

Πανεπιστήμιο Πελοποννήσου

Τμήμα Διοίκησης Επιχειρήσεων & Οργανισμών

Ποσοτικές Μέθοδοι

Δρ. Ιωάννης Ε. Λιβιέρης
Αναπληρωτής Καθηγητής

Μάρτιος 2026

Περιεχόμενα

1 Εισαγωγή στις Ποσοτικές Μεθόδους	1
1.1 Η Ανάγκη για Εκτίμηση και Πρόβλεψη	2
1.1.1 Το Πρόβλημα της Απόφασης υπό Αβεβαιότητα	2
1.1.2 Διαδικασία Εκτίμησης και Πρόβλεψης	2
1.2 Σχέσεις μεταξύ Μεταβλητών	3
1.3 Ιστορική Αναδρομή: Η Έννοια της Παλινδρόμησης	3
1.4 Ερωτήσεις Πολ/λης Επιλογής	4
1.5 Ασκήσεις	6
2 Γραμμική Παλινδρόμηση	9
2.1 Το Μοντέλο Γραμμικής Παλινδρόμησης	9
2.2 Εκτίμηση των Συντελεστών: Μέθοδος Ελαχίστων Τετραγώνων	10
2.3 Ερμηνεία της Γραμμικής Παλινδρόμησης	11
2.3.1 Αποσύνθεση της Συνολικής Μεταβλητότητας	11
2.3.2 Συντελεστής Προσδιορισμού r^2	12
2.4 Υποθέσεις Εφαρμογής της Ανάλυσης Παλινδρόμησης	12
2.4.1 Γραμμικότητα	13
2.4.2 Ανεξαρτησία	13
2.4.3 Κανονικότητα	13
2.4.4 Ίση Διασπορά (Ομοσκεδαστικότητα)	13
2.5 Ερωτήσεις Πολ/λης Επιλογής	13
2.6 Ασκήσεις	15
3 Ανάλυση Χρονολογικών Σειρών I: Συνιστώσες και Τάση	19
3.1 Εισαγωγή στις Χρονολογικές Σειρές	19
3.2 Σκοπός της Ανάλυσης Χρονολογικών Σειρών	20
3.3 Συνιστώσες Χρονολογικής Σειράς	20
3.4 Υποδείγματα Αποσύνθεσης	21
3.4.1 Προσθετικό Υπόδειγμα	21
3.4.2 Πολλαπλασιαστικό Υπόδειγμα	21
3.5 Εκτίμηση της Τάσης	21

3.5.1	Μέθοδοι Εκτίμησης	21
3.5.2	Παραμετρικά Υποδείγματα Τάσης	22
3.6	Δείκτης Εποχικότητας	22
3.7	Πρόβλεψη με Τάση και Εποχικότητα	23
3.8	Ερωτήσεις Πολ/λης Επιλογής	23
3.9	Ασκήσεις	26
4	Ανάλυση Χρονολογικών Σειρών II: Εξομάλυνση και Πρόβλεψη	29
4.1	Εξομάλυνση Χρονολογικής Σειράς	29
4.2	Κινητός Μέσος	30
4.2.1	Επιλογή πλάτους k	30
4.2.2	Πλεονεκτήματα και Μειονεκτήματα	30
4.3	Απλή Εκθετική Εξομάλυνση	30
4.3.1	Σημασία του Συντελεστή α	31
4.4	Διπλή Εκθετική Εξομάλυνση	31
4.5	Επιλογή Μεθόδου Πρόβλεψης	32
4.6	Αυτοσυσχέτιση	32
4.7	Ερωτήσεις Πολ/λης Επιλογής	33
4.8	Ασκήσεις	35

Εισαγωγή στις Ποσοτικές Μεθόδους

Οι επιχειρήσεις λειτουργούν σε ένα περιβάλλον αστάθειας και αβεβαιότητας που δημιουργεί διαρκώς την ανάγκη για εκτιμήσεις ή/και προβλέψεις. Σχεδόν κάθε σύγχρονη επιχείρηση — μικρή ή μεγάλη, ιδιωτική ή δημόσια — κάνει εκτιμήσεις και προβλέψεις προκειμένου να πάρει αποφάσεις. Εκτιμήσεις και προβλέψεις χρειάζονται στην οικονομία, στο εμπόριο, στην παραγωγή, στη λήψη κυβερνητικών αποφάσεων, και σε πολλούς άλλους τομείς.

Συχνά τίθενται ερωτήματα όπως: Ποια είναι τα αναμενόμενα έσοδα του κράτους τα επόμενα δύο χρόνια; Αν αυξηθεί η διαφημιστική δαπάνη κατά 10%, πώς θα μεταβληθούν οι πωλήσεις; Ποιοι παράγοντες επηρεάζουν τις μηνιαίες πωλήσεις; Μπορούμε να προβλέψουμε τον τζίρο της επιχείρησης το επόμενο έτος; Τα ερωτήματα αυτά απαντώνται με τη χρήση **ποσοτικών μεθόδων**.

Ορισμός 1.1. Ποσοτικές Μέθοδοι (Quantitative Methods) είναι ένα σύνολο μαθηματικών και στατιστικών τεχνικών που χρησιμοποιούνται για τη *μοντελοποίηση*, την *ανάλυση* και την *πρόβλεψη* φαινομένων, με σκοπό τη λήψη τεκμηριωμένων αποφάσεων στη διοίκηση επιχειρήσεων.

Η ανάπτυξη των ποσοτικών μεθόδων στηρίζεται σε μαθηματικές και στατιστικές αρχές, ενώ η εφαρμογή τους αφορά κάθε πεδίο ανθρώπινης δραστηριότητας που σχετίζεται με αριθμούς και μετρήσεις. Σε αυτό το μάθημα εξετάζουμε δύο βασικές οικογένειες μεθόδων: την *ανάλυση παλινδρόμησης* και την *ανάλυση χρονολογικών σειρών*.

1.1 Η Ανάγκη για Εκτίμηση και Πρόβλεψη

1.1.1 Το Πρόβλημα της Απόφασης υπό Αβεβαιότητα

Στην επιχειρηματική πραγματικότητα, οι αποφάσεις λαμβάνονται συχνά υπό συνθήκες αβεβαιότητας: το μέλλον δεν είναι γνωστό και τα στοιχεία είναι ελλιπή. Για να περιορίσουμε αυτή την αβεβαιότητα, χρησιμοποιούμε ποσοτικές μεθόδους που εκμεταλλεύονται τα ιστορικά δεδομένα και τις σχέσεις μεταξύ μεταβλητών.

Παράδειγμα 1.1. Ένας Δήμος αντιμετωπίζει το πρόβλημα των διαθέσιμων θέσεων στάθμευσης στην περιοχή ευθύνης του. Προκειμένου να προγραμματίσει έργα για την επόμενη πενταετία, οι υπεύθυνοι θέλουν να γνωρίζουν κατά πόσο θα αυξηθεί η ζήτηση θέσεων στάθμευσης, λαμβάνοντας υπόψη την αναμενόμενη αύξηση του πληθυσμού. Ο προσδιορισμός της σχέσης μεταξύ *πληθυσμού* (μεταβλητή εισόδου) και *ζήτησης* (μεταβλητή εξόδου) γίνεται σημαντικό εργαλείο για τη λήψη της απόφασης. ■

Παράδειγμα 1.2. Η εταιρεία Brandway θέλει να ανοίξει ένα νέο κατάστημα. Μέχρι τώρα, τα διευθυντικά στελέχη επέλεξαν τοποθεσία με υποκειμενικά κριτήρια (φτηνό ενοίκιο, υποκειμενική εκτίμηση για την περιοχή). Ο νέος εμπορικός διευθυντής επιθυμεί να λάβει αποφάσεις με *αντικειμενικό* τρόπο: να μετρήσει τα χαρακτηριστικά των καταναλωτών σε υποψήφιας περιοχές και να προβλέψει τις πωλήσεις με βάση αυτά τα χαρακτηριστικά. ■

1.1.2 Διαδικασία Εκτίμησης και Πρόβλεψης

Η ποσοτική προσέγγιση στη λήψη αποφάσεων ακολουθεί μια συγκεκριμένη μεθοδολογία:

1. **Διατύπωση του πραγματικού προβλήματος:** Σαφής ορισμός του τι θέλουμε να εκτιμήσουμε ή να προβλέψουμε.
2. **Συλλογή δεδομένων:** Συγκέντρωση των κατάλληλων ιστορικών ή πειραματικών δεδομένων.
3. **Μορφοποίηση θεωρητικού ή μαθηματικού μοντέλου:** Επιλογή της κατάλληλης ποσοτικής μεθόδου (π.χ. γραμμική παλινδρόμηση, ανάλυση χρονοσειράς).
4. **Επίλυση του προβλήματος:** Εφαρμογή του μοντέλου στα δεδομένα και εξαγωγή της εκτίμησης ή πρόβλεψης.
5. **Επικύρωση και εφαρμογή της λύσης:** Αξιολόγηση της ακρίβειας του μοντέλου και αξιοποίηση των αποτελεσμάτων για τη λήψη απόφασης.

Παρατήρηση 1.1. Η ποσοτική μεθοδολογία δεν αντικαθιστά την κρίση του στελέχους, αλλά την εμπλουτίζει με αντικειμενικά στοιχεία. Ακόμη και το καλύτερο μοντέλο δεν είναι τέλειο – πάντα υπάρχει κάποιο σφάλμα πρόβλεψης που οφείλεται σε τυχαίους ή αδύνατο να μετρηθούν παράγοντες.

1.2 Σχέσεις μεταξύ Μεταβλητών

Μια θεμελιώδης ιδέα στις ποσοτικές μεθόδους είναι ότι πολλά φαινόμενα εξαρτώνται από άλλα. Π.χ. οι πωλήσεις εξαρτώνται από τους δυνητικούς πελάτες στην περιοχή, ή ο χρόνος παραμονής σε νοσοκομείο εξαρτάται από την ηλικία και τη διάγνωση.

Ορισμός 1.2. Μεταβλητές εισόδου (input variables) ή **ανεξάρτητες μεταβλητές** είναι οι μεταβλητές που χρησιμοποιούνται για να εξηγήσουν ή να προβλέψουν μια άλλη μεταβλητή. Συμβολίζονται συνήθως με x .

Μεταβλητές εξόδου (output variables) ή **εξαρτημένες μεταβλητές** είναι οι μεταβλητές που θέλουμε να εξηγήσουμε ή να προβλέψουμε. Συμβολίζονται συνήθως με y .

Για παράδειγμα, αν μελετάμε πώς οι πωλήσεις ενός καταστήματος (y) σχετίζονται με τον αριθμό δυνητικών πελατών στην περιοχή (x), τότε:

- **Ανεξάρτητη μεταβλητή:** αριθμός δυνητικών πελατών (x).
- **Εξαρτημένη μεταβλητή:** ετήσιες πωλήσεις (y).

Η φύση της σχέσης μεταξύ δύο μεταβλητών μπορεί να είναι:

- **Καμία σχέση:** Η μεταβολή του x δεν επηρεάζει συστηματικά το y .
- **Θετική γραμμική σχέση:** Όταν x αυξάνεται, y τείνει να αυξάνεται επίσης, με σταθερό ρυθμό.
- **Αρνητική γραμμική σχέση:** Όταν x αυξάνεται, y τείνει να μειώνεται, με σταθερό ρυθμό.
- **Καμπυλόγραμμη σχέση:** Η σχέση υπάρχει, αλλά δεν είναι γραμμική (π.χ. σχήμα U ή εκθετική αύξηση).

