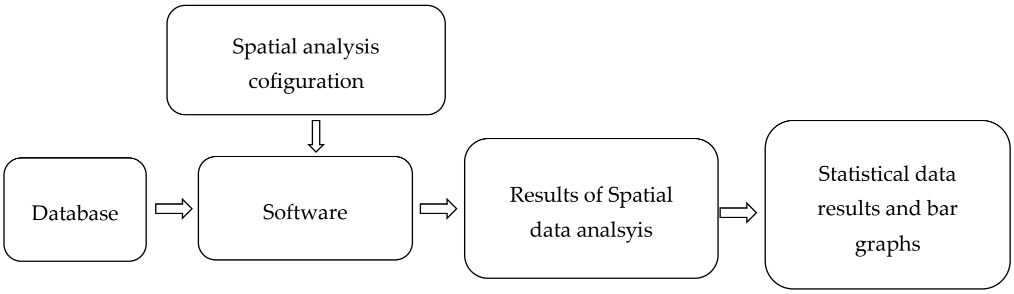
The Musa analysis tool – thecnical design

Designed by: Andres Galarza-Villamar* and John Galarza-Rodrigo

The Musa analysis tool is a software program written in the C# programming language. It is a tool that was developed for the analysis and interpretation of data obtained during field experiments with the Musa game. The program has four parts: an interface for loading information from a file with an [.xlsx] extension, a panel that uses a color code to display the data, a form-type interface for selecting the type of spatial analysis in combination with the type of actions of the participants to be analyzed, and a section of bar graphs to display the numerical interpretations of the results obtained from the selected analysis (Figure 1). The main objective behind the development of this program was the need for a customizable, compact, and easy-to-use tool for processing and analyzing the experimental data.



The variables analyzed in the experiment are the decisions made by participants, which consist of three possible actions: Cut Flower, Uproot Yellow Mat, Uproot Red Mat. These decisions are causal and are directly related to the rules of infection of the mat within the experiment. Appearance of an infection is the resultant of the conditions created by a combination of specific variables and game movement rules. At the same time, the objective is to observe only specific variables of interest, namely the decisions made by game participant.

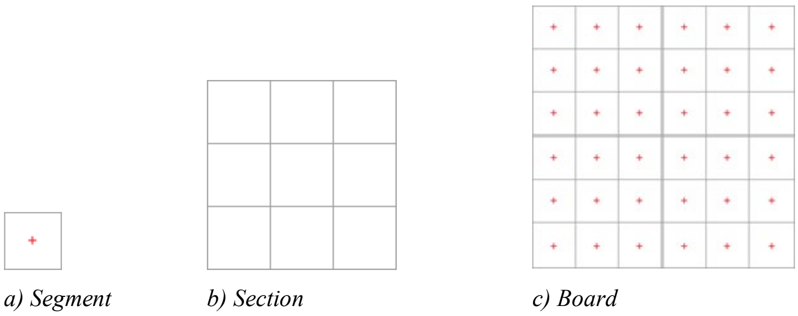


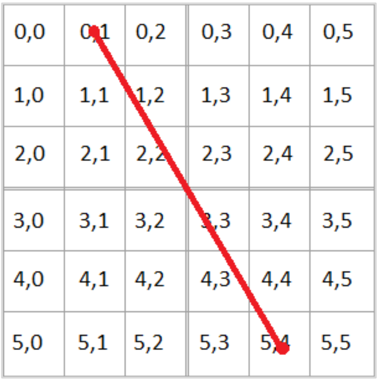
Figure B1. Board components used for software development, where a board has 4 sections, and a section has 9 segments.

The data is collected on a flat and uniform 6 x 6 segment board, that is divided into four sections that are assigned to the 4 participants. In order to simplify calculations, we assume that each segment corresponds in size to a 1 x 1 unit of distance. Likewise, its point of interest will be in the central position of each segment, 0.5 units in X and 0.5 units in Y with respect to the upper left corner of each segment. In calculations, the notation used for the position of each segment corresponds to its location by row (R) and column (C) within according to the displayed matrix (Figure 2). The expression of the general position of each segment is: GP= (R - 1, C - 1), where R and C are values between 1 and 6. However, to take the center of each segment as the position of interest for a calculation the expression is: PI= (GP_x + 0.5, GP_y + 0.5).

