UiO : Universitetet i Oslo

CANDIDATE

184114

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

Correlation = $\frac{-7}{\sqrt{10}\cdot\sqrt{10}}$ = -7/10 = -0.7.
Depression severity and satisfaction with life have a moderate to strong, negative, correlation.

The assumption is that there is a linear relationship between depression severity and satisfaction with life.

Words: 36

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

var(X+2Y) = 2var(X+Y)
= 2(var(X)+var(Y)+2cov(X,Y)
= 2(2+3+2*1)
=2(2+3+2)
=2(7)
= 14

Words: 35

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

a = $5\left(\frac{1^2}{10}\right) = 0.5$

Coefficient alpha of 0.5 indicates weak reliability. Outside of acceptable range.

Coefficient alpha is a lower bound of the reliability if the true model is a single factor model.

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 (1):
If, in order to solve some of the problems, the individuals would have to have knowledge outside of the three factors, the test scores would be influenced by too much. This could, for example, be the case if the 15-year-olds have to know traffic rules in order to get the right answers for some question. The test would "test" more than it was supposed to do.

Too little (2):
If the test only cover some parts of the mathematics construct, and does not include every aspect relevant for the suggested test score use, then the test is influenced by too little. The test would test "less" than it was supposed to.

Words: 114

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

Cut score for X = 20.

eq(Y) = 1.2X + 6
(30 = 1.2X + 6) * 10
300 = 12X + 60
300-60 = 12X
240 = 12X
20 = X
X = 20

Words: 36

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 model fit indices indicates acceptable fit.

Scores on tests are always influenced by sampling variation and measurement error. This is why the sum score correlation is lower than the estimated factor correlation. If the test was repeated indefinitely, the sum score correlation would become closer to the factor correlation, if the estimated factor correlation reflects the true model.

Words: 59

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

Using the sum score is in this case not recommended.

This is because item 1-3 have a higher factor loading towards the general factor than the subfactor of reading.

The general factor cold be covering more than just reading and writing.

Words: 41

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 results of the pilot study could be used to check for uni-dimensionality and model fit.
If the model does not fit the data, then maybe the test would need rework before working on cut-off values.

If the model fits the data as expected, then the framework could be used as an expert review, and cut-off could be proposed based on this.
The expected level of Norwegian reading proficiency from the framework could be compared with the pilot results, to make sure students actually fall within this range.
If all students for some reason are below expected level, the cut-offs might need to be re-evaluated.

Sampling variation and measurement error will always influence test scores. The cut-off scores should reflect this, and maybe be somewhat lower than the framework describes as the expected level.

Words: 134

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 order to achieve the intended purpose of the scale scores, the instrument has to be a valid measure of satisfaction of life.
Experts could be consulted with information about specific items and aspects to include in the scale.
One evidence source could be to collect data from already established scales for satisfaction of life. Data could be collected on the proposed new scale and a "gold standard" test, to see how well they correlate. This could also be done with a similar type of scale, where the results are not expected to be closely correlated. Maybe satisfaction with life is moderately correlated with happiness. High correlation with a scale which correlation is supposed to be moderate could indicate that the instrument actually measures something else.

After deciding on items and scale length, as well as any other scales to include, a pilot study could be
conducted to collect data. This pilot study would also need to collect information of the respondents gender. The data collected would preferably be from a representative sample.
This collected data could be used to perform a confirmatory factor analysis, to check if the uni-dimensional theory holds, using indices such as model fit. The data could be grouped in to gender, and measurement variance between groups can be checked.

If correlation with the other scales are in line with what was theorized, the model shows acceptable fit, and no differences between gender is found, this could be considered as evidence for using the scale-scores in the national study.

If the study indicate differences between gender, the factors associated with high satisfaction of life could be different between genders. Test results could still be used, but they should then always be used with this in mind, and give conclusions/suggestions split between genders.

Words: 296

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

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

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. Relevant evidence category is internal consistency, to check model fit and uni-dimensionality.

2. Moderate to high residual correlation (both negative and positive) could indicate that the single factor model does not fit well.

3. Item contribution = (Factor loading)^2 / Error variance.

Item 1 = 4/4 = 1

Item 2 = 9 / 2 = 4.5

Item 3 = 1/4 = 0.25

Item 4 = 4/5 = 0.20

Item 5 = 4/1 = 4

So, item 2 contributes most, and item 4 contributes least.

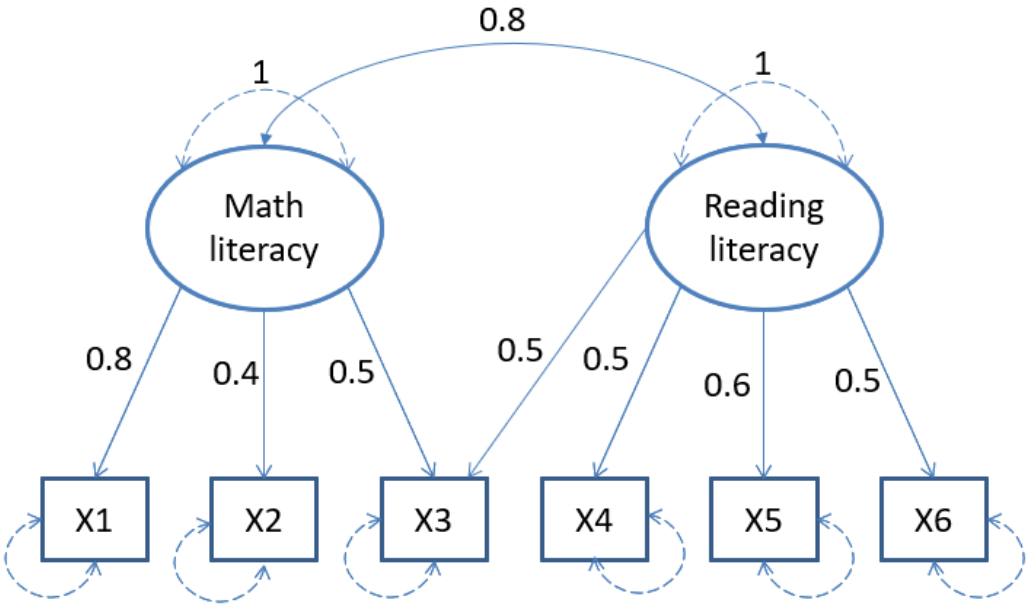
4. Linear factor model should perhaps not be used because of the high residual correlation.

Words: 100

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. $X_3 = \lambda_3 M + \lambda_3 R + E_3$
Where $\lambda_3 M$ = factor loading for Math literacy, for item X3 (M = Math literacy), $\lambda_3 R$ = factor loading for Reading literacy, for item X3 (R = Reading literacy) and E_3 = error variance for item X3.

2. cov(X3, X4)

$= COV(\lambda_3 M + \lambda_3 R + E3,\ \lambda_4 R + E4)$

$= COV(\lambda_3 M + \lambda_3 R + E3,\ \lambda_4 R) + COV(\lambda_3 M + \lambda_3 R + E3,\ E4)$

$= COV(\lambda_3 M, \lambda_4 R) + COV(\lambda_3 R, \lambda_4 R) + COV(E3,\ \lambda_4 R) + COV(\lambda_3 M, E4) + COV(\lambda_3 R, E4) + COV(E3, E4)$

$= \lambda_3 \lambda_4 COV(M, R) + \lambda_4 \lambda_3 VAR(R) + 0 + 0 + 0 + 0$

$= 0.5 \cdot 05(0.8) + 05 \cdot 05(1)$

$= 0.2 + 0.25$

$= 0.45$

Words: 71

Answered.