

O Estabelecimento de Categorias

OS PRINCÍPIOS DE CLASSIFICAÇÃO COMO BASE PARA
CONJUNTOS DE CATEGORIAS

Num levantamento de bairro, uma amostra de população devia responder à seguinte pergunta: "Que tipos de pessoas vivem neste bairro?" O objetivo da pergunta não era obter informação objetiva a respeito do bairro (o que poderia ter sido obtido, de maneira mais exata e com menos esforço, através de números de recenseamento), mas verificar até que ponto as pessoas do bairro pensavam em função de distinções de grupos étnicos. Eis algumas das respostas:

Quase que sô pessoas pobres como eu.

Pessoas de côr e nós.

Negros, italianos, judeus e muitos outros.

Existe muito o que dizer sobre as pessoas que moram por aqui. Mas não quero complicações. A gente precisa viver onde pode.

Pessoas de côr.

É um bairro "duro". Aqui vivem todos os tipos de pessoas.

Sei que existem alguns irlandeses.

Muitos judeus.

Um grande número de negros.

Realmente não sei; sou nôvo aqui.

Evidentemente, para que várias centenas de respostas como essas possam ser organizadas de modo a serem usadas para responder às perguntas da pesquisa, precisam ser agrupadas em certo número de categorias. A fim de decidir quais são as categorias significativas, é preciso escolher algum princípio de classificação. A pergunta da pesquisa ou as hipóteses - se for

2. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

muladas - apresentam a base para a escolha dos princípios de classificação. Suponha-se que esse estudo estivesse voltado para a consciência da existência do grupo negro; o princípio adequado de classificação seria um princípio baseado em referência explícita aos negros. Esse princípio de classificação imediatamente sugere duas categorias.

Menciona negros

Não menciona negros

Essas duas categorias formam um "conjunto de categorias". Um conjunto de categorias deve atender a algumas regras básicas:

1. O conjunto de categorias deve ser derivado de um único princípio de classificação.
2. O conjunto de categorias deve ser exaustivo; vale dizer, deve ser possível colocar qualquer respostas numa das categorias do conjunto.
3. As categorias do conjunto devem ser mutuamente exclusivas; não deve ser possível colocar determinada resposta em mais de uma categoria do conjunto.

Um conjunto pode ser composto por mais de duas categorias, desde que essas regras não sejam violadas. Por exemplo, um exame superficial das respostas enumeradas acima mostra que "não menciona negros" inclui muitas respostas diferentes. A pessoa que não deseja falar sobre os vizinhos por temor de perturbações ou aquela que disse que era nova no bairro são muito diferentes daquelas que mencionam outros grupos, e não os negros. Isso indica a necessidade de subcategorias para distinguir entre os diferentes tipos de respostas que não mencionam os negros. Por exemplo, pode parecer psicologicamente significativo distinguir entre pessoas que não mencionam negros, mas mencionam outros grupos, e pessoas que não mencionam nem os negros nem outros grupos. Nesse caso, o conjunto estaria composto por três categorias:

Menciona negros (com ou sem menção de outros negros)

Não menciona negros, mas menciona outros grupos

Não menciona negros, nem outros grupos

É preciso decidir quanto a essa extensão. A observação das respostas reais sugere que esse conjunto de categorias ainda não faz justiça ao seu conteúdo. Os grupos não-negros mencionados apresentam várias diferenças significativas: algumas pessoas descrevem os grupos através de termos econômicos (pessoas pobres); outras fazem referencia a atributos pessoais - ("duros"), outros mencionam grupos religiosos ou nacionais (judeus, irlandeses). Cada um desses atributos, escolhidos pelas pessoas para descrever os grupos de que estão conscientes, apresenta um possível princípio de classificação que poderia conduzir a outro conjunto de categorias, mais ou menos como no caso dos negros. Evidentemente, em princípio é possível usar todos esses atributos para o estabelecimento de conjuntos de categoria. Na prática, isso é frequentemente pouco econômico e pouco compensador, pois nem todos esses princípios de classificação se ligam aos objetivos da pesquisa.

Se, como o supusemos, o estudo se interessa pela consciência da presença dos negros, poderia parecer que todos esses outros princípios de classificação não são significativos. Todavia, essa opinião se baseia em noção muito restrita da tarefa de análise. O fato de certa porcentagem mencionar os negros ainda não é uma indicação suficiente de sua consciência dos negros. Pode ser que qualquer outro grupo étnico do bairro tivesse sido mencionado da mesma forma. O fato de haver ou não uma consciência específica da presença de negros só pode ser decidido se se fizer uma codificação da menção de outros grupos raciais, nacionais ou religiosos. No entanto, a classificação das respostas através de menção de status econômico ou atributos pessoais de pessoas do bairro provavelmente não teria qualquer função.

A fim de apresentar essas classificações adicionais de re

4. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

ferencias a outros grupos étnicos, é preciso estabelecer outros conjuntos de categorias. No entanto, cada um desses conjuntos deve obedecer às regras apresentadas na página 441. A lista seguinte, que constitui o "código" para a pergunta, poderia ser o resultado final:

Grupos raciais

1. Negros: mencionados; não mencionados
2. Outros grupos raciais: mencionados; não mencionados

Grupos de nacionalidade

3. Irlandeses: mencionados; não mencionados
4. Italianos: mencionados; não mencionados
5. Outros grupos nacionais: mencionados; não mencionados

Grupos religiosos

6. Judeus: mencionados; não mencionados
7. Católicos: mencionados; não mencionados
8. Protestantes: mencionados; não mencionados
9. Outros grupos religiosos: mencionados; não mencionados
10. Menção dos grupos acima enumerados
Menciona um ou mais; não menciona qualquer deles, mas menciona outros grupos humanos; não menciona qualquer deles, afirmando "eu não conheço"; não responde à pergunta

Observa-se que, embora estejamos lidando com as respostas a apenas uma pergunta, nossos interesses específicos de resposta nos levaram a dez conjuntos de categorias para classificar as pessoas. Cada uma das pessoas que respondem pode ser colocada em uma das duas categorias de cada um dos nove primeiros conjuntos, e em uma das quatro categorias do conjunto final.

Como o fato de não assinalar um "mencionado" específico pode ser considerado como suposição do correspondente "não mencionado", a aparência do código pode ser simplificada eliminando-se todas as categorias de "não mencionado"; no entanto, cada um dos nove primeiros conjuntos continua a ser um conjunto de duas categorias, uma das quais indicada por um sinal e a outra por ausência de sinal. Todavia, essa simplificação nem sempre é boa. O fato de não assinalar uma determinada categoria "mencionado" pode representar um esquecimento na codificação. Esses esquecimentos podem ser desestimulados (e certamente identificáveis) através da exigência de que a colocação de uma pessoa numa categoria "não mencionado" deva exigir um ato tão positivo quanto a sua colocação numa categoria "mencionado". Além disso, para fazer uma pequena antecipação, quando se faz a codificação de tais dados para a tabulação por máquina, é muito desejável que todas as categorias tenham uma identificação explícita. As máquinas mais usadas contam, separadamente, o número de casos em cada categoria. A soma de contagens em todas as categorias de um conjunto deve ser igual ao número total de casos; se a categoria "não mencionado" e a "mencionado" são explicitamente codificadas, podemos verificar se todos os casos foram contados.

Várias qualidades desse código devem ser indicadas. Em primeiro lugar, para criar os dez conjuntos de categorias, foram usados quatro critérios fundamentais de classificação: menção de grupos raciais versus não menção de grupos raciais; menção de grupos nacionais versus não menção de grupos nacionais; menção de grupos religiosos versus não menção de grupos religiosos; menção de qualquer desses tipos de grupos versus não menção de qualquer desses grupos. Deve-se notar que cada um desses quatro critérios de classificação permite certo número de conjuntos de categorias. Assim, "menção de grupos nacionais versus não menção de grupos nacionais" inclui os seguintes conjuntos de categorias: menção de irlandeses versus não menção de irlandeses; menção de italianos versus não menção de italianos; menção de outros grupos nacionais versus não men-

6. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

ção de outros grupos nacionais.

Em segundo lugar, as categorias em cada conjunto são mutuamente exclusivas, e também incluem todas as possibilidades significativas para esse conjunto de categorias, de forma que cada resposta possa ser colocada, sem ambiguidade, numa categoria ou outra; por exemplo, determinada pessoa menciona ou não menciona os negros. (Pode ter também mencionado outros grupos raciais, bem como judeus e irlandeses. Cada um desses elementos de sua resposta é considerado separadamente nesse código, mas cada um deles pode ser claramente colocado no seu conjunto significativo de categorias).

Em terceiro lugar, o código total é exaustivo na medida em que existe um lugar para cada resposta possível (sobretudo porque existe a categoria residual que "aceita tudo", que é a de número 10). Todavia, em quarto lugar, não é exaustiva em função de todos os possíveis princípios de classificação que poderiam ser aplicados às respostas. Além dos já eliminados como não significativos para o problema da pesquisa, não inclui, por exemplo, as distinções entre a maneira, depreciativa ou não, de fazer referência aos grupos.

Finalmente, deve-se discutir a inclusão na lista das categorias "católicos" e "protestantes". A tabulação posterior mostrou que 3 por cento das pessoas mencionaram os católicos e 0 por cento os protestantes. Na realidade, isso tinha inclusão desses dois grupos decorreu, em parte, do conhecimento da composição real do bairro e, em parte, da expectativa e interesse na demonstração de que esses dois grupos, embora tivessem muitos membros no bairro, não eram considerados como tão separados ou distintos quanto os outros grupos.

