

Ποσοτικές Μέθοδοι στη Διοίκηση

ΕΙΣΑΓΩΓΗ

Δρ. Ιωάννης Λιβιέρης
Αναπληρωτής Καθηγητής
Τμήμα Διοίκησης Επιχειρήσεων και Οργανισμών



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΕΛΟΠΟΝΝΗΣΟΥ
UNIVERSITY of the PELOPONNESE

Γραμμική παλινδρόμηση

Εξίσωση Γραμμικής παλινδρόμησης

- $y_i = \beta_0 + \beta_1 x_i + e_i$

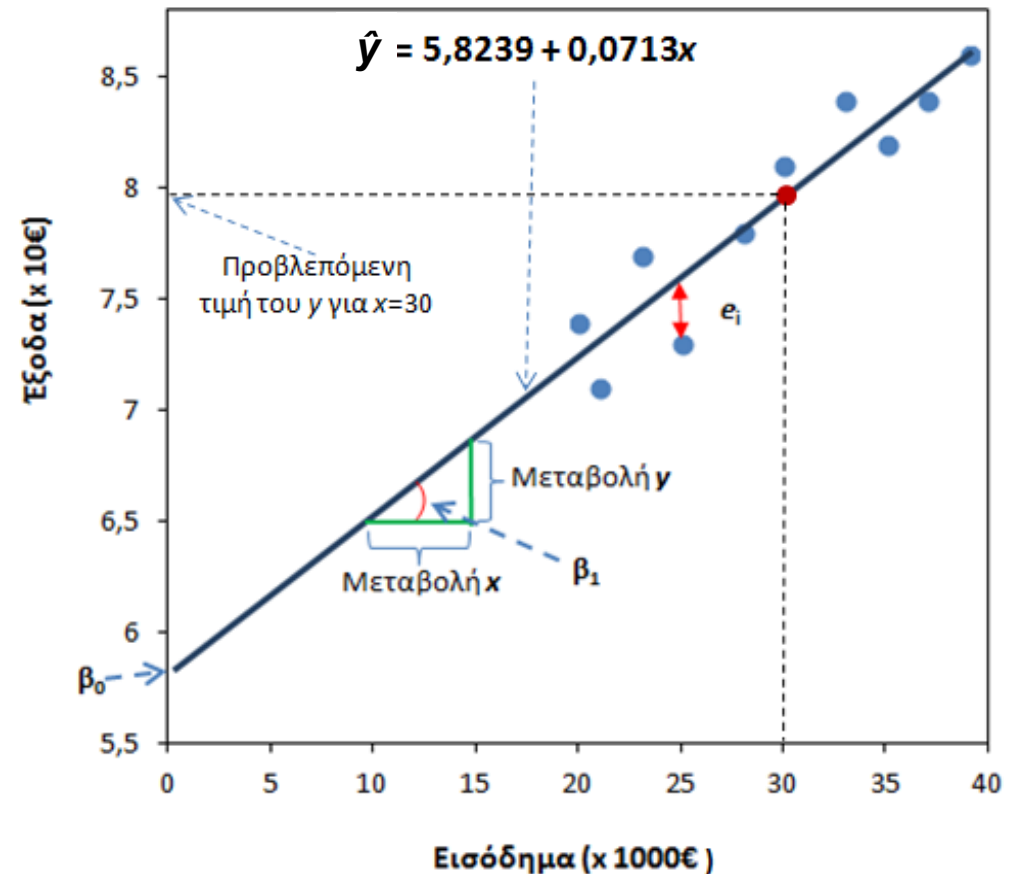
όπου

- β_0 : σταθερά ή σημείο της τομής του άξονα y για $x = 0$
- β_1 : κλίση της γραμμής παλινδρόμησης
- e_i : τυχαίο σφάλμα ή κατάλοιπο της y για την παρατήρηση i
- y_i : τιμή της εξαρτημένης μεταβλητής για την παρατήρηση i
- x_i : τιμή της ανεξάρτητης μεταβλητής για την παρατήρηση i

Γραμμική παλινδρόμηση

Υπόδειγμα Γραμμικής Παλινδρόμησης

- Η εξίσωση $y_i = \beta_0 + \beta_1 x_i$ αντιπροσωπεύει μια ευθεία γραμμή.
- Η κλίση (slope) γραμμής, β_1 , εκφράζει την αναμενόμενη μεταβολή της y για μια μοναδιαία μεταβολή της x . Με άλλα λόγια αντιπροσωπεύει την μέση μεταβολή της y που αντιστοιχεί σε μια μοναδιαία μεταβολή της x .
- Το β_0 εκφράζει την μέση αναμενόμενη τιμή της y όταν το x ισούται με 0.
- Το e_i εκφράζει τη διαφορά μεταξύ της πραγματικής τιμής y_i και της τιμής που προκύπτει από την εξίσωση παλινδρόμησης (δηλ. την απόκλιση από τη γραμμή παλινδρόμησης ή τη γραμμή πρόβλεψης).

