Sensorveiledning MAE4011 - Principles of measurement

Björn Andersson

January 27, 2022

Exam 7/1 2022

Task 9

i Correlation for the sum scores is affected by measurement error, so the magnitude of the correlation between the constructs should be higher. The correlation measures the linear relationship between the constructs. (+1p)

ii Here, we have a lower-bound estimate of the reliability of the scores. When we adjust for attenuation, the correlation increases in magnitude to -0.8. However, this depends on a single factor model being appropriate for the respective scale items. (+1p)

Task 10

i The 95% confidence interval is 10 +/- 1.96*2 and in repeated measurements with the same test, the true score will be in such an interval 95% of the time. (+1p)

ii The assumptions are that there are a large amount of items on the test and that the item scores are independent conditional on the underlying factor. (+1p)

Task 11

$$\begin{aligned}
\text{Cov}(X_2, X_3) &= 0.4 * 0.8 * 1 = 0.32 \\
\text{Cov}(X_2, X_4) &= 0.4 * 0.5 * 0.5 = 0.1 \\
\text{Cov}(X_3, X_4) &= 0.8 * 0.5 * 0.5 = 0.2
\end{aligned}$$

Up to 2p depending on level of completion.

Task 12

$$\mathrm{Var}(Y) = E(Y^2) - (E(Y))^2 = 6 - 2^2 = 6 - 4 = 2$$

Up to 2p depending on level of completion.

Task 13

When there is no shift in the construct over time (+1p) and if the measurements at the two time points are independent (+1p). Partial credit, up to 1p, can be obtained if instead mentioning a single factor model being appropriate and that sufficiently large samples are required.

Task 14

The variance of each observed score can be found by

$$\sigma^2_{X_j} = \mathrm{Var}(X_j) = \lambda^2 + \Psi^2_j.$$

Thus, $\sum_{j=1}^{3} \sigma^2_{X_j} = 1 + 1 + 1 + 0.5 + 1 + 1.5 = 6$. Up to +1p depending on level of completion.

Hence, coefficient alpha is $(3/2) * (1 - 0.5) = 1.5 * 0.5 = 0.75$. Coefficient alpha is equal to the reliability coefficient for the sum score when the factor loadings are all identical (which is the case here). (+1p)

Task 15

- Equal reliability: Tests do not have the same lower-bound estimate. (+1p)
- Equating symmetry: Linear equating is symmetric. (+1p)
- Population invariance: The estimated linear equating functions are the same for both groups, with computation. (+2p)

Task 16

Answer should reflect the following to obtain full marks:

- Description of how to collect data from the target population, evaluate the fit of a single factor model. Use scores in relation to status of individuals, decide diagnostic criterion based on the scores. (+1p)
- Reflection on content-oriented evidence: Operationalization should be developed with indicators adhering to the theory of how panic disorder manifests. Expert panel could rate items with respect to this criterion. (+1p)
- Reflection on internal structure: Reliability to demonstrate that the scores are not unduly influenced by random error variance. (+1p)

- Reflection on relationships to other variables: Convergent and discriminant evidence with similar and different constructs measured using other tests. That is, scores on a test for the panic-disorder construct should correlate strongly with other tests also intended to measure panic disorder, and less strongly with tests intended to measure other constructs. (+1p)

Comments on other aspects, such as consequences, can compensate for lack of one or more and allow for full marks in an overall assessment of the answer.

Task 17

- The study concerns validity evidence with respect to relationships with other variables. There exists a weak relationship between test scores and future job performance. (+1p)
- Only teachers that successfully pass the exam can be included in the study, so the use of the scores is disconnected from this particular study. (+1p)
- Reflection that for example content evidence and internal structure evidence is missing. (+1p)
- Overall appraisal that the regression analysis provides very limited evidence in itself that the test score use for pass/fail decisions is valid.

Tasks 18-21

18, 19 and 20

These are scored together since 19 and 20 are related.

