

O uso da Estatística na pesquisa em Ciências Humanas no Brasil: o caso da Demografia*

Paulo de Martino Jannuzzi**

Introdução

A baixíssima frequência de utilização de métodos estatísticos nas pesquisas em Ciências Humanas no Brasil é um fato bem conhecido na comunidade acadêmica, tanto por estatísticos como por cientistas sociais. Uma rápida revisão dos principais periódicos nacionais na área é reveladora neste sentido: nem mesmo em artigos que tratam de análises empíricas de dados sociais as técnicas estatísticas básicas – como testes de hipóteses acerca de médias ou computação de medidas de associação ou independência entre variáveis –, que poderiam agregar um conhecimento adicional sobre o fenômeno social tratado, são referidas ou empregadas.

Meu objetivo, aqui, é discutir algumas possíveis motivações para este quadro, a partir de um levantamento de evidências de uso de métodos estatísticos na análise de processos sociodemográficos. Objetivamente, faço uma avaliação do uso de modelagem estatística, com destaque para os modelos de regressão, na pesquisa demográfica no Brasil. A delimitação do escopo da análise à Demografia deve-se ao fato de que, estando ela na fronteira entre a Estatística e as Ciências Sociais, as evidências aí encontradas poderão ser

reveladoras da questão mais geral para a qual este trabalho procura contribuir.

Evidências do uso de modelos estatísticos

Todo trabalho de levantamento ou análise da produção acadêmica em uma determinada área de conhecimento lida sempre com a dificuldade de identificação e obtenção de uma amostra representativa da bibliografia corrente e acumulada daquela área. No caso de se tratar de uma área multidisciplinar – como a da Demografia – tais problemas se potencializam, dada a diversidade de fontes e periódicos em que se pode encontrar artigos pertinentes.

Historicamente, a primeira revista de divulgação de trabalhos em Demografia foi a *Revista Brasileira de Estatística*, editada pelo IBGE desde 1940. Nas últimas duas décadas, a pesquisa demográfica ganhou novos espaços de divulgação com o início da realização bianual dos Encontros Nacionais de Estudos Populacionais, promovidos pela Associação Brasileira de Estudos Populacionais (ABEP), com a criação da *Revista Brasileira de Estudos de População* e com a publicação de textos seriados e relatórios técnicos de vários centros de pesquisa social ou demográfica no país (*Novos Estudos Cebrap*, *Textos NEPO*, *Informe Demográfico* da Fundação SEADE, relatórios de pesquisa do Cedeplar, textos para discussão do IPEA, publicações da Bemfam etc.). Há ainda outros canais de divulgação não específicos da área que trazem artigos no campo da Demografia, como as revistas *São Paulo em Perspectiva* (Fundação SEADE) e *Estudos Afro-Asiáticos*

* Trabalho apresentado na IV Escola de Modelos de Regressão, organizada pela Associação Brasileira de Estatística, Águas de São Pedro, SP, fevereiro de 1995.

** Analista de projetos da Diretoria Adjunta de Análise Sócio-Econômica da Fundação Sistema Estadual de Análise de Dados (SEADE) e doutorando em Demografia no Instituto de Filosofia e Ciências Humanas (IFCH) da Universidade Estadual de Campinas (Unicamp).

(Cândido Mendes), os Simpósios Nacionais de Probabilidade e Estatística (SINAPES), organizados pela Associação Brasileira de Estatística (ABE), e a *Revista Brasileira de Ciências Sociais*, editada pela Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais (Anpocs).

Dada a diversidade de fontes e o estágio atual de minha investigação, neste trabalho apresentarei apenas os resultados de um levantamento bibliográfico realizado nos nove *Anais dos Encontros Nacionais de Estudos Populacionais*, de 1978 a 1994. Vale observar, também, que o levantamento limitou-se a registrar o uso de modelos estatísticos. Não se tratou de fazer um balanço do "estado da arte" da quantificação na área, já que

Trata-se, pois, de um levantamento parcial, uma amostra não representativa da diversidade de enfoques tratados na área. No entanto, ainda assim, tal levantamento serve para delinear algumas questões sobre a intensidade e forma de uso de métodos estatísticos na pesquisa demográfica no Brasil.

Os resultados da pesquisa bibliográfica são apresentados na Tabela 1. Como se pode observar, o emprego de técnicas de modelagem estatística sempre foi residual nos Encontros da ABEP, mas teve um salto significativo no último Encontro. Afinal, dos 18 trabalhos apresentados entre 1990 e 1994, 12 o foram em 1994. Em termos das técnicas empregadas, as aplicações com modelos de risco foram as que mais cresceram.

