HAL
open science

Sobre os métodos quantitativos na pesquisa em ciências humanas: riscos e benefícios para o pesquisador,

Jean-Claude Regnier, Jorge da Rocha Falcão

To cite this version:

Jean-Claude Regnier, Jorge da Rocha Falcão. Sobre os métodos quantitativos na pesquisa em ciências humanas: riscos e benefícios para o pesquisador,. Revista Brasileira de Estudos Pedagógicos, 2002, 81 (198), pp.229-243. 10.24109/2176-6681.rbep.81i198.937 . halshs-00363433

HAL Id: halshs-00363433

https://shs.hal.science/halshs-00363433

Submitted on 13 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Régnier, JC, Da Rocha Falcão, J.T., (2002) Sobre os métodos quantitativos na pesquisa em ciências humanas: riscos e benefícios para o pesquisador, *Revista brasileira da Educação*, (198)

SOBRE OS MÉTODOS QUANTITATIVOS NA PESQUISA EM CIÊNCIAS HUMANAS: riscos e benefícios para o pesquisador[1]

Jean-Claude Régnier
Université Lumière Lyon-II - FRANÇA

Jorge Tarcísio da Rocha Falcão
Universidade Federal de Pernambuco - BRASIL

> A observação científica é sempre uma observação polêmica; ela confirma ou invalida uma tese anterior, um esquema prévio, um plano de observação; ela mostra demonstrando; hierarquiza as aparências; transcende o imediato; reconstrói o real após haver reconstruído seus esquemas.
>
> **Gaston Bachelard**

Resumo
O presente artigo se propõe a refletir criticamente acerca do uso da *quantificação* em pesquisas em ciências humanas, chamando a atenção para alguns riscos que permeiam tal atividade, como também para determinadas situações em que tais métodos podem ser caracterizados como *amplificadores culturais* úteis e justificados. Nesse sentido, analisam-se aspectos inerentes à atividade de *quantificação*, como a interconexão estreita com a categorização e a modelização fundada em determinada visão teórica. Busca-se ainda explorar aspectos conceituais centrais à quantificação, como a redução de informação em prol da detecção de tendências gerais e estruturas, as opções de tal redução em função das escalas de medida mobilizadas, a distinção entre a descrição e a inferência, bem como entre o fortuito e o sistemático. Procura-se demonstrar, em relação a este último aspecto, que a quantificação possibilita apenas uma aproximação probabilística, o que limita seu poder à detecção da verossimilhança, e não da verdade.

Palavras-chave: análise de dados, mensuração, quantificação.

About so-called quantitative research methods in human sciences: risks and gains for the researcher

Abstract
This paper is aimed to discuss critically quantification as a methodological initiative in human sciences, focusing on important risks inherent to this activity, but also stressing its utility as a social tool and cultural amplifier in the context of specific situations. In this context, basic aspects concerning quantification are analyzed, particularly those referring to its connections with theoretically guided activities like modeling and categorization. Additionally, some important issues in the activity of quantification itself are described and analyzed: the search for general structures and tendencies in a set of data, measurement scales, distinction and complementarity between description and inference, and finally the interest of distinguishing random from systematic sources of variation. In the context of this last distinction, argument is provided concerning the limitation of quantitative methods in detecting verisimilitude, instead of truth.

Key words: data analysis, measuring, quantitative data-analysis.

---

[1] O Presente trabalho tornou-se possível graças à cooperação internacional estabelecida no âmbito do projeto Capes-Nordeste de apoio a programas de pós-graduação e pesquisa da região.

1. Introdução

O esforço de quantificação, em ciências humanas, é freqüentemente apresentado como uma iniciativa fundamental no sentido de se aceder à construção de um corpo de conhecimentos dito científico. Quantificar, neste caso, implica em mobilizar um sistema de *medidas* que, enquanto objeto matemático, se caracteriza por determinado conjunto de propriedades abstratas, e utilizar este sistema escolhido como referencial para a abordagem (mensuração) de determinado fenômeno. Uma vez constituído um corpo de medidas, abrem-se inúmeras possibilidades de trabalho:

• Um conjunto de *medidas* pode ser **descrito** de várias maneiras, seja de forma analítica, através de freqüências das características escolhidas dos fenômenos estudados, seja de forma sintética, através de *medidas-resumo* unidimensionais (como a *média*, a *moda* e a *mediana*) ou multidimensionais, como é o caso do *centro de tipicalidade* das análises fatoriais (Cibois, 1983).

• Tal descrição pode abranger mais de um grupo de medidas, verificando-se tendências de *ligação* ou *concomitância* entre dois ou mais conjuntos de medidas obtidos, ou mesmo avançando-se na direção da proposição de um modelo matemático-algébrico que relacione um conjunto (ou mais) de medidas com outro, em termos de uma *função matemática*. O caso habitual considera dois conjuntos de medidas numéricas, verificando-se tendência de relação funcional de tipo *linear* representada por um gráfico de pontos distribuídos ao longo de uma reta (esse conjunto de pontos é denominado *nuvem estatística*).

• Depois da descrição pode-se passar a um procedimento inferencial-probabilístico, no qual se procura estabelecer propriedades válidas não somente para o conjunto de dados considerado como amostra efetivamente analisada, mas para um conjunto maior não completamente analisado, o *universo* de observação, de onde tal amostra foi retirada. A utilização desse procedimento *indutivo* expõe-nos ao risco de erros quanto à validade das propriedades enunciadas pelo pesquisador sobre determinada totalidade; não obstante, este tipo de procedimento constitui-se em poderosa ferramenta de pesquisa quando empregado com critério. Esse procedimento apóia-se particularmente em teorias probabilísticas que permitem controlar o risco de erro quanto à validade de determinada conjectura, falando-se de nível de confiança ou, de forma complementar, nível de risco. Nesse sentido os *testes de hipóteses* ou *estimação* constituem-se em ferramentas básicas para entender esse *raciocínio estatístico.* Tentaremos fornecer indícios, mais adiante, de que esse raciocínio conduz a conclusões em

termos basicamente de *verossimilhança* e *plausibilidade*, e não em termos de *verdade*, no que diz respeito ao exame de hipóteses explicativas propostas pelo pesquisador.

As possibilidades acima, enquanto ferramentas culturais e objetos de reflexão e desenvolvimento, são agregadas no domínio matemático da ***estatística***. Tal domínio nasceu, historicamente, como ferramenta de controle fiscal do cidadão pelo estado: a raiz latina do adjetivo ***statisticum*** é a mesma do substantivo Estado (Droesbeke & Tassi, 1990). Os mesopotâmios, por exemplo, fizeram uso de procedimentos para descrição e cálculo de imposto em função da área cultivável. Os egípcios antigos, por sua vez, tabularam detalhadamente observações como a altura das cheias anuais do Nilo e a produtividade agrícola, em função da qual se estabelecia o imposto devido ao faraó; de posse de tais dados, os egípcios foram capazes de propor uma tabela de cálculo de imposto diretamente a partir da altura máxima anual do Nilo, num raciocínio cujo arcabouço conceitual antecipa as idéias de *co-variação* e *co(r)-relação*, só muito depois explicitadas matematicamente.