Το πρώτο βήμα για τη μελέτη της σχέσης δύο μεταβλητών είναι η κατασκευή ενός **διαγράμματος διασποράς** (scatter plot), το οποίο αποτυπώνει γραφικά τα ζεύγη τιμών (x_i, y_i) και επιτρέπει την οπτική εκτίμηση του είδους της σχέσης.

1.3 Ιστορική Αναδρομή: Η Έννοια της Παλινδρόμησης

Ο όρος *παλινδρόμηση* (regression) εισήχθη από τον **Φρανσις Γαλτον** (1822–1911), ο οποίος μελετούσε την κληρονομικότητα. Διαπίστωσε ότι, αν και υπήρχε τάση οι ψηλοί γονείς να έχουν ψηλά παιδιά και οι κοντοί γονείς να έχουν κοντά παιδιά, το μέσο ύψος των παιδιών τείνει να «παλινδρομεί» — δηλαδή να κινείται — προς το μέσο ύψος του συνολικού πληθυσμού.

Με άλλα λόγια: το ύψος των παιδιών ασυνήθιστα ψηλών ή ασυνήθιστα κοντών γονέων τείνει να πλησιάζει τον μέσο όρο του πληθυσμού. Αυτή η παρατήρηση επιβεβαιώθηκε αργότερα από τον **Καρλ Πιερσον** (1857–1936), ο οποίος ανέλυσε περισσότερες από χίλιες οικογένειες.

Παρατήρηση 1.2. Σήμερα, ο όρος «παλινδρόμηση» έχει αποκτήσει πολύ ευρύτερη σημασία: αναφέρεται σε κάθε μέθοδο που εκτιμά τη σχέση μεταξύ μιας εξαρτημένης και μιας ή περισσότερων ανεξάρτητων μεταβλητών, ανεξάρτητα από το φαινόμενο της «παλινδρόμησης προς τον μέσο».

Παράδειγμα 1.3. Ένα νοικοκυριό με εισόδημα 20.000€ ξοδεύει 74€ σε τηλεφωνικές κλήσεις, ένα με 25.000€ ξοδεύει 73€, ενώ ένα με 30.000€ ξοδεύει 81€. Τα δεδομένα αυτά υποδηλώνουν ότι υπάρχει κάποια *θετική* σχέση μεταξύ εισοδήματος και εξόδων τηλεφώνου. Μπορούμε να το εκτιμήσουμε με εξίσωση παλινδρόμησης της μορφής $\hat{y} = 5,82 + 0,07 \cdot x$, όπου x το εισόδημα (σε χιλιάδες €) και y τα έξοδα τηλεφώνου (σε δεκάδες €). ■

Η ανάλυση παλινδρόμησης επιτρέπει να απαντήσουμε δύο θεμελιώδη ερωτήματα:

1. Υπάρχει *στατιστικά σημαντική* σχέση μεταξύ x και y ;
2. Γνωρίζοντας μια τιμή του x , μπορούμε να *εκτιμήσουμε* ή να *προβλέψουμε* την αντίστοιχη τιμή του y ;

1.4 Ερωτήσεις Πολ/λης Επιλογής

Ερωτήσεις Πολ/λης Επιλογής 1.1. Ποιος από τους παρακάτω δεν αποτελεί βήμα της διαδικασίας εκτίμησης/πρόβλεψης;

- Διατύπωση του πραγματικού προβλήματος
- Συλλογή δεδομένων
- Αυθαίρετη επιλογή αποτελέσματος
- Επικύρωση και εφαρμογή της λύσης

Ερωτήσεις Πολ/λης Επιλογής 1.2. Ποιος ερευνητής εισήγαγε πρώτος τον όρο «παλινδρόμηση»;

- Καρλ Πιερσον
- Φρανσις Γαλτον
- Ροναλντ Φισσερ
- Ωίλλιαμ Γοσσετ

Ερωτήσεις Πολ/λης Επιλογής 1.3. Σε μια ανάλυση παλινδρόμησης, οι «ετήσιες πωλήσεις ενός καταστήματος» αποτελούν:

- Ανεξάρτητη μεταβλητή
- Εξαρτημένη μεταβλητή
- Παράμετρο του μοντέλου
- Σφάλμα εκτίμησης

Ερωτήσεις Πολ/λης Επιλογής 1.4. Το διάγραμμα διασποράς (scatter plot) χρησιμοποιείται για :

- Την παρουσίαση αθροιστικών συχνοτήτων
- Την οπτική εκτίμηση της σχέσης μεταξύ δύο μεταβλητών
- Τον υπολογισμό του μέσου όρου
- Την ανάλυση χρονοσειράς

Ερωτήσεις Πολ/λης Επιλογής 1.5. Αν η καμπύλη σχέσης μεταξύ x και y έχει σχήμα U , τότε η σχέση χαρακτηρίζεται ως :

- Θετική γραμμική
- Αρνητική γραμμική
- Καμπυλόγραμμη
- Καμία σχέση

Ερωτήσεις Πολ/λης Επιλογής 1.6. Ποιο είναι το πρώτο βήμα στη διαδικασία ανάλυσης παλινδρόμησης :

- Εκτίμηση των συντελεστών του μοντέλου
- Επικύρωση του μοντέλου
- Κατασκευή διαγράμματος διασποράς
- Πρόβλεψη νέων τιμών

Ερωτήσεις Πολ/λης Επιλογής 1.7. Σύμφωνα με τον Γαλτον, το ύψος των παιδιών πολύ ψηλών γονέων τείνει :

- Να είναι ακόμη μεγαλύτερο από το ύψος των γονέων
- Να κινείται προς τον μέσο ύψος του πληθυσμού
- Να παραμένει σταθερό
- Να μειώνεται κάτω από το μέσο ύψος

Ερωτήσεις Πολ/λης Επιλογής 1.8. Ποιο από τα παρακάτω δεν ανήκει στις ποσοτικές μεθόδους που καλύπτει αυτό το μάθημα ;

- Γραμμική παλινδρόμηση
- Ανάλυση χρονολογικών σειρών
- Ανάλυση παραγόντων (Factor Analysis)
- Εκθετική εξομάλυνση

Ερωτήσεις Πολ/λης Επιλογής 1.9. Σε ποιο από τα παρακάτω το αποτέλεσμα της πρόβλεψης δεν είναι ποτέ τέλεια ακριβές;

- Μόνο αν τα δεδομένα είναι λίγα
- Πάντα, λόγω τυχαίων ή μη μετρήσιμων παραγόντων
- Μόνο αν χρησιμοποιηθεί λάθος μοντέλο
- Μόνο σε μη γραμμικές σχέσεις

Ερωτήσεις Πολ/λης Επιλογής 1.10. Ποια μεταβλητή συμβολίζεται συνήθως με x στην ανάλυση παλινδρόμησης;

- Η εξαρτημένη μεταβλητή
- Η ανεξάρτητη μεταβλητή
- Το σφάλμα εκτίμησης
- Ο συντελεστής προσδιορισμού

1.5 Ασκήσεις

Άσκηση 1.1 (Αναγνώριση μεταβλητών). Για κάθε ένα από τα παρακάτω ζεύγη μεταβλητών, προσδιορίστε ποια είναι ανεξάρτητη (x) και ποια εξαρτημένη (y), και εκτιμήστε αν αναμένετε θετική, αρνητική ή καμία σχέση μεταξύ τους.

- i. Ώρες μελέτης ανά εβδομάδα και βαθμός στην εξέταση.
- ii. Θερμοκρασία εξωτερικού χώρου και κατανάλωση ηλεκτρικής ενέργειας για θέρμανση.
- iii. Αριθμός διαφημίσεων σε ΜΜΕ και πωλήσεις προϊόντος.
- iv. Ηλικία αυτοκινήτου (σε χρόνια) και αξία μεταπώλησης (σε €).

Λύση:

- i. x = ώρες μελέτης, y = βαθμός. Αναμένεται *θετική* σχέση: περισσότερη μελέτη τείνει να οδηγεί σε υψηλότερο βαθμό.
- ii. x = θερμοκρασία, y = κατανάλωση ηλεκτρισμού. Αναμένεται *αρνητική* σχέση: όσο πιο χαμηλή η θερμοκρασία, τόσο μεγαλύτερη η κατανάλωση για θέρμανση.
- iii. x = αριθμός διαφημίσεων, y = πωλήσεις. Αναμένεται *θετική* σχέση: περισσότερη διαφήμιση τείνει να αυξάνει τις πωλήσεις.
- iv. x = ηλικία αυτοκινήτου, y = αξία μεταπώλησης. Αναμένεται *αρνητική* σχέση: παλαιότερα αυτοκίνητα έχουν χαμηλότερη αξία.

■

Άσκηση 1.2 (Ανάγκη για πρόβλεψη). Μια αλυσίδα σούπερ μάρκετ θέλει να αποφασίσει αν θα ανοίξει νέο κατάστημα σε μια περιοχή. Τα διαθέσιμα στοιχεία για κάθε υπάρχον κατάστημα είναι: ο αριθμός νοικοκυριών στην περιοχή, το μέσο εισόδημα των κατοίκων, και οι ετήσιες πωλήσεις.

- i. Ποιος είναι ο σκοπός της ποσοτικής ανάλυσης σε αυτή την περίπτωση;
- ii. Ποιες μεταβλητές θα είναι ανεξάρτητες και ποια εξαρτημένη;
- iii. Τι είδους δεδομένα χρειάζεται να συλλέξει η εταιρεία;

Λύση:

- i. Σκοπός: να εκτιμηθούν οι αναμενόμενες ετήσιες πωλήσεις για ένα *υποψήφιο* νέο κατάστημα, ώστε να αξιολογηθεί η οικονομική βιωσιμότητά του πριν από την επένδυση.
- ii. Ανεξάρτητες: αριθμός νοικοκυριών (x_1) και μέσο εισόδημα κατοίκων (x_2). Εξαρτημένη: ετήσιες πωλήσεις (y).
- iii. Χρειάζονται ιστορικά δεδομένα από τα *υπάρχοντα* καταστήματα (αριθμός νοικοκυριών, μέσο εισόδημα, πωλήσεις), ώστε να εκτιμηθεί το μοντέλο και στη συνέχεια να εφαρμοστεί στα χαρακτηριστικά της νέας περιοχής.



Άσκηση 1.3 (Εντοπισμός είδους σχέσης). Για κάθε μια από τις παρακάτω περιπτώσεις, εκτιμήστε το είδος της σχέσης (θετική γραμμική, αρνητική γραμμική, καμπυλόγραμμη, καμία σχέση) και αιτιολογήστε την απάντησή σας.

- i. Ο αριθμός ωρών εκπαίδευσης ενός υπαλλήλου και η παραγωγικότητά του (μέχρι ένα ορισμένο επίπεδο, μετά σταθεροποιείται).
- ii. Η τιμή ενός προϊόντος και η ζήτησή του.
- iii. Ο αριθμός χαρτιών σε μια εκτύπωση και ο τόνος που καταναλώνεται.

