MULTIVARIATE ANALYSIS – A GREAT TOOL FOR A MULTIVARIATE WORLD

Carlos A. Coelho, Full Professor of Statistics PhD
Department of Mathematics, NOVA University of Lisbon, Portugal
e-mail: cmac@fct.unl.pt

Abstract: In this short report the author tries to introduce in simple terms some Multivariate Analysis techniques and show their usefulness in the analysis of econometric data, namely in studies which aim at studying and comparing different regions inside a given country.

Key words: likelihood ratio tests, equality of mean vectors, covariance structures

JEL: C02, C12.

1. Introduction

In a world where multiple variables are commonly measured on given sets of subjects, Multivariate Analysis is a natural set of tools and techniques that may be useful in the analysis of data.

There are several Multivariate Analysis tools and techniques that may be useful in helping to assess the existence of differences between two or among several areas of a given country, or even in helping to assess the existence of some gradient or trend, even in situations where such trend is not linear.

One of the most powerful set of techniques available in Multivariate Analysis is the set of likelihood ratio testing techniques, the likelihood being the model or expression for the distribution of the underlying random vectors, seen as a function of the parameters in the distribution.

In every test in Statistics there is a null hypothesis to be tested, versus a given alternative hypothesis. A likelihood ratio test is a test whose associated statistic is built as the ratio of the expression of the likelihood under the null hypothesis by the expression of the likelihood under the alternative hypothesis.

2. Likelihood ratio tests in Multivariate Analysis

Likelihood ratio tests in Multivariate Analysis may be built to test differences between two or among several mean vectors pertaining to a given set of variables of interest that were measured on a set of observation units, in different regions of a country, in different institutions, hospitals, etc. (Kshirsagar, 1972, Chap. 5 and 9; Anderson, 2003, Chap. 5 and 8; Morrison, 2005, Chap. 4 and 5; Muirhead, 2005, Chap. 10; Coelho and Arnold, 2019, Sect. 5.1.1), or even to assess the existence of some trend among such vectors of expected values, trend that does not need to be linear and which may even be somehow unsuspected. This latter assessment may be done using what is called a Growth Curve model (Rao, 1958; Potthoff and Roy, 1964; Kshirsagar, 1972, Sec. 5.5; Kshirsagar and Smith, 1995; Coelho and Arnold, 2019, Subsect. 5.1.4.6) or what some authors call more recently a Bilinear Regression model (von Rosen, 2018), or even yet to test hypotheses on a matrix of expected values (Muirhead, 2005, Chap.10; Coelho and Arnold, 2019, Susect. 5.1.4). A somewhat sometimes forgotten set of tests which may provide extremely useful insights is the Profile Analysis (Kshirsagar, 1972, Sec. 5.4; Morrison, 2005, Chap. 4 and 5; Coelho and Arnold Subsect. 5.1.3, 5.1.7).

Also tests for covariance structures may be built, some of which may be turned into much more than that as it is the case of the test of independence of two or more sets of variables. This latter test is the underlying test in the Canonical Analysis and in the Generalized Canonical Analysis models, and under this view it may be seen as the underlying test for most linear models, including the test of equality of mean vectors and the tests in the Growth Curve model (Kshirsagar 1972, Chap. 7; Coelho, 1992; Sengupta, 2014; Coelho and Arnold, Subsect 5.1.9 and 5.1.11).

Likelihood ratio tests may be developed for more elaborate or more embracing covariance structures (Olkin and Press, 1969; Olkin, 1973; Chao and Gupta, 1991; Moschopoulos, 1992; Cardeño and Nagar, 2001; Marques and Coelho, 2012; Coelho and Marques, 2013, Coelho and Roy, 2020), some of which have this test as a particular case (Correia et al., 2018).

Also, lately, tests of equality of (several) mean vectors have been developed for situations where the covariance matrices of the underlying random vectors have some given structure of interest, both in cases where this structure may be or not a block version (Coelho, 2017, 2018). These tests have increased power for situations where these are the correct structures for the covariance matrices.

Although the exact distributions of all these likelihood ratio statistics are usually seen as mostly intractable, lately it has been shown that many of them have a rather simple expression for their exact distributions (Coelho and Arnold, 2019, Chap. 5) or that otherwise, sharp and very manageable near-exact distributions may be developed for these statistics, which allow for the easy computation of sharp near-exact p-values and quantiles (Coelho, 2004; Coelho et al., 2010; Coelho and Marques, 2010; Marques and Coelho, 2012; Coelho and Marques, 2013; Correia et al., 2018).

These near-exact distributions are asymptotic distributions, which besides being asymptotic for increasing sample sizes, as it happens with common asymptotic distributions, they are also asymptotic for increasing number of variables involved, opposite to what happens with common asymptotic distributions, which degrade their performance when the number of variables involved increases.

These distributions are built by leaving most of the exact distribution of the statistic untouched and approximating then asymptotically the smaller part of the distribution which makes the whole distribution intractable. This has to be done in such a way that when we join back together the two parts, the one left untouched and the asymptotic approximation obtained for the remaining part, a tractable near-exact distribution is obtained.

