UiO : Universitetet i Oslo

CANDIDATE

184109

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

$P_{X,Y} = -\frac{-7}{\sqrt{10}\cdot\sqrt{10}} = -0.7$

There's a strong relationship between the two scales. It's negative, meaning that a High score on one scale indicates a low score on the other scale. It's a linear estimate of the relationship between the two scale sum scores. Since we're using sum scores from the tests, we are assuming a single factor model for both scales.

This relationship is affected by sampling variance, and the reliability of the two scales.

Words: 75

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

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

Words: 45

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

α = m* $\frac{\lambda^2}{Var(Y)}$

$\alpha = 5 \cdot \frac{1}{10}$ = 0,5

Coefficient alpha assumes a single factor model. Coefficient alpha is equal to the reliability of the sum score when items are true score equivalent. The value might be indicative of a single factor not being a great fit, where the factor isn't accounting for a substantial amount of the variance in the item scores. The measurement error might also affect the estimated Coefficient alpha.

Words: 70

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 little/ construct underrepresentation: test not covering all parts of the curricula of math, science and reading for the target population.

Too much/ construct irrelevance: Bad conditions during submission of answers, eg: bad lighting in room.

Words: 36

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

$eg(Y) = 1.2X + 6$
30 = 1.2X + 6
30 - 6 = 1.2X
$\frac{30-6}{1.2} = X$
X = 20

The cut score for passing according to the equating function is 20. Indicating the test X is harder than test Y.

Words: 41

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 estimated relationship between factors is always higher than the estimated relationship between sum scores. This is because of measurement error that is inherent to the observed sum score, that the model implied sum score (the factors), have adjusted for.

Words: 40

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

$$\frac{(\sum \lambda_G)^2}{(\sum \lambda_G)^2 + (\sum \lambda_R)^2 + (\sum \lambda_W)^2} = \frac{81}{81+1+1} = 81/83$$

Yes I would recommend using the sumscore as a testiment to the proficiency of the subject being measured. This is because the general factor is dominant, meaning that it can account for more than 70 % of the variance from the items.

Words: 64

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

Based on the responses, first the items should be ordered from least difficult to most difficult.

Then each item will be given a response probability (RP), which is the probability of correctly answering the item given a scale score.

According to the bookmarking procedure, these items are now handed back to the experts. The experts go through the ordered item catalog and answer whether a student, with a given scale score keeping the the qualities of Norwegian proficiency in mind, will the student be able to answer the item correctly. They go through the items, until they reach an item where the answer is "no", a student with that scale score has a low probability of answering that item correctly. When the experts have found a scale score covering the wanted qualities of the curriculum through the high probability of answering items that cover this, that scale score is set as the limit for passing. The standard is set to the scale score.

Words: 163

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

Content evidence
Internal structure (uni-dimensionality)
Consequences of testing; intended by the test developer. That the scale is able to tell what factors make you more satisfied.

A sample from the target population, adults answering the satisfaction with life scale.
I would gather gender metrics from the sample. The intended use of the scale, to be able to identify factors in the population that make them satisfied. For that we would need evidence of test scores from SWL scale being able to concurrently predict factors in the population. For that we would need to gather more metrics from the sample; salary, marital status, job status (just som suggestions).

I would have Content Experts evaluate, so that the items are adhering to the underlying theory of the construct.
I would run a T-Test on the responses of the Satisfaction with life scale, to see if the factor means between men and women are significantly different.
I would do a Confirmatory factor analysis to confirm the uni-dimensionality of the scale.
I would do a regression analysis to confirm the scales ability to predict outcome on factors in the population listed in the second section.

Content evidence: that experts deem the items representative of underlying theory of the construct.
Internal structure: For the interpretation and use of the life satisfaction scale to be valid i would expect, the confirmatory factor analysis indicating a good fit. Fit measures rules of thumb for good fit: RMSEA $\leq$ 0.05, SRMR $\leq$0.08, GFI $\geq$0.95). I would expect the T-test to show a significant difference between mean Factor scores for genders. That the distribution on Factor score are significantly different from one another.
Consequences of testing; Results from the regression analysis would indicate that the scale significantly can account for the variance in salary, marital status, job status.

