



Estatísticas, representação e conhecimento

Statistics, representation and knowledge

Oswaldo GUIZZARDI FILHO¹

R E S U M O

As estatísticas são representações, estão no lugar de alguma coisa. Estas representações são produzidas a partir da definição de recortes que procuram simplificar os aspectos do mundo, os quais são o objeto de estudo. Esses recortes têm como finalidade o conhecimento desses aspectos, e expressam-se nas questões constantes do questionário de uma pesquisa, nas categorias que são associadas aos elementos por ela coletados, na definição de um determinado indicador, etc; tais recortes são estabelecidos em detrimento de outros, que também seriam possíveis. O que visualizamos com as estatísticas são números, que representam conjuntos com as mesmas características, expressas em classificações idealizadas pelos pesquisadores que as constróem. Para que as estatísticas possam ser utilizadas, essas escolhas precisam, portanto, estar explícitas, revelando-se nas classificações e conceitos associados à representação, nos segmentos do universo que são escolhidos como objeto de estudo, nos erros que estão associados aos resultados que são divulgados. Se isso não ocorrer, o uso da estatística pode resultar no oposto do que se espera dela, que é o conhecimento.

Palavras-chave: estatísticas, dados estatísticos, representação, conhecimento.

A B S T R A C T

The statistics are representations of something else. These representations are produced from pre-defined segments, chosen to simplify certain aspects of the

¹ Mestre em Ciências da Comunicação. Analista de Projetos Senior da Fundação Seade. Avenida Casper Líbero, 464, 3º andar, 01033-000, São Paulo, SP, Brasil. E-mail: osvaldo@seade.gov.br
Data de recebimento e aprovação 13/9/2004.

world, which are the object of a given study Those segments are intended to comprehend such aspects and are expressed, for instance, in the questions of a questionnaire, in the categories associated to the elements collected by the research and defined by some specific indicator that is unique among several others. The statistics show numbers that represent sets with the same characteristics expressed in the idealized classification built by the researchers. They may only be used if they reveal the researchers' choices, which should be expressed: in classification and concepts associated to the representation; in the segments of the universe chosen as object of study; and in the errors that are associated to the divulged results. If the statistics do not reveal those choices, using them will defeat the very purpose of statistics: to increase knowledge.

Key words: *statistics, statistical data, representation, knowledge.*

INTRODUÇÃO

As estatísticas são uma das formas de se trazer o mundo distante e desconhecido ao alcance daqueles que precisam agir sobre ele. Das pesquisas, dos levantamentos diretos que são hoje os instrumentos básicos das organizações produtoras de estatísticas para a busca do conhecimento sobre este mundo, origina-se uma infinidade de dados que, trabalhados, manuseados, analisados, resultarão em informação e conhecimento.

Para Senra (1998, p.13),

as estatísticas revelam numericamente múltiplos organizados, superando as individualidades e promovendo as individualizações, participando assim ativamente na condução da conduta dos homens; tornando pensável o mundo desconhecido, ausente e distante, acaba por torná-lo governável, vale dizer, as estatísticas fazem-se instrumentos de saber e de poder. De fato, a partir do século XVIII, tiveram papel decisivo na formação da moderna arte de governar, sob a égide do liberalismo, pautado no pensamento econômico, tomado como tecnologia de governo; então, compreende-se que governar é governar os homens, norteando-se suas condutas com vista a uma maior segurança social.

A utilização dos números para, por exemplo, permitir o controle da população não é recente. No Evangelho segundo São Lucas, como bem lembra Senra (1996), está o relato do censo que César Augusto mandou realizar, levando José e Maria a se deslocarem para Belém, onde teria nascido Jesus. Para o autor, no entanto, só recentemente a estatística passou a ser utilizada sistematicamente como instrumento de governo.

O surgimento e o desenvolvimento da estatística como ferramenta de governo ocorreram num período em que se expandia a atividade comercial e eram criados os Estados nacionais, entre os séculos XVI e XVIII, quando, de acordo com Senra (1998), firmou-se a convicção de que era possível a administração da economia de um país.

