UiO : Universitetet i Oslo

CANDIDATE

184117

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:

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

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

From the covariance matrix $\sum$ the covariance between the scales of depression and satisfaction are -7. That means that there is a negative relationship, such that high score on the depression scale are related with a low satisfaction with life, and likewise opposite.
The correlation will be $cor = \frac{-7}{\sqrt{10} \cdot \sqrt{10}} = \frac{-7}{10} = -0.7$, which is a high negative correlation.

The assumptions are based on the possibility to calculate the covariance and correlation. Aka the best measure is when the scales are continuous.
Furthermore the answering of the scales must be at the same time to have a clear understanding of the relationship between the scales. If there is a huge timegap the circumstances can change and therefore two complete off answering the scales.

Words: 120

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

$$V(Z) = V(X - Y) = V(X) + V(Y) - 2cov(X, Y) = 2 + 3 - 2 \cdot 1 = 3$$

$$V(U) = V(X + 2Y) = V(X) + 2^2 V(Y) + 2 \cdot 2cov(X, Y) = 2 + 4 \cdot 3 + 4 \cdot 1 = 2 + 12 + 4 = 18$$

Words: 2

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

The calculation of α is plotting in the information from the test: $m = 5$, $\lambda = 1 \Longrightarrow \lambda^2 = 1$, and $Var(Y) = 10$.

$\alpha = m \cdot \frac{\lambda^2}{Var(Y)} \Longrightarrow \alpha = 5 \cdot \frac{1^2}{10} = 5 \cdot \frac{1}{10} = \frac{5}{10} = 0.5$

A coefficient α at a value of 0.5 is low, and therefore the reliability is low, which means that the scale is not really good. There is a lot of unexplained variation in the single factor model, given a high variance in the error term.

The coefficient α relays on that the factor loadings of the single factor model are similar, and with an nice fit of the model to the data. With a coefficient $\alpha = 0.5$ it will be questioned if all the factor loadings are of the same magnitude, and there for if a single factor model with the same factor loadings is a proper fit to the data. The factor loadings could be different from each other, and therefore the coefficient α is not the best measurement of reliability. Coefficient ω would have been better.

Words: 156

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

A test takes time, and to measure three domains in one-setting is quite a challenging process. To have a reliable measure there is need for a certain amount of items. More domains would require more items, and in the end a fatigue element will affect the respondents. To avoid the fatigue element the number of items in each domain will be reduced.

So an effect that influenced the test scores too much will be fatigue element, because it is the a question of how fresh the respondent are in the end, and not the ability to answer the items.
The effect that influenced the test scores too little will be the reduction in the number of items in each domain. The judgement based on test scores will be based on a smaller sample of items, which will lead to more uncertainty on the estimate of ability, aka standard error of measurement.

Words: 151

Answered.

14 SR6H22

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

To analyze this linear equation function will be reversed, by the following method:
We know the cut off value of test Y, which is 30. That is placed in the linear equating function, and the isolated for X.
$30 = 1.2X + 6 \Leftrightarrow 30 - 6 = 1.2X \Leftrightarrow \frac{24}{1.2} = X \Leftrightarrow X = 20$.
To have the same cut-off score in test X as in test Y, a respondent need a score of at least 20 to pass the cut-off score of 30 in test Y.

Words: 73

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

With two tests, each based on the own sample of items, the sumscores are affected by errors, while the factor scores are not..

Each factor in the two factor model follows the same structure of single factor model:

$$X_i = \mu_i + \lambda_i F_j + \epsilon_i$$

Where the ϵ_i is the error term of that item. when calculated the sumscores the scores is calculated by summing each X_i in the factor. Then the error terms is affecting the sumscores.

For the factor scores the measurement of the factor are independent of the error terms, and therefore a more clearcut measurement of correlation since no error terms to affect the scores.

So the errors are explaining the differences of the correlations.

Words: 116

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

a bifactoral model are representated of the following factor model:

$X_i = \mu_i + \lambda_{Gi}G + \lambda_{Ri}R + \lambda_{Wi}W + \epsilon_i$

where the factor loading of the general λ_{Gi} is affecting all items, while the factor loadings of each bi-factor (reading (R) and writing (W)) are only affecting some of the items. for item 1-3: $\lambda_{Ri} > 0,\ \lambda_{Wi} = 0$, and for item 4-6: $\lambda_{Ri} = 0,\ \lambda_{Wi} > 0$.

The variance of the sumscores $Y_i = \sum X_i$ will be

$Var(Y) = Var\left(\sum_{i=1}^{6} X_i\right) = \sum_{i=1}^{6} var(X_i) = \sum_{i=1}^{6} \left(\lambda_{Gi}^2 Var(G) + \lambda_{Ri}^2 Var(R) + \lambda_{Wi}^2 Var(W)\right)$

Normally the factors are standardized $Var(G) = Var(R) = Var(W) = 1$.

