

A Statistical Modeling for Estimation of Wheat Yield Using Weighted Rainfalls in the Punjab, Pakistan

Abdul Qayyum (Corresponding author)

Department of Statistics G.C University, Lahore, Pakistan
Tele: 92-333-443-2269 E-mail: qayyum_so@hotmail.com

Muhammad Khalid Pervaiz

Department of Statistics G.C. University, Lahore, Pakistan
Tele: 92-333-461-3163 E-mail: dr.khalidpervaiz@hotmail.com

Wasif Khurshid

Additional Secretary (Planning) Agriculture Department
Government of the Punjab, Pakistan
Tel: 92-333-613-8987 E-mail: wasifkhurshid@gmail.com

Received: February 28, 2011 Accepted: March 23, 2011 Published: December 1, 2011
doi:10.5539/jas.v3n4p95 URL: <http://dx.doi.org/10.5539/jas.v3n4p95>

Abstract

In this paper an effort has been made to develop a statistical model for wheat production in the Punjab, Pakistan. All the variables that have either light or strong impact on the yield of wheat, surveyed by Agriculture Department and Meteorological Department, Punjab, have been taken under consideration. The proposed model is independent of time binding and gives a good estimate of wheat production by using all the inputs of the crop in a particular season. A new concept of weighted rainfalls has been incorporated in the model and a detailed comparative study between actual total and weighted rainfalls has been carried out on each divisional and provincial level modeling for the selection of the best acceptable model and to observe the impact of both types of rainfalls on the quality of estimation of yield. The predictive performance of the final selected model has been assessed through various validity tests. For the purpose a data, in hard form, of 25036 cases comprising almost 1.2 million values have been fed and used. The study is very helpful for agricultural planners in respect of the most important staple food of Pakistan. Also many recommendations to farmers can be made about different inputs of the crop.

Keywords: Barani Area, District, Division, Kharif Season, Mund, Rabi Season

1. Introduction

Area of Pakistan is 796,096 sq.km and the population is almost 170 million. The country has four provinces namely Sindh, Punjab, Khyber Pakhtoonkawa and Balochistan. The Punjab being the second largest (area of Punjab is 205,345 sq.k.m) and the most agricultural province of the country, contributes almost 75% to the total wheat production of Pakistan.

There are two zones of the Punjab: lower and upper. In Kharif season (May to September), rice crop is sown in the upper and cotton is sown in the lower Punjab. Wheat crop is sown in the Rabi season (October to April) throughout the province.

There are two major categories of the area in the Punjab: Irrigated and Un-irrigated (Barani). In irrigated area, there is availability of canal as well as tube-well water, so farmers are comparatively less dependent on the rainfalls. But in case of Barani areas, crops are totally dependent on timely rainfalls as there is no availability of any other source of water. The most of the upper Punjab area is hilly and Barani.

The Punjab is administratively and geographically divided into 133 tehsils, 36 districts and 9 zones (each zone is called a Division). Each division comprises 3 to 6 districts having meteorologically identical state. The name of District-Headquarter of a division is same as the division.

A secondary data of wheat production for the year 2005-06 to 2008-09 have been taken from the Crop Reporting Service (CRS), an attached wing of Agriculture Department, Government of the Punjab, solely responsible for handling of all kinds of Agricultural Statistics in the Punjab.

The total number of selected sample villages is 1086 and in each sample village, six randomly selected plots of 15x20 sq ft in three randomly selected fields of wheat have been harvested. Yield of each plot along with all the 16 variables, having 43 categories impacting the yield of the crop, have been recorded. The variables recorded for each field are as follow:

- Source of seed –Urea used
- No of Plough –DAP used
- No of Levels –No of waters
- Pesticides used on crop / seed –Total Rainfalls in the season
- Spray on crop –Weighted Rainfalls in the season
- Variety of wheat –Average Humidity of the season
- Quantity of seed –Average Max/Min Temperature
- Sowing time – Irrigated/Un-irrigated area

In Pakistan, yield of a crop is measured in ‘maund’ (1 Maund = 37.3242 kg) per ‘acre’ (1 Acre = 198x220 sq.ft). A total yield data of 25036 plots all over the Punjab amounting to almost 1.2 million values have been used in the study.

2. Results and Discussion

Various models on different stages and categories of data have been developed and their utility has been discussed through empirical study. The distribution of this section is based on different stages and different categories under discussion.

2.1 Choice of Regression Technique

In the first step of model building, type of regression or model is specified. As the response variable is continuous quantitative and also out of 43 explanatory variables, 15 variables are categorical, therefore ‘Multiple Linear Regression’ technique is proposed for model building. An important assumption of normality of response variable is verified and illustrated by the histogram and P-P plot using yields of 25036 sample plots from all over the province, given in the Figure 1.

Verification of linear relationship between response and explanatory variables is not checked when some of predictors are categorical (Montgomery, 2001). Also distribution of error terms is same as of the response variable.

2.2 Choice of Model With/Without Intercept

Both type of Liner Regression Models with or without intercept are used in model building purpose. The choice is based on the predictive performance of the models. A model possessing better prediction quality is obviously selected for operational use.

Initially two models on Punjab level are developed; one using with intercept and second without intercept term. For each type, further two types of models have been developed; one using weighted rainfalls, (Qayyum, 2010), and second using total rainfalls as one of the regressor variables. The Table1 bears important information of both types of models.

Making comparison of both types of models, it is evident that all are significant as P-value is much less than 0.05. But there is a considerable difference between values of R^2 of both models i.e. model without intercept accounts for 92.4% variation in the yield whereas model with intercept explains only 38.5% and 38.3% variation of the response variable in case of weighted and total rainfalls respectively.

Though value of R^2 in case of model without intercept is much higher than that of model with intercept, but Mean Square Error (MSE) of model without intercept is increased as compared to the value of MSE of model with intercept. While comparing two models, Sum of Squares of Errors is more important predictive

performance parameter as compared to R^2 , (Hahn, 1979), i.e. betterment in the quality of prediction by a model is measured by lower value of MSE rather by higher value of R^2 . So it is evident from the Table 1 that models with intercept, both for weighted and total rainfalls, are giving better prediction quality as compared to the models without constant term. Hence a model without constant term will not be used as a final model for wheat projection. Also a model without intercept increases multicollinearity.

2.3 Divisional Level Modeling

Initially, various independent models for each division of the Punjab have been developed and their important parameters are compared to check their suitability.

As it has been discussed earlier that Pakistan agriculture is mostly based on climatic conditions among which amounts of rainfalls are the most important that cause significant changes in the production of any crop. Considering the importance of rainfalls for wheat production, three types of models for each division have been developed: firstly, a sum of total rainfalls in the whole Rabi season i.e. TotalRains as one of the explanatory variables, secondly weighted rainfalls i.e. WtRains of the season and thirdly individual rainfall of each Rabi month i.e. AllRains have been used as predictors along with other explanatory variables.

Table 2 bears first column of 'Division Name', second column has total sample share from the respective division, third column has three types of rainfall related variables, as mentioned above, using as predictors along with other predictors, fourth column possesses proportion of explained variation R^2 , fifth column has value of Durbin-Watson test as a check for auto-correlation, sixth column has value of MSE, seventh column has P-value against F test for significance of the model, eighth column is about the coefficient of rainfall related variable in the model, ninth column shows number of significant variables in the particular model and last tenth column possesses name of rainfall related insignificant variables.

A detailed comparison of these all information of different divisional models may lead to the best model choice.

Column No seven of Table 2 bears P-value against F test, which shows significance of the models. It is evident from the table that all models are significant; no one is bad fit. Now the best fit out of all good fits has to be selected.

It is common in all the divisional models that R^2 and MSE of WtRains never be less than corresponding R^2 and MSE of TotalRains. In five out of eight divisions, values of R^2 and MSE, in case of WtRains are greater than corresponding values of TotalRains. Only in three divisions i.e. Lahore, Faisalabad and Sargodha, the values are same in both cases.

