Example 6.2 - Ordered-categorical indicator variables, Part I

The data for this example are responses to five items of the LSAT Figure Classification test:

```
library("psych")
head(lsat6)
```

```
##      Q1 Q2 Q3 Q4 Q5
## [1,]  0  0  0  0  0
## [2,]  0  0  0  0  0
## [3,]  0  0  0  0  0
## [4,]  0  0  0  0  1
## [5,]  0  0  0  0  1
## [6,]  0  0  0  0  1
```

For more info on the data, type `?lsat`.

First we take a look at the tetrachoric correlations between items:

```
tet <- tetrachoric(lsat6)
```

```
tet
```

```
## Call: tetrachoric(x = lsat6)
## tetrachoric correlation
##    Q1   Q2   Q3   Q4   Q5
## Q1 1.00
## Q2 0.17 1.00
## Q3 0.23 0.19 1.00
## Q4 0.11 0.11 0.19 1.00
## Q5 0.07 0.17 0.11 0.20 1.00
##
##  with tau of
##    Q1    Q2    Q3    Q4    Q5
## -1.43 -0.55 -0.13 -0.72 -1.13
```

```
round(cor(lsat6), digits = 3)
```

```
##       Q1    Q2    Q3    Q4    Q5
## Q1 1.000 0.074 0.099 0.044 0.024
## Q2 0.074 1.000 0.115 0.062 0.086
## Q3 0.099 0.115 1.000 0.109 0.053
## Q4 0.044 0.062 0.109 1.000 0.099
## Q5 0.024 0.086 0.053 0.099 1.000
```

We see that the tetrachoric correlations, which account for the binary nature of the items, are higher than the Pearson correlations (calculated with `cor()`). This is what we generally see: for binary items, the Pearson correlation underestimates the strength of associations.

The tetrachoric correlation matrix also provides us with item thresholds. Lower values indicate easier items.

We also inspect item means (i.e., proportion of respondents who had the items correct):

```
sort(colMeans(lsat6), decreasing = TRUE)
```

```
##    Q1    Q5    Q4    Q2    Q3
## 0.924 0.870 0.763 0.709 0.553
```

```
sort(tet$tau)
```

```
##         Q1         Q5         Q4         Q2         Q3
## -1.4325027 -1.1263911 -0.7159860 -0.5504657 -0.1332445
```

The thresholds and item means indicate the same ordering of items in terms of difficulty: The most difficult item is Q3, the easiest item is Q1.

Next we perform a CFA:

```
library("lavaan")
model.CFA <- '
  Theta =~ Q1 + Q2 + Q3 + Q4 + Q5
'
fit.CFA <- cfa(model.CFA, data = data.frame(lsat6), ordered = paste0("Q", 1:5))
paste0("Q", 1:5)
```

```
## [1] "Q1" "Q2" "Q3" "Q4" "Q5"
```

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

```
## lavaan 0.6.15 ended normally after 29 iterations
##
##   Estimator                                         DWLS
##   Optimization method                             NLMINB
##   Number of model parameters                          10
##
##   Number of observations                            1000
##
## Model Test User Model:
##                                               Standard      Scaled
##   Test Statistic                                 4.051       4.740
##   Degrees of freedom                                 5           5
##   P-value (Chi-square)                           0.542       0.448
##   Scaling correction factor                                  0.867
##   Shift parameter                                            0.070
##     simple second-order correction
##
## Model Test Baseline Model:
```

```
##
##   Test statistic                                67.171      65.104
##   Degrees of freedom                                10          10
##   P-value                                        0.000       0.000
##   Scaling correction factor                                  1.038
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    1.000       1.000
##   Tucker-Lewis Index (TLI)                       1.033       1.009
##
##   Robust Comparative Fit Index (CFI)                         1.000
##   Robust Tucker-Lewis Index (TLI)                            1.032
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.000       0.000
##   90 Percent confidence interval - lower         0.000       0.000
##   90 Percent confidence interval - upper         0.039       0.043
##   P-value H_0: RMSEA <= 0.050                    0.989       0.982
##   P-value H_0: RMSEA >= 0.080                    0.000       0.000
##
##   Robust RMSEA                                               0.000
##   90 Percent confidence interval - lower                     0.000
##   90 Percent confidence interval - upper                     0.101
##   P-value H_0: Robust RMSEA <= 0.050                         0.690
##   P-value H_0: Robust RMSEA >= 0.080                         0.126
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.036       0.036
##
## Parameter Estimates:
##
##   Standard errors                           Robust.sem
##   Information                                 Expected
##   Information saturated (h1) model          Unstructured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   Theta =~
##     Q1                1.000                               0.389    0.389
##     Q2                1.020    0.358    2.846    0.004    0.397    0.397
##     Q3                1.210    0.447    2.709    0.007    0.471    0.471
##     Q4                0.968    0.352    2.751    0.006    0.377    0.377
##     Q5                0.879    0.352    2.499    0.012    0.342    0.342
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Q1                0.000                               0.000    0.000
##    .Q2                0.000                               0.000    0.000
##    .Q3                0.000                               0.000    0.000
##    .Q4                0.000                               0.000    0.000
##    .Q5                0.000                               0.000    0.000
```

```
##     Theta              0.000                               0.000    0.000
##
## Thresholds:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     Q1|t1            -1.433    0.059  -24.431    0.000   -1.433   -1.433
##     Q2|t1            -0.550    0.042  -13.133    0.000   -0.550   -0.550
##     Q3|t1            -0.133    0.040   -3.349    0.001   -0.133   -0.133
##     Q4|t1            -0.716    0.044  -16.430    0.000   -0.716   -0.716
##     Q5|t1            -1.126    0.050  -22.395    0.000   -1.126   -1.126
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Q1                0.848                               0.848    0.848
##    .Q2                0.842                               0.842    0.842
##    .Q3                0.778                               0.778    0.778
##    .Q4                0.858                               0.858    0.858
##    .Q5                0.883                               0.883    0.883
##     Theta             0.152    0.087    1.743    0.081    1.000    1.000
##
## Scales y*:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     Q1                1.000                               1.000    1.000
##     Q2                1.000                               1.000    1.000
##     Q3                1.000                               1.000    1.000
##     Q4                1.000                               1.000    1.000
##     Q5                1.000                               1.000    1.000
```

Because we declared the items as ordered-categorical, the DWLS (diagonally weighted least squares) estimator was used. This provides standard and robust (robust against deviations from normality) fit indices, by default.

The model fits very well according to all fit measures. All loadings are substantial and significant. The most difficult item is Q3, easiest item is Q1. Also, Q3 is the strongest indicator of the latent trait, Q5 the weakest indicator.