LOGISTIC REGRESSION MODELS FOR PRE-HARVEST WHEAT YIELD ESTIMATION IN WESTERN ZONE OF HARYANA

BY

SUDESH RANI
2014BS20M

Thesis submitted to the Chaudhary Charan Singh Haryana Agricultural University, Hisar in the partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

IN

STATISTICS

COLLEGE OF BASIC SCIENCES & HUMANITIES
CCS Haryana Agricultural University
HISAR - 125004

2017

CERTIFICATE–I

This is to certify that this dissertation entitled, **"Logistic regression models for pre-harvest wheat yield estimation in western zone of Haryana"** submitted for the degree of **Master of Science** in the subject of **Statistics** to the Chaudhary Charan Singh Haryana Agricultural University, Hisar is bonafide research work carried out by **Ms. Sudesh Rani** under my supervision and guidance that no part of this dissertation has been submitted for any other degree.

The assistance and help received during the course of investigation have been fully acknowledged.

Dr. (Mrs.) Urmil Verma
Major Advisor
Principal Scientist (Statistics)
Department of Math, Stat and Physics
CCS Haryana Agricultural University
Hisar-125004

CERTIFICATE –II

This is to certify that this dissertation entitled, **"Logistic regression models for pre-harvest wheat yield estimation in western zone of Haryana"** submitted by **Ms. Sudesh Rani** to the Chaudhary Charan Singh Haryana Agricultural University, Hisar, in partial fulfillment of the requirements for the degree of **Master of Science** in the subject of **Statistics**, has been approved by the student's Advisory Committee after an oral examination on the same.

MAJOR ADVISOR **EXTERNAL EXAMINER**

HEAD OF THE DEPARTMENT

DEAN, POSTGRADUATE STUDIES

Acknowledgements

*Completion of this master dissertation was possible with the support of several people. I would like to express my sincere gratitude to all of them. First of all, I am extremely grateful to my major advisor, **Dr. (Mrs.) Urmil Verma,** Principal Scientist (Statistics), Department of Math, Stat and Physics, CCS HAU, Hisar for her best guidance, valuable suggestions with an encouraging attitude. I consider it as a great opportunity to do my master programme under her guidance and to learn from her research expertise. My heart rises in gratitude with regards for her valuable guidance, close supervision, concrete suggestions and preparation of this manuscript. Her infinite capacity for hard work, remarkable patience and affectionate behavior will always be an illuminating light to lead me in all my pursuits in future too. In my life, she is as the lamplight in the dark phases of life.*

*I would also like to thank my advisory committee members, **Dr. D. R. Aneja**, Senior Scientist, Deptt. of Math, Stat and Physics, **Dr. R. K. Grover**, Assoc. Director O/o DHRM, **Dr. (Mrs) Manju Tonk** , Professor (Math), Department of Math, Stat and Physics and **Dr. H.R. Singal**, Professor (Biochemistry), Deptt. of Chemistry & Biochemistry for the interest taken to give me the light perspective of my research and valuable guidance for my future ventures. I owe my sincere thanks to **Dr. (Mrs.) Veena Manocha,** Prof. and Head , Deptt. Of Mathematics, statistics and Physics, **Dr. B. K. Hooda**, Professor (Stat) and **Dr. lajpat Rai**, Ex-Prof. & Head (retd.) Department of Math, Stat and Physics for providing necessary and sufficient facilities during my work.*

*My deepest gratitude goes to my parents, **Sh.Raj Kumar Nain** and **Smt.Birmati**, my sister Mukesh, my brother Vikash, sister-in-law Sushila and niece Kashish, Varnya, Mannat and Anirudh for their relentless inspiration, encouragement and moral support in every step of mine.*

I extend my sincere thanks and gratitude to my senior Salinder Malik, Suman Verma Nitin Tanwar, Poonam Godara, Ravita, Sudha Bishnoi, Sarita Hooda, Pinki Dagar and Mujahid Khan and juniors for their cooperation, inspiration and valuable suggestions when needed.

No words can appreciate the all round help rendered to me by my batchmates and colleagues, Ekta Hooda, Mohit Nain, Rekha and Bhushan for their valuable support. I am also thankful to my friends Sweety, Pooja, Neha, Pooja, Tanvi , Sumit and Sushila for their cheerful encouragement, help, suggestions and generous co-operation.

My sincere thanks also goes to CCS Haryana Agricultural University, Hisar for providing me an opportunity of higher studies that may be highly helpful in future.

Dated: December, 2016 **Sudesh Rani**

Hisar

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER–I

INTRODUCTION

Timely and effective pre-harvest forecast of crop yield is important for advance planning, formulation and implementation of policies related to the crop procurement, distribution, price structure and import-export decisions etc. These are also useful to farmers to decide in advance their future prospects and course of action. The yield of any crop is affected by technological change and weather variability. It can be assumed that the technological factors will increase crop yield smoothly through time and therefore, year or some other parameters of time can be used to study the overall effect of technology on crop yield. Weather variability both within and between seasons is major uncontrollable source of variability in yields. Therefore, model based on weather and year as explanatory variables can be effectively used for crop yield forecasting.

Weather variables affect the crop differently during different stages of development. Thus, the extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season which, as such, calls for the necessity of dividing the whole crop season into finer intervals and studying crop weather relationships in these intervals. However, doing so increases the number of variables in the model and in turn, a large number of parameters are to be evaluated from the data, however, sufficient number of observations may not be available for precise estimation of these parameters. Thus, a technique based on relatively smaller number of manageable parameters and at the same time taking care of entire weather distribution may solve the purpose.

1.1 Significance of the study

India has one of the best systems in the world to collect, collate and compile data on crop production. These are based on official feedback of area and yield received from the states, which in turn gets from the districts and so on. Area is based on complete enumeration by revenue agencies and yield by crop cutting experiments.The official forecasts (advance estimates) of major cereal and commercial crops are issued by the Directorate of Economics and Statistics, Ministry of Agriculture, New Delhi. However, the final estimates are given a few months after the actual harvest of the crop. Thus, one of the limitations of Department of Agriculture (DOA) yield estimates is timeliness and quality of the statistics. Hence, there is always a considerable scope of improvement in the conventional system.

Crop yield models are abstract presentation of interaction of the crop with its environment and can range from simple correlation of yield with a finite number of variables to the complex statistical models with predictive end. No satisfactory model which has universal validity exists today. Various statistical approaches are in vogue for arriving at crop forecasts. Every approach has its own advantages and limitations. The Box-Jenkins (1976) methodology is a powerful tool for time-series analysis, when the time-sequenced observations in a data series may be statistically dependent or related to each other.

Climate change is a major concern today, and the researchers are engaged in understanding its impact on growth and yield of crops, and also identifying suitable management options to sustain the productivity of crops under projected climate change scenario. Quantitative understanding of crop response to climate requires the development of statistical models for various characteristics of the crop by taking into account its time-series behaviour along with the climatic factors. Winter crops are especially vulnerable to high temperatures during the reproductive stages and their differential responses to rising temperatures can have important consequences for crop yield. The key problem is how to incorporate the pertinent information into the forecasting process and subsequently into the decision making process.

Yield is a complicated system, which is governed by number of factors. It is such a complex quantity that a method of its prediction based on single aspect may fail to explain all the variability. Numerous factors including weather parameters, cropping practices that may vary from district to district are affecting crop yield in a given region. For any operational yield model to be successful for adoption, it is necessary that data should be available much before the harvest of the crop and it should be cost effective. Weather parameters can be easily obtained at the time of maximum vegetative stage of the crop and may prove very effective for developing suitable yield forecasting models.

1.2 Status of wheat crop in Haryana

Wheat is one of the most important cereal crops in India as it forms a major constituent of the staple diet of a large part of the population. There are two crop seasons in Haryana i.e. *rabi* and *kharif.* The major *rabi* crops are wheat, tobacco, gram, linseed, rapeseed and mustard. India is the second largest producer among wheat growing countries of the World (Source: www.mapsofindia.com/ indiaagriculture/). Haryana occupies third place for wheat production among the various states in India (Source: www.agricoop.nic.in/statistics). Haryana is self-sufficient in food grains production and is one of the top contributors of food grains to the central pool. Haryana together with Punjab is called the 'Grain Bowl' of India.

1.3 Objectives of the study

Keeping in view the importance of the subject matter, an attempt has been made to develop wheat yield forecasting models in Haryana with the following objectives:

1. To develop yield forecast models using probabilities obtained through ordinal logistic regression based on weather data.
2. Selection of weather variables based on multiple linear regression and principal component analyses to develop trend-agromet models.
3. To see the comparative performance of the developed models in relation to DOA wheat yield estimates and testing the post-sample validity of the models.

To achieve the objectives, a complete review of work done on and related problems of crop yield forecasting is given in chapter-II. Chapter-III is devoted to technique/ methodology in context of weather-crop yield modeling using three different statistical procedures. The results of the investigation have been compiled and discussed in chapter-IV. The summary and conclusion are given in the last chapter-V.

CHAPTER-II

REVIEW OF LITERATURE

A comprehensive review of relevant literature in any scientific investigation is imperative. Very few studies in Haryana have been conducted related to the work under consideration, however, the studies having direct or indirect link with the present investigation have been reviewed and presented under the following three major heads:

2.1 Multiple linear regression analysis

2.2 Principal Component analysis

2.3 Logit analysis pertaining to crop yield modeling.

2.1 Weather-crop yield modeling based on multiple linear regression analysis

Huda *et al.* (1975) made an attempt to understand how the intensity and distribution pattern of weather parameters at different stages of growth affect the rice yield. The results showed that the second-degree multiple regression equation can profitably be employed in quantifying the relationship between rice yield and weather variables.

Yoshida (1981) observed clear varietal differences for high temperature injuries at different growth stages. The rice plant appeared to be most sensitive to high temperature at flowering. The second most sensitive stage was found about 9 days before flowering.

Rao (1983) studied the influence of climate and technology on wheat yield in Punjab state. The important climatic factors were identified and their role was discussed. The validity of models was judged by their relative performance in estimating the yields over a four years independent data set. The use of state model to estimate the district yields was suggested.

Agarwal (1986) studied various models and examined the effect of weather variables on rice yield at different stages of crop growth and to forecast its yield for Puri district, Orissa. Forecast model revealed that maximum temperature and sunshine hours up to 12th week were appropriate for forecasting rice yield and it explained 91% variation in yield.

Mal and Gupta (2000) attempted an empirical statistical weather-yield model for wheat in the Varanasi district of Uttar Pradesh. The weather period from 1970/71 – 1998/99 in Varanasi were selected for the regression equations. Yield was significantly related to rainfall, temperature, sunshine and rainy days from 51^{st} to 2^{nd}, 4^{th} to 6^{th}, 49^{th} to 2^{nd} and 51^{st} to 3^{rd} weeks respectively.

Verma *et al.* (2003) attempted area and production forecasts of wheat crop at zonal level in Haryana. Stratified random sampling was used for the generation of vegetation indices. Indian Remote Sensing Satellite digital data of IRS-ID, LISS-I, LISS-II, LISS-III sensors were used for the purpose. Acreage estimation was based on crop discrimination using multi-spectral digital remote sensing data and a supervised pattern recognition algorithm. The district-level production forecasts improved significantly using zonal spectral-trend-agrometeorological yield models.

Bazgeer *et al.* (2007) carried out wheat (*Triticum aestivum*) yield prediction using agrometeorological indices in Hamedan district, Iran during 2003-04 and 2004-05. According to correlation coefficients, the standard error of estimate as well as relative deviation of predicted yield from actual yield, the best subset of agrometeorological indices were selected including daily minimum temperature, accumulated difference of maximum and minimum temperatures, growing degree days, accumulated water vapour pressure deficit, sunshine hours and potential evapotranspiration.

Andarzian *et al.* (2008) developed deterministic crop growth model to predict the site-specific yield potential of wheat. The model simulated critical developmental stages and dry matter production as a function of temperature and solar radiation. Comparisons of the simulated and measured output under non-limiting conditions indicated satisfactory performance for predicting anthesis and maturity dates, and fair prediction of dry matter production and grain yield.

Lee *et al.(* 2011) developed wheat regression models to forecast wheat yield and quality. They included spatial lag effects. Wheat yield, protein and test weight level were strongly influenced by weather variables. The forecasting power of the yield and protein models was enhanced by adding the spatial lag effect. Out of sample forecasting tests confirmed the models' usefulness in accounting for the variations in average wheat yield and qualities.

Keong and Keng (2012) established a stepwise multiple linear regression model with monthly oil palm yield as the dependent variable employing agrometeorological variables in cumulated time-lag period prior to harvest as the independent variables. They showed that the percentage available water holding capacity had significant implications on monthly fresh fruit bunch yield per hectare and it was possible to predict monthly oil palm yield nine months ahead with reasonable accuracy.

Pandey *et al*. (2013) developed models for forecasting rice yield at district level on the basis of weather variables. Weekly data of 19 meteorological weeks on seven weather variables over a span of 21 years period (1989-90 to 2009-10) had been used along with the annual rice production data for Faizabad district (eastern Uttar Pradesh). Stepwise regression was used to screen out the important weather variables.

Cai *et al.* (2015) showed the relationships between the heat load and weather conditions in order to determine the input variables and output variable of the future heat load prediction model.

Verma *et al*. (2011, 12, 14) and Goyal and Verma (2015) have used agromet/spectral indices in context of pre-harvest yield forecasting of cotton, sugarcane, mustard and wheat crops in Haryana. Gardae et al. (2015) carried out discriminant function analysis for estimating wheat productivity inVaranasi district of eastern Uttar Pradesh.

2.2 Principal Component Analysis

Principal component analysis was initially given by Pearson (1901) and later extended by Hotelling (1933) on their work on educational psychology. It was further used by many statisticians to solve the complex data matrices by transferring the original set of variables into a smaller set of linear combinations that accounts for most of the variability of the original set.

Jolliffe (1972, 73) suggested five methods for discarding variables and of these; one was the multiple correlation method, two were principal component methods and two were clustering methods. The methods were compared and all were found to be satisfactory for real and artificial data, although none was shown to be overwhelmingly superior over the others.

Agarwal *et al.* (1980) developed a suitable yield forecasting model of rice in Raipul district using yield and weekly weather variables of 25 years. Principal components of weather variables were obtained and used as independent variables in the regression equation. The 11^{th} week after sowing (i.e. 3^{rd} week of August) was found suitable for forecasting rice yield.

Jain *et al*. (1984) made an attempt to forecast the yield of hybrid sorghum [*sorghum bicolor (Linn.)]* by using principal components on the biometrical characters. The results indicated the possibility of forecasting yield one month before the harvest of a crop of 3½ months duration.

Mahajan and Prasad (1985) studied that the total variation in nine characters of 25 rice (*oryza sativa L.*) varieties grown on three different sowing dates could be explained by first three principal components accounting for more than 80 per cent of variation. The first, second and third components explaining 34.9, 26.3, 19.9 percent of variation were associated with the days to 50% flowering, days to maturity and plant height, ear bearing tillers, number of tillers and grain yield respectively.

Marsh (1990) used dummy variables in wage discrimination cases. Rencher (1992) compared various approaches for interpretation of principal components and showed the choice between the correlation and covariance matrices; the conversion of coefficients into correlations, the rotation of the coefficients and the effect of special pattern in the covariance and correlation matrices.

Manly (2000) observed that the principal component analysis reduced a large number of original variables into smaller number of principal components. This can drastically reduce the computing time for the cluster analysis. He further found that the results of a cluster analysis could be rather different with and without the initial principal component analysis.

Rousson and Gasser (2004) proposed alternative methods to produce components, that unlike principal components, were correlated or had non orthogonal loadings and showed that the existing criteria used to evaluate simple components were not adequate and also proposed two new criteria that were more suitable for evaluation of simple components.

Cornillon *et al.* (2008) introduced two new forecasting methods of time series based on spline principal component analysis (PCAIV) with respect to instrumental variables. The first method was straight forward application of PCAIV. The criteria used according to the unknown value that needs to be predicted were differentiated.

Mingione (2011) used principal component analysis to forecast two indices of financial vulnerability. Rymuja *et al.* (2012) studied the use of principal component analysis for the assessment of spring wheat characteristics depending on growth system.

Pang (2011) studied macroeconomic forecasting with many predictors. Principal component analysis was used to estimate factors, and four kind of principal component estimate were examined. Then all these estimates were compared in context of forecasting macroeconomic series at various time horizons. The results suggested that one or two estimated factors can valuably summarize the information from many predictors.

Wang (2012) developed a hierarchical probabilistic principal component analysis based model with a dynamic component selection procedure. A latent variable was introduced to select promising subsets of components based upon the significance of the relationship between the response variable and principal components in the regression step. He illustrated the model using real and simulated examples.

Verma *et al.* (2015) attempted principal component technique for pre- harvest estimation of cotton yield based on plant biometrical characters spread over six successive stages within the growth period of cotton crop. The results indicated the possibility of yield prediction of cotton hybrids RCH-134BG I, RCH-134BG II and bioseeds 6488-BG II, one month ahead of the harvest time.

Winter and Dodou (2016) compared common factor analysis (CFA) with principal component analysis (PCA) loadings for distortion of a perfect cluster configuration. Results showed that nonzero PCA loadings were higher and more stable than nonzero CFA loadings.. The pattern of difference between CFA and PCA was consistent across sample sizes, levels of loadings, principal axis factoring versus maximum likelihood factor analysis, and blind versus target rotation.

2.3 Logistic regression analysis

Jain *et al.* (1992) used logistic growth model to forecast total dry matter at maturity about one month before harvest for wheat and rice crops. Total dry matter as well as head weight at maturity were over estimated in general when partial data were used. Modified growth model worked well to adjust for over estimation in total dry matter and head weight at maturity.

Johnson *et al.* (1996) described the relationship between weather and outbreaks of potato late blight in the semi-arid environment of south-central Washington with linear discriminant and logistic regression analysis and forecasted late-blight outbreaks.

Bender and Grouven (1997) introduced a non-technical proportional odds models for ordinal data. They explained the relations between standard binary logistic regression model and general polytomous logistic model. Special emphasis was given to the model building procedure and on the assessment of goodness-of-fit.

Ghamdi (2002) applied logistic regression to accident-related data collected from traffic police records in order to examine the contribution of several variables to accident severity.

Zhang (2000) considered a goodness-of-fit test for the logistic regression model under a case-control sampling plan on the basis of Kolmogorov-Smirnov-type statistic. He also proposed a bootstrap procedure for finding the P-values of the proposed test.

Peng *et.al. (*2002) demonstrated the preferred pattern for the application of logistic methods with an illustration to a data set in testing of research hypothesis. Recommendations were also offered for appropriate reporting format of logistic regression results and the minimum observation-to-predictor ratio.

Wissmann *et al.* (2007) used categorical variables in the linear regression model. It exposed the diagnostic tool condition number to linear regression models with categorical explanatory variables and analysed how the dummy variables could affect the degree of multicollinearity.

Henderson et al. (2007) conducted research to identify weather variables useful for forecasting late blight in southern Idaho. The objectives of this research were to determine if regional weather variables could be related to the occurrence of late blight in southern Idaho, and determining if disease severity (scale of 0 to 4) could be predicted using variables found to be correlated with the annual occurrence of late blight in the regions of Columbia Basin. Weather data were collected from five locations over a 9-years period (1995 to 2003). A binary logistic regression model (0 = no disease and 1 = disease) indicated that the number of hours with favorable conditions (10ºC ≤ temperature ≤ 27ºC, relative humidity ≥ 80%) in April and May was a significant disease predictor. Logistic regression analysis using an ordinal disease scale (0 = no disease and 4 = severe disease) indicated the amount of precipitation and favorable hours with extended period from April to June as significant disease predictors.

Eddy (2009) performed experiment to determine conditions that were favorable for the rust infection of wheat isolates from the current population and developed models to predict infection events. The results were derived to develop logistic regression models of infection and it was used to model the probability of yield response.

Lin *et al*. (2013) proposed to utilize the group lasso algorithm for logistic regression to construct a risk scoring system for predicting disease in swine. They showed that their scoring system significantly improved the existing scoring system.

Romero (2014) demonstrated that whether the logistic regression includes a moderation term or doesn't depend entirely on the probabilistic structure of the data, especially when the covariates are not independently distributed.

Wan *et al.* (2016) compared population means and variances under a semi parametric density ratio model using logistic regression procedure. They constructed semi parametric estimators of the differences of two population means and variances and derived their asymptotic distributions. They also showed the semi parametric estimators to be asymptotically more efficient than the corresponding non parametric ones.

CHAPTER-III

MATERIAL AND METHODS

Regional dimension of the study area, different statistical methodologies viz. regression analysis, principal component analysis and logistic regression technique adopted for the development of zonal weather-yield models are presented in the subsequent sections of this chapter

3.1 Data description and statistical methodology

The Haryana state comprising of 21 districts is situated between 74^0 25' to 77^0 38' E longitude and 27^0 40' to 30^0 55' N latitude. The total geographical area of the state is 44212 sq. km. Wheat crop is grown in all the districts of the state with varying density. The districts Hisar, Bhiwani, Sirsa and Fatehabad comprising the western zone of the state, have been considered for the model building.

The Department of Agriculture wheat yield estimates for the period 1978-79 to 2014-15 published by Bureau of Economics and Statistics, Haryana have been used for computing the trend based yield. The daily weather data on maximum temperature, minimum temperature, rainfall, sunshine hours and relative humidity of Hisar district were collected for the same period. Year (time) variable was included to take care of the variation between districts within zone as the weather data are not available for all the districts, however, the zonal model utilized the same weather information in the adjoining districts under the zone. Weather data starting from 1st fortnight of November to 1 month before harvest over the period 1978-79 to 2009-10 were utilized for the model building (crop growth period: 1st November to 15th April).

3.1.1 Computation of weather parameters

Maximum temperature, minimum temperature, rainfall, relative humidity and sunshine hours are the significant weather parameters influencing crop growth through different physiological processes and the rate of phenological development. The fortnightly weather data base was prepared from the daily weather data as shown below:

Average Maximum Temperature (TMX) $= \dfrac{\sum_{i=1}^{15} TMXi}{15}$

Average Minimum Temperature (TMN) $= \dfrac{\sum_{j=1}^{15} TMNj}{15}$

Accumulated Rainfall (ARF) $= \sum_{k=1}^{15} ARF_K$

Average Relative humidity (RH) $= \frac{\sum_{l=1}^{15} RH_l}{15}$

Average Sunshine hours (SSH) $= \frac{\sum_{m=1}^{15} SSH_m}{15}$

Where TMX_i = i^{th} day maximum temperature

TMN_j = j^{th} day minimum temperature

ARF_K = k^{th} day rainfall

RH_l = l^{th} day relative humidity

$SSHm$ = m^{th} day sunshine hours

(i,j,k,l,m refer to daily weather data)

3.1.2 Time-trend analysis

Average wheat yield data of all the districts were used to regress against yield by considering time (year) as an independent variable to get the trend equation of the form Tr = a+bt; where Tr = Trend predicted yield, a = Intercept, b = Slope and t = Year

3.2 Zonal weather-yield modeling

The processes determining crop growth are complex and final yield is governed by a number of factors. The main factors affecting crop yield are agricultural inputs and weather parameters. Use of these factors forms a broad category of models for forecasting crop yield. Weather is a principal determinant of yield. The relationship between crop yield and weather parameters is generally carried out with help of multiple regression models. These models are employed for making quantitative crop yield forecast on operational basis. Following second objective of the study, the trend yield along with 45 weather parameters were used in the development of zonal weather- yield models for western zone of Haryana.

The primary purpose of this study was to develop an objective methodology for pre-harvest wheat yield forecasting. So, the district-level wheat yield forecasting were achieved using multiple linear regression, principal component and logistic regression analyses. Hence, the study is categorized using three different statistical procedures mentioned below:

3.2.1 Multiple linear regression

The multiple linear regression is a common statistical method for data analysis. In this method, a dependent (response) variable is related with a set of independent (explanatory) variables which may or may not be inter-related among themselves. The method of least squares is adopted to estimate the regression coefficients. The standard linear regression model considered may be written in the form **Y=Xß+ε**

where **Y** is an (n×l) vector of observations (DOA yields),

X is an (n×p) matrix of known form form (weather variables & trend yield),

ß is a (p×l) vector of parameters,

ε is an (n×l) vector of errors,

and E**(ε)=0**, V**(ε)**= **I**σ^2**,** so the elements of **ε** are uncorrelated.

Since E**(ε)=0,** an alternative way of writing the model is E**(Y)= Xß**

The error sum of squares is then $e'e = (Y - Xb)'(Y - Xb)$

The normal equations $(X'X)b = X'Y$ are solved by least squares technique to provide the solution $b = (X'X)^{-1}X'Y$ and the variance of the estimated coefficient b_i is

$$\text{Var}(b_i)_{\min} = \sigma^2(\mathbf{X'X})^{-1}$$

Again, if some of the predictors are correlated with the predictor variable column $\mathbf{X_i}$, then the variance of b_i is inflated as

$$\text{Var}(b_i) = \sigma^2(\mathbf{X'X})^{-1} \times (1 - R^2_i)^{-1}$$

Where R^2_i is the R^2-value obtained by regressing the i^h predictor variable column $\mathbf{X_i}$ against all the remaining predictors. The greater the linear dependence among the predictor variable column $\mathbf{X_i}$ and the other predictors, the larger the R^2_i value. Thus $\text{VIF}_i = (1 - R^2_i)^{-1}$. If an $\mathbf{X_i}$ column is orthogonal to all other columns of the **X** matrix, VIF_i= 1and hence the variance of b_i is not inflated at all. However, VIFs exceeding 4 indicate further investigation, while VIFs exceeding 10 are signs of serious multicollinearity requiring correction. Under such situations, principal component analysis offers considerable improvement over the least squares estimates.

3.2.2 Principal component analysis

The use and interpretation of a multiple regression model often depends explicitly or implicitly on the estimates of the individual regression coefficients. However, in some situations, the problem of multicollinearity exists when there are near linear dependencies between/among the regressors. Weather variables affect the crop differently during different stages of development. The resultant response is manifested in the final yield. For instance, temperature or rainfall may be correlated positively with yield at one stage and negatively at another. Thus, crop yields depend on climatic fluctuations in a complex way. Therefore, a complex statistical methodology is necessary to describe the observed crop variations.

Principal component analysis (PCA) is a multivariate technique which reduces the data with large number of correlated variables into a substantially smaller set of new variables, through linear combination of the variables that accounts for most of the variation present in the original variables. The linear combinations so selected are called the principal components. Correlation or co-variance matrices are generally used to estimate principal components. When the variables are measured in different units, scale effects can influence the composition of derived components. In such situations, it becomes desirable to standardize the variables. Therefore, correlation matrix is considered to be better as it does not require standardization.

The procedure consists of finding the eigen roots and eigen vectors of the correlation matrix of explanatory variables. Component loadings represent the ordinary correlation between a variable and a component. Generally, the initial component loadings are not readily interpretable, it is usual practice to rotate the factor axes to a simple structure such that each variable loads highly on a single component and has small to moderate loadings on the remaining components.

Let $X_1, X_2, \ldots\ldots X_p$ be the elements of a p-component random vector X. Assuming that the mean vector of X is 0 and variance-covariance matrix $\sum$ which is a real positive definite matrix. Let us suppose that the non-zero eigen values of $\sum$ are $\lambda_1 > \lambda_2 > \ldots\ldots > \lambda_p$ and the corresponding eigenvectors are $\gamma_1, \gamma_2, \ldots\ldots, \gamma_p$. For distinct λ_i's (i = 1, 2, ……p); a (p × p) orthogonal matrix Γ can be formed, where

$$\mathbf{\Gamma} = [\gamma_1, \gamma_2, \ldots \gamma_p]$$

The **Γ** matrix diagonalizes **Σ** matrix such that

$$\Sigma = \Gamma \Lambda \Gamma',$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) = \Gamma' \Sigma \Gamma$

Now, we consider an orthogonal transformation of ***X*** vector to ***Y*** vector by

$$Y = \Gamma' X$$

where $Y_1, Y_2, \ldots, Y_p$ are the p components of Y and are called principal components. From the above transformation, we have

$$E[Y] = \Gamma' E(X) = 0$$

$$V(Y) = E[Y Y'] = E[\Gamma' X X' \Gamma] = \Gamma' \Sigma \Gamma = \mathbf{\Lambda}$$

$$= \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_p)$$

An overall measure of the variability of ***X*** can be taken as tr**Σ** or the generalized variance |**Σ**| if **Σ** > 0. Here

$$\text{tr}\, \mathbf{\Sigma} = \text{tr}\, \mathbf{\Gamma \Lambda \Gamma'} = \text{tr}\, \mathbf{\Gamma \Gamma' \Lambda} = \text{tr}\, \mathbf{\Lambda} = \sum_{i=1}^{p} \lambda_i$$

Again, we have

$$|\Sigma| = |\mathbf{\Gamma\Lambda\Gamma'}| = |\mathbf{\Lambda}| = \prod_{I=1}^{p} \lambda_i \text{ if } \lambda_i\text{'s} > 0$$

Also, it is seen that the measure of variability for the principal component ***Y*** is

$$tr\,\mathbf{\Lambda} = \sum_{i=1}^{p} \lambda_i \text{ or } |\mathbf{\Lambda}| = \prod_{i=1}^{p} \lambda_i$$

This shows that the total variation of ***X*** vector remains the same, even after transformation of ***X*** vector to a vector of principal components ***Y***.

Due to the positive definite matrix Σ, the strict positivity of λ_i is guaranteed. Here Y_1 is the principal component corresponding to λ_1, Y_2 is the component corresponding to λ_2 and so on. The first component Y_1 has the maximum variance, the second largest variance is observed for Y_2, and the i-th component has variance λ_i. The percentage of variation of X explained by i-th principal component is

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \times 100$$

The PCs are extracted in such a way that the first PC explains the largest amount of the total variation in the data set, where of first component is a linear combination of the observed variables with the values of first eigenvectors as weights. The second PC is the weighted linear combination of the original variable, weights being the values of eigenvector corresponding to λ_2. This component is uncorrelated with the first PC and it accounts for the maximum amount of the remaining total variation not already accounted for by first linear combination of the original variables. In general, the j-th [j = 1, 2,…..,p] PC is the weighted linear combination of the X's which has the largest variance of all linear combinations that are uncorrelated with all of the previously extracted PCs.

In PC analysis, one of the most commonly used criteria for solving the number of components problem is the eigen value-one criterion, also known as the Kaiser's (1960) criterion. With this approach, we retain and interpret any component with an eigen value greater than 1. The rationale for this criterion is straightforward. Each observed variable contributes one unit of variance to the total variance in the data set. Any component that displays an eigen value greater than 1 is accounting for a greater amount of variance than had been contributed by one variable. Such a component is therefore accounting for a meaningful amount of variance, and is worthy of being retained.

3.2.2.1 The Scree test

Alternatively, with the Scree test (Cattell, 1966), we plot the eigen values associated with each component and look for a "break" between the components with relatively large eigen values and those with small eigen values. The components that appeared before the break were assumed to be meaningful and retained for rotation; those appearing after the break were assumed to be less important and were not retained.

3.2.2.2 Jollife criterion

These are not the sole criterion for retaining PCs but according to Jollife (1972), only those PC's should be retained whose eigen value ≥.70.

3.2.3 Logistic regression with ordinal response variable

Logistic regression (LR) is an increasingly popular statistical technique used to model the probability of discrete (i.e., binary or multinomial) outcomes. These models also show the extent to which changes in the values of the attributes may increase or decrease the predicted probability of event outcome. LR is the part of a category of statistical models called generalized linear models. This broad class of models includes ordinary regression and ANOVA, as well as multivariate statistics such as ANCOVA and log linear regression. LR allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. However, the independent or predictor variables in LR can take any form. That is, logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The relationship between the predictor and response variables is not a linear function in logistic regression.

If the response variable Y is ordinal, the categories can be ordered in a natural way such as good/moderate/bad. One way to take account of the ordering is the use of cumulative probabilities, cumulative odd and cumulative logits. Considering k+1 ordered categories, these quantities are defined by

$$P(Y \leq i) = p_1 + \ldots + p_i$$

$$odds\ (Y \leq i) = \frac{P(Y \leq i)}{1 - P(Y \leq i)} = \frac{p_1 + \cdots + p_i}{p_{i+1} + \cdots + p_{k+1}}$$

$$logit\ (Y \leq i) = ln\ (\frac{P(Y \leq i)}{1 - P(Y \leq i)})\ ,\ i = 1, \ldots, k$$

The cumulative logistic model for ordinal response data is given by

$$Logit\ (Y \leq i) = \alpha_i + \beta_{i1} x_1 + \ldots + \beta_{ip} x_p\ , \quad i = 1, .., k$$

Thus, we have k model equations and logistic coefficient β_{ij} for each category/covariate combination.

3.2.3.1 Modeling with three groups

For this empirical study, crop years were categorized into three groups viz., adverse, normal and congenial. Using weather variables in these three groups, probabilities were obtained by ordinal logistic regression. These probabilities along with trend predicted yield were used for the development of zonal yield forecast models using stepwise regression procedure.

When dependent variable has an ordinal nature taking three values say zero, one, two; then the ordinal logistic regression model may be given as:

$$P_0 = \frac{\exp(\alpha_1+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp(\alpha_1+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}$$

$$P_0 + P_1 = \frac{\exp(\alpha_2+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp(\alpha_2+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}$$

and $P_0 + P_1 + P_2 = 1$

where P_0 is probability of $Y = 0$, P_1 is probability of $Y=1$ and P_2 is probability of $Y= 2$, α's are the intercepts and β_i's are the regression coefficients.

Zonal yield forecast models were fitted by taking probabilities P_1 and P_2 along with trend yield as regressors. The form of fitted model may be expressed as

$$\text{Yield} = a + b_1P_1 + b_2P_2 + b_3T_r + \varepsilon$$

where a is an intercept, b_i's are the regression coefficients, P_1 and P_2 are the probabilities of $Y=1$ and $Y=2$, T_r is trend yield and ε is error $\sim N(0,\sigma^2)$.

3.3 Regression diagnostics

Once a regression model has been constructed, it is important to confirm the goodness of fit of the model and statistical significance of the estimated parameters. Commonly used checks of goodness of fit include R-squared, pattern of residuals and hypothesis testing. Statistical significance can be checked by an F-test of the overall fit, followed by t-tests of individual parameters. Interpretations of residual diagnostic tests rest heavily on the model assumptions. Although examination of the residuals can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if the model's assumptions are violated.

Thus, examining residuals is a key part of all statistical modeling. Inference for the general linear model makes several assumptions, including independence of errors, normality and homogeneity of variances. Departure from the latter two of these assumptions may hint e.g., to a need for data transformation or removal of outlying observations. Diagnostic plots of residuals are frequently used to assess the validity of these assumptions or to identify possible outliers.

3.3.1 Tests for residual normality

Any graph suitable for displaying the distribution of a set of data may be used for judging the normality of the distribution of a group of residuals. The three most common types are histogram, normal probability plot and residuals versus fitted values plot.The normal probability plot is a more sensitive graph and it should produce an approximately straight line if the points come from a normal distribution.

3.3.2 Plot of residuals versus fitted values

Plotting residuals versus fitted response should produce a distribution of points scattered randomly about 0, regardless of the size of the fitted value. There should be no pattern of the residuals plotted against fitted values.

3.4 Procedures for comparison and validation of the different models

Comparison and validation of different models were made using adjusted R^2, percent relative deviations from real time data and RMSEs of post sample forecast yields.

Coefficient of determination (R^2) is in general use for checking the adequacy of the model and given by the formula

$$R^2 = 1 - \frac{SS_{res}}{ss_t}$$

where ss_{res} and ss_t are the residual sum of square and total sum of square respectively.

$$R^2_{ads} = 1 - \frac{ss_{res}/(n-p)}{ss_t/(n-1)}$$

where $ss_{res}/(n\text{-}p)$ is the residual mean square and $ss_t/(n\text{-}1)$ is the total mean square.

Percent Relative Deviation (RD%) measures the deviation (in percentage) of forecast from the actual yield data i.e. Relative Deviation={(actual yield – forecasted yield)/actual yield}*100

Root Mean Square Error (RMSE) is used as a measure of comparing two models and the formula of RMSE is

$$RMSE = \left[\left\{\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2\right\}\right]^{\frac{1}{2}}$$

where O_i and E_i are the observed and forecasted values of crop yield and n is the number of years for which forecasting has been done.

CHAPTER-IV

RESULTS AND DISCUSSION

In accordance with the objectives formulated, the analysis have been carried out to develop weather –yield models in Hisar, Bhiwani, Sirsa and Fatehabad districts comprising the western zone of Haryana. The fortnightly weather data base were used for the development of zonal models using three different statistical procedures as has been mentioned in section 3.2. Thus, the chapter contains three broad sections describing different zonal weather- yield models developed, respective wheat yield estimates and a comparative view on the forecast results achieved. Section 4.1 deals with the development of zonal yield models based on fortnightly categorization of weather data. Sections 4.2 and 4.3 provide the results achieved and a comparative view of the fitted models for district-level pre-harvest wheat yield forecasting.

To compare the forecasting performance of the developed models, zonal weather-yield models have been developed using trend based yield along with weather parameters/ PC scores/probabilities of response categories by following multiple linear regression, principal component and logistic regression analyses respectively.

4.1 Distribution of weather variables and development of forecast models

The distribution of five weather variables i.e. average maximum temperature (TMX), average minimum temperature (TMN), average relative humidity (RH), accumulated rainfall (ARF) and average sunshine hours (SSH) spread over 11 fortnights falling under 1st fortnight of November to 1st fortnight of April has been given in Appendix. The fortnightly weather data base was prepared from daily weather data for computational use. The average weather values calculated over 1st to 15th November give 1st fortnight weather parameters then average values calculated over 16th to 30th November provide 2nd fortnight weather parameters and proceeding in the next year, the average weather values calculated over 16th to 31st March and 1st to 15th April give the 10th and 11th fortnight weather parameters. The total growth period spread over 11 fortnights turned up with 55 weather parameters i.e. TMX_1, TMX_2,….TMX_{11}; TMN_1, TMN_2,…. TMN_{11}; RH_1,RH_2,….RH_{11}; ARF_1, ARF_2,…ARF_{11}; SSH_1,SSH_2,….SSH_{11} etc.

The forecast models have been developed using data from 1978-79 to 2009-10 while the data from 2010-11 to 2014-15 were used for validation of the fitted models. Fortnightly weather data starting from 1st November to 1 month before harvest were utilized for the model building. Data for the last 1 month of the crop season were excluded as the idea behind the study was to forecast yield at least 4-5 weeks before crop harvest. The forecast models developed for the western zone were used to obtain district-level pre-harvest wheat yield forecasts. Year/time variable was included to take care of the variation between districts within zone as the weather data was not available for all the districts, however, the zonal model utilized the same weather information in the adjoining districts under the zone. The selected models along with adjusted -R^2 under three different statistical procedures are presented in Tables 4.2 and 4.6.

4.1.1 Multiple linear regression analysis

Regression models via step-wise regression method were fitted, taking yield as the dependent variable and 45 weather parameters along with trend yield as the regressors. The developed zonal models shown in Table 4.2 were used to obtain yield forecasts of all the districts under consideration. The model predicted yield along with observed yield (i.e. DOA yield) and percent relative deviation are given in Table 4.7.

4.1.2 Principal component analysis

Principal component method was used for extraction of factors which consists of finding the eigen values and eigen vectors. Kaiser (1958) suggested the dropping of components having eigen roots less than 1. Thus, first nine/ten eigen values of correlation matrix of weather variables suggested nine/ten factor solution and the remaining 16/17 PCs accounted for a smaller amount i.e. less than 15 per cent of total variation. Hence, those components were not considered to be of much practical significance. Eigen vectors and eigen values of different PCs have been presented in Table 4.1. Eigen vectors being the weights were used to compute PC scores for the fortnightly data base. Scree plot based on principal component analysis is shown in Figure 4.1

The varimax rotation method was used to obtain the higher loading displaying weather variables under the retained components. Further, the trend predicted yield along with PC scores/ higher loading displaying weather variables on the retained PCs were used as regressors to obtain the suitable zonal yield forecast models (Table 4.2) and percent deviations from DOA yields are given in Table 4.7.

Table 4.1: Eigen vectors, eigen values and variance (%) explained by different principal components.

Weather parameters	Components												
	1	2	3	4	5	6	7	8	9	10	11	12	13
TMX_1	-0.6	-0.1	-0.1	-0.1	0.3	-0.6	-0.1	0.1	0.2	0.0	-0.1	0.0	-0.1
TMX_2	-0.5	-0.2	0.4	0.0	0.0	-0.2	-0.1	0.3	-0.2	0.1	-0.3	-0.1	0.2
TMX_3	-0.5	0.3	0.6	0.1	0.0	-0.1	0.2	0.1	-0.1	0.1	0.2	0.1	0.0
TMX_4	-0.6	-0.4	0.3	-0.4	0.3	0.0	0.1	-0.1	-0.2	0.0	-0.1	0.0	0.1
TMX_5	-0.1	-0.5	-0.2	0.1	-0.2	0.2	-0.3	0.5	0.2	0.2	0.0	0.0	0.1
TMX_6	-0.2	-0.4	0.0	0.5	0.5	0.0	0.0	-0.2	0.3	0.2	0.2	-0.2	-0.1
TMX_7	-0.1	0.0	-0.3	0.3	0.6	0.2	0.2	0.3	0.0	-0.3	0.3	0.0	0.0
TMX_8	-0.2	0.6	-0.3	-0.1	0.4	0.2	-0.3	0.1	-0.1	-0.1	0.1	-0.2	0.1
TMX_9	-0.6	0.6	-0.4	0.0	0.1	0.0	-0.1	-0.1	0.0	0.3	0.0	-0.1	0.1
TMN_1	0.2	0.5	0.2	-0.1	0.3	0.3	0.1	-0.2	0.2	0.3	-0.3	-0.1	-0.1
TMN_2	0.2	-0.5	0.2	0.1	0.0	0.0	0.3	0.3	-0.2	0.4	-0.2	0.1	-0.2
TMN_3	0.4	0.4	-0.2	0.1	-0.3	-0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.3
TMN_4	0.3	-0.2	0.3	0.4	-0.2	-0.2	0.1	-0.3	-0.5	0.2	0.1	0.0	0.3
TMN_5	0.0	0.3	0.1	0.0	-0.3	0.2	0.5	0.6	0.2	0.1	-0.1	0.0	0.2
TMN_6	0.2	0.4	0.5	0.5	0.2	0.1	-0.1	0.0	0.1	0.3	0.2	0.0	-0.1
TMN_7	0.5	-0.3	0.3	0.2	0.3	0.0	0.2	0.3	0.1	0.1	0.2	0.0	-0.2
TMN_8	0.2	0.3	0.0	0.5	0.5	0.2	0.0	0.1	0.1	-0.3	-0.2	0.1	0.2
TMN_9	0.1	0.4	-0.3	0.0	0.4	0.3	-0.3	0.3	-0.2	0.4	-0.1	0.1	0.1
RH_1	0.6	0.2	0.0	-0.1	0.2	0.5	0.2	-0.1	-0.3	0.0	0.0	0.1	0.1
RH_2	0.6	-0.1	-0.2	-0.2	0.2	0.0	0.2	-0.2	-0.3	0.1	0.2	0.3	0.0
RH_3	0.8	0.1	-0.4	0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.2
RH_4	0.7	0.3	0.1	0.5	-0.3	-0.2	-0.1	0.0	0.0	-0.1	0.1	-0.1	0.1
RH_5	0.0	0.7	0.4	-0.1	0.1	-0.2	0.3	-0.2	0.0	-0.4	-0.1	0.0	0.0
RH_6	0.2	0.8	0.4	0.0	0.0	0.0	-0.2	0.1	-0.1	-0.1	0.0	0.2	0.0
RH_7	0.6	-0.1	0.6	-0.2	0.2	-0.3	-0.2	-0.2	0.2	0.1	0.0	0.1	0.1
RH_8	0.6	-0.1	0.3	0.3	0.3	-0.3	-0.1	-0.1	0.3	-0.1	-0.1	0.1	0.0
RH_9	0.6	-0.4	0.0	-0.1	0.1	0.2	-0.2	0.3	0.0	-0.3	0.0	0.2	0.2
ARF_1	0.2	-0.2	0.4	-0.2	-0.2	0.8	0.0	-0.2	0.2	-0.1	-0.1	0.1	0.0
ARF_2	0.3	-0.2	0.3	-0.2	-0.3	0.7	0.1	-0.2	0.4	0.0	0.0	0.1	-0.1
ARF_3	0.6	0.1	-0.6	0.0	-0.3	-0.1	0.0	-0.1	0.2	-0.1	-0.1	-0.2	0.0
ARF_4	0.3	-0.1	0.0	0.5	-0.1	0.1	0.0	-0.2	-0.5	0.2	-0.3	-0.1	-0.2
ARF_5	-0.2	0.0	0.2	-0.2	-0.1	0.0	0.6	0.2	-0.2	0.0	0.3	-0.1	-0.1
ARF_6	0.1	0.4	0.3	0.2	-0.1	0.0	-0.3	0.4	0.0	0.0	-0.2	0.2	0.1
ARF_7	0.5	-0.4	0.3	-0.3	0.2	-0.3	-0.1	-0.1	0.0	0.1	0.2	0.2	0.2
ARF_8	0.1	-0.3	-0.2	0.4	0.2	-0.2	0.3	-0.1	0.1	-0.2	-0.5	0.3	0.0
ARF_9	0.6	-0.5	0.1	-0.3	0.2	0.0	-0.2	0.1	-0.1	-0.3	0.2	0.0	-0.1
SSH_1	-0.4	-0.5	0.0	0.5	-0.4	-0.2	-0.1	0.2	0.1	-0.2	0.0	0.2	0.0
SSH_2	-0.6	0.0	0.4	0.3	-0.1	0.2	-0.3	-0.2	0.1	0.0	0.0	0.1	0.3
SSH_3	-0.5	-0.1	0.5	0.4	0.0	0.2	-0.2	-0.1	-0.1	-0.3	0.2	0.1	0.0
SSH_4	-0.6	-0.3	-0.1	-0.4	0.3	0.3	0.0	-0.1	0.1	0.1	-0.1	0.3	0.0
SSH_5	0.1	-0.6	-0.2	0.3	-0.2	0.2	-0.4	-0.2	-0.1	0.3	0.1	0.0	0.0
SSH_6	-0.4	-0.6	-0.3	0.2	0.1	0.0	0.3	-0.2	0.2	0.1	0.0	0.1	-0.1
SSH_7	-0.5	0.2	-0.4	0.5	0.0	0.2	0.0	0.0	-0.3	-0.3	0.0	0.2	-0.1
SSH_8	-0.4	0.3	-0.2	-0.4	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	0.2	0.5	-0.3
SSH_9	-0.4	0.5	-0.1	0.2	0.0	-0.1	0.3	-0.2	0.4	0.1	0.2	0.3	0.1
Eigen value	8.41	6.40	4.18	3.57	2.88	2.84	2.42	2.07	1.91	1.74	1.29	1.19	1.03
Percent variance explained	18.70	14.23	9.29	7.93	6.40	6.32	5.38	4.60	4.24	3.86	2.86	2.64	2.29
Percent cumulative variance	18.70	32.92	42.21	50.14	56.54	62.85	68.23	72.84	77.08	80.94	83.80	86.44	88.73

Eigen vectors being the weights were used to compute principal component scores

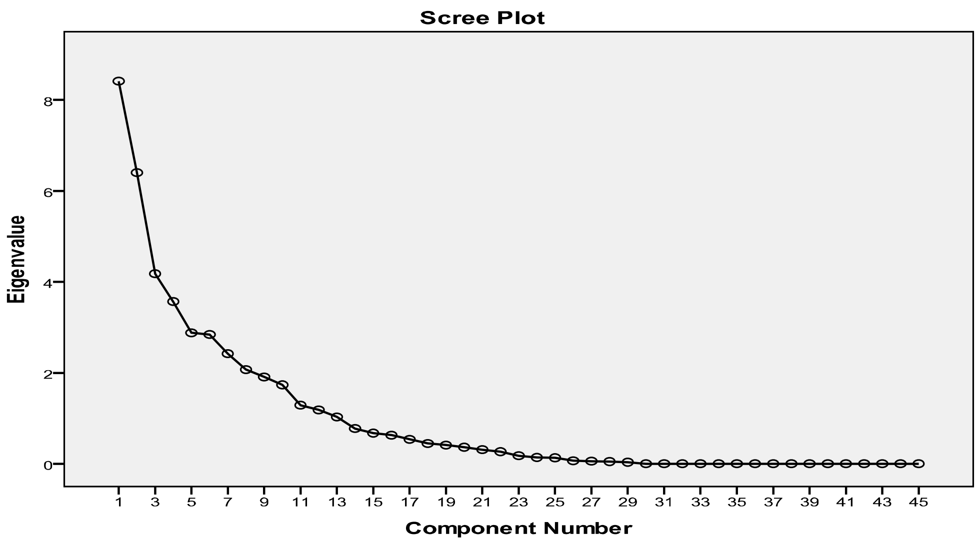


Figure 4.1: Scree plot based on principal component analysis

Zonal weather- yield relationships for Hisar, Bhiwani, Sirsa and Fatehabad districts of Haryana

Table 4.2: Suitable zonal models for district-level wheat yield forecasting.

Model	Yield Forecast Model	adj – R^2	SE
1	Y= -14.51+0.36 Tr -1.19SSH_5 -0.68TMN_5 -1.23SSH_2 +1.61TMX_1 + 0.20RH_4 -0.89SSH_1	0.75	3.49
2	Y= 11.25+0.84Tr +1.49TMX_1-1.35TMX_2+0.60TMN_4 -0.90SSH_5 -1.03SSH_8 -0.49TMN_8	0.87	2.30
3	Y= 13.27+0.87Tr +1.39TMX_1-1.33TMX_2 +0.59TMN_4 -0.79SSH_5 –1.10SSH_8 -0.53TMN_8 -0.04ARF_9	0.89	2.27
4	Y= 17.23+0.51Tr -2.30PC_2+1.70PC_5-1.05PC_9- 0.91PC_4	0.68	4.36
5	Y= 0.07+1.01Tr +0.97PC_9 -0.77PC_5	0.80	3.09

Model 1 & 2 - Weather parameters and trend yield as regressors

Model 3- Weather parameters displaying higher PC loadings and trend yield as regressors

Model 4 & 5 - Principal component scores and trend yield as regressors

Where Y - Model predicted yield (q/ha)

Tr - Trend yield (q/ha)

TMX - Av. Maximum Temperature

TMN - Av. Minimum Temperature

ARF - Accumulated Rainfall

RH - Av. Relative Humidity

SSH - Av. Sunshine Hours

(1,2,3,……,9 refer to different fortnights)

PC_i - i^{th} principal score (i = 1,2,3,……..)

SE - Standard error of the estimate

4.1.3 Logistic regression with ordinal response variable

First of all, the crop years were categorized into three groups viz., adverse, normal and congenial using clustering/classification to make categorical response variable. The results obtained in this regard are presented sequentially.

Clustering/classification for making three groups of DOA wheat yield

Table 4.3: Eigen values and percent variation explained.

Function(s)	Eigen value	% Variance	% Cumulative Variance	Canonical Correlation
1	89.40	93.50	93.50	0.99
2	6.21	6.50	100.00	0.93

First 2 canonical functions were used in the analysis

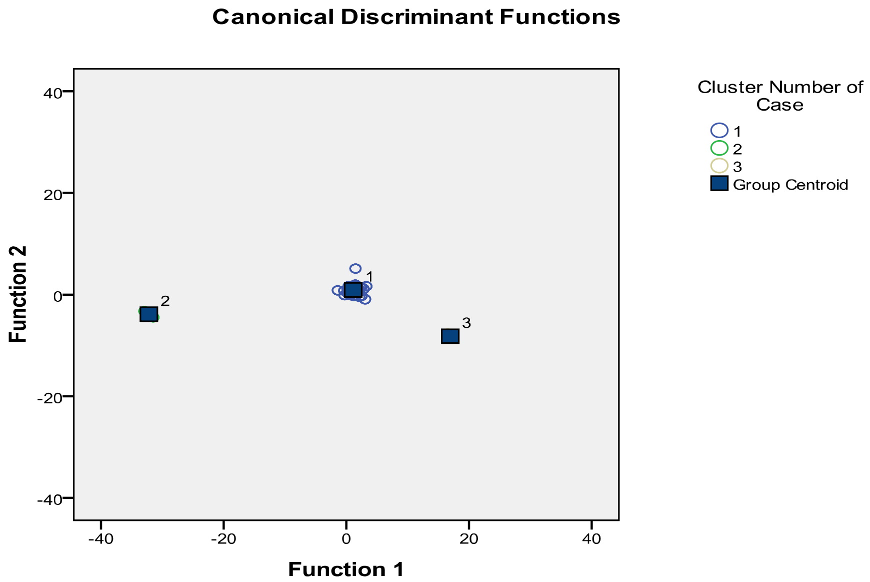


Figure 4.2: Classification results

Table 4.4: Predicted group membership.

		Cluster Number of Case	Predicted Group Membership			Total
			1	2	3	
Original	Count	1	107	0	0	107
		2	0	8	0	8
		3	0	0	8	8
	%	1	100.0	.0	.0	100.0
		2	.0	100.0	.0	100.0
		3	.0	.0	100.0	100.0

100.0% of original grouped cases correctly classified.

Next, the ordinal logistic regression analysis using weather variables in respect of these three groups was carried out to obtain estimated probabilities of response categories.

Table 4.5: Pseudo R-Square values

Cox and Snell	0.60
Nagelkerke	0.98
McFadden	0.97

Link function: Logit

Finally, the zonal yield models were obtained by taking estimated response probabilities along with trend yield as regressors and DOA wheat yield as regressand (Table 4.6). The developed zonal models were used to obtain the district-level wheat yield estimates for the post sample period(s) 2010-11, 2011-12, 2012-13, 2013-14 and 2014-15.

Table 4.6: Parameter estimates and adjusted-R^2 of zonal yield models

Model	Intercept	X_1/ coeff	X_2/ coeff	X_3/ coeff	Adj. R^2	SE
Model-6	-3.73	T_r +1.03	P_2 +208.06		0.83	3.21
Model-7	10.75	T_r +0.69	P_1 +0.30	P_2 -74.10	0.71	5.14

Model 6 & 7 - Estimated response probability and trend yield as regressors
X_1, X_2, X_3 are predictors in the model
Pi - Estimated probabilities for response category 1,2,3

Table 4.7: Comparative view in terms of percent deviations of the forecast yield(s) from DOA yield(s) based on selected zonal weather-yield models

District/ Forecast year	DOA Yield (q/ha)	Model-2		Model-3		Model-6	
		Fitted Yield (q/ha)	RD (%)	Fitted Yield (q/ha)	RD (%)	Fitted Yield (q/ha	RD (%)
Hisar							
2010-11	46.22	48.06	3.97	48.23	4.34	46.43	0.44
2011-12	50.98	45.79	10.18	46.38	9.02	47.01	7.79
2012-13	42.73	47.02	10.05	46.62	9.09	44.40	3.92
2013-14	44.77	41.45	7.41	41.69	6.89	48.20	7.67
2014-15	41.80	47.83	14.42	45.37	8.53	45.59	9.06
Bhiwani							
2010-11	44.65	43.98	1.51	43.98	1.50	41.43	7.22
2011-12	43.06	41.74	3.07	42.16	2.08	42.05	2.34
2012-13	40.55	43.00	6.03	42.42	4.61	39.47	2.66

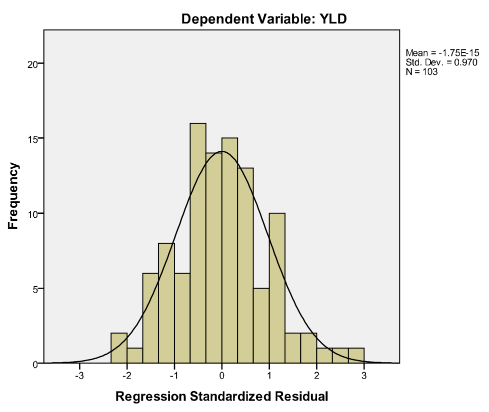
2013-14	42.28	37.45	11.42	37.52	11.27	43.30	2.40
2014-15	38.50	43.85	13.90	41.22	7.07	40.71	5.75
Sirsa							
2010-11	51.30	49.82	2.89	50.07	2.41	48.58	5.30
2011-12	53.57	47.64	11.07	48.31	9.82	49.28	8.00
2012-13	48.42	48.96	1.12	48.63	0.44	46.78	3.39
2013-14	53.47	43.48	18.68	43.80	18.09	50.68	5.22
2014-15	47.06	49.94	6.12	47.57	1.08	48.18	2.37
Fatehabad							
2010-11	50.81	50.53	0.56	50.80	0.02	49.45	2.68
2011-12	54.72	48.35	11.65	49.05	10.37	50.15	8.36
2012-13	46.81	49.66	6.10	49.37	5.46	47.64	1.77
2013-14	53.18	44.18	16.92	44.53	16.27	51.54	3.08
2014-15	48.58	50.64	4.24	48.30	0.58	49.03	0.93

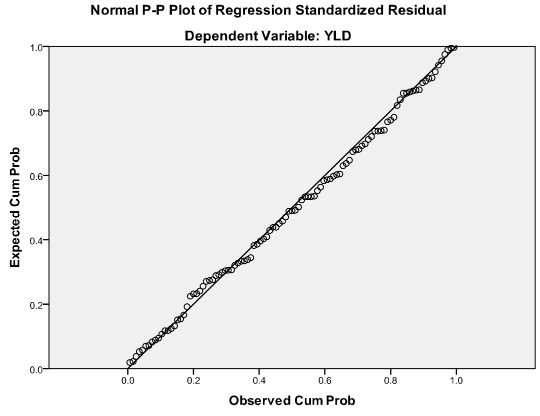
Model 2 - Regression based
Model 3 - PC based
Model 6 - Logit based

$$Percent\ deviation\ (RD\%) = \frac{(DOA\ yield - fitted\ yield)}{DOA\ yield} \times 100$$

4.2 Checking model assumptions

To examine the adequacy of the selected model, the histograms for the residuals (Figure 4.3) were found reasonably symmetric, also the normal-probability plot provided a graphical means of comparing the residual distribution to the normal. The following figure indicates that the residuals have an approximate normal distribution. Except for few outlying observations, the normality assumption is quite reasonable. Residual plot against fitted values for the selected models produced a distribution of the points scattered randomly about 0 in an un-patterned fashion except for few influential observations.





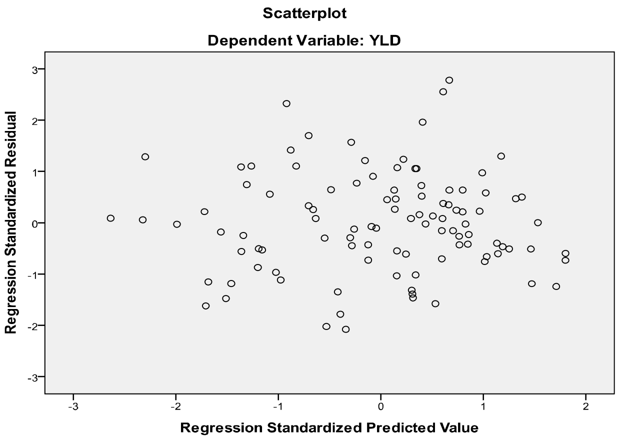


Figure 4.3: Residual diagnostics for selected model

4.3 Comparison of the fitted models

From the selected models, the pre-harvest wheat yield forecasts for the years 2010-11, 2011-12, 2012-13, 2013-14 and 2014-15 have been obtained. The performance of zonal forecast models has been compared on the basis of Adj-R^2, percent deviation of forecast yield from the observed yield and RMSE etc.

- A perusal of the results indicates the preference of using prediction equations based on principal component scores/probabilities of response categories obtained in logit analysis, over the regression models using weather parameters as regressors.
- Trend yield (Tr) has been observed as an important parameter appearing in all the models, indicating that most of the variability in yield is explained by T_r, which is an indication of technological advancement, improvement in fertilizer/insecticide/pesticide/ weedicide used and increased use of high yielding varieties.

The percent deviations of yield forecasts from observed yields (Figure 4.4) are in general of low order except the forecasts for few points. This higher deviation may be due to the reason that the forecast in those cases is based on insufficient data used for prediction purpose. Average absolute percent relative deviations and RMSEs for the post-sample period forecasts based on selected zonal models under different procedures are shown in Tables 4.8 & 4.9.

Table 4.8: Comparative view in terms of average absolute percent deviations based on different procedures.

Districts	Model-2	Model-3	Model-6
Hisar	9.20	7.58	5.78
Bhiwani	7.18	5.31	4.08
Sirsa	7.97	6.37	4.86
Fatehabad	7.89	6.54	3.37

Table 4.9: District-level RMSEs for the post-sample period forecasts based on the alternative models

RMSEs			
Districts	**Model-2**	**Model-3**	**Model-6**
Hisar	4.38	3.54	2.99
Bhiwani	3.47	2.64	1.92
Sirsa	5.40	4.96	2.74
Fatehabad	5.18	4.77	2.30

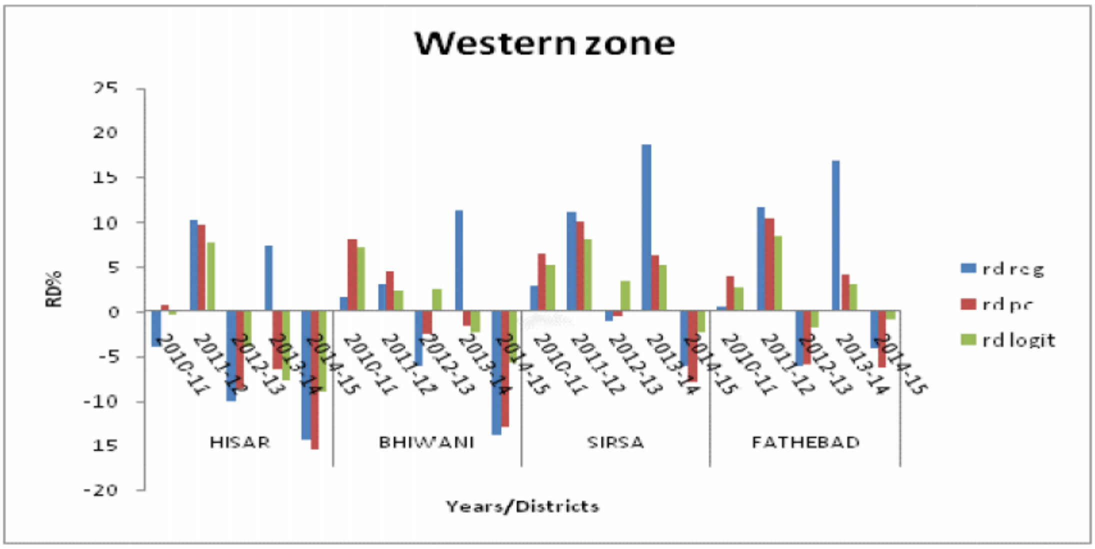


Figure 4.4: Comparative view in terms of percent deviation from the observed yield under different procedures used

The suitable zonal models have been used to obtain district-level wheat yield forecasts in western zone of Haryana. The results show that the pre-harvest wheat yield prediction gives good agreement with real time yield for all the districts under consideration. The other question was to see the usefulness of the zonal model to forecast district-level yield in the state. The results obtained indicate that the district-level pre-harvest wheat yield predictions have improved significantly using zonal weather-yield models. Moreover, the developed models provide reliable forecasts of wheat yield at least one month in advance of the crop harvest, while on the other hand, DOA yield estimates are obtained quite late after the actual harvest of the crop. The percent relative deviations falling within tolerable limits favor the use of developed zonal models for pre-harvest wheat yield forecasting in the western zone of Haryana.

CHAPTER-V

SUMMARY AND CONCLUSION

Commercialization of agriculture and allied sectors urges for intensive research in agriculture and allied sectors. Timely and effective pre-harvest forecast of crop yield is important for advance planning, formulation and implementation of policies related to the crop procurement, distribution, price structure and import-export decisions etc. Regression analysis is the most frequently used statistical technique for investigating and modelling the relationship between variables. Building a regression model is an iterative process. Usually several analyses are required as improvements in the model structure and flaws in the data are discovered.

Crop forecasts can be formed in many different ways. Various statistical approaches are in vogue for arriving at crop forecasts. Every approach has its own advantages and limitations. The method chosen depends on the purpose and importance of the forecasts as well as the costs of alternative forecasting methods. Yield is a complicated system which is governed by number of factors. The main factors affecting the crop yield are agricultural inputs and weather variables.

India has one of the best systems in the world to collect, collate and compile data on crop production. The official estimates of major cereal and commercial crops are issued by Directorate of Economics and Statistics at state/national level. However, the final estimates are given a few months after the actual harvest of the crop. Thus, one of the limitations of the conventional methods is timeliness and quality of the statistics. This makes these statistics unusable for planning and management purposes. Hence, there is always a considerable scope of improvement in the conventional system.

In view of the importance of subject matter, the present investigation has been undertaken with the objectives i) To develop yield forecast models using probabilities obtained through ordinal logistic regression based on weather data, ii) Selection of weather variables based on multiple linear regression and principal component analyses to develop trend-agromet models and iii) To see the comparative performance of the developed models in relation to DOA wheat yield estimates and testing the post-sample validity of the models.

To achieve the objectives & channelize the research work, the study has been divided into five chapters. Introduction of the subject matter is presented in Chapter-I. It explains the need and importance of crop yield forecasting. Chapter-II embodies the review of recent and important developments by prominent researchers for reduction of variables through different statistical procedures i.e. regression analysis, principal component analysis and ordinal logistic regression. Statistical methodology in context of three different statistical procedures used under this study are dealt in Chapter-III. Chapter-IV consists of the results and discussion regarding zonal weather-yield models. It starts with the results achieved during the processing stage of multiple linear regression, principal component analysis and logistic regression analyses; then selecting the zonal models based on fortnightly categorization of weather data and district-level wheat yield forecast obtained and finally to have a comparison between observed (DOA) yield and fitted yield from three different perspectives. Summary and conclusion are given in Chapter-V.

The zonal forecast models have been developed using data from 1978-79 to 2009-10 and the data from 2010-11 to 2014-15 were used for validation of the fitted models. Fortnightly weather data starting from 1st November to 1 month before harvest were utilized. Data for the last 1 month of the crop season were excluded as the idea behind the study was to forecast yield at least 4-5 weeks before crop harvest. Year/time variable was included to take care of the variation between districts within zone as the weather data was not available for all the districts, however, the zonal model utilized the same weather information in the adjoining districts under the zone so that a longer series could be obtained in a relatively shorter period.

First of all, the multiple linear regression models via step-wise regression method were fitted by taking DOA yield as response variable and weather parameters and trend yield as regressors. Secondly, principal component scores along with trend yield were used as regressors to obtain the suitable zonal models. Alternatively, the trend based yield along with higher loading displaying weather variables on PC components were also used as predictors to obtain the same.

Under logistic regression analysis, the crop years were categorized into three groups viz., adverse, normal and congenial using clustering/classification to make categorical response variable. For quantitative forecasting, the zonal yield models were obtained by taking estimated response probabilities and trend yield as regressors and DOA wheat yield as regressand. The performance of the zonal forecast models was compared on the basis of adj-R^2, percent deviation of forecast yield from the observed yield and root mean square error RMSE etc. The adequacy of the fitted models was examined through histogram, normal-probability plot for the residuals and residual plot against fitted values for the selected models under three different statistical procedures used.

The overall results indicate the preference of using prediction equations based on principal component scores/ probabilities of response categories obtained in logit analysis over the regression models using weather parameters as regressors. Trend yield (Tr) has been observed an important parameter appearing in all the models, indicating that most of the variability in yield is explained by T_r, which is an indication of technological advancement, improvement in fertilizer/insecticide/ pesticide/weedicide used and increased use of high yielding varieties. The other question was to see the usefulness of zonal weather-yield models and that provided quite satisfactory results pertaining to district-level pre-harvest wheat yield forecast in the state.

In the end, it is concluded that the district-level yield prediction showed good agreement with DOA yield estimates for all the districts under consideration. Moreover, the selected zonal models provide reliable forecasts of wheat yield about 4-5 weeks in advance of the crop harvest while DOA yield estimates are obtained quite late after the actual harvest of the crop. The percent relative deviations falling within tolerable limits favor the use of developed zonal weather-yield models for district-level pre-harvest wheat yield forecasting in the western zone of Haryana.

Agarwal, R., Jain, R. C., Jha, M. P. and Singh, D. (1980) Forecasting of rice yield using climatic variables, *Ind. J. Agril. Sci.*, 50(9), 680-684.

Agarwal, R., Jain, R. C. and Jha, M. P. (1986) Models for studying rice crop-weather relationship, *Mausam,* 37(1), 67-70.

Andarzian, B., Bakhshandeh, A. M., Bannayan, M., Emam, Y., Fathi, G. and Alami, S. (2008) Wheat Pot : a simple model for spring wheat using monthly weather data, *Biosyst. Eng*., 99, 487–495.

Bazgeer, S., Kamali, Gh. and Mortazavi, A. (2007). Wheat yield prediction through agrometeorological indices for Hamedan, Iran. BIABAN, 12: 33-38.

Bender, R. and Grouven, U. (1997) Ordinal logistic regression in medical research, *Journal of the Royal College of Physicians of London,* 31(5), 546-551.

Cai, Q., Wang, W. and Wang, S. (2015) Multiple regression model based on weather factors for predicting the heat load of a district heating system in Dalian, China—a case study, *The Open Cybernetics & Systemics Journal,* 9, 2755-2773.

Cornillon, P. A., Imam, W. and Lober, M. (2008) Forecasting time series using principal component analysis with respect to instrumental variables, *Computational Statistics & Data Analysis*, 52, 1269-1280.

Draper, N. R. and Smith, H. (1981) Applied Regression Analysis, 3rd Edition, ISBN: 0-471-17082, 736 pages.

Eddy, R. (2009) Logistic regression models to predict stripe rust infections on wheat and yield response to foliar fungicide application on wheat in Kansas, Kansas State University, Kansas, 1-116.

Field, A. (2000) *Discovering Statistics using SPSS for Windows*, London – Thousand Oaks – New Delhi, Sage publications.

Garde, Y. A., Dhekale, B. S. and Singh, S. (2015) Different approaches on pre harvest forecasting of wheat yield, Journal of Applied and Natural Science, 7 (2), 839-843.

Ghamdi, A. S. (2002) Using logistic regression to estimate the influence of accident factors on accident severity, King Saud University, Saudi Arabia, 34, 729-741.

Goyal, M. and Verma, U. (2015) Spectral-weather-crop yield forecasting: Discriminant function analysis, *Journal of Applied Probability and Statistics, 10(1), 1-14*.

Henderson, D., Williams, C. J. and Miller, J. S. (2007) Forecasting late blight in potato crops of southern Idaho using logistic regression analysis, Plant Dis. 91, 951-956.

Huda, A. K. S., Ghildyal, B. P., Tomar, V. S. and Jain, R. C. (1975) Contribution of climate variables in predicting rice yield, *Agril. Meteorology*, 15*,* 71-86.

Jain, R. C., Sridharan, H. and Agarwal, R. (1984) Principal component technique for forecasting sorghum yield. *Indian J. Agric. Sc.*, 54(6), 467-470.

Jain, R. C., Agarwal, R. and Singh K. N. (1992) A within year growth model for crop yield forecasting, *Biom. J.*, 34(7), 789-799.

Johnson, D. A., Alldredge, J. R. and Vakoch, D. L. (1996) Potato late blight forecasting models for the semiarid-environment of south-central Washington, *American Phytopathology*, 86, 480-84.

Jolliffe, I. T. (1972) Discarding variables in a principal component analysis I, Artificial data, *Applied Statistics*, 21, 160-173.

Jolliffe, I. T. (1973) Discarding variables in a Principal Component Analysis II, Real Data, *Applied Statistics*, 21, 160-173.

Keong, Y. K. and Keng W. M. (2012) Statistical modeling of weather-based yield forecasting for young mature oil palm, APCBEE Procedia, 4, 58-65.

Lee, B. H., Kenkel, P. and Brorsen, B.W. (2011) Pre-harvest forecasting of county wheat yield and wheat quality conditional on weather information, Southern Agricultural Economics Association, Corpus Christi Texas, 1-44.

Lin, H., Wang, C., Liu, P. and Holtkamp, D. J. (2013) Construction of disease risk scoring systems using logistic group lasso: application to porcine reproductive and respiratory syndrome survey data, *Journal of Applied Statistics,* 40(4), 736-746.

Mahajan, R. K. and Prasad, A. S. R. (1985) Application of principal component technique in rice. Crop Improv., 12(2),159-164.

Mal, R. K. and Gupta, B. R. D. (2000) Wheat yield forecast models based on meteorological parameters, *Journal of Agro meteorology*, 2(1), 83-87.

Manly, B. F. J. (2000) Multivariate statistical methods a primer, 2nd ed. Chapman and Hall, united state of America.

Marsh, (1990) Using Dummy Variables in Wage Discrimination Cases, University of Notre Dame, 1-27.

Mingione, F. (2011) Forecasting with principal component analysis: an application to financial stability indices for Jamaica, Finanical Stability Department, Research and Economic Programming Division, Bank of Jamaica, 1-27.

Pandey, K. K., Rai, V. N., Sisodia, B. V. S., Bharati, A. K. and Gairola, K. C. (2013) Pre-harvest forecast models based on weather variable and weather indices for Eastern U.P, *Advance in Bioresearch*, 4(2), 118-122.

Pang, Y. (2011) Forecasting using principal components from many predictors, Mathematical Statistics, Stockholm University, SE-106 91, Sweden, 1-40.

Peng, C. Y. J., Lee, K. L. and Ingersoll G. M. (2002) An introduction to logistic regression analysis and reporting, *The Journal of Educational Research*, 96(1), 3-14.

Rao, G.A. (1983). Estimation of wheat yield over Punjab using district and state models. *Mausam*, 34(3): 275-280.

Romero, A. A. (2014) Where do moderation terms come from in binary choice models?, Central European Journal of Economic Modelling and Econometrics, 6, 57-68.

Rousson and Gasser, (2004) Simple Component Analysis, *Applied Statistics*, 53, 539-556.

Verma, U., Aneja, D. R. and Hooda, B. K. (2015) Principal component technique for pre-harvest estimation of cotton yield based on plant biometrical characters. *J. of Cotton Research and Development*, 29 (2), 339-343.

Verma, U., Dabas, D. S., Hooda, R. S., Kalubarme, M. H., Yadav, M. and Sharma, M. P. (2011a) Remote Sensing based wheat acreage and spectral-trend-agrometeorological yield forecasting: Factor Analysis Approach, *Society of Statistics, Computers and Applications,* 9 (1&2), 1-13**.**

Verma, U., Dabas, D. S. and Singh, J. P. **(**2011b**)** Regression, multivariate techniques, time series and mixed modelling in context of pre-harvest wheat, mustard, cotton and sugarcane yield prediction in Haryana using the SAS system, *Research Bulletin*, *Department of Soil Science, CCS Haryana Agricultural University, Hisa, Haryana*, pp.1-154, Verma, U., Piepho, H. P., Ogutu, J. O., Kalubarme, M. H. and Goyal, M. (2012) Multi-level mixed modelling for weather-crop-yield relationships on agro-climatic zone basis in Haryana, *Advances and Applications in Statistics,* 28(1), 1-22.

Verma, U., Piepho, H. P., Ogutu, J. O., Kalubarme, M. H. and Goyal, M. (2014) Develoment of agromet models for district-level cotton yield forecasts in Haryana State, *International J. of agricultural and Statistical Sciences*, 10(1), 59-65.

Verma, U., Ruhal, D. S., Yadav, M., Khera, A. P., Hooda, R. S., Singh, C. P., Kalubarme, M. H. and Hooda, I. S. (2003) Wheat yield modelling using remote sensing and agrometeorological data in Haryana, *J. Indian Society of Agricultural Statistics*, 56 (2), 190-98.

Wan, S., Xu, B. and Zhang B. (2016) Using logistic regression for semiparametric comparison of population mean and variance, Communications in Statistics–Theory and Methods, 45(9), 2485-2503.

Wang, W. (2012) Bayesian principal component regression with data-driven component selection, *Journal of Applied Statistics,* 39(6), 1177-1189.

Winter, J. C. F. and Dodou, D. (2016) Common factor analysis versus principal component analysis: A comparison of loadings by means of simulations, Communications in Statistics—Simulation and Computation, 45, 299–321.

Wissman, M., Toutenburg, H. and Shalabh (2007) Role of Categorical Variables in Multicollinearity in Linear Regression Model, Technical Report 008, 2007, Department of Statistics, University of Munich.

Yoshida, S. (1981) Fundamental of Rice Crop Science, International Rice Research Institute, 933, Manila, Philippines.

Zhang, B. (2000). A goodness of fit test for multiplicative-intercept risk models based on case control data, Statistica Sinica, 10, 839-865.

APPENDIX : Fortnightly weather data of Hisar

Year	Av. Min Temp (^0c)	Av. MaxTemp (^0c)	Av. Relative Humidity (%)	Av. Sunshine (hours)	Acc. Rainfall (mm)
1980					
Jan-I	18.6	3.5	77	7.1	11.0
Jan-II	21.7	4.1	70	8.3	0.2
Feb-I	22.5	4.8	64	9.0	0
Feb-II	28.0	9.8	68	8.7	0
Mar-I	24.7	9.2	60	7.7	18.9
Mar-II	30.3	13.2	54	8.1	2.7
Apr-I	35.7	16.5	41	9.4	0
Apr-II	41.3	21.4	29	9.3	0
Nov-I	27.9	11.1	65	7.6	6.9
Nov-II	28.0	9.9	66	8.3	0
Dec-I	24.3	4.5	66	8.4	0
Dec-II	19.9	8.1	77	4.1	29.3
1981					
Jan-I	19.6	5.6	72	7.7	15.6
Jan-II	20.3	8.8	73	5.8	11.6
Feb-I	21.6	7.5	68	8.0	10
Feb-II	25.9	9.3	65	7.3	11.8
Mar-I	25.7	10.9	66	6.4	4.8
Mar-II	27.4	13.7	68	6.6	37.9
Apr-I	34.8	16.6	44	10.0	0
Apr-II	39.4	20.2	42	10.0	0
Nov-I	25.8	12.6	69	7.3	49.2
Nov-II	25.9	9.0	60	8.7	11.5
Dec-I	23.1	5.2	59	8.8	0
Dec-II	22.1	4.2	63	8.4	0
1982					
Jan-I	20.4	6.4	72	7.1	7.9
Jan-II	19.6	5.2	68	6.8	1.2
Feb-I	18.7	7.1	78	5.4	17.7
Feb-II	21.1	7.3	66	6.8	0
Mar-I	21.1	7.8	74	6.9	27.3
Mar-II	27.2	12.8	70	8.3	14.3
Apr-I	33.1	15.1	58	9.8	25.2
Apr-II	34.0	18.8	61	9.4	28.9
Nov-I	30.0	13.1	53	7.8	0
Nov-II	26.5	10.2	54	7.5	0.4
Dec-I	24.5	7.6	59	7.5	0
Dec-II	20.6	4.6	68	6.8	8.5
1983					
Jan-I	18.9	3.6	71	7.1	0
Jan-II	20.1	5.8	71	6.9	22.9

Feb-I	19.8	4.6	70	7.8	0.4
Feb-II	23.3	8.3	72	8.5	1.5
Mar-I	26.8	11.9	60	7.0	1.0
Mar-II	28.3	11.7	55	8.1	1.6
Apr-I	29.0	14.8	60	6.9	75.5
Apr-II	32.5	17.4	61	9.6	32.9
Nov-I	29.2	7.8	53	8.9	0
Nov-II	27.3	7.5	61	8.6	0
Dec-I	23.6	5.6	62	7.5	0
Dec-II	21.7	5.6	63	7.5	5.5
1984					
Jan-I	19.5	2.7	67	7.3	1
Jan-II	19.8	2.6	60	8.5	0
Feb-I	22.4	4.8	58	9.0	0
Feb-II	20.8	5.6	63	8.6	12.8
Mar-I	29.7	9.3	59	9.1	0
Mar-II	32.8	14.9	56	8.9	0
Apr-I	34.8	15.6	49	9.0	0
Apr-II	38.4	19.7	44	8.8	0
Nov-I	28.7	9.8	60	8.9	0
Nov-II	26.9	8.7	58	8.9	0
Dec-I	25.2	6.9	60	8.0	0
Dec-II	20.6	2.9	59	7.6	0
1985					
Jan-I	18.2	5.0	73	5.6	4.4
Jan-II	21.3	5.9	69	7.5	1
Feb-I	24.1	4.4	59	9.6	0
Feb-II	27.1	6.5	53	9.9	0
Mar-I	31.1	9.6	57	9.5	0
Mar-II	32.2	13.9	57	8.4	6.6
Apr-I	33.1	16.9	55	9.4	5.5
Apr-II	39.6	19.5	34	10.5	0
Nov-I	29.6	11.5	58	9.1	0
Nov-II	26.8	8.0	59	8.9	0
Dec-I	24.0	7.7	63	7.8	2.4
Dec-II	20.8	8.2	72	6.3	5.2
1986					
Jan-I	19.6	2.7	67	8.1	0
Jan-II	20.0	4.9	70	7.4	0.8
Feb-I	21.0	6.8	71	6.7	13.0
Feb-II	22.1	6.6	67	8.6	0.1
Mar-I	29.2	11.5	61	8.4	0.1
Mar-II	27.7	11.5	58	8.9	11.5
Apr-I	34.2	14.9	47	9.9	0
Apr-II	38.2	18.8	43	10.2	0
Nov-I	31.2	12.3	57	9.0	0
Nov-II	27.0	9.0	57	7.3	0

Dec-I	23.4	6.9	64	6.7	0.1
Dec-II	19.2	0.7	61	8.3	0
1987					
Jan-I	21.1	6.2	69	5.0	4.8
Jan-II	20.1	4.9	73	7.6	29.3
Feb-I	24.9	8.0	63	8.6	0
Feb-II	24.5	9.9	69	8.3	22.2
Mar-I	27.2	11.2	65	8.6	4.5
Mar-II	31.3	14.6	65	8.1	8.4
Apr-I	34.5	15.0	52	9.3	4.4
Apr-II	39.2	17.4	51	9.6	0
Nov-I	31.2	10.1	50	9.3	0
Nov-II	29.2	7.6	48	9.2	0
Dec-I	23.1	5.8	58	7.4	6.5
Dec-II	23.0	4.7	66	7.1	0
1988					
Jan-I	22.5	6.9	68	5.5	0
Jan-II	22.0	3.7	61	8.5	1.6
Feb-I	24.9	6.6	60	8.7	8.5
Feb-II	24.7	9.1	60	7.0	1.5
Mar-I	27.5	11.4	64	7.7	11.5
Mar-II	30.0	11.2	47	8.8	0
Apr-I	36.7	14.5	36	10.0	0
Apr-II	39.4	22.1	36	8.6	0.6
Nov-I	30.3	11.6	55	8.6	0
Nov-II	26.0	8.8	61	8.1	0
Dec-I	24.3	4.7	59	8.5	0
Dec-II	21.2	8.1	71	5.4	2.3
1989					
Jan-I	18.1	5.7	74	6.0	16.1
Jan-II	20.2	2.6	59	8.8	0
Feb-I	21.3	4.5	59	7.9	1.0
Feb-II	24.1	5.9	56	9.3	0
Mar-I	27.5	8.6	60	8.9	2.5
Mar-II	28.5	13.6	63	7.7	2.7
Apr-I	32.1	12.1	45	10.4	0
Apr-II	37.2	16.4	36	10.6	0
Nov-I	30.1	10.3	49	9.1	0.1
Nov-II	26.8	10.4	60	8.1	1.0
Dec-I	23.4	5.6	64	7.5	0
Dec-II	18.5	6.4	79	4.2	7.3
1990					
Jan-I	19.6	4.5	71	6.9	0
Jan-II	24.6	7.7	60	8.6	0
Feb-I	22.6	8.9	71	5.9	14.0
Feb-II	20.9	9.4	79	6.2	28.8
Mar-I	25.4	8.6	58	9.9	0

Mar-II	28.0	14.0	53	7.3	6.3
Apr-I	31.3	15.3	48	10.1	4.8
Apr-II	38.8	19.8	34	9.5	0
Nov-I	30.8	10.9	56	9.2	0
Nov-II	26.4	9.2	62	7.9	0
Dec-I	23.8	5.8	66	7.6	0
Dec-II	22.0	6.2	66	6.2	9.5
1991					
Jan-I	17.3	3.0	69	7.1	1.6
Jan-II	21.9	5.0	60	8.3	0
Feb-I	22.5	7.4	65	11.5	4.1
Feb-II	22.5	7.6	69	7.5	12.9
Mar-I	25.6	10.2	64	8.0	0
Mar-II	29.9	12.2	63	8.9	0
Apr-I	31.9	15.8	59	7.9	26.0
Apr-II	35.3	16.8	42	11.1	4.9
Nov-I	29.1	11.8	56	8.3	0
Nov-II	26.9	7.9	57	7.7	0
Dec-I	25.0	6.6	61	7.5	0
Dec-II	20.4	6.6	76	4.5	3.2
1992					
Jan-I	19.7	4.5	67	6.8	3.1
Jan-II	22.0	8.0	71	5.8	8.2
Feb-I	19.2	6.5	76	5.8	22.9
Feb-II	23.1	7.4	65	3.4	0
Mar-I	27.6	9.8	57	6.9	0
Mar-II	27.5	10.9	60	5.0	3.6
Apr-I	32.2	13.7	61	8.9	0
Apr-II	36.8	17.3	48	8.5	8.1
Nov-I	29.7	10.3	67	7.9	0
Nov-II	25.7	9.0	73	6.5	0
Dec-I	23.8	6.2	71	5.7	0
Dec-II	24.1	5.9	62	6.9	0
1993					
Jan-I	19.3	7.4	72	3.6	4.4
Jan-II	20.0	3.9	64	7.3	2.5
Feb-I	26.1	8.2	67	7.2	0
Feb-II	23.9	9.4	62	6.3	17.5
Mar-I	26.2	10.9	74	7.3	2.6
Mar-II	26.9	10.3	58	7.7	0
Apr-I	32.0	13.8	53	7.9	10.7
Apr-II	38.8	18.6	40	10.7	0
Nov-I	31.3	12.1	53	8.0	0
Nov-II	28.0	8.4	57	7.7	0
Dec-I	25.9	5.7	59	7.6	0
Dec-II	22.1	3.1	65	6.1	0

1994					
Jan-I	21.7	7.6	73	4.1	33.2
Jan-II	19.5	5.2	70	6.4	1.3
Feb-I	22.7	6.2	64	7.1	0
Feb-II	23.0	5.9	62	8.0	4.1
Mar-I	28.4	9.2	58	8.9	0
Mar-II	32.8	13.1	52	9.3	0
Apr-I	33.5	13.8	41	8.4	7.4
Apr-II	36.5	17.4	36	9.7	0.6
Nov-I	29.7	11.1	58	8.5	0
Nov-II	27.2	8.7	60	7.3	0
Dec-I	25.4	7.6	62	6.7	0
Dec-II	22.3	5.5	61	6.4	0
1995					
Jan-I	17.9	6.0	75	4.7	48.3
Jan-II	19.3	4.9	68	7.9	0
Feb-I	22.8	8.7	68	6.3	20.7
Feb-II	21.5	7.7	68	7.2	2.7
Mar-I	24.2	7.1	53	8.5	0.2
Mar-II	30.5	13.7	63	7.3	13.0
Apr-I	32.5	14.0	45	9.2	13.5
Apr-II	36.3	19.2	42	8.7	3.5
Nov-I	30.8	10.5	55	8.5	0
Nov-II	28.0	8.4	58	7.5	0.5
Dec-I	23.3	5.7	58	6.8	0
Dec-II	22.6	6.1	63	6.9	0
1996					
Jan-I	20.9	4.6	71	4.2	5.3
Jan-II	19.4	4.1	67	6.9	0
Feb-I	22.9	6.1	71	7.0	16.0
Feb-II	24.3	7.9	73	7.9	19.2
Mar-I	28.4	10.3	74	8.3	0
Mar-II	29.2	13.1	67	7.4	31.6
Apr-I	34.9	12.0	43	10.2	0
Apr-II	38.4	20.0	45	8.5	0
Nov-I	31.4	10.3	51	8.5	2.0
Nov-II	26.0	5.4	58	8.3	0
Dec-I	22.9	1.8	51	7.9	0
Dec-II	23.4	2.8	59	7.4	0
1997					
Jan-I	21.6	0.7	66	6.3	0
Jan-II	18.7	3.1	72	6.2	13.6
Feb-I	21.5	4.0	64	7.8	1.0
Feb-II	25.2	5.6	57	8.6	0
Mar-I	27.8	9.7	58	6.7	7.0
Mar-II	27.0	12.4	67	6.9	48.1
Apr-I	30.2	15.4	59	8.3	21.5

Apr-II	36.3	17.1	62	9.0	1.2
Nov-I	27.1	11.6	68	6.7	0
Nov-II	23.4	8.9	71	4.4	2.0
Dec-I	19.1	8.2	83	3.0	21.5
Dec-II	13.6	5.1	87	1.8	9.5
1998					
Jan-I	18.8	5.0	74	5.6	2.0
Jan-II	19.2	2.7	66	7.0	0
Feb-I	22.8	7.1	68	6.7	10.0
Feb-II	22.2	8.0	72	6.7	16.9
Mar-I	23.4	9.2	70	7.0	27.5
Mar-II	28.8	12.4	56	8.5	0
Apr-I	33.7	16.3	52	8.3	1
Apr-II	38.9	19.5	35	8.8	0
Nov-I	28.5	13.7	64	6.7	5.0
Nov-II	27.5	9.0	58	8.1	0
Dec-I	26.3	6.5	63	7.3	0
Dec-II	17.1	6.3	87	2.4	0
1999					
Jan-I	16.5	7.4	88	1.8	5.5
Jan-II	17.9	7.9	87	3.8	24.7
Feb-I	21.1	7.2	73	7.4	0
Feb-II	24.5	10.1	70	7.2	1.0
Mar-I	28.2	10.7	60	9.5	0
Mar-II	32.2	11.5	49	10.0	0
Apr-I	38.4	15.6	32	10.1	0
Apr-II	41.1	16.9	25	10.8	0
Nov-I	32.3	11.0	45	8.2	0
Nov-II	28.4	6.9	51	7.9	0
Dec-I	24.7	4.3	61	6.5	0
Dec-II	22.5	3.1	67	5.8	0
2000					
Jan-I	16.3	5.5	83	3.0	0
Jan-II	20.3	5.4	69	7.3	8.3
Feb-I	17.8	7.3	84	4.2	10.7
Feb-II	23.1	4.6	71	9.1	0
Mar-I	26.9	8.0	59	8.3	0
Mar-II	30.9	10.9	53	8.9	0
Apr-I	36.5	13.4	37	10.4	0
Apr-II	41.0	21.5	29	8.6	0
Nov-I	32.1	10.6	50	8.0	0
Nov-II	27.0	7.2	59	7.2	0
Dec-I	25.1	2.6	53	8.8	0
Dec-II	23.8	3.5	62	7.7	0
2001					
Jan-I	15.4	3.3	84	4.0	15.0
Jan-II	20.9	2.7	66	7.6	0

Feb-I	23.3	2.8	59	9.3	0
Feb-II	25.9	8.8	62	7.9	9.2
Mar-I	28.7	7.8	55	9.1	0
Mar-II	31.2	11.1	56	8.2	0
Apr-I	35.2	15.4	40	9.0	3.3
Apr-II	36.0	17.5	43	9.0	43.3
Nov-I	31.3	11.2	52	8.6	0
Nov-II	28.2	7.6	61	8.6	0
Dec-I	25.2	7.9	67	7.7	0
Dec-II	20.5	4.8	75	6.2	0
2002					
Jan-I	20.3	5.1	78	5.9	0
Jan-II	19.3	3.9	74	6.7	0
Feb-I	20.4	3.3	74	7.6	25.0
Feb-II	24.9	10.0	70	7.6	10.2
Mar-I	25.9	9.7	64	8.6	2.1
Mar-II	32.7	13.3	54	9.7	0
Apr-I	36.3	16.2	47	8.8	0
Apr-II	40.2	20.5	32	8.7	0
Nov-I	29.2	12.6	66	7.3	0
Nov-II	28.2	6.9	57	8.4	0
Dec-I	24.6	5.3	60	7.8	0
Dec-II	23.1	7.5	72	5.9	10.0
2003					
Jan-I	14.0	2.7	88	3.7	1.5
Jan-II	18.6	4.8	83	5.3	4.9
Feb-I	20.8	4.3	73	7.8	1.3
Feb-II	23.8	11.3	72	7.4	20.7
Mar-I	24.7	8.7	62	7.7	2.0
Mar-II	31.4	13.3	62	8.3	0
Apr-I	36.2	15.6	47	9.7	0
Apr-II	38.1	19.2	30	8.8	0
Nov-I	30.4	11.1	54	8.3	0
Nov-II	24.3	5.7	63	7.7	0
Dec-I	24.5	9.0	70	7.1	8.6
Dec-II	15.7	5.8	89	3.0	0
2004					
Jan-I	17.3	5.1	82	5.0	0
Jan-II	17.5	7.5	87	3.7	16.9
Feb-I	20.7	4.0	71	8.1	0
Feb-II	27.1	9.6	69	8.8	0
Mar-I	29.2	10.2	65	9.6	0
Mar-II	34.1	13.1	57	9.4	0
Apr-I	38.7	17.9	38	7.8	0
Apr-II	38.0	21.4	45	6.4	21.0
Nov-I	29.5	7.9	58	8.6	0
Nov-II	28.2	10.8	67	5.8	0

Dec-I	24.1	6.0	68	6.6	0
Dec-II	20.9	6.5	75	4.2	2.0
2005					
Jan-I	19.5	4.4	73	5.7	13.0
Jan-II	17.2	3.6	75	4.2	9.2
Feb-I	20.1	9.0	79	5.3	56.5
Feb-II	20.9	5.6	71	7.8	0
Mar-I	20.1	9.0	79	5.3	56.5
Mar-II	22.8	6.7	69	8.3	0
Apr-I	34.0	12.9	37	9.2	0
Apr-II	36.6	17.5	41	8.3	16.5
Nov-I	29.5	11.3	60	7.8	3.2
Nov-II	27.3	7.7	56	7.9	0
Dec-I	22.8	2.4	65	8.0	0
Dec-II	20.2	3.2	70	5.0	0
2006					
Jan-I	19.6	2.5	68	6.9	0
Jan-II	22.0	5.4	66	6.3	0
Feb-I	26.9	8.7	68	7.4	0
Feb-II	28.0	11.4	76	6.5	0
Mar-I	27.0	10.9	72	6.8	26.7
Mar-II	28.5	11.8	69	9.4	0.5
Apr-I	36.2	15.1	51	9.2	0
Apr-II	37.9	20.8	44	8.2	0
Nov-I	31.2	13.2	64	5.6	0
Nov-II	25.7	8.7	78	7.3	0
Dec-I	21.8	5.9	70	5.9	2.2
Dec-II	21.9	5.2	71	7.1	3.8
2007					
Jan-I	17.6	1.4	69	6.2	0
Jan-II	22.5	5.1	68	6.9	0
Feb-I	22.4	8.7	89	4.9	75.3
Feb-II	23.2	8.9	78	7.8	10.8
Mar-I	24.0	9.7	70	7.3	44.3
Mar-II	29.8	12.7	65	9.2	0
Apr-I	36.2	15.3	49	9.9	2.0
Apr-II	40.1	20.1	39	9.9	0
Nov-I	30.8	12.3	60	6.2	0
Nov-II	28.0	8.9	63	6.4	0
Dec-I	21.6	7.0	70	3.7	3.8
Dec-II	21.4	2.7	62	7.6	0
2008					
Jan-I	19.6	4.6	65	6.1	3.5
Jan-II	17.6	1.9	63	5.7	0
Feb-I	17.2	2.2	65	6.2	0.9
Feb-II	26.2	6.6	57	8.4	0
Mar-I	31.1	11.1	61	7.8	0

Mar-II	33.6	13.4	54	8.8	0
Apr-I	32.5	16.9	57	7.5	14.0
Apr-II	38.7	17.5	31	10.2	0
Nov-I	31.3	12.1	58	6.7	0
Nov-II	26.0	8.5	65	6.4	3.2
Dec-I	25.1	7.3	65	6.4	0
Dec-II	21.3	6.1	77	4.9	0.8
2009					
Jan-I	18.6	3.3	74	6.0	0
Jan-II	21.5	8.2	76	5.9	10.3
Feb-I	23.2	7.1	72	7.2	6.1
Feb-II	25.1	7.7	65	8.6	0
Mar-I	28.8	9.6	62	9.0	0
Mar-II	30.0	13.9	67	7.4	4.1
Apr-I	34.2	16.3	62	8.5	24.9
Apr-II	37.3	18.3	30	9.8	0
Oct-II	32.3	12.0	55	8.0	0
Nov-I	28.6	13.3	68	3.0	0
Nov-II	25.7	6.4	64	6.7	0
Dec-I	24.5	7.4	63	5.8	0
Dec-II	21.8	2.7	63	6.9	0
2010					
Jan-I	14.9	5.6	87	2.6	11.5
Jan-II	19.6	6.2	80	5.0	0
Feb-I	24.2	6.5	70	6.7	7.6
Feb-II	27.7	8.4	66	8.5	0
Mar-I	31.5	14.2	59	8.8	2.5
Mar-II	38.1	19.2	50	8.7	0
Apr-I	39.8	17.8	40	8.7	0
Apr-II	42.4	23.0	29	7.1	0
Nov-I	29.5	13.4	60	5.4	0
Nov-II	26.0	9.6	62	4.9	0
Dec-I	22.7	3.7	64	7.1	0
Dec-II	20.0	5.4	77	4.9	43.6
2011					
Jan-I	13.7	4.6	82	3.2	0
Jan-II	19.8	3.7	65	6.7	0
Feb-I	23.6	8.4	71	7.0	18.7
Feb-II	21.6	7.8	77	6.7	16.1
Mar-I	24.0	9.1	74	7.2	7.9
Mar-II	32.3	13.3	63	8.7	4.6
Apr-I	32.3	14.3	62	8.5	10.3
Apr-II	36.0	18.4	63	9.1	24.9
Nov-I	31.1	12.1	60	5.4	0
Nov-II	28.1	10.3	67	4.9	0
Dec-I	25.7	8.9	66	7.1	0
Dec-II	20.7	2.0	70	4.9	0

2012					
Jan-I	18.4	5.4	73	4.1	0
Jan-II	18.5	4.2	74	5.6	14.4
Feb-I	20.0	5.1	64	6.5	0
Feb-II	22.3	5.5	63	7.4	0
Mar-I	26.8	8.8	59	8.4	0
Mar-II	30.6	12.3	56	7.8	0
Apr-I	34.6	18.3	55	8.5	7.3
Apr-II	33.8	17.8	56	8.2	26.0
Nov-I	28.8	10.9	65	7.6	0
Nov-II	26.0	7.5	65	6.4	0
Dec-I	23.6	7.3	73	6.4	4.4
Dec-II	18.2	4.8	78	5.9	1.1
2013					
Jan-I	15.6	2.9	81	4.2	0
Jan-II	19.6	5.4	73	6.3	43.0
Feb-I	21.7	7.8	76	6.3	16.4
Feb-II	21.3	10.2	80	5.5	16.3
Mar-I	28.1	10.5	69	8.8	30.5
Mar-II	28.7	13.4	70	7.9	0.6
Apr-I	33.6	15.4	53	9.0	1.5
Apr-II	36.4	18.9	42	8.5	0.8
Nov-I	26.0	11.8	68	4.9	9.4
Nov-II	27.1	8.9	65	7.1	0
Dec-I	25.1	7.7	67	6.4	0
Dec-II	18.7	6.0	77	5.3	0
2014					
Jan-I	17.8	2.9	77	5.2	0
Jan-II	18.2	8.1	89	7.0	2.0
Feb-I	20.1	7.6	82	5.1	5.6
Feb-II	21.7	7.6	80	6.9	6.9
Mar-I	23.9	10.0	79	7.0	24.6
Mar-II	28.5	14.3	67	7.3	22.4
Apr-I	32.6	15.5	57	9.6	8.5
Apr-II	35.6	18.6	51	9.7	7.9
Nov-I	29.2	12.3	59	4.3	0
Nov-II	27.3	8.1	57	7.3	0
Dec-I	24.7	7.6	67	5.0	9.0
Dec-II	14.6	4.6	89	6.6	0
2015					
Jan-I	29.2	12.3	59	6.5	0
Jan-II	27.3	8.1	57	7.9	0
Feb-I	24.7	7.6	67	6.7	9.0
Feb-II	14.6	4.6	89	3.1	0
Mar-I	15.5	6.7	88	2.1	6.6

ABSTRACT

Title of Thesis	:	**Logistic regression models for pre-harvest wheat yield estimation in western zone of Haryana**
Name of degree holder	:	**Sudesh Rani**
Title of degree	:	Master of Science
Admission No.	:	2014BS20M
Name and address of Major advisor	:	**Dr. (Mrs.) Urmil Verma** Principal Scientist (Statistics) Department of Math, Stat and Physics CCS Haryana Agricultural University, Hisar-125004 (Haryana) India.
Degree awarding University/Institute	:	CCS Haryana Agricultural University, Hisar-125004 (Haryana) India.
Year of award of degree	:	2016
Major subject	:	Statistics
Total number of pages in thesis	:	29+iii+X
Number of words in the abstract	:	394

Keywords: Trend yield, Maximum temperature, Minimum temperature, Rainfall, Sunshine hours, Relative humidity, Multiple linear regression, Principal component scores, Ordinal logistic regression and zonal weather- yield models.

Timely and effective pre-harvest forecast of crop yield is important for advance planning, formulation and implementation of policies related to the crop procurement, distribution, price structure and import-export decisions etc. These are also useful to farmers to decide in advance their future prospects and course of action. The statistical modelling approaches viz., multiple linear regression, principal component analysis and ordinal logistic regression were used to achieve pre-harvest wheat yield forecasting in Hisar, Bhiwani, Sirsa and Fatehabad districts comprising the western zone of Haryana. The zonal weather-yield forecast models have been developed using data from 1978-79 to 2009-10 and the data from 2010-11 to 2014-15 were used for validation of the fitted models. Fortnightly weather data starting from 1st November to 1 month before harvest i.e. 1st fortnight of March over the period 1978-79 to 2009-10 were utilized for the model building. Data for the last 1 month of the crop season were excluded as the idea behind the study was to forecast yield at least 4-5 weeks in advance of the crop harvest. Year/time variable was included to take care of the variation between districts within zone as the weather data were not available for all the districts, however, the zonal model utilized the same weather information in the adjoining districts under the zone so that a longer series could be obtained in a relatively shorter period.

The predictive performance(s) of the contending models were observed on the basis of Adj-R^2, percent deviations of wheat yield forecasts in relation to the observed yield(s) and root mean square error(s) as well. The overall results indicate the preference of using prediction equations based on principal component scores/ probabilities of response categories obtained in logit analysis over the regression models using weather parameters as predictors. Trend yield (Tr) has been observed an important parameter appearing in all the models, indicating that most of the variability in yield is explained by T_r, which is an indication of technological advancement, improvement in fertilizer/ insecticide/ pesticide/ weedicide used and increased use of high yielding varieties. The other question was to see the usefulness of zonal weather-yield models and that provided quite satisfactory results pertaining to district-level wheat yield forecast(s) in the state. The percent relative deviations falling within tolerable limits favor the use of developed zonal weather-yield models for district-level pre-harvest wheat yield estimation in the western zone of Haryana.

MAJOR ADVISOR **SIGNATURE OF STUDENT**

HEAD OF THE DEPARTMENT

CURRICULUM VITAE

Name	:	Sudesh Rani
Date of birth	:	14.11.1993
Place of birth	:	Hisar
Mother's name	:	Smt. Birmati
Father's name	:	Sh. Raj Kumar Nain
Permanent address	:	H.No. 458, 12 Quarter road, Hisar (Haryana) Pincode- 125 004
Telephone	:	-
Mobile	:	9896590078
E-mail	:	nainsudesh1193@gmail.com

Academic qualifications

Degree	University/ Board	Year of passing	Percentage of marks	Subjects
Secondary	HBSE	2008	90.20%	Hindi, English, Mathematics, Science, Social Science
Senior Secondary	HBSE	2010	75.40%	English, Physics, Chemistry, Mathematics, Home Science
B.Sc.(Hons.) Agri.	College of Agriculture,CCS HAU, Hisar	2014	70.20%	All Agricultural subjects

Co-curricular activities	:	-
Medals/ Honours received	:	-
List of Publication	:	1

Sudesh and Verma, U. (2016) Using logistic regression to predict wheat yield in western zone of Haryana, *International Journal of computer and Mathematical Sciences*, **5**(12), 26-32.

UNDERTAKING OF COPYRIGHT

I **Sudesh Rani,** Admission No. **2014BS20M** undertake that I give copyright of my thesis entitled,**" Logistic regression models for pre-harvest wheat yield estimation in western zone of Haryana"** to the Chaudhary Charan Singh Haryana Agricultural University, Hisar.

I also undertake that the patent, if any, arising out of the research work conducted during the programme shall be filed by me only with due permission of the competent authority of CCS HAU, Hisar.

SIGNATURE OF THE STUDENT