ΠΟΣΟΤΙΚΕΣ ΜΕΘΟΔΟΙ ΥΠΟΣΤΗΡΙΞΗΣ ΑΠΟΦΑΣΕΩΝ

Τμήμα Διοίκησης Συστημάτων Εφοδιασμού
Γεωπονικό Πανεπιστήμιο Αθηνών

Γραμμική Παλινδρόμηση

Ανάλυση Παλινδρόμησης

Στατιστική τεχνική που χρησιμοποιείται για την εύρεση της σχέσης μεταξύ μιας εξαρτημένης και μιας ή περισσοτέρων ανεξάρτητων μεταβλητών. Στοχεύει στον προσδιορισμό ενός μοντέλου το οποίο θα χρησιμοποιηθεί για την εκτίμηση των τιμών της εξαρτημένης μεταβλητής όταν μας δίνονται οι τιμές της/ των ανεξάρτητης/ανεξάρτητων μεταβλητών.

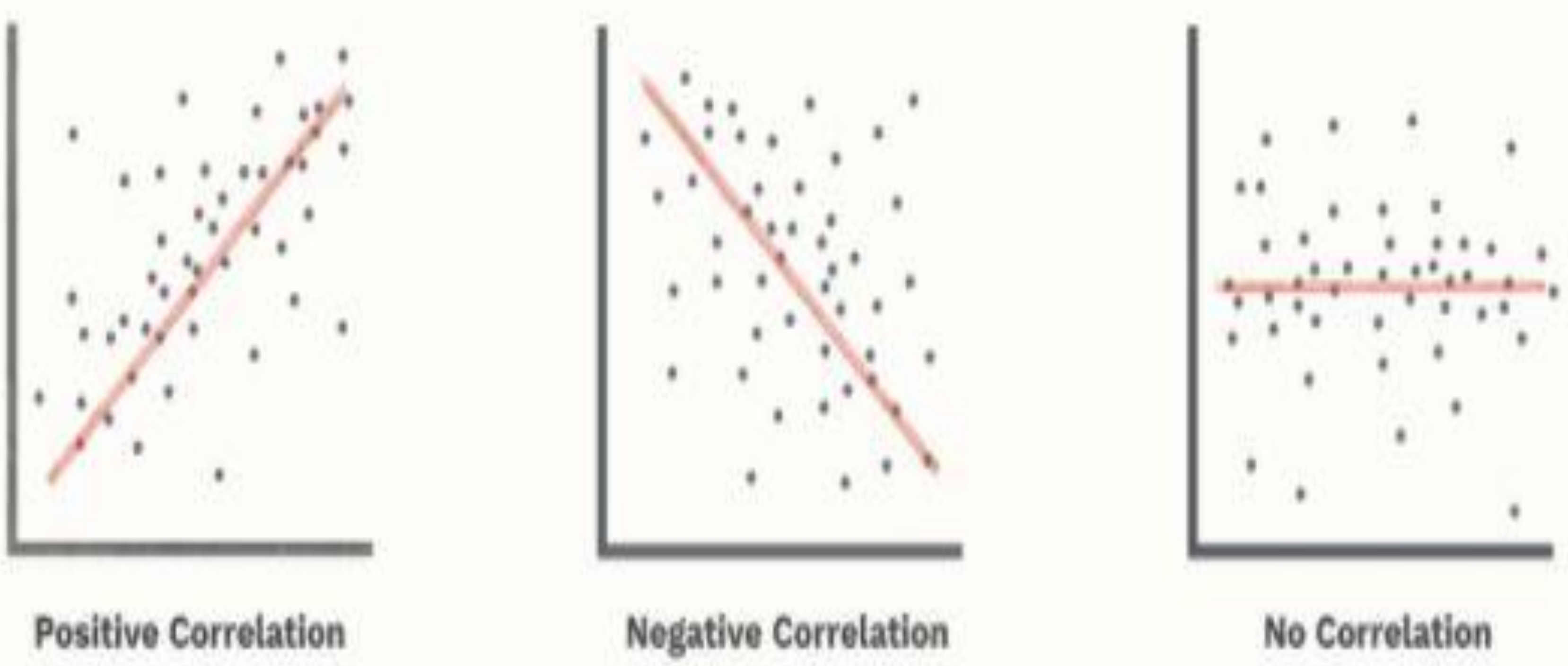
- Όταν η ανεξάρτητη μεταβλητή είναι μία και η σχέση της με την εξαρτημένη μεταβλητή είναι γραμμική τότε εφαρμόζεται απλή γραμμική παλινδρόμηση για τον προσδιορισμό του μοντέλου.
- Όταν έχουμε περισσότερες από μία ανεξάρτητες μεταβλητές και η σχέση καθεμιάς από αυτές με την εξαρτημένη μεταβλητή είναι γραμμική τότε εφαρμόζεται πολλαπλή γραμμική παλινδρόμηση για τον προσδιορισμό του μοντέλου.

