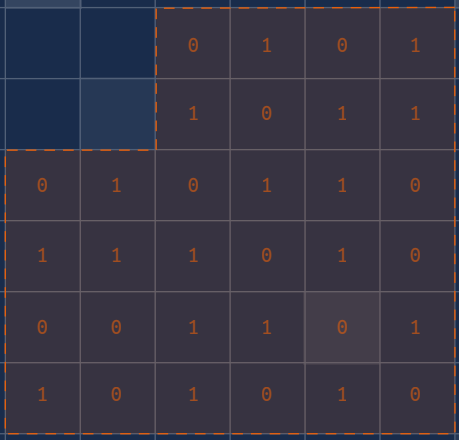
PESQUISA QUANTITATIVA

NA CIÊNCIA DA INFORMAÇÃO



Bianca Savegnago de Mira
Marcelo Schiessl
Milton Shintaku

Editora Ibict

Editora
Ibict

Pesquisa Quantitativa na Ciência da Informação

PRESIDÊNCIA DA REPÚBLICA

Luiz Inácio Lula da Silva
Presidente da República

Geraldo José Rodrigues Alckmin Filho
Vice-Presidente da República

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

Luciana Santos
Ministra da Ciência, Tecnologia e Inovação

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

Tiago Emmanuel Nunes Braga
Diretor

Carlos André Amaral de Freitas
Coordenador de Administração – COADM

Ricardo Medeiros Pimenta
Coordenador de Ensino e Pesquisa em Informação para a Ciência e Tecnologia – COEPI

Henrique Denes Hilgenberg Fernandes
Coordenador de Planejamento, Acompanhamento e Avaliação – COPAV

Cecília Leite Oliveira
Coordenadora Geral de Informação Tecnológica e Informação para a Sociedade – CGIT

Washington Luís Ribeiro de Carvalho Segundo
Coordenador Geral de Informação Científica e Técnica – CGIC

Alexandre Faria de Oliveira
Coordenador Geral de Tecnologias de Informação e Informática – CGTI

Milton Shintaku
Coordenador de Tecnologias para Informação – COTEC

Gustavo Silva Saldanha
Divisão de Editoração Científica - DIECI

Editora
Ibict

Pesquisa Quantitativa na Ciência da Informação

Bianca Savegnago de Mira
Marcelo Schiessl
Milton Shintaku

Brasília
2026

Esta obra é licenciada sob uma licença Creative Commons – Atribuição CC BY 4.0 , sendo permitido que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original.

© 2026 Editora Ibict

CONSELHO EDITORIAL
Gustavo Silva Saldanha
Luana Farias Sales
Milton Shintaku
Franciele Garcês
Leyde Klébia Rodrigues da Silva
Stella Moreira Dourado
Daniel Strauch

COMITÊ EDITORIAL
Tiago Braga
Milton Shintaku
Henrique Denes
Cecília Leite Oliveira
Ricardo Pimenta
Leda Cardoso Sampson Pinto
Carlos André Amaral de Freitas
Marcel Souza
Alexandre Oliveira
Washington Segundo
Alexandre Faria de Oliveira
Emanuelle Torino

COMITÊ CIENTÍFICO
Ania Rosa Hérnandez Quintana
Fernanda do Valle
María Arminda Damus
Martha Sabelli
Natalia Duque Cardona
Vinicios Menezes
Carlos Alberto Ávila Araújo
Murilo Silveira
Sonia Caregnato
Marcia Regina

EQUIPE TÉCNICA

Normalização:
Ingrid Torres Schiessl

Diagramação:
Alisson Eugênio Costa

Capa:
Alisson Eugênio Costa

Dados Internacionais de Catalogação na Publicação (CIP)

M671p Mira, Bianca Savegnago de
Pesquisa Quantitativa na Ciência da Informação / Bianca Savegnago de Mira, Marcelo Schiessl, Milton Shintaku. --
Brasília: Editora Ibict, 2026.
109 p. : il.
Modo de acesso: WWW
Publicação digital (e-book) no formato PDF. [1,24 MB]
ISBN 978-65-82912-62-7
DOI: 10.22477/9786582912627

1. Pesquisa Quantitativa. 2. Ciência da Informação. I. Schiessl, Marcelo. II. Shintaku, Milton. III. Instituto Brasileiro de Informação em Ciência e Tecnologia. IV. Título.

CDU 002:303.023

Como citar:
MIRA, Bianca Savegnago de; SCHIESSL, Marcelo; SHINTAKU, Milton. **Pesquisa Quantitativa na Ciência da Informação**. Brasília: Editora Ibict, 2026. 108p.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto Brasileiro de Informação em Ciência e Tecnologia ou do Ministério da Ciência, Tecnologia e Inovação.

Endereço:
Ibict - Instituto Brasileiro de Informação em Ciência e Tecnologia
Setor de Autarquias Sul (SAUS), Quadra 05, Lote 06, Bloco H – 5° andar. CEP: 70.070-912 - Brasília, DF.

Sumário

Apresentação 7

Prefácio 9

1 Introdução 14

1.1 Estatística 16

1.2 Estudos estatísticos na Ciência da Informação 18

2 Conceitos gerais 22

2.1 População ou Universo 23

2.2 Amostra 23

2.2.1 Técnicas de Amostragem Probabilística 25

2.2.2 Técnicas de Amostragem Não Probabilística 26

2.2.3 Erro Amostral e Cálculo do Tamanho Mínimo de Amostra 27

2.3 Variável 29

3 Como fazer um estudo quantitativo na Ciência da Informação 32

3.1 Planejando o estudo 34

3.1.1 Definindo Objetivo 35

3.1.2 Definindo População e Amostra 36

3.1.3 Definindo o Instrumento de Coleta 43

3.1.4 Definindo o Modelo de Apresentação dos Dados 45

3.1.5 Definindo o Modelo de Análise 46

3.2 Coletando os dados 48

3.2.1 Origem dos dados 49

3.2.2 Ética e proteção dos dados 49
3.2.3 Procedimentos de coleta na prática 51
3.2.4 Instrumentos, pré-teste e estudo-piloto 56
3.2.5 Execução e qualidade da coleta 58

3.3 Tratando os dados 59
3.3.1 Organizando a base de dados 59
3.3.2 Verificando erros e inconsistências 61
3.3.3 Codificando e padronizando os registros 62
3.3.4 Tratando dados ausentes e registros problemáticos 63
3.3.5 Documentando as transformações 65

3.4 Conhecendo os dados 67
3.4.1 Frequências, proporções e tabelas 67
3.4.2 Medidas de posição 71
3.4.3 Medidas de dispersão 75
3.4.4 Distribuição dos dados 78

3.5 Representando os dados 82
3.5.1 Organizando dados em tabelas e quadros 82
3.5.2 Escolhendo o gráfico adequado 85
3.5.3 Utilizando outras formas de representação 88

4 Erros frequentes 91

4.1 Incoerência entre problema, objetivos e dados 92

4.2 Generalizar além dos limites da amostra 94

4.3 Falhas na coleta dos dados 95

4.4 Utilizar tratamento ou análise incompatível com os dados 96

4.5 Produzir representações ou interpretações equivocadas 97

5 Considerações Finais 99

Referências 102

Sobre os autores 107

APRESENTAÇÃO

Este trabalho resulta de estudos desenvolvidos no âmbito de projeto firmado entre a Secretaria Especial de Saúde Indígena (Sesai), do Ministério da Saúde, e o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), vinculado ao Ministério da Ciência, Tecnologia e Inovação. Entre as atividades realizadas no projeto, encontram-se estudos com indicadores que requerem métodos quantitativos orientados pela Ciência da Informação.

Como resultado, surgiu o questionamento acerca da escassez de obras voltadas aos estudos quantitativos na Ciência da Informação, embora essa abordagem esteja presente em muitas pesquisas da área. Nesse contexto, a percepção dessa lacuna, somada à realização de diferentes estudos quantitativos amparados pela Ciência da Informação, motivou o desenvolvimento desta obra.

Sem a pretensão de esgotar o tema, esta obra apresenta fundamentos teóricos e orientações práticas para a realização de pesquisas quantitativas no âmbito da Ciência da Informação. Para tanto, está organizada em cinco capítulos. Após a introdução, apresentam-se os conceitos necessários à compreensão dos estudos quantitativos. Em seguida, descrevem-se as etapas de sua realização, desde o planejamento e a coleta até o tratamento, o conhecimento e a representação dos dados. Posteriormente, discutem-se erros frequentes, com o intuito de alertar o leitor para desafios e decisões que podem comprometer a pesquisa. Por fim, as considerações finais retomam os principais pontos abordados e reforçam a importância do rigor metodológico na produção e na interpretação dos resultados.

Assim, espera-se apoiar pesquisadores que assumem os desafios inerentes aos estudos quantitativos ou mistos, auxiliando-os no planejamento e na execução de suas pesquisas.

Bons estudos!

Os autores

PREFÁCIO

A obra “Pesquisa Quantitativa na Ciência da Informação” traz uma contribuição de extrema importância à área da Ciência da Informação ao conectar os principais conceitos da Estatística às pesquisas de enfoque quantitativo sob o ponto de vista de objetos de investigação da Ciência da Informação.

A princípio, importa ressaltar que, mesmo que essa obra não tenha pretensão de esgotar o tema (conforme destacado pelos autores), é estabelecido um diálogo necessário de assuntos, tradicionalmente relacionados às áreas das ciências exatas, porém, de ampla aplicação às Ciências Sociais Aplicadas, e mais especificamente, à Ciência da Informação. Embora esses assuntos não sejam inéditos na área nem recentes, a escassez de material que os sistematize no contexto da Ciência da Informação, para além dos Estudos Métricos da Informação, justifica e reforça a importância da obra.

Os assuntos são fundamentados de forma clara e didática a fim de situar o leitor, com maior ou menor afinidade em Estatística e aos métodos quantitativos, dando luz a meios de compreender a aplicação e o uso de métodos para desenvolvimento de pesquisas na área. Assim, a obra demonstra a importância do planejamento de pesquisa, sob enfoque da pesquisa quantitativa, com a delimitação de problemas e objetivos e a escolha dos métodos que irão compor a metodologia da pesquisa. Embora a discussão entre a natureza de uma pesquisa (qualitativa, quantitativa ou mista) seja algo de possível superação, o uso de métodos e técnicas de caráter quantitativo advindos da Estatística são de extrema importância e fazem parte da práxis científica da área de Ciência da Informação.

Com isso, essa obra engloba 5 capítulos ancorados em: conceitos gerais e definições da estatística aplicada à Ciência da Informação; Planejamento de pesquisa com enfoque quantitativo na Ciência da Informação, bem como erros frequentes; Fundamentos da estatística descritiva.

Para além da Introdução no Capítulo 1, que traz um contexto histórico da Estatística e de sua relação com a Ciência da Informação, tem-se no Capítulo 2 – "Conceitos gerais" a apresentação dos conceitos de população, amostra, técnicas de amostragem e cálculo do tamanho mínimo da amostra, em que os autores discorrem sobre elementos fundamentais tanto para a fase de planejamento metodológico de uma pesquisa de cunho quantitativo, quanto para execução da pesquisa em si. Além disso, são tratados os níveis de mensuração de variáveis, ponto de atenção aos pesquisadores, pois, as variáveis de um estudo definem os métodos a serem escolhidos no momento de análise.

No Capítulo 3 – "Como fazer um estudo quantitativo em Ciência da Informação", maior capítulo da obra, os autores detalham o fazer científico, a partir da aplicação da Estatística, na área de Ciência da Informação em cinco partes que contemplam: i) O planejamento de pesquisa, trazendo um olhar atento a delimitação de problema e objetivos de pesquisa, a delimitação da população e amostra, instrumentos de coleta de dados, além de modelos de apresentação e análise de dados. Vale destacar que a etapa de planejamento, frequentemente subestimada, é de suma importância para que pesquisadores evitem retrabalho e tenham clareza do percurso que decidiram seguir para obtenção de resultados. ii) A coleta de dados que discute questões caras à pesquisa de enfoque quantitativo como origem dos dados, ética, procedimentos e instrumentos para coleta e execução e qualidade na coleta de dados. Sabe-se que coletar dados não se resume simplesmente a recuperar dados de forma desorganizada, visto que, uma coleta inadequada pode comprometer o andamento de qualquer pesquisa, e nesse sentido, os autores expõem formas de estruturar todo esse processo. iii) O tratamento de dados que versa sobre meios para organização da base de dados da pesquisa verificando inconsistências e padronizando registros. É sabido que na Ciência atual, bem como na Ciência da Informação, à medida que se lida cada

vez mais com dados, o tratamento tem papel fundamental para que resultados consistentes sejam obtidos. iv) Conhecendo os dados traz os fundamentos basilares da estatística descritiva e retoma conceitos como cálculo de frequências absoluta e relativa, medidas de tendência central como média, mediana e moda, e dispersão como amplitude total, variância e desvio padrão. Além disso, também são apresentados os quartis de uma distribuição para discutir o comportamento da distribuição dos dados. Estes fundamentos representam, não só a base da Estatística descritiva, mas também, o alicerce para aproximação e compreensão do objeto científico investigado. v) A representação de dados explora a organização, enquanto resumo de dados, em suas formas tabulares e gráficas. São descritas formas de representar unidades de análise a partir de tabelas de frequência e de gráficos de barras, histograma, gráfico de linhas, dispersão, *boxplot* e cartograma, todos amplamente utilizados na Ciência da Informação. O capítulo apresenta, de forma assertiva e precisa, como relacionar e representar variáveis de estudo, a partir do nível de mensuração (quantitativa ou qualitativa) de cada uma, par a par. Compreender o comportamento de variáveis e representá-las de maneira adequada favorece a comunicação da Ciência ao clarificar informações e resultados.

O Capítulo 4 – "Erros frequentes" traz algo incomum a obras de cunho metodológico: a exploração de erros frequentes que podem ser cometidos. Este capítulo reforça a importância de toda obra ao auxiliar o leitor e pesquisador a compreender possíveis erros gerados ao longo do percurso metodológico de uma pesquisa amparada na aplicação da Estatística. Como destacado pelos próprios autores, erros ocorrem não só por cálculos incorretos, mas também por problemas de planejamento e compreensão do cenário investigado. Importa ressaltar que a existência de cada vez mais ferramentas para a realização de cálculos e a geração de resultados auxilia a reduzir a ocorrência de erros manuais, e nesse ponto os autores sinalizam erros como incoerência entre problema, objetivos e dados, generalizações indevidas

a partir da amostra, falhas em procedimentos de coleta de dados, tratamento e análise incompatível com os dados, e, a representação e interpretações indevidas. Ao evidenciar possíveis erros a serem cometidos, demonstra-se a importância dos capítulos anteriores no que diz respeito à compreensão dos conceitos gerais e o fazer científico a partir da Estatística.

O Capítulo 5 – "Considerações Finais" realiza um fechamento da obra, situando-a com um guia que fornece um processo sequencial e caminhos para elaboração de pesquisas quantitativas ou mistas na área da Ciência da Informação. Assim, a obra se encerra reconhecendo que este material não substitui obras especializadas em Estatística, porém, é relevante admitir o seu potencial em conectar pesquisadores da área de Ciência da Informação à Estatística e fortalecer ainda mais o rol de materiais voltados para pesquisas quantitativas que embasarão o desenvolvimento de conhecimentos futuros.

Por fim, a obra "Pesquisa Quantitativa na Ciência da Informação" revela-se como um importante marco na literatura científica da Ciência da Informação fortalecendo a área e os estudos dos métodos quantitativos e da Estatística aplicada à área, pois, condensa de forma clara, objetiva e didática os caminhos aos quais os pesquisadores poderão recorrer no desenvolvimento de suas pesquisas.

Boa Leitura!

Rafael Gutierres Castanha

1. INTRODUÇÃO

A contagem constitui uma das formas mais antigas de representação de quantidades e está, em parte, relacionada à noção de posse. Em diferentes culturas e épocas, criaram-se sistemas de numeração para quantificar objetos, animais, pessoas e outros elementos da realidade. Algumas dessas representações, embora adotassem a base decimal, não possuíam inicialmente um símbolo para o zero, isto é, para representar a ausência de quantidade.

Durante muito tempo, os números foram utilizados principalmente para representar quantidades. Posteriormente, passaram a apoiar a medição de extensões, áreas, tempo e espaço, que exigiu a definição de unidades de medida. Assim, a quantificação tornou-se parte das atividades humanas a ponto de Nogueira (2011), em sua revisão, mencionar estudos segundo os quais crianças já seriam capazes de distinguir pequenas quantidades, mesmo antes de desenvolver a fala.

Com a contagem, também se tornou necessário reunir e diferenciar os elementos contados. Um pastor, por exemplo, poderia conhecer a quantidade total de animais de seu rebanho e, ao mesmo tempo, identificar quantos eram filhotes, fêmeas e assim por diante. Dessa forma, a quantificação passou a abranger totalidades e subdivisões, que podem ser representadas por conjuntos utilizados em diferentes campos do conhecimento.

Nesse contexto, a Estatística desenvolveu-se com forte base matemática, mas com caráter aplicado. Enquanto a Matemática investiga estruturas, relações e modelos abstratos, a Estatística voltou-se historicamente para a produção e a análise de informações sobre populações, riquezas e atividades do Estado. Contudo, sua aplicação ultrapassou os limites da administração pública e passou a apoiar investigações em diferentes áreas do conhecimento.

1.1. Estatística

Morfologicamente, o termo estatística é formado por estado (no sentido político) seguido do sufixo (t)ica, significando ciência. Logo é uma ciência do estado, tanto que "Estado" e "Estatística" têm a mesma origem na palavra latina "Status", mas a segunda parte tendo o mesmo sufixo de matemática. Por isso, historicamente, a estatística tem seu nome criado pelo estudioso alemão Gottfried Achenwall (1719 - 1772) para se referir a dados a serem utilizados pelo governo e organizações afins.

Bayer *et al*. (2004) relatam que ainda no século XVIII a Estatística era vista como: "o estudo quantitativo de certos fenômenos sociais, destinados à informação dos homens de Estado". Esta visão ainda perdura, tanto que no Brasil, em 1936, foi criado o Instituto Brasileiro de Geografia e Estatística (IBGE), alinhado a esse entendimento. Entretanto, a Estatística não se restringe apenas à questão de estado e estudos demográficos de populações de países.

