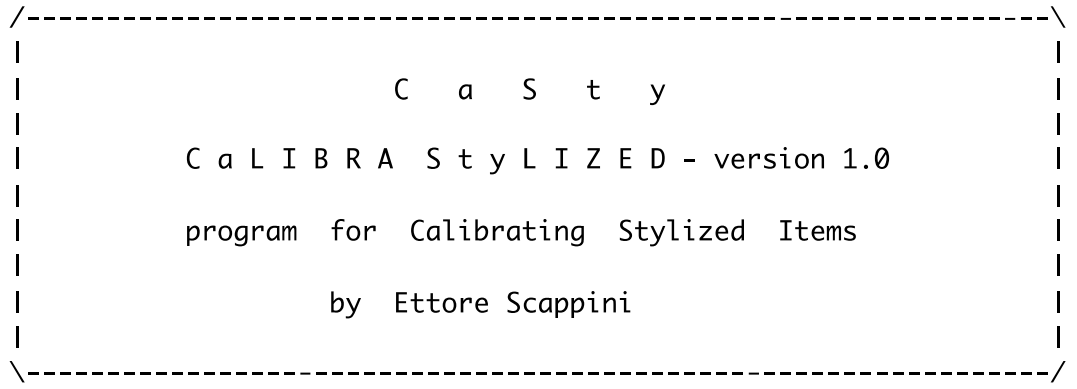


User Manual for CaSty.1.0.exe

Introduction

The CaSty.exe program aims to allow you to combine the information collected from the diaries with that from the stylized items in a very simple way, with the limit that both must be completed by each individual detected within a single survey. The data needed to obtain useful statistics are relatively few and the indispensable commands are very simple. Finally, the program provides tables and graphs containing the information typically needed to present a research report or paper.

This guide will develop with the presentation of examples and commands. When not indicated, the examples will be taken from the simulated data included in the '1.SimpleExample.txt' file. In the final part, the list of commands is presented in alphabetical order.

It should be noted that the current version, CaSty.1.0.exe, works only with the Windows operating system.

To start the procedures, copy the program to a directory in which you want to produce the output and in which there is a file called CaStyInp.txt and launch the program by double clicking on the icon.

Please note: If errors occur, send the data file and the error code to the following email: ettore.scappini@unibo.it

Preliminary information

For the correct functioning of the program and therefore to be able to obtain calibrated data, the program requires little basic information. The file containing the database and the input options is by default 'CaStyInp.txt', *the name cannot be changed*, and must include at least two records of equal length. When the program is launched, the 'CaStyLog.txt' file is created by default. This file is normally not strictly necessary as most of the information is also reported in the output excel format file (see below in the text). It becomes useful in the case of serious errors, in this circumstance the excel format file may not be generated, otherwise the CaStyLog.txt file is created when the program is started, thus allowing you to better understand what caused the interruption of the procedures.

We first point out that the values of an array and the multiple options of the commands must always be separated by a ';', see below in the text.

The minimum information necessary for the functioning of the program consists of two records.

A first record, in which the measured presence values broken down by the options associated with the stylized item used are entered. This record does not necessarily have to be sorted – for example the values can be in ascending or unsorted order, e.g. 8; 20; 12; 30; 50; 70; 90 –, but the sequence must have a correspondence with that derived from the encodings or modalities present in the item. The latter, the stylized presences, have to be ordered in an ascending monotonic way, but not necessarily a strictly ascending monotonic way.

In the indicated file the measured presence is necessarily placed in the first useful data record (even if not at the beginning), e.g.:

(remember that the data after the # is considered a comment)
Measurement presence (necessary data)
8; 20; 12; 30; 50; 70; 90

The second data record, in which the stylized presence values are inserted. This record must be sorted in non-descending order with a sequence corresponding to that of the

measured presence. In the indicated file, for example, this information is placed in the second useful data record, e.g.:

Stylized presence (not really necessary data)
0; 6; 8; 30; 60; 90; 100

This record is not strictly essential, in its absence the program will duplicate the first record indicated (Measurement presence).

Finally, the third (or second) data record in which the distribution in absolute values of the sample broken down by the options associated with the stylized item used is entered. This record must not be sorted in ascending order, but the sequence must correspond to that of the measured presence.
In the indicated file, for example, this information is placed in the third useful data record, e.g. :

Sample (N) (necessary data)
110;80;60;250;100;300;100

These are the necessary data for the correct functioning of the program. For example, if we insert the two data records indicated in the CaStyInp.txt file as follows:

8; 20; 12; 30; 50; 70; 90
110; 80; 60; 250; 100; 300; 100

we will be able to correctly generate an output. We specify that, in this case, some default options will be applied which can be omitted but which we now indicate with their commands:

Model: 1 Uniform; 6 Linear; 11 new Uniform; 16 new Linear (default)
Model = 16

Name file output (default)
Output = CaStyOut.xlsx

We emphasize that, as far as possible, the program is transparent in performing the various operations for which it is built: the various steps of the checks and calculations it is performing are written to the CaStyLog.txt file, while in the output file are reported the data, commands, steps of checks and calculations and also the main warnings.