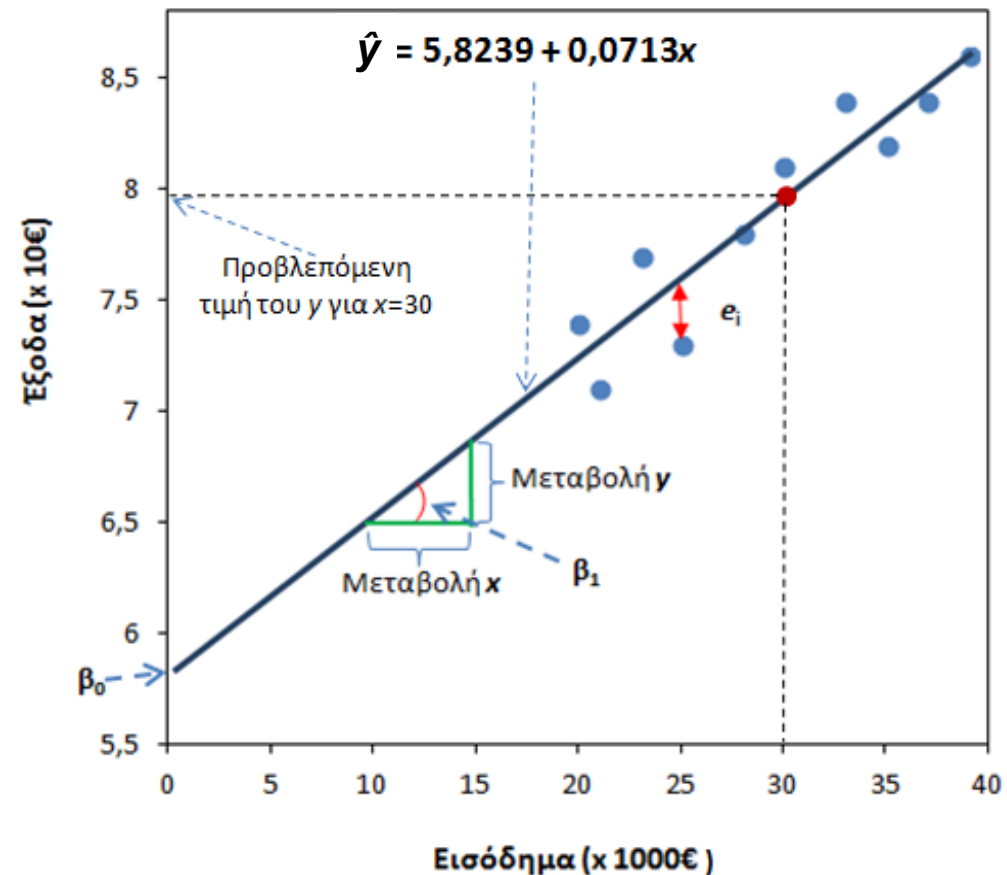


Γραμμική παλινδρόμηση

Υπόδειγμα Γραμμικής Παλινδρόμησης

- Η κλίση (slope) γραμμής, $\beta_1 = 0.0713$, εκφράζει την αναμενόμενη μεταβολή της \hat{y} για μια μοναδιαία μεταβολή της x . Με άλλα λόγια
 - αν το x από 20 γίνει 21 η μέση μεταβολή της \hat{y} θα είναι 0.0713.
 - αν το x από 20 γίνει 28 η μέση μεταβολή της \hat{y} θα είναι $8 \cdot 0.0713 \approx 0.57$
- Το β_0 εκφράζει την μέση αναμενόμενη τιμή της \hat{y} όταν το x ισούται με 0. Δηλαδή, αν $x = 0$ τότε $\hat{y} = 5.82$
- Το e_i εκφράζει τη διαφορά μεταξύ της πραγματικής τιμής y_i και της τιμής \hat{y}_i που προκύπτει από την εξίσωση παλινδρόμησης (δηλ. την απόκλιση από τη γραμμή παλινδρόμησης ή τη γραμμή πρόβλεψης).

Π.χ. για το $x = 20$, $y = 7.4$ και $\hat{y} = 5.8239 + 0.0713 \cdot 20 \approx 7.2$, επομένως $e \approx 0.2$



Ανάλυση παλινδρόμησης

Παράδειγμα

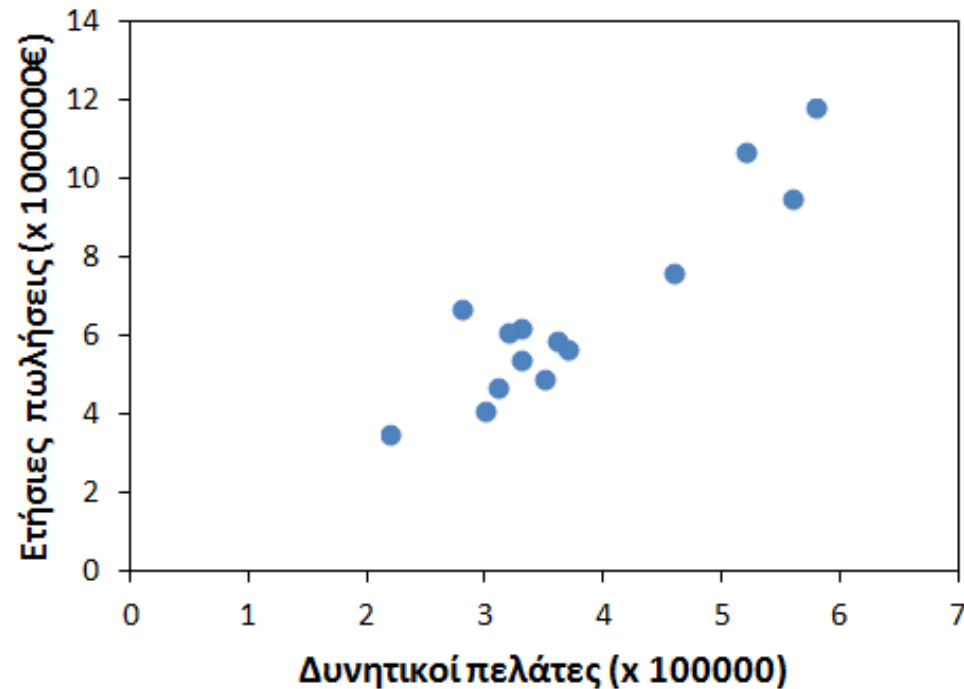
- Ο διευθυντής της Brandway επιθυμεί να προβλέψει τις ετήσιες πωλήσεις για όλα τα νέα καταστήματα, με βάση τον αριθμό των δυνητικών πελατών (π.χ. κάτοικοι συγκεκριμένου ηλικιακού εύρους που διαμένουν σε απόσταση έως 30 λεπτά από το κατάστημα της περιοχής τους).
- Για να εξεταστεί η σχέση μεταξύ του αριθμού δυνητικών πελατών και των ετήσιων πωλήσεων των καταστημάτων, συλλέχτηκαν δεδομένα από δείγμα 14 καταστημάτων .

