

A ESTATÍSTICA COMO LINGUAGEM DE INVESTIGAÇÃO

Manuel Gonçalves Henriques Gameiro*



A perspectiva em que a estatística é entendida (e utilizada)

como “ferramenta” supletiva do processo de investigação deixa de predominar para dar lugar a outra em que é assumida como um sistema lógico gerador de significados - uma “linguagem de investigação”

0 - INTRODUÇÃO

A preocupação fundamental da estatística é transformar o conjunto caótico dos **dados** (os dados em bruto) em **informação** útil, de modo a chegar a conclusões válidas e a facilitar a tomada de decisões adequadas, minimizando, tanto quanto possível, o erro.

Todo o investigador, face a qualquer tipo de problema e em função das suas questões de investigação e/ou hipóteses, necessita colher dados, recorrendo para tal a técnicas e instrumentos diversos.

Quando se utiliza a metodologia quantitativa, esses dados são habitualmente numerosos e difíceis de decifrar. De tal modo que, se não resulta-

rem de um planeamento cuidado e enquanto não forem sujeitos a um tratamento adequado, apresentam-se de uma forma confusa e quase sem sentido. Não raramente, para o investigador menos experiente, alguns deles evidenciam, mesmo, nesta altura, a sua irrelevância ou invalidade para os objectivos em vista: aquilo que alguns autores designam por “lixo estatístico”.

A organização e apuramento, a classificação, o tabelamento, o cálculo de medidas de síntese (de tendência central, de dispersão, quantílicas, de curtose, de simetria) e a análise do ajustamento a determinados modelos de distribuição de frequências, permitem já a sua “leitura”, integrando os diversos sinais e símbolos num quadro com “sintaxe” e “semântica” específicas.

Baseados nesta informação e utilizando técnicas mais complexas, torna-se então possível determinar a força e o sentido das relações (por vezes da influência) entre variáveis. E, sendo os dados, na maior parte dos casos, resultantes do estudo de uma amostra e não da totalidade da população, é ainda possível avaliar até que ponto é muito ou pouco provável que as relações observadas resultem do acaso de serem aqueles e não outros os elementos da amostra seleccionada (significância dos testes de hipóteses), elaborar estimativas para a população e previsões para ocorrências futuras.

Estas últimas, como outras operações estatísticas mais avançadas, baseiam-se em teorias complexas, de elevado nível de abstracção (p. ex. a da teoria das probabilidades) e recorrem a processos e códigos bastante elaborados e singulares, exigindo cálculos complicados e de alcance difícil para um investigador sem particular formação e motivação matemática. Hoje, todavia, a facilidade de acesso e

* Professor - Adjunto na Escola Superior de Enfermagem Dr. Ângelo da Fonseca

Ora, ambas as situações exigem que se considere a operacionalização das variáveis de modo a conseguir um nível de medição adequado, que se construa um instrumento fiável e que se estabeleça um plano de amostragem que permita a selecção de um elenco representativo da população. Tarefas em que a visão estatística pode dar um apoio relevante.

Mesmo assim, é comum os dados depois de recolhidos apresentarem singularidades que originam confusão e dificuldades não previstas, resultantes da sua própria natureza, das características dos elementos observados, das características dos observadores ou de defeitos ou lacunas nos instrumentos utilizados. Nestes casos, QUIVY; CAMPENHOUDT (1992) recomendam como fundamental o estabelecimento de dois "circuitos de retroacção", um entre a "análise das informações" - resultados - e a "construção do modelo de análise" - metodologia - e, outro, entre a "análise da informações" e a "observação" - colheita de dados -. E, sugerem que, face às "divergências" eventualmente observadas, pode ser necessário repensar os conceitos de base, elaborar novas hipóteses ou, mesmo, repetir o processo de colheita de dados.