Neste caso, a "categorização de dados ausentes" - isto é, a apresentação de uma categoria para uma resposta cuja ocorrência não se espera - pode conduzir a um resultado óbvio, isto é, a demonstração de que os membros de grupos dominantes ou socialmente aceitos frequentemente não são percebidos através de sua participação nesses grupos. No entanto, o princípio

aqui existente está longe de ser óbvio, e é muito frequentemente esquecido. Baseia-se na noção de que as idéias para categorização sempre devem vir de duas fontes: um conhecimento íntimo das provas existentes e conhecimento geral e análise prévia dos possíveis tipos de respostas, baseada em considerações teóricas, lógicas ou práticas. A aplicação dessa regra ao estabelecimento de categorias conduz à possível descoberta de ausência significativa de alguma resposta, o que poderia, de outra forma, passar despercebido.

Existem tipos de estudo e tipos de dados para os quais o estabelecimento de categorias é ainda mais simples que neste exemplo. Quando se usam perguntas para assinalar, ou qualquer forma de escala de avaliação, aplicada numa situação de observação, o analista dos dados praticamente não precisa dar atenção ao problema do estabelecimento de categorias. Seus dados foram pré-categorizados nos estágios anteriores da pesquisa. Isso também é verdade em todos os estudos em que as observações são registradas em categorias prescritas. Essa pré-categorização, quando realizada como parte do desenvolvimento de instruções para pesquisa, é naturalmente, muito semelhante ao processo empregado no estabelecimento de categorias depois da coleta de dados.

DEFINIÇÃO DE CATEGORIAS COMPLEXAS

Em algumas situações, o estabelecimento de categorias é consideravelmente mais difícil e exige muito mais tempo que o exemplo do bairro, acima apresentado. O que, neste caso, tornou o estabelecimento de categorias um processo relativamente fácil foi o fato de as respostas serem relativamente simples e nítidas; as categorias poderiam ser facilmente definidas de maneira não-ambígua. Embora as categorias devam ser sempre definidas dessa maneira, a tarefa é muito mais difícil com alguns tipos de conteúdo que com outros.

Por exemplo, Merton e seus colaboradores (estudo inédito)

8. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

perguntaram a senhoras brancas: "O que acha que os moradores de côr sentem por viverem na mesma comunidade com brancos?" E perguntaram a senhoras negras: "O que acha que os moradores brancos sentem por viverem na atribuição de atitudes muito favoráveis até atribuição de atitudes muito desfavoráveis. Eis alguns exemplos disso:

Desejam viver aqui, de forma que possam dizer que são iguais.guais.

Parecem muito amistosos conosco.

Gostam da idéia.

Penso que não se incomodam: nós nos entendemos.

Alguns gostam disso, outros não.

Pensam que isso os rebaixa.

Não estou em contato com eles, e por isso não sei.

Detestam isso.

De acordo com a discussão anterior, não seria difícil criar um conjunto simples de categorias, baseado no princípio de classificação de atitude favorável versus atitude desfavorável atribuída à outra raça. Neste caso, seria necessário uma categoria para atitude favorável, uma para desfavorável, outra para neutra, além de uma categoria residual. No entanto, entre as atitudes favoráveis e desfavoráveis que são atribuídas, aparecem certas sutilezas de sentido. Uma pessoa que diz: "Desejam viver aqui, de forma que possam dizer que são iguais", revela algo diferente daquela que diz: "Gostam da idéia". De forma semelhante, aquela que diz: "Pensam que isso os rebaixa" pode querer dizer coisa diversa de quem diz: "Detestam isso". Qual é o critério distintivo?. Aparentemente, em ambos os casos algumas das pessoas que respondem atribuem o suposto sentimento da outra raça a um valor que, segundo acreditam, é aceito por esse grupo. Por isso, a atribuição de tais valores foi usada como outro princípio de classificação. Criou-se um conjunto de categorias em que uma categoria representava a atribuição de atitude favorável da outra raça a um valor material, social ou espiritual, ou a um benefício que obtinham com a situação (ou a atribuição de sua atitude desfavorá

vel a perda ou desvantagem correspondentes), e outra categoria representava uma ausência de afirmação sobre valor. Para simplificação da operação subsequente de codificação, os dois conjuntos de categorias foram combinados da seguinte maneira:

Atitude favorável atribuída aos membros da outra raça, explicada através de valores ou benefícios que obtêm com o fato de viverem no mesmo conjunto residencial com a raça da pessoa que responde.

Atitude favorável atribuída à outra raça, sem explicação explícita em termos de valor

Atitude neutra ou acomodatória atribuída à outra raça.

Atitude desfavorável atribuída aos membros da outra raça, e explicada em função de valores ou benefícios que obtêm com a segregação.

Atitude desfavorável atribuída à outra raça, sem explicação explícita em termos de valor.

Outras respostas, ausência de resposta, não sabe.

Antes de usar uma lista de tais categorias, é necessário especificar, tão exatamente quanto possível, o conteúdo de cada categoria. Isso é feito através da adição de uma ou duas sentenças explicativas a cada categoria e da exemplificação de seu sentido.

Por exemplo, no caso da primeira categoria da lista acima apresentada, a sentença explicativa era a seguinte: Tais respostas supõem que a outra raça obtém algo com o fato de viver com a raça aqui, de forma que possam dizer que são iguais."

O trabalho com essas categorias complexas exige muito esforço e cuidado na classificação. Mesmo quando as categorias são cuidadosamente estabelecidas, seu emprego apresentará mais problemas que o uso de categorias definidas de forma restrita e exata. Se uma pessoa responde, por exemplo: "Gostam de tudo que encontram aqui, e sabem por que", não se sabe se essa afirmação supõe ou não um benefício. É preciso estabelecer outras regras, a fim de lidar com tais respostas. Neste caso, uma regra adequada para a primeira categoria poderia ser

10. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

a seguinte: "Esta categoria aplica-se apenas às afirmações que especificam a natureza do valor ou benefício".

SELEÇÃO DE PRINCÍPIOS DE CLASSIFICAÇÃO PARA A CATEGORIZAÇÃO DE MATERIAL ASSISTEMÁTICO

Surgem problemas especiais na categorização de material as sistemático - por exemplo, em protocolos de observação, histórias de caso, discursos de agitadores, entrevistas assistemáticas, etc. Já aludimos a essas dificuldades nos capítulos anteriores, ao discutir a observação participante e o emprego, na pesquisa, de documentos pessoais e de conteúdo de comunicação.

Num estudo que empregue instrumentos assistemáticos para a obtenção de dados significativos para hipóteses ou problemas de pesquisa claramente formulados, os princípios adequados para a classificação de respostas são claramente prescritos pela natureza do estímulo ou pelas perguntas e respostas. No entanto, ao trabalhar com provas assistemáticas, o primeiro problema é decidir quanto aos aspectos do material que devem ser categorizados - isto é, quais os princípios de classificação que devem ser usados para o estabelecimento de conjuntos de categorias.

O estabelecimento de princípios de classificação é diffi-cil, sobretudo, em estudos exploratórios, pois tais estudos, por definição, não partem de hipóteses explícitas. No momento da coleta de dados, o pesquisador não sabe quais são os aspectos que se revelarão mais importantes. Por isso, geralmente precisa coligir um grande número de dados; assim, na análise, tem o problema de lidar não apenas com material assistemático, mas também com grande quantidade de tal material, grande parte do qual pode não ter sifnificação para seu objetivo.

O primeiro passo na análise de dados de um estudo explorátório é criar hipóteses de trabalho que permitam princípios de classificação. Geralmente, o pesquisador lê cuidadosamente todo o seu material e mantém-se alerta para indicações nos dados. Existem vários processos que podem ajudá-lo em sua tare

fa. Um dêles é estudar, quando possível, o material de um grupo que apresente contraste com aquêles que estão estudando, a fim de obter idéias a respeito de diferenças importantes entre os dois. Por exemplo, num estudo sobre delinquência, é adequado ler registros de casos em instituições sociais, não apenas a respeito de delinquentes, mas também de outros jovens que estejam sob os cuidados de tal instituição. Vale dizer, contrasta casos que diferem, nitidamente, na característica que está sendo pesquisada, a fim de ver quais as outras diferenças concomitantes que podem ser percebidas nos casos contrastantes.

O pesquisador pode separar seus casos em grupos que parecem comuns, e depois perguntar a si mesmo o que o levou a pensar que os colocados num único grupo sejam semelhantes. Por exemplo, Chein et al. (1952), num estudo sobre opiniões de eminentes educadores judeus e assistentes sociais a respeito de alguns problemas de educação de judeus, consideraram adequado selecionar as pessoas, inicialmente, em função de sua visão total do sentido de ser judeu. O pesquisador pode verificar que agrupou os seus casos a partir de características comuns; pode, depois, examiná-los, a fim de verificar se os que têm características semelhantes passaram por experiências semelhantes. Pode descobrir que seu agrupamento se baseia em experiências semelhantes; depois, pode rever os casos, a fim de verificar se tais experiências semelhantes parecem ter levado a consequências semelhantes.

Outro processo que pode estimular a formulação de hipóteses de trabalho é observar questões que parecem surpreendentes diante de expectativas do senso comum ou da teoria, e depois procurar explicações para o fenômeno surpreendente. Por exemplo, Lambert e Bressler (1957), num estudo sobre estudantes hindus nos Estados Unidos, observaram que esses estudantes frequentemente pareciam mais perturbados por perguntas, a respeito de problemas hindus, propostas por norte-americanos relativamente bem informados, que por estereótipos ou pela ignorância dos que não tinham informação a respeito da Índia;

12. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

mesmo quando as perguntas não eram feitas com intenção inamistosa, os estudantes pareciam interpretá-las como crítica hostil. Essa observação conduziu à hipótese de que, para pessoas de países coloniais ou que foram coloniais, criam-se algumas "áreas sensíveis." Essas "áreas sensíveis" são aspectos da cultura nacional que, historicamente, constituíram objeto de crítica hostil por parte do país ou países dominantes e foram usadas como desculpa para a manutenção do país na situação de subordinação. Qualquer referencia a essas áreas, qualquer que seja a intenção da pessoa que as apresenta, traz consigo as conotações históricas de hostilidade. No entanto, o estranho (nesse caso, o norte-americano) precisa ter algum conhecimento do país de origem do estudante para ter ciência dessas áreas sensíveis; por isso, as perguntas de pessoas relativamente bem informadas apresentam maior tendência para atingir tais áreas. As perguntas de pessoas sem qualquer informação, que geralmente não se referem a essas áreas sensíveis, são atribuídas à ignorância; as perguntas referentes às áreas sensíveis, que só podem ser propostas por pessoas com alguma informação a respeito do país, são interpretadas como provas de hostilidade.

No entanto, mesmo com hipóteses claras, a análise de material assistemático apresenta problemas especiais. Como por definição, o material não é o do tipo em que a mesma pergunta pode ser respondida por todos, ou em que as mesmas observações são feitas a respeito de todas as pessoas, sempre existe a possibilidade de que a informação sobre determinado aspecto não exista em alguns documentos. De outro lado, tende a haver muito material que não é diretamente significativo para as hipóteses. Além disso, existe o problema de decidir quanto ao tamanho de unidades de material a que devem ser aplicadas as categorias. Se os dados foram obtidos com um instrumento sistemático - por exemplo, um questionário -, geralmente cada pergunta apresenta uma unidade natural para a categorização, embora alguns conjuntos de categorias possam aplicar-se a unidades maiores, - por exemplo, o questionário como um todo. No entanto, quando lidamos com material assistemático, não existi

tem essas unidades "naturais" e adequadas. Por exemplo, se o pesquisador emprega registros de casos mantidos por agências de serviço social, pode categorizar cada afirmação feita pelo cliente, ou cada sessão do cliente com o assistente social, ou todo o registro do caso. Precisa decidir qual dessas unidades é mais adequada para as respostas às suas perguntas específicas de pesquisa.

Codificação: A Categorização dos Dados

A codificação é o processo técnico pelo qual são categorizados os dados. Através da codificação, os dados brutos são transformados em símbolos - geralmente numerais - que podem ser tabulados e contados. No entanto, a transformação não é automática; exige julgamento por parte de um codificador.

Frequentemente, o julgamento que coloca uma resposta numa categoria é feito por uma pessoa que não é aquela que tem o título oficial de "codificador." Muitas vezes, é a pessoa que responde que coloca sua resposta numa categoria. Isso ocorre em muitas perguntas de tipo de levantamento de opinião e de escolha múltipla; por exemplo, quando o indivíduo se limita a uma resposta "sim", "não", "não sei", ou "concordo" "discordo", "indeciso", ou quando indica por um sinal sua posição numa escala de avaliação. Ou, então, a pessoa que colige os dados pode categorizar no momento em que os colige. Naturalmente é isso que ocorre quando um entrevistador ou um observador empregam uma escala de avaliação para descrever o comportamento de uma pessoa.

Existem muitas vantagens no fato de o entrevistador ou o observador categorizarem os dados. De um lado, estão numa posição que permite observar a situação e o comportamento do indivíduo. Assim, têm mais informação que o codificador - que trabalha com o registro escrito - para basear um julgamento. Outra vantagem é que a categorização por quem colige os dados economiza tempo e dinheiro.

Apesar dessas vantagens, geralmente a categorização de da-

14. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

dos complexos é feita por codificadores, depois da coleta de dados. Esse processo dá tempo para reflexão; os julgamentos momentâneos de um entrevistador ou de um observador podem não ser tão esclarecidos quanto os feitos com mais tempo para decisão. Os julgamentos dos que coligem dados podem ser coloridos por aspectos pouco importantes - por exemplo, a aparência e os gestos da pessoa que responde, seu sotaque, suas respostas a perguntas anteriores, etc. Além disso, se cada entrevistador ou observador categorizam apenas os dados que coligem, isso tende a aumentar a imprecisão. Existe tendência para a criação de um quadro de referencia para o material que estamos codificando. Ainda que, inicialmente, os que coligem dados fossem perfeitamente coerentes no seu emprego de categorias - uma suposição provavelmente errada -, tenderiam a criar, depois de algum tempo, diferentes quadros de referencia adequados para seus limitados materiais, o que tornaria suas categorizações imprecisas. No escritório, é mais fácil obter e verificar um quadro comum de referencia.

PROBLEMAS DE PRECISÃO NA CODIFICAÇÃO²

Muitas coisas podem contribuir para que os julgamentos dos codificadores³ se tornem imprecisos. Tais fatores podem surgir dos dados que devem ser categorizados, das categorias que devem ser aplicadas, dos codificadores, etc. Consideremos, rapidamente, alguns desses fatores e algumas formas de impedir sua atuação.

DIFICULDADES QUE SURGEM DOS DADOS. Muitas das dificuldades que ocorrem na codificação resultam de inadequações dos dados. Frequentemente, os dados não apresentam suficiente informação significativa para codificação precisa. Isso pode ser consequência de processos inadequados de coleta de dados - perguntas mal formuladas, observadores não-treinados, etc. No entanto, talvez seja mais frequente poder-se corrigir facilmente as dificuldades através de organização cuidadosa dos dados.

Quando o entrevistador ou o observador entregam seu mate-

rial, ainda existe a possibilidade de eliminar muitas dificuldades potenciais. Muitos problemas de codificação podem ser evitados através do exame cuidadoso dos dados logo depois de sua coleta e, se necessário, de um interrogatório sistemático do entrevistador ou do observador. O processo de verificação dos dados a fim de melhorar sua qualidade para a codificação é geralmente denominado revisão.⁴

A revisão não apenas ajuda a evitar problemas posteriores de codificação; pode também melhorar nitidamente a qualidade da coleta de dados, pois chama a atenção para pontos em que os entrevistadores ou observadores não compreenderam as instruções, não estão registrando os dados com suficiente minúcia, etc. Para que tenha essa função, a revisão deve ser feita durante o pré-teste da ficha de entrevista ou de observação e durante o treinamento de entrevistadores ou observadores, bem como durante todo o período de coleta dos dados. Em qualquer caso, para que a revisão afaste os problemas de codificação, deve ser feita enquanto os entrevistadores ou observadores ainda estão disponíveis para a discussão.

Cada ficha de entrevista ou observação deve ser verificada quanto a:

1. Complementação. Todos os itens devem ser preenchidos. Uma parte em branco depois de uma pergunta numa ficha de entrevista pode significar "não sei", "recusa-se a responder" ou que a pergunta não era aplicável, ou que a pergunta foi omitida por engano, etc. Para muitos objetivos, é importante

2 Para uma discussão geral dos problemas de precisão, ver o Capítulo 5.

3 Aqui, falamos de pessoas que codificam os dados depois de estes terem sido coligidos, não de pessoas que respondem ou que coligem dados.

4 Para uma discussão minuciosa do processo de revisão de levantamento em grande escala, ver Parten (1950, Cap. 13).

16. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

que sejamos capazes de distinguir entre esses vários sentidos potenciais.

2. Legibilidade. Se o codificador não pode decifrar a caligrafia do entrevistador ou do observador, ou as abreviações e símbolos que emprega, a codificação é impossível. É simples verificar a legibilidade no momento em que o material é entregue, e fazer com que seja reescrito se isso fôr necessário, mas frequentemente é muito demorado, para o codificador, decifrar a caligrafia ou encontrar de novo o entrevistador, depois do início da codificação.

3. Compreensibilidade. Frequentemente, um comportamento registrado ou uma resposta parecem perfeitamente compreensíveis para o entrevistador ou observador, mas não para outra pessoa. O entrevistador conhece o contexto em que ocorreu a resposta ou o comportamento, mas isso não ocorre com o codificador, que por isso não pode saber exatamente o que foi que a pessoa fez, nem entender o sentido de sua resposta. A qualidade da codificação será muito melhorada pelo interrogatório sistemático do entrevistador ou observador, a fim de afastar confusões e ambiguidade.

4. Coerência. Incoerências marcante em determinadas ficha de observação ou de entrevista não apenas criam problemas na codificação; podem indicar erros na coleta ou no registro de dados. Por exemplo, se, numa entrevista sobre relações entre negros e brancos, a resposta a uma pergunta é que a pessoa que responde não conhece qualquer família negra, mas um comentário posterior mostra que essa pessoa faz e recebe visitas de uma família negra, existe uma necessidade óbvia de verificar imediatamente essa incoerência. Muitas vezes, o entrevistador, durante o interrogatório, esclarece a incoerência. Se não é capaz de fazê-lo, pode ser desejável estabelecer contato com a pessoa que responde, desde que o aspecto seja importante para a análise dos dados.

5. Uniformidade. De modo geral, as instruções adequadas para os entrevistadores ou observadores permitirão processos uniformes para a coleta e o registro de dados; no entanto, é necessário verificar a uniformidade com a qual tais instru

ções foram seguidas. Por exemplo, se a renda é registrada em unidades diferentes das especificadas nas instruções - por exemplo, em função de ordenado mensal e não semanal - e se as unidades de registro não são claramente indicadas, a codificação pode ser errada.

6. Respostas inadequadas. Às vezes, uma resposta não é pertinente aos objetivos da pesquisa; isso tende a ocorrer se uma pergunta não é claramente apresentada ou não é formulada de forma inteligível. Para o codificador, é bom que essas respostas sejam separadas das respostas adequadas, pois, provavelmente, as categorias criadas para codificar as respostas não serão aplicáveis.

