CBCS
http://agrobiol.sggw.waw.pl/cbcs

Communications in Biometry and Crop Science
Vol. 1, No. 1, 2006, pp. 32–34

International Journal of the Faculty of Agriculture and Biology,
Warsaw Agricultural University, Poland

NOTE

Note on modern path analysis in application to crop science

Marcin Kozak [1]*, Manjit S. Kang [2]

[1] Department of Biometry, Warsaw Agricultural University, Nowoursynowska 159, 02-776, Warsaw, Poland.
[2] Department of Agronomy and Environmental Management, Louisiana State University Agricultural Center, Baton Rouge, LA 70803-2110, USA.
* Corresponding author: Marcin Kozak, E-mail: m.kozak@omega.sggw.waw.pl

CITATION: Kozak, M., Kang, M.S. (2006). Note on modern path analysis in application to crop science. *Commun. Biometry Crop Sci.* 1 (1), 32-34.

Received: 23 March 2006, Accepted: 11 May 2006, Published online: 1 June 2006

ABSTRACT
In this note, we discuss path analysis and its applications to crop science investigations – almost all of these applications are based on the methodology that Sewall Wright developed in the 1920s, although the path analysis method has been intensively developed during the last three decades. Via this paper, we point out that new methodology of estimation and testing in path analysis is fairly advanced and is applied in many biological disciplines and that it should also be extended to crop science.
Key Words: *causality; maximum likelihood; estimation; testing*.

INTRODUCTION

Path analysis is a basic method that enables drawing inferences about causal structure of data. Sewall Wright (1921, 1934) originally developed the method; later on, it was expanded and applied by many authors. However, as Shipley (2002, sec. 3.2) points out, scientists somehow ignored the concept of path analysis for many years, mostly because it was beyond the statistical thinking of the times. However, since the 1980s, path analysis has found applications in numerous biological problems.

A quick survey of core crop science journals would suggest that crop scientists commonly accept path analysis. Usually, when some kind of causal analysis is to be applied, path analysis is used. We do not provide a list of such papers, because it would be too long. However, the paper by Dewey and Lu (1959) must be cited here, because it laid the basis for further applications of path analysis in plant sciences. In fact, as of May 8, 2006, this paper has been cited in at least 115 other research publications (Google Scholar citation result; click here to check the current citation result of the paper) and is still a basis for most recent applications.

It must be pointed out, however, that the path analysis method of Wright has witnessed significant changes during the last few decades. For that reason, applications based on Dewey and Lu's paper would appear to be out-of-date. As is the case in scientific disciplines, theory is always ahead of applications; it must be explored enough to be properly applied. It is crucial, however, not to ignore theories that are developed enough to be applied, and that have been used in other scientific disciplines. Here the meaning of exchanging experience/information between various disciplines of science is perfectly visible.

Jöreskog (1970) provided a new theory for estimation and testing in path analysis. It is based on the maximum likelihood approach, where parameters are estimated to minimize the difference between elements of estimated and observed variance-covariance matrices (irrespective of whether data are standardized or not). Thus, Jöreskog's work laid the basis for further development of path analysis and structural equation modeling (SEM) – the method that path analysis is a part of. Shipley (2002) provided an excellent introduction and guide to path analysis and SEM. He addressed his book to biologists, so also to crop scientists; in fact, some of the examples in the book are related to agronomy. It is worth noting that a journal named *Structural Equation Modeling* exists to report work in this area. In this paper, we focus only on path analysis, without discussing other elements of structural equation modeling.

We must point out here that the likelihood approach to path analysis cannot be applied to the situation in which multiple regression is expressed as a path model (which is an often used approach; see, e.g., Aasmo Finne et al., 2000; Seker et al., 2003; Das et al., 2004; or Ssago et al., 2004). As Shipley (2002, sec. 4.3) explains, in such a situation the number of unique elements of the variance-covariance matrix of the variables is equal to the number of free parameters to be estimated, so there are zero degrees of freedom to test the model. Therefore, the application possibilities of the likelihood-based path analysis are limited to models in which there are more elements in the variance-covariance matrix than free parameters to be estimated (this assumption is met in most diagrams that are more complex than the one of multiple regression expressed as a path analysis model). However, the causal structure provided by Dofing and Knight (1992), subsequently used, for instance, by Donaldson et al. (2001) or Maman et al. (2004), cannot be estimated via the likelihood-based approach, because of the reasons stated above.