Através dos números, conseguia-se uma linguagem que permitia comparar e medir a população, a produção e a circulação de mercadorias, bem como os recursos que estavam disponíveis para essa produção, possibilitando dimensionar e controlar o fluxo de produção e circulação de riqueza.

Para Senra (1998), pouco a pouco evoluiu-se da elaboração de quadros sintéticos e descritivos para a confecção de quadros detalhados, usados na administração dos negócios públicos, baseados em números que, ao longo

do tempo, acabam por ser, definitivamente, associados às estatísticas.

Porcaro (2000) vem reforçar a íntima relação existente entre o surgimento da estatística, tal como a entendemos hoje, o Estado e os negócios. De acordo com a autora, os registros de nascimentos e mortes, que se iniciaram no começo do século XVII, na Inglaterra, são considerados as primeiras pesquisas estatísticas. Com a expansão do comércio, começaram a ser gerados registros sobre as transações realizadas, sendo que o primeiro Departamento de Estatísticas Oficiais foi criado em 1695, na Inglaterra, com o objetivo de contabilizar as quantidades e os valores das mercadorias comercializadas, refletindo o crescimento do interesse no uso de informações numéricas para o controle da situação econômica e social (PORCARO, 2000).

Durante muito tempo, as fontes das estatísticas foram os registros administrativos, ou seja, aquelas anotações regularmente feitas pelas empresas (para o acompanhamento de seus empreendimentos e para a prestação de contas ao fisco), pelos governos (para o controle da produção, do comércio e da arrecadação de tributos) e pelas paróquias (para o registro dos nascimentos, casamentos e mortes, que resultaram no chamado “registro civil”, que hoje tem papel fundamental para a observação, por exemplo, do crescimento vegetativo da população, que corresponde ao saldo entre nascimentos e óbitos).

Um longo período foi necessário para que fossem desenvolvidas e implementadas metodologias para a pesquisa de dados estatísticos, que permitissem o controle de todo seu fluxo de execução, desde a coleta até a produção dos resultados finais. É a partir do final da Segunda Guerra Mundial, com o planejamento se impondo para a reconstrução das economias dos países afetados pelo conflito, que a produção de estatísticas se consolidou em organizações

especializadas (SENRA, 1998). Nelas, passam a ser elaborados não só os censos, mas também os agregados da contabilidade nacional, que buscam medir, por exemplo, a produção e a renda, a poupança, os investimentos efetuados num determinado período, fornecendo aos governantes instrumentos para o balizamento de suas ações.

É também, a partir desse período, que se desenvolvem as pesquisas amostrais, possibilitando a realização dos levantamentos em prazo menor do que aquele exigido pela coleta de dados sobre o universo objeto de observação, com custos também menores. Hoje, no Brasil, por exemplo, não são mais levados a campo os Censos Econômicos, que ocorriam de cinco em cinco anos, até 1985, demandando um tempo enorme para o processamento dos dados coletados e para sua divulgação. O Instituto Brasileiro de Geografia e Estatística (IBGE) e outros organismos brasileiros produtores de estatísticas, como a Fundação Seade, de São Paulo, vêm realizando, ao longo dos últimos anos, pesquisas com a mesma natureza dos censos, mas baseadas em amostras. Se, por um lado, essa solução implica prejuízos do ponto de vista da desagregação espacial dos dados num território com as dimensões do brasileiro – não são mais produzidos números relativos à indústria, comércio e serviços nos municípios –, por outro, ganha-se no tempo necessário para que os resultados venham a público e no volume de recursos exigidos para a realização dos levantamentos.

Estatísticas, recortes e representação

Esse processo de refinamento ocorre naqueles lugares que Latour (2000) chama de centrais de cálculo, onde são trabalhadas todas as anotações, todas as coleções de objetos coletados no mundo e que vão permitir seu conhecimento à distância.

