多變量分析 HW3

劉昱維, 吳冠瑋, 廖永賦, 謝靖惟, 黃奎鈞

Q1: 9.12

$$\mathbf{S_n} = \frac{23}{24}\mathbf{S} = \begin{pmatrix} 0.01061 & 0.00768 & 0.00782 \\ 0.00768 & 0.00615 & 0.00575 \\ 0.00782 & 0.00575 & 0.00649 \end{pmatrix}$$

(a) Specific Variances

$\mathbf{S_n} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}$, where $diag(\mathbf{S_n}) = diag(\hat{\mathbf{L}}\hat{\mathbf{L}}^T) + diag(\hat{\mathbf{\Psi}})$
Hence, $\hat{\mathbf{\Psi}} = diag(\hat{\mathbf{\Psi}}) = diag(\mathbf{S_n} - \hat{\mathbf{L}}\hat{\mathbf{L}}^T)$

$$\hat{\mathbf{\Psi}} = \begin{pmatrix} 0.000166 & 0.000000 & 0.000000 \\ 0.000000 & 0.000495 & 0.000000 \\ 0.000000 & 0.000000 & 0.000639 \end{pmatrix}$$

(b) Communalities

$$\sigma_{ii} = \ell_{i1}^2 + \ell_{i2}^2 + \ \dots \ + \ell_{im}^2 + \psi_i$$
$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \ \dots \ + \ell_{im}^2 = \ell_{i1}^2$$
$$h_1^2 = 0.0104,\ h_2^2 = 0.0057,\ h_3^2 = 0.0059$$

(c) Proportion of variance explained by the factor

$$\frac{s_{11} + s_{22} + \ \dots \ + s_{pp}}{h_i^2} = 0.9441$$

(d) Residual Matrix

$$\mathbf{S_n} - \hat{\mathbf{L}}\hat{\mathbf{L}}^T - \hat{\mathbf{\Psi}} = \begin{pmatrix} 0.000000 & -0.000166 & -0.000164 \\ -0.000495 & 0.000000 & -0.000493 \\ -0.000637 & -0.000637 & 0.000000 \end{pmatrix}$$

Q2: 9.32

(a) S PC

	RC1	RC2	RC3
YrHgt	0.566	0.734	0.87
FtFrBody	**31.179**	**83.935**	**24.022**
PrctFFB	0.601	1.488	2.742
Frame	0.296	0.39	0.415
BkFat	0.014	-0.006	-0.055
SaleHt	1.021	0.934	0.82
SaleWt	**122.551**	**39.068**	**-17.48**

(b) S ML

	ML1	ML2	ML3
YrHgt	-0.001	0.581	1.629
FtFrBody	**21.802**	**84.458**	**31.383**
PrctFFB	-0.253	2.155	1.056
Frame	0.019	0.307	0.817
BkFat	0.033	-0.015	-0.027
SaleHt	0.498	0.841	1.532
SaleWt	**119.136**	**33.925**	**38.8**

(c) R PC

	RC1	RC3	RC2
YrHgt	**0.941**	0.27	-0.082
FtFrBody	0.447	**0.794**	0.205
PrctFFB	0.262	**0.859**	-0.295
Frame	**0.938**	0.219	-0.028
BkFat	-0.231	-0.339	**0.812**
SaleHt	**0.833**	0.419	0.109
SaleWt	0.352	0.43	**0.722**

(d) R ML

	ML1	ML2	ML3
YrHgt	**0.941**	0.286	0.164
FtFrBody	0.414	**0.505**	**0.553**
PrctFFB	0.231	**0.947**	0.212
Frame	**0.891**	0.251	0.18
BkFat	-0.256	**-0.514**	0.273
SaleHt	**0.755**	0.269	0.434
SaleWt	0.253	-0.05	**0.879**

