Listy Biometryczne – Biometrical Letters
Vol. 41 (2004), No. 2, 59-69

New concept of yield component analysis

Marcin Kozak

Department of Biometry, Warsaw Agricultural University,
Nowoursynowska 159, 02-776 Warsaw, Poland;

SUMMARY

The paper contains considerations on investigations on an influence of multiple characters developing at the same time (i.e. co-related) on their product. In agronomy such problem is called a "Yield Component Analysis". A new approach to the question is presented and two different forms of estimators are compared using simulation studies. Afterwards, recommendations regarding the appropriate (although biased) estimators of the influence of the components on the yield are given.

KEY WORDS: multiplicative predictors, yield components, Monte Carlo study.

1. Introduction

In many investigations in biological and agricultural sciences we meet a situation in which a final effect, i.e. a dependent variable, is a product of its causes, i.e. predictor variables. Although an analysis of an influence of the predictors on the response variable seems quite easy in such a case, an inquisitive look at the problem causes some doubts. There is no natural and univocal method that could be applied to the problem when we aim to analyze the importance of the multiplicative predictors in affecting the response.

The question has been deeply investigated especially in agricultural sciences. It is called a "Yield Component Analysis", for it is the commonly considered problem in analyzing the influence of the components on the yield, both per plant or area unit. Kozak (2002) and Kozak and Mądry (2004) gave a definition of the components as the variables which product gives the yield.

In what follows we will call the problem the yield component analysis; the dependent variable will be also called the yield, and the causes – the components. Certainly, we do not resolve ourselves into such question; it is widened to all biological and other problems that can be represented by the multiplicative function (1) – see below. The appropriate method of the yield component analysis should be, first of all, convenient in the interpretation (as, for instance, the standardized multiple linear regression), and secondly, it should take into account (at least approximately) a mathematical deterministic multiplicative function of the response variable and its predictors given by

$$Y = f(X_i) = \prod_{i=1}^{k} X_i, \tag{1}$$

where Y is the yield, and $X_i, i = 1, ..., k$, are the yield components.

The form of the relationship (1) may make us think that the influence of the predictors (components) is obvious; unfortunately, it is not so easy and natural. Therefore many statistical methods have been proposed to study the quantity of the influence of the components on the yield. Fraser and Eaton (1983) gave a comprehensive description of the methods used in yield component analysis; Kozak (2002) did the same, but he concentrated mainly on the methods proposed after 1983.

The most often statistical method used in yield component analysis is a linear regression analysis, especially in a standardized scale (also called a path analysis). It requires an assumption on linearity of the relationship between the yield and its components. Therefore, the interpretation is approximate and biased; sometimes a value of an estimator of a coefficient of determination of the linear model can be quite big (almost 100%), but sometimes it is not satisfying, even less than 80% (Kozak, 2004); (and, as it is shown in the paper, it can be near 0%). Note that the real determination is 100%, because the relationship (1) is deterministic. Moreover, knowing the real form of the relationship (1), which is not too complex, we would like to use it instead of approximating it with the linear model.

Therefore, several statistical more or less appropriate nonlinear approaches have been proposed. Some of them used a logarithmic transformation of the function (1), see e.g. Eaton and Kyte (1978), Hardwick and Andrews (1980), Piepho (1995), of the form

$$\ln Y = \sum_{i=1}^{k} \ln X_i. \tag{2}$$

The logarithmic transformation (2) simplifies the relationship, because from the multiplicative function (1) we obtain the additive one (2). Still the analysis cannot be satisfactory enough, for the interpretation regards the logarithmic scale. Obtained coefficients, (specific for the particular method), need to be interpreted with respect to such transformation. Hence, when using the transformation (2), one has to remember that he makes the logarithmic yield component analysis, not the original one. The results obtained by using such procedures cannot be directly compared.