Most of the tests developed for real random variables may also be developed for complex random variables (Coelho et al., 2015; Coelho and Arnold, 2019, Sec. 5.1.5, 5.1.6, 5.1.7, 5.1.8, 5.1.10, 5.1.12, 5.1.14, 5.1.19, 5.1.20). These tests may be useful in spectral analysis of time series to uncover periodicities.

Conclusion

There are a number of likelihood ratio tests available in Multivariate Analysis which may be a much adequate and useful set of tools to help in the analysis of econometrical data, namely when trying to compare different regions of a country, assessing and testing the significance of differences among them or assessing the existence of gradients or trends among these regions.

Moreover, most likelihood ratio tests in Multivariate Analysis are invariant for scale and/or location changes, which is a much desirable feature, mainly in econometric applications.

Literature

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, J. Wiley & Sons, Hoboken, New Jersey.

Cardeño, L., Nagar, D. K. (2001). Testing block sphericity of a covariance matrix, *Divulgaciones Matematicas*, 9, 25-34.

Chao, C. C., Kupta, A. K. (1991). Testing of homogeneity of diagonal blocks with blockwise independence, *Communications in Statistics – Theory and Methods*, 20, 1957-1969.

Coelho, C. A. (1992). *The Generalized Canonical Analysis*, Ph. D. Thesis, The University of Michigan, Ann Arbor, MI, U.S.A.

Coelho, C. A. (2004). The Generalized Near-Integer Gamma distribution: a basis for 'near-exact' approximations to the distribution of statistics which are the product of an odd number of independent Beta random variables, *Journal of Multivariate Analysis*, 89, 191-218.

Coelho, C. A. (2017). The Likelihood Ratio Test for Equality of Mean Vectors with Compound Symmetric Covariance Matrices. In: Gervasi, O.,Murgante, B., Misra, S., Borruso, G., Torre, C.M., Rocha, A.M.A.C., Taniar, D., Apduhan, B.O., Stankova, E., Cuzzocrea, A. (eds.) *Computational Science and Its Applications*. Lecture Notes in Computer Science 10408, vol. V, pp. 20–32, Springer, New York.

Coelho, C. A. (2018). Likelihood ratio tests for equality of mean vectors with circular covariance matrices. In: Oliveira, T.A., Kitsos, C., Oliveira, A., Grilo, L.M. (eds.) *Recent Studies on Risk Analysis and Statistical Modeling*, pp. 255–269. Springer, New York.

Coelho, C. A., Arnold, B. C. (2019). *Finite Form Representations for Meijer G and Fox H Functions – Applied to Multivariate Likelihood Ratio Tests Using Mathematica®, MAXIMA and R*. Lecture Notes in Statistics. Springer, Cham.

Coelho, C. A., Arnold, B. C., Marques, F. J. (2010) Near-exact distributions for certain likelihood ratio test statistics. *Journal of Statistical Theory and Practice*, 4, 4, 711-725.

Coelho, C. A., Arnold, B. C., Marques, F. J. (2015). The exact and near-exact distributions of the main likelihood ratio test statistics used in the complex multivariate normal setting, *Test*, 24, 386-416.

Coelho, C. A., Marques, F. J. (2010) Near-exact distributions for the independence and sphericity likelihood ratio test statistics, *Journal of Multivariate Analysis*, 101, 583-593.

Coelho, C. A., Marques, F. J. (2013). The multi-sample block-scalar sphericity test: exact and near-exact distributions for its likelihood ratio test statistic, *Communications in Statistics – Theory and Methods*, 42, 1153-1175.

Coelho, C. A., Roy, A. (2020). Testing the hypothesis of a doubly exchangeable covariance matrix. Metrika, 83, 45-68.

Correia, B. R., Coelho, C. A., Marques, F. J. (2018). Likelihood ratio test for the hyper-block matrix sphericity covariance structure – characterization of the exact distribution and development of near-exact distributions for the test statistic. *REVSTAT*, 16, 365-403.

Kshirsagar, A. M. (1972). *Multivariate Analysis*, Marcel Dekker, New York.

Kshirsagar, A. M., Smith, W. B. (1995). *Growth curves*, Marcel Dekker, New York.

Marques, F. J., Coelho, C. A. (2012). Near-exact distributions for the likelihood ratio test statistic of the multi-sample block-matrix sphericity test, *Applied Mathematics and Comput*ation, 219, 2861-2874.

Morrison, D. F. (2005). *Multivariate Statistical Methods,* 4th ed., Duxbury Press, Duxbury.

Moschopoulos, P. G. (1992). The hypothesis of multisample block sphericity. *Sankhya*, 54, 260-270.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, 2nd ed., J. Wiley & Sons, Hoboken, New Jersey.

Olkin, I. (1973). Testing and Estimation for structures which are circularly symmetric in blocks. In: D. G. Kabe and R. P. Gupta, eds., *Multivariate Statistical Inference*, 183-195, North Holland, Amsterdam.

Olkin, I., Press, S. J. (1969). Testing and estimation for a circular stationary model, *Annals of Mathematical Statistics*, 40,1358-1373.

Potthoff, R. F., Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. Biometrics, 14, 1-17.

von Rosen, D. (2018). *Bilinear Regression Analysis – An Introduction*, Springer, Lecture Notes in Statistics 220, Cham.

SenGupta, A. (2014). Generalized Canonical Variables. In: *StatsRef – Statistical Reference Online*, Wiley, New York. http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat02670/abstract