第２章

多元回歸分析

迴歸分析（regression analysis）在研究一個或多個自變數對依變數的影響情況，它多用於預測、估計與解釋的統計方法；所謂的預測、估計即是以一個或多個預測變數來描述一個特定效標變數的分析方法（陳順宇，2000）。迴歸分析也是機器學習的一種監督式學習（supervised learning）技術，旨在由訓練資料中學到或建立一個模型，並依此模型推測新的實例。

迴歸分析適用於自變數（independent variable，又稱為**預測變數**，predictor）及依變數（dependent variable，又稱為**效標變數**，criterion）均為計量的變數（含等距變數及比率變數）的分析。如果自變數及依變數各為一個，稱為**簡單迴歸**；如果有多個自變數，一個依變數，稱為**多元迴歸**或**複迴歸**（multiple regression）；如果自變數及依變數均為多個，則是**多變量多元迴歸**（multivariate multiple regression）。

假使自變數是定性的變數（如為名義變數或次序變數），應將該變數轉換為**虛擬變數**（dummy variable）；如果依變數是二類的名義變數，通常會進行二分的**邏輯斯迴歸分析**（binary logistic regression analysis）或 probit **機率迴歸分析**；假使依變數是多類別的名義變數，通常會進行**區別分析**（discriminant）或**多項式邏輯斯迴歸分析**（multinomial logistic regression analysis）；如果依變數是次序變數，則可進行**次序性邏輯斯或機率迴歸分析**。

2.1　迴歸的意義

1885 年 Francis Galton（1982-1911）與 Karl Pearson（1857-1936）在其「Regression towards Mediocrity in hereditary Stature」研究中，發現身高高的父母，其子女之平均身高低於父母的平均身高；反之，身高矮的父母，其子女之平均身高高於父母的平均身高，發現子代有趨向全體平均身高的現象，當時以「regression」一詞表示這樣的效應，表示兩極端身高會「迴歸」到平均數的現象。

圖 2-1 散布圖的 X 軸是父母身高，以（父親身高 + 母親身高 × 1.08）/ 2 代表（因為父親平均身高是母親的 1.08 倍）。Y 軸是子女身高，如果是男性以原始身高代表，女性則乘以 1.08 倍。圖中的虛線代表父母身高與子女身高相同，斜率為 1；直線則是迴歸線，斜率為 0.73。由於迴歸線的斜率小於 1，所以子代的身高不會等於父母

的身高。而兩線相交的地方是（175.82, 175.85），父母身高超過 175.82 公分（平均數），則子代平均身高比其父母矮；反之，父母身高不到 175.82 公分，子代平均身高會比其父母高。

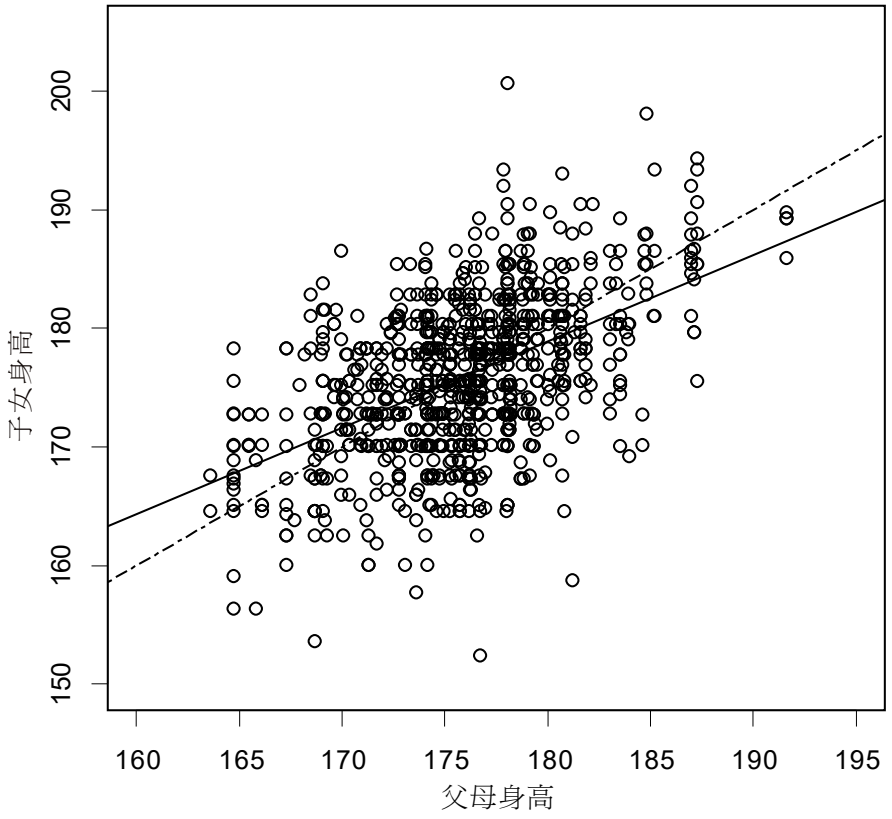


圖 2-1　迴歸現象

迴歸分析與**變異數分析**（analysis of variance）是研究者經常使用的統計方法。而迴歸分析主要的用途有二：一為**解釋**，二為**預測**。解釋的功能主要在於說明預測變數與效果變數間的**關聯強度**及**關聯方向**；預測的功能則是使用迴歸方程式，利用已知的自變數來預測未知的依變數。例如：研究者可以利用高中生在學校的各科畢業成績為預測變數，而以其大學入學成績為效標變數，來建立迴歸方程式，以解釋哪些科目對大學入學成績最有預測作用，及其總預測效果如何。如果其他條件相同，則可利用今年度尚未參加大學入學考試的高中應屆畢業生的各科畢業成績，以預測他們參加入學考試的成績。實務上，兩種取向並未嚴格區分，經常合併使用。

2.2　簡單迴歸

為了確認兩個變數間的關係，可以透過如圖 2-2 的散布圖。由圖中可看出圓點大致呈左下到右上分布，因此兩個變數間有正相關。計算其相關係數值 Pearson $r = 0.70$。

許多時候，研究者還會想進一步了解，是否能用其中一個自變數 X（對智慧型手機的使用態度）預測依變數 Y（行為意圖），此時，就會使用迴歸分析。當只有一個自變數且為線性關係模型，稱為簡單迴歸分析，它是多元迴歸分析的基礎。

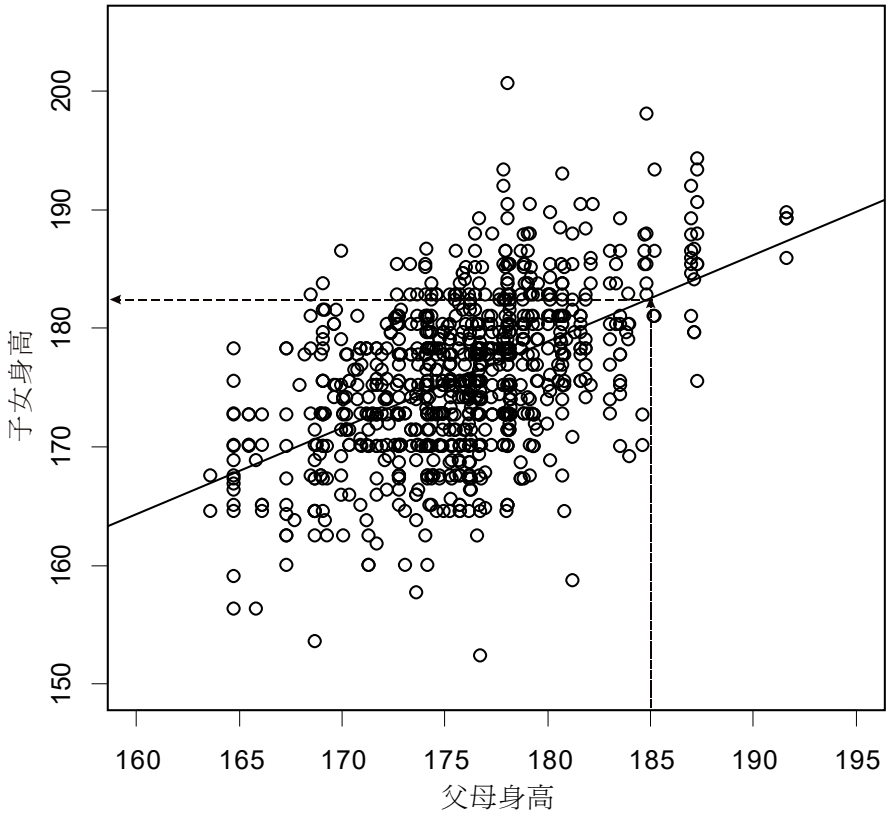


圖 2-2　散布圖

2.2.1　未標準化迴歸係數

簡單線性迴歸模型可以表示為：

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$　　公式 2-1

其中，

1. y_i 是第 i 名個體在依變數 Y 的值，Y 是隨機變數。
2. x_i 是第 i 名個體在自變數 X 的值，X 的值是已知的，不是隨機變數。
3. β_0 與 β_1 是模型的參數，通常是未知的，它反映了由 x 的變化而引起 y 的變化。
4. ε 是隨機誤差，也是隨機變數，代表不能由 x 解釋 y 的其他因素，ε 的平均數 $E(\varepsilon)=0$，變異數 $\sigma_\varepsilon^2=\sigma^2$（謝宇，2013）。

