Checking the Significance of Correlation Coefficient from the Regression Analysis using Wheat Yield.

Ali Akbar Pirzado¹*, Suresh Kumar Suthar², Velo Suthar², Naeem Ahmed Qurashi², Imran Khan Jatoi¹, Suhail Aslam Khaskeli³, Rashid Ali Chandio⁴, Komal Aarain² and Moula Bux Peerzado⁵

¹Department of Statistics, Shaheed Benazir Bhutto University, Shaheed Benazirabad, Pakistan  
²Department of Statistics, Sindh Agriculture University Tandojam, Pakistan  
³Department of General Faculty, Shaheed Benazir Bhutto University, Shaheed Benazirabad  
⁴English Language Development Centre, Mehran University of Engineering and Technology, Pakistan  
⁵Department of Agriculture Economics, Sindh Agriculture University Tandojam, Pakistan  
*Corresponding: ali.akbar@sbbusba.edu.pk.

Article Received 06-02-2021, Revised 23-02-2021, Accepted 20-03-2021.

Abstract

The wheat is the 2nd largest cereal crop in agriculture sector of Pakistan after the livestock sector. Current research presents estimates regression model and Correlation coefficient. The major objectives of the present study were to apply correlation and multiple regression analysis models. Correlation is a statistical technique which shows the level to which or more variables change together. While regression Analysis explores the dependence of one variable on one or more variables and provides the functional relation that is used for the estimating or predicting the significance of the independent variables. The Secondary data of wheat crop was used in this study. The data were collected from the wheat section, Agriculture Research Institute (ARI) Tandojam, regarding various factors influencing on the wheat production. The survey result show that yield has significant positive correlation with the parameters such as the number of tillers per plant, number of seed per spike, length of spike in cm, the number of spikes let per, tiller per hectare and plant height cm. Correlation and regression analysis were associated with the assumption that both variables are covering relations. Correlation coefficient is a measure of linear association between two variables. The correlation coefficient values lie between +1 and always included. It is suggested that the more factors can be studied and different statistical techniques, agronomics practices and different doses of fertilizer can be applied for the high production of wheat crop.

Keywords: Significance, Correlation coefficient, Regression Analysis, Model

Introduction

The wheat is primarily belonging to the Poaceae family. The wheat is the 2nd largest cereal crop in agriculture sector of Pakistan after the livestock sector. In the crop year 2019 and 2020 crop subsector contributed 8.9% to agriculture sector and overall, 1.6 in the GDP of Pakistan. Wheat is the grain crop it provides about half of the calories (48%) great in the daily diet of the people (GOP, 2018). Moreover, as majority of the farmers are face problem of money either they are incognizant of proper production technology or cannot have the funds for to go behind improved production practices (Hussain et al. 2012; Sarfraz et al., 2020). Therefore, analysis of determinants of wheat productivity is very most important. The Punjab province contributes 78% to the wheat production Pakistan. In the province, about 39% of the whole cropped region is to be paid to this crop every year (GOP, 2019). While to explore factor the specific objective of the study was discuss as under.
01. To study the Correlation and Regression Analysis model for wheat Yield.
02. To check the extent of wheat yielding components using the Correlation Coefficient.

**Materials and Method**

The major objectives of the present study were to apply correlation and multiple regression analysis models. The Secondary data were of wheat crop was used in this study. Variety of wheat TD-01 was collected from wheat Section, Agriculture Research Institute, Tando jam. The collected secondary data of wheat crop were analyzed with the help of Statistical Package for Social Sciences (SPSS). Followed by (Pirzado et al., 2021). The collected data were being comprised of about twelve independent variables, and one dependent variable. Independent variables included. Number of tillers/ plant m², (X₁) amount of seed/kernels for every spike, (X₂) period of the spike over in cm, (X₃) number of spikelet for every spike, (X₄) tiller for every hac, (X₅) days to heading in cm, (X₆) Plant tallness in cm, (X₇) lodging score for every hac in cm, (X₈) number of grains per spike, (X₉) seed list 1000 grains weight clinched alongside grams, (X₁₀), N Kg/ha, (X₁₁). Furthermore, per Kg/ha (X₁₂) and seed yield Kg/ha, (Y) were considered as the dependent variable.