Therefore, let us seek some other methods. Sparnaaij and Bos (1993) proposed an interesting approach to the yield component analysis, and Kozak (2002) modified it. It works on the original scale of the variables; however, the method does not regard the components, but the characters that are a source of creating the components; Sparnaaij and Bos (1993) called them the primary characters. Of course it makes the interpretation on the components only approximate. Moreover, the method regards only the variables developing sequentially. For details see the mentioned papers.

Basing on formulas for an expected value of the product of two normal variables given by Lu (1961), Hühn (1987) gave a following formula for a coefficient of variation of the yield

$$v_y = \frac{\sqrt{v_1^2 + v_2^2 + 2\rho_{12} \cdot v_1 \cdot v_2 + \left(1 + \rho_{12}^2\right) \cdot v_1^2 \cdot v_2^2}}{1 + \rho_{12} \cdot v_1 \cdot v_2}, \tag{3}$$

where $v_y = \sigma_y \left(\mu_y\right)^{-1}$is the coefficient of variation of the yield Y, v_1,v_2 are the coefficients of variation of the components X_1, X_2, and ρ_{12} is the coefficient of correlation between them.

Hühn (1987) considered only two components, and he said that a formula for three components is quite ease to obtain (unfortunately, in such a case it would be biased, in the contrary to (3)). The interpretation of the influence of the components on the stability of the yield can be made on a basis of (3). Note that the influence of the coefficients of variation of particular components is not direct, so the interpretation is not easy; we just might conclude, that the yield component with the bigger coefficient of variation has bigger influence on the yield. First, it is obvious (see (1)), and secondly, one may think it is too poor interpretation.

Moreover, some yield components are co-related, and the relationship between the other ones can be treated rather as a cause-and-effect relationship. In the latter case the first component influences the following him components; it should be also taken into account. In this paper we consider just the first case, i.e. the case of the co-related yield components.

Afterwards, we see a need of an appropriate method of the yield component analysis. It should take into account two aspects; the first one is the mathematical form of the relationship (1); secondly, the method should enable an easy enough and direct interpretation of the importance of the components in the yield formation. An aim of the paper is to present a new approach to the problem. Moreover, two different estimators are presented (one of them is known) and they are compared using simulation studies.

2. Description of the problem and the method

Consider a following relationship between the yield and its predictors:

$$y = f(X_1 = x_1, ..., X_k = x_k) = f(x_1, ..., x_k) = \prod_{i=1}^{k} x_i, \tag{4}$$

where y is the value of the response variable Y (yield) and $x_i, i = 1, ..., k$, is the value of the i-th predictor X_i (the component).

Our aim is to obtain a parameter of the influence of the i-th component on the yield, which would be interpreted as an expected change of the function (4) value caused by increasing the value of the i-th component X_i by a unit of its standard deviation σ_i. The most appropriate and convenient from an interpretative point of view would be presenting this change of Y in units of its standard deviation (such interpretation is appreciated by researchers). Therefore, our parameter for the i-th predictor variable, say θ_i, would take a form

$$\theta_i = E\left[\frac{f(x_1, ..., x_i + \sigma_i, ..., x_k) - f(x_1, ..., x_i, ..., x_k)}{\sigma_y}\right], \quad i = 1, ..., k. \tag{5}$$

Let us introduce a change of the value of (4) after increasing $X_i, i = 1, ..., k$, by the unit of its standard deviation:

$$\begin{aligned} \Delta_i &= f(x_1, ..., x_i + \sigma_i, ..., x_k) - f(x_1, ..., x_i, ...x_k) = \\ &= (x_i + \sigma_i) \prod_{j=1, j \neq i}^{k} x_j - \prod_{j=1}^{k} x_j = \sigma_i \prod_{j=1, j \neq i}^{k} x_j, \end{aligned} \tag{6}$$

where σ_i stands for the population standard deviation of the component X_i.