Συντελεστής Συσχέτισης

$$r = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum(Y_i - \bar{Y})^2 \sum(X_i - \bar{X})^2}}$$

$$r = \frac{\sum Y_i X_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{[\sum {X_i}^2 - \frac{(\sum X_i)^2}{n}][\sum {Y_i}^2 - \frac{(\sum Y_i)^2}{n}]}}$$

$$-1 \leq r \leq 1$$


Positive Correlation
Negative Correlation
No Correlation

Απλή Γραμμική Παλινδρόμηση

Χ, Υ οι δύο μεταβλητές

$$Y_i = a_0 + a_1 X_i + \varepsilon_i, i = 1,2, \dots, n$$

όπου Y_i, X_i οι παρατηρήσεις εξαρτημένης και ανεξάρτητης μεταβλητής, a_0 η σταθερά της παλινδρόμησης, a_1 η κλίση της ευθείας της παλινδρόμησης και ε_i το σφάλμα.

Υποθέσεις του Απλού Γραμμικού Μοντέλου

1. Η σχέση μεταξύ των μεταβλητών Y και X είναι γραμμική
2. Οι τιμές της ανεξάρτητης μεταβλητής X είναι σταθερές (όχι τυχαίες). Ο μόνος τυχαίος παράγοντας είναι το σφάλμα (διαταρακτικός όρος)
3. Τα σφάλματα είναι ανεξάρτητα μεταξύ τους
4. Τα σφάλματα ακολουθούν την κανονική κατανομή με μέσο 0 και διακύμανση σ^2 δηλαδή $\varepsilon \sim \mathrm{N}(0, \sigma^2)$

Μέθοδος Ελαχίστων Τετραγώνων

$$Y_i = a_0 + a_1 X_i + \varepsilon_i, i = 1,2, \dots, n$$

$$\hat{Y} = a_0 + a_1 X$$

$$\hat{Y}_i = Y_i + \varepsilon_i$$

Ζητάμε λοιπόν να ελαχιστοποιήσουμε το άθροισμα των τετραγώνων των σφαλμάτων της παλινδρόμησης.

Δηλαδή:

$$SSE = \sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y_i})^2$$

ή

$$SSE = \sum_{i=1}^{n} (Y_i - a_0 - a_1 X_i)^2$$

Οπότε η ελαχιστοποίηση των παραπάνω συναρτήσεων οδηγεί στις κανονικές εξισώσεις:

$$a_0 = \frac{\sum Y_i}{n} - a_1 \frac{\sum X_i}{n} = \bar{Y} - a_1 \bar{X}$$

$$a_1 = \frac{\sum[(Y_i - \bar{Y})(X_i - \bar{X})]}{\sum[(X_i - \bar{X})^2]}$$

$$a_1 = \frac{\sum Y_i X_i - \frac{\sum Y_i \sum X_i}{n}}{\sum {X_i}^2 - \frac{(\sum X_i)^2}{n}}$$

Διακυμάνσεις Εκτιμητριών

- Για το σταθερό όρο $Var(a_0) = \sigma^2 \frac{\sum {X_i}^2}{n \sum (X_i - \bar{X})^2}$
- Για την κλίση $Var(a_1) = \frac{s^2}{\sum (X_i - \bar{X})^2}$
- Για τη συνδιακύμανση

$$Cov(a_0, a_1) = \sigma^2 \frac{(-\bar{X})}{\sum (X_i - \bar{X})^2}$$

Ορίζουμε ως:

$$s^2 = MSE = \frac{SSE}{n-2}$$

όπου $SSE = \sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y_i})^2$

ή

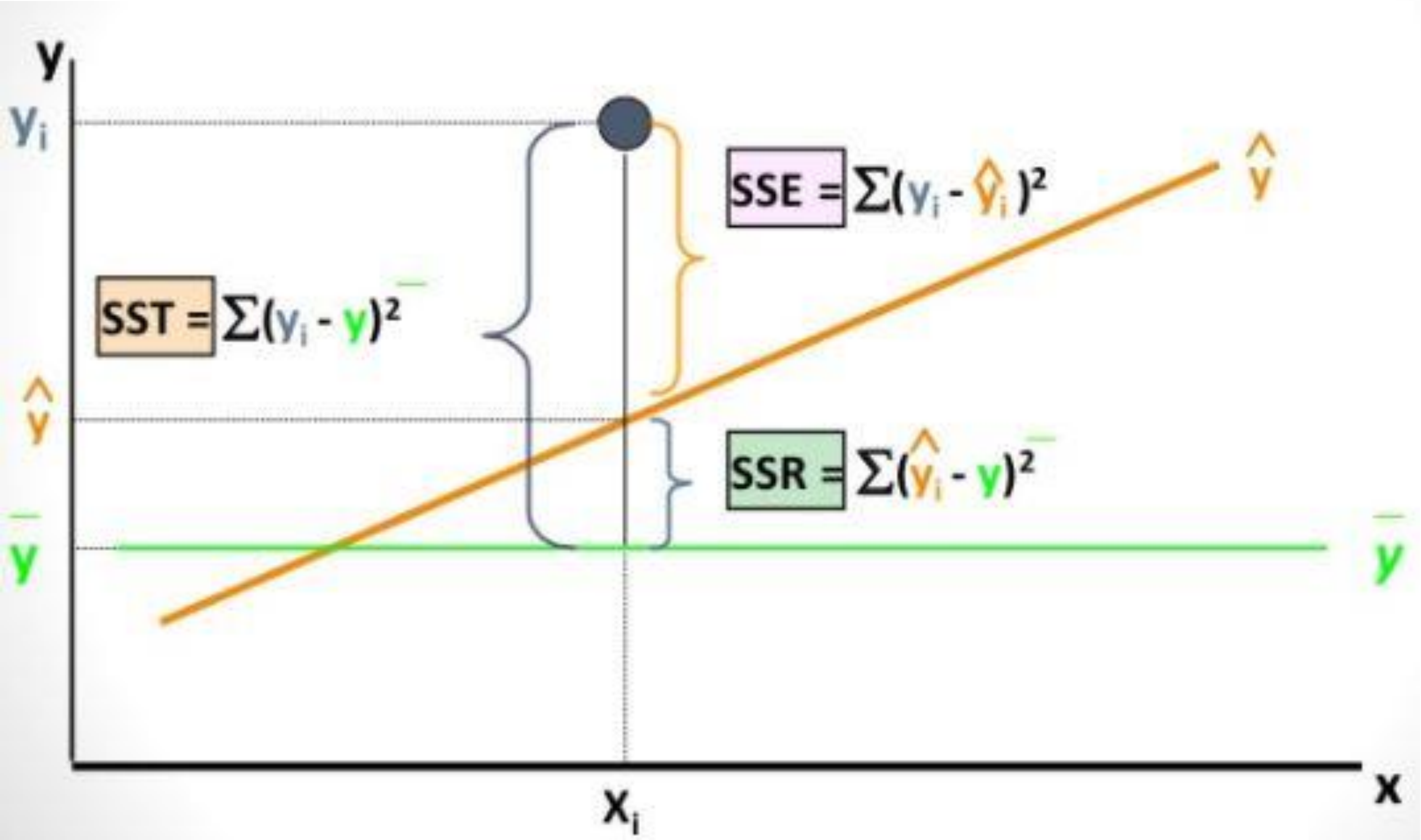
$$SSE = \sum_{i=1}^{n} (Y_i - a_0 - a_1 X_i)^2$$

Οπότε καταλήγουμε:

$$s^2(a_0) = s^2 \frac{\sqrt{\sum {X_i}^2}}{n \sum (X_i - \bar{X})^2}$$

$$s^2(a_1) = \frac{s^2}{\sum (X_i - \bar{X})^2}$$

Ερμηνεύσιμη και Ανερμήνευτη Απόκλιση



SST=SSE+SSR

SST: Συνολικό Άθροισμα Τετραγώνων

SSE: Άθροισμα Τετραγώνων των Καταλοίπων

SSR: Άθροισμα Τετραγώνων της Παλινδρόμησης

Οπότε από τα παραπάνω ορίζουμε το συντελεστή προσδιορισμού R^2

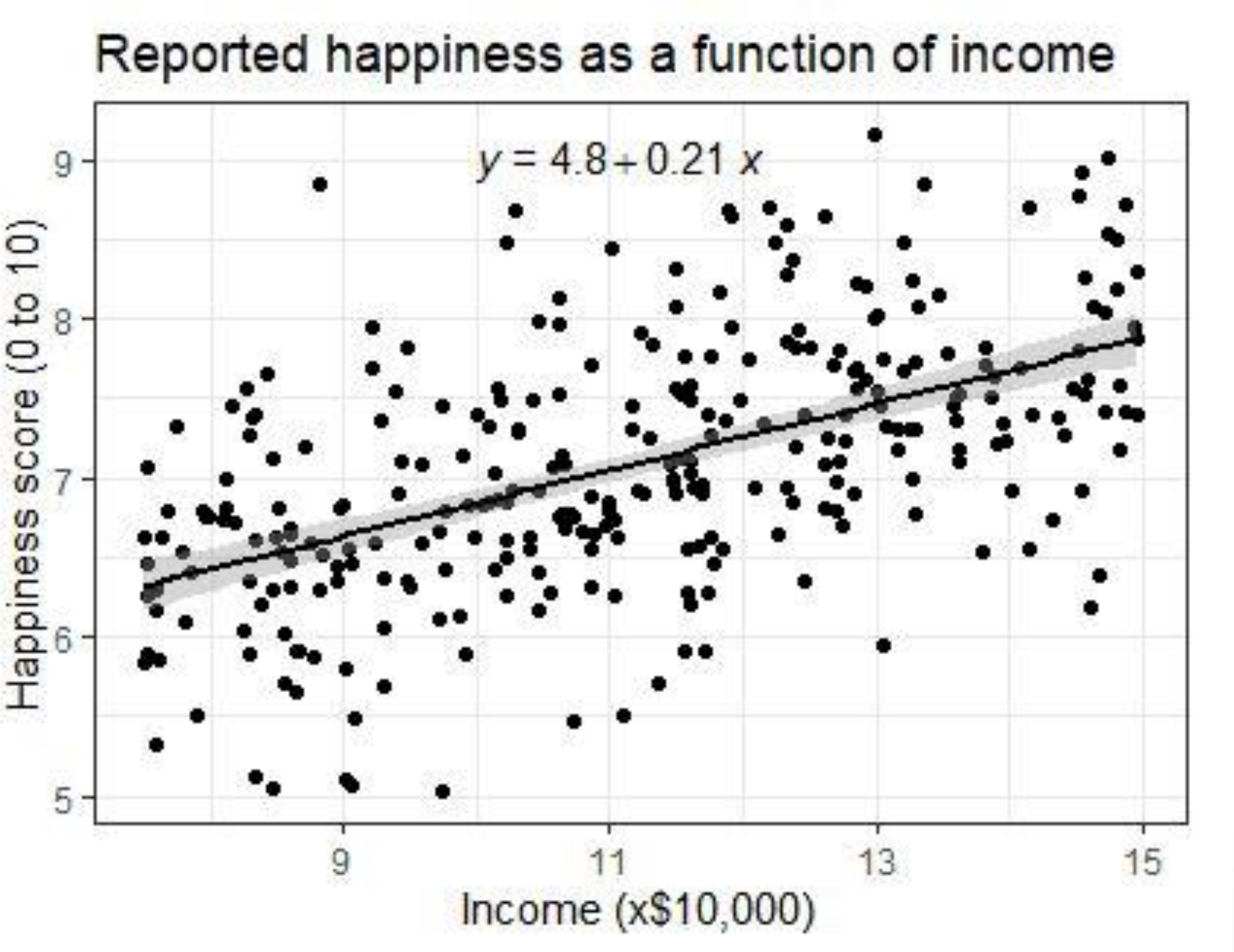
$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

με $0 \leq R^2 \leq 1$

$$R^2 = \frac{[\sum(Y_i - \bar{Y})(X_i - \bar{X})]^2}{\sum(Y_i - \bar{Y})^2 \sum(X_i - \bar{X})^2}$$

$$R^2 = \frac{\left(\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right)^2}{\left[\sum X_i{}^2 - \frac{(\sum X_i)^2}{n}\right]\left[\sum Y_i{}^2 - \frac{(\sum Y_i)^2}{n}\right]}$$

Παράδειγμα



Reported happiness as a function of income
y = 4.8 + 0.21 x
Happiness score (0 to 10)
9
8
7
6
5
9
11
13
15
Income (x$10,000)

Στο παράδειγμα έχω

$$R^2 = 0{,}475$$

Αυτό σημαίνει ότι το 47,5% της μεταβλητότητας των τιμών της Y εξηγείται από τη μεταβλητότητα της X.

Επίσης έχουμε:

$$Y = 4.8 + 0.21X + \varepsilon$$

Άρα μια αύξηση μίας μονάδας (x10000$) εισοδήματος καταλήγει σε αύξηση 0.21 (ή 21%) της ευτυχίας.

Για να είναι το συμπέρασμα έγκυρο πρέπει ο συντελεστής να είναι στατιστικά σημαντικός, δηλαδή να έχει p-value ≤ 0.1 (επίπεδο σημαντικότητας 10% τουλάχιστον, συνήθως ζητάμε p-value ≤ 0.05).

Ενδεικτική Βιβλιογραφία:

1. ΠΟΣΟΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΑΠΟΦΑΣΕΩΝ, Παπαδόγγονας, ΕΚΔΟΣΕΙΣ ΤΣΟΤΡΑΣ, 2020

Ερωτήσεις ???

Ευχαριστώ για την προσοχή σας