2 - OS MODELOS ESTATÍSTICOS

O objecto da estatística não são os indivíduos em si mas grupos (de preferência numerosos) que apresentem algumas características em comum. E, as técnicas estatísticas só têm sentido face à **variabilidade** dos elementos nos valores ou categorias que apresentam em cada uma dessas características - variáveis - e do grau de incerteza que sempre

existe em relação a uma nova observação. A utilização da estatística justifica-se pelo carácter não determinístico dos fenómenos em estudo.

O processo de tratamento estatístico dos dados inicia-se por uma série de operações no sentido de definir um ou mais modelos representativos do conjunto estudado. Esses modelos, que constituem as estruturas básicas do "discurso estatístico", apresentam versões simplificadas, "ilustrando" determinados aspectos e "evitando grande número de detalhes que talvez sejam irrelevantes para o problema" (STEVENSON, 1986), podem

representar as variáveis individualmente ou o cruzamento e associação entre duas ou mais variáveis e a sua expressão pode ser:

- **Tabelar** (tabelas de distribuição de frequências, tabelas de contingência, etc);
- **Gráfica** (histogramas, curvas de distribuição, diagramas de dispersão, etc) ou
- **Númerica** (média, mediana, variância, coeficiente de correlação, etc).

Embora sempre parcial, e incompleto cada modelo permite comunicar uma ideia ou conceito, fazer comparações com outros, de origem similar ou standertizados (distribuição normal, distribuição binomial, distribuição exponencial, etc) e, assim, fornecer informações inteligíveis e servir de apoio na tomada de decisões e na avaliação dos seus efeitos.

No entanto, nesta forma privilegiada de expressão estatística, os

dados perdem a sua natureza de pertença individual e passam a constituir elementos "anónimos" de uma série que adquire uma organização própria e característica do conjunto, enquanto tal. Isto é, os modelos estatísticos representam os grupos (amostras ou populações) e não os indivíduos.

Sugerir, por exemplo, que alguma das medidas de tendência central - média mediana ou moda - caracterizam o indivíduo "tipo" ou "normal" é totalmente destituído de sentido. Contudo, é uma tentação comum, particularmente

com a média aritmética, mesmo quando a expressão dessa medida a desaconselha - a média do nº de filhos dos casais que

habitam num determinado bairro pode muito naturalmente ser de 1.70 e, não podemos apelidar de normal um casal (mesmo desse bairro) que tenha tal nº de filhos. Para além do mais, estaríamos a negar a maior justificação para utilizarmos a estatística descritiva: a variabilidade individual dentro dos grupos.

Todavia, sem reduzir os elementos de um conjunto amplo e variado às medidas calculadas a partir dele, é possível fazer estimativas individuais de valores desconhecidos em determinadas variáveis desde que se conheça para esse indivíduo o valor de uma ou mais variáveis que se saiba estarem fortemente correlacionadas com aquela e cuja fórmula de função tenha sido previamente calculada através de métodos estatísticos, bem como, determinar a probabi-

Os modelos estatísticos representam os grupos (amostras ou populações) e não os indivíduos

lidade de um indivíduo com um valor específico pertencer a um determinado grupo cujos parâmetros foram previamente identificados. De qualquer forma, esses processos de inferência singular não deixam de ter um certo grau de risco de erro (em parte, calculável) e incerteza como em qual-

Os resultados não são por si inferência, mas tão só referências para as decisões inferenciais

quer outro método de inferência estatística em que os dados do grupo observado, de facto, nunca se ajustam de um modo perfeito ao modelo estatístico teórico standart correspondente, no qual que se basearam os respectivos cálculos .

3 - A ESTATÍSTICA E OS RESULTADOS

Os resultados de um trabalho de investigação são mais do que o conjunto já organizado dos dados colhidos. Para se poderem designar por resultados têm que ter significado próprio no sentido de darem respostas às questões de investigação e/ou provarem (corroborando ou não) as relações supostas entre as variáveis consideradas na formulação das hipóteses.

