CFA_SimDesign

```r
library(SimDesign)
library(lavaan)
library(tidyr)
```

Step 1 — Defining conditions (sample size and model)

```r
Design <- createDesign(sample_size = c(50, 100, 500, 1e3, 1e4, 1e5),
                       model_type = c("true", "misspecified"))
```

Step 2 — Define generate, analyse, and summarise functions

```r
# Generate function: Simulate data based on the true model
Generate <- function(condition, fixed_objects) {
    N <- condition$sample_size

    # True model: Two factors, 8 indicators each
    true_model <- '
        f1 =~ 0.7*y1 +
        0.7*y2 +
        0.7*y3 +
        0.7*y4 +
        0.7*y5 +
        0.7*y6 +
        0.7*y7 +
        0.7*y8

        f2 =~ 0.7*y9 +
```

```
    0.7*y10 +
    0.7*y11 +
    0.7*y12 +
    0.7*y13 +
    0.7*y14 +
    0.7*y15 +
    0.7*y16
'


# Misspecification: Add cross-loadings
misspecified_model <- '
    f1 =~ 0.7*y1 +
    0.7*y2 +
    0.7*y3 +
    0.7*y4 +
    0.7*y5 +
    0.7*y6 +
    0.7*y7 +
    0.7*y8 +
    0.3*y9 # added cross-loading

    f2 =~ 0.7*y9 +
    0.7*y10 +
    0.7*y11 +
    0.7*y12 +
    0.7*y13 +
    0.7*y14 +
    0.7*y15 +
    0.7*y16 +
    0.3*y1 # added cross-loading
'

# Generate data based on the condition
if(condition$model_type == "true") {
    data <- simulateData(true_model, sample.nobs = N)
} else {
    data <- simulateData(misspecified_model, sample.nobs = N)
}

return(data)
```

```
}


# Analyse function: Fit the true model and extract fit measures
Analyse <- function(condition, dat, fixed_objects) {
    true_model <- '
        f1 =~ y1 + y2 + y3 + y4 + y5 + y6 + y7 + y8
        f2 =~ y9 + y10 + y11 + y12 + y13 + y14 + y15 + y16
    '

    fit <- cfa(true_model, data = dat)

    # Extract fit measures
    fit_measures <- fitMeasures(fit, c("cfi", "tli", "rmsea", "srmr", "chisq"))
    return(fit_measures)
}


# Summarise function: Calculate the average fit measures from n replications
Summarise <- function(condition, results, fixed_objects) {
    summary_measures <- colMeans(results)
    return(summary_measures)
}
```

```
# Make sure it doesn't try resuming a simulation
SimClean()

# Step 3 --- Collect results by looping over the rows in design
library(SimDesign)
library(lavaan)

res <- runSimulation(design=Design, replications=100,
                     generate=Generate, analyse=Analyse, summarise=Summarise)
```

```
Design: 1/12;   RAM Used: 62.3 Mb;   Replications: 100;   Total Time: 0.00s
 Conditions: sample_size=50, model_type=true


Design: 2/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 4.09s
```

```
 Conditions: sample_size=100, model_type=true


Design: 3/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 8.09s
 Conditions: sample_size=500, model_type=true


Design: 4/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 12.08s
 Conditions: sample_size=1000, model_type=true


Design: 5/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 16.38s
 Conditions: sample_size=10000, model_type=true


Design: 6/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 22.12s
 Conditions: sample_size=1e+05, model_type=true


Design: 7/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 45.44s
 Conditions: sample_size=50, model_type=misspecified


Design: 8/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 52.94s
 Conditions: sample_size=100, model_type=misspecified


Design: 9/12;   RAM Used: 78.8 Mb;   Replications: 100;   Total Time: 01m 0.23s
 Conditions: sample_size=500, model_type=misspecified


Design: 10/12;   RAM Used: 78.9 Mb;   Replications: 100;   Total Time: 01m 7.60s
 Conditions: sample_size=1000, model_type=misspecified


Design: 11/12;   RAM Used: 78.9 Mb;   Replications: 100;   Total Time: 01m 15.24s
 Conditions: sample_size=10000, model_type=misspecified


Design: 12/12;   RAM Used: 78.9 Mb;   Replications: 100;   Total Time: 01m 20.64s
 Conditions: sample_size=1e+05, model_type=misspecified
```

```
Simulation complete. Total execution time: 01m 51.22s
```

```r
res <- res %>% dplyr::select(-COMPLETED, -SEED)

# ------------------------------------------------------
# ------------------------------------------------------
print(res)
```

```
# A tibble: 12 x 10
   sample_size model_type       cfi     tli      rmsea      srmr    chisq
         <dbl> <chr>          <dbl>   <dbl>      <dbl>     <dbl>    <dbl>
 1          50 true         0.89669 0.88426 0.055521   0.10270     123.07
 2         100 true         0.96744 0.97400 0.025384   0.072316    111.04
 3         500 true         0.99557 0.99826 0.0085409  0.032497    105.55
 4        1000 true         0.99833 0.99997 0.0045251  0.022673    103.05
 5       10000 true         0.99976 0.99989 0.0018654  0.0073577   106.12
 6      100000 true         0.99998 1.0000  0.00052441 0.0023223   104.07
 7          50 misspecified 0.86934 0.85193 0.063645   0.10913     127.54
 8         100 misspecified 0.94071 0.93262 0.040156   0.083298    122.90
 9         500 misspecified 0.96078 0.95430 0.034933   0.054521    168.11
10        1000 misspecified 0.96418 0.95827 0.033450   0.048682    219.91
11       10000 misspecified 0.96309 0.95700 0.034296   0.044508   1315.6
12      100000 misspecified 0.96344 0.95741 0.034130   0.043800  12102.
# i 3 more variables: REPLICATIONS <dbl>, SIM_TIME <chr>, RAM_USED <chr>
```

```r
true_model <- '
    f1 =~ y1 + y2 + y3 + y4 + y5 + y6 + y7 + y8
    f2 =~ y9 + y10 + y11 + y12 + y13 + y14 + y15 + y16
'

simulated_data <- simulateData(true_model, sample.nobs = 1000)

fit <- cfa(true_model, data = simulated_data)

summary(fit, fit.measures = TRUE)
```

```
lavaan 0.6-18 ended normally after 26 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
```

```
  Number of model parameters                         33

  Number of observations                           1000

Model Test User Model:

  Test statistic                                100.847
  Degrees of freedom                                103
  P-value (Chi-square)                            0.542

Model Test Baseline Model:

  Test statistic                               6951.272
  Degrees of freedom                                120
  P-value                                         0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                     1.000
  Tucker-Lewis Index (TLI)                        1.000

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -24950.843
  Loglikelihood unrestricted model (H1)      -24900.419

  Akaike (AIC)                                49967.685
  Bayesian (BIC)                              50129.641
  Sample-size adjusted Bayesian (SABIC)       50024.831

Root Mean Square Error of Approximation:

  RMSEA                                           0.000
  90 Percent confidence interval - lower          0.000
  90 Percent confidence interval - upper          0.016
  P-value H_0: RMSEA <= 0.050                     1.000
  P-value H_0: RMSEA >= 0.080                     0.000

Standardized Root Mean Square Residual:

  SRMR                                            0.020

Parameter Estimates:
```

