Factor Analysis

Son Nguyen

Principal Components Analysis

Principal Component

Variance due to what one intended to measure
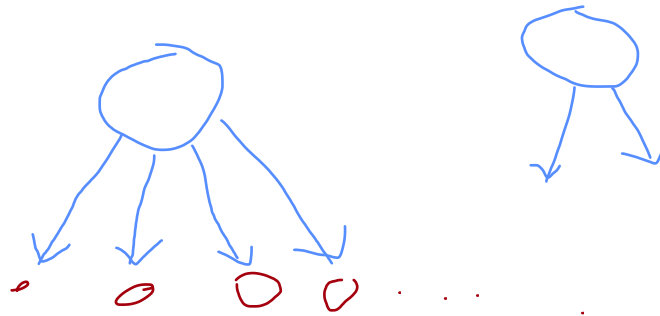
Other sources of variance, including measurement error
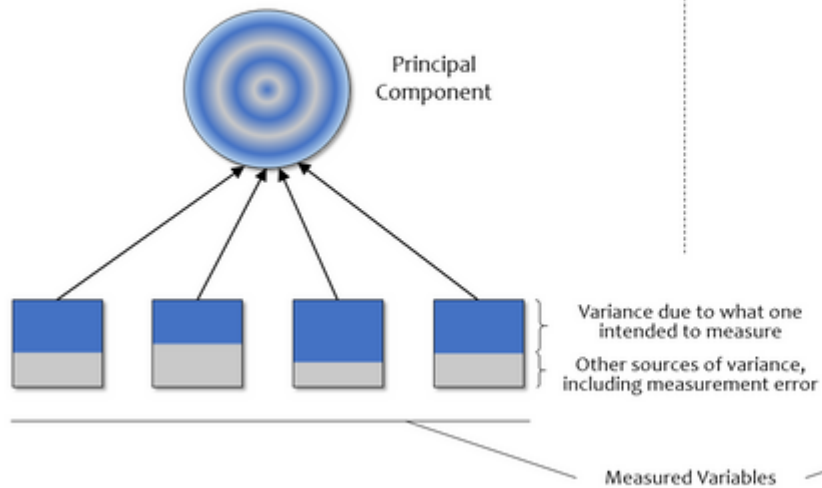
Exploratory Factor Analysis

Latent Variable

Measured Variables
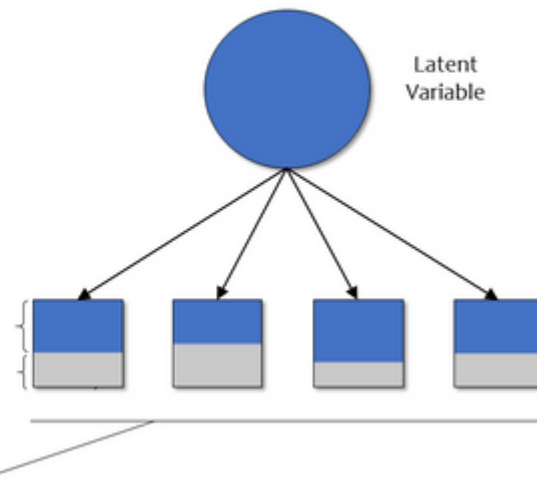
```
library(tidyverse)
df = read_csv('FactorAnalysis.csv')

fa = factanal(df, factors = 4)
```

```
fa$loadings
```

```
##
## Loadings:
##                 Factor1 Factor2 Factor3 Factor4
## Academic record  0.750   0.177   0.209   0.189
## Appearance       0.264   0.299   0.646   0.217
## Communication    0.170   0.813   0.249   0.239
## Company Fit      0.539   0.513   0.221   0.192
## Experience       0.704           0.197   0.296
## Job Fit          0.621   0.394   0.234   0.197
## Letter           0.253   0.236           0.800
## Likeability      0.250   0.291   0.697   0.174
## Organization     0.154   0.907   0.311
## Potential        0.886   0.213   0.261   0.123
## Resume           0.275   0.107   0.304   0.903
## Self-Confidence  0.255   0.189   0.876
##
##                 Factor1 Factor2 Factor3 Factor4
## SS loadings       2.908   2.256   2.181   1.817
## Proportion Var    0.242   0.188   0.182   0.151
## Cumulative Var    0.242   0.430   0.612   0.764
```

Loadings

- The loadings are the contribution of each original variable to the factor

- Variables with a high loading are well explained by the factor

```
fa$uniquenesses
```

```
## Academic record        Appearance     Communication        Company Fit        Experience
##      0.32660908        0.37645433        0.19040778        0.36062557        0.37725928
##          Job Fit            Letter        Likeability      Organization         Potential
##      0.36546646        0.23660719        0.33608162        0.05233760        0.08684769
##          Resume  Self-Confidence
##      0.00500000        0.12408274
```

Uniquenesses

- Uniquenesses: Uniqueness is the variance that is 'unique' to the variable and not shared with other variables

- Uniqueness sometimes referred to as noise

- The greater 'uniqueness' the lower the relevance of the variable in the factor model

- Values more than 0.6 are usually considered high

```
df ← read_csv("https://userpage.fu-berlin.de/soga/data/raw-data/food-texture.csv")
df = df[,c(2:6)]
food_fa ← factanal(df, factors = 2)
```

```
n ← 100
R ← matrix(c(1, 0.8, .9 , 0, 0, 0,
             .8, 1,   .9 , 0, 0, 0,
             .9, 0.9, 1 , 0, 0, 0,
             0,  0,   0 , 1, 0.8,  .9,
             0,  0,   0 , .8, 1, .9,
             0,  0,   0 , .9, .9, 1),
           nrow = 6, ncol = 6, byrow = TRUE)

mu ← c(Tennis = 100, Table_tennis = 100, Pickle_ball = 100, Soccer = 100, Football = 100, Baseball =
df = as_tibble(mvtnorm::rmvnorm(n, mean = mu, sigma = R))
factanal(df, 2)
```

```
##
## Call:
## factanal(x = df, factors = 2)
##
## Uniquenesses:
##       Tennis Table_tennis  Pickle_ball       Soccer     Football     Baseball
##        0.149        0.172        0.005        0.188        0.137        0.038
##
## Loadings:
##              Factor1 Factor2
## Tennis        0.919
## Table_tennis  0.909
## Pickle_ball   0.997
## Soccer                0.900
```