

i Instructions

Dear student,

Some guidelines for the MAE4011 exam:

- The exam consists of a combination of multiple choice questions and open-response questions.
- It is totally fine to respond in bullet points, as long as no important information gets lost.
- For the computational tasks, it is sufficient to show how you *would* obtain the respective coefficients. You will find formulas in the provided external resource in the left corner.
- Read the questions carefully and write down if you do not understand something about a question. I will visit the exam site for during the first two hours of the exam to answer clarifying questions

Good luck!

Esther

1 Classical Test Theory II

Which out of the following is assumed to hold for congeneric multi-item measures? (1 point)

Select one alternative:

- ☐ The error scores of the items are uncorrelated.
- ☐ All items have the same error score variance.
- ☐ All items have identical item difficulties.
- ☐ All items have equal true score variance.

Maximum marks: 1

2 Classical Test Theory I

We denote with X the observed score of a test, with T its true score and with E its error score. Which of the following does **not** follow from the axioms of classical test theory? (1 point)

Select one alternative:

☐ $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$

☐ $\rho_{XE} = 0$

☐ $\sigma_{XT} = \sigma_T^2$

☐ $T = E(X)$

Maximum marks: 1

3 Classical Test Theory III

X and X' are scores on two parallel tests. If $\rho_{XT_X} = 0.70$ and $\rho_{XY} = 0.30$, what are $\rho_{X'T_{X'}}$, $\rho_{X'Y}$, and $\rho_{X'E_{X'}}$? (2 points)

Fill in your answer here

Maximum marks: 2

4 Operationalization

A psychological researcher is interested in studying empathy. For measuring empathy, she is deciding between two measures: one self-report measure and one physiological measure. As the self-report measure, the researcher considers the Interpersonal Reactivity Index (IRI) questionnaire, where individuals rate their empathy based on statements like, "I often have tender, concerned feelings for people less fortunate than me," using a Likert scale. As a physiological measure, the researcher considers measuring heart rate variability (HRV) while participants view emotional stimuli (e.g., videos of people in distress). HRV tends to decrease when individuals engage in empathy or feel emotionally connected. Name and briefly explain one advantage and one disadvantage of each empathy measure. (4 points)

Fill in your answer here

Format
|
B
I
U
x₂
x²
|
I_x
|
📄
📋
|
↶
↷
↺
|
1=
2=
3=
4=
|

≡
≡
≡
≡
|
Ω
📊
|
✎
|
Σ
|
✖

Words: 0

Maximum marks: 4

5 Variance-Covariance

We have the following variance-covariance matrix between three items X_1 , X_2 , and X_3 forming the scale $Y = X_1 + X_2 + X_3$ and an external criterion Z .

1. Obtain the sample variance of the scale score Y
2. Obtain the sample covariance between the scale score Y and the external criterion Z
3. What would be the sample covariance between the scale score Y and the external criterion Z if we would downweigh the contribution of X_3 to the scale score, i.e., define $Y = X_1 + X_2 + 0.5X_3$?

(It is sufficient to write down how you would the respective coefficients. You do not have to provide the exact numbers.) (3 points)

	X_1	X_2	X_3	Z
X_1	1.00	0.70	0.80	0.50
X_2	0.70	1.00	0.60	0.40
X_3	0.80	0.60	1.00	0.50
Z	0.50	0.40	0.50	1.00

Fill in your answer here

Maximum marks: 3

6 Test Scores

A test developer is developing a new theoretical exam for a driver's license. He is debating whether to use norm-based or reference-based scoring for test score interpretation. Provide an example interpretation of both norm-based and reference-based scores in this context. Which scoring method would be more appropriate for this exam, and why? (3 points)