If we want to present the change of Y in the units of its standard deviation σ_y, $\Delta_{i\sigma}$, we just have to divide the Δ_i by σ_y, i.e.

$$\Delta_{i\sigma} = \Delta_i \sigma_y^{-1}. \tag{7}$$

Considering $\Delta_{i\sigma}$ and (5), our parameter takes a form of an expected value of (7), i.e.

$$\begin{aligned} \theta_i &= E(\Delta_{i\sigma}) = E\left(\frac{\Delta_i}{\sigma_y}\right) = E\left(\frac{\sigma_i}{\sigma_y} \prod_{j=1, j \neq i}^{k} x_j\right) = \\ &= \frac{\sigma_i}{\sigma_y} E\left(\prod_{j=1, j \neq i}^{k} x_j\right), i = 1, ..., k. \end{aligned} \tag{8}$$

Afterwards, a final form of the parameter θ_i is

$$\theta_i = \frac{\sigma_i}{\sigma_y}\mu_{y_i}, \tag{9}$$

where μ_{y_i} is a population mean of a variable $Y_i = \prod_{j=1,j\neq i}^{k} X_j$; note that in a bi-variate case, i.e. with two predictors, $\theta_i = \frac{\sigma_i \mu_j}{\sigma_y}$, $i = 1, 2$, $j \neq i$.

Let us consider the bi-variate case first. An interpretation of the parameters $\theta_i, i = 1, 2$, is quite simple and convenient. The given θ_i informs about the expected change of the yield (in the units of its standard deviation) after increasing the i-th predictor variable by its standard deviation. It is a common interpretation in the standardized regression problems; moreover, it is much-desired interpretation by the biologists and agriculturalists (it is a very important type of interpretation, for we often need to know on which variable we should concentrate during a plants' vegetation). The value of θ_i informs about the importance of the i-th variable in determining the function value; these parameters are comparable, i.e. if θ_1 is bigger than θ_2, we conclude that the variable X_1 has the bigger influence on the function (1), i.e. on the yield, than the variable X_2.

The interpretation in the general multivariate case is the same as in the bi-variate case. It is still convenient from practical point of view. The following section contains considerations on a problem of estimation of the effects of the multiplicative variables on their product.

3. Estimation

In what follows two different estimators of the parameters θ_i are presented. They are biased; therefore, in a next section we will compare them using a Monte Carlo study.

First estimator of θ_i, say $\hat{\theta}_{1i}$, is a natural estimator for such parameter, i.e. constructed by replacing in (9) the particular parameters with their estimators:

$$\hat{\theta}_{1i} = \frac{s_i}{s_y}\hat{\mu}_{y_i}, \tag{10}$$

where $s_i = \hat{\sigma}_i$ is the estimator of σ_i, $s_y = \hat{\sigma}_y$ is the estimator of σ_y, and $\hat{\mu}_{y_i}$ is a sample mean of a variable Y_i (in the bi-variate case $\hat{\theta}_{1i} = \frac{s_i \bar{x}_j}{s_y}, i = 1, 2, j \neq i$).

Note that the estimator $\hat{\theta}_{1i}$ is biased, because

$$E\left(\hat{\theta}_{1i}\right) = E\left(\frac{s_i \hat{\mu}_{y_i}}{s_y}\right) \neq \frac{\sigma_i \mu_{y_i}}{\sigma_y}. \tag{11}$$

Second proposed estimator of θ_i, say $\hat{\theta}_{2i}$, has a following form:

$$\hat{\theta}_{2i} = \frac{s_i}{s_y} b_i, \tag{12}$$

where b_i is a least-square estimator of a partial regression coefficient for the i-th predictor variable from a linear regression model $E(Y|X_1, ..., X_k)$ (cf. Draper and Smith, 1998).

It is easy to note that the formula (12) for the estimator is similar to standardized partial regression coefficients from the linear model $E(Y \mid X_1, ..., X_k)$. Such form of the estimator makes us approximating the relationship (1) with the linear model. Although it can make the estimation biased and inappropriate (see the introduction), applications of the estimator (12) can be found in many papers (e.g. Kang *et. al.*, 1991, Rozbicki and Mądry, 1998, Jag-Shoran *et. al.*, 2000, Kumar and Kumar, 2000, and many others); that is why we decided to consider it in the paper and to compare with $\hat{\theta}_{1i}$. The properties, i.e. the bias and mean-square error of the estimators, are studied in a following section.