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

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    y1                1.000
    y2                0.896    0.042   21.329    0.000
    y3                0.975    0.043   22.443    0.000
    y4                0.986    0.044   22.449    0.000
    y5                0.989    0.044   22.638    0.000
    y6                0.941    0.043   21.733    0.000
    y7                0.912    0.043   21.339    0.000
    y8                0.911    0.043   21.427    0.000
  f2 =~
    y9                1.000
    y10               1.015    0.050   20.281    0.000
    y11               1.003    0.050   20.141    0.000
    y12               1.059    0.051   20.604    0.000
    y13               1.045    0.050   20.947    0.000
    y14               1.066    0.051   20.812    0.000
    y15               1.013    0.049   20.775    0.000
    y16               1.062    0.049   21.649    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2               -0.014    0.037   -0.379    0.705

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .y1               0.909    0.048   18.944    0.000
   .y2               1.010    0.051   19.898    0.000
   .y3               0.986    0.051   19.371    0.000
   .y4               1.008    0.052   19.368    0.000
   .y5               0.982    0.051   19.265    0.000
   .y6               1.040    0.053   19.721    0.000
   .y7               1.044    0.052   19.893    0.000
   .y8               1.027    0.052   19.856    0.000
   .y9               0.966    0.049   19.615    0.000
   .y10              1.066    0.054   19.800    0.000
```

```
.y11               1.070    0.054   19.870    0.000
.y12               1.088    0.055   19.626    0.000
.y13               0.987    0.051   19.425    0.000
.y14               1.057    0.054   19.507    0.000
.y15               0.961    0.049   19.528    0.000
.y16               0.878    0.046   18.947    0.000
 f1                1.145    0.086   13.277    0.000
 f2                0.956    0.078   12.257    0.000
```