon_t = \Upsilon_t - \Upsilon_{t-1}}$

Δηλαδή η μη στάσιμη σειρά $\{\Upsilon_t\}$ γίνεται στάσιμη παίρνοντας τις πρώτες διαφορές, οι οποίες είναι AR(1).

Από το μοντέλο $w_t = \delta + \alpha_1 w_{t-1} + \varepsilon_t$, μπορεί να γίνει πρόβλεψη της διαφοράς $w$ και στη συνέχεια πρόβλεψη της $\Upsilon$.

- Η άριστη πρόβλεψη της $w$ την περίοδο $\mathrm{t} = \mathrm{T} + 1$ είναι:

$$\widehat{w}_{\mathrm{T}+1} = \mathrm{E}_{\mathrm{T}}(\delta + \alpha_1 w_{t-1} + \varepsilon_t) \quad \text{ή}$$

$$\widehat{\Upsilon}_{\mathrm{T}+1} = \delta + \alpha_1 w_{\mathrm{T}}$$

- Η πρόβλεψη για την αρχική χρονοσειρά $\Upsilon$ για την περίοδο $t = T + 1$ είναι:

$$\widehat{\Upsilon}_{T+1} = \Upsilon_T + \widehat{w}_{T+1}$$
$$= \Upsilon_T + \delta + \alpha_1 w_T$$
$$= \Upsilon_T + \delta + \alpha_1(\Upsilon_T - \Upsilon_{T-1})$$
$$= \delta + (1 + \alpha_1)\Upsilon_T - \alpha_1 \Upsilon_{T-1}$$

- Γενικά, η πρόβλεψη για $h$ περιόδους μπροστά είναι:

$$\widehat{\Upsilon}_{T+h} = \Upsilon_T + \widehat{w}_{T+1} + \widehat{w}_{T+2} + \cdots + \widehat{w}_{T+h}$$

## Διάστημα εμπιστοσύνης

Γύρω από την προβλεπόμενη τιμή μπορεί να διατυπωθεί ένα διάστημα εμπιστοσύνης με βάση τη διακύμανση του σφάλματος πρόβλεψης.

Αν παραστήσουμε με $\sigma_{\mathrm{h}}^2$ τη διακύμανση του σφάλματος πρόβλεψης, το $1-\alpha$ διάστημα εμπιστοσύνης για την αληθινή τιμή της χρονοσειράς $\mathrm{h}$ περιόδους μπροστά είναι:

$$\boxed{\widehat{\mathrm{Y}}_{\mathrm{T+h}} - \mathrm{z}_{\frac{\alpha}{2}}\sigma_{\mathrm{h}} \leq Y_{\mathrm{T+h}} \leq \widehat{\mathrm{Y}}_{\mathrm{T+h}} + \mathrm{z}_{\frac{\alpha}{2}}\sigma_{\mathrm{h}}}$$

Η διακύμανση είναι συνάρτηση της συνάρτησης διακύμανσης του διαταρακτικού όρου $\varepsilon_{\mathrm{t}}$, η οποία είναι άγνωστη. Οπότε, πρέπει να γίνει εκτίμηση της από το άθροισμα των τετραγώνων των καταλοίπων:

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_t^2}{\mathrm{T}-\mathrm{p}-\mathrm{q}}}$$

## Αξιολόγηση προβλέψεων

Το διάστημα εμπιστοσύνης της πρόβλεψης είναι ένα καλό μέτρο αξιολόγησης της πρόβλεψης το οποίο μας δείχνει την ακρίβεια της πρόβλεψης.
Μπορούμε επίσης να συγκρίνουμε τις προβλεπόμενες τιμές με τις πραγματικές τιμές για την περίοδο για την οποία έχουμε στοιχεία. Για τη σύγκριση αυτή, δηλαδή για την αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου, έχουν προταθεί διάφορα κριτήρια, από τα οποία τα κυριότερα είναι τα ακόλουθα:

**α) Ρίζα του μέσου του τετραγώνου του σφάλματος (root mean square error)**

$$\boxed{\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^f - Y_t^a)^2}}$$

M: αριθμός χρονικών περιόδων
$Y_t^a$: παρατηρούμενη τιμή
$Y_t^f$: προβλεπόμενη τιμή

## β) Μέσο απόλυτο σφάλμα (mean absolute error)

$$\mathrm{MAE} = \frac{1}{\mathrm{M}}\sqrt{\sum_{t=1}^{\mathrm{M}} |Y_t^f - Y_t^a|}$$

M: αριθμός χρονικών περιόδων
$Y_t^a$: παρατηρούμενη τιμή
$Y_t^f$: προβλεπόμενη τιμή

## γ) Μέσο απόλυτο ποσοστιαίο σφάλμα (mean absolute percentage error)

$$\mathrm{MAPE} = \frac{1}{\mathrm{M}}\sqrt{\sum_{t=1}^{\mathrm{M}} \left|\frac{Y_t^f - Y_t^a}{Y_t^a}\right|}$$

## δ) Συντελεστής ανισότητας του Theil (Theil's inequality coefficient)

$$\mathrm{U} = \sqrt{\frac{\frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^f - Y_t^a)^2}{\frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^a)^2}}$$

Ο συντελεστής $\mathrm{U}$ είναι ανεξάρτητος από τις μονάδες μέτρησης και για αυτό είναι περισσότερο κατάλληλος για τη σύγκριση μοντέλων, σε αντίθεση με τα άλλα κριτήρια που εξαρτώνται από τις μονάδες μέτρησης.

- Αν οι προβλεπόμενες τιμές συμπίπτουν απολύτως με τις πραγματικές, η τιμή του $\mathrm{U}$ είναι μηδέν.
- Όταν $\mathrm{U} > 1$, οι προβλέψεις είναι πολύ κακές.
- Αν $\mathrm{U} = 1$, τότε όλες οι προβλέψεις είναι μηδέν. Η περίπτωση αυτή έχει περισσότερο νόημα όταν για τον υπολογισμό του $\mathrm{U}$ χρησιμοποιούνται οι μεταβολές από την προηγούμενη περίοδο και όχι οι αρχικές τιμές. Σε αυτήν την περίπτωση, $\mathrm{U} = 1$ σημαίνει ότι οι προβλεπόμενες μεταβολές είναι μηδέν, δηλαδή συνέχιση της υπάρχουσας κατάστασης.

Ο συντελεστής U μπορεί να διασπαστεί σε 3 συνιστώσες, κάθε μια από τις οποίες εκφράζει μια πηγή ή αιτία της ανακρίβειας των προβλέψεων. Η διάσπαση μπορεί να γίνει ως εξής: Υψώνουμε στο τετράγωνο και τα δύο μέρη του συντελεστή U, οπότε:

$$\boxed{\mathrm{U}^2 = \frac{\frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^f - Y_t^a)^2}{\frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^a)^2}}$$

ή ισοδύναμα

$$\boxed{\mathrm{U}^2 = \frac{(\bar{Y}^f - \bar{Y}^a)^2}{\mathrm{A}} + \frac{(\sigma^f - \sigma^a)^2}{\mathrm{A}} + \frac{2(1-\rho)\sigma^f\sigma^\alpha}{\mathrm{A}}}$$

όπου $\mathrm{A} = \frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^a)^2$

Ο 3ος όρος της σχέσης του $\mathrm{U}^2$ παριστάνει το μη συστηματικό (τυχαίο) παράγοντα που δεν μπορεί να αποφευχθεί, ενώ οι δύο πρώτοι όροι παριστάνουν συστηματικά σφάλματα που πρέπει να αποφεύγονται.

