

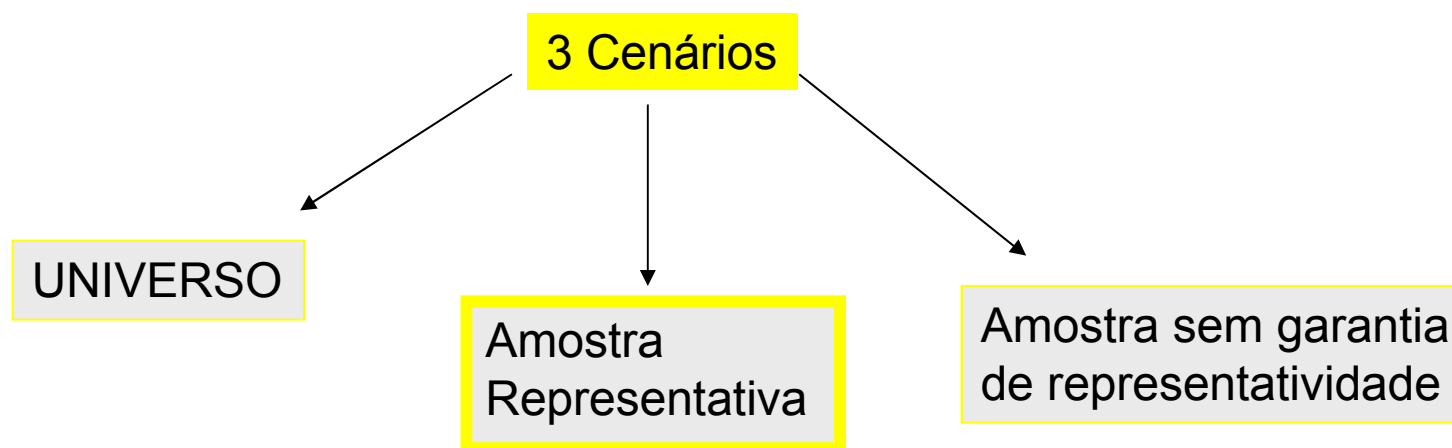
Programa de Doutoramento em Educação

Amostragem

**Seminário de Investigação
Métodos e Técnicas
de Recolha e Tratamento de Dados**

Universo e amostra

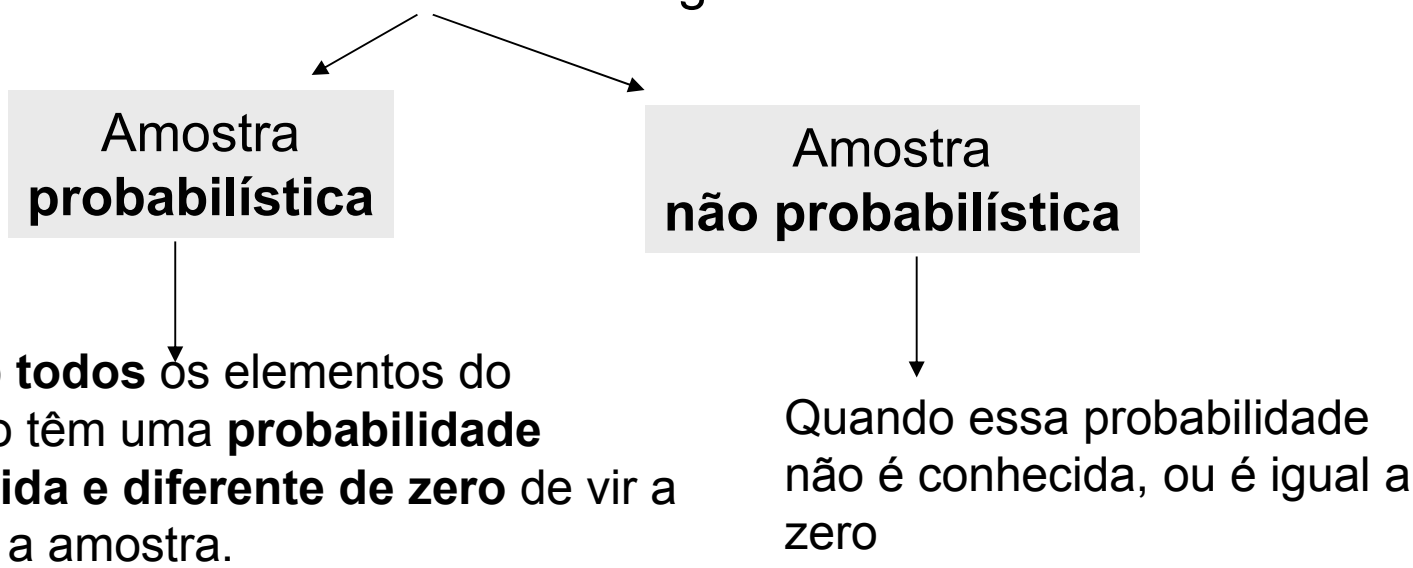
Quem inquirir / observar?



