ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά – Χρηματοοικονομικά Μαθηματικά

2015

Μοντέλο τυχαίων επιδράσεων

Ενδιαφερόμαστε για παράγοντες με μεγάλο αριθμό πιθανών σταθμών. Αν επιλέξουμε τυχαία k από αυτές τις στάθμες τότε λέμε ότι ο παράγοντας είναι τυχαίος.

Το γραμμικό μοντέλο είναι:

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}, \mathrm{i} = 1,2,\ldots,\mathrm{k};\ \mathrm{j} = 1,2,\ldots,\mathrm{n}$$

όπου τα α_i και ε_{ij} είναι τυχαίες μεταβλητές.

Αν η α_i έχει διασπορά σ_a^2 και είναι ανεξάρτητη από την ε_{ij} η διασπορά οποιαδήποτε παρατήρησης είναι:

$$V(Y_{ij}) = \sigma_\alpha^2 + \sigma^2$$

Μοντέλο τυχαίων επιδράσεων

Η μαθηματική ελπίδα μιας παρατήρησης Y_{ij} είναι:

$$E\left(Y_{ij}\right) = \mu_{.}$$

όπου προκύπτει από:

$$Y_{ij} = \mu_{j} + \varepsilon_{ij} \quad \Rightarrow \quad \begin{aligned} E\left(Y_{ij}\right) &= E\left(\mu_{j}\right) + E\left(\varepsilon_{ij}\right) \\ &= \mu_{.} + 0 \\ &= \mu_{.} \end{aligned}$$

Η διακύμανση μιας παρατήρησης Y_{ij} είναι:

$$V\left(Y_{ij}\right) = \sigma^{2}\left(\mu_{j}\right) + \sigma^{2}\left(\varepsilon_{ij}\right) = \sigma_{\alpha}^{2} + \sigma^{2}$$

Μοντέλο τυχαίων επιδράσεων

- Οι Y_{ij} ακολουθούν κανονική κατανομή αφού είναι γραμμικοί συνδυασμοί των ανεξάρτητων τυχαίων μεταβλητών $μ_j$ και $ε_{ij}$.
- Σημειώνουμε ωστόσο πως οι Y_{ij} δεν είναι ανεξάρτητες διότι ένας αριθμός διαφορετικών Y_{ij} έχουν την ίδια συνιστώσα $μ_j$.

Μοντέλο τυχαίων επιδράσεων

Το μοντέλο ονομάζεται μοντέλο τυχαίων επιδράσεων.

Για να κάνουμε έλεγχο υποθέσεων στο μοντέλο αυτό θα πρέπει τα $\{\varepsilon_{ij}\}$ να είναι ανεξάρτητες τυχαίες μεταβλητές με $\varepsilon_{ij} \sim N(0, \sigma^2)$, τα $\{\alpha_i\}$ να είναι ανεξάρτητες τυχαίες μεταβλητές με $\alpha_i \sim N(0, \sigma_\alpha^2)$, και οι α_i και ε_{ij} ανεξάρτητες μεταξύ τους.

Στην περίπτωση αυτή η υπόθεση

$$\sum_{i=1}^{k} a_i = 0$$

δεν ισχύει λόγο ανεξαρτησίας των α_i

Μοντέλο τυχαίων επιδράσεων

- άμεση εκτίμηση της διασποράς των μ_j εκτιμάται από τον λόγο

$$\frac{\sigma_{\alpha}^{2}}{\sigma_{\alpha}^{2}+\sigma^{2}}$$

- Σημειώνουμε τα κύρια χαρακτηριστικά του λόγου:

1. ο λόγος παίρνει τιμές μεταξύ του 0 (όταν $\sigma^2 = \infty$) και 1 (όταν $\sigma^2 = 0$)
2. ο παρονομαστής είναι $\sigma^2(Y_{ij})$
3. ο λόγος μετρά την αναλογία της συνολικής διακύμανσης της Y_{ij} που ερμηνεύεται από τη διακύμανση του μ_j.

Μοντέλο τυχαίων επιδράσεων

- Η σχέση

$$SST = SST_{tr} + SSE$$

ισχύει

Δεν έχει έννοια στο συγκεκριμένο μοντέλο να ελέγξουμε τις υποθέσεις σχετικά με τις επιδράσεις μεμονωμένων σταθμών ($\beta_i \neq 0$).

Επομένως η υπόθεση γίνεται

$$H_0 : \sigma_a^2 = 0$$

$$H_1 : \sigma_a^2 > 0$$

Μοντέλο τυχαίων επιδράσεων

Αν $\sigma_\alpha^2 = 0$ τότε όλες οι στάθμες είναι ταυτόσημες

Αν $\sigma_\alpha^2 > 0$ τότε υπάρχει μεταβλητότητα μεταξύ των διαφορετικών σταθμών

Επομένως $\dfrac{SSE}{\sigma^2} \sim x_{n-k}^2$

και κάτω από την H_0 $\Rightarrow$ $\dfrac{SST_{tr}}{\sigma^2} \sim x_{k-1}^2$

Επειδή οι τυχαίες μεταβλητές είναι ανεξάρτητες ισχύει (κάτω από την H_0) $\sigma_\alpha^2 = 0$