Tabela 1
Emprego de modelos estatísticos nos trabalhos apresentados nas edições bianuais do Encontro Nacional de Estudos Populacionais

	1978-82	1984-88	1990-94	1994
Testes de Indep./Associação	-	2	3	2
Modelos de Regressão	3	7	9	6
Modelos de Risco	-	1	6	4
Total de Trabalhos	54	256	207	78

outras técnicas da Estatística e da Demografia Formal não foram alvo desta investigação. Além disso, a utilização do modelo estatístico deveria estar explicitada no artigo. Dos três primeiros Encontros (1978, 1980 e 1982) só foram analisados os textos integrais contidos nos *Anais*, não sendo considerados os resumos de trabalhos apresentados espontaneamente (ainda que alguns pudessem ter alguma indicação do emprego de métodos estatísticos).

Chama a atenção o fato de que, embora a análise de dados em tabelas bi ou trivariadas seja muito comum, o emprego de medidas de associação e testes de independência entre variáveis é baixo. Quanto aos modelos de regressão, estes têm sido usados na predição ou ajuste de parâmetros de modelos demográficos de projeções populacionais, de mortalidade ou fecundidade. Também têm sido empregados com fins de "inferência causal" (1), especialmente

(1) "Inferência causal" no sentido estreito proporcionado pela análise de regressão: verificação da existência de uma variação concomitante ou covariância entre a variável explicada e as variáveis explicativas. É interessante notar que mesmo na pesquisa sobre estrutura de causalidade entre variáveis demográficas e sociais a *Path Analysis* é pouco empregada.

na área de fecundidade. Em um caso e no outro há pouca discussão sobre a validação dos pressupostos, detecção de pontos atípicos e análise de resíduos. A aplicação de outras técnicas multivariadas, como análise fatorial e análise de agrupamentos, é ainda menos comum.

Tal quadro é bastante contrastante com o que se pode verificar do "estado da arte" da aplicação de métodos estatísticos nos principais periódicos de Demografia e Estudos de População nos EUA e Europa (*Demography, Population and Development Review, Population Studies, Genus* etc.). Sem entrar no mérito da qualidade dos trabalhos aqui desenvolvidos, é forçoso reconhecer que a quantificação estatística na pesquisa demográfica brasileira é ainda muito baixa, comparativamente.

Desafios e perspectivas para o uso de métodos estatísticos na Demografia e Ciências Humanas

A discussão sobre os limites e potencialidades da quantificação na pesquisa social nunca teve, no Brasil, a proeminência que já desfrutou no passado na Sociologia norte-americana. Os principais protagonistas desse debate continuam, cada qual, em seus castelos de areia. Estatísticos, alguns descrentes da "cientificidade" da pesquisa nas Ciências Sociais (Popper, 1975), outros descontentes com a baixa motivação de alunos de Ciências Sociais para apreender os fundamentos matemáticos das técnicas, e outros, ainda, receosos da aplicação de métodos em dados "tão

malcomportados" ou "tão anarquicamente coletados", não conseguem criar um fórum de debates sobre questões metodológicas com sociólogos. Entre estes, a crítica à "reificação das cifras" e ao "fetichismo das técnicas" ou a decepção com a utilização de alguns métodos acabam por distanciá-los ainda mais da Estatística.

Minha posição é a de que, embora não seja imprescindível na análise de processos sociais, a aplicação de métodos estatísticos pode ser uma ferramenta útil na exploração das características dos dados e das relações entre variáveis. Pode, inclusive, dar uma sustentação mais forte aos argumentos propostos e iluminar saídas para a crise paradigmática de alguns campos do conhecimento. Como defendem Blalock (1973) e Torgerson (1958), a mensuração e a aplicação de métodos quantitativos nas Ciências Humanas são encaminhamentos metodológicos importantes para se descobrir relações entre fenômenos sociais e se avançar no teste e formulação de hipóteses sobre os fenômenos e a estrutura de causalidade. Mas não deve escapar ao analista que a contrapartida exigida à aplicação dos métodos pode ser muito custosa: o conjunto de dados deve exibir propriedades estatísticas "desejáveis", qualidade menos freqüente em se tratando de dados sociais (Souza, 1993) (2).

