Noções de Amostragem

Rui Brites
(e-mail: rui.brites@iscte.pt)

Noções de amostragem*

Em Estatística, amostra é o conjunto de elementos extraídos de um conjunto maior, chamado População. É um conjunto constituído de indivíduos (famílias ou outras organizações), acontecimentos ou outros objectos de estudo que o investigador pretende descrever ou para os quais pretende generalizar as suas conclusões ou resultados.

******Amostra (estatística)*** **. In Infopédia [Em linha]. Porto: Porto Editora, 2003-2008. [Consult. 2008-01-15]. Disponível na www: <URL: http://www.infopedia.pt/$amostra-(estatistica)>.**

Principais razões para se trabalhar com uma amostra:

- A população é infinita, ou considerada como tal, não podendo portanto ser analisada na íntegra;
- Custo excessivo do processo de recolha e tratamento dos dados, como resultado da grande dimensão da população ou da complexidade do processo de caracterização de todos os elementos da população;
- Tempo excessivo do processo de recolha e tratamento dos dados, conduzindo à obtenção de informação desactualizada;
- As populações são dinâmicas, de onde resulta que os elementos ou objectos da população estão em constante renovação, de onde resulta a impossibilidade de analisar todos os elementos desta população;

Se a constituição da amostra obedecer a determinadas condições, a análise das características da amostra pode servir para se fazerem inferências sobre a população.

A dimensão da amostra é significativamente inferior à dimensão da população, de forma a justificar a constituição da amostra. A amostragem é, por sua vez, um conjunto de procedimentos através dos quais se selecciona uma amostra de uma população. Pode-se dividir as técnicas de amostragem em vário tipos:

- Amostragem probabilística - procedimento em que todos os elementos da população têm uma probabilidade conhecida e superior a zero de integrar a amostra;
- Amostragem não probabilística:
- Amostragem intencional - amostragem não probabilística subordinada a objectivos específicos do investigador;
- Amostragem não intencional - amostragem não probabilística regida por critérios de conveniência e/ou de disponibilidade dos inquiridos.

Amostra em estudos qualitativos

A Selecção das Pessoas Interrogadas

- Nos estudos qualitativos interroga-se um número limitado de pessoas.
- O critério que determina o valor da amostra passa a ser a sua adequação aos objectivos da investigação.
- Os indivíduos são escolhidos em função do seu carácter exemplar.

Amostra em estudos qualitativos (cont.)

A questão da grandeza da amostra

A amostra depende:

- *da heterogeneidade do público*;
- *do grau de complexidade dos objectivos da investigação;*
- *da saturação do "campo"* (mais inquiridos idênticos não trazem mais informação, mas a mesma informação)

As escolhas em matéria de amostragem apelam à perspicácia e ao bom senso

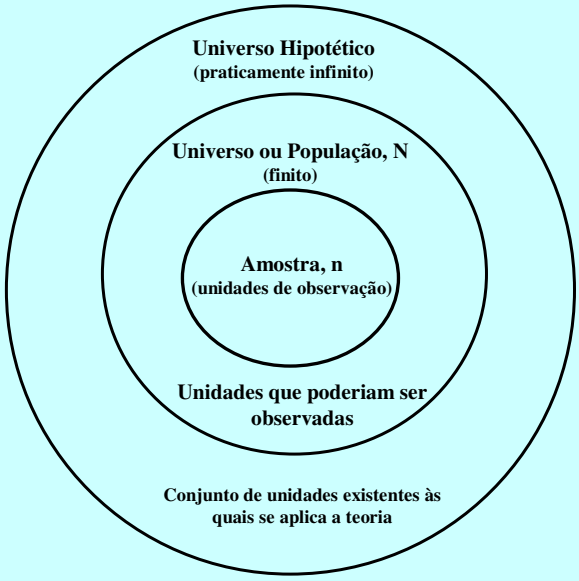
Questões prévias ao processo de amostragem de natureza quantitativa

1. Definição clara dos objectivos do estudo →	O que se pretende saber/conhecer melhor/compreender
2. Quem deverá ser entrevistado →	População alvo* e população a inquirir
3. Quantos deverão ser entrevistados →	Dimensão da amostra
4. Como serão seleccionados →	Método de selecção da amostra (escolha das unidades amostrais)

