User Manual
MCDA Statistical Analysis Software

Manual Version: 1.0
Software Version: 1.0
Last update: January 10, 2026
Compatibility: Python 3.8+, Streamlit 1.28+
This manual is subject to updates as new functionalities are added to the tool. For suggestions or issue reports, please contact the developers.

Developers:

MSc Eng. Anderson Portella
Prof. Dr. Miriam Rocha
Eng. Robson Souza
Prof. Dr. Marcos dos Santos
Prof. Dr. Carlos Francisco Simões Gomes

Introduction

The MCDA (Multi-Criteria Decision Analysis) Statistical Analysis Tool is a web-based application developed to support researchers, students, and professionals in conducting advanced statistical analyses within the context of Multi-Criteria Decision Analysis (MCDA).

This manual aims to guide users, in a detailed and step-by-step manner, in the use of the software, developed through an academic partnership between Fluminense Federal University (UFF), the Federal University of Rio Grande do Norte (UFRN), and the Military Institute of Engineering (IME).

The software provides an integrated environment for statistical analysis applied to MCDA, enabling the validation, exploration, and interpretation of decision matrices, and offers:

- Comprehensive statistical analyses for decision data
- Bilingual interface (Portuguese/English)
- Multi-format result export
- Intelligent AI-based analysis (OpenAI)
- End-to-end processing from raw data to finalized reports

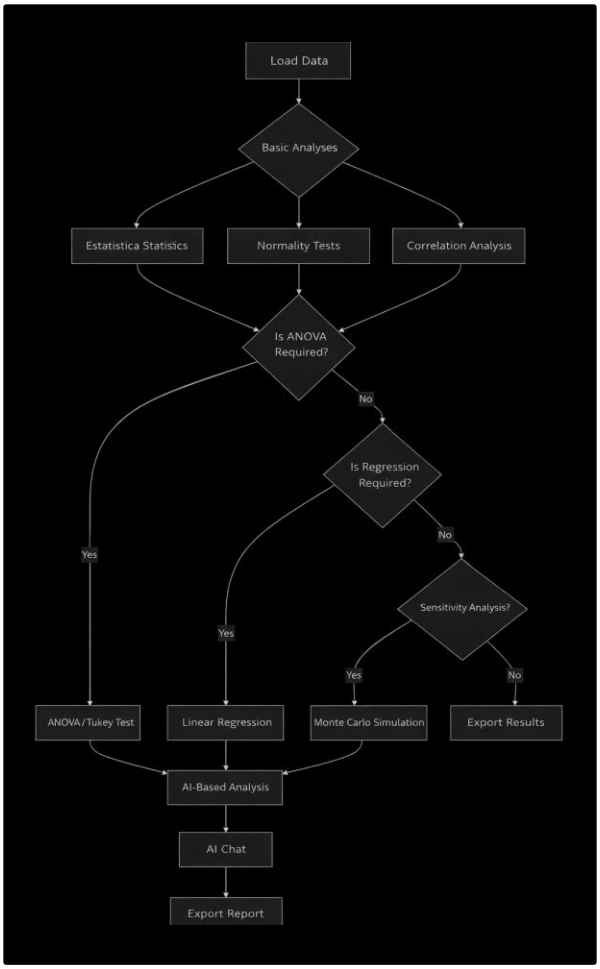
Sumário

1. Software Overview

The MCDA Statistical Analysis software was designed to support students, researchers, and professionals during the statistical pre-processing stage of MCDA matrices. Its primary purpose is to provide statistical support to improve understanding of data structure prior to, or in support of, multi-criteria decision-making methods.

1.1 Available Features

- ✅ Data upload (Excel, CSV)
- ✅ Descriptive statistical analyses
- ✅ Normality and correlation tests
- ✅ ANOVA and non-parametric tests
- ✅ Linear regression with diagnostics
- ✅ Monte Carlo simulation for sensitivity analysis
- ✅ Export to Excel, PDF, and JSON
- ✅ Automatic AI-based analysis
- ✅ Interactive AI chat



1.2 Tips and Best Practices

Data Quality:

1. Check for missing values before analysis
2. Standardize scales when necessary
3. Document all transformations applied
4. Verify test assumptions

Result Interpretation:

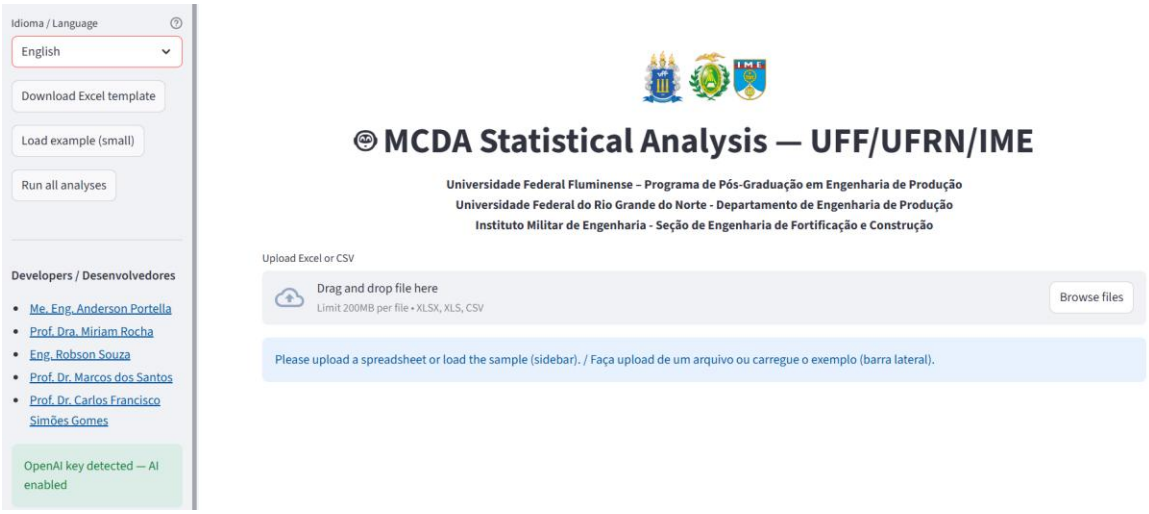
1. Contextualize p-values with effect size
2. Consider practical significance in addition to statistical significance
3. Use graphical outputs to complement tables
4. Document analytical limitations

Use of AI:

1. Provide context in chat queries
2. Request references to specific results
3. Validate recommendations with domain expertise
4. Use AI as a complement, not a substitute, for expert judgment

2. Initial Screen and Data Preparation

Upon accessing the software, users are directed to the initial screen, which consolidates all functionalities required to begin the analysis. From this screen, users can select the interface language, download data templates, load sample datasets, or upload their own files.



2.1 Language Selection

The side menu allows users to select Portuguese or English. The selected language automatically updates all textual elements of the interface, ensuring accessibility for both national and international users.

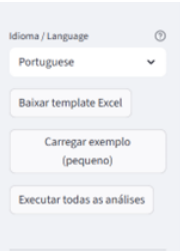
2.2 Data Structure

The data must be organized in a tabular format, containing a column that identifies the alternatives, numeric columns corresponding to the criteria (for example, *Cost*, *Efficiency*, and *Risk*), and, optionally, a categorical group column.

The user may choose to download an Excel template to create their own decision matrix (by clicking *"Download Excel Template"* in the sidebar), load a sample dataset provided by the developers, or upload their own file, provided that it follows the required structure of a payoff matrix, using the menu on the left side of the screen.

File Upload

- Accepted formats: .xlsx, .xls, .csv
- Recommended structure:
    - Column identifying the alternatives
    - Numeric columns for criteria
    - Categorical column for groups (optional)



	A	B	C	D	E
1	**Alternative**	**Cost**	**Efficiency**	**Risk**	**Group**
2	A1	100	0,8	2	G1
3	A2	110	0,85	1,8	G1
4	A3	95	0,78	2,2	G2
5	A4	130	0,9	1,2	G2

3. Initial Statistical Results

After data loading, the software automatically generates descriptive statistics, normality tests, and initial correlation analyses. These outputs provide a preliminary understanding of the behavior of the analyzed criteria.

Data preview

	Alternative	Cost	Efficiency	Risk	Group
0	A1	120	0.8	2	G1
1	A2	95	0.9	1.5	G1
2	A3	110	0.85	1.8	G2
3	A4	130	0.75	2.2	G2

3.1 Estatística Descritiva

Mean

- **What it is:** Average value of the data
- **Formula:** Σx_i / n
- **Interpretation:** Measure of central tendency of the data
- **Example:** Mean cost across alternatives

Median

- **What it is:** Value that divides the data set into two equal parts
- **When to use:** When the data contain outliers
- **Advantage:** Not affected by extreme values

Mode

- **What it is:** The value that appears most frequently in a data set
- **When to use:** Recommended when identifying the most common or recurring value, especially for categorical data or when there are significant repetitions
- **Advantage:** Can be applied to qualitative variables and is not influenced by extreme values (outliers)

Standard Deviation

- **What it is:** Measure of data dispersion
- **Formula:** $\sqrt{[\Sigma(x_i - \text{mean})^2 / (n - 1)]}$
- **Interpretation:** The larger the value, the more dispersed the data