Basic information for using the program

What the program does

In the first part of the procedures the program checks the accuracy of the data. It checks the length of the two, or three data arrays, highlighting the presence of any anomalies that will be signaled with a special warning or error code.
The program then goes on to check the type of data indicated in the sample array. If they are expressed in absolute values it percentages them, otherwise if expressed as a percentage, it checks that the sum is exactly 100 and, in the event that this check is not passed, it readjusts the values proportionally to the error in order to bring the sum to 100. Finally, it checks that the data in the measured presence vector are ordered in a strictly monotonic way. If this check is not verified, it merges the unsorted values in a weighted way, using proximity as a criterion. Then it also meets the corresponding values of the other vector/s.

E.g., hypothetical input data:
8; 20; 12; 30; 50; 70; 90
110; 80; 60; 250; 100; 300; 100

The program checks if array 1 is ordered in a strictly monotonic way. In this case it is not, therefore it checks the distance between the unsorted value and the adjacent ones. It combines the two values using the minimum distance as a criterion. In the hypothetical case presented, the unsorted value '20' will be combined in a weighted manner to the

closest value '12'. As a last step, the sample values will also be merged. If there are three vectors, the second vector will also be rearranged as the first.
Adjusted data:

8; 16.571; 30; 50; 70; 90
11; 14; 25; 10; 30; 10

At this point the program repeats the operation to check the correct ordering of the measured presence vector. If this check is not passed, the merging operation is restarted. Operation that will be repeated as many times as necessary until the measured presence array is obtained, ordered in a strictly monotonous manner.
Note. Only after verifying or re-adjusting the measured presence vector does the program perform a check on the Stylized presence vector. If this is not ordered in a 'non-decreasing' way, it is not necessary for it to be strictly monotone increasing, the program gives an error code and stops.
Only at this point does the calculation of the actual calibration begin.

How it calibrates data

The program procedures provide for calibration using the models described in the article to which we refer. There are four possible calibration models:
‘new Linear’ model also denoted by the letters ‘nL’, *this is the default model*;
‘new Uniform’ model also denoted by the letters ‘nU’;
‘Linear’ model also denoted by the letter ‘L’;
‘Uniform’ model or also denoted by the letter ‘U’.

What it produces as output

The program’s output is a file in excel format which can be assigned any name with or without an extension, this by default will always be set equal to 'xlsx'. The useful command is the **Output** command which, if omitted, has 'CaStyOut.xlsx' as a default value, or any valid file name can be assigned, e.g.:
Output = ValidNameFile.xlsx

The content of the **Output** file is divided into three parts placed in three different sheets with non-editable names.

The first sheet

The first sheet, called 'DataInfo', together with the CaStyLog.txt file, provides the transcription of the input file 'CaStyInp.txt' and the actual reading of the commands by the program highlighting any warnings, which highlight aspects that may require attention, errors, as a result of which the program stops, and invalid commands. In the latter case, the command is still read but it is signaled with a warning indicating an *unrecognized command*.

The second sheet

The second sheet (and possibly also the fourth or sixth or eighth if there are two or three or even four calibration models) is called 'Tables&Graphs|' to which a suffix equal to the name of the model is added. The names of the sheets indicated are, depending on the calibration model or models calculated, the following:

‘Tables&Graphs|nL’
‘Tables&Graphs|nU’
‘Tables&Graphs|L’
‘Tables&Graphs|U’

These sheets contain the data control and rearrangement tables, the output tables with the calibrated values useful for the analysis and two graphs.
There are two control tables, the first reproduces the distribution of the three variables for the options of the stylized item (Presence, Stylized and Sample) of the input data. The second, if present, the data rearranged with the indication in the title of the type of changes made, eg:

new Linear Distribution									
	Original data					Modify options values & percentag			
	Option/s	(sence (p_i)	lized (s_i)	(N)		Option/s	(sence (p_i)	lized (s_i)	%
	1	8,000	0,000	110		1	8,000	0,000	11,000
	2	20,000	6,000	80		2	16,571	6,857	14,000
	3	12,000	8,000	60		3	30,000	30,000	25,000
	4	30,000	30,000	250		4	50,000	60,000	10,000
	5	50,000	60,000	100		5	70,000	90,000	30,000
	6	70,000	90,000	300		6	90,000	100,000	10,000
	7	90,000	100,000	100					
						Total	45,700	51,460	100,000
	Total	45,700	51,460			(N)			(1000)
	(N)			(1000)					

The data table (s) with the calibrated values can be from one to three.