Ο αριθμητής στην αρχική σχέση του $U^2$ είναι το κριτήριο MSE, το οποίο μπορεί να διασπαστεί ως εξής:

$$\boxed{\mathrm{MSE} = \frac{1}{\mathrm{M}}\left[\left(Y_t^f - \bar{Y}^f\right) - \left(Y_t^a - \bar{Y}^a\right) + \left(\bar{Y}^f - \bar{Y}^a\right)\right]^2}$$

ή ισοδύναμα

$$\boxed{\mathrm{MSE} = \underbrace{\left(\bar{Y}^f - \bar{Y}^a\right)^2}_{\text{μέτρο της μεροληψίας (bias)}} + \underbrace{\left(\sigma^f - \sigma^a\right)^2}_{\text{μέτρο της άνισης μεταβλητότητας των } \Upsilon^f \text{ και } \Upsilon^\alpha} + \underbrace{2(1-\rho)\sigma^f\sigma^\alpha}_{\text{μέτρο της ατελούς συμμεταβλητότητας των } \Upsilon^f \text{ και } \Upsilon^\alpha}}$$

$\bar{Y}^f$: μέσος των προβλεπόμενων τιμών, $\bar{Y}^\alpha$: μέσος των πραγματικών τιμών

$\sigma^f$: τυπική απόκλιση των $\Upsilon^f$, $\sigma^\alpha$: τυπική απόκλιση των $\Upsilon^\alpha$

$\rho$: συντελεστής συσχέτισης των $\Upsilon^f$ και $\Upsilon^\alpha$

Η σχέση $\boxed{\mathrm{MSE} = (\bar{Y}^f - \bar{Y}^a)^2 + (\sigma^f - \sigma^a)^2 + 2(1-\rho)\sigma^f \sigma^\alpha}$

γράφεται ως:

$$\boxed{1 = \frac{(\bar{Y}^f - \bar{Y}^a)^2}{\mathrm{MSE}} + \frac{(\sigma^f - \sigma^a)^2}{\mathrm{MSE}} + \frac{2(1-\rho)\sigma^f \sigma^\alpha}{\mathrm{MSE}}}$$

**Άσκηση 1**

Έστω ένα μοντέλο AR(1) με συντελεστή $\alpha_1 = 0.6$, μέσο όρο $\mu = 9$ και διακύμανση του λευκού θορύβου $\sigma^2 = 0.1$ και η τελευταία παρατήρηση είναι $\Upsilon_{100} = 8.9$. Θέλουμε να προβλέψουμε τις τιμές $\Upsilon_{101}, \Upsilon_{102}$ και να βρούμε τα 95% διαστήματα εμπιστοσύνης των προβλέψεων.

Λύση

Το μοντέλο AR(1) γράφεται: $\Upsilon_t = \delta + 0.6\Upsilon_{t-1} + \varepsilon_t$ και

$$\mu = \frac{\delta}{1-\alpha_1} \Leftrightarrow 9 = \frac{\delta}{1-0.6} \Leftrightarrow \delta = 9(1-0.6) \Leftrightarrow \delta = 3.6$$

Η πρόβλεψη h περιόδους μπροστά είναι: $\widehat{\Upsilon}_{\mathrm{T+h}} = \delta + \alpha_1 \widehat{\Upsilon}_{\mathrm{T+h-1}}$

Η διακύμανση του σφάλματος πρόβλεψης είναι:

$$\mathrm{Var}(\hat{\varepsilon}_{\mathrm{T+h}}) = \sigma^2(1 + \alpha_1^2 + \alpha_1^4 + \cdots + \alpha_1^{2(\mathrm{h}-1)})$$

Οπότε:

$$\widehat{\Upsilon}_{T+1} = \delta + \alpha_1 \Upsilon_T$$

$$\widehat{\Upsilon}_{101} = 3.6 + 0.6 \times 8.9 = 8.94$$

$$\sigma_1^2 = \mathrm{Var}(\hat{\varepsilon}_{101}) = \sigma^2 = 0.1$$

$$\sigma_1 = \sqrt{0.1} = 0.316$$

$$\widehat{\Upsilon}_{T+2} = \delta + \alpha_1 \widehat{\Upsilon}_{T+1}$$

$$\widehat{\Upsilon}_{102} = 3.6 + 0.6 \times 8.94 = 8.964$$

$$\sigma_2^2 = \mathrm{Var}(\hat{\varepsilon}_{102}) = \sigma^2(1 + \alpha_1^2) = 0.1(1 + 0.6^2)$$

