

IV. Sondagens, amostras

Sondar

Trata-se de selecionar uma amostra para fins de análise. Os critérios da seleção dependem de uma decisão estratégica e podem ser muito variáveis. Pode-se notadamente escolher amostras contrastadas, tal como se sugeriu para o teste prévio de um questionário ou para selecionar “casos” a fim de registrar histórias de vida. A questão é saber em que medida a amostra pode ser tida como *representativa*, mas pode ser o caso de representar os contrastes sociais ou um “conjunto”. É contudo esta última concepção que tende a prevalecer em razão dos usos estatísticos popularizados pelas grandes pesquisas nacionais (ou internacionais) e pelas pesquisas de opinião. Quase sempre associada à pesquisa por questionário, a sondagem estatística se destina então a dar fundamento a extrapolações a partir de resultados ou hipóteses advindos de outros métodos e implica amostras de “tamanho suficiente”. Como é a mais conhecida, começaremos por ela.

- *O que é uma amostra de tamanho suficiente?* — Seria ótimo poder responder antes de decidir o método da pesquisa: em um tempo limitado, com meios reduzidos, para uma

dissertação, por exemplo, costuma ser melhor aprofundar métodos mais exploratórios (observações ou entrevistas) do que visar uma exploração estatística com uma quantidade demasiado pequena de questionários. O conjunto das grandes amostras dos institutos de pesquisa (geralmente em torno de mil) serve freqüentemente de referência — mas é mais do que pode esperar um pesquisador isolado. É entretanto somente *a posteriori* que se pode avaliar a representatividade estatística da amostra, em função do conjunto coligido, sem dúvida, mas também da distribuição das respostas (cf. representatividade das sondagens).

Sem condições de aplicar uma medida, pode-se no entanto tentar antecipar suas conclusões por uma aproximação ponderada. É necessário imaginar, por exemplo, o futuro quadro que cruzará duas variáveis, origem social e instrução: para um quadro cruzado desse tipo, considera-se às vezes prudente ter uma quantidade de casos igual ou superior a 20. Mesmo que se formule a hipótese (favorável) de respostas distribuídas de maneira relativamente homogênea entre os diferentes itens, não se tarda a ver os limites estatísticos de uma amostra de 100 pessoas: cruzar duas variáveis, cada uma com três itens (instrução superior, secundária, primária; classe alta, média, popular...) determina nove casos cujos conjuntos serão, por conseguinte, em quatro casos ao menos, inferiores a vinte!

AS SONDAgens ALEATÓRIAS

A amostra é constituída por sorteio a partir de uma "base de sondagem", lista precisa da população de referência.

Procedimentos de sondagem aleatória simples

Considere-se uma sondagem de proporção definida (sondagem de 1:10) operando sobre uma base de sondagem (uma lista de estado civil, uma relação de associados, uma lista de estudantes inscritos num mesmo curso...) de 4.000 pessoas. É o seguinte o procedimento habitual:

- 1) decide-se uma classificação que ordene esses indivíduos: pode ser a própria ordem das listas ou relações;
- 2) sendo a sondagem de 1:10, sorteia-se um número inferior ou igual a 10: esse número designa o primeiro indivíduo a sondar (poderá ser, por exemplo, o sexto da lista);
- 3) a partir desse número, que designa o primeiro sondado, escolhem-se os seguintes de 10 em 10 (se o primeiro sondado for o sexto da lista, os seguintes serão o décimo sexto, o vigésimo sexto etc.).

Em vez desse procedimento, pode-se também utilizar tabelas preestabelecidas de números aleatórios, prática antiga um tanto obsoleta, ou recorrer às funções que, nos computadores e calculadoras (cf. tecla "random" de algumas calculadoras), propõem séries de números produzidos segundo uma lei uniforme.

Sondagens aleatórias por conglomerados (em uma ou n etapas)

Quando a unidade estatística sorteada é um grupo, fala-se de *sondagem por conglomerado*. Esse grupo pode ser uma família, uma aglomeração, um estabelecimento... O questionário diz respeito ao conjunto do grupo; ele pode ser aplica-

do a todos os membros dele ou apenas a alguns, definidos previamente (o chefe da família, por exemplo) ou em um segundo momento. Se, no interior dos primeiros conglomerados sorteados, proceder-se a novos sorteios, fala-se então de *sondagem em várias etapas*. Por exemplo:

1ª etapa: sorteio dos estados;

2ª etapa: sorteio dos municípios nos estados;

3ª etapa: sorteio das residências nos municípios.

É esse o procedimento das “pesquisas domiciliares” do IBGE, por exemplo. Existe também a sondagem em duas ou *n* etapas nas pesquisas eleitorais de boca de urna (1ª. etapa: sorteio das seções de voto; 2ª. etapa: sorteio dos eleitores que pertencem a essas seções), nas pesquisas “de passageiros” das companhias de transporte, nas medições de audiência etc.

Quando a unidade de sondagem é uma área, um quarteirão, fala-se de sondagem aureolar.

Sondagens aleatórias estratificadas

A *sondagem estratificada*, por sua vez, sempre é uma sondagem com múltiplos sorteios (seja em uma ou em várias etapas). Divide-se a população (a base) em subconjuntos ou *estratos* (categoria socioprofissional*, faixas etárias, sexo...) e opera-se uma sondagem em cada um deles: a taxa nesse caso pode ser diferente.

Far-se-á, por exemplo, uma sondagem de 1:20 entre os operários e de 1:10 entre os executivos, menos numerosos, para ter um número suficiente de questionários de executivos...

* Doravante abreviada CSP.

Se as taxas diferem, deve-se então, por ocasião da exploração, tratar cada estrato separadamente. Se se quer um tratamento estatístico global, destina-se a cada um uma ponderação inversamente proporcional à sua taxa de sondagem.

No exemplo precedente, os questionários dos operários terão uma ponderação dupla em relação à dos questionários dos executivos; eles serão contados duas vezes.

Na escala da população francesa, as sondagens aleatórias estratificadas são utilizadas quase exclusivamente pelo INSEE. Depois de cada recenseamento (pesquisa exaustiva com todos os indivíduos da população), o INSEE procede também a sondagens entre os questionários recolhidos (sondagens no vigésimo, no quinto...) para publicar os primeiros resultados.

Reposição de amostra

Como a distribuição das não-respostas não é aleatória, em sociologia as reposições de amostra são de uso relativamente limitado. A proposta do método de reposição de amostra é, com efeito, restabelecer a estrutura inicial de uma amostra estratificada caso ela tenha sido fortemente afetada por taxas de não-retornos elevados e distribuídos de maneira desigual segundo os estratos.

Se metade dos ex-alunos da seção A de uma escola respondeu ao questionário que lhe foi enviado, dois terços da seção B e 95% da seção C, os questionários podem então ser contados duas vezes se provêm de A e ser modificados por um coeficiente multiplicador igual a 1,5 para os da seção B e a 100/95 para os da seção C.

A reposição postula que, no interior de cada categoria ou estrato, os que não responderam teriam dado o mesmo tipo de respostas que aqueles que responderam. Esse postulado contradiz e elimina a hipótese de construção social dos não-retornos.

Representatividade e erro aleatório

- *Representatividade* — Em um primeiro sentido, a representatividade de uma amostra, quer seja ela aleatória ou intencional, depende, como vimos, de seu tamanho, de sua confiabilidade e da precisão da base de sondagem e do rigor do protocolo de coleta das informações. Mas é sobretudo a representatividade estatística das sondagens que a palavra evoca, representatividade cuja medida se fundamenta na lei dos grandes números e no cálculo das probabilidades e supõe, portanto, um procedimento de sondagem aleatória.

- *Limite de confiança, intervalo de confiança* — A questão a que se pode responder é a seguinte: qual é a probabilidade para que p , a proporção observada na amostra, deva-se ao acaso? Mais precisamente, com que risco de erro (denominado *limite de confiança*) pode-se dizer que P , a proporção na população, é igual a p , a proporção observada na amostra, mais ou menos um certo intervalo (chamado de *intervalo de confiança* ou *margem de erro*)?

- *A tabela da lei normal reduzida* — O valor do limite de confiança ou do intervalo de confiança pode ser calculado a

partir de uma ou outra das fórmulas a seguir. Elas se referem à tabela dos desvios reduzidos (z ou u).

Lemos, por exemplo, que, para z igual a 1,645, então 0,10, ou seja, 10% dos valores, são superiores ao valor $x + 1,645$ e inferiores a $x - 1,645$. Há dez chances em 100 (limite de confiança) para que os valores observados sejam exteriores ao intervalo de confiança $\pm 1,645$; e há 90% de chances de que eles se inscrevam entre esses dois limites extremos.

LEI NORMAL REDUZIDA:
PROBABILIDADE DE QUE O DESVIO SEJA ULTRAPASSADO;
PROVA BILATERAL

Desvio reduzido z	Probabilidade P	Desvio reduzido z	Probabilidade P
0	1	1,960	0,05
0,126	0,90	2,054	0,04
0,253	0,80	2,170	0,03
0,385	0,70	2,326	0,02
0,524	0,60	2,576	0,01
0,674	0,50	3,291	0,001
0,842	0,40	3,891	0,0001
1,036	0,30	4,417	0,00001
1,282	0,20	4,892	0,000001
1,645	0,10	5,327	0,0000001
1,695	0,09	5,731	0,00000001
1,751	0,08	6,109	0,000000001
1,812	0,07	-	-
1,881	0,06	-	-

- *Estimar os limites e os intervalos* — A fórmula mais completa supõe conhecer (além de p , a proporção na amostra, q ,

a proporção complementar, e n, o tamanho da amostra) o valor da taxa de sondagem f.

$$P = p \pm z \sqrt{(1-f) \left(\frac{pq}{n-1} \right)}$$

O valor de z se lê na tabela dos desvios reduzidos:

Por exemplo, no limite de confiança 0,02, z é igual a 2,326; no limite 0,05 é igual a 1,96.

Se $n = 400$ e $p = 0,25$, no limite de confiança 0,05 o intervalo de confiança passa de mais ou menos 0,0329 (3,29%) se a população-mãe é de mil pessoas a 0,0416 (4,16%) se ela é de cem mil.

Uma fórmula aproximada não leva em conta a taxa de sondagem.

$$P = p \pm z \sqrt{\frac{pq}{n}}$$

No exemplo precedente ($n = 400$, $p = 25\%$), no limite de confiança 0,05, segundo essa fórmula o intervalo de confiança é igual a $\pm 0,042$ (4,2%).

$$P = 0,25 \pm 1,96 z \sqrt{\frac{0,25 * 0,75}{400}} = 0,042.$$

Não levar em consideração a taxa de sondagem quase não altera o resultado.

Na tabela que segue, é apresentada uma leitura mais rápida do intervalo de confiança (ou margem de erro, aqui anotada como "e"), fundada na equação:

$$e = 2 \sigma - 2 z \sqrt{\frac{pq}{n}}$$

No limite de confiança 0,05 (5%), lê-se na tabela que, para uma amostra de tamanho $n = 400$ e para $p = 25\%$, o valor do intervalo de confiança (margem de erro) se lê na interseção da linha $n = 400$ e da coluna $p = 25\%$: ele é igual a 4,3. $P = 25\% \pm 4,3\%$.

MARGEM DE ERRO (E) NO LIMITE DE CONFIANÇA 5%

p → n ↓	2%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
	ou 98%	ou 95%	ou 90%	ou 85%	ou 80%	ou 75%	ou 70%	ou 65%	ou 60%	ou 55%	
100	2,6	4,4	6	7,2	8	8,6	9,2	9,6	9,8	10	10
200	2	3,1	4,3	5,1	5,7	6,1	6,5	6,8	6,9	7,1	7,1
250	1,7	2,8	3,8	4,5	5	5,4	5,8	6	6,2	6,3	6,3
300	1,6	2,5	3,5	4,2	4,6	5	5,3	5,6	5,7	5,8	5,8
350	1,5	2,3	3,2	3,8	4,2	4,55	4,9	5,1	5,2	5,3	5,3
400	1,4	2,2	3	3,6	4	4,3	4,6	4,8	4,9	5	5
500	1,1	2	2,7	3,2	3,6	3,9	4,1	4,3	4,4	4,5	5
600	1,1	1,8	2,4	3	3,3	3,5	3,8	3,9	4	4,1	4,1
700	1	1,7	2,3	2,7	3	3,3	3,5	3,65	3,7	3,8	3,8
800	1	1,5	2,1	2,5	2,8	3	3,2	3,35	3,4	3,5	3,5
900	0,9	1,45	2	2,4	2,6	2,85	3	3,15	3,2	3,1	3,3
1000	0,8	1,4	1,8	2,25	2,5	2,65	2,85	3	3	2,9	3,1
1200	0,8	1,3	1,7	2,1	2,3	2,5	2,65	2,8	2,85	2,6	2,9
1500	0,7	1,15	1,5	1,85	2,1	2,25	2,4	2,5	2,55	2,35	2,6
1800	0,6	1,05	1,4	1,7	1,9	2,02	2,15	2,25	2,3	2,25	2,35
2000	0,6	1	1,3	1,6	1,8	1,95	2,1	2,15	2,2	2	2,25

• Representatividade e taxa de sondagem — A representatividade de uma amostra depende mais de seu efetivo do que da taxa de sondagem. Por exemplo, segundo a última tabela, no limite 0,5 e para $p = 25\%$, o intervalo de confiança passa de 4,3 ($n = 400$) a 3 se "n" se eleva a 700 e a 8,6 se desce a 100.

Resulta daí que uma amostra pode ter uma representatividade satisfatória para uma população mas não para este ou aquele de seus componentes (determinada classe, determinada circunscrição...), uma vez que n aí é menor.

Disso decorre também que há muito poucas razões de se considerar uma amostra de tamanho n mais representativa se ela é retirada de uma população de tamanho reduzido ($N = 2.000...$) do que se a base de sondagem for muito maior ($N = 2.000.000...$).

Qualquer que seja o modo de estimativa, ele coloca a questão daquilo que se denominou a significação estatística, questão particularmente importante quando se trata de comparar proporções oriundas de uma amostra (ou de várias). Os desvios entre proporções podem com efeito ser "estatisticamente não significativos" considerados o limite e o intervalo de confiança que a eles correspondem.

Se, na seqüência de uma pesquisa pré-eleitoral, o candidato A obtém 45% das intenções de voto e o candidato B 55%, será que podemos, sem correr riscos, dar este último como futuro vencedor? No limite de confiança 0,05, tanto para um como para o outro o intervalo de confiança é igual a ± 5 : podemos então dizer com 95 por cento de chances de acertar que o candidato A obterá entre 40% e 50% dos votos e o candidato B entre 50% e 60%. A vitória não está garantida, um segundo turno ainda é possível.

AS SONDAgens INTENCIONAIS

Com esse nome agrupamos procedimentos de sondagem que visam assegurar a representatividade da amostra confor-

mando sua estrutura às características da população. A representatividade depende diretamente do número, da confiabilidade (exatidão), da precisão e da pertinência (adequação aos objetivos da pesquisa) das informações disponíveis sobre o conjunto da população de referência, que é denominado *população-mãe*.

O método das cotas

Trata-se do método mais empregado pelos grandes organismos franceses de estudos e de sondagens (com exceção do INSEE). Supõe uma informação confiável, precisa e rica sobre a população-mãe. Em um dado setor geográfico, decide-se por uma estrutura de amostra que seja, segundo vários

EXEMPLO DE UMA FOLHA DE COTAS (14 ENTREVISTAS)

Cidade: Bairro:		
Sexo	Homem	7 *****
	Mulher	7 *****
Idade	15-24	2 **
	25-34	3 ***
	35-44	3 ***
	45-54	3 ***
	55 e mais	3 ***
CSP do chefe de família	Operário, contramestre, pessoal de serviço	5 *****
	Artesão, pequeno comerciante	3 ***
	Executivo de nível médio, empregado	3 ***
	Executivo de nível superior, prof. liberal, grande comerciante	1 *
	Inativos, aposentados	2 **

critérios, tão próxima quanto possível daquela da população de referência (sexo, idade, profissão, bairro de residência...).

É o pesquisador, em seguida, que se encarrega de interrogar um número determinado de pessoas no setor designado com antecedência; ele deve escolhê-las de tal maneira que sua amostra apresente as características estruturais definidas previamente: um número prédefinido de homens e de mulheres, de representantes de cada faixa etária e de cada grupo socioprofissional etc.

As pesquisas de rua

Com esse nome agrupamos as pesquisas, no mais das vezes encomendadas e de opinião, pesquisas de viabilidade ou de satisfação, que recaem sobre um *público* cuja composição não é bem identificada: os transeuntes na rua estão satisfeitos com o projeto de urbanização levado a cabo pela prefeitura? O que pensam os usuários do serviço de atendimento da delegacia? O sociólogo, desprovido de qualquer base de sondagem, pode fazer algo melhor do que ir colocar suas questões na rua ou no serviço de atendimento? E pode ele interrogar “ao acaso”, como se costuma dizer, mas de um jeito que nada tem a ver com os procedimentos estritos da sondagem aleatória e que significa, ao contrário, sem procedimento nem controle? Contra esse “ao acaso”, o sociólogo raciocina e constrói hipóteses: que esse público não tem a mesma composição segundo as horas, segundo os dias, nem talvez em uma ponta e em outra da mesma rua... Essas hipóteses desembocam em um plano de observação (pare-

cem verificadas?), em entrevistas com informantes (os comerciantes, o pessoal do serviço de atendimento...) e em seguida em um plano de sondagem. É o próprio questionário que irá permitir, *a posteriori*, estabelecer sua validade com a condição de que nele sejam colocadas as questões relativas às hipóteses: quem vem quando? com que frequência? só ou acompanhado? com que objetivo?

Nessas condições, aquilo que, no início, era apenas um “estudo” com uma amostragem mal definida acaba por ganhar o estatuto de pesquisa sociológica.

As sondagens bola de neve

Diz-se que uma sondagem é bola de neve quando se pede às primeiras pessoas interrogadas que indiquem outras que se encaixem na pesquisa: o mesmo pedido será feito a essas e assim por diante. A sondagem bola de neve tem um objetivo exaustivo em um dado território (não se trata de constituir uma amostra) e na maioria das vezes é feita de entrevistas semidirigidas. Ela supõe que o primeiro grupo de pessoas interrogadas tenha sido selecionado com base em outros critérios.

O método foi com frequência utilizado nas pesquisas sobre as “elites”. O grupo ao qual inicialmente se aplicou a pesquisa é então composto de pessoas selecionadas em razão da posição que ocupam no topo de uma ou de várias hierarquias institucionais (políticos eleitos, bispos, presidentes de câmara de comércio...). Pede-se a cada membro desse grupo que indique quais são, na sua área, as pessoas que contam,

que têm poder, autoridade. Com base nessas informações, completa-se o grupo inicial; busca-se o diálogo com esses novos citados e, ao longo dessa segunda etapa da pesquisa, pode-se apresentar a eles a questão formulada às primeiras “elites”: quem conta, na opinião deles, em sua área? Pode-se considerar terminada a pesquisa quando todos os nomes citados (ou quase todos) correspondem a pessoas que já foram entrevistadas.

O método é pertinente se se trata de estudar *formas de poder e de redes* (de solidariedade, de amizade, com base profissional, sexual, étnica...) e, de maneira mais geral, o que Pierre Bourdieu [1980] denomina o *capital social*.

Segunda Parte

O TRATAMENTO: ANÁLISE E SÍNTESE

Para designar os métodos de tratamento, prevaleceu a palavra análise: análise de conteúdo, análise multivariada, análise dos dados etc. Todos, entretanto, associam análise e síntese.

A análise, é verdade, parece com freqüência marcar o primeiro momento deles, aquele em que se isolam unidades significantes (temas, figuras, variáveis...) abstraídas de seu contexto para operar com elas a comparação termo a termo. Mas isolar, abstrair, já é comparar; é propor diferenças e semelhanças, portanto estabelecer relações. E relacionar engaja a síntese. Ela é seu ato elementar e fundamental. "Passar em revista" uma pluralidade de relações e extrair delas alguma ordem é, no sentido próprio do termo, teorizar.

Como o que está em causa é o tratamento, pode-se lançar a hipótese de que a perspectiva analítica prevaleceu na medida mesmo em que se privilegiaram as verificações do elo entre os elementos de informação recolhidos. Mesmo estando situada no centro da verificação no processo de pesquisa, a questão do tratamento não se encontra todavia nela circunscrita e qualquer método de tratamento comporta um momento mais exploratório, geralmente no início, e uma dinâmica própria para fazer emergirem questões novas.

Em face de dados diversos, numerosos, heterogêneos, para que se possa "passá-los em revista" é preciso fazer aparecer pistas (o que procurar?), um sentido, uma ordem, é preciso imaginar o que as liga entre si, seja para construir

hipóteses ou problemáticas provisórias, seja para prepará-las: isso é particularmente importante no início da pesquisa, mas também a cada vez que emerge uma informação, uma questão nova. A mesma exigência se faz quando, analisados os elos termo a termo, trata-se de operar a síntese dessas ligações parciais, síntese globalizante ou síntese estruturada de sua configuração de conjunto.

Assim, entre os métodos de tratamento, alguns têm objetivos mais sintéticos, então nós os agrupamos no início, quando se trata de elaborar hipóteses, e no final desta parte, na conclusão da apresentação dos métodos de tratamento estatístico.

V. Ordenar, classificar

Guias de observação e guias de entrevista visam constituir conjuntos de informação homogêneos e organizados segundo uma ordem construída para favorecer a comparação. Mas a pesquisa documental, a observação e a entrevista recolhem sempre uma pluralidade de informações heterogêneas, reúnem “uma multidão de fenômenos dados isoladamente, difusos e discretos, que podem ser encontrados ora em grande número, ora em pequeno número e não absolutamente por locais” [Weber, 1904], e que se trata de ordenar, no início, no decorrer ou até mesmo no final da pesquisa.

HIPÓTESES E DEFINIÇÕES

O método ideal-típico

A citação de Max Weber nos dá a introdução ao método que ele preconiza para “ordenar” tais fenômenos: a elaboração de tipos ideais. Traduzido do alemão, o termo “ideal” indica que o “conceito-limite” funciona como um modelo possível para o conhecimento. Ele tem por objetivo “formar o julgamento de imputação”, “guiar a elaboração das hipóte-