(+1p):

$$\begin{aligned} E(X_1 + X_4) &= E(\mu_1 + \lambda_1 F_1 + \epsilon_1 + \mu_4 + \lambda_4 F_2 + \epsilon_4) \\ &= E(\mu_1) + E(\lambda_1 F_1) + E(\epsilon_1) + E(\mu_4) + E(\lambda_4 F_2) + E(\epsilon_4) \\ &= E(\mu_1) + \lambda_1 E(F_1) + 0 + E(\mu_4) + \lambda_4 E(F_2) + 0 \\ &= 3 + \lambda_1 * 0 + 0 + 4 + \lambda_4 * 0 + 0 \\ &= 7 \end{aligned}$$

(+1p): We have

$$\begin{aligned} \mathrm{Var}(X_1 + X_4) &= \mathrm{Var}(X_1) + \mathrm{Var}(X_4) + 2 * \mathrm{Cov}(X_1, X_4) \\ &= 6 + 2.5 + 2 * 1 \\ &= 10.5 \end{aligned}$$

This is because:

$$\begin{aligned} \mathrm{Var}(X_1) &= \mathrm{Var}(\mu_1 + \lambda_1 F_1 + \epsilon_1) \\ &= 0 + \lambda_1^2 \mathrm{Var}(F_1) + \mathrm{Var}(\epsilon_1) \\ &= 4 * 1 + 2 \\ &= 6 \end{aligned}$$

$$\begin{aligned}
\text{Var}(X_4) &= \text{Var}(\mu_4 + \lambda_4 F_2 + \epsilon_4) \\
&= 0 + \lambda_4^2 \text{Var}(F_2) + \text{Var}(\epsilon_4) \\
&= 1 * 1 + 1.5 \\
&= 2.5
\end{aligned}$$

(+1p):

$$\begin{aligned}
\text{Cov}(X_1, X_4) &= \text{Cov}(\mu_1 + \lambda_1 F_1 + \epsilon_1, \mu_4 + \lambda_4 F_2 + \epsilon_4) \\
&= \text{Cov}(\mu_1 + \lambda_1 F_1 + \epsilon_1, \mu_4) + \text{Cov}(\mu_1 + \lambda_1 F_1 + \epsilon_1, \lambda_4 F_2) + \text{Cov}(\mu_1 + \lambda_1 F_1 + \epsilon_1, \epsilon_4) \\
&= 0 + 0 + 0 + 0 + \text{Cov}(\lambda_1 F_1, \lambda_4 F_2) + 0 + 0 + 0 + 0 \\
&= \lambda_1 \lambda_4 \text{Cov}(F_1, F_2) \\
&= 2 * 1 * 0.5 \\
&= 1
\end{aligned}$$

Partial credit can be obtained for answers which are not entirely correct.

21

The reliability coefficient for each item score is equal to the true score variance divided by the observed score variance.

$$\rho_{X_1} = \frac{\sigma_T^2}{\sigma_{X_1}^2} = \frac{\lambda_1^2}{\sigma_{X_1}^2} = \frac{4}{6}$$

$$\rho_{X_4} = \frac{\sigma_T^2}{\sigma_{X_4}^2} = \frac{\lambda_4^2}{\sigma_{X_4}^2} = \frac{1}{2.5}$$

So X_1 has higher precision of measurement. Full marks also if looking at the information λ_j^2/Ψ_j^2. (+1p) If instead looking at the factor loadings only, max +0.5p.

Task 22

i E.g. shorter testing time. (Max 1p.)

ii E.g. reduced content coverage. (Max 1p.)

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

Task 23

Select items with basis in the item information:

- X_1: $\lambda_1^2/\Psi_1^2 = 1/1 = 1$
- X_2: $\lambda_2^2/\Psi_2^2 = 4/1 = 4$

- X_3: $\lambda_3^2/\Psi_3^2 = 9/2 = 4.5$
- X_4: $\lambda_4^2/\Psi_4^2 = 4/1 = 4$
- X_5: $\lambda_5^2/\Psi_5^2 = 1/3$
- X_6: $\lambda_6^2/\Psi_6^2 = 1/1 = 1$

So we select items 2, 3, and 4. (+1p) If instead looking at the factor loadings, maximum 0.5p.

The reliability of the sum score is given by coefficient omega

$$\begin{aligned}
\omega &= \frac{\sigma_T^2}{\sigma_T^2 + \sigma_\epsilon^2} \\
&= \frac{(\sum_{j=1}^3 \lambda_j)^2}{(\sum_{j=1}^3 \lambda_j)^2 + \sum_{j=1}^3 \Psi_j^2} \\
&= \frac{(2+3+2)^2}{(2+3+2)^2 + 1 + 2 + 1} \\
&= \frac{49}{49+4} \\
&\approx 0.92.
\end{aligned}$$

(+1p) If there is some minor computational error, full marks can still be obtained.

Task 24

- Reflection on sufficient reliability for the intended usage. (+1p)
- Reflection on coverage in terms of item content that is reasonable. (+1p)

Other aspects that are reasonable can compensate for lack of the above for a maximum of total 2p for the task.