Course: Multivariate statistics (AUT23)

Chapter 10: Structural equation modelling

11.16 Time to practice on your own

11.16.1 STM with multiple latent variables

We want to create a structural model that relates multiple latent variables using social science data describing the effect of student background on academic achievement.

We create a measurement model by defining each latent variable:

- Adjustment: measured by motivation, harmony, and stability.
- Risk: measured by verbal, negative parent psychology, and socioeconomic status.
- Achievement: measured by reading, arithmetic, and spelling.

We also define the regression path where achievement will be the combination of adjustment and risk.

- ➢ Show the code

Next, we fit the model to the data using sem() from lavaan. Followed by a summary, including model fit. How do you interpret these results?

- ➢ Show the code

Note that further analysis using these data can be found in the Introduction to structural equation modeling (SEM) in R with Lavaan from the UCLA: Statistical Consulting Group.

11.16.2 SEM framework

Based on the following flowchart (original can be found here), state whether the following statements are true or false:

- SEM encompasses a broad range of linear models and combines simultaneous linear equations with latent variable modeling.
- ➢ Solution

- Multivariate regression means that there is always more than one exogenous predictor in my model.
- ➢ Solution

- Structural regression models the regression paths only among latent variables.
- ➢ Solution

Chapter 10: Structural equation modelling (answers)

11.16 Time to practice on your own

11.16.1 STM with multiple latent variables

We want to create a structural model that relates multiple latent variables using social science data describing the effect of student background on academic achievement.

We create a measurement model by defining each latent variable:

- Adjustment: measured by motivation, harmony, and stability.
- Risk: measured by verbal, negative parent psychology, and socioeconomic status.
- Achievement: measured by reading, arithmetic, and spelling.

We also define the regression path where achievement will be the combination of adjustment and risk.

➢ Show the code

```
dat <- read.csv("https://stats.idre.ucla.edu/wp-content/uploads/2021/02/worland5.csv")

mod <- '

 # measurement model

 adjust =~ motiv + harm + stabi

 risk =~ verbal + ppsych + ses

 achieve =~ read + arith + spell

 # regression

 achieve ~ adjust + risk

'
```

Next, we fit the model to the data using sem() from lavaan. Followed by a summary, including model fit. How do you interpret these results?

➢ Show the code

```
fit <- lavaan::sem(mod, data=dat)

summary(fit, fit.measures=TRUE)

## lavaan 0.6.15 ended normally after 130 iterations

##

##   Estimator                          ML

##   Optimization method                NLMINB

##   Number of model parameters             21
```

```
##
##   Number of observations                           500
##
## Model Test User Model:
##
##   Test statistic                              148.982
##   Degrees of freedom                               24
##   P-value (Chi-square)                          0.000
##
## Model Test Baseline Model:
##
##   Test statistic                             2597.972
##   Degrees of freedom                               36
##   P-value                                       0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                   0.951
##   Tucker-Lewis Index (TLI)                      0.927
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)           -15517.857
##   Loglikelihood unrestricted model (H1)   -15443.366
##
##   Akaike (AIC)                             31077.713
##   Bayesian (BIC)                           31166.220
##   Sample-size adjusted Bayesian (SABIC)    31099.565
##
```

```
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.102
##   90 Percent confidence interval - lower         0.087
##   90 Percent confidence interval - upper         0.118
##   P-value H_0: RMSEA <= 0.050                    0.000
##   P-value H_0: RMSEA >= 0.080                    0.990
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.041
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   adjust =~
##     motiv             1.000
##     harm              0.884    0.041   21.774    0.000
##     stabi             0.695    0.043   15.987    0.000
##   risk =~
##     verbal            1.000
##     ppsych           -0.770    0.075  -10.223    0.000
##     ses               0.807    0.076   10.607    0.000
##   achieve =~
```

```
##    read           1.000
##    arith          0.837    0.034   24.437    0.000
##    spell          0.976    0.028   34.338    0.000
##
## Regressions:
##                 Estimate  Std.Err  z-value  P(>|z|)
##   achieve ~
##    adjust         0.375    0.046    8.085    0.000
##    risk           0.724    0.078    9.253    0.000
##
## Covariances:
##                 Estimate  Std.Err  z-value  P(>|z|)
##   adjust ~~
##    risk          32.098    4.320    7.431    0.000
##
## Variances:
##                 Estimate  Std.Err  z-value  P(>|z|)
##   .motiv         12.870    2.852    4.512    0.000
##   .harm          31.805    2.973   10.698    0.000
##   .stabi         57.836    3.990   14.494    0.000
##   .verbal        46.239    4.788    9.658    0.000
##   .ppsych        68.033    5.068   13.425    0.000
##   .ses           64.916    4.975   13.048    0.000
##   .read          11.372    1.608    7.074    0.000
##   .arith         37.818    2.680   14.109    0.000
##   .spell         15.560    1.699    9.160    0.000
##    adjust        86.930    6.830   12.727    0.000
##    risk          53.561    6.757    7.927    0.000
##   .achieve       30.685    3.449    8.896    0.000
```

Note that further analysis using these data can be found in the [Introduction to structural equation modeling (SEM) in R with Lavaan]() from the UCLA: Statistical Consulting Group.

11.16.2 SEM framework

Based on the following flowchart (original can be found [here]()), state whether the following statements are true or false:

- SEM encompasses a broad range of linear models and combines simultaneous linear equations with latent variable modeling.

➢ Solution

True: Multivariate regression and path analysis are simultaneous equations of observed variables; factor analysis is a latent variable model, and structural regression combines the concepts of path analysis with factor analysis.

- Multivariate regression means that there is always more than one exogenous predictor in my model.

➢ Solution

False: Multivariate regression indicates more than one endogenous variable. You can certainly have only one exogenous predictor of multiple endogenous variables.

- Structural regression models the regression paths only among latent variables.

➢ Solution

True: Structural regression defines relationships between latent variables and path analysis defines relationships between observed variables.