Memória (2004) argumenta que, embora a Estatística tenha raízes na coleta de informações demográficas para o Estado, seu entendimento se ampliou na modernidade. A área compreende métodos destinados à coleta, à organização, ao tratamento e à análise de dados numéricos ou categóricos, sujeitos a diferentes fontes de variação. Para isso, utiliza recursos matemáticos e computacionais, além do raciocínio indutivo empregado na inferência estatística.

Mesmo sem consenso, os estudos estatísticos podem ser classificados como observacionais ou experimentais, conforme a forma de produção dos dados e o grau de interferência sobre o fenômeno investigado. Essa classificação reforça o entendimento da Estatística como uma ciência ampla e voltada aos estudos quantitativos.

Nos estudos observacionais, o pesquisador observa e mede as variáveis sem intervir deliberadamente sobre a exposição, o tratamento ou

as condições do fenômeno. Esses estudos podem envolver contato direto, entrevistas e acompanhamento dos participantes. Sterrantino (2024) relata que os estudos observacionais podem ser de caso-controle e transversais, cortes de base populacional e variantes. Em todos os casos, os estudos observacionais mantêm relação com o período em que ocorrem, e a possibilidade de generalização depende do desenho do estudo, da forma de seleção e participação, da qualidade das medidas e da população efetivamente representada.

Já nos estudos estatísticos experimentais há um planejamento mais elaborado, na medida em que se pode alterar um, ou mais, elementos no cenário, de forma controlada para verificar seus efeitos. Com isso, há um monitoramento do ambiente em que o estudo está ocorrendo, de forma a verificar impactos, entre outros. Nesse sentido, diferentemente das pesquisas observacionais, nas pesquisas experimentais há interferência do pesquisador no contexto investigado. Dessaune Filho (1987) relata que a estatística experimental se fundamenta em três pilares, sendo: repetição, utilização de mais de um cenário para o estudo, com a qualidade de unidades dependentes do grau de precisão; casualização, garantir tratamentos de forma que tenha estimativas não tendenciosas, e; controle local, controlar as variações nos dados, de forma isenta ao tratamento, de forma a minimizar erros.

Em relação à abrangência dos estudos estatísticos, podem-se coletar informações em toda uma população ou somente em uma amostra. Quando o estudo examina todas as entidades de um grupo, realiza-se um levantamento censitário. Essa estratégia elimina o erro decorrente da seleção amostral, mas pode exigir tempo e recursos incompatíveis com as condições da pesquisa.

Quando o universo de pesquisa é muito amplo ou o levantamento censitário é inviável, pode-se selecionar uma amostra representativa da população. Para isso, utilizam-se técnicas como a amostragem aleatória simples, estratificada ou por conglomerados. Conforme Santos

(2018), mais que o tamanho da amostra, o procedimento de seleção e os recursos disponíveis devem ser considerados conjuntamente.

Já em relação à forma de análise, os estudos estatísticos podem ser categorizados como descritivos ou inferenciais. Na primeira opção, descritiva, os dados são coletados e dispostos em gráficos, tabelas e outras formas de apresentação para descrever um cenário de forma sintetizada e organizada. Já no estudo inferencial, utilizam-se informações amostrais para estimar características da população ou testar hipóteses, respeitados os pressupostos do desenho e das técnicas adotadas.

Nesse contexto, os estudos estatísticos não possuem divisões rígidas e podem ser categorizados de várias maneiras. Em todos os casos, o planejamento e a execução devem seguir procedimentos sistemáticos que permitam avaliar a precisão e os limites dos resultados. A Estatística constitui, portanto, uma ferramenta metodológica cuja aplicação depende do problema e do campo do conhecimento.

1.2. Estudos estatísticos na Ciência da Informação

A Ciência da Informação se insere nas ciências sociais aplicadas, juntamente com o direito, economia, administração, comunicação e outras disciplinas. Entretanto, um dos aspectos que diferenciam a Ciência da Informação das demais áreas está relacionado ao seu principal objeto de estudo: a informação. Com isso, a Ciência da Informação se apresenta com aspectos transversais, na medida em que a informação está presente em todas as atividades humanas. Da mesma forma, a Ciência da Informação é interdisciplinar, agregando teorias, metodologias e práticas de outras áreas.

Historicamente, a Ciência da Informação nasceu da chamada explosão informacional devido ao crescimento científico e tecnológico decorrente

da Segunda Guerra Mundial. Com isso, nasceu a necessidade de criar métodos, técnicas, padrões e outros para selecionar, tratar, armazenar e recuperar informações para que pudessem ser disseminadas e utilizadas. Juntamente com isso, alguns estudos existentes foram absorvidos pela recente ciência, como o caso das métricas da informação.

Pode-se afirmar que o primeiro uso da Estatística para atuação com informação está na bibliometria, nascida na década de 1920, mesmo que tenha recebido essa denominação por Paul Otlet em 1934 e ter sido popularizado por Alan Pritchard, em 1969 (Momesso; Noronha, 2017). Como relata Pritchard (1969, p. 349), a primeira forma de nomear a bibliometria foi de "bibliografia estatística", mas o autor não considerou o termo preciso e sugeriu bibliometria, como sendo a "... aplicação da matemática e de métodos estatísticos a livros e outros meios de comunicação". Com isso, nota-se que na sua essência a bibliometria traz a Estatística como a ferramenta de análise.

Desde então, diferentes métricas foram desenvolvidas para examinar a produção, a circulação e o uso da informação. Nesse contexto, o Quadro 1 sintetiza algumas das abordagens métricas utilizadas na Ciência da Informação e seus objetos predominantes.

Quadro 1 - Abordagens métricas utilizadas na Ciência da Informação

Abordagem	Objeto predominante	Caracterização
Bibliometria	Livros, documentos e comunicação científica registrada	Utiliza métodos estatísticos e matemáticos para quantificar documentos e processos de comunicação escrita (Otlet, 1934; Pritchard, 1969).
Cientometria	Produção e atividade científica	Examina o desenvolvimento da ciência por meio da mensuração das atividades, da produtividade e das publicações científicas (Silva; Bianchi, 2001).
Informetria	Informação registrada em diferentes formatos e contextos	Investiga os aspectos quantitativos da informação, sem se restringir aos documentos bibliográficos ou à comunidade científica (Macias-Chapula, 1998).
Patentometria	Patentes e informações tecnológicas	Analisa características e usos das patentes como documentos e fontes de informação tecnológica (Narin, 1994; Speziali; Nascimento, 2020).
Arquivometria	Fundos, documentos e atividades arquivísticas	Aplica métodos quantitativos ao estudo de arquivos e dos fenômenos relacionados aos seus fundos documentais (Pinto, 2011).
Cibermetria	Conteúdos e processos de comunicação no ciberespaço	Examina quantitativamente os conteúdos, as relações e os processos comunicativos desenvolvidos no ambiente digital (Aguillo, 2007).
Webometria	Recursos, estruturas e relações presentes na Web	Estuda quantitativamente os recursos de informação e as estruturas da Web com base em abordagens bibliométricas e informétricas (Björneborn, 2004).
Webmetria	Acesso e utilização de páginas e serviços na Web	Mensura aspectos relacionados ao uso e ao tráfego de sites, como acessos, visitas e interações registradas (Gouveia; Araújo, 2020).
Altmetria	Atenção e circulação da produção científica na Web social	Utiliza registros produzidos em plataformas digitais e redes sociais como indicadores alternativos de atenção e circulação da produção científica (Barcelos; Maricato, 2023).

Fonte: elaboração própria (2026), com base nos autores indicados no quadro.

Evidentemente, mesmo que as pesquisas métricas tenham maior afinidade com a Estatística, pode-se utilizar essa metodologia em diversos outros estudos. No caso de levantamento, por exemplo, em muitos casos, dados demográficos são levantados e podem fazer uso da Estatística para a sua análise. Assim, o entendimento das possibilidades e formas de atuação com a estatística torna-se útil na Ciência da Informação.

2. CONCEITOS GERAIS

Para a compreensão desta obra, alguns conceitos precisam ser discutidos. Mesmo nas ciências exatas, há questões que merecem atenção diante da possibilidade de interpretações ambíguas. Assim, selecionaram-se conceitos amplamente utilizados nos capítulos, de forma a apresentar o entendimento dos autores.

2.1. População ou Universo

A população é o conjunto de indivíduos ou objetos que o estudo estatístico abrange. Os indivíduos são os elementos que detêm as informações que se deseja coletar e sobre os quais, a partir do estudo, busca-se extrair conclusões válidas. À coleta de dados de toda população dá-se o nome de levantamento censitário.

Medidas estatísticas calculadas a partir de dados de toda a população são conhecidas como “parâmetros”, enquanto medidas calculadas a partir de dados amostrais são chamadas de “estimativas”. Portanto, quando o estudo abrange todos os elementos da população, os resultados obtidos descrevem diretamente as características desse conjunto.

2.2. Amostra

Amostra é uma parte da população selecionada para compor o estudo. Sua utilização ocorre, principalmente, quando o levantamento censitário não é possível ou quando não há recursos suficientes para realizá-lo. O ato de selecionar os elementos da população que integrarão a amostra é chamado de amostragem. Espera-se que essa parte preserve as características da população relevantes para o estudo, possibilitando a obtenção de resultados sem que seja necessário analisar todos os seus elementos.

Trabalhar com uma amostra é interessante quando a população é muito grande ou quando os recursos são escassos. A redução da

quantidade de elementos e, por consequência, dos dados coletados apresenta como vantagens a economia de recursos financeiros e do tempo necessário para a realização do estudo, a melhoria na operacionalização dos dados, a possibilidade de evitar erros de transcrição e a maior atenção aos casos individuais. Por outro lado, a utilização de uma amostra pode não ser necessária quando a população é facilmente mensurável ou quando o estudo exige elevada precisão.

As medidas calculadas a partir dos dados amostrais são chamadas de estimativas. Essas estimativas tendem a se aproximar dos parâmetros populacionais quando a amostra é selecionada por meio de uma técnica adequada e possui quantidade de elementos suficiente para representar as características relevantes da população. Essa quantidade não é definida apenas pela percepção de que a amostra é grande ou pequena, mas depende de fatores como o tamanho e a diversidade da população, a precisão pretendida e a técnica de amostragem utilizada. Entretanto, uma amostra grande, quando selecionada por um processo inadequado ou enviesado, pode produzir um resultado tão impreciso quanto uma amostra pequena e não representativa.

Desse modo, a definição da amostra envolve duas decisões relacionadas: quantos elementos serão selecionados e como essa seleção será realizada. Para orientar essas decisões, foram desenvolvidas técnicas de amostragem e métodos para calcular o tamanho mínimo da amostra. A adequação dessas escolhas influencia a possibilidade de utilizar os resultados obtidos na amostra para formular conclusões sobre a população.

Mesmo quando a amostra é selecionada adequadamente, existe o chamado erro amostral, que corresponde à diferença entre uma medida obtida na amostra e o parâmetro populacional correspondente. Esse erro decorre do fato de se observar apenas uma parte da população e tende a diminuir com o aumento do tamanho da amostra. Entretanto, ele não deve ser confundido com o viés de

seleção, provocado pela forma inadequada de escolher os elementos. Enquanto o erro amostral pode ser reduzido pelo aumento da amostra, o viés de seleção pode permanecer, independentemente da quantidade de elementos selecionados.

2.2.1 Técnicas de Amostragem Probabilística

As técnicas de amostragem são divididas em técnicas probabilísticas e não probabilísticas. Nas técnicas de amostragem probabilística, todos os elementos da população têm probabilidade conhecida e maior que zero de serem selecionados para a amostra, mediante sorteio. Essas probabilidades não precisam ser iguais, salvo em desenhos como a amostragem aleatória simples. As técnicas probabilísticas permitem que os resultados obtidos a partir da amostra sejam generalizados para a população, mas para isso é necessário que a amostra seja grande o suficiente e que seja composta pelas mesmas características gerais apresentadas pela população, ou seja, deve ser representativa em relação à população. As amostras não probabilísticas não oferecem as mesmas bases para calcular o erro amostral e realizar inferências estatísticas para a população, pois as probabilidades de seleção dos elementos não são conhecidas.

Como exemplo de técnicas de amostragem probabilística tem-se: amostragem aleatória simples; amostragem sistemática; amostragem estratificada e amostragem por conglomerado. Na amostragem aleatória simples cada elemento tem a mesma probabilidade de ser selecionado, é uma técnica que funciona como um sorteio em que os elementos são selecionados individualmente. O sorteio pode ser realizado com ou sem reposição dos elementos selecionados.

Na amostragem sistemática os elementos são listados e define-se um intervalo que será repetido para a seleção de cada indivíduo. Por exemplo, ao listar os elementos define-se que será selecionado um indivíduo a cada 10 para compor a amostra. É possível associar a técnica de amostragem aleatória simples à técnica de amostragem

sistemática, nesse caso o primeiro elemento pode ser selecionado por amostragem aleatória simples e o intervalo ocorre a partir dele para obter os demais elementos.

A amostragem estratificada é realizada a partir da divisão da população em estratos (grupos), definidos, por exemplo, pela renda, pelo bairro, pela faixa etária, pelo cargo ocupado, entre outros critérios. Os estratos correspondem a subdivisões da população, ou seja, subpopulações. É uma técnica recomendada principalmente quando se observa heterogeneidade acentuada dentro de uma população em relação a determinada(s) característica(s).

Na amostragem por conglomerados a população é subdividida em conglomerados que também podem ser entendidos como subpopulações distintas. Esse tipo de amostragem é indicado quando há a possibilidade de dividir a população em um grande número de pequenas subpopulações ou conglomerados. Os conglomerados podem ser constituídos por bairros, residências, famílias, escolas, lotes, entre outros. Todos os membros dos conglomerados selecionados se tornam elementos da amostra.

2.2.2 Técnicas de Amostragem Não Probabilística

Nas técnicas de amostragem não probabilística, a escolha dos elementos que comporão a amostra depende do julgamento do pesquisador ou de outros critérios que não garantem uma seleção aleatória. Diferentemente do que ocorre com as técnicas de amostragem probabilística, nas técnicas não probabilísticas não há base estatística para realizar inferências sobre a população a partir da amostra, visto que existe maior possibilidade de viés de seleção, uma vez que a escolha dos elementos não é realizada de forma aleatória.

As técnicas de amostragem não probabilística são vantajosas quando a pesquisa tem uma população muito específica e de difícil acesso. Além disso, podem proporcionar custos mais baixos, maior agilidade

na coleta de dados e uma operação mais simples. Como exemplos de técnicas de amostragem não probabilísticas tem-se: amostragem por conveniência, amostragem acidental, amostragem por julgamento, amostragem por quotas e amostragem por bola de neve.

A amostragem por conveniência é construída a partir de elementos de fácil acesso. Esses elementos fazem parte da população de interesse, mas, por algum motivo, o pesquisador consegue coletar seus dados com maior facilidade e, por essa razão, eles serão selecionados para compor a amostra.

Na amostragem acidental os elementos são selecionados de forma não planejada com base na sua disponibilidade ou acessibilidade, sem que haja um critério específico para sua seleção. Nesse tipo de amostragem é comum que o pesquisador responsável pela coleta de dados se posicione em algum local onde tenha acesso aos indivíduos e consiga abordá-los para realizar a coleta.

Já na amostragem por julgamento, o pesquisador elenca critérios para a seleção dos elementos e, a partir desses critérios, seleciona quais indivíduos irão compor a amostra. A amostragem por quotas é feita a partir da divisão da população em subgrupos, o pesquisador seleciona uma quantidade predeterminada de elementos de cada um desses subgrupos.

No caso da amostragem por bola de neve, parte-se de um participante inicial que fará a indicação dos próximos. Os indicados indicam os próximos participantes e, assim, sucessivamente, o tamanho da amostra cresce como uma "bola de neve".

2.2.3 Erro Amostral e Cálculo do Tamanho Mínimo de Amostra

Conforme mencionado anteriormente, denominamos as medidas estatísticas extraídas a partir de dados de toda população de

"parâmetros", enquanto medidas obtidas por meio de dados amostrais são chamadas de "estimativas". O erro amostral corresponde à diferença entre o valor obtido por meio de uma estatística amostral e o verdadeiro valor do parâmetro que se deseja estimar. Para determinar o tamanho da amostra, o pesquisador deve estabelecer previamente o erro amostral tolerável, isto é, a diferença máxima que está disposto a admitir entre a estimativa obtida a partir da amostra e o parâmetro populacional de interesse.

O cálculo de uma primeira aproximação para o tamanho mínimo da amostra pode ser realizado mesmo sem conhecer o tamanho da população. Para isso, pode-se utilizar a seguinte fórmula:

$$n_0 = 1/E_0^2 \quad (1)$$

Essa expressão fornece uma primeira aproximação simplificada do tamanho amostral para estimação de proporções, sob máxima variabilidade e nível de confiança próximo de 95%. Ela não se aplica indistintamente a qualquer variável ou desenho amostral.

Em que
n: número de elementos da amostra
n_0: uma primeira aproximação para o tamanho da amostra
E_0: erro amostral tolerável

Faremos a seguinte suposição para exemplificar o cálculo: deseja-se realizar um estudo estatístico por amostragem para avaliar as características de uma população de leitores de determinada cidade, tais como quantidade de livros lidos em um ano, os gêneros dos livros lidos e o grau de escolaridade, entre outras. Qual deve ser o tamanho mínimo de uma amostra aleatória? Admitiremos, neste caso, que o erro amostral não deve ser superior a 5%. Assim, nosso erro amostral tolerável, E_0, será igual a 0,05.

Aplicando o cálculo (1) obtemos:

$$n_0 = 1/(0{,}05)^2 = 400 \text{ leitores}$$

Caso o número de elementos da população (N) seja conhecido, podemos adotar a seguinte fórmula:

$$n = (N \cdot n_0) / (N + n_0) \quad (2)$$

Utilizando o exemplo anterior como base, vamos supor que temos a informação de que a cidade possui 600 leitores, ou seja, a população do estudo é formada por esses 600 indivíduos identificados como leitores. Sabendo o tamanho da população e tendo calculado a primeira aproximação da amostra, podemos calcular o tamanho mínimo da amostra pelo cálculo (2):

$$n = (600 \cdot 400) / (600 + 400) = 240 \textit{ leitores}$$

Ao comparar os dois exemplos observamos que o cálculo (2) realiza um ajuste no resultado obtido pelo cálculo (1), ambos apresentam o tamanho mínimo para a amostra, ou seja, não há impedimento em trabalhar com uma amostra que possua quantidade de elementos superior ao tamanho mínimo calculado. No entanto, para manter o nível de precisão estabelecido e as características probabilísticas da amostra, não é indicado trabalhar com uma quantidade de elementos inferior ao tamanho mínimo calculado e utilizar os resultados para realizar inferências sobre a população a partir da amostra.

2.3. Variável

São chamadas de variáveis as características observáveis ou passíveis de medição, sob as mesmas condições, em cada elemento da população. Cada variável, observada ou medida, deve gerar um único resultado para cada elemento da população. As variáveis são obtidas a partir do estudo da população e são classificadas de acordo com o tipo de mensuração, podendo ser qualitativa (também chamada de categórica) ou quantitativa.

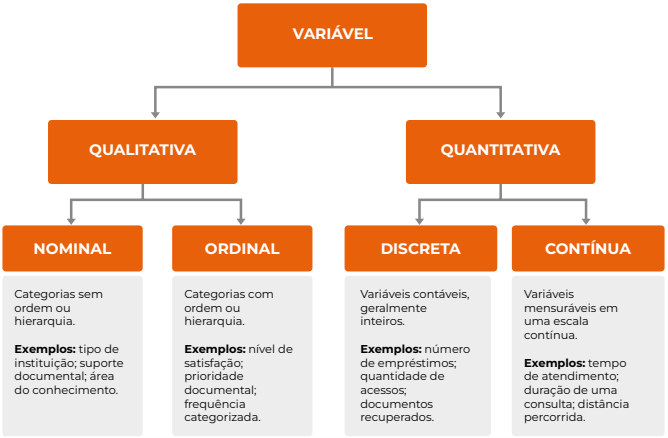
As variáveis qualitativas são expressas por atributos, categorias ou qualidades, enquanto as variáveis quantitativas são expressas por

números ou quantidades. As variáveis qualitativas são subdivididas em variáveis nominais e ordinais. Uma variável qualitativa nominal não exibe ordenação entre as categorias, alguns exemplos são: cor dos olhos, gênero literário, cor da pele, entre outros.

Já uma variável qualitativa ordinal apresenta ordenação/hierarquização entre as categorias, é possível citar como exemplos: escolaridade, classe social, nível de acesso ou sigilo em documentos de um acervo, entre outros. As variáveis quantitativas são subdivididas em discretas e contínuas. Uma variável quantitativa discreta assume valores pertencentes a um conjunto finito e enumerável como, por exemplo, número de filhos ou número de exemplares.

Uma variável quantitativa contínua pode assumir qualquer valor entre dois limites, nesse caso valores fracionários fazem sentido. Um exemplo é a altura de um indivíduo, o valor pode ser representado de acordo com a precisão da medida, com uma quantidade maior ou menor de casas decimais. A Figura 1 classifica e exemplifica os tipos de variáveis.

Figura 1 - Classificação de variáveis



Fonte: elaboração própria (2026).

Cabe aqui destacar que uma população ou amostra é composta por uma série de variáveis e nem todas participam da pesquisa, pois nem todas vão contribuir para atender aos objetivos do estudo. Cada pesquisa, mesmo que utilize uma mesma população ou amostra, precisa escolher as variáveis que comporão a sua pesquisa, ante as possibilidades existentes.

3. COMO FAZER UM ESTUDO QUANTITATIVO NA CIÊNCIA DA INFORMAÇÃO

A escolha para fazer um estudo quantitativo, ou mesmo misto com dados quantitativos, nem sempre é uma opção, mas uma determinação oriunda dos objetivos. Na maioria dos casos, o problema de pesquisa determina o objetivo, o que o pesquisador quer fazer com o problema. Da mesma forma, a metodologia depende dos objetivos, como o pesquisador atenderá os objetivos. Pesquisadores que realizam estudos com dados quantitativos podem fazer uso de alguns passos que ajudam no seu desenvolvimento.

De forma simplificada, a pesquisa quantitativa pode ser organizada em etapas que se iniciam com o planejamento do estudo e conduzem à representação dos resultados. Como mostra a Figura 2, esse percurso compreende o planejamento, a coleta, o tratamento, o conhecimento e a representação dos dados. A organização dessas atividades contribui para que as decisões tomadas em cada etapa permaneçam articuladas aos objetivos da pesquisa (Figura 2).

Figura 2 - Etapas de desenvolvimento de um estudo quantitativo



Fonte: elaboração própria (2026).

Esse fluxo pode ser adaptado a diferentes tipos de pesquisa. Em todos os casos, é necessário planejar o estudo, coletar e tratar os dados, conhecer suas características e selecionar formas adequadas de representação. Nos estudos quantitativos, cada uma dessas etapas deve considerar o tipo de variável, as unidades analisadas e os procedimentos necessários para produzir resultados coerentes com os objetivos estabelecidos.

O rigor científico, relatado na metodologia, deve ser seguido conforme um protocolo bem estruturado. Assim, um estudo quantitativo,

mesmo nas ciências sociais e humanas, busca responder problemas que requerem essa precisão. Em muitos casos, como nos casos das métricas da informação ou mesmo em estudos demográficos, requer-se o uso de dados quantitativos, que fornecem precisão, para gerar os seus resultados.

3.1. Planejando o estudo

Planejar é o ato de estruturar ações com a finalidade de obter sucesso. Com isso, todos os dias as pessoas planejam processos, mesmo que nem sempre se deem conta. No caso específico de pesquisas, o planejamento torna-se essencial para o êxito. Tanto que, em sua grande maioria, em grande parte das instituições, exige-se a elaboração de um projeto antes da realização da pesquisa, no qual se descreve sucintamente como o estudo será conduzido.

Quando se trabalha com dados quantitativos e busca-se tirar conclusões ou testar hipóteses, precisa-se de um plano claro, bem estruturado, viável e que seja metodologicamente válido. Planejar as ações que serão adotadas desde a coleta dos dados até as escolhas de técnicas estatísticas adequadas auxilia na transparência dos procedimentos de pesquisa e na obtenção de resultados confiáveis. Dessa forma, podem-se reduzir inconsistências, retrabalho e decisões incompatíveis com os dados produzidos.

Iniciar um estudo pela escolha de uma técnica estatística ou de um programa de análise pode inverter a lógica da pesquisa. Um bom planejamento é organizado e sistematizado e deve ser elaborado a partir da definição do objetivo do estudo; da seleção da população e, se necessário, da amostra; da escolha do método de coleta e das técnicas estatísticas adequadas. Por essa razão, o plano deve ser suficientemente detalhado para orientar a execução, mas também permitir ajustes justificados e documentados quando surgirem situações não previstas.

3.1.1 Definindo Objetivo

Metodologicamente, a seleção da forma do estudo, seja qualitativo, quantitativo ou misto, depende do problema, ou seja, o que será realizado para atendê-lo. Por isso, dizemos que o problema indica qual o objetivo do estudo. Assim, o estudo apresenta um objetivo geral e pode conter objetivos específicos relacionados às diferentes etapas da pesquisa. Pode-se definir, por exemplo, um objetivo específico para o levantamento da literatura.

Elaborar o objetivo de um estudo estatístico demanda compreensão daquilo que se busca a partir do que será analisado. O objetivo da análise estatística não é necessariamente o mesmo objetivo da pesquisa como um todo. Nesse sentido, o pesquisador pode lançar mão da seguinte pergunta: O que busco descobrir ou comprovar a partir da análise de dados? A resposta guiará a elaboração do objetivo. A definição da população estudada é dada a partir do objetivo.

Com isso, como a primeira etapa do planejamento do estudo, procure estabelecer o objetivo da etapa quantitativa da pesquisa, com descrição clara, precisa e factível. Pode-se estabelecer objetivos relacionados à medição de fenômenos, para: identificar padrões e tendências, testar hipóteses, verificar causas e efeitos, tratar dados e tantos outros. Por exemplo, levantar dados sobre publicações científicas de uma determinada disciplina, para verificar as tendências dos estudos, quais os temas estudados em uma linha do tempo, de modo a identificar o crescimento de determinados temas, ou mesmo quais os temas com interesse recente.

Um exemplo desta etapa pode ser visto no seguinte exemplo:

- **objetivo:** Verificar as publicações brasileiras que tratam dos temas da ciência aberta.

Esse exemplo hipotético trará resultados quantificáveis, tanto na quantidade de publicações, quanto de artigos publicados, autores,

relações de coautoria, quais e quantos temas são abordados e outros mais. Note que esse objetivo, voltado à parte quantitativa do estudo, não deve contrapor ao geral da pesquisa, mas complementar, contribuir para a sua completude.

3.1.2 Definindo População e Amostra

Com o objetivo estabelecido, passa-se para a etapa de definição da população ou amostra. Para tanto, será necessário analisar alguns pontos a fim de determinar se irá utilizar a população ou uma amostra. Em todos os casos, requer procurar responder à questão: Essa população ou amostra é adequada aos objetivos da pesquisa? A resposta a essa pergunta deve atender a uma primeira reflexão sobre a questão, inclusive sobre a questão da seleção de população ou amostra.

Pode-se estruturar o processo de definição da população ou amostra em três passos analíticos: Selecionar qual a unidade de análise, estabelecer critérios para compor a população ou amostra e verificar a sua viabilidade. Seguindo esses passos aumenta-se a probabilidade de sucesso no processo de coleta de dados para atendimento aos objetivos traçados na pesquisa, que dependam das informações oriundas da população ou amostra.

Ao planejar um estudo quantitativo é importante selecionar qual tipo de entidade comporá a população ou a amostra. Nota-se que a unidade nem sempre é humana, de perfil demográfico, podendo ser obras literárias/científicas, organizações, documentos e tantos outros. Com essa definição será possível começar a verificar se é possível utilizar a população no estudo ou será preciso utilizar uma amostra.

Segue-se criando os critérios que serão utilizados para determinar quais elementos podem ou não ser incluídos na população ou amostra. Entretanto, não se deve confundir ter critérios para formação de população com critérios para formação de amostra. Os critérios para inclusão das entidades compõem-se de uma lista de características que devem

ser atendidas para compor a população ou amostra. Da mesma forma, pode-se estabelecer critérios para exclusão, caso seja necessário.

Por fim, verificar a viabilidade de se utilizar a população ou se é preciso fazer uso da amostragem. Após identificar a unidade, estabelecer critérios de inclusão e exclusão cabe a pergunta, a população ficou muito grande? Caso a resposta seja afirmativa, pode-se fazer uso de amostra, tendo o cuidado para que esse conjunto tenha as mesmas características da população, seguindo os mesmos critérios de inclusão e exclusão, além das técnicas para formação de amostras.

Um exemplo prático pode ser entendido no exemplo:

- **Objetivo:** Verificar as publicações brasileiras que tratam dos temas da ciência aberta.
- **Unidade:** publicações científicas ou técnicas
- **Critérios de inclusão e exclusão:** Periódicos de acesso aberto publicados no Brasil, exceto os que cobram *Article Processing Charge* (APC)
- **Viabilidade:** A quantidade de periódicos presentes no universo pode aumentar ou diminuir a depender dos critérios de inclusão e exclusão. Caso a quantidade de elementos do universo impossibilite o estudo é possível adotar critérios temporais e geográficos, por exemplo, para diminuir o universo ou, alternativamente, trabalhar com uma amostra.

Assim, a definição da população ou amostra precisa atender ao objetivo a ser alcançado pela parte da pesquisa quantitativa. Nesse caso, seguindo etapas, de forma a assegurar o cumprimento desse objetivo, não como uma receita, mas como orientação, podendo ser ajustado conforme as características do estudo, do tipo de dados, entre outros pontos que influenciam na realização do estudo.

3.1.2.1 Definindo a Técnica de Amostragem

Após definir qual é a população de interesse, o pesquisador deve decidir se trabalhará com todos os elementos da população ou se optará por trabalhar com uma amostra. Caso opte por trabalhar com amostra é necessário realizar a amostragem, que nada mais é do que a seleção dos elementos da população que farão parte do estudo. Essa seleção deve ser planejada e realizada a partir da técnica de amostragem que mais se adequar ao estudo.

Se o pesquisador deseja realizar inferências sobre os parâmetros populacionais a partir das estimativas amostrais, deve optar pelo uso de técnicas de amostragem probabilísticas. Além disso, é importante que seja feito o cálculo para o tamanho mínimo da amostra e que a quantidade de elementos seja igual ou superior à obtida pelo cálculo. São exemplos de técnicas de amostragem probabilística: a amostragem simples aleatória, sistemática, a estratificada e a por conglomerados. Esses exemplos foram tratados conceitualmente no capítulo 2. A seguir, serão exemplificadas as mesmas técnicas utilizando exemplos hipotéticos aplicados à Ciência da Informação.

Técnica: Amostragem Aleatória Simples

- **Objetivo do estudo**: Classificar os tipos de dados científicos presentes nas teses brasileiras defendidas na área da Ciência da Informação em uma janela de dois anos
- **População**: Todas as teses defendidas na área da CI no intervalo selecionado. Vamos supor que o total de teses defendidas seja igual a 600.
- **Aplicação da Técnica**: Por meio do cálculo amostral (cálculo apresentado no capítulo 2), foi obtido um tamanho mínimo de amostra de 240 teses, de modo a garantir um nível adequado de precisão às estimativas. Para selecionar os elementos da amostra pela técnica de amostragem aleatória simples devem ser sorteadas 240 teses ou mais dentre as 600 que formam a

população. Observe que, nesse processo, todos os registros possuem a mesma probabilidade de serem selecionados para compor a amostra.

Técnica: Amostragem sistemática

- **Objetivo do estudo**: Analisar a presença de dados de pesquisa depositados em um repositório institucional em um período de 3 anos.
- **População**: A população é formada por todos os registros depositados nesse repositório dentro desse período; consideramos 1000 registros.
- **Aplicação da Técnica**: Define-se uma amostra de 300 registros. Nessa técnica, os elementos da população são ordenados em uma lista e a seleção ocorre em intervalos regulares, considerando *k* como fator de sistematização, correspondente ao valor do intervalo sistemático; *N* como o tamanho da população; *n* como tamanho da amostra; e *r* como um número entre 1 e *k*. O intervalo sistemático pode ser obtido pelo cálculo: $k = N/n$. Nesse caso temos: k=1000/300 ≅ 3,33, arredondado para k = 3. Vale destacar que esse arredondamento afeta o tamanho final da amostra. Com k = 3, o número de elementos selecionados será aproximadamente 1000/3 ≅ 333 registros, número ligeiramente acima dos 300 inicialmente planejados. Em seguida, é sorteado um número "*r*" entre 1 e k para determinar o primeiro elemento da amostra. Supondo que o número sorteado tenha sido 3; dessa forma o terceiro elemento da população será o primeiro a ser selecionado. O segundo elemento selecionado será o que ocupa a posição $k+r$, ou seja, o sexto elemento. O terceiro será aquele que ocupa a posição $2k+r$, e assim sucessivamente, de modo que o i-ésimo elemento da amostra ocupe a posição $(i - 1) \times k + r$. Ressalta-se que a amostragem sistemática pressupõe uma lista ordenada da população que não apresente padrão cíclico ou correlação com a variável de interesse. Se um desses casos for observado, o intervalo fixo de seleção da amostra pode inserir viés.

Técnica: Amostragem estratificada

- **Objetivo do estudo**: Analisar o perfil de egressos de programas de pós-graduação em Ciência da Informação brasileiros.

- **População**: Todos os egressos dos programas de pós-graduação em Ciência da Informação brasileiros.
- **Aplicação da Técnica**: De acordo com a ANCIB (2026) existem 38 programas de pós-graduação em Ciência da Informação no território nacional. Os programas dividem-se em: 3 na região Norte, 11 na região Nordeste, 16 na região Sudeste, 2 na região Centro-Oeste e 6 na região Sul. Ou seja, observa-se heterogeneidade em relação à distribuição geográfica. Se coletarmos uma amostra considerando uma única região, ela não será representativa em nível nacional. Dessa forma, é importante que os elementos selecionados para a amostra representem, de forma proporcional ao observado na população, os egressos de cada uma dessas regiões, considerando a região do programa ao qual cada egresso pertenceu. Supondo que 30% da população seja egressa da região Nordeste, assim, 30% dos elementos da amostra deverão ser selecionados dentre os egressos dos programas localizados nessa região. Da mesma forma, deverá ser respeitada a proporção das demais regiões. Ao distribuir os elementos da amostra entre os estratos de acordo com a proporção observada na população realizamos a chamada alocação proporcional. A técnica de amostragem estratificada pode ser associada a outra técnica de amostragem probabilística. Poderíamos, por exemplo, selecionar os elementos de cada região por meio de amostragem aleatória simples.

Técnica: Amostragem por conglomerados

- **Objetivo do estudo**: Compreender o perfil dos usuários de bibliotecas do estado de São Paulo.
- **População**: Todos os usuários de bibliotecas do estado de São Paulo.
- **Aplicação da técnica**: Se buscássemos selecionar uma amostra dessa população por conglomerados, poderíamos definir as bibliotecas do estado como os conglomerados a serem sorteados. Os elementos da amostra seriam os usuários das bibliotecas sorteadas. Nesse caso, cada biblioteca representaria um conglomerado, formado por seus usuários. As bibliotecas de cada região poderiam ser selecionadas a partir de amostragem aleatória simples e, posteriormente, todos os usuários

das bibliotecas selecionadas comporiam a amostra.

A esse processo dá-se o nome de amostragem por conglomerados em um estágio. Se necessário poderia ser aplicada uma nova técnica de amostragem para selecionar os elementos em cada biblioteca, nesse caso teríamos uma amostragem por conglomerados em dois estágios sendo o primeiro composto pelo sorteio das bibliotecas e o segundo, pela seleção dos usuários das bibliotecas sorteadas na amostra.

Técnica: Amostragem por conveniência

- **Objetivo do estudo**: Avaliar a percepção dos alunos de biblioteconomia sobre o curso da instituição em que estão inseridos.
- **População**: Todos os alunos de biblioteconomia dessa instituição.
- **Aplicação da técnica**: Supondo que o pesquisador seja professor do curso de biblioteconomia da mesma instituição, poderá selecionar seus próprios alunos como elementos da amostra. O contexto propicia a conveniência na seleção dos elementos da amostra pelo pesquisador. Nesse caso, o critério de seleção é a facilidade de acesso do pesquisador aos elementos. Como a amostra se restringe especificamente aos alunos desse professor, é possível que haja viés de seleção. Assim como nas demais técnicas não probabilísticas, o resultado não é passível de generalização.

