ΙΕΚ Ξάνθης – Ειδικότητα Μηχανογραφημένου Λογιστηρίου

Σημειώσεις για το μάθημα Στατιστική ΙΙ

Επαμεινώνδας Διαμαντόπουλος - Μαθηματικός

0.1 Εισαγωγή

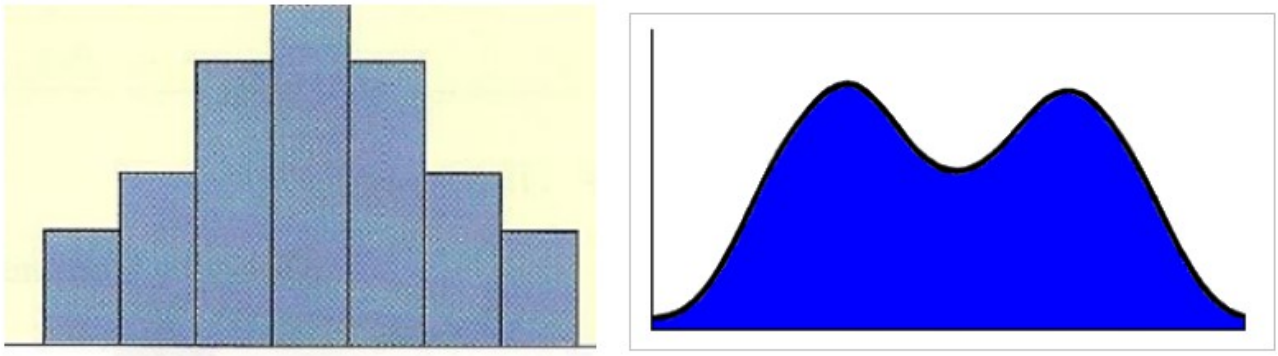
Οι σημειώσεις αυτές αποτελούν τη φυσική συνέχεια των σημειώσεων που δόθηκαν στο πρώτο εξάμηνο σπουδών. Μελετούνται κατά σειρά τα εξής θέματα :

1. Ασυμμετρία, κύρτωση και εφαρμογές.
2. Ροπές κατανομής συχνοτήτων.
3. Παλινδρόμηση
   - Γραμμική παλινδρόμηση
   - Μη γραμμική παλινδρόμηση
4. Συνδιακύμανση δύο ποσοτικών μεταβλητών
5. Συντελεστής συσχέτισης Pearson, Spearman και Kendall
6. Συσχέτιση ποιοτικών μεταβλητών. Δοκιμασία X^2.
7. Χρονοσειρές
   - Περιοδικές μεταβολές
   - Κυκλικές μεταβολές
   - Μέθοδοι προσδιορισμού της μακρόχρονης τάσης
   - Εποχιακές κινήσεις.
8. Αριθμοδείκτες
   - Ιδιαίτεροι αριθμοδείκτες
   - Συνθετικοί αριθμοδείκτες
   - Σταθμικοί συνθετικοί αριθμοδείκτες
   - Τιμάριθμος.
9. Ιδανικός τύπος του Fisher.

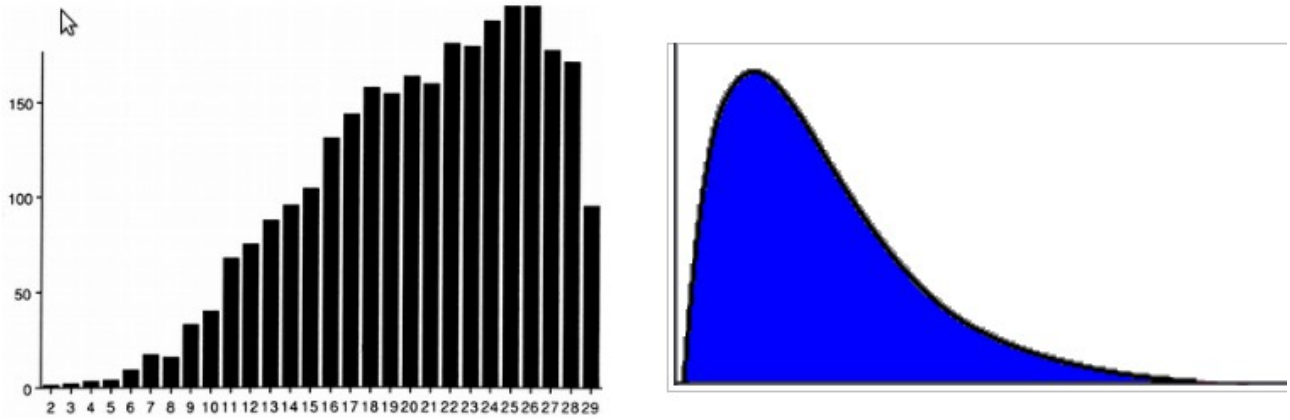
Κεφάλαιο 1 Ασυμμετρία και κυρτότητα

1.1 Ασυμμετρία μίας κατανομής (Skewness).

Μία κατανομή συχνοτήτων (ή σχετικών συχνοτήτων) ονομάζεται συμμετρική όταν είναι φανερό πως υπάρχει ένας κατακόρυφος άξονας ο οποίος λειτουργεί ως καθρέπτης της μισής κατανομής στην άλλη μισή. Χαρακτηριστικό παράδειγμα είναι η κανονική κατανομή χωρίς αυτό να σημαίνει πως δεν μπορεί να είναι συμμετρική μία περισσότερο “ανώμαλη” κατανομή. Στο διάγραμμα 1, (σελίδα 2) παρουσιάζονται κάποια παραδείγματα συμμετρικών κατανομών.



Εικόνα 1: Παραδείγματα συμμετρικών κατανομών. Είναι φανερό πως και στις δύο περιπτώσεις υπάρχει μία κατακόρυφη γραμμή (η οποία ταυτίζεται με τη μέση τιμή αλλά και τη διάμεσο) η οποία χωρίζει σε δύο συμμετρικά μέρη την κατανομή.



Εικόνα 2: Παραδείγματα ασύμμετρων κατανομών.

Σημείωση για την εικόνα 2: Στην αριστερή εικόνα υπάρχουν κάποιες δυσανάλογα μικρές παρατηρήσεις, για τις οποίες η απόσταση από τη μέση τιμή είναι δυσανάλογα μεγάλη και αρνητική, κατά συνέπεια $\gamma < 0$ και λέμε ότι η κατανομή είναι λοξή προς τα αριστερά ή

αρνητικά ασύμμετρη. Αντίθετα, στη δεξιά εικόνα η “ουρά” προς τα δεξιά δίνει θετικό πρόσημο στο συντελεστή ασυμμετρίας (γ > 0) και λέμε ότι η κατανομή είναι λοξή προς τα δεξιά ή θετικά ασύμμετρη.

1.2 Αριθμητική εκτίμηση της ασυμμετρίας μίας κατανομής

Αν $x_{1,} x_{2,} \cdots, x_v$ είναι οι παρατηρήσεις του δείγματός μας τότε η μέση τιμή (ή αριθμητικός μέσος) υπολογίζεται ως εξής $\bar{x} = \frac{1}{v} \sum_{i=1}^{v} x_i$. Ο αριθμητικός μέσος έχει την ιδιότητα να επηρεάζεται δυσανάλογα από τις πολύ μεγάλες ή τις πολύ μικρές παρατηρήσεις του δείγματός μας. Πράγματι, αυτό γίνεται εύκολα κατανοητό αν κάποιος αναλογιστεί πως η πρόσθεση ενός δυσανάλογα μεγάλου αριθμού στον αριθμητή του κλάσματος του ορισμού θα τον αυξήσει υπερβολικά ενώ η αύξηση στον παρονομαστή θα είναι μόνο μία μονάδα! Το γεγονός αυτό αποτελεί και το λόγο που ο συντελεστής ασυμμετρίας ορίζεται ως :

$$\gamma = \frac{\frac{1}{v} \sum_{i=1}^{v} (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{v} \sum_{i=1}^{v} (x_i - \bar{x})^2} \right)^3} = \frac{1}{s^3 v} \sum_{i=1}^{v} (x_i - \bar{x})^3$$

Τύπος 1: Συντελεστής ασυμμετρίας

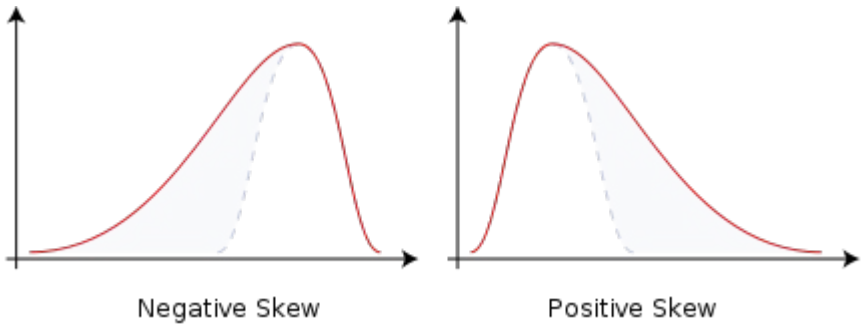
Ο τύπος που δίνει το συντελεστή ασυμμετρίας φαίνεται ιδιαίτερα περίπλοκος, ωστόσο οι παρακάτω παρατηρήσεις κάνουν πιο καθαρό το τοπίο :

1. Ο συντελεστής ασυμμετρίας πρέπει να “ποσοτικοποιεί” το γεγονός ότι κάποιες από παρατηρήσεις απέχουν υπερβολικά προς τη μία μεριά σε σχέση με μέση τιμή της κατανομής.
2. Οι αποστάσεις των παρατηρήσεων από τη μέση τιμή δεν είναι παρά οι διαφορές $x_1 - \bar{x}, x_2 - \bar{x}, \cdots, x_v - \bar{x}$.
3. Το άθροισμα $\sum_{i=1}^{v} (x_i - \bar{x})$ είναι πάντα ίσο με το 0 άρα δεν μπορεί να χρησιμοποιηθεί για την ανίχνευση πιθανής ασυμμετρίας.
4. Το άθροισμα $\sum_{i=1}^{v} (x_i - \bar{x})^2$ ναι μεν εκφράζει το συνολικό μέγεθος της απόστασης των παρατηρήσεων από τη μέση τιμή αλλά δεν δίνει κάποια πληροφορία σχετικά με την συμμετρία της κατανομής καθώς όλοι οι όροι είναι θετικοί.

5. Το άθροισμα $\sum_{i=1}^{v}(x_i-\bar{x})^3$ διατηρεί το πρόσημο των διαφορών ενώ συγκεντρώνει και το μέγεθος της ασυμμετρίας σε μία αριθμητική ποσότητα. Η κανονικοποίηση του δείκτη ασυμμετρίας με τη διαίρεση με το πλήθος v και την τυπική απόκλιση s ακολουθεί με φυσικό τρόπο καθώς ο δείκτης πρέπει να μπορεί να έχει κοινή ερμηνεία σε όλες τις κατανομές ανεξαρτήτως διακύμανσης. Ο λόγος για τον οποίο το στατιστικό 1 αποτελεί το κατάλληλο μέτρο ασυμμετρίας αποδεικνύεται και περισσότερο αυστηρά αλλά η απόδειξη ξεφεύγει από τους σκοπούς αυτών των σημειώσεων.

Ο συντελεστής ασυμμετρίας γ ερμηνεύεται με δύο τρόπους : Ο πρώτος είναι από το πρόσημό του και ο άλλος από το μέγεθος της απόλυτης τιμής του.

Όταν γ > 0 τότε η κατανομή λέγεται λοξή προς τα δεξιά, ενώ όταν γ < 0 τότε η κατανομή λέγεται λοξή προς τα αριστερά (Διάγραμμα 1, σελίδα 4)



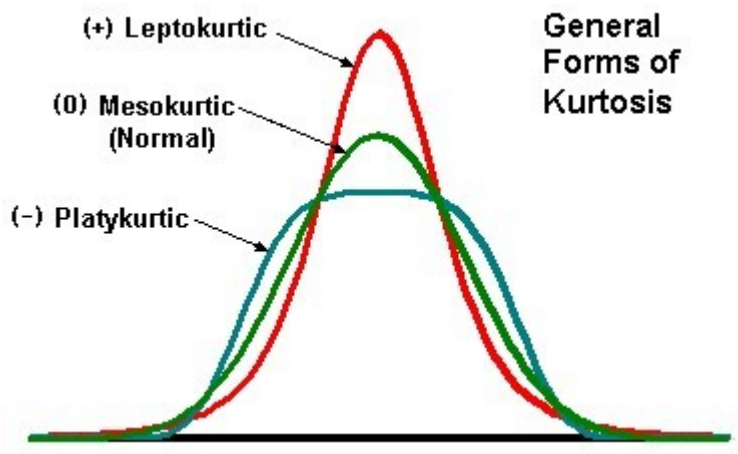
Διάγραμμα 1: Παραδείγματα με γ < ο (αρνητική συμμετρία) και γ > 0 (θετική ασυμμετρία) αντίστοιχα

1.3 Κυρτότητα μίας κατανομής (Kurtosis).

Η κανονική κατανομή είναι το μέτρο και για την κυρτότητα μίας κατανομής. Αν κάποια κατανομή έχει περισσότερο “οξεία” κορυφή από αυτή της κανονικής κατανομής τότε ονομάζεται Λεπτόκυρτη, ενώ όταν έχει περισσότερο “πλατιά” κορυφή τότε ονομάζεται Πλατύκυρτη (Διάγραμμα 2, σελίδα 5). Η κυρτότητα μίας κατανομής υπολογίζεται με το συντελεστή κυρτότητας του Pearson ο οποίος δίνεται από τον εξής τύπο :

$$\alpha=\frac{\frac{1}{v}\sum_{i=1}^{v}(x_i-\bar{x})^4}{\left(\sqrt{\frac{1}{v}\sum_{i=1}^{v}(x_i-\bar{x})^2}\right)^4}=\frac{1}{s^4 v}\sum_{i=1}^{v}(x_i-\bar{x})^4$$

Τύπος 2: Συντελεστής κυρτότητας.



Διάγραμμα 2: Οι τρεις εκδοχές κυρτότητας. Στη λεπτόκυρτη κατανομή είναι α < 3, στη μεσόκυρτη είναι α = 3, ενώ στην πλατύκυρτη είναι α > 3.

Αποδεικνύεται πως στην κανονική κατανομή ο παραπάνω συντελεστής είναι ίσος με τον αριθμό 3. Στην περίπτωση της πλατύκυρτης κατανομής από τον τρόπο που ορίζεται ο συντελεστής κυρτότητας μπορούμε να καταλάβουμε πως θα παίρνει τιμές μεγαλύτερες του 3 (σκεφθείτε το γιατί!) ενώ στην περίπτωση της λεπτόκυρτης κατανομής θα παίρνει τιμές μικρότερες του 3.

Παρατήρησεις

1. Σε πολλά στατιστικά λογισμικά, πριν την εμφάνιση της τιμής της παραμέτρου γίνεται η αφαίρεση με το 3 και ως εκ τούτου η ερμηνεία της τιμής πρέπει να γίνει με ανάλογο τρόπο (α = 0 για την κανονική, α > 0 αν είναι πλατύκυρτη και α < 0 αν είναι λεπτόκυρτη). Ο χρήστης ενός στατιστικού προγράμματος είναι καλό να γνωρίζει τον τρόπο υπολογισμού για να είναι σε θέση να εκτιμήσει και το αποτέλεσμα.

2. Ο προσεκτικός αναγνώστης ίσως πρόσεξε πως ο συντελεστής ασυμμετρίας μπορεί να γραφεί

$$\gamma=\frac{1}{v}\sum_{i=1}^{v}\left(\frac{x_i-\bar{x}}{s}\right)^3=\frac{1}{v}\sum_{i=1}^{v}z_i^3 \quad ,$$

ενώ ο συντελεστής κυρτότητας

$$\alpha=\frac{1}{v}\sum_{i=1}^{v}\left(\frac{x_i-\bar{x}}{s}\right)^4=\frac{1}{v}\sum_{i=1}^{v}z_i^4 \quad ,$$

όπου $z_i=\frac{x_i-\bar{x}}{s}$ οι τυποποιημένες τιμές της μεταβλητής.

3. Αν τα δεδομένα προέρχονται από ένα δείγμα το οποίο θα χρησιμοποιηθεί για την εκτίμηση παραμέτρων του πληθυσμού τότε στους τύπους 1 και 2 (σελίδα 3 και 4 αντίστοιχα), ο λόγος $\frac{1}{v}$ πρέπει να αντικατασταθεί από το $\frac{1}{v-1}$ καθώς αποδεικνύεται πως οι συντελεστές που θα προκύψουν με το νέο παρονομαστή θα αποτελούν καλύτεροι εκτιμητές για τις αντίστοιχες παραμέτρους του πληθυσμού. Για λόγους απλοποίησης όπου εμφανίζονται ανάλογοι τύπου θα εμφανίζεται ο παρονομαστής ν.

1.4 Πότε η συμμετρία και η κυρτότητα της κατανομής διαφέρει σημαντικά από την κανονική;

Ο συντελεστής ασυμμετρίας γ είναι ίσος με 0 όταν οι παρατηρήσεις είναι συμμετρικά τοποθετημένες γύρω από τη μέση τιμή, γεγονός ασυνήθιστο στις πραγματικές έρευνες. Καθώς, ο ερευνητής πρέπει να είναι σε θέση να πιστοποιήσει τη συμμετρία της κατανομής αν θα εφαρμόσει παραμετρικές δοκιμασίες, προκύπτει το ερώτημα : πόσο μεγάλος πρέπει να είναι ο συντελεστής ασυμμετρίας για να θεωρήσουμε την κατανομή μη κανονική;

Μία απάντηση που μπορεί να χρησιμοποιηθεί στην πράξη είναι η εξής : Αν η απόλυτη τιμή του συντελεστή ασυμμετρίας είναι μικρότερη από το διπλάσιο του τυπικού σφάλματος της ασυμμετρίας (Standard Error of Skewness) τότε θεωρείται πως η συμμετρία της κατανομής δεν διαφέρει σημαντικά από την κανονική.

Ανάλογος πρακτικός κανόνας ισχύει και για την κυρτότητα μίας κατανομής : Αν η απόλυτη τιμή του συντελεστή κυρτότητας απέχει από τον αριθμό 3 το πολύ το διπλάσιο του τυπικό σφάλματος της κυρτότητας (Standard Error of Kurtosis) τότε θεωρείται πως η κυρτότητα της κατανομής δεν διαφέρει σημαντικά από την κανονική.

Στην πράξη, οι συντελεστές ασυμμετρίας και κυρτότητας χρησιμοποιούνται για μία πρώτη επισκόπηση της κατανομής. Αν, από την ανάγνωση των συντελεστών δεν προκύπτει καθαρό συμπέρασμα τότε ο ερευνητής πρέπει να προχωρήσει σε άλλες μεθόδους όπως οι δοκιμασίες Kolmogorov – Smirnof και Shapiro – Wilk.

1.5 Οι συντελεστές ασυμμετρίας και κυρτότητας ως στιγμές μίας τυχαίας μεταβλητής.

Η αναμενόμενη τιμή EX (expected value) μίας διακριτής τυχαίας μεταβλητής X ορίζεται να είναι το άθροισμα όλων των πιθανών τιμών που παίρνει η μεταβλητή πολλαπλασιασμένο με την εκάστοτε πιθανότητα να ληφθεί η τιμή αυτή.

Στην πράξη, αν Χ είναι τυχαία μεταβλητή και $x_{1,} x_{2,} \cdots, x_v$ οι διαφορετικές τιμές της, τότε η αναμενόμενη τιμή μίας συνάρτησης του Χ υπολογίζεται ως η μέση τιμή των τιμών της συνάρτησης αυτής στις διαφορετικές τιμές που λαμβάνει η Χ. Για παράδειγμα, σύμφωνα με τα παραπάνω,

$$EX = \frac{x_1 + x_2 + \cdots + x_v}{v} = \bar{x}$$

Η ΕΧ ονομάζεται πρώτη στιγμή της τυχαίας μεταβλητής Χ. Γενικότερα, η κ στιγμή (k-th moment) της τυχαίας μεταβλητής Χ ορίζεται να είναι

$$\kappa_\kappa = E\, X^\kappa$$

ενώ αν μ = ΕΧ είναι η αναμενόμενη (μέση) τιμή τότε η κ-κεντρική στιγμή (k-th central moment) της τυχαίας μεταβλητής Χ ορίζεται να είναι

$$\mu_\kappa = E[(X - EX)^\kappa] = E[(X - \mu)^\kappa]$$

Για παράδειγμα, αν Χ είναι τυχαία μεταβλητή και $x_{1,} x_{2,} \cdots, x_v$ οι διαφορετικές τιμές της τότε

$$E(X-\mu)^2 = \frac{1}{v}\sum_{i=1}^{v}(x_i - \bar{x})^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots (x_v-\bar{x})^2}{v} = s^2$$

Ο αναγνώστης μπορεί εύκολα να επαληθεύσει πως με την ορολογία αυτή, ο τύπος 1, σελίδα 3 του συντελεστή ασυμμετρίας γράφεται

$$\gamma = \frac{1}{s^3 v}\sum_{i=1}^{v}(x_i - \bar{x})^3 = \frac{\mu_3}{s^3}$$

ενώ ο τύπος 2, σελίδα 4 του συντελεστή κυρτότητας γράφεται

$$\alpha = \frac{1}{s^4 v}\sum_{i=1}^{v}(x_i - \bar{x})^4 = \frac{\mu_4}{s^4}$$

Στη βιβλιογραφία ορίζονται και άλλες παραλλαγές των στιγμών μίας τυχαίας μεταβλητής. Αναφέρονται χαρακτηριστικά οι ν-οστές αντίστροφες στιγμές $E(X^{-v})$ και οι ν-οστές λογαριθμικές στιγμές ως $E(\ln^v X)$.

1.6 Ροπογεννήτρια μίας τυχαίας μεταβλητής Χ.

