Chapter 3. Factor analysis

1. Dimension reduction

- In marketing research an attitude may be measured from different directions using different variables.
- One may group similar variables into factors by creating a new factor or variable based on more than one question and such technique is termed analysis of interdependence.

Example

- A survey of 24 customers on their percentages of expenditure on 9 major categories is conducted
  - o Food X_1, Transport X_2, Living X_3, Communication X_4, Entertainment X_5, Other X_6, Clothing X_7, Education X_8 and Saving X_9

- Many of these attitude- and needs-based expenditures measures similar or interrelated constructs.
  - o Complicate a model and lead to invalid conclusions because some data are over-represented while other under-represented.
  - o Classes like Transport X_2 and Communication X_4 may contain considerable redundancy.
  - o Factors not directly observable but measuring similar constructs or observable variables should be identified

- Variables from questions should be combined to create new variables (factors / components)
  - o Reduce the number of variables to a more manageable set while retain as much information as possible.
  - o Less correlated variables makes statistical analysis like regression easier and facilitates interpretation.
    - ▪ Especially with less correlated variables, parameter estimates will be more stable.

- Dimension reduction
  - o Identify the separate dimensions of the structure of the observed variables
  - o Determine the extent to which each variable is explained by each dimension

- Application of dimension reduction
  - o Summarization
    - ▪ Derives underlying dimensions that, when interpreted and understood, describe the data in a much smaller number of concepts than the original individual variables
  - o Data reduction
    - ▪ Calculating scores for each underlying dimension and substituting them for the original variables
  - o Perceptual map
    - ▪ The reduced dimensions usually reveal the perception of the customers, companies or individuals in a marketing research
    - ▪ Plotting the individuals based on their scores on the perception dimensions produces a perceptual map which facilitates grouping/segmenting/positioning individuals.

2. Factor analysis

- Describe the covariance (or correlation) relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors
  - Variables can be grouped by their correlations or covariances.
  - All variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group.
  - Each group of variables represents a single underlying construct, or factor.

3. Orthogonal factor model

a) Definition

- Observed random vector **X** with p components
  - Mean $\boldsymbol{\mu}$
  - Covariance matrix $\boldsymbol{\Sigma}$

- Assume **X** is linearly dependent upon 2 components
  - $F_1, \ldots, F_m$
    - m unobservable random variables
    - Common factors
  - $\varepsilon_1, \ldots, \varepsilon_p$
    - p additional sources of variation
    - Errors or specific factors

- Definition of factor analysis model

$$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1$$
$$\vdots$$
$$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p$$

  - In matrix form

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon} \qquad (\textit{c.f. } \mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon} \text{ in regression})$$

  - The coefficient ℓ_{ij} is the loading of the ith variable on the jth factor
  - **L** is the matrix of factor loadings

- Assumptions
  - p deviations, $X_i - \mu_i$, expressed in terms of $m + p$ ($> p$) random variables, F_j and ε_i
    - no unique solution
  - Assume
    - $E(\mathbf{F}) = \mathbf{0}$
    - $Cov(\mathbf{F}) = E(\mathbf{F}\mathbf{F}^T) = \mathbf{I}$
      - Unit variances and uncorrelated
    - $E(\boldsymbol{\varepsilon}) = \mathbf{0}$
    - $Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & & & 0 \\ & \psi_2 & & \\ & & \ddots & \\ 0 & & & \psi_p \end{bmatrix}$
    - is a diagonal matrix
    - **F** and $\boldsymbol{\varepsilon}$ are independent, so $\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = \mathbf{0}$

b) Properties

- Covariance structure for orthogonal factor model

- Covariance matrix for $\mathbf{X}$
  - $Cov(\mathbf{X}) = \mathbf{LL}^T + \mathbf{\Psi}$
  - $Var(X_i) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i$
  - $Cov(X_i, X_k) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}$

- Covariance between $\mathbf{X}$ and $\mathbf{F}$
  - $Cov(\mathbf{X}, \mathbf{F}) = \mathbf{L}$
  - $Cov(X_i, F_j) = \ell_{ij}$

- Partition of the variance of $\mathbf{X}$
  - The ith communality
    - Variance of the ith variable contributed by the m common factors is
$$h_i^2 = \ell_{i1}^2 + \cdots + \ell_{im}^2,\ i = 1, \ldots, p$$
    - How much of the variance of x_i can be explained by all the m factors
  - Specific variance (or uniqueness)
    - Variance due to specific factor (specific to X_i)
$$\psi_i = \sigma_{ii} - h_i^2$$
    - How much of the variance of x_i cannot be explained by all the m factors
  - Total variances, out of all p variances, explained by the jth factor
$$\sum_{i=1}^{p} \ell_{ij}^2 = \ell_{1j}^2 + \cdots + \ell_{pj}^2,\ j = 1, \ldots, m$$
    - Proportion of total variance explained
$$\sum_{i=1}^{p} \ell_{ij}^2 \Big/ \sum_{i=1}^{p} \sigma_{ii}$$

4. Estimation

a) Principal component method

- Covariance matrix $\mathbf{\Sigma}$ with eigenvalue-eigenvector pairs (λ_i, $\mathbf{v}_i$) where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$, $\|\mathbf{v}_j\| = 1$
$$\mathbf{\Sigma} = \lambda_1\mathbf{v}_1\mathbf{v}_1^T + \lambda_2\mathbf{v}_2\mathbf{v}_2^T \ldots + \lambda_p\mathbf{v}_p\mathbf{v}_p^T$$
$$= \begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1 & \sqrt{\lambda_2}\mathbf{v}_2 & \cdots & \sqrt{\lambda_p}\mathbf{v}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1^T \\ \sqrt{\lambda_2}\mathbf{v}_2^T \\ \vdots \\ \sqrt{\lambda_p}\mathbf{v}_p^T \end{bmatrix} = \mathbf{LL}^T$$
  - Factor model with as many factor as variables ($m = p$)
  - The jth column of the loadings matrix is $\sqrt{\lambda_j}\,\mathbf{v}_j$
  - The covariance matrix and the specific variances matrix are
$$\mathbf{\Sigma} = \mathbf{LL}^T + \mathbf{0}$$

