Multiple Linear Regression Model – Part II

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 3 – Lecture 3C

Fitted Values and Residuals:

The fitted values associated with $\boldsymbol{y}$:

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

H matrix is very important. **H**: "hat" matrix

The residual is

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

Sum of Squares and Mean Squares:

$$SS_{\text{Res}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e}$$

$$SS_{\text{Res}} = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

$$= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$SS_{\text{Res}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

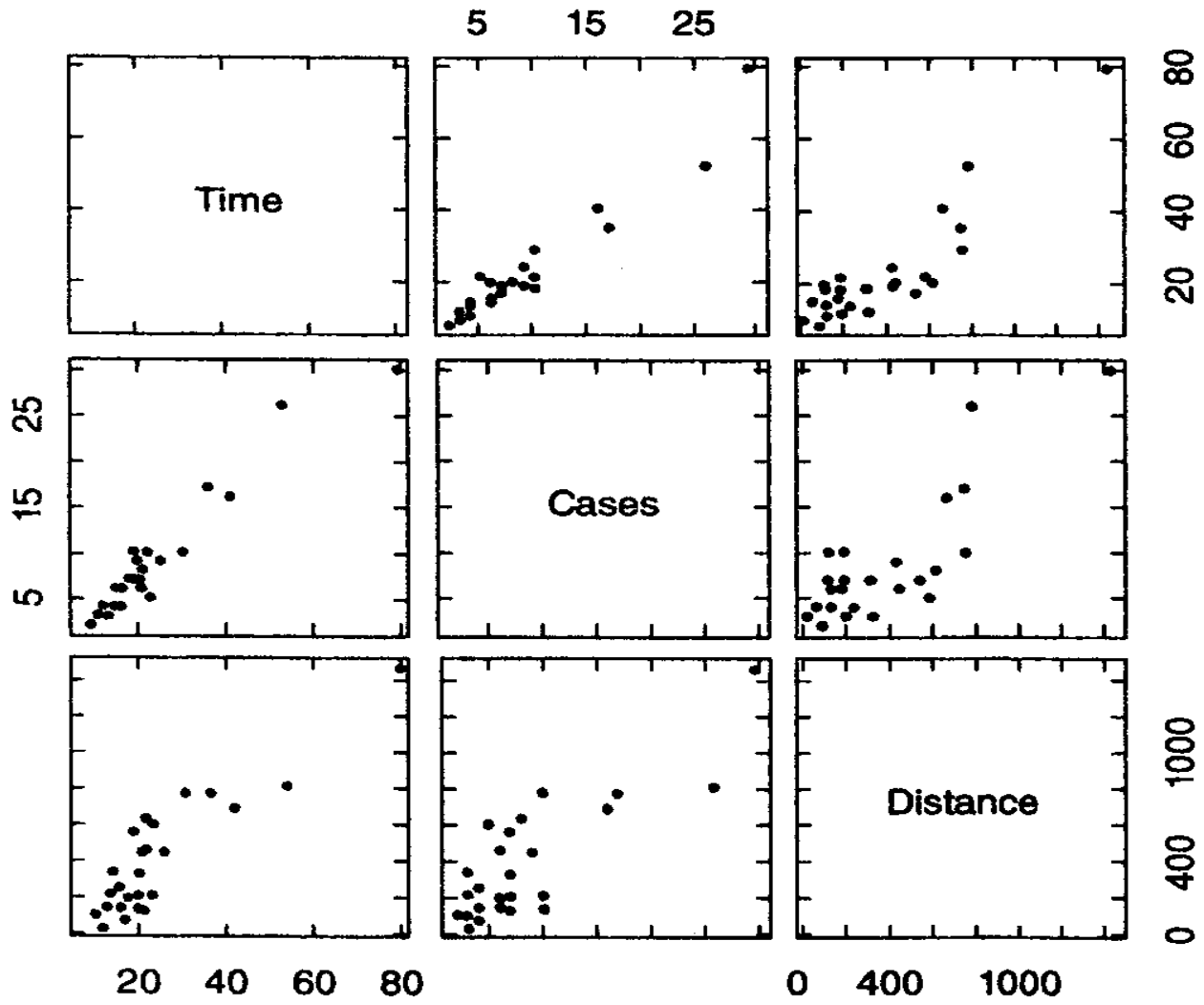
$$MS_{\text{Res}} = \frac{SS_{\text{Res}}}{n-p} \qquad \hat{\sigma}^2 = MS_{\text{Res}} \qquad \boldsymbol{p} = k+1$$

Ex 3.1 The Delivery Time Data (page 74 of Textbook)

Regress y (deliver time) on

x_1 (# of cases of product stocked)
x_2 (distance walked by a route driver)

Scatter Diagrams (page 76 of Textbook)



Least-Squares Estimation

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 150 \end{bmatrix} = \begin{bmatrix} 25 & 219 & 10,232 \\ 219 & 3,055 & 133,899 \\ 10,232 & 133,899 & 6,725,688 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 10.75 \end{bmatrix} = \begin{bmatrix} 559.60 \\ 7,375.44 \\ 337,072.00 \end{bmatrix}$$

Least-Squares Estimation

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

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \begin{bmatrix} 25 & 219 & 10,232\\ 219 & 3,055 & 133,899\\ 10,232 & 133,899 & 6,725,688\end{bmatrix}^{-1}\begin{bmatrix} 559.60\\ 7,375.44\\ 337,072.00\end{bmatrix}$$

$$= \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367\\ -0.00444859 & 0.00274378 & -0.00004786\\ -0.00008367 & -0.00004786 & 0.00000123\end{bmatrix}\begin{bmatrix} 559.60\\ 7,375.44\\ 337,072.00\end{bmatrix}$$

$$= \begin{bmatrix} 2.34123115\\ 1.61590712\\ 0.01438483\end{bmatrix}$$

$$\hat{y} = 2.34123 + 1.61591x_1 + 0.01438x_2 \quad \Rightarrow \text{Table 3.3}$$