DIFICULDADES DECORRENTES DAS CATEGORIAS. O valor da categorização dos dados depende inteiramente da adequação das categorias empregadas. As categorias devem ser bem definidas, de um ponto de vista conceitual, e devem ser significativas para os objetivos da pesquisa. Como se indicou na página 41, cada conjunto de categorias deve também atender a certas exigências formais: deve ser baseado num único princípio de classificação, as categorias devem ser mutuamente exclusivas e exaustivas. No entanto, mesmo que as categorias satisfaçam a todas essas condições, a codificação será imprecisa se as categorias não forem claramente definidas em função de indicadores aplicáveis aos dados imediatos. Geralmente, as categorias são definidas através de exemplos dos dados. Os exemplos devem incluir não apenas respostas típicas de categoria, mas também, se possível, respostas que ajudem a distinguir os limites entre categorias semelhantes.

TREINAMENTO DE CODIFICADORES. É evidente que a precisão da codificação é influenciada pela competência dos codificadores. Geralmente, o treinamento de codificadores obedece às seguintes etapas: (1) os vários códigos são explicados e ilustrados com exemplos do material a ser codificado. (2) Todos os codificadores praticam numa amostra dos dados. Os problemas que surgem são discutidos pelos codificadores, como grupo, com o supervisor, a fim de chegar a definições e procesos.

18. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

tos comuns. (3) Frequentemente, por causa da prática de codificação, as categorias são revistas, a fim de que se tornem mais aplicáveis ao material e a fim de que se registrem, por escrito, os processos e definições obtidos durante a codificação preliminar. (4) No momento em que, no período de prática, surge um número relativamente pequeno de problemas, os codificadores trabalham em parte idêntica dos dados, sem se consultarem e sem consultarem o supervisor. Depois, a coerência ou precisão da codificação é calculada, a fim de verificar se é possível começar imediatamente a codificação.

Ao calcular a coerência da codificação, é possível usar a codificação do supervisor, como um critério para verificar os diferentes codificadores, ou usar alguma medida que reflita a coerência do grupo como um todo, ou comparar cada codificador com cada um dos outros. De acordo com os resultados da verificação de precisão, é possível decidir pela eliminação das categorias que parecem excessivamente imprecisas para que tenham valor, ou gastar mais tempo no treinamento, ou eliminar os codificadores mais incoerentes, ou usar processos especiais com os itens mais difíceis (por exemplo, fazer com que os analistas mais especializados trabalhem com os itens mais difíceis, ou fazer com que duas ou mais pessoas, trabalhando em conjunto, analisem os itens mais difíceis). (5) Depois do início da codificação do material, são necessárias verificações periódicas de coerência, a fim de que se possa ter certeza de que os codificadores não se tornam descuidados quando adquirem mais experiência ou de que não criam métodos idiosincráticos para enfrentar novos problemas no material. O supervisor precisa insistir que todos os novos problemas devem ser discutidos com ele. Para assegurar uniformidade, quaisquer decisões posteriores ao início da codificação devem ser imediatamente comunicadas a todos os codificadores.

CÁLCULO DA PRECISÃO DA CODIFICAÇÃO

Evidentemente, a coerência ou adequação na colocação de determinado tipo de resposta em determinada categoria terá uma grande influência no resultado da análise. Por isso, é impor

tante verificar a precisão de codificação e aumentar, tanto quanto possível, o acordo entre os codificadores. É difícil estabelecer, como padrão, qualquer nível determinado de precisão que deva ser atingido. Diferentes tipos de material apresentam graus diferentes de dificuldade para a obtenção de precisão; de modo geral, quanto mais estruturado o material a ser codificado, e quanto mais simples as categorias usadas, - maior será a precisão. Além disso, diferentes objetivos de pesquisa podem exigir diferentes padrões de acordo. Em qualquer caso, é desejável não apenas verificar a precisão da codificação, mas tornar essa informação acessível ao leitor, ao apresentar o estudo. Essa prática torna-se cada vez mais comum.

Existem vários métodos para calcular a precisão da codificação. Como os métodos mais cuidadosos exigem uma apresentação estatística mais minuciosa que a que seria adequada aqui, o leitor interessado deverá procurar os artigos representativos desses trabalhos.⁵ Deve-se notar, no entanto, que talvez o erro mais frequente nas verificações de precisão seja o cálculo de coeficientes de precisão inadequados para o uso dos dados. Assim, im processo comum - mas errado - consiste em descrever apenas a precisão de categorias gerais, mas omitir as verificações de precisão das subcategorias. Por exemplo, se na análise vamos empregar subcategorias de "dar início ao comportamento" - por exemplo, "iniciar comportamento que é seguido por outros membros de um grupo" e "iniciar comportamento que não é seguido por outros membros do grupo" -, seria errado realizar ou descrever verificações de precisão apenas na categoria geral de "dar início a comportamento."