- Too many factors as a factor model
  - Dimensions not reduced ($m = p$)
  - Not useful

- When the last $p - m$ eigenvalues are small, neglect the contribution of

$$\lambda_{m+1}\mathbf{v}_{m+1}\mathbf{v}_{m+1}^T + \cdots + \lambda_p\mathbf{v}_p\mathbf{v}_p^T$$

  - Reduce to

$$\mathbf{\Sigma} \approx \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$$

    - where

$$\mathbf{L} = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1 & \sqrt{\lambda_2}\mathbf{v}_2 & \cdots & \sqrt{\lambda_m}\mathbf{v}_m \end{bmatrix}$$
$$\mathbf{\Psi} = diag\left(\mathbf{\Sigma} - \mathbf{L}\mathbf{L}^T\right)$$

- Apply to a data set
  - Use the sample covariance matrix, **S**
  - When the units of the variables are not commensurate, applied to the standardized data, i.e. using the sample correlation matrix **R**

Example

- Consider covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- For the eigenvalues

$$|\mathbf{\Sigma} - \lambda\mathbf{I}| = \left\| \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\| = 0$$

$$\left\| \begin{matrix} 1-\lambda & -2 & 0 \\ -2 & 5-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{matrix} \right\| = 0$$

$$[(1-\lambda)(5-\lambda) - 4](2-\lambda) = 0$$

  - The eigenvalues are $\lambda_1 = 5.83$, $\lambda_2 = 2.00$, and $\lambda_3 = 0.17$.
- The eigenvectors are

$$\mathbf{\Sigma}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$$

$$\begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ v_{21} \\ v_{31} \end{bmatrix} = 5.83 \begin{bmatrix} 1 \\ v_{21} \\ v_{31} \end{bmatrix}$$

$$\begin{bmatrix} 1-2v_{21} \\ -2+5v_{21} \\ 2v_{31} \end{bmatrix} = \begin{bmatrix} 5.83 \\ 5.83v_{21} \\ 5.83v_{31} \end{bmatrix}$$

$$\begin{cases} 1-2v_{21} = 5.83 \\ 2v_{31} = 5.83v_{31} \end{cases}$$

$$\begin{cases} v_{21} = \dfrac{1-5.83}{2} \\ v_{31} = 0 \end{cases}$$

$$\begin{cases} v_{21} = -2.415 \\ v_{31} = 0 \end{cases}$$

$$c = \begin{bmatrix} 1 & -2.415 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -2.415 \\ 0 \end{bmatrix} = 6.832$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{6.832}} \begin{bmatrix} 1 \\ -2.415 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.383 \\ -0.924 \\ 0 \end{bmatrix}$$

  - therefore

$$\mathbf{v}_1^T = \begin{bmatrix} 0.383 & -0.924 & 0 \end{bmatrix}$$

  - Similarly, other eigenvectors are

$$\mathbf{v}_2^T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{v}_3^T = \begin{bmatrix} 0.924 & 0.383 & 0 \end{bmatrix}$$

- Consider a 1-factor solution
$$\mathbf{L} = \sqrt{\lambda_1}\mathbf{v}_1 = \begin{bmatrix} 0.924 \\ -2.230 \\ 0 \end{bmatrix}, \boldsymbol{\Psi} = diag(\boldsymbol{\Sigma} - \mathbf{L}\mathbf{L}^T) = \begin{bmatrix} 0.146 & 0 & 0 \\ 0 & 0.025 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$
- Consider a 2-factor solution
$$\mathbf{L} = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{v}_1 & \sqrt{\lambda_2}\mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 0.924 & 0 \\ -2.230 & 0 \\ 0 & 1.414 \end{bmatrix}, \boldsymbol{\Psi} = diag(\boldsymbol{\Sigma} - \mathbf{L}\mathbf{L}^T) = \begin{bmatrix} 0.146 & 0 & 0 \\ 0 & 0.025 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
  - X3 loads on F2 only as it is not correlated with other 2 variables
  - Communality
    $h_1^2 = (0.924)^2 + 0^2 = 0.854$
    $h_2^2 = (-2.230)^2 + 0^2 = 4.975$
    $h_3^2 = 0^2 + 1.414^2 = 2$
  - Proportion of total variance explained
    Factor 1: $[(0.924)^2 + (-2.230)^2 + 0^2]/(1 + 5 + 2) = 5.828/8 = 0.7286$
    Factor 2: $1.414^2/8 = 2/8 = 0.25$
  - Correlation between factors and observed variables
    Corr(F1,X1) = Cov(F1,X1)/√(Var(F1) Var(X1)) = 0.924/1/1 = 0.924
    Corr(F1,X2) = -2.230/1/√5 = -0.997
    Corr(F2,X3) = 1.414/1/√2 = 1

- Number of factors
  - Latent Root Criterion
    - Applied to standardized variables
    - Only factors having eigenvalues greater than 1 are considered significant.
  - A Prior Criterion
    - The analyst already knows the number of factors to extract before undertaking the factor analysis.
    - Useful to testing a theory or hypothesis about the number of factors to be extracted
  - Percentage of Variance Criterion
    - Based on the cumulative percentages of the variance extracted by successive factors.
    - In the social sciences, 60% (or even less) of the total variance.
  - Scree Test Criterion
    - By plotting the eigenvalues of the successive factors.
    - The point at which the curve first begins to straighten out is considered to indicate the maximum number of factors

b) Maximum likelihood method
- Assume $\mathbf{F}_j$ and $\boldsymbol{\varepsilon}_j$ are jointly normal
  - Then $\mathbf{X}_i \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T + \boldsymbol{\Psi}$
  - The likelihood function for the multivariate normal distribution is given as
$$L(\mu, \Sigma|\mathbf{X}) = (2\pi)^{-(n-1)p/2}|\Sigma|^{-(n-1)/2} \exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\right)\right)$$
$$\times (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{n}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})\right)$$
  - Restriction for the uniqueness of the estimate
$$\mathbf{L}^T\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}, \text{ a diagonal matrix}$$
    - No statistical meaning

- Maximum likelihood estimate (MLE) is obtained by maximizing $L(\mu, \Sigma|\mathbf{X})$ and must be obtained by numerical method

- Consider the standardized values
  - If the MLE based on $\mathbf{X}$ are $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$, then the MLE based on $\mathbf{Z} = diag(\mathbf{S})^{-1/2}\mathbf{X}$ are
$$\hat{\mathbf{L}}_z = diag(\mathbf{S})^{-1/2}\hat{\mathbf{L}}$$
$$\hat{\mathbf{\Psi}}_z = diag(\mathbf{S})^{-1/2}\hat{\mathbf{\Psi}}diag(\mathbf{S})^{-1/2} = diag(\mathbf{S})^{-1}\hat{\mathbf{\Psi}}$$
  - Does not hold for principal component factor method.