4. Monte Carlo studies on the estimators

A choice of the appropriate estimator should be based on the mentioned properties of the estimators, i.e. the bias and MSE (mean-square error). Analytical (even approximate) satisfactory formula cannot be found; therefore we will carry out the Monte Carlo study to investigate the empirical bias and MSE of the estimators. On the basis of the studies, conclusions and recommendations regarding the estimators can be made.

Two experiments have been carried out. The first one regarded the bi-variate and the second one the three-variate case. Different values of correlation coefficients between the components and their coefficients of variation have been considered (preliminary investigations showed that we should pay attention only to the coefficients of variation of the predictors, and not necessarily to their means and standard deviation; see also Kozak, 2004).

An artificial population consisting of 100,000 units has been created in each case of the study; such big population should approximately imitate an infinite one. For each unit values of k components were generated according to a given distribution. The predictor variables were generated with given characteristics, i.e. coefficients of variation denoted by cv (controlled by a value of their standard deviations; means were constant and equaled 10) and the coefficients of correlation between them. Values of the effect variable in the population units were evaluated as a product of the predictors in the particular unit, according to (4).

In the bi-variate case the following characteristics of predictors have been used: $cv = \{0.2, 0.5\}, r = \{-0.5, 0, 0.5\}$. It gave us eighteen mutually different combinations $(cv \times r)$, for a sequence of the variables is not important. For each artificial population the parameters $\theta_i, i = 1, 2$, have been evaluated, and for thousand times ($R = 1000$) a simple sample of size $n = 100$ have been drawn; in each such iteration the estimators were evaluated.

On the basis of the results the empirical relative bias, coefficient of variation and mean-square error have been assessed using following formulas:

$$RB\left(\bar{\hat{\theta}}_{ji}\right) = \frac{\hat{\theta}_{ji} - \theta_i}{\theta_i}, j = 1, 2, i = 1, ..., k, \tag{15}$$

$$cv\left(\hat{\theta}_{ji}\right) = \frac{\sqrt{MSE\left(\hat{\theta}_{ji}\right)}}{\bar{\hat{\theta}}_{ji}}, \tag{16}$$

$$MSE\left(\hat{\theta}_{ji}\right) = \frac{1}{(R-1)} \sum_{g=1}^{R} \left(\hat{\theta}_{jig} - \bar{\hat{\theta}}_{ji}\right)^2, \tag{17}$$

where $RB\left(\hat{\theta}_{ji}\right)$ is the relative empirical bias and $MSE\left(\hat{\theta}_{ji}\right)$ is the empirical mean-square error of $\hat{\theta}_{ji}$, $cv(\hat{\theta}_{ji})$ is the empirical coefficient of variation of $\hat{\theta}_{ji}$, θ_i is the known value of the i-th parameter, $\bar{\hat{\theta}}_{ji} = R^{-1} \sum_{g=1}^{R} \hat{\theta}_{jig}$ is the mean value of the j-th estimator ($j = 1, 2$) of the i-th parameter, and $\hat{\theta}_{jig}$ is the g-th value of $\hat{\theta}_{ji}$, $g = 1, ..., 1000$.

Results of the study, i.e. the characteristics (15) and (16), are presented in Table 1; the table contains also values of the coefficient of determination (R^2) of the population linear model approximating the relationship (1) in the particular combination.

The values of the coefficient of determination varied from 76% up to almost 99%; see also Kozak (2004). The analysis showed that both estimators were biased, and they usually overestimated the parameters' value. The relative biases of the estimators were not big (not bigger than 1%) and usually quite similar. The coefficients of variation of both estimators were also quite similar and increased when increasing the variation of the variables. This study on the bi-variate case does not allow us to conclude that any of two presented estimators is better, i.e. less biased or more efficient. Note that even for the weak linear relationship between the response variable and its predictors, the bias of the estimators based on the partial regression coefficients was still small.

