UiO : Universitetet i Oslo

CANDIDATE

184106

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

We assume that there is a linear relationship between the depression severity scale and the scale scores of satisfaction with life. The correlation between these is -0.7. This implies that there is a strong negative linear relationship between depression and satisfaction with life.

$$cor(x,\ y) = \frac{cov(x,y)}{\sqrt{var(x)}\cdot\sqrt{var(y)}} = \frac{-7}{\sqrt{10}\cdot\sqrt{10}} = \frac{-7}{10} = -0.7$$

Words: 55

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

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

$$var(U) = var(X + 2Y) = var(X) + 2^2 var(Y) + 2(2 \cdot 1)cov(X, Y)$$
$$var(U) = 2 + 4(3) + 4(1) = 18$$

Words: 44

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

$$\alpha = 5 \cdot \frac{1^2}{10} = 0.5$$

Assuming a single factor model with uniform factor loadings, the coefficient alpha is equal to the reliability of the sum scores. Hence, this model has 50% reliability.

Words: 34

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: anxiety of the examinee

too little: too easy or too difficult tests

Words: 14

Answered.

14 SR6H22

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

$$30 = 1.2X + 6$$
$$X = \frac{30-6}{1.2} = \frac{24}{1.2} = 20$$

Given the linear equating function for the cut score of test Y, the cut score for pass in terms of test X is 20 as shown the calculation above.

Words: 43

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

An acceptable model fit has GFI greater than or equal to 0.95, RMSEA greater than or equal to 0.06 and SRMR greater than or equal to 0.08. With these cut-off points, the model fit is acceptable to the data collected. The correlation of the sum scores is lower than the factor correlation due to the consideration of measurement errors in the computation of the factor correlation.

Words: 66

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 weight of a factor loading is relevant in accessing how well a model fits a given dataset. A factor loading of 0.7 or higher measures the item well whiles a factor loading less than or equal to 0.3 does not measure the factor well enough. Given that items 2 and 3 of the Reading test and items 5 and 6 of the writing tests have factor loading of 0.5. This is an indication that these items moderately measure the corresponding factors.

The

Words: 83

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

Assign scale scores to each of the items in the construct and set cut score points for pass and fail. Test descriptors are used to determine the competence or proficiency of the examinees who attain the cut scores. A group of subject experts are consulted to assess the validity of the scores and the corresponding description of examinees who attain such scores. The test is conducted on a group of individuals who meet the criteria of the construct and a systematic approach is used to collect information from a such a test and possible adjustments are made in the pass/fail cut-off scores.

Words: 102

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

Evidence based on the content of the survey. This considers the degree to which the items in the survey measure the construct, satisfaction with life. Given that this is a unidimensional construct with expected differences based on gender, the response rate of both genders should be considered. The proportion of males to females within the sample should reflect the response rate of both genders. Researchers should include items which draw the distinction and account for the differences between male and female respondents.

A good model should have a goodness of fit index greater than or equal to 0.95, RMSEA greater than or equal to 0.05 and SRMR of at least 0.08. It could be considered as acceptable if it has a GFI of at least 0.9 and RSMEA of 0.1

Words: 130

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. Evidence based on content and evidence based on the response processes are relevant in this analysis.

2. Yes. there is almost no correlation between the items.

3. The item with the most contribution to the reliability of sum scores has the highest ratio of the squared factor loading and the error variance while the opposite contributes the least.

Item	Factor Loading	Error Variance	$\frac{\lambda_i^2}{\psi^2}$
1	2	4	$\frac{2^2}{4} = 1$
2	3	2	$\frac{3^2}{2} = \frac{9}{2} = 4.5$
3	1	4	$\frac{1^2}{4} = \frac{1}{4} = 0.25$
4	2	5	$\frac{2^2}{5} = \frac{4}{5} = 0.8$
5	2	1	$\frac{2^2}{1} = 4$

From the table above, item 2 contributes the most while item 3 contributes the least