Técnica: Amostragem acidental

- **Objetivo do estudo**: Analisar o perfil dos visitantes de um determinado museu.
- **População**: Todos os visitantes do museu.
- **Aplicação da técnica**: O pesquisador se posiciona em um local em que há circulação de visitantes e os aborda de acordo com a disponibilidade dos visitantes.

Técnica: Amostragem por julgamento

- **Objetivo do estudo**: Compreender os hábitos de leitura de crianças entre 8 e 10 anos.

- **População**: Leitores com idade entre 8 e 10 anos.
- **Aplicação da técnica**: O pesquisador selecionará os elementos da amostra a partir de seu julgamento; no entanto, deve obedecer a critérios de seleção. Nesse caso, os critérios seriam: ser leitor e ter idade entre 8 e 10 anos. O pesquisador ainda poderia julgar ser mais adequado, respeitando os critérios de seleção, incluir na amostra crianças indicadas por professores como leitoras frequentes, pela possibilidade de fornecerem informações e relatos mais ricos e detalhados. A amostragem por julgamento não necessariamente visa à facilidade de acesso, como no caso da amostragem por conveniência. Na amostragem por julgamento, há uma avaliação sobre quais são os elementos mais adequados para o estudo.

Técnica: Amostragem por quotas

- **Objetivo do estudo**: Investigar o nível de satisfação de pesquisadores sobre uma plataforma de metadados.
- **População**: Todos os pesquisadores que utilizam a plataforma de metadados escolhida para o estudo.
- **Aplicação da técnica**: Os pesquisadores (elementos da população) podem ser divididos em subgrupos de acordo com o país de origem. Sendo assim, a seleção da amostra considerará a proporção de pesquisadores cadastrados na plataforma de acordo com o país de origem, selecionando proporcionalmente, por conveniência ou por julgamento, os elementos da amostra dentro de cada subgrupo. Ou seja, não há seleção aleatória dentro das quotas.

Técnica: Amostragem por bola de neve

- **Objetivo do estudo**: Mapear a circulação de informações científicas informais entre pesquisadores da Ciência da Informação.
- **População**: Todos os pesquisadores da área da Ciência da Informação
- **Aplicação da técnica**: Inicia-se identificando dois pesquisadores conhecidos na área; após coletar os dados necessários, solicita-se que cada pesquisador indique outros dois colegas de sua rede.

Os novos participantes indicam os próximos, e assim sucessivamente. O processo se repete continuamente até que o critério de parada seja atingido. Nesse caso, o critério de parada poderia ser a saturação das indicações, ou seja, os indicados passam a indicar participantes que já fazem parte da amostra, ou ainda a obtenção do número de participantes necessário para atingir o tamanho amostral desejado. Essa técnica tende a gerar amostras enviesadas porque está alicerçada na rede de relacionamento entre os participantes. Pesquisadores que não se relacionam com os participantes, mesmo fazendo parte da população de interesse, tendem a não ser indicados.

3.1.3 Definindo o Instrumento de Coleta

Seguindo o processo de planejamento do estudo, passa-se a selecionar, entre todos os instrumentos disponíveis, qual o instrumento de coleta a ser utilizado no estudo. Evidentemente, seguindo a estrutura proposta, a coleta de dados deve atender ao objetivo traçado, ou seja, responder: Como se conseguem obter os dados da população ou amostra, para atender ao objetivo? Assim, o instrumento de coleta de dados depende do tipo de dado a ser coletado, ou seja, do tipo de unidade de dado.

De uma forma simplista, alguns dos tipos de coleta de dados que podem gerar dados quantitativos são: survey (levantamento), entrevista, observação, levantamento documental e bibliográfico, entre outros. Evidentemente, a seleção do instrumento é dependente da unidade de dado e objetivo do estudo, pois responde à pergunta sobre qual o tipo de instrumento consegue coletar os dados que atendam a esses dois pontos, da seguinte forma:

- **Survey**: questionário com perguntas abertas e fechadas que são convertidas em dados quantitativos, como no caso dos levantamentos demográficos (idade, formação etc.)
- **Entrevista**: Conjunto estruturado de perguntas que visam coletar

respostas que podem ser convertidas em dados quantificáveis, como no caso das entrevistas de intenção de voto.

- **Observação**: marcação de ocorrência de algum fenômeno, como determinar o número de participantes de um evento
- **Levantamento Bibliográfico:** busca quantificar documentação científica, como quantidade de artigos ou livros, ou mesmo referências bibliográficas utilizadas, entre outros em documentos avaliados pelos pares
- **Levantamento Documental:** próximo ao levantamento bibliográfico, mas em documentos não avaliados pelos pares, podendo também incluir os sites e outros recursos na web.

Nesse caminho, o planejamento do estudo segue uma sequência de dependências, no qual o instrumento de coleta de dados precisa atender ao objetivo e a população ou amostra, ou seja, qual tipo de ferramenta é capaz de coletar o tipo de unidade da população ou amostra para atender aos objetivos. Por esse motivo, a definição do tipo de instrumento de coleta de dados não é aleatória, mas segue um planejamento que busca os resultados desejados à pesquisa.

Um exemplo prático e hipotético seria:

- **Objetivo**: Verificar as publicações brasileiras que tratam dos temas da ciência aberta.
- **Unidade**: Publicações técnico-científicas
- **Critérios de inclusão e exclusão:** Periódicos de acesso aberto publicados no Brasil, exceto aqueles que cobram *Article Processing Charge* (APC).
- **Viabilidade**: A quantidade de periódicos presentes no universo pode aumentar ou diminuir a depender dos critérios de inclusão e exclusão. A inclusão de critérios temporais e geográficos, por exemplo, diminuiria a quantidade de elementos do universo; alternativamente, seria possível optar pelo uso de amostra.

- **Instrumento de pesquisa:** Pesquisa bibliográfica, visto que precisa coletar informações sobre as publicações técnico-científicas.

Com a definição do instrumento de coleta de dados, seguindo as orientações, tem-se a maior parte do plano de pesquisa. Entretanto, torna-se necessário chamar a atenção sobre o ponto: assegurar que esse instrumento atenda à coleta das variáveis da amostra ou universo, pois em alguns casos, pode-se requerer o uso de mais de uma ferramenta, tendo o cuidado de evitar sobreposições. Assim, se apenas um instrumento não for capaz de atender aos objetivos e variáveis, utilize mais que um instrumento de coleta de dados.

3.1.4 Definindo o Modelo de Apresentação dos Dados

Ainda durante o planejamento, deve-se prever como os dados coletados serão organizados e apresentados. Essa definição depende dos objetivos do estudo, do tipo de variável e da informação que se pretende comunicar. Alguns resultados exigem a apresentação de valores exatos, enquanto outros são mais bem compreendidos por meio da visualização de comparações, distribuições, relações ou mudanças ao longo do tempo.

Um mesmo estudo pode produzir diferentes conjuntos de dados e, por isso, utilizar mais de uma forma de apresentação. Tabelas podem ser previstas quando for necessário consultar valores exatos; gráficos, quando se pretende facilitar a visualização de padrões ou diferenças; e quadros, quando houver informações predominantemente textuais que precisem ser comparadas. A escolha específica de cada representação será tratada no item 3.5.

Um exemplo prático e hipotético seria:

- **Objetivo**: Verificar as publicações brasileiras que tratam dos temas da ciência aberta.
- **Unidade**: publicações técnico-científicas

- **Critérios de inclusão e exclusão:** Periódicos de acesso aberto publicados no Brasil, exceto aqueles que cobram *Article Processing Charge* (APC).
- **Viabilidade**: A quantidade de periódicos presentes no universo pode aumentar ou diminuir a depender dos critérios de inclusão e exclusão. A inclusão de critérios temporais e geográficos, por exemplo, diminuiria a quantidade de elementos do universo; alternativamente, seria possível optar pelo uso de amostra.
- **Instrumento de pesquisa:** Pesquisa bibliográfica, visto que precisa coletar informações sobre as publicações técnico-científicas
- **Modelo de apresentação dos dados:** os resultados podem exigir diferentes formas de apresentação. A quantidade de revistas por área do conhecimento pode ser comparada por meio de um gráfico de barras, enquanto a quantidade de artigos publicados por ano pode ser apresentada em um gráfico de linhas, permitindo acompanhar sua evolução ao longo do tempo.

Cabe destacar que, nessa etapa, as formas de apresentação são previstas com base nos dados que se espera obter. Após a coleta e o tratamento, essa escolha deve ser confirmada ou ajustada conforme as características efetivamente observadas na base.

3.1.5 Definindo o Modelo de Análise

A análise dos dados geralmente recebe destaque na pesquisa. Entretanto, quando os dados não possuem qualidade, os resultados do estudo podem ser comprometidos. Da mesma forma, a seleção do modelo de análise dos dados também é vital para o sucesso do estudo. Por isso, definir antecipadamente os modelos a serem utilizados pode facilitar a realização da análise.

De forma simplificada, os principais modelos de análise de dados quantitativos são apresentados a seguir:

- **Tabulação cruzada**: utilizada para verificar relações entre dados distintos, mas com alguma conexão em formato tabular (matriz), para perceber tendências que a visão isolada não possibilita.
- **Análise de Tendências**: utilizada para encontrar padrões e propor indicações comportamentais futuras, verificar como uma variável se comporta, podendo verificar as suas variações.
- **Análise conjunta:** possibilita a avaliação de preferências em pesquisas de mercado ou estudos similares, possibilitando analisar mais de um experimento ou processo ao mesmo tempo.
- **Turf (*Total Unduplicated Reach and Frequency*)**: utilizada para identificar preferências entre itens, evitando sobreposições. É empregada principalmente em pesquisas de mercado e na simulação de cenários com diferentes opções.
- **Análise de Lacunas:** utilizada para comparar o desempenho real com um desejado, buscando atingir as metas até que os dois cenários se igualem.
- **Análise de texto:** Esse método não deve ser confundido com o mesmo nome para estudos qualitativos, relacionado ao conteúdo. Nesse caso, o texto é quantificado mediante a conversão de palavras, frases e outras estruturas em dados numéricos para análise quantitativa.
- **MaxDiff:** utilizado para estabelecer preferências entre itens dispostos em uma lista. Diferentemente das escalas do tipo Likert, o método solicita aos participantes que indiquem, entre as opções apresentadas, aquelas que consideram mais e menos relevantes.

Evidentemente, o modelo de análise dependerá do objetivo e dos tipos de dados, ou seja, o que quer obter dos dados coletados. Assim, todas as definições são interdependentes, e o planejamento da pesquisa ajuda a selecionar o método de análise entre as opções existentes. Com isso, busca-se ter melhores resultados na pesquisa.

Um exemplo prático e hipotético seria:

- **Objetivo**: Verificar as publicações brasileiras que tratam dos temas da ciência aberta.
- **Unidade**: publicações técnico-científicas
- **Critérios de inclusão e exclusão:** Periódicos de acesso aberto publicados no Brasil, exceto aqueles que cobram *Article Processing Charge* (APC).
- **Viabilidade**: A quantidade de periódicos presentes no universo pode aumentar ou diminuir a depender dos critérios de inclusão e exclusão. A inclusão de critérios temporais e geográficos, por exemplo, diminuiria a quantidade de elementos do universo; alternativamente, seria possível optar pelo uso de amostra.
- **Instrumento de pesquisa:** Pesquisa bibliográfica, visto que precisa coletar informações sobre as publicações técnico-científicas
- **Modelo de apresentação dos dados:** a pesquisa bibliográfica coleta vários conjuntos de dados e pode-se selecionar modelos diferentes para cada tipo de dados:
    - **Quantidade de revistas por área de conhecimento:** gráfico de barras;
    - **Quantidade de artigos publicados por ano:** gráfico de linhas.
- **Modelo de análise:** Análise de tendência, na busca de tendências

Possivelmente, muitos estudos não apresentam o modelo de análise e dados utilizados na pesquisa, visto que para muitos é intrínseco aos processos do estudo. Entretanto, ter um modelo de análise ajuda na condução da pesquisa, facilitando como a pesquisa decorre. Da mesma forma, ajuda os avaliadores a entender as seleções efetuadas na pesquisa.

3.2. Coletando os dados

A coleta de dados é a etapa em que as informações necessárias para responder ao problema investigado são obtidas e registradas de

forma sistemática. Sua execução está vinculada ao planejamento, pois os objetivos, as variáveis e as unidades de análise orientam o que será coletado, de quem ou de qual fonte, em que momento e por quais procedimentos. Assim, a coleta não se limita à reunião de informações disponíveis, mas exige decisões coerentes com o problema e com o planejamento do estudo.

3.2.1 Origem dos dados

A origem dos dados está relacionada à maneira como eles são obtidos. Os dados primários são produzidos especificamente para o estudo, mediante procedimentos definidos pelo estudante ou pesquisador. Já os dados secundários são recuperados de documentos, bases, arquivos ou sistemas preexistentes, muitas vezes constituídos para outras finalidades. De acordo com Fowler Jr. (2014), antes de iniciar uma coleta específica, deve-se verificar se as informações necessárias já estão disponíveis e se são adequadas aos objetivos da pesquisa.

Na Ciência da Informação, essas origens podem ser utilizadas isoladamente ou de forma combinada. Uma pesquisa sobre o uso de uma biblioteca, por exemplo, pode relacionar respostas de questionários aos registros de empréstimos da instituição. Entretanto, a disponibilidade de informações não assegura sua adequação ao estudo. Conforme Neuendorf (2017), a constituição da população documental exige critérios explícitos de seleção, considerando a abrangência e as limitações das fontes.

3.2.2 Ética e proteção dos dados

A coleta de dados deve respeitar princípios éticos que protejam os participantes e assegurem o uso responsável das informações. Antes de iniciá-la, é necessário identificar os possíveis riscos, selecionar apenas os dados necessários e definir medidas para preservar a privacidade, a autonomia e a dignidade das pessoas envolvidas. Na Ciência da Informação, esses cuidados abrangem tanto o contato com

participantes quanto o uso de documentos, imagens e registros que contenham informações pessoais. Conforme Brasil (2016), as pesquisas em Ciências Humanas e Sociais devem observar os princípios relativos ao consentimento, à confidencialidade e à proteção da identidade dos participantes.

Nesse contexto, os participantes devem ser informados, em linguagem clara, sobre os objetivos, os procedimentos, os possíveis riscos e benefícios e a utilização prevista para os dados. Também precisam saber que podem recusar ou interromper sua participação sem prejuízo. Antes da coleta, deve-se verificar com o orientador, quando for o caso, as formas adequadas de esclarecimento e registro, pois o consentimento constitui um processo contínuo de comunicação, e não apenas a assinatura de um documento.

Além disso, a proteção dos dados exige atenção à identificação direta ou indireta dos participantes. Mesmo sem nomes ou documentos, a combinação de informações como idade, função, instituição e localidade pode permitir o reconhecimento de uma pessoa. Por isso, devem ser coletados somente os dados necessários e, quando possível, os identificadores devem ser substituídos por códigos e mantidos separadamente do conjunto utilizado na análise. De acordo com Brasil (2018), o tratamento de dados pessoais deve observar os princípios da finalidade, da necessidade, da segurança e da prevenção.

Por fim, devem ser definidos e documentados o acesso, o armazenamento, o compartilhamento, o período de conservação e o descarte dos dados. Arquivos digitais podem exigir senhas, cópias de segurança e controle de acesso, enquanto documentos físicos devem permanecer em local protegido. O projeto e o relatório devem registrar essas medidas, as pessoas autorizadas a consultar os dados e eventuais alterações, permitindo avaliar a coerência entre o tratamento realizado e as condições comunicadas aos participantes.

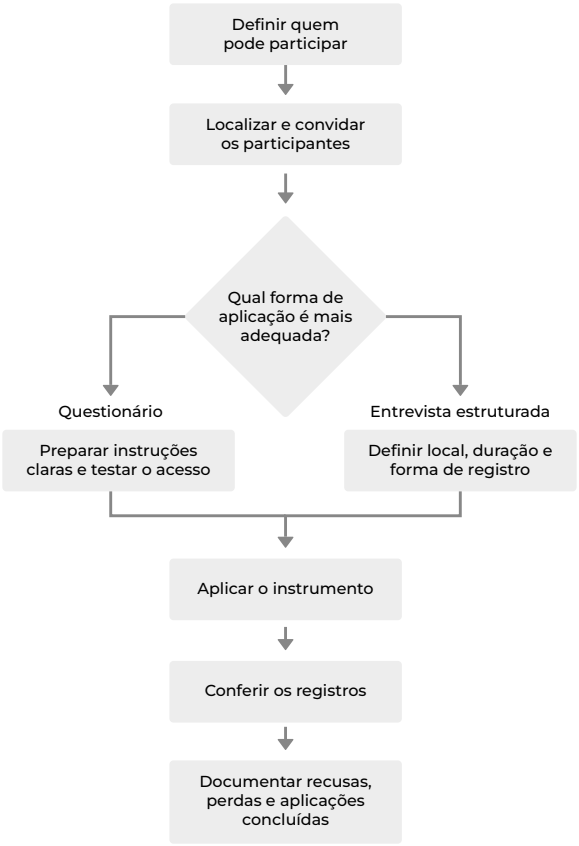
3.2.3 Procedimentos de coleta na prática

Definidos os dados necessários e observados os cuidados éticos e de proteção, deve-se escolher o procedimento adequado para obtê-los. A coleta pode envolver o contato com participantes, a observação estruturada ou a extração de informações de documentos e sistemas. Essa escolha deve considerar o problema investigado e as condições de acesso às pessoas, às situações ou às fontes.

3.2.3.1 Coleta com participantes

A coleta com participantes permite obter informações diretamente das pessoas relacionadas ao fenômeno investigado, por meio de questionários ou entrevistas estruturadas. Antes de iniciá-la, deve-se definir quem pode participar, como essas pessoas serão localizadas e convidadas e qual forma de aplicação é mais adequada. Conforme Fowler Jr. (2014), essa escolha deve considerar as características dos participantes, as possibilidades de contato, os recursos disponíveis e os efeitos do procedimento sobre a qualidade das respostas. A Figura 3 apresenta as principais decisões e atividades dessa coleta.

Figura 3 - Etapas da coleta de dados com participantes



Fonte: elaboração própria (2026).

