Μάθημα 4ο 30/10/17

Παρασκευή 3/11/17 12:00-15:00
Αναπλήρωση της Δευτέρας 6/11/17

Πολλαπλή Γραμμική Παλινδρόμηση (Π.Γ.Π)

$Y = \beta_0 + \beta_1 X + \varepsilon$ φτωχό

παράμετροι, σφάλματα

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \iff Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

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

εξαρτημένη μεταβλητή ή απόκριση

ανεξάρτητες μεταβλητές ή επεξηγηματικές

$$\iff \underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad (*)$$

Μορφή δεδομένων στο μοντέλο ΠΓΠ

Πειραματικές μονάδες	Y	X_1	X_2	$\cdots$	X_p
1	Y_1	X_{11}	X_{12}	$\cdots$	X_{1p}
2	Y_2	X_{21}	X_{22}	$\cdots$	X_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	Y_n	X_{n1}	X_{n2}	$\cdots$	X_{np}

$(*)$ $\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ όπου:

$$\underset{n \times 1}{\underset{\sim}{Y}} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \underset{n \times (p+1)}{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}, \quad \underset{(p+1) \times 1}{\underset{\sim}{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \underset{n \times 1}{\underset{\sim}{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

διάσταση

→ πίνακας σχεδιασμού (design matrix)

Επομένως, το μοντέλο της π.γ.π. είναι $\boxed{\underset{\sim}{Y} = X \underset{\sim}{\beta} + \varepsilon}$

$\underset{\sim}{Y}$ διάνυσμα παρατηρήσεων

$\underset{\sim}{\beta}$ διάνυσμα παραμέτρων

$\underset{\sim}{\varepsilon}$ διάνυσμα σφαλμάτων

$\underset{\sim}{X}$ πίνακας σχεδιασμού

Εκτιμητές Ελαχίστων τετραγώνων των παραμέτρων β του μοντέλου

ΕΕΤ στηρίζονται στην ελαχιστοποίηση.

$$S(\varepsilon)=\sum_{i=1}^{n}\varepsilon_i^2=\varepsilon'\varepsilon=(Y-X\beta)'(Y-X\beta)=(Y'-\beta'X')(Y-X\beta)=$$

(ανάστροφο)

$$=Y'Y-Y'X\beta-\beta'X'Y+\beta'X'X\beta$$

Αλλά

- $Y'X\beta=\beta'X'Y$ $\quad (1\times n)\cdot n\times(p+1)\cdot(p+1)\times 1 = 1\times(p+1)\cdot(p+1)\times n\cdot n\times 1$

$$\Rightarrow S(\varepsilon)=Y'Y-2\beta'X'Y+\beta'X'X\beta$$

Οι ΕΕΤ των β θα προκύψουν από ελαχιστοποίηση της $S(\varepsilon)$ ως προς β

$$\frac{\partial}{\partial\beta}(\beta'a)=a,\quad \frac{\partial}{\partial\beta}(\beta'A\beta)=2A\beta \quad \text{όπου } \frac{\partial}{\partial\beta}=\nabla\beta=\left(\frac{\partial}{\partial\beta_1},\dots,\frac{\partial}{\partial\beta_p}\right)$$

$$\frac{\partial}{\partial\beta}S(\varepsilon)=-2X'Y+2X'X\beta$$

υποψήφιο ακρότατο θα προκύψει ως εξής:

$$\frac{\partial}{\partial\beta}S(\varepsilon)=0\Rightarrow \underbrace{X'X\beta=X'Y}_{((p+1)\times n\cdot n\times(p+1)\cdot(p+1)\times 1=(p+1)\times n\cdot n\times 1)} \xrightarrow{\text{Αν } \exists \text{ ο } (X'X)^{-1}} (X'X)^{-1}\cdot X'X\cdot\beta=(X'X)^{-1}X'Y$$

γραμμικό σύστημα p+1 εξισώσεων με p+1 αγνώστους, ΚΑΝΟΝΙΚΕΣ ΕΞΙΣΩΣΕΙΣ

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

Αν ο $(X'X)$ είναι αντιστρέψιμος τότε οι ΕΕΤ είναι $\hat{\beta}=(X'X)^{-1}X'Y$

ΠΑΡΑΤΗΡΗΣΗ: Αν ο $X'X$ είναι αντιστρέψιμος, τότε το μοντέλο λέγεται **πλήρους βαθμίδας**.