A similar Monte Carlo study has been conducted for the three-variate case. Certainly, two many combinations could be considered in such a case; therefore just several

Table 1. Results of the Monte Carlo study for bi-variate case – empirical characteristics (RB – relative bias, cv – coefficient of variation) of the estimators (Est.) $\hat{\theta}_{1i}$ and $\hat{\theta}_{2i}$, and coefficient of determination of the linear model $E(Y|X_1, X_2)$

Est.	cv(X)	r_{12}	$RB(\theta)$ x 100	$cv(\theta)$	R^2
$\hat{\theta}_{1i}$	0.2, 0.2	–0.5	0.9, 0.8	8.7, 7.4	–
$\hat{\theta}_{2i}$	0.2, 0.2	–0.5	0.8, 0.8	8.1, 7.3	0.952
$\hat{\theta}_{1i}$	0.2, 0.5	–0.5	0.7, 0.1	15.1, 11.2	–
$\hat{\theta}_{2i}$	0.2, 0.5	–0.5	0.5, 0.1	17.9, 7.5	0.938
$\hat{\theta}_{1i}$	0.5, 0.5	–0.5	0.6, 0	10.4, 9.4	–
$\hat{\theta}_{2i}$	0.5, 0.5	–0.5	0.7, –0.3	11.5, 9.6	0.762
$\hat{\theta}_{1i}$	0.2, 0.2	0	0, 0.5	7.24, 7.30	–
$\hat{\theta}_{2i}$	0.2, 0.2	0	0, 0.4	7.2, 7.3	0.981
$\hat{\theta}_{1i}$	0.2, 0.5	0	0.8, 0.1	12.1, 10.6	–
$\hat{\theta}_{2i}$	0.2, 0.5	0	0.6, 0.1	15.0, 9.8	0.967
$\hat{\theta}_{1i}$	0.5, 0.5	0	0.9, 0.9	8.1, 8.7	–
$\hat{\theta}_{2i}$	0.5, 0.5	0	0.2, 0.4	8.6, 8.0	0.888
$\hat{\theta}_{1i}$	0.2, 0.2	0.5	0.2, 0.3	6.0, 5.5	–
$\hat{\theta}_{2i}$	0.2, 0.2	0.5	0.3, 0.2	6.5, 5.5	0.984
$\hat{\theta}_{1i}$	0.2, 0.5	0.5	0.6, 0.2	7.7, 9.3	–
$\hat{\theta}_{2i}$	0.2, 0.5	0.5	0.9, 0	9.7, 8.2	0.969
$\hat{\theta}_{1i}$	0.5, 0.5	0.5	0.4, 0.1	7.2, 7.4	–
$\hat{\theta}_{2i}$	0.5, 0.5	0.5	0.7, –0.2	8.1, 8.4	0.905

of them have been chosen to be presented in the paper – they pretty well characterized the problem. Results and a design of the experiment are presented in Table 2.

In this case we got more interesting results. Note that in the three-variate case we might get very small value of coefficients of the linear determination (in two cases we got almost zero – see table 2). It influenced the values of the coefficients of variation of the estimators $\hat{\theta}_{2i}$, which employ the linear regression analysis. The variation of the estimators $\hat{\theta}_{2i}$ was almost always worse than the variation of $\hat{\theta}_{1i}$. The biases of both estimators were still similar, although in some combinations the bias of $\hat{\theta}_{1i}$ was smaller. The results of the investigation show that in the three-variate case the estimator $\hat{\theta}_{1i}$ is markedly better when the determination of the linear approximation of the relationship (1) is small.