Λύση:

- i. *Καμπυλόγραμμη* (αύξηση που σταθεροποιείται): αρχικά η εκπαίδευση αυξάνει την παραγωγικότητα, αλλά υπάρχει ανώτατο επίπεδο — δεν αυξάνεται απεριόριστα. Η σχέση δεν είναι γραμμική.
- ii. *Αρνητική* σχέση: όταν η τιμή αυξάνεται, η ζήτηση μειώνεται (νόμος της ζήτησης). Η σχέση μπορεί να είναι γραμμική ή ελαφρά καμπυλόγραμμη.
- iii. *Θετική γραμμική* σχέση: κάθε εκτυπωμένη σελίδα καταναλώνει σταθερή ποσότητα τόνερ, οπότε η κατανάλωση αυξάνεται γραμμικά με τον αριθμό σελίδων.



Άσκηση 1.4 (Διαδικασία εκτίμησης). Μια εταιρεία λογιστις θέλει να προβλέψει τον χρόνο παράδοσης (σε ώρες) δεμάτων με βάση την απόσταση διαδρομής (σε χλμ). Περιγράψτε πλήρως τα βήματα της διαδικασίας εκτίμησης που θα πρέπει να ακολουθηθούν.

Λύση :

- i. **Διατύπωση:** Θέλουμε να εκτιμήσουμε τον χρόνο παράδοσης (y , ώρες) ως συνάρτηση της απόστασης (x , χλμ).
- ii. **Συλλογή δεδομένων:** Καταγραφή ιστορικών αποστολών με τα ζεύγη (απόσταση, χρόνος παράδοσης) για αντιπροσωπευτικό δείγμα.
- iii. **Κατασκευή διαγράμματος διασποράς:** Οπτική εξέταση για επιβεβαίωση ότι υπάρχει (γραμμική) σχέση.
- iv. **Επιλογή και εκτίμηση μοντέλου:** Εφαρμογή γραμμικής παλινδρόμησης για εκτίμηση της εξίσωσης $\hat{y} = b_0 + b_1x$.
- v. **Έλεγχος ποιότητας:** Αξιολόγηση του συντελεστή προσδιορισμού r^2 και έλεγχος υποθέσεων.
- vi. **Πρόβλεψη:** Χρήση της εξίσωσης για εκτίμηση χρόνου παράδοσης σε νέα αποστολή γνωστής απόστασης.



Γραμμική Παλινδρόμηση

Στο προηγούμενο κεφάλαιο εισαγάγαμε την έννοια της σχέσης μεταξύ μεταβλητών. Σε αυτό το κεφάλαιο εξετάζουμε αναλυτικά τη **γραμμική παλινδρόμηση** (linear regression), τη βασικότερη ποσοτική μέθοδο για τη μοντελοποίηση γραμμικών σχέσεων. Σκοπός της είναι η εκτίμηση μιας ευθείας γραμμής που αντιπροσωπεύει τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής y και μιας ανεξάρτητης μεταβλητής x .

2.1 Το Μοντέλο Γραμμικής Παλινδρόμησης

Ορισμός 2.1. Το **μοντέλο γραμμικής παλινδρόμησης** (simple linear regression model) ορίζεται ως:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \dots, n$$

όπου:

- y_i : η τιμή της εξαρτημένης μεταβλητής για την i -οστή παρατήρηση.
- x_i : η τιμή της ανεξάρτητης μεταβλητής για την i -οστή παρατήρηση.
- β_0 : η *τεταγμένη επί τη αρχή* (intercept), η τιμή του y όταν $x = 0$.
- β_1 : η *κλίση* (slope), η μεταβολή του y για κάθε μονάδα αύξησης του x .
- e_i : το *σφάλμα* ή *κατάλοιπο* (residual, error), που εκφράζει την απόκλιση της πραγματικής τιμής από την εκτιμώμενη.

Στην πράξη, δεν γνωρίζουμε τις πραγματικές παραμέτρους β_0 και β_1 . Τις εκτιμούμε από τα δεδομένα και συμβολίζουμε τις εκτιμήσεις με b_0 και b_1 . Η εκτιμώμενη ευθεία παλινδρόμησης είναι:

$$\hat{y} = b_0 + b_1 x$$

Παρατήρηση 2.1. Η τιμή \hat{y} (διαβάζεται «y καπέλο») είναι η εκτιμώμενη ή προβλεπόμενη τιμή του y για δεδομένη τιμή x , βάσει της εκτιμώμενης ευθείας. Το κατάλοιπο (residual) για κάθε παρατήρηση ορίζεται ως $e_i = y_i - \hat{y}_i$.

2.2 Εκτίμηση των Συντελεστών: Μέθοδος Ελαχίστων Τετραγώνων

Η πιο διαδεδομένη μέθοδος εκτίμησης της ευθείας παλινδρόμησης είναι η **μέθοδος ελαχίστων τετραγώνων** (Ordinary Least Squares, OLS).

Ορισμός 2.2. Η μέθοδος ελαχίστων τετραγώνων εκτιμά τους συντελεστές b_0 και b_1 επιλέγοντας εκείνες τις τιμές που ελαχιστοποιούν το άθροισμα των τετραγώνων των καταλοίπων:

$$\text{ΣΣΕ} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \rightarrow \min$$

Η επίλυση του προβλήματος ελαχιστοποίησης δίνει τους εξής τύπους:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{και} \quad b_0 = \bar{y} - b_1 \bar{x}$$

όπου $\bar{x} = \frac{1}{n} \sum x_i$ και $\bar{y} = \frac{1}{n} \sum y_i$ είναι οι δειγματικοί μέσοι όροι.

Παρατήρηση 2.2. Η ευθεία ελαχίστων τετραγώνων πάντα διέρχεται από το σημείο (\bar{x}, \bar{y}) .

Παράδειγμα 2.1. Η εταιρεία Brandway διαθέτει 14 καταστήματα. Τα στοιχεία δυνητικών πελατών (x , σε εκατοντάδες χιλιάδες) και ετήσιων πωλήσεων (y , σε εκατομμύρια €) δίνονται παρακάτω:

Κατάστημα	Δυνητικοί πελάτες (x)	Ετήσιες πωλήσεις (y)
1	3,7	5,7
2	3,6	5,9
3	2,8	6,7
4	5,6	9,5
5	3,3	5,4
6	2,2	3,5
7	3,3	6,2
8	3,1	4,7
9	3,2	6,1
10	3,5	4,9
11	5,2	10,7
12	4,6	7,6
13	5,8	11,8
14	3,0	4,1

Η εφαρμογή της μεθόδου ελαχίστων τετραγώνων δίνει: $b_1 \approx 1,670$ και $b_0 \approx 0,964$. Άρα η εξίσωση παλινδρόμησης είναι:

$$\hat{y} = 0,964 + 1,670 \cdot x$$

Αν ένα νέο κατάστημα βρίσκεται σε περιοχή με 4 (= 400.000) δυνητικούς πελάτες, τότε οι προβλεπόμενες πωλήσεις είναι: $\hat{y} = 0,964 + 1,670 \cdot 4 = 7,644$ εκ. €.

2.3 Ερμηνεία της Γραμμικής Παλινδρόμησης

2.3.1 Αποσύνθεση της Συνολικής Μεταβλητότητας

Για να αξιολογήσουμε πόσο καλά η εκτιμώμενη ευθεία αντιπροσωπεύει τα δεδομένα, χρησιμοποιούμε τρία μέτρα μεταβλητότητας:

Ορισμός 2.3. • **Συνολικό άθροισμα τετραγώνων** (SST – Total Sum of Squares): Μετρά τη συνολική μεταβλητότητα του y γύρω από το μέσο \bar{y} :

$$\text{ΣΣΤ} = \sum_{i=1}^n (y_i - \bar{y})^2$$

• **Άθροισμα τετραγώνων παλινδρόμησης** (SSR – Regression Sum of Squares): Μετρά τη μεταβλητότητα που εξηγεί το μοντέλο:

$$\text{ΣΣΡ} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

- **Άθροισμα τετραγώνων σφαλμάτων** (SSE – Error Sum of Squares): Μετρά τη μεταβλητότητα που δεν εξηγεί το μοντέλο:

$$\text{ΣΣΕ} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Ισχύει πάντα: $\text{ΣΣΤ} = \text{ΣΣΡ} + \text{ΣΣΕ}$.

2.3.2 Συντελεστής Προσδιορισμού r^2

Ορισμός 2.4. Ο συντελεστής προσδιορισμού (coefficient of determination) r^2 ορίζεται ως:

$$r^2 = \frac{\text{ΣΣΡ}}{\text{ΣΣΤ}} = 1 - \frac{\text{ΣΣΕ}}{\text{ΣΣΤ}}$$

και εκφράζει το ποσοστό της συνολικής μεταβλητότητας του y που εξηγείται από την ανεξάρτητη μεταβλητή x . Παίρνει τιμές $0 \leq r^2 \leq 1$.

Παρατήρηση 2.3. Ακραίες τιμές:

- Αν $r^2 = 1$: Όλες οι παρατηρήσεις βρίσκονται ακριβώς πάνω στην ευθεία παλινδρόμησης. Τέλεια γραμμική σχέση.
- Αν $r^2 = 0$: Η ανεξάρτητη μεταβλητή δεν εξηγεί καθόλου τη μεταβλητότητα του y .
- Τυπικά, επιθυμούμε $r^2 \geq 0,70$ για να θεωρείται ικανοποιητική η προσαρμογή, αν και αυτό εξαρτάται από το πεδίο εφαρμογής.

Παράδειγμα 2.2. Για τα δεδομένα της Brandway, το $\text{ΣΣΤ} = 75,31$, $\text{ΣΣΡ} = 63,87$, $\text{ΣΣΕ} = 11,44$. Άρα:

$$r^2 = \frac{63,87}{75,31} \approx 0,8479$$

Ερμηνεία: Το 84,79% της μεταβλητότητας των ετήσιων πωλήσεων εξηγείται από τη μεταβλητότητα των δυνητικών πελατών. Το υπόλοιπο 15,21% οφείλεται σε άλλους παράγοντες. ■

2.4 Υποθέσεις Εφαρμογής της Ανάλυσης Παλινδρόμησης

Για να είναι έγκυρα τα αποτελέσματα της ανάλυσης παλινδρόμησης, πρέπει να ισχύουν τέσσερις βασικές υποθέσεις. Αυτές αφορούν τη φύση των καταλοίπων $e_i = y_i - \hat{y}_i$.

2.4.1 Γραμμικότητα

Η σχέση μεταξύ x και y πρέπει να είναι *γραμμική*. Εξετάζεται με το διάγραμμα διασποράς των δεδομένων (x έναντι y) και με το διάγραμμα καταλοίπων (x έναντι e): αν υπάρχει συστηματικό μοτίβο (π.χ. καμπύλη) στα κατάλοιπα, η υπόθεση παραβιάζεται.

2.4.2 Ανεξαρτησία

Τα κατάλοιπα e_i πρέπει να είναι *ανεξάρτητα* μεταξύ τους, δηλαδή η τιμή ενός καταλοίπου δεν επηρεάζει τα υπόλοιπα. Αν τα δεδομένα είναι χρονολογικά, η υπόθεση αυτή παραβιάζεται συχνά (αυτοσυσχέτιση). Ελέγχεται γραφικά με το διάγραμμα καταλοίπων έναντι της σειράς παρατήρησης.