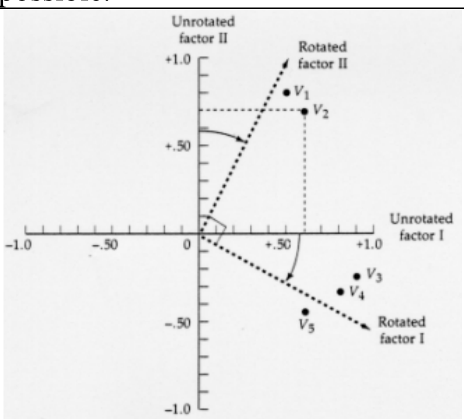
Example

- Covariance matrix
$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$
- Correlation matrix
$$\mathbf{\Gamma} = \begin{bmatrix} 1 & -0.894 & 0 \\ -0.894 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
- 2-factor solution for $\mathbf{\Gamma}$ by PCFA
  - $\lambda_1 = 1.894$, $\lambda_2 = 1.00$
  - $\mathbf{v}_1^T = \begin{bmatrix} 0.707 & -0.707 & 0 \end{bmatrix}$
  - $\mathbf{v}_2^T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$
  - Factor loadings = Corr(F,X)
$$\mathbf{L} = \begin{bmatrix} 0.973 & 0 \\ -0.973 & 0 \\ 0 & 1 \end{bmatrix}$$
  - *c.f.* 2-factor solution for covariance matarix,
$$\mathrm{Corr}(\mathbf{F}, \mathbf{X}) = \begin{bmatrix} 0.924 & 0 \\ -0.997 & 0 \\ 0 & 1 \end{bmatrix}$$

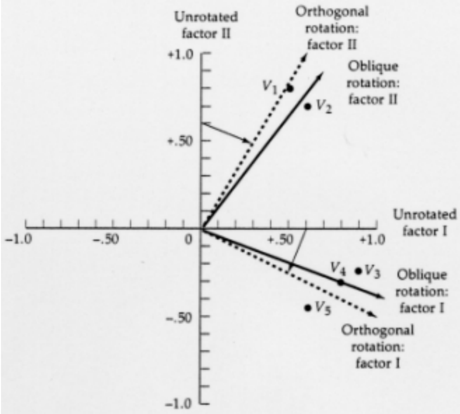
- Test for the number of factors
  - Test the adequacy of the m common factor model
$$H_0: \mathbf{\Sigma}_{p\times p} = \mathbf{L}_{p\times m}\mathbf{L}_{m\times p}^T + \mathbf{\Psi}_{p\times p}$$
$$H_1: \Sigma \text{ any other positive definite matrix}$$
  - Under H_0, MLE for Σ is $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}$
  - Under H_1, MLE for Σ is $\mathbf{S}_n$, sample covariance matrix
  - Likelihood ratio test, reject H_0 at α level if
$$n \cdot \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{\alpha,[(p-m)^2-p-m]/2}$$

5. Factor rotation

- Uniqueness of $\mathbf{L}$
  - Let $\mathbf{T}$ be any $m \times m$ (nonrandom) orthogonal matrix so that
    $$\mathbf{T}\mathbf{T}^T = \mathbf{T}^T\mathbf{T} = \mathbf{I}$$
  - Define rotated loadings and factors are
    $$\mathbf{L}^* = \mathbf{L}\mathbf{T} \text{ and } \mathbf{F}^* = \mathbf{T}^T\mathbf{F}$$
  - We have
    $$E(\mathbf{F}^*) = \mathbf{0}$$
    $$Cov(\mathbf{F}^*) = \mathbf{I}$$
    $$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi} = \mathbf{L}^*\mathbf{L}^{*T} + \mathbf{\Psi}$$
    - Impossible to distinguish the loadings $\mathbf{L}$ from $\mathbf{L}^*$
      - Factor solution is not unique
      - Problem in solving for unique solution
      - Provides the rationale for factor rotation so as to obtain a meaningful loadings and factors.
    - Initial (original) loadings may not be readily interpretable, it is usual practice to rotate them until a "simpler structure" is achieved
  - Since $\mathbf{L}\mathbf{L}^T = \mathbf{L}^*\mathbf{L}^{*T}$ and $\mathbf{\Psi}$ do not change
    - Communality and uniqueness do not change after rotation
  - Orthogonal rotation
    - Keep $Cov(F_i^*, F_j^*) = 0$
    - Criteria: Varimax, equamax, quartimax
    - Kaiser (normal) varimax, maximize
      $$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*4} - \left(\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*2}\right)^2 / p\right]$$
      $$\tilde{\ell}_{ij}^* = \hat{\ell}_{ij}^* / \hat{h}_i^2$$
      $V \propto$ variance of loadings for the factors. Maximizing V corresponds to spreading out the loadings as much as possible.



