

PrefShare: a Software for Estimating the Market Share for New Products or Services with a Split Questionnaire Survey

Simone Balmelli and Francesco Moresino

Abstract

This manual proposes a quick start guide for the software PrefShare. This software permits one to estimate the market share for new products using a split questionnaire survey.

1 INTRODUCTION

PrefShare permits one to estimate the market share for new products using a split questionnaire survey. It implements the method proposed by Balmelli and Moresino (2018). The compressed file *Prefshare.zip* distributed with the supplementary materials contains all files needed to run the software.

PrefShare is a free software distributed under the terms of the GNU General Public License as published by the Free Software Foundation (see <http://www.gnu.org/licenses/>). It is distributed without any warranty; without even the implied warranty of merchandability or fitness for a particular purpose.

This manual is organized as follows. Section 2 proposes a quick start guide illustrated with an example whereas Section 3 recapitulates the inputs and outputs of all programs.

2 QUICK START GUIDE

PrefShare offers three programs to practitioners who desire to implement the method proposed by Balmelli and Moresino (2018). The first program permits one to design the questionnaire, the second one computes the coefficient of the logistic regression from the results of the survey and finally the third one computes the share of preference from the logistic regression. These programs are written in the language Octave, one of the free alternative to Matlab.

These three programs can be found in the supplementary materials accompanying the article Balmelli and Moresino (2018). The directory *PrefShare* contains four folders: the three programs are located in the folder *run_program*, the input and output data are placed in the folders *input* and *output*, finally the folder *functions* contains the library of functions used by the programs.

2.1 Experimental design

Given the salient attributes envisaged in the preliminary product analysis, the desired number of profiles to be addressed to each respondent and the desired number of respondent to be interviewed, the program *Experimental_design.m* computes two files. The first file provides

a D -optimal basis, that consists of the list of profiles retained for the survey. The second file contains the attribution matrix, which is the list of profiles to be addressed to each respondent.

To illustrate how this program functions, let us take a $4 \times 3^4 \times 2^3$ factorial example. To run the program, we must open the file *Experimental_design.m* with the software Octave and then, in the tab "Editor", click on the icon "Save File and Run" (see Figure 1a). Alternately, we can run the command "Experimental_design" in the tab "Command Window" (see Figure 1b). Then, a first dialog box appears (see Figure 2a). If we chose the option "Use the default settings", the program will use the default settings highlighted in blue in Figure 1c and no other dialog boxes are needed. We must notice that, in this case, v must be a divisor of kb . If we choose the option "Enter new settings", we will have to enter data manually. A dialog box will ask us to enter the number of levels for each attributes (see Figure 2b). In our example with the $4 \times 3^4 \times 2^3$ factorial example, we have eight attributes, the first one has 4 levels, the four following ones 3 attributes and the three last ones 2 attributes. Afterwards, we must enter the number of competition's profiles; for our example we chose 2 (see Figure 2c). Subsequently, we have to enter the design for each competition's profile, *i.e.* the attribute level for each attribute. Obviously, the level of each attribute must be between 1 and the number of levels of the considered attribute. In our example, the first profile has the minimal level for each attributes, whereas the second one the maximal level for each attribute (see Figures 2d-2e). Then, we are asked if we wish to impose a limit for the number of different profiles in the basis (see Figure 2f). Let us recall that the best results are obtained when the number of different profiles in the basis is as big as possible. However, for logistical reasons, it is sometimes more convenient to impose a limit on this number. In our example, we choose not to put a limit. Then, we have to enter the desired number of respondents and the desired number of card that will be addressed to each respondent. In our example, we wish to interview 300 respondents and present to each of them 10 profiles (see Figure 2g). At this point, let us recall that balanced designs are optimal, but not all combinations of parameters give a balanced design. For this reason, the program propose a balanced design that gives the best compromise with the wish of the user. In our example, the program proposes us to interview 287 respondents and present each of them 9 profiles (see Figure 2h). Suppose that we absolutely wanted 10 profiles. In this case, we have the possibility to change the importance weights of the parameters. In our example, we give to the number of profiles a huge weight (99 in this case) to force the program to find a solution where the number of profiles matches our wish (see Figure 2i). With these new weights, the program proposes us to interview 259 respondents and present each of them 10 profiles (see Figure 2j). We accept this proposition and, finally, we are asked to give the number of independent tries the program will use to compute the D -optimal basis (see Figure 2k). Let us recall that, unless trying all possibilities, which is obviously not possible for most real cases, it is necessary to use a heuristic to compute the D -optimal basis. For each try, the heuristic will randomly select a starting point. Obviously, the higher the number of tries, the better the optimal basis and the higher the computational time. Based on our observations, we recommend to use a value of 100 for the number of tries.

