

Applied multivariate analysis-HW8

ID : 111024517 Name : 鄭家豪

Problem 1

All the cumulative variance explained

展示指定方法的所有組合的 cumulative variance explained(3 factors):

- PC method(rounding to 4 digits):

	Varimax			Quartimax		
	F1	F2	F3	F1	F2	F3
SS loadings	3.0746	2.7559	2.4977	3.1359	2.6720	2.5203
Proportion Var	0.2196	0.1968	0.1784	0.2240	0.1909	0.1800
Cumulative Var	0.2196	0.4165	0.5949	0.2240	0.4149	0.5949

- PF method(rounding to 4 digits):

	Varimax			Quartimax		
	F1	F2	F3	F1	F2	F3
SS loadings	2.7066	2.2964	2.0874	2.8109	2.1688	2.1106
Proportion Var	0.1933	0.1640	0.1491	0.2008	0.1549	0.1508
Cumulative Var	0.1933	0.3574	0.5065	0.2008	0.3557	0.5065

- MLE method(rounding to 4 digits):

	Varimax			Quartimax		
	F1	F2	F3	F1	F2	F3
SS loadings	2.6849	2.2940	2.1075	2.8068	2.1611	2.1185
Proportion Var	0.1918	0.1639	0.1505	0.2005	0.1544	0.1513
Cumulative Var	0.1918	0.3556	0.5062	0.2005	0.3549	0.5062

(a)

這裡我選取的是 PC method with varimax rotation，以下是其 loadings(空白處代表對應的 loading 值過小):

Loadings:

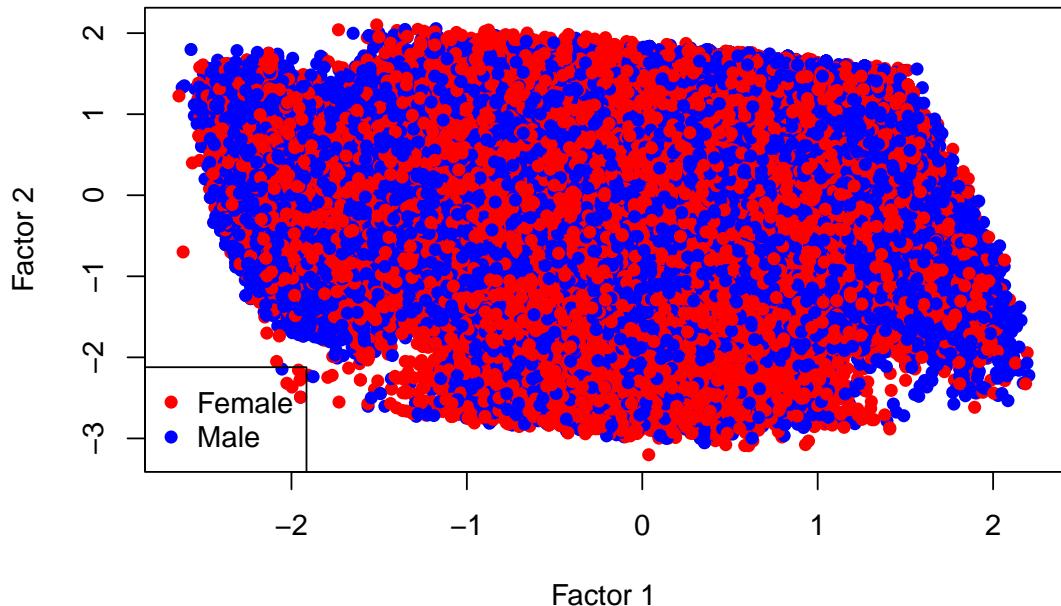
	RC1	RC3	RC2
Inflight.wifi.service	0.187	0.127	0.802
Departure.Arrival.time.convenient			0.666
Ease.of.Online.booking			0.887
Gate.location			0.671
Food.and.drink		0.827	
Online.boarding	0.485	0.108	0.407
Seat.comfort		0.849	
Inflight.entertainment	0.760	0.454	
On.board.service		0.782	
Leg.room.service		0.611	
Baggage.handling		0.821	
Checkin.service	0.165	0.378	
Inflight.service		0.837	
Cleanliness		0.880	

