UiO : Universitetet i Oslo

CANDIDATE

184108

TEST

MAE4011 1 Principles of Measurement

Subject code	MAE4011
Evaluation type	Individuell skriftlig prøve
Test opening time	20.12.2022 09:00
End time	20.12.2022 13:00
Grade deadline	--
PDF created	26.12.2022 19:33
Created by	Tony Clifford Austin Tan

9 SR1H22

A scale to measure depression severity was developed and data were collected from a large group of students, along with the scores of an existing scale for satisfaction with life.

You observed the following covariance matrix for the scores of the two scales, where X denotes the depression severity scale scores and Y denotes the satisfaction with life scale scores:

$$\mathbf{\Sigma} = \begin{pmatrix} 10 & -7 \\ -7 & 10 \end{pmatrix}.$$

Based on these observations, how would you characterize the relationship between depression severity and satisfaction with life?

State the assumptions made in the interpretations of the relationship.

Fill in your answer here

$$correlation = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{-7}{\sqrt{10}\sqrt{10}} = -0.7$$

The correlation shows a relative strong negative relationship between the depression severity scale scores and the satisfaction with life scale scores.
The correlation assumes the linear relationship between the two scale scores.
The correlation coefficient is influenced by measurement error and reliability of each scale.
The sampling variance could be an influential factor as well.

Words: 59

Answered.

11 SR3H22

X and Y are two random variables where $\mathbf{Var}(X) = 2$, $\mathbf{Var}(Y) = 3$ and $\mathbf{Cov}(X, Y) = 1$.

1. Calculate $\mathbf{Var}(Z)$, where $Z = X - Y$. Show your work.
2. Calculate $\mathbf{Var}(U)$, where $U = X + 2Y$. Show your work.

Fill in your answer here

1. Var(Z) = Var(X-Y) = Var(X)+Var(Y)-2*Cov(X,Y) = 2+3-2*1=3

2. Var(U) = Var(X+2Y) = Var(X) + Var(2Y) + 2Cov(X,2Y) = 2+2*2*3+2*2*1=18

Words: 20

Answered.

12 SR4H22

Let m be the number of items on a test. For a five-item test, the common factor loading λ was 1 and the variance of the sum score Y was 10. Compute coefficient alpha

$\alpha = m\frac{\lambda^2}{\mathrm{Var}(Y)}$

and interpret it. State the assumptions underlying the interpretation.

Fill in your answer here

$\alpha = 5 \cdot \frac{1^2}{10} = 0.5$

The coefficient alpha is the lower bound of the reliability of the sum score, which means the reliability of the sum score is either equal to or higher than 0.5 in this case. The number 0.5 shows that the sum score is not very reliable if we use the cut-off value of 0.8 to judge the reliability.

Under the assumption of single factor model, the coefficient alpha can be used as an estimate of the reliability of total scores of the underlying construct with the same factor loadings of all items. We also assume that there is no sampling variance since there is only a single sample is used to do the estimation.

Words: 114

Answered.

13 SR5H22

The *Standards for Educational and Psychological Testing* (2014) state that it is useful to consider ways in which the test scores can be influenced by either (1) too much or (2) too little.

A three-domain test is administered for the purpose of measuring Norwegian 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges. The test is a low-stakes test for the respondents since individual assessment is not of interest.

Provide **one example** of a way in which the test-scores might be influenced by too much, and **one example** of how the test-scores might be influenced by too little.

Fill in your answer here

(1) Too much: the test scores might be influenced by the anxiety or illness of the students when they are taking the test.

(2) Too little: there might be insufficient coverage or content sampling of all aspects of reading, mathematics and science knowledge and skills in the test.

Words: 48

Answered.

14 SR6H22

For two tests of reading comprehension, X and Y, the linear equating function was estimated to be $\mathbf{eq}(Y) = \mathbf{1.2}X + \mathbf{6}$. The cut score for passing test Y was determined to be 30.
Give the cut score for pass in terms of the test X scores, based on the estimated equating function. Present and explain how the result was obtained.

Fill in your answer here

If we insert 30 into the linear equating function,
30 = 1.2*X+6, where we can solve X = 20

The cut score for pass in terms of the test X is 20, which is equivalent to the cut score of 30 in test Y.