Ανάλυση παλινδρόμησης

Κατάστημα	Δυνητικοί πελάτες (x 100000)	Ετήσιες πωλήσεις (x 1000000€)
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

Γραμμική παλινδρόμηση

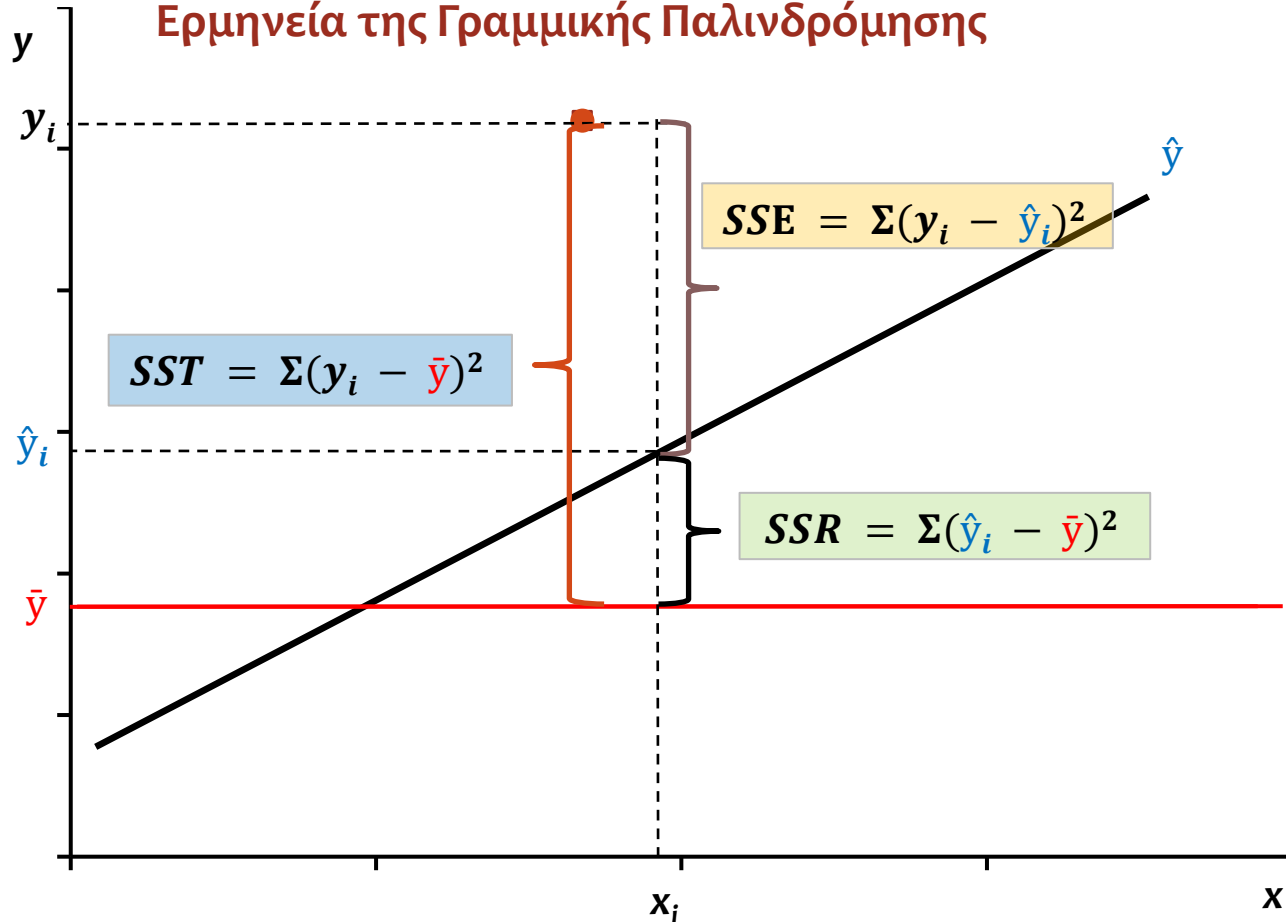
Διάγραμμα διασποράς (Scatter plot)



Μια πρώτη εικόνα της σχέσης μεταξύ δυνητικών πελατών και ετησίων πωλήσεων

Γραμμική παλινδρόμηση

Ερμηνεία της Γραμμικής Παλινδρόμησης



- **SST:** Η απόκλιση των δεδομένων από το «χειρότερο μοντέλο» (μέση τιμή)
- **SSE:** Η απόκλιση των δεδομένων από το «καλύτερο μοντέλο» (ευθεία παλινδρόμησης)
- **SSR:** Η διαφορά ανάμεσα στο «χειρότερο» και στο «καλύτερο μοντέλο»