TABULAÇÃO

A tabulação é uma parte do processo técnico na análise estatística dos dados. A operação essencial na tabulação é a

⁵ Ver Guetzkow (1950); Lasswell, Leites e colaboradores (1949); Robinson (1957); Schutz (1952); Scott (1955).

contagem, para determinar o número de casos que estão nas várias categorias. O termo marginais é comumente usado para fazer referência a simples contagens das frequências com que ocorrem as várias categorias em cada conjunto; por exemplo, o número de pessoas que não frequentaram escola depois do curso primário, o número das que frequentaram o ginásio, mas não se formaram, etc. Os termos tabulação cruzada e desdobramento são frequentemente empregados para indicar a tabulação do número de casos que ocorrem juntamente em duas ou mais categorias - por exemplo, tabulação do número de casos que apresentam elevada educação e pequena renda. A tabulação cruzada é um passo essencial para a descoberta ou verificação de relações entre as variáveis nos dados. Mais adiante, neste capítulo, discutiremos esse emprego de tabulação cruzada.

A tabulação pode ser feita inteiramente à mão, ou pode ser feita à máquina. Os dois métodos têm vantagens e de desvantagens. Em resumo, a tabulação manual é geralmente menos dispendiosa e menos demorada quando o número de casos é pequeno ou relativamente pequeno, quando é pequeno o número de conjunto de categorias a ser contado, quando não é necessário fazer grande número de tabulações cruzadas. À medida que aumenta o número de casos ou de tabulações cruzadas, o uso de tabulações à máquina se torna cada vez mais econômico. O número de tabulações cruzadas é, talvez, o fator mais importante para a determinação da relativa eficiência de um ou de outro processo de tabulação.

Geralmente, a eficiência da tabulação manual tende a ser subestimada. Com técnicas adequadas, a tabulação manual pode ser muito rápida e exata. Uma das técnicas mais eficientes usa, para cada caso, um pequeno cartão de código que pode ser facilmente escolhido e contado. Num cartão como esses podem ser colocados tantos resultados de conjuntos de categorias quanto num cartão de 80 colunas, comumente usado na tabulação à máquina. Com o emprego de alguns sinais - como cores, grifos, etc. - é fácil distinguir os códigos, e os cartões podem ser eficientemente separados. Parten (1950)⁶ diz que um escri

turário relativamente pouco qualificado pode separar, em seis pilhas, 1.000 desses cartões em menos de cinco minutos. A contagem é também quase tão rápida se se empregam cartões pequenos: 1.000 cartões podem ser contados em menos de cinco minutos, desde que se empregue uma boa técnica. Nessa velocidade a separação e a contagem manuais podem, em pequenos intervalos de tempo, rivalizar em eficiência com a tabulação à máquina, mesmo para o tratamento de vários milhares de casos.

Tanto a tabulação à máquina quanto a tabulação manual pressupõem que os dados tenham sido codificados e que a codificação tenha sido verificada. Geralmente, os resultados de cada indivíduo são transcritos num cartão para tabulação manual ou numa folha, a partir da qual um marcador de cartões os transcreve (com furos) em cartões para tabulação à máquina. Na realidade, esta exige mais operações de escritório e outras operações especializadas que a tabulação manual. A perfuração de cartões, a verificação das tabulações à máquina, a transposição dos resultados das formas de tabulação à máquina para tabelas - tais etapas não são exigidas na tabulação manual. Com um grande número de casos ou com muitas tabulações cruzadas, a rapidez das máquinas compensa, com vantagem, o tempo gasto nessas operações.

Existem vários tipos de máquinas; nos últimos anos os desenvolvimentos nesse campo têm sido extremamente rápidos. Algumas máquinas apenas separam e contam os cartões; outras separam, contam e imprimem os resultados; outras são capazes de realizar operações estatísticas mais complexas. Estas máquinas são extremamente complexas; precisam ser "programadas" para determinada operação, por um especialista nesse trabalho. Se já não existe um programa, uma programação pode ser muito dispendiosa.

O custo do uso de uma máquina é muito elevado se considera

6 Ver o seu Capítulo 15 para uma discussão minuciosa dos processos de tabulação.

22. MÉTODOS DE PESQUISA NAS RELAÇÕES SOCIAIS

mos os preços de horas. No entanto, se existe grande quantidade de dados, que exigem tratamento estatístico complexo, a rapidez da máquina pode compensar, com vantagem, a despesa; em alguns casos, permite análises que, de outro modo, seriam impossíveis. Por exemplo, com algumas centenas de casos, o cálculo de todas as intercorrelações entre uma centena de variáveis⁷ é um trabalho que poderia exigir de um estatístico, usando uma máquina comum de calcular, muitos meses de trabalho; um computador eletrônico, uma vez regulado adequadamente para a operação pode completar o trabalho em aproximadamente uma hora.⁸

A rapidez, mesmo das máquinas mais simples, apresenta a tentação de "ficar doido" e realizar a tabulação cruzada de cada variável com cada uma das outras, na esperança de encontrar alguma relação, em vez de planejar a análise, através da consideração do provável valor de cada operação. Esse processo é extremamente indesejável. Mesmo sem falar da despesa, provoca alguns erros que discutiremos mais adiante neste capítulo (página 474).