*Designa-se por população alvo a totalidade dos elementos sobre os quais se deseja obter determinado tipo de informação

7

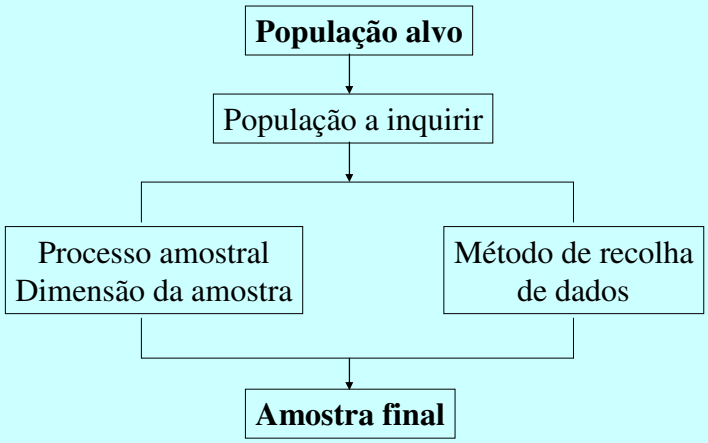
Representação de uma Amostra*



* in: Bravo, Sierra: **Técnicas de Investigación Social**, Madrid, Editorial Paraninfo, 1989

8

Desenvolvimento de um plano amostral*



* in: Reis, Melo, Andrade e Calapez: **Estatística aplicada, vol. 2**, Lisboa, Sílabo, 1999

9

Métodos de selecção da amostra*

Métodos probabilísticos (amostragem casual)

- Amostragem aleatória simples
- Amostragem sistemática
- Amostragem estratificada
- Amostragem por clusters
- Amostragem multi-etapas
- Amostragem multifásica

Métodos não probabilísticos (amostragem dirigida)

- Amostragem por conveniência
- Amostragem intencional
- Amostragem snowball
- Amostragem sequencial
- Amostragem por quotas

* in: Reis, Melo, Andrade e Calapez: **Estatística aplicada, vol. 2**, Lisboa, Sílabo, 1999

10

Métodos de selecção da amostra (cont)*

Métodos probabilísticos - amostragem aleatória

Aleatória simples	**Cada elemento da população tem a mesma probabilidade de ser seleccionado**
Casual sistemática	**Cálculo do rácio K=N/n; selecção aleatória do primeiro elemento da população e sequencial dos restantes**
Estratificada	**Separação dos elementos da população em estratos e selecção aleatória dos elementos dentro de cada estrato**
Clusters	**A população encontra-se dividida em clusters que são seleccionados aleatoriamente, constituindo as unidades amostrais**
Multi-etapas	**Idêntico ao anterior mas em que as unidades amostrais são seleccionadas aleatoriamente dentro de cada cluster**
Multi-fásica	**Numa 1ª fase recolhem-se dados sobre determinadas características dos respondentes (comportamentos e frequência de consumos, variáveis demográficas, etc.) e da sua disponibilidade para responder novamente a um inquérito. É então retirada desta fase uma sub-amostra que será inquirida na 2ª fase.**

* in: Reis, Melo, Andrade e Calapez: **Estatística aplicada, vol. 2**, Lisboa, Sílabo, 1999

Amostra estratificada*

Universo

Área / Sector	Norte (Braga e Porto)	Centro (Coimbra, Aveiro e Leiria)	Sul (Lisboa, Setúbal e Santarém)	Outros (Restantes distritos)	Total
Produto A	180	160	310	200	**850**
Produto B	150	70	230	50	**500**
Produto C	1260	550	700	190	**2700**
Produto D	1070	610	600	170	**2450**
Total	**2 660**	**1390**	**1840**	**610**	**6500**

Amostra (10% da população)

Área / Sector	Norte (Braga e Porto)	Centro (Coimbra, Aveiro e Leiria)	Sul (Lisboa, Setúbal e Santarém)	Outros (Restantes distritos)	Total
Produto A	18	16	31	20	**85**
Produto B	15	7	23	5	**50**
Produto C	126	55	70	19	**270**
Produto D	107	61	60	17	**245**
Total	**266**	**139**	**184**	**61**	**650**

* Adaptado de: Reis, Melo, Andrade e Calapez: **Estatística aplicada, vol. 2**, Lisboa, Sílabo, 1999