Γραμμική Παλινδρόμηση

Ερμηνεία της Γραμμικής Παλινδρόμησης

- Στην απλή γραμμική παλινδρόμηση, συνήθως, με τη μέθοδο των ελαχίστων τετραγώνων προσδιορίζονται εκείνες οι τιμές των συντελεστών β_0 και β_1 που ελαχιστοποιούν το άθροισμα των τετραγωνικών αποκλίσεων (e_i^2) γύρω από τη γραμμή παλινδρόμησης.
- Ο συντελεστής $\beta_0 = -1.2088$ αντιπροσωπεύει την προβλεπόμενη μέση τιμή της y όταν x ισούται με 0.
- Ο συντελεστής κλίσης $\beta_1 = 2,0742$ σημαίνει ότι σε κάθε αύξηση της x κατά μία μονάδα μέτρησής της, η προβλεπόμενη μέση τιμή της y αυξάνεται κατά 2,0742 μονάδες μέτρησής της. Με άλλα λόγια για κάθε αύξηση του αριθμού των δυνητικών πελατών κατά 1 εκατομμύριο, οι προβλεπόμενες πωλήσεις αναμένεται κατά μέσο όρο να αυξηθούν κατά 2.0742 εκατομμύρια ευρώ.

Γραμμική Παλινδρόμηση

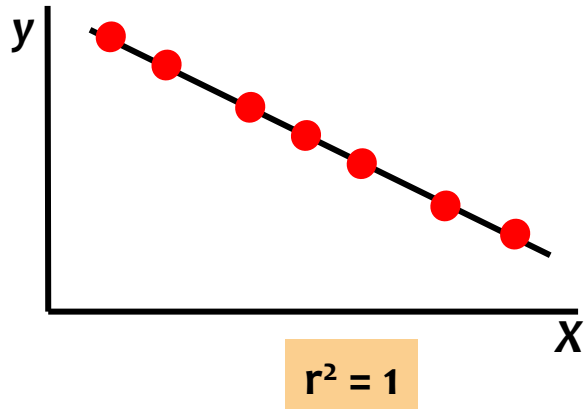
Ερμηνεία Συντελεστή Προσδιορισμού (r^2 ή R^2) – Coefficient of determination

- Ο συντελεστής προσδιορισμού μετρά το ποσοστό της μεταβλητότητας της εξηρημένης μεταβλητής y που εξηγείται από τις μεταβολές της ανεξάρτητης μεταβλητής x . Επομένως, το 84,79% της μεταβλητότητας των ετήσιων πωλήσεων εξηγείται από τη μεταβλητότητα των δυνητικών πελατών (ή το 84,79% των ετήσιων πωλήσεων οφείλεται στην επίδραση του αριθμού των δυνητικών πελατών).
- Το 15,21% της μεταβλητότητας των ετήσιων πωλήσεων των καταστημάτων του δείγματος οφείλεται σε άλλους παράγοντες, εκτός από τον αριθμό των δυνητικών πελατών.

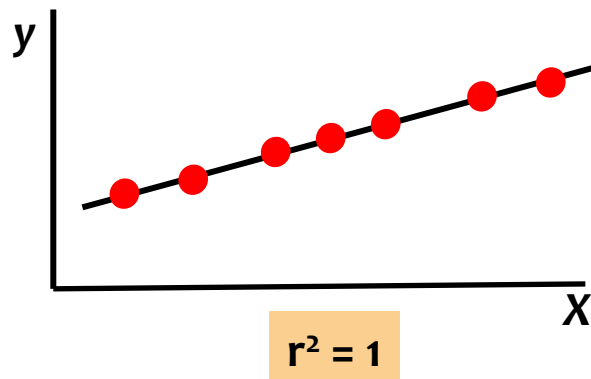
ΈΞΟΔΟΣ ΣΥΜΠΕΡΑΣΜΑΤΟΣ	
<i>Στατιστικά παλινδρόμησης</i>	
Πολλαπλό R	0,9208
R Τετράγωνο	0,8479
Προσαρμοσμένο R Τετράγωνο	0,8352
Τυπικό σφάλμα	0,9993
Μέγεθος δείγματος	14

Γραμμική Παλινδρόμηση

Ερμηνεία Συντελεστή Προσδιορισμού (r^2 ή R^2) – Coefficient of determination



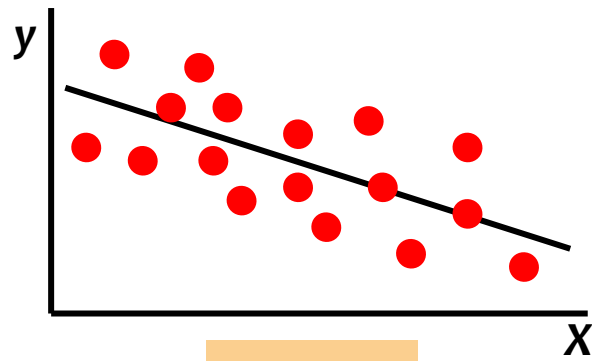
Τέλεια γραμμική σχέση μεταξύ x και y 100% της μεταβλητότητας του y εξηγείται από την μεταβλητότητα του x .



