

WEIGHTING COMPLEX SAMPLES

IAN-AF Databases

Tutorial using software SPSS e R



Table of Contents

Nota introdutória	3
1. Software SPSS	4
2. Software R	17

References

- [1] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-roject.org/.
- [2] T. Lumley (2017) "survey: analysis of complex survey samples". R package version 3.32.
- [3] T. Lumley (2004) Analysis of complex survey samples. Journal of Statistical Software. 9(1): 1-19



Introductory Notes

In the National Food and Physical Activity Survey, IAN-AF 2015-2016, participants were randomly selected from the National Register of Users of the National Health Service, based on a two-stage complex sampling process. The sampling process proceeded as follows:

- Primary Health Units were randomly selected in each Territorial Unit for Statistical Purposes (NUTS II). In each region, the sampling was weighted taking into account the number of individuals. The number of Primary Health Units selected was 21 in the North region, Centre and Metropolitan Area of Lisbon, 12 in the Algarve and Alentejo regions and six in the Autonomous Regions of Madeira and the Azores.
- ii. Individuals registered in each Primary Health Units were randomly selected, with a fixed number of elements by sex and age group.

To estimate the results according to the IAN-AF 2015-2016 complex sample design, at national and regional level (NUTS II), the results are weighted according to a created variable. The sample weights represent how many individuals of the Portuguese population (in number) each individual of the sample represents. The calculation of sample weights included the following criteria:

- i. initial weighting to compensate for the different probabilities of selection of each Primary Health Units;
- weighting to compensate for the different probabilities of selection of each individual in each Primary Health Units, by sex and age group (considering the individuals in each Primary Health Units, in the closest recruitment wave);
- iii. correction of the initial weights for the non-response bias.

At the end, in order to correct data for non-response bias of both first and second interview, two weight variables were created. The first, *Ponderador1*, is used for data collected in the first interview, and the second weight variable, *Ponderador2*, is used for data collected in the second interview. Thus, all estimates referring to the domains Physical Activity and Nutritional Status must use the weight variable *Ponderador1*, while the domain Food must use the weight variable *Ponderador2*.

Next, we present a brief tutorial on how to use the SPSS and R [1] software in order to obtain weighted estimates according to the complex sampling design of the IAN-AF 2015-2016, using the SPSS and R software [1].

1. Software SPSS



In order to obtain weighted estimates according to the IAN-AF 2015-2016 complex sampling design in SPSS, first it is necessary to create a file that indicates the complex sampling design used. To do it so, it is mandatory to have the variables "PSU", "NUT" and the respective weighting variable, which can be found in the sociodemographic database. Thus, it is always necessary to merge the sociodemographic database with the database containing the variables under study.

