DETERMINING THE RELATIONSHIP BETWEEN GRAIN YIELD AND ITS ATTRIBUTING TRAITS IN BARLEY USING STRUCTURAL EQUATION MODELLING

RAM NIWAS*, VINAY KUMAR, O P SHEORAN AND YOGENDER KUMAR[1]

Department of Mathematics and Statistics
[1]Department of Genetics and Plant Breeding
CCS Haryana Agricultural University, Hisar-125 004 (Haryana), India
(e-mail : rniwas@hau.ac.in)

(Received : 5 November 2020; Accepted : 22 December 2020)

SUMMARY

Development of barley cultivars that achieve high yields despite the short growing season is essential for increasing barley production in India. The present study focuses on characterizing the causal relationship between grain yield and various components characteristics using the structural equation modelling with latent variables in barley crop. The data on grain yield and its attributing characters on 87 genotypes of barley (*Hordeum vulgare* L.) were taken for studying the relationships between them. A structural equation model that characterize the complex phenomenon and biological processes with less number of assumptions was used to study and describe the causal relationship between measured variables such as crop characteristics, crop phenology, canopy traits, yield and its components along with the latent variables.

Key words : Structural equation models, maximum likelihood estimation, latent variables

Barley (*Hordeum vulgare* L.) is popularly known as "*Jau*" in Hindi and one of the major cereal grain crops after Rice, Wheat and Maize. It is cultivated in a wide variety of habitats, such as rainfed, irrigated, dry land, saline / alkaline soil, marginal fields, areas vulnerable to drought, hill regions, and marginal / coastal areas vulnerable to flooding in the country. The changing climate scenario in the country has made it a viable crop for the near future in terms of temperature, rainfall and crop span (Raikwar, 2015). During the 2019-20 crop season in India, the region under barley was 0.62 million hectares with production and average productivity of 1.59 million tonnes and 25.73 q/ha, respectively. On 12,200 hectares, Haryana state produced an output average of 44,000 tonnes. In Punjab (37.67 kg / ha), the average crop productivity is highest in barley, followed by Haryana (3607 kg / ha), Uttar Pradesh (2956 kg / ha and) Rajasthan (2884 kg / ha) (ICAR-IIWBR, 2020).

Development and growth of barley crop, is a complex non-linear process which includes many factors. Grain yield is a trait resulting from morpho-physiological processes during its growth and development stage. The interaction of both direct (genetic, physiological and biological) and indirect (habitat and cultivation *etc.*) factors influences the grain yield. These direct and indirect factors play an important role in estimating the grain yield per plant at a given level (Gozdowski *et al.*, 2007). Statistical analysis of yield and its attributing characters allows researchers to understand the biological mechanisms which are important for any breeding programme. The relationships between yield and its components have been analysed through several statistical methods *viz.* linear multiple regression, path analysis, sequential yield component analysis, principal component analysis and factor analysis (Kumar *et al.*, 2018). The correlation between crop yield and yielding factors can also be analyzed using structural equation models (Kozak *et al.*, 2007), which are regarded as an important statistical tool designed to study and describe cause and effect relationships.

Structural equation modelling characterizes the complex phenomenon and biological processes with less number of assumptions, is also regarded as an important statistical procedure to study and describe the causal relationship between measured variables (crop characteristics, crop phenology, canopy traits, yield and its components) along with the latent variables. Lamb *et al.* (2011) initially used this technique for crop analysis and compared with the "first generation" multivariate statistical method PCA and cluster analysis (CA). Mankowski *et al.* (2016) employed structural equation modelling to assess the relationship between grain yield per plant and its components in double haploid spring barley lines (*Hordeum vulgare* L.). Further, Zheng *et al.* (2017) studied the application of structural equation modelling in analyzing the relationship between agronomic characters and yield of winter wheat. Nazmi (2013) reported structural equation modelling to study the relationships between soil properties and yield components of wheat and Zhang *et al.* (2014) used the same approach for Canadian flax (*Linum usitatissimum* L.).

The aim of the present study was to characterize the causal relationship between grain yield and various components using the structural equation modelling in barley genotypes.

MATERIALS AND METHODS

The experimental materials consisted of eighty seven genotypes of barley (*Hordeum vulgare* L.) evaluated at the Barley Research Area of the Department of Genetics and Plant Breeding at CCS Haryana Agricultural University, Hisar, Haryana. The quantitative traits of barley were collected from the *rabi* season 2016–2017 include days to heading, days to maturity, plant height (cm), spike length (cm), number of tillers per meter, number of grains per spike, 1000 grain weight (g), grain yield (kg/plot), biological yield (kg/plot), harvest index (%). The structural equation model has been developed to define and characterise the relationships between yield traits and grain yield in barley crop.

Principal component method of factor analysis was employed to identify the factors which contribute to the yield and its related parameters. The latent variables were obtained by empirical grouping of the exogenous and endogenous variables based on significant factor loading from exploratory factor analysis. Each implied dimension (factor) suggested by factor analysis were then tested using maximum likelihood confirmatory factor analysis and subsequently used for conceptualizing and development of the structural equation model. A recursive structural equation model with latent variables including more complex relationships among the analysed variables was developed which fit well to the data.

The measurement model for each dimension in the form of standard factor analytical model is given by

$$y= \Lambda y \ \eta + \varepsilon \qquad (1)$$

for latent endogenous variables with $E(\varepsilon\varepsilon') = \Theta_\varepsilon$ and

$$x = \Lambda_x \xi + \delta \qquad (2)$$

for latent exogenous variables with $E(\delta\delta') = \Theta_\delta$ We also define $E(\varepsilon\delta) = \Theta_{\delta\varepsilon}$ and $E(\xi\xi') = \Phi$, where

y is a *p x 1* vector of observed indicators of the dependent (endogenous) latent variable η

x is a *q x 1* vector of observed indicators of the independent (exogenous) latent variables ξ

η is a *m x 1* random vector of latent dependent or endogenous variables

ξ is a *n x 1* random vector of latent independent or exogenous variables

ε is a *p x 1* vector of measurement error in y

δ is a *q x 1* vector of measurement error in x

Λy is a *p x m* matrix of coefficients of regression of y on η and

Λ_x is a *q x n* matrix of coefficients of regression of x on ξ

The implied covariance/correlation matrix $\Sigma(\theta)$ is given by E(xx') or E(yy') for measurement models with the assumptions

$$E(x) = E(\delta) = 0 \text{ and } E(\xi\delta') = E(\delta\xi') = 0, \text{ then}$$

$$\Sigma(\theta) \ \Lambda_x \Phi \Lambda'_x + \Theta_\delta \qquad (3)$$

Then the structural part of the model is given by

$$\eta=B\eta + \Gamma\xi + \zeta \qquad (4)$$

We also define $E\ (\zeta\zeta) = \psi$, where

B is a *m x m* coefficient matrix that relates endogenous variables to each other

Γ is a *m x n* coefficient matrix that relates endogenous variables to exogenous variables and ξ is a *m x 1* vector of errors (residuals).

In the present study, initially an attempt was made for estimating and fitting separate measurement models and then, in the second stage, a pooled model (with all measurement models together) was fitted and tested for goodness of fit. The model parameters were estimated by the maximum likelihood estimation method and the adequacy of the assumed model has been evaluated by considering multiple criteria.

RESULTS AND DISCUSSION

For the purpose of developing structural equation models, ten yield and its components were identified and the description of these attributes has been presented in Table 1.

TABLE 1
Codes and description of the variables of barley genotypes

Code	Description	Symbols
DH	Days to heading	x_1
DM	Days to maturity	x_2
PH	Plant height (cm)	x_3
SL	Spike length (cm)	x_4
TM	Number of tillers per meter	x_5
GS	No. of grains per spike	x_6
TGW	1000 grain weight (g)	y_1
GYP	Grain yield (kg/plot)	y_2
BYP	Biological yield (kg/plot)	y_3
HI	Harvest index (%)	y_4

The structural equation model of the data has been hypothesized on the basis of the three latent variables as suggested by the preliminary exploratory factor analysis and then further improved by freeing the elements of residual matrices and adding or deleting the attribute(s) to the latent variables as suggested by the largest modification indices. The model parameters have been re-estimated after every improvement. Finally, the model which converged to the optimum solution with acceptable fit statistics has been obtained. The three factors solutions indicate that there does not appear a "simple structure" in the data. The complex relationships among latent variables and their error terms of yield attributing characters have been identified and tested through structural equation modelling. It can be revealed from the factor analysis that three factors have been identified crucial for grain yield and other attributing characters. Out of three latent variables, phenological parameters ($?_1$) and grain parameters ($?_2$) have been taken as exogenous whereas grain parameters ($?_1$) as endogenous latent variables. Initially, a simple recursive structural equation model with these three latent variables has been formulated on the basis of structure suggested by exploratory factor analysis in barley grain yield. Next, the model parameters have been estimated by maximum likelihood method and finally the model is tested for goodness of fit.

The exogenous measurement model using (2) has been specified by the matrix equation (5) reproduced below:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ y_1 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} & 0 \\ \lambda_{21}^{(x)} & 0 \\ \lambda_{31}^{(x)} & 0 \\ 0 & \lambda_{42}^{(x)} \\ 0 & \lambda_{52}^{(x)} \\ 0 & \lambda_{62}^{(x)} \\ 0 & \lambda_{72}^{(y)} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \end{pmatrix} \tag{5}$$

It can be seen from (5) that, three exogenous latent variables ξ_1 and ξ_2 have been measured by various attributes of grain yield. The latent variable ξ_1 (phenological parameters) has been measured by days to heading (DH), days to maturity (DM) and plant height (PH) with positive factor loading. The second exogenous latent variable ξ_2 (grain parameters) has positive significant loading on indicator variables like spike length (SL), number of tillers per meter (TM), number of grains per spike (GS) and 1000 grain weight (TGW). While the endogenous measurement model using (1) has been formulated as:

$$\begin{pmatrix} y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} \\ \lambda_{21}^{(y)} \\ \lambda_{31}^{(y)} \end{pmatrix} (\eta_1) + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \tag{6}$$

The endogenous measurement model has only one latent variable η_1. The latent variable η_1 (yield parameters) has been measured by grain yield per plot (GYP), biological yield per plot (BYP) and harvest index (HI). The structural equation model has been formulated by using (4) and is as given below:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{21}\xi_2 + \zeta_1 \tag{7}$$

where the latent error terms have the following covariance matrix

$$\psi = \text{var}\ (\zeta_1) \tag{8}$$

The path diagrams for final models which fit well to data for establishing the relationship between yield and its components for barley crop with estimated coefficients has been presented in Fig. 1.

The measured variables DH, DM and PH are termed as phenological parameter and regression weights of these parameters are positive towards this latent variable (ξ_1). A positive correlation was observed among the phenological traits which were also reported by Kumar *et al.* (2013). These traits also showed a

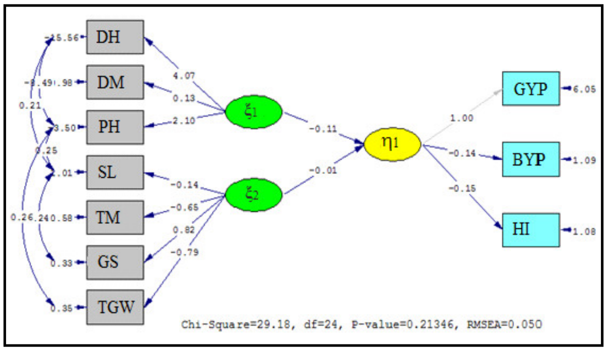

Fig. 1. Path diagram for model of barley to establish relationship between yield and its components with coefficient estimates.

negative direct effect towards grain yield. These results has also supported by Kumar *et al*. (2013) and Yadav *et al*. (2014). The justification that path coefficients i.e. γ_{11} is -0.11, which may be due to the reason as the vegetative phase of the crop had more food reserves diverted towards the plant height. The grain parameters *viz*., ear length (Kumar *et al*., 2018), grains per spike (Yadav *et al*., 2015), tillers per meter and 1000 grain weight (Kumar *et al*., 2013) also exhibited negative direct effect to the grain yield in barley. The figure 1 depicted that the measured variables SL, TM, GS and TGW were formed as grain parameter, also (γ_{12}) shows the negligible value of path coefficient as -0.01, might be due to fuzzy type or weak seed which are more in number but their weight is less. The measurement variables TM, GS and TGW have significant loading on the latent variable (ξ_2). The positive correlation exists between latent variables i.e. phenological parameter (ξ_1) and grain parameter (ξ_2). The yield parameters exhibited positive significant correlation among themselves. These results are in agreement with the results of Kumar *et al.* (2018).

The model parameters have been estimated with the maximum likelihood technique using two different residual correlation matrix specifications. Initially, the restricted model has been estimated by setting the off-diagonal elements of Θ_δ and Θ_ε to zero. The Chi-square value of the restricted model ($\chi^2_{(s-t)df}$ = (N-1)F[S, Σ (θ)] has been obtained as 259.53 (d.f = 45) with GFI {Goodness of fit Index, (GFI = 1-F_t/F_n = 1 – χ^2_t/χ^2_n)} as 0.94 and SRMR {Standardized Root Mean Square Residual, ($S_{ij} - \sigma_{ij}/(S_iS_j)=r_{ij} - \sigma_{ij}/(S_iS_j)$} as 0.089 indicating that the model does not fit well to the data.

These models have been further improved by relaxing the zero restrictions of off-diagonal elements in the correlation matrices of error terms. The error terms which are significantly correlated have been identified on the basis of standardized residual and modification indices (Sorbom, 1989). On removing the zero restriction on several off-diagonal elements of Θ_δ, Θ_ε and $\Theta_{\delta\varepsilon}$ as given in (9), (10) and (11), respectively produced a Chi-square value as 29.18 (d.f. = 24), GFI = 0.98 and SRMR=0.09 indicating that the fit is good to establish relationship between yield and its components for barley crop.

TABLE 2

Maximum likelihood estimates for structural equation model for barley crop in Haryana

Parameter	Estimate (S.E.)	Standardized Estimates	Parameter	Estimate (S.E.)	Standardized Estimates
$\lambda^{(x)}_{11}$	4.07 (8.45)	4.07	θ^δ_{33}	-3.50 (18.32)	-3.77
$\lambda^{(x)}_{21}$	0.13 (0.27)	0.13	θ^δ_{44}	1.01(0.15)	0.98
$\lambda^{(x)}_{31}$	2.10 (4.36)	2.18	θ^δ_{55}	0.58 (0.10)	0.58
$\lambda^{(x)}_{42}$	-0.14 (0.13)	-0.14	θ^δ_{66}	0.33 (0.10)	0.33
$\lambda^{(x)}_{52}$	-0.65 (0.10)	-0.66	θ^δ_{77}	0.35 (0.10)	0.36
$\lambda^{(x)}_{62}$	0.82 (0.10)	0.82	θ^δ_{31}	-8.49 (35.39)	-8.81
$\lambda^{(x)}_{72}$	-0.79 (0.10)	-0.80	θ^δ_{41}	0.21 (0.09)	0.20
$\lambda^{(x)}_{11}$	1.00 (0.00)	0.93	θ^δ_{43}	0.09 (0.25)	0.25
$\lambda^{(x)}_{21}$	-0.14 (0.10)	-0.14	θ^δ_{64}	0.24 (0.08)	0.23
$\lambda^{(x)}_{31}$	-0.15 (0.12)	-0.16	θ^δ_{73}	0.26 (0.07)	0.28
γ_{11}	-0.11 (0.23)	-0.11	θ^δ_{11}	6.05 (4.20)	5.18
γ_{12}	-0.01 (0.02)	-0.01	θ^δ_{22}	1.09 (0.19)	1.09
ϕ_{21}	0.03 (0.07)	-	θ^δ_{33}	1.08 (0.19)	1.12
Var (ζ_1)	-4.90 (4.11)	-4.90	θ^δ_{11}	0.55 (0.25)	0.51
θ^δ_{11}	-15.56 (68.74)	-15.55	θ^δ_{22}	-0.03 (0.09)	-0.03
θ^δ_{22}	0.98 (0.17)	0.98	θ^δ_{33}	-0.36 (0.11)	-0.38
$\chi^2_{(df=24)}$			29.18	(P=0.21346)	
GFI			0.98		
SRMR			0.09		

$$\Theta_{\delta}=\begin{pmatrix} \theta_{11}^{\delta} & & & & & & \\ 0 & \theta_{22}^{\delta} & & & & & \\ \theta_{31}^{\delta} & 0 & \theta_{33}^{\delta} & & & & \\ \theta_{41}^{\delta} & 0 & \theta_{43}^{\delta} & \theta_{44}^{\delta} & & & \\ 0 & 0 & 0 & 0 & \theta_{55}^{\delta} & & \\ 0 & 0 & 0 & \theta_{64}^{\delta} & 0 & \theta_{66}^{\delta} & \\ 0 & 0 & \theta_{73}^{\delta} & 0 & 0 & 0 & \theta_{77}^{\delta} \end{pmatrix} \qquad (9)$$

$$\Theta_{\varepsilon}=\begin{pmatrix} \theta_{11}^{\varepsilon} & & \\ & \theta_{22}^{\varepsilon} & \\ & & \theta_{33}^{\varepsilon} \end{pmatrix} \qquad (10)$$

$$\Theta_{\delta\varepsilon}=\begin{pmatrix} \theta_{11}^{\delta\varepsilon} & 0 & 0 \\ 0 & \theta_{22}^{\delta\varepsilon} & 0 \\ 0 & 0 & \theta_{33}^{\delta\varepsilon} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad (11)$$

The final model parameters along with their estimates and standard error have been presented in Table 2. It can be revealed from Table 2 that the estimated factor loadings are statistically significant at 5 percent level of significance.

The estimates in Table 2 indicated that the exogenous variable ξ_1 has a negative influence on the endogenous latent variable η_1 and exogenous latent variables ξ_2 indicates a negative influence on $\eta_{1.}$ Also the exogenous latent variables ξ_1 and ξ_2 have positive correlation.

CONCLUSION

The structural equation model with latent variables is more efficient than the ordinary path analysis in explaining the relationships between yield and its attributing traits as it is supported by the awareness of growth physiology, yield and development of barley crop. Beside these, it facilitate to study the internal recursive relationships (correlations) between the exogenous and endogenous variables as well as the latent variables. Earlier also the same approach was applied to different crops like pearl millet, sorghum grain, winter wheat, wild oat, lowland rice and grass pea to study the complex relationships between various traits. Further scope of work includes the SEM model of yield and its attributing traits with weather, soil parameters. This can also extend and compare the results with Bayesian approach.

REFERENCES

Gozdowski, D., M. Kozak, M.S. Kang and Z. Wyszy Ski, 2007 : Dependence of grain weight of spring barley genotypes on trials of individual stems. *J. Crop Improve.*, **20**(1/2): 223-233.

ICAR-IIWBR, 2020 : Director's Report of AICRP on Wheat and Barley 2019-2020, Ed: G.P. Singh. ICAR-Indian Institute of Wheat and Barley Research, Karnal, Haryana, India. pp.76.

Kozak, M., P.K. Singh, R.M. Verma and D.K. Hore, 2007 : Causal mechanism for determination of grain yield and milling quality of lowland rice. *Field Crops Res.,* **102**: 178-184.

Kumar, Y., N. Kumar, O.P. Bishnoi and S. Devi, 2018 : Estimation of genetic parameters and character association in barley (*Hordeum vulgare* L.) under irrigated condition. *Forage Res.,* **44** (1): 56-59.

Kumar, Y., R.A.S. Lamba, S.R. Verma and R. Niwas, 2013 : Genetic variability for yield and its components in barley (*Hordeum vulgare* L.). *Forage Res.,* **39** (2): 67-70.

Kumar, Y., R. Niwas, O.P. Bishnoi and N. Kumar, 2018 : Principal component and factor analysis in six rowed barley (*Hordeum vulgare* L.) genotypes. *Forage Res.*, **44** (1): 38-42.

Lamb, E.G., S.J. Shirtliffe and W.E. May, 2011 : Structural equation modelling in the plant sciences: An example using yield components in oat. *Canadian J. Plant Sci.,* **91**: 603-619.

Laxman, V. Singh, Y.P.S Solanki and A.S. Redhu, 2014 : Phenological development, grain growth rate and yield relationships in wheat cultivars under late sown condition. *Indian J. Plant Physiol.,* **19** (3) : 222-229.

Mankowski, D. R., J. Kozdoj and M. Janaszek-Mankowska, 2016 : Structural equation model as a tool to assess the relationship between grain yield per plant and yield components in doubled haploid spring barley lines (*Hordeum vulgare L.*). *Plant Breed. Seed Sci.*, **73**: 63-77.

Nazmi, L., 2013 : Modelling for relationships between soil properties and yield components of wheat using multiple linear regression and structural equation modelling. *Adv. Environmen. Biol.*, **7**(2): 235-242.

Raikwar, R. S., 2015 : Generation mean analysis of grain yield and its related traits in barley (*Hordeum vulgare* L.). *Electron. J. Plant Breed.,* **6**(1): 37-42.

Yadav, N., S.R. Verma and S. Singh, 2015 : Studies of genetic variability and trait association for grain yield and its components in two rowed and six rowed barley *(Hordeum vulgare* L.). *Bioinfolet,* **12** (2 B): 521-524.

Zhang, T., E. G. Lamb, B. Soto-Cerda, S. Duguid, S. Cloutier, G. Rowland, A. Diederichsen and H. M. Booker, 2014 : Structural equation modelling of the Canadian flax (*Linum usitatissimum* L.) core collection of multiple phenotypic traits. *Can. J. Plant Sci.*, **94**:1-8.

Zheng, Lifei, Yifei Shang, Xuejun Li, Hao Feng and Yongsheng Wei, 2017 : Structural equation model for analyzing relationship between d and agronomic traits in winter wheat. *Acta Agron. J.*, **43** (9): 1395-1400.