2.4.3 Κανονικότητα

Τα κατάλοιπα e_i πρέπει να ακολουθούν την *κανονική κατανομή*. Ελέγχεται με :

- Ιστόγραμμα των καταλοίπων (πρέπει να μοιάζει με κωδωνοειδή καμπύλη).
- Θηκόγραμμα (boxplot) των καταλοίπων.
- Διάγραμμα κανονικής πιθανότητας (normal probability plot): τα σημεία πρέπει να βρίσκονται κοντά σε ευθεία γραμμή.

2.4.4 Ίση Διασπορά (Ομοσκεδαστικότητα)

Η διασπορά των καταλοίπων πρέπει να παραμένει *σταθερή* για όλες τις τιμές του x . Αν η διασπορά αυξάνεται (ή μειώνεται) με το x , έχουμε ετεροσκεδαστικότητα. Ελέγχεται γραφικά με το διάγραμμα \hat{y} έναντι e : αν το «φάκελος» των καταλοίπων διευρύνεται, η υπόθεση παραβιάζεται.

Παρατήρηση 2.4. Αρχή παρεμβολής: Η εκτιμώμενη εξίσωση παλινδρόμησης είναι αξιόπιστη μόνο για τιμές του x που ανήκουν στο διάστημα $[\min(x_i), \max(x_i)]$ του δείγματος. Η εκτίμηση έξω από αυτό το διάστημα (εξωπολίτωση) ενέχει υψηλό κίνδυνο σφάλματος.

2.5 Ερωτήσεις Πολ/λης Επιλογής

Ερωτήσεις Πολ/λης Επιλογής 2.1. Στη μέθοδο ελαχίστων τετραγώνων, ελαχιστοποιούμε :

- Το άθροισμα των καταλοίπων $\sum e_i$
- Το άθροισμα των τετραγώνων των καταλοίπων $\sum e_i^2$
- Τη μέση τιμή των y_i
- Τον συντελεστή προσδιορισμού r^2

Ερωτήσεις Πολ/λης Επιλογής 2.2. Αν $r^2 = 0,84$, τότε :

- Το 84% της μεταβλητότητας του y εξηγείται από το x
- Η κλίση b_1 της ευθείας ισούται με 0,84
- Η συσχέτιση r ισούται με 0,84
- Το 84% των παρατηρήσεων βρίσκεται πάνω στην ευθεία

Ερωτήσεις Πολ/λης Επιλογής 2.3. Ποιο από τα παρακάτω δεν αποτελεί υπόθεση εφαρμογής της γραμμικής παλινδρόμησης :

- Γραμμικότητα
- Ανεξαρτησία καταλοίπων
- Κανονικότητα καταλοίπων
- Ο μέσος όρος του x ίσος με μηδέν

Ερωτήσεις Πολ/λης Επιλογής 2.4. Η εξίσωση παλινδρόμησης $\hat{y} = 2 + 3x$ ερμηνεύεται ως :

- Για κάθε μονάδα αύξησης του y , το x αυξάνεται κατά 3
- Για κάθε μονάδα αύξησης του x , το y αυξάνεται κατά 3
- Το y είναι τριπλάσιο του x
- Το x είναι τριπλάσιο του y μείον 2

Ερωτήσεις Πολ/λης Επιλογής 2.5. Η ευθεία ελαχίστων τετραγώνων πάντα διέρχεται από :

- Την αρχή των αξόνων $(0, 0)$
- Το σημείο (\bar{x}, \bar{y})
- Τη μέγιστη τιμή του y
- Τη μέγιστη τιμή του x

Ερωτήσεις Πολ/λης Επιλογής 2.6. Αν $r^2 = 0$, τότε :

- Η ευθεία παλινδρόμησης είναι οριζόντια
- Ο x δεν εξηγεί καθόλου τη μεταβλητότητα του y
- Όλα τα κατάλοιπα ισούνται με μηδέν
- Το ΣΣΤ ισούται με μηδέν

Ερωτήσεις Πολ/λης Επιλογής 2.7. Ισχύει: $\Sigma\Sigma\Gamma = \Sigma\Sigma\rho + \Sigma\Sigma\epsilon$. Αν $\Sigma\Sigma\Gamma = 100$ και $\Sigma\Sigma\epsilon = 20$, ποιο είναι το r^2 ;

- 0,20 0,25 0,80 1,25

Σημείωση: $\Sigma\Sigma\rho = 100 - 20 = 80$, άρα $r^2 = 80/100 = 0,80$.

Ερωτήσεις Πολ/λης Επιλογής 2.8. Ετεροσκεδαστικότητα σημαίνει ότι:

- Τα κατάλοιπα δεν ακολουθούν κανονική κατανομή
 Η διασπορά των καταλοίπων δεν είναι σταθερή
 Τα κατάλοιπα δεν είναι ανεξάρτητα
 Η σχέση x - y δεν είναι γραμμική

Ερωτήσεις Πολ/λης Επιλογής 2.9. Ποιο διάγραμμα χρησιμοποιείται για τον έλεγχο της κανονικότητας των καταλοίπων;

- Διάγραμμα διασποράς (x έναντι y)
 Διάγραμμα κανονικής πιθανότητας (normal probability plot)
 Ιστόγραμμα του x
 Θηκόγραμμα του y

Ερωτήσεις Πολ/λης Επιλογής 2.10. Η χρήση εξίσωσης παλινδρόμησης για πρόβλεψη έξω από το διάστημα $[\min(x_i), \max(x_i)]$ ονομάζεται:

- Παρεμβολή
 Εξωπολάτωση
 Κανονικοποίηση
 Στάθμιση

2.6 Ασκήσεις

Άσκηση 2.1 (Εκτίμηση συντελεστών παλινδρόμησης). Δίνονται τα παρακάτω δεδομένα για τον αριθμό τηλεφωνικών κλήσεων (x) και τον χρόνο εξυπηρέτησης σε λεπτά (y) για 6 κλήσεις:

Κλήση	x	y
1	2	5
2	3	8
3	5	11
4	4	9
5	6	14
6	7	16

- i. Υπολογίστε \bar{x} , \bar{y} , b_1 και b_0 .
- ii. Γράψτε την εξίσωση παλινδρόμησης.
- iii. Εκτιμήστε το y για $x = 5,5$.

Λύση:

$$i. \bar{x} = (2 + 3 + 5 + 4 + 6 + 7)/6 = 27/6 = 4,5. \quad \bar{y} = (5 + 8 + 11 + 9 + 14 + 16)/6 = 63/6 = 10,5.$$

$$\begin{aligned} \sum (x_i - \bar{x})(y_i - \bar{y}) &= (-2,5)(-5,5) + (-1,5)(-2,5) + (0,5)(0,5) + (-0,5)(-1,5) + (1,5)(3,5) + (2,5)(5,5) \\ &= 13,75 + 3,75 + 0,25 + 0,75 + 5,25 + 13,75 = 37,5 \end{aligned}$$

$$\sum (x_i - \bar{x})^2 = 6,25 + 2,25 + 0,25 + 0,25 + 2,25 + 6,25 = 17,5$$

$$b_1 = \frac{37,5}{17,5} \approx 2,143, \quad b_0 = 10,5 - 2,143 \cdot 4,5 \approx 0,857$$

- ii. $\hat{y} = 0,857 + 2,143 \cdot x$.
- iii. Για $x = 5,5$: $\hat{y} = 0,857 + 2,143 \cdot 5,5 \approx \mathbf{12,64}$ λεπτά.



Άσκηση 2.2 (Συντελεστής προσδιορισμού). Σε ανάλυση παλινδρόμησης με $n = 20$ παρατηρήσεις βρέθηκε $\Sigma\Gamma = 400$ και $\Sigma\Sigma\epsilon = 80$.

- i. Υπολογίστε το $\Sigma\Sigma P$ και τον r^2 .
- ii. Ερμηνεύστε τον r^2 .
- iii. Αν η τυπική απόκλιση των καταλοίπων (τυπικό σφάλμα εκτίμησης) είναι $s_e = \sqrt{\Sigma\Sigma\epsilon/(n-2)}$, υπολογίστε την.

Λύση:

- i. $\Sigma\Sigma P = \Sigma\Gamma - \Sigma\Sigma\epsilon = 400 - 80 = 320$. $r^2 = 320/400 = \mathbf{0,80}$.
- ii. Το 80% της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής εξηγείται από τη μεταβλητότητα της ανεξάρτητης. Η προσαρμογή είναι ικανοποιητική.

iii. $s_e = \sqrt{80/(20 - 2)} = \sqrt{80/18} = \sqrt{4,444} \approx 2,108$.

Άσκηση 2.3 (Πρόβλεψη και αξιολόγηση μοντέλου). Μια εταιρεία μελετά τη σχέση μεταξύ των δαπανών διαφήμισης (x , σε χιλιάδες €) και των πωλήσεων (y , σε εκατοντάδες χιλιάδες €). Από δείγμα 10 μηνών εκτιμήθηκε η εξίσωση: $\hat{y} = 3,2 + 0,8x$, με $r^2 = 0,76$.

- i. Ερμηνεύστε τον συντελεστή $b_1 = 0,8$.
- ii. Αν τον επόμενο μήνα η εταιρεία ξοδέψει 15 χιλιάδες € σε διαφήμιση, ποιες οι εκτιμώμενες πωλήσεις;
- iii. Αξιολογήστε την ποιότητα του μοντέλου με βάση τον r^2 .
- iv. Αν τα ιστορικά δεδομένα αφορούν δαπάνες από 5 έως 20 χιλιάδες €, τότε θα ήταν επικίνδυνο να χρησιμοποιήσετε το μοντέλο;

Λύση:

- i. Κάθε επιπλέον χιλιάδα ευρώ σε διαφήμιση σχετίζεται με αύξηση πωλήσεων κατά 0,8 εκατοντάδες χιλιάδων €, δηλαδή κατά 80.000 €, κατά μέσο όρο.
- ii. $\hat{y} = 3,2 + 0,8 \cdot 15 = 3,2 + 12 = 15,2$ (δηλαδή 1.520.000 €).
- iii. $r^2 = 0,76$: το 76% της μεταβλητότητας των πωλήσεων εξηγείται από τη διαφήμιση. Η προσαρμογή είναι αρκετά ικανοποιητική, αν και το 24% εξαρτάται από άλλους παράγοντες.
- iv. Το μοντέλο είναι επικίνδυνο να χρησιμοποιηθεί για τιμές έξω από το $[5, 20]$: π.χ. πρόβλεψη για $x = 0$ ή $x = 50$, καθώς δεν γνωρίζουμε αν η γραμμική σχέση ισχύει εκτός αυτού του εύρους.

Άσκηση 2.4 (Έλεγχος υποθέσεων παλινδρόμησης). Μετά την εκτίμηση ενός μοντέλου γραμμικής παλινδρόμησης, εξετάζεται το διάγραμμα καταλοίπων (x έναντι e_i) και παρατηρείται ότι για μικρές τιμές του x τα κατάλοιπα είναι αρνητικά, για μέσες τιμές θετικά και για μεγάλες τιμές ξανά αρνητικά (σχήμα ανεστραμμένου U).