Variance

- **What it is:** Square of the standard deviation
- **Use:** Applied in advanced statistical calculations

Minimum and Maximum

- **What it is:** Minimum and maximum values of the criteria

Coefficient of Variation (CV)

- **What it is:** Standard deviation relative to the mean
- **Formula:** (Standard Deviation / Mean) × 100%
- **Interpretation:**
    - CV < 15% = low variability
    - CV > 30% = high variability

Skewness

- **What it is:** Measures the symmetry of the distribution
- **Values:**
    - 0 = symmetric distribution
    - 0 = right-skewed distribution
    - < 0 = left-skewed distribution

Kurtosis

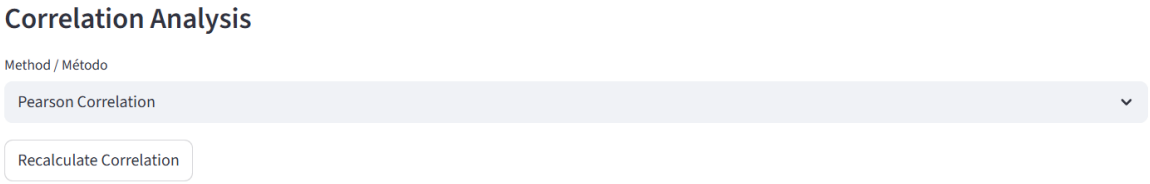
- **What it is:** Measures the “peakedness” of the distribution
- **Interpretation:**
    - 3 = normal (mesokurtic)
    - 3 = more peaked (leptokurtic)
    - < 3 = flatter (platykurtic)

4.0 Execution of Statistical Analyses

The software provides **additional statistical analysis tools** to deepen the analysis of the **payoff matrix data**, thereby improving the overall user experience.

4.1 Correlation Analysis

The **Correlation Analysis** functionality allows the evaluation of the degree of association between the analyzed criteria, supporting the understanding of the relationships among the study variables.



The user may select the correlation method in the **Method / Método** field, choosing between:

• Pearson Correlation

- **Purpose:** Identifies linear relationships between continuous numerical variables, especially when the data exhibit an approximately normal distribution.
- **Use:** For normally distributed data
- **Interpretation:**
    - −1 to −0.7: Strong negative
    - −0.7 to −0.3: Moderate negative
    - −0.3 to 0.3: Weak or no correlation
    - 0.3 to 0.7: Moderate positive
    - 0.7 to 1: Strong positive

• Spearman Correlation

- **Purpose:** Recommended when the data do not follow a normal distribution, contain outliers, or when monotonic relationships based on ranking are to be evaluated, regardless of linearity.
- **Use:** For non-parametric or ordinal data
- **Advantage:** Does not require normality
- **Application:** Suitable when outliers or non-normal distributions are present

After selecting the method, clicking **Recalculate Correlation** automatically updates the correlation coefficients and the associated graphical representations.

The results are displayed in **tabular form** and through a **heatmap**, allowing the user to:

- identify strong, moderate, or weak relationships between criteria;
- detect potential redundancies among variables;

- assess the risk of double counting information in multicriteria models.

This analysis is essential to ensure greater **conceptual consistency** in the criteria structure prior to decision making.

4.2 ANOVA and Tukey Test

The **ANOVA** functionality allows the user to evaluate whether there are **statistically significant differences between the means** of a numerical criterion considering categorical groups.

One-Way ANOVA

- **Purpose:** Compare means across 3 or more groups
- **Assumptions:**
    1. Normality (Shapiro–Wilk)
    2. Homogeneity of variances (Levene)
    3. Independence of observations

Levene's Test

- **Purpose:** Verify homogeneity of variances
- **Interpretation:** $p > 0.05$ = homogeneous variances

Tukey HSD Test

- **When to use:** After a significant ANOVA result ($p < 0.05$)
- **Objective:** Identify which groups differ from each other
- Select the **group column** (categorical)
- Select the **numeric variable** for analysis
- Click **Run ANOVA**

When clicking **Run ANOVA**, the software performs the statistical test and presents:

- the ANOVA result (statistical significance);
- when applicable, the **Tukey test**, which identifies which pairs of groups differ significantly.

⚠️ **Important note:**
The **Alternative** option is not recommended as a grouping variable when there is only **one observation per alternative**, as ANOVA requires multiple observations per group to produce statistically valid results.

Therefore, this functionality should be used as an **exploratory analysis tool**, helping the user understand data heterogeneity and the validity of groupings prior to applying multicriteria decision-making methods.