Hoje é cada vez mais difícil justificar a não utilização da Estatística pela complexidade operacional e requisitos computacionais exigidos. Francis (1981) relaciona e avalia mais de 60 pacotes estatísticos disponíveis para *mainframes* e microcomputadores por volta de 1980

(2) A respeito do "mal comportamento" dos dados sociais, em contraposição ao "enquadramento" dos dados provenientes da Física e Astronomia, Alves (1981:98) não poderia ser mais feliz: "O rigor das ciências da natureza não se deve, em absoluto, a que elas sejam mais rigorosas e seus métodos mais precisos. Acontece que o bicho com que elas lidam é muito doméstico, manso, destituído de imaginação, faz sempre as mesmas coisas, numa rotina enlouquecedora, freqüenta os mesmos lugares. Tanto assim que é possível prever onde estarão Terra, Sol e Lua daqui a 100.000 anos."

nos EUA. Somente para micros, em abril de 1985, nos EUA, havia pelo menos 40 grandes programas comerciais de análise estatística (Siegel, 1985). Inspirado no SPSS – Statistical Package for Social Sciences –, um grupo de pesquisadores chegou a desenvolver na Universidade Federal do Rio de Janeiro, para uso de alunos e pesquisadores em Ciências Humanas, um *software* (PECS) voltado para a análise estatística básica de dados sociais (Helt *et al.*, 1985).

Mas em que pese a ampliação das facilidades computacionais nas universidades, a “banalização” da operação dos *softwares* estatísticos e a produção de manuais estatísticos para interlocutores externos não afeitos ao formalismo matemático dos textos clássicos da área (3), ainda há algumas dificuldades a superar na disseminação destas técnicas para a investigação social. A correta utilização das ferramentas estatísticas exige, por parte do cientista social, um comprometimento maior com o formalismo matemático subjacente às mesmas e requer, por parte do estatístico, um conhecimento mais abrangente e substantivo da questão social em estudo e uma nova postura epistemológica da ciência. Se de um requer-se a superação da “aversão aos números” (Ruas, 1994), do outro exige-se um comprometimento menos fetichista e “reificador” com o conjunto de dados.

O dado coletado não é a realidade objetiva. A realidade não é passível de ser entendida de forma neutra ou objetiva, como querem alguns; ela é captada sempre a partir do conhecimento anterior e da visão de mundo do observador. Tal como o pescador, o cientista precisa de uma rede ou um anzol (teoria) para que, lançados ao mar (realidade), tragam

espécimes (evidências empíricas) para uma investigação mais detalhada (Alves, 1981). Naturalmente, os espécimes fígados ou recolhidos na rede estão determinados, *a priori*, pelo tamanho do anzol ou largura da malha. As amostras – mesmo as coletadas pelos mais rigorosos critérios probabilísticos – são sempre um subconjunto enviesado da realidade. Além disso, é preciso desmistificar o poder “exploratório” das técnicas estatísticas: elas desvendam aquilo que de alguma forma o conhecimento substantivo do problema em questão sugere como provável ou não. Como bem observa Kuhn (1978), os achados inusitados ou descobertas casuais em meio às experiências nas Ciências Naturais só vieram à tona porque havia uma teoria subjacente que os tornava previsíveis ou completamente impossíveis de ocorrer.

Uma solução pragmática para estas questões epistemológicas da ciência e as disputas interdisciplinares entre a “reificação das cifras” e a “frieza dos números” é a clarividente síntese de Bourdieu (*apud* Paula, 1987):

[“se por um lado] quem fala de dados não pode apenas falar de dados, precisa falar também da realidade em que foram produzidos, saber da realidade em que estão sendo analisados [p. 80], [por outro lado] [...] as interpretações são apenas hipóteses, mesmo quando a evidência é grande; precisam de uma verificação operacional, que pode ser fornecida pela estatística. Os números têm valor probatório, porém só sabem revelar as relações que lhes é dado pesquisar; por outro lado, têm valor heurístico: permitem descobrir relações em que não se teria pensado. As estatísticas são para o cientista social o que a experiência é para o físico: opõem à hipótese a resistência do dado,

(3) Neste sentido, ver Dillon e Goldstein (1984) e a série *Métodos Quantitativos para Cientistas Sociais*, da Fundação Sage, nos EUA (Asher, 1983; Liebetrau, 1983; Kim e Mueller, 1976, entre outros). Apostilas de alguns minicursos oferecidos no âmbito dos Simpósios Nacionais de Probabilidade e Estatística, bianuais, também são bons exemplos nacionais de popularização do “saber estatístico” em outras platéias.

obrigando-o assim a formular novas hipóteses" (p. 76).

Também não se deve deixar de observar que o desenvolvimento da terceira linha de pesquisas a que Preston (1978) se refere no seu texto "The next fifteen years in demographic analysis" – o estudo sócio-econômico-biológico do comportamento das variáveis demográficas – supõe um comprometimento maior com o emprego de modelos estatísticos.

