Acta Neuropsychiatrica

STATISTICALLY SPEAKING

Multivariate analysis of variance

It is not unusual for statistical methods to include a MANOVA – a multivariate analysis of variance. MANOVA has many sides to it and in this issue we briefly describe those most likely to be encountered. Usage varies somewhat by discipline and subject matter, but four usages – ordered according to a wild guess about how often they are used in the psychiatric and psychological literature of interest to readers of this journal – are as follows:

1 The analysis of repeated-measures data
2 The use of a MANOVA to control the type I error rate in subsequent non-multivariate analyses
3 The use of a MANOVA in its fullest sense to analyse data
4 The use of discriminant analysis (a part of the results from a MANOVA) to allocate subjects to groups

MANOVA

As with a univariate ANOVA, a MANOVA has two or more groups and we are interested in differences between those groups. In a study with four groups, for example, the groups might reflect the four combinations resulting from two different drugs at either low dosage or high. In most research, subjects in these groups will be measured on more than one outcome (or dependent variable), and here we might measure speed and accuracy when performing a task. Our analysis of the data will attempt to answer questions about how drug type and dosage level affect these outcomes.

An ANOVA will ask these questions about each of the outcomes *one by one*. A MANOVA will ask these questions about *combinations* of the outcomes. An ANOVA might tell us that the two drugs differ on speed but not accuracy. A MANOVA might tell us that the two drugs differ on the combination made up of $0.5 \times \text{speed} - 1.6 \times \text{accuracy}$, where MANOVA has found that combination to be the one on which the two drugs most differ. It might also tell us that the dosage levels differ on $0.3 \times \text{speed} + 0.75 \times \text{accuracy}$; and that there is a drug–dosage interaction on yet another combination.

The capacity to find such combinations (on which groups maximally differ) should be the attraction of MANOVA, as it would appear to be addressing the complexity of our data. Yet apart from reporting whether there are significant multivariate tests for main effects and interaction, researchers tend to avoid the full MANOVA results and, in short, usually revert to testing and discussing their measures one by one – why? Primarily, it would seem, because of the difficulty in explaining (or interpreting) these combinations. How can we explain the finding that the groups differ on $0.5 \times \text{speed} - 1.6 \times \text{accuracy}$? If there are many measures things are even more difficult: consider four neuropsychological tests (A–D) and we have a difference on $0.23 \times \text{A} - 1.8 \times \text{B} - 0.09 \times \text{C} + 2.45 \times \text{D}$. If A and D are frontal tasks, perhaps we can think of this as suggestive of frontal (A + D) versus non-frontal (B + C); or perhaps not. Some of the difficulty with such combinations arises because MANOVA is drawing upon all sources of variation to try and differentiate the groups. Even allowing for that, and for attempts by authors to describe ways of simplifying combinations into more interpretable forms, MANOVA remains in retreat.

Repeated measures

The analysis of repeated-measures data – where an outcome for a subject is measured on two or more occasions over time or under two or more conditions – is now routinely analysed using MANOVA, or the multivariate model, rather than the earlier univariate ANOVA model. To do this, the repeats (e.g. a subject's anxiety pre-treatment, post-treatment and at follow-up) are treated as if they were multiple variates (just as speed and accuracy were multiple variates) and a MANOVA is carried out to compare groups (e.g. placebo vs. drug). However, there is one important difference. It would make little sense to say that drug and placebo differed on a combination of occasions which was say $0.25 \times$ 'pre' $- 1.8 \times$ 'post' $+ 2.93 \times$ 'follow-up'. Hence, the combinations of occasions are restricted to those of a type, usually called contrasts, which are meaningful as differences between conditions or occasions (and indeed groups as well).

If we have multiple measures made at each occasion or under each condition the situation is usually called *doubly multivariate* and is somewhat more complex since our variates represent different measures made at different times. Researchers will almost always only be interested in those combinations of variates where the implied combination of measures is the same for each occasion (to avoid comparing speed − accuracy under drug A and speed + accuracy under drug B) and where the implied differences between occasions too are meaningful. Naturally any researcher who does not like interpreting combinations of variates will want to avoid the double multivariate model.

https://doi.org/10.1111/j.1601-5215.2010.00488.x Published online by Cambridge University Press

The MANOVA model is not feasible when there is missing data, a frequent occurrence in research with repeated measures. One relatively recent solution is to use *mixed models* (also *mixed-effects models*, *multilevel models* and alas other names) which allow (amongst other flexibilities) subjects to have missing data. Indeed, it is likely that this will be the expected approach in the future.

Discriminant analysis

The combinations produced by a MANOVA are called *discriminant functions* and they are central to the statistical method called discriminant function analysis (DFA). The aim of DFA is to use these combinations – which we know maximally separate groups – as the basis for classifying future individuals, whose group is unknown, into their most likely groups (in essence, to 'diagnose'). Even though these functions will look no more or less interpretable than the ones discussed above (they are exactly the same after all) researchers have thought it more comfortable in using them as their classificatory purpose seems to require less interpretation or even none at all.

The use of DFA, however, seems to be rapidly declining. Where once a DFA might have been used researchers now tend to use logistic regression (LR) because LR generally performs the task better than DFA, particularly when the measures are not continuous and normal. Thus a researcher wanting to use a broad range of data (continuous and categorical, normal or not) to classify first-episode psychoses into 'likely schizophrenia' versus 'likely bipolar/mania' will prefer the less restrictive LR model. In addition, the contribution of measures in predicting group membership seems clearer in LR as the weights in the regression are readily expressed as odds ratios. If a researcher's data is not best analyzed by LR there are a number of alternatives to both LR and DFA.

Type I errors

Researchers wish to compare bipolar I and bipolar II patients on a large number of different measures, but are aware that carrying out a large number of *t*-tests is likely to inflate the number of type I errors (falsely concluding that there is a difference), so they use the results of an initial MANOVA as a hurdle. Since the question tested by the MANOVA is whether there is any combination of measures (including measures taken singly) for which the groups differ, if the MANOVA is non-significant no further analyses are done. However, if it is significant then univariate *t*-tests (or ANOVAs when there are more than two groups) are done for all the measures. There is general acceptance that this approach – sometimes called *protected tests* – is not a good way to proceed since it often fails to 'protect' the researcher; there exist better ways of addressing the problem of many tests and, though they are less mechanical than simply doing a preliminary MANOVA, researchers fortunately are taking them up.

Dusan Hadzi-Pavlovic[1,2]

[1]School of Psychiatry, University of New South Wales, Kensington, NSW, Australia; and

[2]Black Dog Institute, Prince of Wales Hospital, Randwick, NSW, Australia

Dusan Hadzi-Pavlovic,
Black Dog Institute Building,
Prince of Wales Hospital,
Hospital Road,
Randwick,
NSW 2031,
Australia.
Tel: +61 2 9382 3716;
Fax: +61 2 9382 3712;
E-mail: d.hadzi-pavlovic@unsw.edu.au

Acta Neuropsychiatrica 2010: 22: 257–258
© 2010 John Wiley & Sons A/S
DOI: 10.1111/j.1601-5215.2010.00488.x

https://doi.org/10.1111/j.1601-5215.2010.00488.x Published online by Cambridge University Press