UiO : Universitetet i Oslo

CANDIDATE

184111

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

cov(X,Y)/σx * σY = (-7)/$\sqrt{10}$*$\sqrt{10}$=-7/10= -0,70
There is linear negative high correlation between depression and satisfaction with life scales: a high score in depression is associated with a low score in satisfaction with life. This assuming there is no sampling error (assuming the large data was representative of the target population). Another possible factor intervening in this result, is the measurement error.

Words: 62

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(Z)= var(X) + var(Y) - 2*cov(X,Y)
Var(Z)= 2 + 3 - 2*1
Var(Z)= 2 + 3 - 2
Var(Z)= 3

2. var(U) = var(X + 2Y)
var(U) = var(X) + var(2Y) + 2*cov(X,2Y)
var(U) = 2 + 2^2 var(Y) + 2*2*cov(X,Y)
var(U) = 2 + 4*3 + 4*1
var(U) = 2+12 + 4
var(U) = 18

Words: 58

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

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

and interpret it. State the assumptions underlying the interpretation.

Fill in your answer here

α=5 * (1^2 /10) = 5*(1/10)=5/10=1/2=0,5
We assume the single factor model fits well and that the sum score is a proxy of the attribute intended to be measured. Since there is a common factor loading for all items, there was tau-equivalency as well. In this case, alpha coefficient is an unbiased estimator of the reliability of the sum score. Nevertheless, the α=0,5 is not consider acceptable reliability coefficient.
We also assume the data collected to compute this reliability coefficient was representative of the population.

Words: 86

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: for example when the test measures something that is beyond the purpose of the test (i.e. beyond: reading, mathematics and science knowledge and skills to meet real-life challenges): such as anxiety or motivation of the students. In this case, the test-scores will be influenced by too much.

2. too little: If the test under represents the constructs that is intended to measured (i.e. the test doesn't include some aspects of the constructs), then we would say the test scores might be influenced by too little. This would happen for example if the aspects for the construct of mathematics are geometry, algebra and arithmetic and the test content doesn't include algebra.

Words: 113

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

eq(Y)=1.2*X + 6
30=1.2*X + 6
30-6=1,2*X
24=1.2*X
24/1.2=X
20=X

The linear equation function showed that the cut-off score for passing the test X is 20. This means that a passing score of 30 in test Y is equivalent to a passing score of 20 in test X.
The result was obtained based on the linear equation: eq(Y)=aX + b; where eq(Y)=30; a=1.2 and b=6 for this case.

Words: 68

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

Since the model shows a good fit according to the fit indices (GFI=0.95; RMSEA is below 0.6 and SRMR is below 0.08), then the estimated correlation between the factors (0,5) is appropriate. The sum scores correlation between a test of mathematics and a test of interest in mathematics is lower (0,2) than the correlation between the factors (0,5) due to measurement error.

Words: 62

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

A bifactor model can be used to evaluate unidimensionality. In this case this is done by fitting a model with two subfactors (reading and writing) and a general factor. If the variance explained by the general factor is higher than 0.7, we could argue a unidimensional model is predominant.

By calculting the ECV=

$3^2 + 1^2 + 2^2 + 1^2 + 1^2 + 1^2 / (3^2 + 1^2 + 2^2 + 1^2 + 1^2 + 1^2) + (0.5^2 + 0,5^2 + 1^2) + (1^2 + 0.5^2 + 0.5^2)=$

17/17+1.5+1.5=17/20=0.85

Calculating the ECV we get that the variance explained by the common factor is 0.85 , which is above the threshold of 0.7, hence yes, I would recommend using a sum score.

Words: 121

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

Since the sample is representative of the population, the items can be organized by level of difficulty in an item-map. The experts (Norwegian teachers) could use a bookmark method, which organizes the item from easiest to hardest, corresponding to levels of difficulty of Norwegian reading proficiency. After that, along with the established framework of the expected Norwegian proficiency; the experts would describe and define clearly the minimum passing standard of Norwegian reading proficiency for which they would assign the cut-off value in the test.

Words: 84

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. Describe evidences

1.1 evidence of internal structure: this evidence evaluates the intra-construct relationships. Such as dimensionality, reliability and different item functioning for different groups.

1.2 evidence based on test content: this evidence evaluates the degree of alignment between the content of the test and the attribute being measured

2. Describe data collected to conduct the validity study: I would like to collect a representative sample (a random sample) of the target population, that is, in the country the national scale the test will be applied.

3. Describe analyses.

3.1 evidence of internal structure

3.1.1 dimensionality: since the theory suggest satisfaction with life is a unidimensional construct, I would conduct a CFA to confirm this and evaluate if a single factor model fits the data well.

3.1.2 reliability. if the single factor mode fits to the data in the previous step, I would compute omega coefficient to find out how reliable is the test.

3.1.3 DIF: Since the theory suggest to expect differences based on gender, I would conduct a differential item functioning for the two groups of men and women. To evaluate if the mean of the factor is the same for the two groups for people in the same ability level.

3.2 evidence based on test content:

expert judgement: I would gather a group of experts in the construct 'satisfaction with life' and make them evaluate if the items (and the test content in general) actually reflect/align to the attribute appropriately.

