Original Research Article

Analysis of Simple Correlation Coefficient by using Theoretical and MS-Excel Procedure for Gerebera (Plant Height and Plant Spread)

V. Sekhar1*, V. Srinivasa Rao2, K. Umakrishna1 and Ch. Satyavani3

1Department of Statistics, 3College of Horticulture, Venkataramannagudem, Dr. YSR Horticultural University, West Godavari (District), Andhra Pradesh (State), India, PIN-534101
2Department of Statistics, Agricultural College, Bapatla, Acharya NG Ranga Agricultural University, Guntur (District), Andhra Pradesh (State), India, PIN-522101

*Corresponding author

A B S T R A C T

Introduction

Simple correlation is a measure of degree of linear relation between two continuous variables, X and Y. When a bivariate population is under consideration, there is generally a need to study the simultaneous between two variables. Karl Pearson in 1890 introduced correlation formula. Range of correlation coefficient is -1 to +1 (1, 2, 3).

Materials and Methods

Karl Pearson Correlation coefficient is denoted by r.

\[ r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right)\left(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right)}} \]

This work was attempted to find out simple correlation coefficient value between plant height and plant spread of 10 Gerebera plants. Correlation coefficient is applicable not only for Plant Height Vs Plant spread and it is applicable for plant yield Vs plant height, plant yield Vs infestation, fertilizer application Vs yield, price Vs demand of commodities, etc.

Keywords

Simple correlation coefficient, Karl Pearson correlation

Article Info

Accepted: 10 February 2019
Available Online: 10 March 2019
If \( r = 0 \), then there is no relation between two variables \( X \) and \( Y \).

If \( r = +1 \), then there exists perfect positive correlation between two variables \( X \) and \( Y \).

If \( r = -1 \), then there exists perfect negative correlation between two variables \( X \) and \( Y \).

If \( r \) lies between \(-1\) and \( < 0 \), then there exists negative correlation between two variables \( X \) and \( Y \).

If \( r \) lies between \( > 0 \) and \( < 1 \), then there exists positive correlation between two variables \( X \) and \( Y \).

Table 1: If we consider the following data belongs to plant height and plant spread of 10 Gerbera plants.

<table>
<thead>
<tr>
<th>Plant height (cm)</th>
<th>Plant spread (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.93</td>
<td>62.69</td>
</tr>
<tr>
<td>43.30</td>
<td>65.30</td>
</tr>
<tr>
<td>41.43</td>
<td>64.50</td>
</tr>
<tr>
<td>41.36</td>
<td>58.69</td>
</tr>
<tr>
<td>55.03</td>
<td>79.13</td>
</tr>
<tr>
<td>46.50</td>
<td>77.11</td>
</tr>
<tr>
<td>39.23</td>
<td>58.21</td>
</tr>
<tr>
<td>41.96</td>
<td>65.08</td>
</tr>
<tr>
<td>39.84</td>
<td>61.56</td>
</tr>
<tr>
<td>40.90</td>
<td>67.90</td>
</tr>
</tbody>
</table>

Then analysis procedure of correlation coefficient is hereunder.

Prepare the following table:

<table>
<thead>
<tr>
<th>( X )</th>
<th>( Y )</th>
<th>( X^2 )</th>
<th>( Y^2 )</th>
<th>( XY )</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.93</td>
<td>62.69</td>
<td>1594.40</td>
<td>3930.04</td>
<td>2503.21</td>
</tr>
<tr>
<td>43.30</td>
<td>65.30</td>
<td>1874.89</td>
<td>4264.09</td>
<td>2827.49</td>
</tr>
<tr>
<td>41.43</td>
<td>64.50</td>
<td>1716.44</td>
<td>4160.25</td>
<td>2672.24</td>
</tr>
<tr>
<td>41.36</td>
<td>58.69</td>
<td>1710.65</td>
<td>3444.52</td>
<td>2427.42</td>
</tr>
<tr>
<td>55.03</td>
<td>79.13</td>
<td>3028.30</td>
<td>6261.56</td>
<td>4354.52</td>
</tr>
<tr>
<td>46.50</td>
<td>77.11</td>
<td>2162.25</td>
<td>5945.95</td>
<td>3585.62</td>
</tr>
<tr>
<td>39.23</td>
<td>58.21</td>
<td>1538.99</td>
<td>3388.40</td>
<td>2283.58</td>
</tr>
<tr>
<td>41.96</td>
<td>65.08</td>
<td>1760.64</td>
<td>4235.41</td>
<td>2730.76</td>
</tr>
<tr>
<td>39.84</td>
<td>61.56</td>
<td>1587.23</td>
<td>3789.63</td>
<td>2452.55</td>
</tr>
<tr>
<td>40.90</td>
<td>67.90</td>
<td>1672.81</td>
<td>4610.41</td>
<td>2777.11</td>
</tr>
</tbody>
</table>

