UiO : Universitetet i Oslo

CANDIDATE

184120

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

The relationship between depression and satisfaction with life should have a negative corelation.
The formula for corelation is: cor(x,y)/sd x sd y and the formula for sd is the root of the variance of the parameter. However, I don't know what the root of 10 is, but if I did, I would fill the formulas out with the numbers from the covariance matrix, as it is like this:

var (x)	cov (x,y)
cov (y,x)	var (y)

Still, the covariance is negative, so I believe the relationship is negative, as it should. I still don't know weather its significant, but that`s what it should be, because someone who scores high on satisfaction with life is assumed to score low on depression and opposite.

Words: 122

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

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

Var(U) = Var(X + bY) = Var(X) + b^2Var(Y) + b2Cov(X,Y)
= 2 + 2^2 * 3 + 2 * 1
= 2 + 12 + 2
Var(U) = 16

Words: 53

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

5 * 1 / 10 = 0.5

Coefficient alpha measures internal consistency of the items of a test. 0.5 is not that good as it should be higher. That means that the items don't really measure the same underlying factor/attribute as they are assumed to do in a test measuring the same underlying construct.

Words: 54

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

Too much: The test measures reading as one of the domains. However, it might also accidentally measure reading skills for the mathematics and the science domains. This can be because the test is administered to measure these domains in real life challenges, and therefore one can assume that the questions are of scenarios, written out in text. This way, it is reading that is measured, not mathematics and science to the extent which these two are supposed to be measured.

Too little: This is the opposite of too much, as the mathematics and science can be measured to a less extent and not reflect the actual reality that it was intended to. This will return an inaccurate measure of the two, as items with text may be challenging for some to read, who could answer wrong, but would actually have quite strong mathematics skills. This can influence for instance the mean of knowledge for the two domains (possibly reduce it), as it is supposedly used to measure Norwegian 15-year olds on a large scale, and for whatever intended use the results are meant for, this may get unsuited consequences for the actual need because it doesn't reflect the full reality.

Words: 200

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) = 1.2X + 6
eq(30) = 1.2X + 6
X = 30 - 6 - 1.2
X = 24 - 1.4
X = 22.8

The cut score for pass in terms of the X test scores is 22.8.

Words: 39

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

0.2 might be because even though some score high on interest, they might not actually be good at math (and the opposite). There fore the low correlation.

0.5 might be because the two factors might measure the same underlying attribute, to some extent, as 0.5 is some correlation, even though a bit weak.

Words: 53

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 factor loadings are okay, but not super satisfactory high. They should preferably be higher, so Im doubtful, but I`ll say yes.

Words: 22

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

Use criterion validity to validate the cut score.

Words: 8

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

- Construct validity, response procedure (Only people who are high or low to an extreme extent may prioritize to answer, or people who are very aware. Might loose many in the mid section, because they are not inclined to answer, and the results become biased.
- A survey to women and men in an interval of age (ie. not include children as it is gender that we have assumptions about and want to check for).
- Confirmatory factor analysis
- High internal consistency, low RMSE and some more that i cannot remember, but they are the ones from the CFA.

Words: 96

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

2. Yes, because of high factor loadings for all the items, indicating that the items measure the factor well so that a low or high score on the item actually reflects a low or high score on F. However, the error variance is also quite big for some of the items, so I`d suggest to cut those items out first.

3. Item 5, as it has the lowest error variance.

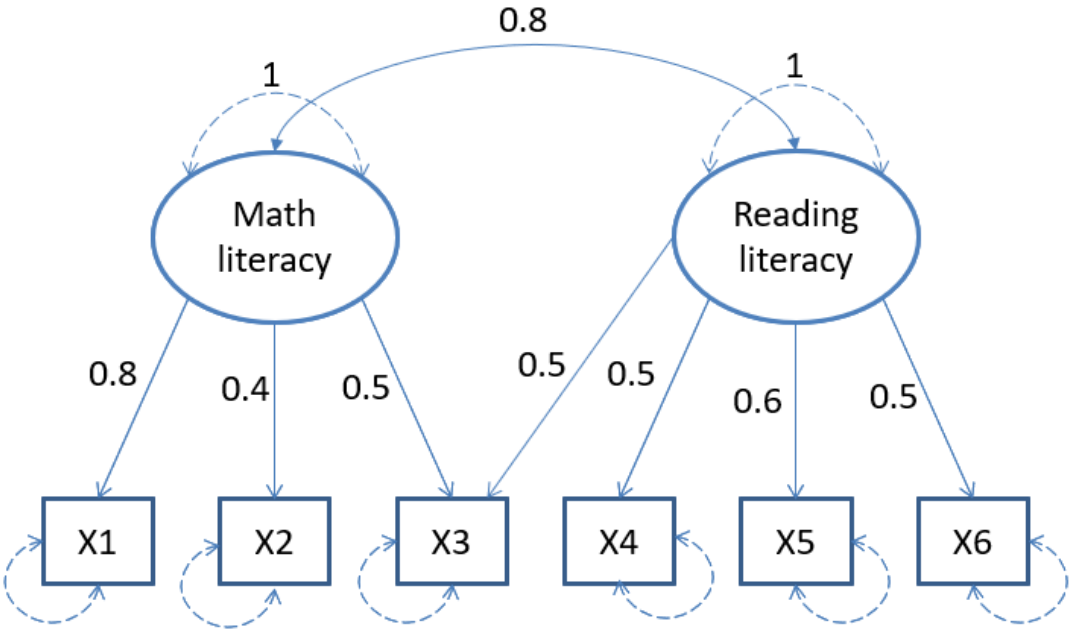
4. High correlation of residuals between item 3 and 2.

Words: 85

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. I do not know.
2. $Cov(X_3, X_4) = lamda_3, lamda_4$
   $= 0.5 * 0.5$
   $= \underline{0.25}$ I did the covariance of X_3 and X_4 between the items concerning only reading literacy.

Words: 32

Answered.