- Αν γενικότερα ο $X'X$ δεν είναι αντιστρέψιμος, τότε το μοντέλο λέγεται μη-πλήρους βαθμίδας, οι ΕΕΤ δεν είναι μοναδικοί, δίνονται όμως από τη σχέση $\tilde{\beta}=(X'X)^{-}X'Y$ όπου ο $(X'X)^{-}$ είναι ένας γενικευμένος αντίστροφος του $X'X$ (δεν θα τη μελετήσουμε)

Εκτιμώμενο μοντέλο π.γ.π.

$$\hat{\underset{\sim}{Y}} = X \hat{\underset{\sim}{\beta}}$$

Υπόλοιπα: $\underset{\sim}{e} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = \underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}$

ΙΔΙΟΤΗΤΕΣ των ΕΕΤ $\hat{\underset{\sim}{\beta}}$

Υποθέσεις για τα σφάλματα

1) $E(\varepsilon_i) = 0$, $i = 1, \ldots, n \Leftrightarrow E(\underset{\sim}{\varepsilon}) = 0$
2) $Var(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$
3) $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i, j = 1, \ldots, n$, $i \neq j$
4) $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$

Μέση τιμή τυχαίου διανύσματος $\underset{\sim}{W}$

Έστω τ.δ. $\underset{\sim}{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix}$, $E(\underset{\sim}{W}) = \begin{pmatrix} E(W_1) \\ \vdots \\ E(W_n) \end{pmatrix}$

$Cov(W_i, W_j) = E(W_i W_j) - EW_i EW_j$

$$Var(\underset{\sim}{W}) \overset{ορ}{=} \underset{n \times n}{\Sigma} = \begin{pmatrix} VarW_1 & \cdots & Cov(W_i, W_j) \\ & VarW_2 & \\ Cov(W_j, W_i) & \cdots & VarW_n \end{pmatrix}$$

οπότε $\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$, $E(\underset{\sim}{\varepsilon}) = \begin{pmatrix} E(\varepsilon_1) \\ \vdots \\ E(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \underset{\sim}{0}$

Ιδιότητες $Var(\underset{\sim}{W})$

1) Ο $Var(\underset{\sim}{W})$ συμμετρικός
2) Ο $Var(\underset{\sim}{W})$ θετικά ημιορισμένος

οπότε:

Υποθέσεις για τα σφάλματα $\underset{\sim}{\varepsilon}$

1) $E(\underset{\sim}{\varepsilon}) = 0$
2) $Var(\underset{\sim}{\varepsilon}) = \sigma^2 I_n = \sigma^2 \begin{pmatrix} 1 & \cdots & 0 \\ & 1 & \\ 0 & \cdots & 1 \end{pmatrix}$
3) $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$

$$Var(\underset{\sim}{\varepsilon}) = \begin{pmatrix} Var(\varepsilon_1) & & Cov(\varepsilon_i, \varepsilon_j) \\ & Var(\varepsilon_2) & \\ & & \ddots & \\ Cov(\varepsilon_j, \varepsilon_i) & & Var(\varepsilon_n) \end{pmatrix}$$

Συνέπειες για $\underset{\sim}{Y}$: $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

1) $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$
2) $Var(\underset{\sim}{Y}) = \sigma^2 \cdot I_n$ συνέπεια του 2) από τις υποθέσεις.
3) $\underset{\sim}{Y} \sim N_n(X\underset{\sim}{\beta}, \sigma^2 I_n)$

Υπό τις υποθέσεις (1),(2) για τα σφάλματα $\underset{\sim}{\varepsilon}$ έχουμε τις εξής **ιδιότητες** για ΕΕΤ $\underset{\sim}{\hat{\beta}}$

1) $E(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{\beta}$ (Αμεροληψία $\underset{\sim}{\hat{\beta}}$)

ΑΠΟΔΕΙΞΗ: $E(\underset{\sim}{\hat{\beta}}) = E[(X'X)^{-1} \cdot X'\underset{\sim}{Y}] = (X'X)^{-1}X' E(\underset{\sim}{Y}) = (X'X)^{-1}X'\underbrace{X\underset{\sim}{\beta}}_{=E(\underset{\sim}{Y})} = I_{p+1}\underset{\sim}{\beta} = \underset{\sim}{\beta}$

2) $Var(\underset{\sim}{\hat{\beta}}) = \sigma^2 (X'X)^{-1}$

ΑΠΟΔΕΙΞΗ: Αν $\underset{\sim}{W}$ ένα τυχαίο διάνυσμα και A είναι πίνακας κατάλληλης διάστασης ώστε να ισχύουν οι πράξεις, τότε $Var(A\underset{\sim}{W}) = A\,Var(\underset{\sim}{W})A'$.