	RC1	RC3	RC2
SS loadings	3.075	2.756	2.498
Proportion Var	0.220	0.197	0.178
Cumulative Var	0.220	0.416	0.595

- Factors(RC1,RC3,RC2) 解釋:

1. RC1: loading 值較大 (以大於 0.6 為標準) 的變數，分別是 Food.and.drink、Seat.comfort、Inflight.entertainment、Cleanliness，這些變數有機上搭乘體驗的涵義在，代表 RC1 是類似這種意義的 factor。
2. RC3: loading 值較大的變數，分別是 On.board.service、Leg.room.service、Baggage.handling、Inflight.service，這些都是具有機上服務的變數，代表 RC3 是類似這種意義的 factor。
3. RC2: loading 值較大的變數，分別是 Inflight.wifi.service、Departure.Arrival.time.convenient、Ease.of.Online.booking、Gate.location，這些變數有個共同性: 便利性。代表 RC2 很有可能是便利相關意義的 factor。

(b)



從以上的圖，可能有幾個觀點：

1. 均衡性: Factor scores 分佈均勻，可能表示樣本中的個體在這兩個 factor(RC1,RC3) 上的表現相對一致，以性別的角度來看，沒有明顯的差異。
2. 多樣性: 反映了樣本中的多樣性或變化程度，即用男女來區分不太能看出明確的分群，可能有其他的特徵在其中。

Problem 2

By eigen-decomposition, S is a matrix in terms of $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$

$$S = \sum_{i=1}^p \hat{\lambda}_i \hat{e}_i \hat{e}_i^T$$

$$\text{We know } \hat{L} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix},$$

$$\text{then } S - \hat{L} \hat{L}^T = \hat{\lambda}_{m+1} \hat{e}_{m+1} \hat{e}_{m+1}^T + \dots + \hat{\lambda}_p \hat{e}_p \hat{e}_p^T$$

$$\stackrel{\text{say}}{=} \hat{P}_{(2)} \hat{\Lambda}_{(2)} \hat{P}_{(2)}^T, \text{ where } \hat{P}_{(2)} = [\hat{e}_{m+1} \dots \hat{e}_p], \hat{\Lambda}_{(2)} = \text{diag}[\hat{\lambda}_{m+1} \dots \hat{\lambda}_p]$$

Thus, the sum of all the entries of $S - \hat{L} \hat{L}^T$ is equivalent to

$$\text{tr}[(S - \hat{L} \hat{L}^T)(S - \hat{L} \hat{L}^T)^T] = \text{tr}[\hat{P}_{(2)} \hat{\Lambda}_{(2)} \hat{P}_{(2)}^T \hat{P}_{(2)} \hat{\Lambda}_{(2)} \hat{P}_{(2)}^T]$$

$$= \text{tr}[\hat{P}_{(2)} \hat{\Lambda}_{(2)} \hat{\Lambda}_{(2)}^T \hat{P}_{(2)}^T] = \text{tr}[\hat{\Lambda}_{(2)} \hat{\Lambda}_{(2)}^T \hat{P}_{(2)}^T \hat{P}_{(2)}]$$

$$= \text{tr}[\hat{\Lambda}_{(2)} \hat{\Lambda}_{(2)}^T] = \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2$$

And, \because the diagonal elements of $S - (\hat{L} \hat{L}^T + \hat{P})$ are zero.

\therefore (sum of squared entries of $S - (\hat{L} \hat{L}^T + \hat{P})$)

$$\leq (\text{sum of squared entries of } S - \hat{L} \hat{L}^T) = \sum_{j=m+1}^p \hat{\lambda}_j^2 \quad \#$$