SEM 1: Confirmatory Factor Analysis

Week 4 - Ordered categorical data

Sacha Epskamp

2020

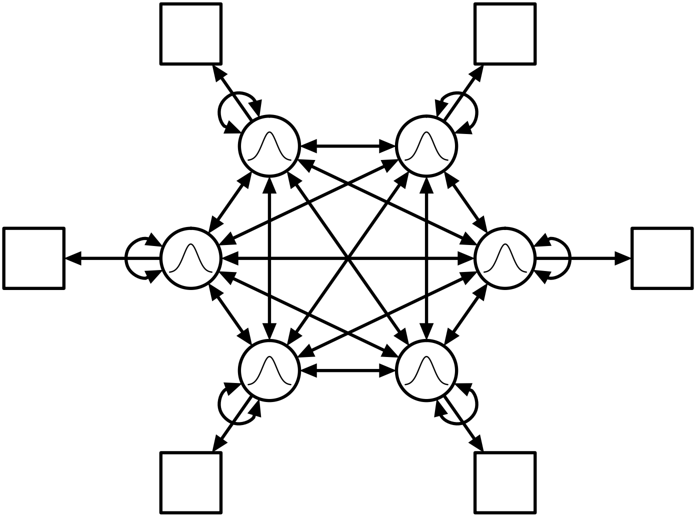
I see myself as someone who is talkative

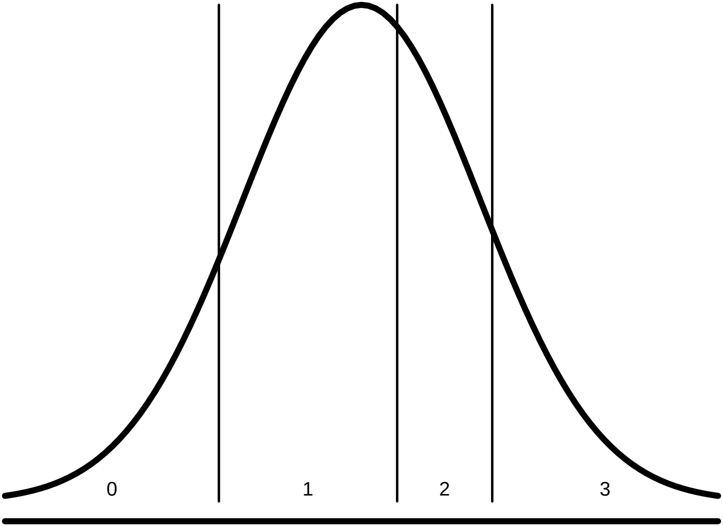
Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree Strongly
1	2	3	4	5

- If data is ordinal and consists of only a few levels of measurement data cannot be assumed normal
  - Roughly less than five categories (Rhemtulla, Brosseau-Liard, & Savalei,2012).
- In this case a typical solution is to use threshold models
  - Polychoric correlation if both variables are ordinal
  - Polyserial correlation if one item is ordinal and the other is continuous
- Estimation not via maximum likelihood but typically via (diagonally) weighted least squares:

$$F_{\mathrm{WLS}} = (\boldsymbol{s} - \boldsymbol{\sigma})^{\top} \boldsymbol{W}^{-1} (\boldsymbol{s} - \boldsymbol{\sigma})$$

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika 49*, 115-132.




0
1
2
3

```r
set.seed(1)
# Setup:
sampleSize <- 1000
cor <- 0.5
thresh1 <- c(-2,0,2)
thresh2 <- c(-1,0.5,1.6)

# Generate data:
library("mvtnorm")
corMat <- matrix(c(1,0.5,0.5,1),2,2)
Data <- as.data.frame(rmvnorm(sampleSize, sigma = corMat))
names(Data) <- c("y1","y2")

# Make catagorical:
Data[,1] <- as.numeric(cut(Data[,1],breaks = c(-Inf,thresh1,Inf)))
Data[,2] <- as.numeric(cut(Data[,2],breaks = c(-Inf,thresh2,Inf)))
```

```r
head(Data,5)

##   y1 y2
## 1  2  2
## 2  2  3
## 3  3  2
## 4  3  3
## 5  3  2
```

```r
table(Data)

##    y2
## y1    1   2   3   4
##   1  17  16   2   0
##   2 123 266  77   8
##   3  20 233 168  42
##   4   2   6  13   7
```

Note: zeroes in marginal crosstables might be problematic..

```r
# Pearson correlation:
cor(Data[,1], Data[,2])

## [1] 0.4076942

# Polychoric correlation:
library("lavaan")
lavCor(Data, ordered = c("y1","y2"))

##    y1    y2
## y1 1.000
## y2 0.499 1.000
```

No thresholds:

```
Model <- '
f1 =~ y1
f2 =~ y2
'
fit <- cfa(Model, Data, std.lv = TRUE)
parameterEstimates(fit)

##   lhs op rhs   est    se      z pvalue ci.lower ci.upper
## 1  f1 =~  y1 0.613 0.014 44.721      0    0.586    0.640
## 2  f2 =~  y2 0.778 0.017 44.721      0    0.744    0.812
## 3  y1 ~~  y1 0.000 0.000     NA     NA    0.000    0.000
## 4  y2 ~~  y2 0.000 0.000     NA     NA    0.000    0.000
## 5  f1 ~~  f1 1.000 0.000     NA     NA    1.000    1.000
## 6  f2 ~~  f2 1.000 0.000     NA     NA    1.000    1.000
## 7  f1 ~~  f2 0.408 0.026 15.463      0    0.356    0.459
```

With thresholds:

```
Model <- 'y1 ~~ y2'
fit <- cfa(Model, Data, std.lv = TRUE,
           ordered = c("y1","y2"))
parameterEstimates(fit)

##    lhs  op rhs    est    se       z pvalue ci.lower ci.upper
## 1   y1  ~~  y2  0.499 0.029  17.143  0.000    0.442    0.556
## 2   y1   |  t1 -1.812 0.075 -24.079  0.000   -1.959   -1.664
## 3   y1   |  t2  0.023 0.040   0.569  0.569   -0.055    0.100
## 4   y1   |  t3  1.911 0.081  23.524  0.000    1.752    2.070
## 5   y2   |  t1 -0.986 0.048 -20.753  0.000   -1.079   -0.893
## 6   y2   |  t2  0.476 0.041  11.519  0.000    0.395    0.557
## 7   y2   |  t3  1.580 0.064  24.653  0.000    1.455    1.706
## 8   y1  ~~  y1  1.000 0.000      NA     NA    1.000    1.000
## 9   y2  ~~  y2  1.000 0.000      NA     NA    1.000    1.000
## 10  y1 ~*~  y1  1.000 0.000      NA     NA    1.000    1.000
## 11  y2 ~*~  y2  1.000 0.000      NA     NA    1.000    1.000
## 12  y1  ~1      0.000 0.000      NA     NA    0.000    0.000
## 13  y2  ~1      0.000 0.000      NA     NA    0.000    0.000
```