Esses registros, amostras, questionários, mapas e formulários, que permitem a transferência do universo de observação para as centrais de cálculo – o que Latour chama de inscrições –, precisam ser manuseados, filtrados e classificados, para que possam, efetivamente, resultar em informações. Ou seja, a partir da infinidade de dados coletados, é necessária a realização de um trabalho que os vá circunscrevendo e extraindo deles elementos que, finalmente, resultem em informação.

Esse é o caminho que vai, num primeiro momento, do questionário para os dados, números isolados, referentes, por exemplo, à quantidade de alunos matriculados num nível de ensino, ou de hospitais e leitos dedicados a determinada especialidade médica. Já em fases posteriores desses dados trabalhados, originar-se-ão nova informação e novo conhecimento, resultantes, por exemplo, de um estudo que constate que o número de leitos hospitalares posto à disposição da população é insuficiente para o atendimento de suas necessidades, considerados os padrões estabelecidos por organismos internacionais, como a Organização Mundial da Saúde (OMS).

Latour (2000) descreve esse trajeto como o passar por inscrições de graus subseqüentes, correspondendo os números registrados no instrumento de coleta às inscrições de primeiro grau. As de segundo grau são geradas, por exemplo, pelo apontamento, a partir do questionário, do sexo, da idade dos que residem num determinado domicílio, dos gastos com alimentação, habitação ou saúde de uma família, do valor produzido e dos custos de empresas dedicadas a uma atividade econômica específica – os microdados. As inscrições de terceiro grau correspondem às totalizações que são feitas a partir dos apontamentos anteriores, de modo que se chegue à população por sexo ou idade de uma localidade, ao valor produzido e ao que é consumido para essa produção num determinado ramo da indústria – aos dados estatísticos, enfim.

As de quarto grau resultam nos gráficos, tabelas e porcentagens elaborados a partir dos resultados do trabalho anterior, e assim sucessivamente, até as inscrições de enésima ordem.

O que se faz ao longo desse processo, desde o momento em que se estabelecem as questões que constarão do questionário de uma pesquisa, às categorias que serão associadas aos elementos por ela coletados, até a definição de um determinado indicador, é estabelecer recortes, que, sucessivamente, procuram simplificar o objeto de estudo, normalmente complexo, com a finalidade de torná-lo compreensível. A busca dessa simplificação avançou muito nos últimos anos, com a produção dos chamados indicadores-síntese, como o Índice de Desenvolvimento Humano (IDH), cuja metodologia foi criada pelo Programa das Nações Unidas para o Desenvolvimento (PNUD). Esses indicadores pretendem condensar, num único número, diversas características das populações de regiões distintas, apontando aquelas de maior e de menor desenvolvimento. No caso do IDH, divulgado desde 1990, as dimensões levadas em conta são a expectativa de vida ao nascer, a alfabetização das pessoas com mais de 15 anos de idade, a taxa de escolaridade e o Produto Interno Bruto (PIB) *per capita*. Para os municípios brasileiros, o IDH pode ser calculado a cada dez anos, com base nos dados dos Censos Demográficos.

No Estado de São Paulo, a Fundação Seade, a pedido da Assembléia Legislativa, desenvolveu metodologia para a produção do Índice Paulista de Responsabilidade Social (IPRS). Com base nas mesmas dimensões utilizadas pelo IDH, mas com metodologia e fontes de dados distintos, o IPRS pode ser calculado em intervalos menores de tempo, já se dispendo de resultados para os anos de 1997 e 2000. Por meio de sua metodologia, os municípios são agregados em grupos com características socioeconômicas semelhantes, evitando-se o cálculo de um índice médio para cada um deles (SÃO PAULO..., 2003).

Esses recortes são definidos em detrimento de outros que também seriam possíveis, revelando a orientação teórica, política, ou até moral daqueles que desenham a pesquisa, ou que estabelecem as classificações que serão utilizadas para “enquadrar” seus resultados, de forma que eles possam ser colocados uns contra os outros, como as cores que podem ser associadas à pele das pessoas, as atividades em que devem ser alocadas as empresas, os tipos de produtos ou serviços que elas fornecem, as ocupações exercidas pelos indivíduos, seu estado civil.