公式 2-1 是以母體參數表示，在樣本中公式改為：

$$Y=b_0+b_1X+e \qquad \text{公式 2-2}$$

在 X 與 Y 的散布圖中，我們希望找到一條適配直線（如圖 2-3 粗直線），使其具有最佳不偏估計式（best linear unbiased estimator, BLUE）的特性。這條直線即簡單迴歸方程式，公式為：

$$\hat{Y}=b_0+b_1X \qquad \text{公式 2-3}$$

其中 b_0 是**常數項**（constant），又稱為**截距**（intercept），b_1 是迴歸的**原始加權係數**，又稱為**斜率**（slope），$\hat{Y}$ 是由 X 所預測的數值，與真正的 Y 變數有差距，差距（**殘差**，residual）$e=Y-\hat{Y}$。在圖中至少有 7 名個體的 X = 10，使用迴歸模型預測，得到 $\hat{Y}$ = 11.15，但是，他們真正的 Y 值從 6 ~ 15 都有，因此會有殘差。

迴歸分析經常使用**普通最小平方法**（ordinary least squares method, OLS）將殘差的平方和 $\sum_{i=1}^{n}e_i^2$ 最小化以求解，求解後：

$$b_1=\frac{CP_{XY}}{SS_X}=\frac{S_{XY}}{S_X^2}=r_{XY}\frac{S_Y}{S_X} \qquad \text{公式 2-4}$$

$$b_0=\bar{Y}-b_1\bar{X} \qquad \text{公式 2-5}$$

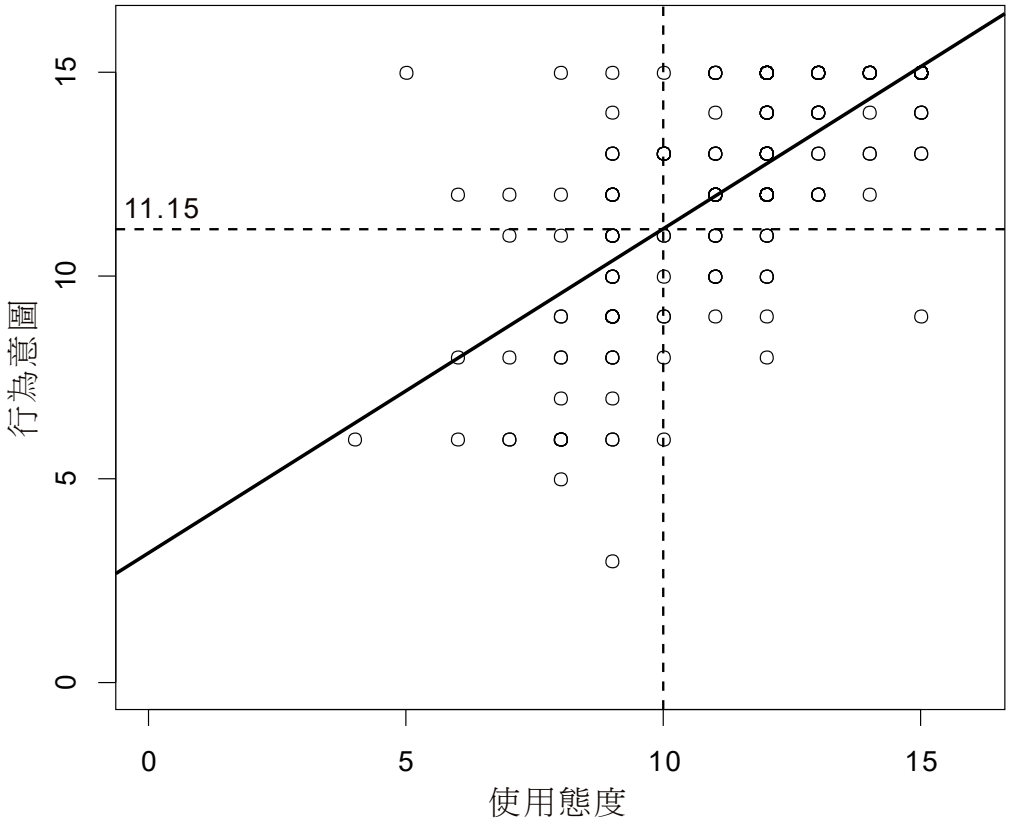


圖 2-3 簡單迴歸適配線

使用 R 進行分析

命令稿 2-1 中共有 4 個指令，分別說明如下：

1. 第 1 個指令讀入 C 碟中 R 資料夾下的 tam292.csv 數據檔，存入 tam 物件。（注：資料可自行設定存放路徑）
2. 由於 tam 資料框架中的變數太多，為了簡化，只選擇第 19～22 行共 4 個變數，存入 reg 物件。
3. 第 3 個指令以 lm() 函數設定 m1 模型，~ 前為效標變數 B（行為意圖），~ 後為預測變數 A（使用態度），數據來自 reg 物件。如果寫成 (m1<−lm(B~A, data=reg)) 則可以同時列出 m1 簡要內容。
4. 第 4 個指令列出 m1 簡要內容，如果要列出詳細內容，可以寫成：summary(m1)。

分析後得到迴歸模型為：

$$B = 3.1885 + 0.7964 \times A$$

當變數 A = 0 時，變數 B = 3.1885；A 每增加 1 分，B 增加 0.7964 分。

命令稿 2-1　簡單迴歸

```
> tam<-read.csv("C:/R/tam292.csv")
> reg<-cbind(tam[,19:22] )
> m1<-lm(B~A, data=reg)
> m1
##
##  Call:
##  lm(formula = B ~ A, data = reg)
##
##  Coefficients:
##  (Intercept)            A
##       3.1885       0.7964
```

2.2.2　平移後未標準化迴歸係數

截距 b_0 是預測變數 X 等於 0 時所求得的 $\hat{Y}$ 值，然而，許多研究中，X 變數並不會等於 0。例如：以父母身高預測子女身高，父母身高不可能等於 0，因此截距並無意義。此時，可以把 X 變數減去其平均數（平移或中心化，centering），再進行迴歸分析，求得的截距就是依變數 Y 的平均數，而斜率則不變。

使用 R 進行分析

命令稿 2-2 中有 4 個指令，分別說明如下：

1. 第 1 個指令以 scale() 函數將 reg 物件中的 A 變數減去平均數，得到 AC 變數。括號引數 scale=F，表示只將變數減去平均數，不除以標準差。
2. 第 2 個指令以 lm() 函數設定 m2 模型，預測變數改為 AC。
3. m2 模型為：B = 12.4589 + 0.7964 × AC。斜率與 m1 模型相同，截距 12.4589 會等於 Y 變數的平均數。指令 2、3 可以合寫為 (m2<−lm(B~AC, data=reg))。

4. 第 4 個指令以 mean() 函數計算 *Y* 變數的平均數，結果為 12.4589，等於迴歸模型中的截距。

命令稿 2-2　簡單迴歸——自變數平移

```
> reg$AC<-scale(reg$A, scale=F)
> m2<-lm(B~AC, data=reg)
> m2
##
##  Call:
##  lm(formula = B ~ AC, data = reg)
##
##  Coefficients:
##  (Intercept)            AC
##      12.4589        0.7964
> mean(reg$B)
##  [1] 12.4589
```