Against the rain variable AllRains, independent rainfall of each Rabi month has been used as a predictor in the model. It is important to highlight that as the variation in the rainfalls of each Rabi month decreases, statistically they become insignificant and necessarily they are dropped out from the study of modeling by the software. But it is illogical that rainfalls are insignificant in estimating wheat production.

In all the divisional models, against the variable AllRains, rainfalls of some of Rabi months were proved insignificant. For instance, in Gujranwala model, rainfalls of month Nov, Dec, Jan and Feb were proved insignificant, wherein it is the prime period of Rabi season when the crop needs timely watering. Similarly, in all the other divisional models, rainfalls of main Rabi months were proved insignificant. Though, in some cases, they have a little higher value of R^2 as compared to the case of WtRains, but a model without rainfalls cannot be preferred as a final operational model for yield estimation of wheat.

As reasoned in the last paragraph, dropping the variable AllRains, a selection has to be made between the models of TotalRains and WtRains. On comparison of both types of the models, it is evident that WtRains variable proved insignificant only in one case of Rawalpindi division when TotalRains and AllRains are also insignificant because of minor variation in the rainfalls. In rest of all the other cases of divisional models, WtRains was proved as a significant contributor to the yield. Like, in Gujranwala model, both TotalRains and AllRains were insignificant but WtRains was significant.

Another important comparison was made in the eighth column of Table 2 with title 'Coeff in Model'. It bears the coefficients of TotalRains and WtRains variables in their respective models. In all the cases, value of coefficient of WtRains is more than that of TotalRains, which depicts that per unit change in the yield of wheat against one unit change in WtRains is more than that of TotalRains, i.e. WtRains contributes more significantly as compared to TotalRains. For example, in case of Bahawalpur model, both WtRains and TotalRains are significant in their own models and their coefficients are 0.066 and 0.019, respectively. It can be interpreted that against 1 mm increase in weighted rainfall causes 66 gm increase in the yield of wheat per area of 300 ft² where as this

increase in case of TotalRains is 19gm. Importantly, because of better contribution, MSE in case of WtRains is less than that of TotalRains.

Fifth column of Table 2 shows value of Durbin-Watson test for auto correlation check. In all the cases except the models against Rawalpindi and Sargodha Divisions, its value is more than 1 and many researchers have used 1+ value of Durbin-Watson test as an acceptable value.

2.4 Provincial Level Modeling

After discussion of divisional level modeling, various models on the Punjab level have been constructed. On comparison of divisional and provincial level models, the best one is to be selected.

Like divisional models, all the models on provincial level are also significant, i.e. no one is bad fit as shown in the column no 9 of Table 3. Following the divisional level pattern, models of three types of Punjab's areas were developed: first for irrigated, second for un-irrigated and thirdly by combining both irrigated and un-irrigated areas. Total 90% of Punjab area is irrigated and only remaining 10% is un-irrigated having significant difference in production of wheat. So same parameters for comparison of modeling have been used as in case of divisional level modeling.

Fourth column of Table 3 shows that value of R^2 against WtRains is more than that of TotalRains and more importantly MSE of WtRains is less than that of TotalRains in all level modeling.

Here also individual rainfalls of important months of Rabi are insignificant. Like in case of Irrigated area model, rainfalls of Nov, Dec and Mar are insignificant, in case of Un-irrigated area model rainfalls of almost all the Rabi months are insignificant and in the overall model of Punjab, rainfalls of Nov and Mar are insignificant, which is not parallel to the ground reality. So model using individual rainfalls of Rabi months as predictors is not suitable for projection of yield of wheat though it has slightly better value of R^2 and MSE.

In provincial level modeling, two important criteria, Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC), (Gujarati, 2003), for the selection of competing models have been used. These both criteria impose a harsher penalty on involving more number of regressor in the model than Adj R^2 . Comparing AIC and SIC, later imposes more harsher penalty on number of regressor in the model than first one. A model with the lowest value of AIC / SIC is preferred among its competing models. Main advantage of these criteria is that the model with least value of AIC / SIC is also useful for not only in-sample but also out-sample forecasting performance, (Gujarati, 2003)

Fifth and sixth column of Table 3 bear values of AIC and SIC for each corresponding model, respectively. In all the WtRain models, values of AIC and SIC are lower than the model with TotalRain, i.e. these both criteria also indicate like R^2 and MSE that model with WtRain is better than TotalRain. Also these criteria support the final selection of the model on Punjab level using WtRain as a regressor.

Excluding the model of AllRains, in all the rest of models, WtRains model is better than TotalRains in all respects. For example, in both types of models of Irrigated and Un-irrigated areas, R^2 is higher, MSE is less and Durbin-Watson test value is more than 1 in case WtRains as compared to TotalRains. So on the basis of maximum value of R^2 , least value of MSE and lower value of AIC / SIC, the WtRains model against overall Punjab area is the best one and is selected as the final acceptable model for the estimation of wheat production.

By applying different regression diagnostics, the quality of the final model can be improved i.e. value of R^2 can be increased from 0.386, MSE can be decreased from 5.234 and ultimately better estimates would be achieved.

2.5 Final Selection of Model for Wheat Projection

An ANOVA table and other important information of the proposed model before applying any regression diagnostic is given in the Table 4.

As the value of P is 0.000, which depicts that the model is significant / appropriate. The value of MSE is 5.234, R^2 is 0.386 and value of Durbin Watson test for autocorrelation is 1.015 that shows no any serious problem of autocorrelation in the response variable.

Now some more assumptions of Multiple Linear Regression are verified and regression diagnostics are applied so that quality of the proposed model may be improved and consequently a final version of operational model would be achieved.

2.5.1 Constant Variance

In Multiple Linear Regression, an important assumption is that the response variable must have a constant variation. Figure 2 shows variation behavior in the data.

Figure 2 shows that there is a fixed or constant variance in the response variable, i.e. yield of wheat is equally dispersed throughout the sample data of 25036 plots. It is depicted by a rectangular shape of response variable in its graphical presentation, i.e. there is no upward / downward / cyclical trend in the data.

2.5.2 Outliers

The Figure 3 shows a diagrammatic presentation of values of standardized residuals by fitting a Multiple Linear Regression Model on the given data.

As depicted in Figure 3, there are some values out of 25036 observations that are beyond the limits of ± 3 , which are outliers. By omitting these outliers from the data, the Figure 4 is obtained. (**Fig: 2.4 here**)

Now these total observations are 24902, i.e. 134 (only 0.535%) values have been dropped being the outliers. By removing the outliers, the quality of projection by the proposed model is improved as depicted in the ANOVA Table 5.

The new improved model is also a good fit as value of $P < 0.05$, value of MSE is reduced to 4.941, R^2 is increased to 0.403 and value of Durbin Watson test is increased to 1.029. These are the indicators that exclusion of outliers from that data has improved the quality of the model.

2.5.3 Influential / Leverage Values

An influential or leverage value may have a great impact on the overall quality of prediction by the model. Also illogical signs of coefficients of predictor's variables, which cannot be interpreted with reference to the real ground situation, unsustainable coefficients and dropping an important variable being statistically insignificant, are all due to the influential values.

Accounting for the importance of influential values, two techniques have been used to detect them: Cook's Distance and Covariance Ratios.

Total 760 (3.04%) out of 25036 observations are discarded being the influential values on the basis of their covariance ratios. The ANOVA Table 6 shows the changes in the proposed model.

This model is also significant or a good fit as P-value is less than 0.05. Values of both MSE and R^2 are improved and they are 4.199 and 0.449 respectively. Value of R^2 is directly proportional to the variation in regressor variables, (Hahn, 1973). As in the study 21 regressor variables are categorical, so value of R^2 cannot be increased substantially from its current value. Result of Durbin Watson test is also better. On the whole, by extracting influential values using covariance ratio technique, the quality level of the model is improved. Where as values of R^2 for Multi Sensor Estimates (MSE) and QC_Coop estimate using Multiple Regression Analysis for Corn Yield and Rainfalls are 0.24 and 0.22, respectively, (Westcott, 2003)

2.5.4 Normality of Residuals

One of the assumptions of Multiple Linear Regression is that errors terms must be normally distributed, i.e. there must be no particular trend in the distribution of error terms. The Figure 5 shows histogram and P-P plot of error terms.