- i. Ποια υπόθεση εφαρμογής παραβιάζεται και γιατί;
- ii. Τι θα προτείνατε για να βελτιωθεί το μοντέλο;

Λύση:

- i. Παραβιάζεται η υπόθεση γραμμικότητας. Το συστηματικό μοτίβο (ανεστραμμένο U) στα κατάλοιπα υποδηλώνει ότι η σχέση x - y δεν είναι γραμμική αλλά καμπυλόγραμμη.
- ii. Θα πρέπει να επεκταθεί το μοντέλο σε *πολυωνυμική παλινδρόμηση* (π.χ. $\hat{y} = b_0 + b_1x + b_2x^2$) ή να γίνει μετασχηματισμός μιας εκ των μεταβλητών (π.χ. λογαριθμικός).

Ανάλυση Χρονολογικών Σειρών Ι: Συνιστώσες και Τάση

Στα προηγούμενα κεφάλαια εξετάσαμε τη σχέση μεταξύ δύο μεταβλητών σε *στατικό* χρόνο (ζυγιστικά δεδομένα). Σε αυτό το κεφάλαιο εισάγουμε μια θεμελιώδη διαφορετική κατηγορία δεδομένων: τις **χρονολογικές σειρές**, στις οποίες μια μεταβλητή παρακολουθείται *στο χρόνο*.

3.1 Εισαγωγή στις Χρονολογικές Σειρές

Ορισμός 3.1. **Χρονολογική σειρά** (time series) ή **χρονοσειρά** είναι το σύνολο των τιμών που παίρνει μια μεταβλητή σε *διαδοχικές χρονικές στιγμές ή περιόδους*. Συμβολίζεται ως x_t ή $x(t)$, $t = 1, 2, \dots, T$.

Παραδείγματα χρονολογικών σειρών: ημερήσιος αριθμός εισαγωγών σε νοσοκομείο, μηνιαίες πωλήσεις εκδοτικού οίκου, ετήσια θερμοκρασία Βόρειου Ημισφαιρίου, εβδομαδιαία κλεισίματα χρηματιστηρίου.

Παράδειγμα 3.1. Παρακάτω δίνεται ο ημερήσιος αριθμός εισαγωγών ηλικιωμένων (+65 ετών) σε ένα νοσοκομείο για ένα διάστημα δύο μηνών (Ιανουάριος-Φεβρουάριος 1991):

Ημερομηνία	Αριθμός εισαγωγών/ημέρα
1/1/1991	8
2/1/1991	17
3/1/1991	15
4/1/1991	21
...	...
28/2/1991	18

Κάθε τιμή συμβολίζεται ως $x_0 = 8$, $x_1 = 17$, $x_2 = 15$, κ.λπ. ■

3.2 Σκοπός της Ανάλυσης Χρονολογικών Σειρών

Η ανάλυση χρονολογικών σειρών επιδιώκει τέσσερις βασικούς στόχους:

1. **Περιγραφή:** Πώς μεταβάλλονται οι διαδοχικές τιμές της χρονοσειράς; Υπάρχει τάση, εποχικότητα, κυκλικότητα;
2. **Κατανόηση-Εξήγηση:** Για ποιο λόγο μεταβάλλονται με αυτόν τον τρόπο; Ποιοι παράγοντες τις επηρεάζουν;
3. **Πρόβλεψη:** Μπορούμε να εκτιμήσουμε την τιμή της χρονοσειράς για μια μελλοντική χρονική στιγμή;
4. **Έλεγχος:** Τι μπορούμε να κάνουμε ώστε να «ελέγξουμε» τις μελλοντικές τιμές της χρονοσειράς;

3.3 Συνιστώσες Χρονολογικής Σειράς

Στην κλασική προσέγγιση ανάλυσης χρονοσειρών, υποθέτουμε ότι κάθε τιμή x_t μπορεί να αποσυντεθεί σε τέσσερις συνιστώσες.

Ορισμός 3.2. Οι τέσσερις συνιστώσες μιας χρονολογικής σειράς είναι:

- **Τάση $T(t)$ (Trend):** Η μακροχρόνια «γενική κίνηση» της χρονοσειράς. Μπορεί να είναι ανοδική, καθοδική ή σταθερή. Αν οι τιμές μεταβάλλονται γύρω από μια σταθερή τιμή, η χρονοσειρά λέγεται *στάσιμη* (stationary).
- **Κυκλική συνιστώσα $C(t)$ (Cyclical):** Κυματοειδής μεταβολή γύρω από την τάση με περίοδο *μεγαλύτερη* του έτους (π.χ. οικονομικοί κύκλοι 4-10 ετών). Οι κυκλικές επαναλήψεις δεν εμφανίζονται με σταθερό χρονικό βήμα.
- **Εποχική συνιστώσα $S(t)$ (Seasonality):** Κυκλική συνιστώσα με περίοδο ίση ή μικρότερη του έτους. Προκαλείται από εποχικούς παράγοντες (π.χ. αυξημένες πωλήσεις βιβλίων τον Σεπτέμβριο).
- **Τυχαία συνιστώσα $I(t)$ (Irregular, Noise):** Η ακανόνιστη τυχαία μεταβλητότητα που δεν

μπορεί να αποτελέσει μέρος καμίας άλλης συνιστώσας.

3.4 Υποδείγματα Αποσύνθεσης

Υπάρχουν δύο βασικοί τρόποι συνδυασμού των τεσσάρων συνιστωσών.

3.4.1 Προσθετικό Υπόδειγμα

$$x_t = T(t) + C(t) + S(t) + I(t)$$

Στο προσθετικό υπόδειγμα, *όλες* οι συνιστώσες εκφράζονται στην ίδια μονάδα μέτρησης με τη μεταβλητή x . Χρησιμοποιείται όταν η εποχική συνιστώσα παραμένει *σταθερή σε μέγεθος* ανεξαρτήτως επιπέδου τάσης.

3.4.2 Πολλαπλασιαστικό Υπόδειγμα

$$x_t = T(t) \cdot C(t) \cdot S(t) \cdot I(t)$$

Στο πολλαπλασιαστικό υπόδειγμα, μόνο η T εκφράζεται στην ίδια μονάδα με x : οι υπόλοιπες συνιστώσες είναι *δείκτες* (αδιάστατοι αριθμοί γύρω από την τιμή 1). Χρησιμοποιείται όταν η εποχική συνιστώσα *μεταβάλλεται ανάλογα* με την τάση (μεγαλώνει όταν η τάση είναι ανοδική, μικραίνει όταν είναι καθοδική).

Παρατήρηση 3.1. Το πρώτο βήμα στην ανάλυση χρονολογικής σειράς είναι πάντα η γραφική παράσταση των τιμών: το γράφημα X_t αποκαλύπτει τα χαρακτηριστικά της (τάση, εποχικότητα) και καθορίζει την κατάλληλη μέθοδο ανάλυσης.

3.5 Εκτίμηση της Τάσης

3.5.1 Μέθοδοι Εκτίμησης

Υπάρχουν τρεις βασικές μέθοδοι εκτίμησης της τάσης $T(t)$:

1. **Χειροκίνητη χάραξη:** Σχεδιασμός ευθείας ή καμπύλης «με το χέρι» στο γράφημα. Χρησιμοποιείται μόνο όταν δεν υπάρχουν καλύτερα μέσα.
2. **Ανάλυση παλινδρόμησης** (παραμετρική εκτίμηση): Εφαρμογή γραμμικής ή πολυωνυμικής παλινδρόμησης με ανεξάρτητη μεταβλητή τον χρόνο t .
3. **Κινητοί μέσοι** (μη παραμετρική εκτίμηση): Εξομάλυνση βραχυχρόνιων διακυμάνσεων (βλ. Κεφάλαιο 4).

3.5.2 Παραμετρικά Υποδείγματα Τάσης

Χρησιμοποιώντας ανάλυση παλινδρόμησης με $t = 1, 2, \dots, T$:

- **Γραμμικό:** $x_t = \beta_0 + \beta_1 \cdot t + e_t$
- **Εκθετικό:** $x_t = \beta_0 \cdot \beta_1^t + e_t$
- **Πολυωνυμικό:** $x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \dots + e_t$

Η επιλογή του κατάλληλου υποδείγματος βασίζεται στο γράφημα της χρονοσειράς και στη σύγκριση των δεικτών ποιότητας (π.χ. r^2 , ΜΣΕ).

3.6 Δείκτης Εποχικότητας

Ορισμός 3.3. Ο δείκτης εποχικότητας (seasonal index) παρουσιάζει την έκταση της εποχικής επίδρασης σε μια συγκεκριμένη χρονική περίοδο. Μετρά κατά πόσο η μέση τιμή μιας ορισμένης περιόδου είναι μεγαλύτερη ή μικρότερη από τη συνολική μέση τιμή.

Αν D_i είναι η μέση τιμή της i -οστής περιόδου (π.χ. μήνα) και \bar{D} η συνολική μέση τιμή:

$$S_i^{(\times)} = \frac{D_i}{\bar{D}} \quad (\text{πολλαπλασιαστικός δείκτης}) \quad \text{και} \quad S_i^{(+)} = D_i - \bar{D} \quad (\text{προσθετικός δείκτης})$$

Παρατήρηση 3.2. Οι πολλαπλασιαστικοί δείκτες εποχικότητας S_i αθροίζουν στον αριθμό των περιόδων (π.χ. 12 για μηνιαία δεδομένα). Οι προσθετικοί αθροίζουν στο μηδέν.

Παράδειγμα 3.2. Ο παρακάτω πίνακας δείχνει τις μηνιαίες πωλήσεις (σε χιλιάδες €) ενός εκδοτικού οίκου για 4 έτη:

Έτος	Ιαν	Φεβ	Μαρ	Απρ	Μάι	Ιουν	Ιουλ	Αυγ	Σεπ	Οκτ	Νοε	Δεκ
1	196	188	192	164	140	120	112	140	160	168	192	200
2	200	188	192	164	140	122	132	144	176	168	196	194
3	196	212	202	180	150	140	156	144	164	186	200	230
4	242	240	196	220	200	192	176	184	204	228	250	260

Η συνολική μέση μηνιαία τιμή είναι $\bar{D} = 182,08$. Ο δείκτης εποχικότητας Ιανουαρίου (πολλαπλασιαστικός):

$$D_{\text{Iav}} = \frac{196 + 200 + 196 + 242}{4} = 208,5 \quad \Rightarrow \quad S_{\text{Iav}}^{(\times)} = \frac{208,5}{182,08} \approx 1,15$$

Ερμηνεία: Τον Ιανουάριο, οι πωλήσεις είναι κατά 15% υψηλότερες από τον ετήσιο μέσο όρο. Αντίθετα, ο Ιούλιος έχει δείκτη $S_{\text{Iουλ}} = 0,79$, δηλαδή οι πωλήσεις είναι 21% χαμηλότερες από τον μέσο. ■

3.7 Πρόβλεψη με Τάση και Εποχικότητα

Ορισμός 3.4. Η ΜΣΕ (Mean Squared Error) αξιολογεί την ακρίβεια πρόβλεψης:

$$\text{ΜΣΕ} = \frac{\sum (\hat{x}_t - x_t)^2}{n}$$

Μικρότερη ΜΣΕ σημαίνει καλύτερη προσαρμογή του μοντέλου.

