Guillaume Pagnier

April 4th, 2019

Multivariate Midterm

1. What simple but powerful form of data reduction and integration underlies any multiple

regression equation, the forming of principal component/factor scores, and contrasts in

ANOVA? (a) Name it, (b) describe what it does mathematically, (c) describe what it means

conceptually, and (d) briefly explain what role it plays in each of the three abovementioned

statistical applications (multiple regression, PCA, and contrasts).

1. Linear combinations.
2. Linear combinations represent summing up different variables to end up with an aggregated variable. This is done by assigning a weight (constant) to each variable and then combining it with other variables to create a ‘weighted combination’ for lack of a better term.
3. Conceptually, a linear combination allows for data reduction as you can take many different vectors or variables, assign an appropriate weight to each of them and essentially reduce those variables to a singular weighted combination.

2. Reliability and validity come in several forms. (a) Describe two types of reliability and (b)

provide examples (real or invented) of conditions/situations/applications for which you would

use each. (c) Then describe two types validity and (d) provide examples of situations or

applications in which you would use each. (Note that this material was covered only partially in

the lecture itself. Refer to the additional lecture handout pages, the readings, and use other

resources, if appropriate, to answer the question.)

1. Two types of reliability would be 1) Cronbach’s alpha (describing internal consistency) and 2) interrater agreement (describing external consistency).
2. I would use Cronbach’s alpha if I was designing a survey that gauged an individual’s extraversion using a battery of questionnaires. A high Cronbach’s alpha would indicate that my questionnaires were consistent with one another and I could trust that my battery of surveys is reliable. I would look at interrater agreement if I was scoring a video for facial expressions of fear. This is a relatively subjective analysis and it’s possible there are individual differences in my RAs in classifying what a fear expression looks like. In this case, if the scores my RAs give the same trial are similar, than I have high interrater agreement and the data would be reliable.
3. Two types of validity would be 1) content validity and 2) criterion validity.

3. Some people claim that exploratory data analysis (EDA) is like cheating; they argue that

looking at your data before running your significance tests biases your testing strategy and

therefore the interpretation of significance tests. Write a critical analysis of this claim, both (a)

discussing what may be correct or incorrect about it and (b) making a counterargument by

pointing to the strengths of EDA.

4. (a) Name the type of matrix that PCA’s L matrix is. (b) Specify what all the elements (entries)

of the matrix mean conceptually. (c) Explain (verbally or formally) how SPSS, or any other

statistical program, moves from Rxx to L.

5. Write a dialogue (feel free to be funny) between a fanatic proponent of PCA and a fanatic

proponent of factor analysis. Let each person state at least *two clearly distinct* arguments for the

supremacy of their preferred technique and finish with your own (fanatic or not) conclusion.