The program will then write three files in the folder *output*. The first file, named *D_basis.csv*, provides a D -optimal basis, that consists of the list of profiles retained for the survey (see Figure 3a). In this figure, we can see the 23 first profiles of the basis that contains, in our example, 2590 profiles. The second file is named *questionnaire_split.csv* and contains the attribution matrix, which is the list of profiles to be addressed to each respondent (see Figure 3b). In this figure, we can see the attributions for the 23 first respondents. Let us recall that in our example, there are 259 respondents and each of them will be present 10 profiles. Each line represents a respondent and contains the information of the 10 profiles (named cards in this

file) that will be presented to her. Finally, a third file, named *parameters.csv*, is also saved in the folder *output*. This last file contains the parameters of the problem and is used by the two other programs (*Logistic_regression.m* and *Preference_share.m*) as input data.

2.2 Logistic regression

Given the matrix of preferences revealed through the survey, the program *Logistic_regression.m* computes the coefficient of the logistic regression estimating the market share in function of the product profile. The input preferences matrix has the same structure as the attribution matrix. Concretely, if the respondent has preferred a given profile to the best competitor profile, the corresponding element of the matrix is filled with 1 and with 0 otherwise. The program provides a file containing the coefficient of the logistic regression.

To illustrate how this program functions, let us take again our $4 \times 3^4 \times 2^3$ factorial example. This program requires four input files. Three of them, namely *D_basis.csv*, *questionnaire_split.csv* and *parameters.csv*, are the outputs of the previous step and must stay in the folder *output*. The fourth one, namely *preference_data.csv*, must be placed in the folder *input*. This file contains the preferences matrix revealed with the survey and has the same structure as the attribution matrix contained in the file *questionnaire_split.csv*. Concretely, if the respondent has preferred a given profile (card) to the best competitor profile, the corresponding element of the matrix is filled with 1 and with 0 otherwise. In Figure 4a, we can see the choices of the 24 first respondents. We can see, for instance, that respondent 2 has preferred the third profile to the competition. Once the input files are ready, we can run the program *Logistic_regression.m*. Similarly as in the previous section, using the Octave environment, we can either run the program in one click or with a command line. In this last case, the command "Logistic_regression" must be run in the the tab "Command Window". The program will write in the folder *output* the file *regression_coefficients.csv* (see Figure 4b). This files contains the regressions coefficients of the logistic regression estimating the market shares. Note that we used a classical coding scheme, where for each attribute one level is removed in order to avoid collinearity. This file is used as input data for the next step.

2.3 Estimation of the market share

Finally, given the coefficient of the logistic regression, the program *Preference_share.m* computes the share of preference for any product profiles we want to study.

To illustrate how this program functions, let us take again our $4 \times 3^4 \times 2^3$ factorial example. This program requires three input files. Two of them, namely *regression_coefficients.csv* and *parameters.csv*, are the outputs of the previous steps and must stay in the folder *output*. The third one, named *profiles.csv*, contains the different profiles that we want to study (see Figure 5a). The columns of this file are the same as the columns of the file *D_basis.csv*, except that the first one has been removed. In our example, we have eight columns representing the eight attributes, and, as we decided to study four different profiles, we have four lines. Similarly as in the previous sections, using the Octave environment, we can either run the program in one click or with a command line. In this last case, the command "Preference_share" must be run in the the tab "Command Window". The program will write in the folder *output* the file *preference_share.csv*. This file contains the different profiles as defined in the input data with the corresponding estimation of the market share (see Figure 5b). We can see, for instance, that the market share of the third profile would be 41%.

3 INPUTS-OUTPUTS AT A GLANCE

This section proposes a recapitulation of the inputs and outputs for the three programs, namely *Experimental_design.m*, *Logistic_regression.m* and *Preference_share.m*. All output files are written in the folder *output* whereas all input files must be placed in the the folder *input*.

3.1 *Experimental_design.m*

For this program, the input data has to be entered as parameters either using the dialog boxes provided by the program or by writing the data directly in the program.

Input file: none.

Parameters:

number of levels for each attribute,
number of respondents,
number of cards per respondent,
number of regression design profiles,
concurrence profiles,
number of independent tries for the algorithm.