Fill in your answer here

Format
|
B
I
U
 x_2
 x^2
 I_x
|

|

|
 $\frac{1}{2}$
 $\frac{3}{4}$
 $\frac{1}{2}$
 $\frac{3}{4}$
|

|
 Ω

|

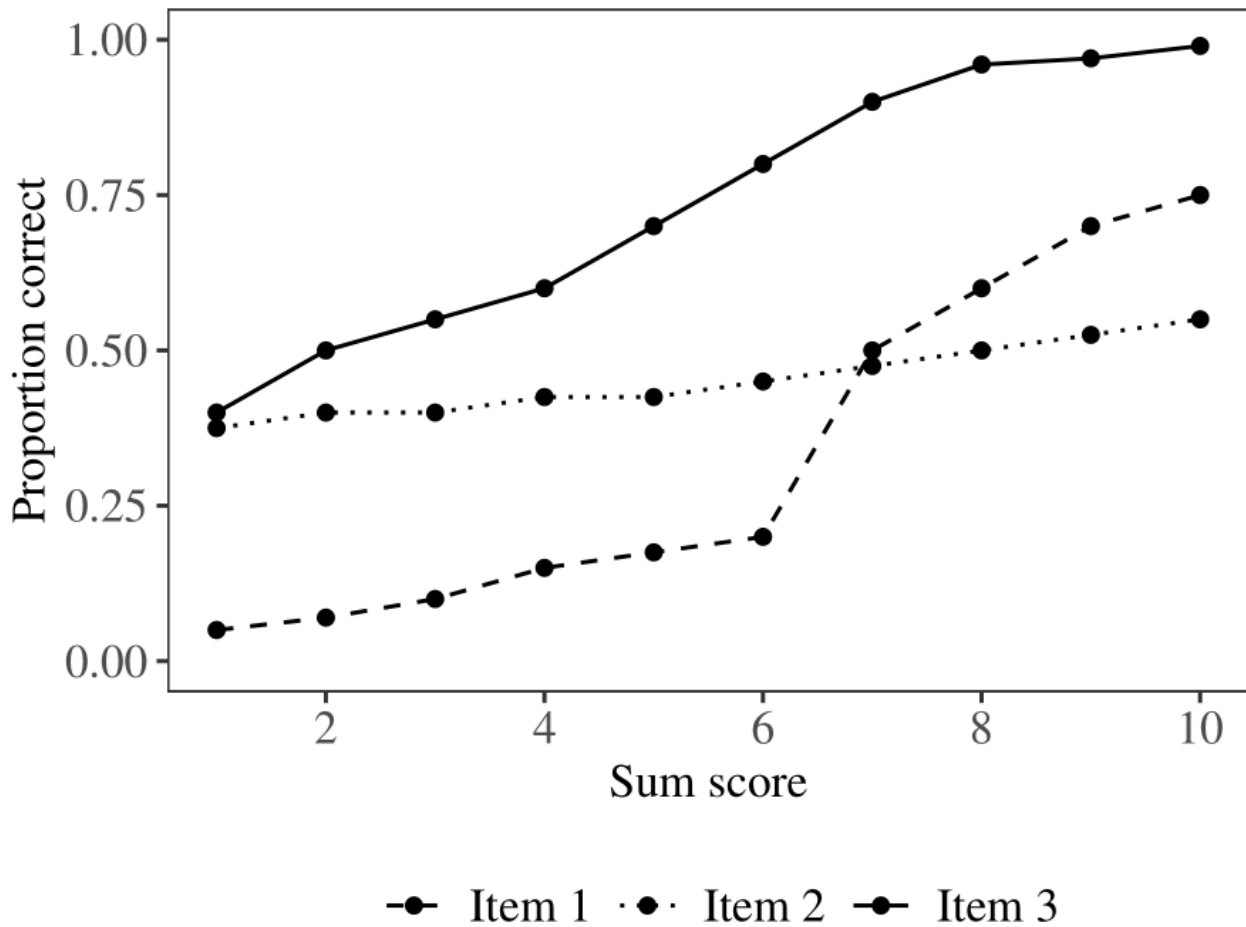
|
 Σ
|

Words: 0

Maximum marks: 3

7 Item Statistics

The figure displays empirical item characteristic curves for three items. Which of the items would you (a) remove from the assessment and (b) use if you want to construct a test differentiating between examinees with high ability? (2 points)



Fill in your answer here

Maximum marks: 2

8 Quality Criteria

In a study aimed to gather validity evidence for a law school admissions test, researchers examine the correlation between test scores and first-year grades of admitted students. They observe that the correlation is strong for students from private high schools but weak for students from public high schools.

- What is this type of phenomenon referred to as?
- Which quality criterion for educational and psychological tests is being violated in this scenario? (2 points)

Fill in your answer here

Maximum marks: 2

9 Reliability I

You have a 60-item test with an estimated reliability of .85. Assuming that the items are parallel, how would you estimate the reliability of a 30-item version of this test? It is sufficient to write down how you would compute reliability. You do not have to provide the exact number. (1 point)

Fill in your answer here

Maximum marks: 1

10 Reliability II

You have a 20-item test with an estimated reliability of .10. If you expanded it to an 80-item test by adding parallel items, what reliability would you expect? It is sufficient to write down how you would compute reliability. You do not have to provide the exact number. (1 point)

Fill in your answer here

Maximum marks: 1

11 Reliability III

Which of the following is **not** an assumption of Cronbach's α ? (1 point)

Select one alternative:

- ☐ Errors of measurement are uncorrelated between items.
- ☐ The items have equal error variances.
- ☐ All items measure a single, unidimensional construct.
- ☐ The items have equal correlations with the true score of the construct.

Maximum marks: 1

12 Reliability IV

A researcher is developing a new test to measure passive vocabulary knowledge of Norwegian as a foreign language. She administers the test twice to recent migrants to Norway, the first time two weeks after arrival and the second time six months later. Using the test scores from the two measurement occasions, the researcher estimates the test-retest correlation coefficient. Explain under which conditions this approach is appropriate for estimating the reliability of the sum scores and under which conditions another approach for estimating the reliability of the sum scores should be chosen. (2 points)

Fill in your answer here

Format
|
B
I
U
 x_2
 x^2
 I_x
|

|

|

|

|

|

|

|

Words: 0

Maximum marks: 2

13 Interrater Reliability

A student mental health center wants to evaluate their diagnostic procedure for screening for testing anxiety. As part of the evaluation process, the center's researchers want to evaluate agreement in diagnoses given by their two psychologists for 50 students. Obtain Cohen's κ from the information provided in the table. It is sufficient to write down how you would compute Cohen's κ . You do not have to provide the exact number. (2 points)