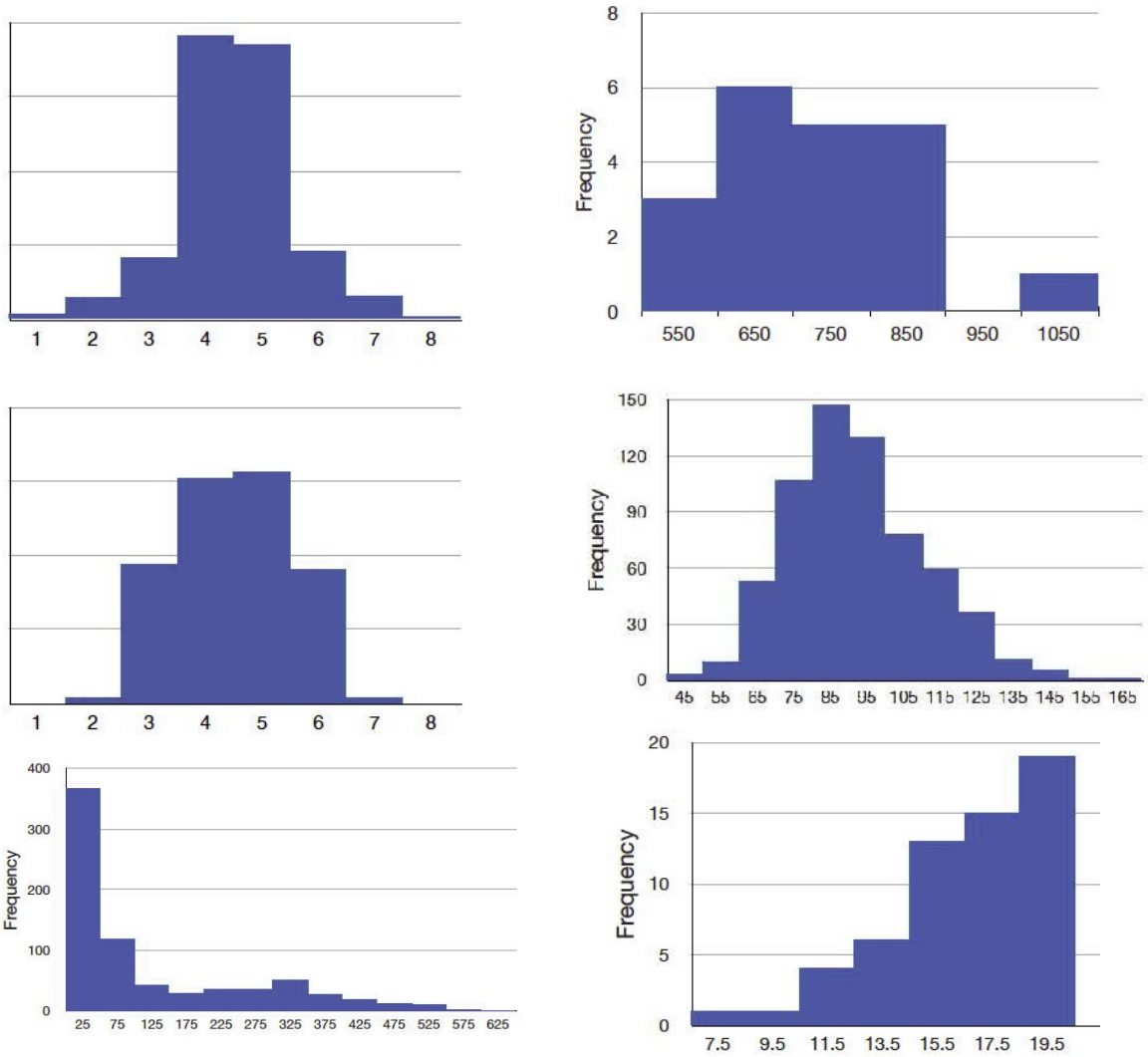
Η ροπογεννήτρια της τυχαίας μεταβλητής Χ ορίζεται να είναι η συνάρτηση

$$M(t) = E\, e^{tX} = E\left(1 + Xt + \frac{X^2}{2!}t^2 + \cdots\right) = \left(1 + tEX + \frac{t^2}{2!}EX^2 + \cdots\right) = \sum_{k=0}^{\infty}\frac{t^k}{k!}EX^k = \sum_{k=0}^{\infty}\frac{t^k}{k!}\kappa_\kappa$$

Συμπεραίνουμε πως : *στη σειρά Taylor της ροπογεννήτριας ο συντελεστής του λόγου* $\frac{t^k}{k!}$, $k \in \mathbb{N}$, *είναι απλά η κεντρική ροπή* $\kappa_\kappa = E\,X^\kappa$.

Ασκήσεις

1. Να χαρακτηρίσετε τις παρακάτω κατανομές ως προς την κυρτότητα και τη συμμετρία τους.



2. Δίνεται η τυχαία μεταβλητή X η οποία παίρνει τις ισοπίθανες τιμές $x_1 = 1$, $x_2 = 6$, $x_3 = 8$ και $x_4 = 17$, δηλαδή $P(X = x_1) = P(X = x_2) = P(X = x_3) = P(X = x_4) = 1/4$. Να υπολογιστούν :

(α) η αναμενόμενη τιμή της μεταβλητής X.

(β) η 2η, η 3η και η 4η κεντρική στιγμή της μεταβλητής X.

(γ) ο συντελεστής ασυμμετρίας της κατανομής

(δ) ο συντελεστής κυρτότητας της κατανομής

(ε) ποιο το συμπέρασμα για την κατανομή των τιμών;

Κεφάλαιο 2 Παλινδρόμηση

Με τον όρο “ανάλυση παλινδρόμησης” περιγράφονται όλες οι στατιστικές διαδικασίες που έχουν ως στόχο την ανίχνευση της σχέσης δύο ή περισσοτέρων μεταβλητών αλλά και την καταγραφή μοντέλου πρόβλεψης. Πιο συγκεκριμένα η ανάλυση παλινδρόμησης προσπαθεί να καταλάβει τον τρόπο που μεταβάλλεται η αναμενόμενη (μέση) τιμή μίας εξαρτημένης μεταβλητής όταν αλλάξει η τιμή μίας ή περισσοτέρων ανεξάρτητων μεταβλητών. Το αποτέλεσμα μίας τέτοιας διαδικασίας είναι ένα μοντέλο πρόβλεψης της μορφής

$$Y = f(X_1, X_2, ..., X_\kappa)$$

Η συνάρτηση f μπορεί να έχει οποιαδήποτε μορφή ωστόσο ο ερευνητής πρέπει να προσπαθεί να εντοπίζει το απλούστερο δυνατό μοντέλο, δηλαδή ένα της μορφής

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_\kappa X_\kappa$$

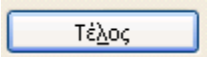
και να το εγκαταλείπει μόνο όταν αυτό δεν είναι ικανοποιητικό στην πρόβλεψή του.

Στις επόμενες παραγράφους θα μελετήσουμε ένα παράδειγμα δημιουργίας γραμμικού μοντέλου, χρησιμοποιώντας δεδομένα προσδόκιμου ζωής ανδρών και γυναικών του έτους 2005 σε 30 τυχαία επιλεγμένες χώρες. Η διαδικασία ξεκινά από την οπτική αναγνώριση μίας σχέσης (αν υπάρχει) και αυτό γίνεται με το διάγραμμα διασποράς (scatterplot).

2.1 Διάγραμμα διασποράς (Scatterplot)

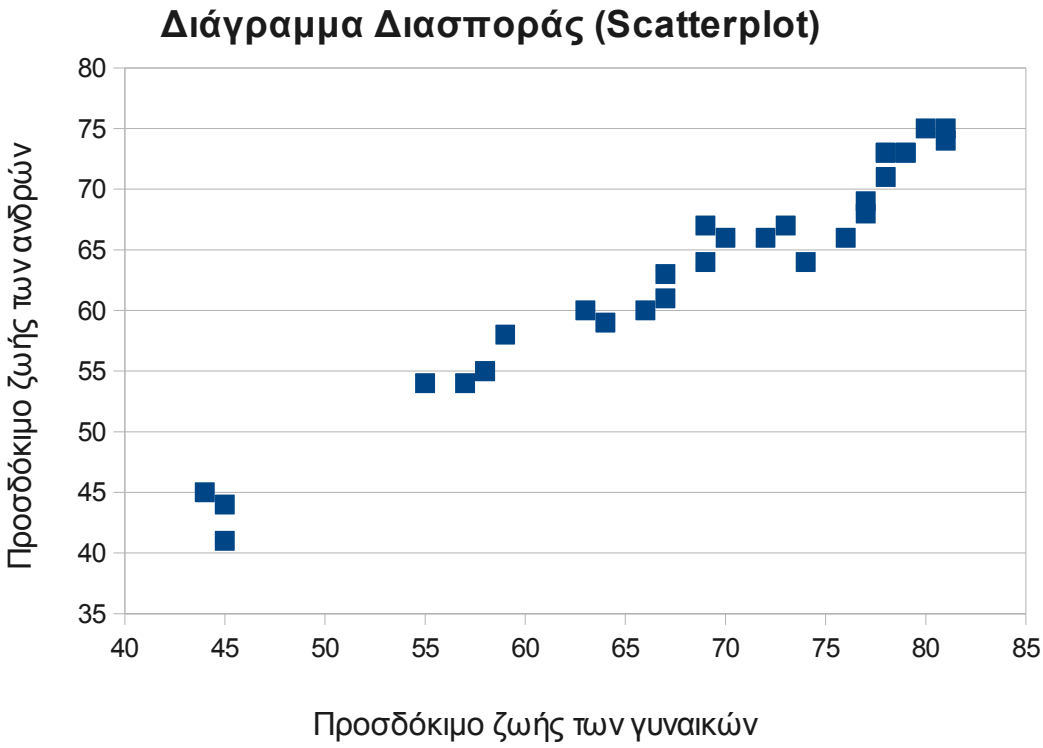
Το διάγραμμα διασποράς είναι το κατάλληλο γράφημα που δημιουργούμε ως πρώτο βήμα για να μελετήσουμε τη σχέση που υπάρχει μεταξύ δύο συνεχών αριθμητικών μεταβλητών, ιδιαίτερα αν αποσκοπούμε στη δημιουργία μοντέλου πρόγνωσης των τιμών της μίας μεταβλητής από την άλλη. Με το διάγραμμα διασποράς και μια έμπειρη στατιστική ματιά ανιχνεύεται η σχέση που ενδεχομένως να υπάρχει μεταξύ των δύο μεταβλητών.

Το διάγραμμα διασποράς δημιουργείται στο LibreOffice Calc, όπως όλα τα γραφήματα επιλέγοντας ***Εισαγωγή*** → ***Διάγραμμα*** είτε επιλέγοντας το εικονίδιο στη γραμμή εργαλείων. Στον οδηγό διαγράμματος που εμφανίζεται στο πρώτο βήμα (**1. Τύπος διαγράμματος**) επιλέγουμε τη δημιουργία διαγράμματος τύπου “**XY (Διασπορά)**” Στο δεύτερο βήμα (**2. Περιοχή δεδομένων**) επιλέγουμε την περιοχή των δεδομένων μας (**B3:C32** στην περίπτωση του παραδείγματος) και προσέχουμε να αποεπιλέξουμε τις επιλογές “**Πρώτη γραμμή σαν ετικέτα**” και “**Πρώτη στήλη σαν ετικέτα**” (Οι ονομασίες των αξόνων τοποθετούνται σε επόμενο βήμα). Προχωρούμε στην τελευταία επιλογή (**4. Στοιχεία διαγράμματος**), αποεπιλέγουμε την επιλογή “**Προβολή υπομνήματος**” και εισάγουμε τον τίτλο του διαγράμματος και τις ονομασίες των αξόνων έχοντας στο μυαλό μας πως το Calc θεωρεί την πρώτη στήλη πάντα σαν τετμημένη για κάθε σημείο (δηλαδή X) και τη δεύτερη

σαν τεταγμένη (δηλαδή Y). Έτσι, στην περίπτωση των δεδομένων του πίνακα 4 ο άξονας X ονομάζεται “Προσδόκιμο ζωής των γυναικών” ενώ ο άξονας Y “Προσδόκιμο ζωής των ανδρών”. Επιλέγοντας 

αφήνουμε τον οδηγό και το διάγραμμα είναι έτοιμο και έχει τοποθετηθεί στο φύλλο εργασίας του Calc. (Διάγραμμα 3).

Στην περίπτωση όπου το διάγραμμα εμφανίζεται αρκετά “κενό” μπορούμε να αλλάξουμε το κατώτερο όριο εμφάνισης τιμών στον έναν ή και στους δύο άξονες κάνοντας διπλό κλικ πάνω στο γράφημα και ξανά διπλό κλικ πάνω στον άξονα (Στο διάγραμμα διασποράς 3 τοποθετήθηκε ως κατώτερη τιμή του άξονα Y ο αριθμός 35). Επιπλέον, με περιήγηση πάνω στα μικρά τετράγωνα εμφανίζεται ο αύξων αριθμός και το ζεύγος τιμών από τις οποίες προήλθε.



Διάγραμμα 3: Διάγραμμα Διασποράς

Παρατηρήσεις

1. Από την παρατήρηση του διαγράμματος 3 φαίνεται πως τα σημεία του διαγράμματος είναι κοντά σε μία ευθεία. Αυτό σημαίνει πως η εξίσωση παλινδρόμησης πρέπει να είναι γραμμική!

2. Το διάγραμμα διασποράς μας δίνει την πληροφορία πως οι δύο μεταβλητές συνδέονται με γραμμικό τρόπο αλλά δεν εξηγεί ποια θα έχει το ρόλο της εξαρτημένης και ποια το ρόλο της ανεξάρτητης. Αυτό το αποφασίζει μόνος του ο ερευνητής.

2.2 Συνδιακύμανση

Η συνδιακύμανση (covariance) είναι μία στατιστική ποσότητα η οποία ποσοτικοποιεί το είδος των μεταβολών που εμφανίζονται στις τιμές μίας συνεχούς τυχαίας μεταβλητής όταν μία άλλη μεταβάλλεται.

Αν X και Y είναι δύο τυχαίες μεταβλητές τότε η συνδιακύμανση ορίζεται ως

$$Cov(X,Y)=E[(X-EX)(Y-EY)]=E[(X-\mu_X)(Y-\mu_Y)]=E(XY)-EX\cdot EY$$

Στην περισσότερο συνηθισμένη περίπτωση όπου δεν γνωρίζουμε τις θεωρητικές κατανομές συχνοτήτων αλλά έχουμε ένα δείγμα τιμών των τυχαίων μεταβλητών τότε υπολογίζουμε τη δειγματική συνδιακύμανση ως

$$s_{XY}^2=\frac{1}{v}\sum_{i=1}^{v}(x_i-\bar{x})(y_i-\bar{y})$$

Παράδειγμα

Δίνεται ο εξής πίνακας με τις παρατηρήσεις των τ.μ. X και Y.

X	1	2	3
Y	3	5	10

Υπολογίζουμε $\bar{x}=\frac{1+2+3}{3}=2$ και $\bar{y}=\frac{3+5+10}{3}=6$ και

$$s_{XY}=\frac{1}{3}[(1-2)(3-6)+(2-2)(5-6)+(3-2)(10-6)]=\frac{1}{3}(3+0+4)=\frac{7}{3}=2{,}33$$

2.2.1 Ερμηνεία της συνδιακύμανσης

Το πρόσημο της τιμής της συνδιακύμανσης δείχνει την τάση στη γραμμική σχέση των δύο μεταβλητών. Πιο συγκεκριμένα, αν οι μεγαλύτερες τιμές της μίας αντιστοιχούν στις μεγαλύτερες τιμές της άλλης και το ίδιο συμβαίνει για τις μικρότερες τιμές, δηλαδή όταν οι δύο μεταβλητές εμφανίζουν παρόμοια συμπεριφορά τότε η συνδιακύμανση είναι θετική. Στην αντίθετη περίπτωση όπου οι μεγαλύτερες τιμές της μίας μεταβλητής αντιστοιχούν στις μικρότερες της άλλης τότε η συνδιακύμανση είναι αρνητική. Εκτός από το πρόσημο, η συνδιακύμανση μπορεί να ερμηνευθεί και ως προς την απόλυτη τιμή της χωρίς όμως αυτό να είναι εύκολο καθώς το μέγεθος της εξαρτάται από το είδος των μονάδων που χρησιμοποιούνται αλλά και από το μέγεθος του δείγματος. Περισσότερο καθαρά συμπεράσματα για το είδος και την ισχύ της γραμμικής σχέσης προκύπτουν από την χρήση της κανονικοποιημένης εκδοχής της συνδιακύμανσης που είναι ο συντελεστής συσχέτισης του Pearson.

2.3 Συντελεστής συσχέτισης Pearson

Ο συντελεστής συσχέτισης του Pearson είναι το κατάλληλο στατιστικό για την ανίχνευση της γραμμικής σχέσης δύο ποσοτικών μεταβλητών, για συνεχείς ή αριθμητικές διακριτές μεταβλητές . Συμπληρώνει το διάγραμμα διασποράς υπό την έννοια ότι αντιστοιχεί μια συγκεκριμένη αριθμητική τιμή στην οπτική παρατήρηση της προσαρμογής του διαγράμματος διασποράς σε μία ευθεία.

Αν X και Y είναι δύο τυχαίες μεταβλητές τότε ο συντελεστής συσχέτισης ορίζεται ως

$$\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X \sigma_Y}=\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Όταν δεν γνωρίζουμε τις θεωρητικές κατανομές συχνοτήτων αλλά έχουμε ένα δείγμα τιμών των τυχαίων μεταβλητών τότε υπολογίζουμε το συντελεστή συσχέτισης ως

$$r_{XY}=\frac{\sum_{i=1}^{v}(x_i-\bar{x})(y_i-\bar{y})}{v\, s_X, s_Y}=\frac{\sum_{i=1}^{v}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{v}(x_i-\bar{x})^2 \sum_{i=1}^{v}(y_i-\bar{y})}}$$

Παράδειγμα

Με τα δεδομένα της παραγράφου 2.2, σελίδα 11 υπολογίζουμε

$$\sum_{i=1}^{3}(x_i-\bar{x})(y_i-\bar{y})=(1-2)(3-6)+(2-2)(5-6)+(3-2)(10-6)=7 \quad ,$$

$$\sum_{i=1}^{3}(x_i-\bar{x})^2=(1-2)^2+(2-2)^2+(3-2)^2=2 \quad \text{και} \quad \sum_{i=1}^{3}(y_i-\bar{y})^2=(3-6)^2+(5-6)^2+(10-6)^2=26$$

άρα ο συντελεστής συσχέτισης του Pearson είναι

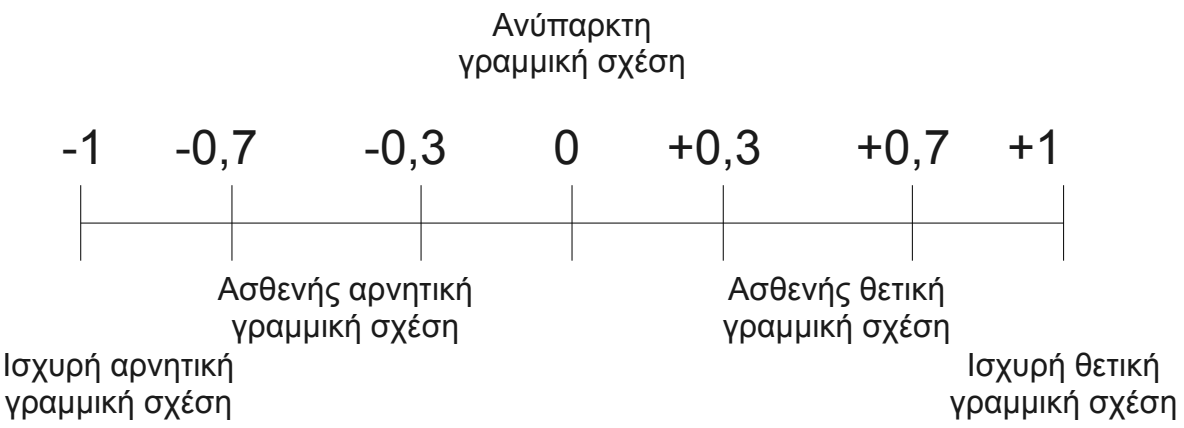
$$r_{XY}=\frac{7}{\sqrt{2\cdot 26}}=0{,}97$$

2.3.1 Προϋποθέσεις υπολογισμού

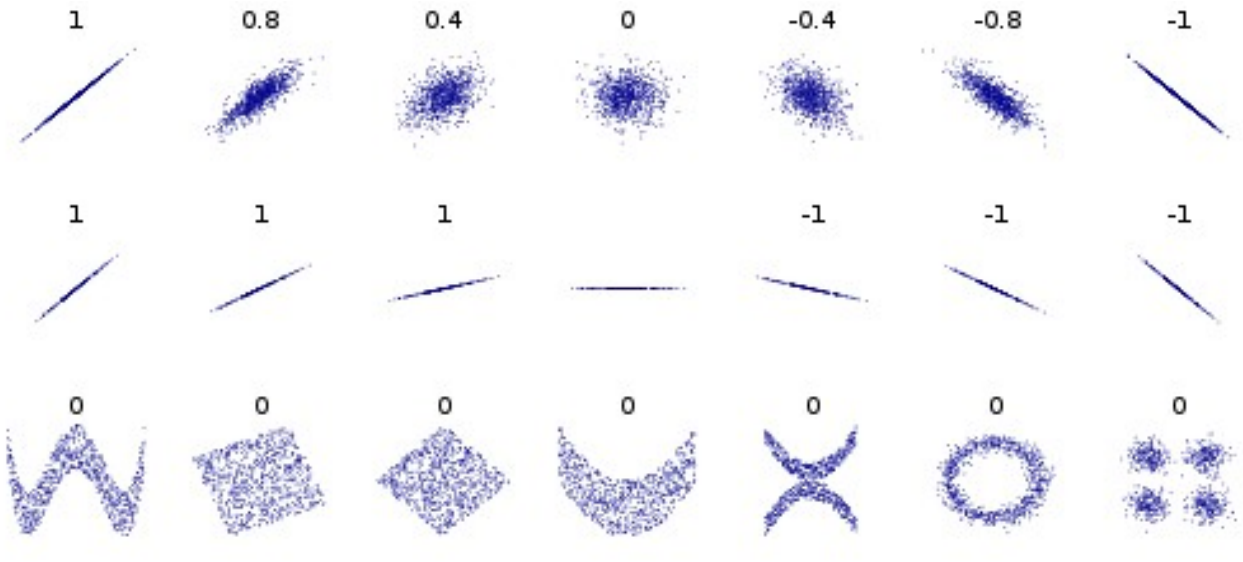
Ο συντελεστής Pearson είναι παραμετρικός, δηλαδή προϋποθέτει πως οι τιμές των δύο μεταβλητών προέρχονται από κανονικούς πληθυσμούς. Η κανονικότητα των πληθυσμών τεκμηριώνεται χρησιμοποιώντας το ιστόγραμμα ή κάποιο από τα διαγράμματα p – p και q – q. Αν υπάρχουν σοβαρές αμφιβολίες για την υπόθεση αυτή ή οι μεταβλητές είναι διακριτές με σχετικά λίγες τιμές τότε καλύτερα να υπολογιστεί ο συντελεστής Spearman ο οποίος είναι το μη παραμετρικό ανάλογο του συντελεστή Pearson.

2.3.2 Αξιολόγηση του συντελεστή Pearson

Ο συντελεστής Pearson, ο οποίος συμβολίζεται συνήθως **r** ή **R**, παίρνει πάντα τιμές μεταξύ -1 και +1 και ανάλογα με την τιμή συνάγουμε το είδος και την ισχύ της γραμμικής σχέσης μεταξύ των μεταβλητών. Στο Σχέδιο 1 παρουσιάζεται συνοπτικά η αξιολόγηση του συντελεστή.



Σχέδιο 1: Αξιολόγηση του συντελεστή συσχέτισης



Διάγραμμα 4: Ενδεικτικά διαγράμματα διασποράς και ο αντίστοιχος συντελεστής Pearson

Παράδειγμα

Με τα δεδομένα της παραγράφου 2.2, σελίδα 11 βρήκαμε $r_{XY}=0{,}97$ κάτι που μας οδηγεί να συμπεράνουμε πως οι μεταβλητές X και Y είναι ισχυρά θετικά γραμμικά συσχετισμένες.

2.3.3 Υπολογισμός συντελεστή συσχέτισης με το Calc

Οι τιμές των μεταβλητών πρέπει να είναι τοποθετημένες ανά ζεύγη σε δύο, όχι κατ΄ ανάγκη συνεχόμενες, στήλες του Calc. Για παράδειγμα στον πίνακα της επόμενης σελίδας (Εικόνα 4) παρουσιάζονται δεδομένα για το προσδόκιμο ζωής ανδρών και γυναικών σε τριάντα χώρες του κόσμου το έτος 1995. Αυτές τοποθετήθηκαν στις στήλες A, B, C ενός φύλλου του Calc από την γραμμή 1 έως την 32 (Τα δεδομένα βρίσκονται στις στήλες 3 έως 32).