is View Date Transfer Output Mage	🚂 InsAlim JANAFcompleta, v2.sav (DataSet1) - IBM SPSS Statistics Data Editor – 🗗 X																	
Produ Produ Produ Production	<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	<u>T</u> ransform	Analyz	Direct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> ti	lities Add- <u>o</u>	ns <u>W</u> indov	w <u>H</u> elp								
Condentation Descriptive Statistics Value 1 10410-02-052 Congoare Maans Consolitation C				Re	ports		•	*		2			ABG					
2: Prostandort 19230 0 Cuggars Hand 2: Foreitan 2: Foreitan 2: Foreitan 2: Foreitan 2: Foreitan 2: Foreitan 3: Foreitan 4: For				Dg	scriptive Statistics		•				1 🗐 🕚							
IAN_D Classe idade Seco Ponderadori EscoProp Classe_Escol AgF/menos Decom 1 101-01-05-0541 General lanes Model 20 1 0 2110-44 4 2 1 5 3 101-01-05-0541 General lanes Model 22 1 0 2110-44 6 2 0 1 4 101-01-05-0541 General lanes Model 22 1 0 2110-44 6 2 0 1 5 101-01-05-05403 General lanes 41 1 0 2453.76 3 2 0 1 0 2 0 1 0 1 0 2 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	12 : Ponderad	ior1 5	29.06	Cu	istom Ta <u>b</u> les		•									Visible: 79 o	f 79 Varia	ables
1 101010-05-054 Generalized Linear Model 2 1 0 2110.44 6 3 0 2 10101-005-058 Connella 22 1 0 2110.44 6 3 0 4 10101-005-0588 Connella 22 1 0 2110.44 6 3 0 5 10101-006-0508 Connella 22 1 0 2463.78 3 2 0 1 0 6 10101-006-06308 Connella 53 1 0 2463.78 3 2 0 1 0 <td></td> <td>IAN</td> <td>_ID</td> <td>Co</td> <td>mpare Means</td> <td></td> <td>۶.</td> <td>idade</td> <td>Classes</td> <td>ldade</td> <td>Sexo</td> <td></td> <td>Ponderador1</td> <td>EscolProp</td> <td>Classe_Escol</td> <td>AgF7menos</td> <td>Dicoto</td> <td>m</td>		IAN	_ID	Co	mpare Means		۶.	idade	Classes	ldade	Sexo		Ponderador1	EscolProp	Classe_Escol	AgF7menos	Dicoto	m
1 10440404584 20 1 0 21104 4 4 1 3 1044040588 Mard Models 223 1 0 21104 5 2 0 5 1044040588 Qarrelate 222 1 0 21104 5 2 0 5 1044040588 Qarrelate 222 1 0 248378 3 2 0 6 10440405483 Ragresion 41 1 0 248378 3 2 0 7 10161406593 Nural Netgoris 65 1 0 248378 6 3 0 10 10414045.138 Saje 54 1 0 248378 6 3 0 11 1040406.639 Remained Reduction 42 0 248378 6 3 0 0 12 1040406.635 Saje 54 1 0 248378 6 3 0 0 13 10404046.6163 Matemaninet Reduction 47 <	1	1 01 01 0 05	054	Ge	eneral Linear Model		•	2	0	1		2	2110 44	4	2		4	
2 10/10/10/40/201 Mage Models 22 1 0 21/0.44 5 2 0 3 10/10/10/40/201 Bayession 41 1 0 24/378 3 2 0 6 10/10/10/40/201 Eggines 58 1 0 24/378 1 1 0 7 10/10/06/203 Carrelation 58 1 0 24/378 4 2 0 9 10/10/10/205/124 Dimension Reduction 42 1 0 24/378 4 2 0 10 10/10/10/205/124 Dimension Reduction 42 1 0 24/378 3 2 0 11 10/10/10/205/124 Dimension Reduction 42 1 0 24/378 3 2 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 0 1 1 0 1 0 1 0 1 <td>2</td> <td>1.01.01.0.05</td> <td>054</td> <td>Ge</td> <td>eneralized Linear Mode</td> <td>els</td> <td>•</td> <td>2</td> <td>0 2</td> <td>1</td> <td></td> <td>0</td> <td>2110.44</td> <td>4</td> <td></td> <td></td> <td>1</td> <td></td>	2	1.01.01.0.05	054	Ge	eneralized Linear Mode	els	•	2	0 2	1		0	2110.44	4			1	
3 10/10/10/00/0000 Contrate 22 1 0 21/00/4 6 3 0 5 10/10/10/00/0000 Bgression 41 1 0 24/07/4 3 2 0 5 10/10/10/00/0000 Bgression 41 1 0 24/07/4 3 2 0 7 10/10/40/00/0000 Bgression 41 1 0 24/07/4 1 1 0 7 10/10/40/00/000 Neural Ntegers 63 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 24/07/4 1 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 0 1 1 0 <td>2</td> <td>1.01-01-0-05</td> <td>0001</td> <td>Mi</td> <td>xed Models</td> <td></td> <td>•</td> <td>2</td> <td>ა ი</td> <td>1</td> <td></td> <td>0</td> <td>2110.44</td> <td>5</td> <td>2</td> <td></td> <td>0</td> <td>- 1</td>	2	1.01-01-0-05	0001	Mi	xed Models		•	2	ა ი	1		0	2110.44	5	2		0	- 1
a biological Begression 2 c	4	1.01.01.0.05	125	<u>C</u> 0	prrelate		•	2	2	1		0	2110.44	6	3		0	- 1
5 10/10/10/06/000 Loglinear 1 0 2450.76 1 1 0 7 10/10/10/06/003 Casas 65 1 0 2450.76 1 1 0 0 9 10/10/10/06/033 Casas 1 0 2450.76 1 1 0 0 9 10/10/10/06/053 Casas 4 2 0 2450.78 3 2 0 10 10/10/10/07/065 Rynametric Tests 66 2 0 533.06 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1	5	1-01-01-0-05	008	Re	gression		•	4	1	1		0	2463 78	3	2		0	
7 14101-08-093 Neural Netgoris 65 1 0 243.78 6 3 0 8 1011-09-06-124 Dimension Reduction 42 1 0 243.78 4 2 0 10 101-09-06-124 Dimension Reduction 42 1 0 243.78 4 2 0 11 10-10-06-136 Sigle 54 1 0 243.78 6 3 0 12 101-10-07-065 Forecasting 77 1 2 0 53.90.6 1 1 0 13 101-10-07-065 Survial 47 1 1825.72 2 1 0 14 10-10-16-040 Multiple Response 55 1 1 1825.72 2 1 0 15 101-10-16-040 Comple Samples 9 9 1 1 1825.72 2 1 0 10 101-10-16-040 Comple Samples 9 1 1 1825.72 2 1 0 1 1	6	1.01.01.0.06	.050	Lo	glinear		۶.	5	8	1		0	2463.78	1	1		0	- 1
8 1-01-0-06-109 Classify + 40 1 0 2463.78 4 2 0 9 1-01-0-06-124 Dimension Reduction + 42 1 0 2463.78 3 2 0 10 10-10-0-06-18 Scie + 42 1 0 2463.78 3 2 0 11 10-10-0-06-16 Monparametric Tests 66 2 0 5329.06 1 1 0 12 10-10-0-60.16 Surval + 71 2 0 5329.06 1 1 0 13 10-10-0-16.03 Wighte Response 53 1 1825.72 2 1 0 15 10-10-1-6.036 Wighte Imputation + 48 1 1 1825.72 4 2 0 16 10-10-1-6.036 Wighte Imputation - 48 1 1 1825.72 4 2 0 19 10-10-1-6.075 Simulation Quality Control Tim Generity Bastrop Control Quality Control	7	1-01-01-0-06	093	Ne	aural Net <u>w</u> orks		•	6	3	1		0	2463.78	6	3		0	- 1
9 101010-06-124 Dimension Reduction + 42 1 0 2463.78 3 2 0 10 10101-06-136 Scale + 54 1 0 2463.78 6 3 0 11 10101-06-766 Bionparametric Tests + 65 2 0 5329.66 1 1 0 12 10101-06-766 Burwal + 47 1 1825.72 2 1 0 13 10101-06-1603 Wutsile Response + 58 1 11825.72 4 2 0 15 10101-06-1603 Wutsile Imputation + 48 1 1825.72 4 2 0 16 10101-06-1607 Complex Samples + 1 1 1825.72 4 2 0 1 10 10101-06-1607 Gaily Control - Prepare for Anajesis 1 1 1825.72 4 2 0 1 10 10101-06-1607 Gaily Control - Prequencles 1 </td <td>8</td> <td>1-01-01-0-06</td> <td>109</td> <td>CI</td> <td>assi<u>f</u>y</td> <td></td> <td>۶.</td> <td>4</td> <td>0</td> <td>1</td> <td></td> <td>0</td> <td>2463.78</td> <td>4</td> <td>2</td> <td></td> <td>0</td> <td></td>	8	1-01-01-0-06	109	CI	assi <u>f</u> y		۶.	4	0	1		0	2463.78	4	2		0	
101 101-01-0-06-136 Scale	9	1-01-01-0-06	124	Di	mension Reduction		۲.	4	2	1		0	2463.78	3	2		0	
11 101-01-007.066 Monparametric Tests 65 2 0 5329.06 1 1 0 12 101-01-07.095 Forecasting 71 2 0 5329.06 1 1 0 13 101-01-06.036 Multiple Response 53 1 1 1825.72 2 0 0 15 101-01-06.036 Multiple Imputation 48 1 1 1825.72 2 0 0 17 10-01-01-06.036 Multiple Imputation 48 1 1 1825.72 2 0 0 18 10-01-01-06.040 Complex Samples 9 Sate at Sample. 1 1 1825.72 2 0 0 19 10-01-01-05.040 Quality Control 9 Perpare for Analysis 1 1 1825.72 4 2 0 20 10-01-01-05.040 Quality Control 9 Perpare for Analysis 1 1 1825.72 4 2 0 21 10-01-01-05.050 1 1 100 2 </td <td>10</td> <td>1-01-01-0-06</td> <td>136</td> <td>So</td> <td><u>a</u>le</td> <td></td> <td>۶.</td> <td>5</td> <td>4</td> <td>1</td> <td></td> <td>0</td> <td>2463.78</td> <td>6</td> <td>3</td> <td></td> <td>0</td> <td></td>	10	1-01-01-0-06	136	So	<u>a</u> le		۶.	5	4	1		0	2463.78	6	3		0	
12 101010-07-095 Forecasing 71 2 0 5329.06 1 1 0 13 101010-16016 Survial Mutiple Response 63 1 1825.72 2 1 0 14 101010-16034 Mutiple Response 68 1 1825.72 2 1 0 15 101010-16035 Mutiple Imputation 48 1 1825.72 2 1 0 18 101010-16036 Complex Samples 9 9 9 9 1 1 1825.72 4 2 0 1 19 101010-16040 Complex Samples 9 9 9 9 1 1 1825.72 4 2 0 1 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	11	1-01-01-0-07	065	N	onparametric Tests		•	6	5	2		0	5329.06	1	1		0	
13 101-01-0-16-016 Sunival 47 1 1825.72 2 1 0 14 1-01-01-0-16-034 Multiple Response 53 1 1825.72 3 2 0 15 1-01-01-0-16-035 Multiple Imputation 48 1 1825.72 2 1 0 16 1-01-01-0-16-035 Multiple Imputation 48 1 1825.72 2 1 0 17 1-01-01-0-16-075 Signulation Quality Control 1 1 1825.72 4 2 0 19 1-01-01-0-16-075 Signulation Quality Control 1 1 1825.72 4 2 0 20 1-01-01-0-16-089 Prepare for Analysis 1 1 1825.72 4 2 0 21 1-01-01-0-16-075 Signulation Quality Control 1 1 0 21 1 1 0 21 0 1 0 21 0 1 0 21 0 1 0 22 0 1	12	1-01-01-0-07	095	Fo	recasțing		۶.	7	1	2		0	5329.06	1	1		0	
14 101-01-0-16-034 Multiple Response 53 1 1 1825.72 3 2 0 15 101-01-0-16-035 Wultiple Imputation 48 1 1 1825.72 4 2 0 16 101-01-0-16-035 Multiple Imputation 48 1 1 1825.72 4 2 0 18 101-01-0-16-035 Multiple Response • 9 get a Sample 1 1 1 1825.72 6 3 0 18 101-01-0-16-075 Imputation • 48 1 1 1825.72 4 2 0 20 10-101-0-16-075 Imputation • 9 get a Sample 1 1 1825.72 4 2 0 21 10-101-0-16-075 Imputation · Impect of Malysis 1 1 1 0 2 22 10-101-1-06-075 Impatial and Temporal Modeling · Impatial and Temporal Model 1 0 243.78 4 2 0 23 10-101-106-035 1 1.01 Impatial and Temporal	13	1-01-01-0-16	016	<u>S</u> u	irvival		•	4	7	1		1	1825.72	2	1		0	
15 101-01-0-16-035 Witsping Value Analysis 58 1 1 1825.72 4 2 0 16 1-01-01-0-16-036 Multiple Imputation + 48 1 1 1825.72 2 1 0 17 1-01-01-0-16-040 Complex Samples + 48 1 1 1825.72 4 2 0 18 1-01-01-0-17-096 Quality Contol + Imputation + 48 1 1 1825.72 4 2 0 20 1-01-01-0-17-096 Quality Contol + Imputation + Imputation - Impu	14	1-01-01-0-16	034	Mu	Itiple Response		•	5	3	1		1	1825.72	3	2		0	
16 101-01-0-16-036 Muttiple Imputation 48 1 1 1825.72 2 1 0 17 101-01-0-16-040 Complex Samples 1 1 1825.72 6 3 0 18 101-01-0-16-075 Sigmulation Quality Control 1 1 1825.72 4 2 0 20 101-01-0-18-089 Quality Control Prepare for Analysis 2 1 1917.49 1 1 0 21 101-01-0-18-089 RCC Curge 1 0 2110.44 3 2 0 22 101-01-1-06-035 1 1.01 Battos 1 0 2463.78 2 1 0 24 101-01-1-06-035 1 1.01 Qeneral Linear Model 1 0 2463.78 4 2 0 1 0 2463.78 4 2 0 1 0 2463.78 4 2 0 1 0 2463.78 4 2 0 1 0 2463.78 3 2 0	15	1-01-01-0-16	035	😼 Mis	ssing Value Anal <u>v</u> sis			5	8	1		1	1825.72	4	2		0	
17 101-01-0-16-040 Complex Samples iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	16	1-01-01-0-16	036	Mu	ultiple Imputation		•	4	8	1		1	1825.72	2	1		0	
18 101-01-0-16-075 Image: Simulation Image: Simulation </td <td>17</td> <td>1-01-01-0-16</td> <td>040</td> <td>Co</td> <td>omplex Samples</td> <td></td> <td>•</td> <td>Select a Select a</td> <td>Sample</td> <td>1</td> <td></td> <td>1</td> <td>1825.72</td> <td>6</td> <td>3</td> <td></td> <td>0</td> <td></td>	17	1-01-01-0-16	040	Co	omplex Samples		•	Select a Select a	Sample	1		1	1825.72	6	3		0	
19 101-01-0-17-096 Quality Control Image: Control of the part of the transmission of the part of the part of the transmission of the part	18	1-01-01-0-16	075	🖶 Sir	nulation			Branara (les Applynin	1		1	1825.72	4	2		0	
20 101-01-0-18-089 Image: Compension of the second se	19	1-01-01-0-17	096	Q	ality Control			Prepare 1	or Analysis	2		1	1917.49	1	1		0	
21 101011105075 Image: Constant of the constant o	20	1-01-01-0-18	089	R	C Curve			Erequence	cies	2		1	1917.49	1	1		0	
22 101-01-1-05-106 geada and Temporal modeling 1 0 2110.44 4 2 0 23 101-01-1-06-010 1 1.01 Image: Costabs 1 0 2463.78 2 1 0 24 101-01-1-06-035 1 1.01 Image: Costabs 1 0 2463.78 2 0 0 25 101-01-1-06-036 1 1.01 Image: Costabs 1 0 2463.78 4 2 0 0 26 101-01-1-06-089 1 1.01 Image: Costabs 1 0 2463.78 4 2 0 0 27 101-01-1-06-089 1 1.01 Image: Costabs 1 0 2463.78 3 2 0 0 28 101-01-1-06-089 1 1.01 Image: Costabs 1 0 2463.78 3 2 0 1 0 2 0 1 0 2 1 0 2 1 0 2 1 0 2 1	21	1-01-01-1-05	075		no ourge	deling		bescripti	ves	1		0	2110.44	3	2		0	
23 101011.106.010 1 101 Image: Regression and the second s	22	1-01-01-1-05	106	의	auarand remporarmo	deling	r	🔡 <u>C</u> rosstab	S	1		0	2110.44	4	2		0	
24 101011106:055 1 101 Image: General Linear Model 1 0 2463.78 4 2 0 25 101011106:067 1 101 Image: Linear Model 1 0 2463.78 4 2 0 26 101011:06:069 1 101 Image: Linear Model 1 0 2463.78 1 1 0 27 101011:06:098 1 101 Image: Cog Regression 1 0 2463.78 3 2 0 28 101:01:1:06:098 1 101 Image: Cog Regression 1 0 2463.78 4 2 1 28 101:01:1:06:119 1 1.01 Image: Cog Regression 1 0 2463.78 4 2 1 0 29 1:01:01:1:07:027 1 1.01 71 2 0 5329.06 0 1 0 Image: Cog Regression is ready Im	23	1-01-01-1-06	010		1	1.01		😰 <u>R</u> atios		1		0	2463.78	2	1		0	
25 1-01-01-1-06-067 1 1 0 2463.78 4 2 0 26 1-01-01-1-06-089 1 1.01 Image: Control of the second seco	24	1-01-01-1-06	035		1	1.01		General L	Linear Model	1		0	2463.78	4	2		0	
26 1-01-011-06-089 1 1.01 Qrdinal Regression 1 0 2463.78 1 1 0 27 1-01-01-06-098 1 1.01 Qrdinal Regression 1 0 2463.78 3 2 0 28 1-01-01-06-098 1 1.01 Cog Regression 1 0 2463.78 4 2 1 29 1-01-01-027 1 1.01 71 2 0 5329.06 0 1 0 Vetw Data View IBM SPSS Statistics Processor is ready Unicode 0FF Page for Analysis	25	1-01-01-1-06	067		1	1.01		Logistic F	Regression	1		0	2463.78	4	2		0	
27 1-01-01-1-06-098 1 1.01 O 2463.78 3 2 0 28 1-01-01-06-019 1 1.01 Cog Regression 1 0 2463.78 4 2 1 29 1-01-01-07-027 1 1.01 71 2 0 3329.06 0 1 0 x Data View repare for Analysis IBM SPSS Statistics Processor is ready Unicode:0FF	26	1-01-01-1-06	089		1	1.01			Pegression	1		0	2463.78	1	1		0	
28 1-01-01-1-06-119 1 1 0 2463.78 4 2 1 29 1-01-01-1-07-027 1 1.01 71 2 0 5329.06 0 1 0 Data View repare for Analysis	27	1-01-01-1-06	098		1	1.01			egression	1		0	2463.78	3	2		0	
29 1-01-01-107-027 1 1.01 71 2 0 5329.06 0 1 0 7 Data View Variable View Trepare for Analysis	28	1-01-01-1-06	119		1	1.01		Cox Regr	ression	1		0	2463.78	4	2		1	
Trepare for Analysis	29	1-01-01-1-07	027	_	1	1.01	_	7	1	2		0	5329.06	0	1		0	
Data View Variable View repare for Analysis IBM SPSS Statistics Processor is ready Unicode OFF IBM SPSS Statistics Processor is ready Unicode OFF Image: Compare to the state of	_	•		_			_			***		_						
repare for Analysis IBM SPSS Statistics Processor is ready Unicode :OFF Unicode :OFF Unicode :OFF Unicode :OFF	Data View	/ariable View																
🔳 O 🗂 🦳 🚬 🚓 🍙 💷 🍙 🙆 🕿	Prepare for Ar	Prepare for Analysis IEM SPSS Statistics Processor is ready Unicode: OFF Unicode: OFF																
		repare tor Analysis																