Words: 44

Answered.

15 SR7H22

Item scores on a test of mathematics and a test of interest in mathematics were given to the same group of students. A two-factor model with correlated factors (one factor measured by the mathematics test items and the other by the interest in mathematics items) was estimated, yielding the model fit indices:

GFI	0.95
RMSEA	0.05
SRMR	0.06

The correlation between the sum scores of the respective tests was 0.2 while the estimated factor correlation was 0.5. Explain why there is a difference in the factor correlation and the sum score correlation in this context.

Fill in your answer here

If we compare the given values with the cut-off values of each index measuring the model fit (GFI >=0.95, RMSEA < 0.05, SRMR <0.08), we found the two-factor model fits well. Then it is appropriate to use the estimated factor correlation as the correlation between the two underlying constructs. The correlation between the sum scores of the respective tests is lower than the factor correlation since there is measurement error. If the reliability of the two test scores are validated, we can adjust correlation by removing attenuation between the two sum scores.

Words: 92

Answered.

16 SR8H22

A bifactor model with one general factor and two subfactors (all factors independent) was estimated for an Norwegian test with two subdomains (reading and writing), yielding the following factor loading estimates:

Item	General	Reading	Writing
1	3	0.5	0
2	1	0.5	0
3	2	1	0
4	1	0	1
5	1	0	0.5
6	1	0	0.5

The model fit was judged to be satisfactory.

In a previous study, the sum score was used. Based on the estimated factor loadings, would you recommend doing this? Justify your answer.

Fill in your answer here

The bifactor model can be used to evaluate whether a single factor model is appropriate. If the general factor dominates, which means the variance substantially explained by the general factor instead of the specific factors, then we can prove the unidimensionality of the test.

Common variance $\sum_{j=1}^{6} \lambda_G^2$ = 3^2 + 1^2 +2^2 +1^2 +1^2 +1^2 =17

Common variance + subfactors variance $\sum_{j=1}^{6} \lambda_G^2 + \sum_{j=1}^{3} \lambda_R^2 + \sum_{j=4}^{6} \lambda_W^2$ = 17 + 0.5^2 +0.5^2 +1^2 +1^2 +0.5^2 +0.5^2 = 20

The ECV (Explained Common Variance) = $\frac{17}{20} = 0.85$, which is higher than the cut-off value 0.7.

Therefore, a single factor model is appropriate, and the underlying construct is unidimensional so we can use the sum score to do estimation like reliability etc..

Words: 114

Answered.

17 LR1H22

You have been asked to assist a group of teachers of Norwegian as a foreign language to find the appropriate cut-score for a test of Norwegian reading proficiency.

As part of the process, the test was piloted with a representative sample of the intended population and the results are available to you. In addition, an established framework describes the expected level of Norwegian reading proficiency.

Give a brief outline of how a standard-setting procedure could be used to find the cut-score for pass/fail on the Norwegian reading proficiency test.

Fill in your answer here

We can use bookmarking method to find the cut-score.
First, all test questions are ranked from the least difficult question to the most difficult question, which form the ordered items book.
Then an expert panel needs to go through each item and identify the relationship between items and the performance-level descriptor, which corresponds to the expected level of Norwegian reading proficiency. An RP (response probability) needs to be decided before the standard-setting meeting.
After that, the judges will follow the ordered book to check whether the threshold students have at least the given RP of correctly answering each test question. If the answer is "yes", then they will move to the next one until there is a "no". Then the bookmark is placed at the last "yes" item.
The exercise will normally go with 3 rounds until they finally integrate the cut-score for pass/fail on the Norwegian reading proficiency test.

Words: 150

Answered.

18 LR2H22

A scale is being developed to measure satisfaction with life with the intended purpose to use the scale in national survey to identify which factors are associated with high satisfaction of life in the population. The scale consists of Likert items. According to the underlying theory of satisfaction with life, it is a unidimensional attribute. The theory also states that satisfaction with life is expected to have differences based on gender.