Na Ciência da Informação, uma pesquisa sobre a utilização de um arquivo universitário, por exemplo, pode encaminhar um questionário aos usuários ou aplicar as mesmas perguntas por meio de entrevistas estruturadas. Em ambos os casos, devem ser mantidos os critérios de seleção e de registro, além dos cuidados com consentimento, privacidade e proteção dos dados apresentados na subseção sobre ética e proteção dos dados.

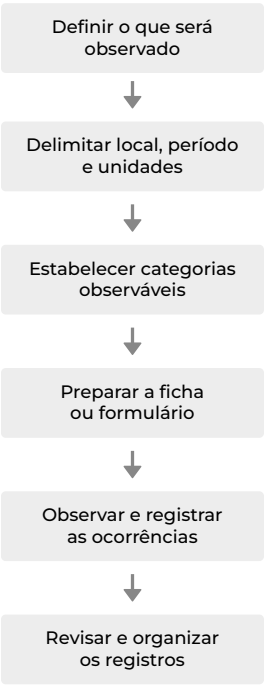
3.2.3.2 Observação estruturada

A observação estruturada consiste no acompanhamento e no registro de comportamentos, eventos ou situações segundo categorias e procedimentos previamente estabelecidos. Outras denominações podem ser encontradas na literatura, especialmente observação sistemática. Neste guia, adota-se a expressão observação estruturada para diferenciá-la da amostragem sistemática, que se refere à seleção das unidades que integrarão uma amostra. Conforme Glazier (1985), a observação é estruturada quando categorias predeterminadas orientam o registro das ocorrências durante períodos definidos.

Antes de iniciar a coleta, deve-se definir o que será observado, em quais locais e períodos e quais unidades serão consideradas. Em seguida, o fenômeno investigado precisa ser traduzido em categorias que possam ser reconhecidas durante a observação. Uma categoria como "uso do serviço de referência", por exemplo, ainda é muito ampla e pode ser desdobrada em ocorrências observáveis, como solicitação de informação, orientação para uma fonte, encaminhamento para outro setor e conclusão do atendimento. Cada categoria deve apresentar uma definição clara, para que situações semelhantes sejam registradas da mesma forma. Características que não podem ser diretamente observadas, como satisfação, intenção ou motivação, não devem ser inferidas apenas pelo comportamento das pessoas.

Para orientar a coleta, pode-se utilizar uma ficha, um checklist ou um formulário digital com as categorias e os campos necessários ao registro. As ocorrências devem ser anotadas durante a observação ou logo após seu término, mediante descrições objetivas, diferenciando os fatos observados das interpretações do observador. Também devem ser registradas situações que possam ter afetado a coleta, como interrupções, mudanças no ambiente ou dificuldade para acompanhar ocorrências simultâneas. A Figura 4 sintetiza esse processo, desde a delimitação do que será observado até a revisão dos registros produzidos.

Figura 4 - Etapas da observação estruturada



Fonte: elaboração própria (2026).

Na Ciência da Informação, a observação estruturada pode ser utilizada, por exemplo, para estudar o atendimento em uma biblioteca. Pode-se definir períodos de observação e registrar o tipo de solicitação apresentada, o recurso consultado, a necessidade de encaminhamento e a conclusão do atendimento. Glazier (1985) considera essa técnica adequada para uso por bibliotecários acadêmicos, quando combinada com entrevistas ou outros procedimentos de coleta, na investigação do comportamento e das necessidades informacionais de grupos específicos. O autor também menciona sua aplicação na observação de atividades de ensino, reuniões de docentes, reuniões curriculares e trabalhos desenvolvidos por grupos específicos.

3.2.3.3 Documentos, bases, registros e rastros digitais

A coleta em documentos e sistemas deve seguir um protocolo que permita localizar novamente as fontes e reproduzir a extração realizada. Para isso, deve-se registrar a base ou o acervo consultado, o período abrangido, os termos de busca, os campos pesquisados, os filtros aplicados e a data da consulta. Quando possível, deve-se preservar a expressão ou a estratégia de busca utilizada e conservar uma cópia do conjunto recuperado, pois bases atualizadas continuamente podem apresentar resultados diferentes em acessos posteriores. A Figura 5 sintetiza o processo de coleta em documentos e sistemas, desde a definição das fontes até o registro das limitações identificadas.

Figura 5 - Processo de coleta em documentos e sistemas



Fonte: elaboração própria (2026).

Em sistemas digitais, podem ser coletados rastros como pesquisas efetuadas, acessos, cliques, downloads e consultas a registros, além de transações como empréstimos. Esses dados, entretanto, representam apenas os eventos que o sistema foi programado para registrar. Em uma pesquisa sobre o uso de um catálogo, por exemplo, podem ser extraídos os termos pesquisados, a data da consulta, a quantidade de resultados e os itens acessados. A ausência de uma ação nos registros não permite concluir, por si só, que ela não tenha ocorrido. Os procedimentos utilizados na extração e as limitações do sistema devem ser documentados para orientar a análise dos dados.

3.2.4 Instrumentos, pré-teste e estudo-piloto

Os instrumentos de coleta transformam as variáveis definidas no planejamento em perguntas, categorias, campos ou procedimentos de registro. Conforme a técnica adotada, podem assumir a forma de questionário, roteiro de entrevista, ficha de observação ou formulário para extração de dados de documentos e sistemas. Ao elaborá-los, deve-se relacionar cada item aos objetivos do estudo e prever como as respostas ou ocorrências serão registradas e analisadas. A inclusão de perguntas ou campos sem função definida aumenta o trabalho de coleta e pode produzir dados que não contribuem para responder ao problema investigado.

Após a elaboração da primeira versão, o instrumento deve ser revisado para verificar a clareza das instruções, a ordem dos itens, a adequação da linguagem e o tempo necessário para sua aplicação ou seu preenchimento. Nos questionários e roteiros de entrevista, é preciso examinar se as perguntas podem ser compreendidas de maneira consistente e se as alternativas permitem registrar as respostas esperadas. Conforme Fowler Jr. (2014), a elaboração desses instrumentos envolve decidir o que será medido, formular as perguntas e testá-las antes da coleta definitiva. Nas fichas de observação e nos formulários de extração, as categorias e os campos devem estar definidos de modo que ocorrências semelhantes sejam registradas segundo os mesmos critérios.

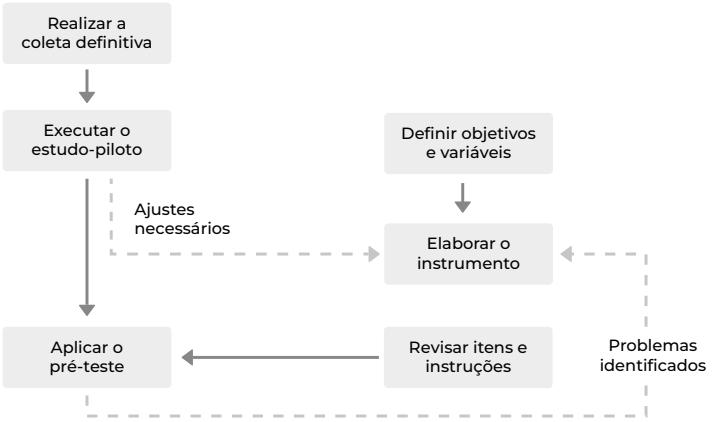
Neste guia, o pré-teste corresponde à avaliação do instrumento, enquanto o estudo-piloto representa uma execução reduzida do procedimento completo. O pré-teste consiste na aplicação do instrumento a pessoas ou unidades com características semelhantes às que integrarão o estudo. Sua finalidade é identificar perguntas ambíguas, instruções pouco claras, alternativas insuficientes, categorias difíceis de aplicar e problemas no registro das informações. Durante sua realização, devem-se observar as dificuldades encontradas e registrar o

tempo de aplicação. Quando houver participantes, também podem ser solicitados comentários sobre as perguntas e instruções. Os problemas identificados devem orientar a revisão do instrumento e, quando as alterações forem significativas, uma nova aplicação deve ser realizada.

O estudo-piloto, por sua vez, permite testar o procedimento completo em condições próximas às da pesquisa, sendo especialmente útil quando a coleta envolve várias etapas, fontes ou integrantes da equipe. Pode-se verificar o acesso aos participantes ou às fontes, a aplicação do instrumento, o registro, o armazenamento e a preparação dos dados para análise. Ao final, devem ser documentados os problemas encontrados e os ajustes realizados antes da coleta definitiva.

Como mostra a Figura 6, a preparação do instrumento não termina em sua primeira versão. O pré-teste e o estudo-piloto podem revelar problemas que exigem ajustes e nova verificação antes da coleta definitiva.

Figura 6 - Processo de elaboração e avaliação do instrumento de coleta



Fonte: elaboração própria (2026).

3.2.5 Execução e qualidade da coleta

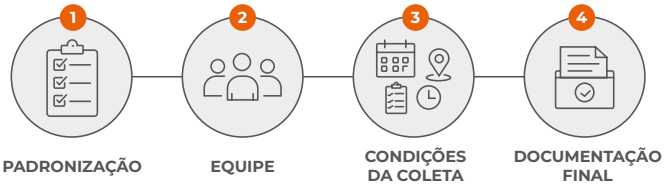
Depois de testados o instrumento e os procedimentos, a coleta definitiva deve ser executada de forma padronizada e documentada, visto que falhas nessa etapa podem comprometer a qualidade dos dados obtidos. Conforme Fowler Jr. (2014), a aplicação de procedimentos uniformes favorece a obtenção de informações comparáveis e reduz as variações produzidas pela forma como o instrumento é aplicado ou o dado é registrado. Assim, as informações devem ser obtidas e registradas segundo critérios comuns.

Nesse contexto, quando a coleta envolve uma equipe, todos os responsáveis devem receber orientações sobre a aplicação dos instrumentos, a sequência dos procedimentos e o tratamento de situações não previstas. Essa preparação deve ser complementada pelo acompanhamento da execução, mediante a revisão dos instrumentos preenchidos e a observação de parte das coletas.

Além da preparação da equipe, devem ser acompanhadas as condições em que os dados são obtidos, incluindo período, local, fonte e ocorrências capazes de afetar os registros. Alterações no instrumento, interrupções ou indisponibilidade de fontes podem produzir diferenças que não decorrem do fenômeno investigado. Por essa razão, essas ocorrências precisam ser registradas e consideradas na avaliação dos dados.

Por fim, a descrição metodológica deve distinguir os procedimentos previstos daqueles efetivamente realizados, incluindo alterações, perdas e medidas corretivas. Essa documentação permite avaliar a qualidade dos resultados e reconhecer diferenças que possam afetar sua comparabilidade ou a reprodução do estudo. A Figura 7 sintetiza os componentes relacionados à execução e à qualidade da coleta.

Figura 7 - Componentes relacionados à execução e à qualidade da coleta



Fonte: elaboração própria (2026).

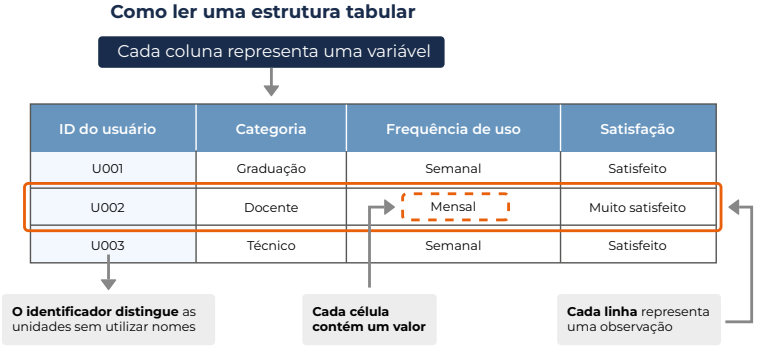
3.3. Tratando os dados

Concluída a coleta, os registros obtidos ainda não estão necessariamente prontos para análise. Respostas de questionários, fichas de observação, documentos e dados extraídos de sistemas podem apresentar formatos diferentes, campos incompletos, duplicações ou formas inconsistentes de registro. O tratamento consiste em organizar, verificar e padronizar essas informações, preservando sua relação com as unidades e as variáveis definidas no planejamento. Nessa etapa, não se devem modificar os dados para adequá-los aos resultados esperados; as correções e transformações precisam seguir critérios explícitos e documentados. Ao final, deve-se obter uma base consistente e adequada à exploração e à análise.

3.3.1 Organizando a base de dados

A organização da base começa pela definição do que cada linha, coluna e célula representa. Em uma estrutura tabular, cada coluna deve corresponder a uma variável, cada linha deve representar uma observação e cada célula deve conter um único valor. Também se deve incluir um identificador que permita distinguir as unidades analisadas sem depender de nomes ou de outros dados pessoais. Conforme Wickham (2014), essa organização facilita a manipulação e a análise porque mantém uma estrutura consistente entre observações e variáveis.

Figura 8 - Elementos de uma estrutura tabular de dados



Fonte: elaboração própria (2026).

A unidade representada em cada linha depende do fenômeno investigado. Em um questionário aplicado a usuários de uma biblioteca, por exemplo, cada linha pode representar um participante e as colunas podem registrar sua categoria de usuário, frequência de utilização e nível de satisfação. Em registros de empréstimos, por outro lado, cada linha pode representar uma transação, de modo que os identificadores do usuário e do item se repitam sempre que ocorrer um novo empréstimo. Portanto, antes de montar a base, é necessário definir claramente qual evento, pessoa, documento ou objeto corresponde a uma observação.

Quando a pesquisa utiliza diferentes tipos de unidades, pode ser necessário organizá-las em tabelas separadas e relacioná-las por meio de identificadores comuns. Dados sobre usuários, itens do acervo e empréstimos, por exemplo, podem compor três tabelas distintas, ligadas pelos códigos do usuário e do item. Essa separação evita reunir, em uma mesma estrutura, registros produzidos em níveis diferentes, como características dos usuários e ocorrências de empréstimos. Antes de consolidar as fontes, deve-se verificar, portanto, se os registros representam a mesma unidade e apresentam o mesmo nível de detalhamento.

3.3.2 Verificando erros e inconsistências

Depois de organizada, a base deve ser examinada para identificar registros que possam comprometer a análise. Essa verificação pode começar pelos identificadores, procurando duplicações indevidas ou unidades registradas mais de uma vez. Em seguida, devem-se conferir os valores aceitos para cada variável. Uma escala de satisfação definida entre 1 e 5, por exemplo, não pode conter o valor 7; da mesma forma, uma quantidade de empréstimos não pode ser negativa. Planilhas e programas estatísticos permitem aplicar filtros, ordenar valores e criar regras que sinalizem automaticamente registros fora dos limites estabelecidos.

Também é necessário verificar a coerência entre variáveis relacionadas. Em pesquisa sobre uso de biblioteca, uma pessoa não pode declarar que nunca utiliza o serviço e, simultaneamente, informar frequência semanal. Em registros de circulação, a data de devolução não pode ser anterior à data do empréstimo. Também se comparam totais, datas, categorias e identificadores com os instrumentos preenchidos ou com as fontes das quais os dados foram extraídos. Essas conferências ajudam a localizar erros de preenchimento, digitação, importação ou associação entre registros.

Também devem ser sinalizados os valores que se afastam dos padrões observados na base. Sua identificação, entretanto, indica apenas a necessidade de conferência e não comprova a existência de erro. Conforme Van den Broeck *et al*. (2005), deve-se distinguir a detecção de um registro suspeito do diagnóstico de sua verdadeira condição. Nessa etapa, o valor identificado deve ser comparado com o instrumento original, com as regras definidas para a variável e com outras informações relacionadas ao registro. As decisões sobre sua manutenção, correção ou exclusão serão tratadas no item 3.3.4.

3.3.3 Codificando e padronizando os registros

Após a verificação de erros e inconsistências, os registros devem ser codificados e padronizados para que informações equivalentes sejam representadas da mesma forma. A codificação consiste em atribuir rótulos ou códigos às categorias de uma variável, enquanto a padronização busca uniformizar grafias, formatos, unidades de medida e convenções de registro. Esses procedimentos não devem alterar o significado dos dados, mas permitir que valores equivalentes sejam identificados e analisados conjuntamente.

Em uma base sobre usuários de uma unidade de informação, por exemplo, respostas como “Graduação”, “graduacao”, “GRAD” e “aluno de graduação” podem representar a mesma categoria. Nesse caso, pode-se adotar a forma padronizada “Graduação”. Também é possível atribuir códigos, como “1 = Graduação”, “2 = Pós-graduação” e ”3 = Docente”. Esses números funcionam apenas como rótulos: o código 3 não representa uma quantidade maior que o código 1 e, portanto, não permite operações como soma ou média.

A padronização também deve abranger datas, unidades de medida, nomes institucionais e respostas recorrentes. Em uma mesma variável, não se devem misturar registros como “Sim”, “S” e “verdadeiro”, nem datas representadas por convenções diferentes. Deve-se escolher uma forma única para cada tipo de informação e aplicá-la a toda a base. O Quadro 2 apresenta alguns exemplos.

Quadro 2 - Exemplos de codificação e padronização dos registros

Registro original	Problema identificado	Registro padronizado
Graduação; graduacao; GRAD	Grafias diferentes para a mesma categoria	Graduação
Sim; S; verdadeiro	Formas diferentes de resposta positiva	Sim
30/08/26; 2026-08-30; 30 ago. 2026	Formatos diferentes de data	30/08/2026
15 minutos; 0,25 hora	Unidades de medida diferentes	15 minutos

Fonte: elaboração própria (2026).

Para evitar alterações arbitrárias, devem-se definir previamente as categorias, os códigos e as regras de conversão que serão utilizados. Também é necessário distinguir valores ausentes de respostas negativas ou valores iguais a zero: um campo não preenchido não significa necessariamente "não", "nenhum" ou "zero". Ao final, deve-se verificar se as regras foram aplicadas de forma uniforme e se os valores originais podem ser recuperados quando for necessário conferir as transformações realizadas.

3.3.4 Tratando dados ausentes e registros problemáticos

Os dados ausentes correspondem a informações que deveriam ou poderiam ter sido registradas, mas não estão disponíveis na base. Essa ausência pode ocorrer porque uma pergunta não foi respondida, um campo não se aplicava à unidade observada, a informação não constava na fonte ou houve falha durante a coleta ou a importação. Essas situações não possuem o mesmo significado e, sempre que for possível identificá-las, devem ser representadas por categorias distintas.