E, esse significado, como a seguir procuramos demonstrar, não pode retirar-se directamente do resultado das operações estatísticas. Estes resultados permitem identificar, descrever e elucidar estruturas e relações entre variáveis mas, o significado destas cabe ao investigador atribuí-lo, consi-

derando o modelo teórico que previamente construiu e em função do qual realizou as várias opções metodológicas, incluindo as relativas ao processo de análise estatística (BERQUÓ; 1981; QUIVY; CAMPENHOUDT, 1992).

3.1- Os limites da inferência

Na definição de KERLINGER (1980), "inferência é uma proposição ou generalização derivada pelo raciocínio, de outras proposições ou da evidência". De acordo com esta definição, depreende-se que a segurança e validade de uma inferência dependem sobretudo das premissas em que se baseia e do sistema lógico utilizado na decisão.

Ora, em investigação quantitativa, o sistema lógico em questão passa pela selecção, cálculo e interpre-

distribuição populacional) em que se baseiam. Os resultados só são relevantes na medida em que todos esses factores são coerentes.

E, dado o carácter não determinístico (aleatório) dos fenómenos com que trabalhamos, as técnicas e formulas estatísticas não passam de aplicações mais ou menos complexas da teoria das probabilidades. Consequentemente, os resultados não são per si inferência, mas tão só referências para as decisões inferenciais.

O investigador a quem, como já referimos, cabem em última análise as inferências não pode utilizar tais resultados como universais e absolutos. Antes, deve ponderá-los e interpretá-los ciente dos seus limites:

- **Límite substantivo:** a generalização dos resultados não pode



Figura 2 - O processo de investigação e a validade dos resultados

tação dos resultados de determinados conjuntos de medidas estatísticas e, nos casos específicos dos testes de hipóteses e das outras operações inferenciais, algumas das premissas são já resultado de operações estatísticas (médias, variâncias, frequências, postos, etc.) e outras constituem os pressupostos de carácter teórico-metodológico (plano de amostragem - tipo e dimensão da amostra - nível de medição das variáveis e respectivos modelos de

ultrapassar a fronteira da população representada pela amostra que de facto se estudou e que muitas vezes não coincide com aquela em que se identificou o problema, dado que as contingências do estudo real podem ter originado desvios no plano de amostragem estabelecido (impossibilidade de acesso, recusa de resposta) ou a população não é possível de definir em extensão. Nestes casos, os resultados baseados em tais amostras

só podem ser extrapolados para uma "população fictícia", definida à posteriori e identificada com as características dos elementos estudados.

- **Limite qualitativo:** os resultados dos processos estatísticos inferenciais encerram sempre um certo grau de incerteza e erro. Baseando-se em cálculos de probabilidades, a sua expressão não pode ser de certeza ou de verdade mas, no máximo, de uma "forte convicção" e os erros podem ser minimizados mas são inevitáveis.

Os erros que interferem com as inferências dos processos de investigação condicionando a sua validade, são de dois tipos:

- **Erros aleatórios:** são aqueles que se devem ao acaso, derivam de fontes desconhecidas e não são directamente controláveis pelo investigador, podendo distorcer os dados em qualquer sentido. A minimização deste tipo de erro passa pelo

aumento do tamanho da amostra, pelo aumento da precisão dos instrumentos de medida e por se ter em conta a significância nos testes de hipóteses e os intervalos de confiança nas estimações de parâmetros.

