Sensorveiledning MAE4011 - Principles of measurement

Björn Andersson

December 22, 2023

Exam 18/12 2023

Task 9 (2p)

i Answer the question correctly based on the result, i.e. the constructs are moderately to strongly linearly related, adjusted Pearson correlation 0.8 (Max 1p.)

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

- Independent measurement errors.
- Coefficient alpha is an appropriate measure of the reliability, which in turn depends on a single factor model with equal or highly similar factor loadings fitting well to the data.

Task 10 (2p)

SEM is 3 and the observed score is 20. The 95% confidence interval for the true score is then

$$20 \pm 1.96 \times 3 \approx (14, 26).$$

The interpretation is that if the student answered the same test multiple times, an interval like the above would cover the true score 95% of the time.

Statement of the assumptions:

- Item scores measure a single construct.
- A large number of item scores with independent error terms.

Max 1p for computation and interpretation and max 1p for statement of assumptions.

Task 11 (2p)

From the graph, we have $X_1 = \lambda_1 \times \text{Math} + \epsilon_1$, $X_3 = \lambda_3 \times \text{Math} + \epsilon_3$, $X_4 = \lambda_4 \times \text{Reading} + \epsilon_4$, where $\lambda_1 = 0.5$, $\lambda_3 = 0.8$, $\lambda_4 = 0.5$, $\text{Var}(\text{Math}) = \text{Var}(\text{Reading}) = \text{Var}(X_1) = \text{Var}(X_2) = \text{Var}(X_4) = 1$, and $\text{Cov}(\text{Math}, \text{Reading}) = 0.5$. We obtain

$$\frac{\text{Cov}(X_1, X_3)}{\sigma_{X_1}\sigma_{X_3}} = \frac{\text{Cov}(X_1, X_3)}{1 \times 1} = \lambda_1\lambda_3\text{Var}(\text{Math}) = 0.5 \times 0.8 \times 1 = 0.4$$

and

$$\frac{\text{Cov}(X_1, X_4)}{\sigma_{X_1}\sigma_{X_4}} = \frac{\text{Cov}(X_1, X_4)}{1 \times 1} = \lambda_1\lambda_4\text{Cov}(\text{Math}, \text{Reading}) = 0.5 \times 0.5 \times 0.5 = 0.125.$$

For full marks (2p), correct presentations or explanations of the computation are required. Only the final answers gives 1p.

Task 12 (2p)

We have that

$$\begin{aligned}
\text{Var}(Y) &= E[(Y - E(Y))^2] \\
&= E(Y^2 + E(Y)^2 - 2YE(Y)) \\
&= E(Y^2) + E(Y)^2 - 2E(Y)E(Y) \\
&= E(Y^2) + E(Y)^2 - 2E(Y)^2 \\
&= E(Y^2) - E(Y)^2.
\end{aligned}$$

We thus have

$$\text{Var}(Y) = E(Y^2) - E(Y)^2 = 12 - 3^2 = 12 - 9 = 3.$$

Correct derivation or presentation of the equality $\text{Var}(Y) = E(Y^2) - E(Y)^2$ gives up to 1p and correct computation gives up to 1p.

Task 13 (2p)

i This is appropriate when there is no memorization or learning and when there are uncorrelated errors between measurement time points. (Max 1p.)

ii The results are valid for the post-intervention time point given that there is no change in error variance or in the true score variance. (Max 1p.)

Task 14 (4p)

i Propose to compute the difficulty of items from the observed data via a model or via proportion correct. (Max 1p)

ii Describe construction of an item map that connects the items to the scale by matching the difficulty of each item to the scale scores. (Max 1p.)

iii Decide on what it means to have low, medium, and high performance via the data collected or via expert opinion. Interpret the meaning of these performance descriptors by considering the skills of the items that have difficulty levels that match the scale scores for low, medium and high performance. (Max 2p.)

Task 15 (4p)

i Equal reliability: Lower-bound reliability is not equal, but is similar. (Max 1p.)

ii Symmetry: Linear equating is symmetrical. (Max 1p.)

iii Population invariance: Compute equating function for boys and girls and compare them.

- Girls: $\frac{11}{11}X + (30 - \frac{11}{11}26) = X + 4$
- Boys: $\frac{9}{9}X + (27 - \frac{9}{9}23) = X + 4$

The same equating function is obtained, so population invariance of the equating transformation holds between boys and girls. (Max 2p.)

Task 16 (6p)

i Evidence sources: internal structure, relationship to other variables, content of items, response processes (Max 2p.)

ii Description of study, data and analysis: random sample from the target population, give the scale to these individuals, expert panel to assess items based on content, in-person observations of respondent behaviors. Estimation of a multiple factor model and estimation of correlations between sum scores of measures. (Max 2p.)

iii Results: multidimensional model in accordance with the theoretical framework, expert evaluation consistent with theory, at least moderately high reliability of domain scores (Max 2p.)

Task 17 (4p)

The following aspects can, for example, give credit:

- The study concerns validity evidence with respect to relationships with other variables. There exists a moderate relationship between test scores and future job performance.
- Only nurses that successfully pass the exam can be included in the study, and the study doesn't look at the classification into pass/fail in particular. Hence, the use of the scores is somewhat disconnected from the study done.
- Reflection that for example content evidence is missing.

- Reflection that the sample is large and fairly representative, since the response rate is 76%.
- Overall appraisal that the regression analysis provides some, but limited, evidence that the test score use for pass/fail decisions is valid.

Other relevant aspects may also qualify for points up to a maximum of 4p in total.

Task 18 (6p)

i Benefit: for example shorter testing time which reduces test fatigue. Drawback: for example lack of proper content coverage or too low reliability. (Max 2p.)

ii Present a reasonable way to select items, either by considering content or by considering the item information/standardized factor loadings. (Max 2p.)

iii Impact on reliability, with a numerical illustration via the omega reliability coefficient. Very high reliability is not needed because the scores will be aggregated and individual-level assessment is not the purpose. Impact on content coverage with reflection to specific situation. (Max 2p.)