Output files: *D_basis.csv*, *questionnaire_split.csv* and *parameters.csv*.

3.2 *Logistic_regression.m*

Before running this program, we must have run the program *Experimental_design.m*. In particular, the three files *D_basis.csv*, *questionnaire_split.csv* and *parameters.csv* must be in the folder *output*.

Input file: *preference_data.csv*.

Parameters: none.

Output file: *regression_coefficients.csv*.

3.3 *Preference_share.m*

Before running this program, we must have run both the programs *Experimental_design.m* and *Logistic_regression.m*. In particular, the two files *regression_coefficients.csv* and *parameters.csv* must be in the folder *output*.

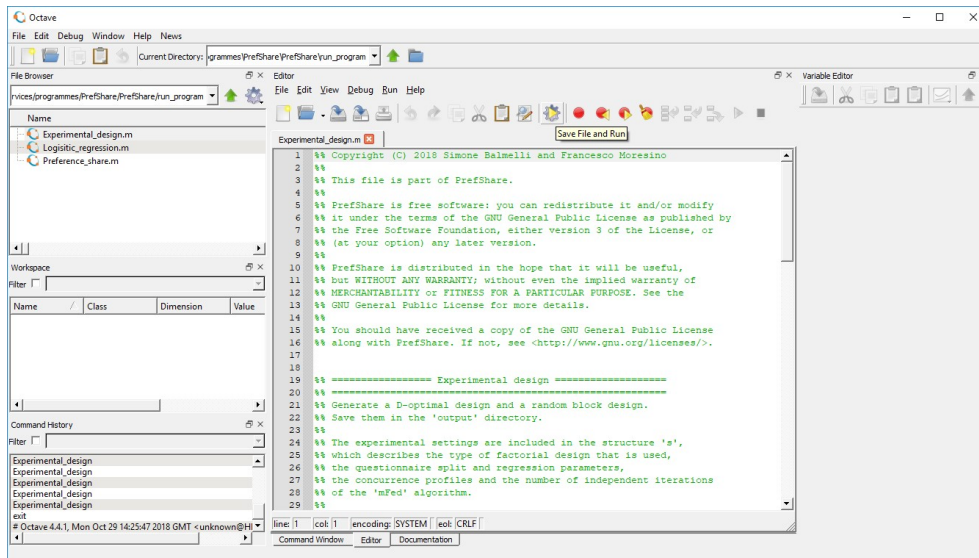
Input file: *profiles.csv*

Parameters: none.

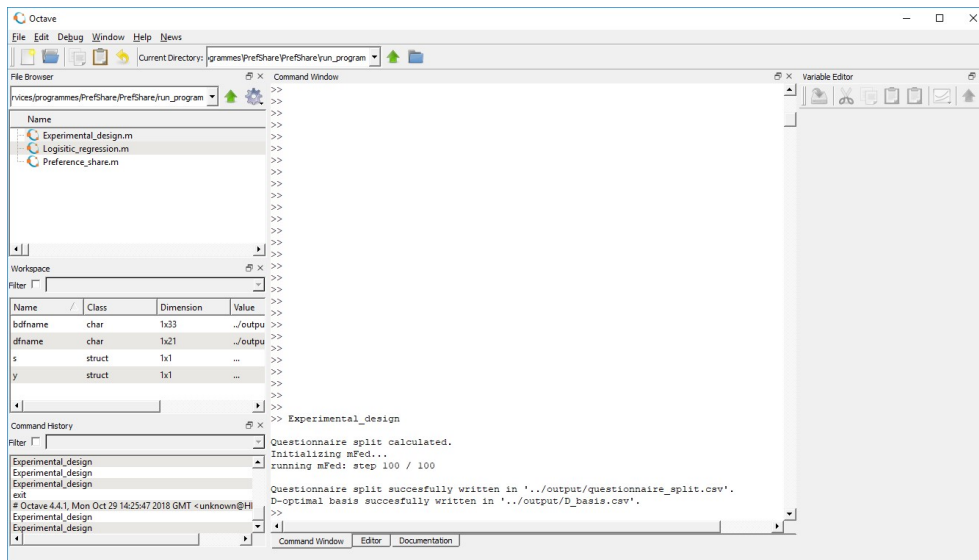
Output file: *preference_share.csv*.