4.3 Linear Regression

The Linear Regression functionality allows the user to analyze and quantify the relationship between a dependent variable (Y) and one or more independent variables (X).

Regression Model

- **Formula:** $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \varepsilon$
- **Objective:** Predict the dependent variable (Y) based on predictors (X)

Evaluation Metrics

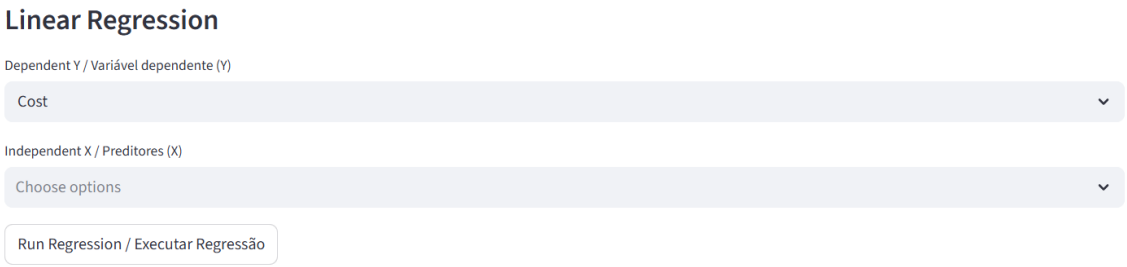
- **R^2:** Proportion of explained variance (0 to 1)
- **Adjusted R^2:** R^2 penalized by the number of predictors
- **MSE:** Mean Squared Error (lower = better)
- **MAE:** Mean Absolute Error

Diagnostics

- **Breusch–Pagan Test:** Heteroskedasticity ($p > 0.05$ = OK)
- **Durbin–Watson:** Autocorrelation (≈ 2 = OK)
- **VIF (Variance Inflation Factor):** Multicollinearity (VIF < 5 = OK)
- **Cook's Distance:** Identifies influential observations

In the Dependent Y / Variável dependente (Y) field, the user selects the criterion to be explained or predicted, such as:

- Cost
- Efficiency
- Risk



In the Independent X / Preditores (X) field, the explanatory variables are selected (e.g., *Efficiency* and/or *Risk*), which act as potential drivers of the dependent variable.

When clicking Run Regression, the software estimates the linear model and presents:

- model coefficients;
- direction of relationships (positive or negative);

- statistical metrics of fit and significance.

This functionality is particularly useful for exploratory analysis, enabling the user to understand interdependencies among criteria, identify approximate causal relationships, and support decisions regarding the structuring of the multicriteria model.

Interpretation of Results

- **p-values:** Compare with α = 0.05
- **Confidence intervals:** Check whether they include zero
- **Graphs:** Analyze patterns and trends

4.4 Monte Carlo Simulation

The **Monte Carlo Simulation** functionality allows the evaluation of the **robustness and sensitivity** of multicriteria analysis results under uncertainty in the input data.

Purpose

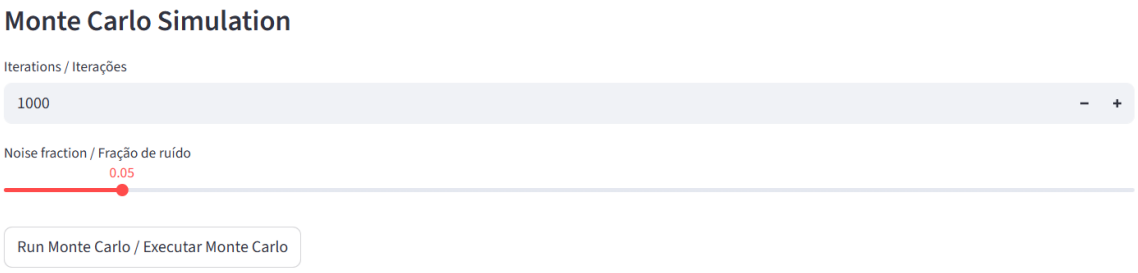
- Evaluate sensitivity of results to variations
- Test robustness of alternatives
- Quantify uncertainty

Parameters

- **Iterations:** Number of simulations (recommended: 1000+)
- **Noise fraction:** Perturbation intensity (0.05 = 5%)

Results

- **Mean score:** Average ranking of alternatives
- **Standard deviation:** Ranking stability



The Iterations parameter defines the number of random simulations to be performed. Higher values increase statistical stability.

The Noise fraction parameter controls the magnitude of random variation applied to the original criteria values, simulating real-world uncertainty such as:

- market fluctuations;
- data imprecision;

- operational variability.