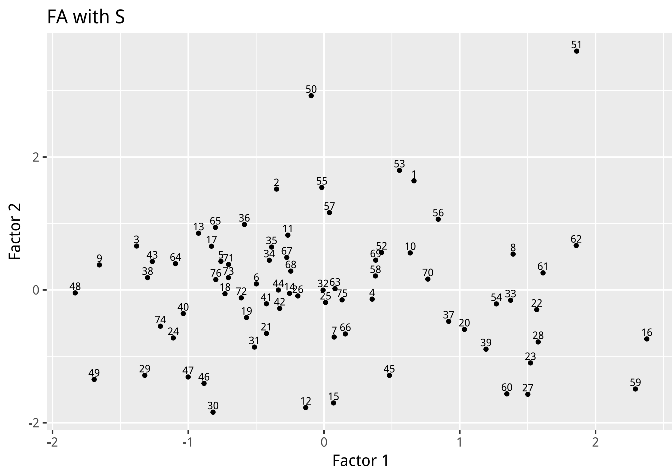
(e) Compare S & R

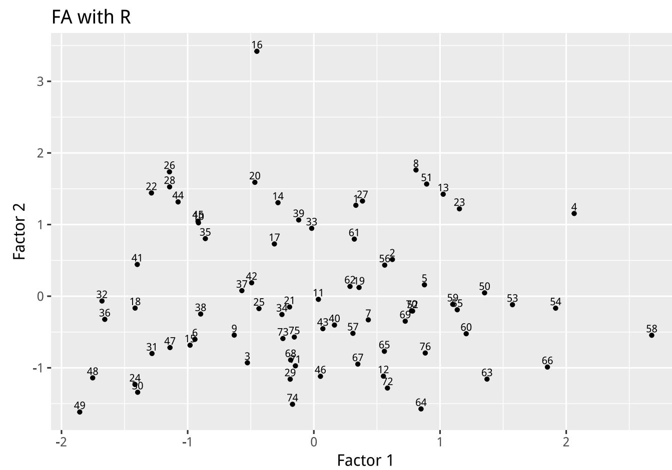
The results obtained using covariance matrices are hard to interpret here. Since there are three kinds of units used: **pound**, **inch**, and **self-defined scales(1-8)**, the factor loadings on some variables are very large and others small.

Interpretation of the factors is straightforward using results obtained from the covariance matrix.
By (c) and (d), 1. factor 1 has larger loadings on **YrHgt**, **Frame**, and **SaleHt**, which are all related to the **height**s of the bulls. 2. Factor 2 has large loadings on **FtFrBody** and **PrctFFB** and is negatively related to **BkFat** in both methods(PC & ML) of factor analysis. It might be called a **lean** factor. 3. In both PC & ML method of factor analysis, the loadings of factor 3 is large on **SaleWt**. The loading is also large on **BkFat** from the PC method, and the loading is medium on **FtFrBody** from the ML method. Factor 3 might be related to the **weight**s of the bulls.

(f) scatter plots of factor2 vs factor1 in (a) and (c)

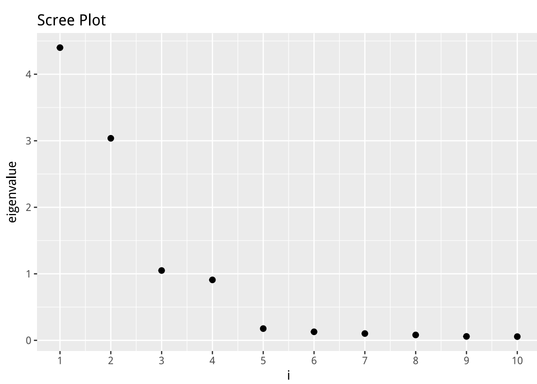
The point, **51**, in the scatter plot on the left and the point, **16**, in the scatter plot on the right seem to be outliers.





Q3

(a) Appropriate number of factors



By the unity criterion we use 4 factors[1].

	RC4	RC1	RC2	RC3
X1	0.131	-0.215	**0.905**	-0.044
X2	**0.953**	0.212	0.05	-0.023
X3	0.293	**0.899**	-0.165	-0.043
X4	0.027	-0.091	**0.955**	0.054
X5	**0.933**	0.261	0.111	-0.076
X6	-0.083	-0.043	0.041	**0.994**
X7	0.228	**0.918**	-0.23	-0.028
X8	0.057	-0.224	**0.935**	0.051
X9	**0.935**	0.276	0.073	-0.048
X10	0.273	**0.896**	-0.199	-0.016

[1] Although **factor 4** is less than 1, it’s eigenvalue is very close to **factor 3**, hence we retain **factor 4**.

