Sensorveiledning MAE4011 - Principles of measurement

Björn Andersson

March 12, 2024

Exam 21/2 2024

Task 9 (3p)

The Pearson correlations between X and Y and Y and Z should be computed based on the covariance matrix given:

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{2}{\sqrt{10 \times 10}} = 0.2.$$

and

$$\text{Cor}(Y, Z) = \frac{\text{Cov}(Y, Z)}{\sqrt{\text{Var}(Y)\text{Var}(Z)}} = \frac{8}{\sqrt{10 \times 10}} = 0.8.$$

i Answer the question correctly based on the result, i.e. interest in mathematics is weakly positively linearly related with mathematics proficiency (Pearson correlation 0.2) and mathematics proficiency is strongly positively linearly related with reading comprehension (Pearson correlation 0.8) (Max 2p.)

ii Statement of the assumptions: (Max 1p.)

- Independent measurement errors.
- High reliability of all test scores.
- Other things that make sense.

Task 10 (2p)

The question was written incorrectly in the exam, it should have said observed sample variance of the sum score equal to 500 and the formula for the coefficient alpha estimator as

$$\hat{\alpha} = m^2 \frac{\hat{\lambda}^2}{\hat{\sigma}_Y^2},$$

which gives $\hat{\alpha}^{\text{est}} = 0.8$. Computing the coefficient as presented instead gives

$$\hat{\alpha} = 0.4.$$

The correct expression of coefficient alpha gives the reliability coefficient of the sum scores, with a factor model that is specified to have equal factor loadings. Hence, the estimated coefficient alpha gives the estimated reliability coefficient of the sum scores. The item has been graded consistent with how the item was written in the exam. Using the formula presented in the exam suggests that the reliability of the sum scores is low. The assumption underlying this interpretation is that the single factor model with equal factor loadings fits the data well. Max 1p for computation and interpretation and max 1p for statement of assumption.

Task 11 (2p)

i Too much: E.g., anxiety, illness, on-site disturbance (loud environment), cheating. (Max 1p.)

ii Too little: E.g., insufficient content coverage, too easy or too difficult questions for specific skills. (Max 1p.)

If listing multiple examples, only give full credits if all examples are appropriate.

Task 12 (3p)

a) We plug in the cut score value 20 in the equating function, yielding

$$20 = 0.8X + 4 \iff 16 = 0.8X \iff X = 20.$$

Thus, based on the equating function $eq(Y) = 0.8X + 4$, the cut score for test X should be set to 20. (2p)

b) A student with a score of 20 just makes the cut score. Hence, the student passes the test. (1p)

Task 13 (4p)

a) The model fits well according to commonly used criteria, RMSEA $<$ 0.06 and SRMSR $<$ 0.08. (1p)

b) The residual correlation matrix would give more details about the model fit. Other aspects can also give points if they are relevant and make sense. (1p)

c) The sum score correlation does not account for measurement error, and hence the correlation is lower than the correlation between the constructs because of this. (1p)

d) The factor correlation directly gives the linear relationship between the two constructs, while the sum score correlation gives the linear relationship between the scores of the constructs. Hence, the factor correlation better represents the relationship between the constructs. (1p)

Task 14 (6p)

i Evidence sources: internal structure, relationship to other variables, content of items (up to +2p)

ii Description of data: random sample from the target population, give the scale to these individuals, record gender of respondents, expert panel to assess items based on content (up to +1p)

iii Analyses: Estimate unidimensional factor model, compute mean scores for gender groups and home ownership groups. (up to +1p)

iv Results: unidimensional model in accordance with the theoretical framework, no differences in factor means with respect to gender but differences by home ownership status, expert evaluation consistent with theory, sufficiently reliable scores (up to +2p)

Task 15 (4p)

a)

The bifactor model can be used to assess if a single factor model is approximately appropriate. If a general factor dominates (explains substantially more variance than the specific factors do), then a single factor model can be judged as appropriate. We can compare the variance explained by the general factor, which is

$$\sum_{j=1}^{6} \lambda_{Gj}^2 = 4^2 + 2^2 + 3^2 + 2^2 + 1^2 = 16 + 4 + 9 + 4 + 4 + 1 = 38$$

to the variance explained by the all factors, which is

$$\sum_{j=1}^{6} \lambda_{Gj}^2 + \sum_{j=1}^{6} \lambda_{Sj}^2 = 38 + 4 + 1 + 1 + 1 + 4 + 1 = 50.$$

This yields 38 / 50 = 0.76, which is higher than the lowest recommended threshold (0.7). Hence, using the sum score is supported based on the relatively small amount of multidimensionality that is present. (2p)

b)

An alternative method of scoring is to use the bifactor model directly to produce general scores and subscores. Benefit: For example, increased measurement accuracy and more detailed scoring in multiple domains. Downside: For example, interpretation of subscores can be difficult since the subfactors are independent of the general factor. (2p)

Task 16 (4p)

Required components for acceptable responses:

- i Identify a suitable approach, e.g. the bookmarking method and briefly describe the procedure. (Max +2p)
- ii Suggest a procedure for sorting items by difficulty. (Max +1p)
- iii Outline how an expert panel should judge the appropriate cut-score. (Max +1p)

Task 17 (6p)

a (2p)

$$X6 = \lambda_{X6} \times R + E_6$$

R is the reading literacy factor and E_6 is the error term, both of which are random variables. λ_{X6} is the factor loading and $\Psi_6^2 = \text{Var}(E_6)$ is the variance of the error term, both of which are parameters of the model. λ_{X6} says how sensitive the item is at measuring the latent factor R and Ψ_6^2 indicates how much random error there exists for item 6. Correct equation +1p, correct interpretation of all components +1p. It is not necessary to include the mean parameter for the item score in the description.

b (2p)

We can first note that

$$X_3 = 0.8 \times M + 0.5 \times R + E_3$$

and

$$X_4 = 0.5 \times R + E_4.$$

We then have that

$$\begin{aligned}
\text{Cov}(X_3, X_4) =& \text{Cov}(0.8 \times M + 0.5 \times R + E_3, 0.5 \times R + E_4) \\
=& \text{Cov}(0.8 \times M, 0.5 \times R + E_4) + \text{Cov}(0.5 \times R, 0.5 \times R + E_4) \\
& + \text{Cov}(E_3, 0.5 \times R + E_4) \\
=& \text{Cov}(0.8 \times M, 0.5 \times R) + 0 + \text{Cov}(0.5 \times R, 0.5 \times R) + 0 + 0 + 0 \\
=& 0.8 \times 0.5\text{Cov}(M, R) + 0.5 \times 0.5\text{Var}(R, R) \\
=& 0.4 * 0.5 + 0.25 \times 1 \\
=& 0.2 + 0.25 = 0.45.
\end{aligned}$$

Entirely correct derivation gives 2p, even if there is some mistake in the computation. Otherwise, up to 1.5p depending on level of completion.

c (2p)

The reliability coefficient is given by the ratio of the true score variance and the observed score variance. The true score variance is given by the square of the factor loading. The observed score variance can be read off the figure and is equal to 1. Hence, the reliability coefficient for $X2$ is

$$\rho_{X2,X2'} = \frac{0.4^2}{1} = 0.16.$$

Entirely correct reasoning gives 2p, even if there is some mistake in the computation. Otherwise, up to 1.5p depending on level of completion.