第 15 章　數學附錄

一　利用 demeaned form 求 $\hat{\beta},\hat{\gamma}$

設複迴歸模型為

$$Y=\alpha+\beta X+\gamma Z+\varepsilon \qquad (15A.1)$$

$$\overline{Y}=\alpha+\beta\overline{X}+\gamma\overline{Z}+\overline{\varepsilon} \qquad (15A.2)$$

由式(15A.1)，(15A.2)可得：

$$Y-\overline{Y}=\beta(X-\overline{X})+\gamma(Z-\overline{Z})+(\varepsilon-\overline{\varepsilon})$$

$$y=\beta x+\gamma z+\varepsilon' \qquad (15A.3)$$

式(15A.3)稱為 Y 的 demeaned form 的迴歸模型

現欲利用 OLS 解式(15A.3)中的 β 與 γ 。令估計的迴歸方程式為：

$$\hat{y}=\hat{\beta}x+\hat{\gamma}z$$

求 $SSE=\sum_{i=1}^{n}(y-\hat{\beta}x-\hat{\gamma}z)^2$ 最小

$$\frac{\partial SSE}{\partial\hat{\beta}}=0 \quad \sum xy=\hat{\beta}\sum x^2+\hat{\gamma}\sum xz \qquad (15A.4)$$

$$\frac{\partial SSE}{\partial\hat{\gamma}}=0 \quad \sum zy=\hat{\beta}\sum xz+\hat{\gamma}\sum z^2 \qquad (15A.5)$$

根據式(15A.4)，(15A.5)解 $\hat{\beta},\hat{\gamma}$ 可得：

$$\hat{\beta}=\frac{\sum xy\sum z^2-\sum xz\sum zy}{\sum x^2\sum z^2-(\sum xz)^2}, \quad \hat{\gamma}=\frac{\sum xy\sum xz-\sum x^2\sum zy}{\sum x^2\sum z^2-(\sum xz)^2}$$

二　$\hat{\alpha},\hat{\beta},\hat{\gamma}$ 的分配

$$V(\hat{\beta})=\frac{\sum z^2}{\sum x^2\sum z^2-(\sum xz)^2}\sigma^2, \quad V(\hat{\gamma})=\frac{\sum x^2}{\sum x^2\sum z^2-(\sum xz)^2}\sigma^2$$

將式(16A.3)代入 $\hat{\beta}$ 得

$$\hat{\beta}=\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}[\sum z^2\sum x(\beta x+\gamma z+\varepsilon')-\sum xz\sum z(\beta x+\gamma z+\varepsilon')]$$
$$=\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}[\beta\sum x^2\sum z^2+\gamma\sum xz\sum z^2+\sum z^2\sum x\varepsilon'-\beta(\sum xz)^2]$$
$$-\gamma\sum xz\sum z^2-\sum xz\sum z\varepsilon'$$
$$=\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}[\beta[\sum x^2\sum z^2-(\sum xz)^2]+\sum z^2\sum x\varepsilon'-\sum xz\sum z\varepsilon']$$
$$=\beta+\frac{\sum z^2\sum x\varepsilon-\sum xz\sum z\varepsilon}{\sum x^2\sum z^2-(\sum xz)^2} \qquad (16A.6)$$

由(15A.6)式可知 $\hat{\beta}$ 為 ε 的線性組合， ε 為常態分配，則 $\hat{\beta}$ 亦為常態分配。

$\hat{\beta}$ 之不偏性

$$E(\hat{\beta})=\beta+E[\frac{\sum z^2\sum x\varepsilon-\sum xz\sum z\varepsilon}{\sum x^2\sum z^2-(\sum xz)^2}]$$
$$=\beta+\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}E[\sum z^2\sum x\varepsilon-\sum xz\sum z\varepsilon]$$
$$=\beta+\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}(0+0)=\beta$$