Um campo vazio não deve ser automaticamente preenchido com zero ou com uma resposta negativa. Em uma pesquisa com usuários,

por exemplo, a ausência de resposta sobre a quantidade de visitas à biblioteca não significa que nenhuma visita tenha ocorrido. Da mesma forma, a falta da data de devolução pode indicar que o item ainda está emprestado, que o registro não foi atualizado ou que a informação foi perdida. Antes de tratar o valor ausente, deve-se investigar sua origem e considerar o contexto em que o dado foi produzido.

Quadro 3 - Exemplos de dados ausentes e seus possíveis tratamentos

Situação encontrada	Significado possível	Forma de tratamento
Campo de idade vazio	Participante não respondeu	Registrar como “não informado”
Número de empréstimos igual a zero	Nenhum empréstimo realizado	Manter o valor zero
Data de devolução vazia	Item ainda emprestado	Registrar como “não se aplica no momento”
Campo inexistente em parte dos documentos	Informação não constava na fonte	Registrar como “não disponível”
Resposta ilegível	Valor coletado, mas impossível de identificar	Registrar como “não foi possível determinar”

Fonte: elaboração própria (2026).

Além dos dados ausentes, a verificação da base pode revelar duplicações, datas contraditórias, respostas incompatíveis com o instrumento e valores que exigem investigação. Depois da conferência descrita no item 3.3.2, deve-se definir o tratamento de cada situação com base nas informações disponíveis. Um número elevado de acessos a um repositório, por exemplo, pode ser confirmado como utilização efetivamente intensa, identificado como atividade automatizada ou reconhecido como erro de registro. Cada uma dessas condições exige uma decisão diferente.

Conforme o resultado da investigação, o registro pode ser mantido, corrigido com base na fonte original, classificado como ausente

ou excluído da análise. Quando não for possível confirmar o valor correto, não se deve substituí-lo por uma estimativa escolhida arbitrariamente. Técnicas de imputação, que atribuem valores aos dados ausentes, exigem critérios metodológicos compatíveis com o tipo de variável e com a análise pretendida. Em qualquer situação, a decisão e sua justificativa devem ser documentadas, preservando-se uma cópia da base original.

3.3.5 Documentando as transformações

As alterações realizadas durante o tratamento devem ser documentadas para que se possa compreender como os registros originais deram origem à base utilizada na análise. Essa documentação deve abranger correções, padronizações, recodificações, exclusões, criação de variáveis e tratamento de dados ausentes. Conforme Van den Broeck *et al*. (2005), as modificações feitas durante a limpeza dos dados precisam ser registradas para permitir sua avaliação e reprodução.

Para preservar os dados coletados, recomenda-se manter pelo menos duas versões da base: uma original, sem alterações, e outra destinada ao tratamento e à análise. Quando o processo envolver várias etapas, podem-se conservar também versões intermediárias identificadas pela data ou por um número sequencial. A base original não deve ser sobrescrita, pois poderá ser necessária para verificar registros ou refazer o tratamento com critérios diferentes.

Além das versões da base, deve-se elaborar um registro das transformações realizadas. Para cada alteração, é necessário indicar a variável ou o registro afetado, o problema identificado, a decisão adotada e a justificativa correspondente. Quando houver consulta à fonte original, essa informação também deve ser registrada.

Quadro 4 - Exemplo de registro das transformações realizadas na base

Data	Registro ou variável	Situação encontrada	Transformação realizada	Justificativa
02/09/2026	Perfil do usuário	“GRAD” e “graduacao”	Padronização para “Graduação”	Registros representam a mesma categoria
02/09/2026	Data de devolução – registro 154	Data anterior ao empréstimo	Correção de 05/08 para 15/08	Conferência no sistema de empréstimos
03/09/2026	Satisfação – registro 208	Valor 7 em escala de 1 a 5	Classificação como dado ausente	Não foi possível determinar a resposta correta
03/09/2026	Identificador 327	Registro repetido	Exclusão de uma ocorrência	Duplicação confirmada na importação

Fonte: elaboração própria (2026).

Também é importante manter um dicionário de dados com o nome e a descrição das variáveis, os tipos de valores admitidos, as unidades de medida e os códigos utilizados. Caso seja criada uma variável, como a faixa etária derivada da idade, devem-se registrar tanto a fórmula ou regra empregada quanto os limites adotados para formar as categorias.

Ao final do tratamento, a base, o dicionário de dados e o registro das transformações devem permanecer associados. Esse conjunto permite conferir as decisões tomadas, corrigir eventuais problemas e repetir o processo em uma nova versão dos dados. A documentação, portanto, não é apenas um registro administrativo: ela integra o método da pesquisa e oferece transparência às etapas que antecedem a análise.

3.4. Conhecendo os dados

Depois de organizada, verificada e padronizada, a base pode ser explorada para que se conheçam suas principais características. Essa etapa consiste em examinar como os valores estão distribuídos, quais categorias aparecem com maior frequência, quais valores representam melhor o conjunto e quanto as observações variam entre si. Diferentemente do tratamento, seu propósito não é corrigir os registros, mas descrevê-los e identificar aspectos que deverão ser considerados nas análises posteriores.

A forma de explorar os dados depende do tipo de variável e dos objetivos estabelecidos no planejamento. Variáveis qualitativas podem ser resumidas por frequências e proporções, enquanto variáveis quantitativas permitem calcular medidas que representam a posição e a dispersão dos valores. Tabelas e gráficos também podem auxiliar na identificação de concentrações, diferenças, assimetrias e valores incomuns. Esses resultados descrevem o comportamento dos dados, mas ainda não explicam suas causas ou implicações, que serão discutidas na etapa de interpretação.

3.4.1 Frequências, proporções e tabelas

Para conhecer uma variável, pode-se começar pela contagem de quantas vezes cada valor ou categoria aparece na base. Essa contagem é denominada frequência absoluta. Em uma pesquisa com 40 usuários de uma biblioteca universitária, por exemplo, pode-se verificar quantos são estudantes de graduação, estudantes de pós-graduação, docentes, técnicos ou usuários externos. A frequência permite reconhecer as categorias mais e menos presentes no conjunto analisado.

Entretanto, a contagem isolada nem sempre possibilita comparar grupos ou pesquisas com tamanhos diferentes. Para isso, pode-se calcular a frequência relativa, dividindo a quantidade de ocorrências

de uma categoria pelo total de registros válidos. Quando o resultado é multiplicado por 100, obtém-se o percentual:

$$Percentual = \frac{frequência\ da\ categoria}{total\ de\ registros\ válidos} x100$$

A Tabela 1 apresenta a distribuição dos 40 usuários considerados no exemplo.

Tabela 1 - Distribuição dos usuários por categoria

Categoria de usuário	Frequência	Percentual
Graduação	18	45%
Pós-graduação	10	25%
Docente	6	15%
Técnico	4	10%
Usuário externo	2	5%
Total	**40**	**100%**

Fonte: elaboração própria (2026).

A leitura da Tabela 1 permite afirmar que os estudantes de graduação constituem a categoria mais frequente, com 18 usuários, correspondentes a 45% do conjunto analisado. A apresentação conjunta da frequência e do percentual oferece duas informações complementares: a primeira mostra a quantidade observada, enquanto a segunda indica a participação dessa quantidade no total.

No cálculo dos percentuais, deve-se informar qual total foi utilizado como referência. Se a variável apresentar respostas ausentes, pode-se calcular o percentual apenas entre os registros válidos, desde que a quantidade de ausências também seja apresentada. Em uma base com 40 participantes e duas respostas ausentes, por exemplo, os percentuais das categorias podem ter como referência as 38 respostas válidas, mas as duas ausências não devem ser ocultadas.

Também é necessário observar se cada unidade pode pertencer a apenas uma categoria ou selecionar várias respostas. Em uma pergunta sobre os motivos da utilização da biblioteca, uma mesma pessoa pode indicar estudo, acesso à internet e empréstimo de livros. Nesse caso, a soma das frequências pode superar o número de participantes e a soma dos percentuais pode ultrapassar 100%. Isso não representa um erro, desde que a tabela informe que eram permitidas respostas múltiplas.

A Tabela 1 é um bom exemplo de tabela simples. Uma tabela simples é uma ferramenta que organiza os dados de forma clara e concisa. Esse tipo de tabela permite fácil visualização e análise. É a tabela ideal para apresentar dados que não exigem muito detalhamento ou relacionamentos complexos. Pode ser utilizada para representar variáveis qualitativas ou quantitativas discretas por meio da distribuição de frequência. Especificamente no caso de variáveis quantitativas contínuas, recomenda-se que os dados sejam agrupados em intervalos de classe.

A Tabela 2 apresenta a distribuição de frequência dos usuários por tempo de permanência na biblioteca.

Tabela 2 - Distribuição dos usuários por tempo de permanência na biblioteca

Tempo de Permanência (min)	Frequência	Percentual
0 ⊢ 30	9	22,5%
30 ⊢ 60	20	50%
60 ⊢ 90	5	12,5%
90 ⊢ 120	5	12,5%
120 ⊢ 150	1	2,5%
Total	**40**	**100%**

Fonte: elaboração própria (2026).

Uma variável quantitativa contínua pode assumir valores muito distintos ou com muitas casas decimais. Se listássemos cada valor na tabela individualmente não seria prático nem informativo. Nesses casos podemos agrupar os dados em intervalos de classe, na Tabela 2 o tempo de permanência foi agrupado em intervalos de classe de 30 em 30 minutos. Assim, a tabela apresenta a frequência de ocorrências dentro de cada intervalo.

Para construir os intervalos de classe é necessário encontrar o valor máximo e o valor mínimo do conjunto de dados e calcular a diferença entre os dois. O valor obtido pela diferença é chamado de amplitude. O resultado da amplitude deve ser dividido pelo número de classes que se deseja. O resultado da divisão é o intervalo de classe que será utilizado.

No conjunto de dados da Tabela 2 o menor valor é 0 e o maior valor é 150. Calculando a amplitude temos:

$$Amplitude = Valor\ Máximo - Valor\ Mínimo$$
$$Amplitude = 150 - 0 = 150$$

Definiremos cinco classes, sendo assim o intervalo de classe será obtido pela divisão da amplitude pela quantidade de classes:

$$Intervalo\ de\ Classe = Amplitude/número\ de\ classes$$
$$Intervalo\ de\ Classe = 150/5 = 30$$

Cada classe deve apresentar um intervalo de 30 minutos. Os valores apresentados em cada classe são os extremos de classe e a notação utilizada mostra como o intervalo deve ser interpretado. Por exemplo, na primeira classe temos o intervalo 0 ⊢ 30, isso significa que o intervalo é fechado à esquerda e aberto à direita, ou seja, pertencem à classe os valores iguais a 0 e menores que 30. Ou seja, o valor do extremo inferior pertence à classe, mas o valor do extremo superior não; este último será incluído na classe seguinte. Existem outras formas de

definir os intervalos, como, por exemplo, a regra de Sturges, cabe ao pesquisador selecionar a forma que melhor se aplica ao seu estudo.

Além das tabelas de classificação simples exemplificadas pelas Tabelas 1 e 2 também podemos representar os dados por meio de tabelas duplas, também chamadas de tabelas de dupla entrada ou tabelas de contingência. O uso de tabelas duplas é recomendado quando estamos trabalhando com variáveis qualitativas e é necessário apresentar, em uma única tabela, a variação de valores de mais de uma variável, ou seja, quando é necessário conjugar variáveis. A Tabela 3 é uma tabela de contingência que conjuga as variáveis: categoria de usuário e tempo de permanência na biblioteca.

Tabela 3 - Categoria de usuário e tempo de permanência na biblioteca

Categoria de usuário	Permanência inferior a 30 min	Permanência superior a 30 min	Total
Graduação	12	6	18
Pós-graduação	4	6	10
Docente	3	3	6
Técnico	1	3	4
Usuário externo	0	2	2
Total	20	20	40

Fonte: elaboração própria (2026).

No exemplo da Tabela 3 a frequência é mostrada em valor absoluto, mas cabe salientar que tabelas de contingência também podem apresentar as frequências relativas em porcentagem.

3.4.2 Medidas de posição

Depois de conhecer as frequências, pode-se buscar um valor que represente a posição central dos dados. Para isso, são utilizadas

principalmente a média, a mediana e a moda. Embora essas medidas sejam frequentemente apresentadas juntas, cada uma resume o conjunto de maneira diferente e sua escolha depende do tipo de variável e da distribuição dos valores.

A média aritmética é obtida pela soma dos valores x_i, dividida pela quantidade de observações válidas. Sendo representada pela fórmula:

$$Média = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Em que

n : *número de observações*

x_i : *valores individuais das observações*

Em cinco atendimentos realizados por uma unidade de informação, por exemplo, foram registrados os tempos de 4, 5, 5, 6 e 20 minutos. A soma desses valores é 40; dividindo-a pelos cinco atendimentos, obtém-se a média de 8 minutos:

$$Média = \frac{4+5+5+6+20}{5} = 8$$

A mediana corresponde ao valor central depois que os dados são colocados em ordem. No exemplo, a sequência já está ordenada — 4, 5, 5, 6 e 20 — e o valor central é 5 minutos. A mediana indica, portanto, que metade dos atendimentos durou até 5 minutos e a outra metade durou pelo menos esse tempo. Diferentemente da média, ela é pouco influenciada pelo atendimento de 20 minutos, que se distancia dos demais. Em conjuntos maiores executamos um cálculo para encontrar a mediana que considera posição e valor. A posição pode ser calculada da seguinte forma:

$$posição = \frac{n+1}{2}$$

O resultado obtido é a posição da mediana e não seu valor. Do exemplo temos 5 observações, logo, $n = 5$, sendo assim:

$$posição = \frac{5+1}{2} = 3$$

Ou seja, o elemento que ocupa a terceira posição (após ordenação) será o valor da mediana que nesse caso será igual a 5. Caso haja um número par de observações, calculamos a posição da mesma forma e a mediana será obtida pela média entre a posição anterior e a posterior.

Supondo que agora temos 6 atendimentos registrados nos tempos de 4, 5, 5, 6, 8, 20 minutos. O cálculo da posição seria:

$$posição = \frac{6+1}{2} = 3{,}5$$

O valor da mediana será obtido pela média calculada entre os valores da posição 3 e 4:

$$mediana = \frac{5+6}{2} = 5{,}5$$

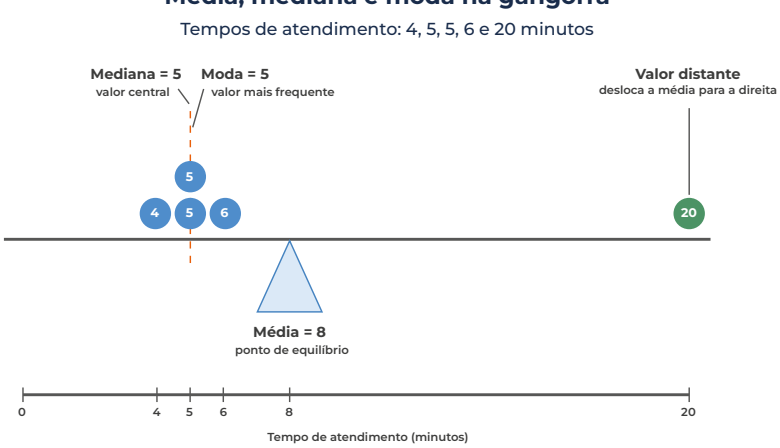
Assim, o valor da mediana de acordo com as 6 observações será igual a 5,5.

A moda é o valor ou a categoria que aparece com maior frequência. No exemplo, o tempo de 5 minutos ocorre duas vezes e constitui a moda. Essa medida também pode ser utilizada com variáveis qualitativas. Em uma pesquisa sobre a finalidade de uso de um arquivo, por exemplo, se “consulta para pesquisa acadêmica” for a resposta mais frequente, essa categoria será a moda. Um conjunto pode apresentar mais de uma moda ou não possuir um valor que se repita.

As diferenças entre as três medidas podem ser visualizadas na Figura 9. Na representação, os tempos de atendimento são posicionados

sobre uma gangorra como pesos de mesmo tamanho. A média corresponde ao ponto de equilíbrio, localizado em 8 minutos, pois o valor de 20 minutos exerce maior influência por estar distante dos demais. A mediana permanece em 5 minutos por ocupar a posição central dos valores ordenados, enquanto a moda também é igual a 5 minutos porque esse é o tempo registrado com maior frequência.

Figura 9 - Representação intuitiva da média, da mediana e da moda



Fonte: elaboração própria (2026).

Enquanto a Figura 9 permite visualizar o comportamento das medidas, a Tabela 4 reúne seus resultados numéricos e sintetiza o que cada uma representa no conjunto analisado.

Tabela 4 - Comparação das medidas de posição

Medida	Resultado	O que representa no exemplo
Média	8 minutos	Ponto de equilíbrio dos tempos de atendimento
Mediana	5 minutos	Valor central dos tempos ordenados
Moda	5 minutos	Tempo registrado com maior frequência

Fonte: elaboração própria (2026).

A leitura conjunta da figura e da tabela mostra que uma mesma base pode ser resumida por valores diferentes. Embora quatro dos cinco atendimentos tenham durado entre 4 e 6 minutos, o atendimento de 20 minutos deslocou a média para 8 minutos. A mediana e a moda permaneceram em 5 minutos e, nesse caso, representam melhor a concentração dos tempos observados.

Por isso, não se deve escolher automaticamente uma única medida. A média utiliza todos os valores e é adequada para variáveis quantitativas, mas pode ser influenciada por valores muito afastados. A mediana representa a posição central e pode descrever melhor distribuições assimétricas. A moda, por sua vez, identifica o valor ou a categoria mais frequente e pode ser utilizada com variáveis qualitativas. Sempre que necessário, pode-se apresentar mais de uma medida para oferecer uma visão mais completa dos dados.

3.4.3 Medidas de dispersão

As medidas de posição indicam um valor representativo, mas não mostram quanto os dados variam. Dois conjuntos podem apresentar a mesma média e, ainda assim, ter comportamentos bastante diferentes. Para demonstrar essa situação, podem-se comparar os tempos de atendimento registrados em duas unidades de informação:

- Unidade A: 6, 7, 8, 9 e 10 minutos;
- Unidade B: 2, 4, 8, 12 e 14 minutos.

Nos dois conjuntos, a média é igual a 8 minutos. Entretanto, os tempos da Unidade A estão próximos desse valor, enquanto os da Unidade B encontram-se mais afastados. As medidas de dispersão permitem quantificar essa diferença, indicando quanto os valores se concentram ou se espalham em torno de uma medida de posição.

