Advanced Data Analysis for Psychological Science

Exam simulation

Luca Menghini

2023-12-14

Content

This is a simulation of the exam for the course entitled "*Advanced Data Analysis for Psychological Science*" (University of Padova, Academic year 2023-24). As clarified in the `1-intro.pdf` slides and the `exam-information.pdf` file on Moodle, both the final exam and the current simulation are written and consist of **31 multi-choice questions** to be answered in **40 minutes**.

Questions

1. **When can we talk about 'nested data structures'?**

   a) When multiple individuals are nested within clusters
   b) When multiple clusters are nested within individual
   c) When multiple clusters and individuals are nested within each other
   d) When multiple individuals are nested within statistical units

2. **Which linear model (LM) assumption is violated with clustered data?**

   a) Linearity
   b) Normality
   c) Independence of observations
   d) Independence between errors and predictors

3. **Which of the following is NOT a nested data structure?**

   a) School principals within schools
   b) Students within schools
   c) Classes within schools
   d) Teachers within schools

edit: this question was debated in class because it assumes that a school can only have one school principal. If we consider that schools can have multiple school principals, then all the options would be right. In the actual exam, you will not find such ambiguous questions.

4. **You are hired by a school principal to assess whether teachers' self-efficacy (x) impact on students' achievement (y). Which is the cluster variable?**

   a) Teacher self-efficacy
   b) Students' classes
   c) Student self-efficacy
   d) Teacher's classes

5. **Which of the following is NOT a potential consequence of using a statistical model that neglects the local dependencies in the data?**

   a) Biased random effects
   b) Biased standard errors
   c) Biased p-value
   d) Biased confidence intervals

6. **Which of the following parameters is a fixed effect?**

   a) Intercept
   b) Residual variance
   c) Cluster variability around the intercept
   d) None of the other options

7. **You want to investigate the relationship between child attachment (independent variable `x1`) and health (dependent variable `y`) in families with 3 or more children, controlling for child's age (`x2`) and the number of children in each family (`x3`). Which of the following is NOT a random effect?**

   a) Differences between families in the mean health of their children
   b) Differences between families in the relationships between attachment and health
   c) Mean relationship between attachment and health
   d) None of the other options

8. **Considering the research described in question #7, which of the following variables is at the cluster level (level 2)?**

   a) Attachment
   b) Health
   c) Number of children
   d) Age

9. **Considering the research described in question #7, which of the following variables should be included in a model to estimate the random intercept?**

   a) Attachment and family
   b) Health and family

c) Attachment and child

d) Health and child

10. **Considering the research described in question #7, you got a level-2 correlation between child attachment and health of 0.30 and a level-1 correlation of -0.10. What does that mean?**

    a) A negative correlation is estimated between family means of attachment and health

    b) A positive correlation is estimated between attachment and health within each family

    c) Neither (a) nor (b) are true

    d) Both (a) and (b) are true

11. **How can we compute the correlation between two variables at the within-cluster level?**

    a) By correlating the two variables expressed in their original scale

    b) By correlating the cluster means of the two variables

    c) By correlating the cluster-mean-centered values of the two variables

    d) By correlating the grand-mean-centered values of the two variables

12. **Which dataset should you use to compute between-cluster correlations?**

    a) The long-form dataset

    b) The wide-form dataset

    c) Both the long- and the wide-form dataset can be used

    d) Neither the long- nor the wide-form dataset can be used

13. **The variable `x` shows an intraclass correlation coefficient (ICC) of 0.30. What does that mean?**

    a) `x` varies more within than between clusters

    b) `x` varies more between than within clusters

    c) `x` equally varies between and within clusters

    d) `x` only varies within clusters

14. **For a given variable, both the residual variance and the variance of the random intercept are equal to 20. How much is the ICC?**

    a) 0.20

    b) 20

    c) 0.50

    d) 1

The following questions are related to a study where a group of healthy **adolescents** (`subjID`) were involved in a 10-day ambulatory protocol where they reported they daily levels of **study load** (`SL`, 7-point Likert scale), day-time **anxiety** (`Anx`, 7-point Likert scale), evening **relaxation** (`Relax`, 7-point Likert scale), and evening **blood pressure** (`BP`, measured in mmHg). The following results were obtained from two alternative models predicting `BP` by the other variables while controlling for the adolescent **smoking status** (`Smoker`, Yes/No):