When clicking Run Monte Carlo, the software generates multiple possible scenarios and evaluates the stability of alternative performance.

This analysis verifies whether the top-ranked alternative remains consistent under different conditions, providing a strong basis for risk analysis and robust decision making.

4.5 Normality Test

Shapiro–Wilk Test

- **Purpose:** Verify whether data follow a normal distribution
- **Hypotheses:**
  - H_0: Data are normally distributed ($p > 0.05$)
  - H_1: Data are not normally distributed ($p \leq 0.05$)
- **Recommendation:** $n < 50$ observations

4.6 Non-Parametric Test

The **Kruskal–Wallis Test** is used to compare the median of a numerical variable across two or more groups without assuming normality.

- **Alternative:** When ANOVA assumptions are violated
- **Use:** Non-parametric or non-normal data

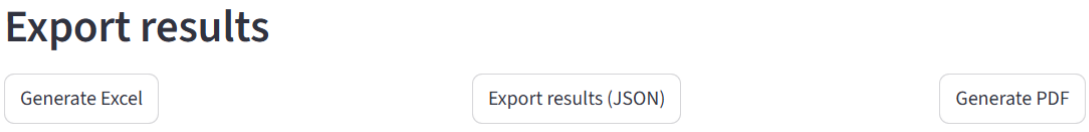
5. Exporting Results

The software provides multiple options for exporting results, enabling their use in reports, scientific articles, and academic activities.

Available options:

- **Generate Excel:** Exports results as spreadsheets for further analysis and archiving
- **Export Results (JSON):** Intended for system integration or advanced technical use
- **Generate PDF:** Produces a consolidated report suitable for documentation and presentation

These options enhance the **reproducibility and transparency** of the analyses performed.



6. Interaction with AI

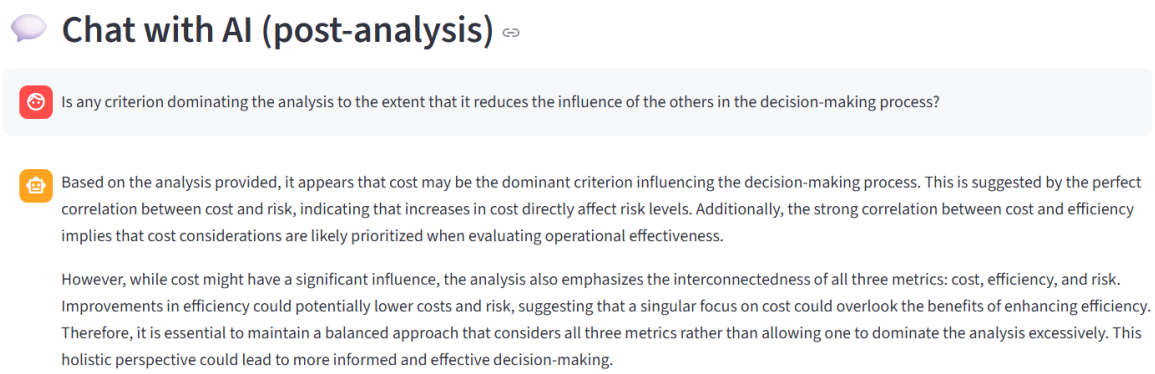
Another available feature is AI-Based Result Analysis, which automatically generates a textual interpretation of the computed statistical results.

When clicking Generate AI Analysis, the software produces:

- a concise interpretation of descriptive statistics;
- significant correlations among criteria;
- identification of relevant patterns;
- key statistical insights;
- ANOVA/Test results (significance);
- regression model performance;
- sensitivity of alternatives (Monte Carlo);
- practical, data-driven recommendations for decision making.

Main advantages include:

- interpretative support for novice users;
- increased productivity in report writing;
- translation of statistical results into accessible, structured language.



After generating the AI analysis, users may interact directly with the system through an **integrated chat**, asking questions about the obtained results.

Within this environment, users may:

- question whether any criterion is dominating the analysis;
- explore practical implications of the results.
- request detailed explanations or additional recommendations.
- explore hypothetical scenarios;
- request further clarification regarding correlations, variability, or risks.

The system responds contextually, relying exclusively on results computed by the software itself, reinforcing the **didactic, exploratory, and explanatory** nature of the tool.

7. Frequently Asked Questions (FAQ)

What is the difference between Alternative and Group?
Alternative represents individual options; *Group* represents categorical groupings.

Which correlation method should be used?
Pearson for normally distributed data and linear relationships; Spearman for non-normal data.

What is ANOVA used for?
To test statistical differences between groups.

Why are no results displayed?
There may be insufficient data or missing parameter selections.