0,0	0,1	0,2	0,3	0,4	0,5
1,0	1,1	1,2	1,3	1,4	1,5
2,0	2,1	2,2	2,3	2,4	2,5
3,0	3,1	3,2	3,3	3,4	3,5
4,0	4,1	4,2	4,3	4,4	4,5
5,0	5,1	5,2	5,3	5,4	5,5

Figure B2. Notation used for the position of each segment

Example 1 (Figure 3). The distance between two random points A and B, located at the position GPA (0,1) and GPB (5,4), will be given by: $D = \sqrt{(PI_{Ax} - PI_{Bx})^2 + (PI_{Ay} - PI_{By})^2}$

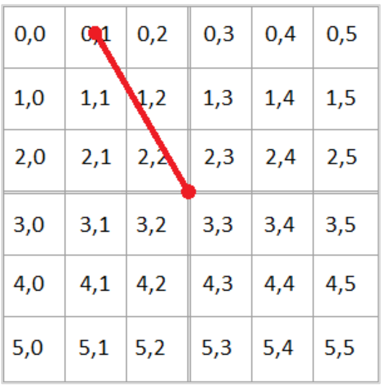


$$D = \sqrt{((0 + 0.5) - (5 + 0.5))^2 + ((1 + 0.5) - (4 + 0.5))^2} = 5.83$$

Figure B3. Example 1, distance calculation between two random points PGA (0,1) and PGB (5,4).

For measurements between a random point and the center of the board, the central point (Pc) is taken as a reference and expressed as Pc: (R / 2, C / 2). The central point is a position of interest for measurements because it represents the onlyposition on the game board where the board sections of all four players meet, and hence decisions of each player near thecentral point can directly affect the other players in the game.

Example 2 (Figure 4). The distance between a random point A located at PGA (0,1) and the central position (Pc) for a 6x6 board, where Pc = (6/2, 6/2) is given by: $D = \sqrt{(PI_{Ax} - Pc_x)^2 + (PI_{Ay} - Pc_y)^2}$



$$D = \sqrt{((0 + 0.5) - 3)^2 + ((1 + 0.5) - 3)^2} = 2.91$$

Figure B4. Example 2, distance calculation between a random point PGA (0,1) and Pc.

Each distance measured in the experiment corresponds with a distance between a PI (Point of Interest) of a segment(corresponding with the player's actions), and a PI of a second segment (corresponding with a direct value of the board) in one moment in time (Game Round), or the Pc position (Center position).

Said measurements are normalized to a scale of values between 0 and 1, meaning a value of 0 for positions outside border of the board, and 1 for the position that is exactly in the center of the point of interest/segment, which is taken as a reference for the measurement.

We call the distance given in values between 1 and 0 the normalized distance or Dn, which is given by: $Dn = \frac{(Dm-D)}{Dm}$, where Dm is the value of the maximum possible distance between two ends of the board, which is the direct result of the maximum diagonal distance of the board $Dm = \sqrt{R^2 + C^2}$, except for calculations where the only reference is the Central Position (Pc), in which case Dm has the value of the maximum distance between two points of interest of one of the sections of the board, which is maximum half the diagonal of the board $Dm = \frac{\sqrt{F^2+C^2}}{2}$.

