ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά – Χρηματοοικονομικά Μαθηματικά

2015

Εκτιμηση παραμέτρων

Γενική μορφή

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + + \beta_{p-1} X_{ip-1} + \varepsilon_i$$

$$i = 1, 2, \ldots, n$$

$$\mathbf{Y} = \mathbf{Xb} + \boldsymbol{\varepsilon}$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right) \Rightarrow E\left(\varepsilon\right) = 0, V\left(\varepsilon\right) = \sigma^2 I$$

Εκτιμηση παραμέτρων

Γενική μορφή εξισώσεων

$$X'X\beta = X'Y$$

Εκτιμηση ελάχιστων τετραγώνων

$$\widehat{\beta} = \left(X'X\right)^{-1} X'Y$$

$$SST_r + SSE = SST_{tot}$$

$$\sum_j n_j\left(\bar{y}_j - \bar{y}\right)^2 + \sum_{ij}\left(y_{ij} - \bar{y}_j\right)^2 = \sum_{ij}\left(y_{ij} - \bar{y}\right)^2$$

$$\left(\widehat{Y}'\widehat{Y} - n\bar{Y}^2\right) = \widehat{\beta}'X'Y - n\bar{Y}^2$$

$$\left(\widehat{Y}'\widehat{Y} - n\bar{Y}^2\right) + \left(\widehat{Y}'\widehat{Y} - \widehat{\beta}X'Y\right) = Y'Y - n\bar{Y}^2$$

Εκτιμηση παραμέτρων

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_{p-1} x_{i,p-1},\ i = 1,2,\ldots,\nu$$

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \ldots + \hat{\beta}_{p-1} x_{1,p-1} \\ \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \ldots + \hat{\beta}_{p-1} x_{2,p-1} \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_{\nu 1} + \ldots + \hat{\beta}_{p-1} x_{\nu,p-1} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{\nu 1} & \cdots & x_{\nu,p-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{bmatrix}$$

$$\hat{\mathbf{Y}} = X\,\hat{\boldsymbol{\beta}}$$

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_{p-1} x_{i,p-1}),\ i = 1,2,\ldots,\nu$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\nu \end{bmatrix} \qquad \hat{\boldsymbol{\varepsilon}} = \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_\nu \end{bmatrix} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_\nu - \hat{Y}_\nu \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\nu \end{bmatrix} - \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_\nu \end{bmatrix}$$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}$$

Εκτιμηση παραμέτρων

$$\hat{\mathbf{Y}} = X\hat{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{Y}$$

$$\hat{\mathbf{Y}} = \boxed{X(X'X)^{-1}X'}\,\mathbf{Y} = \boxed{H}\,\mathbf{Y}$$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - H\,\mathbf{Y} = (I_{\nu} - H)\,\mathbf{Y}$$

Εκτιμηση παραμετρων

Πρόταση 4.3.5 *Στο στατιστικό μοντέλο πολλαπλής παλινδρόμησης* $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *μ* $\sigma^2(\boldsymbol{\varepsilon}) = \sigma^2 I_\nu$,

Η μέση τιμή του αθροίσματος τετραγώνων SSE δίνεται από τον τύπο

$$E(SSE) = (\nu - p)\sigma^2,$$

και επομένως η ποσότητα

$$\frac{1}{\nu - p}\sum_{i=1}^{\nu} \hat{\varepsilon}_i^2 = \frac{SSE}{\nu - p}$$

αποτελεί αμερόληπτη εκτιμήτρια της παραμέτρου σ^2.

