Chapter 11

Factor Analysis

11.1 Introduction

Factor analysis gives an approximation of the dispersion matrix in terms of $m < p$ unobservable random quantities called *factors*. Typically factor analysis is useful if the p random variables can be placed into a few groups of variables with fairly high correlation such that the variables within the group are not highly correlated with variables outside of the group. Let m be the number of groups. Then the hope is that the kth group can be explained by the kth factor. For example, if the $p = 6$ random variables consist of three head measurements and height, arm length and leg length, then perhaps the three head measurements are highly correlated and the three other measurements are highly correlated. Then there would be $m = 2$ groups corresponding to a "head measurement" factor and a "length" factor.

Some notation is needed before presenting the model. When the eigenvalue λ_i of $\boldsymbol{\Sigma}$ is unique, there are two standardized eigenvectors: $\boldsymbol{e}_i$ and $-\boldsymbol{e}_i$. The literature sometimes states that the standardized eigenvectors are "unique up to sign." Assume $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$. If $\hat{\boldsymbol{\Sigma}} \stackrel{P}{\to} c\boldsymbol{\Sigma}$ for some positive constant c, then by the spectral decomposition theorem, $\hat{\boldsymbol{\Sigma}} = \sum_{i=1}^p \hat{\lambda}_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^T \stackrel{P}{\to} c \sum_{i=1}^p \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T = c\boldsymbol{\Sigma}$, and $\hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^T \stackrel{P}{\to} \boldsymbol{e}_i \boldsymbol{e}_i^T$ for $i = 1, ..., p$ by Theorem 6.2 since $\boldsymbol{e}_i \boldsymbol{e}_i^T = (-\boldsymbol{e}_i)(-\boldsymbol{e}_i)^T$.

The factor analysis approximation of the dispersion matrix $\boldsymbol{\Sigma} \approx \boldsymbol{\Sigma}_P$ uses the first m terms of the spectral decomposition of $\boldsymbol{\Sigma}$ and a diagonal matrix $\boldsymbol{\Psi}$ so that the approximation is exact for the diagonal elements: $\boldsymbol{\Sigma}_{ii} = \boldsymbol{\Sigma}_{P,ii}$. Let the ith column of the $p \times m$ matrix $\boldsymbol{L}$ be $\sqrt{\lambda_i}\boldsymbol{e}_i$ where $m < p$.

Then $\boldsymbol{L} = \left[\sqrt{\lambda_1}\boldsymbol{e}_1 \quad \sqrt{\lambda_2}\boldsymbol{e}_2 \quad \ldots \quad \sqrt{\lambda_m}\boldsymbol{e}_m \right]$. Then $\boldsymbol{\Sigma} = \sum_{i=1}^m \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T + \sum_{i=m+1}^p \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T = \boldsymbol{L}\boldsymbol{L}^T + \sum_{i=m+1}^p \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^T \approx \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi} \equiv \boldsymbol{\Sigma}_P$ where $\boldsymbol{\Psi} = diag(\psi_1, ..., \psi_p)$ and $\boldsymbol{\Sigma}_{ii} = \boldsymbol{\Sigma}_{P,ii}$. Hence $(\boldsymbol{L}\boldsymbol{L}^T)_{ii} + \psi_i = \boldsymbol{\Sigma}_{ii}$.

Definition 11.1. The *orthogonal factor analysis model* is $\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon}$ where the $p \times 1$ random vector $\boldsymbol{x} = (X_1, ..., X_p)$, the $p \times m$ *matrix of factor loadings* $\boldsymbol{L} = ((l_{ij}))$, the $m \times 1$ random vector of *common factors* is $\boldsymbol{F} = (F_1, ..., F_m)^T$ and the $p \times 1$ error vector is $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_p)^T$. The ϵ_i are called errors or *specific factors*. The dispersion structure is $\boldsymbol{\Sigma} \approx \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi} = \boldsymbol{\Sigma}_P$ with equality for the diagonal elements. Hence $\boldsymbol{\Sigma}_{ii} = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i = h_i^2 + \psi_i$ where $h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2$ is called the ith *communality*. The model has $X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \cdots + l_{im}F_m + \epsilon_m$ for $i = 1, ..., p$. The *loading* of the ith variable on the jth factor $= l_{ij}$.