  - Oblique rotation
    - Rotated factors may be dependent, $Cov(F_i^*, F_j^*) \neq 0$
    - Non-orthogonal rotation of the coordinate system such that the rotated axes (no longer perpendicular) pass (nearly) through the clusters of variables.
    - More realistic and useful than orthogonal rotation.
    - Criteria: Oblimmin, promax

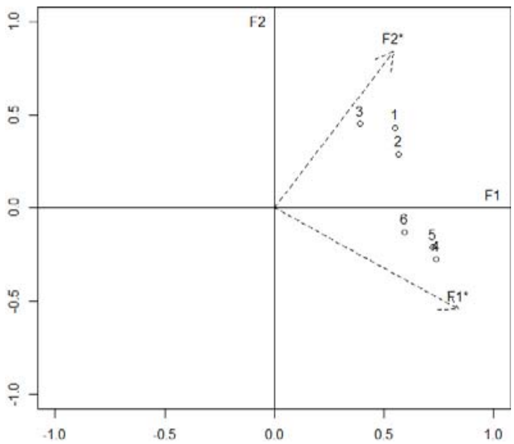


Example

- Examination scores in 6 subjects for 22 students
- Maximum likelihood estimates of the factor loadings are

	Factor loadings	
Variable	F_1	F_2
1. Gaelic	0.553	0.429
2. English	0.568	0.288
3. History	0.392	0.450
4. Arithmetic	0.740	-0.273
5. Algebra	0.724	-0.211
6. Geometry	0.595	-0.132

  - First factor
    - All variables have positive loadings
    - Reflects the overall response
    - General intelligence
  - Second factor
    - Half the loadings are positive and half are negative
    - Bipolar factor
    - A contrast of math and nonmath subjects
- Factor loadings pairs plot



  - Points are the variables with loadings as coordinates

- Rotate the coordinate systems from F_1 and F_2 to F_1^* and F_2^*
  - Rotation matrix based on varimax rotation method

$$\mathbf{T} = \begin{bmatrix} 0.840 & 0.543 \\ -0.543 & 0.840 \end{bmatrix}$$

- The rotated loadings ($\mathbf{L}^* = \mathbf{LT}$) are

Variable	Factor loadings	
	F_1^*	F_2^*
1. Gaelic	0.232	0.660
2. English	0.321	0.551
3. History	0.085	0.591
4. Arithmetic	0.770	0.173
5. Algebra	0.723	0.215
6. Geometry	0.572	0.213

- The factors F_1^* and F_2^* are still uncorrelated (independent if normally distributed)
- The math variables load highly on F_1^* and have negligible loadings on F_2^*
    - F_1^* is the mathematical-ability factor
- The verbal variables have high loadings on F_2^* and moderate-to-small loadings on F_1^*
    - F_2^* is the verbal-ability factor

6. Factor scores

- Factor scores are the estimated values of the common factors

$$\hat{\mathbf{f}}_j = \text{estimate of the value, } \mathbf{f}_j \text{, attained by } \mathbf{F}_j \text{ (}j\text{th case)}$$

    - Used for diagnostic purposes
    - Observations under the reduced dimensions, from $\mathbf{x}_j$ to $\hat{\mathbf{f}}_j$
    - Inputs to a subsequent analysis

- Regression
    - It is known that if

$$\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right)$$

then

$$E(\mathbf{X}_2|\mathbf{X}_1) = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1)$$

    - Consider in a factor model,

$$\begin{pmatrix}\mathbf{X}\\ \mathbf{F}\end{pmatrix} \sim N\left(\begin{pmatrix}\boldsymbol{\mu}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{LL}^T + \boldsymbol{\Psi} & \mathbf{L}\\ \mathbf{L}^T & \mathbf{I}\end{pmatrix}\right)$$

then

$$E(\mathbf{F}|\mathbf{X}) = \mathbf{L}^T(\mathbf{LL}^T + \boldsymbol{\Psi})^{-1}(\mathbf{X} - \boldsymbol{\mu}).$$

    - Therefore, we may estimate the factor scores by

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}^T\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}).$$

    - So as to reduce a possible incorrect determination of the number of factors, we use **S** (or **R**) instead of $\hat{\boldsymbol{\Sigma}}$ and thus,

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}^T\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

- Weighted least squares (Bartlett's method)

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{Lf} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\Psi}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}) = \boldsymbol{\Psi}^{-1/2}\mathbf{Lf} + \boldsymbol{\Psi}^{-1/2}\boldsymbol{\varepsilon}$$

(c.f. $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$, where $\hat{\mathbf{B}} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{Y}$)

$$\hat{\mathbf{f}} = \left(\mathbf{L}^\mathbf{T}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Psi}^{-\frac{1}{2}}\mathbf{L}\right)^{-1}\mathbf{L}^\mathbf{T}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Psi}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$$

$$\hat{\mathbf{f}}_j = \left(\hat{\mathbf{L}}^T\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}\right)^{-1}\hat{\mathbf{L}}^T\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$$

- Rotated factors
    - If rotated loadings

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$$

then

$$\hat{\mathbf{f}}_j^* = \mathbf{T}^T\hat{\mathbf{f}}_j.$$

Example (eg1)

- In a consumer brand perception survey, consumer ratings of brands with regard to 9 perceptual adjectives were collected on a scale where 1 is least and 10 is most.
- The adjective are

Adjective	Description
Perform	Brand has strong performance
Leader	Brand is a leader in the field
Latest	Brand has the latest products
Fun	Brand is fun
Serious	Brand is serious
Bargain	Brand products are a bargain
Value	Brand products are a good value
Trendy	Brand is trendy
Rebuy	I would buy from Brand again

- Apply factor analysis to explore the perception and study the relationship with the brand.

Solution

- Correlation matrix of the 9 variables
    - Perform, serious, and leader are positively correlated.
    - Trendy and latest are positively correlated.
    - Bargain, value and rebuy are positively correlated.

```
> print(cor(x=brand.ratings[,1:9]),digits=2)
        perform leader latest    fun serious bargain  value  trendy rebuy
perform  1.0000  0.500 -0.122 -0.256  0.3592  0.0571  0.102  0.0087  0.31
leader   0.5002  1.000  0.027 -0.290  0.5712  0.0331  0.118  0.0665  0.21
latest  -0.1224  0.027  1.000  0.245  0.0100 -0.2544 -0.343  0.6276 -0.40
fun     -0.2563 -0.290  0.245  1.000 -0.2811 -0.0666 -0.145  0.1280 -0.24
serious  0.3592  0.571  0.010 -0.281  1.0000 -0.0027  0.024  0.1210  0.18
bargain  0.0571  0.033 -0.254 -0.067 -0.0027  1.0000  0.740 -0.3505  0.47
value    0.1019  0.118 -0.343 -0.145  0.0238  0.7396  1.000 -0.4345  0.51
trendy   0.0087  0.067  0.628  0.128  0.1210 -0.3505 -0.435  1.0000 -0.30
rebuy    0.3067  0.209 -0.397 -0.237  0.1807  0.4674  0.506 -0.2982  1.00
```

