

Análise de dados

Imagine uma pesquisa que tenha feito 30 registros de observação de crianças, de meia hora cada um deles, onde estão por escrito os inúmeros comportamentos e atitudes observados nessas crianças. Imagine ainda outra pesquisa com 120 questionários preenchidos de 25 questões em cada um ou então 50 entrevistas com 20 questões abertas em cada uma. É um enorme conjunto de dados, não?

O pesquisador poderia ir lendo cada um dos registros (questionários e entrevistas), para ir extraindo suas conclusões. Mas é bem provável que, ao ler o último, já tenha se esquecido dos conteúdos principais do primeiro. Neste sentido é que se propõe a etapa de *Análise de Dados*, onde o pesquisador deverá organizar os dados chamados brutos, tornando-os legíveis, prontos para oferecerem uma explicação, uma resposta do problema de pesquisa.

Organizar os dados envolve, basicamente, quatro etapas: rever as categorias estabelecidas na previsão de análise, tabular os dados, representá-los em tabelas ou figuras e descrever as relações entre os dados.

5.1. Rever as categorias da previsão de análise

Lembre-se que, antes de iniciar a coleta de dados, o pesquisador estabelecia quais seriam as possíveis relações entre as variáveis em estudo (p. 19) e, em seguida, estabelecia categorias para os valores que as variáveis poderiam assumir (p. 20).

Após coletar os dados referentes a essas variáveis, seus valores e relações, seja através dos questionários, entrevistas ou observação,

o pesquisador deverá rever as categorias previstas em função dos dados obtidos, e manter, eliminar, modificar ou criar novas categorias.¹

Suponha, por exemplo, uma pesquisa que investigue sobre a procedência de rendas familiares (utilizando questionários) e tenha estabelecido apenas duas categorias: provenientes do “cabeça-do-casal”; provenientes de outras fontes, externas aos membros da família.

Ao fazer uma questão aberta para obter esse dado, o pesquisador percebe que “cabeça-do-casal” é um termo ambíguo e obsoleto: muitos casais faziam piadas, “brincavam” sobre quem seria o “cabeça-do-casal”, e os dados mostravam que a renda familiar ora era procedente de ambos (marido e mulher), ora de um deles não considerado tradicionalmente o “cabeça-do-casal”, ou ainda daquele considerado tradicionalmente o “cabeça-do-casal”. Assim, os dados revelaram que, além das categorias serem insuficientes, uma delas era pouco clara. Os dados sempre são o voto de Minerva para o julgamento do pesquisador acerca da adequação de suas categorias.

5.2. Tabular os dados

Tabular significa *contar* o número de vezes que apareceu uma determinada categoria (quando se trata de uma análise quantitativa; trataremos mais adiante da análise qualitativa).

Um estudo observacional sobre interação mãe-criança elabora, por exemplo, as seguintes categorias de observação (Ramos, 1979) quanto ao desempenho verbal da mãe e criança (tipo de verbalização):

- declaração;
- nomeação;
- perguntas que exigem respostas verbais;
- ordens que exigem respostas verbais;
- vocativos;
- exclamações;
- negação.

A tabulação do número de vezes que tais categorias apareceram poderia ser feita para cada sessão de observação e poderia ser apresentada em quadros de tabulação, tal como o apresentado a seguir:

¹ Quando se trata do método de observação, é preciso realizar observações cursivas (corridas, não-categorizadas) antes de se estabelecer categorias.

| Tipos de respostas verbais | Sessões | | |
|----------------------------|----------|----------|----------|
| | sessão 1 | sessão 2 | sessão 3 |
| — declaração | ☑ | ☑☑☑ | ☑☑ |
| — nomeação | ☑☑☑☑ | ☑☑ | ☑☑☑☑ |

Assim, com a tabulação feita, tem-se a idéia da frequência de aparecimento dos dados coletados, e também uma idéia geral da maneira como eles respondem ao problema de pesquisa.

5.3. Representação dos dados em figuras e tabelas

Para uma melhor visualização dos dados coletados e já tabulados, isto é, para que se possa “enxergar” rapidamente que tendência, que porção do total de dados cada categoria ocupa, os pesquisadores criaram recursos de representação denominados tabelas e figuras.

Você já deve ter visto muitas tabelas e figuras em artigos de revistas e jornais, e é bem provável que, sem os mesmos, você não tivesse compreendido muito bem o texto.