$Var(\underset{\sim}{\hat{\beta}}) = Var(\underbrace{(X'X)^{-1}X'}_{A\,(p+1)\times n}\underset{\sim}{Y}) = (X'X)^{-1}X'(Var(\underset{\sim}{Y}))[(X'X)^{-1}X']' =$ $\boxed{\begin{matrix}(AB)' = B'A' \\ (A^{-1})' = (A')^{-1}\end{matrix}}$

$= (X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1} =$

$= \sigma^2 \cancel{(X'X)^{-1}}\cancel{(X'X)}(X'X)^{-1} =$

$= \sigma^2 (X'X)^{-1}$.

Για να δω πόσο μεγάλος είναι ο πίνακας

• trace • ιδιοτιμές • ορίζουσα

3) **Διακύμανση της πρόβλεψης $\hat{Y}_0$ για τιμή $\underset{\sim}{X_0} = (1, X_{01}, X_{02}, .., X_{0p})'$ των ανεξ. μεταβλητών**

Έστω $X_0' = (1, X_{01} .. X_{0p})$ ένα γνωστό διάνυσμα για δεδομένες τιμές X_{0i} των ανεξάρτητων μεταβλητών X_i, $i = 1, ..., p$.

Η εκτιμώμενη τιμή του Y είναι: $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_{01} + \dots + \hat{\beta}_p X_{0p} = \underset{\sim}{X_0'}\underset{\sim}{\hat{\beta}}$

Η πρόβλεψη $\hat{Y}_0$ είναι καλή αν η $Var(\hat{Y}_0)$ είναι μικρή.

Θα αποδείξω ότι $Var(\hat{Y}_0) = \sigma^2 \cdot \underset{\sim}{X_0'}(X'X)^{-1}\underset{\sim}{X_0}$

$Var(\hat{Y}_0) = Var(\underset{\sim}{X_0'}\underset{\sim}{\hat{\beta}})$

Ισχύει: $Var(\underset{\sim}{a}'\underset{\sim}{W}) = \underset{\sim}{a}'Var(\underset{\sim}{W})\underset{\sim}{a} = \sum_{i=1}^{n} a_i^2 Var W_i + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} a_i a_j Cov(W_i, W_j)$

οπότε: $Var(\hat{Y}_0) = X_0' Var(\underset{\sim}{\hat{\beta}}) X_0 = \underset{\sim}{X_0'}[\sigma^2 (X'X)^{-1}]\underset{\sim}{X_0} = \sigma^2 \underset{\sim}{X_0'}(X'X)^{-1}\underset{\sim}{X_0}$

4) **Ανάλυση Διακύμανσης στο μοντέλο ΠΓΠ**

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ του μοντέλου π.γ.π.

$$SS_{tot} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum (Y_i^2 - 2\bar{Y}Y_i + \bar{Y}^2) = \sum Y_i^2 - 2\bar{Y}\sum Y_i + n\bar{Y}^2 =$$

$$= \sum Y_i^2 - 2\bar{Y} \cdot n\bar{Y} + n\bar{Y}^2$$

$$\boxed{SS_{tot} = \underset{\sim}{Y}'\underset{\sim}{Y} - n\bar{Y}^2}$$

$$SS_{res} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \underset{\sim}{e}'\underset{\sim}{e} = (\underset{\sim}{Y} - \underset{\sim}{\hat{Y}})'(\underset{\sim}{Y} - \underset{\sim}{\hat{Y}}) = (\underset{\sim}{Y} - X\underset{\sim}{\hat{\beta}})'(\underset{\sim}{Y} - X\underset{\sim}{\hat{\beta}}) =$$

$$= (\underset{\sim}{Y}' - \underset{\sim}{\hat{\beta}}'X')(\underset{\sim}{Y} - X\underset{\sim}{\hat{\beta}}) = \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'X\underset{\sim}{\hat{\beta}} - \underset{\sim}{\hat{\beta}}'X'\underset{\sim}{Y} + \underset{\sim}{\hat{\beta}}'X'X\underset{\sim}{\hat{\beta}} =$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - 2\underset{\sim}{\hat{\beta}}'X'\underset{\sim}{Y} + \underset{\sim}{\hat{\beta}}'X'X\underset{\sim}{\hat{\beta}}$$