Revelam também o momento em que se vive. No final do século XIX, por exemplo, de acordo com o Relatório do Anno de 1898, que corresponde ao atual Anuário Estatístico do Estado de São Paulo, as mortes eram associadas a moléstias como “marasmo senil”, “morfhêa”, “beri-beri”, “cachesia palustre”, “coqueluche”, “sarampão”, “cancro”, “syphillis”, “tuberculose”, “amollecimento cerebral” (REPARTIÇÃO DE ESTATÍSTICA..., 1900). De acordo com informações fornecidas pela Dra. Rita Barradas, da Santa Casa de São Paulo, a caquexia ou cachesia corresponde ao emagrecimento pronunciado, que era freqüente em muitas síndromes febris como malária, parasitoses anemiantes, tuberculose etc. No século XIX, como muitas doenças não tinham seus agentes causadores claramente identificados, conservava-se a determinação usada pela medicina das espécies (baseada na classificação de sintomas maiores). O amolecimento cerebral ou encefalomalacia também é um sinal e não uma doença, que pode ser provocado por uma série de doenças, geralmente infecciosas. Os filhos, nos registros de nascimento, eram classificados como legítimos ou ilegítimos, e nos registros de casamento informava-se se os cônjuges eram primos, tios e sobrinhos, ou se não tinham parentesco. Já os eleitores exerciam profissões como as de agricultor, artista, letrado, militar, operário, industrial, comerciante, jornalista

e clérigo. A produção das indústrias era classificada em agrícola, extrativa e zootécnica. Dados sobre matrículas e freqüência às escolas, movimento das bibliotecas, jornais publicados, hospícios de alienados e hospitais de caridade apareciam num capítulo denominado “Estatística Moral”.

Se, ao longo desse caminho de produção de sucessivas inscrições, são obtidos elementos que possibilitam um melhor entendimento do mundo para aqueles que detêm os conhecimentos para que esses registros sejam corretamente interpretados, para outros eles não passam de números sem significação. Isso está associado ao que Wersig (1993) chama de credibilidade, fragmentação e racionalização do conhecimento: os resultados dessas pesquisas ou desses registros só podem ser corretamente interpretados por aqueles que possuem o ferramental cognitivo necessário para isso. Para os demais, só resta acreditar, o que fica muito claro, por exemplo, no que diz respeito aos diversos índices de inflação, ao apontarem comportamentos médios de preços quase incompreensíveis para o cidadão comum, que percebe movimentos dos preços totalmente diferentes quando precisa recorrer ao supermercado, à feira, à farmácia, ao dentista. O termo Produto Interno Bruto (PIB), para muitos, parece referir-se a uma entidade quase mística.

Produzir estatísticas significa compartimentar o mundo, dividi-lo em categorias que o tornem compreensível, pelo menos do ponto de vista dos especialistas que definem, por exemplo, a metodologia de uma determinada pesquisa, cujos resultados deverão espelhar os aspectos que se procura caracterizar.

Um bom exemplo dessas diferentes possibilidades de modelização ou de recorte encontra-se nas disparidades existentes entre a metodologia de medição do desemprego da Pesquisa de Emprego e Desemprego (PED), realizada pela Fundação Seade e pelo Departamento Intersindical de Estatísticas e Estudos

Sócio-econômicos (DIEESE), e aquela da Pesquisa Mensal de Emprego (PME), do IBGE. A PED considera três categorias de desemprego, correspondendo o desemprego total à soma dos números correspondentes a esses três grupos (SEADE, 1995, p.24). Essas categorias são as seguintes:

Desemprego Aberto: pessoas sem trabalho nos 7 últimos dias e com procura de trabalho efetiva nos 30 dias anteriores ao da entrevista.

Desemprego Oculto pelo Trabalho Precário: pessoas que realizaram, nos últimos 30 dias, trabalhos precários – algum trabalho remunerado irregular ou trabalho não-remunerado em ajuda a negócios de parentes – e que procuraram substituir este trabalho nos 30 dias anteriores ao da entrevista ou que, não tendo procurado neste período, o fizeram sem êxito até 12 meses atrás.

