Tutorial 4 (STAT8019)

ZHANG Yan

February 8, 2021

Differentiate different methods:

(1) Factor Analysis and Linear Regression

Orthogonal factor model:

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon},\ Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$$

Goal: Try to figure out the structure of the distribution of $\boldsymbol{X}$, $P(\boldsymbol{X}) = P(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{L}, \boldsymbol{\Psi})$.

Regression model:

$$Y = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{x} + \varepsilon,\ Var(\varepsilon) = \sigma^2$$

Goal: Try to figure out the linear relationship between Y and $\boldsymbol{X}$ based on the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$, $P(Y|\boldsymbol{x}) = P(Y|\boldsymbol{x}, \beta_0, \boldsymbol{\beta}, \sigma^2)$.

(2) Principal Component Analysis and Probabilistic PCA

PCA is a model-free method based on the following compression-reconstruction linear transformations:

$$\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{L},\ \boldsymbol{D} \approx \boldsymbol{D}^* = \boldsymbol{Z}\boldsymbol{L}^T$$

where $\boldsymbol{D} \in R^{n\times p}$, $\boldsymbol{Z} \in R^{n\times m}$, $\boldsymbol{L} \in R^{p\times m}$, $\boldsymbol{L}^T\boldsymbol{L} = \boldsymbol{I}$.

Goal: Try to find $\boldsymbol{L}$, or more generally, a compression-reconstruction method, so that we loss the least information in reconstructed data $\boldsymbol{D}^*$.

Probabilistic PCA model:

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon},\ Cov(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$$

It's different from PCA.

(3) Principal Component Method and Maximum Likelihood Method

They are two estimation methods for factor model.

Principal component method is based on eigen-decomposition :

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T = \boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{V}^T \approx \boldsymbol{L}\boldsymbol{L}^T.$$

Maximum likelihood method added a new assumption to the factor model: $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are jointly normal, and so is $\boldsymbol{X}$. And it estimate $\boldsymbol{L}$ and $\boldsymbol{\Psi}$ based on MLE.

Q1:

a).

Assume

$$\boldsymbol{Z}^* = \begin{bmatrix} Z_1^* \\ Z_2^* \\ Z_3^* \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.7 \\ 0.5 \end{bmatrix} F_1 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \boldsymbol{L}F_1 + \boldsymbol{\varepsilon},$$

here we try to prove $Cov(\boldsymbol{Z}^*) = Cov(\boldsymbol{Z}) = \boldsymbol{\Sigma}$.

From the independence assumption between F_1 and $\boldsymbol{\varepsilon}$, we can see that $Cov(\boldsymbol{Z}^*) = \boldsymbol{L}Var(F_1)\boldsymbol{L}^T + Cov(\boldsymbol{\varepsilon}) = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}$.

So we only need to check $\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi} = \boldsymbol{\Sigma}$.

$$\begin{bmatrix} 0.9 \\ 0.7 \\ 0.5 \end{bmatrix} \begin{bmatrix} 0.9 & 0.7 & 0.5 \end{bmatrix} + \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} =$$

$$\begin{bmatrix} 0.81 & 0.63 & 0.45 \\ 0.63 & 0.49 & 0.35 \\ 0.45 & 0.35 & 0.25 \end{bmatrix} + \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} = \begin{bmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{bmatrix} = \boldsymbol{\Sigma}$$

b).

Remember the definition of communality: variance of the ith variable contributed by all the common factors.

So here we have $h_i^2 = l_{i1}^2$, $i = 1, 2, 3$.

	Z_1	Z_2	Z_3
h_i^2	0.81	0.49	0.25
$h_i^2/Var(Z_i)$	81%	49%	25%

c).

Here we try to calculate $Cor(Z_i, F_1) = Cov(Z_i, F_1)/\sqrt{Var(Z_i)Var(F_1)}$, $i = 1, 2, 3$.

Refer to $\boldsymbol{\Sigma}$ and the definition of factor, we know $Cor(Z_i, F_1) = Cov(Z_i, F_1) = Cov(l_{i1}F_1 + \varepsilon_i, F_1) = l_{i1}Var(F_1) = l_{i1}$. (Normally, $l_{i1}/Std(Z_i)$)

	Z_1	Z_2	Z_3
$Cor(Z_i, F_1)$	0.9	0.7	0.5

Since Z_1 has the greatest correlation with F_1, it carries the greatest weight in 'naming' the common factor. (When we want to 'name' or interpret a factor, we mainly count on the variables with the biggest positive or negative correlations with the factor.)

Q2:

a).

