HUDM6122 Homework_09

Chenguang Pan

2023-04-26

0.0 Github Address

All my latest homework can be found on Github: https://github.com/cgpan/hudm6122_homeworks . Thanks for checking if interested.

1.0 Exercise 1

The matrix below shows the correlations between ratings on nine statements about pain made by 123 people suffering from extreme pain. Each statement was scored on a scale from 1 to 6, ranging from agreement to disagreement. The nine pain statements were as follows:
MY SOLUTION:

To solve the question, I chose to use the package `lavaan`. This question only provides the correlation matrix, we **have to make a assumption that all variables are standardized**. Then, the covariance's value equals to the correlation and the variance of a centered and standardized variable is equal to one.

```
> # import the correlation matrix first
> corr <- '
+ 1
+ -0.04 1
+ 0.61  -0.07 1
+ 0.45  -0.12 0.59  1
+ 0.03  0.49  0.03  -0.08 1
+ -0.29 0.43  -0.13 -0.21 0.47  1
+ -0.3  0.3 -0.24 -0.19 0.41  0.63  1
+ 0.45  -0.31 0.59  0.63  -0.14 -0.13 -0.26 1
+ 0.30  -0.17 0.32  0.37  -0.24 -0.15 -0.29 0.4 1
+ '
>
> cov_stded <- getCov(corr,
+                     names = c("Q1","Q2","Q3","Q4","Q5",
+                               "Q6","Q7","Q8","Q9"))
> head(cov_stded)
      Q1    Q2    Q3    Q4    Q5    Q6    Q7    Q8    Q9
Q1  1.00 -0.04  0.61  0.45  0.03 -0.29 -0.30  0.45  0.30
Q2 -0.04  1.00 -0.07 -0.12  0.49  0.43  0.30 -0.31 -0.17
Q3  0.61 -0.07  1.00  0.59  0.03 -0.13 -0.24  0.59  0.32
Q4  0.45 -0.12  0.59  1.00 -0.08 -0.21 -0.19  0.63  0.37
Q5  0.03  0.49  0.03 -0.08  1.00  0.47  0.41 -0.14 -0.24
Q6 -0.29  0.43 -0.13 -0.21  0.47  1.00  0.63 -0.13 -0.15
```

The complete correlation matrix looks good. Next, I fitted the SEM model.

```
> library(lavaan)
> model_01 <- 'DocRes =~ Q1 + Q3 + Q4 + Q8
+              PatRes =~ Q2 + Q5 + Q6 + Q7
+              DocRes ~~ PatRes'
> model_01_fit <- sem(model_01, sample.cov = cov_stded, sample.nobs = 123)
> summary(model_01_fit)
lavaan 0.6.15 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        17

  Number of observations                           123

Model Test User Model:

  Test statistic                                63.749
  Degrees of freedom                                19
  P-value (Chi-square)                           0.000

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  DocRes =~
    Q1                1.000
    Q3                1.209    0.169    7.158    0.000
    Q4                1.131    0.165    6.873    0.000
    Q8                1.134    0.165    6.884    0.000
  PatRes =~
    Q2                1.000
    Q5                1.116    0.240    4.649    0.000
    Q6                1.572    0.296    5.304    0.000
    Q7                1.381    0.266    5.198    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  DocRes ~~
    PatRes           -0.107    0.044   -2.409    0.016

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .Q1                0.551    0.083    6.640    0.000
   .Q3                0.347    0.070    4.990    0.000
   .Q4                0.427    0.073    5.812    0.000
   .Q8                0.425    0.073    5.790    0.000
   .Q2                0.714    0.100    7.114    0.000
   .Q5                0.645    0.095    6.816    0.000
   .Q6                0.305    0.087    3.521    0.000
```

```
   .Q7                 0.461    0.086    5.394    0.000
    DocRes             0.441    0.114    3.868    0.000
    PatRes             0.278    0.097    2.854    0.004
```

The results show that the correlation between the two latent variables is $-.107$, $p = .016$. The standard error is .044. Therefore, the 95% confidence interval of this correlation is $-.107 \pm 1.96 \times .044 = [-.193, -.021]$.

2.0 Exercise 2

For the stability of alienation example, fit the model in which the measurement errors for anomia in 1967 and anomia in 1971 are allowed to be correlated.

MY SOLUTION:

This dataset shown in the 7.4.1 section is included in the package `MVA`. I do not know why the authors did not provide an easier way to find the data rather than hiding it in the package!! I found the covariance data and save it as an csv file.

```
> # import the data
> cov_df <-read.csv("alien_cov_size932.csv")
> cov_m <- as.matrix(cov_df)
> # convert the covariance matrix into correlation matrix
> cor_m <- cov2cor(cov_m)
> round(cor_m,2)
     Anomia67 Powles67 Anomia71 Powles71  Educ   SEI
[1,]     1.00     0.66     0.56     0.44 -0.36 -0.30
[2,]     0.66     1.00     0.47     0.52 -0.41 -0.29
[3,]     0.56     0.47     1.00     0.67 -0.35 -0.29
[4,]     0.44     0.52     0.67     1.00 -0.37 -0.28
[5,]    -0.36    -0.41    -0.35    -0.37  1.00  0.54
[6,]    -0.30    -0.29    -0.29    -0.28  0.54  1.00
```

Based on the description from the book, the sample size is 932. Next, I specified the model.

```
> # specificy the model
> model_alien <- '# measurement models
+                 SES =~ 1*Educ + SEI
+                 Alienation67 =~ 1*Anomia67 + Powles67
+                 Alienation71 =~ 1*Anomia71 + Powles71
+
+                 # structural model
+                 Alienation67 ~ SES
+                 Alienation71 ~ Alienation67 + SES
+
+                 # add correlation
+                 Anomia67 ~~ Anomia71
+                 '
> # fit the model
> model_alien_fit <- sem(model_alien,
+                        sample.cov = cov_m,
+                        sample.nobs = 932)
> # check the result
> summary(model_alien_fit, fit.measure = T)
```

```
lavaan 0.6.15 ended normally after 78 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        16

  Number of observations                           932

Model Test User Model:

  Test statistic                                 6.366
  Degrees of freedom                                 5
  P-value (Chi-square)                           0.272

Model Test Baseline Model:

  Test statistic                              2133.833
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.999
  Tucker-Lewis Index (TLI)                       0.998

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -15213.740
  Loglikelihood unrestricted model (H1)     -15210.558

  Akaike (AIC)                               30459.481
  Bayesian (BIC)                             30536.878
  Sample-size adjusted Bayesian (SABIC)      30486.063

Root Mean Square Error of Approximation:

  RMSEA                                          0.017
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.051
  P-value H_0: RMSEA <= 0.050                    0.943
  P-value H_0: RMSEA >= 0.080                    0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.011

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Latent Variables:
```

```
                    Estimate  Std.Err  z-value  P(>|z|)
  SES =~
    Educ               1.000
    SEI                5.163    0.421   12.263    0.000
  Alienation67 =~
    Anomia67           1.000
    Powles67           1.027    0.053   19.319    0.000
  Alienation71 =~
    Anomia71           1.000
    Powles71           0.971    0.049   19.660    0.000

Regressions:
                    Estimate  Std.Err  z-value  P(>|z|)
  Alienation67 ~
    SES               -0.549    0.053  -10.296    0.000
  Alienation71 ~
    Alienation67       0.617    0.050   12.428    0.000
    SES               -0.212    0.049   -4.302    0.000

Covariances:
                    Estimate  Std.Err  z-value  P(>|z|)
 .Anomia67 ~~
   .Anomia71           1.886    0.240    7.869    0.000

Variances:
                    Estimate  Std.Err  z-value  P(>|z|)
   .Educ               2.718    0.515    5.276    0.000
   .SEI              266.579   18.160   14.679    0.000
   .Anomia67           5.067    0.371   13.669    0.000
   .Powles67           2.209    0.317    6.962    0.000
   .Anomia71           4.807    0.395   12.184    0.000
   .Powles71           2.681    0.329    8.141    0.000
    SES                6.872    0.656   10.469    0.000
   .Alienation67       4.700    0.432   10.868    0.000
   .Alienation71       3.862    0.343   11.262    0.000
```

This model looks good. Please note, here I use the variance covariance matrix to fit the model. If we standardize everything in this model, the loadings will become correlation coefficients.

3.0 Exercise 3

Meyer and Bendig (1961) administered the five Thurstone Primary Mental Ability tests, verbal meaning (V), space (S), reasoning (R), numerical (N), and word fluency (W), to 49 boys and 61 girls in grade 8 and again three and a half years later in grade 11. The observed correlation matrix is shown below. Fit a single-factor model to the correlations that allows the factor at time one to be correlated with the factor at time two.