**Statistical Analysis**

**Pearson’s Correlation Coefficient:**

Correlation is a statistical technique which shows the level to which or more variables change together. A positive correlation indicates the extent to which those variables increase or decrease.

\[ r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}} \]

**Regression:** The word regression was introduced by English biometrician, Sir Francis Galton in 1871. Regression Analysis explores the dependence of one variable on one or more variables and provides the functional relation that is used for the estimating or predicting the significance of the independent variables. A random variable was understood by dependent variable whereas the predetermined values were assumed by the independent variables i.e. They are special non- randomly. The regression is separated in two types:

1) Simple linear Regression.

2) Multiple Linear Regressions.

**Simple linear Regression:** The linear dependence of two permanent variables examines by Simple Linear Regression one is the response variable (which is usually denoted by “Y”) and other is the predictor (which is mostly represented by “X”). It is possible that the predictor variable to predicted from the response variable for the improved accuracy. In geometrical sense, regression provides the line that best fit the data. This line can then be used to this simple linear regression model can be written in the following mathematical form.

\[ Y_i = \alpha + \beta X_{1,i} + \mu \]

**Multiple linear regressions:** A multiple linear regression model contain further than one predictor variables. The multiple linear regression models of predictor’s variables is given below mathematically.

\[ Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots \beta_n X_{n,i} + \mu \]

**Results**

(Table. 01) shows the descriptive results. Summary Statistics indicates that the mean and std. deviation for Seed Yield (Kg/ha) is 19.63 and 5.458, for number of tillers per plant in meter square was 12.619 and 2.479, while for the number of seed/ kernels per spike was 56.14 and 10.413, for the length of spike in centimeter was 12.262 and 0.736, whereas for the Number of spikelets per spike was 19.382 and 1.775 respectively. The mean and std. deviation for the Tiller per hac was 4491.673 and 36.675, for the Days to heading it was 72.097 and 1.445, likewise for the Plant height in cm is 58.109 and 4.428, for the Lodging score per hac was 69.476
and 1.99, while for the Number of grains per spike is 308.845 and 1184.416, whereas for the seed index for 1000 grains weight in grams was 1.744 and 0.271, for the plot yield grams it was 277.619 and 51.65, respectively. The mean of nitrogen per hectar kilograms is 161.772 with the standard deviation is 6.685 and the mean of phosphorus is 69.275 having standard deviation is 3.229, respectively.

### Table-1 Summary of statistics of different variables:

<table>
<thead>
<tr>
<th>Name of variables</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed Yield (Kg/ha)</td>
<td>13</td>
<td>32</td>
<td>19.63</td>
<td>5.458</td>
</tr>
<tr>
<td>Number of tillers per plant m²</td>
<td>8.00</td>
<td>17.00</td>
<td>12.619</td>
<td>2.479</td>
</tr>
<tr>
<td>Number of seed/kernels per spike</td>
<td>33</td>
<td>76</td>
<td>56.14</td>
<td>10.413</td>
</tr>
<tr>
<td>length of spike in cm</td>
<td>10.90</td>
<td>13.50</td>
<td>12.262</td>
<td>0.0736</td>
</tr>
<tr>
<td>Number of spikelet per spike</td>
<td>16.00</td>
<td>23.00</td>
<td>19.382</td>
<td>1.775</td>
</tr>
<tr>
<td>Tiller per hack</td>
<td>4398.67</td>
<td>4589.98</td>
<td>4916.73</td>
<td>36.675</td>
</tr>
<tr>
<td>Days to heading cm</td>
<td>70.00</td>
<td>75.00</td>
<td>72.097</td>
<td>1.445</td>
</tr>
<tr>
<td>Plant height cm</td>
<td>49.800</td>
<td>65.800</td>
<td>58.109</td>
<td>4.428</td>
</tr>
<tr>
<td>Lodging score per hack cm</td>
<td>65.00</td>
<td>73.00</td>
<td>69.476</td>
<td>1.99</td>
</tr>
<tr>
<td>Number of grains per spike</td>
<td>40.90</td>
<td>5478.00</td>
<td>308.845</td>
<td>1184.416</td>
</tr>
<tr>
<td>Seed index 1000 grains weight in gms</td>
<td>1.23</td>
<td>2.06</td>
<td>1.744</td>
<td>0.271</td>
</tr>
<tr>
<td>Plot yield (g)</td>
<td>181.00</td>
<td>377.00</td>
<td>277.619</td>
<td>51.65</td>
</tr>
<tr>
<td>Nitrogen per hac Kg</td>
<td>147.00</td>
<td>168.00</td>
<td>161.772</td>
<td>6.685</td>
</tr>
<tr>
<td>Phosphorus per hac Kg</td>
<td>58.20</td>
<td>73.00</td>
<td>69.275</td>
<td>3.227</td>
</tr>
</tbody>
</table>