Η ποιοτική ανάλυση των μεταβλητών φανερώνει πως οι μεταβλητές “Προσδόκιμο ζωής ανδρών” και “Προσδόκιμο ζωής γυναικών” είναι θετικά συσχετισμένες. Πράγματι, όσο μεγαλύτερο είναι το προσδόκιμο ζωής για το ένα από τα δύο φύλα σε μία χώρα ανάλογα μεγάλο περιμένουμε να είναι και για το άλλο φύλο. Υπολογίζοντας τον συντελεστή Pearson θα αποκτήσουμε μια ποσοτική μέτρηση του γεγονότος αυτού ενώ η επιβεβαίωση της ισχυρής γραμμικής σχέσης θα ανοίξει τον δρόμο για την αξιόπιστη πρόβλεψη της τιμής της μίας μεταβλητής από την άλλη χρησιμοποιώντας την εξίσωση της ευθείας γραμμικής παλινδρόμησης.

Ο συντελεστής γραμμικής συσχέτισης υπολογίζεται με τη συνάρτηση **PEARSON()** η οποία χρειάζεται δύο ορίσματα, το πρώτο από το οποίο θα είναι οι τιμές της μίας μεταβλητής και το δεύτερο οι τιμές της άλλης (Εικόνα 3). Είναι φανερό πως πρέπει οι δύο ομάδες δεδομένων να είναι ίσες στο πλήθος! Αν τοποθετήσουμε ομάδες διαφορετικού πλήθους τότε το αποτέλεσμα θα είναι **Σφάλμα: 502.**

Συντελεστής Γραμμικής Συσχέτισης Pearson	Συνάρτηση
0,98	PEARSON(B3:B32;C3:C32)

Εικόνα 3: Συντελεστής συσχέτισης Pearson

2.4 Συντελεστής συσχέτισης Spearman

Ο συντελεστής συσχέτισης Spearman είναι το μη παραμετρικό ανάλογο του συντελεστή Pearson υπό την έννοια πως υπολογίζεται χρησιμοποιώντας την τάξη κάθε στοιχείου δηλαδή τη σειρά κατάταξης του στα ταξινομημένα συνολικά δεδομένα. Συμβολίζεται συνήθως με το γράμμα ρ. Λόγω του τρόπου ορισμού του είναι καλύτερος εκτιμητής της συσχέτισης δύο διατακτικών μεταβλητών δηλαδή μεταβλητών που παίρνουν τιμές ακέραιους αριθμούς όπως συμβαίνει στις μεταβλητές με κλίμακα Liker.

Δυστυχώς το Calc δεν έχει συνάρτηση άμεσου υπολογισμού του συντελεστή Spearman αλλά εύκολα μπορούμε να το υπολογίσουμε χρησιμοποιώντας τον τύπο ορισμού του

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad .$$

Στον παραπάνω τύπο το n είναι το πλήθος των ζευγών παρατηρήσεων, ενώ το d είναι η διαφορά της τάξης της μίας παρατήρησης ενός ζεύγους από την άλλη, αριθμοί οι οποίοι υψώνονται στο τετράγωνο και αθροίζονται για να μας δώσουν τον αριθμητή του κλάσματος.

Δεδομένα		
Χώρα	**Προσδόκιμο ζωής γυναικών**	**Προσδόκιμο ζωής ανδρών**
Αίγυπτος	63	60
Αυστρία	79	73
Αφγανιστάν	44	45
Βέλγιο	79	73
Βολιβία	64	59
Δομινικανή Δημοκρατία	70	66
Ελ Σαλβαδόρ	69	64
Ελλάδα	80	75
Ζάμπια	45	44
Ινδία	59	58
Ισημερινός	73	67
Καμερούν	58	55
Καναδάς	81	74
Κίνα	69	67
Κουβέιτ	78	73
Λευκορωσία	76	66
Λιθουανία	77	68
Μαλαισία	72	66
Μποτσουάνα	66	60
Νησιά Μπαρμπάντος	78	73
Νιγηρία	57	54
Νικαράγουα	67	61
Ολλανδία	81	75
Περού	67	63
Ρωσία	74	64
Σενεγάλη	58	55
Σομαλία	55	54
Τανζανία	45	41
Τσεχία	77	69
Χιλή	78	71

Εικόνα 4: Δεδομένα προσδόκιμου ζωής έτους 1995

Για κάθε μία από τις δύο στήλες των δεδομένων δημιουργούμε μια στήλη στην οποία αποθηκεύονται οι τάξεις των παρατηρήσεων χρησιμοποιώντας τη συνάρτηση **RANK()**. Οι υπόλοιπες ενέργειες είναι απλές!

Ο συντελεστής Spearman αξιολογείται με τον ίδιο τρόπο με αυτόν του Pearson (Σχέδιο 1). Απόλυτη τιμή του συντελεστή μεγαλύτερη από 0,7 συνήθως αξιολογείται ως ισχυρή γραμμική σχέση, μεταξύ 0,3 και 0,7 ως ασθενής γραμμική σχέση ενώ μεταξύ 0 και 0,3 ως μη γραμμική σχέση.

Συντελεστής Συσχέτισης Spearman	Συνάρτηση	
0,98	1-6*I9/(I10*(I10^2-1))	
		Συνάρτηση
Άθροισμα d^2	91	SUM(F3:F32)
Πλήθος Παρατηρήσεων	30	COUNT(E3:E32)

Εικόνα 5: Υπολογισμός συντελεστή συσχέτισης Spearman

RANK(B3;B$3:B$32)	RANK(C3;C$3:C$32)	(D3-E3)^2	(D3-E3)^2
Τάξη Προσδόκιμου Ζωής Γυναικών	Τάξη Προσδόκιμου Ζωής Ανδρών	Διαφορά	Διαφορά^2
22	20	2	4
4	4	0	0
30	28	2	4
4	4	0	0
21	22	-1	1
15	13	2	4
16	16	0	0
3	1	2	4
28	29	-1	1
23	23	0	0
13	11	2	4
24	24	0	0
1	3	-2	4
16	11	5	25
6	4	2	4
11	13	-2	4
9	10	-1	1
14	13	1	1
20	20	0	0
6	4	2	4
26	26	0	0
18	19	-1	1
1	1	0	0
18	18	0	0
12	16	-4	16
24	24	0	0
27	26	1	1
28	30	-2	4
9	9	0	0

Εικόνα 6: Υπολογισμός της τάξης των παρατηρήσεων τα οποία τοποθετήθηκαν στις στήλες D,E και F

2.5 Γραμμική παλινδρόμηση.

Το πρώτο βήμα για τη μελέτη της σχέσης δύο συνεχών μεταβλητών είναι ο υπολογισμός του συντελεστή γραμμικής συσχέτισης Pearson. Το επόμενο βήμα είναι η εύρεση της ευθείας γραμμικής παλινδρόμησης της μίας μεταβλητής πάνω στην άλλη, ιδιαίτερα αν μεταξύ των μεταβλητών υπάρχει ισχυρή γραμμική συσχέτιση (συντελεστής μεγαλύτερος κατά απόλυτη τιμή από το 0,7).

Η ευθεία γραμμικής παλινδρόμησης είναι η ευθεία η οποία βρίσκεται όσο το δυνατόν “κοντύτερα” στα σημεία του διαγράμματος διασποράς. Το “κοντύτερα” είναι υποκειμενικό και μπορεί να οριστεί με διάφορους τρόπους. Η πλέον συνηθισμένη επιλογή είναι ο εντοπισμός της ευθείας για την οποία το άθροισμα των κάθετων αποστάσεων όλων των σημείων από αυτήν είναι ελάχιστο Η ευθεία αυτή ονομάζεται και ευθεία ελαχίστων τετραγώνων.

Η εξίσωσή της ευθείας ελαχίστων τετραγώνων, όπως και η εξίσωση κάθε ευθείας έχει τη μορφή

Y = α*X+β.

Στην παραπάνω εξίσωση το X ονομάζεται ανεξάρτητη μεταβλητή (independent variable) ενώ το Y εξαρτημένη (dependent variable) διότι οι τιμές του λαμβάνονται από τις τιμές του X.

Οι συντελεστές του γραμμικού μοντέλου υπολογίζονται από τα ζεύγη τιμών των μεταβλητών X και Y ως εξής :

$$\alpha=\frac{\sum_{i=1}^{v}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{v}(x_i-\bar{x})^2} \quad \text{και} \quad \beta=\bar{y}-\alpha\cdot\bar{x}$$

Παράδειγμα

1. Με τα δεδομένα της παραγράφου 2.2, σελίδα 11 είναι $\bar{x}=2$, $\bar{y}=6$, $\sum_{i=1}^{3}(x_i-\bar{x})^2=2$ και $\sum_{i=1}^{3}(x_i-\bar{x})(y_i-\bar{y})=7$ και υπολογίζουμε $\alpha=\frac{7}{2}=3{,}5$ ενώ $\beta=6-3{,}5\cdot 2=-1$ άρα το γραμμικό μοντέλο είναι

Y = 3,5 * X -1.

Η απόφαση για το ρόλο που θα αναλάβει κάθε μεταβλητή στην εξίσωση της ευθείας γραμμικής παλινδρόμησης είναι αποκλειστικά του ερευνητή (δεν υπάρχει σωστή και λάθος επιλογή) και καθορίζεται από την εσωτερική σχέση που έχουν οι δύο μεταβλητές, δηλαδή την συμπεριφορά της μίας ως “αιτίας” και της άλλης ως “αιτιατό” στο συγκεκριμένο εννοιολογικό πλαίσιο που αυτές ορίζονται.

2. Το παράδειγμα αφορά τα δεδομένα της εικόνας 4, σελίδα 15 και θα το λύσουμε με τις συναρτήσεις που προσφέρει το Calc. Θεωρούμε ως εξαρτημένη μεταβλητή **Y** το προσδόκιμο ζωής των γυναικών (τοποθετημένα στη στήλη **B** στα δεδομένα της εικόνας 4) και ως ανεξάρτητη μεταβλητή **X** το προσδόκιμο ζωής των ανδρών (τοποθετημένα στη στήλη **C** στα δεδομένα της εικόνας 4). Επιζητούμε την εξίσωση της ευθείας

Προσδόκιμο ζωής γυναικών = α * Προσδόκιμο ζωής ανδρών + β

Είναι φανερό πως χρησιμοποιώντας την τελευταία εξίσωση είναι δυνατή η πρόβλεψη του προσδόκιμου ζωής των γυναικών σε μία χώρα για την οποία δεν θα το γνωρίζουμε αλλά θα γνωρίζουμε το αντίστοιχο προσδόκιμο των ανδρών.

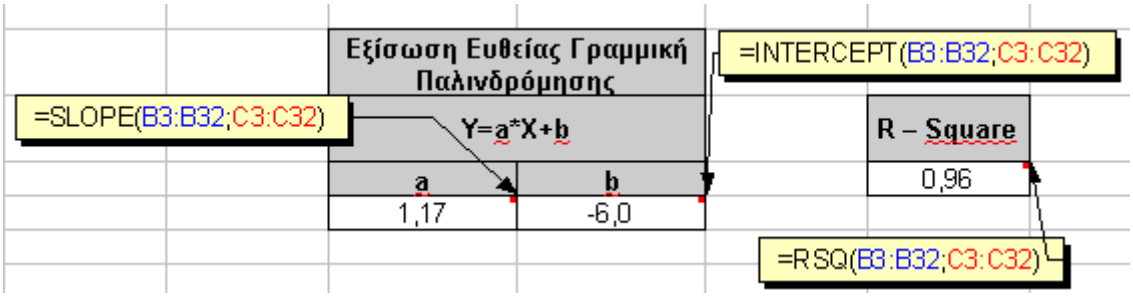
Το Calc προσφέρει έτοιμες συναρτήσεις για τον υπολογισμό των συντελεστών **α** και **β** της εξίσωσης της ευθείας. Η εφαρμογή των συναρτήσεων αυτών στα δεδομένα του προσδόκιμου ζωής (Εικόνα 4) παρουσιάζεται στην εικόνα 7.

Προσέξτε πως και στις δύο συναρτήσεις **SLOPE()** και **INTERCEPT()** πρώτα τοποθετούνται τα δεδομένα της εξαρτημένης μεταβλητής (Y) και μετά τα δεδομένα της ανεξάρτητης μεταβλητής (X).

Αντίστροφη τοποθέτηση θα οδηγούσε στον υπολογισμό των συντελεστών της εξίσωσης

Προσδόκιμο ζωής ανδρών = a * Προσδόκιμο ζωής γυναικών + b

το οποίο έχει ενδιαφέρον αλλά δεν είναι ο σκοπός της υποτιθέμενης έρευνας μας.



Εικόνα 7: Υπολογισμός συντελεστών της ευθείας γραμμικής παλινδρόμησης

Από τα εξαγόμενα των συναρτήσεων προκύπτει πως η ζητούμενη εξίσωση είναι η

Προσδόκιμο ζωής γυναικών = 1,17 * Προσδόκιμο ζωής ανδρών – 6

Εφαρμογή : Να βρεθεί το προσδόκιμο ζωής των γυναικών σε μία χώρα στην οποία οι άνδρες έχουν προσδόκιμο ζωής τα 65 χρόνια.

Υλοποίηση : Απλά αντικαθιστούμε όπου **X (Προσδόκιμο ζωής ανδρών) = 65** στην εξίσωση της ευθείας γραμμικής παλινδρόμησης και υπολογίζουμε

Αναμενόμενο προσδόκιμο ζωής γυναικών = 1,17 * 70 – 6 = 75,9 έτη

Προσέξτε πως το παραπάνω αποτέλεσμα δεν σημαίνει πως σε όλες τις χώρες όπου οι άνδρες θα ζούνε 70 χρόνια κατά μέσο όρο, οι γυναίκες θα ζούνε 75,9. Ο αριθμός που υπολογίσαμε είναι το αναμενόμενο προσδόκιμο ζωής των γυναικών, δηλαδή το μέσο προσδόκιμο ζωής των γυναικών σε όλες τις χώρες στις οποίες οι άνδρες ζούνε 70 χρόνια.

2.5.1 Αξιολόγηση μοντέλου γραμμικής παλινδρόμησης

Η ευθεία γραμμικής παλινδρόμησης δίνει πρόβλεψη της άγνωστης τιμής της εξαρτημένης μεταβλητής. Η ποιότητα της πρόβλεψης αυτής ελέγχεται από την τιμή του συντελεστή προσδιοριστίας (Coefficient of determination) **R^2 (R – Square)** το οποίο είναι καλό να υπολογίζεται μαζί με του συντελεστές της εξίσωσης (Εικόνα 7) και το οποίο ερμηνεύεται ως το ποσοστό της μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής που προσδιορίζεται από τις τιμές της ανεξάρτητης μεταβλητής, και είναι καλό να είναι μεγάλο δηλαδή κοντά στη μονάδα.

Πιο αναλυτικά, είναι

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$$

όπου SS_{err} είναι το άθροισμα τετραγώνων των σφαλμάτων του μοντέλου στις παρατηρήσεις της εξαρτημένης μεταβλητής

$$SS_{err} = \sum (y_i - \hat{y}_i)^2$$

ενώ SS_{tot} είναι το συνολικό άθροισμα τετραγώνων των αποστάσεων των παρατηρήσεων της εξαρτημένης μεταβλητής από τη μέση της τιμή

$$SS_{tot} = \sum (y_i - \bar{y})^2$$

Ιδανική είναι η περίπτωση όπου $SS_{err} = 0$, περίπτωση κατά την οποία το μοντέλο προβλέπει επακριβώς τις τιμές της εξαρτημένης μεταβλητής από τις οποίες υπολογίστηκε. Σπάνια συμβαίνει αυτό, στην πράξη ένας συντελεστής μεγαλύτερος από 80% θεωρείται ικανοποιητικός ενώ αν είναι μεγαλύτερος από 90% τότε έχουμε καλό λόγο για να θεωρήσουμε πως η πρόβλεψη των τιμών του μοντέλου θα είναι αξιόπιστη για κάθε πρακτική χρήση. Αν παρ' ελπίδα η τιμή του R^2 είναι μικρή τότε πρέπει να αναζητήσουμε περισσότερο περίπλοκο μοντέλο (μη γραμμικό ή γραμμικό με περισσότερες μεταβλητές)

Παράδειγμα

1. Με τα δεδομένα της παραγράφου 2.2, σελίδα 11 είναι $\bar{x} = 2$, $\bar{y} = 6$, και το μοντέλο πρόβλεψης της Y από τη X είναι **Y = 3,5 * X -1** (σελίδα 17). Υπολογίζουμε :

X	1	2	3
Y	3	5	10
$Y-\bar{Y}$	-3	-1	4
$(Y-\bar{Y})^2$	9	1	16
$\hat{Y}$	2,5	6	9,5
$Y-\hat{Y}$	0.5	-1	0.5
$(Y-\hat{Y})^2$	0.25	1	0.25

Από τον πίνακα έχουμε $SS_{err}=\sum(y_i-\hat{y}_i)^2=0.25+1+0.25=1.5$ και

$SS_{tot}=\sum(y_i-\bar{y})^2=9+1+16=26$. Κατά συνέπεια

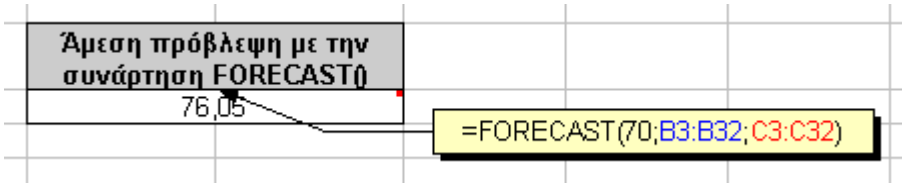
$$R^2=1-\frac{SS_{err}}{SS_{tot}}=1-\frac{1.5}{26}=0.942=94{,}2\%$$.

2. Το παράδειγμα αφορά τα δεδομένα της εικόνας 4, σελίδα 15 και θα το λύσουμε με τη συνάρτηση **RSQ()** που προσφέρει το Calc Στην περίπτωση του παραδείγματος μας είναι 0,96 ή 96% (Εικόνα 7). Αυτό σημαίνει πως το 96% της μεταβλητότητας του προσδόκιμου ζωής των γυναικών Y ορίζεται από το προσδόκιμο ζωής των ανδρών X ενώ το υπόλοιπο 4% (=100%-96%) της μεταβολής της Y ορίζεται από άλλους παράγοντες εκτός από την μεταβλητή X κάτι που στην πράξη σημαίνει πως το μοντέλο μας μάλλον κάνει καλά την δουλειά του!

Μία επιπλέον χρήσιμη παρατήρηση είναι πως η αξιοπιστία της πρόβλεψης που θα προκύψει από την εξίσωση της ευθείας γραμμικής παλινδρόμησης είναι ανάλογη της αξιοπιστίας του υπολογισμού των συντελεστών α και β. Καθώς η μεταβλητότητα των δεδομένων επηρεάζει το μέγεθος των συντελεστών αυτών είναι καλό μετά τον υπολογισμό των α και β να προχωρούμε και σε στατιστικούς ελέγχους οι οποίοι θα διαβεβαιώνουν πως τα α και β δεν είναι μηδέν!
Πιο συγκεκριμένα, για το συντελεστή α,
Ερευνητική Υπόθεση H_1: Ο συντελεστής α στην παραπάνω εξίσωση είναι διάφορος από το μηδέν.
Στατιστική Υπόθεση H_0: Ο συντελεστής α στην παραπάνω εξίσωση είναι ίσος με το μηδέν.
και ανάλογες υποθέσεις για το συντελεστή β.
Είναι επιθυμητή η απόρριψη των στατιστικών υποθέσεων και στους δύο ελέγχους καθώς η αποδοχή κάποιας από τις δύο θα σημαίνει πως υπάρχει τόσο μεγάλη μεταβλητότητα στις τιμές που δεν μπορούμε να υπολογίσουμε με ασφάλεια τον αντίστοιχο συντελεστή. Δυστυχώς, οι έλεγχοι αυτοί ενώ είναι δεδομένοι σε περισσότερο ειδικά λογισμικά (R-Project ή SPAW) στο Calc δεν υπάρχουν ανάλογες διαδικασίες!

2.5.2 Άμεση πρόβλεψη με γραμμικό μοντέλο με συνάρτηση του Calc

Αν δεν επιθυμούμε να βρούμε την εξίσωση της ευθείας αλλά θέλουμε απευθείας να την εφαρμόσουμε μπορούμε να χρησιμοποιήσουμε την συνάρτηση **FORECAST()** η οποία υπολογίζει άμεσα την πρόβλεψη βάσει του γραμμικού μοντέλου χωρίς να το εμφανίζει!

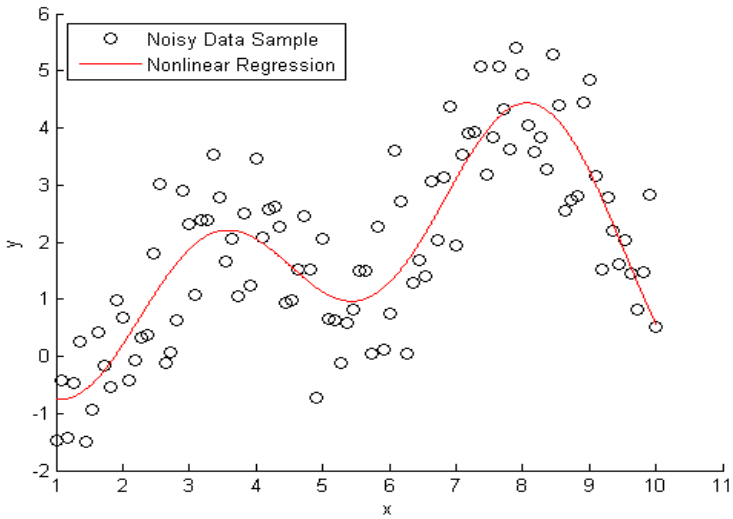


Εικόνα 8: Άμεση πρόβλεψη

Στην εικόνα Σφάλμα: Δεν βρέθηκε η πηγή παραπομπής παρουσιάζεται η εφαρμογή της συνάρτησης. Είναι χαρακτηριστικό πως η τιμή που προκύπτει από την εφαρμογή της συνάρτησης **FORECAST()** είναι λίγο διαφορετική από την τιμή που υπολογίσαμε με το γραμμικό μας μοντέλο της παραγράφου 2.5. Η διαφορά οφείλεται στο γεγονός της αναγκαστικής στρογγυλοποίησης που εφαρμόσαμε στους συντελεστές a και b της εξίσωσης την οποία δεν πράττει στον ίδιο βαθμό η συνάρτηση **FORECAST()**.

2.6 Μη γραμμική παλινδρόμηση

Πολλές φορές, ιδιαίτερα στις βιολογικές επιστήμες το διάγραμμα διασποράς συνηγορεί σε μία μη γραμμική σχέση.