\( \Sigma X = 429.48 \) \( \Sigma Y = 660.17 \) \( \Sigma X^2 = 18646.61 \) \( \Sigma Y^2 = 44030.26 \) \( \Sigma XY = 28614.49 \)
Correlation coefficient between X & Y is

\[ r = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left( \sum X_i^2 - \frac{\left( \sum X_i \right)^2}{n} \right)} \sqrt{\left( \sum Y_i^2 - \frac{\left( \sum Y_i \right)^2}{n} \right)}} \]

\[ = \frac{28614.49 - \frac{(429.48)(660.17)}{10}}{\sqrt{(18646.61 - \frac{(429.48)^2}{10}) \left( 44030.26 - \frac{(660.17)^2}{10} \right)}} \]

\[ = \frac{28614.49 - 28352.98}{\sqrt{(18646.61 - 18445.31)(44030.26 - 43582.44)}} \]

\[ = \frac{261.51}{\sqrt{201.30}(447.81)} = \frac{261.51}{300.2435} = 0.8710 \]

The value of coefficient of determination \((r^2) = (0.8710)^2 = 0.7586\)

Test of significance of correlation coefficient: t-test conducted

\[ H_0: \text{The population correlation coefficient } \rho = 0 \]

Statistic:

\[ t = \frac{|r| \sqrt{n - 2}}{\sqrt{1 - r^2}} = \frac{|0.8710| \sqrt{10 - 2}}{\sqrt{1 - 0.8710^2}} = \frac{|0.8710| \sqrt{8}}{\sqrt{0.2414}} = \frac{0.8710 \times 2.8214}{0.4913} = 2.4574 = 5.0019^{**} \]

t-calculated value is 5.0019

Calculation of correlation coefficient value in MS-Excel

In MS-Excel sheet, enter data (plant height, plant spread) as below picture.
In B13 cell, type formula
= CORREL (A1:A11, B1: B11)
automatically above formula gives correlation coefficient value

**Results and Discussion**

According to our problem, correlation coefficient(r) value is 0.8710. It indicates that, there exists positive correlation between plant height and plant spread. i.e. plant height and plant spread both variables are deviate in same direction.

\[ r^2 \times 100 = 0.8710^2 \times 100 = 0.7586 \times 100 = 75.86 \approx 76. \]

It indicates out of 100 Gerebera plants, 76 gerebera plants have positive correlation between Plant height and plant spreads. Remaining 24 plants may be negative or zero.

From test of significance of correlation coefficient, t-calculated value is 5.0019

t-tabulated value at 5% level of significance with (n-2) = (10-2) = 8 d.f. is 2.3060

t-tabulated value at 1% level of significance with 8 d.f. is 3.3554

Hence, t-calculated value 5.0019 is greater than t-tabulated value 2.3060 and also greater than 3.3554. So we reject our Null Hypothesis at 1% level of significance. i.e. The population correlation coefficient \( \rho \neq 0 \).

It indicates, there exists highly significant correlation between Plant height and Plant spread of Gerebera plants.

**References**

4. Practical manual of STCA-101 (for class use only)- Dr. V. Srinivasa Rao and R. Srinivasulu, Agricultural College, Bapatla, 2013
5. MS-Excel: Statistical Procedures- Cini Varghese, IASRI, Library Avenue, New Delhi-110012

**How to cite this article:**

doi: [https://doi.org/10.20546/ijcmas.2019.803.113](https://doi.org/10.20546/ijcmas.2019.803.113)