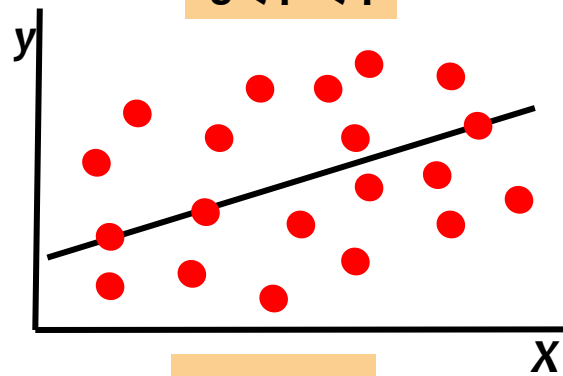
Μόνο κατά σύμπτωση και με μικρό αριθμό παρατηρήσεων μπορεί να συμβεί στις στατιστικές σχέσεις

Γραμμική Παλινδρόμηση

Ερμηνεία Συντελεστή Προσδιορισμού (r^2 ή R^2) – Coefficient of determination



$$0 < r^2 < 1$$



$$0 < r^2 < 1$$

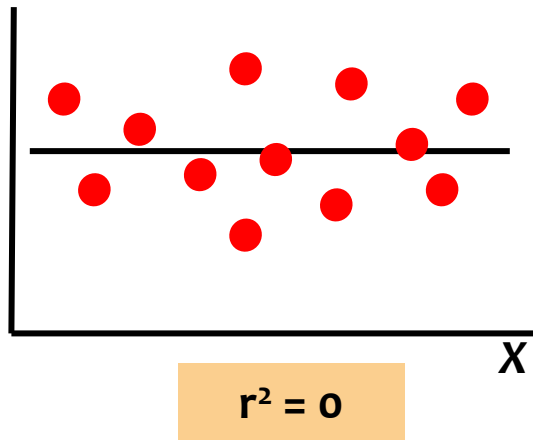
Πιο αδύναμες γραμμικές σχέσεις μεταξύ x και y

Δεν εξηγείται το 100% της μεταβλητότητας του y από την μεταβλητότητα του x

Όσο πιο διεσπαρμένα είναι τα σημεία του διαγράμματος διασποράς γύρω από τη γραμμή παλινδρόμησης τόσο πιο μικρό είναι το r^2

Γραμμική Παλινδρόμηση

Ερμηνεία Συντελεστή Προσδιορισμού (r^2 ή R^2) – Coefficient of determination



Ισχυρή ένδειξη μη ύπαρξης γραμμικής σχέσης μεταξύ x και y

Οι τιμές του y δεν εξαρτώνται από τις τιμές του x .

Η μεταβλητότητα του y δεν εξηγείται από την μεταβλητότητα του x .

Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Γραμμικότητα

- Η σχέση μεταξύ x και y είναι γραμμική, ισχύει δηλαδή στον πληθυσμό η εξίσωση

$$y = \beta_0 + \beta_1 x + e$$

Ανεξαρτησία των Σφαλμάτων

- Οι τιμές των σφαλμάτων, e_i , είναι στατιστικά ανεξάρτητες (δηλαδή, η τιμή του σφάλματος για μια παρατήρηση δεν εξαρτάται από την τιμή του σφάλματος για μια άλλη παρατήρηση)

Κανονικότητα Σφάλματος

- Οι τιμές σφαλμάτων κατανέμονται κανονικά για κάθε δεδομένη τιμή του x

Ίσες διασπορές (διακυμάνσεις) των σφαλμάτων (ομοσκεδαστικότητα)

Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Γραμμικότητα

Γραμμικότητα

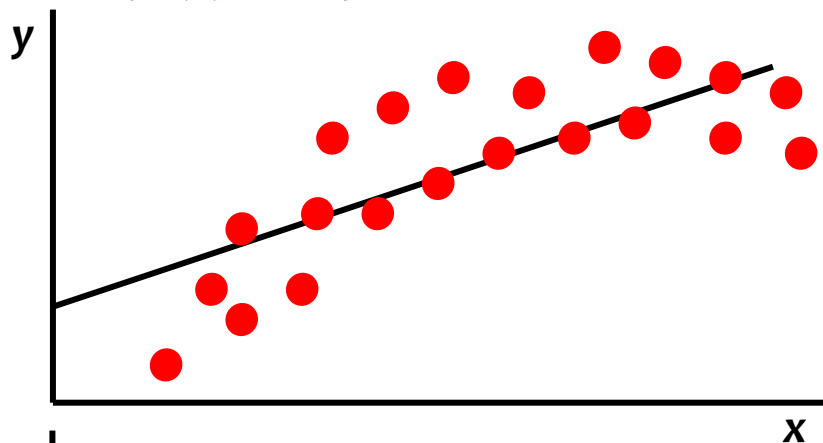
- Η σχέση μεταξύ x και y είναι γραμμική, ισχύει δηλαδή στον πληθυσμό η εξίσωση
$$y = \beta_0 + \beta_1 x + e$$