- **Erros sistemáticos:** resultam de enviesamentos, desvios dos dados num determinado sentido e em que a falta de representatividade das amostras, os erros de construção ou limitações dos instrumentos de colheita de dados, a tendenciosidade dos observadores, os processos de categorização/classificação dos dados e a

escolha incorrecta das medidas e/ou provas estatísticas são as principais causas. O conheci-

prova em contrário, a hipótese que vai ser colocada à prova é a de nulidade, isto é, testa-se a

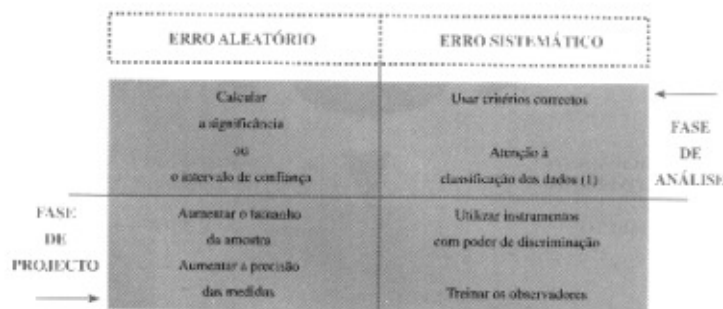


Figura 3 - Estratégias de minimização dos erros (adaptado de HULLEY; CUMMINGS, 1993)

mento profundo da problemática em estudo e a experiência de investigação permitem detectar e reduzir este tipo de erros mais frequente no investigador com menos prática.

Resumindo, baseado em amostras o investigador não pode

demonstrar (ou refutar) uma hipótese de forma absoluta. As hipóteses podem sair mais fortalecidas ou enfraquecidas de um

trabalho de investigação mas nunca definitivamente provadas ou rejeitadas.

3.2 - A interpretação dos resultados - a estatística como "julgamento"

Do ponto de vista metodológico, o processo de inferência estatística aproxima-se muito do inquérito judicial.

Supondo que possa existir qualquer tipo de relação entre variáveis, assim como no processo judicial em que o arguido deve ser considerado inocente até

hipótese de que não existe relação, não existe correlação, não existe diferença, etc.

Do mesmo modo, assim como se têm que reunir provas ou indícios e estes têm de ser suficientes e razoáveis para se atribuir culpabilidade, também a inferência estatística passa pela aplicação de testes, aceitando-se a hipótese alternativa se os valores calculados indicarem que probabilidade da distribuição observada ser resultado do acaso na amostragem é inferior a um valor previamente estabelecido como critério de rejeição da hipótese nula (nível de significância). Só nestes casos é razoável considerar significativas as relações verificadas entre as variáveis: se é muito pouco provável que as associações detectadas na amostra se devam ao acaso de serem aqueles e não outros os seus elementos constituintes então, é muito provável que se devam às relações, existentes de facto, entre as variáveis, na população.

Mesmo assim, os testes estatísticos, como na maior parte dos casos os indícios de crimes não testemunhados, não confirmam com certeza absoluta a existência ou não das relações hipotéticas. A inferência, como o veredito de

As hipóteses podem sair mais fortalecidas ou enfraquecidas de um trabalho de investigação mas nunca definitivamente provadas ou rejeitadas

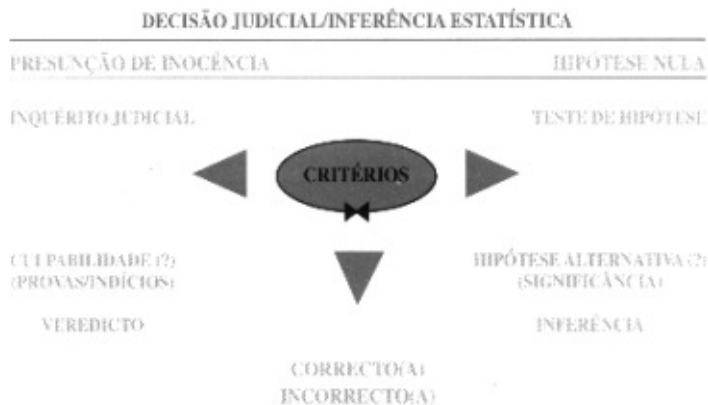


Figura 4 - A inferência estatística como "juízo"

um juiz, não está livre de equívoco. Será correcta se concluir que há relação entre as variáveis quando de facto essa relação existe na população ou se concluir que não há quando de facto ela não existe na população. Em contrapartida, a inferência será incorrecta se concluir que há relação quando ela não existe na população (erro tipo I - falso positivo) ou se concluir que não há relação quando ela de facto existe na população (erro tipo II - falso negativo).

