ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Στατιστική II

Ενότητα 11: Παλινδρόμηση I

Γεώργιος Κ. Τσιώτας
Τμήμα Οικονομικών Επιστημών

Περιεχόμενα

Συμμεταβολή Δύο Τυχαίων Μεταβλητών

Ορισμός

Για τον προσδιορισμό του μεγέθους αλλά και του προσήμου συμμεταβολής δύο τ.μ. χρησιμοποιούμε την συνάρτηση συνδιακύμανσης. Αυτή για τις X και Y ορίζεται ως:

$$Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y).$$

Ιδιότητες

1. Εάν $Cov(X, Y) > 0$ έχουμε θετική συμμεταβολή
2. Εάν $Cov(X, Y) < 0$ έχουμε αρνητική και εάν $Cov(X, Y) \to 0$ έχουμε σχεδόν μηδενική συμμεταβολή

Συμμεταβολή Δύο Τυχαίων Μεταβλητών

Συσχέτιση

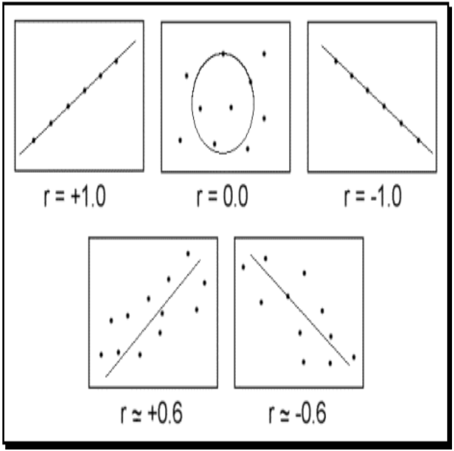
Ένα σχετικό μέτρο συμμεταβολή δύο τ.μ. αποτελεί ο συντελεστής συσχέτισης $\rho(X, Y)$. Αυτός για τις X και Y ορίζεται ως:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}}.$$

Ιδιότητες Συσχέτισης

1. $-1 \leq \rho(X, Y) \leq 1$
2. Εάν $\rho(X, Y) > 0$ έχουμε θετική συσχέτιση.
3. Εάν $\rho(X, Y) < 0$ έχουμε αρνητική συσχέτιση.
4. Εάν $\rho(X, Y) \to 0$ έχουμε σχεδόν μηδενική συσχέτιση.

Αποτύπωση της Συσχέτισης 2 μεταβλητών



Η ΄Εννοια της Παλινδρόμησης

Η ΄Εννοια της Παλινδρόμησης

Η σχέση μεταξύ δυο ή περισσοτέρων μεταβλητών μπορεί να εκφραστεί μέσω μιας συνάρτησης. Η εξίσωση (ή συνάρτηση) παλινδρόμησης εκφράζει την στατιστική εξάρτηση μεταξύ της y (ερμηνευομένης μεταβλητης) και των $x_1, \ldots x_k$ (ερμηνευτικών μεταβλητων) μέσω μιας εξίσωσης της λεγόμενης *εξίσωσης παλινδρόμησης*. Αυτή μπορεί να εκφράζεται ως:

$$y = f(x_1, \ldots x_k).$$

Η Έννοια της Απλής Γραμμικής Παλινδρόμησης

Η σχέση μεταξύ δυο μεταβλητών μπορεί να εκφραστεί μέσω μιας γραμμικής συνάρτησης. Η εξίσωση (ή συνάρτηση) παλινδρόμησης εκφράζει στην στατιστική εξάρτηση μεταξύ της y (ερμηνευομένης μεταβλητης) και της x (ερμηνευτικής μεταβλητης) μέσω μιας γραμμικής εξίσωσης της λεγόμενης *εξίσωσης παλινδρόμησης*. Αυτή μπορεί να εκφράζεται ως:

$$y = \alpha + \beta x,$$

όπου οι α και β πρέπει να εκτιμηθούν ως $\hat{\alpha}$ και $\hat{\beta}$, δεδομένου του δείγματος μας.

Η Έννοια της Απλής Γραμμικής Παλινδρόμησης (ς υν.)

Έτσι, για τις μεταβλητές y και x από δείγμα:

$$\{(y_1, x_1), \ldots, (y_n, x_n)\},$$

θα εκτιμήσουμε την:

$$\hat{y} \equiv E(y \mid x) = \hat{\alpha} + \hat{\beta} x,$$

όπου $\hat{y}$ η μέση εκτίμηση της y δεδομένου του x.

Εκτίμηση μέσω Μεθοδου Ελαχίστων τετραγώνων

Θέτουμε ως σφάλμα (ή κατάλοιπο) εκτίμησης το:

$$\epsilon_i = y_i - \alpha - \beta x_i$$

για παρατηρήσεις i από 1 εως n. Εκτιμούμε τα $\hat{\alpha}$ και $\hat{\beta}$ ελαχιστοποιώντας το άθροισμα των τετραγωνων των σφαλμάτων εκτίμησης της y_i, δηλ. της:

$$S(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{n} \hat{\epsilon}_i^2 \equiv \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_{i,1})^2$$

ως προς $\hat{\alpha}$ και $\hat{\beta}$.

Εκτίμηση μέσω Μεθοδου Ελαχίστων τετραγώνων (συν.)

Από συνθήκες πρώτης τάξης ($\partial S(\hat{\alpha},\hat{\beta})/\partial\hat{\alpha} = 0$,$\partial S(\hat{\alpha},\hat{\beta})/\partial\hat{\beta} = 0$) , έχουμε:

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{i=1}^{n} y_i}{n} - \hat{\beta}\frac{\sum_{i=1}^{n} x_i}{n}, \\ &\equiv \bar{y} - \hat{\beta}\bar{x}. \end{aligned}$$

Εκτίμηση μέσω Μεθοδου Ελαχίστων τετραγώνων (συν.)

Επίσης:

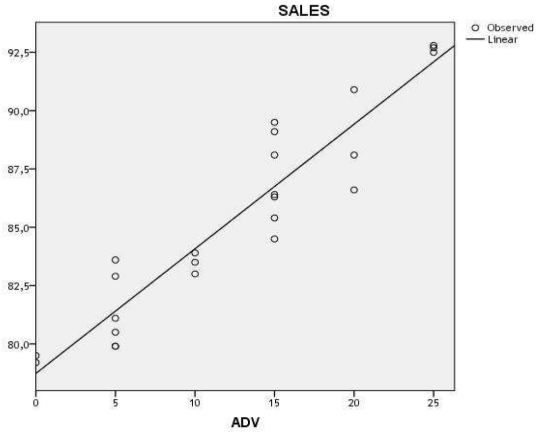
$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Παράδειγμα εκτίμησης απλής γραμμικής παλινδρόμηση

Χρησιμοποιώντας δεδομένα για Y-πωλήσεις ($SALES$) και X-δαπάνες για διαφημιση (ADV) (σε χιλ. €), προσδιορίσαμε την εξίσωση παλινδρόμησης:

$$\hat{y}_i = 78,73 + 0,534 x_i$$

Παράδειγμα εκτίμησης απλής γραμμικής παλινδρόμηση (συν.)



Τέλος Ενότητας