

# **CogStat Manual**

**Attila Krajcsi**

---

# **CogStat Manual**

by Attila Krajcsi

Publication date 2014, 22 January

Copyright © 2014 Eötvös Loránd Tudományegyetem, Pedagógiai és Pszichológiai Kar, Pszichológiai Intézet, Kognitív Pszichológiai Tanszék

## **CogStat Manual**

Free, open source statistical package for simple statistical analysis

TÁMOP 4.1.2.A/1-11/1-2011-0018

## Table of Contents

1. Introduction to CogStat .....	1
1.1. Quick overview of the CogStat functions .....	1
2. Installing and starting CogStat .....	2
2.1. Installation and running CogStat on Windows .....	2
Installation .....	2
Starting CogStat .....	2
2.2. Installation and running CogStat on Linux .....	2
Installation .....	2
Starting CogStat .....	3
2.3. Installation and running CogStat on Mac .....	3
3. Handling data .....	4
3.1. How to handle data in CogStat .....	4
3.2. How should the data be arranged? .....	4
3.3. How should the data files are saved? .....	5
4. Exploring variables .....	6
4.1. Frequencies .....	6
4.2. Histogram .....	6
4.3. Descriptives .....	6
4.4. Normality .....	7
4.5. Testing central tendency .....	7
5. Exploring variable pairs .....	8
6. Pivot table .....	9
7. Compare variables .....	10
8. Compare groups .....	11
9. Handling the results .....	12

# 1. Introduction to CogStat

CogStat is a free, open source statistical software which enables quick and efficient statistical analysis, mostly optimized for (but not limited to) cognitive psychological research.

How is CogStat different?

- CogStat focuses on professional problems instead of statistical problems
  - The menus reflect professional questions, not the statistical procedures
  - Based on the professional question the statistical analysis is chosen by CogStat
  - CogStat automatically checks whether assumptions are met
- CogStat presents the results in both numerical and graphical form
  - It is more easy to understand what the results mean
  - Less chance to misunderstand the numerical results
- Clear output
  - Only relevant information is displayed
  - Appropriate results are displayed in APA format
- Free, open source, cross-platform
  - Free: no license fee
  - Open source: CogStat can be extended with some computer programming experience
  - Cross-platform: available on all three popular operating systems

## 1.1. Quick overview of the CogStat functions

The most important capabilities of CogStat include:

- Reading data
  - Data should be handled in spreadsheet software, and can be imported to CogStat either via the clipboard or via simple text file.
- Statistical analysis
  - Explore variable
    - Frequency, distribution, box plot, checking normality, checking central tendency
  - Explore variable pair
    - Scatter plot, regression equation, correlation, mosaic plot
  - Pivot table
  - Comparing variables
    - One-way comparisons for interval, ordinal and nominal variables
  - Comparing groups
    - One-way comparisons for interval, ordinal and nominal variables
- Saving results to pdf files

## 2. Installing and starting CogStat

Currently CogStat does not have a usual installation software, because the development had different focuses so far. Thus, installation is somewhat uncomfortable, but luckily it has to be performed only once per computer.

During installation Python (a computer language) and some Python extensions should be installed.

### 2.1. Installation and running CogStat on Windows

#### Installation

- On Windows you can download Python and many Python modules in a single install package. Such a package is pythonxy used for scientific computations.
  - Download pythonxy from here: <https://code.google.com/p/pythonxy/>
    - The version 2.7.5.1 might include a bug which stops CogStat. If there is no newer version, use the version 2.7.5.0. The latter can be found in the Download section on the Mirror link of the latest version.
  - Install pythonxy. While installing it, check statsmodel in the Choose components window, because it is unchecked by default.
- Install *configobj*. In the console (Start > Accessories > Command line) type *easy\_install configobj*, then press Enter.
- Download CogStat from its website: <https://sites.google.com/site/cogstatprogram/>, and unpack it to a folder.