For practical reasons, the round values of 1 and 0 will are be represented for normalized distance (Dn) in the real measurements. Given the rules of the game these values are impossible. The minimum possible value (Vmnp) will be given by $Vmnp = \frac{\sqrt{R^2+C^2}-\sqrt{(R-1)^2+(C-1)^2}}{\sqrt{R^2+C^2}}$, and the maximum possible value (Vmxp) will be given by $Vmxp = \frac{\sqrt{R^2+C^2}-1}{\sqrt{R^2+C^2}}$. Likewise, the Vmnp and the Vmxp when the only reference is the Central Position will be $Vmnp = \frac{\sqrt{R^2+C^2}-\sqrt{(R-1)^2+(C-1)^2}}{\sqrt{R^2+C^2}}$ and $Vmxp = \frac{\sqrt{R^2+C^2}-\sqrt{2}}{\sqrt{R^2+C^2}}$. Since the value of 0 is not possible in the calculation, it has been reserved for distance measurements that do not meet the action criteria between both points. That is, one of the points of interest within the measurement does not exist.

Measurement methodology

For the experiment, there are eight states of the banana mat that are represented by codes: Dead (5), Intervened (6), Green Flower (1), Green (2), Yellow (3), Yellow Uprooted (31), Red (4), Red Uprooted (41), which for purposes of interpretation were cataloged using the following color codes (Figures 5 and 6).

Empty	Red
Green with flower	Cut red
Green (without flower)	Intervened
Yellow	Death
Cut Yellow	

Figure B5. Card colours codes as used in software. Note that the Green with flower refers to the White card in the game.

0.17	0.31	0.40	0.40	0.31	0.17
0.31	0.50	0.63	0.63	0.50	0.31
0.40	0.63	0.83	0.83	0.63	0.40
0.40	0.63	0.83	0.83	0.63	0.40
0.31	0.50	0.63	0.63	0.50	0.31
0.17	0.31	0.40	0.40	0.31	0.17

Figure B6. Example of the Initial state of the boardgame in the Musa analysis tool. It shows the values of the Normalized Distance (Dn) of each segment concerning the Central Point (Pc) of all types of card stages for a board in the initial round.

As mentioned, the events of interest for the analysis are the actions of the players, thereby considerating the circumstances on the board when said decision is made. The actions of the players to consider were: Cut Flower, Uproot Yellow Mat, Uproot Red Mat. Likewise, the events occurring randomly or controlled by game rules were considered, namely: New Yellow Mat, New Red Mat. These events were identified through an algorithm. It compared the board conditions from one round with another round and then quantified and grouped these by their Normalized Distance (Dn) in different analyzes (Figure 6).

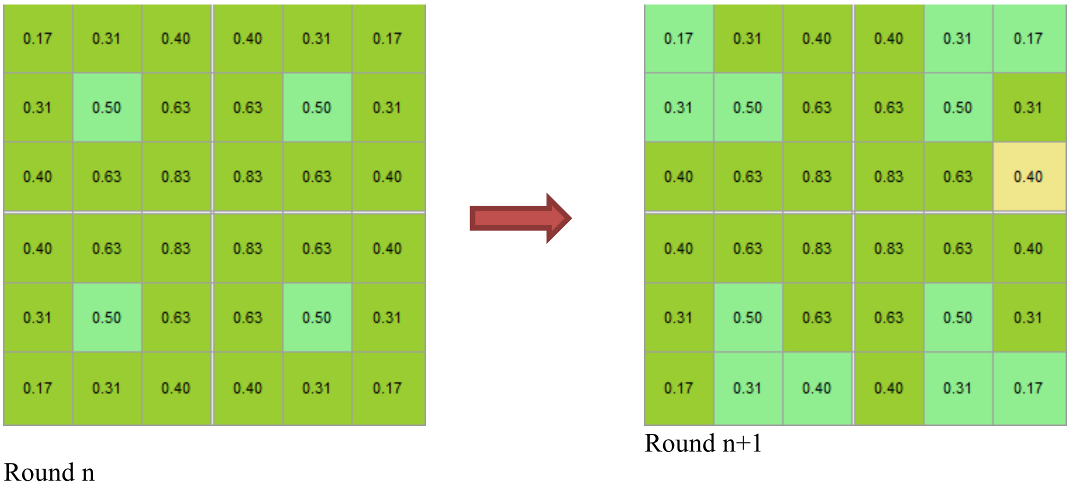


Figure B7. Example: 6 Green Flower mats changed to Green (6 Cut Flower actions). 1 Green Flower Mat changed to Yellow (1 New Yellow Mat event), measurements relative to the centre.

