

7. ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ (Multicollinearity)

Η ερμηνεία ενός προβλήματος με τη χρησιμοποίηση της μεθόδου αναλύσεως της πολλαπλής παλινδρόμησης επιτυγχάνεται καλύτερα όταν οι ανεξάρτητες μεταβλητές που αποτελούν το μοντέλο είναι μεταξύ τους ασυσχέτιστες. Όταν υφίσταται έντονες συσχετίσεις μεταξύ των μεταβλητών είναι δύσκολο, αν όχι αδύνατο, να αξιολογηθεί η ουσιαστική προσφορά μιας συγκεκριμένης ανεξάρτητης μεταβλητής επί της εξαρτημένης μεταβλητής που οφείλεται αποκλειστικά στη συγκεκριμένη ανεξάρτητη μεταβλητή. Όταν οι ανεξάρτητες μεταβλητές δεν είναι ορθογώνιες μεταξύ τους είναι ενδεχόμενο οι εκτιμώμενοι συντελεστές παλινδρόμησης να είναι εξαιρετικά ασταθείς και οι τιμές τους να υφίστανται δραματικές αλλαγές όταν κάποια νέα μεταβλητή προστίθεται ή απομακρύνεται ή όταν συμβαίνουν μικρές μεταβολές στα δεδομένα του προβλήματος.

Η κατάσταση η οποία δημιουργείται όταν υπάρχουν ισχυρές συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών στην πολλαπλή παλινδρόμηση ονομάζεται **πολυσυγγραμμικότητα (multicollinearity)**.

Στις περιπτώσεις που το πρόβλημα αυτό υφίσταται θα πρέπει κανείς να είναι ιδιαίτερα προσεκτικός στην ερμηνεία όλων των εκτιμητριών που προκύπτουν από το μοντέλο αυτό.

Υπάρχουν μια σειρά από προειδοποιητικές ενδείξεις που αν ο ερευνητής τις προσέξει είναι δυνατόν να αντιληφθεί ότι υπάρχει πολυσυγγραμμικότητα. Η πιο σημαντική από αυτές είναι ο πίνακας των συντελεστών συσχέτισεως (*Correlation Matrix*) των ανεξάρτητων μεταβλητών. Αν στον πίνακα αυτόν υπάρχουν μεγάλες θετικές ή αρνητικές τιμές θα έχουμε μια ένδειξη ότι οι αντιστοιχες ανεξάρτητες μεταβλητές που χρησιμοποιούνται στο μοντέλο έχουν μεταξύ τους ισχυρό βαθμό συσχέτισης. Το στατιστικό συμπέρασμα που προκύπτει στις περιπτώσεις αυτές είναι ότι κάποιες από τις μεταβλητές συνεισφέρουν ελάχιστα ή καθόλου στην πρόβλεψη της εξαρτημένης μεταβλητής οπότε και θα πρέπει να απομακρυνθούν από το μοντέλο. Εάν, παρ' όλα αυτά, ο ερευνητής είναι βέβαιος ότι ο καθορισμός των

ανεξάρτητων μεταβλητών έγινε σωστά θα πρέπει να εξετάσει δύο άλλες ενδείξεις για το κατά πόσον υπάρχει πολυσυγγραμμικότητα.

α) Εάν τα πρόσημα ορισμένων συντελεστών στην παλινδρόμηση είναι αντίθετα από αυτά που θα περίμενε κανείς λόγω της φύσης του προβλήματος και

β) Εάν σημαντικοί συντελεστές της παλινδρόμησης εμφανίζονται να έχουν μεγάλες τιμές στις τυπικές αποκλίσεις τους.

Οποιαδήποτε από τις δύο αυτές ενδείξεις θα πρέπει να προβληματίσει τον ερευνητή και να τον οδηγήσει σε μια σοβαρή έρευνα για το κατά πόσον υφίσταται πολυσυγγραμμικότητα.

Ο καθορισμός εκείνων των γραμμικών συνδυασμών των παραμέτρων β που μπορούν να εκτιμηθούν με ακρίβεια είναι εξαιρετικά δύσκολο στην περίπτωση που υφίσταται πολυσυγγραμμικότητα. Παρότι, εν γένει, δεν είναι δυνατόν να εξαληφθεί τελείως το πρόβλημα αυτό υπάρχει μια διαδικασία με την οποία ο ερευνητής μπορεί να εργασθεί με ένα μοντέλο που προκύπτει από το αρχικό με μετασχηματισμό των αρχικών μεταβλητών σε ένα σύνολο άλλων μεταβλητών που είναι ασυσχέτιστες μεταξύ τους. Η μεθοδολογία αυτή ονομάζεται **ανάλυση κυρίων συνιστωσών** (*principal component analysis*). Η τεχνική αυτή είναι μια παρά πολύ ισχυρή τεχνική τόσο στον εντοπισμό της πολυσυγγραμμικότητας όσο και ως μεθοδολογία που οδηγεί στον καθορισμό εκείνων των γραμμικών συνδυασμών των συντελεστών παλινδρόμησης που μπορούν να εκτιμηθούν με ακρίβεια (βλέπε π.χ. Ι. Πανάρετου και Ε. Ξεκαλάκη: Πολυμεταβλητή Στατιστική Ανάλυση. Τόμος Ι Αθήνα 1993).