Η συνολική διαδικασία πρόβλεψης με τάση και εποχικότητα συνοψίζεται σε τρία βήματα:

1. **Βήμα 1:** Υπολογισμός της μελλοντικής τάσης χρησιμοποιώντας την εκτιμώμενη εξίσωση τάσης.
2. **Βήμα 2:** Πολλαπλασιασμός (ή πρόσθεση) της τιμής τάσης με τον δείκτη εποχικότητας της αντίστοιχης περιόδου.
3. **Βήμα 3** (προαιρετικά): Πολλαπλασιασμός (ή πρόσθεση) με τον δείκτη κυκλικότητας, αν αυτός εκτιμάται.

Παράδειγμα 3.3. Συνεχίζοντας το παράδειγμα του εκδοτικού οίκου, εκτιμάται πολυωνυμικό υπόδειγμα 2ου βαθμού για την τάση (χρησιμοποιώντας τους ετήσιους μέσους όρους $\bar{x}_1 = 164$, $\bar{x}_2 = 168$, $\bar{x}_3 = 180$, $\bar{x}_4 = 216$):

$$\hat{T}(t) = 180,75 - 23,717t + 8,0833t^2$$

Για το 5ο έτος ($t = 5$): $\hat{T}(5) = 180,75 - 118,585 + 202,08 \approx 264,25$.

Πρόβλεψη για τον Ιούλιο του 5ου έτους:

- Πολλαπλασιαστικό: $264,25 \times 0,79 \approx 209$ χιλιάδες €.
- Προσθετικό ($S_{\text{Ιουλ}}^{(+)} = -38,08$): $264,25 + (-38,08) \approx 226$ χιλιάδες €.

Σύγκριση ΜΣΕ: Προσθετικό (77,01) ° Πολλαπλασιαστικό (92,39), άρα το προσθετικό υπόδειγμα εφαρμόζεται καλύτερα σε αυτά τα δεδομένα. ■

3.8 Ερωτήσεις Πολ/λης Επιλογής

Ερωτήσεις Πολ/λης Επιλογής 3.1. Ποιο είναι το *πρώτο* βήμα στην ανάλυση μιας χρονολογικής σειράς;

- Εκτίμηση των δεικτών εποχικότητας
- Γραφική παράσταση της χρονοσειράς
- Επιλογή του υποδείγματος (προσθετικό ή πολλαπλασιαστικό)
- Υπολογισμός κινητών μέσων

Ερωτήσεις Πολ/λης Επιλογής 3.2. Ποια συνιστώσα χρονοσειράς αντιπροσωπεύει επαναλαμβανόμενες κυμάνσεις με περίοδο μεγαλύτερη του έτους;

- Τάση
- Κυκλική
- Εποχική
- Τυχαία

Ερωτήσεις Πολ/λης Επιλογής 3.3. Στο προσθετικό υπόδειγμα $x_t = T + C + S + I$, το άθροισμα των μηνιαίων προσθετικών δεικτών εποχικότητας ισούται με:

- 1 12 0 T

Ερωτήσεις Πολ/λης Επιλογής 3.4. Αν ο δείκτης εποχικότητας Αυγούστου (πολλαπλασιαστικός) είναι $S_{\text{Αυγ}} = 0,75$, τότε:

- Τον Αύγουστο οι τιμές είναι κατά 25% χαμηλότερες από τον ετήσιο μέσο
- Τον Αύγουστο οι τιμές είναι κατά 75% χαμηλότερες από τον ετήσιο μέσο
- Τον Αύγουστο οι τιμές είναι κατά 25% υψηλότερες από τον ετήσιο μέσο
- Δεν υπάρχει εποχικότητα τον Αύγουστο

Ερωτήσεις Πολ/λης Επιλογής 3.5. Ποιο υπόδειγμα επιλέγεται όταν η εποχική συνιστώσα αυξάνεται ανάλογα με την τάση;

- Προσθετικό
- Πολλαπλασιαστικό
- Γραμμικό
- Εκθετικό

Ερωτήσεις Πολ/λης Επιλογής 3.6. Μια χρονοσειρά είναι *στάσιμη* (stationary) όταν:

- Έχει έντονη εποχικότητα
- Οι τιμές μεταβάλλονται γύρω από μια σταθερή αναμενόμενη τιμή
- Η τάση είναι γραμμικά αυξητική
- Η τυχαία συνιστώσα είναι μηδενική

Ερωτήσεις Πολ/λης Επιλογής 3.7. Το κριτήριο ΜΣΕ χρησιμοποιείται για:

- Τον υπολογισμό της τάσης
- Τον προσδιορισμό των δεικτών εποχικότητας
- Τη σύγκριση ακρίβειας διαφορετικών μοντέλων πρόβλεψης
- Τον έλεγχο κανονικότητας

Ερωτήσεις Πολ/λης Επιλογής 3.8. Ο δείκτης εποχικότητας D_i/\bar{D} χρησιμοποιείται στο:

- Πολλαπλασιαστικό υπόδειγμα
- Προσθετικό υπόδειγμα
- Και στα δύο
- Σε κανένα από τα δύο

Ερωτήσεις Πολ/λης Επιλογής 3.9. Ποιο παραμετρικό υπόδειγμα τάσης εκφράζει αύξηση με *σταθερό απόλυτο ρυθμό*;

- Γραμμικό: $x_t = \beta_0 + \beta_1 t$
- Εκθετικό: $x_t = \beta_0 \cdot \beta_1^t$
- Τετραγωνικό: $x_t = \beta_0 + \beta_1 t + \beta_2 t^2$
- Κανένα από τα παραπάνω

Ερωτήσεις Πολ/λης Επιλογής 3.10. Σε ποιο βήμα της πρόβλεψης χρησιμοποιείται ο δείκτης εποχικότητας;

- Βήμα 1 (υπολογισμός τάσης)
- Βήμα 2 (προσαρμογή για εποχικότητα)
- Βήμα 3 (προσαρμογή για κυκλικότητα)
- Σε κανένα βήμα

3.9 Ασκήσεις

Άσκηση 3.1 (Αναγνώριση συνιστωσών). Εξετάστε τις παρακάτω χρονοσειρές και αναγνωρίστε ποιες συνιστώσες πιθανώς υπάρχουν (τάση, κυκλική, εποχική, τυχαία):

- i. Μηνιαίες πωλήσεις χριστουγεννιάτικων στολιδιών για 10 χρόνια.
- ii. Ετήσιο ΑΕΠ χώρας για 30 χρόνια, με εμφανείς οικονομικές κρίσεις κάθε 7-10 χρόνια.
- iii. Ημερήσια θερμοκρασία για ένα χρόνο.
- iv. Εβδομαδιαίες πωλήσεις ενός σούπερ μάρκετ χωρίς εποχικά προϊόντα.

Λύση:

- i. *Εποχική* (Νοέμβριος-Δεκέμβριος) + *τάση* (αν η αγορά αναπτύσσεται) + *τυχαία*. Πολύ έντονη εποχικότητα.
- ii. *Τάση* (ανοδική μακροπρόθεσμα) + *κυκλική* (οικονομικοί κύκλοι 7-10 ετών) + *τυχαία*. Πιθανώς αμελητέα εποχική συνιστώσα σε ετήσια δεδομένα.
- iii. *Εποχική* (θερμό καλοκαίρι, κρύος χειμώνας) + *τυχαία* (καιρικές μεταβολές). Η τάση είναι ουσιαστικά μηδενική σε ένα μόνο χρόνο.
- iv. *Τάση* (αν η επιχείρηση αναπτύσσεται) + *τυχαία*. Χωρίς εποχικότητα εκ ορισμού. Πιθανή εβδομαδιαία κυκλικότητα (π.χ. Σαββατοκύριακο).

Άσκηση 3.2 (Δείκτης εποχικότητας). Ένα ξενοδοχείο παρουσιάζει τα παρακάτω τριμηνιαία έσοδα (σε χιλιάδες €) για 3 έτη:

Έτος	Α' τρίμηνο	Β' τρίμηνο	Γ' τρίμηνο	Δ' τρίμηνο
1	80	120	200	100
2	90	130	220	110
3	100	140	240	120

- i. Υπολογίστε τον τριμηνιαίο μέσο για κάθε τρίμηνο και τον συνολικό μέσο.
- ii. Υπολογίστε τον πολλαπλασιαστικό δείκτη εποχικότητας για κάθε τρίμηνο.
- iii. Ερμηνεύστε τον δείκτη του Γ' τριμήνου.

Λύση:

- i. Τριμηνιαίοι μέσοι:
 - $D_A = (80 + 90 + 100)/3 = 90$
 - $D_B = (120 + 130 + 140)/3 = 130$
 - $D_\Gamma = (200 + 220 + 240)/3 = 220$

- $D_{\Delta} = (100 + 110 + 120)/3 = 110$

Συνολικός μέσος: $\bar{D} = (90 + 130 + 220 + 110)/4 = 550/4 = 137,5$.

- ii. Δείκτες εποχικότητας: $S_A = 90/137,5 = 0,655$, $S_B = 130/137,5 = 0,945$, $S_{\Gamma} = 220/137,5 = 1,600$, $S_{\Delta} = 110/137,5 = 0,800$. Έλεγχος: $0,655 + 0,945 + 1,600 + 0,800 = 4,00$.
- iii. $S_{\Gamma} = 1,60$: Το Γ τρίμηνο (καλοκαίρι) τα έσοδα είναι κατά 60% *υψηλότερα* από τον ετήσιο τριμηνιαίο μέσο. Αυτό αντικατοπτρίζει την έντονη τουριστική εποχή.



Άσκηση 3.3 (Εκτίμηση τάσης και πρόβλεψη). Τα ετήσια έσοδα (σε εκατομμύρια €) μιας εταιρείας για 5 χρόνια είναι:

Έτος (t)	Έσοδα (x_t)
1	10
2	13
3	17
4	22
5	28

- i. Χρησιμοποιώντας γραμμική παλινδρόμηση, εκτιμήστε την εξίσωση τάσης.
- ii. Προβλέψτε τα έσοδα για $t = 6$ και $t = 7$.
- iii. Σχολιάστε κατά πόσο η γραμμική τάση είναι ο κατάλληλος τύπος για αυτά τα δεδομένα.

Λύση:

i. $\bar{t} = 3$, $\bar{x} = (10 + 13 + 17 + 22 + 28)/5 = 90/5 = 18$.

$$\sum (t_i - \bar{t})(x_i - \bar{x}) = (-2)(-8) + (-1)(-5) + (0)(-1) + (1)(4) + (2)(10) = 16 + 5 + 0 + 4 + 20 = 45$$

$$\sum (t_i - \bar{t})^2 = 4 + 1 + 0 + 1 + 4 = 10$$

$$b_1 = 45/10 = 4,5, \quad b_0 = 18 - 4,5 \cdot 3 = 4,5. \quad \text{Εξίσωση: } \hat{x}_t = 4,5 + 4,5t.$$

- ii. $\hat{x}_6 = 4,5 + 4,5 \cdot 6 = \mathbf{31,5}$ εκ. €. $\hat{x}_7 = 4,5 + 4,5 \cdot 7 = \mathbf{36,0}$ εκ. €.
- iii. Τα πραγματικά δεδομένα (10, 13, 17, 22, 28) παρουσιάζουν *επιταχυνόμενη* αύξηση (οι αυξήσεις 3, 4, 5, 6 μεγαλώνουν). Αυτό υποδηλώνει ότι ένα *τετραγωνικό* ή *εκθετικό* υπόδειγμα ίσως ταιριάζει καλύτερα από το γραμμικό.