The F-value in the (Table-2) shows that the independent variables statistically significantly predict the dependent variable because the calculated value is much larger than the α-quintile of the F-distribution i.e. (2, 7) =3.794, P<0.05.

### Table- 2. Analysis of variance (ANOVA):

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>d.f</th>
<th>SS</th>
<th>MS</th>
<th>F-Value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>13</td>
<td>536.025</td>
<td>41.2327</td>
<td>4.45</td>
<td>0.0205</td>
</tr>
<tr>
<td>Residuals</td>
<td>8</td>
<td>74.198</td>
<td>9.2748</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>610.223</td>
<td>----------</td>
<td>--------</td>
<td>------</td>
</tr>
</tbody>
</table>

The results of above table describes that the coefficient of multiple correlation (R= 0.936) reveals that there is positive relationship between the dependent and independent variables. The significance of coefficient of determination (R² = 0.87) indicates that 86% of the variances in the dependent variable is explained by the fitted regression model which is good enough to comprise that the model performs well in term of prediction. Similarly, the value of adjusted R² (0.645) shows with the present adjustment of the parameter in the
regression model the model accounts nearly for same variation in the data set without adjusting parameter.

**Table- 3. R, R², Adjusted - R² and STD. Error Estimated Model**

<table>
<thead>
<tr>
<th>R</th>
<th>R- Square</th>
<th>Adjusted R- Square</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.936</td>
<td>0.8784</td>
<td>0.6808</td>
<td>3.045</td>
</tr>
</tbody>
</table>

The table- 4 shows that the parameters of wheat the constant of standard error are (121.538). The parameter of wheat number of tiller per plant meter square standard error is (0.569), Standardized Coefficients Beta is (0.300), T value is (-0.759) and with positively significant is (0.472) The parameter of Number of seed/kernels’ per spike standard error is (0.135), standardized coefficient beta is (0.124), T-value of (1.162) with weakly significant (0.283*). The parameter of wheat is nitrogen per hector kilogram with standard error is (0.154), standardized coefficient beta is (0.236), is T-value is (-0.663) with the significant of (0.140). The parameter of Phosphorus per hac Kg with standard error is (0.363), standardized coefficient of beta is (-0.072), T-value is this parameter is (1.255), with the significant of (0.250)