Data often does not have this structure, so an important question in whether the factor analysis structure is reasonable. Note that if $\boldsymbol{\Sigma}$ is the covariance matrix, then $V(X_i) = \sigma_{ii} = \boldsymbol{\Sigma}_{ii} = h_i^2 + \psi_i$. $\boldsymbol{L}, \boldsymbol{F}, \boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ are unobservable. When $\boldsymbol{\Sigma}$ is the covariance matrix, assume that $E(\boldsymbol{F}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{F}) = \boldsymbol{I}_m$, $E(\boldsymbol{\epsilon} = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$ and that $\boldsymbol{F}$ and $\boldsymbol{\epsilon}$ are independent. Then $\text{Cov}(\boldsymbol{x}, \boldsymbol{F}) = \boldsymbol{L}$ or $\text{Cov}(X_i, F_j) = l_{ij}$, and $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi} = \boldsymbol{\Sigma}_P$.

Let the ith column of the $p \times m$ matrix $\hat{\boldsymbol{L}}$ be $\sqrt{\hat{\lambda}_i}\hat{\boldsymbol{e}}_i$ where $m < p$. Then $\hat{\boldsymbol{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\boldsymbol{e}}_1 \quad \sqrt{\hat{\lambda}_2}\hat{\boldsymbol{e}}_2 \quad \ldots \quad \sqrt{\hat{\lambda}_m}\hat{\boldsymbol{e}}_m \right]$. Then $\hat{\boldsymbol{\Sigma}} = \sum_{i=1}^m \hat{\lambda}_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^T + \sum_{i=m+1}^p \hat{\lambda}_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^T = \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \sum_{i=m+1}^p \hat{\lambda}_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^T \approx \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}} \equiv \hat{\boldsymbol{\Sigma}}_P$ where $\hat{\boldsymbol{\Psi}} = diag(\hat{\psi}_1, ..., \hat{\psi}_p)$ and $\hat{\boldsymbol{\Sigma}}_{ii} = \hat{\boldsymbol{\Sigma}}_{P,ii}$. Hence $(\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T)_{ii} + \hat{\psi}_i = \hat{\boldsymbol{\Sigma}}_{ii}$.

Definition 11.2. The *principal component factor analysis* uses the approximation $\hat{\boldsymbol{\Sigma}} \approx \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}}$. $\hat{\boldsymbol{L}}$ is called the *matrix of estimated factor loadings.* The ith *estimated communality* $\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2$ for $i = 1, .., p$. The kth column $\sqrt{\hat{\lambda}_k}\hat{\boldsymbol{e}}_k$ of $\hat{\boldsymbol{L}}$ gives the estimated factor loadings for factor F_k. These estimated factor loadings do not change as m is increased. If $\boldsymbol{\Gamma}$ is an orthogonal matrix, then $\hat{\boldsymbol{L}}^* = \hat{\boldsymbol{L}}\boldsymbol{\Gamma}$ is also a matrix of estimated factor loadings, and $\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T = \hat{\boldsymbol{L}}^*(\hat{\boldsymbol{L}}^*)^T$. The communalities are unaffected by the choice of $\boldsymbol{\Gamma}$.