Atualmente, pode-se facilmente verificar que a estatística continua sendo um instrumento absolutamente central no contexto do exercício do poder de estado. Governar implica basicamente em fixar objetivos, obter informações pertinentes, controlar politicamente o acesso e divulgação dessas informações e tomar decisões em conformidade com os aspectos anteriores. No que diz respeito à obtenção de informações, o papel das ferramentas estatísticas é central: é praxe nos países ocidentais a montagem de agências governamentais especificamente encarregadas de monitorar a cena sócio-econômico-demográfica[2].

Mas apesar desse lado oficial (ou até justamente por isso...) nem sempre a estatística tem gozado de boa fama: há quem se refira aos estatísticos profissionais como os trapaceiros da matemática. Glass & Stanley (1970, p.1), por exemplo, cunharam o termo "*statisticulation*"[3], para definir a arte de mentir com a ajuda da estatística, mantendo-se a aparência de objetividade e racionalidade; J.-L. Besson, por sua vez, organizou com outros colaboradores excelente volume de reflexões acerca do que denominou de "ilusão das estatísticas" (Besson, 1995). Neste contexto particular, Benjamin Disraeli (citado por Angers, 1991) propôs a existência de três tipos de mentiras com graus de gravidade crescentes: as mentiras usuais, as mentiras terríveis e a estatística; exemplo clássico em apoio à perspectiva do ex-primeiro ministro britânico: propor que alguém ponha a cabeça em um geladeira e os

---

[2] É o caso, no Brasil, do IBGE (**Instituto Brasileiro de Geografia e Estatística** http://www.ibge.gov.br/)

pés num forno, justificando tal proposta nos seguintes termos: "*na média, você vai se sentir bem...*". A. F. Siegel (1988, p.13), por sua vez, faz uma bem-humorada alusão ao fato de que, se alguém submeter um problema qualquer à análise de dois estatísticos profissionais, com certeza obterá no mínimo três respostas diferentes...

Além desses percalços, é preciso ainda admitir que, apesar de provocar várias controvérsias quanto às utilizações eventualmente inadequadas, a estatística tem tido um importante papel no contexto do debate epistemológico acerca da cientificidade das ciências humanas, conforme aludido já na abertura das presentes reflexões. É frequente, nesta comunidade de pesquisa, o apelo a métodos apoiados em testes estatísticos, mobilizados como "selo de qualidade científica", algo que confere respeitabilidade (cientificidade) aos resultados de uma pesquisa, distinguindo-os do conhecimento dito de "senso comum" ou mesmo do conhecimento especulativo. Nesse sentido, o par *método experimental* e *teste de hipótese assistido pela estatística* representa para muitos pesquisadores a única possibilidade de encaminhamento metodológico para a construção de conhecimento científico em qualquer domínio do conhecimento. Tal perspectiva, hegemônica em vários contextos sociais de prática científico-profissional em ciências humanas, tem contudo uma contrapartida radical, representada pela recusa sem nuances de utilização da categorização e dos procedimentos estatísticos na construção de conhecimento. No contexto das presentes reflexões, não nos cabe aprofundar o debate acerca dos critérios de cientificidade gerais ou específicos das ciências humanas. Paralelamente, não pretendemos subsidiar o debate simplista do *pró* ou *contra* o uso da estatística no âmbito destas ciências, pois tal opção, colocada em tais termos, pouco acrescenta à formação e à prática dos pesquisadores em ciências humanas em geral. Nosso objetivo, aqui, consiste em oferecer alguns subsídios à reflexão acerca dos riscos e ganhos inerentes à utilização dos métodos ditos *quantitativos* no âmbito das ciências acima aludidas, de forma a se avançar em termos da formação metodológica e da prática científica acima aludidas.

2. Quantificação em ciências humanas: aspectos básicos a considerar

Em termos gerais, propomos que a idéia de *quantificação* abrange um conjunto de procedimentos, técnicas e algoritmos destinados a ***auxiliar*** o pesquisador a extrair de seus dados ***subsídios*** para responder à(s) pergunta(s) que o mesmo estabeleceu como objetivo(s) de

---

[3] Neologismo em língua inglesa resultante da combinação das palavras *statistics* e *speculation* (especulação).

trabalho. Siegel (op.cit.) numa perspectiva semelhante, propõe que a estatística seria "(...) a *arte* de fazer inferências sobre dados imperfeitos buscando-se a mensagem escondida em meio ao ruído" (itálicos nossos). Nesse sentido, faz parte de um esforço maior de análise de dados, e sua justificativa básica poderia ser resumida nos seguintes termos: *a informação que não pode ser diretamente "visualizada" a partir de uma massa de dados poderá sê-lo se tais dados sofrerem algum tipo de transformação que permita uma observação de outro ponto-de-vista*. Assim, propomos desde já que se tenha da quantificação a perspectiva de um *amplificador cultural*, no sentido proposto por J. Bruner (Bruner, 1983; 1996).

A estatística, enquanto subcampo da matemática aplicada, abrange um conjunto de ferramentas auxiliares à *quantificação*, abrangendo a exploração sistematizada de uma massa de dados através da descrição frequencial e dos resumos. Sua utilização por parte do pesquisador em ciências humanas exige a aceitação do recurso à noção de ***mensuração***, que por sua vez não tem sentido a não ser em presença de uma atividade prévia de ***categorização***. Medir implica portanto em acoplar ao fenômeno observado um sistema classificatório; em outras palavras, medir pressupõe necessariamente ***modelizar***, uma vez que qualquer sistema de categorização / mensuração traz em seu bojo, implícita ou explicitamente, um determinado ***modelo teórico*** acerca do objeto observado. Assim, quando Piaget, por exemplo, ao examinar determinadas produções de seus sujeitos, categoriza-os como *conservativos*, *intermediários* e *não-conservativos* (Piaget, 1970), apoia-se para tal categorização numa perspectiva específica acerca do desenvolvimento cognitivo; esta perspectiva fornecerá os critérios diferenciadores a partir dos quais as produções das crianças serão classificadas. Em conseqüência, nenhuma categorização é rigorosamente descritiva, no sentido de captura dos objetos "tal como eles são na realidade", conforme discute F. Halbwachs (Halbwachs, 1974, pp. 38-60). Este mesmo autor chama a atenção para o fato de que, mesmo no contexto de ciências tais como a física, "o papel de representação desempenhado pelos modelos não é de descrição de uma situação real" (Halbwachs, op.cit., p. 47). Para A.-M. Drouin, o que conta, em termos da avaliação de um modelo, não é a fidelidade a um real hipotético "em si"[4], mas a eficácia descritiva, explicativa ou preditiva de tal modelo cuja pertinência é validada no confronto com o real tal qual o mesmo se nos afigura (Drouin, 1988, p.11).