2.2.3　標準化迴歸係數

迴歸分析使用的變數，單位經常不同。即使相同的構念，也會因為測量工具不同，而有不同的分數。為了同一研究中不同變數或是不同研究中相同構念的比較，可以分別將 *X*、*Y* 變數化為 *Z* 分數（平均數為 0，標準差為 1），此時：

$$b_1 = r_{XY}\frac{S_Y}{S_X} = r_{XY}\frac{1}{1} = r_{XY}$$

$$b_0 = \overline{Y} - b_1\overline{X} = 0 - 0 = 0$$

求得的迴歸方程式為 $Z_{\hat{Y}} = b_1^* Z_X$，b_1^* 為**標準化加權係數**估計值，在簡單迴歸中，$b_1^* = r_{XY}$。一般而言，原始的迴歸方程式比較適合直接使用，而標準化迴歸方程式常用在比較預測變數的重要性（前提是沒有嚴重的共線性問題）。美國心理學會（Wilkinson et al., 1999）建議：一般情形下，原始及標準化迴歸係數都要呈現在研究結果中。不過，如果是純粹應用性的研究，只要列出原始係數；而純粹理論性的研究，則只要列出標準化係數。

使用 R 進行分析

命令稿 2-3 中有 6 個指令，分別說明如後：

1. 第 1 個指令使用 scale() 函數將 reg 物件中的 A 變數標準化，存入 reg 物件中，命名為 ZA。
2. 第 2 個指令以同樣的方式將 B 變數標準化，存入 reg 物件中，命名為 ZB。
3. 第 3 個指令以 lm() 函數建立線性模型，依變數為 ZB，自變數為 ZA，結果存於 m3 物件。
4. 列出 m3 結果。截距雖然顯示為 2.478e-16（2.478×10^{-16}），實際上為 0，斜率為 0.7039（7.039×10^{-1}）。
5. 如果要簡化，則使用第 5 個指令以 round() 函數將 m3 物件中的係數 $coefficients 取到小數第 6 位（位數可自訂），分別為 0 及 0.703861。
6. 第 6 個指令以 cor() 函數計算 reg 物件中 B 及 A 兩變數的相關，為 0.703861，與標準化迴歸係數相同。

命令稿 2-3　簡單迴歸——標準化

```
> reg$ZA<-scale(reg$A)
> reg$ZB<-scale(reg$B)
> m3<-lm(ZB~ZA, data=reg)
> m3
##
##  Call:
##  lm(formula = ZB ~ ZA, data = reg)
##
##  Coefficients:
##  (Intercept)            ZA
##    2.478e-16     7.039e-01
> round(m3$coefficients, 6)
##  (Intercept)           ZA
##     0.000000     0.703861
> cor(reg$B, reg$A)
##  [1] 0.703861
```

2.2.4 整體檢定——*F* 檢定

迴歸分析會比照變異數分析將 *SS*（離均差平方和，sum of squares）拆解，並做成摘要表（表 2-1）。

表 2-1 變異數分析摘要表

變異來源	*SS*	*df*	*MS*	*F*	*p*
迴歸	$\Sigma(\hat{Y}-\bar{Y})^2$	*k*	$SS_{迴歸}/df_{迴歸}$	$MS_{迴歸}/MS_{殘差}$	pf(F, df1, df2, lower=F)
殘差	$\Sigma(Y-\hat{Y})^2$	*n*−k−1	$SS_{殘差}/df_{殘差}$		
總和	$\Sigma(Y-\bar{Y})^2$	*n*−1	$SS_{總和}/df_{總和}$=*Y* 的變異數		

當研究者不知道預測變數 X 而想預測 Y，最好的方法就是使用 $\bar{Y}$，因為 $\Sigma(Y-\bar{Y})=0$，而 $\Sigma(Y-\bar{Y})^2 \Rightarrow \min$，$\Sigma(Y-\bar{Y})^2$ 就是 Y 變數的離均差平方和（SS_Y），一般稱為 SS_{total}。SS_{total} 的自由度是樣本數 n 減 1。

如果知道 X 而想預測 Y，最好的方法就是使用模型所得預測值 $\hat{Y}$（因為 $\hat{Y}=b_0+b_1X_1$），$\Sigma(\hat{Y}-\bar{Y})^2$ 表示使用 $\hat{Y}$ 而不用 $\bar{Y}$ 預測 Y 而減少的錯誤，$SS_{reg}=\Sigma(\hat{Y}-\bar{Y})^2$。$SS_{reg}$ 的自由度是預測變數數目 k。

殘差平方和 $\Sigma e^2=\Sigma(Y-\hat{Y})^2$，是使用迴歸方程式不能預測到 Y 的部分，也就是知道 X 而預測 Y，但仍不能減少的錯誤，$SS_{res}=\Sigma(Y-\hat{Y})^2$。$SS_{res}$ 的自由度是 $n-k-1$。

計算後會得到 $SS_{total}=SS_{reg}+SS_{res}$，同樣地，$df_{total}=df_{reg}+df_{res}$，把 *SS* 除以各自的自由度就是 *MS*（mean square，平均平方和），而計算所得 $F=MS_{reg}/MS_{res}$。計算所得的 *F* 如果大於或等於 *F* 分配的臨界值〔用 qf(α, df1, df2, lower=F) 計算，如 qf(.05, 1, 290, lower=F) 等於 3.874〕或是 $p \le \alpha$，則應拒絕 H_0，表示自變數可以顯著預測依變數，這也是迴歸分析的整體檢定。

使用 R 進行分析

命令稿 2-4 有 2 個指令，分別說明如下：

1. 第 1 個指令先以 lm() 建立線性模型，存入 m1 物件。資料來自 reg 物件，依變數為 B，自變數為 A。

2. 第 2 個指令再以 anova() 函數列出 m1 的 *SS*，分別是 953.44 及 971.07，$F(1, 290) = 284.74$，$p < 0.001$，因此至少有一個自變數（在此模型只有 A）可以顯著預測依變數 B。

命令稿 2-4　變異數分析摘要表

```
> m1<-lm(B~A, data=reg)
> anova(m1)
##  Analysis of Variance Table
##
##  Response: B
##             Df Sum Sq Mean Sq F value    Pr(>F)
##  A           1 953.44  953.44  284.74 < 2.2e-16 ***
##  Residuals 290 971.07    3.35
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.2.5　個別檢定——*t* 檢定

個別係數的檢定採用 *t* 檢定（$H_0 : \beta_1 = 0$），將斜率減去假設母體的斜率（通常是 0）除以斜率的標準誤，就可得到 *t* 值：

$$t_1 = \frac{b_1 - \beta_1}{se(b)}$$

計算所得 *t* 值如果大於或等於臨界值 *t* 值，或是 $p \leq \alpha$，就表示斜率顯著不為 0。

使用 R 進行分析

命令稿 2-5 中先以 summary() 函數列出模型 m1 的摘要，斜率為 0.7964，它的標準誤為：

$$se(b) = \sqrt{\frac{MS_{res}}{SS_X}} = \sqrt{\frac{3.35}{1503.243}} = 0.0472$$

計算所得 *t* 值為：

$$t = \frac{0.7964 - 0}{0.0472} = 16.874$$

P 值小於 0.001，因此斜率顯示不為 0，也就是預測變數 A 可以顯著預測 B 變數。截距的 *t* 檢定為：

$$t_0 = \frac{b_0 - \beta_0}{\sqrt{MS_{res}\left(\frac{1}{N} + \frac{\bar{X}^2}{SS_X}\right)}} = \frac{3.1885 - 0}{0.5597} = 5.696$$

由於截距要在所有預測變數都為 0 時，才有意義，因此，研究者通常比較不關心截距的檢定。

第 2 個指令以 var() 函數計算 A 變數的變異數，再乘以 $n - 1$ 即為 SS_X，結果為 1503.243。

命令稿 2-5　係數檢定