Desemprego Oculto pelo Desalento: pessoas sem trabalho e com necessidade de trabalhar, porém sem procura efetiva de trabalho por desestímulo do mercado de trabalho ou por circunstâncias fortuitas, mas que apresentaram procura ativa de trabalho, de pelo menos 15 dias, nos últimos 12 meses.

Já a PME (INSTITUTO BRASILEIRO..., 2003, p.5) leva em consideração o desemprego aberto, definindo como desempregadas:

as pessoas sem trabalho na semana de referência, mas que estavam disponíveis para assumir um trabalho nessa semana e que tomaram alguma providência efetiva para conseguir trabalho no período de referência de 30 dias², sem terem tido qualquer trabalho ou após terem saído do último trabalho que tiveram nesse período.

As estatísticas são, portanto, instrumentos de representação, estão no lugar de alguma coisa. Das pesquisas, dos registros administrativos utilizados para a produção de estatísticas e das inscrições que vão sendo realizadas nos centros produtores, originam-se, portanto, dados que farão referência a elementos do mundo que nos cerca, superando as individualidades, como diz Senra. O que visualizamos com os dados estatísticos são números que representam conjuntos com as mesmas características, expressas em classificações idealizadas pelos pesquisadores que as constroem: população por sexo, idade ou cor; produção de aeronaves; óbitos por causa; consumo de fertilizantes; valor adicionado pela indústria de alimentos e bebidas. À medida que se avança no processo de inscrições, passam a ser elaborados tabelas, gráficos e mapas que, ao estabelecerem relações entre os dados, são capazes de gerar novas informações, modificando o conhecimento de seus usuários sobre o assunto que é objeto de pesquisa. Essas informações não são necessariamente elaboradas nos próprios institutos de estatística, podendo ser produzidas pelos próprios usuários, a partir dos dados levantados pelas instituições produtoras.

No entanto, para que as estatísticas possam assumir suas características de informação, é necessário que se estabeleça um processo que possui duas faces, as quais devem se manifestar tanto em sua coleta, quanto em sua disseminação e uso.

Por um lado, precisam estar registrados e muito claros os conceitos e as categorias associados a cada uma das pesquisas, bem como os procedimentos necessários para a verificação da consistência das respostas obtidas, de modo que sejam captados e classificados com a menor distorção possível os fenômenos que se pretende registrar. Por outro,

² Antes de outubro de 2001 o período de referência para a procura de trabalho era de 7 dias.

aqueles que buscam ou recebem as estatísticas devem ter condições de interpretá-las corretamente para usá-las de forma apropriada. Para que isso ocorra, é necessário que se revele a outra face: os usuários precisam estar informados sobre os recortes que foram feitos, o que os resultados da pesquisa contemplam e não contemplam, qual o significado dos conceitos utilizados. A Pesquisa Industrial Anual (PIA), realizada pelo IBGE, por exemplo, só levanta informações para empresas com cinco ou mais pessoas ocupadas, estando, portanto, as microempresas fora de seu universo de avaliação (INSTITUTO BRASILEIRO..., 2002). Se a preocupação do pesquisador é a análise da estrutura da indústria a partir de dados como o valor da produção industrial, certamente os resultados da PIA serão suficientes para os seus propósitos, já que a produção está concentrada nas grandes empresas. Porém, se o seu foco é o emprego industrial, os dados da PIA terão de ser utilizados com um pouco mais de cuidado, já que as microempresas, em alguns setores da indústria, são importantes do ponto de vista da ocupação de mão-de-obra. Informações de 2000, do Cadastro Central de Empresas (CEMPRE), revelam que, para o total das indústrias extrativas e de transformação, a participação das empresas com até quatro pessoas ocupadas no total do pessoal ocupado era de 8%. Na divisão que agrega os fabricantes de confecções e artigos do vestuário, esse percentual era de 15%, enquanto na de edição, impressão e reprodução de gravações, chegava a 17% (INSTITUTO BRASILEIRO..., 2003).