REFERENCES

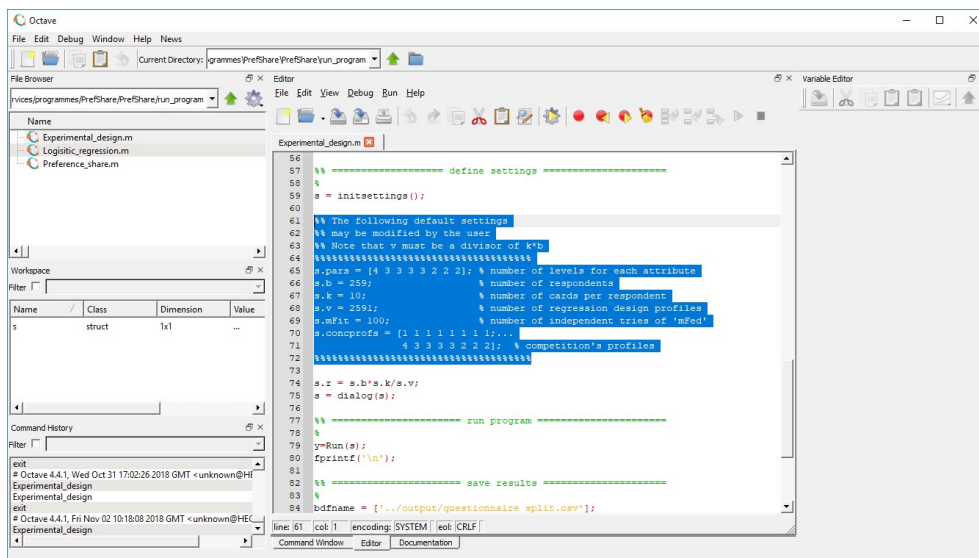
Balmelli, Simone and Francesco Moresino. 2020. “Estimating the Market Share for New Products With a Split Questionnaire Survey.” Submitted.



(a) Running the program in one click.



(b) Running the program with a command.



(c) Default data that can be changed by the user.

Figure 1: Octave's environment for the program *Experimental_design.m*.