```
> summary(m1)
##
##  Call:
##  lm(formula = b ~ a, data = reg)
##
##  Residuals:
##      Min       1Q   Median       3Q      Max
##  -7.3561 -0.7453 -0.1345  1.2547  7.8295
##
##  Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##  (Intercept)   3.1885     0.5597   5.696    3e-08 ***
##  a             0.7964     0.0472  16.874   <2e-16 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Residual standard error: 1.83 on 290 degrees of freedom
##  Multiple R-squared:  0.4954,    Adjusted R-squared:  0.4937
##  F-statistic: 284.7 on 1 and 290 DF,  p-value: < 2.2e-16
> var(reg$A)*(length(reg$A)-1)
##  [1] 1503.243
```

2.2.6 效果量 R^2

在計算迴歸的**效果量**（effect size）時，一般會使用**消減錯誤比例**（proportional reduction in error, PRE）。$PRE = \frac{E_1 - E_2}{E_1}$，$E_1$ 是不知道 X 變數而直接預測 Y 變數時的錯誤，也就是 SS_{total}；E_2 是知道 X 而預測 Y 的錯誤，也就是 SS_{res}，因此迴歸分析的效果量公式為：

$$R^2 = PRE = \frac{SS_{total} - SS_{res}}{SS_{total}} = \frac{SS_{reg}}{SS_{total}} = 1 - \frac{SS_{res}}{SS_{total}}$$

R^2 稱為**決定係數**（coefficient of determination），是預測變數對依變數的解釋力，也是 Y 的實際值與預測值 $\hat{Y}$ 之 Pearson 相關 $r_{Y\hat{Y}}$ 的平方。如果要應用到不同的群體時，R^2 會有縮減（shrinkage）現象，一般統計軟體常用公式 2-6 計算調整 R^2：

$$\hat{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-k-1} \qquad \text{公式 2-6}$$

在命令稿 2-4 中，迴歸及殘差的 SS 分別為 953.44 與 971.07，因此 R^2 為：

$$R^2 = \frac{953.44}{953.44 + 971.07} = 0.4954$$

調整 R^2 為：

$$\hat{R}^2 = 1 - (1 - 0.4954)\frac{292-1}{292-1-1} = 0.4937$$

以上計算結果均與命令稿 2-5 中的 R-squared 相同。

2.3 多個預測變數的多元迴歸模型

建立迴歸模型時，很少只用一個自變數，而會使用兩個以上的自變數，以更準確預測依變數，此時稱為**多元迴歸分析**（或**複迴歸分析**）。

以下概要說明多元迴歸分析相關概念。

2.3.1 迴歸係數之求解

多元迴歸分析主要在建立以下的模型：

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

一般統計軟體在進行多元迴歸分析時，是以矩陣形式運算，其原始係數解為：

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

標準化係數使用 $\mathbf{R}^{-1}\mathbf{r}$ 求得，其中 $\mathbf{R}^{-1}$ 是自變數間的相關矩陣，$\mathbf{r}$ 是自變數與依變數間的相關矩陣（行向量）。

使用 R 進行分析

命令稿 2-6 共有 6 個指令，分別說明如下：

1. 第 1 個指令以行合併函數 cbind() 將常數 1 及 reg 中的 1 ~ 3 行合併成 X 物件，其中包含 E（認知易用性）、U（認知有用性）、A（使用態度）三個自變數。
2. 第 2 個指令將 X 物件轉為矩陣。
3. 第 3 個指令將 reg 物件的第 4 行存為 Y 物件，為依變數 B。
4. 第 4 個指令在解 $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$，其中轉置矩陣的函數為 t()，solve() 函數在求反矩陣，%*%為矩陣相乘運算符號。求解後置於 Beta 物件。
5. 第 5 個指令列出 Beta 物件內容，得到多元迴歸模型為：
   B = 0.7748435 + 0.1167639 × E + 0.3697811 × U + 0.5056607 × A
6. 如果要計算標準化的係數，可以先使用 scale() 函數將原始資料標準化為 sreg，再重複步驟 1 ~ 5 即可。其中第 1 個指令不加常數項，reg 物件改為 sreg。

命令稿 2-6 以矩陣解原始係數

```
> X<-cbind(1, reg[, 1:3])
> X <- as.matrix(X)
> Y<-reg[,4]
> Beta<- solve(t(X) %*% X) %*% t(X) %*% Y
```

```
> Beta
##          [,1]
##  1 0.7748435
##  E 0.1167639
##  U 0.3697811
##  A 0.5056607
> sreg<-scale(reg)
```

命令稿 2-7 有 2 個指令，分別說明如下：

1. 第 1 個指令先以 lm() 函數設定線性模型 m4，自變數為 E、U、A，依變數為 B，資料來自 reg 物件。
2. 第 2 個指令再以 summary() 列出摘要，係數與命令稿 2-6 相同，各斜率均顯著不等於 0，截距則與 0 沒有不同。

命令稿 2-7　以 lm() 函數解原始係數

```
> m4<-lm(B~E+U+A, data=reg)
> summary(m4)
##
##  Call:
##  lm(formula = B ~ E + U + A, data = reg)
##
##  Residuals:
##      Min       1Q   Median       3Q      Max
##  -7.0745 -0.8176 -0.0511   1.0926   4.5557
##
##  Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##  (Intercept)  0.77484    0.58780   1.318  0.18848
##  E            0.11676    0.04195   2.783  0.00574 **
##  U            0.36978    0.05543   6.671  1.3e-10 ***
##  A            0.50566    0.05590   9.046  < 2e-16 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Residual standard error: 1.656 on 288 degrees of freedom
##  Multiple R-squared:  0.5895,    Adjusted R-squared:  0.5852
##  F-statistic: 137.8 on 3 and 288 DF,  p-value: < 2.2e-16
```

2.3.2 整體檢定

整體檢定在同時檢定所有預測變數是否能顯著預測依變數，它的統計假設為：

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

$$H_1 : \text{至少有一個 } \beta \text{ 不等於 } 0$$

使用 R 進行分析

在 R 中如果直接以 anova() 函數列出模型摘要，並無法進行整體檢定。我們可以透過 update() 比較只含常數項 1 之模型與完整模型，間接進行整體檢定。

命令稿 2-8 中列出的 SS_{total} 為 1924.51，SS_{res} 為 709.06，SS_{reg} 為 1134.5，$F(3, 288) = 137.85$，$p < 0.001$，結果與命令稿 2-7 最後一列相同，因此 3 個迴歸係數中，至少有 1 個不等於 0。

命令稿 2-8 整體檢定

```
> anova(update(m4, ~1), m4)
## Analysis of Variance Table
##
## Model 1: B ~ 1
## Model 2: B ~ E + U + A
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    291 1924.51
## 2    288  790.06  3    1134.5 137.85 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.3.3 個別檢定

個別檢定在檢定控制了其他變數後，某個預測變數是否能顯著預測依變數。它的統計假設為：

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

在 R 中，直接使用 summary(模型) 即可列出模型中個別係數的檢定。由命令稿

2-7 中看出：3 個迴歸係數的 p 值均小於 0.01，因此均顯著不等於 0。

使用 R 進行分析

個別係數的檢定，也可以透過係數的信賴區間來判斷。命令稿 2-9 以 confint() 函數列出模型 m4 中的係數信賴區間，內定為 level=0.95，如果要改變區間大小，可以另外設定。由結果可看出：除了截距外，其他係數的 95%信賴區間都不包含 0，因此與 0 有顯著不同。

命令稿 2-9　係數的信賴區間

```
> confint(m4)
##                    2.5 %     97.5 %
##  (Intercept) -0.38208003 1.9317670
##  e            0.03419524 0.1993325
##  u            0.26067730 0.4788849
##  a            0.39564160 0.6156798
```

2.3.4　標準化迴歸係數

由於 lm() 函數只提供未標準化迴歸係數，如果要計算標準化迴歸係數，可以使用 lm.beta 程式套件，或是使用 scale() 函數先將每個變數標準化。

使用 R 進行分析

命令稿 2-10 有 3 個指令，分別說明如下：

1. 第 1 個指令載入 lm.beta 程式套件。
2. 第 2 個指令以 lm.beta() 函數將 m4 模型標準化後存在 m4.1 物件。
3. 第 3 個指令以 summary() 函數列出模型 m4.1 摘要。Standardized 這一行即為標準化迴歸係數。
4. 第 4 個指令是以 scale() 函數先將每個變數標準化後，再進行線性迴歸分析。

命令稿 2-10　標準化係數

```
> library(lm.beta)
> m4.1<-lm.beta(m4)
```

```
> summary(m4.1)
##  Call:
##  lm(formula = B ~ E + U + A, data = reg)
##
##  Residuals:
##      Min       1Q   Median       3Q      Max
##  -7.0745 -0.8176 -0.0511   1.0926   4.5557
##
##  Coefficients:
##              Estimate Standardized Std. Error t value Pr(>|t|)
##  (Intercept)  0.77484       0.00000    0.58780   1.318  0.18848
##  E            0.11676       0.12459    0.04195   2.783  0.00574 **
##  U            0.36978       0.32596    0.05543   6.671  1.3e-10 ***
##  A            0.50566       0.44690    0.05590   9.046  < 2e-16 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Residual standard error: 1.656 on 288 degrees of freedom
##  Multiple R-squared:  0.5895,    Adjusted R-squared:  0.5852
##  F-statistic: 137.8 on 3 and 288 DF,  p-value: < 2.2e-16
> lm(scale(B)~scale(E)+scale(U)+scale(A), data=reg)
```

