

BABBIE, Earl. **Métodos de Pesquisas de Survey**. Belo Horizonte: Editora UFMG, 2001, 519 p.

Resumo realizado, conjuntamente, por Carlos Milani, Karine Oliveira, Naiana Guedes, Rafael Portinho e Sheila Cunha

Este resumo visa a apresentar, sinteticamente, as idéias e experiências do Professor Earl Babbie, acerca do método de pesquisa de *survey*. Este método é uma das ferramentas utilizadas no projeto de pesquisa sobre as relações entre capital social e desenvolvimento local na Bahia. Daí o interesse em realizar este resumo. Foram, para este fim, selecionados alguns capítulos do livro (capítulos 3, 13, 14, 15, 16, 17, 18 e 19), sobretudo aqueles que dizem respeito à análise dos dados, uma vez que, no momento atual do projeto, os questionários (para indivíduos e para Organizações) já foram concebidos, testados e aplicados. O resumo traz uma breve definição do método, bem como algumas técnicas que são utilizadas na condução do projeto de pesquisa, sobretudo no caso de Pintadas.

PESQUISA DE SURVEY COMO MÉTODO DAS CIÊNCIAS SOCIAIS

Surveys são muito semelhantes a censos, mas deles se diferenciam porque examinam somente uma amostra da população (enquanto o censo geralmente implica uma enumeração da população toda). *Surveys* são muito utilizados em pesquisas políticas (para avaliação da intenção de voto, por exemplo). Um dos primeiros usos políticos do *surveys* foi feito em 1880, quando Marx enviou pelo correio 25 mil questionários a trabalhadores franceses a fim de averiguar de que forma eram explorados pelos seus patrões. Weber também teria utilizado métodos de *survey* para o seu estudo sobre a ética protestante. No século XX, o *US Bureau of Census* fez importantes contribuições aos campos de amostragem e coleta de dados. Firms comerciais também desenvolveram pesquisas de opinião (Gallup, por exemplo). Algumas universidades norte-americanas também contribuíram para o desenvolvimento de *surveys* (sobretudo os professores Samuel Stouffer e Paul Lazarsfeld). Paul Lazarsfeld, por exemplo, projetou e realizou estudos de painéis, reentrevistando um grupo de respondentes em ocasiões diferentes durante uma campanha política, acompanhando assim mudanças nas intenções de voto ao longo do tempo. Além disso, Paul Lazarsfeld ajudou a criar o *Bureau for Applied Social Research* na Universidade Columbia (Nova Iorque).

Características gerais da pesquisa de *survey*:

- Os dados facilitam a aplicação cuidadosa do pensamento lógico;
- Sempre que o pesquisador procura explicar as razões para e as fontes de eventos, características e correlações observados, a pesquisa deve assumir uma postura determinística baseada nas relações de causa e efeito;
- *Surveys* amostrais são realizados para entender-se a população maior da qual a amostra foi inicialmente selecionada. Análises explicativas em pesquisas de *survey* visam a desenvolver proposições gerais sobre o comportamento humano;
- Replicar um achado entre subgrupos diferentes (por sexo, por idade, etc.) fortalece a certeza de que ele representa um fenômeno geral na sociedade;

- Com a pesquisa, busca-se o máximo de compreensão com o menor número de variáveis possível;
- A conceituação e a medição de variáveis estão no âmago da prática científica. Se as variáveis não são adequadamente conceituadas e medidas, correlações observadas entre elas podem não fazer sentido. Assim, se observadores independentes discordam sobre como variáveis foram medidas, podem discordar logicamente da conclusão geral;
- Nunca esquecer que o ato de medir é um dos problemas da pesquisa em *survey*. A simples presença de pesquisadores pode afetar os entrevistados. Pedir uma opinião pode cristalizar uma opinião que existia. Alguns respondentes podem formar opiniões na hora. Por isso, é necessário que o cientista social tenha uma compreensão mais sofisticada da medição e da conceituação.

AS LÓGICAS DA MEDIÇÃO E DA ASSOCIAÇÃO

O *survey* reparte-se entre a descrição e a explicação. Os analistas do *survey* medem variáveis e depois as associações entre elas. O capítulo 13 do livro distingue a “imagem tradicional” do método científico com a “imagem alternativa” (mais concreta - ciência na prática), considerando a noção de Paul Lazarsfeld sobre a "interpermutabilidade de índices".

A IMAGEM TRADICIONAL - baseia-se numa seqüência de passos:

1. Construção da Teoria - dedução de uma teoria abstrata, construção teórica a partir de uma teoria sociológica existente (relacionamento de conceitos para formar uma teoria).
2. Derivação das hipóteses teóricas - com base na teoria abstrata, derivam-se hipóteses teóricas.
3. Operacionalização de conceitos - especificação de indicadores empíricos - representantes dos conceitos teóricos. Os indicadores são específicos, enquanto que os conceitos são abstratos. A operacionalização é a especificação desses conceitos na pesquisa - converter a hipótese teórica em empírica.
4. Coleta de Dados Empíricos - coleta de dados relacionados aos indicadores empíricos. Aplicação de questionário.
5. Teste Empírico de Hipóteses - teste estatístico da hipótese. Averiguar se há correlação entre as variáveis para confirmar ou não a hipótese empírica e, por conseguinte, saber se a hipótese teórica é aceita ou rejeitada.

A hipótese na imagem tradicional é representada pela função $Y = f(X)$, onde os valores de Y são determinados (causados) pelos valores de X.

Há dois problemas básicos na imagem tradicional da ciência:

1. Os conceitos teóricos raramente permitem operacionalizações não-ambíguas, ou seja, os indicadores não são absolutamente fiéis ao conceito representado

- imperfeição dos indicadores (imprecisão empírica). A busca de melhores indicadores é infinita.
- 2. As associações empíricas entre as variáveis quase nunca são perfeitas. O uso de indicadores obtém associações parciais. "Quase todas as variáveis são empiricamente relacionadas entre si 'em algum grau'. Especificar o 'grau' que representa a aceitação da hipótese e o 'grau' que representa sua rejeição é arbitrário" (p. 330).

O autor critica a posição determinista da imagem tradicional da ciência afirmando que a medição e associação são conceitos inter-relacionados – trabalhados simultaneamente e que não existem passos fixos e ordenados.

Diz, ainda que o subitem "A intercambialidade de Índices", discutido por Paul Lazarsfeld, trata justamente sobre essa inter-relação entre medição e associação, tentando trazer uma solução parcial dos dois problemas acima citados. Essa discussão de Lazarsfeld reconhece que qualquer conceito tem vários indicadores possíveis, ou seja, na função $Y = f(X)$, o "X" é $x_1, x_2, x_3...$ A hipótese teórica é aceita como uma proposição geral se for confirmada por *todos* os testes empíricos específicos. Essa intercambialidade de índices, cujas associações podem ser paradoxais, favorece a uma compreensão mais abrangente da natureza da associação. Muda-se a perspectiva da pergunta: em vez de perguntar se as variáveis estão associadas, pergunta-se *como* estão associadas.

CONSTRUINDO E COMPREENDENDO TABELAS

Quando do estudo de *survey*, temos a possibilidade de trabalhar com alguns modos de análise:

UNIVARIADA (*utilização de apenas uma variável*): Descreve a amostra do *survey* e a população "trabalhada"

a) Distribuições:

- Listar atributos individualmente
- Listar o nº de "entes" em cada faixa desejada
- Listar as frequências dos dados: escolher a base (=100%) – nº total dos respondentes ou o nº de pessoas que responderam (não inclui os que não responderam)

b) Tendência central (Média):

- Moda: valor mais frequente; maior nº de "entes" que apresentam o dado que se pretende estudar
- Mediana: valor do "meio" – dados exatos ou agrupados
- Média aritmética: (dado X nº de "entes"); nº total de "entes"

c) Dispersão (compensar o cálculo do valor central):

- Amplitude: distância entre o valor mais alto e o mais baixo
- Desvio Padrão: Capítulo 5

- Desvio quartílico (ou amplitude semi-interquartílica): (valor mais alto do quartil mais alto – valor mais baixo do quartil mais baixo): 2
- d) Variáveis contínuas: passam continuamente de um ponto a outro
- e) Variáveis discretas: passam bruscamente de uma categoria a outra
- f) Descrição de subgrupos: representam 100% (amostra total); razões descritivas; razões comparativas (efeito causal sobre a variável de descrição):
Fundindo categorias de respostas: quando é pequena a porcentagem de entrevistados (as respostas extremas): somar das frequências brutas, recomputar as porcentagens para evitar conclusões errôneas e formular justificativas para tais questões
Os “não sei”: podem confundir os resultados se formados por uma porcentagem significativa; devem ser informados os números para que o leitor tire suas conclusões.

BIVARIADA (associação de duas variáveis): Objetiva a predição, além de temas explicativos

- a) Conhecer a lógica das variáveis independentes e dependentes; explicar valores da variável dep. em função dos valores da variáveis independentes.
- Ordem temporal: variável independente “vem” antes
 - Ordem quase-temporal: bases lógicas
- b) Construção de tabelas (percentuadas arbitrariamente)
- Amostra dividida em valores (subgrupos) da(s) variável(eis) independente(s)
 - Subgrupos em termos dos valores das variáveis dependentes
 - Subgrupos da variável independente comparados em função da var. dep.
 - Tabela percentuada de forma a não comprometer a interpretação
 - Tabelas bivariadas (ou tabelas de contingência ou tabulações cruzadas):
 - i) Cabeçalho e títulos
 - ii) Conteúdo, valores e categorias claramente expostos (as respostas devem vir completas no relato do texto)
 - iii) Indicação da base das porcentagens
 - iv) Não há necessidade dos n^os brutos
 - v) Indicação das omissões
- c) Multivariada (Capítulo 15: associação simultânea de diversas variáveis)
- Existe mais de uma variável independente (podem ter efeito cumulativo)
 - Variável dependente explicada em função de mais de uma variável independente
 - i) Dividir a amostra total em subgrupos de acordo com as variáveis independentes
 - ii) Subgrupos descritos em função da variável dependente

- iii) Comparação entre os subgrupos
- iv) A apresentação de tabelas com intersecções entre diversas variáveis independentes pode facilitar a interpretação dos dados.

O MODELO DE ELABORAÇÃO

Também conhecido como “método de interpretação” e “método de Lazarsfeld”, esse modelo tem o intuito de elaborar uma relação empírica entre duas variáveis por meio da inserção simultânea e controlada de variáveis adicionais, fornecendo uma análise mais apropriada do *survey*.

O modelo de elaboração encontra suas raízes na II Guerra Mundial, quando Samuel Stouffer realizou uma pesquisa social no exército dos Estados Unidos, objetivando relacionar quais fatores afetavam a efetividade dos soldados em combate. Encontrando anomalias em seus dados, Stouffer utilizou os conceitos de “grupo de referência” e de “privação relativa”, que indicam tendências dos indivíduos em opinarem de acordo com “seus pares”.

Formalmente o modelo de elaboração foi desenvolvido por Paul Lazarsfeld, que juntamente com Patrícia Kendall fez a revisão metodológica dos estudos de Stouffer.

Para o efetivo entendimento de uma análise de elaboração devemos verificar se a variável de teste (ou de controle) antecede às outras duas variáveis ou se é interveniente (afeta) a estas. Dessa forma teremos diferentes relações lógicas no modelo multivariado. O paradigma da elaboração é, assim, constituído pela seguinte lógica:

- Observamos empiricamente duas variáveis;
- Verificamos a relação original entre as variáveis, mantendo as demais variáveis constantes e atentando para as possíveis relações parciais;

Daí teremos quatro possibilidades:

- Se a relação original se repetir nas relações parciais, ela foi replicada: a relação existente entre as duas variáveis é genuína e generalizável;
- Se somente em uma das relações parciais tivermos a repetição da relação original, teremos uma especificação: uma variável causa outra;
- Se a relação original não é demonstrada nas relações parciais e a variável de controle é antecedente às outras duas variáveis, o resultado é de explicação: anula-se uma relação não-genuína;
- Se a relação original não é demonstrada nas relações parciais e a variável de controle é interveniente junto às outras duas variáveis, tem-se uma interpretação: tem-se os meios pelos quais uma variável é causa da outra.

É importante atentarmos que o próprio pesquisador deve decidir arbitrariamente sobre o fato de uma relação parcial ser ou não mais fraca (ou mais forte) que a original. Da mesma forma deve-se ter cuidado quando se usa mais de uma variável de teste devido ao nível de complexidade na análise do *survey*.

ESTATÍSTICA SOCIAL

A matemática e a estatística são utilizadas para trabalhar dados usados em pesquisas empíricas, tornando possível uma análise eficaz destes. Dentre alguns tipos de estatística, estão a descritiva e a inferencial.

ESTATÍSTICA DESCRITIVA

Possibilita a apresentação de dados quantitativos de forma manejável, viabilizando a descrição das variáveis isoladamente ou a partir de associações, através da “redução de dados” e das “medidas de associação”, dentre demais técnicas. A redução de dados facilita a gestão dos mesmos, através de cálculos que permitam a diminuição da quantidade de valores obtidos em pesquisas quantitativas. As formas utilizadas para realizar tal procedimento englobam a elaboração de matrizes, categorias e códigos, medidas de tendência como moda, mediana e média, além de medidas de dispersão, como desvio padrão. No caso específico do resumo de medidas de associação entre duas variáveis, podem-se verificar alternativas baseadas no *modelo de redução proporcional de erro* (RPE), que se adapta a diferentes níveis de medida dentre as formas nominal, ordinal ou intervalo. São elas:

Variáveis nominais - dentre as medidas utilizadas, encontra-se a *lambda*, que prevê os valores de uma das variáveis a partir dos valores conhecidos de uma outra variável. A redução de erros remete à proporção de erros que seriam cometidos baseando-se na distribuição geral.

Variáveis Ordinais - neste caso, pode-se utilizar a medida *gamma* (γ), que funciona realizando previsões de valores para determinada variável, conhecendo os valores de outra, a partir do arranjo ordinal de valores. Os valores de uma variável correspondem aos valores a serem obtidos por uma outra variável positivamente ou negativamente, baseando-se em comparações emparelhadas.

Variáveis de Intervalo ou de Razão - Uma medida de associação adequada para esse tipo de variável é a correlação *produto-momento de Pearson* (r), que adivinha o valor de uma variável conhecendo os de uma outra, refletindo o quão exata é essa previsão. O cálculo é baseado na *variação total*.

ESTATÍSTICA INFERENCIAL

Sabe-se que possibilita a realização de inferências a partir de achados univariados e multivariados amostrais, e por isso podem ser Inferências Univariadas ou Inferências Multivariadas,

Inferências Univariadas - observa-se que as mesmas medidas de resumo usadas para descrever a amostra estruturada são aproveitadas para fazer afirmativas mais amplas sobre a população. Há de se reforçar que o erro padrão é essencial na estimativa do erro amostral, e uma afirmação sobre o erro amostral necessita o nível de confiança e o intervalo de confiança. Para que seja possível realizar inferências:

- a) A amostra deve ser retirada da população sobre a qual se está inferindo;
- b) Com relação aos tipos de amostragem utilizáveis, pode-se verificar que a estatística inferencial supõe a utilização da aleatória simples; o fato da amostragem com reposição quase não ser usada, mesmo que seu uso seja suposto pela estatística, não representa um problema; a sistemática é mal usada que a aleatória, mas também não há nisso problema sério; a estratificada, por aumentar a representatividade, da mesma forma não resulta em um problema; a amostragem por conglomerado pode representar um problema devido às pequenas estimativas de erro amostral; e, por fim, a amostragem de esquina já não possibilita a estatística inferencial;
- c) A estatística inferencial não considera “erros não-amostrais”, porque trata apenas do erro de amostragem.

Inferências Multivariadas - podem ser visualizadas dentro dos Testes de Significância.

Através dos Testes de Significância Estatística, em caso de não ocorrer associações entre as variáveis na população maior, é possível estimar a probabilidade de uma associação resultar de erro de amostragem normal. Os testes realizam suposições por *surveys* reais, mas também são utilizáveis ao analisar e interpretar dados. É importante ressaltar que de relações observadas em dados colhidos de populações inteiras não se devem ter sua significância estatística calculada, pois esta se refere à probabilidade de as relações entre as variáveis serem produto de erro de amostragem. Se as suposições de amostragem usadas na computação dos intervalos de confiança não se enquadrarem no desenho da amostra, não há legitimidade nos testes de significância, dos quais também constituem uma base.

TÉCNICAS MULTIVARIADAS AVANÇADAS

As técnicas multivariadas avançadas, caracterizam-se pela complexidade na análise e apresentação de dados. Serão apresentadas aqui, algumas técnicas sob a perspectiva do modelo de elaboração. São elas: análise de regressão, análise de trajetória, análise fatorial, análise de variância, análise discriminante e modelos lineares.

ANÁLISE DE REGRESSÃO

Técnica para estabelecer a equação de regressão representando a linha geométrica que mais se aproxima da distribuição de pontos. Esta equação é valiosa tanto descritiva como inferencialmente. Ela faz uma descrição matemática entre as variáveis e permite inferir os valores de Y quando temos os valores de X em outras palavras, determina a função específica relacionando Y com X.

Ex.: $Y = f(x)$. Significa que valores de Y podem ser explicados em termos de variações nos valores de X.

Formas de análise de regressão

Regressão linear - associação linear perfeita entre duas variáveis. A relação entre duas variáveis é descrita pela equação $X=Y$ (equação de regressão) . Traçando-se uma linha sobre o encontro dessas duas variáveis esta formará uma linha reta que representa com a menor quantidade de discrepância a localização real dos pontos num gráfico (linha de regressão). Temos, então, para a equação de regressão linear simples a forma $Y= a+bx$, onde Y é o valor (estimado) da variável dependente, a é o valor constante e b é outro valor numérico, que é multiplicado por X , o valor da variável independente .

Tem importante uso descritivo e valor inferencial, uma vez que pode ser usada para prever outros números de valores (valor de $X= 3,5$ logo valor de $Y =3,5$)

Regressão múltipla - meio utilizado para analisar situações em que uma variável dependente é afetada simultaneamente por diversas variáveis independentes.

Regressão parcial – examina os efeitos de diversas variáveis independentes, com o efeito de cada uma delas expresso separadamente, quando os efeitos das demais são mantidos constantes.

Regressão curvilínea – permite que a linha de regressão não seja reta ,desta forma os valores das variáveis independentes podem ser elevados a potências maiores que 1 (ao quadrado ou ao cubo etc). Em alguns casos Eça pode propiciar melhor compreensão das relações empíricas que qualquer modelo linear.

Precauções com a análise de regressão: deve-se ter cuidado na avaliação dos resultados nas análises de regressão pois esta se baseia em amostragem aleatória simples, ausência de erros não mostrais e dados de intervalos contínuos que podem levar a extrapolação (estimando casos além da faixa de observações).

ANÁLISE DE TRAJETÓRIA

Apresenta graficamente as relações causais entre variáveis, baseia-se na análise de regressão e supõe que os valores de uma variável (dependente) são causados pelos valores de outra variável (independente).

ANÁLISE FATORIAL

È usada para descobrir padrões de variações nos valores de diversas variáveis. Por meio de um computador, são criadas dimensões artificiais (fatores) que se correlacionam altamente com diversas das variáveis reais.

ANÁLISE DE VARIÂNCIA

Os casos são estudados em grupos representando uma variável independente, e o grau de diferenciação entre os grupos é analisado em termos de uma variável dependente. Pretende determinar se diferenças entre grupos pode ter ocorrido razoavelmente em amostragem aleatória simples ou se representam uma relação pura entre variáveis envolvidas.

ANÁLISE DISCRIMINANTE

Através de uma dimensão hipotética que separa duas categorias da variável dependente procura explicar a sua variação desta. Com a atribuição de escores às pessoas, com base na dimensão já citada, permite prever seus valores na variável dependente.

ANÁLISE LOG-LINEARES

Técnica baseada na especificação de modelos que descrevem as inter-relações complexas entre variáveis nominais com mais de dois atributos cada e na comparação entre frequências esperadas e observadas de células de tabela. Envolve cálculos logarítmicos bem elaborados.

O RELATÓRIO DA PESQUISA DE SURVEY

Regras básicas na preparação de um relatório de pesquisa de *survey*:

- Excelente conhecimento da língua!
- Pensar no destinatário e na ampla gama de discursos possíveis em função do destinatário do relatório (quem lerá o relatório?);
- Comunicar um corpo de idéias e dados;
- O relatório científico deve ser visto como uma contribuição ao conhecimento científico em geral;
- O relatório deve estimular e orientar futuros trabalhos.

Tipos de relatório:

- Alguns relatórios podem focalizar a exploração de um tópico. Neste caso, o relatório mostra o caminho para pesquisas mais refinadas sobre o assunto. A pesquisa exploratória visa a desenvolver, esclarecer e modificar conceitos e idéias;
- O relatório pode ter o objetivo descritivo. Neste caso, é importante distinguir a descrição que se aplica apenas à amostra daquela que se aplica, por inferência, à população em geral. No caso das inferências, deve-se explicitar qual a margem de erro;
- Muitos relatórios têm o objetivo explicativo: aqui se mostram relações de causa e efeito entre as variáveis;
- Alguns relatórios (sobretudo aqueles que se destinam a responsáveis políticos ou gestores organizacionais) podem ter o objetivo de propor ação.