- Principal component factor model is employed
    - Standardized values are considered, i.e. PCFA applies to the correlation matrix

- Number of factors
    - The 1st 4 factors explained 76% of the total variance

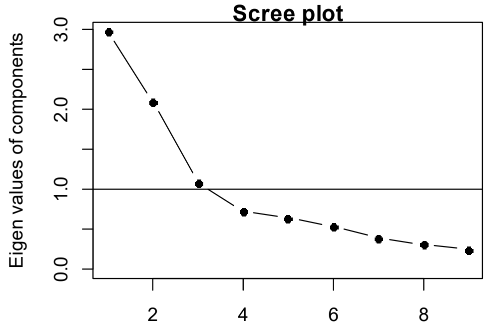
```
                      PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9
SS loadings          2.98 2.10 1.08 0.73 0.64 0.53 0.39 0.31 0.24
Proportion Var       0.33 0.23 0.12 0.08 0.07 0.06 0.04 0.03 0.03
Cumulative Var       0.33 0.56 0.68 0.76 0.84 0.89 0.94 0.97 1.00
```

- The variance explained (eigenvalue) is 2.97 by the 1st factor, 2.10 by the 2nd factor, 1.08 by the 3rd factor and 0.73 by the 4th factor.
- The 1st factor explains 2.98/9 = 33% of the total variance

- The 2nd factor explains 2.10/9 = 23% of the total variance
- The 3rd factor explains 1.08/9 = 12% of the total variance
- The 4th factor explains 0.73/9 = 8% of the total variance
- The additional variance explained by the 5th factor is only 7%
- Therefore 4 factors are selected (if they are meaningful).

o Latent root criteria
  - Eigenvalues for the 3rd factor > 1 while that of 4th factors are <1
  - 3 factors are suggested

o Scree plot shows the eigenvalue for the number of factors.
  - An elbow is observed at 3
  - 2 factors are suggested



- Initial loadings for a 3-factor solution
  - o Factor pattern shows the initial loadings

```
> fit
Principal Components Analysis

Standardized loadings (pattern matrix) based upon correlation matrix
          PC1   PC2   PC3   h2   u2 com
perform  0.41  0.61 -0.04 0.54 0.46 1.8
leader   0.36  0.76  0.10 0.71 0.29 1.5
latest  -0.64  0.29  0.55 0.80 0.20 2.4
fun     -0.43 -0.36  0.43 0.51 0.49 2.9
serious  0.28  0.74  0.04 0.62 0.38 1.3
bargain  0.69 -0.32  0.51 0.83 0.17 2.3
value    0.77 -0.27  0.38 0.82 0.18 1.8
trendy  -0.61  0.46  0.39 0.73 0.27 2.6
rebuy    0.76  0.02  0.13 0.59 0.41 1.1

                       PC1  PC2  PC3
SS loadings           2.98 2.10 1.08
Proportion Var        0.33 0.23 0.12
Cumulative Var        0.33 0.56 0.68
```

  - $Z_{perform} = 0.41F_1 + 0.61F_2 - 0.04F_3 + \varepsilon_1$
  - $Z_{leader} = 0.36F_1 + 0.76F_2 + 0.10F_3 + \varepsilon_2$
  - …

o Communalities
  - Perform: $(0.41)^2 + (0.61)^2 + (-0.04)^2 = 0.54$
    - ♦ 54% of the variance of perform are explained by the 3 factors.
  - Leader: $(0.36)^2 + 0.76^2 + 0.10^2 = 0.71$
    - ♦ 71% of the variance of leader are explained by the 3 factors.

- 51% to 83% of the variances of the variables (attributes) are explained by the 3 factors
  - Converted to percentages as the variances of the standardized variables are all equal to one
- Uniqueness = 1 – communality

- Rotated loadings
  - Varimax rotation is employed

```
> fitr
Standardized loadings (pattern matrix) based upon correlation matrix
          RC2   RC1   RC3   h2   u2 com
perform  0.72  0.10 -0.09 0.54 0.46 1.1
leader   0.83  0.10  0.09 0.71 0.29 1.1
latest  -0.02 -0.20  0.87 0.80 0.20 1.1
fun     -0.53  0.06  0.48 0.51 0.49 2.0
serious  0.78  0.01  0.09 0.62 0.38 1.0
bargain -0.03  0.91 -0.08 0.83 0.17 1.0
value    0.04  0.88 -0.21 0.82 0.18 1.1
trendy   0.16 -0.33  0.77 0.73 0.27 1.5
rebuy    0.32  0.62 -0.32 0.59 0.41 2.0

                       RC2  RC1  RC3
SS loadings           2.24 2.16 1.76
Proportion Var        0.25 0.24 0.20
Cumulative Var        0.25 0.49 0.68
```

  - $Z_{perform} = 0.72F_1^* + 0.10F_2^* - 0.09F_3^* + \varepsilon_1$
  - $Z_{leader} = 0.83F_1^* + 0.10F_2^* + 0.09F_3^* + \varepsilon_2$
  - …

  - Rotation matrix

```
> fitr$rot.mat[c(1,2,3),c(2,1,3)]
           [,1]       [,2]       [,3]
[1,]  0.4046359  0.7239827 -0.5586759
[2,]  0.9134025 -0.2903499  0.2852944
[3,] -0.0443367  0.6257363  0.7787736
```

LT = L*

$$\begin{bmatrix} 0.41 & 0.61 & -0.04 \\ \vdots & \vdots & \vdots \\ 0.76 & 0.02 & 0.13 \end{bmatrix} \begin{bmatrix} 0.40 & 0.72 & -0.56 \\ 0.91 & -0.29 & 0.29 \\ -0.04 & 0.63 & 0.78 \end{bmatrix} = \begin{bmatrix} 0.72 & 0.10 & -0.09 \\ \vdots & \vdots & \vdots \\ 0.32 & 0.62 & -0.32 \end{bmatrix}$$

  - Total variance explained by the rotated factor
    - 1st factor: $(0.72)^2 + (0.83)^2 + \ldots + (0.32)^2 = 2.24$
      - 2.24/9 = 25% of the total variances are explained
    - 2nd factor: $(0.10)^2 + (0.10)^2 + \ldots + 0.62^2 = 2.16$
      - 2.16/9 = 24% of the total variances are explained
    - 3rd factor: $(-0.09)^2 + \ldots + (0.59)^2 = 1.76$
      - 1.76/9 = 20% of the total variances are explained
    - 25% + 24% + 20 = 68% of the total variances are explained by the 3 factors
      - Exactly the same as that by the initial factors

  - Are the communalities the same as those by initial factors?