With this information in mind, do the following:

- Describe what evidence sources you want to consider in order to evaluate the validity of the scale scores for their intended purpose
- Describe the data you would like to collect to conduct the validity study
- Describe the analyses you would do in the validity study
- Outline what results you would consider as evidence supporting the validity of using the scale scores in the national survey

Fill in your answer here

In order to evaluate the validity of the scale scores for their intended purpose, evidence based on the contents and evidence based on internal structure should be considered.

A random sample can be chosen from the target population (around the country) with recording the demographic characteristics (e.g. gender) of all sampling participants, and asking the participants to do the test independently. Then we can get the item scores and test scores with this sample.

CFA should be used in order to find the evidence based on internal structure. We can fit a single factor model with the scale. Then we can check the Chi-squared, GFI, RMSEA, and SRMR to see if the model fits well. If they are all match the cut-off values(Chi-squared's p-value large than 0.05, GFI >=0.95, RMSEA < 0.05, SRMR <0.08), then we have the evidence to say the single factor fits well, which corresponds to the unidimensionality of the underlying attribute with the theoretical framework.

If the single factor fits well, then we can calculate the coefficient omega to estimate the reliability of the sum score to check the precision of the measurement.

We also need to check the mean difference of sum score between men and women. A t-test can be applied to provide evidence that whether there is significant difference of the mean scores between men and women. If the p-value is less than 0.05, then we can reject the null hypothesis that there is no difference between the mean scores.

An expert panel can be employed to check the content of items to see if they are actually measuring all aspects of satisfaction of life corresponding to the underlying theory.

Words: 277

Answered.

19 # LR3H22

The following output was obtained from estimating a single factor model to five 4-category Likert scale items from a scale measuring the environmental awareness of 15-year olds in Norway.

Item	Factor loading	Error variance
1	2.00	4.00
2	3.00	2.00
3	1.00	4.00
4	2.00	5.00
5	2.00	1.00

The residual correlation matrix was

$$\boldsymbol{\Sigma}_{\text{res}} = \begin{pmatrix} \mathbf{0.000} & & & & \\ \mathbf{0.026} & \mathbf{0.000} & & & \\ \mathbf{0.017} & \mathbf{-0.035} & \mathbf{0.000} & & \\ \mathbf{-0.014} & \mathbf{0.072} & \mathbf{-0.019} & \mathbf{0.000} & \\ \mathbf{-0.025} & \mathbf{-0.039} & \mathbf{0.020} & \mathbf{0.009} & \mathbf{0.000} \end{pmatrix}.$$

Address the following in your response:

1. What validity evidence categories from the Standards for Educational and Psychological Testing are relevant in this analysis? (1p)
2. Based on your appraisal, does the single factor model fit well?
3. Assume that a single factor model is appropriate for the analysis of the five item scores. Which item contributes the most to the reliability of the sum score and which item contributes the least? Justify your answers. (1p)
4. From the description of the items above and the results of the estimated model, give **one reservation** against the use of the linear factor model in this case. (1p)

Fill in your answer here

1. Internal structure

2. Yes. The residual correlation matrix measures the difference between the standardized model-implied covariance and the standardized sample covariance. According to the cut-off value (<0.1), we can say the model fits well.

3. We can calculate the information of each item to see the contribution of each item to the reliability of the sum score.

$\frac{\lambda_1^2}{\psi_1^2} = \frac{4}{4} = 1$

$\frac{\lambda_2^2}{\psi_2^2} = \frac{9}{2} = 4.5$

$\frac{\lambda_3^2}{\psi_3^2} = \frac{1}{4} = 0.25$

$\frac{\lambda_4^2}{\psi_4^2} = \frac{4}{5} = 0.8$

$\frac{\lambda_5^2}{\varphi_5^2} = \frac{4}{1} = 4$

Based on the calculation results, we can rank the contribution of each item from the highest to the lowest:

item 2, item 5, item 1, item 4, item 3.

Therefore, item2 contribute the most while item3 contribute the least.