The first basic table, always present, contains all the information deemed useful for evaluating the effectiveness of the calibration and the calibrated values. Following the table, additional observations have been added which can be simply informative or even indicative on aspects to be paid attention to and / or warning / s.

Table, new Linear Distribution, nL(x)									
Option/s (i)	1	2	3	4	5	6	Total		
Measured presence (p_i) %	8,000	16,571	30,000	50,000	70,000	90,000	45,700		
Stylized presence (s_i) %	0,000	6,857	30,000	60,000	90,000	100,000	51,460		
Sample %	11,0	14,0	25,0	10,0	30,0	10,0	100,0		
(N)	(110)	(140)	(250)	(100)	(300)	(100)	(1000)		
Cumulative population %									
Stylized CF($S \leq s_i$)	11,0	25,0	50,0	60,0	90,0	100,0			
Calibrated nL($X \leq s_i$)	0,0	4,7	37,5	62,5	95,0	100,0			
Error Index (EI)	∞	430	33	-4	-5				
Retro-cumulative population %									
Stylized CF($S \geq s_i$)	100,0	89,0	75,0	50,0	40,0	10,0			
Calibrated nL($X \geq s_i$)	100,0	95,3	62,5	37,5	5,0	0,0			
Error Index (EI)		-7	20	33	700	∞			
Error Index (EI) = (CF - nL(x) / nL(x))*100									
NOTE: the Cumulative calibrated value nL(x) is calculated from the corresponding Stylized value.									
For example: in option 2 the stylized presence s_i = 6.857 then the corresponding nL($X \leq s_i$) = 4.7 .									
NOTE: the Meaured presence array was rearranged.									
NOTE: the parameter XxMin is missing, the default value was assigned: 0.000									
NOTE: the parameter XxMax is missing, the default value was assigned: 100.000									

In the event that the measured presence values are not ordered and therefore it is necessary to make one or more mergers, the program will calculate a second supplementary table containing the merged values and not present in the base table.

Table with the Original stylized presence distribution not present in the Table Base		
Option/s (i)	**7**	**8**
Measured presence (p_i) %	6,000	8,000
Cumulative population %		
Calibrated nL($X \leq p_i$)	4,1	5,5
Retro-cumulative population %		
Calibrated nL($X \geq p_i$)	95,9	94,5

The third table (or the second, if present) contains additional values that have been requested to calibrate with the AddStyVal command.
In this case if the command is:

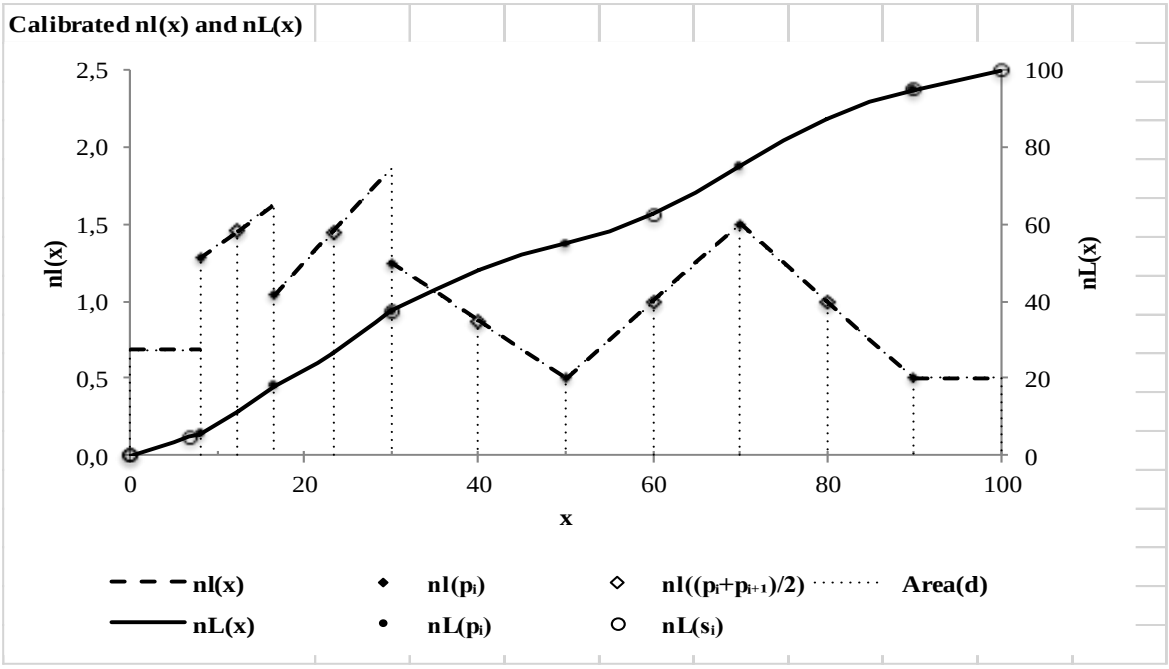
AddStyVal = *6.5; 16.5; 37.5; 55; 62.5; 75.0; 85.0; 95*