Predictors	*b (SE)*	*CI*	*p*	*b (SE)*	*CI*	*p*
(Intercept)	115.12 (2.32)	110.56 – 119.67	**<0.001**	117.58 (3.09)	111.52 – 123.64	**<0.001**
SL	1.25 (0.31)	0.64 – 1.86	**<0.001**	0.70 (0.55)	-0.39 – 1.78	0.208
Anx	0.42 (0.33)	-0.22 – 1.07	0.199	-0.47 (0.80)	-2.04 – 1.11	0.563
Smoker [Yes]	-0.80 (3.43)	-7.53 – 5.93	0.816	-0.80 (3.43)	-7.54 – 5.93	0.815
SL × Anx				0.19 (0.16)	-0.12 – 0.49	0.226
Random Effects						
σ^2	103.00			102.88		
τ_{00}	251.20 $_{subjID}$			251.89 $_{subjID}$		
Observations	824			824		
Marginal R^2 / Conditional R^2	0.015 / 0.714			0.016 / 0.715		
AIC	6517.045			6519.460		

15. **Which are the random effects?**

    a) Random intercept and smoking status
    b) Random intercept and random slope
    c) Random intercept and residual variance
    d) Random intercept and fixed intercept

16. **How many parameters are included in the first model (shown on the left)?**

    a) Four: 1 intercept + 3 slopes (SL, Anx, Smoker)
    b) Five: 1 intercept + 3 slopes (SL, Anx, Smoker) + 1 residual variance
    c) Six: 1 intercept + 3 slopes (SL, Anx, Smoker) + 1 residual variance + 1 random intercept
    d) Seven: 1 intercept + 3 slopes (SL, Anx, Smoker) + 1 residual variance + 1 random intercept + 1 interaction between SL and Anx

17. **Which is the best model according to the Akaike information criterion (AIC)**?

    a) The first model (on the left)
    b) The second model (on the right)
    c) The two models are equivalent
    d) The AIC does not allow to compare two models

18. **What is the interpretation of the `Smoker` coefficient in the first model?**

    a) For a one-unit increase in the smoking status, blood pressure is predicted to *decrease* by 0.80 mmHg
    b) For a one-unit increase in the smoking status, blood pressure is predicted to *increase* by 0.80 mmHg
    c) On average, smokers are predicted to show a blood pressure 0.80 mmHg *lower* than non-smokers
    d) On average, smokers are predicted to show a blood pressure 0.80 mmHg *higher* than non-smokers

19. **What is the interpretation of the `SL` × `Anx` coefficient in the second model?**

    a) For a one-unit increase in *study load*, the relationship between *anxiety* and blood pressure is predicted to increase by 0.19
    b) For a one-unit increase in *anxiety*, the relationship between *study load* and blood pressure is predicted to increase by 0.19
    c) Both (a) and (b) are true
    d) Neither (a) nor (b) are true

20. **What percentage of blood pressure variance is explained by the two models, considering both fixed and random effects?**

    a) 1.5% by the first model, 71.4% by the second model
    b) 1.5% by the first model, 1.6% by the second model
    c) 71.4% by the first model, 71.5% by the second model
    d) 1.6% by the first model, 71.5% by the second model

21. **Which of the following statements on structural equation models (SEM) is FALSE?**

    a) A SEM is a multivariate model formalized by a system of equations
    b) A SEM can include both observed and latent variables
    c) A SEM cannot include only one 'dependent' variable
    d) A SEM can include multiple 'dependent' variables

22. **Which of the following statements on latent variables (LV) is TRUE?**

    a) A latent variable is always estimated with an error component
    b) A latent variable cannot have an error component
    c) A latent variable is always estimated by using one or more observed variables

d) A latent variable cannot be estimated by using one or more observed variables

23. **Which of the following statements on endogenous variables is TRUE?**

a) An endogenous variable is always estimated with an error component
b) An endogenous variable cannot have an error component
c) An endogenous variable cannot be latent
d) An endogenous variable is not directly 'caused' from inside the model

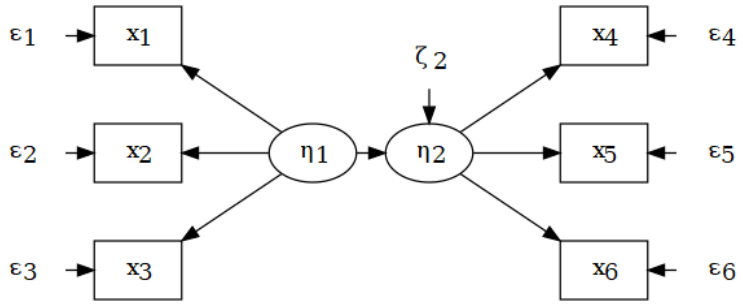
24. **Which of the following statements on confirmatory factor analysis (CFA) is FALSE?**

a) A CFA model is a SEM that only includes the measurement part of the model
b) A CFA model should have at least one latent variable
c) A CFA model should have at least one observed variable
d) A CFA model is a SEM that only inlcudes the structural part of the model