A amplitude é obtida pela diferença entre o maior e o menor valor do conjunto. Na Unidade A, o maior tempo é 10 minutos e o menor é 6,

resultando em uma amplitude de 4 minutos. Na Unidade B, a diferença entre 14 e 2 produz uma amplitude de 12 minutos:

$$Amplitude = valor\ máximo - valor\ mínimo$$

Embora seja simples, a amplitude considera apenas os dois valores extremos e não mostra como as demais observações estão distribuídas. Para considerar todos os valores, pode-se utilizar a variância, calculada a partir das diferenças entre cada observação e a média. Como valores abaixo e acima da média produzem diferenças negativas e positivas, essas diferenças são elevadas ao quadrado antes de serem somadas. No exemplo, considerando os cinco atendimentos como o conjunto descrito, a variância é calculada por:

$$Variância\ populacional = \frac{\sum(x_i - \bar{x})^2}{n}$$

O cálculo anterior deve ser utilizado quando trabalhamos com os dados de toda população, caso o estudo esteja utilizando dados obtidos pela amostragem deve-se dividir a soma dos desvios ao quadrado pelo número de observações menos 1 (n − 1):

$$Variância\ amostral = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

A variância da Unidade A é igual a 2, enquanto a da Unidade B é igual a 20,8. Entretanto, como as diferenças são elevadas ao quadrado, a variância é expressa em unidades também elevadas ao quadrado. Para retornar à unidade original dos dados, calcula-se o desvio-padrão, obtido pela raiz quadrada da variância:

$$\text{Desvio-padrão} = \sqrt{\text{variância}}$$

Assim, o desvio-padrão é de aproximadamente 1,41 minuto na Unidade A e 4,56 minutos na Unidade B. Quanto menor o desvio-padrão, maior a concentração dos valores em torno da média; quanto maior, mais dispersos estão os dados.

Tabela 5 - Comparação das medidas de dispersão dos tempos de atendimento

Medida	Unidade A	Unidade B
Média	8 minutos	8 minutos
Valor mínimo	6 minutos	2 minutos
Valor máximo	10 minutos	14 minutos
Amplitude	4 minutos	12 minutos
Variância	2	20,8
Desvio-padrão	1,41 minuto	4,56 minutos

Fonte: elaboração própria (2026).

A Tabela 5 mostra que a igualdade entre as médias não significa que os atendimentos apresentem o mesmo comportamento. Na Unidade A, os tempos estão concentrados próximos de 8 minutos; na Unidade B, variam de 2 a 14 minutos. Por isso, uma medida de posição deve ser acompanhada por uma medida de dispersão sempre que a variação dos dados for relevante para o estudo.

Quando os dados correspondem a uma amostra utilizada para estimar características de uma população, a variância amostral é calculada com uma adaptação no denominador, utilizando-se $n - 1$ em lugar de n. Planilhas e programas estatísticos podem oferecer as duas formas de cálculo. Por essa razão, deve-se verificar qual medida foi utilizada e registrar essa informação na metodologia.

3.4.4 Distribuição dos dados

Além de conhecer os valores representativos e sua dispersão, é importante observar como as ocorrências estão distribuídas ao longo da escala. A distribuição mostra quais valores aparecem, com que frequência e em quais partes do conjunto eles se concentram. Essa observação permite reconhecer formatos que não são revelados apenas pela média ou pelo desvio-padrão.

Uma distribuição é considerada aproximadamente simétrica quando os valores se espalham de maneira semelhante nos dois lados de sua região central. Nas distribuições assimétricas, os dados se prolongam mais em uma das direções. Tempos de atendimento, quantidade de acessos e número de empréstimos, por exemplo, podem apresentar concentração em valores baixos ou intermediários e algumas ocorrências muito elevadas, formando uma distribuição assimétrica à direita. Em outras situações, a concentração pode ocorrer nos valores mais altos, com prolongamento em direção aos menores, caracterizando uma assimetria à esquerda.

A relação entre média e mediana pode oferecer uma indicação inicial desse formato. Em distribuições aproximadamente simétricas, essas medidas tendem a apresentar valores próximos. Quando há valores elevados distantes dos demais, a média pode ser deslocada para cima e tornar-se maior que a mediana. Quando os valores mais afastados se encontram na parte inferior da escala, pode ocorrer o contrário. Essa relação deve ser considerada apenas como um indício e precisa ser examinada em conjunto com os valores observados e suas representações gráficas.

Outra forma de conhecer a distribuição consiste em dividi-la em quatro partes por meio dos quartis. Depois de ordenados os valores, o primeiro quartil (Q_1) indica o ponto abaixo do qual se encontram aproximadamente 25% das observações; o segundo quartil (Q_2) corresponde à mediana; e o terceiro quartil (Q_3) indica o ponto abaixo do

qual se encontram aproximadamente 75% das observações. Para obter a posição dos quartis podemos utilizar os seguintes cálculos:

$$Q_1 = \frac{1}{4} \cdot (n + 1)$$

$$Q_2 = \frac{1}{2} \cdot (n + 1)$$

$$Q_3 = \frac{3}{4} \cdot (n + 1)$$

Em que:

Q_1, Q_2, Q_3: *indicam a posição de cada quartil*
n: *número de observações*

A diferença entre o terceiro e o primeiro quartil é denominada intervalo interquartil:

$$IIQ = Q_3 - Q_1$$

Para compreender a aplicação dessas medidas, considere agora outro conjunto, formado pelos tempos, em minutos, de onze atendimentos realizados em uma unidade de informação: 4, 5, 5, 5, 6, 6, 7, 8, 8, 9 e 20. A Tabela 6 apresenta as principais medidas utilizadas para descrever esse conjunto.

Tabela 6 - Resumo da distribuição dos tempos de atendimento

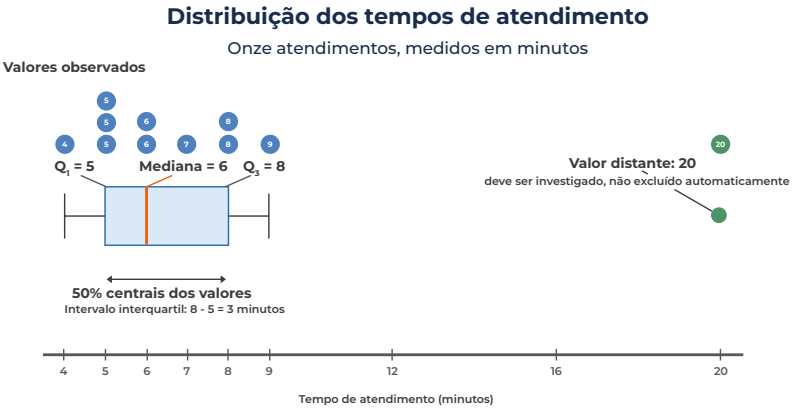
Medida	Resultado	Posição na distribuição
Valor mínimo	4 minutos	Menor tempo observado
Primeiro quartil (Q_1)	5 minutos	Aproximadamente 25% dos valores estão até esse ponto
Mediana (Q_2)	6 minutos	Divide os valores ordenados em duas partes
Terceiro quartil (Q_3)	8 minutos	Aproximadamente 75% dos valores estão até esse ponto
Valor máximo	20 minutos	Maior tempo observado
Intervalo interquartil	3 minutos	Faixa ocupada pela metade central dos valores

Fonte: elaboração própria (2026).

Como mostra a Tabela 6, os valores centrais da distribuição concentram-se entre 5 e 8 minutos. O primeiro quartil (Q_1) indica que aproximadamente 25% dos atendimentos duraram até 5 minutos, enquanto o terceiro quartil (Q_3) mostra que aproximadamente 75% duraram até 8 minutos. A diferença entre esses dois valores corresponde ao intervalo interquartil: $8 - 5 = 3$ minutos. Essa medida representa a faixa ocupada pelos 50% centrais dos registros e sofre menos influência de valores muito afastados.

A Figura 10 permite visualizar essa distribuição. Os círculos mostram os tempos registrados e suas frequências, enquanto a caixa representa o intervalo entre Q_1 e Q_3. A linha interna assinala a mediana, e os prolongamentos laterais indicam a extensão dos valores que permanecem próximos do conjunto central.

Figura 10 - Distribuição dos tempos de atendimento e intervalo interquartil



Fonte: elaboração própria (2026).

Na Figura 10, a maior parte dos atendimentos está concentrada entre 4 e 9 minutos, enquanto o registro de 20 minutos aparece afastado dos demais. Sua presença amplia a extensão da distribuição e exerce maior influência sobre a média e a amplitude. A mediana e o intervalo interquartil, por sua vez, são menos afetados porque descrevem a posição e a dispersão dos valores centrais. A representação gráfica permite reconhecer esse comportamento, mas não determina, por si só, se o registro está correto ou contém algum erro.

Quando a investigação realizada durante o tratamento confirma que o valor representa uma ocorrência válida, ele deve permanecer na base e sua influência precisa ser considerada na escolha e na interpretação das medidas estatísticas. Caso tenha sido corrigido, classificado como ausente ou excluído, a decisão deve constar no registro das transformações. A leitura conjunta da tabela e da figura permite, assim, reconhecer a posição dos dados, sua dispersão, o formato da distribuição e a presença de valores capazes de afetar sua descrição.

3.5. Representando os dados

Conforme previsto no planejamento, depois de conhecer as características do conjunto, é necessário confirmar ou ajustar as formas escolhidas para apresentar os resultados. Tabelas, quadros e gráficos permitem organizar as informações, destacar padrões e facilitar comparações que seriam difíceis de perceber em uma sequência de números. Essas representações, entretanto, não substituem a análise, pois selecionam e evidenciam determinados aspectos dos dados.

A escolha da representação deve considerar o tipo de variável, o objetivo da análise e a mensagem que se pretende comunicar. Uma tabela pode ser mais adequada quando é necessário consultar valores exatos; um quadro, quando se pretende organizar e comparar informações predominantemente textuais; e um gráfico, quando se deseja visualizar diferenças, distribuições, relações ou mudanças ao longo do tempo. Em todos os casos, títulos, rótulos, unidades de medida, fontes e notas devem fornecer as informações necessárias para uma leitura autônoma.

Representar os dados também exige cuidado para não induzir interpretações equivocadas. Escalas inadequadas, categorias desordenadas, excesso de elementos visuais e ausência de informações importantes podem ocultar diferenças ou exagerar resultados. Por isso, deve-se priorizar a clareza, a precisão e a correspondência entre os dados analisados e a forma utilizada para apresentá-los.

3.5.1 Organizando dados em tabelas e quadros

As tabelas apresentam dados de forma ordenada, permitindo consultar valores exatos e comparar categorias, períodos ou grupos. Diferentemente da base utilizada no tratamento e na análise, na qual cada linha costuma representar uma unidade e cada coluna uma variável, a tabela de apresentação reúne apenas as informações necessárias para comunicar determinado resultado. Portanto, não se

deve transferir toda a base de dados para o texto, mas selecionar os valores relacionados ao objetivo da análise.

Antes de construir uma tabela, deve-se definir o que será comparado e qual informação ocupará suas linhas e colunas. Em uma pesquisa sobre os canais utilizados para solicitar atendimento em uma unidade de informação, por exemplo, as linhas podem representar os canais identificados e as colunas, as frequências absoluta e relativa. Essa organização permite verificar tanto a quantidade de ocorrências quanto a participação de cada categoria no conjunto.

Tabela 7 - Solicitações de atendimento segundo o canal utilizado

	Frequência	Percentual
Presencial	42	42%
E-mail	28	28%
Aplicativo de mensagens	20	20%
Telefone	10	10%
Total	**100**	**100%**

Fonte: elaboração própria (2026).

Na Tabela 7, os canais foram organizados da maior para a menor frequência, o que facilita a comparação. Observa-se que o atendimento presencial concentrou 42% das solicitações, seguido pelo e-mail, com 28%. A coluna de frequência informa o número de registros, enquanto o percentual permite avaliar a participação de cada canal no total analisado. A apresentação conjunta dessas medidas ajuda a dimensionar as diferenças sem exigir que o leitor realize cálculos adicionais.

Para que a tabela possa ser compreendida de forma autônoma, seu título deve indicar o que está sendo apresentado e, quando necessário, o local e o período de referência. As linhas e colunas precisam ter rótulos claros, as unidades de medida devem ser identificadas e a fonte deve aparecer abaixo da estrutura. Notas podem ser acrescentadas para explicar abreviações, critérios de agrupamento ou

situações específicas, mas não devem substituir informações essenciais do título ou dos cabeçalhos.

Os mesmos cuidados de identificação e clareza devem ser aplicados aos quadros. Entretanto, como seu conteúdo é predominantemente textual, é necessário evitar descrições excessivamente longas nas células. Um quadro pode ser utilizado, por exemplo, para comparar instrumentos de coleta segundo sua finalidade, o tipo de informação produzida e suas limitações. Nesse caso, as colunas devem representar os mesmos critérios de comparação para todos os itens apresentados.

Quadro 5 - Comparação entre instrumentos de coleta de dados

Instrumento	Finalidade	Informação produzida	Limitação que exige atenção
Questionário	Obter respostas padronizadas de um conjunto de participantes	Respostas a perguntas abertas ou fechadas	Pode apresentar perguntas não respondidas ou interpretadas de maneiras diferentes
Roteiro de entrevista estruturada	Orientar a realização de perguntas previamente definidas	Respostas registradas segundo a mesma sequência	A forma de condução pode influenciar as respostas
Ficha de observação	Registrar comportamentos, eventos ou situações observáveis	Ocorrências classificadas em categorias predeterminadas	Depende da clareza das categorias e da uniformidade do registro
Formulário de extração de dados	Retirar informações de documentos, bases ou sistemas	Campos selecionados das fontes consultadas	As informações podem estar ausentes, incompletas ou registradas em formatos diferentes

Fonte: elaboração própria (2026).

No Quadro 5, os instrumentos de coleta estão organizados segundo os mesmos critérios de comparação: finalidade, informação produzida e limitação que exige atenção. Como seu conteúdo é predominantemente textual e não apresenta resultados numéricos, o quadro permite comparar as características dos instrumentos e reconhecer os cuidados relacionados à sua utilização.

Também se devem evitar tabelas e quadros excessivamente extensos ou carregados de detalhes. Quando houver muitas categorias ou informações complementares, pode-se apresentar no corpo do texto apenas o conteúdo relevante para a discussão e disponibilizar a relação completa em apêndice. Se a finalidade principal for mostrar visualmente diferenças, tendências ou padrões numéricos, um gráfico poderá complementar ou substituir uma tabela. A decisão depende, portanto, da pergunta que a representação deve ajudar a responder.

3.5.2 Escolhendo o gráfico adequado

Os gráficos permitem visualizar diferenças, tendências, distribuições e relações que podem ser difíceis de identificar em uma tabela ou em uma sequência de valores. Entretanto, cada tipo de gráfico organiza os dados de maneira diferente. Por isso, sua escolha deve partir da pergunta que se pretende responder, considerando também o tipo das variáveis e o aspecto que se deseja destacar.

Entre os gráficos mais utilizados, o gráfico de barras permite comparar categorias, enquanto o gráfico de linhas é adequado para acompanhar mudanças ao longo do tempo. Para observar como os valores de uma variável quantitativa estão distribuídos, pode-se recorrer ao histograma. Já o gráfico de dispersão permite examinar a relação entre duas variáveis quantitativas. Essas possibilidades não esgotam os tipos de gráficos existentes, mas atendem a situações frequentes nos estudos quantitativos desenvolvidos na Ciência da Informação. O Quadro 6 sintetiza essas escolhas.

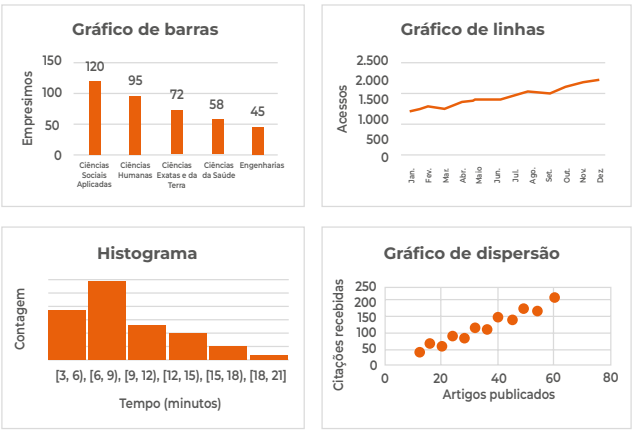
Quadro 6 - Escolha do gráfico conforme o objetivo da representação

Pergunta que orienta a representação	Tipo de variáveis indicadas	Gráfico indicado	Exemplo
Como as categorias se comparam?	qualitativas	Barras (simples ou duplas)	Empréstimos por área do conhecimento
Como os valores mudaram ao longo do tempo?	quantitativas	Linhas	Acessos mensais a um repositório
Como uma variável quantitativa está distribuída?	quantitativas contínuas	Histograma	Tempos de atendimento agrupados em intervalos
Existe relação entre duas variáveis quantitativas?	quantitativas	Dispersão	Número de publicações e quantidade de citações

Fonte: elaboração própria (2026).

Como mostra o Quadro 6, a escolha do gráfico depende principalmente do aspecto que se deseja observar. A Figura 11 apresenta exemplos construídos com dados fictícios da Ciência da Informação e permite reconhecer visualmente a estrutura e a finalidade de cada representação.

Figura 11 - Exemplos de gráficos aplicados à Ciência da Informação



Fonte: elaboração própria (2026).

No gráfico de barras, a altura de cada barra representa a quantidade de empréstimos de uma área do conhecimento. A disposição das categorias em uma linha de base comum permite perceber que as Ciências Sociais Aplicadas apresentam a maior quantidade de empréstimos, enquanto as Engenharias registram a menor. Esse tipo de gráfico é indicado para comparar categorias distintas. Quando não houver uma ordem própria entre elas, pode-se organizá-las por frequência para facilitar a leitura. O gráfico de barras pode ser de barras simples ou barras duplas e é recomendado para a representação de variáveis qualitativas.

O gráfico de linhas apresenta a quantidade mensal de acessos a um repositório institucional. Os meses aparecem em ordem cronológica, e a ligação entre os valores orienta a leitura da trajetória. Apesar de pequenas reduções em alguns períodos, observa-se crescimento dos acessos entre janeiro e dezembro. Esse gráfico é adequado para evidenciar mudanças, oscilações e tendências ao longo do tempo. É utilizado para representar variáveis quantitativas.