Ανάλυση:

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

Άρα, $\boxed{SS_{res} = \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{\hat{\beta}}'X'\underset{\sim}{Y}}$

$$SS_{reg} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = \sum (\hat{Y}_i^2 - 2\bar{Y}\hat{Y}_i + \bar{Y}^2) = \sum \hat{Y}_i^2 - 2\bar{Y} \cdot \sum \hat{Y}_i + n\bar{Y}^2 =$$

$$= \underset{\sim}{\hat{Y}}'\underset{\sim}{\hat{Y}} + n\bar{Y}^2 - 2\bar{Y}\sum_{i=1}^{n} \hat{Y}_i \overset{\underset{\sim}{\hat{Y}} = X\underset{\sim}{\hat{\beta}}}{=\!=\!=} \underset{\sim}{\hat{\beta}}'X'X\underset{\sim}{\hat{\beta}} + n\bar{Y}^2 - 2\bar{Y}\sum_{i=1}^{n} \hat{Y}_i \overset{\underset{\sim}{\hat{\beta}}'X'X\underset{\sim}{\hat{\beta}} = \underset{\sim}{\hat{\beta}}'X'\underset{\sim}{Y}}{=\!=\!=\!=\!=}$$

$$= \underset{\sim}{\hat{\beta}}'X'\underset{\sim}{Y} + n\bar{Y}^2 - 2\underset{\sim}{\bar{Y}}\sum_{i=1}^{n} \hat{Y}_i \quad (1)$$

Ανάλυση:

$$2\bar{Y}\sum_{i=1}^{n} \hat{Y}_i$$

Το $\underset{\sim}{\hat{\beta}}$ ικανοποιεί τις κανονικές εξισώσεις $X'X\underset{\sim}{\beta} = X'\underset{\sim}{Y}$

$$(X'X)\underset{\sim}{\hat{\beta}} = X'\underset{\sim}{Y} \Rightarrow X'\underset{\sim}{Y} - X'X\underset{\sim}{\hat{\beta}} = \underset{\sim}{0} \Rightarrow X'(\underset{\sim}{Y} - X\underset{\sim}{\hat{\beta}}) = \underset{\sim}{0} \Rightarrow [X'(\underset{\sim}{Y} - X\underset{\sim}{\hat{\beta}})]' = \underset{\sim}{0} \Rightarrow$$

$$\Rightarrow (\underset{\sim}{Y} - X\underset{\sim}{\hat{\beta}})'X = \underset{\sim}{0} \Rightarrow (\underset{\sim}{Y} - \underset{\sim}{\hat{Y}})'X = \underset{\sim}{0} \Rightarrow (Y_1 - \hat{Y}_1, Y_2 - \hat{Y}_2, \ldots, Y_n - \hat{Y}_n) \cdot \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix} = \underset{\sim}{0}$$

$$\Rightarrow \sum_{i=1}^{n}(Y_i-\hat{Y}_i)=0 \Rightarrow \sum Y_i - \sum \hat{Y}_i = 0 \Rightarrow \sum Y_i = \sum \hat{Y}_i \Rightarrow \sum \hat{Y}_i = n\bar{Y}. \quad (2)$$

Από τις (1), (2): $SS_{reg} = \hat{\underset{\sim}{\beta}}' X' \underset{\sim}{Y} + n\bar{Y}^2 - 2\bar{Y} \cdot n\bar{Y} \Rightarrow$

$$\boxed{SS_{reg} = \hat{\underset{\sim}{\beta}}' X' \underset{\sim}{Y} - n\bar{Y}^2}$$

Ισχύει: $SS_{tot} = SS_{reg} + SS_{res}$

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ

Πηγή μεταβλητότητας	SS	β.ε	MS	F-πηλίκο
Μοντέλο π.γ.π.	SS_{reg}	p	$MS_{reg} = \frac{SS_{reg}}{p}$	
Υπόλοιπα	SS_{res}	$n-p-1$	$MS_{res} = \frac{SS_{res}}{n-p-1}$	
Ολική	SS_{tot}	$n-1$		

Εκτιμητής της σ^2

$E(MS_{res}) = \sigma^2$

Συντελεστής προσδιορισμού ή προσαρμοστικότητας

$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}}$ $<$ R^2 καθαρός αριθμός, $0 \leq R^2 \leq 1$

$R^2\text{-adjusted} = 1 - \frac{MS_{res}}{MS_{tot}}$