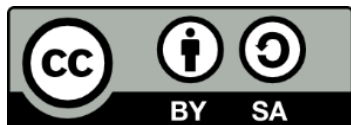


# ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΜΕ ΧΡΗΣΗ Η/Υ

## Ενότητα 9: Απλή γραμμική παλινδρόμηση

Κωνσταντίνος Ζαφειρόπουλος  
Τμήμα Διεθνών και Ευρωπαϊκών Σπουδών



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Μακεδονίας» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



- Σχέση αιτίου αποτελέσματος
- Δύο ποσοτικές μεταβλητές
- $Y$  εξαρτημένη,  $X$  ανεξάρτητη
- Ευθεία ελαχίστων τετραγώνων
- Συντελεστές του μοντέλου
- Εκτίμηση της  $Y$
- Σφάλματα-υπόλοιπα-κατάλοιπα

# Περισσότερα για την παλινδρόμηση

- Είναι ένα μοντέλο πρόβλεψης
- Προβλέψεις για παρατηρήσεις έξω από το μοντέλο δεν είναι ακριβείς, μπορούμε όμως να κάνουμε προβλέψεις για τα ενδιάμεσα σημεία
- $Y$  εξαρτημένη μεταβλητή,  $X$  ανεξάρτητη
- Αν έχουμε πολλές ανεξάρτητες μεταβλητές δεν έχουμε απλή, αλλά πολλαπλή παλινδρόμηση
- Ευθεία ελαχίστων τετραγώνων (least square) ή γραμμή παλινδρόμησης ή γραμμική παλινδρόμηση (linear regression) είναι μία και μοναδική γραμμή, που ελαχιστοποιεί το άθροισμα των τετραγώνων των αποστάσεων των  $y$  από την ευθεία

# Συντελεστής προσδιορισμού

- 
- Στην απλή παλινδρόμηση είναι το τετράγωνο του  $r$
- Παίρνει τιμές στο  $[0, 1]$
- Ισοδύναμα στο 0-100%
- Εκφράζει το ποσοστό της διασποράς της  $Y$  που ερμηνεύεται από το σύνολο των ανεξάρτητων μεταβλητών (από το μοντέλο παλινδρόμησης)

- Το σημείο με συντεταγμένες  $(\bar{x}, \bar{y})$  ή αλλιώς κέντρο βάρους (centroid) του νέφους βρίσκεται πάνω στην ευθεία
- $\hat{y}$ : εκτιμώμενη τιμή του  $y$  ή εκτίμηση ή πρόβλεψη (όσο πιο κοντά είναι στο  $y$ , τόσο το καλύτερο)
- $e$ : είναι το σφάλμα ή υπόλοιπο (residual) ή κατάλοιπο, δηλαδή η διαφορά της προβλεπόμενης από την παρατηρούμενη τιμή ( $e = y - \hat{y}$ )
- Ο μέσος όρος των σφαλμάτων είναι μηδέν.

# Ερμηνεία της διασποράς της $Y$ από το μοντέλο

- $y - \bar{y} = (y - \hat{y}) + (\hat{y} - \bar{y})$
- Οι πραγματικές τιμές των  $y$  και οι προβλεπόμενες τιμές των  $\hat{y}$  έχουν τον ίδιο μέσο όρο
- $(y - \bar{y})^2 = (y - \hat{y})^2 + (\hat{y} - \bar{y})^2$
- $\sum (y - \bar{y})^2 = \sum (y - \hat{y})^2 + \sum (\hat{y} - \bar{y})^2$
- $\sum (y - \bar{y})^2$  ή  $\sum (y - \hat{y})^2$  θέλουμε να είναι το μικρότερο δυνατόν
- συντελεστής προσδιορισμού ή  $\frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2}$
- $F$ : αν η παλινδρόμηση είναι καλή, θα πρέπει να μας δίνει ένα μεγάλο  $F$  (το **significance** είναι μικρότερο από 0,05)



- Συντελεστής προσδιορισμού
- Απονα της παλινδρόμησης
- = δείκτες καλής προσαρμογής
- Συντελεστές του μοντέλου
- Στατιστική σημαντικότητα των συντελεστών των ανεξάρτητων μεταβλητών

# Παραδείγματα

- Παραδείγματα απλής παλινδρόμησης
- Σε συνδυασμό με το συντελεστή συσχέτισης
- Ερμηνεία

# Τέλος 9ης Ενότητας



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΜΕ ΧΡΗΣΗ Η/Υ  
Τμήμα Διεθνών και Ευρωπαϊκών Σπουδών