Κεφάλαιο 15

Ανάλυση της Συνδιακύμανσης

15.1 Ανάλυση της Συνδιακύμανσης

Ανάλυση της Συνδιακύμασης (ANCOVA) είναι ο όρος που χρησιμοποιείται για την ανάλυση ενός γραμμικού μοντέλου όταν κάποιες από τις ανεξάρτητες μεταβλητές είναι παράγοντες και κάποιες συνεχείς. Όπως και με την ανάλυση της διακύμανσης ενδιαφερόμαστε στη σύγκριση μέσων όρων ανάμεσα στα επίπεδα του παράγοντα, αλλά αναγνωρίζουμε το γεγονός ότι η συνεχής μεταβλητή έχει επίδραση στην εξαρτημένη. Το παρακάτω παράδειγμα δείχνει πως εφαρμόζεται η μέθοδος.

Έστω τα ότι έχουμε τα ακόλουθα δεδομένα:

- Y_{jk}: η βαθμολογία για τρεις διαφορετικές μεθόδους, A, B, C, και
- x: η ικανότητα μάθησης πριν την διδασκαλία για 7 μαθητές.

Για να εξετάσουμε κατά πόσον υπάρχουν διαφορές μεταξύ μεθόδων δοθέντος της x, θεωρούμε το πλήρες μοντέλο

$$E(Y_{jk}) = \mu_j + \gamma x_{jk}, \quad j = 1, 2, 3, \quad k = 1, \dots, 7$$

και το μοντέλο

$$E(Y_{jk}) = \mu + \gamma x_{jk}$$

Παρατηρούμε ότι $j = 1$ αντιστοιχεί στη μέθοδο A, $j = 2$ στη μέθοδο B και $j = 3$ στη μέθοδο C.

Έστω,

$$\mathbf{Y_j} = \begin{bmatrix} Y_{j1} \\ \vdots \\ Y_{j7} \end{bmatrix}$$

και

$$\mathbf{x_j} = \begin{bmatrix} x_{j1} \\ \vdots \\ x_{j7} \end{bmatrix}$$

Το πλήρες μοντέλο δίνεται από

$$E(\mathbf{Y}) = X\boldsymbol{\beta},$$

με

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \\ \mathbf{Y_3} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \gamma \end{bmatrix},$$

και

$$X = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{x_1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{x_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{x_3} \end{bmatrix}$$

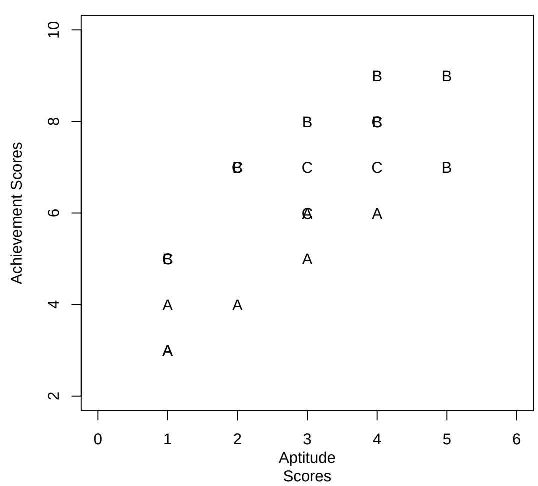
Η ίδια ανάλυση μπορεί να γίνει και για το μικρότερο μοντέλο.

Στη συνέχεια παρουσιάζεται ο τρόπος που επιτυγχάνεται η ανάλυση της συνδιακύμανσης στην R. Στην αρχή παρουσιάζεται το διάγραμμα διασπορών της ικανότητας μάθησης πριν τη διδασκαλία συναρτήσει της βαθμολογίας για τις τρεις διαφορετικές μεθόδους (Σχήμα 15.1). Από το διάγραμμα φαίνεται ότι υπάρχουν διαφορές ανάμεσα στις τρεις μεθόδους όταν ληφθεί υπόψη η ικανότητα μάθησης πριν την διδασκαλία, και αυτό είναι που θα εξεταστεί με την ανάλυση συνδιακύμασης. Αυτή εφαρμόζεται με τη εντολή που χρησιμοποιείται και για τη γραμμική παλινδρόμηση, δηλαδή την `lm()`. Για να υπολογιστεί ο πίνακας συνδιακύμανσης χρησιμοποιείται η εντολή `anova()` και όρισμα το αντικείμενο της γραμμικής παλινδρόμησης με μερικές ανεξάρτητες μεταβλητές παράγοντες και μερικές συνεχείς. Από τον πίνακα συνδιακύμανσης συμπεραίνεται ότι η ικανότητα μάθησης πρίν τη διδασκαλία (x) επηρεάζει τη βαθμολογία (Y), και ότι υπάρχουν διαφορές ανάμεσα στις τρεις όταν πάρουμε υπόψη την ικανότητα μάθησης.

```
> y <- c(6,4,5,3,4,3,6, 8,9,7,9,8,5,7, 6,7,7,7,8,5,7)
> x <- c(3,1,3,1,2,1,4, 4,5,5,4,3,1,2, 3,2,2,3,4,1,4)
```

```
> m <- gl(3,7)
> plot(x[m==1], y[m==1], pch="A", xlim=c(0,6), ylim=c(2,10),
+ xlab="Aptitude Scores", ylab="Achievement Scores")
> points(x[m==2], y[m==2], pch="B")
> points(x[m==3], y[m==3], pch="C")
> anova(z <- lm(y~x+m))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 36.575  36.575  60.355 5.428e-07 ***
m          2 16.932   8.466  13.970 0.0002579 ***
Residuals 17 10.302   0.606
---
Signif. codes:  0 ‘‘***’’ 0.001 ‘‘**’’ 0.01 ‘‘*’’ 0.05 ‘‘.’’ 0.1 ‘‘ ’’ 1
```



Σχήμα 15.1: Διάγραμμα διασπορών της ικανότητας μάθησης συναρτήσει της βαθμολογίας για τις 3 μεθόδους.