Μια άλλη προσέγγιση είναι να υπολογισθούν οι k συντελεστές προσδιορισμού των παλινδρομήσεων κάθε μιας από τις ανεξάρτητες μεταβλητές στις υπόλοιπες $k-1$ ανεξάρτητες μεταβλητές. Εκείνες οι μεταβλητές που εμφανίζουν υψηλό συντελεστή προσδιορισμού θα πρέπει να θεωρηθεί ότι είναι συγγραμικές με τουλάχιστον μια από τις υπόλοιπες μεταβλητές. Στη συνέχεια θα πρέπει να υπολογισθεί η ομάδα εκείνων των μεταβλητών που έχουν υψηλή πολυσυγγραμμικότητα και μια ή περισσότερες από αυτές τις μεταβλητές μέσα στη συγκεκριμένη ομάδα θα πρέπει να

απομακρυνθούν πριν προχωρήσει κανείς σε ανάλυση παλινδρόμησης για την αρχική εξαρτημένη μεταβλητή.

7.1 Αμφικλινής Παλινδρόμηση (Ridge Regression)

Μια εναλλακτική μέθοδος για την αντιμετώπιση της πολυσυγγραμικότητας είναι η μέθοδος της **αμφικλινούς παλινδρόμησης** (*ridge regression*). Η μέθοδος αυτή προτάθηκε από τον Hoerl (1962) και αναλύθηκε με λεπτομερή τρόπο από τους Hoerl & Kennard (1970).

Η διαδικασία που προτείνεται με την μέθοδο αυτή έχει ως στόχο να ξεπεράσει την δυσκολία που δημιουργείται από την ύπαρξη συσχέτισης μεταξύ μεταβλητών. Σύμφωνα με την μέθοδο αυτή προστίθεται μία σταθερά Θ στα στοιχεία της διαγωνίου του πίνακα $\mathbf{Z}'\mathbf{Z}$ όπου:

$$\mathbf{Z} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mk} \end{bmatrix}$$

($\mathbf{Z}$ είναι δηλαδή ο πίνακας που έχει ως στήλες του τις παρατηρήσεις των ανεξαρτήτων μεταβλητών $X_1, X_2, \dots, X_k$).

Οι εκτιμήτριες των συντελεστών παλινδρόμησης της μεθόδου αυτής δίνονται από τον τύπο:

$$\hat{\beta}^* = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}'\mathbf{Y}$$

Δεν είναι στους στόχους του παρόντος συγγράμματος η παραπέρα ανάπτυξη της μεθόδου αμφικλινούς παλινδρόμησης. Θα πρέπει όμως να τονίσουμε ότι οι εκτιμητές που προκύπτουν από την μέθοδο αυτή είναι πάντοτε μεροληπτικοί, αλλά οι διασπορές τους είναι αισθητά μικρότερες από αυτές των εκτιμητριών ελαχίστων τετραγώνων. Περισσότερα για την μέθοδο της αμφικλινούς παλινδρόμησης, όπως επίσης και για εφαρμογές της στους κλάδους της γεωπονίας και της χημικής μηχανικής μπορεί ο αναγνώστης να βρει στην εργασία των Marquardt & Snee (1975). Για τους λόγους επίσης που οδηγούν στην υιοθέτηση της μεθόδου ο αναγνώστης μπορεί να ανατρέξει στην εργασία του Mulled (1976).

Παρατήρηση: Η μέθοδος της αμφικλινούς παλινδρόμησης εισήγαγε, για πρώτη φορά στο βιβλίο αυτό, εκτιμήτριες στην μέθοδο παλινδρόμησης άλλες από αυτές των ελαχίστων τετραγώνων. Οι λόγοι που οι εκτιμήτριες ελαχίστων τετραγώνων χρησιμοποιούνται περισσότερο στις μεθόδους παλινδρόμησης οφείλονται στο ότι είναι απλές όσον αφορά τον υπολογισμό τους, γεωμετρικά αισθητικές και με δεδομένο ότι κάποιες υποθέσεις ισχύουν με απόλυτο τρόπο είναι βέλτιστες για μια σειρά από κριτήρια. Οι εκτιμήτριες ελαχίστων τετραγώνων έχουν χρησιμοποιηθεί με επιτυχία για περισσότερο από 160 χρόνια. Τα τελευταία χρόνια, κυρίως λόγω της πληθώρας των ηλεκτρονικών υπολογιστών διαφόρων μορφών που είναι διαθέσιμοι, έχουν αναπτυχθεί και χρησιμοποιούνται μια σειρά από άλλες εκτιμήτριες κυρίως για να καλύψουν κάποιες συγκεκριμένες αδυναμίες της μεθόδου ελαχίστων τετραγώνων. (Για παράδειγμα οι εκτιμήτριες που προαναφέραμε στην μέθοδο της αμφικλινούς παλινδρόμησης αναπτύχθηκαν για να ξεπερασθεί το πρόβλημα που δημιουργείται στις εκτιμήτριες της παλινδρόμησης από το πρόβλημα της πολυσυγγραμικότητας).