2.3.5 繪製迴歸係數圖

迴歸係數通常不需要以圖表示，如果要像第 11～14 章繪製迴歸係數圖，除了使用 lavaan 程式套件外，也可以使用 psych 程式套件進行。

使用 R 進行分析

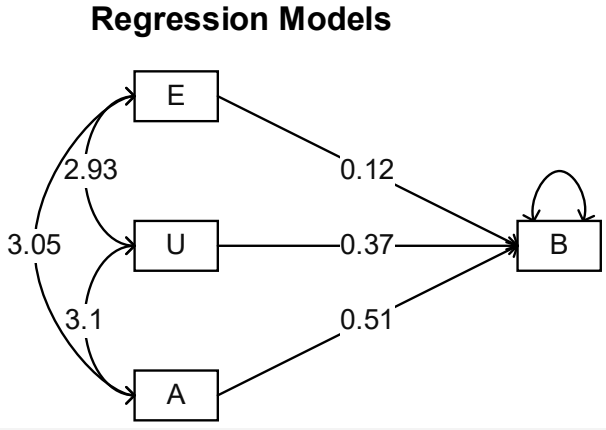
命令稿 2-11 有 7 個指令，分別說明如下：

1. 第 1 個指令載入 psych 程式套件。
2. 第 2 個指令以 setCor() 函數設定迴歸模型，語法與 lm() 相似，~ 前是效標變數，~ 後是預測變數，如果要以原始係數繪出，設定 std=FALSE。分析後將結果存在 m4.1 物件，此時，會立即繪出模式圖，係數設定為小數後 2 位，分別是 0.12、0.37、0.51。預測變數間的雙向箭頭代表關聯程度，未標準化時為共變數。如果要改變小數位，可再以第 5 個指令設定。

3. 第 3 個指令列出 m4.1 模型，報表中包含原始迴歸係數、係數標準誤、t 值、p 值、95%信賴區間、VIF 值、R^2、$\hat{R}^2$。
4. 第 4 個指令列出 m4.2 模型，3 個變數的 VIF 值分別為 1.41、1.68、1.71，均未大於 10，代表 3 個自變數的多元共線性問題並不嚴重。
5. 第 5 個指令設定 std=TRUE（也是默認的設定），將結果存於 m4.2 物件。標準化係數分別為 0.12、0.33、0.45。預測變數間為 Pearson 相關係數，介於 0.47 ~ 0.60 間，顯示共線性問題不嚴重。
6. 第 6 個指令以 lmCor.diagram() 函數將 m4.1 物件（原始迴歸係數）繪出，設定小數位為 3。
7. 第 7 個指令以 lmCor.diagram() 函數將 m4.2 物件（標準化迴歸係數）繪出，設定小數位為 3。

命令稿 2-11　繪製迴歸係數圖

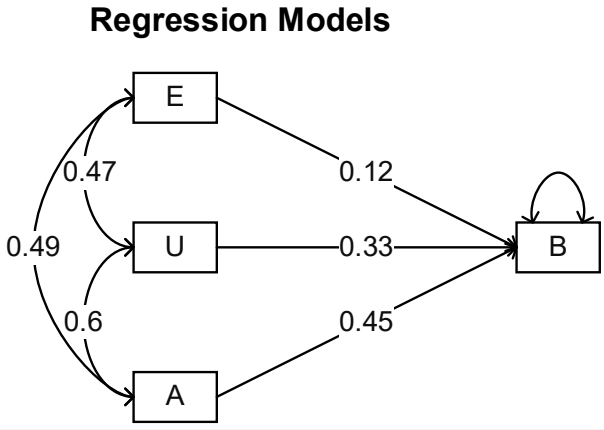
```
> library(psych)
> m4.1<-setCor(B~E+U+A, data=reg, std=FALSE)
```




```
> m4.1
##  Call: setCor(y = B ~ E + U + A, data = reg, std = FALSE)
##
##  Multiple Regression from raw data
##
##   DV =  B
##
```

```
##              slope   se    t       p lower.ci upper.ci   VIF
##  (Intercept)  0.77 0.59 1.32 1.9e-01    -0.38     1.93 36.78
##  E            0.12 0.04 2.78 5.7e-03     0.03     0.20 22.37
##  U            0.37 0.06 6.67 1.3e-10     0.26     0.48 51.47
##  A            0.51 0.06 9.05 2.3e-17     0.40     0.62 46.78
##
##  Residual Standard Error =  1.66  with  288  degrees of freedom
##
##   Multiple Regression
##       R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2        p
##  B 0.77 0.59 0.69 0.48         0.59      0.04    137.85   3 288 2.19e-55
```

```
> m4.2<-setCor(B~E+U+A, data=reg, std=TRUE)
```




```
> m4.2
##  Call: setCor(y = B ~ E + U + A, data = reg, std = TRUE)
##
##  Multiple Regression from raw data
##
##   DV =  B
##              slope   se    t       p lower.ci upper.ci  VIF
##  (Intercept)  0.00 0.04 0.00 1.0e+00    -0.07     0.07 1.00
##  E            0.12 0.04 2.78 5.7e-03     0.04     0.21 1.41
##  U            0.33 0.05 6.67 1.3e-10     0.23     0.42 1.68
##  A            0.45 0.05 9.05 2.3e-17     0.35     0.54 1.71
[省略部分報表]
> lmCor.diagram(m4.1,digits=3)
> lmCor.diagram(m4.2,digits=3)
```

2.3.6 效果量 R^2

由命令稿 2-7 可看出：3 個預測變數對依變數的解釋力 $R^2 = 0.5895 = 58.95\%$，調整後 R^2 為：

$$\hat{R}^2 = 1 - (1 - 0.5895)\frac{292 - 1}{292 - 3 - 1} = 0.5852 = 58.52\%$$

依據 Cohen（1988）的經驗法則，多元迴歸分析 R^2 值之小、中、大的效果量分別是 0.02、0.13，及 0.26，本範例為大的效果量。

使用 R 進行分析

命令稿 2-12 有 6 個指令，分別說明如下：

1. 第 1 個指令以 predict() 函數將 m4 的預測值存入 reg 物件中，命名為 predB。
2. 第 2 個指令以 residuals() 函數將 m4 的預測值存入 reg 物件中，命名為 resB。
3. 第 3 個指令以 cor() 函數計算 reg 物件中 B 及 predB 的積差相關，得到 r = 0.7678，此即為 B 與 E、U、A 三個變數的多元相關 R。
4. 第 4 個指令計算多元相關的平方，得到 R^2 = 0.5895，就是決定係數。
5. 第 5 個指令計算 $\sqrt{1 - R^2}$ = 0.6407224，此稱為疏離係數（coefficient of alienation），它是依變數 B 與殘差的相關係數。
6. 第 6 個指令以 cor() 函數計算依變數 B 與殘差的相關，得到 0.6407224，與指令 5 的結果一致。

命令稿 2-12　計算 R^2 及疏離係數

```
> reg$predB<-predict(m4)
> reg$resB<-residuals(m4)
> cor(reg$B, reg$predB)
##   [1] 0.7677726
> cor(reg$B, reg$predB)^2
##   [1] 0.5894748
> (1-cor(reg$B, reg$predB)^2)^.5
##   [1] 0.6407224
> cor(reg$B, reg$resB)
##   [1] 0.6407224
```

2.4　虛擬變數的多元迴歸分析

在多元迴歸分析中，預測變數的性質通常是**等距變數**或**比率變數**（二者合稱**計量性資料**）。當預測變數為**名義變數**或**次序變數**之**非計量性資料**時，不可以直接投入分析，必須轉換成**虛擬變數**，以 0、1 代表。轉換成虛擬變數時，虛擬變數的數目必須是**水準**（level）數減 1，以避免**線性相依**（linearly dependent）的情形。

2.4.1　兩個類別的預測變數

假設研究者想以受訪者的「性別」，預測他對智慧型手機的「行為意圖」。分析時可以把「女性」編碼為 0，「男性」編碼為 1（相反亦可），再設定線性模型即可。

使用 R 進行分析

命令稿 2-13 中先以 lm() 函數設定模型 m5，依變數為 B，自變數為 GENDER，再以 summary() 函數列出 m5 摘要，得到迴歸模型為：

$$B = 12.2381 + 0.6260 \times GENDER$$

由於女性編碼為 0，因此代入迴歸模型後得到「行為意圖」分數為：

$$B = 12.2381 + 0.6260 \times 0 = 12.2381$$

可知，截距代表女性的「行為意圖」平均分數，此時女性是參照組。而男性代碼為 1，代入迴歸模型後得到「行為意圖」分數為：

$$B = 12.2381 + 0.6260 \times 1 = 12.8641$$

因此男性對使用智慧型手機的「行為意圖」平均分數為 12.8641，兩性的「行為意圖」平均分數相差 0.6260 分。模型中斜率的 t 值為 1.998，$p = 0.0467$，所以受訪者的「性別」可以顯著預測他對智慧型手機的「行為意圖」。不過，R^2 僅有 0.01357，並不高。