25. **Which of the following statements on the observed variance-covariance matrix (S) is FALSE?**

a) S is the starting point of path analysis, but it is not the starting point of confirmatory factor analysis
b) S is a symmetrical matrix (i.e., same covariance values above and below the main diagonal)
c) The main diagonal of S includes the observed variances
d) When the variables are standardized, S is a correlation matrix

The following questions are related to the model represented below:



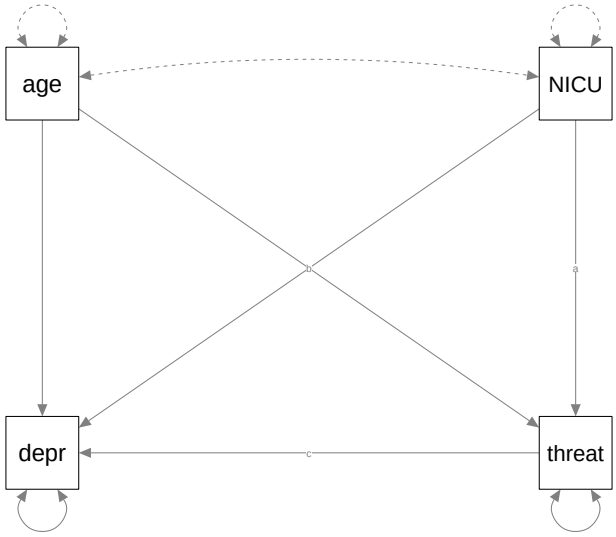
26. **How many endogenous variables are included in the model?**

a) 2
b) 4
c) 6
d) 7

27. **How many parameters are estimated by the model?**

a) 6
b) 7
c) 14
d) 15

The following questions are related to a model of the "*Pregnancy during pandemics*" data where young mother postnatal **depression** (`depr`, measured on a Likert scale) is predicted by the mother **fear of being infected** by COVID-19 (`threat`, measured on a Likert scale), the need to go into **intensive care unit** (`NICU`, Yes/No), and mothers' **age** (`age`, measured in years). Moreover, the model is used to test whether `threat` mediates the relationship between `NICU` and `depr`. The path diagram and the summary table of the models are reported below:



	lhs	op	rhs	est	se	z	pvalue
1	depr	~	threat	0.06	0.00	22.40	0.00
2	depr	~	age	-0.08	0.02	-4.87	0.00
3	depr	~	NICU	1.01	0.24	4.25	0.00
4	threat	~	NICU	4.38	1.18	3.73	0.00
5	threat	~	age	0.11	0.08	1.35	0.18
6	depr	~~	depr	25.44	0.50	50.86	0.00
7	threat	~~	threat	625.31	12.30	50.86	0.00
11	indirect.effect	:=	a*b	4.42	1.58	2.80	0.01
12	total.effect	:=	c+(a*b)	4.48	1.58	2.84	0.00

28. **How many exogenous variables are included in the model?**

    a) 1
    b) 2
    c) 3
    d) 4

29. **How can we interpret the parameter estimate reported in the second row of the table (i.e., `depr ~ age`)?**

    a) Depression and age are predicted to negatively covary by -0.08
    b) For a one-unit increase in depression, age is predicted to decrease by 0.08 years
    c) For a one-year increase in age, depression is predicted to decrease by 0.08 units

d) Age is predicted to mediate the relationship between fear of being infected and depression

30. **What is indexed by the parameter estimate reported in the sixth row of the table (i.e., `depr ~~ depr`)**

    a) The estimated variance of depression
    b) The estimated residual variance of depression
    c) The estimated covariance between depression and the other variables
    d) The estimated direct effect of `threat` on depression

31. **How can we interpret the results on the mediation between `NICU` and `depr`?**

    a) `threat` fully mediates the relationship between `NICU` and `depr`
    b) `NICU` is not directly related to `depr`
    c) `threat` is not directly related to `depr`
    d) `threat` partially mediates the relationship between `NICU` and `depr`