- Factor 1
  - Perform, leader, serious load highly and positively (fun negatively)
  - Leader factor
- Factor 2
  - Bargain, value, rebuy load highly and positively
  - Value factor
- Factor 3
  - Latest and trendy load highly and positively
  - Latest factor

- Factor scores
  - The ratings are obtained for 10 brands, a, b, …, j
  - Based on the factor scores, one can reveal the differences among brands
  - Employing the regression method

$$\hat{\mathbf{f}}_i = \hat{\mathbf{L}}^T \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$$

  - Factor score coefficients (weights^T) are

$$\hat{\mathbf{L}}^T \mathbf{S}^{-1}$$

  - When correlation matrix is used, i.e. standardized values are used,

$$\mathbf{S} = \mathbf{R}$$
$$\mathbf{x}_i - \bar{\mathbf{x}} = \mathbf{z}_i$$

```
> fitr$weights
                RC2          RC1         RC3
perform  0.32220484 -0.007812555 -0.02301553
leader   0.37462977  0.038590336  0.10790099
latest   0.01751890  0.125126337  0.55892722
fun     -0.23463257  0.196561000  0.34513466
serious  0.35773089 -0.010863999  0.07936525
bargain -0.06516430  0.506320402  0.19499200
value   -0.03059865  0.448147777  0.09457163
trendy   0.10279176  0.012303175  0.45441432
rebuy    0.10712385  0.256173984 -0.04570612

> head(fitr$scores)
            RC2        RC1        RC3
[1,] -0.8640183  1.8417025  0.7136232
[2,] -1.4328558 -1.4820205 -1.3650147

> head(scale(brand.ratings[,1:9]))
     perform leader latest   fun serious bargain value trendy rebuy
[1,]   -0.78  -0.16   0.59  0.70   -0.84   1.778  1.11  -0.44  0.89
[2,]   -1.09  -1.31  -0.71  0.34   -1.20  -1.222 -1.39  -1.17 -0.68
```

  - Obs. 1

$$f_1^* = 0.32Z_{perform} + 0.37Z_{leader} + \cdots + 0.11Z_{rebuy}$$
$$= 0.32(-0.78) + 0.37(-0.16) + \cdots + 0.11(0.89) = -0.86$$

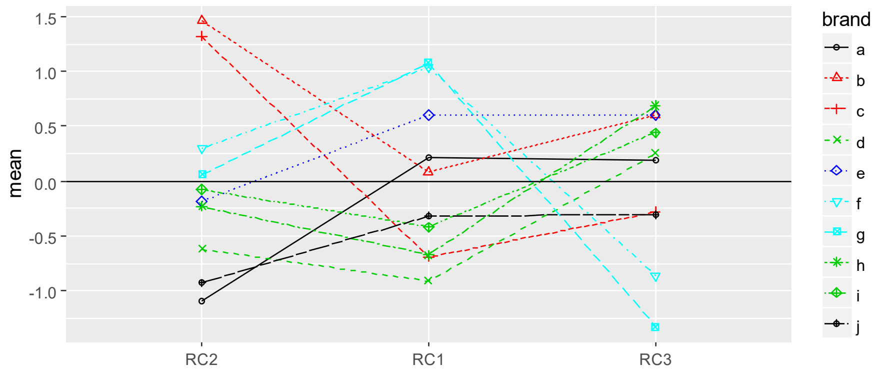
  - Obs. 2

$$f_2^* = -0.01(-1.09) + 0.04(-1.31) + \cdots + 0.26(-0.68) = -1.48$$

- Means scores for each brand

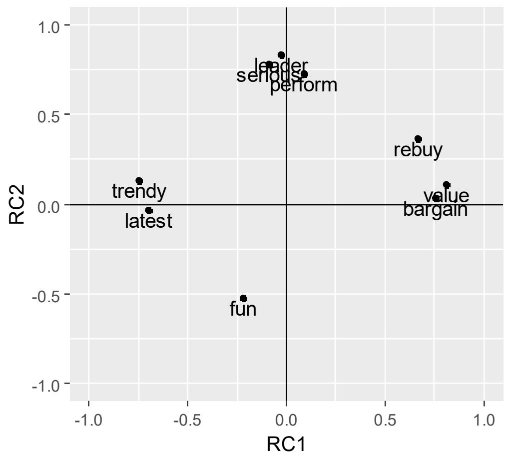
```
> tb
  brand         RC2         RC1        RC3
1     a -1.09626675  0.20901475  0.1946609
2     b  1.45928881  0.07966015  0.6063110
3     c  1.32029981 -0.69078498 -0.2835684
4     d -0.61772538 -0.91066342  0.2555178
5     e -0.18654770  0.60439712  0.6027842
```

```
6       f  0.29538558  1.04216033 -0.8678992
7       g  0.05864475  1.07961147 -1.3266879
8       h -0.22979231 -0.67510683  0.6866718
9       i -0.07598665 -0.42151091  0.4392064
10      j -0.92730016 -0.31677767 -0.3069966
```