Remember the factor model:

$$\boldsymbol{X}_{p\times 1} - \boldsymbol{\mu}_{p\times 1} = \boldsymbol{L}_{p\times m}\boldsymbol{F}_{m\times 1} + \boldsymbol{\varepsilon}_{p\times 1},$$

Normally we assume $m \le p$ and the loading matrix $\boldsymbol{L}$ has rank m. Then by manipulating the above equation, we have:

$$\begin{aligned}\boldsymbol{\varepsilon} &= \boldsymbol{X} - \boldsymbol{\mu} - \boldsymbol{L}\boldsymbol{F},\\ \boldsymbol{F} &= (\boldsymbol{L}^T\boldsymbol{L})^{-1}\boldsymbol{L}^T(\boldsymbol{X} - \boldsymbol{\mu} - \boldsymbol{\varepsilon}),\end{aligned}$$

So the existence of linear combination of the attributes for $\boldsymbol{F}$ is equivalent to the existence of that for $\boldsymbol{\varepsilon}$. So we only need to check whether we can get $\boldsymbol{\varepsilon} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$, which is equivalent to check the properties of $\boldsymbol{\varepsilon}$:

$$\begin{aligned}E[\boldsymbol{\varepsilon}] &= \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b} = \boldsymbol{0},\\ Cov(\boldsymbol{\varepsilon}) &= \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T = \boldsymbol{\Psi}.\end{aligned}$$

Because $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are both positive semidefinite, $\boldsymbol{A}$ and $\boldsymbol{b}$ exist if and only if $Rank(\boldsymbol{\Sigma}) \ge Rank(\boldsymbol{\Psi})$. And this condition almost always holds because $\boldsymbol{\Sigma}$ is normally nonsingular.

b).

Check the definition of rotation: $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{T}$, where $\boldsymbol{T}\boldsymbol{T}^T = \boldsymbol{I}$. As the communality $\boldsymbol{h^{*2}} = Diag(\boldsymbol{L}^*\boldsymbol{L}^{*T}) = Diag(\boldsymbol{L}\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{L}^T) = Diag(\boldsymbol{L}\boldsymbol{L}^T) = \boldsymbol{h^2}$ didn't change before and after rotation, we have the following equations:

$$\begin{aligned}h_1^2 &= 0.976^2 + 0.139^2 = 0.025^2 + l_{12}^{*2},\\ h_3^2 &= l_{31}^2 + 0.032^2 = 0.131^2 + 0.972^2,\\ h_4^2 &= 0.535^2 + 0.739^2 = l_{41}^{*2} + 0.405^2.\end{aligned}$$

So, given they are all positive, $l_{12}^* = 0.986$, $l_{31} = 0.980$, $l_{41}^* = 0.818$.

c).

Since standardized data were adopted, the variability in Flavour $\sigma_3^2 = 1$, and the proportion of variance of Flavour cannot be explained by the common factors should be $\psi_3/\sigma_3^2 = (1 - h_3^2)/1 = 3.8\%$.

d).

Proportion of total variances explained by the first factor:

$$\frac{\sum_{i=1}^5 l_{i1}^2}{5} = \frac{0.976^2 + 0.15^2 + 0.98^2 + 0.535^2 + 0.146^2}{5} = \frac{2.244}{5} = 44.8\%.$$

Proportion of total variances explained by the 2 factors:

$$\frac{\sum_{j=1}^2\sum_{i=1}^5 l_{ij}^2}{5} = 44.8\% + \frac{0.139^2 + 0.86^2 + 0.032^2 + 0.739^2 + 0.963^2}{5} = 89.5\%.$$

e).

	Taste	Money's worth	Flavour	Good for snack	Nutrition
F1	0.025	0.873	0.131	0.818	0.974
F2	0.986	0.005	0.972	0.405	-0.016

Use 0.5 or 0.6 as threshold to determine whether a correlation is large or small.

Factor 1 represents Money's worthiness, Good for snack and Nutrition. Factor 2 represents Taste and Flavour.

f).

H_0: $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}$ vs H_1: $\boldsymbol{\Sigma} =$ any other positive definite matrix.

Likelihood ratio test statistic $R = n\ln(|\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}}|/|\boldsymbol{S}_n|) = 150\ln(3.56/3.48) = 3.409$.

Critical value $\chi^2_{\alpha,[(p-m)^2-p-m]/2} = \chi^2_{0.05,[(5-2)^2-5-2]/2} = \chi^2_{0.05,1} = 3.8415$.

As $R < \chi^2_{0.05,1}$, we do not reject H_0 at the 5% level of significance, i.e. the 2-factor model is adequate.