Elementos a considerar na construção de uma amostra

1. DIMENSÃO da amostra (ver mais à frente)

2. MÉTODO de amostragem



Requisito: listagem com identificação de todos os elementos do universo

Elementos a considerar na construção de uma amostra

Vantagens das amostras probabilísticas:

- Permitem garantir a representatividade da amostra
- Permitem medir o erro associado à inferência

Principais tipos de amostras probabilísticas:

1. Amostra aleatória simples

Cada elemento da população tem **igual probabilidade** de ser selecionado

Forma de seleção: tabelas de números aleatórios; recurso a software que permita a seleção aleatória de casos (ex: SPSS/PASW)

2. Amostra estratificada

Quando se conhecem as características da população relacionadas com as variáveis que se pretendem estudar.

Procedimento:

- 1) Definição de estratos na população;
- 2) Seleção de uma amostra aleatória simples em cada estrato.

- Relativamente ao universo, os estratos podem ser:

Não proporcionais

ou

Proporcionais

Amostra estratificada proporcional (exemplo)

Sexo	Valores absolutos população (N)			Peso de cada estrato na população (Nij/N)			Distribuição da Amostra (n=700) nij=n*Nij/N		
	M	F	Total	M	F	Total	M	F	Total
Grau									
S/ nível de instrução	608	656	1264	0,09	0,09	0,18	61	66	126
Básico 1	1308	1093	2401	0,19	0,16	0,34	131	109	240
Básico 2	601	595	1196	0,09	0,08	0,17	60	59	120
Básico 3	410	599	1009	0,06	0,09	0,14	41	60	101
Secundário	304	337	641	0,04	0,05	0,09	30	34	64
Superior	188	305	493	0,03	0,04	0,07	19	30	49
Total	3419	3585	7004	0,49	0,51	1,00	342	358	700

Amostra não proporcional (exemplo: Estudo Nacional de Literacia)

	População		Amostra	
	N	%	n	%
Total	6 340 844	100,0	2449	100,0
Sexo				
Masculino	3 096 665	48,8	1195	48,8
Feminino	3 244 179	51,2	1254	51,2
Idades				
15-24 anos	1 548 678	24,4	491	20,0
25-39 anos	2 024 960	31,9	737	30,1
40-54 anos	1 724 248	27,2	687	28,1
55-64 anos	1 042 958	16,5	534	21,8
Nível de Instrução				
S/graú	1 126 236	17,8	449	18,3
Básico 1	2 427 400	38,3	747	30,5
Básico 2	1 118 847	17,6	406	16,6
Básico 3	761 269	12,0	408	16,7
Secundário	561 459	8,9	230	9,4
Superior	345 633	5,5	209	8,5
Actividade				
Activos	3 923 000	61,9	1460	59,6
Não Activos	2 417 844	38,1	989	40,4

Amostras não probabilísticas

Conveniência
Intencional
Snowball

Amostra por quotas
(principal método nas
ciências sociais)

Proporcional

Não proporcional

Embora não existam listagens exaustivas de todos os elementos da população...

- a) a informação disponível permite dividir a população em grupos
- b) e estabelecer o número de pessoas a inquirir em cada um desses grupos ou segmentos
- c) Porém, a seleção dos elementos a inquirir não é aleatória...

Outros métodos de amostragem

Amostragem multi-etapas

Método fundamental nas ciências sociais

Pode ser:

- Probabilística
- “Quase” **Probabilística** (ex: *random route*)
- **Semi-probabilística**
- **Não probabilística**

Itinerário aleatório (exemplificação das etapas do método)

- 1) Seleção aleatória de freguesias (ex: da cidade de Lisboa);
- 2) Seleção aleatória de pontos de partida (com base em mapas);
- 3) A partir de cada ponto de partida, determinação de um caminho quase aleatório, com base num esquema predefinido (esquerda / direita; regras para rotundas, becos sem saída, etc.);
- 4) Os questionários não se realizam na primeira rua mas sim nas seguintes. Nota: os questionários são distribuídos aos entrevistadores organizados em séries de 10 e numerados de 1 a 10. Em cada capa de questionário constam ainda dois números aleatórios (ex: 1 e 5);
- 5) Seleção dos edifícios. São escolhidos todos os edifícios cujos números da porta tenham a terminação dos números aleatórios impressos na capa do questionário (ex: 1, 5, 15, 21, 25);
- 6) Seleção do piso: atribuição de um número de ordem a cada piso com residências. Seleção do piso cruzando o número de ordem do questionário e o número de pisos do edifício (tabela); o número de pisos a seleccionar depende do número de andares do prédio. Em cada piso seleccionado deverá ser realizado um questionário;
- 7) Seleção do alojamento: atribuição de um número de ordem a cada alojamento; seleção aleatória a partir de uma tabela que cruza o número de ordem do questionário com o número de alojamentos no piso;
- 8) Seleção do indivíduo a inquirir. Ordenação de todos os indivíduos residentes no alojamento (em condições de ser inquiridos) por ordem decrescente de idades. Consulta de uma tabela que cruza o número de ordem do questionário com o número de elementos do agregado familiar e permite determinar qual o número de ordem do indivíduo a inquirir.