A revista *Veja*, de 08.07.81, publicando um artigo sobre previsões para as eleições de 82, a partir de uma pesquisa *Veja-Gallup*, apresenta várias “tabelas” que ajudam na visualização do conteúdo a que o texto se refere. Vejamos o texto primeiramente:

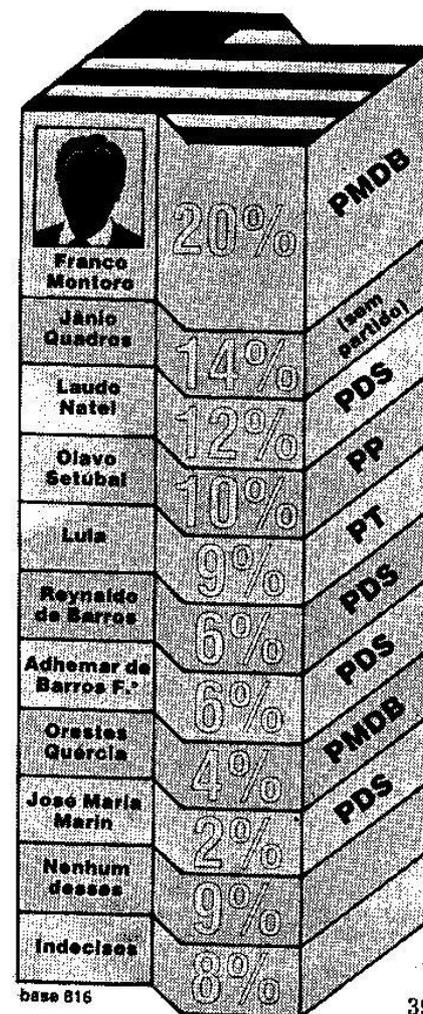
“Ao contrário do que afirma em sucessivas entrevistas, o ex-presidente Jânio Quadros está longe de ser o candidato imbatível dos anos 50: aparentemente arranhado por mais uma renúncia — desta vez, ao PTB — ele não está nos calcanhares do senador Franco Montoro, como sugeriam sondagens de opinião feitas à época do seu barulhento reingresso na cena política. Bem votado no interior, o ex-governador Laudo Natel corre alguns corpos à frente de seus concorrentes do PDS, mas dificilmente terá chances de vitória sem a ajuda das sublegendas. Uma surpresa é o prefeito de São Paulo, Reynaldo de Barros, empatado com seu primo Adhemar de Barros Filho — com quem disputa o espólio eleitoral do ‘ademarismo’ — e bem acima do vice-governador José Maria Marin, que está em campanha há alguns meses e utiliza regularmente a máquina administrativa estadual.

Como Laudo, também Jânio e Montoro têm mais popularidade no interior do que na capital, uma tendência que se inverte nos casos de Reynaldo de Barros, do ex-prefeito Olavo Setúbal e do ex-dirigente sindical Luís Inácio da Silva, o ‘Lula’, outra surpresa registrada pela pesquisa. O índice que alcançou não lhe permite sonhar com o governo estadual. Mas melhora sensivelmente o cacife que poderá levar à mesa de negociações para a qual

Montoro, interessado no apoio do PT, já há algum tempo tenta atraí-lo. Tal cacife decididamente falta ao senador Orestes Quêrcia, que, embora imagine possível uma reedição da virada, que o levou ao Congresso em 1974, estaciona num índice que só lhe permite aspirar à conquista de uma sublegenda do PMDB.” (*Veja*, 670: 15, 1981).

Logo a seguir, no artigo, os autores apresentam a “tabela”. Observem como fica mais fácil a visualização do conteúdo anteriormente colocado:

SAO PAULO
11 159 018 eleitores



Veja, 670: 15, 1981.

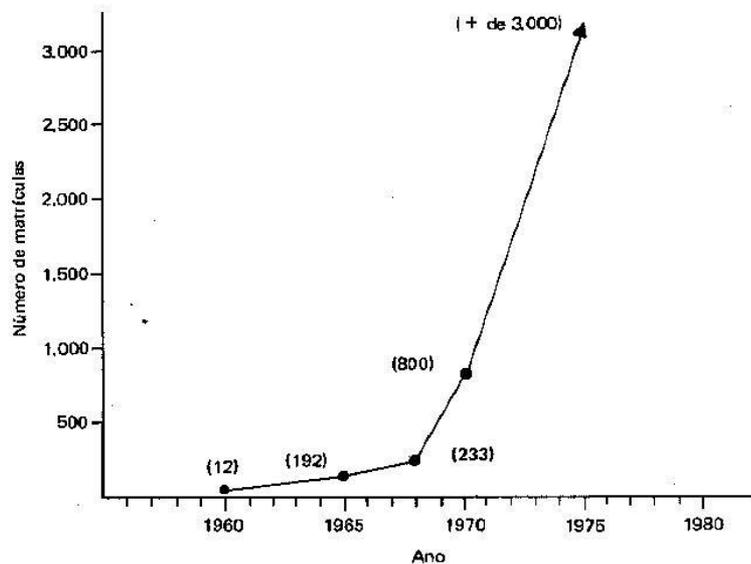


Figura 5.1. Número de matrículas nos Cursos de Psicologia do Estado de São Paulo de 1960 a 1975. (Fontes: Pastore, 1972; Mello, 1971; MEC, 1968; Secretaria da Educação e Cultura, 1975.)

Há várias formas de representação. Uma das mais freqüentes é denominada *gráfico em curva* (Rocha, 1971). Seus componentes são: 2 eixos, um vertical, chamado ordenada, e um horizontal, chamado abscissa. Esses eixos representam as duas variáveis que estão sendo relacionadas. É muito importante que se obedeça a uma proporção entre a ordenada e a abscissa para não se "achatar" demais a linha ou "encolhê-la" também demais. O que se sugere é que a ordenada seja 75% ou 2/3 do tamanho da abscissa.

Vejamos um exemplo.² No gráfico da figura 5.1 temos "número de matrículas" nos cursos de Psicologia do Estado de São Paulo na *ordenada* e, na *abscissa*, temos os anos (de 1960 a 1980).

Uma outra maneira de representar esses dados seria o *histograma*: ao invés de pontos, colocaríamos retângulos justapostos — veja a figura 5.2.

² Por sua simplicidade de apresentação e relevância do tema apresentado, todas as figuras e tabelas analisadas nesta seção foram extraídas do artigo de Botomé, S. A quem nós, Psicólogos, servimos, de fato? *Psicologia*, p. 1-15, 1979.

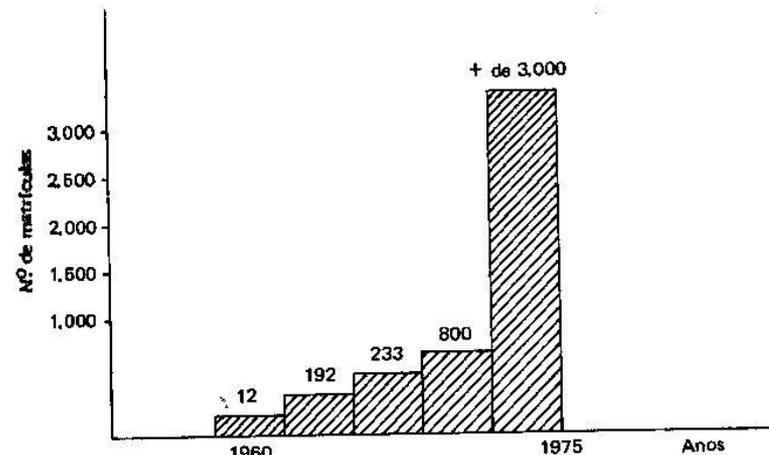


Figura 5.2.

Tanto o histograma como o gráfico em curva servem para representar variáveis ditas contínuas, isto é, sua medida é seqüenciada e tem uma ordem de ocorrência, uma hierarquia (por exemplo: anos, idade, horas, altura etc.) (Kerlinger, 1975).

Quando se trata de variáveis não contínuas ou *categorias nominais* (Rocha, 1971), representam-se os dados em figuras denominadas *diagramas em colunas*, onde os retângulos não são justapostos, porque não há uma seqüência entre os valores da variável representada. Vejamos o exemplo da figura 5.3.

A variável "tipo de serviços em Psicologia" é categórica, isto é, não apresenta uma ordem hierárquica do tipo "maior que" ou "menor que" entre os seus valores, e, por isso, não pode ser representada em um histograma ou diagrama de barras. Deve ser representada em um diagrama de colunas.

Uma terceira maneira de se representar dados é pela construção de *tabelas*. Quando são muitas as curvas e/ou categorias a serem representadas em um gráfico, o que dificultaria sua leitura, opta-se pela tabela, embora as tendências e flutuações de variáveis sejam visualizadas mais facilmente através dos gráficos.