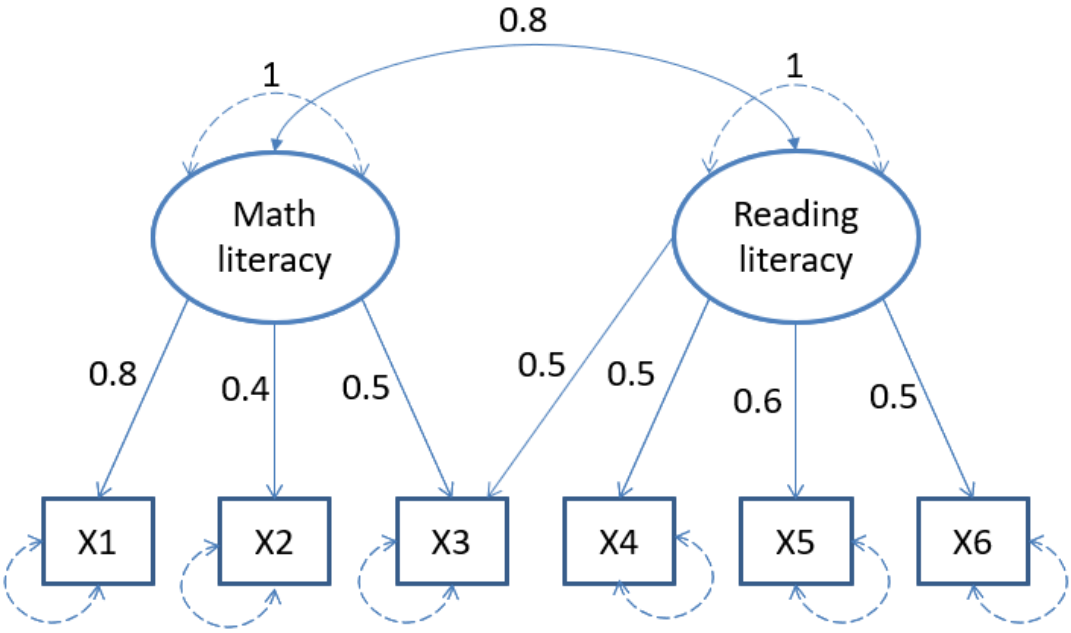
4. Strictly speaking, the 4-category Likert scales for the items are ordinal data which might not be appropriate to be used in a linear factor model.

Words: 128

Answered.

20 LR4H22

A multiple factor model is illustrated in the graph below. The latent variables and the observed variables are all standardized.



Answer the following questions based on the graph.

1. What is the equation which describes the model for the item score **X3**? Write down the equation with an explanation of the parameters and variables included. (2p)
2. What is the covariance between item scores **X3** and **X4** according to the model? Show your work and explain the steps taken. (2p)

Fill in your answer here

1. $X_3 = \mu_3 + \lambda_{M,3}M + \lambda_{R,3}R + E_3$

X_3 is the item score, the indicator
M is the factor of math literacy, R is the factor of reading literacy
E_3 is the error score of item 3,
X_3, M, R, and E_3 are radom variables.
μ_3 is the item difficulty , $\lambda_{M,3}$ is the factor loading of item3 with respect to math literacy
$\lambda_{R,3}$ is the factor loading of item3 with respect to reading literacy
μ_3, $\lambda_{M,3}$, and $\lambda_{R,3}$ are parameters.

2. $X_4 = \mu_4 + \lambda_{R,4}R + E_4$

$$Cov(X_3\ ,\ X_4) = Cov(\mu_3 + \lambda_{M,3}M + \lambda_{R,3}R + E_3\ ,\ \mu_4 + \lambda_{R,4} + E_4)$$
$$= Cov(\lambda_{M,3}M + \lambda_{R,3}R + E_3,\ \lambda_{R,4}R) + Cov(\lambda_{M,3}M + \lambda_{R,3}R + E_3,\ E_4)$$
$$= Cov(\lambda_{M,3}M,\ \lambda_{R,4}R) + Cov(\lambda_{R,3}R,\ \lambda_{R,4}R) + 0 + 0 + 0 + 0$$
$$= \lambda_{M,3} * \lambda_{R,4}\ Cov(M\ ,\ R) + \lambda_{R,3} * \lambda_{R,4}\ Var(R)$$
$$= 0.5 * 0.5 * 0.8 + 0.5 * 0.5 * 1 = 0.45$$

Words: 95

Answered.