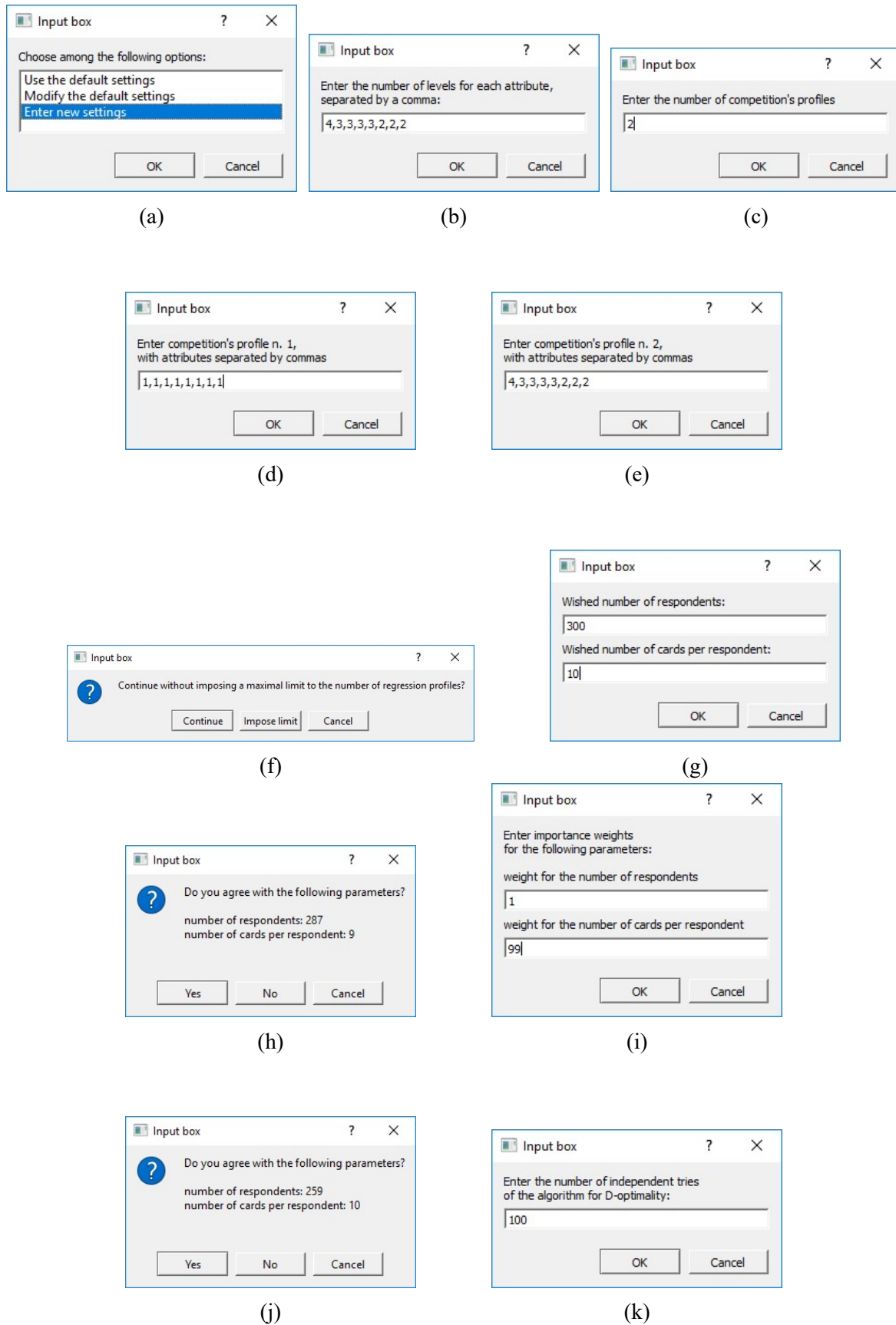


Figure 2: Dialog boxes for the program *Experimental_design.m*.

D_basis.csv - Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	profile	attribute1	attribute2	attribute3	attribute4	attribute5	attribute6	attribute7	attribute8			
2	1	1	1	1	1	1	1	1	2			
3	2	1	1	1	1	1	1	1	2	1		
4	3	1	1	1	1	1	1	2	2			
5	4	1	1	1	1	1	2	1	1			
6	5	1	1	1	1	1	2	1	2			
7	6	1	1	1	1	1	2	2	1			
8	7	1	1	1	1	1	2	2	2			
9	8	1	1	1	1	2	1	1	1			
10	9	1	1	1	1	2	1	1	2			
11	10	1	1	1	1	2	1	2	1			
12	11	1	1	1	1	2	1	2	2			
13	12	1	1	1	1	2	2	1	1			
14	13	1	1	1	1	2	2	1	2			
15	14	1	1	1	1	2	2	2	1			
16	15	1	1	1	1	2	2	2	2			
17	16	1	1	1	1	3	1	1	1			
18	17	1	1	1	1	3	1	1	2			
19	18	1	1	1	1	3	1	2	1			
20	19	1	1	1	1	3	1	2	2			
21	20	1	1	1	1	3	2	1	1			
22	21	1	1	1	1	3	2	1	2			
23	22	1	1	1	1	3	2	2	1			
24	23	1	1	1	1	3	2	2	2			

(a) The file *D_basis.csv*.

questionnaire_split.csv - Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	respondent	card1	card2	card3	card4	card5	card6	card7	card8	card9	card10	
2	1	368	653	690	950	1263	1435	1539	1592	1884	1938	
3	2	45	328	931	1120	1144	1256	1449	1727	1861	1906	
4	3	64	172	227	649	771	899	1114	1360	1867	1993	
5	4	185	598	698	755	2147	2229	2432	2440	2443	2488	
6	5	10	50	326	593	742	804	1030	1791	1919	2066	
7	6	101	269	327	2086	2233	2348	2383	2490	2570	2572	
8	7	336	467	553	717	893	1115	1292	1423	1485	2236	
9	8	174	635	708	1173	1407	1664	1856	2074	2312	2441	
10	9	203	792	898	904	1001	1321	1355	1858	1989	2182	
11	10	305	499	512	633	873	1014	1044	1124	1146	1229	
12	11	111	373	488	503	576	677	1234	1336	1675	1814	
13	12	97	505	592	616	1805	1838	2143	2214	2366	2461	
14	13	70	391	843	869	1079	1808	1830	1918	1974	2172	
15	14	106	697	722	1109	1380	1431	1448	1712	2138	2382	
16	15	173	320	460	579	1222	1889	2266	2285	2318	2384	
17	16	682	1397	2467	2483	2513	2530	2559	2560	2566	2567	
18	17	131	183	273	322	353	376	640	1143	1388	1501	
19	18	73	330	387	519	678	1196	1266	1464	1534	2206	
20	19	18	109	233	566	608	877	901	1200	1319	1715	
21	20	133	163	207	211	296	632	875	880	1021	1100	
22	21	259	405	628	665	958	967	974	1066	1478	1740	
23	22	313	463	764	818	938	1210	1235	1400	1569	1921	
24	23	49	756	1118	1298	1834	1874	2015	2361	2363	2375	