$$V(\hat{\beta})=[E(\hat{\beta})-E(\beta)]^2=E(\hat{\beta}-\beta)^2=E[\frac{\sum z^2\sum x\varepsilon-\sum xz\sum z\varepsilon}{\sum x^2\sum z^2-(\sum xz)^2}]^2$$
$$=\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}\cdot$$
$$E[(\sum z^2)^2(\sum x\varepsilon)^2+(\sum xz)^2(\sum z\varepsilon)^2-2\sum z^2\sum xz\sum x\varepsilon\sum z\varepsilon]$$
$$=\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}\cdot[\sigma^2\sum x^2(\sum z^2)^2+\sigma^2(\sum xz)^2\sum z^2$$
$$-2\sigma^2\sum z^2(\sum xz)^2]$$
$$=\frac{1}{\sum x^2\sum z^2-(\sum xz)^2}\cdot\sigma^2(\sum z^2)[\sum x^2\sum z^2-(\sum xz)^2]$$
$$=\frac{\sum z^2\sigma^2}{\sum x^2\sum z^2-(\sum xz)^2}$$

上式中：

$$E[(\sum z^2)^2(\sum x\varepsilon)^2]=(\sum z^2)^2E[(\sum x\varepsilon)^2]$$
$$=(\sum z^2)^2E[(x_1^2\varepsilon_1^2+\Lambda+x_n^2\varepsilon_n^2)]=(\sum z^2)^2\sigma^2\sum x^2$$

同理可求：

$\hat{\alpha}\sim N(\alpha,\sigma_{\hat{\alpha}}^2)$， $\hat{\gamma}\sim N(\gamma,\sigma_{\hat{\gamma}}^2)$

$$\sigma_{\hat{\gamma}}^2 = \sigma^2[\frac{\sum x^2}{\sum x^2 \sum z^2 - (\sum xz)^2}]$$

三　$\sum(\hat{Y} - \overline{Y})^2 = \hat{\beta}\sum xy + \hat{\gamma}\sum zy$ 公式之導出

$$\sum(\hat{Y} - \overline{Y})^2 = \sum \hat{y}^2 = \sum \hat{y}(y - e) = \sum \hat{y}y - \sum \hat{y}e$$
$$= \sum \hat{y}y = \sum(\hat{\beta}x + \gamma z) \cdot y = \hat{\beta}\sum xy + \hat{\gamma}\sum zy$$

(因 $\sum \hat{y}e = \sum(\hat{\beta}x + \hat{\gamma}z)e = \hat{\beta}\sum xe + \hat{\gamma}\sum ze = 0$)

其中　$\hat{y} = \hat{\beta}x + \hat{\gamma}z$，說明如下：

因　$\hat{Y} = \hat{\alpha} + \hat{\beta}X + \hat{\gamma}Z$，$\overline{Y} = \hat{\alpha} + \hat{\beta}\overline{X} + \hat{\gamma}\overline{Z}$

$\hat{Y} - \overline{Y} = \hat{y} = \hat{\beta}(X - \overline{X}) + \hat{\gamma}(Z - \overline{Z}) = \hat{\beta}x + \hat{\gamma}z$

$\sum xe = 0$，$\sum ze = 0$ 說明如下：

因 $y = \hat{\beta}x + \hat{\gamma}z + e$，$\sum xy = \hat{\beta}\sum x^2 + \hat{\gamma}\sum xz + \sum xe$

與(15A.4)比較 $\sum xe = 0$，同理可證 $\sum ze = 0$

四　$S_{Y|XZ}^2$ 計算公式

① $S_{Y|XZ}^2 = \frac{1}{n-k-1}\sum(Y - \hat{Y})^2 = \frac{1}{n-k-1}[\sum Y^2 - \sum \hat{Y}^2]$

$$= \frac{1}{n-k-1}[\sum Y^2 - \sum \hat{Y}(Y - e)] = \frac{1}{n-k-1}[\sum Y^2 - \sum Y\hat{Y} - \sum \hat{Y}e]$$
$$= \frac{1}{n-k-1}[\sum Y^2 - \sum Y(\hat{\alpha} + \hat{\beta} + \hat{\gamma}Z)] \qquad \Theta\ \hat{Y}e = 0$$
$$= \frac{1}{n-k-1}[\sum Y^2 - \hat{\alpha}\sum Y - \hat{\beta}\sum XY - \hat{\gamma}\sum ZY]$$