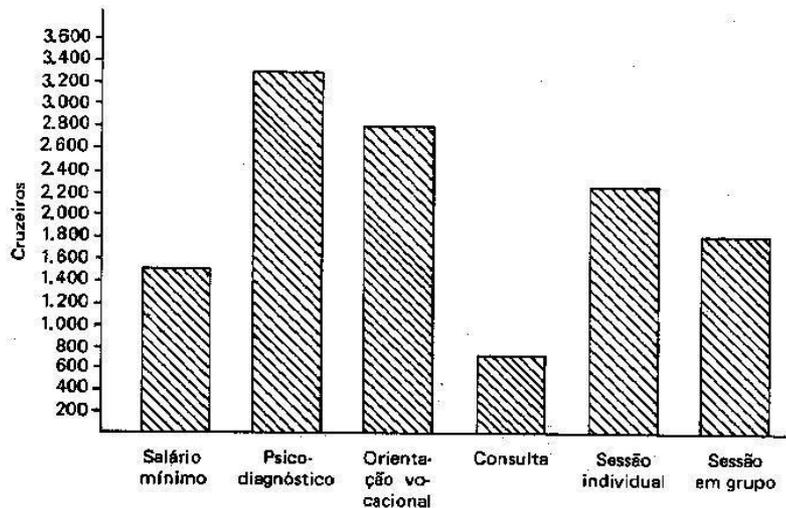


Figura 5.3. Média dos preços dos serviços de psicologia em relação ao salário mínimo no 2.º semestre de 1978, em São Paulo. Os preços foram calculados considerando 40 por cento sobre os preços de 1977, publicados pelo Sindicato dos Psicólogos de São Paulo. Os preços de "sessão individual" e "sessão em grupo" foram calculados considerando quatro sessões por mês — o preço na figura é mensal. (Fonte: *Boletim do Sindicato dos Psicólogos de São Paulo*, 2.º semestre de 1977.)

Uma tabela apresenta os seguintes componentes (acompanhe na tabela 5.1):

a) *título*: indica quais os dados que a tabela contém, a natureza do dado e a unidade numérica empregada. Coloca-se na parte superior da tabela. Na tabela a seguir (extraída de Botomé, 1979), o título é: Preços dos serviços de psicologia na cidade de São Paulo, conforme Boletim Informativo do Sindicato dos Psicólogos de São Paulo (1977).

b) *linha*: é a parte da tabela que contém as informações em horizontal. Na tabela 5.1 a linha é "Serviços: Orientação Vocacional, Psicodiagnóstico, Consulta, Sessão Individual, Sessão em Grupo".

c) *coluna*: é a parte que contém as informações na vertical. Na tabela 5.1 a coluna é: Preços: Média e Variação.

d) *casa, célula ou casela*: é a parte da tabela formada pelo cruzamento de uma linha com uma coluna. Na tabela 5.1 as casas

Tabela 5.1. Preço dos serviços de psicologia na cidade de São Paulo. Modificado do *Boletim Informativo do Sindicato dos Psicólogos de São Paulo* (1977).

| Preços | Média | Variação |
|-----------------------|----------|-------------------------|
| Serviços | | |
| Orientação Vocacional | 2.000,00 | 300,00 a 4.500,00 |
| Psicodiagnóstico | 2.300,00 | 500,00 a 3.500,00 |
| Consulta | 400,00 | — |
| Sessão Individual | 400,00 | — |
| Sessão em Grupo | 300,00 | — |

são as médias de preço e variações em cada um dos tipos de serviços (2.000,00, 300,00 a 4.500,00 etc).

Há, ainda, a fonte da tabela, isto é, de onde foram extraídos os dados. No caso da tabela que estamos analisando, a fonte já está explicitada no próprio título: *Boletim Informativo do Sindicato dos Psicólogos de São Paulo*, 1977.

Essas representações de dados que aqui apresentamos são as mais freqüentemente utilizadas pelos pesquisadores. No entanto, criatividade é uma qualidade bem-vinda em ciência e muitas outras formas de representação gráfica podem ser criadas.

Qualquer que seja a descrição, ela deve obedecer aos seguintes critérios:

- simplicidade — possibilitar uma análise rápida do fenômeno representado na figura ou tabela;
- clareza;
- precisão.

5.4. Descrição das figuras e tabelas

Embora as figuras e tabelas tenham significação própria, independente de um texto que as descreva, o pesquisador deverá ressaltar as informações mais importantes que representam considerando o problema de pesquisa e a previsão de análise e excluindo detalhes.

A seguir você encontra a descrição da figura 5.3.

"Na figura 5.3 podemos ver a relação entre o preço dos serviços dos psicólogos e o salário mínimo em São Paulo. Uma Orientação Vocacional ou um Psicodiagnóstico custam, aproximadamente, dois salários mínimos. Uma terapia individual custará um salário mínimo e meio por mês, e em grupo mais de um salário por mês. Uma consulta isolada custará mais de 30 por cento do salário mínimo" (Botomé, 1979).

A atividade final da etapa de Análise de Dados consiste em relacionar os dados descritos, isto é, comparar os resultados representados nas diferentes figuras e tabelas. Isto permite resumir as principais conclusões e encaminhar respostas ao problema de pesquisa. Vejamos um exemplo, em que se relaciona a figura 5.1 e a figura 5.4.

"A figura 5.4 mostra uma aceleração bastante grande no número de Cursos de Psicologia existentes nesta década. Isto, além de poder explicar a ocupação de tantos psicólogos com o ensino de psicologia (como vimos na figura 5.1) a partir de 1969, também anuncia um grande aumento no pessoal disponível para prestar serviços de psicologia.

A figura 5.1 mostra este problema com mais clareza ao descrever o aumento na procura dos Cursos de Psicologia através do número de matrículas a cada ano.

Parece que, além de aumentar os Cursos de Psicologia, cada um deles também está aumentando a oferta de vagas" (Botomé, 1979).

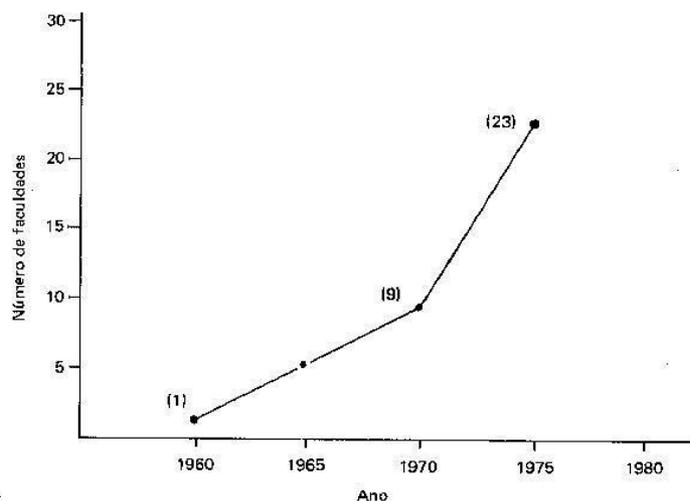


Figura 5.4. Número de Faculdades de Psicologia no Estado de São Paulo, de 1960 a 1975. (Fontes: Pastore, 1972; Mello, 1971; MEC, 1968; Secretaria da Educação e Cultura, 1975.)

* Não encontramos dados sobre 1965.

Tanto as descrições como as relações estabelecidas entre elas devem obedecer aos critérios de linguagem que vimos nas páginas 13 e 33 e serem sempre guiadas pela questão que a pesquisa pretende responder. Afinal, é a única etapa da pesquisa onde os dados ficam disponíveis, de maneira organizada, para que leitores interessados os consultem sempre que o quiserem. Há pesquisadores que consideram a Análise de Dados como a etapa mais estável de todo o processo de pesquisa, no sentido de serem resistentes ao tempo: as interpretações sobre os dados, sua relevância social, seu significado científico podem mudar a qualquer instante, ao passo que os dados dessa pesquisa permanecem sempre os mesmos; inclusive, por sua estabilidade, servem de ponto de partida para eventuais discordâncias com outro pesquisador que interpreta esses dados.

5.5. Análise estatística dos dados

Para a colocação de pontos e valores em figuras e tabelas, o pesquisador pode proceder a algumas medidas, tais como as chamadas "medidas de tendência central" (Kaplan, 1969): média aritmética (ou apenas "média"), mediana etc. Utiliza, portanto, o recurso da *estatística descritiva* (Siegel, 1975), que serve ao propósito de reduzir uma multiplicidade de dados a uma simplicidade manuseável. O que se deseja com a utilização das estatísticas descritivas é um meio simples de caracterizar um conjunto de várias grandezas como um todo, através de uma estimativa única.