命令稿 2-13　以性別為預測變數的迴歸分析

```
> m5<-lm(B~GENDER, data=tam)
> summary(m5)
```

```
##
##  Call:
##  lm(formula = B ~ GENDER, data = tam)
##
##  Residuals:
##      Min       1Q   Median       3Q      Max
##  -9.2381 -1.2381   0.1359   2.1359   2.7619
##
##  Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##  (Intercept)  12.2381     0.1861  65.758   <2e-16 ***
##  GENDER        0.6260     0.3134   1.998   0.0467 *
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Residual standard error: 2.559 on 290 degrees of freedom
##  Multiple R-squared:  0.01357,    Adjusted R-squared:  0.01017
##  F-statistic: 3.991 on 1 and 290 DF,  p-value: 0.04668
```

事實上，同樣的數據，也可以進行兩個獨立樣本平均數 t 檢定。命令稿 2-14 中以 t.test() 函數進行 t 檢定，依變數為 B，自變數為 GENDER，假定兩組變異數相等。由結果中可看出：

1. 女性的「行為意圖」平均數為 12.23810，男性為 12.86408，與前面的計算結果相同。
2. $t(290) = -1.9977$，$p = 0.04668$，其中 p 值與命令稿 2-13 的 Pr(>|t|)相同，t 值則正負相反，但數值相同，這是因為在 t 檢定中分子是以女性的平均數減去男性的平均數，所以結果為負。把−1.9977 取平方，會等於命令稿 2-13 最後一列的 F 值 3.991。

命令稿 2-14　獨立樣本 t 檢定

```
> t.test(B~GENDER, var.equal=T, data=tam)
##
##          Two Sample t-test
##
```

```
##  data:  B by GENDER
##  t = -1.9977, df = 290, p-value = 0.04668
##  alternative hypothesis: true difference in means is not equal to 0
##  95 percent confidence interval:
##   -1.242718230 -0.009246633
##  sample estimates:
##  mean in group 0 mean in group 1
##         12.23810         12.86408
```

由於性別是二分變數，而「行為意圖」是計量變數，兩個變數可以計算點二系列相關。命令稿 2-15 中先以 cor() 函數計算 tam 物件中的 B 與 GENDER 之相關，得到點二系列相關 $r_{pb} = 0.1165$（也是 Pearson's r）。其次，以 cor.test() 函數進行檢定，得到 $t(290) = 1.9977$，$p = 0.04668$，與命令稿 2-13 中的 t 及 p 一致。最後計算 r_{pb} 平方，數值為 0.01357，這就是命令稿 2-13 中 R^2。

命令稿 2-15　相關係數

```
> cor(tam$B, tam$GENDER)
##  [1] 0.1165093
> cor.test(tam$B, tam$GENDER)
##
##          Pearson's product-moment correlation
##
##  data:  tam$B and tam$GENDER
##  t = 1.9977, df = 290, p-value = 0.04668
##  alternative hypothesis: true correlation is not equal to 0
##  95 percent confidence interval:
##   0.001748792 0.228240794
##  sample estimates:
##        cor
##  0.1165093
> cor(tam$B, tam$GENDER)^2
##  [1] 0.01357441
```

2.4.2　三個以上類別的預測變數

如果以受訪者的「年齡層」，預測他對智慧型手機的「行為意圖」。由於有 4 個年

齡層，因此須以 3 個虛擬變數代表，茲以表 2-2 表示之：

表 2-2　虛擬變數的轉換

		虛擬變數		
		factor(AGE)2：20-29 歲	factor(AGE)3：30-39 歲	factor(AGE)4：40 歲以上
原變項	1：19 歲以下	0	0	0
	2：20-29 歲	1	0	0
	3：30-39 歲	0	1	0
	4：40 歲以上	0	0	1

由上表可看出：原來以 1 代表 19 歲以下的受訪者，設定為參照組，經轉換後在 3 個虛擬變數的數值分別是 0、0、0。第 2 組在 3 個虛擬變數的數值分別是 1、0、0。其他兩組轉碼如表所示。經過這樣的轉換後，即可將年齡層當成預測變數。

使用 R 進行分析

命令稿 2-16 中以 lm() 函數設定迴歸模型 m6，依變數為 B，預測為 AGE，加上 factor，直接轉成 3 個虛擬變數。整體檢定 $F(3, 288) = 0.508$，$p = 0.677$，「年齡層」無法預測受訪者的「行為意圖」。迴歸方程式為：

$$B = 12.43939 + (-0.02192) \times \text{factor(AGE)2} + 0.25940 \times \text{factor(AGE)3} + (-0.33939) \times \text{factor(AGE)4}$$

由模型可看出：

1. 年齡層為 1 的受訪者（參照組），「行為意圖」的平均數為 12.43939。
2. 年齡層為 2 的受訪者，「行為意圖」的平均數比第 1 組低 0.02192，12.43939 − 0.02192 = 12.41747。
3. 年齡層為 3 的受訪者，「行為意圖」的平均數比第 1 組高 0.25940，12.43939 + 0.25940 = 12.69879。
4. 年齡層為 4 的受訪者，「行為意圖」的平均數比第 1 組低 0.33939，12.43939 − 0.33939 = 12.10000。
5. 各組「行為意圖」的平均數與第 1 組相比，都未達 0.05 顯著水準之差異。

命令稿 2-16 以年齡層為預測變數的迴歸分析

```
> m6<-lm(B~factor(AGE), data=tam)
> anova(m6)
## Analysis of Variance Table

## Response: B
##              Df  Sum Sq Mean Sq F value Pr(>F)
## factor(AGE)   3   10.13  3.3770   0.508  0.677
## Residuals   288 1914.38  6.6471
> summary(m6)
##
## Call:
## lm(formula = B ~ factor(AGE), data = tam)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6988 -1.1000  0.5606  2.3012  2.9000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.43939    0.31736  39.197   <2e-16 ***
## factor(AGE)2 -0.02192    0.40651  -0.054    0.957
## factor(AGE)3  0.25940    0.42521   0.610    0.542
## factor(AGE)4 -0.33939    0.51662  -0.657    0.512
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.578 on 288 degrees of freedom
## Multiple R-squared:  0.005264,   Adjusted R-squared:  -0.005098
## F-statistic: 0.508 on 3 and 288 DF,  p-value: 0.677
```

同樣的數據，也可以使用變異數分析。命令稿 2-17 有 4 個指令，分別說明如下：

1. 第 1 個指令以 aov() 函數設定 anova 模型 fit1，依變數為 B，自變數為 AGE，由於 AGE 為數值，因此以 as.factor() 函數設定為因子。
2. 第 2 個指令以 summary() 函數列出 fit1 的結果，$F(3, 288) = 0.508$，$p = 0.677$，與命令稿 2-16 中 m6 的結果相同。

3. 第 3 個指令載入 Rmisc 程式套件。
4. 第 4 個指令以 summarySE() 函數列出 AGE 各組之 B 變數的各項統計量，其平均數與前面計算的結果相同。

由上述的分析可以得知：變異數分析是迴歸分析的特例。將變異數分析的自變數轉為虛擬變數，當成迴歸分析的預測變數，最後分析結果是一致的。

命令稿 2-17　變異數分析

```
> fit1<-aov(B~as.factor(AGE), data=tam)
> summary(fit1)
##                    Df Sum Sq Mean Sq F value Pr(>F)
##   as.factor(AGE)    3   10.1   3.377   0.508  0.677
##   Residuals       288 1914.4   6.647
> library(Rmisc)
> summarySE(m="B", g="AGE", data=tam)
##     AGE   N        B       sd        se        ci
##   1   1  66 12.43939 2.572809 0.3166909 0.6324754
##   2   2 103 12.41748 2.471501 0.2435243 0.4830292
##   3   3  83 12.69880 2.783816 0.3055635 0.6078632
##   4   4  40 12.10000 2.405123 0.3802833 0.7691956
```

2.5　預測變數的選擇

多元迴歸模型根據目的的不同，可大致分成三種模型建立方式：同時迴歸分析（simultaneous regression）、逐步迴歸分析（stepwise regression）、與階層式迴歸分析（hierarchical regression）。以下分別簡要論述之。