Both histogram and P-P plot of standardized residuals show that there is no serious issue of abnormality in the error terms. All the points / observations are dispersed along the straight line in the P-P plot that shows that neither of the tail of normal curve is thick and it is also illustrated by the histogram of Regression Standardized Residuals.

2.5.5 Multicollinearity

Multicollinearity refers to the linear dependency in the regressor variables, which makes the regression coefficients unstable, illogical / immature results of statistical inferences and unrealistic signs with the regression coefficients. Due to existence of multicollinearity, model coefficients are dramatically changed with the change of sample.

The most widely used diagnostic for multicollinearity is Variance Inflation Factor (VIF). The value of VIFs should be as low as possible for the indication of absence of multicollinearity. A value of VIF more than 10 shows a serious problem of multicollinearity, Montgomery (2003). Referring Table 7 of Final Model, the last column shows value of VIF against each regressor variable. It is revealed that no value of VIF is greater than 10, rather not greater than 8. It is a clear indication that there is no serious problem of multicollinearity in the model i.e. all explanatory variables are linearly independent.

2.6 The Final Model

Following the initial important Multiple Linear Regression diagnostics, now the final model is presented in the Table 7.

Out of total 37 explanatory variables, 33 variables are statistically proved significant as column six of Table 7 bears P-value of each regressor variable. No P-value is greater than 0.05 i.e. these variables really contribute to the variation in the response variable; Yield of wheat. Also Standard Errors of regression coefficients are nominal i.e. all values of standard errors are less than 1 that shows their stability.

2.6.1 Interpretation of Regression Coefficients

Now each coefficient of regressor variable is independently interpreted.

a) Constant

The model does not pass through the origin i.e. it has a non-zero y-intercept, means in any case yield of wheat will not be nil if all other parameters are fixed except in the case of irregular variation like flood, hailing or fire etc, which cannot be accommodated in the model.

It also logically true that, excluding irregular variations, yield of a plot cannot be zero that's why a model without intercept has not been selected as a final model for wheat projection as earlier discussed in the Section: 2.2

According to the model, keeping all the parameters fixed, yield of a plot will be 2.976 kg per plot of 300 ft² (11.58 m/ac)

b) Un-irrigated

It was discussed earlier that agricultural land of Punjab is divided in to two categories: irrigated and un-irrigated. As it is a generally known belief / experience that production of wheat in irrigated area is more than that of un-irrigated one, the same is depicted by the coefficient of regressor Un-irrigated. Type of irrigated area is used as a base category and coefficient -1.965 depicts that wheat production is decreased on the average by 1.965 kg per plot (7.64 m/ac or 285.30 kg/ac) when category is switched from Irrigated to Un-irrigated.

c) Variety

There are more than six varieties of seed that are mostly applied for wheat in the Punjab. But half of farmers use Inqlab-91 variety in the province, so it has been used as a base category. The coefficient 0.290 of regressor variable Variety shows that production of other varieties is more than Inqlab-91 i.e. production is increased on the average by 0.290 kg per plot (1.13 m/ac or 42 kg/ac) when category of variety is switched from Inqlab-91 to others.

d) Seed_From

In Punjab, seed of wheat is obtained from two sources; government certified seed and other is own farmers home seed. Only 20% farmers use certified seed of wheat. As a big majority of farmers of the province use home seed, so it has been used as a base category. The coefficient 0.415 of variable Seed_From shows that wheat production is increased on the average by 0.415 kg per plot (1.61 m/ac or 60.25 kg/ac) when category of source of seed is switched from Own Home to Certified.

e) DAP Fertilizer

Two major varieties of fertilizers are used in the Punjab over the crop of wheat: DAP and Urea. These both varieties are included in the study as independent regressor variables. The quantity of fertilizer used has been measured in Kg, not in number of bags. The coefficient against the variable DAP shows that wheat production is increased on the average by 0.022 kg per plot (0.086 m/ac or 3.19 kg/ac) against 1 kg increase in the quantity of fertilizer DAP.

f) Urea Fertilizer

Fertilizers Urea and DAP are applied in a specific proportion on the crop of wheat. The fertilizers are rarely used alone because a combination of both fertilizers returns a better result rather than application of any one of them. The coefficient of Urea depicts that production of wheat is increased on the average by 0.015 kg per plot (0.058 m/ac or 2.18 kg/ac) against 1 kg increase in the quantity of Urea.

g) Number of Plough (No_Plough)

Mostly tractors are used for ploughing in the province. The variable has been recorded as the number of times ploughs were made prior to the sowing time of wheat. A well preparation of soil by ploughing causes a better

production of the crop. The coefficient of No_Plough depicts that yield of wheat is increased on the average by 0.173 kg per plot (0.67 m/ac or 25.12 kg/ac) against one count increase in the number of plough.

h) Number of Level (No_Level)

After ploughing, leveling is also used for even distribution of water, seed and fertilizers throughout the field. A poor leveling may cause poor distribution of inputs and consequently a poor production of the crop. The variable No_Level is also measured as a count of number of leveling applied. Its coefficient reveals that yield of wheat is increased on the average by 0.088 kg per plot (0.34 m/ac or 12.78 kg/ac) against one count increase in the number of levels.

i) Number of Water (No_Water)

Variation in the number of waters applied to the crop is proportional to the availability of timely rainfalls. In case of increase in the timely rainfalls, number of water is decreased and vice versa. The variable No_Water is measured as a counter of total number of water applied to the crop from its sowing to maturity level. Its coefficient says that average production of wheat is increased by 0.117 kg per plot (0.46 m/ac or 16.99 kg/ac) against one count increase in the number of water.

j) Disease

The crop of wheat may face various diseases during its life due to a short / excess input or because of any climatic / soil distortion. There is a considerable difference of yields between healthy and diseased crops and it is same depicted here by the coefficient of variable Disease. It has been measured in Yes (diseased) and No (no disease) and later has been used as base category. The coefficient of variable Disease depicts that production of wheat is decreased on the average by 0.262 kg per plot (1.02 m/ac or 38.04 kg/ac) when switching from healthy to diseased crop.

k) Spray

Spray of pesticides for particular diseases or weeds is essential when the crop is facing such a problem in order to get a better production. It is measured with the variable Disease that either the farmer has used any spray on the crop or not. Like Disease variable, it also has been measured in Yes (sprayed) and No (not sprayed), as a base category. Its coefficient depicts that production of the crop is increased on the average by 0.756 kg per plot (2.94 m/ac or 109.76 kg/ac) when switching from no spray to sprayed category. It also emphasizes the utility of pesticides spray on the crop.

l) Weighted Rainfalls (WtRain)

The concept of weighted rainfalls, as discussed earlier, is incorporated in the model and has a coefficient of 0.013, which reveals that against 1 mm increase in the amount of weighted rainfalls causes an increase of 0.013 kg yield of wheat per plot (0.051 m/ac or 1.89 kg/ac). In case of regressor variable TotalRain, its coefficient is 0.003, which reflects that TotalRain causes less increase in the yield as compared to WtRain.

m) Sowing Time (STNov1, STNov2, STDec1, STDec2)

Sowing time of wheat plays an important role in the production of the crop. As its sowing is delayed, the yield is gradually decreased. Same pattern is depicted by the variables related to the sowing time period.

Sowing time is divided in to five categories i.e. up to October (base category), first half of November (STNov1), second half of November (STNov2), first half of December (STDec1) and second half of December and later (STDec2). The Table 8 shows changes in the yield of wheat in different sowing periods with reference to base category.

With in parenthesis positive sign shows direct and negative sign inverse relation between sowing time and quantity of yield. As indicated in the table, first half of November is the best for getting the maximum yield of wheat. As the sowing of the crop is delayed, the yield gradually decreases with respect to the base category of sowing period (up to October) and minimum yield is obtained against the crop sown in the last half of December or later.