MY SOLUTION:

```
> # import the data
> corr <- '
+ 1
+ 0.37 1
+ 0.42  0.33 1
```

```
+ 0.53  0.14 0.38  1
+ 0.38  0.10  0.20  0.24 1
+ 0.81  0.34  0.49  0.58 0.32  1
+ 0.35  0.65 0.20 -0.04 0.11  0.34  1
+ 0.42  0.32 0.75  0.46  0.26 0.46 0.18 1
+ 0.40  0.14 0.39  0.73  0.19 0.55 0.06 0.54 1
+ 0.24  0.15 0.17  0.15  0.43 0.24 0.15 0.20  0.16 1
+ '
>
> cov_stded <- getCov(corr,
+                     names = c("V1","S1","R1","N1","W1",
+                               "V2","S2","R2","N2","W2"))
> cov_stded
     V1   S1   R1    N1   W1   V2    S2   R2   N2   W2
V1 1.00 0.37 0.42  0.53 0.38 0.81  0.35 0.42 0.40 0.24
S1 0.37 1.00 0.33  0.14 0.10 0.34  0.65 0.32 0.14 0.15
R1 0.42 0.33 1.00  0.38 0.20 0.49  0.20 0.75 0.39 0.17
N1 0.53 0.14 0.38  1.00 0.24 0.58 -0.04 0.46 0.73 0.15
W1 0.38 0.10 0.20  0.24 1.00 0.32  0.11 0.26 0.19 0.43
V2 0.81 0.34 0.49  0.58 0.32 1.00  0.34 0.46 0.55 0.24
S2 0.35 0.65 0.20 -0.04 0.11 0.34  1.00 0.18 0.06 0.15
R2 0.42 0.32 0.75  0.46 0.26 0.46  0.18 1.00 0.54 0.20
N2 0.40 0.14 0.39  0.73 0.19 0.55  0.06 0.54 1.00 0.16
W2 0.24 0.15 0.17  0.15 0.43 0.24  0.15 0.20 0.16 1.00
```

Correlation matrix looks good. Next, I specified the model.

```
> # specify the model
> model_spe <- '# the measurement model
+               LV1 =~ V1 + S1 + R1 + N1 + W1
+               LV2 =~ V2 + S2 + R2 + N2 + W2
+
+               # add the correlation
+               LV1 ~~ LV2 '
> # fit the model
> model_fit <- sem(model_spe, sample.cov = cov_stded, sample.nobs = 110)
>
> # extract the results
> summary(model_fit, fit.measure=T)
lavaan 0.6.15 ended normally after 26 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        21

  Number of observations                           110

Model Test User Model:

  Test statistic                               211.828
  Degrees of freedom                                34
  P-value (Chi-square)                           0.000

Model Test Baseline Model:
```

```
  Test statistic                               553.443
  Degrees of freedom                                45
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.650
  Tucker-Lewis Index (TLI)                       0.537

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -1385.002
  Loglikelihood unrestricted model (H1)      -1279.088

  Akaike (AIC)                                2812.004
  Bayesian (BIC)                              2868.714
  Sample-size adjusted Bayesian (SABIC)       2802.353

Root Mean Square Error of Approximation:

  RMSEA                                          0.218
  90 Percent confidence interval - lower         0.190
  90 Percent confidence interval - upper         0.247
  P-value H_0: RMSEA <= 0.050                    0.000
  P-value H_0: RMSEA >= 0.080                    1.000

Standardized Root Mean Square Residual:

  SRMR                                           0.120

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  LV1 =~
    V1                1.000
    S1                0.523    0.122    4.270    0.000
    R1                0.827    0.118    6.991    0.000
    N1                0.871    0.118    7.414    0.000
    W1                0.466    0.123    3.793    0.000
  LV2 =~
    V2                1.000
    S2                0.390    0.113    3.451    0.001
    R2                0.790    0.102    7.759    0.000
    N2                0.735    0.104    7.078    0.000
    W2                0.351    0.114    3.086    0.002

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
```

```
  LV1 ~~
    LV2               0.706    0.115    6.146    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .V1                0.419    0.063    6.604    0.000
   .S1                0.835    0.113    7.405    0.000
   .R1                0.600    0.083    7.215    0.000
   .N1                0.557    0.078    7.128    0.000
   .W1                0.866    0.117    7.411    0.000
   .V2                0.305    0.054    5.629    0.000
   .S2                0.886    0.120    7.382    0.000
   .R2                0.563    0.080    7.054    0.000
   .N2                0.620    0.087    7.159    0.000
   .W2                0.906    0.123    7.390    0.000
    LV1               0.572    0.124    4.596    0.000
    LV2               0.686    0.132    5.198    0.000
```

Since we inputted the correlation matrix here, the `covariance` shown in the result is actually the correlation coefficient between the the factor at two time points, which is .706, $p < .001$.