Table 2. Results of the Monte Carlo study for three-variate case – empirical characteristics (RB – relative bias, cv – coefficient of variation) of the estimators (Est.) $\hat{\theta}_{1i}$ and $\hat{\theta}_{2i}$, and coefficient of determination of the linear model $E(Y|X_1, X_2, X_3)$

Est.	$cv(X)$	$r = \{r_{12}, r_{13}, r_{23}\}$	$RB(\theta) \times 100$	$cv(\theta)$	R^2
$\hat{\theta}_{1i}$	0.5, 0.5, 0.5	-0.5, -0.5, -0.5	4.2, 3.8, 4.5	25.1, 32.0, 17.7	–
$\hat{\theta}_{2i}$	0.5, 0.5, 0.5	-0.5, -0.5, -0.5	6.3, 5.7, 7.3	57.5, 57.3, 33.1	0.0335
$\hat{\theta}_{1i}$	0.5, 0.5, 0.5	0.5, 0.5, 0.5	2.0, 2.0, 2.7	10.5, 8.8, 10.5	–
$\hat{\theta}_{2i}$	0.5, 0.5, 0.5	0.5, 0.5, 0.5	2.5, 0.7, 0.6	14.5, 13.5, 17.7	0.777
$\hat{\theta}_{1i}$	0.5, 0.5, 0.5	0.5, 0, 0.5	1.0, 1.0, 1.4	16.1, 14.6, 13.2	–
$\hat{\theta}_{2i}$	0.5, 0.5, 0.5	0.5, 0, 0.5	0, 0.5, 1.9	17.7, 14.7, 16.0	0.570
$\hat{\theta}_{1i}$	0.2, 0.2, 0.2	-0.5, -0.5, -0.5	2.2, 2.0, 1.5	15.3, 12.0, 10.7	–
$\hat{\theta}_{2i}$	0.2, 0.2, 0.2	-0.5, -0.5, -0.5	-1.1, -0.9, -1.7	34.4, 49.7, 33.1	0.059
$\hat{\theta}_{1i}$	0.2, 0.2, 0.2	0.5, 0.5, 0.5	0.6, 0.6, 0.6	6.6, 7.9, 7.9	–
$\hat{\theta}_{2i}$	0.2, 0.2, 0.2	0.5, 0.5, 0.5	0.3, 0.4, 0.4	8.5, 10.4, 10.2	0.950
$\hat{\theta}_{1i}$	0.5, 0.2, 0.5	-0.5, -0.5, -0.5	1.3, 1.8, 1.4	11.0, 6.1, 12.5	–
$\hat{\theta}_{2i}$	0.5, 0.2, 0.5	-0.5, -0.5, -0.5	-2.2, -5.9, -2.6	25.9, 27.4, 52.4	0.548
$\hat{\theta}_{1i}$	0.5, 0.2, 0.2	-0.5, -0.5, -0.5	0.7, 1.7, 1.8	7.0, 6.7, 6.8	–
$\hat{\theta}_{2i}$	0.5, 0.2, 0.2	-0.5, -0.5, -0.5	0.5, 1.7, 2.3	11.5, 12.6, 17.6	0.835
$\hat{\theta}_{1i}$	0.5, 0.2, 0.2	0.5, -0.5, -0.5	0.1, 0.4, 0.3	4.4, 3.6, 4.9	–
$\hat{\theta}_{2i}$	0.5, 0.2, 0.2	0.5, -0.5, -0.5	0.3, -0.1, -0.3	4.1, 4.5, 6.3	0.955

All computation has been made using the R language; cf. R Development Core Team (2004). The multivariate normal distributions with a specified covariance matrix and vector of means were generated using a function *mvrnorm* from a package *MASS*.

5. Discussion

The considerations presented in the paper are correct when the predictor variables, i.e. the components, satisfy some conditions. First, the variables $X_1, ..., X_k$ should follow a k-variate normal distribution (Lu, 1961).

In almost all applications at least one component is a discontinuous variable. Therefore we should confine to treat it as the continuous one and have to assume its normal distribution. Investigations show that the histogram based on values of such variable is a good approximation of a smooth density of the normal distribution (cf. Kozak, 2002).

Moreover, some of the variables $X_1, ..., X_k$ do not follow the normal distribution, and the whole set of k variables does not follow the k-variate normal distribution. Again, investigations show that we can assume the approximate univariate and multivariate normal distribution of the studied variables.