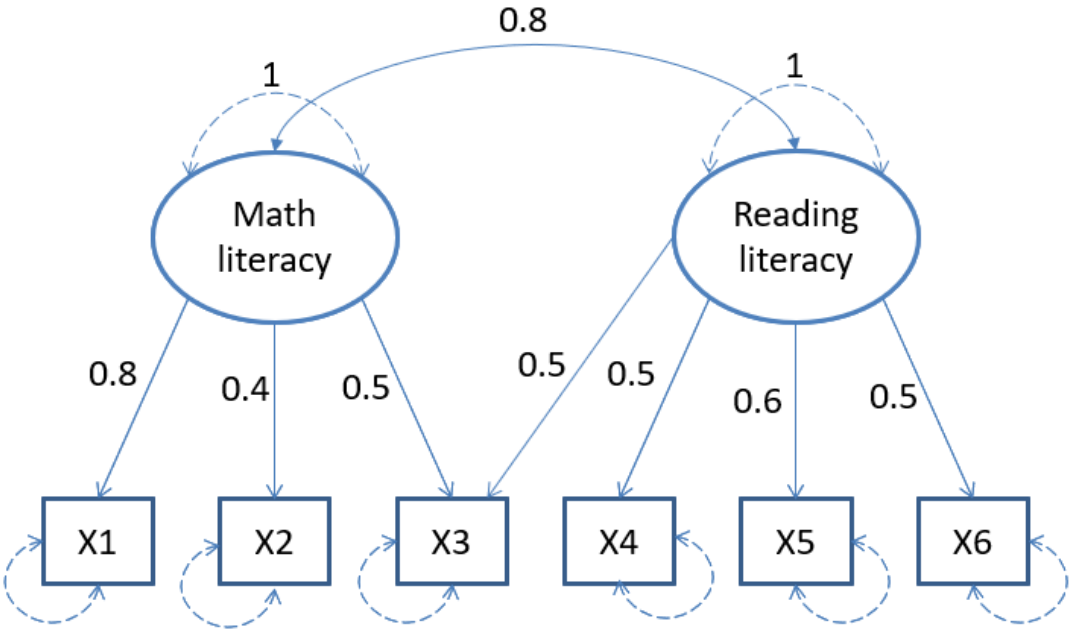
4. It is irrational to compute values from a Likert scale items since these are ordinal measurements.

Words: 133

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.
Item score X3 is described by

$$X3 = \mu_3 + \lambda_{M3}M + \lambda_{R3}R + \epsilon_3$$

where \mu _3 is the level of difficulty associated with item X3
λ_{M3} is the factor loading of the Math literacy for item 3
λ_{R3} is the factor loading of the Reading literacy for item 3
ϵ_3 this is the sampling error associate with the item 3

2.

$$cov(X_3,\ X_4) = cov(X_4,\ X_3)$$
$$cov(X_4,\ X_3) = cov(\lambda_{R4}R + \epsilon_4\ ,\ \lambda_{M3}M + \lambda_{R3}R + \epsilon_3)$$

expanding the terms

$$cov(X_4,\ X_3) = cov(\lambda_{R4}R\ ,\ \lambda_{M3}M + \lambda_{R3}R + \epsilon_3) + cov(\epsilon_4\ ,\ \lambda_{M3}M + \lambda_{R3}R + \epsilon_3)$$

the error term of one test is uncorrelated with the Factor scores and the error terms of a different test. Therefore the covariance between these is 0

$$cov(X_4,\ X_3) = cov(\lambda_{R4}R\ ,\ \lambda_{M3}M) + cov(\lambda_{R4}R\ ,\ \lambda_{R3}R) + cov(\lambda_{R4}R\ ,\ \epsilon_3) + 0$$
$$cov(X_4,\ X_3) = \lambda_{R4}\,\lambda_{M3}cov(R\ ,\ M) + \lambda_{R4}\lambda_{R3}cov(R\ ,\ R) + 0 + 0$$
$$cov(X_4,\ X_3) = \lambda_{R4}\,\lambda_{M3}cov(R\ ,\ M) + \lambda_{R4}\lambda_{R3}var(R) + 0$$

$$cov(X_4,\ X_3)\ =\ 0.5 \cdot 0.5 \cdot 0.8\ +\ 0.5 \cdot 0.5 \cdot 1$$
$$cov(X_4,\ X_3)\ = 0.2\ +\ 0.25\ =\ 0.45$$

Therefore the covariance between item scores X3 and X4 is 0.45

Words: 211

Answered.