				m 📉 😤 🥅	😂 🔜 🖬 🖬					Visible 70 a	- (70) /
12 : Ponderad	011 5329.06	1.0.07	001							VISIBLE: 79 0	or /9 variable
	IAN_ID	NI ta	ampling Wizard					>	Classe_Escol	AgF7menos	Dicotom
1	1-01-01-0-05-054								2		1
2	1-01-01-0-05-081	W	elcome to the Sampling	Wizard					3		0
3	1-01-01-0-05-088	Th	e Sampling Wizard help	s you design and select a compl	lex sample. Your selection	ns will be :	saved to a plan file that	you can use at	2		0
4	1-01-01-0-05-125	an	alysis time to indicate h	ow the data were sampled.					3		0
5	1-01-01-0-06-008	Yo	u can also use the wiza	rd to modify a sampling plan or d	fraw a sample according t	to an exist	ing plan.		2		0
6	1-01-01-0-06-050			Whet would you like to	4-0				1		0
7	1-01-01-0-06-093			what would you like to	007				3		0
8	1-01-01-0-06-109			Design a sample					2		0
9	1-01-01-0-06-124			Choose this option it	f you have not created a				2		0
10	1-01-01-0-06-136		4	plan file. You will hav	ve the option to draw the	File:	plan	Browse	3		0
11	1-01-01-0-07-065			sample.					1		0
12	1-01-01-0-07-095			C Edite comple design					1		0
13	1-01-01-0-16-016			Cont a sample design	n				1		0
14	1-01-01-0-16-034			Choose this option it	f you want to add, remove,				2		0
15	1-01-01-0-16-035			or modify stages of a	an existing plan. You will	File:		Browse	2		0
16	1-01-01-0-16-036			have the option to dra	aw the sample.				1		0
17	1-01-01-0-16-040		- O H, 1	O Draw the sample					3		0
18	1-01-01-0-16-075		\$ \$ \$ (z) \						2		0
19	1-01-01-0-17-096			Choose this option if	f you already have a plan f	ile File:		Browise	1		0
20	1-01-01-0-18-089			and want to draw a s	ample.			Dro <u>w</u> se	1		0
21	1-01-01-1-05-075								2		0
22	1-01-01-1-05-106								2		0
23	1-01-01-1-06-010								1		0
24	1-01-01-1-06-035								2		0
25	1-01-01-1-06-067								2		0
26	1-01-01-1-06-089			< Back Nex	t > Finish Cancel	Help			1		0
27	1-01-01-1-06-098		1 1.	48	1	U	2463.78		, 2		0
28	1-01-01-1-06-119		1 1.	01 37	1	0	2463.78	4	2		1
29	1-01-01-1-07-027		1 1.	01 71	2	0	5329.06	(1		0
Data View Select a Sam	/ariable View ple		•	•	***		IBI	I SPSS Statistics Pr	ocessor is ready	Unicode:OFF 14:2 11/04/	20 2018