The above assumptions should be enough to carry out the appropriate yield component analysis. Certainly, the analysis is always approximate for two reasons; first is mentioned above (regarding the assumptions), and the second one results from the bias of both types of the estimators.

In conclusion, the results of the simulation studies showed that in the bi-variate case both estimators have similar properties (bias and mean-square error), but in the three-variate case $\hat{\theta}_{1i}$, the newly proposed estimator, has been chosen as the better one (especially in a case of big values of the coefficients of variation of the predictors and negative correlation between them). After connecting this conclusion with the simple form of $\hat{\theta}_{1i}$, we can recommend it as the appropriate estimator of the influence of the co-related yield components (or, in general, the multiplicative predictors) on the yield (response variable).

References

Draper N.R., Smith H. (1998). *Applied Regression Analysis.* John Wiley & Sons, New York.

Eaton G.W., Kyte T.R. (1978). Yield component analysis in strawberry. *Journal of the American Society for Horticultural Science* **103**, 578-583.

Fraser J., Eaton G.W. (1983). Applications of yield component analysis to crop research. *Field Crop Abstracts* **36**, 787-796.

Hardwick R.C., Andrews D.J. (1980). Genotypic and environmental variation in crop yield. A method of estimating the interdependence of the components of yield. *Euphytica* **26**, 177-188.

Hűhn M. (1987). Stability analysis of winter-rape (*Brassica napus* L.) by using plant density and mean yield per plant. *J. Agronomy & Crop Science* **159**, 73-81.

Jag Shoran Hariprasad A.S., Lakshmi Kant Mani V.P., Chauhan V.S., Shoran J., Kant L. (2000). Association and contribution of yield attributes to seed yield in wheat under varying environments in north western hills. *Annals of Agricultural Research* **21**, 274-278.

Kang M. S., Tai P., Miller J. (1991). Genetic and phenotypic path analyses in sugarcane: artificially created relationships. *Crop Science* **31**, 1684-1686.

Kozak M. (2002). *Yield component analysis.* Doctoral Thesis, Warsaw Agricultural University, Department of Mathematical Statistics and Experimentation (in Polish).

Kozak M. (2004). On usefulness of linear model in yield component analysis. *Biometrical Colloquium* **34**, 117-126 (in Polish).

Kozak M., Mądry W. (2004). Statistical analysis of multiplicative crop yield components – background of yield modeling. *Postępy Nauk Rolniczych* 5, 13-25 (in Polish).

Kumar M.V.N., Kumar S.S. 2000. Studies on character association and path coefficients for grain yield and oil content in maize (*Zea mays* L.). *Annals of Agricultural Research* **21**, 73-78.

Lu K.H. (1961). The means and variances of the product of two or three normal variables. *Biometrics* **17**, 172.

Piepho H.P. (1995). A simple procedure for yield component analysis. *Euphytica* **84**, 43-48.

R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; URL http://www.R-project.org.

Rozbicki J., Mądry W. (1998). The relationship between winter triticale grain yield, its components and agronomical and morphological traits in variable cultivation and weather conditions. *Biuletyn IHAR* **205/206**, 295-204 (in Polish).

Sparnaaij L.D., Bos I. (1993). Component analysis of complex characters in plant breeding. I. Proposed method for quantifying the relative contribution of individual components to variation of the complex character. *Euphytica* **70**, 225-235.

Received 14 July 2004

Nowa koncepcja analizy komponentów plonu

STRESZCZENIE

W artykule przedstawione są rozważania na temat wpływu cech ilorazowych rozwijających się w tym samym czasie, czyli współzależnych, na cechę będącą ich iloczynem. W agronomii tak postawiony problem nazywany jest "analizą składowych plonu". Autor proponuje nowe podejście do tego problemu oraz za pomocą badań symulacyjnych porównuje dwa różne estymatory. Na podstawie tych badań zaproponowana została postać odpowiedniego (choć obciążonego) estymatora wpływu składowych na plon.

SŁOWA KLUCZOWE: iloczynowe zmienne objaśniające, składowe plonu, badanie Monte Carlo.