Nesse ponto da reflexão, convém enfatizar o seguinte aspecto central: *quantificar* exige *categorizar* e *medir*, três atividades que implicam em *modelizar*, o que nos remete

[4] A busca desse "real", com base em que constituir-se-ia um "conhecimento total", em contato com uma essência ou uma verdade, é considerado por esta autora como "um sonho teológico" (*un rêve théologique*).

necessariamente ao domínio da teoria. A atividade de *quantificação* engaja portanto o pesquisador em um processo complexo, conforme esquematizado na ilustração 1 mais adiante. Essas operações não se processam, na prática, de forma linear e sequencial, de vez que os resultados obtidos podem conduzir o pesquisador a modificar as categorias ou mesmo o modelo com o qual iniciou todo o trabalho. É este o caso, por exemplo, quando a constatação de uma freqüência muito reduzida de uma categoria conduz a agregá-la a outra, ou quando a partição de categorias prevista pelo modelo se mostra inadequada após uma primeira análise dos dados.

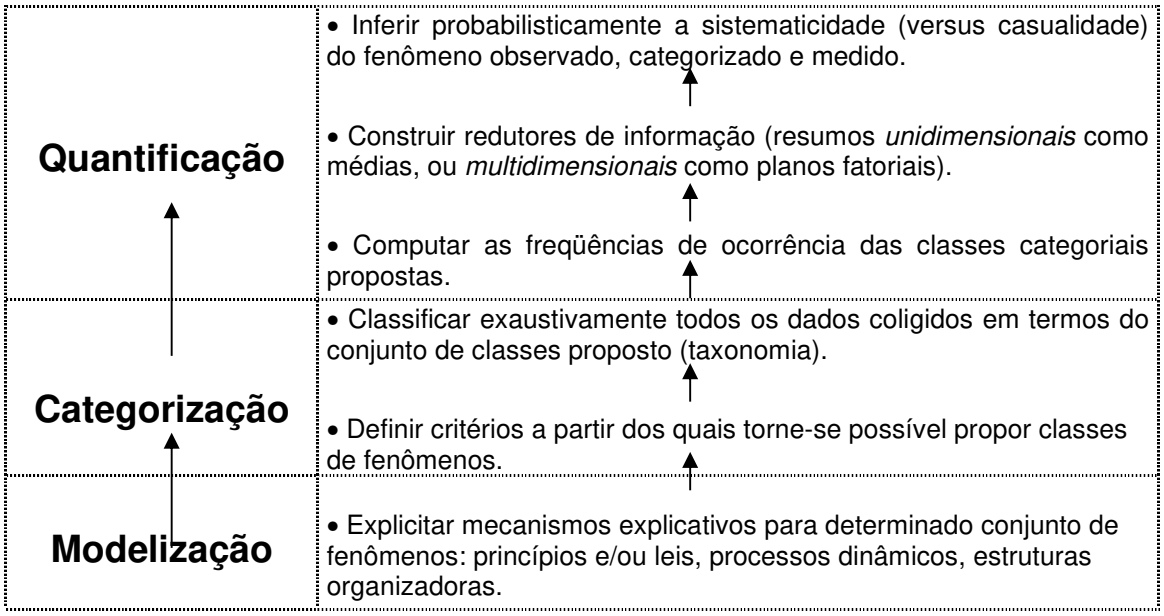

Ilustração 1: explorando esquematicamente as bases epistemológicas da quantificação

Tais atividades representam sem dúvida benefícios em termos de instrumentalização e amplificação do poder do observador, mas implicam em um risco epistemológico que o pesquisador não pode ignorar. Tal risco se caracteriza ao mesmo tempo por conseqüências epistemológicas decorrentes do acoplamento de um sistema abstrato de medidas ao fenômeno que o pesquisador escolhe estudar, ou seja, conseqüências de sua aceitação em categorizar e medir, e, em segundo lugar, conseqüências operacionais na exploração das possibilidades acima referidas, riscos de percurso muitas vezes ignorados no contexto atual, em que poderosas ferramentas informatizadas têm tornado os procedimentos de manipulação de medidas um ritual automático de entrada e saída de dados. O presente artigo busca, portanto, reagir à perda de significado dos procedimentos de quantificação em ciências humanas, elegendo alguns pontos de alto risco que precisam ser considerados criticamente, bem como

seu poder como amplificadores culturais. Tais pontos são analisados ao longo dos tópicos seguintes.

2.1. A categorização e a mensuração

Conforme esboçado anteriormente, o procedimento de categorização implica em assimilar um conjunto de observações sistemáticas e singulares feitas pelo observador (por exemplo: as estratégias de resolução de problemas de uma amostra de 20 sujeitos) a um conjunto mais restrito de classes de informações (categorias) que abarcarão observações unificáveis de acordo com um ou mais critérios (ver ilustração 2 abaixo).

Amostra de protocolos singulares (*dados de base*)																			
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T

↓

Protocolos categorizados		
Categoria 1	Categoria 2	Categoria 3
A C D E T R U	B F G H L Q	O P J K M N S

Ilustração 2: Procedimento básico de categorização

Em conseqüência, trata-se de um procedimento em que inevitavelmente se perde informação, pois dados que a rigor são diferentes passam a conviver sob o mesmo rótulo categorial. Naturalmente, a tal perda corresponde um ganho que a justifica: tem-se com tal procedimento redutor acesso a determinada *estrutura* dos dados, algo que tem, por definição, baixa visibilidade quando considerada a olho nu, ou sob enfoque clínico. Assim, torna-se difícil verificar determinada *tendência* em termos, por exemplo, de diferenciação de desempenho matemático escolar em função do gênero, quando se lida com casos singulares, mas tal tendência se torna acessível quando se perde de vista a informação singular e se agrupam informações, de forma a se obter gráficos como o que vem reproduzido na ilustração 3 mais adiante. Tais considerações nos levam, em seguida, ao exame do tipo de transformação a que é submetida a informação, após categorizada, de forma a servir de suporte à decisão do pesquisador após tratamento estatístico

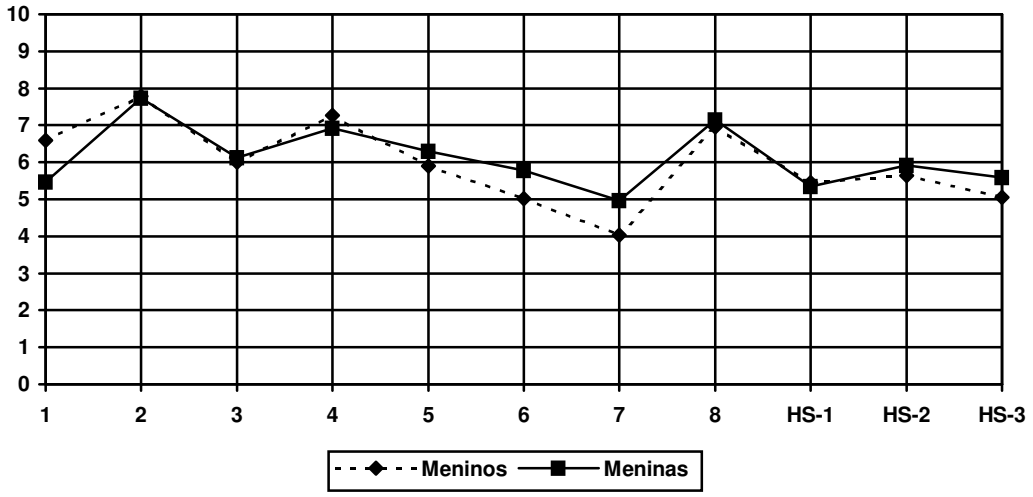

Ilustração 3: Curvas de desempenho de meninos e meninas de escolas públicas e privadas de Recife (n=721), ao longo de 11 anos de escolarização (1ª série do ciclo elementar à terceira série (HS-3) do ciclo médio). Reproduzido de Da Rocha Falcão & Loos, em preparação.