🔒 InsAlim_JANAFcompleta_v2.sav (DataSet1) - IBM SPSS Statistics Data Editor – 🗗 X															
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata <u>T</u> ran	nsform <u>A</u> nal	yze Direct	<u>Marketing</u> ra	ohs <u>U</u>	tilities Add- <u>o</u> r	is <u>W</u> indow <u>H</u> e	p							
2	🖨 🛄 🖬	<u>ר א</u>	E 1	<u>۲</u>	Ϊů.	*, 🖬		1	A 📀 🖣	ABS					
12 : Ponderad	or1 5329.06												Visible: 79 d	of 79 Variable	s
	IAN_ID		NUT	PSU		idade	ClassesIdade		Sexo	Ponderador1	EscolProp	Classe_Escol	AgF7menos	Dicotom	
1	1-01-01-0-05-054		ta Analys	sis Preparation Wiza	rd					4	×	2		1	
2	1-01-01-0-05-081											3		0	Т
3	1-01-01-0-05-088		Stage 1:	Design Variables								2		0	
4	1-01-01-0-05-125		In this pa	nel you con celect	variable	that define atra	la or ductora A com	olo woid	tuoriable mus	the colocted in the	Fratataga	3		0	
5	1-01-01-0-06-008		in uns pa	ner you can select	vallaule:	s ulat deline sua	a of clusters. A sam	ple welg	it variable mus	st be selected in the	ilist stage.	2		0	
6	1-01-01-0-06-050		You can	also provide a labe	I for the	stage that will be	used in the output.					1		0	
7	1-01-01-0-06-093											3		0	
8	1-01-01-0-06-109		V	Velcome		Variables:			<u>S</u> trata:			2		0	
9	1-01-01-0-06-124		. S	tage 1 Design Variable		💑 IAN_ID	-		💦 NUT			2		0	
10	1-01-01-0-06-136			Estimation Meth	od	🛷 idade	de de	•				3		0	
11	1-01-01-0-07-065			Summary		Sexo	Idade					1		0	
12	1-01-01-0-07-095		C	completion		EscolPro	p a					1		0	
13	1-01-01-0-16-016					💑 Classe_	Escol		Clusters:			1		0	
14	1-01-01-0-16-034					AgF7me	nos		🔗 PSU			2		0	
15	1-01-01-0-16-035					Dicotom	icaAgF7menos	•				2		0	
16	1-01-01-0-16-036					A Dicotom	icaAqF7a17					1		0	
17	1-01-01-0-16-040					💰 AgF18a6	4					3		0	
18	1-01-01-0-16-075					뤚 AgF65m	ais		Sample We	ight		2		0	
19	1-01-01-0-17-096					Dicotom	icaAgF65mais	-	Ponder	ador1		1		0	
20	1-01-01-0-18-089					Rend						1		0	
21	1-01-01-1-05-075					& Classes	Rend	_				2		0	
22	1-01-01-1-05-106					💑 Ins01	-	Sta	ge Label:			2		0	
23	1-01-01-1-06-010											1		0	
24	1-01-01-1-06-035					< Bad	Next > Finish	Can	Help			2		0	
25	1-01-01-1-06-067					. 540		Joan				2		0	
26	1-01-01-1-06-089		1	1	.01	46		1	0	2463.7	8 1	1		0	
27	1-01-01-1-06-098		1	1	.01	48		1	0	2463.7	8 3	2		0	
28	1-01-01-1-06-119		1	1	.01	37		1	0	2463.7	8 4	2		1	
29	1-01-01-1-07-027		1	1	.01	71		2	0	5329.0	6 0	1			2
Data View	/ariable View														
Prepare for An	alysis									[BM SPSS Statistics Pro	cessor is ready	Unicode:OFF		
ر م	0 🤤		0 🧿	W	9	R						^ مج	^ d ³ 14:2 11/04/2	2018	