Métodos de selecção da amostra (cont)*

Métodos não probabilísticos - amostragem dirigida

Amostragem por conveniência	**A amostra é seleccionada em função da disponibilidade e acessibilidade dos elementos que constituem a população alvo**
Amostragem intencional	**A escolha dos elementos a incluir na amostra baseia-se na opinião de uma ou mais pessoas que conhecem muito bem as características específicas da população em estudo, que se pretende analisar**
Amostragem *snowball*	**Numa 1ª fase os inquiridos são escolhidos aleatoriamente, sendo, numa segunda fase, os inquiridos adicionais escolhidos com base na informação dos primeiros**
Amostragem sequencial	**Semelhante ao método multi-fásico. A realização da fase seguinte só é decidida depois de analisados os resultados da fase anterior.**
Amostragem por quotas	**Equivalente à amostragem aleatória estratificada. As proporções dos vários sub-grupos reflectem a sua distribuição dentro da população. Cada entrevistador dispõe das características que os entrevistados deverão satisfazer, terminando as entrevistas quando as quotas estiverem preenchidas.**

* in: Reis, Melo, Andrade e Calapez: **Estatística aplicada, vol. 2**, Lisboa, Sílabo, 1999

13

Determinação da margem de erro em função do nº de elementos*

220

DETERMINACION DEL MARGEN DE ERROR (*E*) DE UN RESULTADO, SEGUN EL NUMERO DE ELEMENTOS COMPRENDIDOS EN LA MUESTRA (1.ª COLUMNA) Y DE LOS VARIOS VALORES *p* (INDICADOS EN EL TITULO)

Fórmula: $E = 2\sqrt{\frac{p\,q}{N}}$

Seguridad del 95,5 % (2 sigma)

Número de elementos de la muestra	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
25 . . .	4,0	5,6	6,8	7,8	8,7	12,0	14,3	16,0	17,3	18,3	19,1	19,6	19,8	20,0
50 . . .	2,8	4,0	4,9	5,6	6,2	8,5	10,1	11,4	12,3	13,0	13,5	13,9	14,1	14,2
75 . . .	2,3	3,2	3,9	4,5	5,0	6,9	8,2	9,2	10,0	10,5	11,0	11,3	11,4	11,5
100 . . .	2,0	2,8	3,4	3,9	4,4	6,0	7,1	8,0	8,7	9,2	9,5	9,8	9,9	10,0
150 . . .	1,6	2,3	2,8	3,2	3,6	4,9	5,9	6,6	7,1	7,5	7,8	8,0	8,1	8,2
200 . . .	1,4	2,0	2,4	2,8	3,1	4,3	5,1	5,7	6,1	6,5	6,8	7,0	7,0	7,1
250 . . .	1,2	1,8	2,2	2,5	2,7	3,8	4,5	5,0	5,5	5,8	6,0	6,2	6,2	6,3
300 . . .	1,1	1,6	2,0	2,3	2,5	3,5	4,1	4,6	5,0	5,3	5,5	5,7	5,8	5,8
400 . . .	0,99	1,4	1,7	2,0	2,2	3,0	3,6	4,0	4,3	4,6	4,8	4,9	5,0	5,0
500 . . .	0,89	1,3	1,5	1,8	2,0	2,7	3,2	3,6	3,9	4,1	4,3	4,4	4,5	4,5
600 . . .	0,81	1,1	1,4	1,6	1,8	2,5	2,9	3,3	3,6	3,8	3,9	4,0	4,1	4,1
800 . . .	0,69	0,98	1,2	1,4	1,5	2,1	2,5	2,8	3,0	3,3	3,3	3,4	3,5	3,5
1.000 . . .	0,63	0,90	1,1	1,3	1,4	1,9	2,3	2,6	2,8	2,9	3,1	3,1	3,2	3,2
1.200 . . .	0,57	0,81	0,99	1,1	1,3	1,7	2,1	2,3	2,5	2,7	2,8	2,8	2,9	2,9
1.500 . . .	0,51	0,73	0,89	1,0	1,1	1,6	1,9	2,1	2,3	2,4	2,5	2,5	2,6	2,6
2.000 . . .	0,44	0,61	0,75	0,86	0,90	1,3	1,6	1,8	1,9	2,0	2,1	2,2	2,2	2,2
2.500 . . .	0,40	0,56	0,68	0,78	0,87	1,22	1,4	1,6	1,7	1,8	1,9	2,0	2,0	2,0
3.000 . . .	0,36	0,51	0,62	0,71	0,79	1,1	1,3	1,5	1,6	1,7	1,7	1,8	1,8	1,8
4.000 . . .	0,31	0,44	0,54	0,62	0,69	0,95	1,1	1,3	1,4	1,4	1,5	1,5	1,6	1,6
5.000 . . .	0,28	0,40	0,49	0,56	0,62	0,85	1,0	1,1	1,2	1,3	1,4	1,4	1,4	1,4
7.500 . . .	0,23	0,32	0,39	0,45	0,50	0,69	0,82	0,92	1,0	1,1	1,1	1,1	1,1	1,2
10.000 . . .	0,20	0,28	0,34	0,39	0,44	0,60	0,41	0,80	0,87	0,95	0,95	0,98	0,99	1,0
15.000 . . .	0,16	0,23	0,28	0,32	0,36	0,49	0,59	0,66	0,71	0,75	0,78	0,80	0,81	0,82
25.000 . . .	0,12	0,18	0,22	0,25	0,27	0,38	0,45	0,50	0,65	0,58	0,60	0,62	0,62	0,63
50.000 . . .	0,08	0,11	0,14	0,16	0,17	0,24	0,29	0,32	0,35	0,37	0,48	0,39	0,40	0,40

in: Bravo, Sierra: **Técnicas de Investigación Social**, Madrid, Editorial Paraninfo, 1989

14

Determinação nº de elementos em função da margem de erro *

Para población infinita o muy numerosa

TABLA PRONTUARIO ELABORADA POR TAGLIACARNE, PARA ESTABLECER SUFICIENTEMENTE SEGUROS (SEGURIDAD DEL 95,5 %) QUE EL RESULTADO ESTE COMPRENDIDO DENTRO DEL LIMITE DE ERROR (±) INDICADO AL MARGEN

Intervalo de confianza del 95,5 % (2 sigma)

Fórmula: $N = \frac{4\,p\,q}{E^2}$

Límites de error $\pm 2\sigma$ en % (E)	VALORES PRESUMIBLES DE p Y DE q en % (p + q = 100)													
	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
0,1 . . .	39.600	78.400	116.400	153.600	190.000	360.000	510.000	640.000	750.000	840.000	910.000	960.000	990.000	1.000.000
0,2 . . .	9.900	19.600	29.100	38.400	47.500	90.000	127.500	160.000	187.500	210.000	227.500	240.000	247.500	250.000
0,3 . . .	4.400	8.711	12.933	17.067	21.111	40.000	56.667	71.111	83.333	93.333	101.111	106.667	110.000	111.111
0,4 . . .	2.475	4.900	7.275	9.600	11.875	22.500	31.875	40.000	46.875	52.500	56.875	60.000	61.875	62.500
0,5 . . .	1.584	3.136	4.656	6.144	7.600	14.400	20.400	25.600	30.000	33.600	36.400	38.400	39.600	40.000
0,6 . . .	1.100	2.178	3.233	4.267	5.278	10.000	14.167	17.778	20.833	23.333	25.278	26.667	27.500	27.778
0,7 . . .	808	1.600	2.376	3.135	3.878	7.347	10.408	13.061	15.306	17.143	18.577	19.592	20.204	20.408
0,8 . . .	619	1.225	1.819	2.400	2.969	5.625	7.969	10.000	11.719	13.125	14.219	15.000	15.469	15.625
0,9 . . .	489	968	1.437	1.896	2.346	4.444	6.296	7.901	9.259	10.370	11.235	11.852	12.222	12.346
1,0 . . .	396	784	1.164	1.536	1.900	3.600	5.100	6.400	7.500	8.400	9.100	9.600	9.900	10.000
1,5 . . .	176	348	517	683	844	1.600	2.267	2.844	3.333	3.733	4.044	4.267	4.400	4.444
2,0 . . .	99	196	291	384	475	900	1.275	1.600	1.875	2.100	2.275	2.400	2.475	2.500
2,5 . . .	63	125	186	246	304	576	816	1.024	1.200	1.344	1.456	1.536	1.584	1.600
3,0 . . .	44	87	129	171	211	400	567	711	833	933	1.011	1.067	1.100	1.111
3,5 . . .	32	64	95	125	155	294	416	522	612	686	743	784	808	816
4,0 . . .	25	49	73	96	119	225	319	400	469	525	569	600	619	625
4,5 . . .	20	39	57	76	94	178	252	316	370	415	449	474	489	494
5,0 . . .	16	31	47	61	76	144	204	256	300	336	364	384	396	400
6,0 . . .	11	22	32	43	53	100	142	178	208	233	253	267	275	278
7,0 . . .	8	16	24	31	39	73	104	131	153	171	186	196	202	204
8,0 . . .	6	12	18	24	30	56	80	100	117	131	142	150	155	156
9,0 . . .	5	10	14	19	23	44	63	79	93	104	112	119	122	123
10,0 . . .	4	8	12	15	19	36	51	64	75	83	91	96	99	100
15,0 . . .	2	3	5	7	8	16	23	28	33	37	40	43	44	45
20,0 . . .	1	2	3	4	5	9	13	16	19	21	23	24	25	25
25,0 . . .	0,6	1	2	2	3	6	8	12	12	13	15	15	16	16

in: Bravo, Sierra: **Técnicas de Investigación Social**, Madrid, Editorial Paraninfo, 1989

15

Determinacão nº de elementos em funcão 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			638	441	316	94
2.000			714	476	333	95
2.500		1.250	769	500	345	96
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

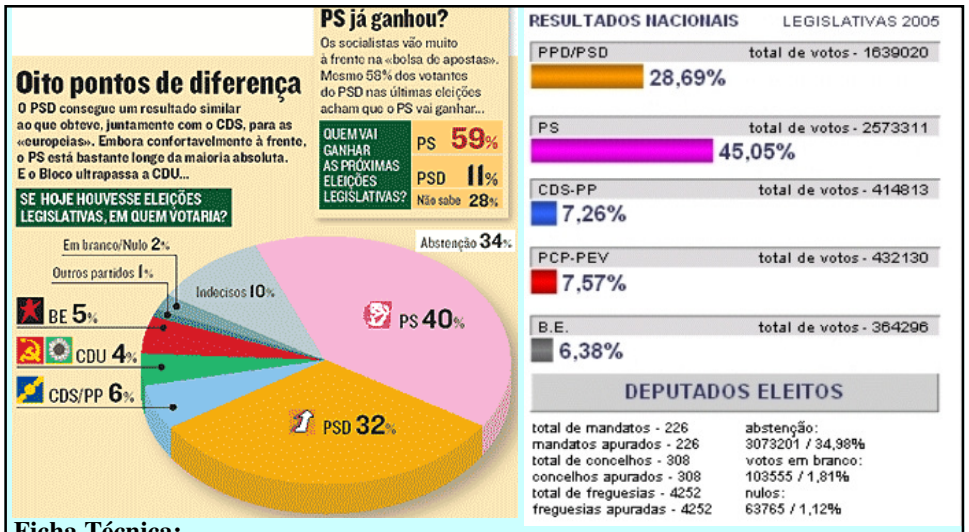
p = proporción (en porcentaje) de los elementos portadores del carácter considerado. Si p es < 50 % la muestra necesaria es más pequeña.

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*.

in: Bravo, Sierra: **Técnicas de Investigación Social**, Madrid, Editorial Paraninfo, 1989

16



Ficha Técnica:

Realização: entre 17 e 21 de Janeiro de 2005. Objectivos da sondagem: opinião sobre os principais problemas do País, avaliação da actuação dos líderes partidários, características importantes num primeiro-ministro, imagem de Pedro Santana Lopes e José Sócrates, dinâmica de vitória eleitoral e intenção de voto. Universo do estudo: indivíduos com 18 ou mais anos de idade, residentes e recenseados no continente, vivendo em lares com telefone fixo. Amostra: total de 800 entrevistas telefónicas, recolhidas por idade, sexo, habitat, região, ocupação e voto nas últimas eleições legislativas. Entidade responsável: TNS/Euroteste tendo por técnico responsável José António Simões Vidal de Oliveira

Cavaco aumenta vantagem apesar de estar mais longe da vitória

DN, 24/11/2005

Público, 24/11/2005