UiO : Universitetet i Oslo

CANDIDATE

184121

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

var(X)= 10 and Var (Y) = 10
Cov(XY)= -7

So, Cor(XY)= -7/(sqrt of 10)(sqrt of 10) = -0,7
So the correlation between depression and satisfaction with life is fairly strong and the relationship is negative meaning that depression severity increases as the life satisfaction level reduces.

Assumptions made here are
- depression and satisfaction with life are in a linear relationship with each other
- there exists sampling variance
- the relationship is influenced by the sampling variability

Words: 78

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)+ 2Cov(XY)
= 2+3+2*1 = 7

2. Var (U) = Var (X+2Y)= Var(X) + 2^2Var(Y)+ 2*2*Cov(XY)
= 2+4*3+2*2*1 = 18

Words: 28

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

Given that, m= 5, lambda = 1 and Var(Y)= 10
So alpha= 5*1^2/10 = 0.5

Coefficient alpha indicates here the reliability of sum-score of the items -which is 0.5 means there exists no good reliability in using sum score under the assumption that in a single factor model all factor loadings are equal.
Assumptions also underlying that the coefficient alpha is associated with measurement errors. this is the lower bound of the reliability, the reliability of factor would be larger than coefficient alpha. we can get higher reliability after attenuation.

Words: 90

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 standards stated the way how test scores influenced by too much can happen when the construct is irrelevant to the purpose. When the test includes irrelevant items that do not focus on the subject related issues and try to measure beyond the purpose.

1) Example of construct irrelevance from the given scenario could be-

while measuring ability to use mathematics knowledge and skill, if the item is overloaded with reading comprehensions (with lots of jargons and complex sentence) that makes the test hard for the test takers to demonstrate their knowledge in mathematics.

On the other hand. test scores influenced by too little is considered as the construct under-representation. If the test fails to reflect the main focus of measurement.

2) Example could be from the case that-

while measuring ability to use mathematics knowledge and skill, the item should contain all domains of mathematics-lets say algebra, calculus and geometry. If the test only contains items related to algebra then the construct would not reflect the mathematics knowledge.Or if the test item contain too few questions related to one specific domain as compared to others would be influenced by too little.

Words: 192

Answered.

14 SR6H22

For two tests of reading comprehension, X and Y, the linear equating function was estimated to be $\mathbf{eq}(Y) = 1.2X + 6$. 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

The linear equating function define the relationship as
Y= 1.2X + 6
so X= (Y - 6)/1.2

Given that, Cut score for passing the test Y = 30
so, cut score for pass in interns of the test X = (30-6)/1.2 = 24/1.2= 20

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

Looking at the GFI, RMSEA and SRMR which are all good that the model fit is well with two factors. But the sum-score correlation is 0.2 because of having the measurement errors. when in the factor model this corelation is attenuated by dividing each individual factors reliability (internal consistency) - factor corelation gets higher to 0.5.

Words: 56

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

Since the bifactor model fits satisfactory, to find the reasonable use of sum-score, we can estimate explained common variance (ECV) for the model and check whether it meets the criteria of unidimensionality.

ECV is expressed as the ratio of sum of variance of factor loadings of General factor and sum of variance of factor loadings for all factors.

Sum of variance of factor loadings of General factor = 9+1+4+1+1+1= 17
Sum of variance of factor loadings of reading factor = 0.25+ 0.25+ 1 = 1.5
Sum of variance of factor loadings of writing factor = 1+0.25+ 0.25 = 1.5

so ECV = 17/(1.5+1.5)= 17/20= 0.85
which is greater than 0.7 (as defined as cut off value for unidimenationity)

So sum score can be used as a measure.of the construct

Words: 129

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

Standard setting is one the item scaling procedures for defining specific level that is most popular in real.
Standard setting procedure in the given case to define the cut-score for pass/fail on the Norwegian reading proficiency test should be as follows:

1. Formation of a subject matter expert panel - in this case, Norwegian language experts including a statistician
who will discuss and develop a descriptor for the Norwegian language proficiency levels.
2. Bookmark setting procedure can be followed in this case, in bookmark setting, cut off value for pass/fail will be defined after the test score is received.
3. Item would be mapped and ordered according to ascending order of item difficulty level
4. expert group will collect data from test scores and through item anchoring, define item based on descriptor on what the test takers are able to do and what not
5. A cut-off value for pass/fail on the Norwegian reading proficiency test.is set by experts' opinion and consensus.

Words: 162

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. From the Standards for Educational and Psychological Testing, the validity evidence sources to be considered in this would be test-criteria relationship to see how the scale identify the independent variables (latent factors) to measure satisfaction with life. As well as evidence sources from internal structure would be necessary to see intra-construct reliability.