$Var(Y) = 3^2 + 0.5^2 + 1^2 + 0.5^2 + 2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0.5^2 + 1^2 + 0.5^2$

$Var(Y) = 9 + 0.25 + 1 + 0.25 + 4 + 1 + 1 + 1 + 1 + 0.25 + 1 + 0.25 = 20$

The practice of using sum scores, will be okay to continue using since the variance are some kind of low compared to just the general factors variance

$Var(Y_G) = 9 + 1 + 4 + 1 + 1 + 1 = 17$. The factor loadings of the bifactors are quite low, and do not affected the variance of the sumscores much.Then the reliability will be all right and in line with the rest.

Words: 138

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

To make an defining cut-off value to be proficiency, multiple steps must be taken.

1: The collected data should come from the targeted population.

2: The collected data and items should be correct specified, should that it measure what is relevant. The construct must not underrepresented, or misrepresented. Aka the construct should be valid.

3: The collected data must fit an appropriate model with a high reliability, such that the measurement are relative precis.

4: In collaboration with expects the specific cut-off are determined by content. Aka what should the respondent master at this level. (Or it could be some norm-cut-off value, e.g. 22%)

5: Figure out what cut-off value that is on the factor. An item map would help to figure out content level. If it is norm based figure out the sumscore of the test, in the population, e.g. what is the sumscore of the 22% in the population have, and then use that as the cut-off value.

6: test this in practice with a test-retest to see if the same respondents fails/passes both test. This is to test how precise the cut-off scores.

Words: 187

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

A scale for satisfaction with life should have multiple items within them, and a question of the gender difference is that if the gender difference are on the scale level or item level.

To have a valid measurement it has to have a high reliability, which means it has to fit to the chosen model. Typically it will be a single factor model.
First an analysis of the items and how well the model fits to the data. This will be done in group of both genders.
The data in the analysis should come from a representative sample of the population, since the measurement should be valid in every group of the population. Especially the gender groups have to be balanced, such that any difference only occurs because the group differs in the factor.

To investigate a gender difference two approaches will be used. The first one is to see if the model are same in both groups. If yes then the scale can be used in both group. If not, then there is a different affect in both models. The difference could be caused by item bias, such that it is only one item that affects differently. To examining item bias, the second approach will be used, which is a common factor analysis run, and comparisons in expected item scores between gender groups, whom have the same factor score. If there is a difference there, it will be problematic to use the same items across genders, and therefore not a valid measurement.

If there is no item bias, and only a difference in level of the factors between genders, then a validation with other scales will be performed. This will be to see if there is concurrent validation (high correlation with scales that measure the same) and discrimination validation (no correlation with scales that measure something different).

Words: 308

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: The evidence of validity is here the internal reliability and the fit of the model to the data. If the sample is the targeted and representative for the targeted population, then will a single factor model fits well. And therefore be a part of the validation process.

2: When looking at the residual correlation matrix, none of the correlations are too high. A correlation higher than of absolute (0.1) will indicate if something is wrong with the single factor model. The highest value is 0.072, which is lower than 0.100.

3: When looking at the contribution to the scale, we will look at the relative contribution of each item, measure as: $\frac{\lambda_i^2}{var(\epsilon_i)}$.

item	relative contribution
1	$\frac{2^2}{4} = \frac{4}{4} = 1$
2	$\frac{3^2}{2} = \frac{9}{2} = 4.5$
3	$\frac{1^2}{4} = \frac{1}{4} = 0.25$
4	$\frac{2^2}{5} = \frac{4}{5} = 0.8$
5	$\frac{2^2}{1} = \frac{4}{1} = 4$

Item 2 is giving most to the reliability since the factor loading are much higher than the error term of that item.

item 3 is giving least to the reliability since the factor loading are much lower than the error term of that item.