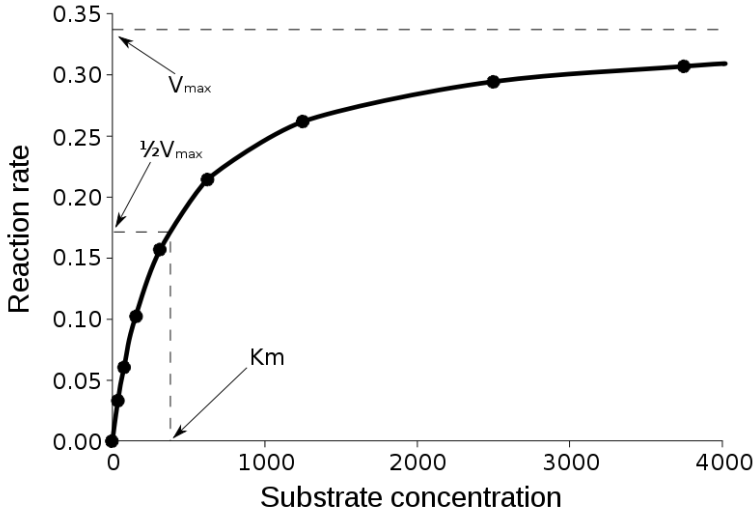


Διάγραμμα 5: Μη γραμμικό μοντέλο (1ο παράδειγμα)

Στην περίπτωση του διαγράμματος 5 το βέλτιστο μοντέλο προκύπτει να είναι

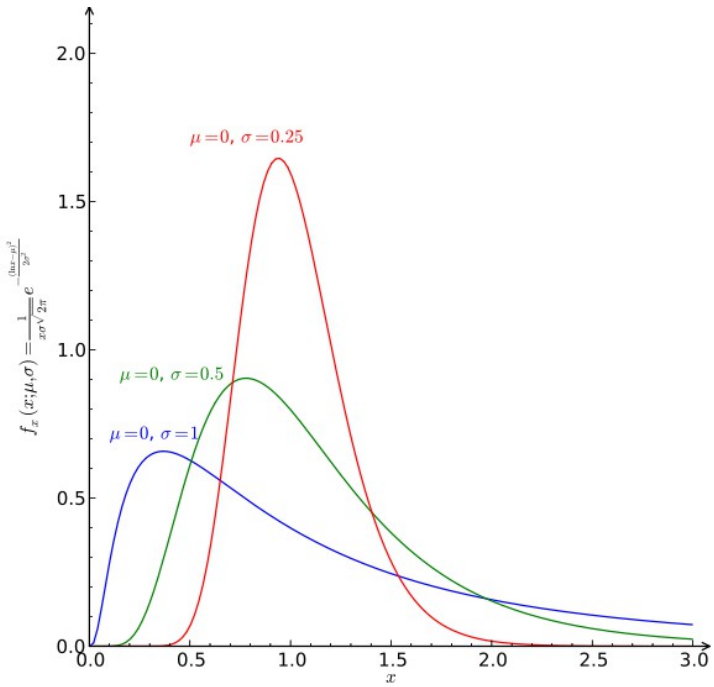
$$Y = 1{,}73 - 0{,}2\,\sigma\upsilon\nu\left(0{,}63\,X\right) - 1{,}41\,\eta\mu\left(0{,}63\,X\right) - 0{,}77\,\sigma\upsilon\nu\left(2\cdot 0{,}63\,X\right) - 1{,}31\,\eta\mu\left(2\cdot 063\,X\right)$$

το οποίο ένα έμπειρο μάτι ίσως τα αναγνωρίζει ως τους πρώτους 4 όρους μίας σειράς Fourier!



Διάγραμμα 6: Μη γραμμικό μοντέλο (2ο παράδειγμα)

Στην περίπτωση του διαγράμματος 6 το βέλτιστο μοντέλο είναι $\upsilon = \frac{V_{max} X}{K_m + X}$. Υπάρχουν απεριόριστες περιπτώσεις για τα μη γραμμικά μοντέλα.



Διάγραμμα 7: Λογαριθμοκανονική κατανομή

Στην πράξη όταν κάποια σχέση φαίνεται να είναι μη γραμμική, προσπαθούμε με κατάλληλη αλλαγή μεταβλητής να την εκφράσουμε ως γραμμική και να βρούμε γραμμικό μοντέλο. Η διαδικασία αυτή

ακολουθείται κατά κανόνα στα βιολογικά ή φυσικά συστήματα που στα οποία τα φαινόμενα εξελίσσονται συνήθως πολλαπλασιαστικά και κατά συνέπεια τα δεδομένα ακολουθούν λογαριθμοκανονική κατανομή (διάγραμμα 7, σελίδα 22). Στην περίπτωση αυτή ο υπολογισμός του λογαρίθμου της εξαρτημένης μεταβλητής συνήθως αρκεί για να κάνει τα δεδομένα κανονικά και να γίνει γραμμική η ανάλυση.

Ασκήσεις

1. Δίνονται οι εξής παρατηρήσεις των μεταβλητών Χ και Υ

Χ	0	2	5	6
Υ	8	8	5	3

(α) Να γίνει το διάγραμμα διασποράς των μεταβλητών Χ και Υ.

(β) Να βρεθεί η συνδιακύμανση των μεταβλητών Χ και Υ. Πως ερμηνεύεται η τιμή που βρέθηκε;

(γ) Να βρεθεί ο συντελεστής συσχέτισης Pearson. Πως ερμηνεύεται η τιμή που βρέθηκε;

(δ) Να βρεθεί γραμμικό μοντέλο πρόβλεψης των τιμών της μεταβλητής Υ από τη μεταβλητή Χ.

(ε) Αν Χ = 4 τότε ποια περιμένουμε να είναι η τιμή του Υ;

(στ) Να βρεθεί ο συντελεστής προσδιοριστίας του γραμμικού μοντέλου που βρέθηκε.

2. Επισκεφθείτε την ιστοσελίδα της ελληνικής στατιστικής αρχής http://www.statistics.gr/

και επιλέξτε ένα κατάλληλο σύνολο δεδομένων και δημιουργήστε ένα γραμμικό μοντέλο πρόβλεψης. Πιο συγκεκριμένα πρέπει να κάνετε τα εξής :

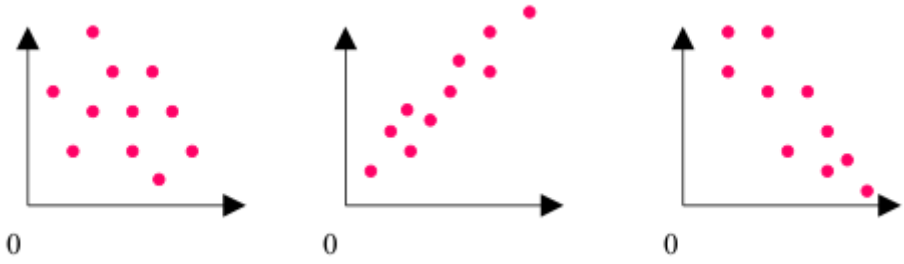
(α) Επιλέξτε ένα σύνολο δεδομένων με τουλάχιστον 2 μεταβλητές.

(β) Αποφασίστε ποια θα είναι η εξαρτημένη και ποια η ανεξάρτητη μεταβλητή.

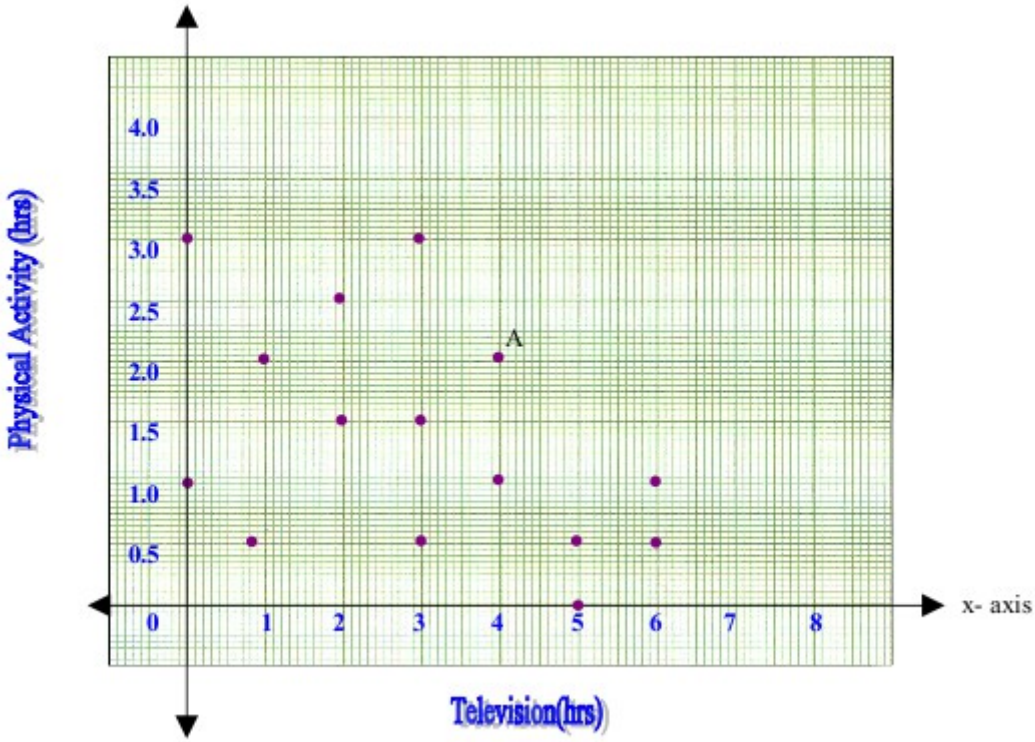
(γ) Δημιουργήστε γραμμικό μοντέλο πρόβλεψης (με το Calc!)

(δ) Προχωρήστε σε πρόβλεψη τιμής της Υ από τη Χ και συγκρίνεται με τα πραγματικά δεδομένα.

3. Βρείτε το πρόσημο και το μέγεθος του συντελεστή συσχέτισης του Pearson που αντιστοιχεί στα παρακάτω διαγράμματα



4. Από το διάγραμμα διασποράς που ακολουθεί να ανακτήσετε τα δεδομένα από τα οποία προήλθε

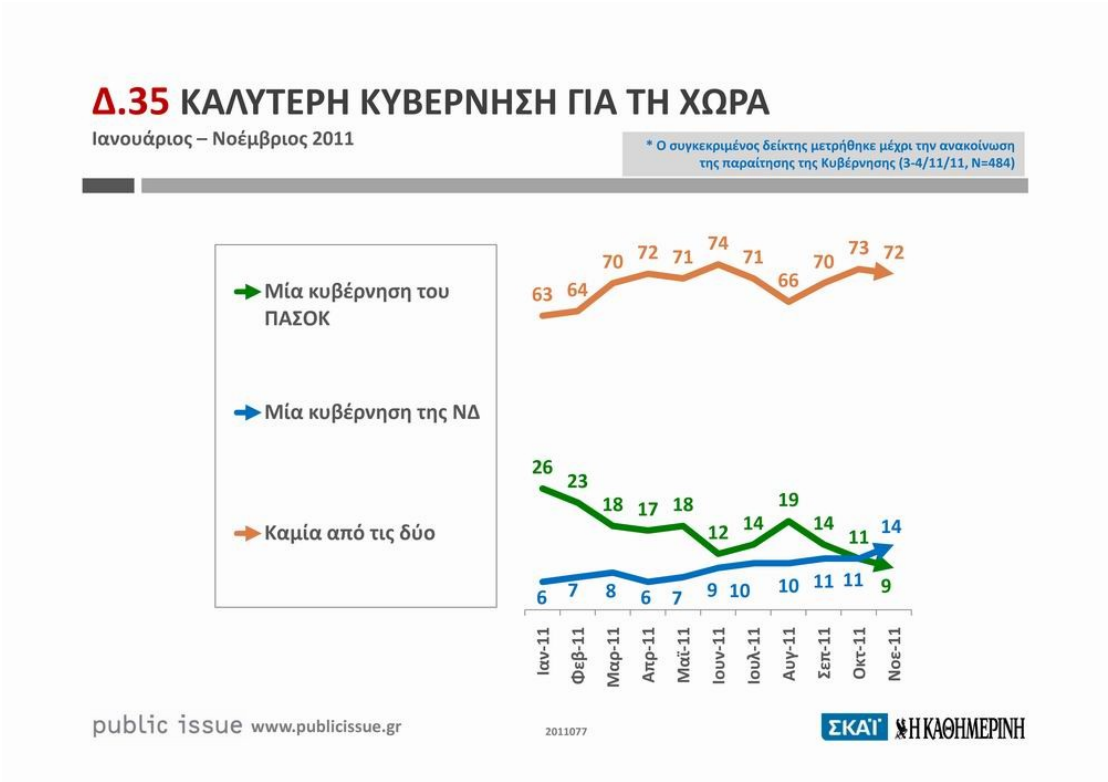


Κεφάλαιο 3 Χρονοσειρές

Στη στατιστική επιστήμη στην οικονομετρία και τα οικονομικά μαθηματικά, μια χρονολογική σειρά είναι μια ακολουθία χρονολογικών δεδομένων, η οποία μετράται συνήθως σε διαδοχικές χρονικές στιγμές οι οποίες απέχουν μεταξύ τους κατά ίσα χρονικά διαστήματα. Παραδείγματα χρονοσειρών είναι η ημερήσια τιμή κλεισίματος του γενικού δείκτη του χρηματιστηρίου ή οι ημερήσιες πωλήσεις ενός προϊόντος στη διάρκεια του τελευταίου έτους.

Η ανάλυση των χρονολογικών σειρών αποσκοπεί στην εύρεση χρήσιμων χαρακτηριστικών της χρονοσειράς που θα επιστρέψει στον ερευνητή να κατανοήσεις καλύτερα το φαινόμενο που παρατηρεί. Επιπλέον, ιδιαίτερα επιθυμητή είναι συνήθως η δημιουργία μοντέλων πρόβλεψης της μελλοντικής συμπεριφοράς της χρονοσειράς με βάση τις τιμές που παρατηρήθηκαν στο παρελθόν.

Οι χρονοσειρές συνήθως περιγράφονται γραφικά με τα χρονοδιαγράμματα, στα οποία ο χρόνος τοποθετείται στον οριζόντιο άξονα σε κατάλληλες μονάδες ενώ η τιμή του χαρακτηριστικού που παρατηρείται τοποθετείται στον κάθετο άξονα σε κατάλληλη κλίμακα. Στο παρακάτω διάγραμμα 8 παρουσιάζεται ένα χρονοδιάγραμμα από μία πρόσφατη δημοσκόπηση :



Διάγραμμα 8: http://www.publicissue.gr/1925/varometro-analysis-nov-2011/

3.1 Τρόποι στατιστικής ανάλυσης χρονοσειρών

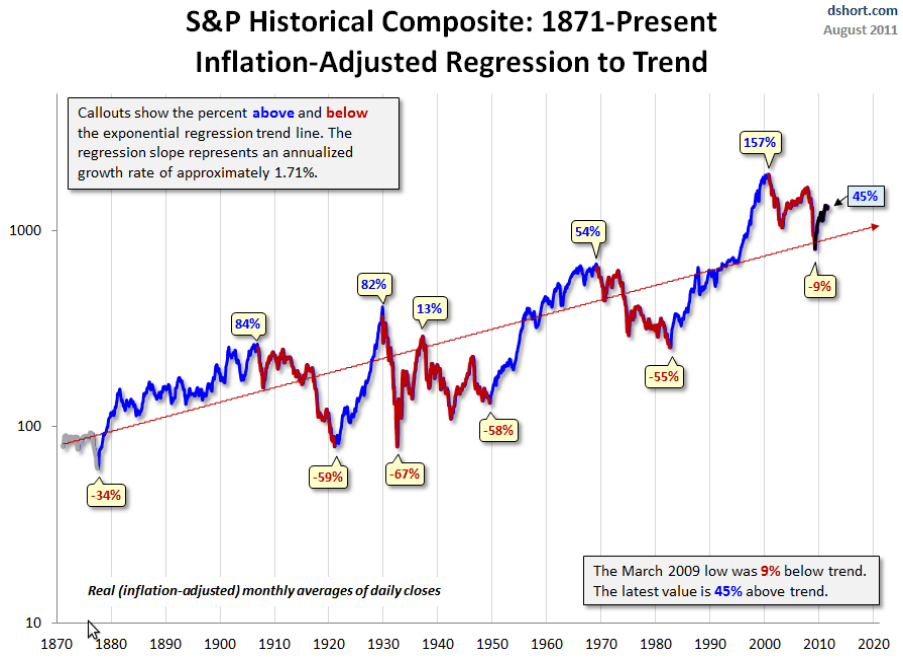
Οι βασικότεροι τρόποι με τους οποίου μπορούμε να επεξεργαστούμε μία χρονοσειρά είναι οι εξής :

1. Ανάλυση της χρονοσειράς στις κυριότερες συνιστώσες
2. Έλεγχος για αυτοσυσχέτιση – ανίχνευση γραμμικής εξάρτησης συνεχόμενων τιμών
3. Φασματική ανάλυση

3.2 Ανάλυση της χρονοσειράς στις κυριότερες συνιστώσες

Η ανάλυση μίας χρονοσειράς στις κυριότερες συνιστώσες της είναι η βασικότερη διαδικασία επεξεργασίας μίας χρονοσειράς. Στόχος είναι η χρησιμοποίηση των διαθέσιμων δεδομένων για την καταγραφή όλων των “επιρροών” που διαμορφώνουν την τελική τιμή της χρονοσειράς. Συνήθως, η ανάλυση γίνεται στις εξής συνιστώσες :

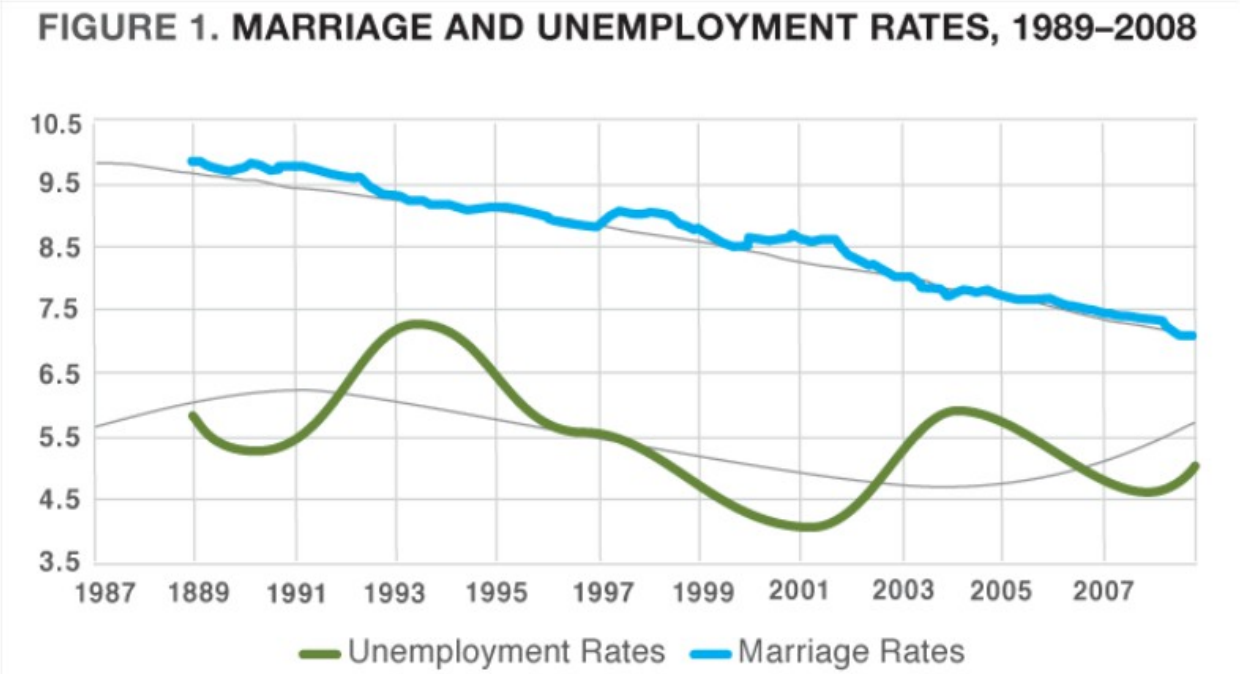
- Τη συνιστώσα της κύριας τάσης T_t (secular trend) η οποία περιγράφει τη μακροχρόνια εξέλιξη της χρονοσειράς (σχεδιάζεται ως μία ευθεία)



Διάγραμμα 9: Η κύρια τάση είναι το βέλος που δείχνει τη γενική πορεία της χρονοσειράς

Παρατήρηση : Δεν είναι απαραίτητο για μία χρονοσειρά να έχει κάποια κύρια τάση αύξησης ή μείωσης των τιμών της.

- Τη συνιστώσα της κυκλικής εναλλαγής C_t (cyclical fluctuation) η οποία περιγράφει επαναλαμβανόμενες μη περιοδικές μεταβολές στη χρονοσειρά.

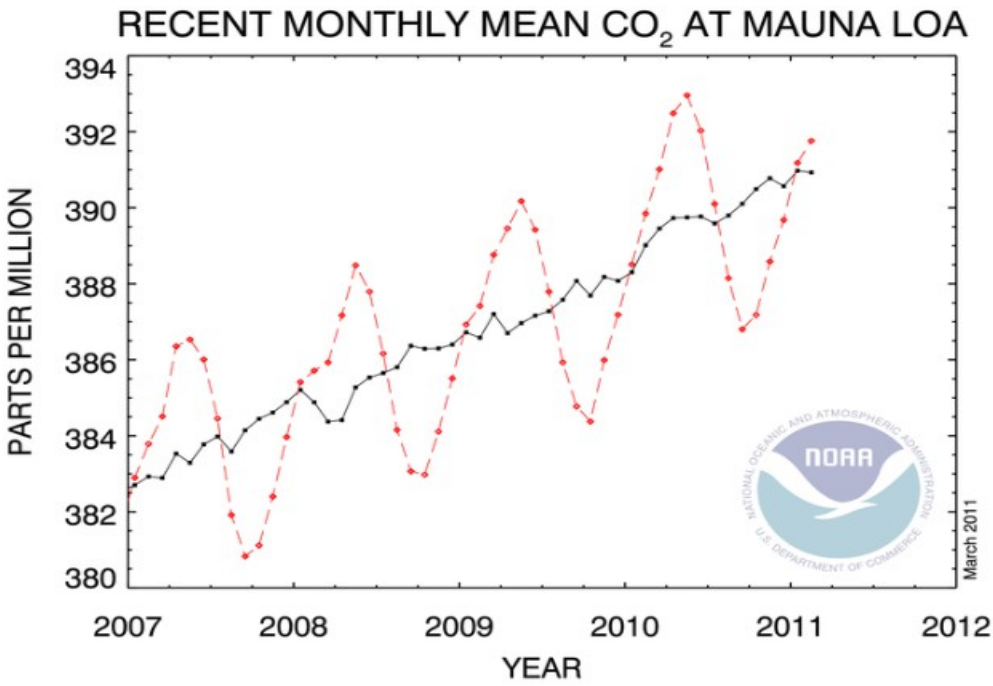


Διάγραμμα 10:Σύγκριση γάμων ανά 1.000 άτομα πληθυσμού και ποσοστού ανεργίας ενηλίκων. Μέθοδος : κινούμενος μέσος 24 μηνών.

Πηγή : http://stateofourunions.org/2009/marriage_and_the_recession.php

Παρατήρηση : Σημειώνεται πως για να χαρακτηριστεί μία εναλλαγή γύρω από τη γραμμή τάσης ως κυκλική εναλλαγή δεν πρέπει να είναι αυτή περιοδική. Δηλαδή, στην πράξη πρέπει τα σημεία της χρονοσειράς να βρίσκονται κάτω από τη γραμμή τάσης για μια σειρά ετών και στη συνέχεια πάνω από τη γραμμή τάσης για την επόμενη σειρά ετών πάνω από τη γραμμή τάσης. Τα δύο χρονικά διαστήματα δεν πρέπει να είναι σταθερά, καθώς αν είναι σταθερά τότε η εναλλαγή ίσως αρμόζει να ονομαστεί περιοδική (ή εποχιακή).