**Table-4 Wheat parameter, Coefficient, STD Error, T and Significant using multiple regression analysis:**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Std Error</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-99.476</td>
<td>115.99</td>
<td>-0.86</td>
<td>0.416</td>
</tr>
<tr>
<td>Number of tillers per plant m²</td>
<td>0.816</td>
<td>0.569</td>
<td>1.43</td>
<td>0.189</td>
</tr>
<tr>
<td>Number of seed/kernels’ per spike</td>
<td>0.081</td>
<td>0.107</td>
<td>0.75</td>
<td>0.472</td>
</tr>
<tr>
<td>length of spike in cm</td>
<td>1.092</td>
<td>1.704</td>
<td>0.64</td>
<td>0.539</td>
</tr>
<tr>
<td>Number of spikelet per spike</td>
<td>1.135</td>
<td>0.5172</td>
<td>2.20</td>
<td>0.059</td>
</tr>
<tr>
<td>Tiller per hac</td>
<td>0.0419</td>
<td>0.024</td>
<td>1.73</td>
<td>0.121</td>
</tr>
<tr>
<td>Days to heading cm</td>
<td>-0.426</td>
<td>0.650</td>
<td>-2.19</td>
<td>0.059</td>
</tr>
<tr>
<td>Plant height cm</td>
<td>0.081</td>
<td>0.203</td>
<td>0.40</td>
<td>0.697</td>
</tr>
<tr>
<td>Lodging score per hack cm</td>
<td>-0.504</td>
<td>0.662</td>
<td>-0.76</td>
<td>0.468</td>
</tr>
<tr>
<td>Number of grains per spike</td>
<td>0.106</td>
<td>0.245</td>
<td>0.43</td>
<td>0.677</td>
</tr>
<tr>
<td>Seed index 1000 grains weight in gems</td>
<td>0.384</td>
<td>3.437</td>
<td>0.11</td>
<td>0.913</td>
</tr>
<tr>
<td>Plot yield gram</td>
<td>-0.365</td>
<td>0.0215</td>
<td>-1.72</td>
<td>0.124</td>
</tr>
<tr>
<td>Nitrogen per hack Kg</td>
<td>0.168</td>
<td>0.168</td>
<td>1.00</td>
<td>0.345</td>
</tr>
<tr>
<td>Phosphorus per hack Kg</td>
<td>-0.138</td>
<td>0.333</td>
<td>0.42</td>
<td>0.688</td>
</tr>
</tbody>
</table>