Έλεγχος Γραμμικότητας

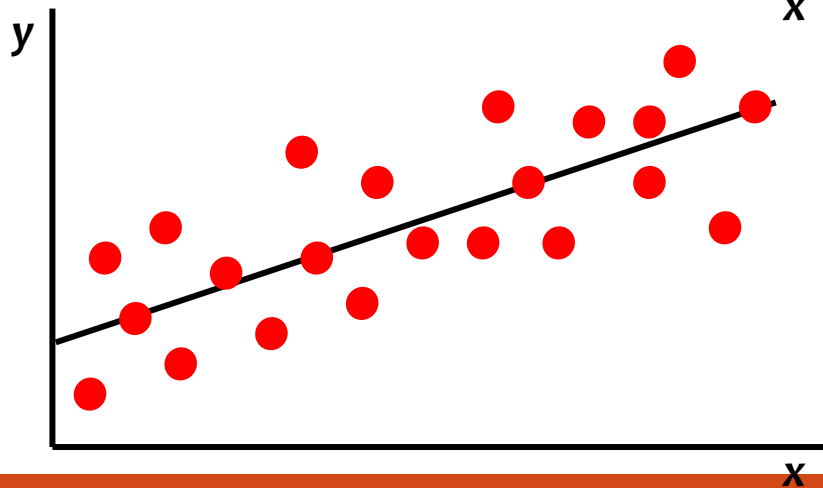
- Για την αξιολόγηση της γραμμικότητας φτιάχνεται διάγραμμα, στο οποίο απεικονίζονται τα κατάλοιπα e_i στον κάθετο άξονα και οι αντίστοιχες τιμές x_i στον οριζόντιο άξονα.
- Αν η σχέση μεταξύ των μεταβλητών x και y είναι γραμμική τότε το διάγραμμα των καταλοίπων έχει τη μορφή μιας τυχαίας διασποράς σημείων (τα σημεία δεν ακολουθούν κάποιο πρότυπο)

Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

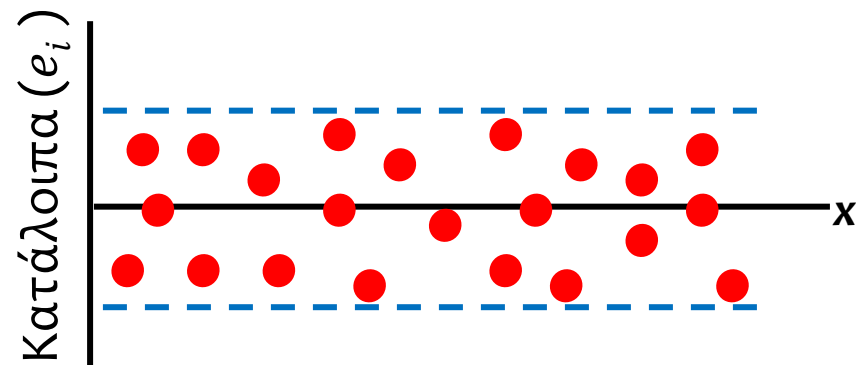
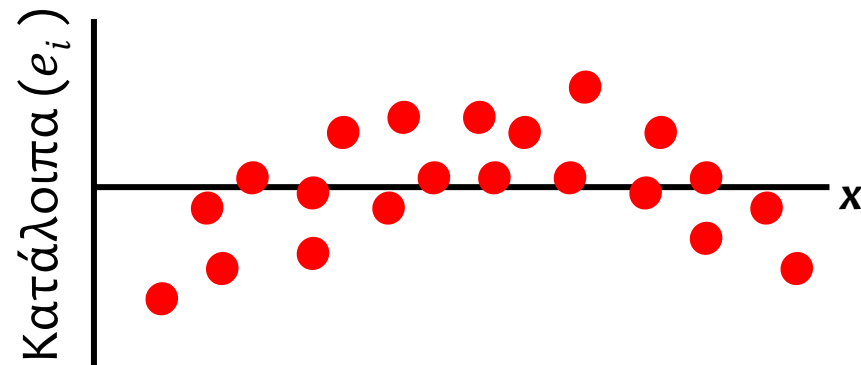
Γραμμικότητα