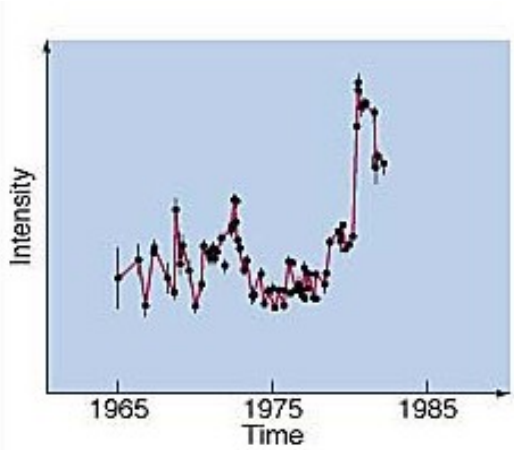
- Την εποχιακή συνιστώσα (seasonal variation) η οποία ορίζεται από τις εποχιακές (περιοδικές) μεταβολές



Διάγραμμα 11: Εποχιακές διακυμάνσεις της παγκόσμιας έκλυσης CO2. Η εποχιακότητα ανά τρίμηνο του έτους είναι εμφανής.

Πηγη : http://joannenova.com.au/2011/04/is-man-made-co2-different-1000-years-try-4-years/

- Τη συνιστώσα του τυχαίου σφάλματος της χρονοσειράς (irregular variations) η οποία περιέχει όλες τις μεταβολές που δεν ανήκουν στις προηγούμενες συνιστώσες.



Διάγραμμα 12: Ακανόνιστες μεταβολές της φωτεινότητας ενός γαλαξία Shefert σε περίοδο 20 ετών.

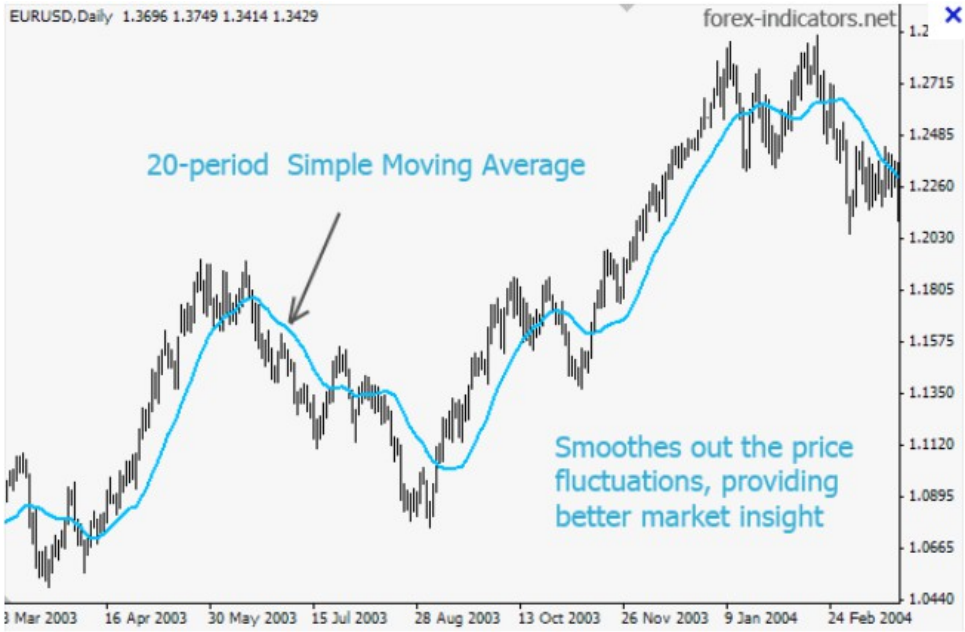
Πηγή : http://obsn3.on.br/~jlkm/astron2e/AT_MEDIA/CH25/CHAP25AT.HTM

3.3 Μέθοδοι προσδιορισμού της κύριας τάσης

Ο προσδιορισμός της κύριας τάσης (αν αυτή υπάρχει) είναι το πρώτο βήμα στη μελέτη μίας χρονοσειράς.

3.3.1 Μέθοδος του κινητού μέσου

Για τον εντοπισμό της κύριας τάσης υπάρχουν πολλές μέθοδοι, απλές και περισσότερο σύνθετες. Ως απλή μέθοδος θεωρείται η μέθοδος του κινητού μέσου κατά την οποία απλά υπολογίζεται η μέση τιμή των τιμών της χρονοσειράς σε ορισμένο χρονικό διάστημα. Το διάγραμμα του κινητού μέσου πολλές φορές φανερώνει τη γενική εξέλιξη της χρονοσειράς.



Διάγραμμα 13: Παράδειγμα υπολογισμού κινούμενου μέσου όρου.
Πηγή : http://forex-indicators.net/trend-indicators

Παράδειγμα

Δίνεται η χρονοσειρά :

Χρόνος	1	2	3	4	5	6	7	8	9	10
Τιμή	3	8	25	20	22	16	14	6	12	15

Να βρεθεί ο κινούμενος μέσος όρος 3 σημείων.

Λύση

Ονομάζουμε x_i τις αρχικές παρατηρήσεις και y_i τις παρατηρήσεις του κινούμενου χρονικού μέσου. Θέλουμε τον κινούμενο μέσο όρο 3 σημείων άρα πρέπει για κάθε μία χρονική στιγμή να υπολογίσουμε τη μέση τιμή των τριών τελευταίων παρατηρήσεων. Είναι κατανοητό πως δεν μπορεί

να υπολογιστεί τέτοιος μέσος όρος για την πρώτη και τη τελευταία παρατήρηση άρα οι υπολογισμοί μας ξεκινούν από την τρίτη παρατήρηση.

Υπολογίζουμε

$y_3=\frac{x_1+x_2+x_3}{3}=\frac{3+8+25}{3}=\frac{36}{3}=12$ και ανάλογα $y_4=\frac{x_2+x_3+x_4}{3}=\frac{8+25+20}{3}=\frac{53}{3}=17{,}7$

$y_5=\frac{x_3+x_4+x_5}{3}=\frac{25+20+22}{3}=\frac{67}{3}=22{,}3$, $y_6=\frac{x_4+x_5+x_6}{3}=\frac{20+22+16}{3}=\frac{58}{3}=19{,}3$,

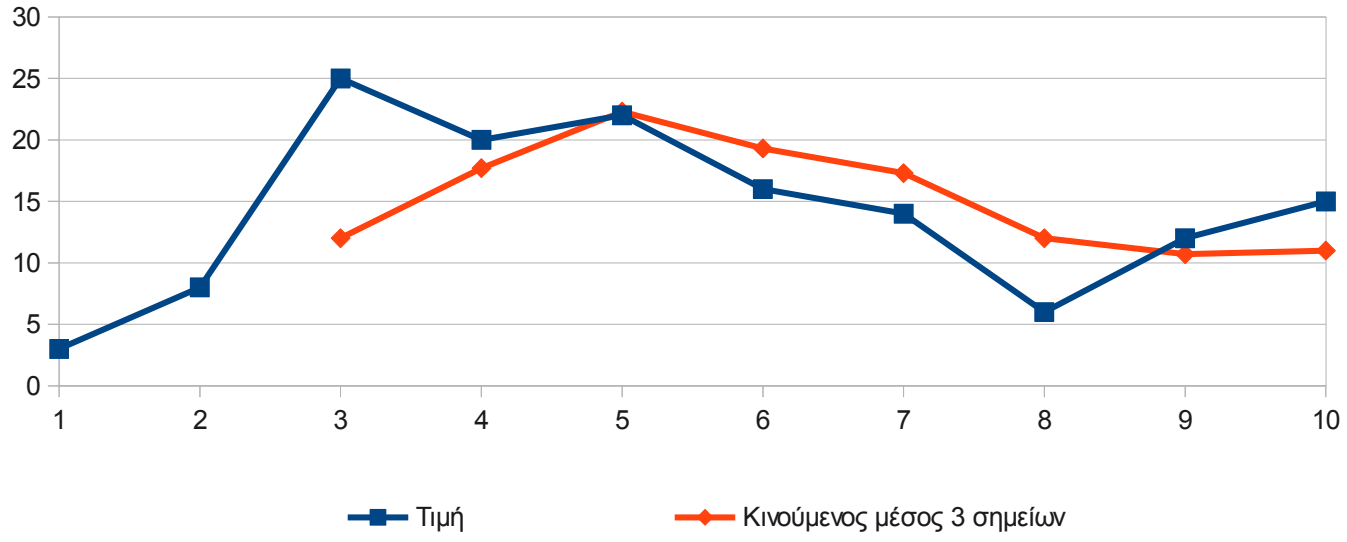
$y_7=\frac{x_5+x_6+x_7}{3}=\frac{22+16+14}{3}=\frac{52}{3}=17{,}3$, $y_8=\frac{x_6+x_7+x_8}{3}=\frac{16+14+6}{3}=\frac{36}{3}=12$,

$y_9=\frac{x_7+x_8+x_9}{3}=\frac{14+6+12}{3}=\frac{32}{3}=10{,}7$, $y_{10}=\frac{x_8+x_9+x_{10}}{3}=\frac{6+12+15}{3}=\frac{33}{3}=11$.

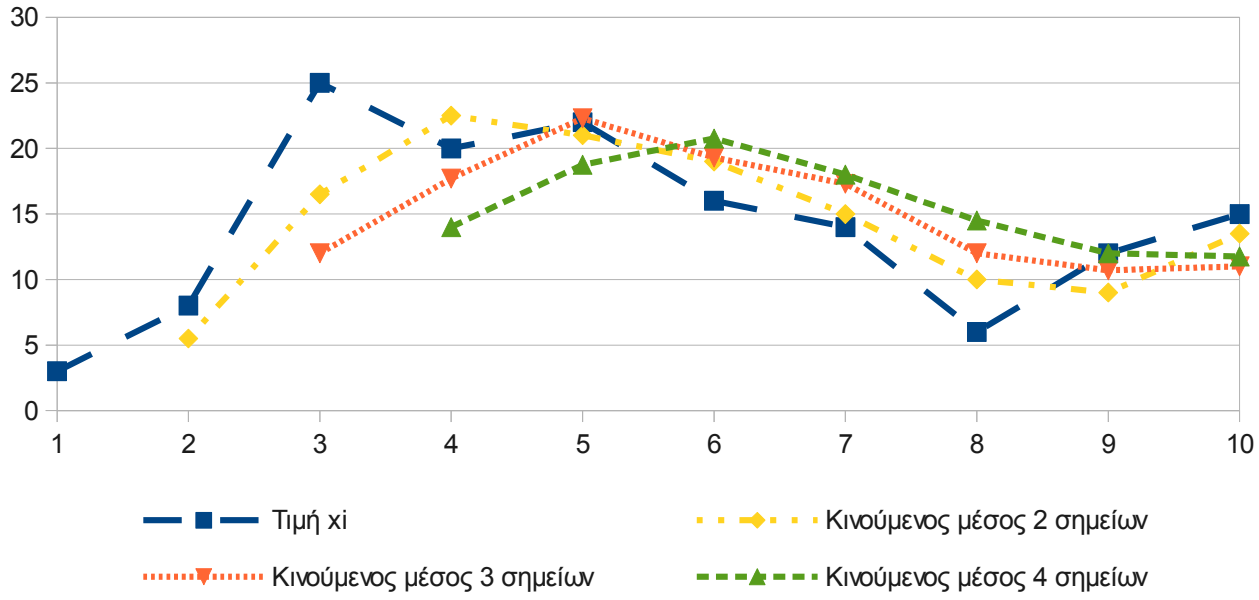
Συμπληρώνουμε τον πίνακα με τις τιμές του κινούμενου μέσου 3 σημείων.

Χρόνος	1	2	3	4	5	6	7	8	9	10
Τιμή	3	8	25	20	22	16	14	6	12	15
Κινούμενος μέσος 3 σημείων			12	17,7	22,3	19,3	17,3	12	10,7	11

Το διάγραμμα 14 παρουσιάζει τις αρχικές τιμές μαζί με τον κινούμενο μέσο όρο 3 σημείων. Στο διάγραμμα 15, σελίδα 31, παρουσιάζεται η ιδιότητα του κινούμενου μέσου όρου να “απορροφά” εύκολα τις μεγάλες διακυμάνσεις της χρονοσειράς με την κατάλληλη αύξηση των στοιχείων της σειράς που χρησιμοποιούνται για τον υπολογισμό του.



Διάγραμμα 14: Διάγραμμα των τιμών και του κινούμενου μέσου τριών σημείων



Διάγραμμα 15: Όσο μεγαλώνει το πλήθος των σημείων που χρησιμοποιείται για τον υπολογισμό του κινούμενου μέσου όρου τόσο πιο ομαλή γίνεται η γραφική παράσταση του.

3.3.2 Μέθοδος της ευθείας των ελαχίστων τετραγώνων

Περισσότερο πολύπλοκη μέθοδος που απαιτεί υπολογιστή είναι η μέθοδος της ευθείας των ελαχίστων τετραγώνων με την οποία απλά υπολογίζεται η ευθεία γραμμικής παλινδρόμησης της εξαρτημένης μεταβλητής πάνω στο χρόνο (δηλαδή ο χρόνος t παίρνει το ρόλο της ανεξάρτητης μεταβλητής στο γραμμικό μοντέλο)

Η εξίσωσή της ευθείας ελαχίστων τετραγώνων, έχει τη μορφή

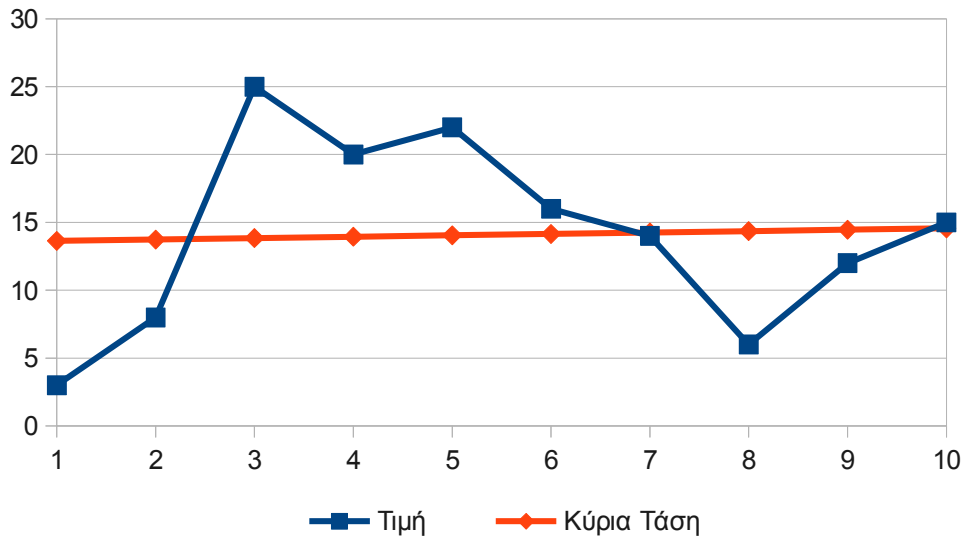
X = α * t+β.

Οι συντελεστές του γραμμικού μοντέλου υπολογίζονται από τις τιμές της μεταβλητής X ως εξής :

$$\alpha=\frac{\sum_{i=1}^{v}(t_i-\bar{t})(x_i-\bar{x})}{\sum_{i=1}^{v}(t_i-\bar{t})^2} \quad \text{και} \quad \beta=\bar{x}-\alpha\cdot\bar{t}$$

Στον πίνακα που ακολουθεί παρουσιάζονται οι απαραίτητοι υπολογισμοί, ενώ στο διάγραμμα 16, σελίδα 32, εμφανίζεται η κύρια τάση των χρονικών δεδομένων του παραδείγματος.

Χρόνος	1	2	3	4	5	6	7	8	9	10	Σύνολο
Τιμή	3	8	25	20	22	16	14	6	12	15	
(xi – x)(ti - t)	50,0	21,4	-27,3	-8,9	-4,0	1,0	-0,1	-20,3	-7,4	4,1	8,5
(ti - t)^2	20,3	12,3	6,3	2,3	0,3	0,3	2,3	6,3	12,3	20,3	82,5
Κύρια Τάση	13,6	13,7	13,8	13,9	14,0	14,2	14,3	14,4	14,5	14,6	



Διάγραμμα 16: Κύρια τάση της χρονοσειράς του παραδείγματος

Ακόμα περισσότερο το ίδιο μπορεί να γίνει προς την κατεύθυνση ενός μη γραμμικού μοντέλου εξέλιξης στο χρόνο όπου ακολουθείται η γενικότερη μέθοδος προσαρμογής καμπύλης στα δεδομένα, καμπύλη που την εντοπίζει ο ερευνητής ανάλογα με τα δεδομένα του προβλήματος.

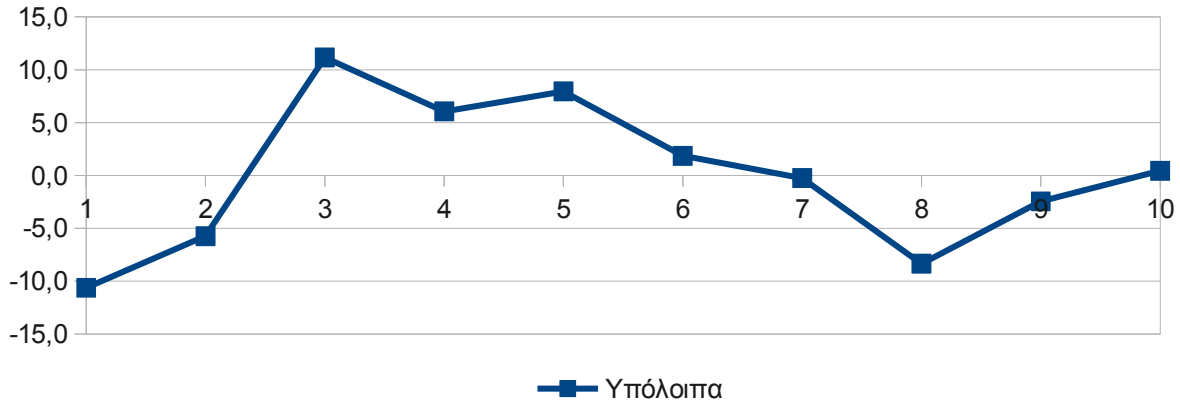
3.4 Προσδιορισμός της συνιστώσας της κυκλικής εναλλαγής

Ο τρόπος με τον οποίο βρίσκουμε την κυκλική μεταβολή – αν αυτή υπάρχει – είναι η μέθοδος των υπολοίπων (ή καταλοίπων) (residuals).

Στον πίνακα που ακολουθεί εμφανίζονται οι υπολογισμοί των υπολοίπων και των σχετικών (ως προς την πραγματική τιμή) υπολοίπων ενώ στα διαγράμματα 17 και 18, σελίδα 33 παρουσιάζεται η εξέλιξη των υπολοίπων και των σχετικών υπολοίπων με την πάροδο των χρονικών στιγμών.

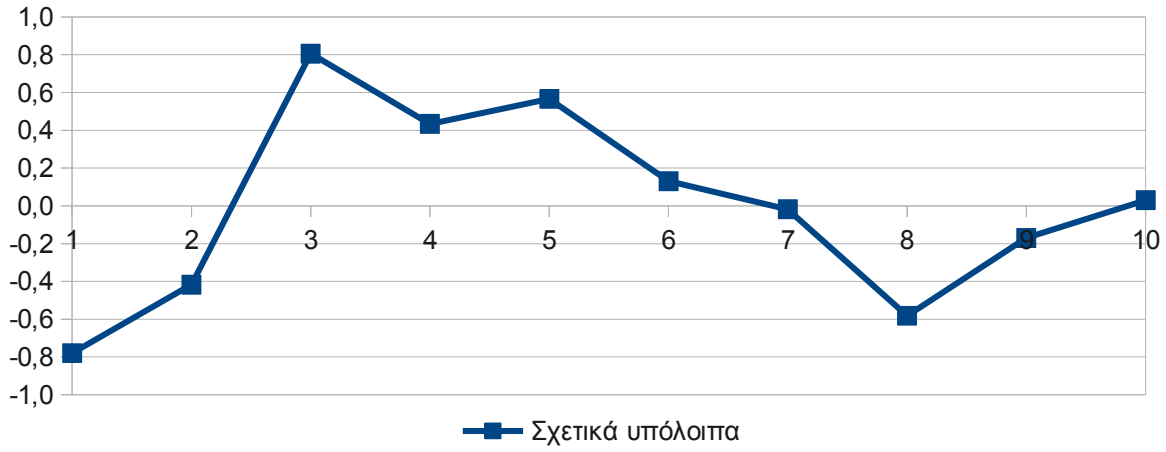
Χρόνος	1	2	3	4	5	6	7	8	9	10
Τιμή	3	8	25	20	22	16	14	6	12	15
Κύρια Τάση	13,6	13,7	13,8	13,9	14,0	14,2	14,3	14,4	14,5	14,6
Υπόλοιπα	-10,6	-5,7	11,2	6,1	8,0	1,8	-0,3	-8,4	-2,5	0,4
Σχετικά υπόλοιπα	-0,8	-0,4	0,8	0,4	0,6	0,1	0,0	-0,6	-0,2	0,0

Πίνακας 3.1: Υπόλοιπο = Τιμή - Κύρια τάση
Σχετικό υπόλοιπο = (Τιμή – Κύρια τάση) / Κύρια τάση



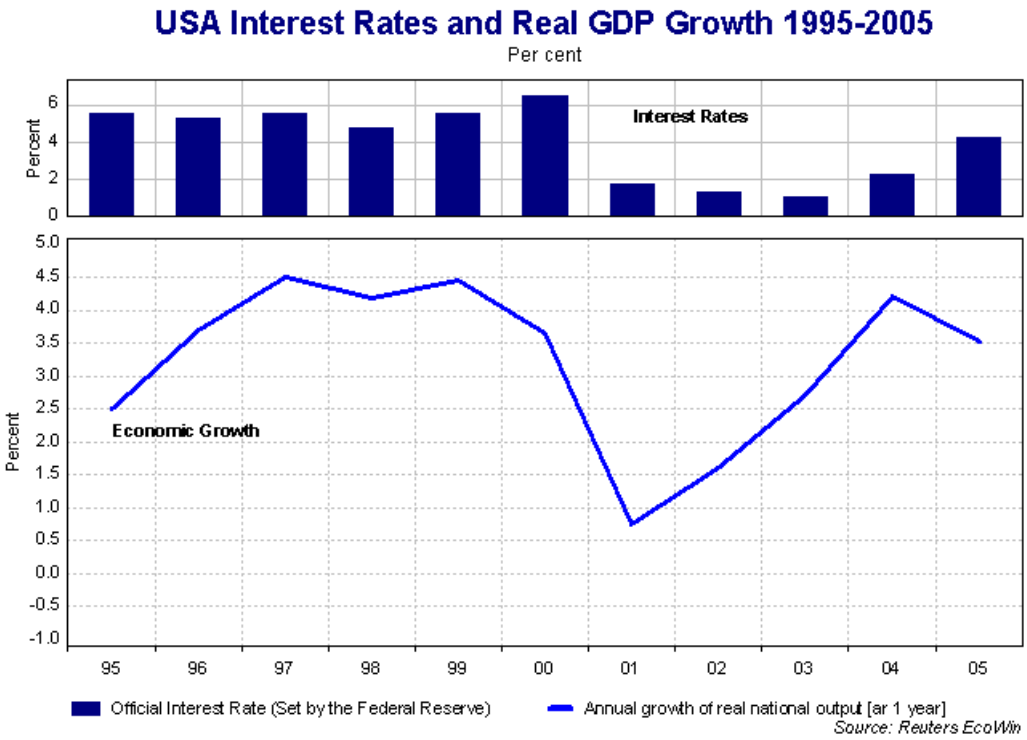
Διάγραμμα 17: Διάγραμμα υπολοίπων της χρονοσειράς

Το διάγραμμα σχετικών υπολοίπων δίνει περισσότερες πληροφορίες σε σχέση με το απλό διάγραμμα υπολοίπων καθώς δεν εξαρτάται από το σύστημα μέτρησης των τιμών της μεταβλητής.