$$\sigma_2 = \sqrt{0.1(1 + 0.6^2)} = 0.368$$

Τα 95% διαστήματα εμπιστοσύνης για κάθε μια πρόβλεψη προκύπτει από την σχέση:

$$\widehat{Y}_{T+h} - z_{\frac{\alpha}{2}}\sigma_h \leq Y_{T+h} \leq \widehat{Y}_{T+h} + z_{\frac{\alpha}{2}}\sigma_h$$

$\sigma_h$: τυπική απόκλιση σφάλματος πρόβλεψης

Οπότε:

$$\widehat{Y}_{101} - 1.96 \times \sigma_1 \leq Y_{101} \leq \widehat{Y}_{101} + 1.96 \times \sigma_1$$

$$8.94 - 1.96 \times \sqrt{0.1} \leq Y_{101} \leq 8.94 + 1.96 \times \sqrt{0.1}$$

$$8.320 \leq Y_{101} \leq 9.560$$

και

$$\widehat{Y}_{102} - 1.96 \times \sigma_2 \leq Y_{102} \leq \widehat{Y}_{102} + 1.96 \times \sigma_2$$

$$8.964 - 1.96 \times 0.368 \leq Y_{102} \leq 8.964 + 1.96 \times 0.368$$

$$8.241 \leq Y_{102} \leq 9.687$$

## Άσκηση 2

Ο πίνακας δείχνει στοιχεία για τις μηνιαίες πωλήσεις μιας εταιρείας και τις προβλέψεις που είχε κάνει για τις μηνιαίες πωλήσεις της.

Πως μπορεί να αξιολογήσει η εταιρεία την απόδοση του μοντέλου πρόβλεψης που χρησιμοποίησε;

| Μήνας | Πωλήσεις | Προβλέψεις | Απόλυτο Σφάλμα | Τετράγωνο Σφάλματος |
|---|---|---|---|---|
| 1 | 220 | - | - | - |
| 2 | 250 | 255 | 5 | 25 |
| 3 | 210 | 205 | 5 | 25 |
| 4 | 300 | 320 | 20 | 400 |
| 5 | 325 | 315 | 10 | 100 |
| 6 | 340 | 320 | 20 | 400 |
| 7 | 350 | 340 | 10 | 100 |
| 8 | 310 | 350 | 40 | 1600 |
| 9 | 300 | 290 | 10 | 100 |
| 10 | 250 | 260 | 10 | 100 |
| | | Άθροισμα = | 130 | 2850 |

**Λύση**

Προκειμένου να υπολογίσει την απόδοση του μοντέλου πρόβλεψης που χρησιμοποιεί η εταιρεία, έγιναν οι παρακάτω υπολογισμοί.

**Μέση Απόλυτη Απόκλιση**

$$MAE = \frac{1}{\mathrm{M}}\sqrt{\sum_{t=1}^{\mathrm{M}} |Y_t^f - Y_t^a|} = \frac{1}{\mathrm{M}}\sqrt{\sum_{t=1}^{\mathrm{M}} |\varepsilon_t|} = \frac{1}{9}\times 130 = 14.4$$

**Μέσο Τετραγωνικό Σφάλμα**

$$\mathrm{MSE} = \frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}}(Y_t^f - Y_t^a)^2 = \frac{1}{\mathrm{M}}\sum_{t=1}^{\mathrm{M}} {\varepsilon_t}^2 = \frac{1}{9}\times 2850 = 316.66$$

**Τυπική Απόκλιση Σφαλμάτων**

$$s = \sqrt{\frac{\sum_{t=1}^{\mathrm{M}} {\varepsilon_t}^2}{M-1}} = \sqrt{\frac{2850}{9-1}} = 18.87$$