		Psychologist 2		
		No anxiety	Anxiety	Total 1
Psychologist 1	No anxiety	17	8	25
	Anxiety	6	19	25
Total 2		23	27	50

Fill in your answer here

Maximum marks: 2

14 Validity I

Which of the following statements is most in line with the validity theory offered by the 2014 Standards for Educational and Psychological Testing? (1 point)

Select one alternative:

- ☐ Content validity is always the first evidence category to consider in validation.
- ☐ Once a test has been validated, it may be used in any number of contexts.
- ☐ Intended use of test scores dictates what validity evidence is needed.
- ☐ All evidence categories are of equal importance, regardless of intended test score use.

Maximum marks: 1

15 Validity II

A researcher designs a behavioral assessment tool to measure impulsivity in adolescents, where adolescents' reactions to provocations are observed in a role-play scenario with a social worker. To gather validity evidence, the researcher uses the test scores of delinquent, imprisoned adolescents to predict whether or not they re-commit the crime they were imprisoned for. However, their study reveals a surprisingly low validity coefficient. Discuss and name examples for how three factors of your choosing out of the following four factors might have contributed to this finding in this specific context: (6 points)

- Construct underrepresentation
- Construct irrelevant variability
- Restricted range
- Unsuitable and/or too distal validity criterion

Fill in your answer here

Format
|
B
I
U
 x_2
 x^2
 $\frac{I_x}{x}$
|

|

|

|

|

|

|

|

|

|

|

Words: 0

Maximum marks: 6

- Describe three evidence sources you want to consider in order to evaluate the validity of the scale scores for their intended purpose.
- Give a brief overview of how you would design a study to evaluate the validity of the scale score use. The answer should include a brief description of the data you would like to analyze and the analysis methods you would suggest using.
- For each source of validity evidence, describe the specific empirical results you would consider as evidence supporting the validity of using the scale scores for their intended purpose.

[illegible]

Words: 0

Maximum marks: 6

17 MTMM

In the context of the Multitrait-Multimethod (MTMM) matrix, which of the following provides strong discriminant validity evidence? (1 point)

Select one alternative:

- ☐ Low correlations between different traits measured using the same method.
- ☐ Low correlations between the same trait measured using different methods.
- ☐ High correlations between the same trait measured using different methods.
- ☐ High correlations between different traits measured using different methods.

Maximum marks: 1

18 Confirmatory Factor Analysis I

A researcher wants to evaluate the properties of a newly developed three item measure, with items Q4, Q9, and Q14. The figure displays the parameter estimates obtained from fitting a unidimensional confirmatory factor analysis model. Use the output to obtain an estimate of McDonald's ω . It is sufficient to write down how you would compute McDonald's ω . You do not have to provide the exact number. (2 points)

	lhs	op	rhs	est	se	z	pvalue	ci.lower	ci.upper	std.lv	std.all
1	F	=~	Q4	1.000	0.000	NA	NA	1.000	1.000	1.138	0.784
2	F	=~	Q9	0.898	0.029	30.855	0	0.841	0.955	1.022	0.720
3	F	=~	Q14	1.008	0.032	31.400	0	0.945	1.071	1.147	0.770
4	Q4	~~	Q4	0.811	0.040	20.060	0	0.732	0.890	0.811	0.385
5	Q9	~~	Q9	0.969	0.038	25.246	0	0.894	1.044	0.969	0.481
6	Q14	~~	Q14	0.902	0.042	21.304	0	0.819	0.985	0.902	0.407
7	F	~~	F	1.295	0.064	20.137	0	1.169	1.421	1.000	1.000
8	Q4	~1		2.636	0.029	90.741	0	2.579	2.693	2.636	1.817
9	Q9	~1		2.233	0.028	78.600	0	2.177	2.289	2.233	1.574
10	Q14	~1		2.956	0.030	99.164	0	2.897	3.014	2.956	1.985
11	F	~1		0.000	0.000	NA	NA	0.000	0.000	0.000	0.000

Fill in your answer here

Maximum marks: 2

19 Confirmatory Factor Analysis II

A researcher wants to evaluate the properties of a newly developed three item measure. She compared two unidimensional confirmatory factor analysis models, one assuming an essentially τ equivalent model (fit.etau), and one assuming a τ equivalent model (fit.tau). The figure displays the results of a comparison between the two models. Which model provides the better fit to the data? Briefly explain your reasoning. (1 point)

Chi-Squared Difference Test

```

          Df   AIC   BIC   Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
fit.etau   2 24494 24535   16.096
fit.tau    4 25129 25158  655.239      639.14 0.35733      2 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Fill in your answer here

Maximum marks: 1

20 Confirmatory Factor Analysis III

Which of the following assumptions and constraints are made in a τ -equivalent confirmatory factor analysis model?

Select one or more alternatives:

- ☐ All item loadings are constrained to be equal
- ☐ All item error variances are constrained to be equal
- ☐ The model is assumed to be unidimensional
- ☐ All item intercepts are constrained to be equal

Maximum marks: 3