6



IAN ID NUT PSU idade Classes/dade Sexo Ponderador1 EscolProp Classe_Escol AgF7menos Dicotom													
	IAN_ID	NUT	PSU	idade	ClassesIdade	Sexo	Ponderador1	EscolProp	Classe_Escol	AgF7menos	Dicotor		
1	1-01-01-0-05-054	ta Analy	sis Preparation Wizard	1	д.		1	×	2		1		
2	1-01-01-0-05-081	- runny	sis rreputation mizard					~	3		0		
3	1-01-01-0-05-088	Stage 1:	Estimation Method						2		0		
4	1-01-01-0-05-125	In this pa	anal you calect a math	d for optimating stan	dard arrora				3		0		
5	1-01-01-0-06-008	in this pa	anel you select a metri	od for esumating stan	dard errors.				2		0		
6	1-01-01-0-06-050	The estin	mation method depen	ts on assumptions a	bout how the sample v	vas drawn.			1		0		
7	1-01-01-0-06-093								3		0		
8	1-01-01-0-06-109		Velcome						2		0		
9	1-01-01-0-06-124		Stage 1	Which of the f	ollowing sample desig	ins should be assum	ed for estimation?		2		0		
10	1-01-01-0-06-136		 Estimation Method 	WR (san	pling with replaceme	nt)			3		0		
11	1-01-01-0-07-065		Summary	If you ch	oose this ontion you w	ill not be able to add a	additional stages Any	sample stages after	1		0		
12	1-01-01-0-07-095		Completion	the curre	ent stage will be ignore	d when the data are a	analyzed.	Sample Stages alter	1		0		
13	1-01-01-0-16-016			🗹 (Use finite population c	orrection (FPC) when	estimating variance u	nder	1		0		
14	1-01-01-0-16-034			\$	simple random sampli	ing assumption			2		0		
15	1-01-01-0-16-035			© Equal W	OR (equal probability s	ampling without repla	acement)		2		0		
16	1-01-01-0-16-036			The next	t panel will ask you to s	specify inclusion prob	abilities or population	sizes.	1		0		
1/	1-01-01-0-16-040								3		0		
18	1-01-01-0-16-075			O <u>U</u> nequal	WOR (unequal proba	bility sampling without	t replacement)		2		0		
19	1-01-01-0-17-096			Joint pro	babilities will be requi	red to analyze sample	e data. This option is a	available in stage 1	1		0		
20	1-01-01-0-18-089			oniy.					1		0		
21	1-01-01-1-05-075								2		0		
22	1-01-01-1-05-106								2		0		
23	1-01-01-1-06-010								1		0		
24	1-01-01-1-00-035			< <u>B</u> ac	k Next > Finish	Cancel Help			2		0		
26	1 01 01 1 06 089		1 1.04	46	4	0	2462 70	4	2		0		
27	1.01.01.1.06.009		1.01	40	1	0	2403.70	1	1		0		
28	1.01.01.1.06.119		1.01	40	1	0	2403.70	3	2		1		
20	1-01-01-1-07-027		1 1.01	71	2	0	5329.06	4			0		
	1		1.01		2		3323.00	0			·		

<u>F</u> ile <u>E</u> di	t <u>V</u> iew	<u>D</u> ata	Transform	n <u>A</u> nal	yze Direc	ct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> n	s <u>W</u> indow	<u>H</u> elp											
🔁 I) 🛄	I	2			r 1	1 *		- 4		 1 ଲ	6									
12 : Ponde	rador1	5329.	06																	Visible: 79 of	79 Varia	bles
		IAN_ID			NUT	PS	U	ida	ade	ClassesIda	ide	Se	xo	Pondera	dor1	EscolProp		Classe_Escol	,	AgF7menos	Dicotor	m
1	1-01-0	1-0-05-054	L .		ta Analy	vsis Preparatic	on Wizard			a'							×		2	1		4
2	1-01-0	1-0-05-081																	3	()	
3	1-01-0	1-0-05-088	1		Stage 1	: Plan Summ	ary												2	()	
4	1-01-0	1-0-05-125			This see		-	f er Th		is the Osmalal									3	()	
5	1-01-0	1-0-06-008	1		This par	lei summanz	tes trie pla	n so iar. Tr	ie next step	is the Complet	ion par	iei.							2	()	
6	1-01-0	01-0-06-050)																1	()	
7	1-01-0	1-0-06-093	1																3	()	
8	1-01-0	01-0-06-109	•		_ _ ► •	Welcome		Sum	mary:										2	()	
9	1-01-0	1-0-06-124	l.			Stage 1	oriables	Stag	le Label	Strata	С	lusters	Weights	Size	Metho	d			2	()	
10	1-01-0	1-0-06-136	i			 Estimatio 	n Method	1	(None)) NUT	PS	SU	Ponderador	1(n/a)	WR				3	()	
11	1-01-0	1-0-07-065				Summary	/												1	()	
12	1-01-0	1-0-07-095			L	Completion													1	()	
13	1-01-0)1-0-16-016	;																1	()	
14	1-01-0	01-0-16-034	ļ.																2	()	
15	1-01-0)1-0-16-035						File:	plano										2	()	
16	1-01-0	01-0-16-036	;																1	()	
17	1-01-0	1-0-16-040)																3	()	
18	1-01-0	01-0-16-075																	2	()	
19	1-01-0	1-0-17-096	i																1	()	
20	1-01-0	1-0-18-089)																1	()	
21	1-01-0	1-1-05-075	i																2	()	
22	1-01-0	1-1-05-106	i																2	()	
23	1-01-0	1-1-06-010)																1	()	
24	1-01-0	1-1-06-035											<u></u>						2	()	
25	1-01-0	1-1-06-067							< <u>B</u> ack	(<u>N</u> ext >	inisn	Cancel	нер						2	()	
26	1-01-0	1-1-06-089)			1	1.01		46		1		0	1	2463.78		1		1	()	
27	1-01-0	1-1-06-098	1			1	1.01		48		1		0	2	2463.78		3		2	()	
28	1-01-0)1-1-06-119)			1	1.01		37		1		0	:	2463.78		4		2	1		
29	1-01-0	1-1-07-027				1	1.01		71		2		0	(5329.06		0		1	()	Ŧ
	1						_	_	_		***	_	_	_)	
Data Viev	variable	eview													IBM	SPSS Statistics	Proce	esor is ready		Inicode:OFF		
r repare 10	analysis.												_	_	IDM -	or oo otaiislitts	THUCE	Joovi is reauly		14:31		
) (ле		Y	o 🦻		9	R										Ŕ		c ^{1/3} 11/04/20	018	



🔄 InsAlim_IAN	IAFcompleta_v2.sav [E	DataSet1]	- IBM SPS	5 Statistics	Data Editor							_	٥	\times
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata <u>T</u> ra	ansform	Analyze	Direct	Marketing <u>G</u> raphs	s <u>U</u> tilities Add- <u>o</u>	ns <u>W</u> indow <u>H</u> elp							
2	🖨 🛄		ช 🎚	š 🛓		h 🐮 🖬			ABS					
12 : Ponderad	lor1 5329.06											Visible: 7	9 of 79 Varia	ables
	IAN_ID		NU	т	PSU	idade	ClassesIdade	Sexo	Ponderador1	EscolProp	Classe_Escol	AgF7menos	Dicoto	m
1	1-01-01-0-05-054			🔄 Analys	is Preparation Wizard					×	2	2	1	-
2	1-01-01-0-05-081										3	3	0	
3	1-01-01-0-05-088			Completi	ng the Analysis Wiza	rd					2	2	0	
4	1-01-01-0-05-125			You have	provided all of the int	formation peeded to a	reate a plan				3	3	0	
5	1-01-01-0-06-008			rounave	provided all of the fill	ormation needed to c	reate a plan.				2	2	0	
6	1-01-01-0-06-050			You can u	use the plan file in an	v Complex Samples a	analysis procedure whe	en vou are readv to ana	alvze the data.		1	1	0	
7	1-01-01-0-06-093								-		3	3	0	
8	1-01-01-0-06-109			_ ► W	/elcome						2	2	0	
9	1-01-01-0-06-124			s s	tage 1 Design Variables	What	do you want to do?				2	2	0	
10	1-01-01-0-06-136				Estimation Method		o your apositions to				3	3	0	
11	1-01-01-0-07-065			L)	Summary	© <u>S</u> an	ve your specifications to	r a pran me d by the Wizerd inte e	ounter window		1	1	0	
12	1-01-01-0-07-095			- • c	ompletion	<u> </u>		1	1	0				
13	1-01-01-0-16-016								1	1	0			
14	1-01-01-0-16-034								2	2	0			
15	1-01-01-0-16-035										2	2	0	
16	1-01-01-0-16-036										1	1	0	
17	1-01-01-0-16-040										3	\$	0	
18	1-01-01-0-16-075										2	2	0	
19	1-01-01-0-17-096										1	1	0	
20	1-01-01-0-18-089										1	1	0	
21	1-01-01-1-05-075										2	2	0	
22	1-01-01-1-05-106					To de	an Main colored attack Fig				2	2	0	
23	1-01-01-1-06-010					10 00	ise this wizard, click Fir	isn.			1	1	0	
24	1-01-01-1-06-035					e Ba		Cancel Help			2	2	0	
25	1-01-01-1-06-067					- <u>D</u> at		Cancer			2	2	0	
26	1-01-01-1-06-089			1	1.01	1 4	6 1	0	2463.78	1	1	1	0	
27	1-01-01-1-06-098			1	1.01	1 4	B 1	0	2463.78	3	2	2	0	
28	1-01-01-1-06-119			1	1.01	1 3	7 1	0	2463.78	4	2	2	1	
29	1-01-01-1-07-027		_	1	1.01	1 7	1 2	0	5329.06	0	1	1	0	T
	1						***							
Data View	/ariable View													
Prepare for A	nalysis				_				IBI	M SPSS Statistics Proc	essor is ready	Unicode:OFF		
, C	· 🗇 🧲		10	9	🗾 😂 🤇						٨	^ d ³ 11/0	4:31	