Amostra não proporcional (exemplo: Públicos da Ciência em Portugal)

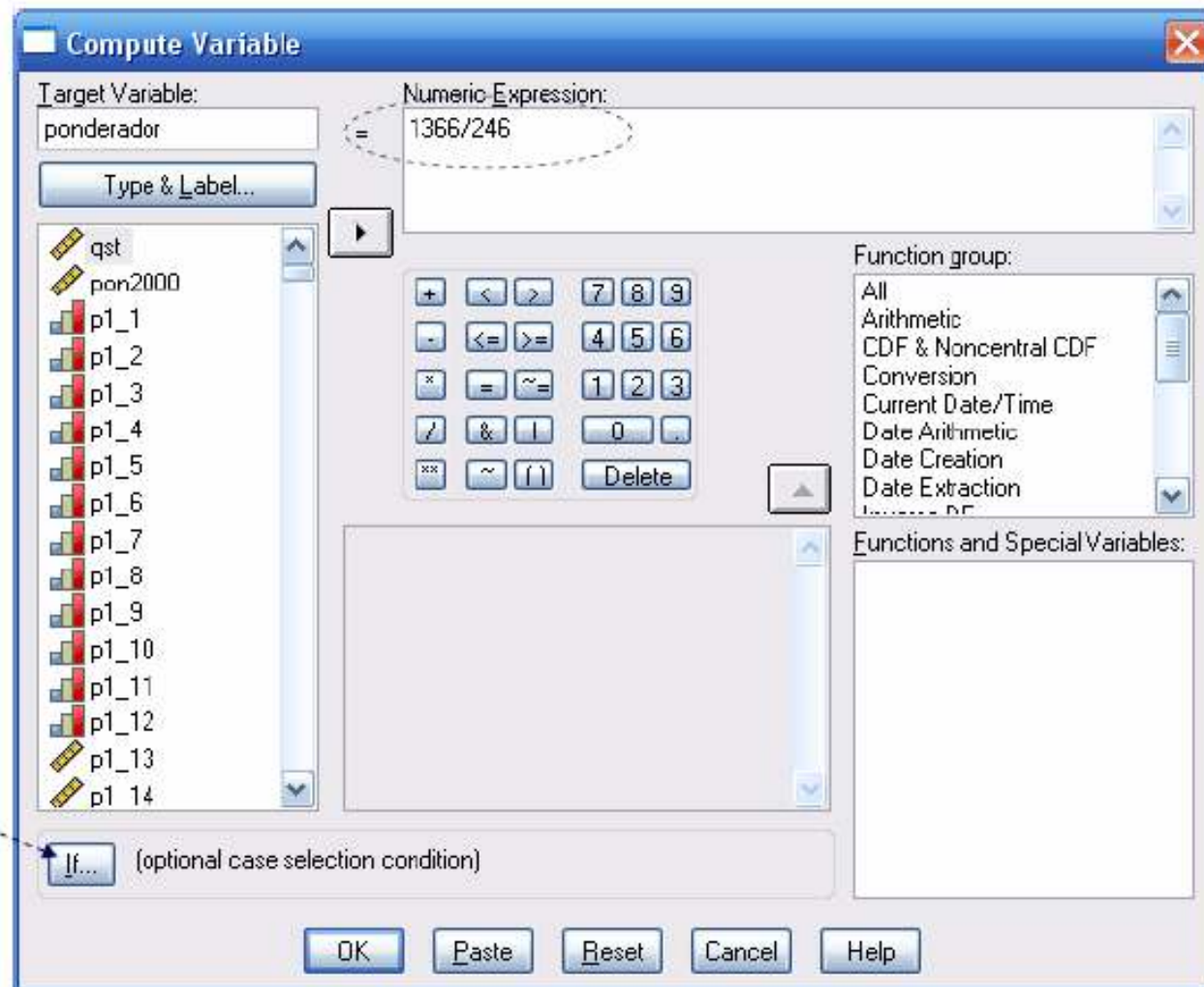
	População		Amostra	
	N (10 ³)	%	n	%
Total	7322,1	100	2057	100
Grau de escolaridade				
Até Básico 2	4863,4	66,4	246	12,0
Básico 3	1116,5	15,3	268	13,0
Secundário	852,1	11,6	543	26,4
Superior	490,1	6,7	1000	48,6

Cálculo de um ponderador (exemplo: Públicos da Ciência em Portugal)