To analyse the distance under specific circumstances, the dashboard has values for as long as there is at least one location that serves as a reference or measurement center for the other dashboard segmentsFor example, distances from a yellow mat. All the segments have numerical values that represent their respective normalized distance (Dn) compared to the variable of interest (Yellow Mat). The location of the Yellow Mat has a value of 0 because there is no second Yellow Mat from which it could obtain its distance value(Figure 8). For investigation and analysis of data, employing the software, all the possible analysis variables were parameterized to allow for selection of different combinations and observe possible trends in the results these combinations give (Figure 9).

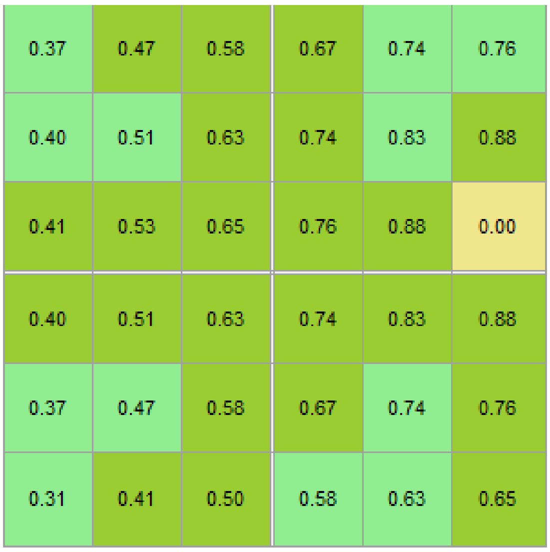


Figure B8. example of numerical values and their respective normalized distance (Dn) compared with the variable of interest (Yellow Mat).