#### Starting CogStat

You can use any of the following:

- Double click on cogstat.py
  - If .py files are not assigned to a Python interpreter (e.g., a file editor appears with the code in it, instead of CogStat running), then right click on cogstat.py, choose Open with > Choose default program, and choose Python from the available list.
- From the folder you unpacked CogStat to, open a command line and type "*c:\python27\python*" *cogstat.py*, and press Enter.

If any of the components is missing, the software warns you, and you can fix your installation.

### 2.2. Installation and running CogStat on Linux

#### Installation

- Install required packages.
  - On Debian and Ubuntu based Linux distribution install the following packages with your usual package manager (some of them may be already installed in your system): *python2.7*, *python-numpy*, *python-pandas*, *python-scipy*, *python-statsmodels*, *python-matplotlib*, *python-qt4*, *r-base*. *python-rpy2*, *r-cran-car*
    - If you use terminal, type: *sudo apt-get install python2.7 python-numpy python-pandas python-scipy python-statsmodels python-matplotlib python-qt4 r-base python-rpy2 r-cran-car*
  - Other distributions were not tested. Let us know about your experiences running CogStat on other Linux distributions.
- Download CogStat from its website: <https://sites.google.com/site/cogstatprogram/>, and unpack it to a folder.

## Starting CogStat

You can use any of the following:

- Double click on run\_cogstat\_on\_linux file, then choose "Run" from the options.
- Open a terminal from the directory you unpacked CogStat to, and type `python cogstat.py`, and press Enter.

If any of the components is missing, the software warns you, and you can check your installation.

## 2.3. Installation and running CogStat on Mac

There is no known limitation to run CogStat on Mac, although it was not tested. Let us know if you have any experience with running CogStat on a Mac.

## 3. Handling data

### 3.1. How to handle data in CogStat

In an unusual way, data cannot be edited in CogStat. Data should be handled in a spreadsheet software (e.g., Microsoft Excel, LibreOffice Calc), and the data can be imported via the clipboard or simple text files:

- As one possible solution the data can be selected in the spreadsheet (Edit > Copy), then the data can be pasted into CogStat (Data > Paste Data)
  - When pasting the data it is important how many decimal places were displayed in the spreadsheet software, because that is the precision which will be available in CogStat.
- As another solution, data can be saved as simple text files (.csv or .txt extensions), and the file can be opened in CogStat (Data > Open data file, or dragging the file to the CogStat window).

### 3.2. How should the data be arranged?

The data should be arranged as in all statistical software: rows are the cases and columns are the variables with the following rules.

- First row should include the variable names
  - If some of the names are missing, CogStat automatically gives the var1, var2, etc. names.
  - If all variable names are missing, the first line should be empty, otherwise incorrectly the data of first case will be used as variable names.
  - Two variables cannot have the same name. If this happens, CogStat will rename the later.
- Second line may include optionally the measurement levels.
  - *Nom*, *ord* and *int* words can be used to denote nominal, ordinal and interval variables. If any other word is in that line, CogStat will handle it as regular data line.
  - If no measurement level is specified, CogStat will set it to *unk* (unkown). Most statistical decisions of CogStat will handle it as interval variable. Thus, if the data include ordinal or nominal variables, those variables should be set to make CogStat able to choose the appropriate procedures.
  - Variables including text are set to nominal, no matter what level was specified in the file.
- The rest of the table are the data themselves.
  - In case of missing data either the cells can be left empty, or the word *nan* can be written.

An example of a data table:

ID	Gender	Creat
nom	nom	int
lcf	1	96
gok	1	121
tf	2	118
trs	1	128
rs	2	99

### 3.3. How should the data files are saved?

The data files should be saved as simple text files. The field delimiters should be tabulators. To save the data in this format follow these instruction:

- In LibreOffice Calc
  - File > Save as...
  - Format: Text CSV (.csv), Edit filter settings on
  - Field delimiter: {Tab}, Text delimiter: "
- In Microsoft Excel
  - File > Save as > Text (Tab delimited)
- In Google Spreadsheet
  - File > Download as > Plain text
- In Gnumeric
  - File > Save as
  - File Type: Text (configurable)
  - Separator: Tab, Quote character: "



## 4. Exploring variables

The following properties are calculated within these categories.