This file will be used to perform all weighted statistical analyses, which must be uniquely made in the **Analyze**

>> (Comp	lex	Samp	es	menu.
------	------	-----	------	----	-------

taseAndrei	a03072018.sav [DataSet1] - IBM S	PSS Statistics Data Editor											- 0	×
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata <u>T</u> ransform	Analyze Graphs Utilities	Extensions	<u>W</u> indow <u>H</u> elp										
👝 🔳		Reports	•											
		Descriptive Statistics	÷											
		Bayesian Statistics	•									Vis	ble: 76 of 76 V	/ariables
	🛷 V1 🛛 🚜 IAN	Ta <u>b</u> les	*	占 X1.5 🛛 💑 X1.6	💑 X10.1	💑 X10.3	💑 X11.1	💑 X12.1	💑 X2.1	💑 X3.1	💑 ХЗ.З	💑 X3.4	💑 X4.1	<u> </u>
1	233 1-08-01-0-05-08	Co <u>m</u> pare Means	•	1 1	1	1	1	2	2	3	3	1	2	
2	234 1-08-01-0-05-10	<u>G</u> eneral Linear Model	•	1 1	1	1	1	1	3	2	2	1	1	
3	235 1-08-01-0-06-00	Generalized Linear Models	; • • •	2 1	1	1	3	1	2	2	2	1	1	
4	236 1-08-01-0-06-03	Mixed Models	•	1 1	3	1	1	1	1	2	1	1	1	
5	237 1-08-01-0-06-03	<u>C</u> orrelate		1 1	1	1	1	1	1	2	2	1	1	
6	238 1-08-01-0-06-05	Regression	•	1 2	1	1	1	1	3	3	2	1	1	
	239 1-08-01-0-06-07	Loglinear	*	2 2	1	1	1	1	3	2	2	2	2	
8	240 1-08-01-0-06-07	Neural Net <u>w</u> orks	•	1 1	1	1	1	2	2	1	1	1	2	
9	241 1-00-01-0-06-06	Classify	*	1 1	1	1	1	1	2	2	2	1	1	
11	242 1-00-01-0-06-10	Dimension Reduction	•	1 3	1	1			2	2	3	1	1	
10	243 1-00-01-0-06-13	Scale		2 4	1	1			3	1	3	1		
12	244 1-00-01-0-07-02	Nonparametric Tests	•	3 1	1	1	4	1		4	J 1	1	1	_
14	245 1-06-01-0-07-04	Forecasting		1 2	1	1		1	2	2	1	1	2	
14	247 1-08-01-0-15-00	Survival		1 2	1	1		2	2	2	2	1	1	
16	248 1-08-01-0-16-0/	Multiple Response	*	1 3	2	1	1	1		2	2	1	1	
17	249 1-08-01-0-16-04	🖾 Missing Value Analysis		1 1		1	1	2	2	2	2	1	2	
18	250 1-08-01-0-16-04	Multiple Imputation		1 2	3	1	1	1	2	2	- 3	. 1	- 1	
19	251 1-08-01-0-16-07	Complex Samples	•	Delegt a Comple	1	2	1	1	2	- 3	2	1	1	
20	252 1-08-01-0-16-0			gelect a sample	1	- 1	1	1	2	2	- 3	1	1	
21	253 1-08-01-0-16-11	Quality Control		Prepare for Analysis.	. 3	1	1	2	3	2	2	2	1	
22	254 1-08-01-0-16-13		,	Erequencies	3	1	1	1	3	2	2	1	1	
23	255 1-08-01-0-17-05		X	Descriptives	2	1	1	1	1	2	2	1	1	
24	256 1-08-01-1-05-01	Spatial and Temporal Mod	eling P	Crosstabs	1	1	1	1	2	2	2	1	1	
25	257 1-08-01-1-05-06	Direct Marketing	-	12 Ratios	1	1	1	1	2	2	2	1	1	
26	258 1-08-01-1-05-09	3 3	2	General Linear Mode	L 1	1	1	1	3	2	2	1	1	
27	259 1-08-01-1-05-11	16 2	2	Logistic Regression	1	1	1	1	1	2	2	1	1	
28	260 1-08-01-1-05-12	29 1	2	Crdinal Regression	. 1	1	1	1	2	2	2	1	2	
29	261 1-08-01-1-05-13	32 3	2	Orumar Regression	1	1	3	1	3	1	2	1	3	-
	4			Cox Regression										
Data View	/ariable View													
Frequencies									IBM SPSS	Statistics Proce	essor is ready	Unicod	e:ON	
م 🖿	Hi 🧲 🧮	D 🕑 🚾 🧿	Ŷ	<u>Ps</u> 🔹 😒							Ŕ	へ d>)) POR	13-07-2018	



1.1. Weighted frequencies

In order to estimate weighted frequencies, one should go to **Analyze** >> **Complex Samples** >> **Frequencies** and select the previously created file.