Ανάλυση Χρονολογικών Σειρών II: Εξομάλυνση και Πρόβλεψη

Στο προηγούμενο κεφάλαιο εξετάσαμε την κλασική αποσύνθεση χρονοσειρών σε συνιστώσες και εκτιμήσαμε την τάση και την εποχικότητα. Σε αυτό το κεφάλαιο παρουσιάζουμε μεθόδους **εξομάλυνσης** χρονοσειρών για την αντιμετώπιση της τυχαίας συνιστώσας, καθώς και τη **συνάρτηση αυτοσυσχέτισης** ως εργαλείο ανάλυσης.

4.1 Εξομάλυνση Χρονολογικής Σειράς

Ορισμός 4.1. Εξομάλυνση (smoothing) ή **λείανση** μιας χρονοσειράς είναι η διαδικασία με την οποία απαλείφουμε τις τυχαίες διακυμάνσεις («ακανόνιστες» ακραίες τιμές) από μια χρονοσειρά, ώστε να αναδειχθούν καλύτερα οι άλλες συνιστώσες της (τάση, εποχικότητα).

Η παρουσία της τυχαίας μεταβλητότητας συγκαλύπτει τις άλλες συνιστώσες και δυσκολεύει την αναγνώρισή τους. Η εξομάλυνση αποτελεί έναν από τους απλούστερους τρόπους για τη μείωση αυτής της τυχαίας μεταβλητότητας.

Οι κύριες τεχνικές εξομάλυνσης είναι:

1. **Κινητός μέσος** (Moving Average).
2. **Απλή εκθετική εξομάλυνση** (Simple Exponential Smoothing).
3. **Διπλή εκθετική εξομάλυνση** (Double Exponential Smoothing).

4.2 Κινητός Μέσος

Ορισμός 4.2. Ο **κινητός μέσος** (moving average) τάξης k ($MA(k)$) υπολογίζεται αντικαθιστώντας κάθε τιμή της χρονοσειράς με τον **αριθμητικό μέσο** k διαδοχικών τιμών (το «παράθυρο» εξομάλυνσης):

$$MA_t(k) = \frac{x_{t-m} + \dots + x_t + \dots + x_{t+m}}{k}$$

όπου $m = (k - 1)/2$ για περιττό k .

Παράδειγμα 4.1. Στη χρονοσειρά εισαγωγών νοσοκομείου (8, 17, 15, 21, 8, 9, 17, 19, 23, 15, 15, 7, 9, 19, 15, ...):

- $MA_3(3/1/1991) = (8 + 17 + 15)/3 = 13,33$
- $MA_3(4/1/1991) = (17 + 15 + 21)/3 = 17,67$
- $MA_5(5/1/1991) = (8 + 17 + 15 + 21 + 8)/5 = 13,80$



4.2.1 Επιλογή πλάτους k

- Για χρονοσειρές *χωρίς* εποχικότητα: μικρό k (2–5) συνήθως επαρκεί.
- Για χρονοσειρές *με* εποχικότητα: επιλέγεται k ίσο με το μήκος του εποχικού κύκλου (π.χ. $k = 12$ για μηνιαία δεδομένα).
- Μεγάλα k αφαιρούν καλύτερα την τυχαία συνιστώσα, αλλά η εξομάλυνση είναι λιγότερο «ευαίσθητη» σε αλλαγές της χρονοσειράς.

4.2.2 Πλεονεκτήματα και Μειονεκτήματα

Ο κινητός μέσος:

- ✓ Απαλείφει την κυκλική, εποχική και τυχαία συνιστώσα.
- × Οδηγεί σε *απώλεια δεδομένων* στην αρχή και στο τέλος της χρονοσειράς (δεν υπολογίζεται για τις πρώτες και τελευταίες m παρατηρήσεις).
- × Παρουσιάζει ευαισθησία στις ακραίες τιμές.

4.3 Απλή Εκθετική Εξομάλυνση

Ο κινητός μέσος δίνει *ισα βάρη* σε όλες τις k τιμές του παραθύρου και μηδενικά βάρη σε παλαιότερες τιμές. Η εκθετική εξομάλυνση επιλύει αυτό το πρόβλημα δίνοντας *φθίνουσα βαρύτητα* σε παλαιότερες τιμές.

Ορισμός 4.3. Η απλή εκθετική εξομάλυνση (Simple Exponential Smoothing) ορίζεται ως:

$$s_t = \alpha \cdot x_t + (1 - \alpha) \cdot s_{t-1}, \quad t \geq 2$$

με $s_1 = x_1$, όπου:

- x_t : η τιμή της χρονοσειράς τη στιγμή t .
- s_t : η εξομαλυμένη τιμή τη στιγμή t .
- α : ο **συντελεστής εξομάλυνσης** ή απόσβεσης, $0 \leq \alpha \leq 1$.

Αναπτύσσοντας αναδρομικά, αποδεικνύεται ότι:

$$s_t = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots + (1 - \alpha)^t x_0$$

Κάθε εξομαλυμένη τιμή λαμβάνει υπόψη *όλες* τις προηγούμενες τιμές της χρονοσειράς, αλλά με βαρύτητα που φθίνει εκθετικά όσο πηγαίνουμε πίσω στον χρόνο.

4.3.1 Σημασία του Συντελεστή α

- Αν $\alpha = 0$: η τρέχουσα παρατήρηση αγνοείται τελείως· $s_t = s_{t-1}$.
- Αν $\alpha = 1$: αγνοούνται όλες οι παρελθοντικές τιμές· $s_t = x_t$.
- **Μικρές τιμές α** : μεγάλη εξομάλυνση (η χρονοσειρά «ισοπεδώνεται»).
- **Μεγάλες τιμές α** : μικρή εξομάλυνση (η εξομαλυμένη τιμή ακολουθεί στενά τα δεδομένα).

Παράδειγμα 4.2. Για τη χρονοσειρά εισαγωγών νοσοκομείου και $\alpha = 0,1$ ή $\alpha = 0,5$:

Ημερομηνία	x_t	s_t ($\alpha = 0,1$)	s_t ($\alpha = 0,5$)
1/1/1991	8	8,00	8,00
2/1/1991	17	8,90	12,50
3/1/1991	15	9,51	13,75
4/1/1991	21	10,66	17,38
5/1/1991	8	10,39	12,69

Με $\alpha = 0,1$: η εξομαλυμένη τιμή μεταβάλλεται αργά και αντικατοπτρίζει μια μακροχρόνια τάση.

Με $\alpha = 0,5$: η εξομαλυμένη τιμή αντιδρά πιο γρήγορα στις πρόσφατες αλλαγές. ■

4.4 Διπλή Εκθετική Εξομάλυνση

Η απλή εκθετική εξομάλυνση δεν είναι αποτελεσματική όταν η χρονοσειρά έχει *τάση*. Στην περίπτωση αυτή χρησιμοποιείται η διπλή εκθετική εξομάλυνση.

Ορισμός 4.4. Η **διπλή εκθετική εξομάλυνση** (Double Exponential Smoothing) ή μέθοδος Holt χρησιμοποιεί δύο εξισώσεις:

$$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} + T_{t-1})$$

$$T_t = \gamma(s_t - s_{t-1}) + (1 - \gamma)T_{t-1}$$

με αρχικές τιμές $s_1 = x_1$ και $T_1 = x_1 - x_0$, όπου:

- s_t : εξομαλυμένη επίπεδο (level).
- T_t : εξομαλυμένη εκτίμηση της τάσης (trend).
- γ : ο συντελεστής εξομάλυνσης της τάσης, $0 \leq \gamma \leq 1$.

Η πρόβλεψη m βήματα μπροστά δίνεται από:

$$\hat{x}_{t+m} = s_t + m \cdot T_t$$

4.5 Επιλογή Μεθόδου Πρόβλεψης

Ο παρακάτω πίνακας συνοψίζει πότε χρησιμοποιείται κάθε μέθοδος:

Μέθοδος	Κατάλληλη για	Ορίζοντας
Ανάλυση τάσης	Σταθερή τάση, χωρίς εποχικότητα	Μακρύς
Προσθετικό υπόδειγμα	Σταθερή εποχική συνιστώσα	Μακρύς
Πολλαπλασιαστικό υπόδειγμα	Μεταβαλλόμενη εποχική συνιστώσα	Μακρύς
Κινητοί μέσοι	Χωρίς τάση και εποχικότητα	Βραχύς
Απλή εκθετική εξομάλυνση	Χωρίς τάση και εποχικότητα	Βραχύς
Διπλή εκθετική εξομάλυνση	Με τάση, χωρίς εποχικότητα	Βραχύς

4.6 Αυτοσυσχέτιση

Ορισμός 4.5. Αυτοσυσχέτιση (autocorrelation) είναι η συσχέτιση (εξάρτηση) μεταξύ των διαδοχικών τιμών μιας χρονοσειράς. Η **συνάρτηση αυτοσυσχέτισης** (Autocorrelation Function, ACF) υπολογίζει τη συσχέτιση μεταξύ της χρονοσειράς x_t και της ίδιας χρονοσειράς μετατοπισμένης κατά k χρονικές περιόδους (**υστέρηση** ή lag k):

$$r_k = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

Παρατήρηση 4.1. Η $A^*\Phi$ παίρνει τιμές $-1 \leq r_k \leq 1$:

- r_1 μεγάλη: κάθε τιμή συσχετίζεται ισχυρά με την αμέσως προηγούμενη (π.χ. ημερήσιες θερμοκρασίες).
- r_{12} μεγάλη (μηνιαία δεδομένα): κάθε μηνιαία τιμή συσχετίζεται ισχυρά με την τιμή του ίδιου μήνα του προηγούμενου έτους – ένδειξη εποχικότητας.
- Αν οι r_k είναι κοντά στο μηδέν για όλα τα k : τυχαία χρονοσειρά (λευκός θόρυβος).

Το **γράφημα αυτοσυσχετίσεων** (correlogram) παρουσιάζει τις τιμές r_k για $k = 1, 2, \dots$ και βοηθά στη διερεύνηση της εσωτερικής δομής της χρονοσειράς.

Παράδειγμα 4.3. Για τη μηνιαία χρονοσειρά επιβατών αεροπορικών εταιρειών 1949–1960 (Box & Jenkins):

- Η r_1 είναι πολύ μεγάλη (ισχυρή συσχέτιση διαδοχικών μηνών).
- Η r_{12} είναι επίσης μεγάλη (εποχικότητα: κορυφή κάθε καλοκαίρι).
- Η Α^οΦ φθίνει αργά: ένδειξη ισχυρής ανοδικής τάσης.