(b) The file *questionnaire_split.csv*.Figure 3: Output files of the program *Experimental_design.m*.

preference_data.csv - Excel

Formules Données Révision Affichage Développeur Acrobat Recherche Connexion Partager

Calibri 11

Standard

Mise en forme conditionnelle Insérer Σ

Mettre sous forme de tableau Supprimer %

Styles de cellules Format 0,00

Cellules Édition

A1 X ✓ f x respondent

	A	B	C	D	E	F	G	H	I	J	K	L
	respondent	card1	card2	card3	card4	card5	card6	card7	card8	card9	card10	
1	1	0	0	1	1	1	0	1	1	0	1	
2	2	1	0	1	1	1	1	1	1	1	0	
3	3	0	0	0	1	1	0	0	1	1	1	
4	4	1	1	0	1	0	1	0	1	1	0	
5	5	1	1	1	1	0	1	1	0	0	1	
6	6	0	0	1	0	1	1	0	1	0	1	
7	7	0	1	1	1	0	1	1	1	0	1	
8	8	0	0	1	1	1	0	0	0	1	0	
9	9	0	0	1	1	1	1	1	0	0	0	
10	10	0	1	0	1	0	1	1	1	0	0	
11	11	1	0	1	0	1	0	1	0	1	0	
12	12	0	0	0	1	0	1	0	1	1	0	
13	13	0	1	1	1	0	1	1	0	0	0	
14	14	0	0	0	1	1	1	0	0	0	1	
15	15	1	0	0	0	0	1	0	0	0	1	
16	16	0	1	0	0	1	1	1	0	0	0	
17	17	1	1	0	1	1	0	0	0	0	1	
18	18	0	1	0	0	0	0	0	0	1	1	
19	19	0	0	1	1	0	0	0	1	0	0	
20	20	0	0	1	1	0	0	0	1	0	0	
21	21	0	1	0	1	0	1	1	1	0	1	
22	22	0	0	0	1	1	1	0	0	0	0	
23	23	1	0	0	1	0	0	0	0	0	0	

preference_data

Prêt

(a) The file *preference_data.csv*.

regression_coefficients.csv - Excel

Formules Données Révision Affichage Développeur Acrobat Recherche Connexion Partager

Calibri 11

Standard

Mise en forme conditionnelle Insérer Σ

Mettre sous forme de tableau Supprimer %

Styles de cellules Format 0,00

Cellules Édition

A1 X ✓ f x intercept_term

	A	B	C	D	E	F	G	H	I	J	K
1	intercept_term	-0.257925									
2	attribute1level2	0.107914									
3	attribute1level3	0.114149									
4	attribute1level4	0.111268									
5	attribute2level2	-0.030247									
6	attribute2level3	0.013933									
7	attribute3level2	0.011714									
8	attribute3level3	0.083644									
9	attribute4level2	0.151086									
10	attribute4level3	0.181336									
11	attribute5level2	0.053259									
12	attribute5level3	0.000227									
13	attribute6level2	-0.063951									
14	attribute7level2	0.079507									
15	attribute8level2	-0.001571									
16											
17											
18											
19											
20											
21											
22											
23											
24											

regression_coefficients

Prêt

(b) The file *regression_coefficients.csv*.Figure 4: Input and output files for the program *Logistic_regression.m*.

profiles.csv - Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	attribute1	attribute2	attribute3	attribute4	attribute5	attribute6	attribute7	attribute8				
2	2	1	1	1	2	2	2	2				
3	4	3	3	3	3	2	2	2				
4	1	2	1	1	1	2	1	2				
5	3	3	3	3	2	1	2	1				
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												

(a) The file *profiles.csv*.

preference_share.csv - Excel

	A	B	C	D	E	F	G	H	I	J	K
1	attribute1	attribute2	attribute3	attribute4	attribute5	attribute6	attribute7	attribute8	preference_share		
2	2	1	1	1	2	2	2	2	0.47932		
3	4	3	3	3	3	2	2	2	0.536552		
4	1	2	1	1	1	2	1	2	0.412487		
5	3	3	3	3	2	1	2	1	0.566578		
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											

(b) The file *preference_share.csv*.Figure 5: Input and output files for the program *Preference_share.m*.