Διάγραμμα 18: Διάγραμμα σχετικών υπολοίπων της χρονοσειράς

Η παρατήρηση της κυκλικής συνιστώσας μίας χρονοσειράς μπορεί να οδηγήσει σε καλύτερη κατανόηση των παραγόντων που την επηρεάζουν και στη λήψη των κατάλληλων αποφάσεων για μεταβολή της μελλοντικής της πορείας. Χαρακτηριστικό παράδειγμα είναι η σχέση τραπεζικών επιτοκίων και ανάπτυξης όπως αποτυπώνεται στο επόμενο διάγραμμα.



Διάγραμμα 19: Παράδειγμα κυκλικής μεταβολής. Πηγή : http://tutor2u.net/economics/revision-notes/a2-macro-cyclical-fluctuations.html

3.5 Προσδιορισμός της εποχιακής συνιστώσας

Η εποχιακή συνιστώσα γίνεται άμεσα εμφανής από το χρονοδιάγραμμα. Με οπτική παρατήρηση γίνεται φανερό αν υπάρχει επανάληψη των τιμών ανά εβδομάδα, μήνα, τρίμηνο, τετράμηνο, εξάμηνο ή ως προς οποιαδήποτε μονάδα μέτρησης έχει νόημα στο πλαίσιο που τίθεται η χρονοσειρά.

Ο πιο απλός τρόπος για να εντοπιστεί η συνεισφορά της εποχικότητας στην τιμή της χρονοσειράς είναι ο υπολογισμός του **κεντρικού** κινούμενου μέσου όρου των γειτονικών όρων σε πλήθος που να καλύπτει μία πλήρη εποχική μεταβολή. Έτσι για παράδειγμα αν έχουμε δεδομένα ημερησίων πωλήσεων και παρατηρούμε εβδομαδιαία επανάληψη τότε για να εκτιμήσουμε αυτή την περιοδικότητα θα υπολογίζαμε για τη Δευτέρα, τη μέση τιμή των πωλήσεων από την προηγούμενη

Παρασκευή έως και την επόμενη Πέμπτη, για την ήμερα Τρίτη, θα υπολογίζαμε από το Σάββατο έως την Παρασκευή κτλ. Ανάλογα, αν συλλέγαμε μηνιαία δεδομένα και παρατηρούσαμε ετήσια περιοδικότητα τότε θα υπολογίζαμε το κινούμενο μέσο όρο 12 μηνών κλπ.

Παρατήρηση : Πρέπει να τονιστεί πως όταν το πλήθος των χρονικών περιόδων είναι άρτιο (όπως συμβαίνει στους 12 μήνες) τότε δεν μπορούμε να πάρουμε ίσο πλήθος περιόδων πριν και μετά, περίπτωση κατά την οποία χωρίζουμε ανισομερώς τα δύο διαστήματα. Το σφάλμα της εκτίμησης θα είναι μικρό.

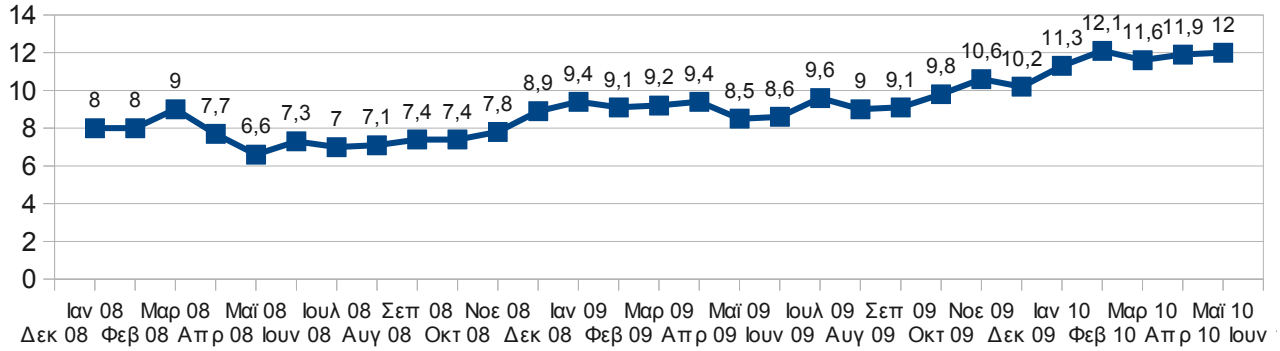
Η διαφορά που θα προκύπτει μεταξύ της πραγματικής τιμής και του κινούμενου μέσου όρου μπορεί να θεωρηθεί η περιοδική (εποχιακή) συνιστώσα της χρονοσειράς.

Παράδειγμα

Στον παρακάτω πινάκα δίνεται το ποσοστό ανεργίας στην Ελλάδα από τον Ιανουάριο του 2008 έως το Μάιο του 2010. Τα ίδια δεδομένα παρουσιάζονται στο διάγραμμα 20,σελίδα 35 που ακολουθεί.

Μήνας	Ποσοστό	Μήνας	Ποσοστό	Μήνας	Ποσοστό
Ιαν 08	8	Ιαν 09	9,4	Ιαν 10	11,3
Φεβ 08	8	Φεβ 09	9,1	Φεβ 10	12,1
Μαρ 08	9	Μαρ 09	9,2	Μαρ 10	11,6
Απρ 08	7,7	Απρ 09	9,4	Απρ 10	11,9
Μαϊ 08	6,6	Μαϊ 09	8,5	Μαϊ 10	12
Ιουν 08	7,3	Ιουν 09	8,6		
Ιουλ 08	7	Ιουλ 09	9,6		
Αυγ 08	7,1	Αυγ 09	9		
Σεπ 08	7,4	Σεπ 09	9,1		
Οκτ 08	7,4	Οκτ 09	9,8		
Νοε 08	7,8	Νοε 09	10,6		
Δεκ 08	8,9	Δεκ 09	10,2		

Να εκτιμηθεί η εποχιακή συνιστώσα της ανεργίας με τη μέθοδο του κινούμενου μέσου όρου.



Διάγραμμα 20: Ποσοστό ανεργίας στην Ελλάδα (Ιανουάριος 2008 έως Μάιος 2010)

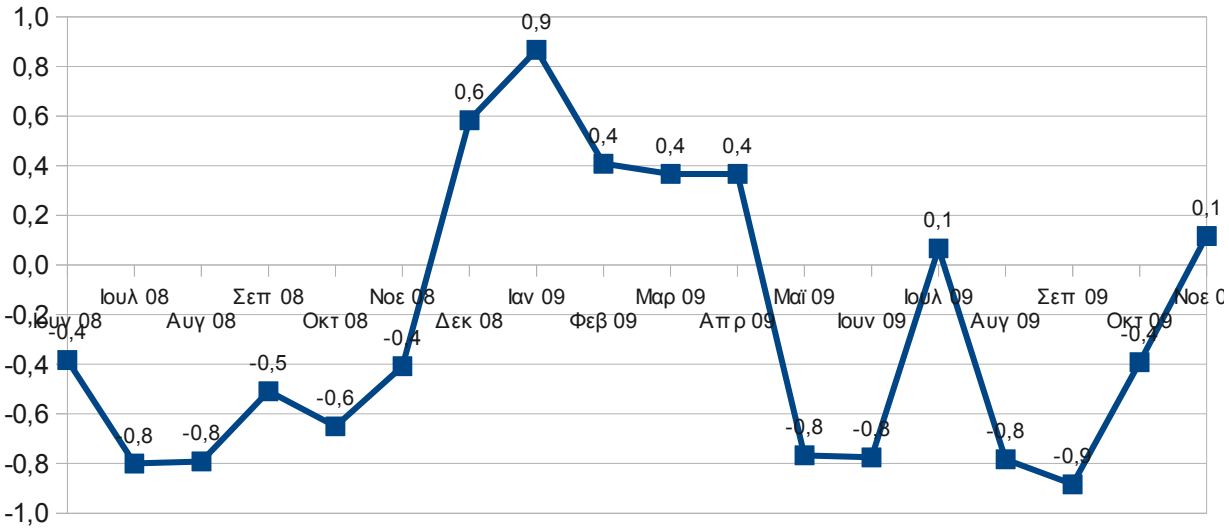
Λύση

Είναι φανερό πως η ανεργία στην Ελλάδα, λόγω της τουριστικής απασχόλησης η οποία συμβαίνει κατά κύριο λόγο το καλοκαίρι, είναι φαινόμενο που μεταβάλλεται σε ετήσια βάση. Με αυτήν την εκτίμηση καταλαβαίνουμε πως ο κινούμενος μέσος όρος πρέπει να υπολογίζεται από το σύνολο των 12 μηνών του έτους. Καθώς το πλήθος είναι ζυγό επιλέγουμε τους 5 προηγούμενους μήνες και τους 6 επόμενους. Αυτό σημαίνει πως η πρώτη τιμή για τον κεντρικό κινούμενο μέσο όρο θα υπολογιστεί για τον 6ο μήνα των δεδομένων δηλαδή για τον Ιούνιο του 2008 ενώ ο τελευταίος μήνας υπολογισμού θα είναι ο Νοέμβριος του 2009.

Οι υπολογισμοί γίνονται εύκολα με κάποιον υπολογιστή και δίνουν τον επόμενο πίνακα

	Ποσοστό	**Κεντρικός κινούμενος μέσος**	**Διαφορά (εποχιακή συνιστώσα της ανεργίας)**
Ιουν 08	7,3	7,7	-0,4
Ιουλ 08	7	7,8	-0,8
Αυγ 08	7,1	7,9	-0,8
Σεπ 08	7,4	7,9	-0,5
Οκτ 08	7,4	8,1	-0,7
Νοε 08	7,8	8,2	-0,4
Δεκ 08	8,9	8,3	0,6
Ιαν 09	9,4	8,5	0,9
Φεβ 09	9,1	8,7	0,4
Μαρ 09	9,2	8,8	0,4
Απρ 09	9,4	9,0	0,4
Μαϊ 09	8,5	9,3	-0,8
Ιουν 09	8,6	9,4	-0,8
Ιουλ 09	9,6	9,5	0,1
Αυγ 09	9	9,8	-0,8
Σεπ 09	9,1	10,0	-0,9
Οκτ 09	9,8	10,2	-0,4
Νοε 09	10,6	10,5	0,1

Στην επόμενη σελίδα παρουσιάζεται το διάγραμμα 21 της διαφοράς του παρατηρούμενου ποσοστού ανεργίας από την τιμή του κεντρικού κινούμενου μέσου το οποίο δεν είναι παρά μία καλή εκτίμηση της εποχιακής συνιστώσας της ανεργίας.



Διάγραμμα 21: Εποχιακή συνιστώσα της ανεργίας

3.6 Συνιστώσα του τυχαίου σφάλματος της χρονοσειράς

Η συνιστώσα του τυχαίου σφάλματος της χρονοσειράς δεν είναι τίποτα άλλο παρά ότι απομένει αν από την τιμή της χρονοσειράς αφαιρέσουμε την κύρια τάση, την κυκλική εναλλαγή και την εποχιακή συνιστώσα.

Το μέγεθος του τυχαίου σφάλματος δίνει το μέγεθος της αξιοπιστίας της πρόβλεψης της εξέλιξης της χρονοσειράς στο μέλλον, χωρίς να υπάρχει ωστόσο κάποιος κανόνας αξιολόγησης της.

3.7 Αυτοδιακύμανση και αυτοσυσχέτιση

Τόσο η αυτοδιακύμανση όσο και η αυτοσυσχέτιση είναι δύο στατιστικές ποσότητες που μας επιτρέπουν να ανιχνεύσουμε τυχόν περιοδικότητα μίας χρονοσειράς με ένα μετρήσιμο τρόπο.

Η αυτοδιακύμανση είναι η συνδιακύμανση της μεταβλητής που ορίζει τη χρονοσειρά με ένα αντίγραφό της χρονικά μετατοπισμένο κατά πλήθος χρονικών στιγμών που επιλέγεται από τον ερευνητή.

Η αυτοσυσχέτιση είναι η κανονικοποιημένη ως προς το πιθανό εύρος τιμών εκδοχή της αυτοδιακύμανσης.

Στην πράξη, υπολογίζουμε την αυτοσυσχέτιση της χρονοσειράς με τον εαυτό της σε όλες τις πιθανές μετατοπίσεις και παρατηρούμε το σύνολο των τιμών που προκύπτουν. Η συνδιακύμανση στα πρώτα βήματα, θα μειώνεται από το 1 μέχρι κάποιο σημείο όπου θα αρχίσει να ανέρχεται ξανά (αν υπάρχει κάποια περιοδικότητα). Στο σημείο όπου η αυτοσυσχέτιση θα ξαναγίνει θετική και ισχυρή, θα ορίζεται η περίοδος της περιοδικής συνιστώσας της χρονοσειράς.

3.7.1 Αυτοδιακύμανση

Θυμίζουμε πως ο τύπος της συνδιακύμανσης δύο μεταβλητών δίνεται από τον τύπο (παράγραφο 2.2, σελίδα 11) :

$$s_{XY}^2=\frac{1}{\nu}\sum_{i=1}^{\nu}(x_i-\bar{x})(y_i-\bar{y})$$ (τύπος συνδιακύμανσης)

Αν x_1, x_2, ..., x_κ είναι οι παρατηρήσεις μίας μεταβλητής στις χρονικές στιγμές 1, 2, ..., κ, τότε οι μετατοπισμένες παρατηρήσεις κατά τ στιγμές $(0\leq\tau\leq\kappa)$ θα είναι οι $x_{1+\tau}$, $x_{2+\tau}$, ..., x_κ . Αν τις τοποθετήσουμε κατά ζεύγη, μαζί με τις αρχικές θα πάρουμε τον εξής πίνακα

Θέση	**1**	**2**	**3**	**...**	**κ-τ-1**	**κ-τ**	**κ-τ+1**	**...**	**κ-1**	**κ**
Αρχική χρονοσειρά	x_1	x_2	x_3	…	$x_{\kappa-\tau-1}$	$x_{\kappa-\tau}$	$x_{\kappa-\tau+1}$	…	$x_{\kappa-1}$	x_κ
Μετατοπισμένη χρονοσειρά	$x_{1+\tau}$	$x_{2+\tau}$	$x_{3+\tau}$	…	$x_{\kappa-1}$	x_κ				

Παρατηρούμε πως υπάρχουν πλέον μόνο κ – τ – 1 ζευγάρια τιμών από τα οποία μπορεί να υπολογισθεί η συνδιακύμανση. Τοποθετώντας στον τύπο της συνδιακύμανσης, στη θέση των τιμών x_i τα αρχικά στοιχεία της χρονοσειράς και στη θέση των τιμών y_i τις μετατοπισμένες παρατηρήσεις παίρνουμε τον τύπο της αυτοδιακύμανσης :

$$C_{XX}=\frac{1}{\kappa-\tau}\sum_{i=1}^{\kappa-\tau}(x_i-\bar{x})(x_{i+\tau}-\bar{x}_\tau)$$ (τύπος δειγματικής συνδιακύμανσης)

όπου $\bar{x}=\frac{1}{\kappa}\sum_{i=1}^{\kappa}x_i$ και $\bar{x}_\tau=\frac{1}{\kappa-\tau}\sum_{i=1+\tau}^{\kappa}x_i=\frac{1}{\kappa-\tau}\sum_{i=1}^{\kappa-\tau}x_{i+\tau}$

Παρατήρηση : Ο παραπάνω τύπος δεν είναι αμερόληπτος εκτιμητής της αυτοδιακύμανσης καθώς ο υπολογισμός των μέσων τιμών της χρονοσειράς και της μετατόπισής της κατά τ μονάδες γίνεται από το ίδιο το δείγμα. Αν με κάποιον άλλο τρόπο είναι γνωστή η μέση τιμή όλης της χρονοσειράς (έστω μ αυτή) τότε εναλλακτικά χρησιμοποιούμε τον τύπο

$$C_{XX}=\frac{1}{\kappa-\tau}\sum_{i=1}^{\kappa-\tau}(x_i-\mu)(x_{i+\tau}-\mu)$$ (αμερόληπτος τύπος συνδιακύμανσης)

Στη γλώσσα της θεωρίας πιθανοτήτων επιπλέον γράφουμε :

$$C_{XX}(t,s)=E[(X_t-\mu_t)(X_s-\mu_s)]=E[X_tX_s]-\mu_t\mu_s$$

όπου η διαφορά των s και t είναι αυτή που υποθέτει ο ερευνητής πως καθορίζει την περιοδική συμπεριφορά της χρονοσειράς.

.

3.7.2 Αυτοσυσχέτιση

Αν x_1, x_2, ..., x_κ είναι οι παρατηρήσεις μίας μεταβλητής στις χρονικές στιγμές 1, 2, ..., κ, τότε η δειγματική αυτοσυσχέτιση με διαφορά (τ) ορίζεται να είναι

$$R_\tau = \frac{1}{(\kappa-\tau)\sigma^2}\sum_{i=1}^{\kappa-\tau}(x_i-\mu)(x_{i+\tau}-\mu) \quad ,$$

όπου μ και σ η μέση τιμή και η τυπική απόκλιση της χρονοσειράς. Καθώς στην πράξη σπάνια γνωρίζουμε τις τιμές αυτές τις αντικαθιστούμε με τις αντίστοιχες δειγματικές ποσότητες, δηλαδή χρησιμοποιούμε τον τύπο

$$R_\tau = \frac{1}{(\kappa-\tau)s^2}\sum_{i=1}^{\kappa-\tau}(x_i-\bar{x})(x_{i+\tau}-\bar{x}_\tau)$$

εκτίμηση που δεν είναι η καλύτερη δυνατή καθώς δεν είναι αμερόληπτη αλλά είναι η μόνη που μπορούμε να υλοποιήσουμε σε αυτήν την περίπτωση.

Παρατηρήσεις

1. Εύκολα αποδεικνύεται πως (τ = t-s = σταθερό) :

$$C_{XX}(\tau) = E[(X(t)-\mu)(X(t+\tau)-\mu)] = E[X(t)X(t+\tau)] - \mu^2 = R_\tau - \mu^2 \quad .$$

2. Αν η χρονοσειρά είναι συνεχής, τότε αναπαριστάται από μία συνεχής συνάρτηση X(t). Σε αυτήν την περίπτωση η αυτοσυσχέτιση ορίζεται ως

$$C(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T X(t)X(t+\tau)\,dt$$

Παράδειγμα 1

Θα χρησιμοποιήσουμε τα δεδομένα ανεργίας 29 συνεχόμενων μηνών του προηγούμενου παραδείγματος. Είναι φανερό πως η κατάλληλη διαφορά για την περιοδικότητα είναι 12 χρονικές στιγμές.

Οι υπολογισμοί γίνονται εύκολα με το LibreOffice Calc και παρουσιάζονται στον πίνακα της επόμενης σελίδας.

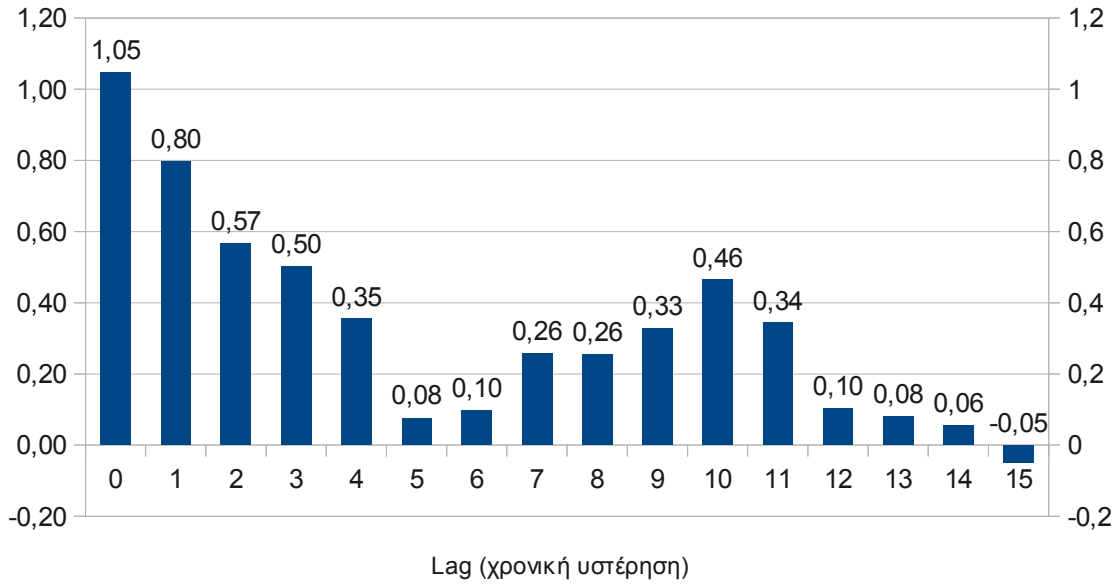
Οι συναρτήσεις που χρησιμοποιούνται είναι οι COVAR() και οι CORREL() σε δύο στήλες που περιέχουν τη χρονοσειρά και τη μετατόπισή της κατά 0 έως και 15 χρονικές στιγμές.

Από το διάγραμμα 23, σελίδα 41, παρατηρούμε πως ισχυρή αυτοσυσχέτιση εμφανίζεται μεταξύ των χρονικών υστερήσεων 10 έως και 13.

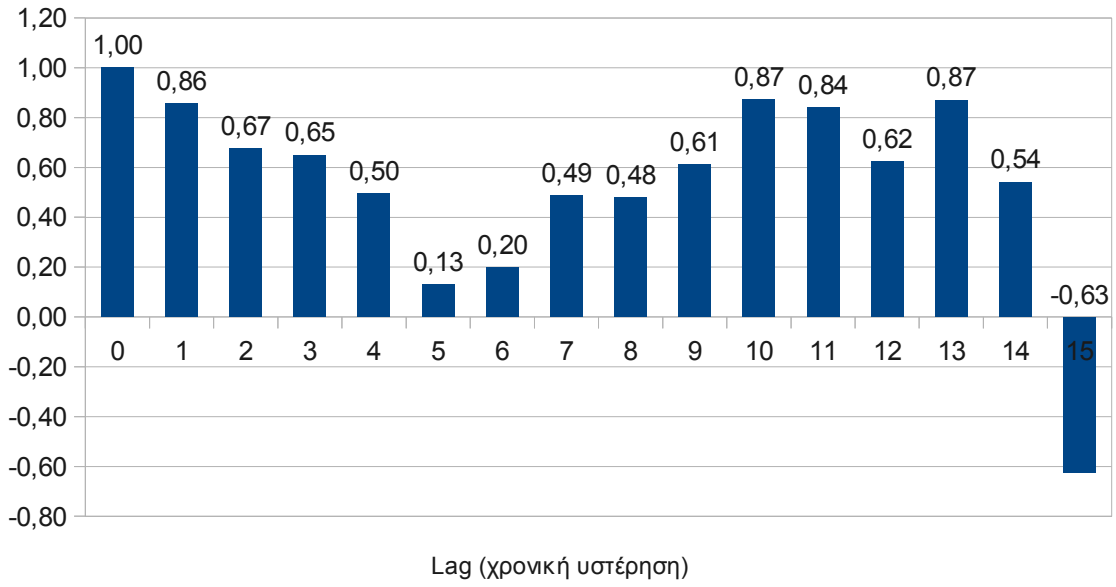
Από την εμπειρία μας και λόγω της φύσης του φαινομένου της ανεργίας, αντιλαμβανόμαστε πως η ορθότερη επιλογή για την περιοδική ερμηνεία του φαινομένου μεταξύ των αριθμών 10, 11, 12 και

13 είναι η υστέρηση 12, δηλαδή η ετήσια περιοδικότητα.