Εκτιμηση παραμετρων

μέσο τετραγωνικό υπόλοιπο

$$\frac{SSE}{\nu - p} \Rightarrow s^2 = MSE = \frac{SSE}{\nu - p} = \frac{1}{\nu - p}\sum_{i=1}^{\nu} \hat{\varepsilon}_i^2 = \frac{1}{\nu - p}\sum_{i=1}^{\nu}(Y_i - \hat{Y}_i)^2$$

Αν χρησιμοποιήσουμε το s^2 στη θέση του σ^2 που εμφανίζεται στον τύπο $\sigma^2(\hat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}$, θα προκύψει ο πίνακας

$$s^2(\hat{\boldsymbol{\beta}}) = s^2(X'X)^{-1}$$

τα στοιχεία του οποίου δίνουν αμερόληπτες εκτιμήτριες των διακυμάνσεων και των συνδιακυμάνσεων των εκτιμητριών ελαχίστων τετραγώνων $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{p-1}$.

Παράδειγμα

Ένα αγρότης διαθέτει $\nu = 10$ όμοια αγροτεμάχια, και ενδιαφέρεται να προσδιορίσει τον τρόπο με τον οποίο επηρεάζεται η παραγωγή Y των αγρών του από την ποσότητα X_1 του λιπάσματος που χρησιμοποιεί και την ποσότητα νερού άρδευσης X_2 κατά τη διάρκεια της καλλιέργειας. Για το σκοπό αυτό κάνει διάφορους συνδυασμούς τιμών (x_1, x_2), $i = 1,2,\ldots,10$ των χαρακτηριστικών X_1, X_2 στους 10 αγρούς και καταγράφει τις αντίστοιχες παραγωγές y_i, $i = 1,2,\ldots,10$. Τα αποτελέσματα των μετρήσεων δίνονται στον παραπάνω πίνακα όπου οι μονάδες μέτρησης για τα χαρακτηριστικά X_1, X_2, Y είναι εκατοντάδες κιλά, εκατοντάδες m^3 και χιλιάδες κιλά αντίστοιχα.

i	Λίπασμα (x_{i1})	Άρδευση (x_{i2})	Παραγωγή (y_i)
1	7	9	673
2	22	10	719
3	12	8	685
4	24	11	730
5	9	5	655
6	15	6	672
7	17	14	701
8	20	13	700
9	10	16	684
10	14	8	681

Παράδειγμα

$$\hat{\beta} = (X'X)^{-1} X'Y$$

$$X = \begin{bmatrix} 1 & 7 & 9 \\ 1 & 22 & 10 \\ 1 & 12 & 8 \\ 1 & 24 & 11 \\ 1 & 9 & 5 \\ 1 & 15 & 6 \\ 1 & 17 & 14 \\ 1 & 20 & 13 \\ 1 & 10 & 16 \\ 1 & 14 & 8 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 673 \\ 719 \\ 685 \\ 730 \\ 655 \\ 672 \\ 701 \\ 700 \\ 684 \\ 681 \end{bmatrix}$$

$$A = X'X = \begin{bmatrix} 10 & 150 & 100 \\ 150 & 2544 & 1548 \\ 100 & 1548 & 1112 \end{bmatrix} \qquad X'\mathbf{y} = \begin{bmatrix} 6900 \\ 104535 \\ 69370 \end{bmatrix}$$

$$A^{-1} = \frac{1}{|A|}\begin{bmatrix} 432624 & -12000 & -22200 \\ -12000 & 1120 & -480 \\ -22200 & -480 & 2940 \end{bmatrix} = \frac{1}{306240}\begin{bmatrix} 432624 & -12000 & -22200 \\ -12000 & 1120 & -480 \\ -22200 & -480 & 2940 \end{bmatrix}$$

Παράδειγμα

Έχουμε διαμορφώσει ένα στατιστικό μοντέλο πολλαπλής παλινδρόμησης για τ
δεδομένα του διπλανού πίνακα, έτσι ώστε να προβλέπουμε την παραγωγή Y ενό
αγρού στον οποίο χρησιμοποιούνται $X_1 = x_1$ εκατοντάδες κιλά λιπάσματος κ
$X_2 = x_2$ εκατοντάδες κυβικά μέτρα νερού άρδευσης. Χρησιμοποιώντας τις $\nu = 1$
μετρήσεις $y_1, y_2, \ldots, y_{10}$ που αφορούσαν 10 αγρούς καταλήξαμε στο μοντέλο

$$\hat{y} = 622.62 + 3.2x_1 + 1.93x_2 .$$

α. Να βρεθεί η τιμή της αμερόληπτης εκτιμήτριας $s^2 = MSE$ της άγνωστη
διακύμανσης $\sigma^2 = V(\varepsilon) = V(Y)$.

