

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΜΕ ΧΡΗΣΗ Η/Υ

Ενότητα 11: Επιλογή μεταβλητών στην παλινδρόμηση



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



Βελτιστοποίηση της παλινδρόμησης

- Αντιμετώπιση της συγγραμμικότητας και
- εισαγωγή του ελάχιστου αριθμού μεταβλητών με τη μέγιστη ερμηνευτική ικανότητα

Συντελεστής μερικής συσχέτισης

- Συντελεστής μερικής συσχέτισης ανάμεσα στις Z και Y ως προς k ανεξάρτητες μεταβλητές X_i , $i=1,..k$, είναι ο συντελεστής συσχέτισης των υπολοίπων των παλινδρομήσεων των Z και Y με τις X_i
- Χρησιμοποιείται συχνά για την επιλογή εισαγωγής μεταβλητών σε μοντέλα παλινδρόμησης

- Ας υποθέσουμε ότι έχουμε μια εξαρτημένη μεταβλητή και 4 υποψήφιες ανεξάρτητες X_1 , X_2 , X_3 και X_4
- Αρχικά υπολογίζονται όλα τα απλά μοντέλα παλινδρόμησης της Y ως προς τις τέσσερις X_i μεταβλητές
- Επιλέγεται για εισαγωγή στο μοντέλο εκείνη η X_i που έχει το μικρότερο p στον έλεγχο σημαντικότητάς της (t-test), έστω ότι είναι η X_1 . Αν το p δεν είναι μικρότερο από μια προκαθορισμένη τιμή η διαδικασία σταματά
- Στη συνέχεια, υπολογίζονται τρία μοντέλα παλινδρόμησης με εξαρτημένη μεταβλητή της Y και ανεξάρτητες δύο μεταβλητές, που η μία είναι η X_1 και η δεύτερη κάποια από τις υπόλοιπες τρεις
- Επιλέγεται ως δεύτερη μεταβλητή αυτή που στο αντίστοιχο μοντέλο έχει το μικρότερο p στον έλεγχο σημαντικότητας (t-test) αρκεί το p να είναι μικρότερο μιας προκαθορισμένης τιμής. Αλλιώς η διαδικασία σταματά. Ας υποθέσουμε ότι επιλέγεται η X_2

- Αυτή τη στιγμή έχουν εισαχθεί δύο ανεξάρτητες μεταβλητές
- Ελέγχεται αν η σημαντικότητα p της X_1 τώρα δεν είναι μικρότερη του προκαθορισμένου ορίου. Αν δεν είναι μικρότερη αφαιρείται η X_1 από το μοντέλο. Ας υποθέσουμε ότι παραμένει
- Υπολογίζονται όλα τα μοντέλα παλινδρόμησης με εξαρτημένη της Y και ανεξάρτητες τις X_1, X_2 που έχουν εισαχθεί ήδη και μία από τις X_3, X_4 , άρα δύο μοντέλα
- Επιλέγεται για εισαγωγή η X_3 ή η X_4 , εκείνη με το μικρότερο p στον έλεγχο σημαντικότητας (t-test) που δεν υπερβαίνει μια προκαθορισμένη τιμή. Αλλιώς η διαδικασία σταματά

- Αυτή τη στιγμή έχουν εισαχθεί τρεις μεταβλητές, η X_1 , X_2 και έστω η X_3 (στο τελευταίο στάδιο)
- Ελέγχεται τώρα αν οι σημαντικότητες των X_1 , X_2 έχουν υπερβεί την προκαθορισμένη τιμή ή είναι μη σημαντικές. Αν ισχύει αυτό αφαιρούνται από το μοντέλο οι αντίστοιχες μεταβλητές. Ας υποθέσουμε ότι δεν συμβαίνει κάτι τέτοιο
- Υπολογίζεται το μοντέλο παλινδρόμησης με εξαρτημένη της Y και ανεξάρτητες τις X_1, X_2, X_3 και την X_4 ως νέα μεταβλητή εφόσον η σημαντικότητα του t-test που την αφορά δεν υπερβαίνει την προκαθορισμένη τιμή. Αλλιώς η διαδικασία σταματά με την εισαγωγή των X_1, X_2, X_3

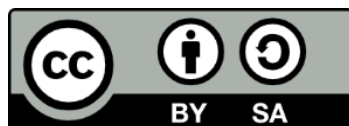
- Αν εισαχθεί και η X_4 ελέγχονται οι σημαντικότητες των X_1, X_2, X_3 για το αν υπερβαίνουν την προκαθορισμένη τιμή
- Στην περίπτωση αυτή αφαιρούνται από το μοντέλο
- Η διαδικασία συνεχίζεται μέχρι να επιλεγούν όλες οι μεταβλητές, ή όταν δεν έχουν εισαχθεί όλες αλλά η επιπλέον εισαγωγή μεταβλητών δημιουργεί μη στατιστικά σημαντικούς συντελεστές μοντέλου
- Μέγιστος αριθμός μοντέλων όσο το πλήθος των ανεξάρτητων μεταβλητών

Επιπρόσθετα

- Ελέγχονται επίσης δείκτες συγγραμικότητας των ανεξάρτητων μεταβλητών (Tolerance) και
- Συντελεστές μερικής συσχέτισης

- Η stepwise μέθοδος είναι χρήσιμη και πρακτική
- Παράγει μοντέλα με μικρό αριθμό μεταβλητών και μεγάλη ερμηνευτική ικανότητα
- Δεν παράγει όμως κατ' ανάγκη το καλύτερο μοντέλο, ή το μοναδικό καλό μοντέλο
- Ο ερευνητής έχει τον πρώτο λόγο στην επιλογή, λόγω της σημασίας τους, των μεταβλητών,
- αρκεί να πληρούνται οι στατιστικές προϋποθέσεις

Τέλος 11ης Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