		Lag (χρονική υστέρηση)															
Μήνας	**Ποσοστό**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Ιουν 08	7,3	7,3															
Ιουλ 08	7	7	7,3														
Αυγ 08	7,1	7,1	7	7,3													
Σεπ 08	7,4	7,4	7,1	7	7,3												
Οκτ 08	7,4	7,4	7,4	7,1	7	7,3											
Νοε 08	7,8	7,8	7,4	7,4	7,1	7	7,3										
Δεκ 08	8,9	8,9	7,8	7,4	7,4	7,1	7	7,3									
Ιαν 09	9,4	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3								
Φεβ 09	9,1	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3							
Μαρ 09	9,2	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3						
Απρ 09	9,4	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3					
Μαϊ 09	8,5	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3				
Ιουν 09	8,6	8,6	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3			
Ιουλ 09	9,6	9,6	8,6	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3		
Αυγ 09	9	9	9,6	8,6	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3	
Σεπ 09	9,1	9,1	9	9,6	8,6	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7	7,3
Οκτ 09	9,8	9,8	9,1	9	9,6	8,6	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1	7
Νοε 09	10,6	10,6	9,8	9,1	9	9,6	8,6	8,5	9,4	9,2	9,1	9,4	8,9	7,8	7,4	7,4	7,1
Αυτοδιακύμανση		**1,05**	**0,80**	**0,57**	**0,50**	**0,35**	**0,08**	**0,10**	**0,26**	**0,26**	**0,33**	**0,46**	**0,34**	**0,10**	**0,08**	**0,06**	**-0,05**
Αυτοσυσχέτιση		**1,00**	**0,86**	**0,67**	**0,65**	**0,50**	**0,13**	**0,20**	**0,49**	**0,48**	**0,61**	**0,87**	**0,84**	**0,62**	**0,87**	**0,54**	**-0,63**



Διάγραμμα 22: Αυτοδιακύμανση του ποσοστού ανεργίας



Διάγραμμα 23: Αυτοσυσχέτιση του ποσοστού ανεργίας

Παράδειγμα 2

Στον επόμενο πίνακα παρουσιάζεται το πλήθος ηλιακών κηλίδων από το έτος 1771 έως το 1870.

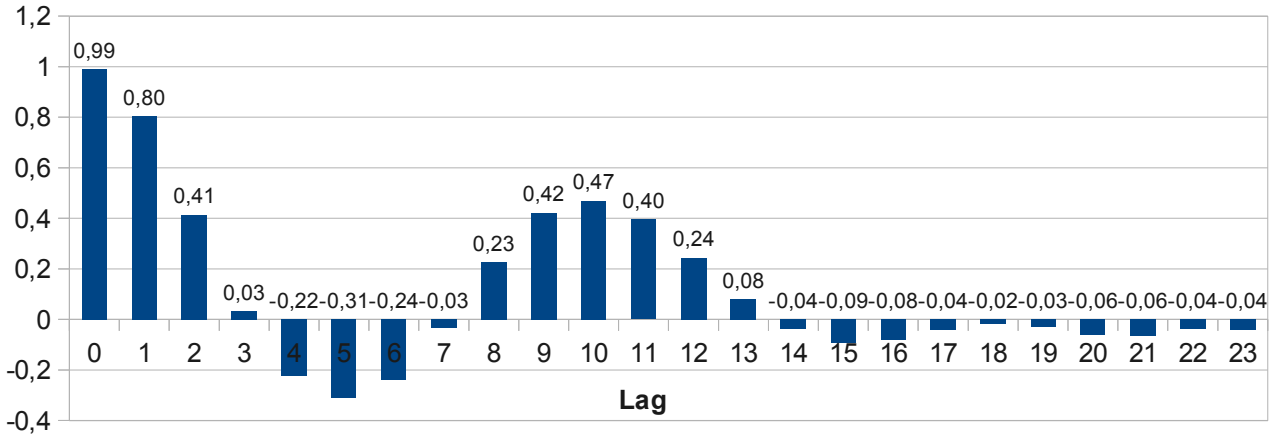
Ηλιακές Κηλίδες									
Έτος	**Πλήθος**	**Έτος**	**Πλήθος**	**Έτος**	**Πλήθος**	**Έτος**	**Πλήθος**	**Έτος**	**Πλήθος**
1771	8	1791	15	1811	47	1831	10	1851	48
1772	17	1792	40	1812	30	1832	24	1852	42
1773	36	1793	62	1813	16	1833	83	1853	28
1774	50	1794	98	1814	7	1834	132	1854	10
1775	64	1795	125	1815	38	1835	131	1855	8
1776	67	1796	96	1816	74	1836	118	1856	2
1777	71	1797	67	1817	101	1837	90	1857	0
1778	48	1798	64	1818	82	1838	67	1858	1
1779	28	1799	54	1819	66	1839	60	1859	5
1780	8	1800	39	1820	35	1840	47	1860	12
1781	13	1801	21	1821	31	1841	41	1861	14
1782	57	1802	7	1822	7	1842	21	1862	35
1783	122	1803	4	1823	20	1843	16	1863	46
1784	138	1804	23	1824	92	1844	6	1864	41
1785	103	1805	55	1825	154	1845	4	1865	30
1786	86	1806	94	1826	126	1846	7	1866	24
1787	65	1807	96	1827	85	1847	14	1867	16
1788	37	1808	77	1828	68	1848	34	1868	7
1789	24	1809	59	1829	38	1849	45	1869	4
1790	11	1810	44	1830	23	1850	43	1870	2

Η χρονική εξέλιξη του πλήθους παρουσιάζεται στο επόμενο διάγραμμα 24. Είναι φανερό πως

υπάρχει περιοδικότητα χωρίς ωστόσο να υπάρχει κάποια γενικότερη αυξητική ή φθίνουσα τάση.

Διάγραμμα 24: Πλήθος ηλιακών κηλίδων

Από το διάγραμμα 25 είναι φανερό ότι η αυτοσυσχέτιση έχει ημιτονοειδή μεταβολή κάτι που οφείλεται στην περιοδικότητα του φαινομένου. Επίσης, η περιοδικότητα είναι ίση με 10 με 11 χρόνια καθώς σε αυτές τις διαφορές εμφανίζεται η μεγαλύτερη θετική συσχέτιση.



Διάγραμμα 25: Αυτοσυσχέτιση του πλήθους ηλιακών κηλίδων με χρονική μετατόπιση από 1 έως και 23 έτη

3.8 Φασματική ανάλυση

Με τον όρο φασματική ανάλυση εννοούνται όλες οι μέθοδοι που προσπαθούν να ερμηνεύσουν τη συμπεριφορά της μέσω των συχνοτήτων που την αποτελούν, δηλαδή μέσω κανονικών περιοδικών συνιστωσών. Μαζί με τη χρονική ανάλυση που περιγράφηκε στις προηγούμενες παραγράφους αποτελούν εργαλεία ανάλυσης χρονοσειρών.

Το κύριο εργαλείο ανάλυσης μίας χρονοσειράς σε στοιχειώδες κανονικές περιοδικές συνιστώσες είναι η θεωρία Fourier σε διακριτή και συνεχής μορφή. Σύμφωνα με τη θεωρία Fourier αν X(t) είναι μία συνεχής περιοδική συνάρτηση με περίοδο T τότε

$$X(t)=\frac{\alpha_0}{2}+\sum_{i=1}^{\infty}\alpha_i \sigma\upsilon\nu\left(\frac{2\pi i}{T}t\right)+\beta_i \eta\mu\left(\frac{2\pi i}{T}t\right) \quad ,$$

όπου $\alpha_i=\frac{2}{T}\int_0^T X(t)\sigma\upsilon\nu\left(\frac{2\pi i}{T}\right)dt \quad , \quad \beta_i=\frac{2}{T}\int_0^T X(t)\eta\mu\left(\frac{2\pi i}{T}\right)dt \quad .$

Θεωρώντας την χρονοσειρά x_1, x_2, ..., x_κ που έχει συλλεχθεί ως διακριτές τιμές μίας συνεχής περιοδικής χρονοσειράς την οποία δεν γνωρίζουμε και αναζητούμε, τότε μπορούμε να βρούμε τη σειρά Fourier που διέρχεται από τις παρατηρήσεις. Αποδεικνύεται ότι στην περίπτωση αυτή :

$$X(t)=A_0+\sum_{i=1}^{\kappa/2}A_i \sigma\upsilon\nu\left(\frac{2\pi i}{T}t\right)+\sum_{i=1}^{\kappa/2-1}B_i \eta\mu\left(\frac{2\pi i}{T}t\right) \quad ,$$

όπου $A_0=\frac{1}{\kappa}\sum_{i=1}^{\kappa}x_i=\bar{x} \quad , \quad A_i=\frac{2}{\kappa}\sum_{q=1}^{\kappa}x_q \sigma\upsilon\nu\left(\frac{2\pi iq}{\kappa}\right) \quad , \quad i=1,2,\ldots,\frac{\kappa}{2}-1 \quad , \quad A_{\frac{\kappa}{2}}=\frac{1}{\kappa}\sum_{q=1}^{\kappa}x_q \sigma\upsilon\nu\, q\pi$

και $B_i=\frac{2}{\kappa}\sum_{q=1}^{\kappa}x_q \eta\mu\left(\frac{2\pi iq}{\kappa}\right) \quad , \quad i=1,2,\ldots,\frac{\kappa}{2}-1$

Η γενική κατεύθυνση είναι πως οι συνιστώσες που ορίζουν την πορεία της χρονοσειράς έχουν αντίστοιχους συντελεστές στατιστικά διάφορους από το μηδέν.

Το θέμα της φασματικής ανάλυσης δεν θα αναπτυχθεί περαιτέρω καθώς ξεφεύγει από τους σκοπούς αυτών των σημειώσεων.

Ασκήσεις

Στον παρακάτω πίνακα δίνονται οι πωλήσεις μίας εταιρείας (σε χιλιάδες ευρώ) τα τελευταία τρία χρόνια.

Έτος	2009				2010				2011			
Τρίμηνο	1	2	3	4	1	2	3	4	1	2	3	4
Πωλήσεις	41	50	54	41	45	60	69	53	55	74	79	63

(i) Να υπολογίσετε κατάλληλους κινητούς μέσους όρους για να προσδιορίσετε την ύπαρξη τάσης στα δεδομένα μας. Στη συνέχεια, να σχεδιάσετε γραφικά τα δεδομένα.

(ii) Να κάνετε το διάγραμμα των αυτοδιακυμάνσεων και των αυτοσυσχετίσεων της χρονοσειράς.

(iii) Να σχολιάσετε τα συμπεράσματα σας.

Κεφάλαιο 4 Αριθμοδείκτες

4.1 Εισαγωγή

Οι αριθμοδείκτες είναι απλές αριθμητικές ποσότητες, οι οποίοι **ορίζονται συνήθως ως λόγοι δύο ποσοτήτων** και χρησιμοποιούνται για την εξαγωγή χρησίμων συμπερασμάτων σε διάφορα οικονομικά προβλήματα.

Ο λόγος της εισαγωγής και της ευρείας χρήσης αριθμοδεικτών είναι η ανάγκη για αναγωγή διαφόρων μεγεθών σε ένα κοινό μέτρο, κάτι που κάνει εύκολη τη σύγκριση τους.

Χαρακτηριστικό παράδειγμα αριθμοδείκτη είναι ο τιμάριθμος τον οποίο χρησιμοποιούμε για να συγκρίνουμε το κόστος ζωής σε δύο διαφορετικές χρονικές στιγμές.

Ο τιμάριθμος και όλοι οι αριθμοδείκτες που ορίζονται από ποσότητες που μετρήθηκαν σε δύο ή περισσότερες χρονικές στιγμές ονομάζονται χρονολογικοί αριθμοδείκτες. Αν για τον υπολογισμό του αριθμοδείκτη χρησιμοποιούνται γεωγραφικά δεδομένα τότε ο αριθμοδείκτης ονομάζεται γεωγραφικός, ενώ αν οι ποσότητες είναι οικονομικά μεγέθη μίας επιχείρησης τότε ο αριθμοδείκτης ονομάζεται οικονομικός.

Πίνακας 4.1: Είδη αριθμοδεικτών

Είδος		Περιγραφή
Χρονολογικοί	Απλός	Εκφράζουν τη χρονική μεταβολή ενός μόνο μεγέθους (π.χ. ανεργία)
	Σύνθετοι	Εκφράζουν και ορίζονται από τη χρονική μεταβολή τιμών μίας ομάδας μεγεθών (π.χ. τιμάριθμος)
Γεωγραφικοί		Ορίζονται από μεγέθη τα οποία μετρήθηκαν σε διαφορετικές γεωγραφικές περιοχές
Οικονομικοί		Ορίζονται από οικονομικά μεγέθη. Οι κυριότεροι είναι : • Αριθμοδείκτες αξίας. • Αριθμοδείκτες ρευστότητας. • Αριθμοδείκτες δραστηριότητας. • Αριθμοδείκτες αποδοτικότητας. • Αριθμοδείκτες κεφαλαίου. • Επενδυτικοί αριθμοδείκτες.

4.2 Υπολογίζοντας έναν αριθμοδείκτη

Συνήθως, οι αριθμοδείκτες παρουσιάζονται ως ποσοστό τοις εκατό για ευκολότερη ερμηνεία του αποτελέσματος. Στους διάφορους τύπους που χρησιμοποιούνται για τον υπολογισμό των αριθμοδεικτών χρησιμοποιείται η εξής σύμβαση :

- Οι αριθμοδείκτες γενικών οικονομικών μεγεθών συνήθως συμβολίζονται με το γράμμα **I** (Index). Αν ο αριθμοδείκτης αποσκοπεί στην σύγκριση δύο ποσοτήτων τις χρονικές στιγμές t1 και t2 γράφουμε το δείκτη ως $\mathbf{I}_{t2,\ t1}$, δηλαδή $\mathbf{I}_{\text{τελική χρονική στιγμή, αρχική χρονική στιγμή}}$. Έτσι, για παράδειγμα ο δείκτης $\mathbf{I}_{2011,\ 2005}$, περιγράφει την μεταβολή μίας ποσότητας από το 2005 έως το 2011.

- Ιδιαίτερα, οι αριθμοδείκτες που αναφέρονται σε μεγέθη τιμής συμβολίζονται με το κεφαλαίο γράμμα **P** (Price), οι αριθμοδείκτες που αναφέρονται σε μεγέθη τιμής συμβολίζονται με το κεφαλαίο γράμμα **Q** (Quantity) ενώ οι αριθμοδείκτες συνολικής αξίας / όγκου συμβολίζονται με το γράμμα **V** (Value ή Volume).

- Τα μεγέθη ποσοτήτων συμβολίζονται με το πεζό λατινικό γράμμα **q** (quantity). Έτσι, το σύμβολο $\mathbf{q_0}$ χαρακτηρίζει την ποσότητα ενός προϊόντος την χρονική στιγμή 0 ενώ το σύμβολο $\mathbf{q_1}$ την ποσότητα τη χρονική στιγμή 1. Γράφουμε :

$$Q_{1,0}=\frac{q_1}{q_0}$$

- Τα μεγέθη τιμών συμβολίζονται με το πεζό λατινικό γράμμα **p** (price). Έτσι, το σύμβολο $\mathbf{p_0}$ χαρακτηρίζει την τιμή ενός προϊόντος την χρονική στιγμή 0 ενώ το σύμβολο $\mathbf{p_1}$ την τιμή του τη χρονική στιγμή 1. Γράφουμε :

$$P_{1,0}=\frac{p_1}{p_0}$$

- Η συνολική αξία ενός εμπορεύματος ποσότητας **q** από τα οποία κάθε μονάδα έχει τιμή **p** θα είναι **p·q**. Γράφουμε :

$$V_{1,0}=\frac{p_1 \cdot q_1}{p_0 \cdot q_0}$$

4.3 Απλοί αριθμοδείκτες

Απλός ονομάζεται ένας αριθμοδείκτης όταν ορίζεται από τη μεταβολή σε ένα μόνο μέγεθος είτε αυτό αφορά ποσότητα είτε κόστος είτε συνολική αξία/όγκος.

Παράδειγμα υπολογισμού απλού χρονολογικού αριθμοδείκτη τιμής

Η τιμή πώλησης του μικρού νερού στα περίπτερα το 2000 ήταν 15 λεπτά ενώ το 2010 ήταν 50 λεπτά. Αν πάρουμε το 2000 ως έτος βάσης (αρχική χρονική στιγμή) τότε η σχετική μεταβολή της τιμής είναι

$$P_{2010,2000} = \frac{p_{2010}}{p_{2000}} = \frac{50}{15} = 3{,}33 = 333\% \quad .$$

Παράδειγμα υπολογισμού απλού χρονολογικού αριθμοδείκτη αξίας

Δίνονται οι τιμές σε (€) και οι ποσότητες που καταναλώθηκαν σε kg, ενός αγαθού για τις χρονιές 2009 και 2010.

Έτος	Τιμή (p)	Ποσότητα (q)
2009	3,5€	160 kg
2010	4,2€	165 kg

Ο σχετικός δείκτης αξίας για το 2010 με βάση το έτος 2009 είναι

$$V_{2010,2009} = \frac{p_{2010} \cdot q_{2010}}{p_{2009} \cdot q_{2009}} = \frac{4{,}2 \cdot 165}{3{,}5 \cdot 160} = \frac{693}{560} = 1{,}2375 = 123{,}75\% \quad .$$

4.4 Σύνθετοι αριθμοδείκτες

Οι σύνθετοι αριθμοδείκτες ορίζονται από μία ομάδα αγαθών για να δείξουν την εξέλιξη ενός γενικότερου φαινόμενου. Οι αριθμοδείκτες αυτού του τύπου είναι περισσότερο χρήσιμοι στην πράξη. Διαχωρίζονται περαιτέρω σε μη σταθμικούς και σταθμικούς σύνθετους αριθμοδείκτες.

Μη σταθμικός ονομάζεται ένας σύνθετος αριθμοδείκτης όταν όλες οι ποσότητες που συμμετέχουν στον υπολογισμό της είναι ισότιμες ως προς το τελικό αποτέλεσμα.

Σταθμικός ονομάζεται ένας σύνθετος αριθμοδείκτης όταν δεν υπάρχει ισότιμη συμμετοχή όλων των προϊόντων στη διαμόρφωση της τιμής του αριθμοδείκτη. Χαρακτηριστικό παράδειγμα σταθμικού σύνθετου αριθμοδείκτη είναι ο *τιμάριθμος* που δείχνει την εξέλιξη του πληθωρισμού.

Παράδειγμα υπολογισμού μη σταθμικού σύνθετου αριθμοδείκτη

Έστω πως παρακολουθούμε τα προϊόντα Α, Β και Γ για τα έτη 2010 και 2011 των οποίων η τιμή καταγράφεται στον παρακάτω πίνακα.

Προϊόν	Τιμή 2010 (p_{2010})	Τιμή 2011 (p_{2011})
Α	13€	12€
Β	27€	31€
Γ	3€	5€

Ο μη σταθμικός αριθμοδείκτης που δείχνει την εξέλιξη των τιμών της ομάδας των τριών προϊόντων είναι

$$P_{2011,2010} = \frac{p_{A,2011}+p_{B,2011}+p_{\Gamma,2011}}{p_{A,2010}+p_{B,2010}+p_{\Gamma,2010}} = \frac{12+31+5}{13+27+3} = \frac{48}{43} = 1{,}116=111{,}6\%$$

Παράδειγμα υπολογισμού σταθμικού σύνθετου αριθμοδείκτη

Έστω πως παρακολουθούμε τα προϊόντα Α, Β και Γ για τα έτη 2010 και 2011 των οποίων η τιμή και η ποσότητα που καταναλώθηκε καταγράφεται στον παρακάτω πίνακα.

