

Mock exam "Multivariate Statistics", fall semester 2023

Linear regression

Multicollinearity (Kprim) (Kprim) - 2

Which of the following statements about multicollinearity are correct, and which are incorrect?

Decide whether the following statements are true or false by clicking the respective box.

	True	False
In the presence of multicollinearity, there is an increase in standard errors of the estimates.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Multicollinearity occurs when researchers include many highly correlated variables.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Collinearity statistics include the Tolerance and the VIF.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Theory-driven and intelligent variable selection can help to avoid multicollinearity.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Linear regression (open-ended) (Essay) - 1

What is the added-value of adjusted R-squared compared to single R-squared?

Tips: think about the number of predictors and the number of cases.

A corrected R-squared takes shrinkage into account by reducing the simple R-squared by a correction value that is larger: the larger the number of predictors and the smaller the number of cases (whereas the simple R-squared increases with each added predictor, the corrected R-squared decreases as more predictors are added).

Logistic regression

Logistic regression (Kprim) (Kprim) - 2

Which of the following statements about logistic regression are correct, and which are incorrect?

Decide whether the following statements are true or false by clicking the respective box.

	True	False
Logistic regression is used to make predictions about a dichotomous dependent variable.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Odds can be defined as the number of times something occurs relative to the number of times it does not occur.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
If the odds ratio of a dummy variable is greater than 1, then the group captured in the dummy variable is predicted to be more likely than the reference group to have something occur.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
When there is exactly a 0.5 probability of something occurring, the log odds are 1.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Logistic regression (open-ended) (Essay) - 2

Linear regression is NOT appropriate when the dependent variable is a dichotomous (or binary) variable. Explain why this is the case by referring to two properties of the logistic regression.

Tips: think about properties such as normal distribution, variance of the residual errors, and predicted values.

Answer: The linear regression model assumes that the outcome Y is continuous, with errors which are normally distributed. If the outcome variable is binary this assumption is clearly violated, and so in general we might expect our inferences to be invalid (or 1 point). With binary data the variance is a function of the mean. This violates one of the standard linear regression assumptions that the variance of the residual errors is constant (or 1 point). For a binary outcome the mean is the probability of a 1, or success. If we use linear regression to model a binary outcome it is entirely possible to have a fitted regression which gives predicted values for some individuals which are outside of the $(0,1)$ range or probabilities (or 1 point).

Moderation/Mediation

Moderation (single choice) (Choice) - 1

Which of the following statements about the interpretation of the moderation effect is correct?

Recall that the regression equation for moderation is: $Y = a + b_1 * X + b_2 * Z + b_3 * XZ + e = a + b_2 * Z + (b_1 + b_3 Z) * X + e$

Decide which of the following statements is correct by clicking on a single box.

The coefficient b_1 can be the main effect of X from a 2X2 ANOVA if X and Z are coded as numeric variables.

The coefficient b_1 is the conditional effect of Z on Y when $X = 0$.

When $X = 0$, the conditional effect of Z reduces to b_3 .

The coefficient b_1 is the estimated difference in Y between two cases in the data that differ by one unit in X but have a value of 0 for Z .

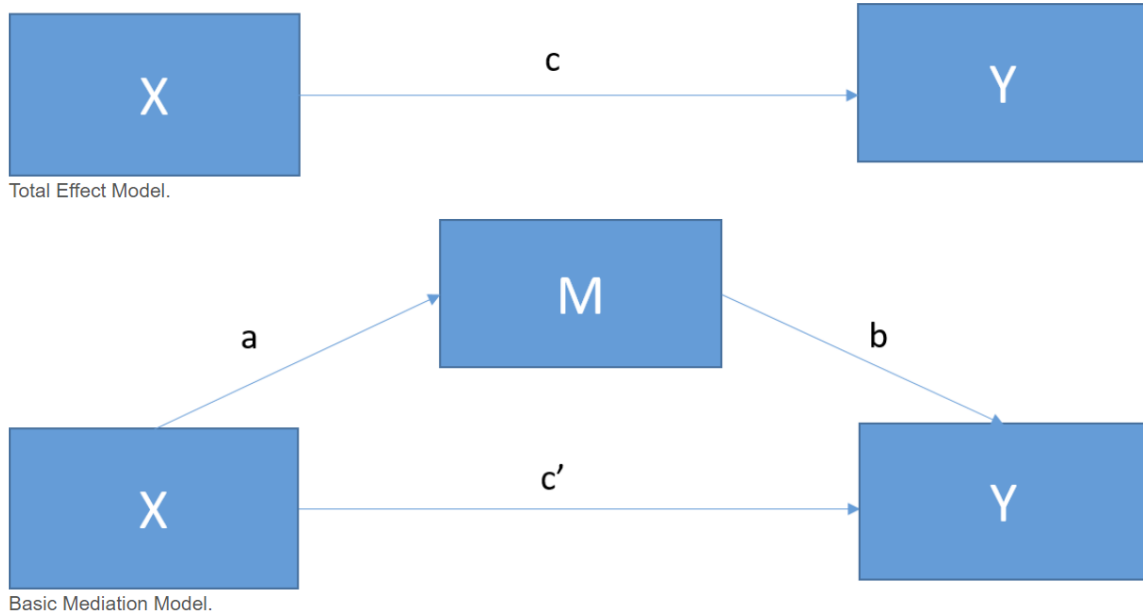
Moderation/Mediation (open-ended) (Essay) - 2

What is conceptionally the difference between moderation analysis and mediation analysis?