② $S_{Y|XZ}^2 = \frac{1}{n-k-1}\sum(y - \hat{y})^2 = \frac{1}{n-k-1}[\sum y^2 - \sum \hat{y}^2]$

$$= \frac{1}{n-k-1}[\sum y^2 - \sum \hat{y}(y - e)] = \frac{1}{n-k-1}[\sum y^2 - \sum y\hat{y} - \sum \hat{y}e]$$
$$= \frac{1}{n-k-1}[\sum y^2 - \sum y(\hat{\beta}x + \hat{\gamma}Z)] \qquad \Theta\ \hat{y}e = 0$$
$$= \frac{1}{n-k-1}[\sum y^2 - \hat{\beta}\sum xy - \hat{\gamma}\sum zy]$$

五　類別資料之迴歸分析(變異數分析與迴歸分析之比較)

設 X 為兩種類別的變數，會影響依變數 Y，兩種類別下的 Y 資料如下：

表 15a1　A、B 兩類資料

類別	Y
A	$Y_{11},...,Y_{1n_1}$
B	$Y_{21},...,Y_{2n_2}$

利用上述資料可進行變異數分析，並以$\dfrac{MSF}{MSE}$比來檢定：

$$H_0:\mu_1=\mu_2$$
$$H_1:\mu_1\neq\mu_2$$

亦可利用迴歸分析檢定 X 是否會影響 Y，在迴歸分析中設迴歸模型為：

$$Y=\alpha+\beta X+\varepsilon$$

$X=1$ ： A 類別， $X=0$ ：其他(B 類別)。

利用 OLS 去估計迴歸模型，可證明：

① $\hat{\alpha}=\overline{Y}_2$， $\hat{\beta}=\overline{Y}_1-\overline{Y}_2$

②檢定 $\begin{matrix}H_0:\beta=0\\ H_1:\beta\neq 0\end{matrix}$，即檢定 $\begin{matrix}H_0:\mu_1=\mu_2\\ H_1:\mu_1\neq\mu_2\end{matrix}$

證明

①表 15a1 的資料可寫為 X 與 Y 的表如下：

表 15a2　X 與 Y 表

	Y	X
$\overline{Y}_1$	Y_{11}	1
	$\vdots$	$\vdots$
	Y_{1n_1}	1
$\overline{Y}_2$	Y_{21}	0
	$\vdots$	$\vdots$
	Y_{2n_2}	0

於是根據表 15a2 的資料及利用 OLS 方法可估計 $\hat{\beta},\hat{\alpha}$ 如下：

$$\hat{\beta}=\frac{\sum xy}{\sum x^2}=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}=\frac{\sum XY-n\overline{X}\overline{Y}}{\sum X^2-n\overline{X}^2}\quad(n=n_1+n_2)$$

$$=\frac{n_1\overline{Y}_1-(n_1+n_2)\dfrac{n_1}{n_1+n_2}\dfrac{n_1\overline{Y}_1+n_2\overline{Y}_2}{n_1+n_2}}{n_1-(n_1+n_2)(\dfrac{n_1}{n_1+n_2})^2}=\frac{n_1\overline{Y}_1-n_1\dfrac{n_1\overline{Y}_1+n_2\overline{Y}_2}{n_1+n_2}}{n_1-\dfrac{n_1^2}{n_1+n_2}}$$

$$=\overline{Y}_1-\overline{Y}_2$$

因 $\sum X^2=n_1$，$\sum XY=\sum_{i=1}^{n_1}Y_{1i}=n_1\overline{Y}_1$，故

$$\overline{X}=\frac{\sum X}{n_1+n_2}=\frac{n_1}{n_1+n_2}，\quad \overline{Y}=\frac{\sum Y_1+\sum Y_2}{n_1+n_2}=\frac{n_1\overline{Y}_1+n_2\overline{Y}_2}{n_1+n_2}$$

又 $\hat{\alpha}=\overline{Y}-\hat{\beta}\overline{X}=\dfrac{n_1\overline{Y}_1+n_2\overline{Y}_2}{n_1+n_2}-(\overline{Y}_1-\overline{Y}_2)(\dfrac{n_1}{n_1+n_2})=\overline{Y}_2$

②証明 $\begin{matrix}H_0:\beta=0\\H_1:\beta\neq 0\end{matrix}$，即是檢定 $\begin{matrix}H_0:\mu_1=\mu_2\\H_1:\mu_1\neq\mu_2\end{matrix}$