These results are very much parallel to the ground reality and according to the instruction / guide line given to the farmers by the Agriculture Department of the Punjab.

n) Variables of Seasonal Temperatures

Production of wheat is extremely dependent on the climatic parameters of the province in which temperatures in the months of the whole Rabi season are very important. Average of maximum and minimum temperatures of each Rabi month have been used as independent regressor variables. Only average maximum temperature of

March (T_Mar_Max) is statistically proved insignificant but rest of all 13 variables regarding temperature are significantly contributing to the yield of the crop.

The Table 9 summarizes the interpretations of the coefficients of temperature variables.

The coefficient against the variable T_Oct_Min (average minimum temperature of October) can be interpreted as against 1°C increase / decrease in the average minimum temperature of October causes a decrease / increase in the yield of wheat by 0.024 kg per plot (0.093 m/ac or 3.485 kg/ac). But relation between average maximum temperature of October (T_Oct_Max) and yield is positive, which can be interpreted that as the variation between maximum and minimum temperatures of month of October increases, the production of wheat also increases and vice versa. Same opposite pattern between maximum and minimum temperatures can be vetted in case of December and February. This pattern of temperatures can be reversed in case of negative temperatures but the data possess no negative temperature, as in the Punjab winter season, temperature rarely goes down from 0°C especially in wheat growing areas.

Temperatures of November, January and April have same directional pattern. Like in case of maximum average temperature of January (T_Jan_Max) yield per plot is increased by 0.085 kg (0.331 m/ac or 12.341 kg/ac) as average maximum temperature is increased by 1°C. Similarly almost same quantity of yield is increased against 1°C increase in average minimum temperature of the month. (T_Jan_Min). It can be collectively interpreted that when wintriness in the month of January increases, yield of wheat is negatively affected. It is also parallel to the real ground situation that in winter season when temperature considerably goes down, a thin layer of frost appears on the top of wheat plants in morning / dawn time, which damages the germination of the crop. It is important to highlight that minimum temperature of the winter season is mostly recorded in the month of January in Punjab.

Average minimum and maximum temperatures of April (T_Apr_Min and T_Apr_Max) have negative coefficients mean that they have inverse relation with the yield of wheat. Coefficient of T_Apr_Min is -0.038, which can be interpreted that as temperature is increased by 1°C, the yield of an experimental plot is decreased by 0.038 kg (0.148 m/ac or 5.517 kg/ac). Same change is occurred in case of maximum temperature of the month.

It is also very much parallel to real ground observations. At the end of March and start of April, wheat crop is in its last maturity level and germination of grain is on its peak as it gains weight and bigger size during this period. If temperature rises unusually during this time slot, germination of grain is instantaneously stopped, it gets dried in minor weight and size and consequently production of the crop is decreased. Same problem occurred in Rabi season 2010-11 when production of wheat was considerably decreased as compared to the previous years as maximum temperature rose to 36°C from an average routine temperature of 25°C during this time period. A moderate temperature during April causes a better production of wheat.

o) Variables of Seasonal Humidity

Average humidity level (proportion of moisture in air) of each Rabi month has been used as an independent regressor variable in the model. Humidity of two Rabi months February and March (Humidity_Feb & Humidity_Mar) were proved statistically insignificant and ultimately dropped out of the study. Humidity is also one of the parameters on which wheat production is based. The Table 10 bears a summary of interpretation of significant humidity variables.

Humidity of October / January has positive and November / December / April has negative impact on the yield of wheat. The major impact of humidity on yield is of January and April. The coefficient against variable Humidity_Jan can be interpreted as against 1% increase in humidity causes 0.032 kg increase in the yield of wheat per plot (0.124 m/ac or 4.646 kg/ac). Similarly maximum change against Humidity_Apr is interpreted as; there is 0.034 kg decrease in the yield per plot (0.132 m/ac or 4.936 kg/ac) against 1% increase in humidity of the month.

A main cause of increase in humidity is rainfall. Rainfall of January has maximum weight i.e. January rainfall has maximum positive impact on the production of wheat as compared to rest of all the Rabi months. Same pattern can be observed in case of January humidity as it also has maximum positive impact on the yield. Both results of weighted rainfall and humidity of January validate each other.

Harvesting of wheat is started from the mid of April and is completed at the end of April or maximum in the first week of May. For harvesting of the crop, a dry and hot weather is essential. When crop is ready for harvesting, a cats and dogs rainfall and increase in humidity cause damage to the crop as it lays down along the ground and production is loosed. Depicting the real ground situation, Humidity_Apr has maximum negative coefficient.

2.7 Validation of the Model

After development of the model and conduction of its all-necessary diagnostics, validity of the model is examined before putting it into an operational form. Validity means its behavior in making prediction of the yield for given set of the data, discrepancies raised in its working form and quality of prediction for the new set of data.

In this section, a comparison between actual yield and projected yield by the model for given set of data on districts, divisional and provincial levels independently for irrigated and un-irrigated areas is made and quality of projection is vetted. Further to observe the behavior of the model for new set of data, wheat data of Sargodha division comprising Sargodha, Khushab, Nianwali and Bhakhar districts for the year 2009-10 is used and comparison between actual and predicted values is made.

2.7.1 Analysis of Magnitudes / Signs of Coefficients of the Model

In the first phase of validation of a model, magnitudes and signs of the coefficients of the model are analyzed / compared with the prior experience, physical theory and their realistic status, (Snee, 1977).

In the whole model no sign of a coefficient of a regressor variable is unrealistic and unstable i.e. they are all according to the physical theories of agricultural science, prior experiences conducted on small scales and general perception regarding the production of wheat. In the individual interpretation of each regressor variable logical meaning of signs has been discussed.

Magnitude wise all coefficients are reasonable and no coefficient has shown an extraordinary increase / decrease. For more comparable and compact comprehension, amount of each coefficient has been expressed in per plot, m/ac and kg/ac and nothing has shown exceptional value.

Stability of the coefficients can be examined by vetting their respective standard error S.E (β) and the value of VIF given in the fourth and last column of Table 7, respectively.

The maximum S.E (β) is 0.076 (only 5.5% variation) against the regressor variable STDec2 (sowing time of wheat in the second half of December), which, in spite of highest standard error, is a reasonable value. All others regressor variables have much less standard errors than 0.076 that reflect a sound stability of the coefficients of the final model.

Similarly, the maximum value of VIF for multicollinearity check is 7.252 against the regressor variable Humidity_Dec, which is an acceptable value as a VIF value greater than 10 reflects a serious problem of multicollinearity, (Montgomery, 2003). Rests of all regressor variables have VIFs much less than 7, which is an indication of stability of the coefficients of the model.

2.7.2 Confirmation Runs

In this phase of validity checking of the model, its predictive performance is examined for the given set of data taking different dimensions like its projection behavior for the whole data, for various divisional data segments, for irrigated / un-irrigated areas and for fresh set of data. A reasonable projection level of estimates reflects an acceptable operational use of the model.

a) Confirmation Runs for the Whole Data

Table 11 bears a comparison of actual average yield of wheat (m/ac) with projected yield by the model on district, divisional and provincial levels independently for irrigated / un-irrigated areas.

The overall Squares of Residuals (SR) shows that the model predictive performance is good. On the Punjab level SR is 0.004 including both irrigated and un-irrigated areas, which is quite good. For irrigated and un-irrigated areas independently, it is 0.000 and 0.051 respectively, which reflects that model predictive performance is better in irrigated case as compared to un-irrigated one.

On divisional level, the highest SR is 1.638 in case of D.G.Khan division; in spite of maximum value, it is acceptable SR. Where as on district level, SSR is in two digits only in case of Narowal and Okara districts out of 35 districts i.e. 11.244 and 13.144 respectively. In rest of all the districts, it is less than 10.

It is observed that bigger value of SR is happened only in the case where the number of data points is comparatively less. For instance Rawalpindi and Chakwal districts mostly comprise un-irrigated areas and a few sample points fall in irrigated areas that's why SR for irrigated area is 23.267 and 76.470 and for un-irrigated area 1.317 and 0.011 respectively. It is clearly revealed from the results that the model is behaving well in case of sufficient number of data points.

b) Confirmation Runs for Yearly Data

The Table 12 contains a comparison of estimates of wheat by the model with its actual production segregated on the basis of different years.

On the Punjab level SR is below 1 in all the years i.e. projection quality of the model for each independent year is also quite reasonable. The minimum value of SR, 0.044 is against 2005-06 and maximum 0.570 is against 2008-09 but the difference between them is minimal i.e. it can be said that the model is equally good for each year. Similarly, wheat estimates by the model, independently for irrigated and un-irrigated areas for each year, are mostly less than 1 i.e. the model is equally a good predictor for both types of areas.

c) Confirmation Runs for Fresh Data

To vet the projection behavior of the model for new set of data, wheat production data of 786 randomly selected plots from Sargodha division comprising four districts namely Sargodha, Khushab, Mianwali and Bhakhar for the year 2009-10 has been used. The Table 13 bears divisional and districts estimates of wheat by the model along with the actual production results.

The main problem occurred with the wheat production in the year 2009-10 that temperature suddenly raised unexpectedly at the end of March causing the grain of wheat got dried in the small size i.e. germination of wheat grain suddenly stopped prior to its routine maturity level and resultantly, the production of the crop decreased unusually. It was a quite rare variation, which usually occurs once in a decade. The model well accommodates all the routine variations of the Rabi season and returns good estimates of wheat irrespective of any particular year and zone.

3. Conclusions

The objective of the study has been achieved. A subjective approach has been converted in to an objective one.

The proposed model is independent of time binding i.e. it might be valid for any year, as it has been observed in model validity Section 2.7. By inputting parameters of a particular year, it returns estimates of wheat for the year. It is important to highlight that in each year, regressor variables involved in the model are not dramatically changed that's why average yield of wheat per acre remains almost consistent with a little bit variation each year. The total production of wheat of the province mainly depends on the total area under cultivation of the crop each year.

The proposed model does not behave ideally in estimating yield in case of irregular variations, which occur rarely. So its predictive performance for a routine year with usual inputs is excellent.

This research work is an effort to develop a statistical model for projection of wheat crop independent of time binding using a concept of weighted rainfalls. Similar research efforts can be conducted for other major and minor crops. These models should be independent of time and capable of yield production of their respective crops in all the years.

As discussed earlier, amount of rainfalls in each month of the season of a crop plays a significant role in its production. But timely rainfall, may be small in amount, contributes a lot in the production of a crop that's why a concept of weighted rainfall has been used in the proposed model for wheat. Similarly weighted pattern for rainfalls, temperature and humidity level can be determined for all the crops.

References

- Gujarati D.N. (2003). *Basic Econometrics*, (4th ed.). New York: McGraw-Hill Publishing.
- Hahn G.J. (1973). The coefficient of determination exposed! *Chem. Technol*, 3, 609-614.
- Hahn G.J. (1979). Fitting regression models with no intercept term, *J. Qual. Technol*, 9(2), 56-61.
- Montgomery D.C., Martin E.W. & Peck E.A. (1980). Interior analysis of the observations in multiple linear regression, *J. Qual. Technol*, 12(3), 165-173.
- Montgomery D.C. (2001). *Introduction to linear regression analysis*, (3rd ed.). New York: Wiley Publishing.
- Qayyum A. & Pervaiz M.K. (2010). Impact of weighted rainfalls on the yield of wheat in the Punjab, Pakistan, *African Journal of Agricultural Research*, Vol. 5(22), pp 3017-3025.
- Snee R.D. (1977). Validation of regression models: Methods and examples, *Technometrics*, 19, pp 415-428. <http://dx.doi.org/10.2307/1267881>
- Westcott N.E, Hollinger S.E. & Kunkel K.E. (2003). Relationship between Normalized Corn Yields and Monthly Rainfall for the Midwestern United States, *Illinois State Water Survey*, Champaign, Illinois.

Table 1. Comparison of models with / without intercepts

Model Type	Models Comparison					
	Rainfalls					
	Weighted			Total		
	R ²	MSE	Model Sig.	R ²	MSE	Model Sig.
With Intercept	0.385	5.241	0.000	0.383	5.258	0.000
Without Intercept	0.924	5.255	0.000	0.924	5.273	0.000

Table 2. Comparison of divisional models of the Punjab

Comparison of All Divisional Models									
Divisions	Sample Points	Rain variables	R ²	DW	MSE	Model Sig.	Coeff. In Model	No. of Sig. Variables	Insig. Rain Variables
Gujranwala	4120	TotalRains	0.383	1.029	4.768	0.000	-	27	TotalRains
		WtRains	0.385	1.030	4.761	0.000	0.003	27	-
		AllRains	0.374	1.011	4.827	0.000	-	19	Nov, Dec, Jan, Feb
Rawalpindi	2016	TotalRains	0.454	0.977	3.189	0.000	0.001	15	-
		WtRains	0.454	0.978	3.189	0.000	0.009	16	-
		AllRains	0.452	0.972	3.200	0.000	-	15	Oct, Nov, Dec, Jan, Feb
Lahore	2988	TotalRains	0.326	1.140	5.263	0.000	-	18	TotalRains
		WtRains	0.326	1.140	5.263	0.000	-	18	WtRains
		AllRains	0.327	1.145	5.252	0.000	-	17	Oct, Nov, Dec, Jan, Feb
Faisalabad	2832	TotalRains	0.257	1.293	5.397	0.000	0.003	13	-
		WtRains	0.257	1.293	5.399	0.000	0.009	13	-
		AllRains	0.258	1.296	5.391	0.000	-	15	Oct, Nov, Jan, Apr
Sargodha	3090	TotalRains	0.458	0.989	3.773	0.000	0.003	20	-
		WtRains	0.458	0.989	3.774	0.000	0.012	20	-
		AllRains	0.457	0.988	3.774	0.000	-	19	Oct, Dec, Feb, Apr
Multan	4794	TotalRains	0.290	1.114	5.310	0.000	0.009	22	-
		WtRains	0.291	1.115	5.300	0.000	0.031	20	-
		AllRains	0.295	1.119	5.279	0.000	-	21	Nov, Jan, Feb, Apr
Bahawalpur	2622	TotalRains	0.319	1.172	4.925	0.000	0.019	15	-
		WtRains	0.321	1.175	4.913	0.000	0.066	16	-
		AllRains	0.322	1.175	4.911	0.000	-	16	Nov, Jan, Mar, Apr
DG Khan	2574	TotalRains	0.307	1.061	4.782	0.000	0.005	16	-
		WtRains	0.309	1.061	4.775	0.000	0.014	18	-
		AllRains	0.311	1.064	4.762	0.000	-	21	Nov, Apr

Table 3. Comparison of provincial models of the Punjab

Comparison of All Provincial Models											
Area	Sample Points	Rain variables	R ²	AIC	SIC	DW	MSE	Model Sig.	Coeff. In Model	No. of Sig. Variable	Insig. Rain Variable
Irrigated	22428	TotalRains	0.245	1.6995	1.7106	1.022	5.464	0.000	0.003	30	-
		WtRains	0.248	1.6962	1.7069	1.024	5.446	0.000	0.013	29	-
		AllRains	0.251	1.6918	1.7051	1.027	5.422	0.000	-	33	Nov, Dec, Mar
Un-irrigated	2608	TotalRains	0.358	1.1368	1.1773	0.957	3.094	0.000	0.003	17	-
		WtRains	0.365	1.1303	1.1798	0.971	3.065	0.000	0.004	20	-
		AllRains	0.380	1.1072	1.1634	0.992	2.996	0.000	-	20	Oct, Nov, Dec, Jan,
Punjab	25036	TotalRains	0.384	1.6594	1.6698	1.013	5.250	0.000	0.003	31	-
		WtRains	0.386	1.6562	1.6669	1.015	5.234	0.000	0.013	32*	-
		AllRains	0.389	1.6514	1.6637	1.019	5.207	0.000	-	36	Nov, Mar