Considerações finais

Procurou-se levantar, nestas notas de pesquisa, algumas evidências empíricas da propalada baixa utilização da Estatística na pesquisa social brasileira. Se na Demografia – onde militam, além de sociólogos, economistas e antropólogos, pesquisadores de formação matemática ou estatística – o uso de métodos estatísticos ainda é residual (porém crescente), não é difícil imaginar o que se passa nas outras disciplinas humanísticas.

A utilização crescente de métodos estatísticos para a análise de dados sociais e demográficos depende da der-

rubada do muro "virtual" que separa o estatístico do sociólogo. Estatísticos mais sensíveis às questões teóricas da pesquisa social e sociólogos mais dispostos a colocar suas premissas teóricas e seus dados empíricos a provas quantitativas poderão viabilizar um salto metodológico significativo na análise de dados sociais e demográficos que se faz no Brasil e revitalizar as desacreditadas disciplinas de Estatística nos cursos de graduação de Ciências Humanas.

Por fim, é prudente esclarecer que não se espera que toda pesquisa no campo das Ciências Humanas, nem a maioria delas, tenha compromisso com a aplicação de métodos estatísticos. O emprego de um modelo estatístico não confere maior confiabilidade ou "cientificidade" às teses e evidências apresentadas, como parecem fazer crer as editoriais de alguns periódicos internacionais da área. A explicitação do uso de um método estatístico em textos de análise de processos sociodemográficos não deveria ser um requisito necessário, muito menos suficiente, para sua avaliação editorial. A questão aqui apontada é que, em várias destas pesquisas, o ferramental estatístico poderia ser muito útil e, no entanto, não é empregado.

Referências bibliográficas

- ALVES, Rubem. *Filosofia da ciência: introdução ao jogo e suas regras*. São Paulo, Brasiliense, 1981.
- ASHER, H. *Causal modeling*. Beverly Hills, Sage Publications, 1983.
- BARBETTA, Pedro. Uma proposta de ensino da Estatística nas Ciências Sociais e Humanas. Comunicação apresentada no 11º Simpósio Nacional de Probabilidade e Estatística, Belo Horizonte, ABE, 1994.
- BLALOCK, H.M. *Introdução à pesquisa social*. Rio de Janeiro, Zahar, 1973.
- DILLON, William R. e GOLDSTEIN, Mathew. *Multivariate analysis: methods and applications*. Nova York, Wiley, 1984.
- FRANCIS, Ivor. *Statistical software: a comparative review*. Nova York, North Holland, 1981.
- HELT, Sonia et al. *Estatística em microcomputadores: uma nova opção para usuários do SPSS*. Rio de Janeiro, LTC, 1985.
- KIM, Jae-On e MUELLER, C.W. *Introduction to factor analysis: what is and how to do it*. Beverly Hills, Sage Publications, 1978.

- KUHN, Thomas S. *A estrutura das revoluções científicas*. São Paulo, Perspectiva, 1978.
- LIEBETRAU, A. *Measures of association*. Beverly Hills, Sage Publications, 1983.
- PAULA, Sérgio Goes de. *Morrendo à Toa: sobre os limites das relações entre mortalidade e desenvolvimento no Brasil*. Tese de Doutorado, Campinas, Instituto de Economia da Unicamp, 1987.
- POLLINGER, K. "Teaching statistics in Sociology". In: MILLER, D.C., *Handbook of research design and social measurement*, Nova York, Longman, 1977, pp. 152-5.
- POPPER, K. "A lógica da investigação científica". In: *Popper*, São Paulo, Ed. Abril Cultural, Coleção Os Pensadores, vol. XLIV, 1975, pp. 263-384.
- PRESTON, Samuel. "The next fifteen years in demographic analysis". In: TAUBER, K.E. et al., *Social demography*, Nova York, Academic Press, 1978, pp. 299-313.
- RUAS, Claudete. O ensino de Estatística nas Ciências Sociais: notas para discussão. Comunicação apresentada no 11º Simpósio Nacional de Probabilidade e Estatística, Belo Horizonte, ABE, 1994.
- SIEGEL, Jay B. *Statistical software for microcomputers: a guide for 40 programs*. Nova York, North Holland, 1985.
- SOUZA, Jorge. *Sobre a natureza das observações e o delineamento observacional*. Brasília, Departamento de Estatística da UnB, mimeo, 1993.
- TORGERSON, Warren S. *Theory and methods of scaling*. Nova York, Wiley, 1958.