2.2. Um exemplo de tratamento de informação categorizada: a avaliação escolar

Vamos considerar uma situação de partida bastante comum, em que um professor deseja avaliar o nível de aprendizagem de seus alunos após determinada atividade de ensino por ele conduzida em sala-de-aula. Um modelo simplificado da situação de avaliação poderia considerar que o professor teria uma decisão a tomar, para cada aluno, em termos de duas únicas possibilidades ou hipóteses: $\mathbf{H_0}$, nível de aprendizagem insuficiente, versus $\mathbf{H_1}$, nível de aprendizagem suficiente. Esse professor decide então recorrer a uma ferramenta auxiliar bastante usual, que é o questionário de exame; este questionário, por si, comporta uma quantidade grande de possibilidades, no contexto das quais o professor exercerá seu poder de escolha com conseqüências inevitáveis: questionário com questões abertas, semi-abertas ou fechadas, com questões independentes (ou não) umas das outras, com nível de dificuldade semelhante ou diferenciado. Seja como for, é ainda importante considerar que se trabalha com a suposição fundamental de que as questões são pertinentes e relevantes enquanto elementos representativos da competência que se deseja avaliar. Por último, o aluno-aprendiz em avaliação poderá responder de forma correta ou incorreta, ou não responder, o que representa, portanto, três possibilidades simples de categorização de suas respostas ao questionário. Consideremos ainda que um questionário tem naturalmente um número limitado de questões, e que tais questões não podem pretender esgotar o conjunto de aspectos que compõem a

competência avaliada, até porque não se sabe ao certo em que consiste tal competência! Veja, portanto, o leitor sobre que nível de decisões prévias, suposições e incerteza se apóia, de saída, o procedimento (inicialmente julgado simples) de avaliação aqui discutido ilustrativamente.

Para prosseguirmos o raciocínio, vamos propor que o professor do exemplo em tela optou por um questionário bastante simples e comum, o *questionário de múltipla escolha*, ou seja, aquele instrumento que se caracteriza pela proposição de uma questão, à qual se seguem múltiplas opções de respostas dentre as quais apenas uma é considerada como sendo correta. Vamos considerar agora uma **variável de decisão D**, que diz respeito simplesmente ao número de respostas corretas apresentadas por determinado aluno para o questionário-exame. Tal variável é nomeada "variável de decisão" porque é com base nela que o professor decidirá onde situar o aluno examinado, se no grupo dos "suficientes" ou naquele dos "insuficientes". Sabemos, por outro lado, que o conjunto de resultados possíveis a serem apresentados pelos alunos será o conjunto de números variando de **0** (zero) a **q** (número total de questões do questionário). Ainda à guisa de simplificação do raciocínio, consideremos que as questões do questionário sejam independentes umas das outras e de igual nível de dificuldade; neste caso, a variável de decisão acima aludida será uma fronteira que demarcará duas regiões ou subconjuntos de resultados: aqueles correspondentes aos alunos que serão considerados *insuficientes* ou reprovados (região esta que irá de zero até um determinado valor), e aqueles correspondentes aos alunos julgados *suficientes* ou aprovados (após este determinado valor até o número máximo de acertos possíveis). Ora, o ponto por onde passará a fronteira acima aludida será função da nossa expectativa acerca do nível de dificuldade do teste: se se trata de um teste julgado “fácil”, nossa fronteira será mais exigente: em um teste com 10 questões, poderíamos estabelecer que somente a partir de 5 acertos consideraríamos o aluno como tendo demonstrado suficiência; por outro lado, se se trata de um teste julgado "difícil", já a partir de 4 pontos poder-se-ia considerar o aluno suficiente. Em função disso, podemos afirmar que a *variável de decisão* com a qual lidamos pode ser modelizada por uma variável matemática do tipo *variável binomial*[5], com parâmetros **q** e π (pi) onde **q** é o número de questões e π é a probabilidade (ou expectativa) de fornecer-se uma resposta correta.

Vamos agora retomar a possibilidade anteriormente referida segundo a qual, diante de

---

[5] Essa variável bem conhecida permite descrever a freqüência do número de *sucessos* numa série de eventos identicamente repetidos. Assim, num jogo de cara e coroa com uma moeda, poderíamos estabelecer a freqüência de aparição do evento “obter 25 vezes coroa” após lançar cem vezes tal moeda.

um aluno qualquer que se dispõe a responder o questionário, nós, avaliadores, poderemos ter duas "posturas" ou "expectativas" que serão traduzidas por duas *hipóteses*: uma primeira hipótese, que chamaremos $\mathbf{H_o}$, será nesse caso a hipótese de **insuficiência de aprendizagem**, enquanto que uma segunda hipótese, $\mathbf{H_1}$, será aquela de **aprendizagem suficiente**. Dois esquemas de modelos poder-se-iam abrir diante de nós:

1. *A probabilidade de um aluno que não aprendeu suficientemente responder corretamente a uma questão do questionário é* ***igual*** *à probabilidade deste mesmo aluno errar; concomitantemente, a probabilidade de que um outro aluno que aprendeu suficientemente responder corretamente a esta mesma questão é* ***igual*** *à probabilidade dele errar*. Ora, se esta expectativa for verdadeira, estaremos diante de um problema grave de avaliação, pois nossa capacidade de distinguir entre alunos que aprenderam e não aprenderam, via questionário, é bastante grosseira: há uma margem probabilisticamente grande de avaliarmos como "bom" aquele aluno que não aprendeu (ou seja, abandonarmos indevidamente a hipótese Ho de não-aprendizagem, o que chamaremos aqui de **erro de primeiro tipo**), e de avaliarmos como "mau" aquele aluno que aprendeu (abandonarmos indevidamente a hipótese H1 de aprendizagem, que denominaremos **erro de segundo tipo**).
2. *A probabilidade de um aluno que não aprendeu suficientemente responder corretamente a uma questão do questionário é* ***menor*** *que a probabilidade deste mesmo aluno errar; concomitantemente, a probabilidade de que um outro aluno que aprendeu suficientemente responder corretamente a esta mesma questão é* ***maior*** *que a probabilidade dele errar*. Este é o caso ideal em termos de utilização de nosso questionário como instrumento de avaliação, pois o risco de avaliarmos como "bom" aquele aluno que não aprendeu (ou seja, abandonarmos indevidamente a hipótese Ho de não-aprendizagem, cometendo um erro de primeiro tipo), bem como de avaliarmos como "mau" aquele aluno que aprendeu (abandonarmos indevidamente H1, cometendo um erro de segundo tipo), é menor do que o mesmo tipo de risco referido no tópico 1 acima. Podemos inclusive afirmar que a probabilidade de uma rejeição indevida de Ho (erro de tipo 1) cairá mais e mais em função da menor probabilidade de acerto no caso do aluno que não aprendeu, e que a probabilidade de rejeição indevida de H1 (erro de tipo 2) cairá mais e mais em função da maior probabilidade de acerto por parte do aluno que aprendeu. Estas relações são ilustradas pelo gráfico abaixo:

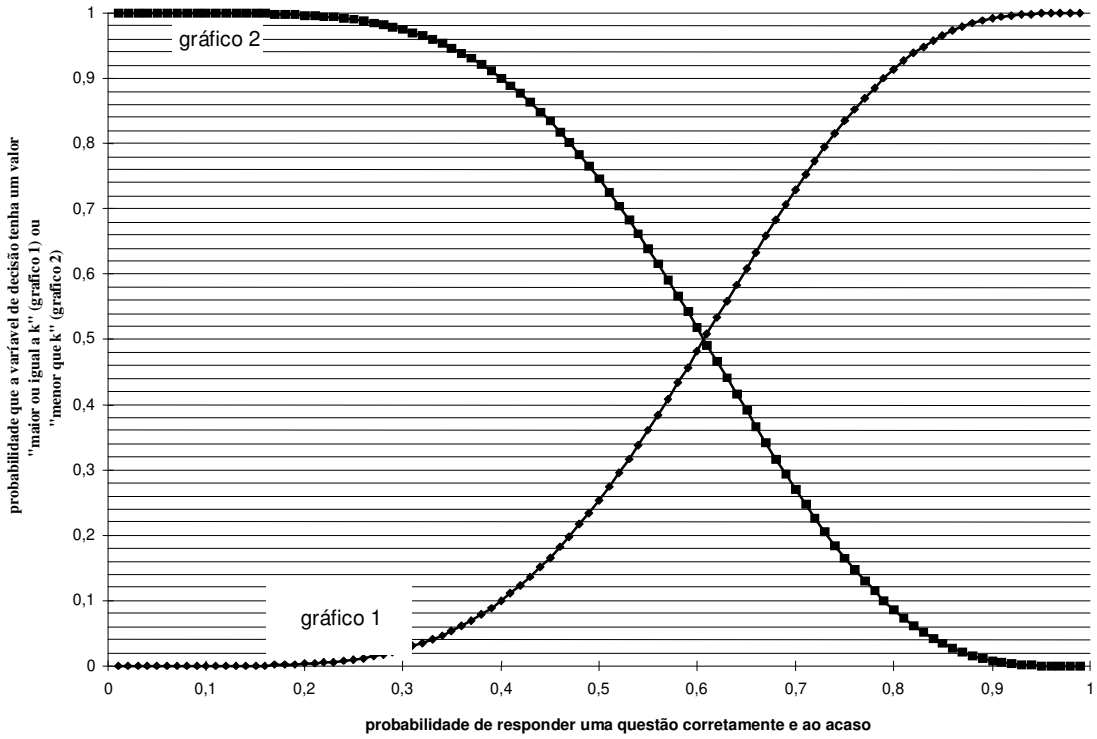

Ilustração 4: Comportamento dos gráficos descritivos das probabilidades de resposta correta a uma questão em função da variação de probabilidade referente ao valor da variável de decisão.

Para construir esses gráficos é necessário especificar previamente o modelo de partida (nosso exemplo baseia-se sobre um questionário de q =10 questões e com o valor crítico k =5) :

- no gráfico 1, α (alfa) é a probabilidade da variável de decisão ter um resultado (número dos sucessos) **maior** que **k** (valor critico) sob a hipótese Ho de responder **corretamente** bem como **por acaso** a cada questão com uma probabilidade π_0

α = Prob{ D>= k sob H_0}

- no gráfico 2, β (beta) é a probabilidade da variável de decisão de ter um resultado (numero dos sucessos) **menor** que **k** (valor critico) sob a hipótese H_1 que um sujeito competente pode responder **corretamente** a cada questão com uma probabilidade π_1

$$\beta = \text{Prob}\{ D< k \text{ sob } H_1\}$$

$$\mathbf{1-\beta} = \text{Prob}\{ D>= k \text{ sob } H_1\}$$

Assim, com esse questionário de 10 questões e o critério de decisão " responder corretamente

a 5 questões ou mais sobre 10 ", sob a hipótese Ho (considerando-se portanto o sujeito incompetente), a probabilidade α fica menor que 0,05 para uma probabilidade π_0 de responder corretamente inferior a 0,34. No nosso modelo um questionário *de múltipla escolha* cujas questões contêm 3 ou mais possibilidades de resposta assegura o respeito dessa condição[6]. Alternativamente sob a hipótese H_1 considerando-se o sujeito competente, a probabilidade β teria um valor inferior a 0,04 para uma probabilidade π_1 de responder corretamente superior a 0,85.

Em outros termos, se pensarmos numa situação em que a margem probabilística de erro de tipo 1 é baixa, ou seja, a probabilidade de rejeitarmos Ho indevidamente é pequena, estaremos numa situação de alto nível de exigência, em que será mais difícil consideramos "suficiente" um aluno de fato "insuficiente"; mas ao estabelecermos um tal nível de exigência, passamos a incorrer num risco maior de considerarmos "insuficiente" um aluno de fato "suficiente", ou seja, aceitar indevidamente Ho. Nessa ordem de idéias, se denominamos α (alfa) probabilidade de erro de tipo1, e β (beta) a probabilidade de erro de tipo 2, podemos então reter que alfa e beta variam de 0 a 1 de forma inversa, e que esses dois tipos de erro relacionam-se entre si conforme ilustrado pelo quadro reproduzido na ilustração 5 adiante. A situação fictícia e altamente simplificada que exploramos acima serve para ilustrar algumas idéias fundamentais em termos de análise de dados categorial-quantitativa em ciências humanas: em primeiro lugar, nem todas as situações se prestam a uma modelização em termos de hipóteses básicas de trabalho, espaço de possibilidades e variáveis de decisão. Não obstante, algumas situações se prestam a uma tal abordagem; estas situações contudo são razoavelmente complexas, porque:

1) Partem de uma pré-análise na qual as escolhas, ênfases e vieses do pesquisador têm papel central na estruturação dos dados: trata-se da etapa de escolha das situações, questões, etc.;
2) Lidam com situações em que não estão em jogo dicotomias do tipo verdadeiro-falso, e sim um continuum probabilístico em que se lida com as idéias de mais provável ou menos provável, com margens de risco variáveis porém inevitáveis. Assim, a análise inferencial de dados nos força a abrir mão da ambição da "verdade" e a nos contentarmos com os vislumbres da "verossimilhança", do que "é" para o que "pode ser", com margem de erro

---

[6] Com 3 modalidades a probabilidade seria de $1/3 \cong 0,34$, com 4 modalidades, seria $1/4 = 0,25$, com 5 seria $1/5 = 0,20$...

variável.

Decisão do pesquisador ▼	**Status de H_o e H_1 (7)** ▼	
	H_o verdadeira (H_1 falsa)	**H_1 verdadeira (H_o falsa)**
Conservar H_o (rejeitar H_1)	Decisão CORRETA **$1 - \alpha$** = Prob{conservar Ho sob a condição H_o verdadeira}	**Decisão INCORRETA (Erro de tipo 2)** β = Prob{conservar H_o sob a condição Ho falsa}
Rejeitar H_o (conservar H_1)	**Decisão INCORRETA (Erro de tipo1)** α = Prob{rejeitar Ho sob a condição H_o verdadeira}	Decisão CORRETA **$1 - \beta$**= Prob{rejeitar Ho sob a condição H_o falsa}

Ilustração 5: Situações possíveis em termos de tipos de decisão tomadas pelo pesquisador em função das hipóteses básicas consideras, e tipos de erro associados (OBS: Prob = Probabilidade).

2.3. As variáveis estatísticas ditas escalas de medida pelo pesquisador em ciências humanas: caracterizações, pressupostos e riscos

A **medida**, conforme discutido acima, diz respeito especificamente à atribuição de um código (número, letra, etc.) a determinado evento de acordo com regras explícitas. Tais regras dirão respeito basicamente aos critérios de atribuição, que por sua vez guardam relação não só com a natureza dos dados, como também com determinadas decisões do examinador.

Aqui nós devemos precisar algumas noções do domínio matemático que permitem basear a modelização da *medida* de forma menos superficial. Essa medida consiste em estabelecer uma ligação entre um conjunto de elementos (unidades estatísticas) escolhidos em função do fenômeno estudado, e um conjunto de elementos simples ou complexos gerados pela modelização (categorização), de tal maneira que uma unidade estatística não pode ser relacionada a mais de um elemento categorial considerado. À guisa de exemplo, poderíamos pensar na representação, em um plano cartesiano cujos eixos dizem respeito respectivamente a uma escala de desempenho em matemática e em ciências, das notas de determinado grupo de estudantes. Esse procedimento de conexão de uma medida a um evento categorizado é chamado, em termos matemáticos genéricos, de estabelecimento de relação funcional ou

7 O status das hipóteses H_o e H_1 são aqui considerados em termos convencionais. O caráter intrínseco de verdade

função, e em termos mais especificamente estatísticos, construção de variável estatístico-probabilística. No seu domínio próprio de trabalho, o estatístico propõe várias "caixas de ferramentas" cada dia mais enriquecidas; nosso objetivo aqui não é expor em detalhes e extensivamente todas essas ferramentas, mas somente discutir algumas noções básicas que permitem refinar a linguagem através da qual se constrói a argumentação do pesquisador em ciências humanas.

As variáveis com as quais lida o pesquisador acima referido podem ser classificadas, em primeira instância, em termos do objetivo básico da medida ou resultado pretendido. Esta primeira partição nos conduz a uma diferenciação em termos de variáveis ditas ***qualitativas*** (que se prestam a uma diferenciação em termos de presença/ausência de determinados atributos ou qualidades), e ***quantitativas*** (que ensejam quantificação da diferenciação, através de atividade básica de contagem). As variáveis qualitativas podem ainda sofrer uma segunda partição, em função do caráter ordenável ou não-ordenável do fenômeno que modelizam; em função desta partição, teremos respectivamente variáveis qualitativas ***ordinais*** e ***nominais***. As variáveis quantitativas, por seu turno, vão sofrer uma partição com base noutro critério, visto que os números (quantidade) se inserem numa relação de ordem[8], o que faz com que tais variáveis comportem a ordenação. O critério de partição, aqui, diz respeito à distinção entre *conjunto contínuo* e *conjunto discreto* de números, o que nos conduzirá respectivamente a variáveis quantitativas contínuas e discretas. A rigor, cabe salientar que o caráter *contínuo* ou *discreto* dos fenômenos quantificados diz respeito antes a uma tomada de decisão (convenção) que a propriedades inerentes ao fenômeno: mesmo aspectos usualmente considerados contínuos, como o tempo, podem ser discretizados, e vice-versa.

Feitas as precisões acima, cabem algumas considerações específicas acerca da utilização pelo pesquisador de escalas de medidas baseadas nas variáveis anteriormenente aludidas.

- A ***escala nominal*** mobiliza variáveis qualitativas que não permitem ordenar pertinentemente os resultados. O recurso a tal escala justifica-se quando é necessário atribuir um rótulo diferenciador ("nome") a classes qualitativamente diversas de fenômenos. É o caso, por exemplo, de classificação por gênero (masculino - feminino), por tipo de participação em determinado desenho experimental (grupos experimental e controle), e assim por diante. Em

---

ou falsidade de cada hipótese não pode, em última análise, ser determinado.

[8] Dados dois números quaisquer e diferentes A e B, é sempre possível estabelecer as relações A maior que B ou B maior que A.

termos de potencialidade para análise de dados, é comum se ouvir o comentário segundo o qual tal escala é a mais "fraca" em termos de veiculação de informação, pois usa apenas a propriedade distintiva das categorias (categoria A ≠ categoria B). Essa afirmação, se bem que compreensível em termos de uma comparação no interior do conjunto de escalas de medida, deve ser considerada com reserva (e sobretudo sem "complexo de inferioridade"...) pelo pesquisador, de vez que o surgimento de ferramentas descritivas poderosas baseadas em dados nominais, como é o caso dos procedimentos informatizados de análise multidimensional[9], ampliaram consideravelmente o poder informativo de tal tipo de dados.

• A ***escala ordinal*** faz apelo a variáveis qualitativas com possibilidade de ordenar os resultados. Sua utilização é pertinente quando se deve atribuir rótulos a partir da verificação da presença de *graus* diversos e ordenados no âmbito de uma mesma variável. Tem-se, neste caso, diferenças qualitativas (como na escala anterior) a que se acrescentam diferenças quantitativas (não necessariamente explícitas, como ver-se-á mais adiante): níveis 1 (fraco), 2 (médio) e 3 (forte) em desempenho em Matemática, ou níveis de desenvolvimento cognitivo geral I (pré-operacional) II (intermediário) e III (operacional). Em relação a esta escala de medida, cabem algumas observações importantes do ponto de vista da pesquisa em ciências humanas em geral, notadamente em psicologia, em que tal escala costuma ser bastante utilizada. Retomemos e enfatizemos este ponto central: a escala de medida ordinal utiliza a propriedade distintiva das categorias, à qual se agrega determinada propriedade hierárquico-sequencial: categoria A < categoria B < categoria C. Muitas vezes, tal ordenação é simbolizada utilizando-se o sistema numérico, pelo simples fato desse sistema de codificação ser amplamente disponível na cultura e já comportar os dois níveis de mensuração (qualitativo e ordenado-sequencial) que se quer utilizar. Nesse caso, ao invés de se utilizar A, B, C e D, utilizam-se os números 1, 2, 3 e 4. Ora, essa última codificação ordinal possui riscos nem sempre bem avaliados por quem dela lança mão. Dois desses riscos são os seguintes:

a) *Risco de interpretação abusiva das diferenças entre categorias codificadas por números*: A *escala numérica* que é mobilizada por uma mensuração ordinal é perfeitamente regular. Em outras palavras, o incremento de 1 para 2, deste para 3, deste para 4 e assim por diante é constante, de forma que toda a série numérica pode ser representada pela expressão algébrica n, n+1, (n+1)+1, ((n+1)+1)+1... ∞ na qual a unidade pode, por convenção, ser substituída por qualquer outro valor. Verifique o leitor que nesta progressão, cada "passo" unitário dado