*Selected model for operational use

Table 4. ANOVA Table & other Information

A N O V A						Explained Variation		Durbin Watson
Model	Sum of Squares	df	Mean Square	F	Sig.	R ²	Adj R ²	
Regression	82099.80	32	2565.619	490.226	0.000	0.386	0.385	1.015
Residuals	130854.20	25003	5.234	-	-	-	-	-
Total	212954.00	25035	-	-	-	-	-	-

Table 5. ANOVA Table & other Information

A N O V A						Explained Variation		Durbin Watson
Model	Sum of Squares	df	Mean Square	F	Sig.	R ²	Adj R ²	
Regression	82875.39	31	2673.40	541.07	0.000	0.403	0.402	1.029
Residuals	122880.90	24870	4.941	-	-	-	-	-
Total	205756.30	24901	-	-	-	-	-	-

Table 6. ANOVA Table & other Information

A N O V A						Explained Variation		Durbin Watson
Model	Sum of Squares	df	Mean Square	F	Sig.	R ²	Adj R ²	
Regression	83085.99	33	2517.76	599.63	0.000	0.449	0.449	1.083
Residuals	101788.00	24242	4.199	-	-	-	-	-
Total	184874.00	24275	-	-	-	-	-	-

Table 7. The final model for operational use

S. No	Variables	MODEL				Collinearity Statistics	
		Unstandardized Coefficients		t	Sig	Tolerance	VIF
		β	Std. Error				
1	Constant	2.976	0.300	9.917	0.000	-	-
2	Unirrigated	-1.965	0.072	-27.382	0.000	0.353	2.834
3	Variety	0.290	0.029	10.135	0.000	0.848	1.180
4	Seed_From	0.415	0.036	11.360	0.000	0.927	1.079
5	DAP	0.022	0.001	36.151	0.000	0.753	1.328
6	Urea	0.015	0.000	30.608	0.000	0.556	1.800
7	No_Plough	0.173	0.011	15.518	0.000	0.609	1.641
8	No_Level	0.088	0.016	5.621	0.000	0.644	1.553
9	No_Water	0.117	0.012	9.496	0.000	0.392	2.549
10	Disease	-0.262	0.049	-5.327	0.000	0.909	1.100
11	Spray	0.756	0.029	25.848	0.000	0.837	1.194
12	WtRain	0.013	0.001	18.410	0.000	0.503	1.987
13	STNov1	0.209	0.061	3.430	0.001	0.215	4.652
14	STNov2	-0.174	0.062	-2.790	0.005	0.190	5.269
15	STDec1	-0.645	0.069	-9.368	0.000	0.282	3.551
16	STDec2	-1.370	0.076	-17.978	0.000	0.397	2.521
17	T_Oct_Min	-0.024	0.007	-3.405	0.001	0.397	2.517
18	T_Oct_Max	0.029	0.008	3.757	0.000	0.422	2.370
19	T_Nov_Min	0.016	0.007	2.330	0.020	0.426	2.346
20	T_Nov_Max	0.030	0.008	3.718	0.000	0.280	3.568
21	T_Dec_Min	0.114	0.011	10.648	0.000	0.279	3.584
22	T_Dec_Max	-0.072	0.010	-7.504	0.000	0.225	4.441
23	T_Jan_Min	0.083	0.013	5.563	0.000	0.257	3.888
24	T_Jan_Max	0.085	0.007	12.261	0.000	0.248	4.038
25	T_Feb_Min	-0.046	0.009	-4.939	0.000	0.286	3.496
26	T_Feb_Max	0.017	0.008	2.090	0.037	0.195	5.135
27	T_Mar_Min	0.074	0.010	7.286	0.000	0.374	2.673
28	T_Apr_Min	-0.038	0.009	-4.072	0.000	0.269	3.711
29	T_Apr_Max	-0.038	0.007	-5.317	0.000	0.340	2.943
30	Humidity_Oct	0.012	0.003	4.570	0.000	0.257	3.885
31	Humidity_Nov	-0.010	0.002	-4.548	0.000	0.235	4.259
32	Humidity_Dec	-0.017	0.003	-5.866	0.000	0.138	7.252
33	Humidity_Jan	0.032	0.002	13.558	0.000	0.224	4.472
34	Humidity_Apr	-0.034	0.002	-20.354	0.000	0.391	2.559

Table 8. Sowing time variables interpretation

Sowing Time Variables	Change in Yield w.r.t Base Category (up to Oct)		
	Per Plot (kg)	Per Acre (m)	Per Acre (kg)
STNov1	(+) 0.204	0.794	29.62
STNov2	(-) 0.183	0.711	26.57
STDec1	(-) 0.647	2.517	93.94
STDec2	(-) 1.380	5.368	200.36

Table 9. Interpretation of temperature variables

Temp Variables	Change in Yield against 1 °C Change in Temperature		
	Per Plot (kg)	Per Acre (m)	Per Acre (kg)
T_Oct_Min	(-) 0.024	0.093	3.485
T_Oct_Max	(+) 0.028	0.109	4.065
T_Nov_Min	(+) 0.014	0.054	2.033
T_Nov_Max	(+) 0.030	0.117	4.356
T_Dec_Min	(+) 0.117	0.455	16.987
T_Dec_Max	(-) 0.069	0.268	10.018
T_Jan_Min	(+) 0.085	0.331	12.341
T_Jan_Max	(+) 0.083	0.323	12.051
T_Feb_Min	(-) 0.048	0.187	6.969
T_Feb_Max	(+) 0.019	0.074	2.759
T_Mar_Min	(+) 0.075	0.292	10.889
T_Apr_Min	(-) 0.039	0.152	5.662
T_Apr_Max	(-) 0.038	0.148	5.517

Table 10. Interpretation of humidity variables

Humidity Variables	Change in Yield Against 1% Change in Humidity		
	Per Plot (kg)	Per Acre (m)	Per Acre (kg)
Humidity_Oct	(+) 0.012	0.047	1.742
Humidity_Nov	(-) 0.010	0.039	1.452
Humidity_Dec	(-) 0.017	0.066	2.468
Humidity_Jan	(+) 0.032	0.124	4.646
Humidity_Apr	(-) 0.034	0.132	4.936