taseAndrei	a03072018.	sav [Data	aSet1] - IBM S	PSS Statistic	s Data Editor													- 0	×	
<u>F</u> ile <u>E</u> dit	<u>V</u> iew E	ata]	<u>T</u> ransform	Analyze	<u>G</u> raphs <u>U</u>	tilities Eg	tensions	Window	<u>H</u> elp											
e		nn-		Re <u>p</u> ort	s		•		A											
				D <u>e</u> scri	ptive Statistic	s	•		⊞ 14											_
				<u>B</u> ayesi	an Statistics		•										Vis	sible: 76 of 76	Variables	5
	🔷 🔗 V1		🚜 IAN	Ta <u>b</u> les			•	6 X1.5	💑 X1.6	💰 X10.1	💑 X10.3	💑 X11.1	💑 X12.1	💑 X2.1	💑 X3.1	💑 ХЗ.З	💑 ХЗ.4	💑 X4.1		
1		233 1-0	8-01-0-05-08	Compa	are Means		•	1		1	1 .	1	2	2	3	3	1	2	: 1	
2		234 1-0	8-01-0-05-10	<u>G</u> enera	al Linear Mod	el	•	1		1	1	1 1	1 1	3	2	2	1	1	1	
3		235 1-0	8-01-0-06-00	Genera	alized Linear	Models		2		1	1 .	3	8 1	2	2	2	1	1	1	l
4		236 1-0	8-01-0-06-03	Mixed I	Nodels		•	1		1	3 .	1	1 1	1	2	1	1	1	-	l
5		237 1-0	8-01-0-06-03	<u>C</u> orrela	ate			1		1	1 .	1	1 1	1	2	2	1	1		l
6		238 1-0	8-01-0-06-05	Regres	sion		•	1		2	1 .	1	1 1	3	3	2	1	1		I
7		239 1-0	8-01-0-06-07	Logline	ear			2		2	1 .	1	1 1	3	2	2	2	2	2	l
8		240 1-0	8-01-0-06-07	Neural	Networks		•	1		1	1 .	1	1 2	2	1	1	1	2	2	l
9		241 1-0	8-01-0-06-08	Classi	tv —			1		1	1	1	1 1	2	2	2	1	1	()	l
10		242 1-0	8-01-0-06-10	Dimen	sion Reducti	on		1		3	1 .	2	2 1	3	2	3	1	1	(l
11		243 1-0	8-01-0-06-13	Scale				1		3	1 .	1	1 1	3	1	3	1	1	·	l
12		244 1-0	8-01-0-07-02	Nonna	rametric Tes'	19		3		1	1 .	2	2 1	2	2	3	1	1	· · · · ·	l
13		245 1-0	8-01-0-07-04	Eoreca	eting			1		1	1 .	1	1 1	1	1	1	1	1	(l
14		246 1-0	8-01-0-15-00	Cupius	aging		, i	1		2	1 .	1	1 1	2	2	1	1	3	;	L
15		247 1-0	8-01-0-15-01	Surviva			, r	1		3	1	1	2	2	2	2	1	1	0	ľ
16		248 1-0	8-01-0-16-04	Mulupi	Response			1		3	2 .	1	1 1	1	3	2	1	1	(I
17		249 1-0	8-01-0-16-04	Missing	Value Analy	SIS		1		1	1 .	1	1 2	2	2	2	1	2	2	l
18		250 1-0	8-01-0-16-06	Mulţipl	a Imputation		•	1		2	3	1	1 1	2	2	3	1	1	(I
19		251 1-0	8-01-0-16-07	Compl	ex Samples		- P.	😑 Select :	a Sample		1 2	2 1	1 1	2	3	2	1	1	(l
20		252 1-0	8-01-0-16-09	🖶 Sįmula	tion			Prepare	e for Analysis	s	1 .	1	1 1	2	2	3	1	1	(l
21		253 1-0	8-01-0-16-11	<u>Q</u> uality	Control		•	123 Freque	ncies		3	1	2	3	2	2	2	1	(l
22		254 1-0	8-01-0-16-13	ROC C	ur <u>v</u> e			Descrip	tivee		3	1	1 1	3	2	2	1	1	(l
23		255 1-0	8-01-0-17-05	Spatial	and Tempor	al Modeling		Descrip			2 .	1	1 1	1	2	2	1	1	(l
24		256 1-0	8-01-1-05-01	Direct I	Mar <u>k</u> eting		•	Crosst	abs		1 .	1	1 1	2	2	2	1	1	(l
25		257 1-0	8-01-1-05-06			-	-	12 Ratios.			1 .	1	1 1	2	2	2	1	1	(l
26		258 1-0	8-01-1-05-09	13		3	2	🔛 <u>G</u> enera	I Linear Mod	lel	1 .	1	1 1	3	2	2	1	1	(l
27		259 1-0	8-01-1-05-11	6		2	2	🔣 Logisti	Regression	n	1 .	1	1 1	1	2	2	1	1	(l
28		260 1-0	8-01-1-05-12	9		1	2	🔣 Ordinal	Regression	L	1 .	1	1 1	2	2	2	1	2	2	l
29		261 1-0	8-01-1-05-13	2		3	2	Cox Re	aression		1	3	3 1	3	1	2	1	3	; -	4
	4	_																		1
Data View V	ariable Vie	w																		
Frequencies		_												IBM SPSS	Statistics Proc	essor is readv	Unico	de:ON		
in requancies			_	_										1.541 01 00	010101001100	occorris ready	Onico	11:35		l
	i	e						R	2.9							a ^r	^ (⊅)) PO	12 07 2010	5	



1	1	I	1	1
ta Com	plex Samples Pla	an for Frequencies	Analysis	×
Plan -				
<u>F</u> ile	HE-FMUP\An	dreiaPadroesIAN	\plan.csaplan	Browse
lf vou	do not have a p	lan file for vour co	mplex sample.	vou can
use th	ne Analysis Prep	paration Wizard to	create one. Ch	oose
Prepa	re for Analysis f	from the Complex	Samples men	u to access
the wi	zard.			
Joint P	robabilities —			
Joint	probabilities are	e required if the p	lan requests un	equal
proba	only work estir	nation. OtherWise	e, uley are ignor	eu.
0 Use	erault file (F	H:\MILTON\Docun	nents\SHE-FM	(plan.sav)
© An <u>i</u>	open dataset			
	baseAndreia03	3072018.sav [Da	taSet1]	
0.0				
© <u>C</u> us	storn me			
File				Browse
	_			
	<u>C</u> ontir	nue Cancel	Help	
	1		1	1

Next, one should select the variable under study and the associated statistics.





	Complex Samples Plan for Frequencies Analysis	×
E	Vai 🔚 Complex Samples Frequencies: Statistics	×
	Cells Cells Population size	
з	Statistics	
	- 🖉 🗹 Standard error 📃 Unweighted count	
	Confidence interval	
	Level(%): 95 📃 Sguare root of design effe	ect
	Coefficient of variation Cumulative values	
	Test of equal cell proportions	
-	Continue Cancel Help	
	OK Paste Reset Cancel Help	

Result:

Sexo.x							
		Estimate	Standard Error	95% Confidence Interval			
			Stanuaru LITU	Lower	Upper		
	0	4739432,770	145329,479	4450795,879	5028069,661		
Population Size	1	4449227,520	126039,458	4198902,276	4699552,764		
	Total	9188660,290	239273,706	8713442,056	9663878,524		
% of Total _	0	51,6%	0,7%	50,2%	53,0%		
	1	48,4%	0,7%	47,0%	49,8%		
	Total	100,0%	0,0%	100,0%	100,0%		



1.2. Test independence/association between 2 categorical variables

In order to test the independence/association between two categorical variables, one should access to the **Analyze >> Complex Samples >> Crosstabs** menu and select the previously created file.

Then, select the variables under hypothesis and the respective statistics.





5		Ζ 1 1 1	3	1
1	Comple	x Samples Crosstabs	×	1
1	Variables:	🔄 Complex Samples Crosstabs: Statistics 🛛 🕹 🗙		1
1	valiables.	- Calla	stics	1
2		Cells	Values	1
1	idade	Population size V Column percent	ons	2
2	💑 GE4	Kow percent I able percent		1
3	💑 class	C Statistics		1
3	Escol	Standard error 🔄 Unweighted count		1
2	ESCOI	Confidence interval		1
1	Saud	Level{%}: 95 Square root of design effect		1
3	MC 🔊	Coefficient of variation Residuals		1
2	🛷 zIMC	Expected values Adjusted residuals		2
2	Class			1
2		Summaries for 2-by-2 Tables		2
3	PROT	☑ Odds ratio		1
2	🖋 FATg	Relative ris <u>k</u>		1
2	CARE	Tast of independence of rows and columns		1
3	ALCO			2
1	ALCO	Continue Cancel Help		1
2	l	OK Paste Reset Cancel Help		1
2		1 1 1	1	1
2		1 1 1 1	1	1

Result:

Sexo.x * Desp

Sovo v			Desp				
	Sex0.X		0	1	Total		
		Estimate	2916200,750	1689662,870	4605863,620		
_	Population Size	Standard Error	119981,932	104059,923	143375,307		
0	0/ithin Course	Estimate	63,3%	36,7%	100,0%		
0	% Within Sexo.x	Standard Error	1,9%	1,9%	0,0%		
		Estimate	53,4%	47,1%	50,9%		
	% within Desp	Standard Error	1,3%	1,7%	0,7%		
		Estimate	2547897,160	1899139,430	4447036,590		
	Population Size	Standard Error	109990,959	108317,206	126295,420		
		Estimate	57,3%	42,7%	100,0%		
1	% Within Sexo.x	Standard Error	2,0%	2,0%	0,0%		
		Estimate	46,6%	52,9%	49 ,1%		
	% Within Desp	Standard Error	1,3%	1,7%	0,7%		
	Desculation Cine	Estimate	5464097,910	3588802,300	9052900,210		
_	Population Size	Standard Error	183758,461	173125,807	234706,467		
Total	0/	Estimate	60,4%	39,6%	100,0%		
	% Within Sexo.x	Standard Error	1,5%	1,5%	0,0%		
	0/	Estimate	100,0%	100,0%	100,0%		
	% within Desp	Standard Error	0,0%	0,0%	0,0%		

Tests of Independence								
Chi-Square Adjusted F df1 df2 Sig.								
Sexo.x * Desp –	Pearson	14,388	6,020	1	92	,016		
	Likelihood Ratio	14,394	6,022	1	92	,016		

The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

Measures of Association				
		Estimate		
Sexo.x * Desp	Odds Ratio	1,286		

Statistics are computed only for 2-by-2 tables with all cells observed.