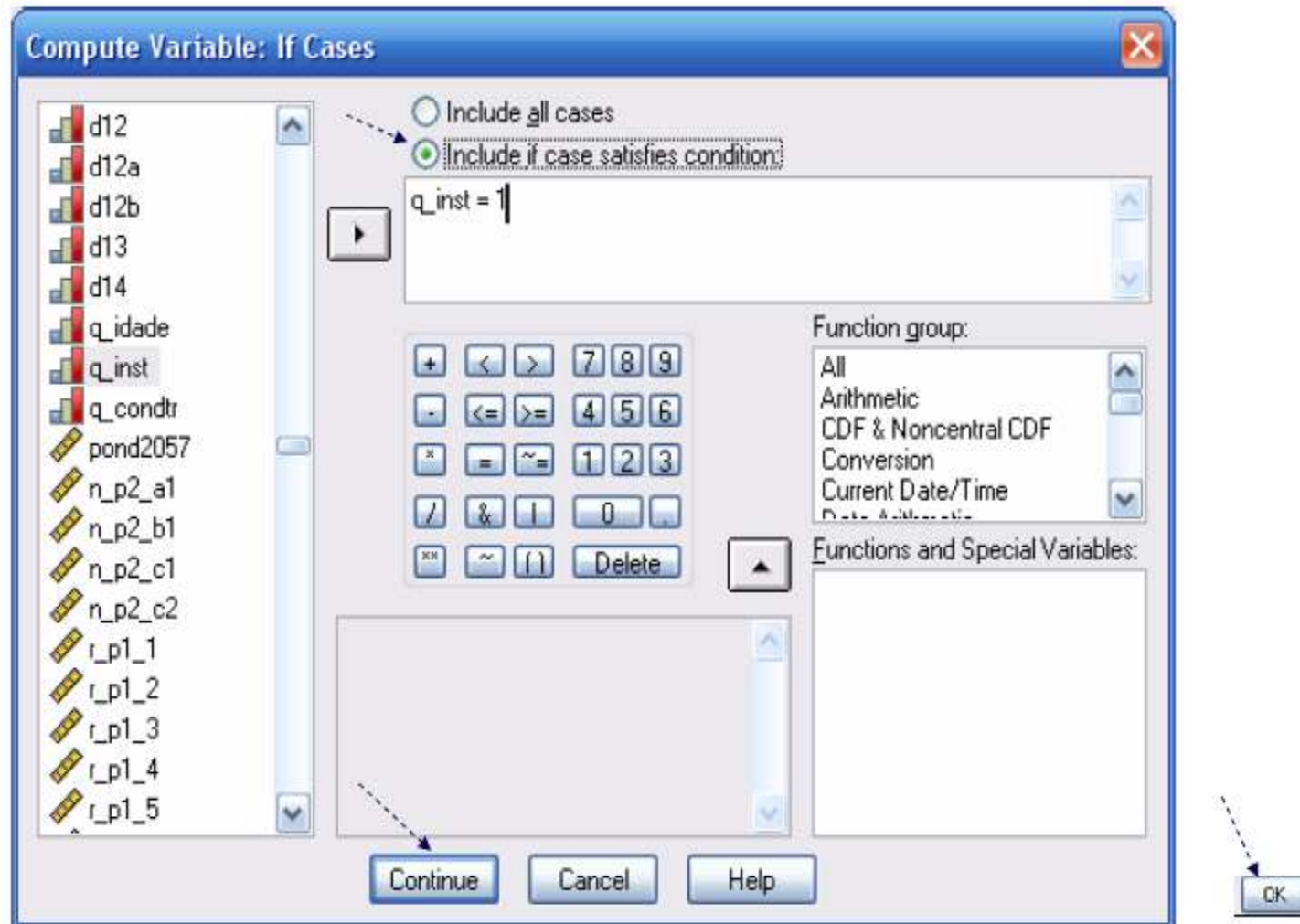
Whi = nº de observações no estrato teórico / nº de observações no estrato observado

	1) "n" Teórico (2057 * peso do estrato na população)	2) "n" Observado	3) Ponderador (whi)
Até básico 2	1366	246	5,553894
Básico 3	314	268	1,170352
Secundário	239	543	0,440842
Superior	138	1000	0,137682

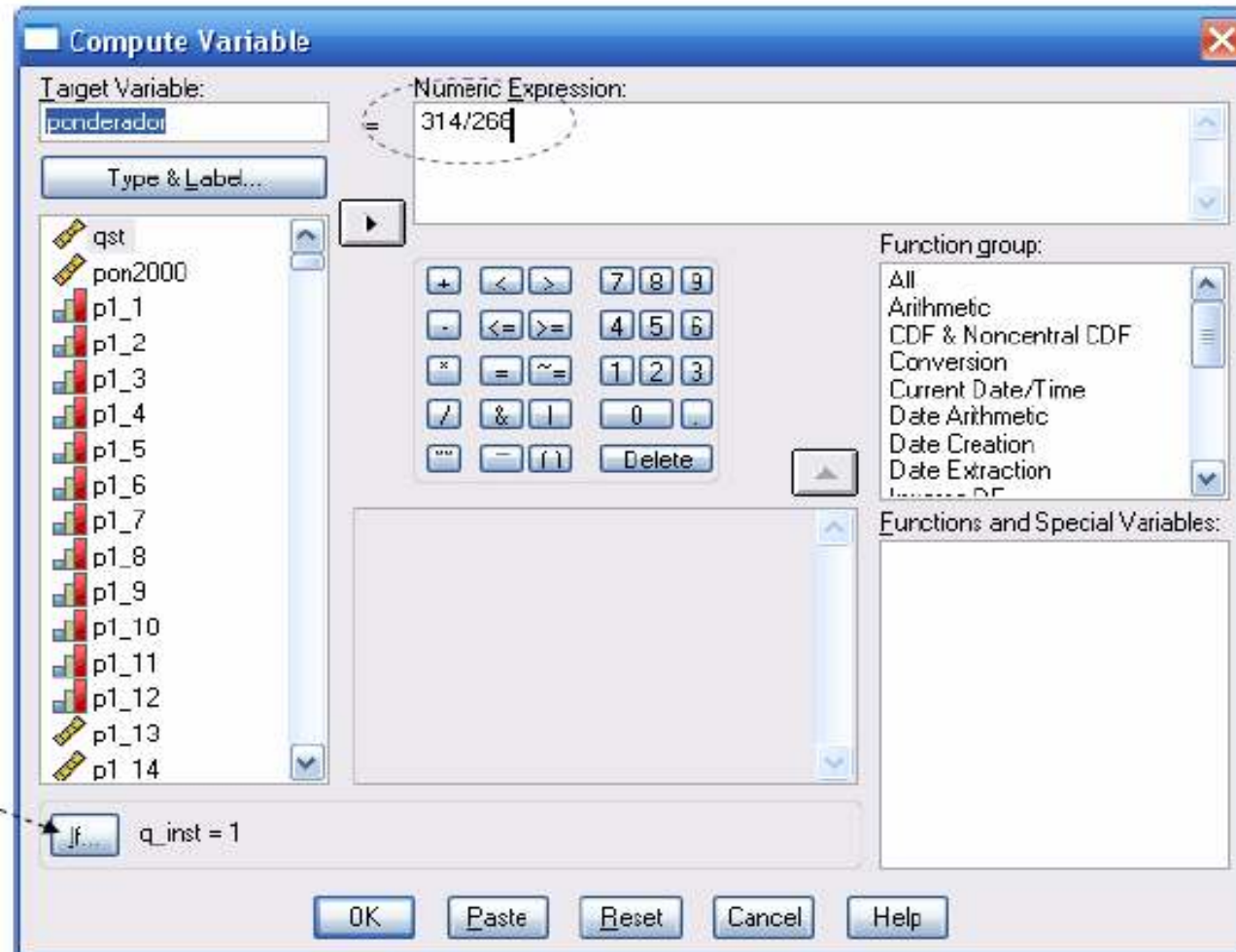
Construção de um ponderador no SPSS/PASW (1)



