A MATRIZ DE DADOS E A METODOLOGIA DA PESQUISA EM CIÊNCIA DA INFORMAÇÃO

Gilda Olinto do Valle Silva
Escola de Comunicação
Universidade Federal do Rio de Janeiro
22290 Rio de Janeiro, RJ

RESUMO

As três principais fases da maioria das pesquisas em Ciência da Informação são vistas aqui como três maneiras de se focalizar uma matriz de dados. Na primeira fase, o pesquisador define uma matriz vazia, ao selecionar variáveis e os casos a serem estudados. Na segunda fase, o pesquisador faz o trabalho de campo que pode ser visto como passos necessários para preencher uma matriz de dados. Na terceira fase da pesquisa a matriz de dados é analisada. Abordando os assuntos e problemas envolvidos em dada uma dessas três fases, pretende-se contribuir para um programa de metodologia de pesquisa que visa o maior rigor e eficiência no processo da pesquisa.

1 – INTRODUÇÃO

A maioria dos estudos empíricos na área de Ciência da Informação recebe, ao menos parcialmente, uma abordagem quantitativa. Isto é, estes estudos trabalham com conceitos medidos através de um esquema classificatório que permite registrar a sua variação. Por exemplo, num estudo de usuário pode-se estar interessado em trabalhar com o conceito de "uso de um serviço de informação" e estabelecer como registrar a variação deste uso, de forma que se possa determinar se o entrevistado utiliza-o ou não, utiliza-o pouco ou muito, etc. Os conceitos assim trabalhados podem ser denominados **variáveis** que são obtidas através de diversos tipos de instrumentos de coleta de dados como questionários, no caso de estudos de usuários, ou fichas, no caso de análise de citações. Esses instrumentos de coleta de dados registram a variações desses conceitos em todos os **casos estudados**, que podem ser indivíduos, instituições, artigos de periódicos, etc.

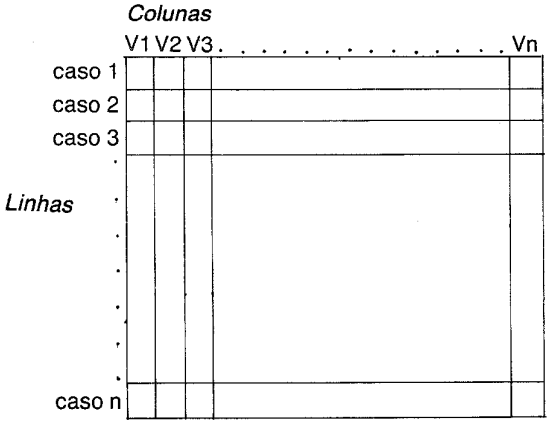
Com o objetivo de contribuir para um maior rigor e eficiência da atividade de pesquisa em Ciência da Informação, enfatiza-se, neste texto, que um curso introdutório de metodologia de pesquisa gira em torno de um conceito básico: matriz de dados. Uma matriz de dados é composta de linhas e colunas: nas colunas estão colocadas todas as variáveis incluídas no estudo (*n* variáveis), nas linhas da matriz de dados estão todos os casos estudados (*n* casos), conforme indicado no Quadro 1.

A matriz de dados contém todas as informações necessárias ao pesquisador. Pode-se dizer, portanto, que a pesquisa e a atividade didática focaliza a matriz de dados sob ângulos diferentes, conforme a fase da pesquisa que está sendo considerada. Destaca-se a seguir três maneiras básicas de se abordar a matriz de dados que correspondem às três principais fases da pesquisa.[1]

Na primeira fase da pesquisa define-se a matriz de dados vazia. É quando se decide sobre os principais conceitos com que se deseja trabalhar, como estes serão mensurados (esses dois tópicos definem as **colunas** da matriz de dados) e que casos serão estudados (a decisão sobre os casos incluídos no estudo define as **linhas** da matriz de dados).

A segunda fase da pesquisa centraliza-se no preenchimento dessa matriz de dados. É quando se faz o trabalho de campo e a codificação dos dados.

Quadro 1 — **Matriz de Dados**



Na terceira fase, o pesquisador volta-se para a análise da matriz já preenchida.

Cada uma dessas fases envolve discussões bastante complexas que podem fornecer assunto para diversos cursos de metodologia de pesquisa. A decisão sobre os casos que serão submetidos ao estudo implica, geralmente, na seleção de uma amostra. Isto significa entrar num ramo da matemática com aspectos metodológicos bastante complexos.

Entretanto, o problema da amostragem pode ser considerado pequeno quando comparado às dificuldades de se abordar, num período de, às vezes, apenas um semestre, os problemas relativos ao equacionamento do problema da pesquisa, à mensuração, à coleta e à análise dos dados.

Para dar subsídios a um programa de metodologia de pesquisa apresenta-se aqui, sumariamente, os assuntos e problemas envolvidos no trabalho didático com as três principais fases da pesquisa, isto é, nas três maneiras de se considerar a matriz de dados.

2 – A MATRIZ DE DADOS VAZIA

A decisão sobre que variáveis serão incluídas na pesquisa e que casos serão estudados é o que se denonima definição de uma matriz de dados vazia.

Esta é geralmente a fase mais difícil em qualquer pesquisa e, também, em um curso de metodologia sobre o assunto.

Aborda-se, inicialmente, questões relativas à definição das colunas da matriz de dados que envolve, fundamentalmente, a discussão de dois tópicos: a identificação dos conceitos básicos envolvidos na pesquisa e o estabelecimento dos critérios utilizados na mensuração. A seguir são apresentadas questões relativas à definição das linhas da matriz de dados que giram em torno do problema da amostragem.

2.1 – *A IDENTIFICAÇÃO DOS CONCEITOS BÁSICOS*

Freqüentemente os cursos de métodos quantitativos são criticados pelo fato de não discutirem a problemática da pesquisa, isto é, o que se deseja compreender no processo de investigação e que fatores podem contribuir para esta compreensão.

De fato esta é uma etapa fundamental em qualquer pesquisa. É a partir do equacionamento do problema que se identificam os conceitos básicos a serem mensurados e que se estabelece como esses conceitos serão, posteriormente, tratados na fase de análise dos dados.

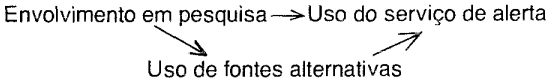
Freqüentemente, especialmente em estudos de usuário, faz-se um caminho na contra-mão. A partir de discussões vagas, elabora-se perguntas de questionários, sem considerar que essas perguntas devem refletir os conceitos básicos com que se deseja trabalhar. Conceitos esses que devem ser identificados antes da elaboração do instrumento de coleta de dados, assim como explorada a relação que possa existir entre eles.

Qualquer que seja o método utilizado pelo professor, para que o aluno experimente em sala de aula a fase da definição do problema de pesquisa, o seu papel é o de fazê-lo refletir (se possível produzindo um texto) sobre o problema e, ao final, fazê-lo identificar os principais conceitos envolvidos e como eles se relacionam. Procura-se, assim, formar um modelo hipotético que se torna a base de todas as demais etapas da pesquisa.

Para utilizar um exemplo bem simples, imagine-se um estudo voltado para a avaliação de um serviço de alerta recentemente introduzido num instituto de pesquisa em Física. A análise cuidadosa do problema permite o equacionamento da seguinte hipótese de trabalho: o serviço está tendo pouca utilidade no instituto (a despeito dos recursos humanos e materiais nele envolvidos). O serviço está sendo valorizado e utilizado apenas pelos profissionais que não estão realmente envolvidos em pesquisas. Os pesquisadores estariam se atualizando através de uma fonte de informação secundária alternativa: o *Physics Abstracts*. O uso regular dessa outra fonte secundária garante a atualização bibliográfica do pesquisador e torna supérflua a prestação do serviço em questão.

Define-se, assim, três fatores básicos envolvidos no problema de pesquisa: o uso do serviço de alerta, o envolvimento em pesquisa e o uso de fontes alternativas de informação. Pelo equacionamento do problema sabe-se, também, que o envolvimento em pesquisa e o uso de fontes alternativas são fatores que podem explicar o uso do serviço.

Entretanto, esses dois fatores explicativos não são independentes, mas relacionados entre si (envolvimento em pesquisa explica uso de fonte alternativa), permitindo a proposição do seguinte modelo hipotético:



Neste modelo "envolvimento em pesquisa" pode ser denominado de fator antecedente, "uso de fonte alternativa" de fator interveniente e "uso do serviço de alerta" fator dependente.

A construção de modelos desse tipo é assunto de difícil abordagem em sala de aula. Os alunos sem experiência prévia de pesquisa têm dificuldade de conceber relações deste tipo apenas como uma esquematização da hipótese de trabalho. Tendem a visualizar tais relações como intrínsecas e generalizáveis para qualquer situação de pesquisa.

2.2 – *OS CRITÉRIOS DE MENSURAÇÃO*

Estabelecidos os principais conceitos envolvidos no estudo e como se relacionam, passa-se a colocar a questão da mensuração.

Aborda-se aqui o problema de se detectar, através de um instrumento de coleta de dados como o questionário, os conceitos em que se está interessado. É a fase da construção de variáveis que requer especial cuidado por parte do professor. Os alunos, que são capazes de desenvolver tarefas bastante complexas em fases posteriores da pesquisa têm, geralmente, dificuldades nestas questões iniciais.

A questão da mensuração é de fato complexa. Considere-se, por exemplo, que cada um dos conceitos básicos envolvidos no exemplo citado pode ser mensurado de diversas maneiras, tanto pelo aspecto da formulação das questões, quanto pelo aspecto das opções de resposta.

Para mensurar, por exemplo, o uso do *Physics Abstracts* pode-se optar por questões genéricas (você costuma consultar o *Physics Abstracts*?), fazer referência ao uso específico desta fonte de informação para a atividade de pesquisa (você consulta o *Physics Abstracts* para as suas atividades de pesquisa?), fazer referência a um incidente crítico (Na última vez que fez uma busca bibliográfica, que fontes de informação utilizou?) etc.

Essas diversas opções podem ser utilizadas simultaneamente, dando origem a diversas questões para a mensuração do mesmo conceito.

Ao lado da complexidade envolvida na decisão sobre as dimensões de cada conceito que serão consideradas na mensuração, há a questão da construção da frase em cada pergunta para evitar a dubiedade e, assim, garantir a validade do instrumento de coleta de dados.

Aos cuidados necessários na formulação das perguntas, seguem-se os cuidados necessários nas opções de respostas oferecidas ao entrevistado quando se utiliza perguntas de tipo fechado. Através dessas opções, estabelece-se o nível de mensuração (nominal, ordinal, intervalar, de razão) de cada variável considerada, com profundas implicações para a posterior análise dos dados.

É importante que o aluno adquira noções básicas sobre níveis de mensuração para saber optar por um nível mais poderoso quando isso lhe for facultado.

Terminada a fase de construção de variáveis estará pronto o questionário, ou outro instrumento de coleta de dados, e estarão definidas as colunas da matriz de dados.

2.3 – *AMOSTRAGEM*

A decisão sobre as linhas da matriz de dados – quem e quantos serão os casos incluídos no estudo – implica numa definição precisa da população que se deseja conhecer e, geralmente, na extração de uma amostra representativa desta população. A amostragem abre um leque de assuntos que podem ser abordados num curso de metodologia de pesquisa.

Em primeiro lugar é útil considerar os aspectos básicos da relação entre amostra e população através dos princípios de probabilidade. Somente com essa base inicial se poderá compreender a importância do rigor no processo de amostragem para garantir a representatividade da amostra, e, ao mesmo tempo, acreditar no uso da amostra. Os procedimentos de estimação de parâmetros e teste de hipótese, que serão utilizados na fase de análise de dados, necessitam também de uma abordagem introdutória à teoria de amostragem.

Além da garantia da representatividade, que é a preocupação básica deste tópico sobre amostragem, deve-se considerar também a questão do tamanho mínimo da amostra: a opção de se trabalhar com amostras muito reduzidas, porque tecnicamente representativas de pequenas populações, pode ser prejudicial à fase de análise dos dados.

Obviamente, um curso introdutório, de carga horária pequena, não permite a extensão deste tópico, nem capacitar o aluno a solucionar problemas de amostragem. Considera-se apenas a seleção da amostra em situações simples e específicas dos exemplos utilizados em sala de aula.

Com o instrumental necessário para a seleção de uma amostra, já se pode definir as linhas da matriz de dados, isto é, os casos a serem incluídos no estudo. E considerando que os tópicos necessários à definição das colunas da matriz já foram abordados, pode-se observar que, a esta altura, já se dispõe dos elementos básicos para identificar uma matriz de dados vazia, especificando as variáveis e os casos a serem estudados.

3 – O PREENCHIMENTO DA MATRIZ DE DADOS

O preenchimento da matriz de dados é o resultado do trabalho de campo e da codificação dos dados.

São bem conhecidos os cuidados que é preciso tomar durante o trabalho de campo. Na aplicação das entrevistas e no envio de formulários procura-se garantir um alto retorno e o correto preenchimento dos itens constantes do instrumento de coleta.

Alguns itens a considerar, num curso de métodos, para o sucesso do trabalho de campo são: a identificação das entrevistas, o controle do entrevistador, o retorno a campo.

Passa-se, então, a instruções sobre a fase da codificação cujo objetivo é transformar cada formulário ou questionário preenchido numa linha da matriz de dados.

O alcance desse objetivo de maneira rápida, após o trabalho de campo, implica em que as perguntas sejam fechadas e, se possível, pré-codificadas. Esta orientação não é, entretanto, imprescindível nem descarta o uso de perguntas abertas que muitas vezes podem ser utilizadas, inclusive para testar a validade de perguntas fechadas.

O uso do livro de códigos é outro passo indispensável num curso de métodos. Ele é utilizado para as perguntas abertas ou quando as opções de resposta, embora previstas, envolvem diversas alternativas que não podem ser incluídas no instrumento de coleta.

Estabelecidas as regras de codificação e feita a codificação dos dados está pronta a matriz de dados, na qual se baseia a fase final da pesquisa. Esta matriz pode ser montada manualmente ou constar de um arquivo em computador.

4 – A ANÁLISE DA MATRIZ DE DADOS

A análise dos dados se inicia com uma leitura vertical das colunas da matriz de dados. Através desta leitura observa-se, isoladamente, cada variável incluída na análise, procurando-se identificar se o seu comportamento é condizente com as questões introduzidas na fase inicial da pesquisa, isto é, as hipóteses levantadas. Esta análise inicial permite, também, verificar se os conceitos mensurados têm um mínimo de variação, se estes se constituem realmente em variáveis que possam ser utilizadas em testes de modelos. No exemplo dado, os resultados das variáveis que medem o conceito "uso do serviço de alerta" são fundamentais. A não variação das medidas utilizadas para mensurar este conceito pode indicar que não se deve prosseguir no teste do modelo proposto.

A capacitação dos alunos para essa leitura vertical da matriz de dados envolve noções gerais de estatística descritiva que são instrumentos utilizados para sintetizar as informações sobre cada variável que está dispersa na matriz de dados. Aborda-se aqui as medidas descritivas (proporções, razões, medidas de tendência central e dispersão) e a elaboração de tabelas e gráficos de freqüência.

Um aspecto importante a considerar na descrição das variáveis é a sua recodificação, que pode ser necessária para a melhor apresentação das principais tendências identificadas. Este procedimento pode ser adotado quando uma variável apresenta, por exemplo, grande concentração de casos em apenas algumas das opções de resposta colocadas pelo pesquisador.

O treinamento em estatística descritiva toma razoável tempo de aula, mas é de fácil absorção pelos alunos.

Assunto mais complexo é introduzido quando se passa da leitura vertical inicial para a leitura parcialmente horizontalizada da matriz de dados. É quando se procura identificar covariações que permitem teste de modelos. No exemplo dado, espera-se que o uso dos serviços de alerta esteja relacionado a envolvimento em pesquisa e, também, ao uso de fontes alternativas de informação. Estas relações se refletirão na covariação dos dados da matriz, o que implica nesta leitura horizontalizada.

O instrumental necessário para analisar essas covariações e, inicialmente, a montagem de tabelas de dupla entrada (para o caso de variáveis medidas a nível nominal) e do diagrama de dispersão (para variáveis medidas em escala de razão) com os testes estatísticos, apropriados a estas distribuições bivariadas. Cita-se como exemplos básicos de testes estatísticos o X^2, (qui-quadrado) as medidas de associação nele baseadas e no r^2 (coeficiente de correlação).

Para a compreensão desses testes estatísticos são necessários as noções de probabilidade e amostragem previamente introduzidas.

Mesmo num curso introdutório é interessante que se aborde a questão das relações entre mais de duas variáveis. Mesmo em exemplos simples, como os que foram utilizados no início deste texto, utiliza-se modelos envolvendo mais do que duas variáveis que necessitam ser trabalhadas simultaneamente. Considera-se, portanto, útil que se introduza, ainda que em linhas gerais, algumas noções de análise de modelos mais complexos, incluindo pelo menos três variáveis. Cita-se aqui a construção de tabelas cruzadas com o uso de variáveis de controle e, também, a introdução e modelos lineares com mais de duas variáveis e o cálculo de coeficientes de correlações parciais.

