**Q1**: Perform a complete **Principal Components Analysis** for this data and interpret the result. Note: The number of PCs must be determined by a formal statistical hypothesis test, while the relationships among objects and variables can be interpreted by using a 2D plot.

先看看資料的樣式

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | Agr | Min | Man | PS | Con | SI | Fin | SPS | TC |
| Belgium | 3.3 | 0.9 | 27.6 | 0.9 | 8.2 | 19.1 | 6.2 | 26.6 | 7.2 |
| Denmark | 9.2 | 0.1 | 21.8 | 0.6 | 8.3 | 14.6 | 6.5 | 32.2 | 7.1 |
| France | 10.8 | 0.8 | 27.5 | 0.9 | 8.9 | 16.8 | 6.0 | 22.6 | 5.7 |
| W. Germany | 6.7 | 1.3 | 35.8 | 0.9 | 7.3 | 14.4 | 5.0 | 22.3 | .6.1 |

總共變數為10個變數，26筆樣本資料，我們要用除了國家名之外的九個變數去做組成分分析(**Principal Components Analysis, PCA**)。9個變數皆為數值型態，但每個變數的範圍都不大相等，因此，再進行分析前要先標準化，藉此消除掉變異程度過大的變數影響分析。

下圖為相關係數圖（解釋）（缺圖）

接下來，我們透過Ｒ計算PCA，得到下列幾種組合以及他們解釋變異的程度

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 | Comp.8 | Comp.9 |
| Agr | -0.5238 | -0.0536 | 0.0487 | -0.0288 | 0.2127 | -0.1533 | 0.0213 | 0.0079 | 0.8064 |
| Min | -0.0013 | -0.6178 | -0.2011 | -0.0641 | -0.1637 | 0.1006 | -0.7257 | 0.0884 | 0.0486 |
| Man | 0.3475 | -0.3551 | -0.1505 | 0.3461 | -0.385 | 0.2882 | 0.4794 | 0.1258 | 0.366 |
| PS | 0.2557 | -0.2611 | -0.5611 | -0.3933 | 0.2952 | -0.3573 | 0.2556 | -0.3412 | 0.0194 |
| Con | 0.3252 | -0.0513 | 0.1533 | 0.6683 | 0.4716 | -0.1304 | -0.2207 | -0.3557 | 0.0826 |
| SI | 0.3789 | 0.3502 | -0.1151 | 0.0502 | -0.2836 | -0.6148 | -0.2294 | 0.3875 | 0.2383 |
| Fin | 0.0744 | 0.4537 | -0.5874 | 0.0516 | 0.2796 | 0.5256 | -0.1875 | 0.1743 | 0.1452 |
| SPS | 0.3874 | 0.2215 | 0.3119 | -0.4122 | -0.2204 | 0.2629 | -0.1913 | -0.5062 | 0.3509 |
| TC | 0.3668 | -0.2026 | 0.3751 | -0.3144 | 0.5129 | 0.124 | 0.0682 | 0.5446 | 0.0721 |
| Standard deviation | 1.86739 | 1.45951 | 1.04831 | 0.99724 | 0.73703 | 0.61922 | 0.47514 | 0.36985 | 0.00675 |
| Proportion of Variance | 0.38746 | 0.23669 | 0.12211 | 0.1105 | 0.06036 | 0.0426 | 0.02508 | 0.0152 | 5.1E-06 |
| Cumulative Proportion | 0.38746 | 0.62415 | 0.74625 | 0.85675 | 0.91711 | 0.95971 | 0.9848 | 0.99999 | 1 |

由上表可知，城市跑出9個components組合，依照每一個成分標準差是否大於1為判斷標準則有component1~component3要保留; 若是依照累積解釋變異程度超過70%的判斷標準，則有component1~component3便能解釋70％以上的變異程度。最後我們依照檢定每一個components是否不顯著(去判斷要使用哪些components。透過permutation test去計算每一個components的p value，在給定顯著水準為0.05下，只有component1以及component2拒絕，

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 | Comp.8 | Comp.9 |
| Pvalue | 0.000 | 0.000 | 0.990 | 0.915 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

而用這個兩個成分所能解釋的總變異數有62%，為可接受範圍，因此我們使用component1以及component2為PCA後的結果，從維度為9(Agr, Min, … ,TC)降維度至2維度（component1以及component2）

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| component | Agr | Min | Man | PS | Con | SI | Fin | SPS | TC |
| 1 | -0.5238 |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |