

3. ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗ ΜΕΘΟΔΟ ΤΗΣ ΠΡΟΟΔΕΥΤΙΚΗΣ ΠΡΟΣΘΗΚΗΣ ΜΕΤΑΒΛΗΤΩΝ (Forward Procedure)

Η μέθοδος της προοδευτικής προσθήκης μεταβλητών χρησιμοποιεί την ακριβώς αντίθετη λογική από αυτήν που αναπτύξαμε στην προηγούμενη ενότητα. Εδώ το μοντέλο αναπτύσσεται με την προσθήκη κάθε φορά μίας ανεξάρτητης μεταβλητής. Το πλεονέκτημα στη μέθοδο αυτή είναι ότι οι πίνακες που πρέπει να αντιστραφούν θα είναι, εν γένει, μικρότεροι από αυτούς που χρησιμοποιούμε στην μέθοδο αποκλεισμού μεταβλητών. Η μέθοδος όμως αυτή παρουσιάζει μειονεκτήματα σε σχέση με την προηγούμενη από την ιδιαιτερότητα που παρατηρείται λόγω των μεγαλύτερων τιμών των συντελεστών συσχέτισης των εξαρτημένων μεταβλητών με τις ανεξάρτητες. Ένα άλλο μειονέκτημα είναι ότι υπάρχει η δυνατότητα μόνο προσθήκης ανεξαρτήτων μεταβλητών οι οποίες, από τη στιγμή που θα ενσωματωθούν στη διαδικασία, διατηρούνται έστω και αν η τιμή της στατιστικής συνάρτησης T του συντελεστή τους παλινδρόμησης σε κάποιο βήμα δεν υπερβαίνει το κρίσιμο επίπεδο. (Οι δυσκολίες αυτές ξεπερνιούνται με την μέθοδο που θα δούμε στην επόμενη ενότητα).

Η μέθοδος επιλογής μοντέλου με την προσθήκη μεταβλητών ακολουθεί τα εξής βήματα:

- 1ο Από ένα σύνολο ανεξαρτήτων μεταβλητών $X_1, X_2, \dots$ που είναι υποψήφιος να περιληφθούν στο μοντέλο διαλέγουμε την μεταβλητή X_1^* που έχει τον μεγαλύτερο συντελεστή συσχέτισης με την εξαρτημένη μεταβλητή Y .
- 2ο Υπολογίζουμε την τιμή της στατιστικής συνάρτησης T για τον έλεγχο της υπόθεσης

$$H_0 : \beta_1^* = 0$$

ότι δηλαδή το Y δεν σχετίζεται με το X_1^* στο απλό γραμμικό μοντέλο. Εάν, με βάση τα στοιχεία του δείγματος, η H_0 δεν απορριφθεί σε κάποιο συγκεκριμένο επίπεδο σημαντικότητας

θα πρέπει να οδηγηθούμε στο συμπέρασμα ότι καμιά από τις ανεξάρτητες μεταβλητές δεν πρέπει να συμπεριληφθεί στο μοντέλο και επομένως σταματάμε την διαδικασία στο σημείο αυτό. Αν $|T|$ υπερβαίνει κάποια κρίσιμη τιμή $t_{n-2, 1-\alpha/2}$ προχωρούμε σε επιλογή μίας δεύτερης μεταβλητής.

- 3ο Για να επιλέξουμε τη δεύτερη μεταβλητή X_2^* υπολογίζουμε τους συντελεστές μερικής συσχέτισης (*partial correlations*) του Y με κάθε μία από τις υπόλοιπες υποψήφιες ανεξάρτητες μεταβλητές διατηρώντας ταυτόχρονα την μεταβλητή X_1^* σταθερά. Η X_2^* θα είναι η μεταβλητή με τον μεγαλύτερο απόλυτο συντελεστή μερικής συσχέτισης. Στη συνέχεια, για να ελέγξουμε την υπόθεση:

$$H_0 : \beta_2^* = 0$$

στο μοντέλο γραμμικής παλινδρόμησης με δύο ανεξάρτητες μεταβλητές X_1^* και X_2^* υπολογίζουμε την τιμή της αντίστοιχης στατιστικής συνάρτησης T . Αν $|T| \leq t_{n-3, 1-\alpha/2}$ για κάποιο επίπεδο σημαντικότητας α θα πρέπει να σταματήσουμε τη διαδικασία επιλογής και να θεωρήσουμε ότι το μοντέλο θα πρέπει να περιέχει μόνο την ανεξάρτητη μεταβλητή X_1^* . Αν η απόλυτη τιμή του T υπερβαίνει την κρίσιμη τιμή προχωρούμε στην επιλογή της τρίτης μεταβλητής.

- 4ο Προχωρούμε με τον ίδιο τρόπο υπολογίζοντας και πάλι διαδοχικά συντελεστές μερικής συσχέτισης μεγαλύτερης τάξης σε κάθε βήμα, προσθέτοντας στο μοντέλο τη μεταβλητή με την μεγαλύτερη απόλυτη τιμή συντελεστή μερικής συσχέτισης με το Y διατηρώντας όλες τις μεταβλητές που έχουν προστεθεί μέχρι εκείνη τη στιγμή στο μοντέλο, σταθερές. Σε κάθε βήμα υπολογίζεται ο συντελεστής συσχέτισης του Y με τις επιλεγείσες μεταβλητές όπως επίσης και η τιμή της στατιστικής συνάρτησης F για τον έλεγχο της συνολικής υπόθεσης ότι όλοι οι συντελεστές παλινδρόμησης είναι μηδέν. Με τον τρόπο αυτό

είναι δυνατόν να παρακολουθούμε τη βελτίωση που παρατηρείται με τη χρησιμοποίηση των διαφορετικών υποσυνόλων ανεξαρτήτων μεταβλητών στην εξήγηση της συνολικής διασποράς. Η διαδικασία θα πρέπει να σταματήσει όταν η τιμή της στατιστικής συνάρτησης $|T|$ για την ανεξάρτητη μεταβλητή που επελέγει τελευταία δεν υπερβαίνει κάποιο προκαθορισμένο κρίσιμο επίπεδο. Το τελικό σύνολο ανεξαρτήτων μεταβλητών που επιλέγεται για να χρησιμοποιηθεί στο μοντέλο είναι αυτό που αποτελείται από όλες τις μεταβλητές, εκτός της τελευταίας, που έχει τιμή της στατιστικής συνάρτησης $|T|$ μη στατιστικά σημαντική. Φυσικά, αναμένεται ότι η τιμή της στατιστικής συνάρτησης F για τον έλεγχο της συνολικής υπόθεσης για μηδενικούς συντελεστές παλινδρόμησης για τις μεταβλητές που έχουν επιλεγεί δεν θα ξεπερνά το συγκεκριμένο κρίσιμο επίπεδο που έχει επιλεγεί.