Rule of thumb 11.1. To use factor analysis, assume the DD plot and subplots of the scatterplot matrix are linear. Want $n > 10p$ for classical factor analysis and $n > 20p$ for robust factor analysis that uses FCH, RFCH or RMVN. For classical factor analysis, use the correlation matrix $\boldsymbol{R}$ instead of the covariance matrix $\boldsymbol{S}$ if $\max_{i=1,...,p} S_i^2 / \min_{i=1,...,p} S_i^2 > 2$. If $\boldsymbol{S}$ is used, also do a factor analysis using $\boldsymbol{R}$. Want the *proportion of the trace explained* by the first m factors $= \sum_{i=1}^m \hat{\lambda}_i / \sum_{j=1}^p \hat{\lambda}_j = \sum_{i=1}^m \hat{\lambda}_i / tr(\hat{\boldsymbol{\Sigma}}) > 0.7$. Want $m < \min(10, p)$. Suppose $(T, \hat{\boldsymbol{\Sigma}})$ is the estimator of multivariate location and dispersion. Make a plot of $D_i(T, \hat{\boldsymbol{\Sigma}}_P)$ versus $D_i(T, \hat{\boldsymbol{\Sigma}})$ with the identity line that has unit slope and zero intercept added as a visual aid. If $\hat{\boldsymbol{\Sigma}}_P$ is an adequate approximation of $\hat{\boldsymbol{\Sigma}}$, then the plotted points should cluster tightly about the identity line.

11.2 Robust Factor Analysis

Robust factor analysis can be done using the FCH, RFCH or RMVN dispersion estimator as $\hat{\boldsymbol{\Sigma}}$. Under (E1) the robust factor analysis has $\hat{\boldsymbol{\Sigma}} \xrightarrow{P} c\boldsymbol{\Sigma}$ while $\boldsymbol{S} \xrightarrow{P} c_X \boldsymbol{\Sigma}$. If the generalized correlation matrix is used as $\hat{\boldsymbol{\Sigma}}$, then the classical and robust methods both satisfy $\hat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\rho}$. The RMVN method is easy to program since it is the classical factor analysis applied to the RMVN subset.

11.3 Summary

1) Factor analysis is use to write $\hat{\boldsymbol{\Sigma}} \approx \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}} = \hat{\boldsymbol{\Sigma}}_F$. Factor analysis clusters variables into groups called factors and suggests that the $m < p$ factors explain the dispersion more simply than $X_1, ..., X_p$. $\hat{\boldsymbol{L}} = [\boldsymbol{L}_1, ..., \boldsymbol{L}_m]$ is the matrix of factor loadings.

2) Factor analysis output is a lot like PCA output, but replace PC1, ..., PCp by Factor 1, ..., Factor m:

Factor 1	Factor 2	$\cdots$	Factor m
$\hat{\boldsymbol{L}}_1$	$\hat{\boldsymbol{L}}_2$	$\cdots$	$\hat{\boldsymbol{L}}_m$

3) To try to explain Factor j, look at entries in $\hat{\boldsymbol{L}}_j$ that are large in magnitude and ignore entries close to zero. Sometimes only one entry is large. Sometimes all of the large entries have approximately the same size and sign, then the Factor is interpreted as an average of these entrees. If all of the large entries have approximately the same size but different signs then the Factor is interpreted as the sum of the variables with the positive sign −

the sum of the variables with a minus sign. Thus if exactly two entries are of similar large magnitude but of different sign, the Factor is interpreted as a difference of the two entrees. If there are $k \geq 2$ large entrees that differ in magnitude, then the Factor is interpreted as a linear combination of the corresponding variables.

4) The proportion of variance explained and cumulative proportion of variance explained are interpreted as for PCA. Use the k factor model if the proportion of the variance explained by the first k Factors is larger than some percentage such as 50%, 60%, 70%, 80% or 90%.

5) For a k factor model, want the degrees of freedom $d \geq 0$ where $d = 0.5(p-k)^2 - 0.5(p+k)$.

6) If the 1 factor model is not adequate, R will give a test for whether a k factor model is sufficient. A k factor model with $pval < 0.05$ is not sufficient: more factors are needed. A k factor model with $pval > 0.05$ is sufficient.

7) Let $\hat{\boldsymbol{\Gamma}}$ be an orthogonal matrix. The $\hat{\boldsymbol{L}}_\Gamma \hat{\boldsymbol{L}}_\Gamma^T = \hat{\boldsymbol{L}}\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}^T\hat{\boldsymbol{L}}^T = \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T$. The varimax and promax rotations seek $\hat{\boldsymbol{\Gamma}}$ such that $\hat{\boldsymbol{L}}_\Gamma = \hat{\boldsymbol{L}}\hat{\boldsymbol{\Gamma}}$ has loadings that are easier to interpret than the loadings of $\hat{\boldsymbol{L}}$. The promax rotation attempts to produce loading with a lot of zeroes.

