$$=\frac{SS_{XY}}{\sqrt{(n-1)S_X^2}\sqrt{(n-1)S_Y^2}}$$

其中 $S_X^2=\dfrac{\sum(X_i-\bar{X})^2}{n-1}$ ，$S_Y^2=\dfrac{\sum(Y_i-\bar{Y})^2}{n-1}$

1. 迴歸分析比相關係數(ρ_{XY})之應用層面更為廣泛，因為相關係數只討論兩個變數間的線性關係，但迴歸分析可討論多個自變數與多個應變數間的關係，且可討論非線性下之關係，更可進行預測的工作。
2. 由於 $\rho_{XY}=\rho_{YX}$（$r_{XY}=r_{YX}$），因此相關係數是無法衡量變數間之因果關係。而迴歸分析中設立模式時，其實已經隱含了變數間因果關係之先驗假設了（但是迴歸分析不能驗證因果關係）。

   $\mathrm{Y}=\beta_0+\beta_1\mathrm{X}+\varepsilon$ 與 $\mathrm{X}=\alpha_0+\alpha_1\mathrm{Y}+\delta$
3. 迴歸分析可應用的空間單位遠較於相關係數為廣。

   e.g. 薪資←→菸酒量

   ↑

   （經濟成長、社會消費型態、政府無菸環境政策）
4. 在相關係數中，X、Y都必須要為非固定變數。但在迴歸分析中，只須Y為非固定變數即可，X可為非固定變數或固定變數。

第二節　迴歸模型(Regression model)

<u>數學模型</u>

$$Y=\beta_0+\beta_1X$$

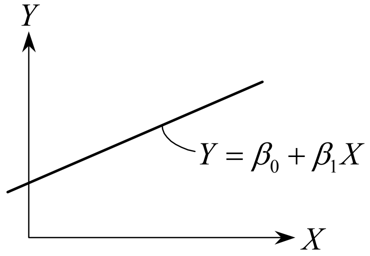

圖1

<u>迴歸模型</u>

$$Y=\beta_0+\beta_1X+\varepsilon \text{ ，} \varepsilon \overset{iid}{\sim} N(0,\sigma^2)$$

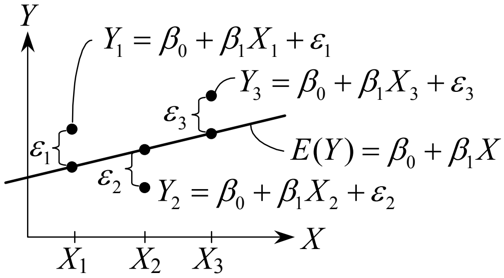

圖2

＜註＞

1. 為什麼模型中要考慮ε誤差項(error term)？
在實務問題中，除了研究者根據文獻探討及個人認定了某些自變數外，可能還有部分的影響變數是沒有被考慮的，而這些變數所產生之影響通常是不大，或它們根本無法控制，或它們會缺乏規則性，以上種種的影響變數都會歸納到ε誤差項中。
e.g. 氣候，雨量，機場移民局電腦大當機…等等。
2. 為什麼假設ε誤差項為隨機變數？
原因是假如以上可能會產生之誤差對於應變數存在一個系統性之影響，即這自變數應該是一個重要之影響因素了，若重要之因素沒有被研究者考慮在計量模型中的話，模型的解釋能力就不足。因此，ε誤差項為隨機變數在實務上也是合理的。
3. 迴歸模型$Y=\beta_0+\beta_1 X+\varepsilon$中那一些是隨機變數?應變數Y，誤差項$\varepsilon$。
為什麼假設自變數X是常數（有的問題裡X是隨機變數）？
由於在迴歸分析的問題裡，應變數Y為研究者欲預測的變數，而X為輸入變數，輸入就表示已知，故假設X為常數也是合理的。
4. $E(Y_i)=E(Y_i\,|\,X_i)=\beta_o+\beta_1 X_i$
稱為母體迴歸線(population regression function)
複習：當X，Y存在線性函數關係時，

$$E(Y\,|\,X=x)=\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)=(\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}\mu_X)+\rho\frac{\sigma_Y}{\sigma_X}x$$

對照以上母體迴歸線，可得知：

截距項：$\beta_0=(\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}\mu_X)$　　斜率項：$\beta_1=\rho\frac{\sigma_Y}{\sigma_X}$

證明：設$E(X\,|\,Y)=a+bY$

$$E(X)=E[E(X\,|\,Y)]=a+bE(Y)$$

$$\Rightarrow \mu_X=a+b\mu_Y \text{ ---(1)}$$

$$E(XY)=E[YE(X\,|\,Y)]=E[aY+bY^2]=aE(Y)+bE(Y^2)$$

$$\Rightarrow \rho\sigma_X\sigma_Y+\mu_X\mu_Y=a\mu_Y+b[\sigma_Y^2+\mu_Y^2] \text{ ---(2)}$$

以上(1)(2)式解聯立，可得

$$a=\mu_X-\rho\frac{\sigma_X}{\sigma_Y}\mu_Y \quad b=\rho\frac{\sigma_X}{\sigma_Y}$$

估計後的迴歸模型(fitted regression model)或
樣本迴歸線(sample regression function)

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$$

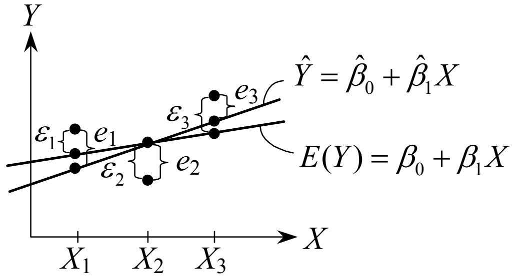

圖3

＜註＞
1. 樣本迴歸線 $\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$ 中那一些是隨機變數?
   應變數之估計值 $\hat{Y}$，迴歸係數之估計值 $\hat{\beta}_0,\hat{\beta}_1$。
2. $\hat{\beta}_i \xrightarrow{估計} \beta_i$，$\hat{Y}_i \xrightarrow{估計} E(Y_i)$，殘差(residual) $e_i=Y_i-\hat{Y}_i \xrightarrow{估計} \varepsilon_i$。

第三節　參數估計

模型：$Y_i=\beta_0+\beta_1X_i+\varepsilon_i$，$i=1,2,\cdots,n$

假設：$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

(1)隨機性且獨立性假設 $Cov(\varepsilon_i,\varepsilon_j)=0$，$i,j=1,2,\cdots,n$，$i\neq j$

(2)平均數為零假設 $E(\varepsilon_i)=0$

(3)變異數齊一性假設 $V(\varepsilon_i)=\sigma^2$

(4)常態性假設 $\varepsilon\sim N$

(5)模型之正確性假設

母體迴歸線：$E(Y_i) \overset{X為常數}{=} E(Y_i\,|\,X_i)=\beta_o+\beta_1X_i$

找 $\hat{\beta}_0$，$\hat{\beta}_1$：

(一)方法一：最小平方法(Ordinary Least Squares，OLS)