Προϊόν	Τιμή 2010 (p_{2010})	Ποσότητα 2010 (q_{2010})	Τιμή 2011 (p_{2011})	Ποσότητα 2011 (q_{2010})
Α	13€	1200 kg	12€	1400 kg
Β	27€	500 kg	31€	700 kg
Γ	3€	2300 kg	5€	2900 kg

Ο σταθμικός αριθμοδείκτης που δείχνει την εξέλιξη των αξίας της ομάδας των τριών προϊόντων είναι

$$V_{2011,2010} = \frac{p_{A,2011}\cdot q_{A,2011}+p_{B,2011}\cdot q_{B,2011}+p_{\Gamma,2011}\cdot q_{\Gamma,2011}}{p_{A,2010}\cdot q_{A,2010}+p_{B,2010}\cdot q_{B,2010}+p_{\Gamma,2010}\cdot q_{\Gamma,2010}} = \frac{12\cdot 1400+31\cdot 700+5\cdot 2900}{13\cdot 1200+27\cdot 500+3\cdot 2300}$$
$$= \frac{53000}{36000} = 1{,}472=147{,}2\%$$

Παράδειγμα υπολογισμού σταθμικού σύνθετου αριθμοδείκτη (τιμάριθμος)

Ο τιμάριθμος είναι ο πιο γνωστός σταθμικός σύνθετος αριθμοδείκτης. Χρησιμοποιείται για τη μέτρηση του φαινομένου του πληθωρισμού. Για τον υπολογισμό του τιμάριθμου χρησιμοποιείται μία μεγάλη ομάδα προϊόντων τα οποία ταξινομούνται σε 12 μεγάλες κατηγορίες όπως παρουσιάζονται στον επόμενο πίνακα Σφάλμα: Δεν βρέθηκε η πηγή παραπομπής. Η ταξινόμηση των ειδών βασίζεται στη διεθνή ταξινόμηση COICOP (Classification Of Individual Consumption by Purpose). Με δειγματοληψία στα νοικοκυριά αποφασίζεται η στάθμιση στις 1000 μονάδες που πρέπει να αντιστοιχεί σε κάθε μία κατηγορία ενώ κάθε μία κατηγορία διαχωρίζεται περαιτέρω σε υποκατηγορίες – συγκεκριμένες υποομάδες προϊόντων για τους οποίους υπολογίζονται και οι ανάλογοι υποδείκτες που δείχνουν τον πληθωρισμό της αντίστοιχης κατηγορίας. Συνολικά υπολογίζονται 140 υποδείκτες οι οποίοι διαμορφώνουν τους 12 δείκτες από τους οποίους υπολογίζεται ο τιμάριθμος. Η δειγματοληψία, διενεργείται συνήθως στα μέσα του μήνα που αφορά ο υπολογισμός και διαρκεί μία εργάσιμη εβδομάδα ενώ έχει μεγαλύτερο χρονικό εύρος για τα προϊόντα που έχουν μεγάλη διακύμανση στη διάρκεια του μήνα όπως π.χ. Το πετρέλαιο ή τα νωπά φρούτα και λαχανικά. Περισσότερες πληροφορίες στο μεθοδολογικό σημείωμα της ΕΛ.ΣΤΑΤ. Διαθέσιμη στην ιστοσελίδα της υπηρεσίας http://www.statistics.gr/portal/page/portal/ESYE/PAGE-themes?p_param=A1701&r_param=DKT90&y_param=MT&mytabs=0

Βασικές ομάδες ειδών ταξινόμησης COICOP/HICP	Συντελεστές στάθμισης (‰) Weights (2007)	Main groups of items of the COICOP/HICP classification
Γενικός Εναρμονισμένος Δείκτης Τιμών Καταναλωτή	**1.000,00**	**Overall Harmonized Index of Consumer Prices**
1. Διατροφή και μη αλκοολούχα ποτά	171,88	1. Food and non-alcoholic beverages
2. Αλκοολούχα ποτά και καπνός	45,21	2. Alcoholic beverages and tobacco
3. Ένδυση – Υπόδηση	94,76	3. Clothing – Footwear
4. Στέγαση	88,65	4. Housing
5. Διαρκή αγαθά – Είδη νοικοκυριού και υπηρεσίες	70,47	5. Durable goods – Household articles and services
6. Υγεία	57,06	6. Health
7. Μεταφορές	130,91	7. Transport
8. Επικοινωνίες	38,15	8. Communication
9. Αναψυχή – Πολιτιστικές δραστηριότητες	52,89	9. Recreation – Cultural activities
10. Εκπαίδευση	22,69	10. Education
11. Ξενοδοχεία – Καφέ – Εστιατόρια	163,62	11. Hotels – Cafés – Restaurants
12. Άλλα αγαθά και υπηρεσίες	63,72	12. Miscellaneous goods and services

Πίνακας 4.2: Ταξινόμηση των ειδών σύμφωνα με την ταξινόμηση COICOP

Το πρόβλημα της σύγκρισης του τιμάριθμου σε δύο χρονικές στιγμές

Ο τιμάριθμος όπως παρουσιάστηκε στην προηγούμενη παράγραφο, βασίζεται σε συγκεκριμένες σταθμίσεις που ορίζονται σε κάθε έτος βάσει των συνηθειών κατανάλωσης αγαθών και προϊόντων που διατίθενται στην αγορά. Καθώς, υπάρχουν μεταβολές στο είδος των προϊόντων που καταναλώνεται σε κάθε χρονική περίοδο ανακύπτει το πρόβλημα της σύγκρισης του τιμάριθμου σε δύο μακρινές χρονικές στιγμές t_0 και t_1 για τις οποίες υπάρχουν διαφορετικές σταθμίσεις σε κάθε ομάδα προϊόντων.

Υπάρχουν δύο βασικές κατευθύνσεις : Η μία είναι να υπολογιστεί η μεταβολή των τιμών στις δύο χρονικές στιγμές με βάση τη στάθμιση της αρχικής χρονικής στιγμής ενώ η άλλη είναι η πραγματοποίηση του υπολογισμού με βάση τη στάθμιση της τελικής χρονικής στιγμής. Στην πρώτη περίπτωση ο τιμάριθμος λέμε ότι υπολογίζεται με τον δείκτη Luspeyres ενώ στην δεύτερη περίπτωση ο δείκτης ονομάζεται δείκτης Paasche.

Πιο αναλυτικά :

Αν $p_{1,0}$, $p_{2,0}$, ...,$p_{k,0}$, είναι οι τιμές από κ προϊόντα και $w_{1,0}$, $w_{2,0}$, ...,$w_{k,0}$, είναι οι αντίστοιχοι συντελεστές στάθμισης την πρώτη χρονική στιγμή ενώ $p_{1,1}$, $p_{2,1}$, ...,$p_{k,1}$ και $w_{1,1}$, $w_{2,1}$, ...,$w_{k,1}$ είναι οι αντίστοιχοι συντελεστές στάθμισης τη δεύτερη χρονική στιγμή τότε ο **δείκτης Luspeyres** ορίζεται να είναι η ποσότητα

$$L_{1,0}=\frac{\sum_{i=1}^{\kappa} p_{i,1} w_{i,0}}{\sum_{i=1}^{\kappa} p_{i,0} w_{i,0}}=\frac{\sum p_1 w_0}{\sum p_0 w_0}$$

ενώ ο **δείκτης Paasche** ορίζεται να είναι

$$P_{1,0}=\frac{\sum_{i=1}^{\kappa} p_{i,1} w_{i,1}}{\sum_{i=1}^{\kappa} p_{i,0} w_{i,1}}=\frac{\sum p_1 w_1}{\sum p_0 w_1}$$

Παρατήρηση

Από τους τελευταίους δύο τύπους επιβεβαιώνεται πως η διαφορά μεταξύ των δεικτών Luspeyres και του Paasche είναι απλά η χρήση διαφορετικών συντελεστών στάθμισης w_i για τον υπολογισμό του σκορ στις δύο χρονικές στιγμές.

Είναι φανερό πως δεν υπάρχει σωστή και λάθος επιλογή στην απόφαση για το ποια βάρη θα χρησιμοποιηθούν για τον υπολογισμό ενός αντιπροσωπευτικού δείκτη. Στην πραγματικότητα, για τον υπολογισμό του τιμάριθμου χρησιμοποιείται ένας συνδυασμός των δύο δεικτών με κάποιον συνδυαστικό υπολογισμό των συντελεστών στάθμισης.

Παράδειγμα

Δίνεται ο παρακάτω πίνακας που περιέχει την αξία και το συντελεστή στάθμισης για δύο προϊόντα Α και Β. Να υπολογιστεί (i) ο δείκτης Luspeyres και (ii) ο δείκτης Paasche.

	2010		2011	
Προϊόν	p_0	w_0	p_1	w_1
A	20	6	24	4
B	8	100	8	160

Λύση

Υπολογίζουμε τις απαραίτητες ποσότητες πριν προχωρήσουμε στους υπολογισμούς των δεικτών.

	2010		2011					
Προϊόν	p_0	w_0	p_1	w_1	p_0w_0	p_1w_0	p_0w_1	p_1w_1
A	20	6	24	4	120	144	80	96
B	8	100	8	160	800	800	1280	1280
Σύνολο					920	944	1360	1376

Από τον παραπάνω πίνακα υπολογίζουμε :

- Δείκτης Luspeyres : $L_{2011,2010}=\frac{\sum p_1 w_0}{\sum p_0 w_0}=\frac{944}{920}=1{,}02=102\%$
- Δείκτης Paasche : $L_{2011,2010}=\frac{\sum p_1 w_1}{\sum p_0 w_1}=\frac{1376}{1360}=1{,}01=101\%$

4.5 Αριθμοδείκτες στην ανάλυση Λογιστικών καταστάσεων

Οι αριθμοδείκτες βρίσκουν ευρεία χρήση στην χρηματοοικονομική ανάλυση επιχειρήσεων και άλλων οργανισμών, καθώς αποτυπώνουν σε μία κατανοητή γλώσσα την οικονομική πραγματικότητα του οργανισμού. Η διαχρονική ανάλυση τέτοιων δεικτών καθιστά ευκολότερη τη διοίκηση και την οργάνωση της επιχείρισης. Η στατική θεώρηση των δεικτών μιας επιχείρησης έχει μικρή πρακτική σημασία.

Παρακάτω παρουσιάζονται οι βασικές κατηγορίες αριθμοδεικτών που χρησιμοποιούνται για την ανάλυση της πιστοληπτικής ικανότητας μίας επιχείρησης.

4.5.1 Αριθμοδείκτες ρευστότητας

Οι αριθμοδείκτες ρευστότητας χρησιμοποιούνται για να αξιολογήσουν την ικανότητα μίας επιχείρησης να ανταποκρίνεται στις βραχυπρόθεσμες υποχρεώσεις της.

Η επιχείρηση πρέπει να έχει τη δυνατότητα:

• Να εξοφλεί τους βραχυπρόθεσμους πιστωτές

• Να καταβάλει τους μισθούς, τους οφειλόμενους τόκους

• Να διατηρεί μια υγιή πιστοληπτική ικανότητα

Από τους αριθμοδείκτες ρευστότητας, οι κυριότεροι είναι:

Αριθμοδείκτης έμμεσης ρευστότητας (current ratio ή γενικής ρευστότητας ή κυκλοφοριακής ρευστότητας ή κεφαλαίου κίνησης)

Ο αριθμοδείκτης αυτός είναι ο πλέον χρησιμοποιούμενος δείκτης. Ο αριθμοδείκτης γενικής ρευστότητας προκύπτει αν διαιρέσουμε το σύνολο των κυκλοφοριακών στοιχείων μίας εταιρίας με το σύνολο των βραχυχρόνιων υποχρεώσεών της.

$$\text{Αριθμοδείκτης έμμεσης ρευστότητας} = \frac{\text{Κυκλοφορούν Ενεργητικό}}{\text{Βραχυπρόθεσμες υποχρεώσεις}}$$

Το κυκλοφορούν ενεργητικό περιλαμβάνει τα αποθέματα, τους εισπρακτέους λογαριασμούς, τα χρεόγραφα και τα χρηματικά διαθέσιμα, ενώ οι βραχυπρόθεσμες υποχρεώσεις αποτελούνται από τους πληρωτέους λογαριασμούς, τα πληρωτέα γραμμάτια, τις βραχυπρόθεσμες υποχρεώσεις, όπως βραχυχρόνια τραπεζικά δάνεια, οφειλόμενους τόκους κ.α.

Όσο μεγαλύτερος είναι ο αριθμοδείκτης γενικής ρευστότητας, τόσο καλύτερη από πλευράς ρευστότητας είναι η θέση της συγκεκριμένης επιχείρησης. Γενικά, ένας αριθμοδείκτης γενικής ρευστότητας κοντά στο 2 θεωρείται συνήθως ικανοποιητικός για μία βιομηχανική ή εμπορική επιχείρηση.

Αριθμοδείκτης άμεσης ρευστότητας (quick ration)

Ο δείκτης άμεσης ρευστότητας υπολογίζεται με το λόγο του κυκλοφορούντος ενεργητικού, πλην των αποθεμάτων προς τις βραχυπρόθεσμες υποχρεώσεις.

$$\text{Αριθμοδείκτης άμεσης ρευστότητας} = \frac{\text{Κυκλοφορούν Ενεργητικό - Απόθεμα}}{\text{Βραχυπρόθεσμες υποχρεώσεις}}$$

Η τιμή του δείκτη άμεσης ρευστότητας θεωρείται ικανοποιητική όταν είναι τουλάχιστον ίση με τη μονάδα.

Αριθμοδείκτης ταμειακής ρευστότητας

Είναι ο αριθμοδείκτης που μας δίνει την εικόνα της επάρκειας και όχι των μετρητών στην επιχείρηση, σε σχέση με τις τρέχουσες λειτουργικές ανάγκες.

$$\text{Αριθμοδείκτης ταμιακής ρευστότητας} = \frac{\text{Διαθέσιμο Ενεργητικό}}{\text{Ληξιπρόθεσμες υποχρεώσεις}}$$

Στον παραπάνω τύπο το διαθέσιμο ενεργητικό ορίζεται ως τα μετρητά, στο ταμείο οι καταθέσεις όψεως, επιταγές κ.λ.π.

Ο αριθμοδείκτης αυτός μας ενημερώνει για το πόσες φορές τα διαθέσιμα περιουσιακά στοιχεία μίας επιχείρησης καλύπτουν τις ληξιπρόθεσμες υποχρεώσεις της. Σε γενικές γραμμές, η συνεχής πτώση των δεικτών ρευστότητας και η διαμόρφωσή τους σε επίπεδο μικρότερο του δείκτη του σχετικού κλάδου, αποτελεί ένα σημάδι κινδύνου για την επιχείρηση.

Παράλληλα όμως, υπάρχει το ενδεχόμενο η εταιρία να έχει βελτιώσει τη διαδικασία μετατροπής των αποθεμάτων της σε πωλήσεις και ρευστό είτε ότι η εταιρία αναγκάζεται να ρευστοποιήσει τα αποθέματά της για την εξόφληση υποχρεώσεων. Γενικά είναι καλό οι δείκτες ρευστότητας να αξιολογούνται σε συνδυασμό και με τους δείκτες δραστηριότητας.

4.5.2 Αριθμοδείκτες δραστηριότητας (κυκλοφοριακής ταχύτητας)

Η χρησιμοποίηση ορισμένων αριθμοδεικτών κυκλοφοριακής ταχύτητας βοηθάει να προσδιορίσουμε το βαθμό μετατροπής ορισμένων περιουσιακών στοιχείων (αποθέματα, απαιτήσεις) σε ρευστό. Γενικά, το ποσοστό των αποθεμάτων που διατηρεί μία επιχείρηση πρέπει να σχετίζεται πάντοτε με το ύψος πωλήσεών της.

Μερικοί από τους πιο σημαντικούς αριθμοδείκτες κυκλοφοριακής ταχύτητας, είναι:

Αριθμοδείκτης ταχύτητας είσπραξης

Ο αριθμοδείκτης ταχύτητας είσπραξης απαιτήσεων δείχνει πόσες φορές, κατά μέσο όρο, εισπράττονται κατά τη διάρκεια της λογιστικής χρήσεως.

$$\text{Αριθμοδείκτης ταχύτητας εισπράξεων απαιτήσεων} = \frac{\text{Καθαρές πωλήσεις}}{\text{Μέσος όρος απαιτήσεων}}$$

Αριθμοδείκτης ταχύτητας κυκλοφορίας αποθεμάτων

Ο δείκτης ταχύτητας κυκλοφορίας αποθεμάτων ορίζεται ως:

$$\text{Αριθμοδείκτης ταχύτητας κυκλοφορίας αποθεμάτων} = \frac{\text{Καθαρές πωλήσεις}}{\text{Μέσο απόθεμα προϊόντων}}$$

Αν διαιρέσουμε τον αριθμό των ημερών του έτους (365) με τον αριθμοδείκτη ταχύτητας κυκλοφορίας αποθεμάτων, τότε θα έχουμε σε αριθμό ημερών το χρόνο που παρέμειναν τα αποθέματα στην επιχείρηση μέχρι να πωληθούν.

Όσο μεγαλύτερος είναι ο αριθμοδείκτης ταχύτητας κυκλοφορίας αποθεμάτων, τόσο πιο αποτελεσματικά λειτουργεί η επιχείρηση.

Αριθμοδείκτης ταχύτητας κυκλοφορίας κεφαλαίου κίνησης

Ο δείκτης ταχύτητας κυκλοφορίας κεφαλαίου κίνησης ορίζεται ως :

$$\text{Αριθμοδείκτης ταχύτητας κυκλοφορίας κεφαλαίου κίνησης} = \frac{\text{Καθαρές πωλήσεις}}{\text{Καθαρό κεφάλαιο}}$$

Όσο αυξάνουν οι πωλήσεις μίας εταιρίας, τόσο περισσότερα κεφάλαια κινήσεως απαιτούνται για αποθέματα και για αυξημένες πιστώσεις προς τους πελάτες της.

4.5.3 Αριθμοδείκτες αποδοτικότητας

Οι δείκτες αυτοί χρησιμοποιούνται για την αξιολόγηση της ικανότητας της επιχείρησης να πραγματοποιεί κέρδη. Δηλαδή, οι δείκτες αυτοί αξιολογούν το βαθμό κερδοφορίας της επιχείρησης.

Βασικοί αριθμοδείκτες αποδοτικότητας είναι:

Αριθμοδείκτης καθαρού περιθωρίου ή καθαρού κέρδους

Ο αριθμοδείκτης αυτός, δείχνει το ποσοστό του καθαρού κέρδους που επιτυγχάνει μία επιχείρηση από τις πωλήσεις της. Ο αριθμοδείκτης ορίζεται ως:

$$\text{Αριθμοδείκτης καθαρού κέρδους} = \frac{\text{Καθαρά κέρδη χρήσης}}{\text{Καθαρές πωλήσεις}}$$

Όσο ο αριθμοδείκτης καθαρού κέρδους είναι μεγαλύτερος, τόσο πιο επικερδής είναι η επιχείρηση.

Αριθμοδείκτης αποδοτικότητας επένδυσης

Ο αριθμοδείκτης αυτός, μετράει την απόδοση των συνολικών περιουσιακών στοιχείων μίας επιχείρησης και αποτελεί είδος ελέγχου της διοικήσεώς της.

Ο αριθμοδείκτης ορίζεται ως:

$$\text{Αριθμοδείκτης αποδοτικότητας επένδυσης} = \frac{\text{Καθαρά κέρδη χρήσης}}{\text{Σύνολο ενεργητικού}}$$

Η αποτελεσματικότητα λειτουργίας μίας επιχείρησης, που δείχνει ο αριθμοδείκτης αυτός, μας δείχνει την ικανότητά της να μπορεί να επιβιώσει οικονομικά και να προσελκύει κεφάλαια που προσφέρονται για επένδυση, δίνοντας την κατάλληλη ανταμοιβή.

Αριθμοδείκτης αποδοτικότητας ίδιων κεφαλαίων

Ο αριθμοδείκτης αυτός απεικονίζει την κερδοφόρα δυναμικότητα της επιχείρησης και μας παρέχει ένδειξη του κατά πόσο πραγματοποιήθηκε ο στόχος πραγματοποίησης ενός ικανοποιητικού αποτελέσματος. Ο αριθμοδείκτης ορίζεται ως εξής:

$$\text{Αριθμοδείκτης αποδοτικότητας ιδίων κεφαλαίων} = \frac{\text{Καθαρά κέρδη χρήσης}}{\text{Σύνολο ιδίων κεφαλαίων}}$$

Κατάλογος περιεχομένων

Εισαγωγή 1

Κεφάλαιο 1 Ασυμμετρία και κυρτότητα 2

1.1 Ασυμμετρία μίας κατανομής (Skewness) 2
1.2 Αριθμητική εκτίμηση της ασυμμετρίας μίας κατανομής 3
1.3 Κυρτότητα μίας κατανομής (Kurtosis) 4
1.4 Πότε η συμμετρία και η κυρτότητα της κατανομής διαφέρει σημαντικά από την κανονική; 6
1.5 Οι συντελεστές ασυμμετρίας και κυρτότητας ως στιγμές μίας τυχαίας μεταβλητής 6
1.6 Ροπογεννήτρια μίας τυχαίας μεταβλητής Χ 7

Κεφάλαιο 2 Παλινδρόμηση 9

2.1 Διάγραμμα διασποράς (Scatterplot) 9
2.2 Συνδιακύμανση 11
2.2.1 Ερμηνεία της συνδιακύμανσης 11
2.3 Συντελεστής συσχέτισης Pearson 12
2.3.1 Προϋποθέσεις υπολογισμού 12
2.3.2 Αξιολόγηση του συντελεστή Pearson 13
2.3.3 Υπολογισμός συντελεστή συσχέτισης με το Calc 14
2.4 Συντελεστής συσχέτισης Spearman 14
2.5 Γραμμική παλινδρόμηση 17
2.5.1 Αξιολόγηση μοντέλου γραμμικής παλινδρόμησης 19
2.5.2 Άμεση πρόβλεψη με γραμμικό μοντέλο με συνάρτηση του Calc 21
2.6 Μη γραμμική παλινδρόμηση 21

Κεφάλαιο 3 Χρονοσειρές 25

3.1 Τρόποι στατιστικής ανάλυσης χρονοσειρών 26
3.2 Ανάλυση της χρονοσειράς στις κυριότερες συνιστώσες 26
3.3 Μέθοδοι προσδιορισμού της κύριας τάσης 29
3.3.1 Μέθοδος του κινητού μέσου 29
3.3.2 Μέθοδος της ευθείας των ελαχίστων τετραγώνων 31
3.4 Προσδιορισμός της συνιστώσας της κυκλικής εναλλαγής 32
3.5 Προσδιορισμός της εποχιακής συνιστώσας 34

3.6 Συνιστώσα του τυχαίου σφάλματος της χρονοσειράς....37
3.7 Αυτοδιακύμανση και αυτοσυσχέτιση....37
3.7.1 Αυτοδιακύμανση....38
3.7.2 Αυτοσυσχέτιση....39
3.8 Φασματική ανάλυση....43

Κεφάλαιο 4 Αριθμοδείκτες....45

4.1 Εισαγωγή....45
4.2 Υπολογίζοντας έναν αριθμοδείκτη....46
4.3 Απλοί αριθμοδείκτες....47
4.4 Σύνθετοι αριθμοδείκτες....47
4.5 Αριθμοδείκτες στην ανάλυση Λογιστικών καταστάσεων....51
4.5.1 Αριθμοδείκτες ρευστότητας....52
4.5.2 Αριθμοδείκτες δραστηριότητας (κυκλοφοριακής ταχύτητας)....53
4.5.3 Αριθμοδείκτες αποδοτικότητας....54

Ευρετήριο πινάκων

Πίνακας 3.1: Υπόλοιπο και Σχετικό υπόλοιπο....33
Πίνακας 4.1: Είδη αριθμοδεικτών....45
Πίνακας 4.2: Ταξινόμηση των ειδών σύμφωνα με την ταξινόμηση COICOP....49

Κατάλογος εικόνων

Εικόνα 1: Παραδείγματα συμμετρικών κατανομών. Είναι φανερό πως και στις δύο περιπτώσεις υπάρχει μία κατακόρυφη γραμμή (η οποία ταυτίζεται με τη μέση τιμή αλλά και τη διάμεσο) η οποία χωρίζει σε δύο συμμετρικά μέρη την κατανομή....2
Εικόνα 2: Παραδείγματα ασύμμετρων κατανομών. 2
Εικόνα 3: Συντελεστής συσχέτισης Pearson....14
Εικόνα 4: Δεδομένα προσδόκιμου ζωής έτους 1995....15
Εικόνα 5: Υπολογισμός συντελεστή συσχέτισης Spearman....16
Εικόνα 6: Υπολογισμός της τάξης των παρατηρήσεων....16
Εικόνα 7: Υπολογισμός συντελεστών της ευθείας γραμμικής παλινδρόμησης....18
Εικόνα 8: Άμεση πρόβλεψη....21

Βιβλιογραφία

1. Wikipedia, η ελεύθερη εγκυκλοπαίδεια
2. http://mathworld.wolfram.com/
3. Κανονισμός Κατάρτισης Ειδικότητας “Ειδικός Μηχανογραφημένου Λογιστηρίου”, Ε.Ο.Π.Π., 2010.
4. Στατιστική – Εφαρμογές στη Διοίκηση – Οικονομία – Επιχειρήσεις, ΥΠΕΠΘ.
5. Στατιστική με το OpenOffice, Ε. Διαμαντόπουλος, 2008.
6. Στατιστική, Ο.Ε.Δ.Β., 1999.

Οι σημειώσεις αυτές συντάχθηκαν για να υποστηρίξουν τις παραδόσεις του μαθήματος Στατιστική ΙΙ της ειδικότητας “Ειδικός Μηχανογραφημένου Λογιστηρίου” του ΙΕΚ Ξάνθης το Α εξάμηνο 2012.
Επαμεινώνδας Διαμαντόπουλος, Μαθηματικός, epdiamantopoulos@yahoo.gr
Ξάνθη, Μάιος 2012