### 4.1. Frequencies

- Values
- Frequencies
- Relative frequencies
- Cumulative frequencies
- Cumulative relative frequencies

### 4.2. Histogram

- For ordinal and interval variables:
  - Histogram
    - Individual data are also displayed on x axis
  - Box plot
- For nominal variables:
  - Histogram

### 4.3. Descriptives

- For all variables:
  - Sample size
  - Number of invalid cases
- For at least interval variables:
  - Mean
  - Standard deviation
  - Skewness
  - Kurtosis
- For at least ordinal variables:

- Median
- Range

## **4.4. Normality**

- Anderson-Darling normality test
- Histogram with normal distribution curve
- Q-Q plot

## **4.5. Testing central tendency**

- For interval variable: one sample t-test
- For ordinal variable: Wilcoxon sign-test (Linux only)

## 5. Exploring variable pairs

For interval and ordinal variables:

- Compute correlation coefficient, and test its deviation from 0.
  - If both variables are interval, then Pearson correlation coefficient is calculated.
  - If the lowest measurement level is ordinal, Spearman correlation coefficient is calculated.
- Calculate the regression equation.
- Display scatter plot with the regression curve.
  - Size of the dots is proportional with the size of the data at that position.

For nominal variables:

- Compute Cramér's V measure of association.
- Compute chi-square test.
- Display contingency table.
- Display mosaic plot.

## 6. Pivot table

With the pivot table one can group a dependent variable as a function of one or more independent variables, then a specific function can be applied to the groups of the dependent variable. To use the pivot table specify:

- The dependent variable
- One or more independent variables. Any of the grouping variable can be used either in rows, columns or pages
- The function (sample size, sum, mean, median, standard deviation, variance)

## 7. Compare variables

After specifying the variables to compare the following results are computed:

1. Descriptives
  - Mean for interval variables
  - Median for ordinal variables
  - Contingency table for nominal variables
2. Graphs
  - For interval and ordinal variables box graph with individual data
    - Individual data are connected in the neighboring variables
  - For nominal variable mosaic plot
3. Hypothesis test
  - For two variables
    - Both variables are interval: choosing paired t-test
      - If normality assumption is met: paired t-test
      - If normality assumption is violated: paired Wilcoxon-test
    - Both variables are ordinal: paired Wilcoxon-test
    - Both variables are nominal: not implemented yet
  - For more than two variables
    - All variables are interval: choosing repeated measured ANOVA
      - Normality assumption is met: not implemented yet
      - Normality assumption violated: Friedman-test
    - All variables are ordinal: Friedman-test
    - All variables are nominal: not implemented yet

## 8. Compare groups

After specifying the grouping and dependent variables the following results are computed:

1. Descriptives
  - Mean for interval variables
  - Median for ordinal variables
  - Contingency table for nominal variables
2. Graphs
  - For interval and ordinal variables box graph with individual data
  - For nominal variable mosaic plot
3. Hypothesis test
  - Two groups are compared
    - For interval variable:
      - Normality and equal variance assumptions are met: two-sample t-test
      - Normality and equal variance assumptions are not met: Mann-Whitney-test
    - For ordinal variable: Mann-Whitney test
    - For nominal variable: Chi-square test
  - More than two groups are compared
    - For interval variable:
      - Normality and equal variance assumptions are met: one-way ANOVA
      - Normality and equal variance assumptions are not met: Kruskal-Wallis test
    - For ordinal variable: Kruskal-Wallis test
    - For nominal variable: Chi-square test

## 9. Handling the results

The results of the analyses can be saved to pdf file.

- Results > Save results or Results > Save results as commands save the output.
  - It always saves the actual content of the output. It overwrites the previous data in the output file.
- Results > Clear deletes the output window.

**Készült az Új Széchenyi Terv TÁMOP 4.1.2.A/1-11/1-2011-0018 sz. projektje keretében.**