Table 11. Comparison of actual and projected districts average yields

Estimates Comparison of Wheat Production in Punjab 2005-09										
Divisions /Districts	Sample Points	Average Yield (m/ac)						Residuals ²		
		Estimates by Model			Actual					
		Irri	Unirri	Total	Irri	Unirri	Total	Irri	Unirri	Total
Gujranwala	4120	30.58	18.28	29.46	31.12	16.96	29.86	0.297	1.719	0.161
Gujranwala	864	32.81	-	32.81	35.20	-	35.20	5.742	-	5.742
Gujrat	576	28.37	17.17	24.57	25.95	15.88	22.60	5.817	1.658	3.905
Sialkot	846	28.58	16.88	28.23	30.79	18.28	30.44	4.882	1.961	4.848
Narowal	598	30.01	20.18	27.63	26.15	18.28	24.28	14.891	3.604	11.244
Hafizabad	594	32.22	15.63	31.98	32.42	15.75	32.20	0.043	0.015	0.046
M B Din	642	30.29	-	30.29	31.45	-	31.45	1.344	-	1.344
Rawalpindi	2016	31.42	16.93	18.26	30.30	17.22	18.42	1.264	0.085	0.025
Rawalpindi	672	26.49	18.47	18.57	21.66	19.62	19.64	23.267	1.317	1.153
Attock	552	30.64	17.51	18.87	32.27	17.69	19.22	2.644	0.033	0.125
Jehlani	228	33.62	17.98	24.58	33.03	15.84	23.08	0.345	4.594	2.266
Chakwal	564	26.25	14.27	14.78	17.51	14.16	14.31	76.470	0.011	0.225
Lahore	2988	33.49	-	33.49	33.99	-	33.99	0.260	-	0.260
Lahore	408	32.89	-	32.89	30.16	-	30.16	7.415	-	7.415
Sheikupura	540	30.96	-	30.96	30.63	-	30.63	0.104	-	0.104
Nankana Sb	432	34.07	-	34.07	34.48	-	34.48	0.167	-	0.167
Kasur	828	33.80	-	33.80	33.52	-	33.52	0.078	-	0.078
Okara	780	34.93	-	34.93	38.56	-	38.56	13.144	-	13.144
Faislabad	2832	32.20	-	32.20	33.08	-	33.08	0.787	-	0.787
Faislabad	1008	32.31	-	32.31	32.48	-	32.48	0.029	-	0.029
Jhang	1062	30.28	-	30.28	32.83	-	32.83	6.492	-	6.492
T.T.Singh	762	34.77	-	34.77	34.24	-	34.24	0.288	-	0.288
Sargodha	3090	29.35	9.28	26.74	27.80	10.50	25.47	2.409	1.501	1.618
Sargodha	936	30.07	-	30.07	29.04	-	29.04	1.063	-	1.063
Khushab	576	25.58	9.62	19.73	25.21	10.99	20.08	0.134	1.886	0.120
Mianwali	624	28.12	8.64	23.88	27.66	9.85	23.83	0.214	1.473	0.002
Bahkhar	954	30.82	9.56	29.60	27.65	10.21	26.63	10.026	0.422	8.833
Multan	4794	33.61	-	33.61	34.15	-	34.15	0.284	-	0.284
Multan	762	31.91	-	31.91	31.77	-	31.77	0.019	-	0.019
Khanewal	882	33.77	-	33.77	34.86	-	34.86	1.186	-	1.186
Vehari	876	31.79	-	31.79	32.94	-	32.94	1.317	-	1.317
Lodhran	666	32.83	-	32.83	33.36	-	33.36	0.280	-	0.280
Sahiwal	906	34.67	-	34.67	33.70	-	33.70	0.931	-	0.931
Pakpattan	702	36.99	-	36.99	38.65	-	38.65	2.759	-	2.759
Bahawalpur	2622	31.73	-	31.73	31.44	-	31.44	0.083	-	0.083
Bahawalpur	648	28.84	-	28.84	31.78	-	31.78	8.671	-	8.671
Bahawalnga	978	33.26	-	33.26	31.75	-	31.75	2.302	-	2.302
R.Y.Khan	996	32.08	-	32.08	30.93	-	30.93	1.317	-	1.317
D.G.Khan	2574	31.35	13.05	31.26	30.05	16.34	29.98	1.678	10.805	1.638
D.G.Khan	534	28.58	13.05	28.21	30.02	16.34	29.66	2.072	10.805	2.083
Rajanpur	534	31.08	-	31.08	30.36	-	30.36	0.518	-	0.518
Layvah	660	31.72	-	31.72	29.23	-	29.23	6.198	-	6.198

Table 12. Comparison of actual and projected yearly divisional average

Year Wise Estimates Comparison of Wheat Production in Punjab											
Years	Divisions	Sample Points	Average Yield (m/ac)						Residuals ²		
			Estimates by Model			Actual					
			Irri	Unirri	Total	Irri	Unirri	Total	Irri	Unirri	Total
2005-06	Gujranwala	1014	29.54	16.27	28.15	30.93	14.65	29.22	1.918	2.631	1.153
	Rawalpindi	522	30.88	14.81	16.17	31.98	15.09	16.58	1.203	0.078	0.167
	Lahore	714	32.31	-	32.31	32.08	-	32.08	0.054	-	0.054
	Faislabad	672	31.02	-	31.02	31.63	-	31.63	0.373	-	0.373
	Sargodha	780	28.49	8.96	26.24	27.25	9.18	25.16	1.540	0.047	1.161
	Multan	1038	32.05	-	32.05	30.74	-	30.74	1.708	-	1.708
	Bahawalpur	702	28.32	-	28.32	27.62	-	27.62	0.490	-	0.496
	D.G.Khan	528	29.18	-	29.18	28.73	-	28.73	0.197	-	0.197
	P U N J A B	5970	30.27	14.26	28.44	30.02	14.22	28.23	0.064	0.002	0.044
2006-07	Gujranwala	1024	30.78	19.98	29.74	30.19	21.69	29.39	0.354	2.956	0.117
	Rawalpindi	534	34.66	19.18	20.59	34.00	20.69	21.89	0.432	2.278	1.668
	Lahore	768	34.02	-	34.02	35.40	-	35.40	1.896	-	1.896
	Faislabad	690	33.09	-	33.09	34.15	-	34.15	1.120	-	1.120
	Sargodha	768	31.80	10.67	29.05	29.04	12.53	26.93	7.585	3.443	4.462
	Multan	1176	34.66	-	34.66	35.43	-	35.43	0.593	-	0.593
	Bahawalpur	498	33.63	-	33.63	35.40	-	35.40	3.133	-	3.133
	D.G.Khan	648	32.29	-	32.29	32.25	-	32.25	0.001	-	0.001
	P U N J A B	6106	32.98	18.05	31.28	33.18	19.66	31.67	0.041	2.581	0.151
2007-08	Gujranwala	1038	29.87	18.40	29.18	30.12	16.07	29.28	0.060	5.448	0.010
	Rawalpindi	522	31.00	16.24	17.95	29.61	14.24	16.00	1.929	4.013	3.783
	Lahore	762	32.35	-	32.35	32.28	-	32.28	0.006	-	0.006
	Faislabad	726	30.62	-	30.62	31.28	-	31.28	0.427	-	0.427
	Sargodha	768	27.02	7.84	24.48	25.00	8.88	22.90	4.076	1.079	2.494
	Multan	1308	31.67	-	31.67	31.44	-	31.44	0.053	-	0.053
	Bahawalpur	708	31.22	-	31.22	29.13	-	29.13	4.380	-	4.380
	D.G.Khan	684	29.41	-	29.41	27.62	-	27.62	3.202	-	3.202
	P U N J A B	6516	30.47	15.09	28.98	29.84	13.56	28.28	0.407	2.349	0.485
2008-09	Gujranwala	1044	32.09	18.64	30.75	33.25	15.46	31.51	1.344	10.101	0.587
	Rawalpindi	438	28.11	17.54	18.32	23.62	18.95	19.29	20.187	1.972	0.946
	Lahore	744	35.28	-	35.28	36.15	-	36.15	0.759	-	0.759
	Faislabad	744	33.99	-	33.99	35.17	-	35.17	1.398	-	1.398
	Sargodha	774	30.16	9.60	27.20	29.95	11.25	27.29	0.043	2.708	0.008
	Multan	1272	35.94	-	35.94	38.52	-	38.52	6.692	-	6.692
	Bahawalpur	714	34.40	-	34.40	34.76	-	34.76	0.128	-	0.128
	D.G.Khan	714	33.65	-	33.65	31.12	-	31.12	6.393	-	6.393
	P U N J A B	6444	33.84	16.24	32.08	34.55	17.00	32.83	0.496	0.575	0.570

Table 13. Comparison of actual and projected average yield for 2009-10 in respect of Sargodha Division

Estimates Comparison of Wheat Production in Sargodha Division 2009-10										
Divisions /Districts	Sample Points	Average Yield (m/ac)						Residuals□		
		Estimates by Model			Actual					
		Irri	Unirri	Total	Irri	Unirri	Total	Irri	Unirri	Total
Sargodha	786	31.75	7.78	29.90	26.51	7.78	24.94	27.458	0.000	24.602
Sargodha	234	31.06	-	31.06	25.42	-	25.42	31.810	-	31.810
Khushab	156	26.90	6.40	22.70	23.98	7.62	20.62	8.526	1.488	4.326
Mianwali	156	35.40	13.93	31.82	28.52	7.13	24.95	47.334	46.240	47.197
Bahkhar	240	32.99	9.20	32.20	27.84	10.53	27.27	26.523	1.769	24.305

