UiO : Universitetet i Oslo

CANDIDATE

184102

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:

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

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

$\frac{-7}{\sqrt{10+\sqrt{10}}} = \frac{-7}{\sim 6{,}67} = \frac{-0.0525}{\sim 0.0500} = \sim 0,0001$

The relationship between depression severity and satisfaction with life is nonexistent because the relationship is so weak that it cannot explain anything.

Words: 28

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.

$$Var(X) + Var(Y) - Cov(X, Y)^2 = Var(Z)$$
$$2 + 3 - 1^2 = 4$$

Var(Z) = 4

2.

$$Var(X) + Var(2Y) + Cov(X, 2Y)^2 = Var(U)$$
$$2 + 6 + 1,6^2 = 10,25$$
$$8 + 2.56 = 10.56$$

Var(U) = 10,56

This is because:
Cov(X) = 0,4
Cov(Y) = 0.6

Words: 49

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

$\alpha = 5\frac{1^2}{10} = 0.10$

The coefficient alpha is weak because variance is very high.

Words: 11

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

The test could be influenced too much by loud distracting noise from outside construction work.

The test could be influenced by too little if the test has an overwhelming majority of questions about mathematics.

Words: 34

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

Y - 6 = 24
24/1.2 = 20
X = 20

Since the pass score for Y is calculated from 1.2X + 6, we only need to subtract the 6 and divide the answer by the weight of X. Hence, X = 20, and the test X has a cut off score for passing of 20.

Words: 56

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

While the items are a good fit for the measure of mathematics and the interest of mathematic items because GFI is >= 0.95, RMSEA is <= 0.05 and SMSR is <= 0.08, mathematics and the interest of mathematic items are measuring two different things which has a weak correlation.

Words: 49

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

While the model fit was judged to be satisfactory, the test measures reading more than it measures writing at a 2 to 1 ratio, therefore I would not recommend to use the sum score.

Reading
I1 = 2,0 * 0,5 = 1,5
I2 = 1,0 * 0,5 = 0,5
I3 = 2,0 * 1,0 = 2,0
= 4

Writing
I4 = 1,0 * 1,0 = 1,0
I5 = 1,0 * 0,5 = 0,5
I6 = 1,0 * 0,5 = 0,5
= 2

Words: 82

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

First, we would need to find out at what level the test-takers are learning Norwegian and therefore we have to assess the teachers. If we cannot evaluate this ourselves, we need experts to help us.

Second, we need to look into the sample and make sure that the test sample actually is representative of the intended target population.

Third, we need to get the framework evaluated by experts to deem it good enough to be able to measure a good level of Norwegian reading proficiency.

Fourth, we need to get experts to define a minimum viable level that if reached represents the ability to read Norwegian.

Words: 106

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

1. Construct validity have to be established in order to make sure that the test measures Satisfaction with life and are not influenced by other constructs that might also be related to the same attribute.

2. Since the scale is supposed to be used to infer something about an entire population, we need to make sure that the data gathered is a good sample which is evenly distributed just as the intended target population will be.

3. You would analyze the construct to look if the model fits the data well as the theory suggests and to analyze the gender differences which the theory suggests there are.

4. If we have a good representable sample, the construct has been been deemed a good model fit by the data, the there are clear differences between gender and we have also tested that the test does not measure any other attributes, then it would be a valid study.

Words: 156

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. Relation to other variables (Criterion).

2. The correlation between item scores residuals are close to nonexistent, but this is not necessarily enough information to be able to justify a single factor model.

3. Item 4 contributes the most, and item 3 and item 5 are tied for contributing the least. Calculations bellow:

$$\textit{Item } 1 \;\; \mathbf{2.00^2 \times 4.00 = 16}$$
$$\textit{Item } 2 \;\; \mathbf{3.00^2 \times 2.00 = 18}$$
$$\textit{Item } 3 \;\; \mathbf{1.00^2 \times 4.00 = 4}$$
$$\textit{Item } 4 \;\; \mathbf{2.00^2 \times 5.00 = 20}$$
$$\textit{Item } 5 \;\; \mathbf{2.00^2 \times 1.00 = 4}$$

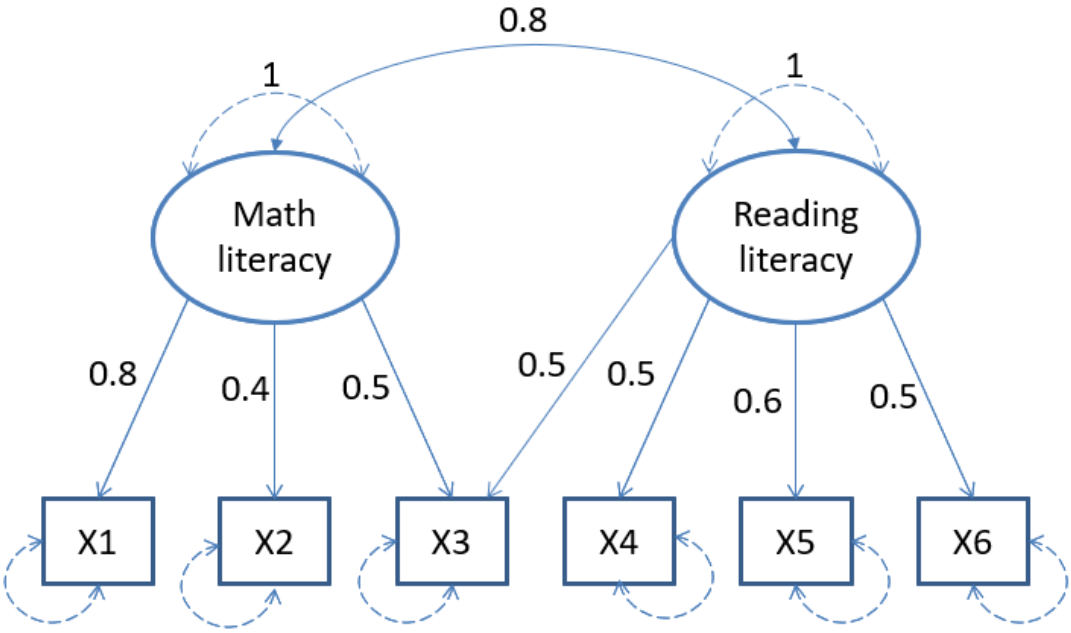
4. While the residuals are low, the item reliability differ very much and therefore suggests that something else is happening that the linear model cannot account for.

Words: 121

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.

$$x3 = \mu M + \mu R + \lambda M + \lambda R + \lambda MR + \in X3$$

X3 = 1 * 1 * 0,5 * 0,5 * 0,8 = 0,2

X3 = the item

μ = uniqueness

λ = factor loading (labda)

Math/Reading (M,R) = Factors

$\in$ = error variance

2.

Factor loading between the items and their factor times the factor loading between the factors times the factor loadings of the factors.

Factor loading between X3 and Math literacy = 0.5
Factor loading between X3 and Reading literacy = 0.5
Factor loading between X4 and Reading literacy = 0.5
Factor loading between Math literacy and Reading literacy = 0.8

Optionally :
Factor loading of Math literacy = 1
Factor loading of Reading literacy = 1

0.5 * 0.5 * 0.5 * 0.8 * 1 * 1 = 0.125 * 0.800 = 0.100

Words: 145

Answered.