29 FEB 72 – 6

LINEAR MODEL

$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_N X_N + \varepsilon$

$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \beta_3 X_{13} + \dots + \beta_N X_{1N} + \varepsilon_1$
$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \beta_3 X_{23} + \dots + \beta_N X_{2N} + \varepsilon_2$
$\vdots$
$Y_M = \beta_0 + \beta_1 X_{M1} + \beta_2 X_{M2} + \beta_3 X_{M3} + \dots + \beta_N X_{MN} + \varepsilon_M$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_M \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1N} \\ X_{21} & \dots & & & \\ \vdots & & & & \\ X_{M1} & \dots & & & X_{MN} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{pmatrix}$$

$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$

WISH TO FIND $\hat{\underline{\beta}}$ – LEAST SQUARES; MINIMIZE LENGTH OF $\hat{\underline{\varepsilon}}$, $\hat{\underline{\varepsilon}} = \underline{Y} - X\hat{\underline{\beta}}$

TAKE COLUMNS OF X AS VECTORS, $X\hat{\underline{\beta}}$ IS LINEAR COMBINATION OF COL. VECTORS OF X

$\forall$ LIN. COM. OF COL. OF X GENERATE SUBSPACE WITH COL. OF X AS BASIS

$\hat{\underline{\varepsilon}}$ IS DIFF. VECTOR BETWEEN $\hat{\underline{Y}}$ AND $\underline{Y}$, $\hat{\underline{Y}}$ MUST LIE IN SUBSPACE OF $X_{\hat{\beta}}$, WISH TO PICK $\hat{\underline{Y}}$ i.e. $\hat{\underline{\beta}}$

$\ni$ $\hat{\underline{\varepsilon}} = \underline{Y} - \hat{\underline{Y}} = \underline{Y} - X\hat{\underline{\beta}}$ IS OF MINIMAL LENGTH, SHORTEST DISTANCE FROM POINT TO 'PLANE' IS PERPENDICULAR, VECTOR $\perp$ TO PLANE IS $\perp$ TO @ VECTOR IN THE PLANE, $\therefore$ $\perp$ TO COL. OF X

$\therefore$ INNER PRODUCTS OF $\hat{\underline{\varepsilon}}$ AND COL. OF X ARE 0

$\therefore$ $X'\hat{\underline{\varepsilon}} = \underline{0}$

HOWEVER: $\underline{0} = X'\hat{\underline{\varepsilon}} = X'(\underline{Y} - \hat{\underline{Y}}) = X'(\underline{Y} - X\hat{\underline{\beta}}) = X'Y - X'X\hat{\beta}$

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

$$E(\hat{\underline{\beta}}) = E((X'X)^{-1}X'\underline{Y}) = (X'X)^{-1}X'E(\underline{Y}) = (X'X)^{-1}X'E(X\beta + \varepsilon) = (X'X)^{-1}X'(X\beta + E(\varepsilon)) = (X'X)^{-1}X'(X\beta + 0)$$
$$= (X'X)^{-1}X'X\beta = \beta$$

$V(\hat{\underline{\beta}}) = \dots = (X'X)^{-1}X'\,V(\underline{\varepsilon})\,X(X'X)^{-1}$ $\qquad V(\underline{\varepsilon}) = \sigma^2 I \Rightarrow = \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} = \sigma^2 (X'X)^{-1} = \sigma^2 (c^{ij})$

$V(\varepsilon) = \sigma^2 I$, $\quad$ [GLS, 2SLS, 3SLS]

$\hat{\beta}_i \sim N(\beta_i, \sigma^2 c^{ii})$ $\quad\therefore\quad$ $\dfrac{\hat{\beta}_i - \beta_{io}}{\sqrt{s^2 c^{ii}}} \sim t_{(M-(N+1))}$ $\quad$ TEST $H_0: \beta_i = 0$ $\;H_1: \beta_i \neq \beta_{io}$

WHERE $s^2 = \dfrac{\sum_{i=1}^{M}(Y_i - \hat{Y}_i)^2}{M-(N+1)} = \dfrac{\sum_{i=1}^{N}\hat{\varepsilon}_i^2}{M-(N+1)} = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{M-(N+1)} = \dfrac{Y'Y - \hat{Y}'Y}{M-(N+1)}$

$Cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 c^{ij}$

$H_0: \forall \beta_i\text{'s} = 0 \Rightarrow \dfrac{\hat{Y}'\hat{Y}/N+1}{\hat{\varepsilon}'\hat{\varepsilon}/M-(N+1)} \sim F_{(N+1,\, M-(N+1))}$

FIXED EFFECTS MODEL, RANDOM EFFECTS MODEL, 'DUMMY VARIABLES' (NON REAL NUMBER VAR.S), MORE COMPLICATED HYPOTHESIS STRUCTURES AND CORRESPONDING F TESTS – EXPERIMENTAL DESIGN

SAMPLE VARIANCE, VARIANCE OF STATISTIC, VARIANCE OF SAMPLE OF STATISTIC, WHAT IS THE DISTRIBUTION? –

TUKEY'S JACKKNIFE – t DISTRIBUTION $\qquad$ GENERAL STRATEGY: NOT TO ASSUME ANY PARTICULAR UNDERLYING DISTRIBUTIONS

EXPLORATORY DATA ANALYSIS (TUKEY) – WHAT ARE THE INDICATORS, WHAT DO THEY INDICATE, IN PRELUDE TO RATHER THAN HOW GOOD OF AN INDICATOR IS IT