**Discussions:**

The aim of the present study was to explore the different parameters of wheat yield in Sindh. Study was based on the data that for last year 2016. Wheat yield data production function was used to examine different factors affecting in the wheat data. The mean of the Seed Yield (Kg/ha was 19.63 cm with standard deviation of 5.458. The mean of the number of tillers per plant in meter square was 12.619 with standard
deviation of 2.479. The mean for the number of seed/kernels per spike was 56.14 with standard deviation 10.413. The mean length of spike in centimeter is 12.262 with standard deviation of 0.736. The mean for the Number of spikelets per spike is 19.382 with standard deviation 1.775. The mean for the tiller per hectar was 4491.673 with standard deviation 36.675. The mean of the Days to heading is 72.097 with standard deviation 1.445. The mean for Plant height in cm was 58.109 cm with standard deviation of 4.428. The mean for the Lodging score per hectar was 69.476 with standard deviation 1.99. The mean for the Number of grains per spike is 308.845 with standard deviation 1184.416. The mean of seed index for 1000 grains weight in grams is 1.744 with standard deviation of 0.271. The mean of plot yield in grams is 277.619 with the standard deviation 51.65. The mean of nitrogen per hectar in kilograms is 161.772 with the standard deviation 6.685 and the mean for the phosphorus was 69.27 with the standard deviation of 3.229. The coefficient of multiple correlation \( R = 0.936 \) reveals that there is positive relationship between the dependent and independent variables. The significance of coefficient of determination \( R^2 = 0.87 \) indicates that 87% of the variances in the dependent variable is explained by the fitted regression model which is good enough to comprise that the model performs well in term of prediction. Similarly, the value of adjusted \( R^2 (0.645) \) shows with the present adjustment of the parameter in the regression model the model accounts nearly for same variation in the data set without adjusting parameter. The regression is a semi-automated process of building a model by successively adding or removing variables based solely on the t-statistics of their estimated coefficients. In classify to take out effect of no effective characteristics in regression model on grain yield, regression was used. The results of regression results showed that the biological yield, harvest index and weight spike per unit with R square of 98.3%, reasonable the greatest of yield changes. A path coefficient is a standardized regression coefficient (beta) showing the direct consequence of an independent variable on a dependent variable in the path model. According to the results showed that harvest index and biomass had positive direct effect on grain yield (0.552** and 0.57**, respectively), whereas number of fertile spikes had a negative direct effect on grain yield (-0.449**). The highest indirect effects on grain yield were observed with harvest index (0.507) on biomass, and biomass (0.491) on harvest index. The highest negative indirect effects on grain yield were observed with harvest index (-0.35) on the number of fertile spikes, and number of fertile spikes (-0.136) on Spike weight per unit area. However, plant height, spikes number per plant, number of grains per spike and thousand-kernel weight showed a non-significant effect on grain yield. Similar results were reported by Singh and Diwivedi (2002) and who revealed that biological yield per plant and harvest index had positive and high direct effect on grain yield. Conversely, Bulman Pand Smith DL (1993) reported that sheath length followed by grains per spike, spike length and 1000-grain weight exhibited the maximum positive direct effect. The parameter of wheat number of tiller per plant meter square standard error is (0.569), Standardized Coefficients Beta is (0.300), T value is (-0.759) and with significant is 0.472. The parameter of Number of seed/kernels per spike standard error is (0.135), standardized coefficient beta is (0.124), T-value of (1.162) with significant (0.283*). The parameter of wheat is length of spike in cm is standard error is (1.258), standardized coefficient, T-value is (0.478) with the significant is (0.647*). Number of spikelet per spike with standard error of this parameter (1.258), standard coefficient beta is (0.430), the t-value is (0.381) and with significant is (0.714*). The parameter of tiller per hectar is standard error is (0.026), standardized coefficients beta is (0.279), and T-value of this parameter is (2.026) and the
significant value is (0.082*). The parameter of wheat is day of heading cm with the standard error is (0.723), the standardized coefficient beta is negative (-0.373), the t – value of this parameter is (1.604) with the significant of (0.153*). The parameter of wheat is plant height cm standard error is (0.247), with standardized coefficient beta is (0.072), T-value of this parameter is (-1.948) with the significant value is (0.092). The parameter of lodging score per hectar centimeter with standard error is (0.724), the standard coefficient of beta is (-0.189), T-value of this parameter is (0.359), with the significant of (0.730). The parameter of number of grains weight in grams with standard error is (0.001), standardized coefficient of beta is (-0.088), T-value is (-0.716) with the significant value is (0.497*). The parameter of wheat is seed index 1000 grains weight with standard error is (4.168), standardized coefficient beta is (0.074), T-value is (-0.400). The parameter of wheat is plot yield gram with standard error is (0.023), with standard coefficients beta is (-0.369), is T-value of this parameter is (0.358) with the significant value is (0.731*). The parameter of wheat is nitrogen per hector kilogram with standard error is (0.154), standardized coefficient beta is (0.236), is T-value is (-0.663) with the significant of (0.140). The parameter of Phosphorus per hac Kg with standard error is (0.363), standardized coefficient of beta is (-0.072), T-value is this parameter is (1.255), with the significant of (0.250) (Ahmadizadeh et al., 2011; Kumar et al., 2021). Correlations coefficients among traits are available in Grain yield explain positive and significant correlation with above position biomass (r =0.87), Grain weight per spike (r = 0.85), number of fertile spikes (r = 0.77) and figure of grains per spike (r = 0.69). Our conclusions were in conformity with the results of details significant and positive correlations among grain yield, biomass yield and grain weight per spike. Other authors reported similar results between grain yield, spikes number and number of grains per spike Singh and Sharma, 1994; Subhani and Khaliq, 1994; Khan et al., 1999; Mohammad et al., 2002; positive and significant correlation was found between grain yield and the number of grains per spike (r = 0.69). Harvest index (HI) also showed a significant positive correlation (0.76) with grain yield. Two major reasons are understood for associations between two traits; 1) due to linkage among genes which affect both the traits simultaneously and 2) effect of genes which means one gene influence more than one trait (Singh, 2009). Significantly negative associations (r= -0.60**) between plant height and number of tillers plant-1 and between plant height and harvest index (r=-0.54**) were recorded the coefficient of determination (r²) revealed that 36.0 and 29.1% of total variability in tillers plant-1 and harvest index, respectively was due to its association with plant height. The regression coefficient (b) indicated that a unit increase in plant height will reduce 0.08 tillers plant-1 and 0.25% decrease in harvest. The correlations between plant height and spike length (r=0.11NS); plant height and grain yield plant-1 (r=0.23NS); tillers plant-1 and spike length (r=0.20NS); tillers plant-1 and grains spike-1 (r=0.23NS) and grains spike-1 and seed index (r=0.45NS) were non-significant (Sharma and Rao, 1989). These non-significant correlations indicated that increase in one character will not cause any significant change in another trait. Positive and significant association (r=0.73**) between tillers plant-1 and grain yield plant-1 exposed that raise in tillers plant-1 will increase correspondingly increase grain yield plant-1. About 53.2% of total variability in grain yield plant-1 was due to its association with tillers plant-1; while regression coefficient indicated that one tiller increase will simultaneously give increase of 0.34 g in grain yield plant-1. Significant and positive association (r=0.57**) between spike length and grains spike-1 indicated that increase in spike length can remarkably increase grains
spike-1. The coefficient of determination (r²) indicated that 32.4% of total variability in grains spike-1 was due to its association with spike length-1; while regression coefficient suggested that one cm increase in spike length will cause an increase of 0.05 grains spike-1. Highly positive association (r=0.71**) between spike length and spikelet’s spike-1 was observed which suggested that increase in spike length, will simultaneously increase spikelet’s spike-1. The r² indicated that 50.41% variability in spikelets spike-1 was due to its relationship with spike length-1. (Kashif and Khaliq 2004; Yamin et al., 2021). Reported that yield components like tillers plant-1 had significantly contributed towards grain yield development. They also noted that grains spike-1 and 1000-grain weight were main contributors to grain yield in wheat. It was also reported that grain yield plant-1 showed significantly positive association with number of productive tillers plant, plant height, 1000- grain weight and spike length at genotypic and phenotypic levels the coefficient of determination (r²) explains that 68.8% of total variability in seed index was due to its association with harvest index; while regression coefficient (b) describes that a unit increase in harvest index will result in an increase of 0.64 g in seed index value. Highly positive correlation (r =0.97**) between seed index and grain yield plant-1 indicated that increase in independent variable will cause simultaneous increase in dependent variable. High percent (94.0%) of total variability in grain yield plant-1 is due to its association with seed index. One percent increase in seed index caused 0.74 g increase in grain yield plant 1. Like present results Balcoh et al., (2012) reported positive association of grain yield plant-1 with seeds spike-1 and 1000-grain weight.