Novamente colocam-se aqui os problemas que Wersig (1993) associa à credibilidade e à fragmentação do conhecimento: as estatísticas resultam de um processo extremamente especializado de concepção e produção, sendo limitado seu alcance em termos do número de pessoas que as utilizam, que, na maioria das vezes, são oriundas de círculos também bastante

especializados de usuários. Nem todos conseguem usá-las levando em consideração suas potencialidades e seus limites.

CONCLUSÃO

As estatísticas são, hoje, um instrumento fundamental para o planejamento e a avaliação dos resultados das ações de governos, empresas e outras organizações, e para o conhecimento dos mais diversos aspectos do mundo que nos cerca. Em nosso dia-a-dia elas são uma presença constante, pois somos bombardeados o tempo todo por uma série de números que procuram nos revelar o crescimento da população, a dimensão da pobreza, o quanto se produziu ou se deixou de produzir num determinado período. São poucos, no entanto, os que conseguem ver nas estatísticas aquilo que elas realmente são: formas de representação e de simplificação do mundo, resultantes de escolhas que implicam no descarte de outras formas também possíveis para essa representação e simplificação: as estatísticas não são um espelho fiel desse mundo.

Para que elas possam ser corretamente utilizadas precisam, portanto, estarem explícitas essas escolhas, que se revelam nas classificações e conceitos associados à representação, nos segmentos do universo que são escolhidos como objeto de estudo, nos erros que estão associados aos resultados que são divulgados. Sem que isso ocorra seu uso pode resultar no oposto do que se espera das estatísticas, que é o conhecimento.

As agências produtoras de estatística necessitam, portanto, ir muito além de sua produção e disseminação. Cabe também a elas um outro papel fundamental, de caráter pedagógico, ao tornar evidentes para os mais diversos tipos de públicos as potencialidades e limites das estatísticas para a compreensão de um mundo cada vez mais complexo.

REFERÊNCIAS

- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Cadastro central de empresas 2000*. Disponível em: <ftp://ftp.ibge.gov.br/Economia_Cadastro_de_Empresas/2000/Tabelas_2000.zip>. Acesso em: 26 fev. 2003.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa industrial anual 2000: empresa*. Rio de Janeiro: IBGE, v.19, n.1, 2002. 260 p.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa mensal de emprego: notas metodológicas*. Disponível em: <http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pme_nova/pmemet1.pdf>. Acesso em 23 set. 2003.
- LATOUR, B. *Ciência em ação: como seguir cientistas e engenheiros sociedade afora*. São Paulo: Unesp, 2000. 438 p.
- PORCARO, R.M. *Produção de informação estatística oficial na (des)ordem social da modernidade*. 2000. 186p. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação/ECO, Universidade Federal do Rio de Janeiro/UFRJ; Rio de Janeiro, 2000.
- REPARTIÇÃO DE ESTATÍSTICA E ARCHIVO DO ESTADO DE SÃO PAULO. *Relatório do ano de 1898*. São Paulo, 1900.
- SÃO PAULO (Estado). Assembléia Legislativa; Fundação Sistema Estadual de Análise de Dados/Seade. *O estado dos municípios 1997-2000: índice paulista de responsabilidade social: Região Metropolitana de São Paulo*. São Paulo: Seade, 2003. 102 p.
- SEADE. *Pesquisa de emprego e desemprego: conceitos, metodologia e operacionalização*. São Paulo: Seade, 1995. 86p.
- SENRA, N.C. *A coordenação da estatística nacional: o equilíbrio entre o desejável e o possível*. 1998. 178p. Tese (Doutorado em Ciência da Informação). Escola de Comunicação/ECO, Universidade Federal do Rio de Janeiro/UFRJ; Rio de Janeiro, 1998.
- SENRA, N.C. Governamentalidade, a invenção política das estatísticas. *INFORMARE – Cad. Prog. Pós-Grad. Ci. Inf.*, Rio de Janeiro, v.2, n.1, p. 88-95, 1996.
- WERSIG, G. Information science: the study of postmodern knowledge usage. *Information Processing & Management*, v.29, n.2, p. 229-239, 1993.