Words: 299

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) Internal structure. Content evidence

2) $\omega = \frac{(\sum \lambda)^2}{(\sum \lambda)^2 + \psi^2}$ = 100/116

So the factor accounts for the majority of the variance in the items indicating that a single factor is a good fit. The discrepencie matrix indicates no item correlation discrepencies above 0.1, which also indicates a good fit of the single factor model.

3. $\frac{\lambda^2}{\psi^2}$ Based on the information of the items, item 2 contributes the most: $\frac{\lambda^2}{\psi^2}$ = 4.5

Item 3 contributes the least: $\frac{\lambda^2}{\psi^2}$ = 0.25

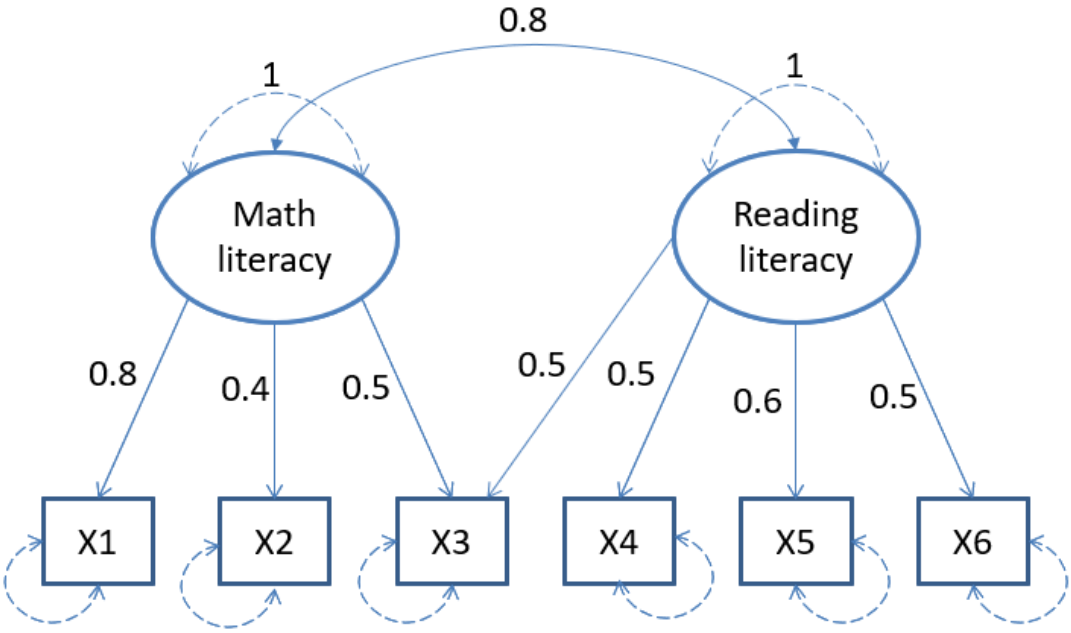
4. The level of an awareness scale is ordinal, the properties of an ordinal scale is that, even though we try, we don't have equal steps between levels of awareness. or rather its not appropriate assigning numerics indicating a numeric step between levels. This should be taken into account when interpreting results - eg. One can't be "double as aware" with a score of two vs. a score of four on factor.

Words: 157

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. $X_3 = \mu_3 + \lambda_{3M}M + \lambda_{3R}R + \epsilon_3$
μ_3 *is the difficulty of the item*. Parameter.
λ_3, is the factor loading. the value of this parameter is the expected difference in item score with 1 factor score difference. Parameter
M, is the factor score on math literacy. Random variable.
R, is the factor score on reading literacy. Random variable.
ϵ_3, is the error score, its the score on item that the factor can't account for. Random variable.

2. This factor model assumes no covariance between errors and factors, and between errors.
$Cov(X_3, X_4) = Cov(\mu_3 + \lambda_{3M}M + \lambda_{3R}R + \epsilon_3,\ \mu_4 + \lambda_{4R}R + \epsilon_4)$
$Cov(\lambda_{3M}M,\ \lambda_{4R}R) + 2Cov(\lambda_{3R}R,\ \lambda_{4R}R)$
$\lambda_{3M} \cdot \lambda_{4R}Cov(M, R) + \lambda_{3R} \cdot \lambda_{4R} \cdot 2Cov(R,\ R)$
$0.5 \cdot 0.5 \cdot 0.8 + 0.5 \cdot 0.5 \cdot 2 \cdot 1 = 0.4$

Words: 124

Answered.