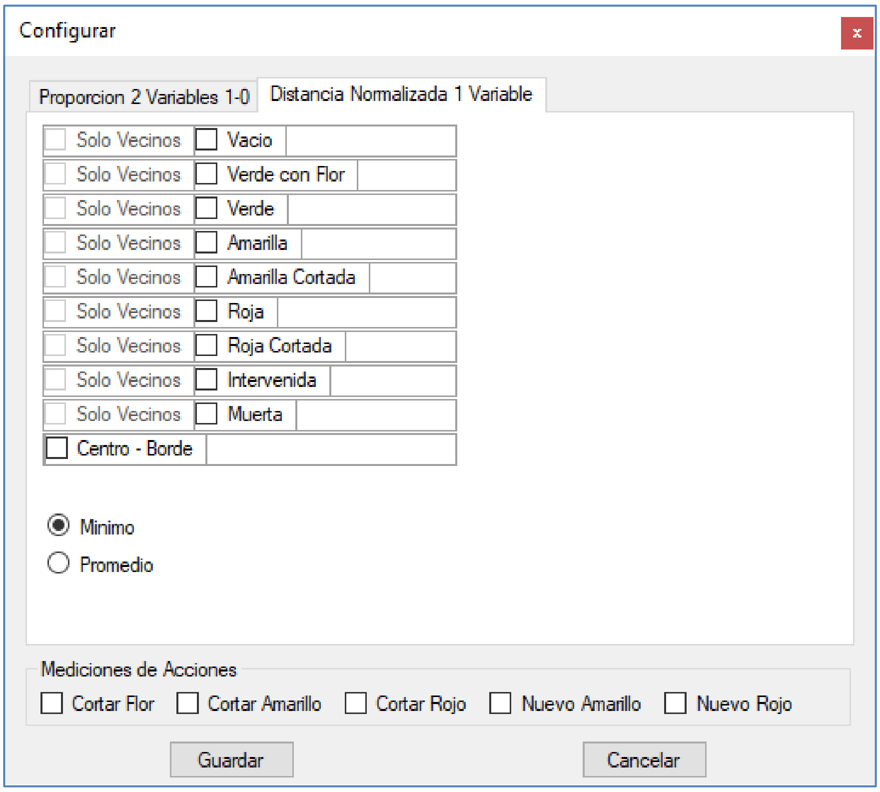


Figure B9. Panel of options for analysis configuration

As a practical example, for the spatial analysis of distance the software is be configured to consider the action to cut flowers. This action is the central reference to measure distances to the Yellow Mat or Red Mat segments only in neighboring sections. The minimum distance was selected to control for cases in which there is more than one Yellow or Red Mat (Figure 10) .

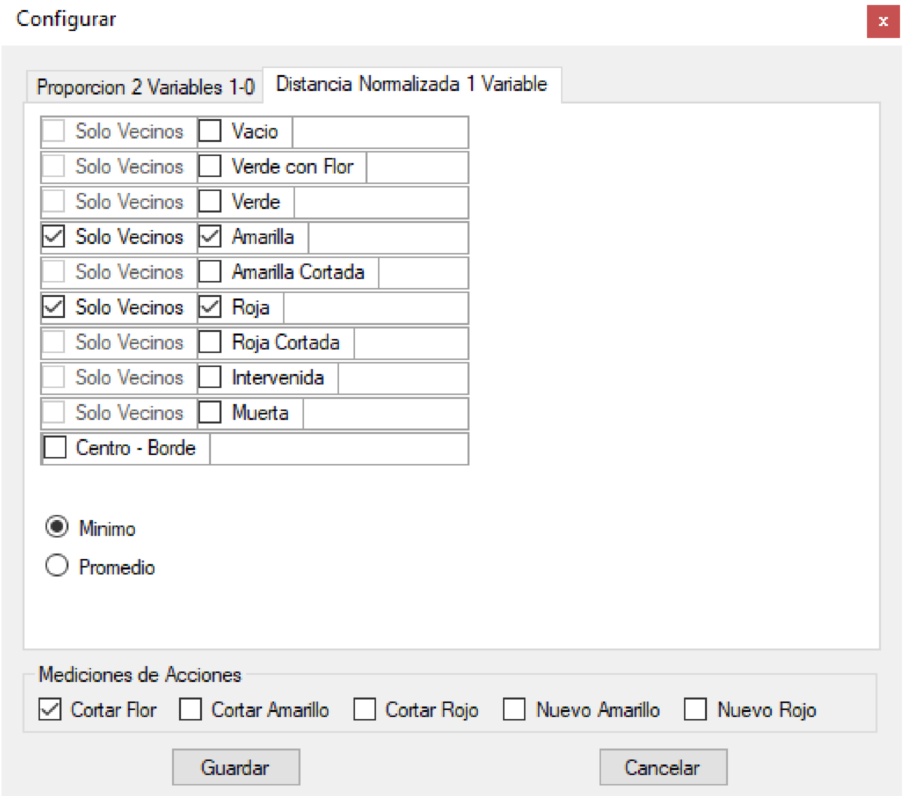


Figure B10. Data analysis configuration window for the example

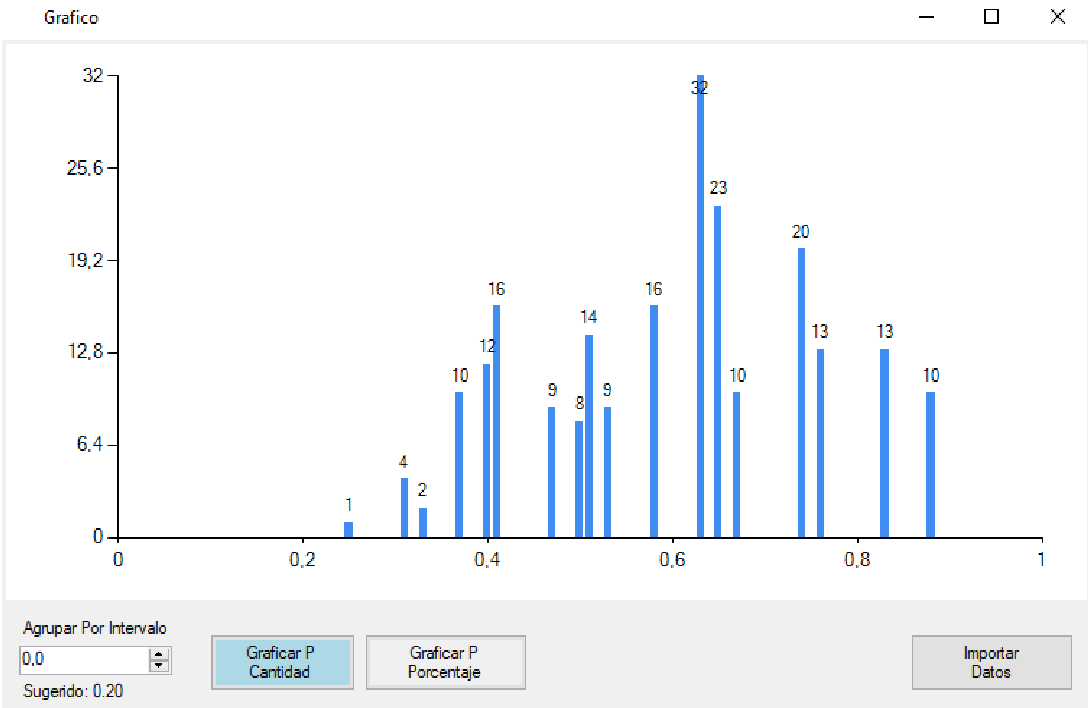


Figure B11. Results based on configuration of Figure 10

In the graph (Figure 11), the axis Y shows the number of Cut Flower actions that players made and the axis Y shows the Normalized Distance (Dn) between 0 and 1 in relation to a diseased mat when making the decision to cut flowers.

Table B1. Number of decisions made at different normalized distances (Dn)

Normalized Distance (Dn)	*Decisions*
0,25	1
0,31	4
0,33	2
0,37	10
0,4	12
0,41	16
0,47	9
0,5	8
0,51	14
0,53	9
0,58	16
0,63	32
0,65	23
0,67	10
0,74	20
0,76	13
0,83	13
0,88	10

We can see that there is a uniform distribution in the example results (Table 1). However, due to the nature of the calculation, there are distances among them that are very similar. For example, distances such as 4.0 and 4.1 or 0.50 and 0.51). For practical reasons and to simplify data analysis these values can be grouped.

To group data, a uniform distance interval is defined for the data of the variable that is to be analyzed. In this particular case, it will be the distance to which all the sampled distance values will approximate. These are defined as Values close to the interval (Vi) and given by: $Vi = Round\left(\frac{V}{I}\right) * I$, where V is the real value, and I is the selected interval (Tables 2 and 3).

Table B2. Example: Values Close to Interval (Vi) of Normalized Distances (Dn) obtained, for an interval of 0.1.

Normalized Distances (Dn)	*Values Close to Interval (Vi)*
0,25	0,3
0,31	0,3
0,33	0,3
0,37	0,4
0,4	0,4
0,41	0,4
0,47	0,5
0,5	0,5
0,51	0,5
0,53	0,5
0,58	0,6
0,63	0,6
0,65	0,7
0,67	0,7
0,74	0,7
0,76	0,8
0,83	0,8
0,88	0,9

Table B3. Decisions Grouped by Vi

Values Close to Interval (Vi)	*Decisions*
0,4	38
0,5	40
0,7	53
0,8	26
0,6	48
0,9	10
0,3	7

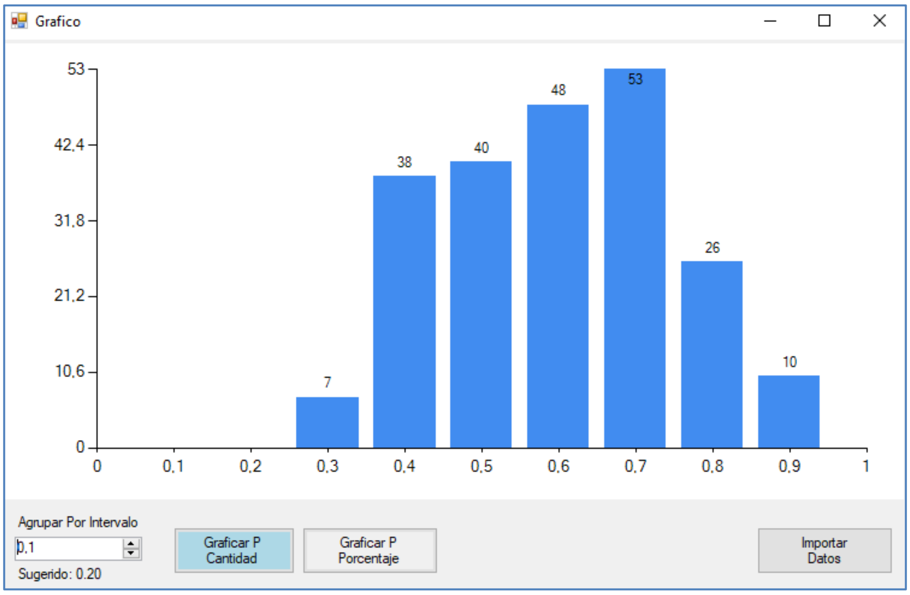


Figure B12. Graph of grouped results with an interval equal to 0.1

Since the interval value can be very relevant for the analysis, an optimal interval value for the data pool can be calculated. For this purpose, an empirical value is calculated defined as: Maximum Interval Value (Vim). The Vim is the quotient of the arithmetic mean of the distances between the Values Near the Interval (Vi) and the Real Value (V), and the selected interval (I), $Vim_I = \frac{\left(\frac{2}{I}\right)\sum_{k=0}^{n}|Vi_k - V_k|}{n}$, where Vi is the value close to the interval I, V is the actual value, I is the selected interval value, and n is the amount of data.

With the initial results obtained through the calculations done as exemplified by Figure 12 as input, the Vim values for the intervals (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) are calculated (Figure 13). For these interval values, we discard all those in which the number of intervals contained in the range from 0 to 1 is less than or equal to 1 since for these the grouped values would give the lowest possible resolution and would not have a significant value for the analysis. The Contained Intervals (Ic) will be given by, $Ic = Floor\left(\frac{1}{I}\right)$, where I is the selected interval (Table 4).

Table B4. Contained Intervals (Ic) values for the intervals (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

Intervals	*Contained Intervals (Ic)*
0,1	10
0,2	5
0,3	3
0,4	2
0,5	2
0,6	1
0,7	1
0,8	1
0,9	1

Ic values less than or equal to 1 are discarded, because the only suitable Interval values for this analysis are the intervals: 0.1, 0.2, 0.3, 0.4 and 0.5. When calculating the Vim values for these intervals we get Table 5.

Table B5. Calculated Vim values

I	*Vim*
0,1	0,511
0,2	0,533
0,3	0,481
0,4	0,444
0,5	0,511

Therefore, when looking for the highest possible uniformity in the distribution for the grouped data, the VI value closest to 0.5 will belong to an interval of best-distributed data. Other statistical criteria could be considered for the selection of a grouped interval for the data, such as the standard deviation of the values with respect to the interval, the selection of the minimum possible interval, the minimum amount of data grouped in the said interval, among others.

Selecting 0.2 as Interval, we can observe in the following graph (Figure 13) a distribution similar to the 0.1 intervals. The difference with figure 12 is greater visibility of possible trends, which is useful for analysis.

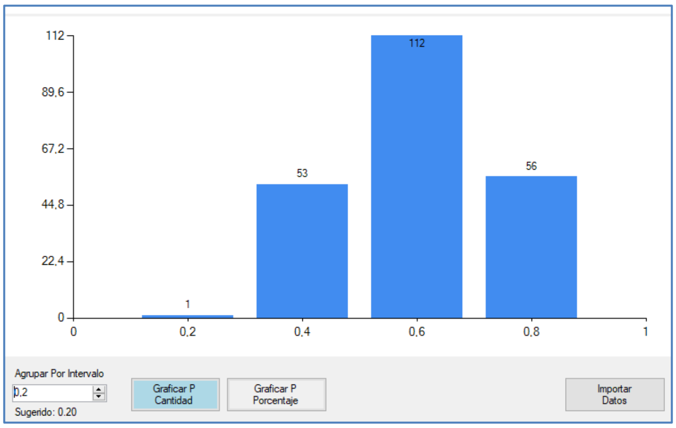


Figure B13. distribution example with an interval of 0.2

Despite observing a trend in the results, the data of these random events must be contrasted with the actions taken. That is, comparing the event (action) with the participant's decision (reaction). This contrast is necessary because the trends observed from a single analysis might be caused by a rule intrinsic to the game and not the players. Therefore, in the second example, two analysis configurations are performed. The first to get the appearance of Red and Yellow Mats on the board in relation to a Neighbouring Mat with Flower. The second analysis is to obtain the number of Red and Yellow Mats that were uprooted in relation to a Neighboring Mat with a Flower (Figures 14 and 15).

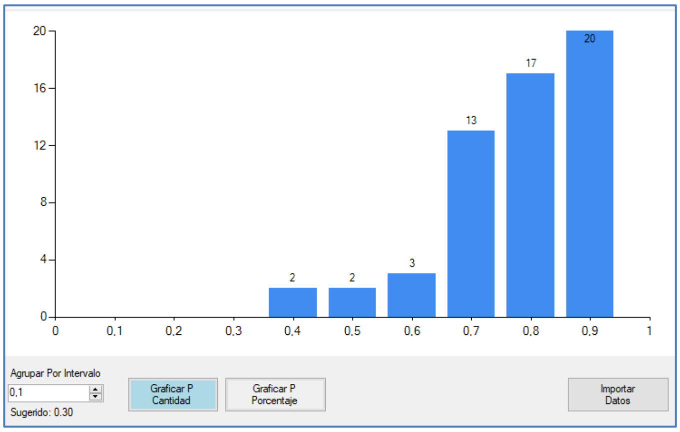


Figure B14. Appearance of Diseased Mats in relation to Healthy Mats

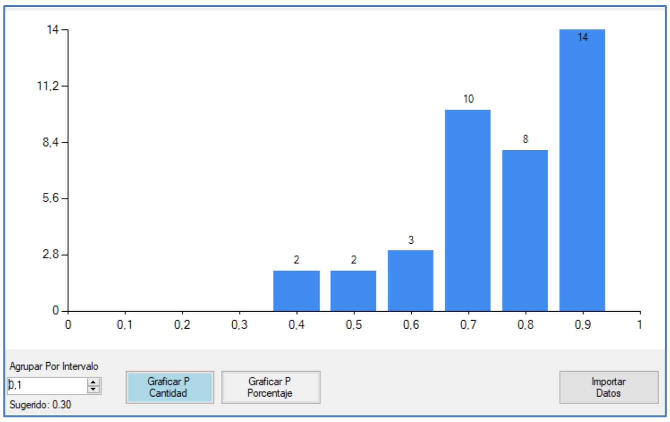


Figure B15. Cutting of Diseased Mats in relation to Healthy Mats

When comparing the results, an increasing trend of uprooting actions in relation to distance is visible which is similar in both graphs and is caused by the appearance of diseased mats (Red or Yellow) (Random Variable). However, by contrasting the information obtained, we can observe a higher frequency in the decision-making of uprooting diseased plants in the farthest distances from the Healthy Plants.