[9] Para uma análise introdutória a esses métodos, ver Bouroche & Saporta, 1980, e Fenelon, 1981

(de uma categoria para a categoria adjacente) representa um mesmo acréscimo ou decréscimo em termos de medida. Isto posto, considere as duas mensurações ordinais e fictícias apresentadas na ilustração 7 a seguir.

ESTUDO 1		**ESTUDO 2**	
Desempenho eleitoral dos candidatos (Votos obtidos)	**Categoria**	**Desempenho em teste de álgebra** (Produções dos alunos)	**Categoria**
RUIM (0 a 1000 votos)	**1**	Não consegue propor equações, restringe-se a tentativas de manipulação aritmética	**1**
MÉDIO (1001 a 2000 votos)	**2**	Propõe equações, mas apenas instancia valores nas mesmas.	**2**
BOM (2001 a 3000 votos)	**3**	Propõe equações e é capaz de manipulá-las de acordo com script-algoritmos adequados	**3**

Ilustração 6: ordenando categorias classificatórias em duas situações distintas

Conforme se pode verificar, temos uma mesma escala ordinal (1 – 2 – 3) , em termos de codificação, utilizada em duas situações, mas se constatam diferenças extremamente importantes ao se observar, em cada caso, a *coisa* medida em interação com o suporte de medida. No que diz respeito à escala de votos, a natureza da coisa medida não sofre constrangimentos quando nos questionamos se de fato tal coisa é incrementada tão regularmente quanto sugere a escala: o incremento de 1 para 2 e de 2 para 3 é o mesmo (em termos de regularidade) tanto em termos de escala quanto em termos da coisa medida; neste caso, a escala reproduz fielmente o comportamento do fenômeno observado. No caso da segunda mensuração, referente ao desempenho em álgebra, a situação não é tão tranqüila assim: nós não podemos em absoluto garantir que a passagem do procedimento 1 para o procedimento 2 represente o "mesmo tanto" de progresso da passagem de 2 para 3. É bastante pertinente pensar que os dois fenômenos observados possam ter a *métrica* hipotética representada na ilustração 8 adiante. No caso do estudo 1, temos uma *métrica* regular, o que não é o caso no estudo 2, em que o incremento de 1 para 2 é provavelmente bastante diverso daquele que se verifica de 2 para 3. Assim, é preciso ter muito cuidado para não "contaminar" o fenômeno estudado com propriedades do código de mensuração escolhido. Tal cuidado não representa um preciosismo de iniciados, mas deve fundamentar uma ação bastante corriqueira no mundo da pesquisa, que é a escolha do tipo de escala de mensuração e, em conseqüência,

do tipo de tratamento estatístico pertinente e adequado.

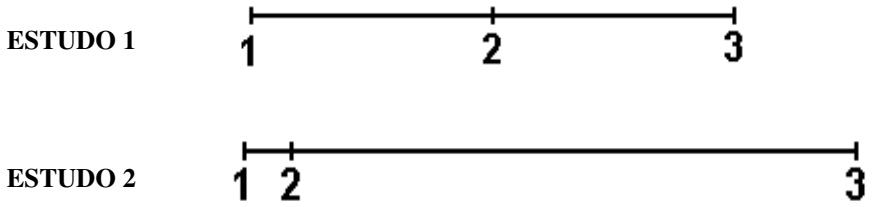

Ilustração 7: aventando diferenças de regularidade de *métrica* entre os estudos 1 e 2

Mas os problemas não se restringem a questões de métrica: há algo ainda mais "grave", epistemologicamente falando, a considerar, conforme discutido no item seguinte.

b) *Risco de proposição indevida de ordenação*: A rigor, muitas vezes o pesquisador não pode garantir se, ao passar de uma categoria para outra (no contexto de uma escala de mensuração ordinal), ele ainda tem o direito de falar de *incremento* de algo, ao invés de falar de coisas completamente distintas, não necessariamente ordenáveis em qualquer hierarquia simples, o que o forçaria a retornar à escala nominal. De fato, a utilização da escala ordinal pressupõe uma progressão em que há um incremento gradativo de uma categoria a outra, um "tanto" a mais da coisa medida. Será que isso sempre faz sentido, do ponto de vista da epistemologia do objeto investigado? Será, por exemplo, que podemos de fato afirmar que os quatro quartis de uma distribuição de QIs (quocientes intelectuais) agregam dados psicológicos *de fato* ordenáveis[10]? Será que os estágios das provas piagetianas são sempre representativos de uma ordenação, no sentido estrito do termo? Naturalmente adentra-se aqui uma discussão de caráter teórico, que foge ao escopo da presente reflexão; não obstante, nosso intuito ao mencionar tal ordem de risco na utilização de determinada escala de medida é enfatizar o quanto o recurso a tais ferramentas força o pesquisador a uma reflexão crítica constante, e em váridos contextos. Tal reflexão se conclui com as considerações seguintes, referentes às escalas numérica e intervalar.

- As escalas ***numérica*** e ***intervalar*** são mobilizadas para modelizar e medir variáveis quantitativas discretas e contínuas. O uso destas escalas pressupõe que é possível **medir** no sentido forte, quantitativo do termo. Conforme observações acima, uma tal escala considera *diferenças* entre medidas numéricas, uma vez que propõe um sistema de medida de métrica

[10] Para uma discussão aprofundada desse aspecto, ver Gould, 1981.

rigorosamente interpretável, o que permite não somente qualificar as diferenças em termos de ordem (10 acertos configuram melhor desempenho que 5 acertos em determinado teste), mas também quantificá-las (10 acertos representam um incremento quantificável em relação a 5 acertos). Ressurge aqui um problema epistemológico importante. Mais uma vez nos confrontamos a uma situação preocupante, no contexto do uso de ferramentas estatísticas em ciências humanas em geral, e particularmente em psicologia. Será que sempre faz sentido aceitar que 10 acertos num teste de avaliação representam *o dobro de desempenho* em relação a 5 acertos? Tais questões têm efetivamente de ser enfrentadas, pois a *escala intervalar* possibilita muitos tratamentos estatísticos dignos de atenção tanto em termos de ganho de informação quanto de riscos de interpretação contaminada pelos artefatos de medida: é o caso, notadamente, dos tratamentos que possibilitam a substituição de todo um conjunto de medidas por um numero reduzido de valores informativos.