Este tipo de erros podem tanto ser imputados a deficiências do projecto de investigação, como aos efeitos do acaso (aleatórios) não controlados pelo investigador, como, ainda, à natureza das variáveis e das associações realmente existentes. A probabilidade de que num estudo se venha a detectar uma associação entre variáveis depende desde logo da magnitude dessa associação e da maior ou menor variação dos valores na população: quanto menor a variabilidade e maior a magnitude da associação na população, mais fácil será de detectar numa amostra.

Todas estas considerações reforçam a ideia já expressa de que aceitar ou rejeitar uma hipótese

nunca deve ser em definitivo e em absoluto e não pode derivar unicamente dos resultados das provas estatísticas. Deve, também e necessariamente, ter em conta os resultados de outras pesquisas, o processo e o contexto da investigação e as implicações das decisões. Resultados divergentes, determinados a partir de metodologias frágeis e desenvolvidas em contextos mal conhecidos, deverão originar inferências cautelosas, tal como cautelosas deverão ser as decisões nelas baseadas.

4 - CONCLUSÃO

Em qualquer domínio da investigação e também na enfermagem em que a abordagem quantitativa é habitual e as metodologias diversas, reconhecer a estatística como uma linguagem da investigação é fundamental para dela se tirar o maior proveito minimizando os erros e aumentando a validade dos resultados.

Nesse sentido, a estatística deve ser considerada como um universo de referências, necessário ao longo de todo o processo da pesquisa, desde elaboração do projecto até à interpretação dos resultados, não se limitando ao seu papel de "alfaia" no tratamento dos dados colhidos.

Todavia, o sentido das interpretações e o valor das inferências baseadas nos resultados das fórmulas e operações estatísticas dependem da adequação e da coerência nas diversas opções metodológicas e, em última análise, derivam da actividade racional, lógica e ponderada do investigador. Este não se pode iludir com a ideia de que esses resultados são significativos por si e utilizá-los indiscriminadamente sem uma prévia análise crítico/interpretativa no quadro mais global dos objectivos, conceitos e métodos utilizados no processo de investigação.

SV

BIBLIOGRAFIA

- BERQUÓ, Elsa Salvatori et al - *Bioestatística*. São Paulo, E.P.U., 1981.
- GRAWVITZ, Madeleine - *Méthodes des Sciences Sociales*. 8^a Ed. Paris, Editions Dalloz, 1990.
- HULLEY, Stephen B.; CUMMINGS, Steven R. - *Design de la Investigación Clínica*. Barcelona, Doyma, 1993.
- KAPLAN, Abraham - *A Conduta na Pesquisa: Metodologia para as Ciências do Comportamento*. São Paulo, E.P.U. 1975.
- KERLINGER, Fred N. - *Metodologia da Pesquisa em Ciências Sociais*. São Paulo, E.P.U., 1980.
- LOPES, A; Rodrigues - *Introdução à Edição Portuguesa*. In D'AINAUT, Louis - *Conceitos e Métodos de Estatística*, vol.1. Lisboa, Fundação Calouste Gulbenkian, 1990.
- PALLÁS, J. M. Arginón; VILLA, J. Jiménez - *Métodos de Investigación Aplicados à la Atención Primaria de Salud*. Barcelona, Doyma, 1991.
- POLIT D.; HUNGLER B. - *Investigación Científica en Ciencias de la Salud*. 2^a ed. México, Interamericana, 1985.
- QUIVY, Raymond; CAMPELHOUDT, Luc Van - *Manual de Investigación em Ciências Sociais*. Lisboa, Grádiva, 1992.
- STEVENSON, William J. - *Estatística Aplicada à Administração*. São Paulo, 1991.