The following table will be obtained:

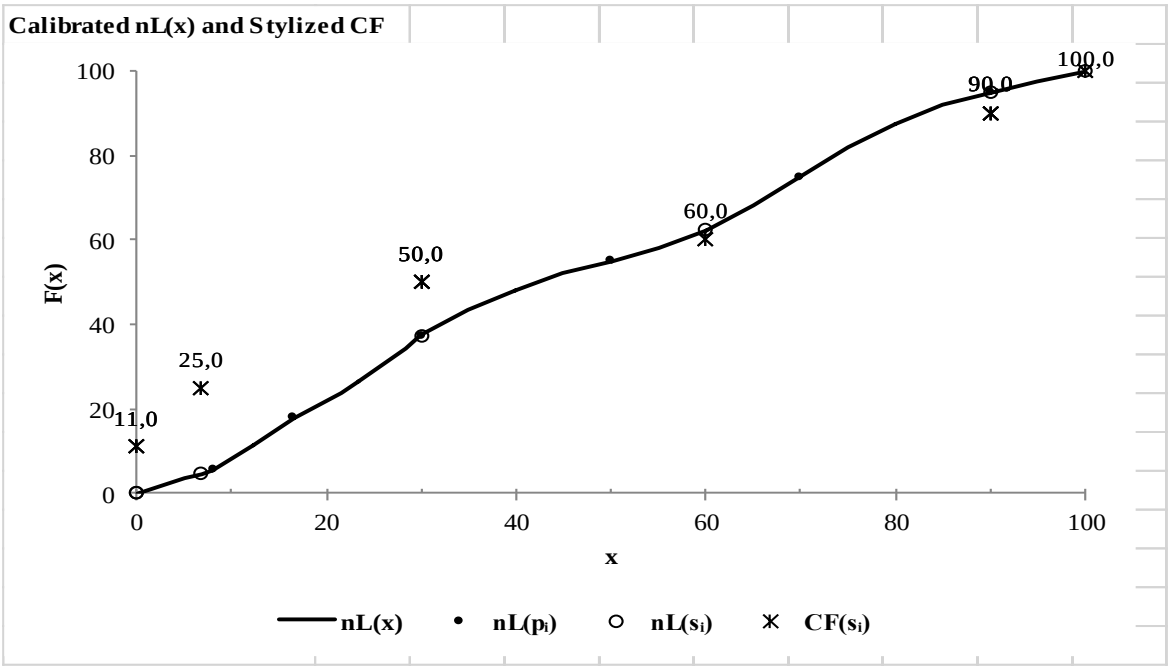
Table with Additional point/s								
Option/s	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Additional presence (x) %	6,500	16,500	37,500	55,000	62,500	75,000	85,000	95,000
Cumulative population %								
Calibrated nL($X \leq x$)	4,5	17,9	45,8	58,1	65,2	81,9	91,9	97,5
Retro-cumulative population %								
Calibrated nL($X \geq x$)	95,5	82,1	54,2	41,9	34,8	18,1	8,1	2,5

Two graphs are created from the calculated calibrated and uncalibrated data.

The first shows the values of f(x) and the calibrated ones F(x), indicated in this case, with nl(x) and nL(x) and, since 'CalibraArea = 1', the relative area.



The second graph shows the calibrated F(x) and uncalibrated (Stylized CF) values.



While it is possible to request only one, two or all four models at the same time for an analysis, it should be noted that the output format, consisting of tables and graphs, cannot be changed.