4.7 Ερωτήσεις Πολ/λης Επιλογής

Ερωτήσεις Πολ/λης Επιλογής 4.1. Σκοπός της εξομάλυνσης μιας χρονοσειράς είναι:

- Η αύξηση της τυχαίας μεταβλητότητας
- Η απάλειψη των τυχαίων διακυμάνσεων για ανάδειξη των άλλων συνιστωσών
- Η εκτίμηση των δεικτών εποχικότητας
- Η πρόβλεψη μακροπρόθεσμης τάσης

Ερωτήσεις Πολ/λης Επιλογής 4.2. Ο κινητός μέσος MA(5) για $t = 5$ υπολογίζεται ως μέσος των:

- Τελευταίων 5 τιμών: x_1, x_2, x_3, x_4, x_5
- Τιμών x_3, x_4, x_5, x_6, x_7 (κεντροποιημένος)
- Μόνο των τιμών x_4 και x_6
- Τελευταίων 5 τιμών ζυγισμένων με α Σημείωση: Για τον κεντροποιημένο MA(5) στη θέση $t = 5$ χρησιμοποιούνται οι x_3, x_4, x_5, x_6, x_7

Ερωτήσεις Πολ/λης Επιλογής 4.3. Με $\alpha = 0,9$ στην απλή εκθετική εξομάλυνση:

- Η εξομαλυμένη τιμή ακολουθεί στενά τα πρόσφατα δεδομένα
- Επιτυγχάνεται μεγάλη εξομάλυνση
- Αγνοούνται πλήρως οι πρόσφατες τιμές
- Η εξομαλυμένη τιμή ισούται με τον αριθμητικό μέσο

Ερωτήσεις Πολ/λης Επιλογής 4.4. Ποιο μειονέκτημα έχει ο κινητός μέσος σε σχέση με την εκθετική εξομάλυνση;

- Είναι πιο πολύπλοκος στον υπολογισμό
- Οδηγεί σε απώλεια δεδομένων στην αρχή και στο τέλος
- Δεν μπορεί να χρησιμοποιηθεί για εποχικά δεδομένα
- Δίνει μεγαλύτερο βάρος στις παλαιότερες τιμές

Ερωτήσεις Πολ/λης Επιλογής 4.5. Η διπλή εκθετική εξομάλυνση χρησιμοποιείται όταν η χρονοσειρά:

- Δεν έχει τάση και δεν έχει εποχικότητα
- Έχει τάση αλλά δεν έχει εποχικότητα
- Έχει εποχικότητα αλλά δεν έχει τάση
- Έχει και τάση και εποχικότητα

Ερωτήσεις Πολ/λης Επιλογής 4.6. Η αυτοσυσχέτιση r_{12} σε μηνιαία δεδομένα μετρά τη συσχέτιση μεταξύ:

- Δύο διαδοχικών μηνών
- Κάθε μήνα και της αντίστοιχης τιμής ένα χρόνο νωρίτερα
- Κάθε τιμής και του μέσου όρου
- Δύο γειτονικών ετών (ετήσια δεδομένα)

Ερωτήσεις Πολ/λης Επιλογής 4.7. Αν οι τιμές της $A^*Φ$ είναι κοντά στο μηδέν για όλα τα lags k , τότε η χρονοσειρά:

- Έχει ισχυρή εποχικότητα
- Έχει ισχυρή τάση
- Είναι ουσιαστικά τυχαία (Λευκός θόρυβος)
- Απαιτεί διπλή εκθετική εξομάλυνση

Ερωτήσεις Πολ/λης Επιλογής 4.8. Για πρόβλεψη μικρής διάρκειας σε χρονοσειρά χωρίς τάση και εποχικότητα, η πιο κατάλληλη μέθοδος είναι:

- Ανάλυση τάσης
- Πολλαπλασιαστικό υπόδειγμα
- Απλή εκθετική εξομάλυνση
- Διπλή εκθετική εξομάλυνση

Ερωτήσεις Πολ/λης Επιλογής 4.9. Αν $\alpha = 0$ στην απλή εκθετική εξομάλυνση, τότε:

- $s_t = s_{t-1}$ (η τρέχουσα τιμή αγνοείται)
- $s_t = x_t$ (αγνοείται το παρελθόν)
- $s_t = (x_t + x_{t-1})/2$
- Η μέθοδος δεν εφαρμόζεται

Ερωτήσεις Πολ/λης Επιλογής 4.10. Η πρόβλεψη m βήματα μπροστά στη διπλή εκθετική εξομάλυνση δίνεται από:

- $\hat{x}_{t+m} = s_t$
- $\hat{x}_{t+m} = s_t + m \cdot T_t$
- $\hat{x}_{t+m} = s_t \cdot T_t^m$
- $\hat{x}_{t+m} = m \cdot s_t + T_t$

4.8 Ασκήσεις

Άσκηση 4.1 (Κινητός μέσος). Δίνεται η χρονοσειρά εβδομαδιαίων πωλήσεων (σε χιλιάδες €) ενός καταστήματος για 10 εβδομάδες:

12 15 14 18 11 13 16 20 17 14

- i. Υπολογίστε τον κινητό μέσο MA(3) για κάθε εβδομάδα που είναι δυνατόν.
- ii. Υπολογίστε τον κινητό μέσο MA(5) για κάθε εβδομάδα που είναι δυνατόν.
- iii. Πόσες τιμές «χάνονται» σε κάθε περίπτωση;

Λύση:

- i. MA(3): υπολογίζεται για εβδομάδες 2-9:

Εβδομάδα	MA(3)
2	$(12 + 15 + 14)/3 = 13,67$
3	$(15 + 14 + 18)/3 = 15,67$
4	$(14 + 18 + 11)/3 = 14,33$
5	$(18 + 11 + 13)/3 = 14,00$
6	$(11 + 13 + 16)/3 = 13,33$
7	$(13 + 16 + 20)/3 = 16,33$
8	$(16 + 20 + 17)/3 = 17,67$
9	$(20 + 17 + 14)/3 = 17,00$

- ii. MA(5): υπολογίζεται για εβδομάδες 3-8: $(12 + 15 + 14 + 18 + 11)/5 = 14,00$, $(15 + 14 + 18 +$

$$(11 + 13)/5 = 14,20, (14 + 18 + 11 + 13 + 16)/5 = 14,40, (18 + 11 + 13 + 16 + 20)/5 = 15,60, \\ (11 + 13 + 16 + 20 + 17)/5 = 15,40, (13 + 16 + 20 + 17 + 14)/5 = 16,00.$$

iii. MA(3): χάνονται 2 τιμές (εβδομάδες 1 και 10). MA(5): χάνονται 4 τιμές (εβδομάδες 1, 2, 9, 10). ■

Άσκηση 4.2 (Απλή εκθετική εξομάλυνση). Χρησιμοποιώντας την ίδια χρονοσειρά (12, 15, 14, 18, 11, 13, 16, 20, 17, 14):

- Υπολογίστε s_t για $t = 1, \dots, 6$ με $\alpha = 0,3$.
- Ποια θα είναι η πρόβλεψη για την εβδομάδα 11;
- Τι παρατηρείτε για την επίδραση του α στην αντίδραση της εξομαλυμένης τιμής σε απότομες μεταβολές;

Λύση:

- $s_1 = 12$. $s_2 = 0,3 \cdot 15 + 0,7 \cdot 12 = 4,5 + 8,4 = 12,90$. $s_3 = 0,3 \cdot 14 + 0,7 \cdot 12,90 = 4,2 + 9,03 = 13,23$.
 $s_4 = 0,3 \cdot 18 + 0,7 \cdot 13,23 = 5,4 + 9,26 = 14,66$. $s_5 = 0,3 \cdot 11 + 0,7 \cdot 14,66 = 3,3 + 10,26 = 13,56$.
 $s_6 = 0,3 \cdot 13 + 0,7 \cdot 13,56 = 3,9 + 9,49 = 13,39$.
- Η πρόβλεψη για την εβδομάδα 11 είναι η τελευταία εξομαλυμένη τιμή s_{10} (υπολογίζεται ολοκληρώνοντας την αναδρομή μέχρι $t = 10$): $s_7 \approx 13,97$, $s_8 \approx 15,78$, $s_9 \approx 15,95$, $s_{10} \approx 15,16$.
Πρόβλεψη: $\hat{x}_{11} = 15,16$.
- Με $\alpha = 0,3$, η εξομαλυμένη τιμή αντιδρά αργά σε απότομες μεταβολές (π.χ. στην πτώση από 18 σε 11). Μεγαλύτερο α (π.χ. 0,7) θα αντιδρούσε πιο γρήγορα. ■

Άσκηση 4.3 (Αυτοσυσχέτιση και ερμηνεία). Υπολογίστε την αυτοσυσχέτιση r_1 για τη χρονοσειρά $x_t = 3, 5, 4, 6, 5, 7, 6, 8$.

- Υπολογίστε \bar{x} και εφαρμόστε τον τύπο.
- Ερμηνεύστε το αποτέλεσμα.

Λύση:

- $\bar{x} = (3 + 5 + 4 + 6 + 5 + 7 + 6 + 8)/8 = 44/8 = 5,5$.
Αριθμητής ($k = 1$, ζεύγη (x_t, x_{t-1}) για $t = 2, \dots, 8$):

$$\sum_{t=2}^8 (x_t - \bar{x})(x_{t-1} - \bar{x}) = (-0,5)(-2,5) + (-1,5)(-0,5) + (0,5)(-1,5) \\ + (-0,5)(0,5) + (1,5)(-0,5) + (0,5)(1,5) + (2,5)(0,5) \\ = 1,25 + 0,75 - 0,75 - 0,25 - 0,75 + 0,75 + 1,25 = 2,25$$

Παρονομαστής: $\sum (x_t - \bar{x})^2 = 6,25 + 0,25 + 2,25 + 0,25 + 0,25 + 2,25 + 0,25 + 6,25 = 18$
 $r_1 = 2,25/18 \approx 0,125$.

- ii. Η $r_1 \approx 0,125$ είναι χαμηλή, δείχνοντας *ασθενή αυτοσυσχέτιση* υστέρησης 1. Η χρονοσειρά έχει ανοδική τάση (τιμές αυξάνονται), αλλά η τιμή κάθε t δεν συσχετίζεται ισχυρά με την αμέσως προηγούμενη.



Άσκηση 4.4 (Επιλογή μεθόδου πρόβλεψης). Για κάθε μια από τις παρακάτω χρονοσειρές, επιλέξτε και αιτιολογήστε την πιο κατάλληλη μέθοδο πρόβλεψης.

- i. Μηνιαίες καταναλώσεις ηλεκτρικής ενέργειας που παρουσιάζουν έντονη εποχικότητα (υψηλή τον χειμώνα, χαμηλή το καλοκαίρι) και σταθερή ανοδική τάση σε βάθος 10 ετών.
- ii. Ημερήσιες πωλήσεις ενός εστιατορίου χωρίς σαφή τάση και εποχικότητα, με βραχύ οριζόντιο πρόβλεψης (1 ημέρα).
- iii. Τριμηνιαία κέρδη εταιρείας που εμφανίζουν συνεχώς αυξανόμενη τάση χωρίς εποχικότητα.

Λύση:

- i. **Πολλαπλασιαστικό υπόδειγμα:** Υπάρχει ανοδική τάση (η εποχική συνιστώσα αναμένεται να αυξάνεται ανάλογα με το επίπεδο) και έντονη εποχικότητα. Πρόβλεψη μακράς διάρκειας.
- ii. **Απλή εκθετική εξομάλυνση:** Χωρίς τάση και εποχικότητα, βραχύς οριζόντιος. Ο κινητός μέσος θα ήταν επίσης αποδεκτός.
- iii. **Διπλή εκθετική εξομάλυνση:** Υπαρξη (μεταβαλλόμενης;) τάσης χωρίς εποχικότητα. Εναλλακτικά, αν η τάση φαίνεται σταθερή, η ανάλυση τάσης με γραμμική παλινδρόμηση θα ήταν εξίσου κατάλληλη.