Μη γραμμική σχέση



Γραμμική σχέση



Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Ανεξαρτησία

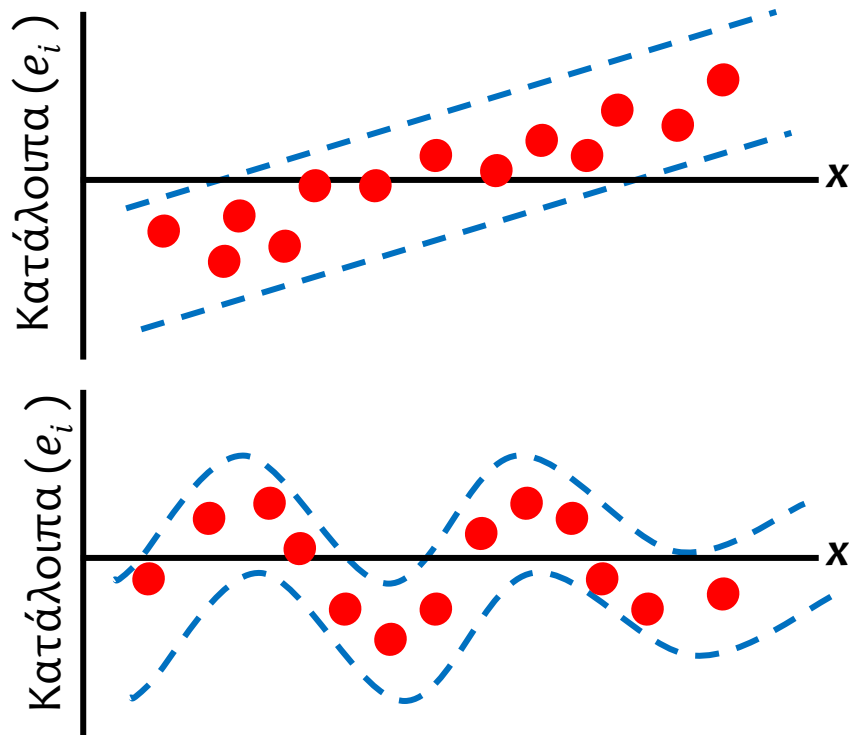
Ανεξαρτησία

- Σημαίνει ότι τα κατάλοιπα, e_i , δεν σχετίζονται με τις τιμές του x

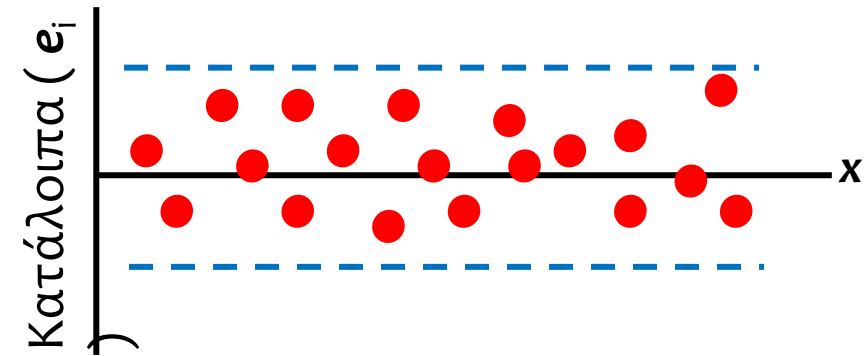
Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Ανεξαρτησία

Μη ανεξάρτητα κατάλοιπα



Ανεξάρτητα κατάλοιπα



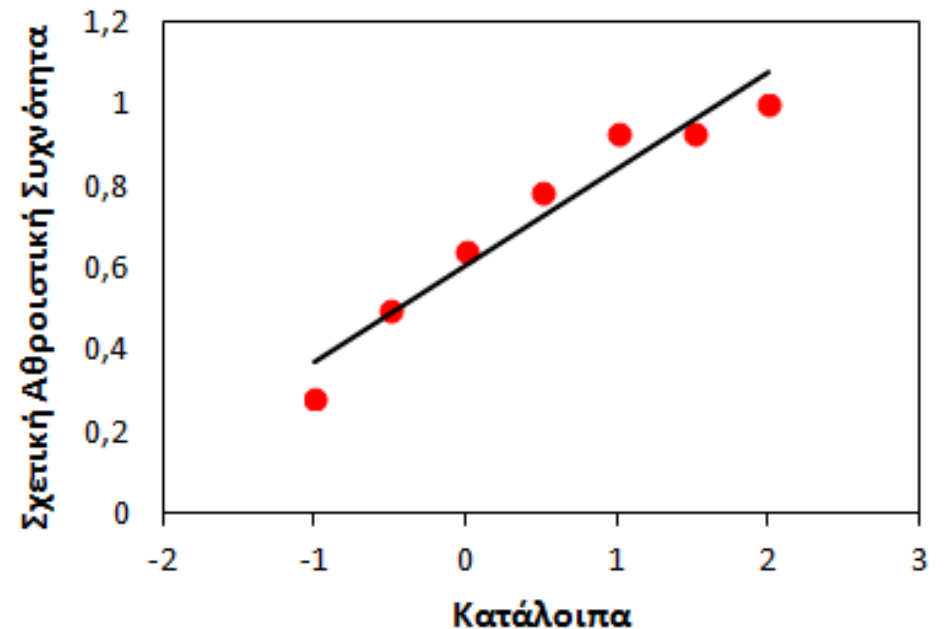
Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Κανονικότητα

Κανονικότητα

- Σημαίνει ότι η κατανομή των καταλοίπων είναι κανονική

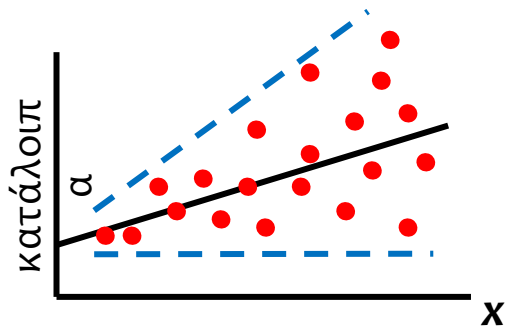
- ✓ Εξετάζετε το Θηκόγραμμα των Καταλοίπων
- ✓ Εξετάζετε το Ιστόγραμμα των Καταλοίπων
- ✓ Κατασκευάζετε ένα Διάγραμμα Πιθανοτήτων των Καταλοίπων