11.4 Complements

Kosfeld (1996) does factor analysis with the DGK estimator.

11.5 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

```
Loadings:
               Factor1 Factor2
height          0.872
arm.span        0.973
forearm         0.938
lower.leg       0.876
weight                  0.961
bitro.diameter          0.803
```

```
chest.girth              0.796
chest.width     0.125   0.611

               Factor1 Factor2
SS loadings      3.375   2.589
Proportion Var   0.422   0.324
Cumulative Var   0.422   0.745
```

11.1*. The above output is for the factor analysis using a correlation matrix of eight physical measurements on 305 girls between ages seven and seventeen.

a) What is the cumulative variance explained by the 2 factors?

b) Which factor has a nonzero loading for weight?

c) Explain Factor 2.

```
factanal(marry,factors=2,rotation="promax")

Uniquenesses:  pop    mmen    mwmn   mmilmen  milwmn
              0.010   0.005   0.005   0.005   0.005

Loadings:Factor1 Factor2
pop      0.986
mmen     1.003
mwmn     1.003
mmilmen          0.965
milwmn           0.958

               Factor1 Factor2
SS loadings      2.995   1.850
Proportion Var   0.599   0.370
Cumulative Var   0.599   0.969
```

11.2. The above output is for a factor analysis of the Hebbler (1847) data from the the 1843 Prussia census. Sometimes if the wife or husband was not at the household, then s/he would not be counted. $X_1 = pop =$ population of the district in 1843, $X_2 = mmen =$ number of married civilian men in the district, $X_3 = mwmn =$ number of women married to civilians in the district, $X_4 = mmilmen =$ number of married military men in the district, and $x_5 = milwmn$ = number of women married to military men in the district.

a) What is the cumulative variance explained by the 2 factors?

b) Explain Factor 1.

c) Explain Factor 2.

```
Uniquenesses:
    age  breadth cephalic   circum   headht   height      len     size  cbrainy
  0.005    0.005    0.005    0.142    0.005    0.303    0.005    0.005    0.366

Loadings:
         Factor1 Factor2 Factor3 Factor4
log(age)          1.026
breadth   0.874           0.461  -0.142
cephalic -0.115           1.020
circum    0.849   0.113
headht                            0.965
height    0.202   0.597           0.204
len       1.109          -0.363  -0.156
size      0.805                   0.231
brainwt   0.642  -0.262           0.296

               Factor1 Factor2 Factor3 Factor4
SS loadings      3.833   1.491   1.389   1.161
Proportion Var   0.426   0.166   0.154   0.129
Cumulative Var   0.426   0.592   0.746   0.875
```

11.3. The above output is for the factor analysis of the Gladstone (1905-6) data. The variables included *log*(*age*) and *height* and 7 head measurements breadth, cephalic, circum, headht, len, size, and brain weight.

a) What is the cumulative variance explained by the 4 factors?

b) Which factor has a nonzero loading for log(age)?

c) Explain Factor 3.

R/Splus Problems

Warning: Use the command *source("G:/mpack.txt")* **to download the programs. See Preface or Section 15.2.** Typing the name of the `mpack` function, eg *ddplot*, will display the code for the function. Use the `args` command, eg *args(ddplot)*, to display the needed arguments for the function.

11.4. The Buxton data has 5 massive outliers in variables len and buxy = height.

a) The *R* commands for this part do a factor analysis on the Buxton data using the sample covariance matrix. Copy and paste the output into *Word*.

i) Which variables have nonzero loadings for factor 1?

ii) Which variables have nonzero loadings for factor 2?

iii) What is the cumulative variance explained by the two factors?

b) The *R* commands for this part do a factor analysis on the Buxton data using the RMVN dispersion matrix. Copy and paste the output into *Word*.

i) Which variables have nonzero loadings for factor 1?

ii) Which variables have nonzero loadings for factor 2?

iii) What is the cumulative variance explained by the two factors?