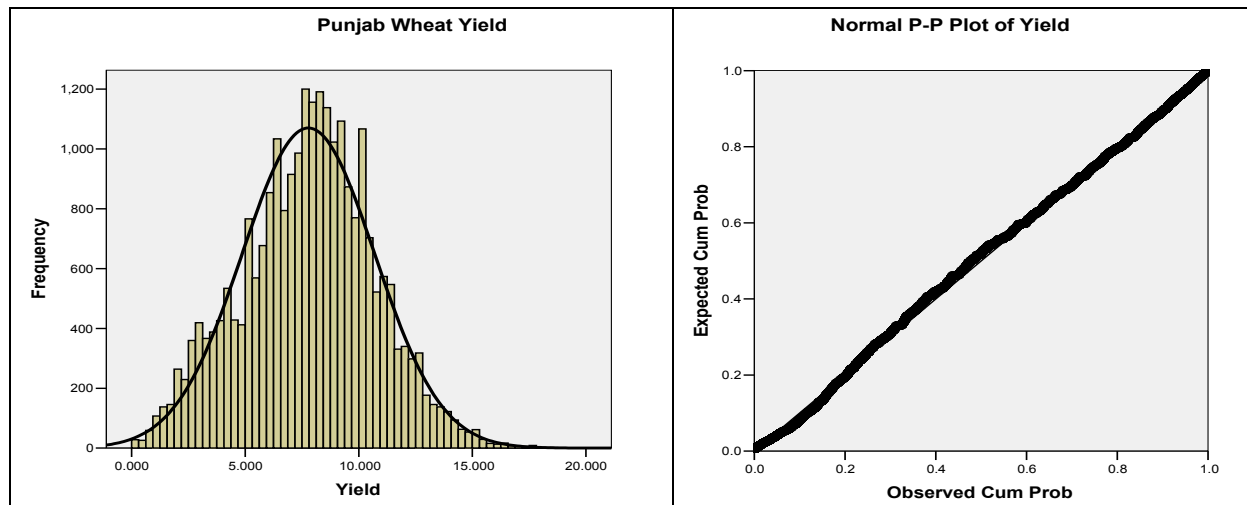


Figure 1. Histogram & P-P Plot of wheat production of Punjab

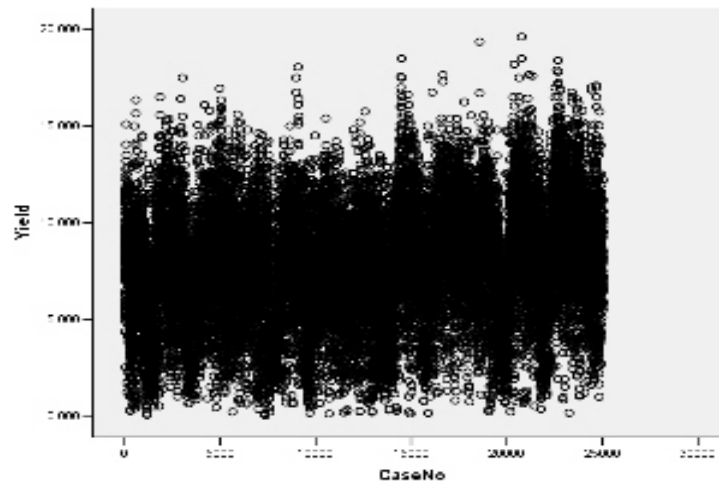


Figure 2. Variation of wheat production in Punjab

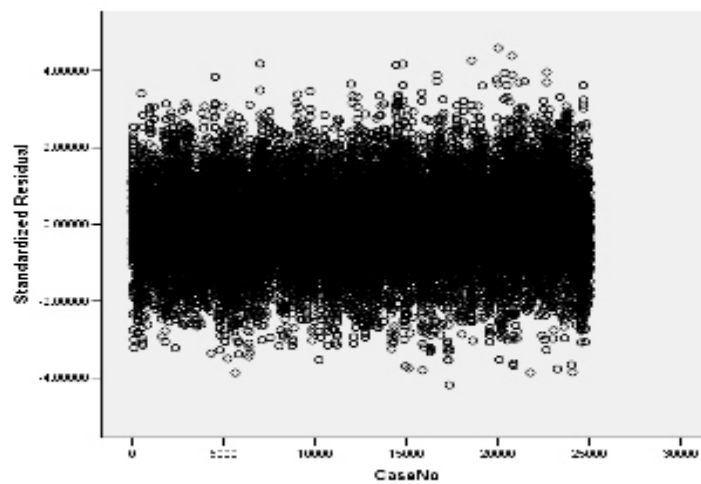


Figure 3. Standardized residuals of wheat estimates

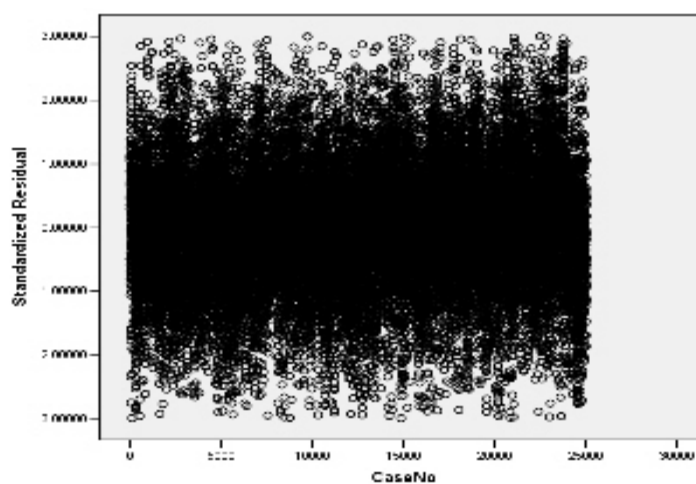


Figure 4. Standardized residuals by excluding outliers

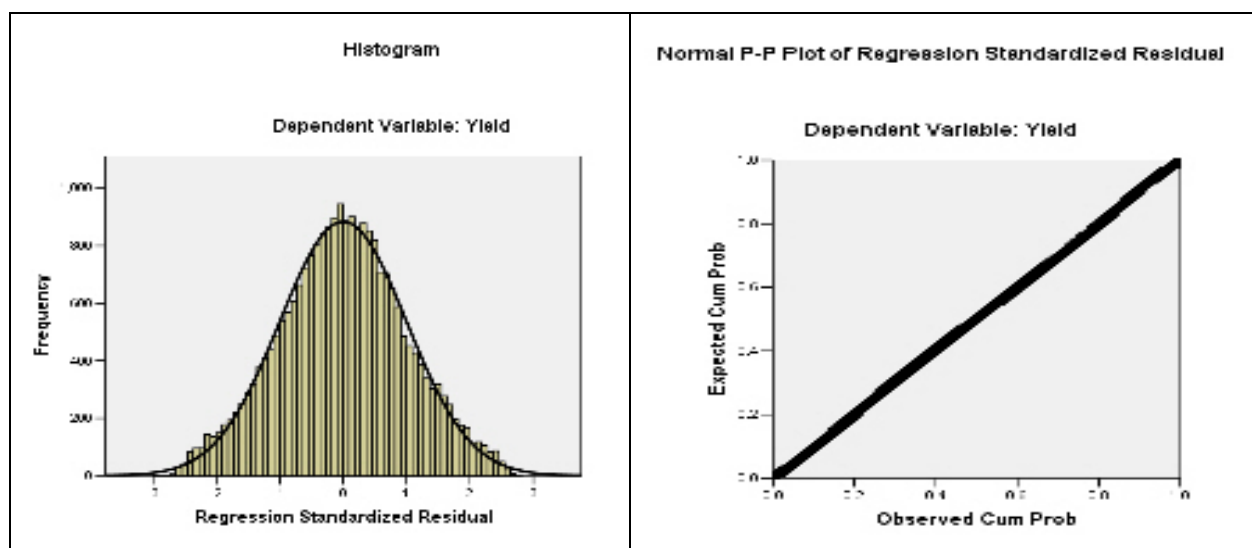


Figure 5. Histogram & P-P Plot of standardized residuals