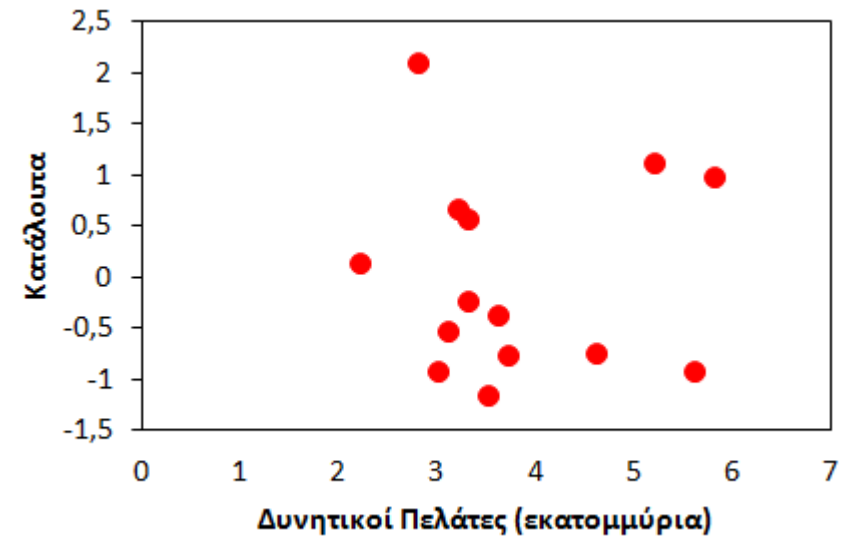
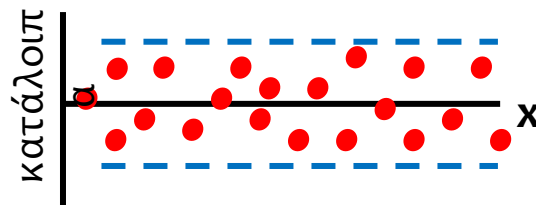
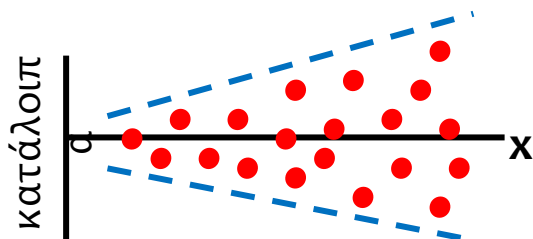
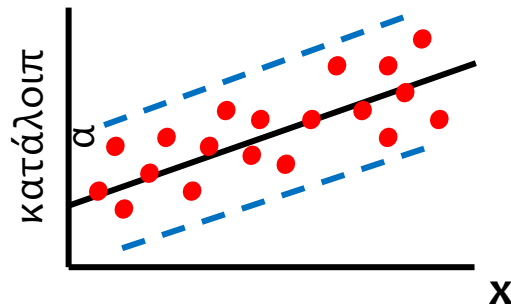
Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Ίση διασπορά

Μη σταθερή διασπορά

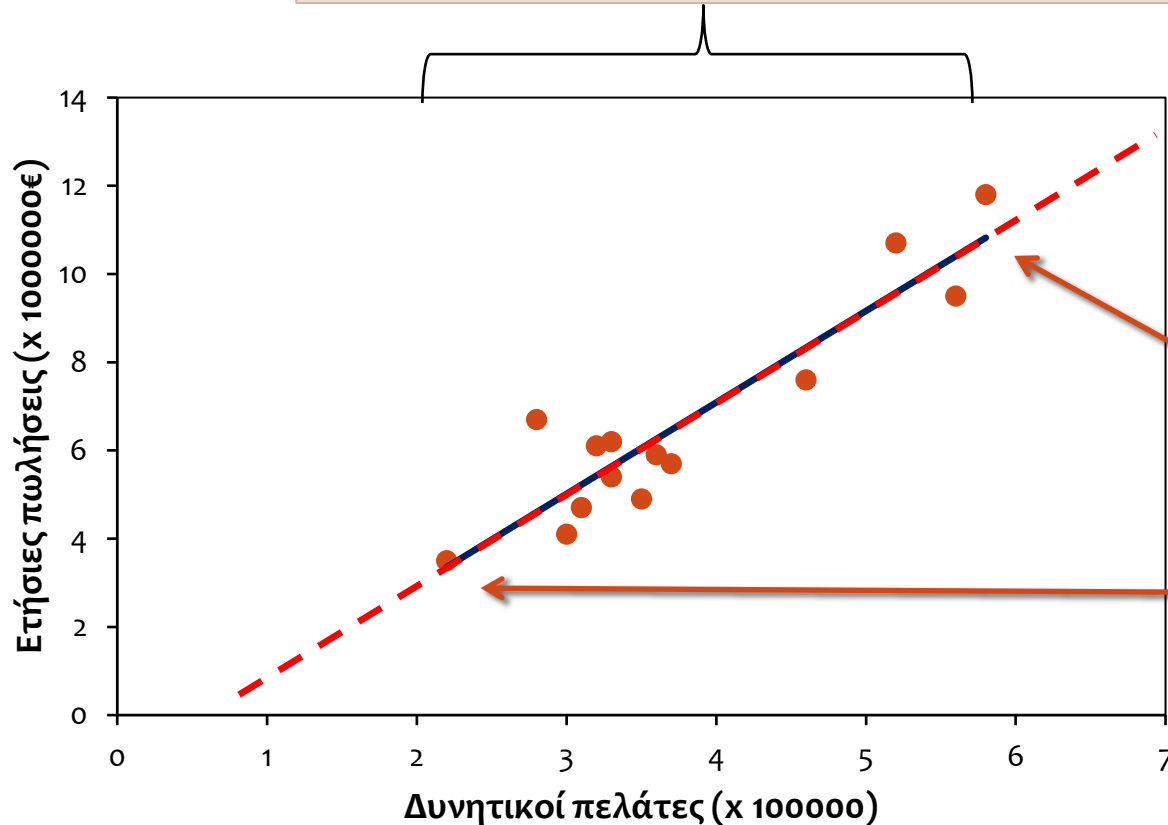


Σταθερή διασπορά



Υποθέσεις εφαρμογής της Ανάλυσης Παλινδρόμησης

Περιοχή παρεμβολής (Περιοχή παρατηρούμενων τιμών x)



Μην προσπαθήσετε να εκτιμήσετε τα y έξω από την περιοχή παρεμβολής