4. Outline what results you'd consider as evidence supporting the validity (of using the scale scores in the national survey)

4.1 evidence of internal structure

4.1.1 dimensionality: if the CFA gives appropriate fit measures (GFI > 0.95; RMSEA <0.05, SRMR < 0.08), I would consider that the scale is unidimensional.

4.1.2 reliability: if the omega coefficient is equal or larger than 0.80 I would consider this scale as reliable, that is, consistent and precise to measure satisfaction with life.

4.1.3 DIF: if there are differences in the factor mean for people of same ability level, we would say DIF is present (aligned with theory). If we don't find differences, then DIF is not present. Regardless of the result, the analyses need to be conducted as theory suggests differences by gender.

4.2 evidence of test-content: if the experts judge the test-content is aligned to the target construct, evidence of test content is provided.

(Considering the sample was representative) All these 4 results together are evidence that the scale can be used in national survey to identify which factors are associated with high satisfaction of life is justified.

Words: 429

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. Considering the evidence presented by this analysis, the AERA evidence category of internal structure is relevant.

2. To evaluate the model fit, we need to look at the residual correlation matrix. Values < 0.1 would indicate acceptable fit. Considering the previously explained, the single factor model fits well.

3. Calculations:
Item 1: 2^2 / 4=4/4=1
Item 2: 3^2 / 2= 9/2=4.5
item 3: 1^2 /4=1/4=0.25
item 4=2^2 /5=4/5=0.8
item 5: 2^2 / 1=4/1=4
answer: Item 2 contributes the most to the reliability of the sum score and item 3 contributes the least.

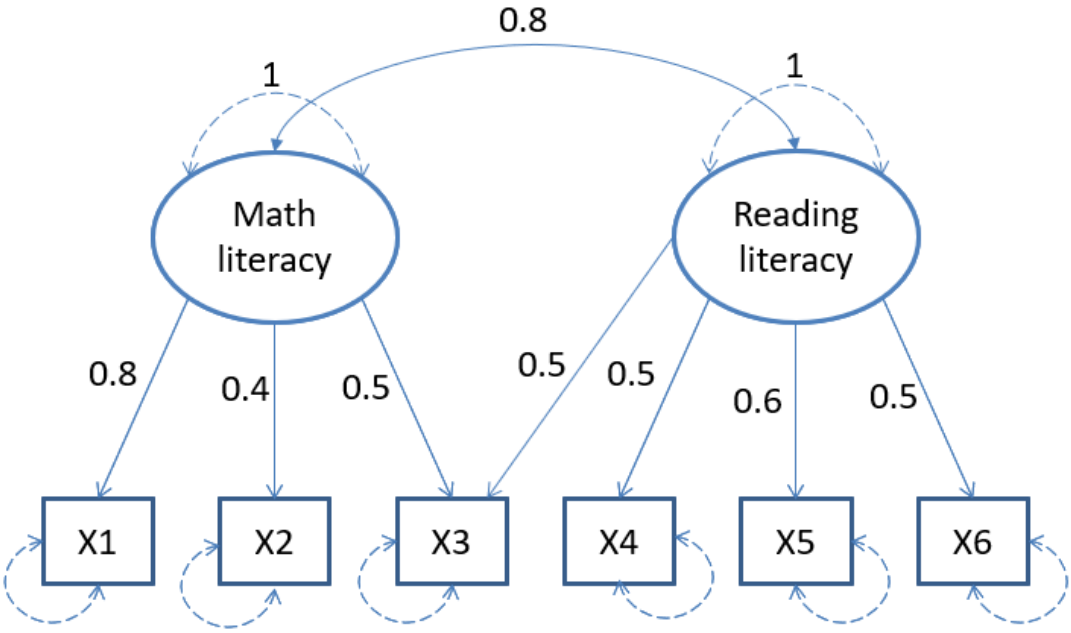
4. Observation against the use of the linear factor model in this case: The items are ordinal.

Words: 110

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. equation:

$X3 = \mu_3 + \lambda_{M3} * M + \lambda_{R3} * R + e_3$

or replacing the values of the factor loadings

$X3 = \mu_3 + 0.5 * M + 0.5 * R + e_3$

explanation:

μ_3 is the intercept for the item score X3, is a parameter

λ_{M3} is the factor loading, how sensitive is the item for measuring the factor math literacy, is a parameter

M is the factor (or attribute) math literacy, is a random variable

λ_{R3} is the factor loading, how sensitive is the item for measuring the factor reading literacy, is a parameter

R is the factor (or attribute) reading literacy, is a random variable

e_3 is the error term of the item score X3, is a random variable

The variance of the error term ψ^2 is the amount of random error, the unique variance. Is a parameter.

2. cov(X3,X4)=0,5*0,8*0,5=0,2

In order to get the covariance between two items, we need to multiply the two item's factor loadings (in reference to each factor in this example): 0.5*0.5. along with the covariance between the factors: 0.8. Hence, the complete calculation is cov(X3,X4)=0.5*0.8*0.5=0.2

Words: 195

Answered.