3. Conclusões

> O estatuto de existência dos objetos experimentais está diretamente relacionado ao fato de que tais objetos são produzidos pelo homem.
>
> **F. Halbwachs**

Os aspectos discutidos brevemente no presente artigo sugerem, em termos gerais, que a *quantificação* em ciências humanas representa uma ferramenta cultural de apoio ao esforço global de construção de conhecimento neste domínio, ferramenta esta que pode, em contextos específicos, prestar serviços de amplificação da observação de valor inegável. Por outro lado, os vários procedimentos mobilizados pela quantificação, por sua própria natureza modeladora-formalizadora, podem efetivamente deformar a observação ao ponto de esterilizá-la por completo. Nesse sentido, a grosseria inata ao procedimento de categorização (que "fabrica" a homogeneidade no interior das categorias a partir da singularidade das produções que a compõem) pode de fato conduzir à constatação daquilo que ela própria criou; por outro lado, torna-se muito mais difícil vislumbrar fenômenos de conjunto que transcendem necessariamente o idiossincrático sem a categorização. A abordagem de tais fenômenos de conjunto ou tendências tem portanto na quantificação uma ferramenta extremamente útil; não obstante, é preciso não perder de vista que a natureza matemática dos algoritmos probabilísticos sobre os quais tais ferramentas se apóiam limita necessariamente as conclusões e vislumbres do pesquisador a proposições no máximo *verossímeis*, de vez que a presunção

do atingimento da verdade encontra-se, por definição, fora do escopo de possibilidades da inferência estatística.

Propõe-se, com as presentes reflexões, que o pesquisador em ciências humanas faça efetivamente uso das ferramentas oferecidas pela *quantificação*, com critério, cuidado crítico e o máximo de respeito epistêmico possível pelas peculiaridades do objeto estudado. Nesse sentido, permitimo-nos aqui lançar mão de exemplo ilustrativo simples porém bastante pertinente[11]: se eu disponho de 20 litros d'água, recebo mais 50 litros e quero saber de quantos litros passo a dispor, é perfeitamente lícito lançar mão de ferramenta matemática oriunda do campo conceitual das estruturas aditivas, e somar as medidas de volume d'água implicadas: 20 litros + 50 litros = 70 litros; mas se agora tenho em mãos problema superficialmente semelhante (mas de fato bem diverso), referente ao fato de dispor de dois volumes iguais de água, um com temperatura de 20ºC e outro com 70ºC, e querendo eu saber a temperatura final da mistura dos dois volumes, não será mais pertinente pensar em ferramentas aditivas para modelizar a situação e resolver o problema.

O uso de ferramentas quantificadoras a partir de categorização fundada em critérios vinculados, por sua vez, a *modelos* específicos, tem sem dúvida nenhuma um custo e um risco que justificam esforços sérios de conhecimento e análise crítica: a mensuração, muitas vezes, conduz a "largar a presa por sua sombra", conforme dito popular francês[12]. Nesse mesmo contexto de idéias, C. Chrétien endossa reflexões de Nietzsche, quando este observa que "(...) se se avaliasse o "valor" de uma música por aquilo que se pode calcular e contar, pelo que se pode traduzir em números... quão absurda seria essa avaliação "científica"!". E o mesmo autor conclui: "Que se haveria apreendido, compreendido, conhecido de uma melodia assim apreciada? Nada, e literalmente nada, daquilo que faz justamente a "música"!..." (Chrétien, 1994). Não obstante, uma análise multidimensional do tipo análise fatorial ou classificação hierárquica (Cibois, op.cit.), a partir de categorização de improvisos de músicos de jazz, poderia auxiliar um eventual estudioso a detectar estruturas temáticas que contribuíssem, por exemplo, para enriquecer a diferenciação entre diferentes modalidades de improviso, ou estilos.

Há sempre uma relação de custo e benefício a negociar em cada situação de análise de dados em psicologia e nas ciências humanas em geral, e isso é válido tanto para as abordagens

---

[11] Tal exemplo, juntamente com outras observações de que muito se beneficiaram o presente artigo, foi proposto pela professora Terezinha Nunes (Brookes University, Inglaterra), durante discussão preliminar das idéias aqui mencionadas.

ditas *qualitativas* quanto *quantitativas*: tal negociação implica em considerar cuidadosamente o objeto em estudo, os modelos teóricos que balizarão tal abordagem, as categorias oriundas da análise de tal objeto a partir de tal modelo, as ferramentas disponíveis para processar a informação bruta assim gerada. O benefício em termos de clarificação conceitual, ao que nos parece, freqüentemente suplanta o custo representado pela ênfase em aspectos secundários, ou pela grosseria representada por tratamentos estatísticos que, em última análise, somam indevidamente temperaturas de dois recipientes de água. Nesse sentido, o pesquisador em ciências humanas se priva de importante ferramenta cultural ao se autoproibir, freqüentemente por motivação ideológica, pressão de grupo corporativo ou ignorância pura e simples, o acesso a tais métodos quantitativos de análise de dados.

Referências bibliográficas

Angers, C. (1991) *Les statistiques: oui, mais...* Paris, Editions Agence d'Arc.

Besson, J.-L. (1995) *A ilusão das estatísticas*. São Paulo, Editora Unesp.

Bouroche, J-M., Saporta, G. (1980) *L'analyse des données.* Paris, Presses Universitaires de France.

Bruner, J.S. (1983). *Savoir faire savoir dire*, Paris, Presses Universitaires de France.

_____ (1996). *The culture of education.* Cambridge, Harvard University Press.

Chrétien, C. (1994) *A ciência em ação*. São Paulo, Papirus.

Cibois, P. (1983) *L'analyse factorielle*. Paris, Presses Universitaires de France.

Da Rocha Falcão, J.T., Loos, H. (1999) *Gender differences in mathematics achievement among Brazilian students: an old issue revisited*. (artigo em preparação).

Droesbeke, J-J., Tassi, P. (1990) *Histoire de la statistique*. Paris, Presses Universitaires de France.

Drouin, A-M. (1988). Le modèle en questions. *Aster* (7), 1-20.

Fenelon, J-P. (1981) *Qu'est-ce que l'analyse des données?* Paris, Lefonen.

Glass, G.V. & Stanley, J.C. (1970) *Statistical methods in education and psychology*. New Jersey, Prentice-Hall.

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

Halbwachs, F. (1974). *La pensée physique chez l'enfant et le savant.* Neuchâtel, Delachaux et Niestlé.

Levin, J. (1987) *Estatística aplicada a ciências humanas*. São Paulo, Habra.

Piaget, J. (1970) *L'Epistémologie Génétique*. Paris, Presses Universitaires de France.

Siegel, A.F. (1988) *Statistics and data analysis: an introduction*. New York, John Wiley & Sons.

---

[12] "*Lâcher la proie par son ombre*", na forma original.