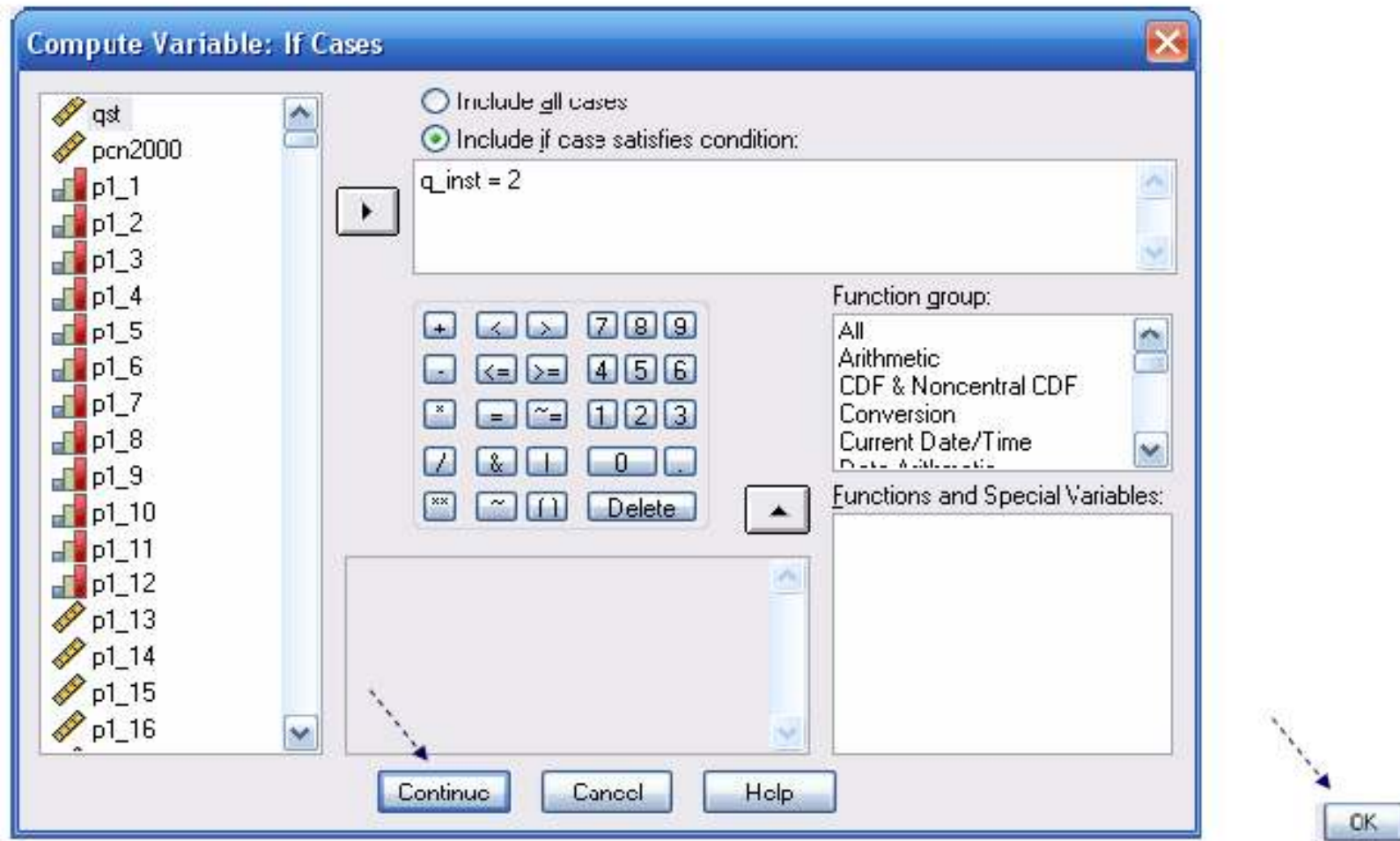
Construção de um ponderador no SPSS/PASW (2)