2.5.1　解釋型迴歸分析——同時迴歸分析

解釋型迴歸分析的目的是釐清研究者關心變數間的關係，以及如何對於依變數的變異提出一套最合理最有解釋力的迴歸模型（邱皓政，2011a）。理論的重要性不僅決定變數的選擇與安排，也影響研究結果的解釋，故在建立迴歸方程式時，除非預測變數間的共線性問題過高，否則每一個預測變數皆會保留在模型中，並不會將未達顯

著的變數排除在外。

本章前面所述的分析方法，都是同時迴歸分析，不再說明 R 語法。

2.5.2 預測型迴歸分析——逐步迴歸分析

逐步迴歸方法是指以最少的變數來達成對依變數最大的預測力。此法是利用各解釋變數與依變數的相關的相對強弱，來決定哪些解釋變數應納入迴歸方程式中，而非依理論的觀點來取捨變數（邱皓政，2011a）。預測型迴歸分析不是以變數的關係的釐清為目的，而是以建立最佳方程式為目標。逐步迴歸分析方法有三種：向前法、向後法、與逐步法。向前法是依次納入淨進入 F 值最大的變數，一直到沒有符合條件的變數為止。向後法則先將所有變數納入迴歸方程式，然後依次剔除淨退出 F 值最大的變數，一直到沒有符合條件的變數為止。逐步法則兼用向前法與向後法反覆進行，一直到沒有變數被選取或剔除為止，它是最常被使用的方法。

然而，學者（Edirisooriya, 1995; Thompson, 1995b）也指出逐步迴歸的許多問題：1.更換研究樣本後，選取的預測變數就可能不同；2.進入迴歸模型中的預測變數不見得是最重要的變數；3.變數進入的順序不代表重要性的順序，因此在應用時宜更加留意。

使用 R 進行分析

命令稿 2-18 先設定只有常數項 1 之 m7.0 模型，再設定含 3 個預測變數的完整模型 m7.f，接著用 step() 函數以 both 法（雙向，也就是逐步迴歸）從 m7.0 到 m7.f 進行逐步迴歸分析，並將結果存在 m7.s 模型。

由結果可看出，預測變數依序加入 A、U、E，最後的 AIC 為 298.64（愈小模型愈好）。如果要檢視最後結果，可以使用 summary(m7.s) 列出模型摘要。

命令稿 2-18 逐步迴歸分析

```
> m7.0<-lm(B~1, data=reg)
> m7.f<-lm(B~1+E+U+A, data=reg)
> m7.s<-step(m7.0, scope=list(upper=m7.f), direction="both")
##
##
```

```
##  Start:  AIC=552.62
##  B ~ 1
##
##         Df Sum of Sq     RSS    AIC
##  + A     1    953.44  971.07 354.88
##  + U     1    821.78 1102.72 392.00
##  + E     1    475.23 1449.28 471.80
##  <none>               1924.51 552.62
##
##  Step:  AIC=354.88
##  B ~ A
##
##         Df Sum of Sq     RSS    AIC
##  + U     1    159.76  811.31 304.39
##  + E     1     58.93  912.13 338.60
##  <none>               971.07 354.88
##  - A     1    953.44 1924.51 552.62
##
##  Step:  AIC=304.39
##  B ~ A + U
##
##         Df Sum of Sq     RSS    AIC
##  + E     1    21.252  790.06 298.64
##  <none>               811.31 304.39
##  - U     1   159.756  971.07 354.88
##  - A     1   291.412 1102.72 392.00
##
##  Step:  AIC=298.64
##  B ~ A + U + E
##
##         Df Sum of Sq     RSS    AIC
##  <none>               790.06 298.64
##  - E     1    21.252  811.31 304.39
##  - U     1   122.076  912.13 338.60
##  - A     1   224.493 1014.55 369.67
```

```
> summary(m7.s)
```

2.5.3　階層迴歸分析

階層迴歸分析類似逐步迴歸分析作法，一樣是分成好幾個步驟，逐步進行迴歸分析。但與逐步迴歸分析不同的是，逐步迴歸分析是由 F 統計量作為預測變數取捨的依據，而階層迴歸分析則是研究者根據理論或研究需要所設定（邱皓政，2011a）。具體言之，在進行預測變數進入迴歸模型的篩選時，有時預測變數的投入是有先後次序關係的，必須依據次序來進行分析。因此，階層迴歸分析的重要任務即是分階段投入變數來進行迴歸分析，且變數投入次序的規劃，必須有理論基礎，並非研究者可自由決定或由電腦自行篩選。

使用 R 進行分析

命令稿 2-19 中先以 lm() 函數設定階層模型，m8.0 只含常數項 1，後面依據科技接受模型（TAM）的順序加上 E、U、A。設定後再以 anova 比較各模型，F 值均達 0.05 顯著水準，表示後一個模型都比前一個模型好。由第 1 個 m8.0 模型中得到 SS_{total} 為 1924.51，加入其他預測變數後 SS_{reg} 分別增加 475.23、434.73、224.49，將它們分別除以 1924.51 後即為增加的 R^2 ，各為 0.247、0.226、0.117。

命令稿 2-19　階層迴歸分析

```
> m8.0<-lm(B~1, data=reg)
> m8.1<-lm(B~1+E, data=reg)
> m8.2<-lm(B~1+E+U, data=reg)
> m8.3<-lm(B~1+E+U+A, data=reg)
> anova(m8.0, m8.1, m8.2, m8.3)
## Analysis of Variance Table
##
## Model 1: B ~ 1
## Model 2: B ~ 1 + E
## Model 3: B ~ 1 + E + U
## Model 4: B ~ 1 + E + U + A
##   Res.Df     RSS Df Sum of Sq       F    Pr(>F)
## 1    291 1924.51
## 2    290 1449.28  1    475.23 173.234 < 2.2e-16 ***
## 3    289 1014.55  1    434.73 158.472 < 2.2e-16 ***
## 4    288  790.06  1    224.49  81.834 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> 475.23/1924.51
## [1] 0.2469356
> 434.73/1924.51
## [1] 0.2258913
> 224.49/1924.51
## [1] 0.1166479
```

2.6 樣本數之決定

進行多元迴歸分析時，每個自變數最少要有 5 個樣本，而且最好有 15～20 個樣本；如果使用逐步法，則更要增加到 50 個樣本，如此迴歸分析的結果才具有類推性，也才可以適用到不同的樣本上（Hair et al., 2019）。

2.7 迴歸診斷

在迴歸診斷方面，大致可分成三部分：1.**殘差**的檢定；2.**離群值**（outlier）及**具影響力觀察值**（influential observation）的檢出；3.**共線性**的檢定。

此部分較為專業，不在本書說明，讀者可參閱陳正昌（2011a）的另一篇著作。

2.8 使用 JASP 分析

在 JASP 中，讀入 reg.csv 資料檔後，在 Regression（迴歸分析）中選擇 Linear Regression（線性迴歸）。接著，將 B 變數選至 Dependent Variable（依變數）中，E、U、A 三個變數選至 Covariates（共變量）中，此時，預設的分析方法是 Enter（圖 2-4）。

在 Statistics（統計量）中，除了已勾選的 Estimates（估計值）及 Model fit（模型適配度外），可再勾選 Confidence intervals（信賴區間）（圖 2-5）。

