Στατιστική ΙΙ

Ενότητα 12: Παλινδρόμηση ΙΙ

Γεώργιος Κ. Τσιώτας
Τμήμα Οικονομικών Επιστημών

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
Ευρωπαϊκή Ένωση
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Περιεχόμενα

Διακύμανση Σφάλματος Εκτίμησης

Περιγραφή:

Αφού εκτιμήσουμε της εξίσωση παλινδρόμησης:

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

δημιουργείται το ερώτημα το πόσο καλή είναι η εν λόγω εξίσωση για την εκτίμηση των $y_1, \ldots, y_n$. Ένα μέτρο το οποίο μας δίνει εκτίμηση της ακρίβειας εκτίμησης είναι:

$$s_{\hat{\epsilon}}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{n-2},$$

η διακύμανση του σφάλματος εκτίμησης.

Παρατήρηση στη Διακύμανση Σφάλματος Εκτίμησης

Η διακύμανση του σφάλματος εκτίμησης ($s^2_{\hat{\epsilon}}$) είναι ένα στατιστικό μέτρο του οποίου αναζητάται η ελαχιστοποίηση. Έτσι, δεδομένης της ανάλυσης των μεταβλητών Y και X, αναζητάται δείγμα το οποίο μπορεί να επιτύχει την ελάχιστη τιμή για τη $s^2_{\hat{\epsilon}}$.

Συντελεστής Προσδιορισμού

Ορίζουμε τα σφάλματα μιας παλινδρόμησης ως:

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$$

Χρειάζεται να ορίσουμε ένα μέτρο `αρίστης" εφαρμογής των δεδομένων μας στην εκτιμόμενη ευθεία ($\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$).

Ο Συντελεστής Προσδιορισμού ορίζεται ως:

$$R^2 = \frac{RSS}{TSS}, \;\; RSS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2, \;\; TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Ερμηνεία του συντελεστή R^2

1. Ο R^2 μετρά το ποσοστό της συνολικής μεταβλητότητας της y που δύναται να ερμηνευτεί απο την x.
2. $0 \leq R^2 \leq 1$, με $R^2 \rightarrow 1$, δηλώνει πολύ καλή προσαρμογή, ενω όταν $R^2 \rightarrow 0$ μη-καλή προσαρμογή.

Διαστήματα Εμπιστοσύνης στην Α.Γ.Π.

Εύρεση Διακύμανσης Εκτιμητριών

Πέραν των σημειακών εκτιμήσεων $\hat{\alpha}$ και $\hat{\beta}$ για τις α και β παραμέτρους, μπορούμε να εκτιμήσουμε τα Διαστήματα Εμπιστοσύνης (Δ.Ε.) αυτών. Η διακύμανση των $\hat{\alpha}$ και $\hat{\beta}$ ορίζονται ως:

$$s_{\hat{\alpha}}^2 = s_{\hat{\epsilon}}^2 \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

και

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

αντίστοιχα (βλ. Ψωινός σελ. $363 - 366$).

Διαστήματα Εμπιστοσύνης στην Α.Γ.Π.(συν.)

Δ.Ε. για $\hat{\alpha}$ και $\hat{\beta}$

Ορίζουμε το $(1-\gamma)100\%$ Δ.Ε. για την πληθυσμιακή α ως:

$$P(\hat{\alpha} - t_{\gamma/2,n-2} \cdot s_{\hat{\alpha}} \leq \alpha \leq \hat{\alpha} + t_{\gamma/2,n-2} \cdot s_{\hat{\alpha}}) = 1-\gamma,$$

και για την β ως:

$$P(\hat{\beta} - t_{\gamma/2,n-2} \cdot s_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{\gamma/2,n-2} \cdot s_{\hat{\beta}}) = 1-\gamma.$$

Όπου $t_{\gamma/2,n-2}$ κριτική τιμή της $t-Student$ κατανομής με $n-2$ βαθμούς ελευθερίας.

Έλεγχοι Υπόθεσης στην Α.Γ.Π.

Έλεγχοι Υπόθεσης για $\hat{\alpha}$ και $\hat{\beta}$

Για ποσοστο σφάλματος γ μπορούμε να έλεγξουμε στατιστικά διάφορες υποθέσεις όπως την:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0.$$

Έλεγχοι Υπόθεσης στην Α.Γ.Π. (συν.)

Χρησιμοποιώντας την $t - Student$ κατανομή,

$$t = \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}},$$

η οποία σύμφωνα με την H_0 υπόθεση γίνεται:

$$t = \frac{\hat{\beta}}{s_{\hat{\beta}}}.$$

Έτσι, για $|t| \leq t_{\gamma/2, n-2}$ αποδεχόμαστε την H_0 (την στατιστική σημαντικότητα της x μεταβλητής).

Παράδειγμα εκτίμησης στην Α.Γ.Π

Αφού εκτιμήσαμε την εξίσωση παλινδρόμησης:

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

δεδομένων των $n = 44$ παρατηρήσεων, εκτιμούμε τα:

$R^2 = 0,885$ $s^2_{\hat{\alpha}} = 0,597^2$ $s^2_{\hat{\beta}} = 0,041^2$ $s^2_{\hat{\epsilon}} = 49,872/42.$

Παράδειγμα εκτίμησης απλής γραμμικής παλινδρόμηση: Δ.Ε και Ε.Υ. για $\hat{\alpha}$ και $\hat{\beta}$

Έτσι τα Δ.Ε διαμορφώνονται ως:

$$P(77,495 \leq \alpha \leq 79,973) = 0,95$$

και

$$P(0,449 \leq \beta \leq 0,619) = 0,95$$

για $t_{0,975,42} \approx 2,02$. Ενώ η $H_0 : \beta = 0$ απορρίπτεται γιατί

$$t = \frac{0,534}{0,041} = 13,02439 > 2,02.$$

Τέλος Ενότητας