The novel methodology of path analysis has been applied in various biological and ecological studies (see Shipley (2002) and the citations there). Insofar as we could determine, the only crop science paper in which the likelihood-based approach to path analysis was employed is that of Guillen-Portal et al. (2006).

An additional issue that we would like to address here is the exploratory character of path analysis. This topic is usually ignored in applications of path analysis: a causal structure is assumed and parameters of the model assumed are estimated. Path analysis, however, gives an important opportunity to explore the causal structure. As the effect of such an analysis, we are able to provide, on the basis of sample data, the best model of causal associations we aim to study. It is good to be aware that the likelihood-based approach to path analysis provides efficient tools for model selection. Nonetheless, note that usually there are many possible models to consider, which makes it unlikely that all of them would be studied (see Shipley, 1997; Shipley, 2002, pp. 228-230). In practice, one studies some of them, especially those that are most worthy. Shipley (1997) addresses the issue of exploratory path analysis.

What needs to be carefully addressed is a problem of applying path analysis to data originating from factorial experiments. In general, it is assumed that the data subjected to path analysis are generated by the same causal structure (see Shipley, 2002). Therefore, caution must be exercised while applying path analysis to data originating from factorial experiments, which is, in fact, a time-honored approach in crop science. For a thorough discussion on this subject, refer to Kozak and Kang (2006).

In conclusion, path analysis is a method that is undoubtedly useful in studying causal structures in all disciplines of science. Crop scientists should take account of its dynamic development in the past two to three decades and try to apply the novel estimation approach in their research. Without this, studying causal associations in crop science will lag behind other disciplines of science in which the new path analysis methodology is being put into practice.

REFERENCES

Aasmo Finne, M., Rognli, O.A., Schjelderup, I. (2000). Genetic variation in a Norvegian germplasm collection of white clover (*Trifolium repens* L.). 3. Correlation and path coefficient analyses of agronomic characters. *Euphytica* 112, 57-68.

Das, M.D., Fuentes, R.G., Taliaferro, C.M. (2004). Genetic variability and trait relationships in switchgrass. *Crop Sci*. 44, 443-448.

Dewey, D.R., Lu, K.H. (1959). A correlation and path-coefficient analysis of components of crested wheatgrass seed production. *Agron. J*. 51, 515-518.

Dofing, S.M., Knight, C.W. (1992). Alternative model for path analysis of small-grain yield. *Crop Sci*. 32, 487-489.

Donaldson, E., Schillinger, W.F., Dofing, S.M. (2001). Straw production and grain yield relationships in winter wheat. *Crop Sci*. 41, 100-106.

Guillen-Portal, F.R., Stougaard, R.N., Xue, Q., Eskridge, K.M. (2006). Compensatory mechanisms associated with the effect of spring wheat seed size on wild oat competition. *Crop Sci*. 46, 935-945.

Jöreskog, K. (1970). A general method for analysis of covariance structures. *Biometrika* 57, 239-251.

Kozak, M., Kang, M.S. (2006). Note on applying statistical methods for pooled data. (Communicated).

Maman, N., Mason, S.C., Lyon, D.J., Dhungana, P. (2004). Yield components of pearl millet and grain sorghum across environments in the Central Great Plains. *Crop Sci.* 44, 2138-2145.

Seker, H., Rowe, D.E., Brink, G.E. (2003). White clover morphology changes with stress treatments. *Crop Sci*. 43, 2218-2225.

Shipley, B. (1997). Exploratory path analysis with applications in ecology and evolution. *Amer. Naturalist* 149(6), 1113-1138.

Shipley, B. (2002). *Cause and correlation in biology. A user's guide to path analysis, structural equations and causal inference*. Cambridge University Press, Cambridge.

Ssago, F., Speijer, P.R., Coyne, D.L., De Waele, D. (2004). Path analysis: a novel approach to determine the contribution of nematode damage to East African Highland banana (*Musa* spp., AAA) yield loss under two crop management practices in Uganda. *Field Crops Res.* 90, 177-187.

Wright, S. (1921). Correlation and causation. *J. Agric. Res*. 20, 557-585.

Wright, S. (1934). The method of path coefficients. *Annals Math. Statist*. 5, 161-215.