(b) Label the Factors

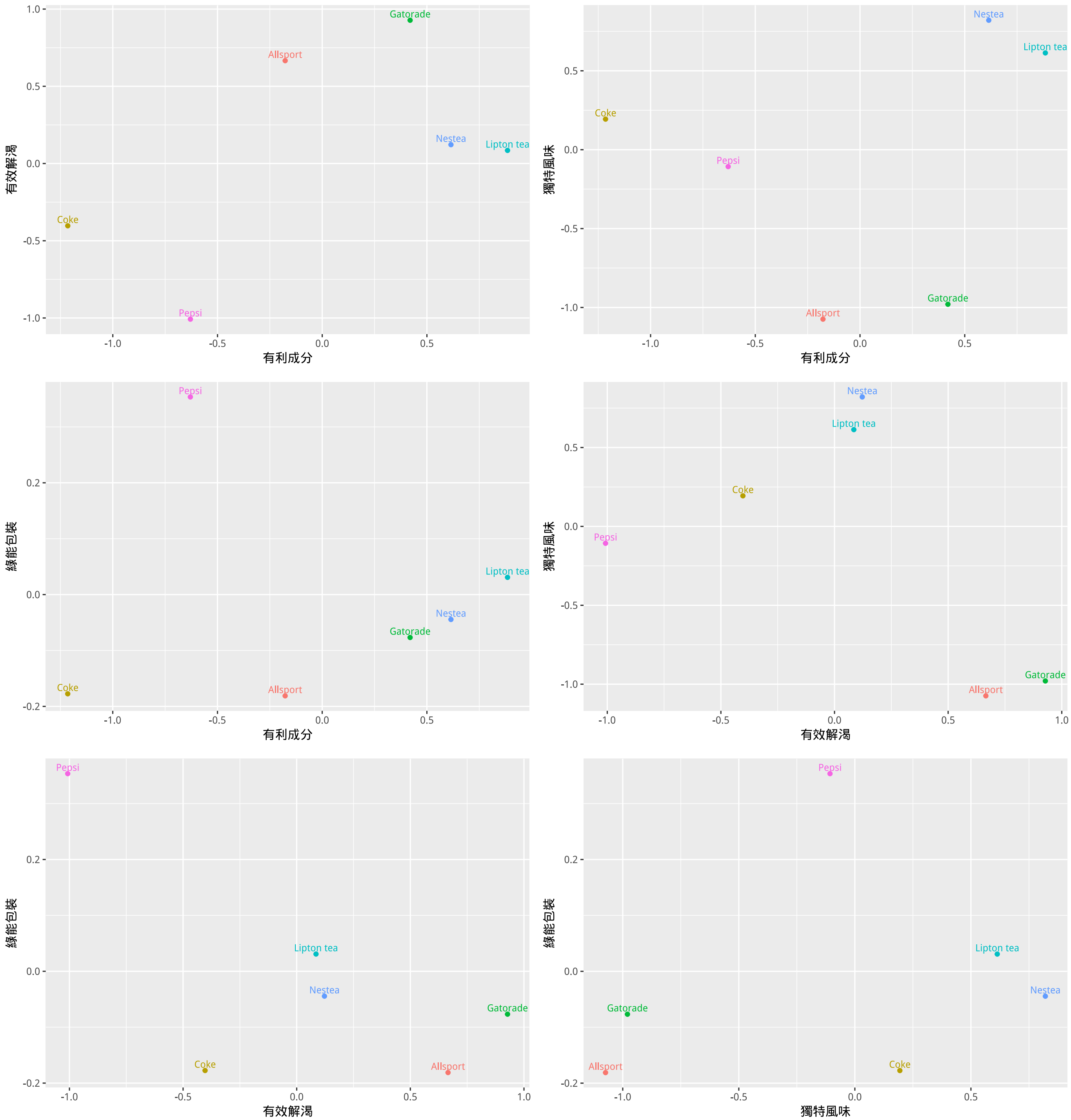
- RC4: 有利成分
  此因素在第 2, 5, 9 題有較高的因素負荷量。這些題目似乎皆與健康有關。飲料能提供營養或避免添加不健康的物質，似與此因素有關。
- RC1: 有效解渴此因素在第 3, 7, 10 題有較高的因素負荷量，這些題目皆與飲料是否**解渴**有關。
- RC2: 獨特風味
  此因素在第 1, 4, 8 題有較高的因素負荷量，這些題目皆強調飲料的**口感**。
- RC3: 綠能包裝
  此因素僅在第 6 題有較高的因素負荷量。第 6 題是一獨特的題目，所有題目中僅其強調**環境因素**。因此一個因素僅在第 6 題有較高的因素負荷量顯得相當合理。

(c) Average factor scores of each brand

Average factor scores

	有利成分	有效解渴	獨特風味	綠能包裝
Pepsi	-0.629	-1.008	-0.107	0.354
Coke	-1.215	-0.403	0.194	-0.177
Gatorade	0.420	0.928	-0.980	-0.077
Allsport	-0.177	0.666	-1.074	-0.181
Lipton tea	0.885	0.085	0.613	0.031
Nestea	0.615	0.122	0.821	-0.044

(d) Interpret the positions of the six brands



在不考慮**綠能包裝**下，可以從各散布圖中看出 3 種飲料類型：

1. Pepsi, Coke: 碳酸飲料
2. Gatorade, Allsport: 運動飲料
3. Lipton tea, Nestea: 茶飲

相同類型的飲料在散布圖中，通常位置會較為接近。例如，茶飲較為健康且有獨特風味，因此出現在出現在右上圖的右上方；碳酸飲料含糖量高，既不健康又不解渴，因此在左上圖中，出現在左下方；運動飲料則在**有效解渴**上，出現在類似的位置 。