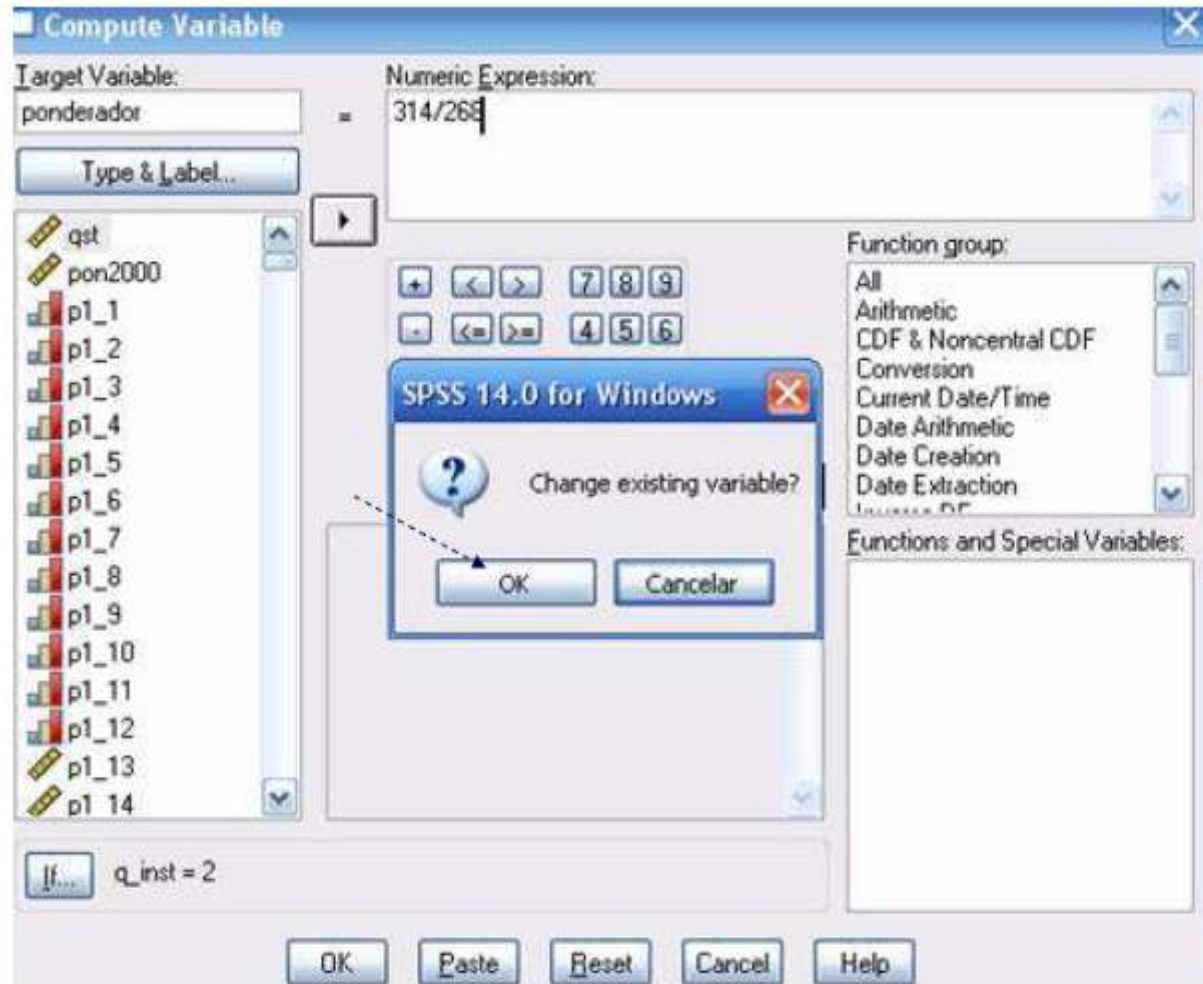
Construção de um ponderador no SPSS/PASW (3)



Construção de um ponderador no SPSS/PASW (4)