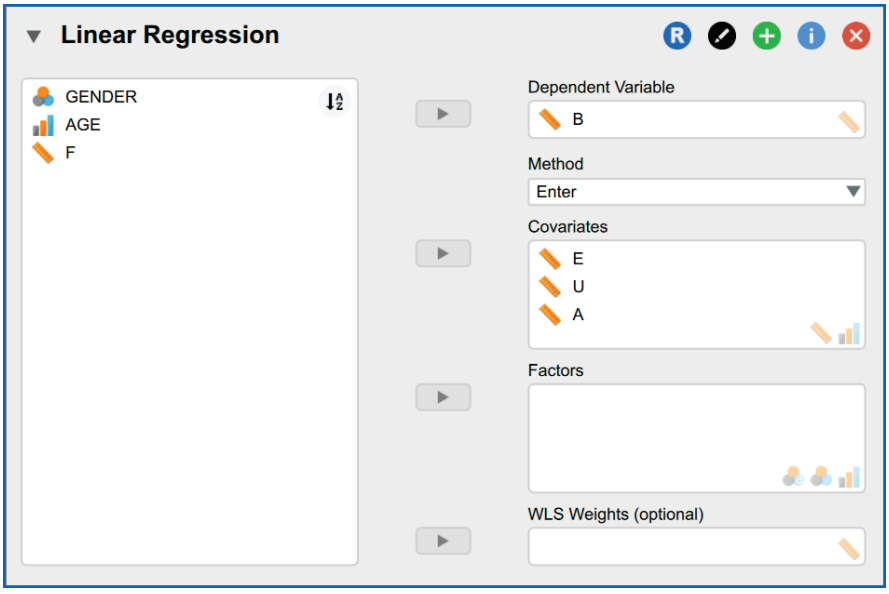


圖 2-4　選擇分析變數——JASP

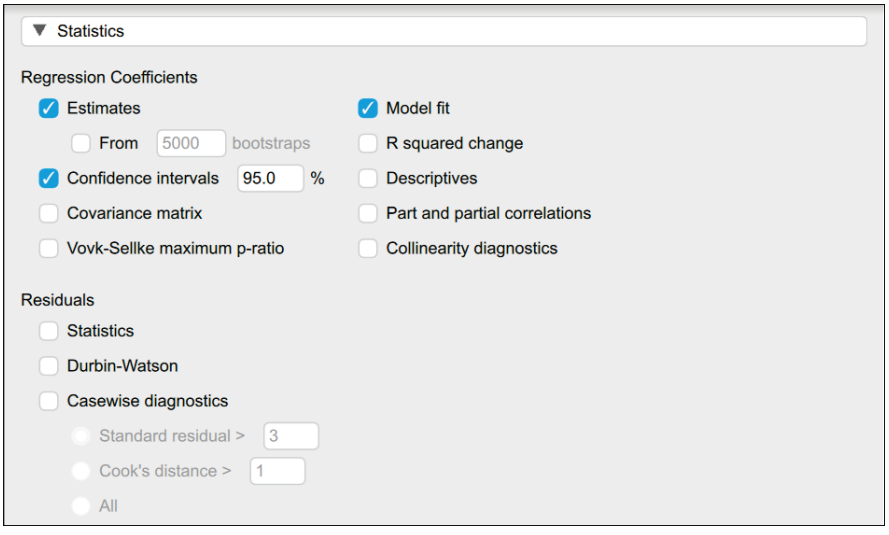


圖 2-5　選擇統計量——JASP

分析後得到 3 個報表。報表 2-1 為模型摘要，$R^2 = .589$，調整後 $R^2 = .585$，RMSE 是迴歸的估計標準誤，為 1.656，代表使用 3 個自變數預測 B 變數，平均會有 1.656 分的差距。此分析的統計假設為 $H_0 : R^2 = 0$ ， $H_1 : R^2 > 0$ 。

報表 2-1　Model Summary - B

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	2.572
H_1	0.768	0.589	0.585	1.656

報表 2-2 是變異數分析摘要表，為整體檢定，$F(3,288) = 137.847$，$p < .001$，3 個預測變數可以聯合預測依變數 B。

報表 2-2　ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1134.448	3	378.149	137.847	2.188×10^{-55}
	Residual	790.059	288	2.743		
	Total	1924.507	291			

Note. The intercept model is omitted, as no meaningful information can be shown.

報表 2-3 是迴歸係數，也包含個別變數的檢定。其中，Unstandardized 一欄是未標準化係數，它們除以 Standard Error（標準誤）之後即為 *t* 值。如果 *t* 的絕對值大於 1.968〔使用 qt(1-.05/2, 288) 計算而得〕，*p* 值小於 .05，或是 95% CI（信賴區間）不含 0，就表示迴歸係數顯著不為 0。Standardized 一欄是標準化係數，分別為 0.125、0.326、0.447。

報表 2-3　Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_0	(Intercept)	12.459	0.150		82.786	2.640×10^{-204}	12.163	12.755
H_1	(Intercept)	0.775	0.588		1.318	0.188	-0.382	1.932
	E	0.117	0.042	0.125	2.783	0.006	0.034	0.199
	U	0.370	0.055	0.326	6.671	1.300×10^{-10}	0.261	0.479
	A	0.506	0.056	0.447	9.046	2.258×10^{-17}	0.396	0.616

2.9　使用 jamovi 分析

在 jamovi 中，讀入 reg.csv 資料檔後，在 Regression（迴歸分析）中選擇 Linear Regression（線性迴歸）。接著，將 B 變數選至 Dependent Variable（依變數）中，E、U、A 三個變數選至 Covariates（共變量）中（圖 2-6）。

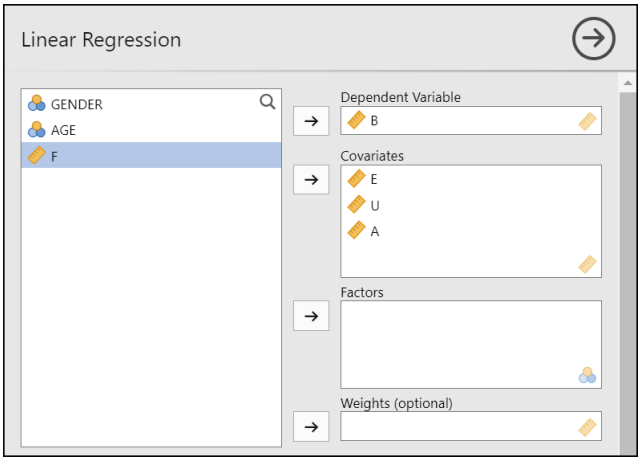


圖 2-6　選擇分析變數——jamovi

在 Model Fit（模型適配）中勾選需要的 Fit Measures（適配量數）及 Overall Model Test（整體模型檢定）（圖 2-7）。

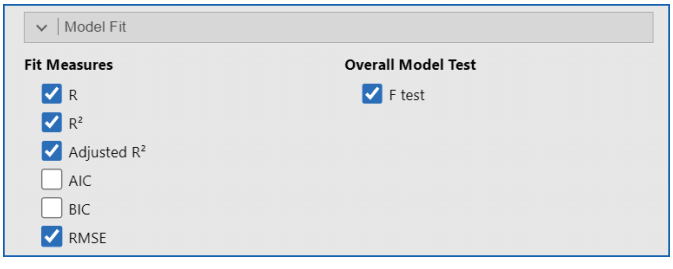


圖 2-7　選擇適配量數——jamovi

在 Model Coefficient（模型係數）中，勾選 Omnibus Test（綜合檢定）、Estimate（估計值）的 Confidence interval（信賴區間），及 Standardized Estimate（標準化估計值）（圖 2-8）。

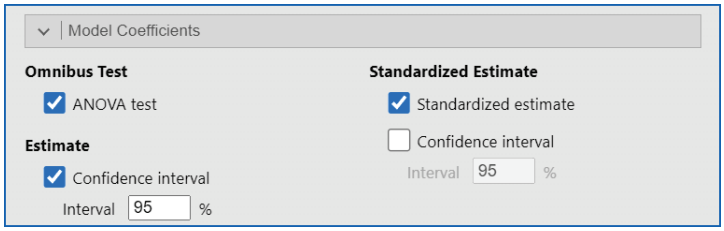


圖 2-8　選擇模型係數——jamovi

分析後得到 3 個報表，其中 2 個與 JASP 相似，此處只呈現 Omnibus ANOVA Test（綜合的變異數分析檢定）。Sum of Squares（SS）一欄是移除某一變數後，*SS* 的改變量。例如：E 變數的 *SS* 是以三個變數同時進入後的迴歸 *SS*（為 1134.448，見報表 2-2）減去只含 U、A 二變數的迴歸 *SS*（為 1113.196，另外分析而得，不呈現報表），1134.448 − 1113.196 = 21.252。Mean Square（MS）為 *SS* / *df*，*F* 為每個變數的 *MS* 除以 Residuals（殘差）的 *MS*。當 *df* = 1 時，$F = t^2$，因此，此處的 *F* 會等於報表 2-3 中 *t* 的平方，而兩個報表中的 *p* 值相同，都小於 .01。

報表 2-4　Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
E	21.252	1	21.252	7.747	0.006
U	122.076	1	122.076	44.500	< .001
A	224.493	1	224.493	81.834	< .001
Residuals	790.059	288	2.743		

Note. Type 3 sum of squares

2.10　分析結論

以對智慧型手機的認知易用性（E）、認知有用性（U）、使用態度（A）對行為意圖（B）進行多元迴歸分析。整體效果是顯著的，$F(3, 288) = 137.85$，$p < .001$，$R^2 = .59$，$\hat{R}^2 = .59$。個別變數分析，三個自變數均可顯著預測依變數，原始迴歸係數分別為 0.12、0.37、0.51，標準化係數分別為 0.12、0.33、0.45，*p* 均小於 0.01。