Estimation of σ^2

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^{25} y_i^2 = 18{,}310.6290$$

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = [2.34123115 \quad 1.61590721 \quad 0.01438483]\begin{bmatrix} 559.60 \\ 7{,}375.44 \\ 337{,}072.00 \end{bmatrix}$$

$$= 18{,}076.90304$$

$$SS_{\text{Res}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

$$= 18{,}310.6290 - 18{,}076.9030 = 233.7260$$

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-p} = \frac{233.7260}{25-3} = 10.6239$$

The Delivery Time Data – Regression Analysis Table

TABLE 3.4 Minitab Output for Soft Drink Time Data

Regression Analysis: Time versus Cases, Distance

The regression equation is
Time = 2.34 + 1.62 cases + 0.0144 Distance

Predictor	Coef	SE Coef	T	P
Constant	2.341	1.097	2.13	0.044
Cases	1.6159	0.1707	9.46	0.000
Distance	0.014385	0.003613	3.98	0.001

S = 3.25947 R - Sq = 96.0% R - Sq (adj) = 95.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5550.8	2775.4	261.24	0.000
Residual Error	22	233.7	10.6		
Total	24	5784.5			

Source	DF	Seq SS
Cases	1	5382.4
Distance	1	168.4

Geometrical Interpretation of Least-Squares Fitting

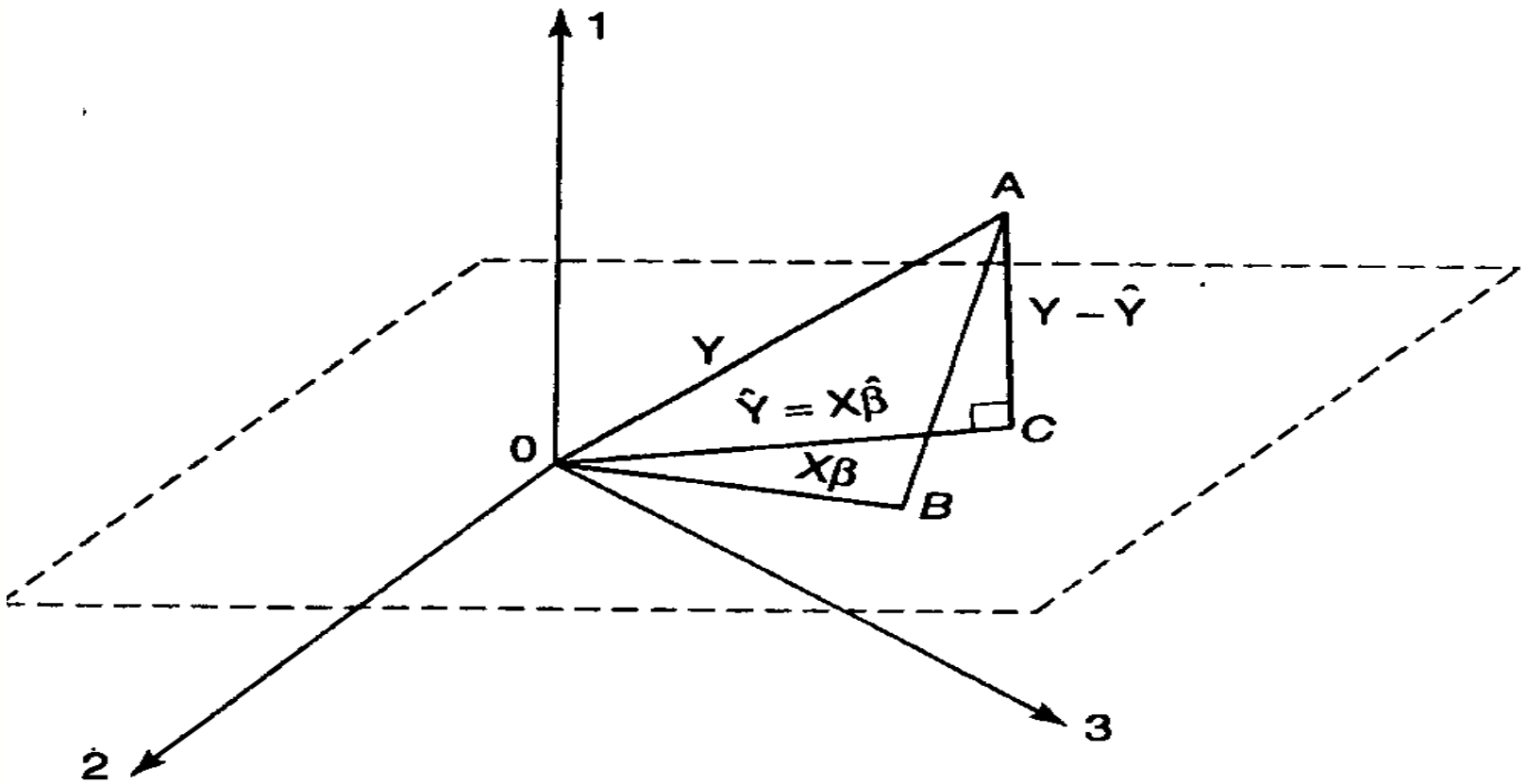


Figure 3.6 A geometrical interpretation of least squares.

JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING