When considering categorical variables, we used the Chi-square test (χ2) as a statistical test to determine the existence of a significant relationship between the dependent variable and each of the independent variables that will be considered in the study. Since this is a non-parametric test, it is not necessary to verify that the variables have a normal distribution.

The hypotheses are:

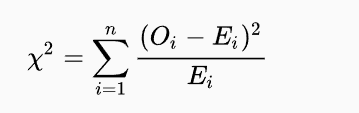
♦ Null hypothesis (Ho): The two variables are independent, which means there is no relationship between them.

♦ Alternative hypothesis (Ha): The two variables are not independent, that is, there is a relationship between them.

The shape of the χ2 distribution is positive and asymmetric.

This probability distribution represents that if *Z* 1 , ..., *Z k,* are independent variables and with Laplace-Gauss normality standards [[1]](#footnote-1)of random variables, then the sum of their squares is distributed according to the χ2 distribution, with a number of k degrees of freedom, designated as the number of random variables or independent responses in each of the questions compared, where by increasing the number of these with normality standards, their distribution is linked to a normal distribution [[2]](#footnote-2).

Now, basing the study on the search for correlation between variables using the χ2 test, the observed frequencies are compared with the expected frequencies under this hypothesis of independence (Ho), where the χ2 value is obtained from the sum of the quotients between the squares of the differences between the observed value and the expected value, with respect to the expected values for each category.



χ2 = Pearson cumulative test statistic, which asymptotically approximates the Chi-square distribution.

= Number of observations of type i

= Expected (theoretical) frequency of type i = (row total x column total) / grand total

= Number of cells in the table

The assumptions that must be met for the value of this statistic to be statistically reliable in the study of independence between variables are:

1. That the variables are categorical. That is, that the values of the observed frequencies come from qualitative variables.
2. That each observation must be independent. That is, the same instance or individual must not belong to another record evaluated in the study.
3. That the expected frequency of each cell must be at least 5. In other words, that each crossing between the possible responses of both variables has an expected value (E) greater than or equal to a frequency of 5.

Once the criteria of the χ 2 distribution and its link to the asymptotic approximation of the Pearson cumulative statistical test to a χ 2 distribution have been described, we begin the search for independent variables that demonstrate contrasting evidence of dependence on our dependent variable.

Rstudio interface, for the statistical analysis of the correlation between variables, implementing the χ2 test statistical function, from the integrated stats library [[3]](#footnote-3), on the instances that have experienced a feeling of bliss or joy after their experience, we obtain the p-values [[4]](#footnote-4)for the set of variables analyzed.

Those variables that meet the assumptions necessary for the implementation of the χ2 hypothesis test and that have obtained statistically significant results, which reject the null hypothesis of independence between variables, with 95% confidence are:

1. Experience of loss of self or of some belief strongly linked to your personality
2. Have you had any previous yoga activity?
3. Have you ever experienced a mystical event?
4. Involuntary movements as a phenomenological experience
5. Phenomenological experience linked to death
6. Breathing practices prior to the phenomenological experience

Cramer's V is a metric used to measure the level of association between two qualitative variables. The function that calculates this coefficient is:

χ2 = Pearson cumulative test statistic, which asymptotically approximates the Chi-square distribution.

n = Total number of instances considered.

r = Number of rows in the contingency table.

c = Number of columns in the contingency table.

The assumption necessary to implement this relationship measure on the variables is:

* + - * 1. The χ2 statistic must be able to be calculated. This means that all the previous criteria necessary for the use of this test must be met [[5]](#footnote-5).

The range of values of this coefficient is [0, 1] [[6]](#footnote-6), where the effect size of the variables is classified as:

|  |  |
| --- | --- |
| **Effect size (ES)** | **Interpretation** |
| **ES ≤ 0.2** | The result is weak. Although the result is statistically significant, the fields are only weakly associated. |
| **0.2 < ES ≤ 0.6** | The result is moderate. The fields are moderately associated. |
| **IN > 0.6** | The result is strong. The fields are strongly associated. |

*Source: IBM Congos Analytics*

The variable with the highest level of association (Cramer's V) is “Prior mystical experience,” which refers to the variable of previous experiences of mystical events that the instances indicate they have experienced at some stage in their lives. This variable obtains a value of 0.2432, which indicates a moderate level of association with respect to the dependent variable, according to the “Cramer's V level of association” table. For the rest of the variables, the results returned with this univariate relationship metric, with respect to the response variable “Feeling of bliss or joy caused by the phenomenological experience, in the non-ordinary state of consciousness, experienced through the use of psychedelics, ” are 0.2044 in the degree of relationship with the variable “Experience of loss of self or of some belief strongly linked to your personality”; 0.1962 with respect to the variable “Previous yoga activity”; 0.1857 for the relationship with the variable “Involuntary movements as a phenomenological experience”; and 0.1797 and 0.1796 for the variables “Phenomenological experience linked to death” and “Breathing practices prior to the phenomenological experience”, respectively.

It is worth mentioning that both the instrument and the study have the sole objective of explaining and using descriptive statistical techniques to represent the responses of the 117 participants. The results obtained by the use of metrics and inferential techniques do not support replicability and reproducibility with different samples and contexts, due to the limitations of the methods and instruments of collection and experimentation of the study.

However, there are variables that do not meet the expected value assumption with a frequency of at least 5 for each cell in the contingency table.

For these cases, the Fisher test [[7]](#footnote-7)provides a more accurate equivalent [[8]](#footnote-8)than the χ2 test when the number of events expected per level of each cell is small, that is, less than 5. Therefore, those variables that demonstrate a significant relationship by this means are evaluated through hypothesis contrast evidence with the Fisher test.

The hypotheses for this statistical test of independence are:

* Null hypothesis (Ho): The variables are independent so one variable does not change between different levels of the other variable.
* Alternative hypothesis (Ha): The variables are dependent, that is, one variable changes between different levels of the other variable.

Where the assumptions that must be met for the correct implementation of this test are:

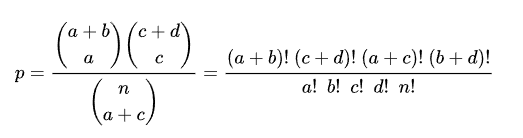
1. Independence among observations. As with Pearson's cumulative statistical test, each observation contributes only to a single level of the categorical variable, represented in the contingency table.
2. Fixed marginal frequencies. That is, there are no alterations in the total frequencies of the rows and columns, when obtaining different combinations between the levels of each variable.

Under these assumptions and the null hypothesis of independence between the variables, the calculation of the exact probability of obtaining the observed data from a contingency table follows a hypergeometric distribution. This distribution describes the probability of obtaining a given number of successes with a sample without replacement for a finite population.

For a general contingency table with two dichotomous variables

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature A** | | | |
| **Feature B** | Present | Absent | Total |
| Present | to | b | a + b |
| Absent | c | d | c + d |
| Total | a + c | b + d | n |

The exact probability of obtaining any specific combination with these conditions is given by the hypergeometric distribution:



Where is the binomial coefficient [[9]](#footnote-9)and the symbol “!” represents the factorial operator.

Now, those variables that showed a p-value lower than the significance level of 0.05, granting a confidence level greater than 95% are:

1. Your connection with nature changed after your psychedelic experience.
2. Your psychedelic experience led you to some important life changes or decisions.
3. Change in mood stability following psychedelic experience
4. Change in social relationships or connection with others
5. Change in the feeling of empathy and compassion, after the psychedelic experience

However, the following variable obtained a p-value of 0.051, which is slightly above the significance level of 0.05, but we have included it for its heuristic potential within the study.

1. Level of satisfaction with your life before the psychedelic experience(s)

The study focused on a small set of variables, the 233 questions presented in the collection instrument invite further investigation into more relationships and descriptions that improve and deepen the analysis after this study. It is important to reemphasize the fact that these results are intended to describe the responses of the instances studied. None of these results should be extrapolated to individuals outside the study. To do this, more sophisticated sampling techniques must be created, as well as experimental control environments that can robustly meet the appropriate standards to infer the results to samples not present in the study.

1. Referring to a continuous probability distribution, where the values belong to real numbers, with a mean of 0 and a standard deviation of 1. [↑](#footnote-ref-1)
2. The fulfillment of these assumptions denotes the relationship of the Chi-square distribution with the normal distribution, as the number of random variables that meet these conditions increases. [↑](#footnote-ref-2)
3. R base library containing a set of statistical metrics for descriptive and inferential analysis. [↑](#footnote-ref-3)
4. Probability of occurrence that favorably reiterates the null hypothesis (Ho). [↑](#footnote-ref-4)
5. Assumptions linked to the χ 2 statistics. [↑](#footnote-ref-5)
6. This allows it to be an alternative to the C contingency test, which only approximates the degree of association to 1, but never reaches this value. [↑](#footnote-ref-6)
7. Statistical method that allows evaluating the relationships between two qualitative variables when the assumption of expected values >= 5 in each of the cells of the contingency table is not met. [↑](#footnote-ref-7)
8. It is an exact calculation of the deviation, rather than an asymptotic (infinite) approximation. [↑](#footnote-ref-8)
9. Combinations that correspond to the number of possible subsets, from the given set. [↑](#footnote-ref-9)