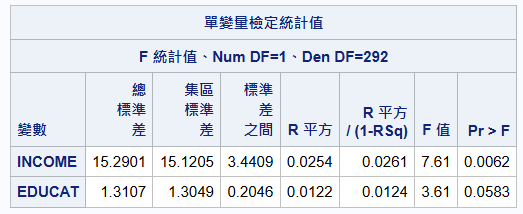
蔡宜諠 711378912 題目：#8.7

[ Use Equal Priors ]

( a ) 利用判別分析由INCOME與EDUCAT評估某人是否有抑鬱症 ( y = CASES)

* Describe univariate test of equal means results



INCOME 的檢定結果：P value 為 0.0052，小於顯著水平 0.05，說明 INCOME 在兩組之間有顯著差異。

EDUCAT 的檢定結果：P value為 0.0583，大於顯著水平 0.05，說明 EDUCAT 在兩組之間沒有足夠證據證明兩組之間有顯著差異。

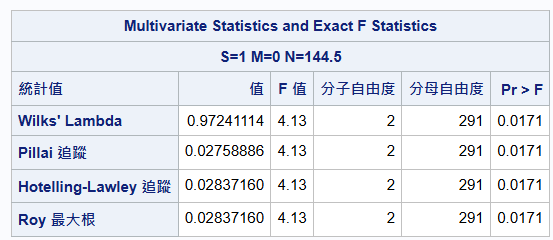
* Describe the practical significance for the discriminant function

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

從此圖可以看出INCOME是更重要的變數，對區分CASES的影響更大。

* Describe the result of multivariate test of equal means for two variables



Wilks' Lambda P value 為 0.0171，小於顯著水平 0.05。這表示 INCOME 和 EDUCAT 的組合在兩組之間有顯著差異。

* Describe the relative importance of variables and existence of collinearity problem

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

從此圖可以看出INCOME 的貢獻更大。這些值沒有接近 1，說明 INCOME 與 EDUCAT 並沒有高度共線性（即共線性問題不明顯）。

* Classification Result

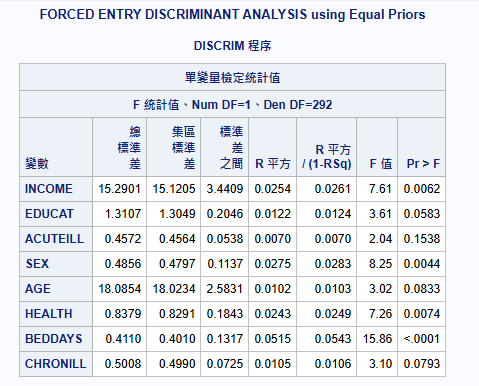
一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

從此圖可以看出CASES = 0，正確分類 112 個，錯誤分類 132 個，正確率為 45.9%；CASES = 1，正確分類 35 個，錯誤分類 15 個，正確率為 70%。整體分類正確率為 ( 112+35 ) / 294= 50%，說明模型對 CASES = 1 的分類表現較好。

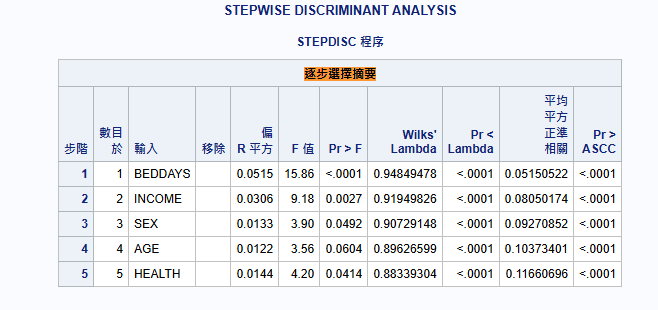
( b ) 增加自變數帶入stepwise

* + Describe univariate test of equal means results.

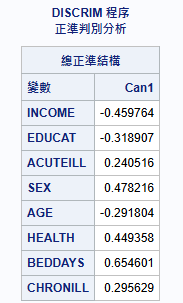


這裡可以看到EDUCAT、ACUTEILL、AGE、CHRONILL的p-value 大於0.05表示沒有足夠證據證明這幾個變數在 CASES = 0 & 1 之間有顯著差異。其餘p-value小於0.05表示這幾個變在CASES = 0 & 1 之間有顯著差異。

* + Describe the practical significance for the discriminant function



BEDDAYS 是第一個被選入模型的變數，具有最高的 R 平方值（0.0515），表明它對判別函數的貢獻最大。INCOME 和 SEX 也顯示出較大的 F 值和顯著的 p-value，表示其對判別函數的重要性。



在總正準結構中，BEDDAYS 的標準化係數最大（0.654601），其次是 HEALTH（0.449358）和 SEX（0.478216），顯示它們對判別函數的貢獻較大。

* + Describe the result of multivariate test of equal means for two variable.