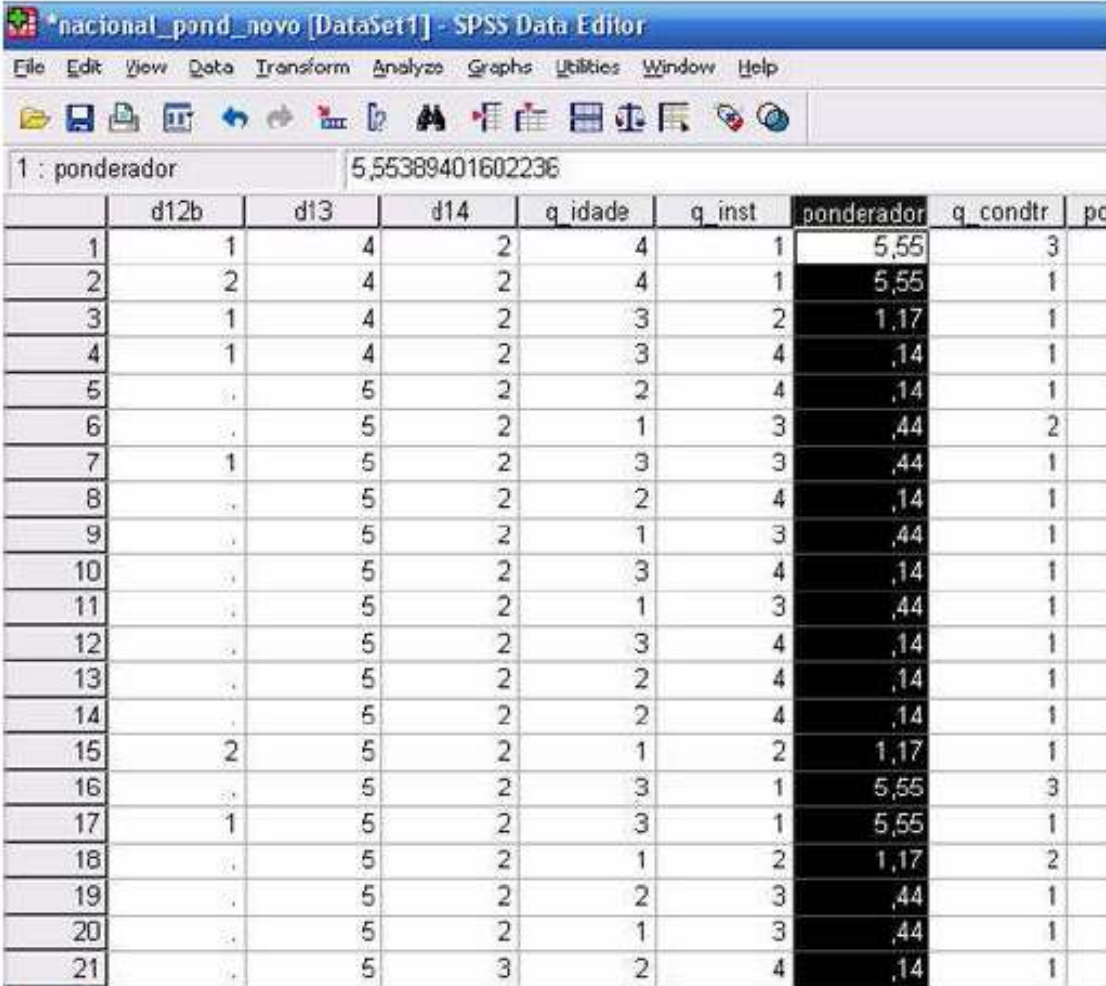


Construção de um ponderador no SPSS/PASW (5)



Repete-se o procedimento para as restantes 2 categorias de escolaridade

Construção de um ponderador no SPSS/PASW (6)



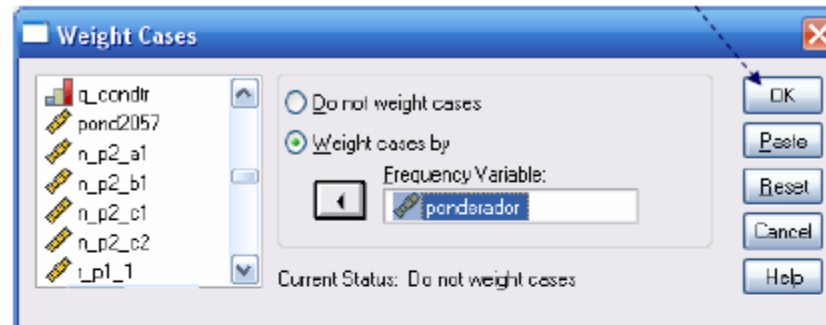
The screenshot shows the SPSS Data Editor window for a dataset named 'nacional_pond_novo'. The 'Data View' tab is active, displaying a table with 21 rows and 9 columns. The columns are: 'd12b', 'd13', 'd14', 'q_idade', 'q_inst', 'ponderador', 'q_condtr', and 'po'. The 'ponderador' column contains numerical values calculated from the other variables. The values range from 1.17 to 5.55. The 'q_condtr' column contains integer values from 1 to 3. The 'po' column contains integer values from 1 to 3.

	d12b	d13	d14	q_idade	q_inst	ponderador	q_condtr	po
1	1	4	2	4	1	5,55	3	
2	2	4	2	4	1	5,55	1	
3	1	4	2	3	2	1,17	1	
4	1	4	2	3	4	,14	1	
5	.	5	2	2	4	,14	1	
6	.	5	2	1	3	,44	2	
7	1	5	2	3	3	,44	1	
8	.	5	2	2	4	,14	1	
9	.	5	2	1	3	,44	1	
10	.	5	2	3	4	,14	1	
11	.	5	2	1	3	,44	1	
12	.	5	2	3	4	,14	1	
13	.	5	2	2	4	,14	1	
14	.	5	2	2	4	,14	1	
15	2	5	2	1	2	1,17	1	
16	.	5	2	3	1	5,55	3	
17	1	5	2	3	1	5,55	1	
18	.	5	2	1	2	1,17	2	
19	.	5	2	2	3	,44	1	
20	.	5	2	1	3	,44	1	
21	.	5	3	2	4	,14	1	

Visualização do ponderador na janela dos dados

Construção de um ponderador no SPSS/PASW (7)

Activação do ponderador



Visualização dos resultados (via *tables of frequencies*)

	<i>Grau de ensino completo</i>		
	<i>Count</i>	<i>%</i>	<i>Unweighted Count</i>
1 Até básico 2	1366	66,4	246
2 Básico 3	314	15,2	268
3 Secundário	239	11,6	543
4 Superior	138	6,7	1000
Total	2057	100,0	2057

