UiO **:** Universitetet i Oslo

CANDIDATE

184113

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:

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

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

There seems to be a linear corelation between the depression severity scores and the satisfaction with life scale scores. This can be observed where the diagonals are symmetrical. There is also a negative relationship between the scales. For example. this would mean that a average positive difference in scores on one of the scales would relate negatively on the other scale.

Words: 61

Answered.

11 SR3H22

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

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

Fill in your answer here

Var(Z)
Z = X-Y
Z = 2-3
Z = -1

Var (U) U = X + 2Y
2+2(3)
2+6
U= 8

Words: 21

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

Since m = 5, λ^2 = 1^2 and Var (Y) = 10
α = 5 * 1/10
α = 0.5

In this case, m is the number of items or 5 in the test, and the factor loadings are 1^2 and variance of Y equals 10. An alpha of 0,5 indicates that there test is reliable and there is a good fit between the model and the observed scores.

Words: 69

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

In this case, the standards are referring to the fact that, reading, mathematics and science knowledge may each be distinct and highly significant entities.Meaning that they each may distinctively and significantly impact the outcome and considering them in a three domain test may be ¨too much¨. This is to state that reliably mapping domain knowledge in vast and encompassing domains such as math and science may necessitate a operationalizing several sub-constructs and items across these three domains. This may lead to a large and disproportionate number of items in these domains as compared to reading as a domain. Therefore item scores may be influenced buy such disproportionally and items to reliably measure the constructs in the three domain test.

(2) too little.

Here the standards are referring to the fact due to testing three vast domains being may lead to having too narrow or too little to reliably represent the domain. This means that the resulting item score may not be reliably reflective of the domain knowledge and skills.

Words: 172

Answered.

14 SR6H22

For two tests of reading comprehension, X and Y, the linear equating function was estimated to be $\text{eq}(Y) = 1.2X + 6$. 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 following equation eq(Y)=1.2X+6 means that , for the two tests of reading comprehension to related.

The scores in test Y would be 1.2 times X scores + 6.

Therefore if the cut off score for Y is 30, then to determine the equivalent cut off score for X we need to find X, when Y = 30.

30= 1.2 X + 6
1.2X =30-6
1.2X=24
X=24/1.2
X=20

Therefore when the Y cut off score is 30, then it goes that the X cut off score for the reading comprehension test is 20

Words: 93

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

The sum score corelation between the test of mathematics and a test of interest in mathematics is weak at 0,2 while the estimated factor corelation is 0,5, which shows a moderate corelation between the entities. The goodness of fit index at 0,95, shows a very good fit. The RMSEA are also within the acceptable limits.

This could be explained by the is positive moderate corelation seen in the estimated factor corelation between scores in interest in mathematics and mathematics test . However, one of the factors, for example -the interest in mathematics among the group of students may observe big variances and due to these there sum scores may be varying. Therefore there is a difference between actual knowledge in mathematics as seen in the scores of the mathematical test and the interest in mathematics in the same group of students. So we can make the statement that while some students ay be proficient in mathematics but may not be interested and the variance in the interest scores of these students account for the difference in the factor corelation and the sum score corelation in this context .

Words: 188

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

> I would not use sum scores in this study based on the estimated factor loadings. This is mainly because using the sum scores would provide a higher sum test cores to reading factor as compared to writing factor when the general factor has been accounted for. The general factor loading on reading items are much higher than that of writing items. The sum scores in the sections would not be indicative of the intended measure assumed to test reading and writing abilities.

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

> The standard setting procedure would outline the items, their factor loading and their error variance. The variance and the corelation will then be estimated.

Words: 24

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 the study of validity I would look at
Construct validity
Concurrent Validity
Conceptual framework

Words: 15

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

$$\Sigma_{res} = \begin{pmatrix} 0.000 & & & & \\ 0.026 & 0.000 & & & \\ 0.017 & -0.035 & 0.000 & & \\ -0.014 & 0.072 & -0.019 & 0.000 & \\ -0.025 & -0.039 & 0.020 & 0.009 & 0.000 \end{pmatrix}.$$

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.
2. The single factor model does fit well in this case as this measures environmental awareness
3. Item 5 contributes the most to the reliability of the sum score and which item 4 contributes the least. This is because of the high factor loadings in item 5 as compared to the error variance. And by the same evaluation item 4 that has a very high error variance and lower factor loading
4. One **reser**vation against the use of the linear factor model in this case would be because of the 4 category linear item scale, meaning that the categories may lead to bias in scores

Words: 102

Answered.

LR4H22

A multiple factor model is illustrated in the graph below. The latent variables and the observed variables are all standardized.
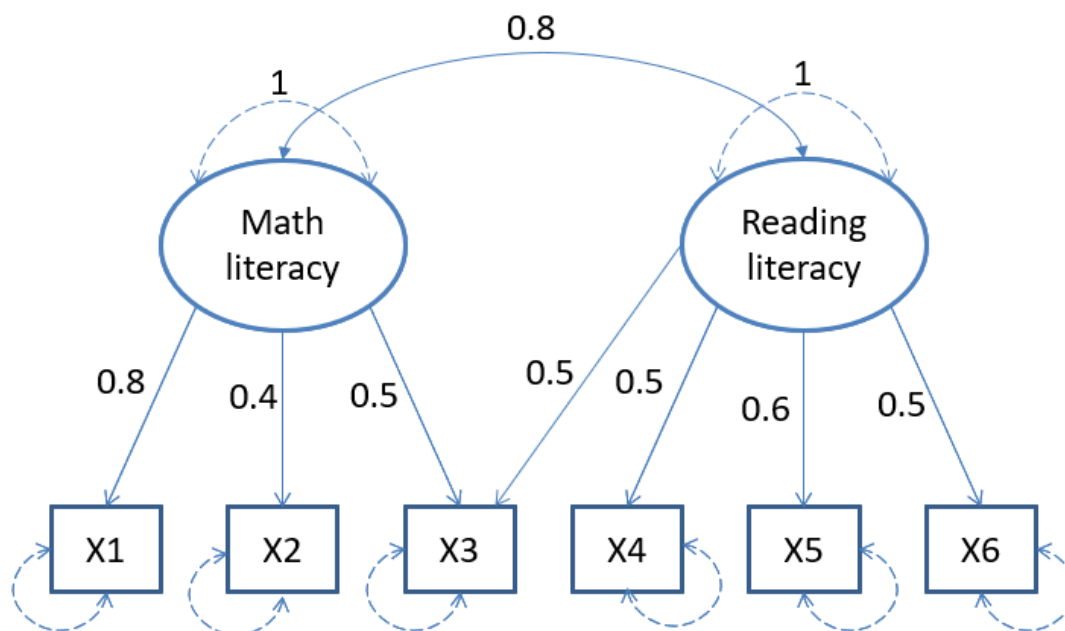
Answer the following questions based on the graph.

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

Words: 0

Answered.