β. Να βρεθεί η τιμή του μέσου αθροίσματος τετραγώνων της παλινδρόμησης MSR.

γ. Να υπολογισθούν τα τυπικά σφάλματα $s(\hat{\beta}_0), s(\hat{\beta}_1), s(\hat{\beta}_2)$ των εκτιμητριώ
ελαχίστων τετραγώνων.

Παράδειγμα

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X'X)^{-1} X'\mathbf{y} = \frac{1}{306240} \begin{bmatrix} 432624 & -12000 & -2200 \\ -12000 & 1120 & -480 \\ -2200 & -480 & 2940 \end{bmatrix} \begin{bmatrix} 6900 \\ 104535 \\ 69370 \end{bmatrix} = \begin{bmatrix} \frac{190671600}{306240} \\ \frac{981600}{306240} \\ \frac{591000}{306240} \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 622.62 \\ 3.2 \\ 1.93 \end{bmatrix}$$

Παράδειγμα

i	Λίπασμα (x_{i1})	Άρδευση (x_{i2})	Παραγωγή (y_i)
1	7	9	673
2	22	10	719
3	12	8	685
4	24	11	730
5	9	5	655
6	15	6	672
7	17	14	701
8	20	13	700
9	10	16	684
10	14	8	681

$SST_{tot}=4642$

$SSE=610.4$

$SST_r=4031.6$

Παράδειγμα

α. Αφού $\nu = 10, p = 3$, η τιμή της αμερόληπτης εκτιμήτριας s^2 του σ^2 θα είναι

$$s^2 = \frac{SSE}{\nu - p} = \frac{610.4}{7} = 87.2\,.$$

β. Η τιμή του MSR είναι ίση με

$$MSR = \frac{SSR}{p-1} = \frac{4031.6}{2} = 2015.8\,.$$

$$\sigma^2(\hat{\beta}_0) = V(\hat{\beta}_0) = \frac{432624\sigma^2}{306240} = 1.4127\sigma^2 \text{ και } \sigma(\hat{\beta}_0) = 1.189\sigma\,,$$

$$\sigma^2(\hat{\beta}_1) = V(\hat{\beta}_1) = \frac{1120\sigma^2}{306240} = 0.0036\sigma^2 \text{ και } \sigma(\hat{\beta}_1) = 0.06\sigma\,,$$

$$\sigma^2(\hat{\beta}_2) = V(\hat{\beta}_2) = \frac{2940\sigma^2}{306240} = 0.0096\sigma^2 \text{ και } \sigma(\hat{\beta}_2) = 0.098\sigma\,.$$

Παράδειγμα

Αντικαθιστώντας το σ^2 με το $s^2 = MSE = 87.2$ στις παραπάνω σχέσεις βρίσκουμε τις αμερόληπτες εκτιμήτριες

$$s^2(\hat{\beta}_0) = 1.4127 \cdot 87.2 = 123.19\,,$$

$$s^2(\hat{\beta}_1) = 0.0036 \cdot 87.2 = 0.314\,,$$

$$s^2(\hat{\beta}_2) = 0.0096 \cdot 87.2 = 0.837$$

των τριών διακυμάνσεων. Για τα αντίστοιχα τυπικά σφάλματα έχουμε,

$$s(\hat{\beta}_0) = \sqrt{123.19} = 11.1\,,$$

$$s(\hat{\beta}_1) = \sqrt{0.314} = 0.56 \text{ και}$$

$$s(\hat{\beta}_2) = \sqrt{0.837} = 0.915\,.$$