	<i>Frequência leitura livros</i>		
	<i>Count</i>	<i>%</i>	<i>Unweighted Count</i>
1 Mais de 2 por mês	180	8,8	319
2 1 ou 2 por mês	301	14,6	535
3 5 ou 6 por ano	291	14,1	436
4 1 ou 2 por ano	327	15,9	385
5 Raramente	616	30,0	299
6 Nunca	340	16,6	80
Total	2056	100,0	2054

Principais fontes de erro

- **Erro amostral** (pode ser medido nas amostras probabilísticas)
- **Outros erros relacionados com o processo de amostragem** (exemplo: base de sondagem incompleta, não-respostas)
- Erros associados ao processo de recolha de informação (formulação das perguntas, técnicas de inquirição)
- Erros decorrentes do processamento dos dados (ex: informatização da informação, codificação)

Bibliografia

- Arozin, Francisco e José Luis Sánchez-Crespo (1986), Métodos y aplicaciones del muestro, Madrid, Alianza Universidad.
- Cochran, William G. (1977), Sampling Techniques, New York, John Wiley and Sons.
- Kalton, Graham (1983), Introduction to Survey Sampling, Series: Quantitative Applications in the Social Sciences, Sage / University Press.
- Meyneau, D. (1985), Les sondages d'opinion, Paris, Lá Découverte.
- Osuna, J.R. (1991), Métodos de muestro, Madrid, CIS.
- Osuna, J.R. (1993), Métodos de muestro. Casos Práticos, Madrid, CIS.
- Tagliacarne, Guglielmo (1986), Pesquisa de Mercado: Técnica e Prática, São Paulo, Editora Atlas S.A.
- Vicente, Paula, Elizabeth Reis e Fátima Ferrão (1996), Sondagens. A amostragem como Factor Decisivo de Qualidade, Lisboa, Sílabo.

Cálculo da Dimensão da Amostra

- Populações finitas

$$n = \frac{(Z_{\alpha/2})^2 \cdot N \cdot p \cdot q}{N \cdot E^2 + (Z_{\alpha/2})^2 \cdot p \cdot q}$$

- Populações infinitas

$$n = \frac{(Z_{\alpha/2})^2 \cdot p \cdot q}{E^2}$$

Legenda:

E - amplitude máxima de erro (diferença entre a estimação e o parâmetro);

α - nível de significância (ou probabilidade de erro)

P - valor estimado da proporção de indivíduos que possui o atributo

q - valor estimado da proporção de indivíduos que não possui o atributo

Z - valor crítico da normal-padrão para uma probabilidade $\alpha/2$

N - dimensão da população

n - dimensão da amostra

Cálculo do erro máximo admissível (E)

- **Populações finitas**

$$E = Z \cdot \sqrt{\frac{N-n}{N} \cdot \frac{p \cdot q}{n}}$$

- **Populações infinitas**

$$E = Z \cdot \sqrt{\frac{p \cdot q}{n}}$$

Determinação do nº de elementos em função da margem de erro

TABLA PARA LA DETERMINACION DE UNA MUESTRA SACADA DE UNA POBLACION FINITA, PARA MARGENES DE ERROR DEL 1, 2, 3, 4, y 5 POR 100, EN LA HIPOTESIS de $p = 50\%$

Margen de confianza del 95,5 por 100

Amplitud de la población	Amplitud de la muestra para márgenes de error abajo indicados					
	± 1 %	± 2 %	± 3 %	± 4 %	± 5 %	± 10 %
500					222	83
1.000				385	286	91
1.500				441	316	94
2.000			638	476	333	95
2.500			714	500	345	96
		1.250				
3.000		1.364	811	517	353	97
3.500		1.458	843	530	359	97
4.000		1.538	870	541	364	98
4.500		1.607	891	549	367	98
5.000		1.667	909	556	370	98
6.000		1.765	938	566	375	98
7.000		1.842	949	574	378	99
8.000		1.905	976	580	381	99
9.000		1.957	989	584	383	99
10.000	5.000	2.000	1.000	588	385	99
15.000	6.000	2.143	1.034	600	390	99
20.000	6.667	2.222	1.053	606	392	100
25.000	7.143	2.273	1.064	610	394	100
50.000	8.333	2.381	1.087	617	397	100
100.000	9.091	2.439	1.099	621	398	100
	10.000	2.500	1.111	625	400	100

À medida que cresce a dimensão da amostra, os custos do processo aumentam e os ganhos de precisão são mínimos.

p = proporción (en porcentaje) de los elementos portadores del carácter considerado. Si p es $< 50\%$ la muestra necesaria es más pequeña.

NOTA.—Cuando no se indica, la cifra significa que la muestra debería tener una amplitud superior a la mitad de la población, lo que equivaldría extender la encuesta al total de la población misma.

Fuente: ARKIN y COLTON, *Tables for Statisticians*.

Bravo, Sierra (1989), *Técnicas de Investigación Social*, Madrid, Editorial Paraninfo.

Programa de Doutoramento em Educação

Amostragem

**Seminário de Investigação
Métodos e Técnicas
de Recolha e Tratamento de Dados**

Susana Henriques, 2012



Adaptado de: Pós-graduação em Análise de Dados em Ciências Sociais (2009-2010),
ISCTE-Instituto Universitário de Lisboa