Principal components and factor analysis

Wan Nor Arifin

Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.

email: wnarifin@usm.my

December 2, 2018

Introduction

Factoring

- *Group* things that have common concept.
- *Simplify*.
- Factoring = Grouping.
- Factor = Construct = Concept.

Orange, motorcycle, bus, durian, banana, car

Anything in common?

Intuitive factoring

Group them

[Orange, durian, banana]

[Motorcycle, bus, car]

into two groups

Name the group

Fruits	Motor vehicle
Orange	Motorcycle
Durian	Bus
Banana	Car

factor out the common concept

Factoring

- find out correlated variables from analysis of correlation matrix.
- manageable for small number of variables.
- impossible for large number of variables.

Factoring

We consider two methods:

- Principal components analysis (PCA).
- Factor analysis.

These are applied to numerical variables.

Some preliminary considerations

Suitability of data for the analysis:

- there must be correlations between the variables.

Judged by:

- Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy (MSA). KMO > 0.7 required.
- Bartlet's test of sphericity. P-value < 0.05 indicates presence of correlations.

Some preliminary considerations

Number of factors to extract:

- Factors with Eigenvalues > 1 (Kaiser's rule).
- Cattell's scree test.
- Parallel analysis.

Principal components analysis

Data reduction

- when we deal with many variables (multivariate) – may need to reduce the number of variables.
- questionnaire – hundred of variables.
- genome study – few thousands variables.
- data reduction – combine/group related variables into smaller sets of variables.
- can reduce burden of analysis and interpretation.

PCA

- a data reduction technique, basically a descriptive method.
- uncover most important principal components from the data.
- group correlated variables $\rightarrow$ uncorrelated principal components.
- many variables $\rightarrow$ few component scores $\rightarrow$ subsequent analysis e.g. multiple linear regression.

Component variances

Basics:

- say we have p variables $\rightarrow$ k components.
- sum of all variable variaces $=$ sum of component variances.
- what PCA does, it extracts out few principal components that can explain (as good as) the variances of all the variables.

PCA

Research questions:

- How many principal components are there?
- Strength of relationship between variables and the components?
- % of variance extracted by the components?

PCA

Applications:

- combine 100 IVs in a multiple linear regression into a number of smaller principal components.
- extract attitude factors from 60 items in a questionnaire analysis.

PCA

Component loading:

- variable-component relationship.
- values > 0.3.

Practical

pca_efa.R

Factor analysis

Factor analysis

- latent variable model analysis.
- group correlated items (in a measurement scale).
- factor out latent (unobserved) factors cause the correlation between the items.

Factor analysis

- in fields like psychology, we cannot observe directly psychological states, thus measured indirectly in form items.
- e.g. depression:
  - depression causes symptoms of depression.
  - depression (latent) is measured indirectly by items representing its symptoms.
  - prove the symptoms are correlated to each other, representing the concept of depression *rightarrow* factor analysis.

Common factor model

Common Factors + Measurement Error

Classification:

- Exploratory factor analysis.
- Confirmatory factor analysis.

Exploratory factor analysis (EFA)

EFA

- explorative method.
- e.g. at early of questionnaire development.
- theory generating.

EFA

Research questions:

- How many factors are there?
- Strength of relationship between items and the factors?
- Factor correlations?
- % variance explained by the extracted factors?

EFA

Applications:

- psychological scales/questionnaires, e.g. personality, depression, stress etc.
- explore the number of common factors in personality items.

EFA

Extraction methods:

- classical: Principal axis factoring.
- other methods: Maximum likelihood, image analysis, alpha analysis.

Factor loading:

- item-factor relationship.
- values > 0.3.

EFA

To simplify EFA results, may need factor rotation.

Types of rotation:

- Orthogonal – uncorrelated factors.
  - Varimax, Quartimax, Equamax.
- Oblique – correlated factors.
  - Oblimin, Promax.

Obtain clear factors and factor loadings.

Confirmatory factor analysis (CFA)

Structural equation modeling (SEM):

- measurement model – CFA.
- structural model – path analysis.

Next lecture.

Practical

pca_efa.R

Bibliography

Gorsuch, R. L. (2014). *Exploratory factor analysis*. New York: Routledge.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. New Jersey: Prentice Hall.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences (5th eds.)*. New York: Taylor & Francis Group.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics (5th ed.)*. USA: Pearson.