The third sheet

The third sheet (or the fifth or seventh or ninth) is called 'DataBase|' to which a suffix equal to the name of the model is added. As the name implies, this or these sheets contain the data base(s) on which the graphs are based. The names of the sheets indicated are, depending on the calibration model or models calculated, the following:

‘DataBase|nL’
‘DataBase|nU’
‘DataBase|L’
‘DataBase|U’

Limits

The program is built to calibrate arrays containing from a minimum of 4 to a maximum of 9 values.

How to use the commands

The commands that must be entered in the procedures are case-sensitive, they must therefore be typed as written taking into account the uppercase or lowercase characters (eg: AddStyVal and not addstyval or ADDSTYVAL). We also specify that all records starting with a hash # or an asterisk * are considered comment records.
All commands must be indicated by separating the command from the option with a '=', for example if the sample used is 1000 cases, to indicate it, we will have to use the Sample command and indicate as follows:

Sample = 1000

Finally, the data in the arrays or the multiple options must be separated by a ';'. If we want to, for example, calibrate in the four possible ways we should use the Model command and indicate as follows:

Comment: 16 –> nL(x) (default); 11 –> nU(x); 6–> L(x); 1 –> U(x)
Model = 1; 6; 11; 16

Glossary of commands (in alphabetic order)

AddStyVal = Additional calibrated values to the stylized values;
Other values for which calibration is required.
Example: **AddStyVal** = *6.5; 16.5; 37.5; 55; 62.5; 75.0; 85.0; 95*
No default value

CalibraArea = Locating the Calibration Area on the graph.
Options: 0, no indication concerning the identified area;
1, indications related to the identified area.
Example: **CalibraArea** = *1*
default value: **CalibraArea** = *0*

Cite = Citing the references.
Options:0, no references;
1, references.
Default value: Cite= 0

CdFCum = Calibration of Cumulative values or retro-cumulated values.
Options: 0, Cumulative F(x);
1, Retro-Cumulative F(x).
Example: **CdFCum** = *1*
default value: **CdFCum** = *0*

DecimalNot = Decimal Notation.
Options: ' . ' or *period* or *dot*, American or Great Britain notation to indicate the decimal place;
' , ' or *comma*, European countries notation to indicate the decimal place.
Example: **DecimalNot** = *'.'*
or just: **DecimalNot** = .
or **DecimalNot** = *dot*
default value: **DecimalNot** = *period*
Note: it is recommended or preferable to use the option period or '.' or dot.

FxMax = in the Graph, Maximum value of the Y2-axis, values of F(x);
useful in case it is necessary to compare several graphs.
Example: **FxMax** = 9*2.5*
default value: **FxMax** = *100*

FxMin = in the Graph, Minimum value of the Y2-axis, values of F(x);
useful in case it is necessary to compare several graphs.
Example: **FxMin** = *10.1*
default value: **FxMin** = *0*

FyMax = in the Graph, Maximum value of the Y-axis, values of f(x);
useful in case it is necessary to compare several graphs.
Example: **FyMax** = *2.5*
No default value

FyMin = in the Graph, Minimum value of the Y-axis, values of f(x);
useful in case it is necessary to compare several graphs.
Example: **FyMin** = *.1*
default value: **FyMin** = *0*

XxMax = Max value of x – p_{I+1} – in the model;
useful in case of fixed Max value of $p_{I+1} < 100$.
Example: **XxMax** = *93.603*
default value: **XxMax** = *100*

XxMin = Min value of x – p_{I+1} – in the model;
useful in case of fixed Min value of $p_0 > 100$.
Example: **XxMin** = *1.290*
default value: **XxMin** = *0*

Model = calibration used Model.
Options: 16, new Linear Calibration (nL);
11, new Uniform Calibration (nU);
6, Linear Calibration (L);
1, Uniform Calibration (U).
Example: **Model** = *1; 6; 11; 16*
default value: **Model** = *16*

Output = Output file.
Example: **Output** = *Output.xlsx* or simple **Output** = *Output*
default value: **Output** = *CaStyOut.xlsx*
Note: *if not present, the program adds to the file name the extension xlsx, and eventually change the extension from xls in xlsx.*

Sample = Number of cases in the sample (N).
Example: **Sample** = *1000*
No default value
Note: if this command is present, it is assumed that the 'sample distribution' vector is in percentages, differently if the sample option is not present it is assumed that the sample distribution is expressed in absolute values or cases.

StepData = improve the definition in the presentation of charts.
Example: **StepData** = 3
default value = None
Note: the range of possible values is between 1e-1 to 5.

Title = *Work title.*
Example: **Title** = -|-|-|-|- Calibra Data -|-|-|-|-
default value: **Title** = Calibra Stylized Items
Note: the length of the text cannot exceed 90 characters.