2. - Collection of data from groups of population to whom the test is administered to.
- The data for male and female participants in the test.
- Likert scale details whether it is 4 or 5 point likert scale

3. The analysis should cover the all use of validity evidence.
confirmatory factor analysis for
- Unidirectionality check and model fit analysis
statistical analyss for
- Internal consistency by analyzing reliability coefficient
- Mean difference of the sum-scores of different gender groups

4. I would expect
- good model fit parameters value
- High reliability coefficient to provide confidence in measurement precision
- Different means for male and female.
- Variance for male and female population to see potential item bias

Words: 174

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. From the Standards for Educational and Psychological Testing, the validity evidence categories relevant in this analysis would be test content and internal structure., where the use and interpretation of score tap to see the subject content knowledge and whether the intra-construct reliability exists

2. .Yes, single factor model can fit well looking at the data given and since the factor contains more than 2/3 items and the residual correlation matrix shows residual for all items are below 0.1 as a cut-off value defined by McDoanld (1999); however for model performance we would be able to look into key indicators like GFI, RMSEA, SRMR etc upon factor analysis. we can also estimate measurement precision to see the model fitting as below:

sum of Var(factor loadings) = 4+9+1+4+4 = 22

sum of error variance = 4+2+4+5+1= 16

omega = 22/(22+16) = 22/38

so the measurement precision is approx 60% or reliability can be 0,6 approx

3. To know which item contributes the most to the reliability of the sum score and which item contributes the least, we can estimate item information as below;

Item information is the ratio of Var(factor loadings) and error variance

Item 1 = 4/4 = 1

item 2= 9/2 = 4.5

item 3 = 1/4 = 0.25

item 4 = 4/5 = 0.8

item 5 = 4/1 = 4

so item 2 contributes the most to the reliability of the sum score and item 3 contributes the least

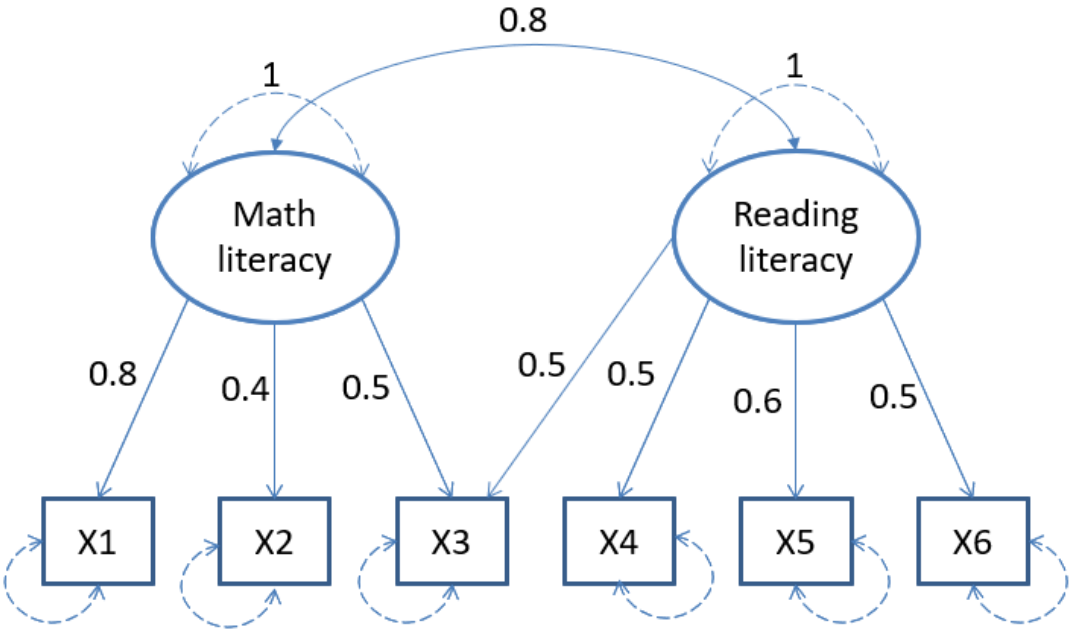
4. All items do not seem to be normally distributed and not with uniform variance. Some itemmay not be in liner relation with outcome. And some items have low measurement precision that linear modelling might put threats to the construct evaluation.

Words: 282

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. The equation for X3 is

$X3 = u3 + L_{3M*}M + L_{3R}R + E3$ OR $L_{3M*}M + L_{3R}R + E3$

where, X3 is the item score. M is the factor score for math literacy, R is the factor score for reading literacy and E3 is the error terms for item X3- all of which are random variables. L denotes for Lambda that is factor loading and L3M and L3R are the factor loading for item 3 associated with factors math and reading literacy- which are parameters. u3 is item difficulty for X3 is a constant.

2. Item equations for X3 and X4 would be

$X3 = L_{3M*}M + L_{3R}R + E3$ and

$X4 = L_{4M*}0 + L_{4R}R + E4 = L_{4R}R + E4$

$Cov(X3,X4) = Cov (L_{3M*}M + L_{3R}R + E3.\ L_{4R}R + E4)$

$= Cov (L_{3M*}M.\ L_{4R}R + E4) + Cov(L_{3R}R, L_{4R}R + E4) + Cov (E3, L_{4R}R + E4)$

$= Cov(L\ L_{4R3M*}M.\ L_{4R}R) + Cov(L_{3M*}M.\ E4) + Cov (L_{3R}R, L_{4R}R) + Cov(L_{3R}R, E4) + 0$

$= L_{3M*}\ L_{4R*}\ Cov(M.R) + 0 + L_{3R*}\ L_{4R*}\ Var\ (R) + 0$

$= L_{3M*}\ L_{4R*}\ Cov(M.R) + L_{3R*}\ L_{4R*}\ Var\ (R)$

$= 0.5 * 0.5 * 0{,}8 + 0.5 * 0.5 * 1$

$= 0.2 + 0.25$

= 0.45

so the covariance between item scores **X3** and **X4** is 0.45

Words: 215

Answered.