Biometrical Letters
Vol. 47 (2010), No. 2, 129-132

Note on two methods of additive yield component analysis

Marcin Kozak

Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland, e-mail: nyggus@gmail.com

SUMMARY

The aim of this note is to show that two existing methods of additive yield component analysis are based on the same statistical background, although the interpretation they offer differs.

Key words: additive yield components, variance decomposition, path analysis.

1. Introduction

Jolliffe and Courtney (1984) defined an additive yield component model as follows:

$$Y = \sum_{i=1}^{k} X_i \,, \tag{1}$$

where Y is the response variable (in agricultural analyses this is usually yield, although it can also be any other response trait that can be analysed with equation (1)) and X_i is the ith (i = 1, …, k) additive component. Additive yield component analysis aims to study which of the additive components are most important in determining the response variable Y, and to describe this influence (Kozak et al. 2006). In social, behavioural and medical sciences the variable (1) is called the composite variable or composite score (e.g. Bentler 2004).

Some agricultural examples where additive yield component analysis can be of help are listed here: (a) root yield of sugar beet is studied as a sum of yields of roots from different fractions; (b) cereal grain yield is studied as a sum of grain yields from plants with one spike, two spikes, and so on; (c) total seed yield of oilseed rape is studied as a sum of seed yield of main stem plus first-order branches, second-order branches, and so on; (d) total biomass yield of sugar beet is studied as a sum of mass of roots, leafblade and petioles; etc.

Because of the specific form of the model (1), additive yield component analysis requires the employment of special methods of analysing the influence of components on yield. In the literature, two propositions can be found. The aim of this paper is to show that they are based on the same statistical background.

2. Results

Kozak et al. (2002) suggested applying Piepho's (1995) approach to study the problem; although concerned with multiplicative yield components analysis, Piepho's approach can be easily adapted to the problem in question because after applying a logarithmic transformation the model analysed is of a similar form as that in equation (1). The interpretation is based on the following decomposition of the variance of the response variable Y into variance and covariance terms of the additive components:

$$\sigma_Y^2 = \sum_{i=1}^{k} \sigma_i^2 + 2\sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} \sigma_{ij} \tag{2}$$

where σ_y is the population standard deviation of Y, σ_i is the population standard deviation of the ith additive yield component X_i, and σ_{ij} is the population covariance between X_i and X_j. Based on (2), further coefficients are defined to facilitate explanation of the share of Xs in the variance of Y. The quality given in equation (2) is in fact commonly known in statistics. Although at first sight this decomposition seems the easiest approach to the problem, Kozak et al. (2006) suggested that this was not a classical approach to studying causal relationships among traits, and proposed an alternative method to serve this purpose. It is based on coefficients similar to those used in path analysis; since path analysis is well-known among agronomists and crop scientists, this similarity can be regarded as an advantage of the approach. Let us recall that the influence (in population) P_i of X_i on Y can be described as (Kozak et al. 2006)

$$P_i = \sigma_i / \sigma_y \,. \tag{3}$$

This is a standardized coefficient with interpretation exactly the same as that of a standardized path coefficient. The interpretation in the approach also uses coefficients for indirect effects. Here we show that both of these methods are based on the same statistical background. Note that the determination coefficient in the model (1) is equal to 100%. Applying the results given by Kozak et al. (2007) for path analysis with all predictors set at the same ontogenetic level (so none of them is a cause or an effect of any other predictor), we can present the decomposition of the determination coefficient R^2 of Y into shares Q_i corresponding to the additive components as follows:

$$R^2 = 100\% = 1 = \sum_{i=1}^{k} Q_i \,, \tag{4}$$

where the Q_i's are given by

$$Q_i = P_{iy}^2 + \sum_{j=1,\, j\neq i}^{k} P_{iy} P_{jy} r_{ij} \,. \tag{5}$$

Using (4) and (5) we can easily show that the adaptation of Piepho's (1995) method and Kozak et al.'s (2006) approach are based on the same statistical background and that in terms of the decomposition (4) they are in fact the same. This can be shown by the following straightforward derivation:

$$1 = \sum_{i=1}^{k} Q_i = \sum_{i=1}^{k} P_i^2 + 2\sum_{i=1}^{k} \sum_{j=1,\, j\neq i}^{k} P_i P_j \rho_{ij} = \sum_{i=1}^{k} \frac{\sigma_i^2}{\sigma_Y^2} + 2\sum_{i=1}^{k} \sum_{j=1,\, j\neq i}^{k} \frac{\sigma_i}{\sigma_Y} \frac{\sigma_j}{\sigma_Y} \frac{\sigma_{ij}}{\sigma_i \sigma_j},$$

from which it follows that

$$1 = \sum_{i=1}^{k} \frac{\sigma_i^2}{\sigma_Y^2} + 2\sum_{i=1}^{k} \sum_{j=1,\, j\neq i}^{k} \frac{\sigma_{ij}}{\sigma_Y^2} \,. \tag{6}$$

Multiplying both sides of (6) by σ_Y^2 yields the following equation:

$\sigma_Y^2=\sum_{i=1}^{k}\sigma_i^2+2\sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k}\sigma_{ij}$, the decomposition of the variance of *Y* that is exactly the same as that used in the adaptation of Piepho's (1995) approach; cf. equation (2).

3. Conclusion

From the above result it follows that both approaches to additive yield component analysis, namely one that adapts Piepho's approach and one that uses coefficients similar to those from path analysis, have the same statistical background. Nonetheless, following Kozak et al.'s (2006) conclusion based on Rencher (1998, p. 210), it is worth noting that the coefficients (3) are concerned with studying the *influence* of additive components on *Y*, whereas the decomposition of the variance of *Y*, offered by both approaches (as has been shown in this paper), is concerned with studying the *shares* of additive components in the determination (variance) of *Y*. To make the interpretation complete, both interpretation tools, namely that concerned with influence and that concerned with shares of components, should be applied in additive yield component analysis.

REFERENCES

Bentler P.M. (2004): Maximal reliability for unit-weighted composites. UCLA Statistics Preprint no. 405, Los Angeles, University of California.

Jolliffe P.A., Courtney W.H. (1984): Plant growth analysis: additive and multiplicative components of growth. Annals of Botany 54: 243–254.

Kozak M., Gozdowski D., Wyszyński Z. (2006): An approach to analyzing a response variable as affected by its additive components: Example for spring barley grain yield. Cereal Research Communications 34(2-3): 981–988.

Kozak M., Kang M.S., Stępień M. (2007): Causal pathways when independent variables are co-related: New interpretational possibilities. Plant, Soil and Environment 53(6): 267–275.

Kozak M., Mądry W., Wyszyński Z. (2002): Method of analysis of the influence of weight of roots from different fractions on sugar beet yield. Fragmenta Agronomica 2 (74): 251–262 (in Polish).

Piepho H.P. (1995): A simple procedure for yield component analysis. Euphytica 84: 43–48.

Rencher A.C. (1998): Multivariate statistical inference and applications. John Wiley and Sons, New York.