利用 t 統計量檢定 β 如下：

$$t=\frac{\hat{\beta}}{S_{\hat{\beta}}}=\frac{(\overline{Y}_1-\overline{Y}_2)-0}{S_{\hat{\beta}}}=\frac{\overline{Y}_1-\overline{Y}_2}{\sqrt{\dfrac{S^2_{Y|X}}{\sum x^2}}}=\frac{\overline{Y}_1-\overline{Y}_2}{\sqrt{S_P^2(\dfrac{1}{n_1}+\dfrac{1}{n_2})}}$$

因 $\hat{\beta}=\overline{Y}_1-\overline{Y}_2$，$S^2_{\hat{\beta}}=\dfrac{1}{\sum x^2}S^2_{Y|X}$ 又

$$S^2_{Y|X}=\frac{1}{n-2}\sum(Y-\hat{Y})^2=\frac{1}{n_1+n_2-2}[\sum Y^2-\alpha\sum Y-\hat{\beta}\sum XY]$$

$$=\frac{1}{n_1+n_2-2}\left[(\sum Y_{1i}^2+\sum Y_{2i}^2-\overline{Y}_2(n_1\overline{Y}_1+n_2\overline{Y}_2)-(\overline{Y}_1-\overline{Y}_2)(n_1\overline{Y}_1)\right]$$

$$=\frac{1}{n_1+n_2-2}[(n_1-1)S_1^2+(n_2-1)S_2^2]=S_p^2$$

且 $\sum x^2=\sum X^2-n\overline{X}^2=\dfrac{n_1n_2}{n_1+n_2}$，$\dfrac{1}{\sum x^2}=\dfrac{n_1+n_2}{n_1n_2}=\dfrac{1}{n_1}+\dfrac{1}{n_2}$

利用變異數分析，F 統計量為：

$$F=\frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\overline{Y}_i-\overline{Y})^2\Big/1}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)^2\Big/n_1+n_2-2}=\frac{(\overline{Y}_1-\overline{Y}_2)^2\dfrac{n_1n_2}{n_1+n_2}}{S_P^2}=\frac{(\overline{Y}_1-\overline{Y}_2)^2}{S_P^2(\dfrac{1}{n_1}+\dfrac{1}{n_2})}=t^2$$ 因

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\overline{Y}_i-\overline{Y})^2=(\overline{Y}_1-\overline{Y}_2)^2(\frac{n_1n_2}{n_1+n_2})$$

因此得證：利用變異數分析的 $F=\dfrac{MSF}{MSE}$ 檢定結果與檢定兩母體的平均數是否相同的檢定完全一致。

六 若 X 分成三個類別，利用二個虛擬變數區分三個類別如下：

$X_1=1$ ： A 類， $X_1=0$ ：其他 ； $X_2=1$ ： B 類， $X_2=0$ ：其他。

迴歸模型建立為：

$$Y=\alpha+\beta X_1+\gamma X_2+\varepsilon，且 E(Y\mid X)=\alpha+\beta X_1+\gamma X_2$$

當 $X_1=1$, $E(Y\mid X)=\alpha+\beta$ A 類 Y 的平均數

當 $X_2=1$, $E(Y\mid X)=\alpha+\gamma$ B 類 Y 的平均數

當 $X_1=0$, $X_2=0$, $E(Y\mid X)=\alpha$ C 類 Y 的平均數

其資料可表為表 15a3，利用 OLS 估計 $\hat{\beta},\hat{\gamma},\hat{\alpha}$ ，可得：

$$\hat{\beta}=\overline{Y}_1-\overline{Y}_3，\ \hat{\gamma}=\overline{Y}_2-\overline{Y}_3，\ \hat{\alpha}=\overline{Y}_3$$

又利用變異數分析方法檢定 $H_0:\mu_1=\mu_2=\mu_3$ 與利用迴歸分析方法檢定 $H_0:\beta=0,\ \gamma=0$ 完全一致。

表 15a3 三類資料

	Y	X_1	X_2
$\overline{Y}_1$	Y_{11}	1	0
	M	M	M
	Y_{1n_1}	1	0
$\overline{Y}_2$	Y_{21}	0	1
	M	M	M
	Y_{2n_2}	0	1
$\overline{Y}_3$	Y_{31}	0	0
	M	M	M
	Y_{3n_3}	0	0