Nos estudos de casos, por exemplo, é freqüente o psicólogo clínico observar por muitos anos uma mesma pessoa registrando qualitativa e quantitativamente comportamentos de naturezas diferentes. Em determinado momento, para que seja possível extrair conclusões a respeito do caso, ele sente a necessidade de resumir esses dados, reuni-los, extraindo medidas representativas. Nesse sentido, as médias percentuais extraídas, recursos de estatística descritiva, são extremamente úteis.

Uma outra maneira de se utilizar a estatística em pesquisa é através da *estatística inferencial* (Siegel, 1975). É a parte da estatística que visa a tirar conclusões sobre um grande número de eventos, com base na observação de apenas parte deles. Fornece meios para se formalizar e padronizar os processos para se tirar tais conclusões.

É, em geral, aplicada em pesquisas com um grande número de sujeitos, e, de seus recursos, pesquisadores extraem os seguintes tipos de conclusões: sobre a probabilidade de hipóteses serem verdadeiras

ou falsas, aceitas ou rejeitadas, sobre a probabilidade de uma classe de acontecimentos ocorrerem, sobre a significância de diferenças entre grupos pesquisados etc.

Segundo Siegel (1975), um problema comum de inferência estatística consiste em determinar, em termos de probabilidades, se as diferenças observadas entre duas amostras significam que sejam realmente diferentes entre si as populações submetidas à amostragem.

Trata-se de um assunto bastante complexo e nos currículos de faculdades de Psicologia há matérias de dedicação especial ao estudo de estatística.

Para os propósitos de nosso livro de introdução à pesquisa, é importante apenas que você saiba ser a estatística (tanto a descritiva como a inferencial) um instrumento freqüentemente utilizado para se apresentar os resultados científicos.

E, mais do que apresentar resultados, quando se trata de realizar pesquisas com grandes números de sujeitos, para se extrair conclusões sobre a população maior, a estatística inferencial fornece, inclusive, regras para a coleta de dados.³

5.6. Análise qualitativa e quantitativa dos dados

É muito comum a impressão de que a pesquisa trata de dados quantitativos, numéricos apenas, e que, portanto, não se adequa muito bem às ciências do homem. Tal impressão nos parece infundada e exige uma reflexão.

Kaplan (1969) denomina a dicotomia que se faz entre qualidade e quantidade de "mística da qualidade" e "mística da quantidade". O que, em geral, se pensa é que ambas são antitéticas ou mesmo alternativas, ou que a análise qualitativa seja mais "rica", mais "profunda" que a análise quantitativa, porque descreve o fenômeno em si, abordando seus aspectos mais relevantes. Na realidade, em ambas há um grau de abstração, de subjetividade. "As quantidades são quantidades de qualidades", e não é o fato de ser "qualitativo" ou "quantitativo" que vai decidir sobre a relevância do fenômeno descrito.

Quando se extrai uma medida numérica, tal como a "porcentagem", está se selecionando apenas um aspecto do fenômeno e uma forma de medi-lo. Existe, portanto, uma abstração. Uma escolha.

³ Para estudo, ver Siegel, S. *Estatística Não Paramétrica para as Ciências do Comportamento*. São Paulo, McGraw-Hill do Brasil, 1975.

Vimos, na página 120, como a conclusão sobre a destinação de verbas para a educação muda em função do tipo de medida que se apresenta. Isto significa, portanto, que o fato de um fenômeno ser apresentado sob a forma numérica, com uma análise quantitativa, não implica que ele seja desprovido de "subjetividade", de valores do pesquisador. Por outro lado, quando se extrai uma medida qualitativa não temos também a garantia de que se descreverá o fenômeno mais "profundamente", de maneira mais "rica".

A decisão de que tipo de análise fazer — qualitativa ou quantitativa — não pode ser tomada com base nesses aspectos, a nosso ver, simplistas e errôneos.

São muitas as pesquisas em Psicologia que empregam medidas numéricas e que são consideradas extremamente relevantes pelo tipo de fenômeno que estudam. E há, por sua vez, inúmeros estudos descritivos, com análises qualitativas extensas e bem elaboradas que trouxeram indiscutíveis contribuições para o avanço da Psicologia (Ramos, 1979).

Temos sempre que analisar a relevância de um estudo à luz dos critérios de relevância social e científica vistos nas páginas 11 e 12.