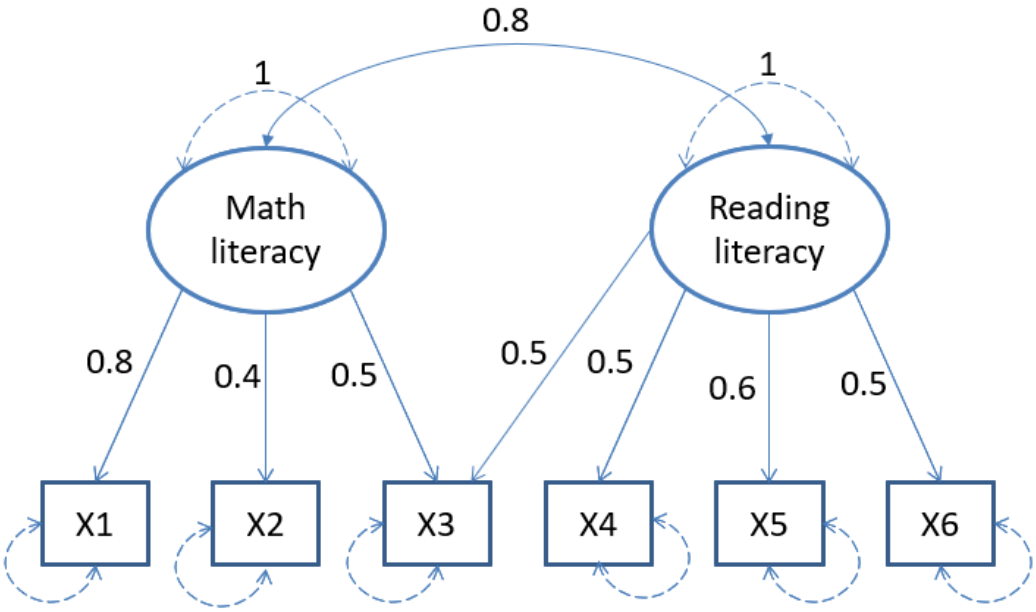
4: One reservation will be the high factor loading of each item. Especially compare to the error terms of the items. It could indicate that there was some problems with the single factor model, and it could be due to other factors affecting the items. Perhaps with a bifactor model would be better.

Words: 226

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: For item 3 the factor model will be:

$$X_3 = \mu_3 + \lambda_{3M} Math + \lambda_{3R} Reading + \epsilon_3 \Longrightarrow X_3 = \mu_3 + 0.5 \cdot Math + 0.5 \cdot Reading + \epsilon_3$$

.

For item 3 there is two factors affecting it: The scale for Math literacy and the scale for Reading literacy. The random variable of the observed score in item 3, X_3, are affected by a item mean, μ_3, which we do not know (and let us assume it is 0), and the factors of math and reading with a factor loading parameter of each 0.5. The higher the factor a respondent have on either factors the higher will the expected score be in item 3 will be. At last there is the error term, which can affect the observed score in both direction.

2: The covariance between X_3 and X_4, is calculated as such:

$$V(X_3 + X_4) = V(X_3) + V(X_4) + 2cov(X_3, X_4)$$

Use the variance of the sum of $X_3 + X_4$ to get the covariance of $cov(X_3, X_4)$.

We ignore the error terms since they will become zero later on, since we are taking expectations of them, where they become 0.

$$\Longrightarrow V(\lambda_{3M} M + \lambda_{3R} R + \lambda_{4R} R) = V(\lambda_{3M} M + \lambda_{3R} R) + V(\lambda_{4R} R) + 2cov(X_3, X_4)$$

From the left hand side: $V(X_3 + X_4)$:

$$V(\lambda_{3M} M + \lambda_{3R} R + \lambda_{4R} R) = E\Big((\lambda_{3M} M + \lambda_{3R} R + \lambda_{4R} R)^2\Big)$$

$$= E(\lambda_{3M}^2 M^2 + \lambda_{3R}^2 R^2 + \lambda_{4R}^2 R^2 + 2\lambda_{3M}\lambda_{3R} MR + 2\lambda_{3M}\lambda_{4R} MR + 2\lambda_{3R}\lambda_{4R} R \cdot R) =$$

$$\lambda_{3M}^2 V(M) + \lambda_{3R}^2 V(R) + \lambda_{4R}^2 V(R)$$

$$+2\lambda_{3M}\lambda_{3R} cov(M, R) + 2\lambda_{3M}\lambda_{4R} cov(M, R) + 2\lambda_{3R}\lambda_{4R} cov(R, R)$$

From the right hand side:
$V(X_3) = \lambda_{3M}^2 \cdot Var(M) + \lambda_{3R}^2 Var(R) + 2 \cdot \lambda_{3M}\lambda_{3R} cov(M,R)$ and
$V(X_4) = \lambda_{4R}^2 Var(R)$
Combining the found numbers to isolate $cov(X_3, X_4)$
$\implies V(X_3 + X_4) - V(X_3) - V(X_4) = 2cov(X_3, X_4)$
$\implies 2\lambda_{3M}\lambda_{4R} cov(M,R) + 2\lambda_{3R}\lambda_{4R} cov(R,R) = 2cov(X_3, X_4)$,
and use the fact that $cov(R,R) = V(R)$:
$\Leftrightarrow \lambda_{3M}\lambda_{4R} cov(M,R) + \lambda_{3R}\lambda_{4R} V(R) = cov(X_3, X_4)$
Plugging in the number from the model:
$cov(X_3, X_4) = 0.5 \cdot 0.5 \cdot 0.8 + 0.5 \cdot 0.5 \cdot 1 = 0,25 \cdot 0.8 + 0.25 \cdot 1 = 0.2 + 0.25 = 0.45$
The covariance between the two items are 0.45.

Words: 236

Answered.