Conclusion:
Correlation and regression analysis were associated with the assumption that both variables are covering relations. Correlation coefficient is a measure of linear association between two variables. The correlation coefficient values lie between +1 and always included. The survey results showed that grain yield had significant positive correlations with properties such as the number of fertile spikes, spike weight per unit area, harvest index, biomass, and thousand grain weights, number of grains per spike and plant height. To remove the effect of a stepwise regression models the performance of inefficient use of regression lines. As a result of the wheat crop coefficient of multiple correlation (r = 0.936), shows that 93% reveals that there is a positive relationship between dependent and independent variables. The significance of the coefficient of determination (R² = 0.87), shows that 87% of the discrepancy between the dependent variable explained by the regression model fitted well enough that the model performs well in terms of forecasting. Similarly, the value of the adjusted R² 0.878 indicates that the regression model parameter model makes the current correction is almost the same variation in the data set without adjusting parameter.

Suggestions:
It is suggested that these types of studies should be conducted with the special concentration on wheat producing area such as Badin, Mirpurkhas and Tandojam etc. It is also suggested that the more factors can be studied, and different statistical technique can apply to produce wheat crops.

References:


strains of Wheat, *tritium aestivum* L. Arid zone Research Institute, (parc), Umerkot.


Agriculture University, Tandojam 2(4):847-851