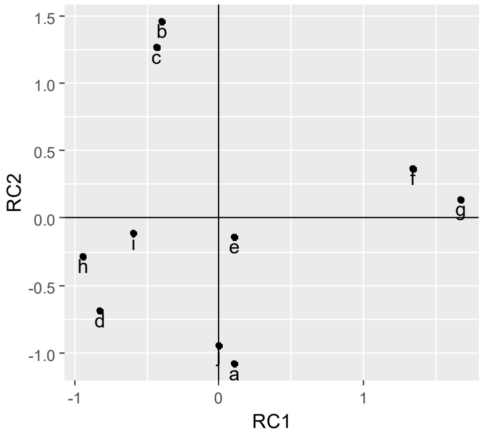


- Brand F, G, high at Value (factor 2)
- Brands B, C high at Leader (factor 1)
- Brands E: high at Latest (factor 3) and high at Value (factor 2)
- Brands A, J: low at Leader (factor 1) ⇒ high at Fun
- Brands D, H, I: high at Latest (factor 3) and low at Value (factor 2)

- 2 factor solution
  - The first 2 factors explain 56 % of the total variances
    - A 2-dimensional perceptual map is produced
  - Rotated loadings
    - Leader, serious, perform are correlated
    - Value, rebuy and bargain are correlated
    - Latest, trendy are correlated
    - Fun on its own

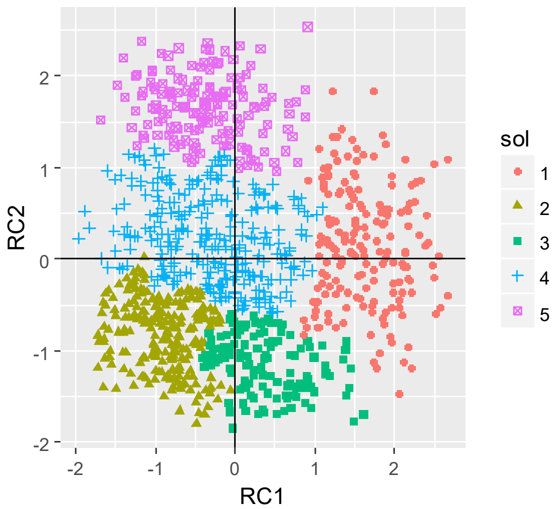


- Factor scores (regression method)



- Comparing the factor loadings plot and the factor scores plot
  - B, C
    - Leader, serious, perform
  - F, G
    - Value, rebuy, bargain
  - A, J
    - Fun
  - D, H, I
    - Latest, trendy
  - E
    - No strong characteristic

- Factor scores for a 5-cluster solution



  - Cluster 5
    - Leader, serious, perform
  - Cluster 1
    - Value, rebuy, bargain
  - Cluster 3
    - Fun
  - Cluster 2
    - Latest, trendy
  - Cluster 4
    - No strong characteristic

Example (eg2)

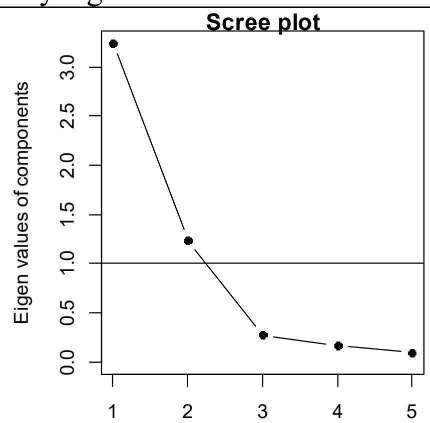
- In a consumer-preference study, a random sample of customers were asked to rate several attributes, Taste (X1), Good buy for money (X2), Flavor (X3), Suitable for snack (X4), Provides lots of energy (X5), of a new product of candy.
- The responses on a 7-point semantic differential scale were tabulated.
- Apply factor analysis by maximum likelihood method to explore the structure of the attributes.

Solution

- Correlation matrix
  - o Taste and Flavor (correlation = 0.899), Good buy for money and Provides lots of energy (correlation = 0.728) form groups
  - o Suitable for snack closer to (Good buy for money and Provides lots of energy) (correlation = 0.730 and 0.712) than (Taste and Flavor) (correlation = 0.623 and 0.584)
  - o Linear relationships between the variables can be explained in terms of 2 (or 3) common factors

```
> cor(pref[2:6])
          x1        x2        x3        x4        x5
x1 1.0000000 0.4021041 0.8991231 0.6232539 0.2030294
x2 0.4021041 1.0000000 0.4499566 0.7299040 0.7282935
x3 0.8991231 0.4499566 1.0000000 0.5839569 0.2066882
x4 0.6232539 0.7299040 0.5839569 1.0000000 0.7121766
x5 0.2030294 0.7282935 0.2066882 0.7121766 1.0000000
```

- Number of factors
  - o Eigenvalues
    - ▪ Based on the preliminary eigenvalues




```
> scp$pc
[1] 3.23798738 1.23163132 0.27413866 0.16811889 0.08812375
```

  - o Scree plot
    - ▪ Elbow at 3rd eigenvalue
    - ▪ 2nd eigenvalue > 1, 3rd eigenvalue < 1
    - ▪ 2 factors are suggested

- 2-factor model

```
> print(fit,digits=3)
Standardized loadings (pattern matrix) based upon correlation matrix
     ML1    ML2    h2     u2  com
x1 0.975 -0.141 0.970 0.0301 1.04
x2 0.515  0.653 0.692 0.3082 1.90
x3 0.908 -0.097 0.835 0.1653 1.02
```

```
x4 0.717  0.554 0.822 0.1785 1.88
x5 0.330  0.856 0.841 0.1588 1.29

                       ML1   ML2
SS loadings           2.664 1.495
Proportion Var        0.533 0.299
Cumulative Var        0.533 0.832

The degrees of freedom for the model are 1  and the objective function was
0.061
The total number of observations was  50  with Likelihood Chi Square =  2.74
6  with prob <  0.0975
```