1.3. Weighted mean

In order to estimate the weighted mean and the respective confidence interval of a continuous variable, one should access to the **Analyze >> Complex Samples >> Descriptives** menu and select the previously created file.

Then, select the continuous variable under study and the respective statistics.

ti Complex Samples Descriptives		×	
Variables: ✓ V1 ♣ IAN_ID ♣ X1.3 ♣ X1.5 ♣ X1.6 ♣ X10.1 ♣ X10.3 ♣ X1.1 ♣ X12.1 ♣ X3.4 ♣ X4.1 ♣ X4.3 ♥ K4.3	Measures: X1.1 Miss Subpopulations: Each combination of categories defines a subpopulation. Reset Cancel Help	tatistics ing Values pptions Summaries Mean t-test Test value Statistics Statistics Statistics Confidence interval	tatistics × Sum t-test Test value
elynica Percent 51,6%		Level (%): 95	Design effect Square root of design effect
48,4% 100,0%		Continue Car	ncel Help

Result:

Univariate Statistics						
Estimate		Estimato	Standard Error	95% Confidence Interval		
			Lower	Upper		
Mean	X1.1	2,14	,027	2,09	2,19	



1.4. Linear Regression

In order to compare weighted mean values or a linear regression for weighted data, one should access to the **Analyze >> Complex Samples >> General Linear Model** menu and select the previously created file.

Then, select the dependent variable and the independent variables, as well as the respective statistics. If a variable is of type categorical, then the variable must be added in "Factors". Otherwise, if a variable is of type continuous, then the variable must be added in "Covariates".



Resultado:

Parameter Estimates^a

Parameter I	Estimato	95% Confidence Interval		Hypothesis Test		
	Estimate	Lower	Upper	t	df	Sig.
(Intercept)	2,129	2,056	2,203	57,592	92,000	,000
[Sexo.x=0]	,020	-,068	,108	,456	92,000	,649
[Sexo.x=1]	,000 ^b	•	•		•	

a. Model: X1.1 = (Intercept) + Sexo.x

b. Set to zero because this parameter is redundant.

2. Software R



In order to obtain weighted estimates according to the IAN-AF 2015-2016 complex sampling design in R, the library "survey" is used [2,3].

```
> install.packages("survey")
```

> library(survey)

When creating the database, it is mandatory to include the variables "PSU", "NUT" and the respective weighting variable, which can be found in the sociodemographic database. Thus, it is always necessary to join the sociodemographic database with the variables under study.

```
# mudar nome das tabelas de acordo com os nomes dos ficheiros exportados
# mudar variável ponderador de acordo com as variáveis a analisar
> base = read.csv2("Tabela_Ponderador_Sociodem.csv", stringsAsFactors = F)
> atvfis = read.csv2("Tabela_AFisica.csv", stringsAsFactors = F)
> b = merge(base, atvfis)
> svdx<-svydesign(id = ~PSU, strata = ~NUT, weights = ~Ponderador1, data = b)
> summary(svdx)
```

Next, some statistical analysis using the indicated library are exemplified. More information about the implemented functions in this library is available in the respective documentation.



2.1. Weighted frequency and mean values of categorical and continuous variables, respectively

The "svymean" function calculates the weighted mean of a variable according to the complex sampling design previously established. If the variable under study is of type "factor", then this function calculates the weighted proportion of each category of the variable.

```
> svymean(~idade, svdx)
    mean SE
idade 42.686 0.3652
> svymean(~factor(Sexo), svdx)
    mean SE
factor(Sexo)0 0.51217 0.0064
factor(Sexo)1 0.48783 0.0064
```

2.2. Statistics on subsets

In order to estimate statistics on subsets defined by a factor, use the "svyby" function.

```
> svyby(~idade, ~Sexo, svdx, svymean)
    Sexo idade se
0 0 42.22272 0.4738476
1 1 42.11595 0.4994525
```

It is also possible to define separately a subset, and proceed as usual.



2.3. Hypothesis tests

t-test for comparison of mean values:

 χ^2 -test to comparison of proportions:

> svychisq(~GE4+Sexo, svdx)

Pearson's X^2: Rao & Scott adjustment

data: svychisq(~GE4 + Sexo, svdx)
F = 4.4883, ndf = 1.9053, ddf = 175.2800, p-value = 0.01385



2.4. Regression models

Linear regression model:

```
> m1=svyglm(IMC ~ Sexo + Idade + factor(EscolClass_Prop) , family=gaussian(), svdx)
> summary(m1)
Call:
svyglm(formula = IMC ~ Sexo + Idade + factor(EscolClass_Prop),
   family = gaussian(), subsvdx)
Survey design:
svdx
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
                        24.445613 0.472124 51.778 < 2e-16 ***
(Intercept)
                                    0.241667 -1.376
Sexo
                        -0.332601
                                                        0.172
Idade
                         0.084928 0.007141 11.894 < 2e-16 ***
                                    0.272237 -5.142 1.63e-06 ***
factor(EscolClass_Prop)2 -1.399916
factor(Escolclass_Prop)3 -2.057181 0.269839 -7.624 2.70e-11 ***
___
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for gaussian family taken to be 20.84462)
Number of Fisher Scoring iterations: 2
> cbind(coef(m1),confint(m1))
                                          2.5 %
                                                     97.5 %
                        24.44561278 23.52026639 25.37095917
(Intercept)
                        -0.33260125 -0.80626059 0.14105808
Sexo
                         0.08492765 0.07093221 0.09892308
Idade
factor(EscolClass_Prop)2 -1.39991563 -1.93349039 -0.86634087
factor(EscolClass_Prop)3 -2.05718129 -2.58605546 -1.52830711
```



Logistic regression model:

```
> m1=svyglm(factor(Desp) ~ factor(GrupoEtario), family=binomial(link = 'logit'), svdx)
> summary(m1)
Call:
svyglm(formula = factor(Desp) ~ factor(GrupoEtario), family = binomial(link = "logit"),
    subsvdx)
Survey design:
svdx
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)
                     0.44697
                                0.14980
                                         2.984
                                                 0.00367 **
factor(GrupoEtario)2 -0.08235
                                0.18099 -0.455 0.65023
                                0.15511 -5.407 5.32e-07 ***
factor(GrupoEtario)3 -0.83873
factor(GrupoEtario)4 -1.15278 0.18788 -6.136 2.30e-08 ***
___
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1.000187)
Number of Fisher Scoring iterations: 4
> cbind(exp(coef(m1)),exp(confint(m1)))
                                  2.5 %
                                           97.5 %
                    1.5601185 1.1636513 2.0916658
(Intercept)
factor(GrupoEtario)2 0.9240598 0.6467305 1.3203127
factor(GrupoEtario)3 0.4309102 0.3187190 0.5825935
factor(GrupoEtario)4 0.3164551 0.2187010 0.4579029
```