Answer: The moderation examines how the relationship between the independent variable and the dependent variable changes for different levels of a moderator (1 point). In other words, the moderation examines the conditions under which there is a relationship between the independent and dependent variable (or 1 point). The Mediation however investigates the process by which a relationship between independent and dependent variable can be explained (1 point).

Mediation (single choice) (Choice) - 1

Which of the following statements about mediation analysis is NOT correct?



Decide which of the following statements is NOT correct by clicking on a single box.

- Mediation analysis should only be performed when this total effect of X, path c, is statistically different from zero
- In mediation analysis, the size of ab is not determined by c or c'.
- The Sobel test has been criticized because it assumes the sampling distribution of ab is normal in form
- In mediation analysis, ab could be large even though c is small

ANOVA

Repeated Measurement ANOVA (single choice) (Choice) - 1

Which of the following about the partitioning of the error term in repeated measures (RM) ANOVA is correct?

Decide which of the following statements is correct by clicking on a single box.

RM ANOVA does not further partition the error term from ANOVA.

The partitioning of the within-group variability in RM ANOVA decreases the value of the F-statistic.

The partitioning of the within-group variability in RM ANOVA reduces the size of the error term compared to ANOVA.

The partitioning of the within-group variability in RM ANOVA decreases the power of the test to detect significant differences between means.

ANOVA (open-ended) (Essay) - 3

Mention and describe the three components of the variance in ANOVA.

Answer: ANOVA estimates 3 sample variances. To find the total amount of variation within our data we calculate the difference between each observed data point and the grand mean. We then square these differences and add them together to give us the total sum of squares (1 point). The model sum of squares requires us to calculate the differences between each participant's predicted value and the grand mean: it is the sum of the squared distances between what the model predicts for each data point and the overall mean of the data (1 point). The final sum of squares is the residual sum of squares, which tells us how much of the variation cannot be explained by the model. The simplest way to calculate SSE is to subtract MSS from TSS, and it is calculated by looking at the difference between the score obtained by a person and the mean of the group to which the person belongs (1 point).

ANOVA (Kprim) (Kprim) - 2

When are post hoc tests useful in one-way ANOVA?

Tips: recall that one-way ANOVA is when there is one independent variable and one dependent variable.

Decide whether the following statements are true or false by clicking the respective box.

	True	False
Post hoc tests are useful when a factor has more than two levels.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Post hoc tests are useful when differences between individual factor levels from two different factors are to be tested.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Post hoc tests are useful when differences between individual factor levels from the same factor are to be tested.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Post hoc tests are useful when the p-value associated with the F-test is statistically significant.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Factor analysis (EFA/CFA)

EFA (single choice) (Choice) - 1

Which of the following statement about exploratory factor analysis (EFA) is correct?
Decide which of the following statements is correct by clicking on a single box.

The KMO (Kaiser-Meyer-Olkin measure of sampling adequacy) describes the amount of variance of one item that is explained by all factors.

The Kaiser's criterion is a measure of whether the data is suitable for an exploratory factor analysis.

In EFA each variable cannot be fully explained by a linear combination of the factors.

Bartlett's test has a significant value when correlations between items are not large enough to be used in factor analysis.

CFA (open-ended) (Essay) - 1

Name two kinds of parameters in a CFA and define them.

Answer: Fixed parameter: Parameters that are assigned a fixed value (or 0.5 point);
Constrained (or restricted) parameter: Parameter that is equal to another parameter (or 0.5 point); Free parameter: Parameter that must be estimated with no previous assigned value (or 0.5 point).

CFA (single choice) (Choice) - 1

In confirmatory factor analysis (CFA), which of the following statement is NOT a mathematical requirement?

Decide which of the following statements is NOT correct by clicking on a single box.

Indicators should not correlate with one another.

There are several interval-scaled items, each of which is normally distributed.

Items that should theoretically load on a factor should not correlate empirically.

Sufficient directly measured items must be available in order to be able to test the assumed model structure made up of items and factors.

SEM

SEM (Kprim) (Kprim) - 2

Which of the following statements about structural equation modelling (SEM) are correct, and which are incorrect?

Decide whether the following statements are true or false by clicking the respective box.

	True	False
SEM is estimated so as to minimize the difference between the observed and estimated covariance matrices	<input checked="" type="checkbox"/>	<input type="checkbox"/>
SEM is estimated so as to minimize the sum of squares between constructs	<input type="checkbox"/>	<input checked="" type="checkbox"/>
It is recommended that each construct be measured by at least three observed variables	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Confirmatory factor analysis is a type of SEM	<input checked="" type="checkbox"/>	<input type="checkbox"/>

MLM

MLM (Kprim) (Kprim) - 2

Imagine you have the following two-level null model specified for a dependent variable Y_{ij} (i : students and j : schools):

- Level 1 (students): $Y_{ij} = b_{0j} + r_{ij}$
- Level 2 (schools): $b_{0j} = g_{00} + u_{0j}$

Which of the following statements are correct, and which are incorrect?

Decide whether the following statements are true or false by clicking the respective box.

	True	False
b_{0j} represents the group-mean of Y	<input checked="" type="checkbox"/>	<input type="checkbox"/>
r_{ij} represents level-2 residuals	<input type="checkbox"/>	<input checked="" type="checkbox"/>
g_{00} represents the grand mean of the variable Y	<input checked="" type="checkbox"/>	<input type="checkbox"/>
u_{0j} represents the group-specific deviation from the grand mean of Y	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Bonus

Bonus (open-ended) (Essay) - 2

Imagine that survey items investigating which factors are related to COVID-19 vaccine hesitancy for adults have been collected and that you are in charge of making a report. Which method(s) of analysis would you use and why?