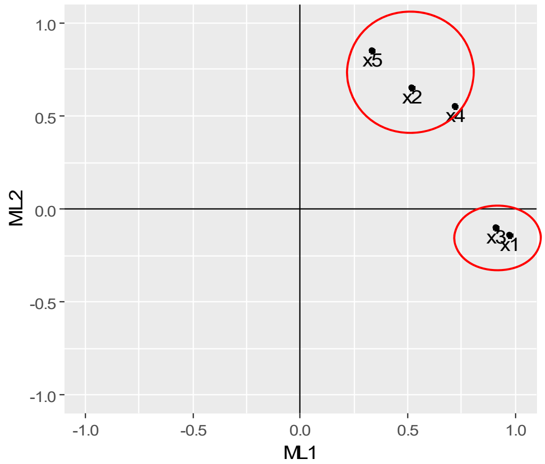
- Test if 2 factors are sufficient
  - $\chi^2 = n \cdot \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|} = 3.039$
    - Bartlett's correction: $\chi^2 = (n - 1 - (2p + 4m + 5)/6) \cdot \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|} = 2.745$
  - $df = \frac{[(p-m)^2 - p - m]}{2} = \frac{[(5-2)^2 - 5 - 2]}{2} = 1$
  - $\chi^2_{0.05,1} = 3.84 > 2.7457$
  - p-value = 0.0975 > 0.05
  - Do not reject H_0 at $\alpha = 5\%$. 2 factors are sufficient.

- Initial loadings
  - Communalities
    - Taste: $0.975^2 + (-0.141)^2 = 0.970$
    - Good buy for money: $0.515^2 + 0.653^2 = 0.692$
    - Communalities are close to 1, except Good buy for money
    - 2 factors account for a large percentage of the variance of each variable

  - Total variance explained in a 2-factor solution
    - Factor 1
      - $0.975^2 + 0.515^2 + 0.908^2 + 0.717^2 + 0.330^2 = 2.664$
      - 2.664/5 = 53.3%
    - Factor 2
      - $(-0.141)^2 + 0.653^2 + (-0.097)^2 + 0.554^2 + 0.856^2 = 1.495$
      - 1.495/5 = 29.9%
    - Total variance explained = 53.3% + 29.9% = 83.2%
    - 2 factors account for a large percentage of the total variance

  - All 5 attributes load quite highly on the first factor
  - Good buy for money, Suitable for snack and Provides lots of energy load highly on the second factor

- Loading plot shows 2 groups of variables



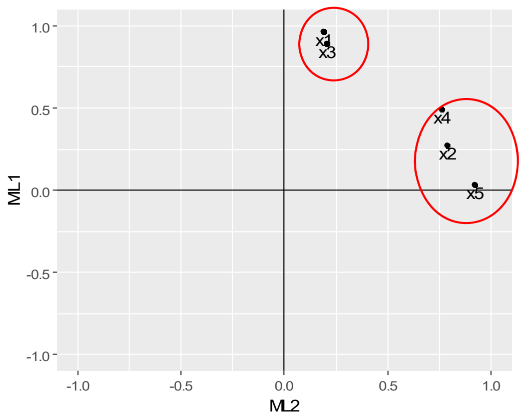
- Varimax rotation

```
> print(fitr,digits=3)
Standardized loadings (pattern matrix) based upon correlation matrix
     ML2   ML1    h2     u2  com
x1 0.186 0.967 0.970 0.0301 1.07
x2 0.786 0.272 0.692 0.3082 1.24
x3 0.206 0.890 0.835 0.1653 1.11
x4 0.759 0.496 0.822 0.1785 1.72
x5 0.917 0.032 0.841 0.1588 1.00

                        ML2   ML1
SS loadings           2.111 2.048
Proportion Var        0.422 0.410
Cumulative Var        0.422 0.832

> fitr$rot.mat[c(1,2),c(2,1)]
          [,1]       [,2]
[1,] 0.3277644  0.9447595
[2,] 0.9447595 -0.3277644
```

- Factor 1
  - Good buy for money (X2), Suitable for snack (X4) and Provides lots of energy (X5) define factor 1 (high loadings on factor 1, small or negligible loadings on factor 2)
  - Nutritional factor
- Factor 2
  - Taste (X1) and Flavor (X3) define factor 2 (high loadings on factor 2, small or negligible loadings on factor 1)
  - Taste factor

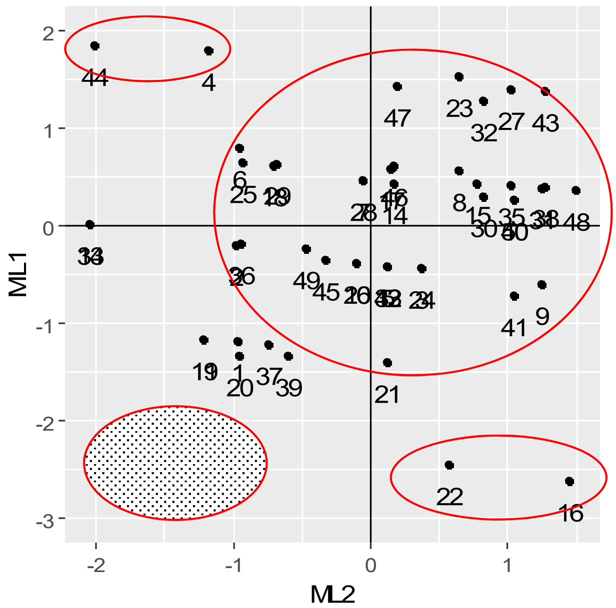


- Calculation of factor scores
  - Factor score coefficients are

$$\left(\hat{\mathbf{L}}^T \mathbf{R}^{-1}\right)^T$$

```
> fitr$weights
           ML2         ML1
x1 -0.21610865  0.89143367
x2  0.22537278 -0.03608151
x3 -0.01599958  0.14405822
x4  0.34233328 -0.01748175
x5  0.55587471 -0.14041265
```

- Perceptual map
  - Plot of the factor scores



  - Most people preferred good at both nutrition and taste

  - There are extreme cases
    - Case 44 and 4
      - Prefer very good at taste even if very bad at nutrition
    - Case 22 and 16
      - Prefer very good at nutrition even if very bad at taste
  - No people preferred product that is bad at both nutrition and taste

  - Are the products in the market serving the majority only?
  - Are there any products serving cases 44 and 4, and cases 22 and 16?
  - Worth to produce product for these 2 groups?