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

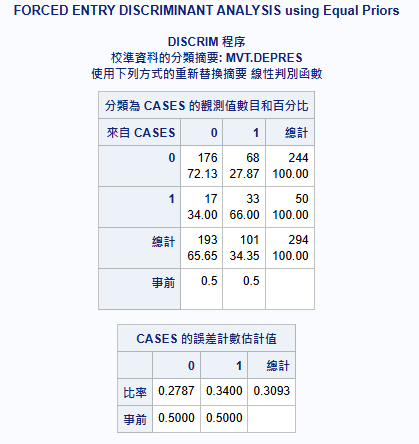
Wilks' Lambda p-value = <0.0001，小於 0.05 表示多變量模型在 CASES = 0 & 1 之間有顯著差異。其他統計量也顯示有顯著。

* + Describe the relative importance of variables and existence of collinearity problem



BEDDAYS 的標準化係數為 0.5635，是最重要的變數。HEALTH 和 SEX 的標準化係數分別為 0.3076 和 0.3245，也有較大的貢獻。變數之間的相關性較低，無顯著共線性問題。

* + Classification Result



從此圖可以看出CASES = 0 的正確分類數為 176，正確率為 70.13%；CASES = 1 的正確分類數為 33，正確率為 66.00%。模型在CASES = 0的時候表現得比較好。整體分類正確率為 71.09%（( 176 + 33 ) / 294）。

( c ) 解釋判別分析的結果 比較 ( a ) vs ( b )

模型 ( a )結果對於CASES = 1的分類表現比較好，總體分類的正確率為50%，模型 ( b ) 結果對於CASES = 0的分類表現比較好，總體分類的正確率為71.09%，模型 ( b ) 較 模型 ( a ) 總體的分類正確率好

[ Use Unequal Priors ]

( a ) 利用判別分析由INCOME與EDUCAT評估某人是否有抑鬱症 ( y = CASES)

* Describe univariate test of equal means results

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

INCOME 的檢定結果：P value 為 0.0052，小於顯著水平 0.05，說明 INCOME 在兩組之間有顯著差異。

EDUCAT 的檢定結果：P value為 0.0583，大於顯著水平 0.05，說明 EDUCAT 在兩組之間沒有足夠證據證明兩組之間有顯著差異。

* Describe the practical significance for the discriminant function

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

從此圖可以看出INCOME是更重要的變數，對區分CASES的影響更大。

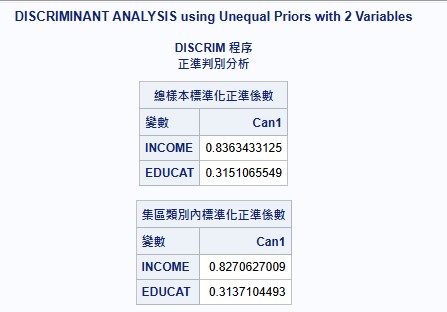
* Describe the result of multivariate test of equal means for two variables

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

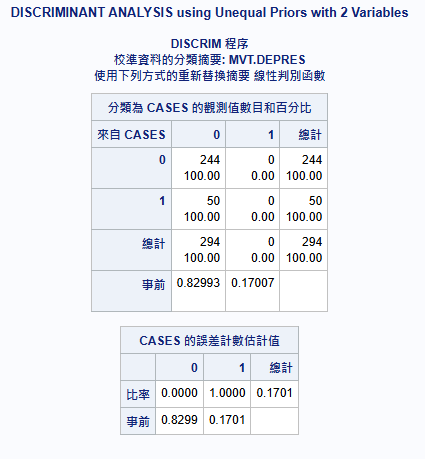
Wilks' Lambda P value 為 0.0171，小於顯著水平 0.05。這表示 INCOME 和 EDUCAT 的組合在兩組之間有顯著差異。

* Describe the relative importance of variables and existence of collinearity problem



從此圖可以看出INCOME 的貢獻更大。這些值沒有接近 1，說明 INCOME 與 EDUCAT 並沒有高度共線性（即共線性問題不明顯）。

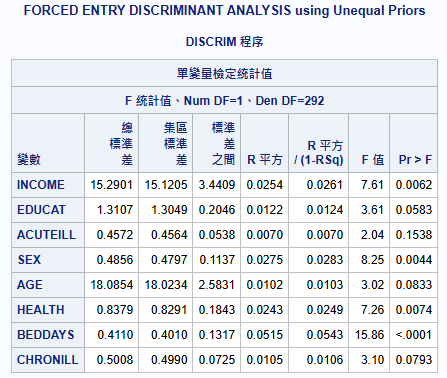
* Classification Result



從此圖可以看出CASES = 0，正確分類 244 個，錯誤分類 0 個，正確率為 100%；CASES = 1，正確分類 0 個，錯誤分類 50 個，正確率為 0%。整體分類正確率為 ( 244+0 ) / 294= 82.99%，正確率很高但該模型把所有都歸類為CASES = 0，因此正確率並不能表示甚麼。

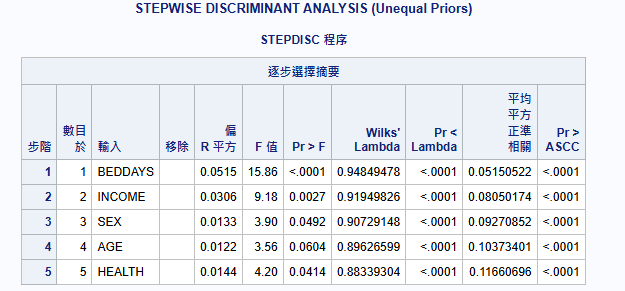
( b ) 增加自變數帶入stepwise

* + Describe univariate test of equal means results.

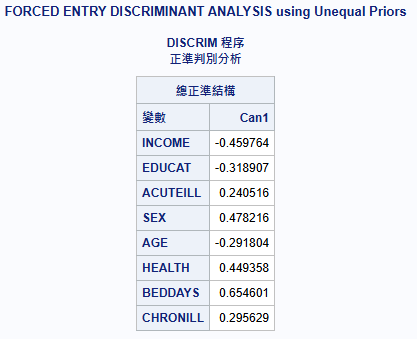


這裡可以看到EDUCAT、ACUTEILL、AGE、CHRONILL的p-value 大於0.05表示沒有足夠證據證明這幾個變數在 CASES = 0 & 1 之間有顯著差異。其餘p-value小於0.05表示這幾個變在CASES = 0 & 1 之間有顯著差異。

* + Describe the practical significance for the discriminant function

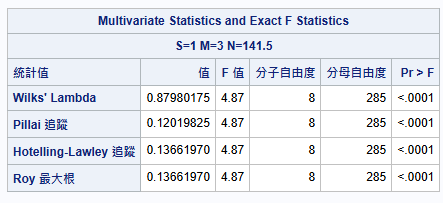


BEDDAYS 是第一個被選入模型的變數，具有最高的 R 平方值（0.0515），表明它對判別函數的貢獻最大。INCOME 和 SEX 也顯示出較大的 F 值和顯著的 p-value，表示其對判別函數的重要性。



在總正準結構中，BEDDAYS 的標準化係數最大（0.654601），其次是 HEALTH（0.449358）和 SEX（0.478216），顯示它們對判別函數的貢獻較大。

* + Describe the result of multivariate test of equal means for two variable.



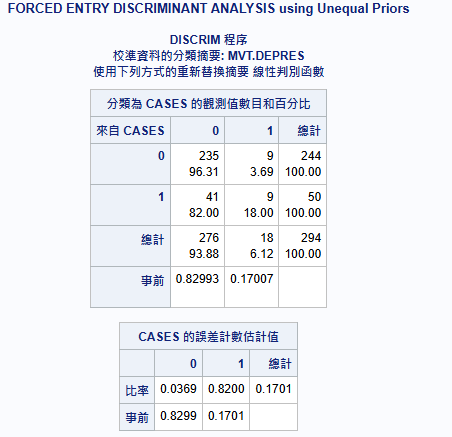
Wilks' Lambda p-value = <0.0001，小於 0.05 表示多變量模型在 CASES = 0 & 1 之間有顯著差異。其他統計量也顯示有顯著。

* + Describe the relative importance of variables and existence of collinearity problem



BEDDAYS 的標準化係數為 0.5635，是最重要的變數。HEALTH 和 SEX 的標準化係數分別為 0.3076 和 0.3245，也有較大的貢獻。變數之間的相關性較低，無顯著共線性問題。

* + Classification Result



從此圖可以看出CASES = 0 的正確分類數為 235，正確率為 96.31%；CASES = 1 的正確分類數為 9，正確率為 18.00%。模型在CASES = 0的時候表現得比較好。整體分類正確率為 82.99%（( 235 + 9 ) / 294）。

( c ) 解釋判別分析的結果 比較 ( a ) vs ( b )

模型 ( a ) 以及 模型 ( b ) 結果對於CASES = 0的分類表現都比較好，總體分類的正確率皆為82.99%，但模型 ( a )把所有的都判定為CASES = 0，因此模型 ( b ) 比較好。

[ Comments on the best model among the aforementioned 4 model ]

Equal Priors ：

1. 模型( a )結果對於CASES = 1的分類表現比較好，總體分類的正確率50%
2. 模型( b )結果對於CASES = 0的分類表現比較好，總體分類的正確率71.09%
3. 模型 (b) 在總體分類表現上優於 模型 (a)，因此更具實用性。

Unequal Priors：

模型 ( a ) 以及 模型 ( b ) 結果對於CASES = 0的分類表現都比較好，總體分類的正確率皆為82.99%。然而，模型 ( a ) 將所有樣本都判定為 CASES = 0，缺乏對 CASES = 1 的區分能力。因此，模型 ( b ) 雖然總體分類正確率相同，但能更好地平衡對兩個類別的區分能力，較 模型 ( a ) 更為優秀。

在 Equal Priors 和 Unequal Priors 的比較中，Unequal Priors 的模型 (b) 是表現最佳的模型，不僅總體正確率高，還能更好地平衡 CASES = 0 和 CASES = 1 的分類表現。