Άρα

$$F^* = \frac{\dfrac{SST_{tr}}{k-1}}{\dfrac{SSE}{n-k}} = \frac{MST_{tr}}{MSE} \sim F_{k-1,n-k}$$

Μοντέλο τυχαίων επιδράσεων

Πλήρη διερεύνηση των αναμενόμενων τιμών

$$E\left(MST_{tr}\right)=\frac{E\left(SST_{tr}\right)}{k-1}=\frac{1}{k-1}E\left(\sum_{i=1}^{k}\frac{Y_{i.}^{2}}{n_{i.}}-\frac{Y_{..}^{2}}{n..}\right)=$$

$$=\frac{1}{k-1}E\left(\frac{1}{n_{i.}}\sum_{i=1}^{k}\left(\sum_{i=1}^{n_{i.}}\mu+\alpha_{i}+\varepsilon_{ij}\right)^{2}-\frac{1}{n..}\left(\sum_{i=1}^{k}\sum_{i=1}^{n_{i.}}\mu+\alpha_{i}+\varepsilon_{ij}\right)^{2}\right)$$

Παρατηρούμε ότι:

$$a_{i}=\sigma_{\alpha}^{2}\Leftarrow \mathrm{E}\left(\alpha_{\iota}\right)=0$$

$$\varepsilon_{\iota.}^{2}=n\sigma^{2}$$

$$\varepsilon_{..}^{2}=kn\sigma^{2}$$

$$\sum_{i=1}^{k}\sum_{i=1}^{n_{i.}}\alpha_{\iota}^{2}=kn^{2}\sigma_{\alpha}^{2}$$

Μοντέλο τυχαίων επιδράσεων

Επομένως

$$E\left(MST_{tr}\right)=\frac{1}{k-1}\left(n_{..}\mu^2+n_{..}\sigma^2+k\sigma^2-n_{..}\mu^2-n_{i.}\sigma_{\alpha}^2-\sigma^2\right)$$

ή

$$E\left(MST_{tr}\right)=\sigma^2+n_{i.}\sigma_{\alpha}^2$$

αφού $n_{..}=kn_{i.}$

Επίσης

$$E\left(MSE\right)=\sigma^2$$

Μοντέλο τυχαίων επιδράσεων

Παρατηρούμε ότι κάτω από την H_0 οι αναμενόμενες τιμές του αριθμητή και παρονομαστής της F^* είναι αμερόληπτοι εκτιμητές της σ^2 ενώ κάτω από την H_1 η αναμενόμενη τιμή του αριθμητή είναι μεγαλύτερη από εκείνη του παρονομαστή.

Κατά συνέπεια θα απορρίπτουμε την H_0 για τις τιμές της F^* που είναι αρκετά μεγάλες.

Επομένως απορρίπτουμε την H_0 αν $F^* > F_{k-1,n-k,\alpha}$

Μοντέλο τυχαίων επιδράσεων

Η διαδικασία υπολογισμού και ανάλυσης της διασποράς των μοντέλων τυχαίας επίδρασης είναι ταυτόσημη με εκείνη για το μοντέλο των σταθερών επιδράσεων

$$MST_{tr} = \sigma^2 + n\sigma_\alpha^2$$

$$MSE = \sigma^2$$

$$\hat{\sigma}^2 = MSE$$

$$\hat{\sigma}_\alpha^2 = \frac{MST_{tr} - MSE}{n}$$

Μοντέλο τυχαίων επιδράσεων

- Αν ισχύει το μοντέλο

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}, \mathrm{i} = 1,2,\ldots,\mathrm{k};\ \mathrm{j} = 1,2,\ldots,\mathrm{n}$$

συχνά ενδιαφερόμαστε για την εκτίμηση του ολικού μ.

Θα υποθέσουμε ότι όλα τα δειγματικά μεγέθη των επιπέδων του παράγοντα είναι ίσα, δηλαδή ότι $n_j = n$

Γνωρίζουμε ότι

$$E\left(Y_{ij}\right) = \mu_{.}$$

Οπότε ένας αμερόληπτος εκτιμητής του $\mu_{.}$ είναι ο

$$\hat{\mu}_{.} = \overline{Y}_{..}$$

Μοντέλο τυχαίων επιδράσεων

- Προκύπτει ότι η διακύμανση του εκτιμητή αυτού είναι:

$$\sigma^2\left(\overline{Y}_{..}\right)=\frac{\sigma_\alpha^2}{r}+\frac{\sigma^2}{n_T}=\frac{n\sigma_\alpha^2+\sigma^2}{n_T}$$

με $n_T = rn$

Ο τύπος δείχνει ότι η διασπορά αποτελείται από 2 συνιστώσες.

Η πρώτη αντιστοιχεί στη διασπορά ενός δειγματικού μέσου

βασιζόμενου σε r παρατηρήσεις από τη δειγματοληψία του πληθυσμού των μ_j, και απεικονίζει τη συνεισφορά της δειγματοληψίας των επιπέδων του παράγοντα.

Μοντέλο τυχαίων επιδράσεων

- Η δεύτερη συνιστώσα αντιστοιχεί στη διασπορά ενός δειγματικού μέσου βασιζόμενου σε n_T παρατηρήσεις από τη δειγματοληψία των πληθυσμών των Y_{ij}, δοθέντος του μ_j και απεικονίζει τη συνεισφορά ανάμεσα στα επίπεδα του παράγοντα.