Canonical correlation analysis: an introduction to a multivariate statistical analysis

1 author:

Nuwan Darshana Wickramasinghe

Rajarata University of Sri Lanka

152 PUBLICATIONS **1,107** CITATIONS

Some of the authors of this publication are also working on these related projects:

Burnout and Work Engagement in Collegiate Cycle Students in Sri Lanka View project

Continued Medical Education

Canonical correlation analysis: an introduction to a multivariate statistical analysis

Nuwan Darshana Wickramasinghe

Department of Community Medicine, Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka

Correspondence: nuwick74@yahoo.com

Background

Multivariate analysis includes statistical techniques that simultaneously analyse multiple measurements on a number of individuals or objects (1-2). In multivariate analysis, a variate is defined as a linear combination of variables with empirically determined weights. While the variables are specific to a particular study, the weights are determined by the multivariate technique used in the study (1).

Canonical correlation analysis (CCA) is a multivariate statistical method (1-3) and the initial analytic framework of the CCA was introduced by Hotelling in mid 1930s (4). CCA explores the relationship between multiple dependent and independent variables. The underlying principle of CCA is to investigate the relationship between the variables by developing a number of independent canonical functions that maximize the correlation between the linear composites (1, 3, 5), known as canonical variates. The canonical correlation coefficient measures the strength of association between the variable sets under concern (1, 6). Canonical loadings measure the simple linear correlation between the independent variables and their respective canonical variates, whereas canonical cross-loadings measure the correlation of each observed variable with the opposite canonical variate (3).

In CCA, variable sets are often categorised as the predictor set (independent variables) and the criterion set (dependent variables); however, causal inferences are not drawn solely based on this analysis (3). As highlighted, CCA is a unique statistical technique in which either set of variables may contain one or more variables and CCA derives weights for each variable set with maximally correlated weighted sums (7).

Applications of CCA

One of the advantages of using CCA is that, being a multivariate technique, it limits the probability of committing Type I error in studies (8). Furthermore, CCA provides an opportunity to explore the complexity of multiple relationships of constructs under investigation (3), especially in psychological, behavioural and educational research. In addition, CCA can be used instead of other parametric tests in many instances (8-9), since many parametric tests can be subsumed by CCA as special cases in the general linear model (3).

As for any multivariate analysis, there are basic assumptions in the CCA. Even though CCA does not make strong normality assumptions, it is recommended that all variables be evaluated for normality and transformed if necessary. More importantly, CCA assumes linear relations among the variables. Further, multicollinearity and homoscedasticity need to be assessed. As highlighted above, CCA is most appropriate when the relationship between multiple variables is examined. Figure 1 illustrates the relationship as in CCA between two sets of hypothetical variables, which consist of three predictor- and three criterion-variables. In order to evaluate the relationship between several predictor and criterion variables simultaneously, the observed variables in each set are combined into one synthetic/ latent predictor and criterion variable separately.
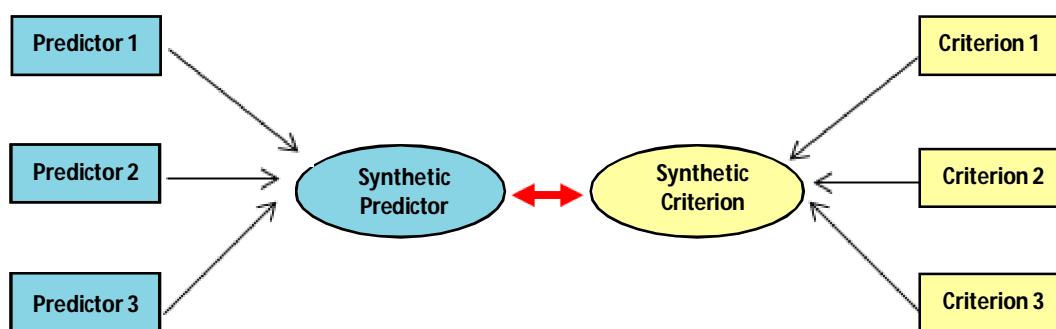
Figure 1. Relationship between a set of predictor variables and a set of criterion variables

For instance, a study conducted to examine the relationship between burnout syndrome (criterion) and organizational commitment (predictor) among health-care professionals has employed CCA [10]. In that study, the three dimensions of burnout, viz., emotional exhaustion, depersonalization and personal accomplishment were considered as one set of variables while sub-dimensions of organizational commitment, viz., affective commitment, continuance commitment and normative commitment were considered as the other variable set. The analysis revealed that in addition to the statistically significant correlation between burnout syndrome and organizational commitment, emotional exhaustion and affective commitment contributed the most towards the explanatory capacity of canonical variables estimated from the sub-dimensions of burnout syndrome and sub-dimensions of organizational commitment respectively (10).

As highlighted in the above study findings, CCA makes provisions not only to extract the relationship between the two sets of variables, but also the relative contributions of each variable to the canonical relationships. Hence, CCA is an important multivariate technique that could be used to explore complex relationships between multiple variables.

References

1. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate data analysis* (7th Edition). Harlow: Pearson Education Ltd., 2014.

2. Anderson TW. *An introduction to multivariate statistical analysis* (3rd Edition). New Jersey: John Wiley & Sons Inc., 2003.

3. Sherry A & Henson RK. Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment* 2005; 84(1): 37-48. DOI: 10.1207/s15327752jpa8401_09

4. Hotelling H. Relations between two sets of variates. *Biometrika* 1936; 28: 321-377. DOI: 10.2307/2333955

5. Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation* 2004; 16(12): 2639-2664. DOI: 10.1162/0899766042321814

6. Alpert MI, Peterson RA. On the interpretation of canonical analysis. *Journal of Marketing Research* 1972; 9(2): 187-192. DOI: 10.2307/3149953

7. Wimmer R. Canonical correlation/factor analysis: Similarities and differences. *Journal of Broadcasting* 1977; 21(2): 211-213. DOI: 10.1080/08838157709363830

8. Thompson B. A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counselling and Development* 1991; 24: 80-95.

9. Henson RK. Demystifying parametric analyses: Illustrating canonical correlation as the multivariate general linear model. *Multiple Linear Regression Viewpoints* 2000; 26(1): 11-19.

10. Enginyurt O, Cankaya S, Aksay K, Tunc T, Koc B, Bas O, Ozer E. Relationship between organisational commitment and burnout syndrome: a canonical correlation approach. *Australian Health Review* 2016; 40(2): 181-187. DOI: 10.1071/AH14177.