Somente com esses instrumentais básicos para a análise da matriz de dados, equipa-se o aluno para exercer com certa autonomia a atividade de pesquisa.

Este conhecimento é também necessário para o diálogo com pessoal técnico em outras áreas, como a estatística e a computação, que possa estar assessorando a pesquisa. O domínio desse instrumental básico permite, também, ao profissional amplo uso de sistemas de análise de dados por computador. Os "pacotes" estatísticos disponíveis são conversacionais e sua utilização extremamente simples, desde que se conheça um pouco de análise de dados, e, ainda, infelizmente, um pouco de inglês.

5 – CONCLUSÃO

O presente texto tem por objetivo salientar que a atividade de pesquisa e a experiência didática focalizam, sob diferentes ângulos, a matriz de dados.

Enfatiza-se, inicialmente, que uma pesquisa de de usuários, assim como qualquer levantamento de dados baseado em formulário, questionário, ou fichas, centraliza-se em torno da definição, preenchimento e análise da matriz de dados.

Conseqüentemente, experiência didática é uma tentativa de transmitir informações necessárias à passagem por essas três etapas. Aborda-se, então, assuntos e problemas envolvidos na transmissão de conhecimentos em cada uma dessas etapas.

Convém salientar que a ênfase dada, no texto, à primeira fase da pesquisa, isto é, a definição da matriz vazia, especificamente a questão da identificação de conceitos básicos, não implica em que um curso de metodologia deva, necessariamente, estender-se mais nessa fase da pesquisa. A ênfase dada, aqui, a essa fase inicial reflete apenas a preocupação em mostrar que é preciso investir bastante em considerações metodológicas antes do início do levantamento dos dados. Além disso, os assuntos abordados, sobre a fase de análise de dados seguem, em linhas gerais, o que vem sendo tratado em livros introdutórios de estatística.

Esse texto reflete apenas uma preocupação em contribuir para a simplificação, a sistematização e o rigor nas atividades de ensino e pesquisa em Ciência da Informação, especificamente estudo de usuário, análise de citações e outros estudos de comunicação científica, em que se obtém uma série de informações sobre indivíduos ou outra unidade de análise, como um artigo de periódico. Uma das preocupações, neste sentido, é facilitar a formação e o uso de bancos de dados de pesquisa que possam ser utilizados, quantas vezes for desejado, por profissionais e estudantes na área de informação.

AGRADECIMENTOS

A autora agradece à professora Regina Célia Montenegro de Lima pelas sugestões e pela revisão cuidadosa do texto.

Artigo recebido em 13 de novembro de 1987.

REFERÊNCIAS BIBLIOGRÁFICAS

1 WONNACOTT, T. H. & WONNACOTT, R. J. *Introdução à estatística*. Rio de Janeiro, Livros Técnicos e Científicos, 1980.

2 *BABBIE, E. R. The practice of social research.* Belmont, Wadsworth, 1979.

3 BLALOCK, H. M. *Theory construction*: from verbal to mathematical formulations. New Jersey, Prentice-Hall, 1969.

4 BUCHA, C. & HARTER, S. P. *Research methods in Librarianship*. New York, Academic Press, 1980.

5 FOX, D. I. & GUIRE, K. E. *Midas*: Michigan interactive data analysis system. Statistical Research Laboratory. University of Mychigan, 1979.

6 NIE, N.H. et alii. *SPSS*: statistical package for the social sciences. New York, McGraw-Hill, 1975.

7 STINCH COMBE, A.L. *Constructing social theories*. New York, Harcourt, Brace & World, 1968.

8 KERLINGER, F. N. *Metodologia da pesquisa em Ciências Sociais*. São Paulo, EPU-EDUSP, 1979.

The data matrix and research methods in Information Science

ABSTRACT

The three main phases of most research projects in Information Science are seen here as three ways of looking at a data matrix. During the first phase the researcher defines an empty matrix by selecting variables and cases to be studied. During the second phase the researcher does his field work which can be seen as steps necessary to fill in a data matrix. In the third phase the researcher analyses a data matrix. Considering the main subjects and problems involved in each phase we intend to contribute to a program in research methods which aims at more rigor and efficiency in the research process.