No histograma, os tempos de atendimento são agrupados em intervalos, e a altura de cada barra indica a quantidade de registros pertencentes a cada faixa. A concentração nos intervalos de menor duração e a diminuição das frequências nos intervalos superiores revelam uma distribuição assimétrica à direita. Diferentemente do gráfico de barras, as colunas do histograma permanecem unidas porque representam intervalos contínuos de uma variável quantitativa.

O gráfico de dispersão representa a relação entre o número de artigos publicados e a quantidade de citações recebidas. Seu uso é indicado para relacionar duas variáveis quantitativas, no exemplo temos o número de artigos publicados no eixo x e o número de citações recebidas no eixo y. Cada ponto corresponde a um periódico e é posicionado de acordo com seus valores nas duas variáveis. O padrão ascendente sugere que periódicos com mais artigos publicados tendem a

apresentar maior quantidade de citações. Essa associação, entretanto, não permite concluir que o aumento das publicações seja, isoladamente, a causa do crescimento das citações.

A leitura conjunta do quadro e da figura mostra que gráficos diferentes respondem a perguntas diferentes. O gráfico de barras compara categorias; o gráfico de linhas acompanha mudanças no tempo; o histograma mostra a distribuição dos valores; e o gráfico de dispersão permite examinar a relação entre variáveis quantitativas. Assim, deve-se escolher o gráfico pela informação que se deseja evidenciar, e não apenas por sua aparência visual.

3.5.3 Utilizando outras formas de representação

Além das tabelas, dos quadros e dos gráficos apresentados, os dados quantitativos podem ser comunicados por meio de outras representações. A escolha depende das características dos dados e da informação que se pretende destacar. Entre essas possibilidades, encontram-se o diagrama de caixa e os mapas temáticos.

O diagrama de caixa, também denominado *boxplot*, permite representar a posição e a dispersão dos valores de uma variável quantitativa, ou da relação entre uma variável quantitativa e uma qualitativa de agrupamento. Conforme apresentado no item 3.4, sua estrutura possibilita visualizar os quartis, a mediana, a extensão dos valores e a ocorrência de registros distantes do conjunto central. Entretanto, sua utilização não se limita à descrição de uma única distribuição, pois também permite comparar diferentes grupos.

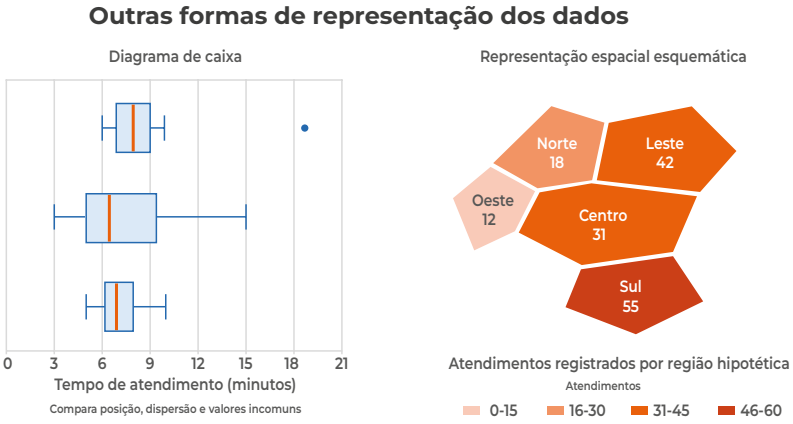
Em um estudo sobre o tempo de atendimento em unidades de informação, por exemplo, pode-se construir um diagrama de caixa para cada unidade. A comparação permite verificar se os tempos apresentam posições centrais semelhantes, se uma unidade possui maior variação e se existem atendimentos com duração incomum. Desse modo, vários aspectos das distribuições podem ser examinados em

uma mesma representação, sem depender apenas da comparação entre médias.

Os mapas, por sua vez, são adequados quando a localização geográfica constitui uma dimensão relevante para o estudo. Eles podem representar, por exemplo, a distribuição de bibliotecas entre municípios, a quantidade de arquivos públicos por unidade da Federação ou a proporção de instituições que oferecem determinado serviço. Nesses casos, cores, símbolos ou diferentes intensidades podem ser utilizados para indicar os valores observados em cada área.

A Figura 12 apresenta exemplos dessas duas formas de representação. No primeiro painel, os diagramas de caixa permitem comparar a posição, a dispersão e a ocorrência de valores incomuns nos tempos de atendimento de três unidades. No segundo, uma representação espacial esquemática, composta por regiões hipotéticas e sem correspondência com limites geográficos reais, utiliza diferentes intensidades de cor para representar a quantidade de atendimentos em cada região.

Figura 12 - Exemplos de outras formas de representação dos dados



Fonte: elaboração própria (2026).

Contudo, a elaboração de um mapa requer atenção à medida representada. A utilização de valores absolutos pode destacar regiões mais populosas ou com maior quantidade de instituições, sem indicar necessariamente maior presença proporcional do fenômeno. Dependendo do objetivo, pode ser mais adequado utilizar taxas, percentuais ou valores ajustados pela população de cada região. A legenda, a escala, as unidades territoriais e a fonte dos dados também precisam ser apresentadas de forma clara.

Portanto, o diagrama de caixa deve ser escolhido quando se pretende examinar ou comparar distribuições, enquanto os mapas são indicados quando se deseja observar a dimensão espacial dos dados. Assim como ocorre com as demais representações, ambos precisam estar acompanhados de uma análise escrita que destaque os padrões identificados e os relacione aos objetivos do estudo.

Em síntese, representar os dados consiste em selecionar a forma que melhor responde à pergunta da pesquisa e favorece a compreensão dos resultados. As tabelas permitem consultar valores exatos; os quadros organizam informações predominantemente textuais; os gráficos evidenciam comparações, tendências, distribuições e relações; os diagramas de caixa sintetizam a posição e a dispersão dos valores; e os mapas permitem observar sua dimensão espacial. Entretanto, nenhuma dessas representações substitui a análise escrita. Cabe ao pesquisador explicar os padrões identificados, relacioná-los aos objetivos do estudo e indicar os limites de sua interpretação.

4. ERROS FREQUENTES

Os problemas encontrados em uma pesquisa quantitativa nem sempre decorrem de cálculos incorretos. Muitas vezes, resultam da falta de correspondência entre o problema investigado, os dados coletados e os procedimentos empregados em seu tratamento, análise e apresentação. Assim, uma operação estatística pode estar matematicamente correta e, ainda assim, produzir uma conclusão inadequada quando aplicada a uma variável, amostra ou situação incompatível com seus pressupostos.

Alguns problemas podem ser corrigidos antes da análise, como registros duplicados, códigos inconsistentes e erros de digitação. Outros, contudo, constituem limitações do estudo, como a baixa participação, a ausência de informações ou as restrições da fonte consultada. Nesses casos, não se deve modificar os dados para aproximá-los do resultado esperado. É necessário reconhecer a limitação, avaliar suas consequências e registrá-la na apresentação da pesquisa.

Nesse sentido, os erros discutidos nesta seção não correspondem apenas a falhas operacionais. Eles podem surgir da ruptura da coerência entre as diferentes etapas da investigação e comprometer o alcance das conclusões, mesmo quando os procedimentos são executados corretamente.

4.1. Incoerência entre problema, objetivos e dados

Um erro frequente ocorre quando os dados coletados não permitem responder ao problema ou cumprir os objetivos da pesquisa. Isso pode acontecer quando as variáveis representam apenas parte do fenômeno, quando o instrumento não registra informações necessárias ou quando a unidade analisada difere daquela definida no planejamento. Nesses casos, aumentar a quantidade de dados ou aplicar técnicas estatísticas mais complexas não corrige a falta de correspondência entre o que se pretendia investigar e o que foi efetivamente observado.

Considere-se, por exemplo, uma pesquisa cujo objetivo seja avaliar se o depósito de artigos em um repositório institucional aumenta sua visibilidade científica. Se forem coletados apenas os números de acessos e downloads, será possível descrever o uso dos documentos, mas não concluir diretamente que houve aumento da visibilidade científica. Para sustentar essa conclusão, seria necessário definir como a visibilidade será representada e coletar dados compatíveis, como citações, menções ou outra medida previamente justificada. Portanto, a conclusão deve permanecer limitada ao aspecto efetivamente observado.

A incoerência também pode surgir durante a construção do instrumento. Se o objetivo for comparar a satisfação de estudantes e docentes com os serviços de uma biblioteca, será necessário registrar tanto a avaliação dos serviços quanto a categoria de cada participante. Caso o questionário pergunte apenas sobre a satisfação, a comparação prevista não poderá ser realizada. Por isso, antes da coleta, deve-se relacionar cada objetivo às variáveis e aos itens responsáveis pela produção das informações necessárias.

Essa correspondência pode ser verificada por meio de três perguntas:

1. Que resultado é necessário para responder ao problema?

2. Quais variáveis permitem produzir esse resultado?

3. Em qual pergunta, campo ou registro cada variável será obtida?

Se alguma variável necessária não estiver vinculada a uma fonte ou a um item do instrumento, existe uma lacuna no planejamento. Em contrapartida, se uma pergunta ou um campo não contribuir para nenhum objetivo, deve-se avaliar se sua coleta é realmente necessária. Tal verificação evita tanto a ausência de informações essenciais quanto a reunião de dados que não serão utilizados.

Quando a incoerência for identificada antes da coleta, podem-se revisar os objetivos, as variáveis ou o instrumento. Se for percebida posteriormente, não se deve ampliar artificialmente o alcance das conclusões. Nesse caso, é necessário reformular a pergunta que pode ser respondida pelos dados disponíveis e registrar a diferença entre o planejamento inicial e o que foi efetivamente realizado. Contudo, mesmo quando os dados são coerentes com os objetivos, as conclusões ainda dependem das unidades incluídas no estudo e da forma como foram selecionadas.

4.2. Generalizar além dos limites da amostra

Um resultado obtido em uma amostra não pode ser automaticamente atribuído a toda a população. O alcance da conclusão depende da forma de seleção das unidades, da participação obtida e da presença dos grupos relevantes para o estudo. Assim, uma amostra numerosa também pode apresentar limitações quando determinados segmentos estão ausentes ou pouco representados.

Em uma pesquisa sobre a satisfação com os serviços de uma biblioteca universitária, por exemplo, um questionário divulgado apenas nas redes sociais da instituição alcançará principalmente as pessoas que acompanham esses canais. Nesse caso, afirmar que “os usuários estão satisfeitos” ultrapassa o que os dados permitem demonstrar. Seria mais adequado informar que “os participantes da pesquisa declararam estar satisfeitos”, restringindo a afirmação ao grupo efetivamente observado.

Além disso, a quantidade total de participantes não revela, por si só, se os diferentes segmentos da população estão adequadamente representados. Uma pesquisa pode reunir muitas respostas e, ainda assim, refletir predominantemente as experiências de apenas uma categoria de usuários. Por isso, a composição da amostra deve ser considerada juntamente com seu tamanho e com o procedimento utilizado para selecionar as unidades.

Portanto, antes de formular uma conclusão, deve-se verificar a quem os resultados efetivamente se aplicam. Quando a amostra não permitir representar toda a população, podem-se utilizar expressões como “entre os participantes”, “no grupo analisado” ou “na amostra estudada”. As limitações da seleção e da participação devem ser consideradas na interpretação, sem invalidar as informações que o estudo conseguiu produzir.

4.3. Falhas na coleta dos dados

Alterações realizadas durante a coleta podem comprometer a comparabilidade dos dados quando afetam apenas parte das unidades pesquisadas. Isso ocorre, por exemplo, quando se oferecem explicações adicionais somente a alguns participantes, modificam-se perguntas após o início da aplicação, adotam-se períodos distintos de observação ou alteram-se termos de busca e filtros durante a extração de registros. Nesses casos, as diferenças encontradas podem decorrer do procedimento utilizado, e não apenas do fenômeno investigado.

Quando uma alteração for inevitável, devem ser registrados o momento em que ocorreu, as unidades afetadas e os motivos da decisão. Também é necessário avaliar se os dados obtidos antes e depois da mudança ainda podem ser analisados conjuntamente. Se a comparabilidade estiver comprometida, pode-se separar os conjuntos, repetir parte da coleta ou restringir as conclusões.

Assim, não se deve corrigir silenciosamente uma falha nem desconsiderar seus possíveis efeitos. A documentação da ocorrência permite avaliar em que medida a mudança interferiu nos registros e fornece os elementos necessários para justificar a decisão adotada. Os procedimentos destinados à prevenção e ao registro dessas situações são apresentados no item 3.2.5.

4.4. Utilizar tratamento ou análise incompatível com os dados

Mesmo quando a coleta foi realizada adequadamente, as decisões tomadas durante o tratamento e a análise podem alterar o significado dos registros ou produzir resultados inadequados. Isso ocorre, por exemplo, quando categorias distintas são reunidas sem justificativa, códigos são interpretados como valores quantitativos ou se aplica uma medida estatística incompatível com o tipo de variável.

Em uma pesquisa sobre o canal utilizado para solicitar atendimento, os códigos 1 = presencial, 2 = e-mail e 3 = aplicativo de mensagens servem apenas para identificar categorias. Calcular a média desses números não produz uma informação interpretável, pois a distância entre os códigos não representa uma diferença quantitativa entre os canais. Nesse caso, podem-se utilizar frequências e percentuais para demonstrar a participação de cada categoria.

Além disso, deve-se preservar o significado dos dados ausentes. Em um questionário sobre a frequência de uso de um repositório, uma resposta em branco não significa necessariamente que o participante nunca utilizou o serviço. A ausência pode decorrer de esquecimento, recusa, dificuldade de compreensão ou falha no preenchimento. Portanto, substituí-la automaticamente por zero reduziria artificialmente a frequência observada e modificaria os resultados.

Antes de realizar os cálculos, deve-se verificar o tipo de variável, o significado dos códigos e as decisões adotadas durante o tratamento. Quando houver agrupamentos, exclusões ou substituições, seus critérios precisam ser justificados e documentados. Dessa forma, o procedimento escolhido deve representar as características dos dados, sem criar relações ou diferenças que não estavam presentes nos registros originais.

4.5. Produzir representações ou interpretações equivocadas

Tabelas e gráficos podem apresentar valores corretos e, ainda assim, induzir a conclusões equivocadas. Isso ocorre quando a forma de representação altera visualmente as diferenças, omite informações necessárias à comparação ou destaca apenas parte dos resultados. O erro, portanto, não se encontra somente nos cálculos, mas também na maneira como os dados são apresentados e interpretados.

Em um gráfico sobre empréstimos realizados em duas bibliotecas, por exemplo, os valores de 950 e 1.000 empréstimos parecerão muito diferentes se o eixo vertical começar em 900. A redução da escala amplia visualmente uma diferença de apenas 50 empréstimos. Embora nem todo eixo precise começar em zero, sua escolha deve ser claramente indicada e não pode produzir uma impressão desproporcional à variação observada.

Por outro lado, comparações também podem ser equivocadas quando utilizam apenas números absolutos para grupos de tamanhos diferentes. Constatar que uma biblioteca realizou mais empréstimos do que outra não significa necessariamente que seus usuários utilizam mais intensamente o acervo. Para tal comparação, pode ser necessário relacionar a quantidade de empréstimos ao número de usuários, ao período observado ou ao tamanho da comunidade atendida. Assim, a base utilizada no cálculo deve acompanhar o resultado para que a diferença seja corretamente compreendida.

Outro erro consiste em transformar uma associação observada em explicação causal. Se unidades de informação com maior oferta de treinamentos apresentarem maior utilização das bases de dados, não se pode concluir, apenas por essa relação, que os treinamentos causaram o aumento. Características como o tamanho da instituição, o perfil dos usuários e a disponibilidade de recursos também podem

estar relacionadas aos dois resultados. Logo, a existência de associação não demonstra, por si só, uma relação de causa e efeito.

Antes de apresentar uma tabela, um gráfico ou uma conclusão, deve-se verificar se os elementos selecionados representam os dados sem distorção, se as comparações utilizam bases equivalentes e se a interpretação permanece dentro do que o estudo permite afirmar. Títulos, escalas, unidades de medida, fontes e períodos devem fornecer as informações necessárias para uma leitura autônoma. A representação deve facilitar a compreensão dos resultados, sem ampliar, ocultar ou explicar diferenças que os dados não sustentam.

Os erros apresentados ao longo desta seção podem ocorrer em diferentes etapas e produzir efeitos cumulativos. Dados adequados aos objetivos podem conduzir a conclusões frágeis quando a amostra não sustenta a generalização. Da mesma forma, uma coleta bem executada pode ser prejudicada por procedimentos incompatíveis de tratamento, análise ou representação.

Portanto, a qualidade de um estudo quantitativo não depende de uma decisão isolada, mas da coerência mantida entre o problema investigado, os dados produzidos, os procedimentos adotados e o alcance das conclusões. A revisão dessas relações ao longo da pesquisa permite identificar limitações, corrigir problemas quando possível e comunicar os resultados de maneira compatível com as evidências obtidas.

5. CONSIDERAÇÕES FINAIS

A Estatística constitui uma ferramenta metodológica relevante para a Ciência da Informação, especialmente nas pesquisas quantitativas e nos estudos mistos. Sua contribuição, no entanto, não se limita à realização de cálculos, visto que abrange um conjunto de decisões relacionadas ao planejamento do estudo, à produção dos dados e à interpretação dos resultados. Desse modo, a qualidade da pesquisa depende mais da coerência mantida entre o problema investigado, os objetivos, as variáveis, as unidades de análise e os procedimentos adotados do que da quantidade de técnicas utilizadas.

Ao longo deste guia, foram apresentados, inicialmente, os conceitos necessários à compreensão de um estudo quantitativo, como população, amostra, variável, erro amostral e formas de amostragem. Esses conceitos fundamentam as decisões práticas desenvolvidas posteriormente. Afinal, definir de quem ou de qual fonte os dados serão obtidos, o que será observado e até onde os resultados poderão ser generalizados exige que se compreendam as características e os limites do conjunto investigado.

Na dimensão prática, o desenvolvimento da pesquisa foi organizado como um processo sequencial. O planejamento orienta a coleta; a coleta produz os registros que serão tratados; o tratamento prepara a base para sua exploração; e o conhecimento das frequências, das medidas de posição, da dispersão e da distribuição permite escolher formas adequadas de representação. Portanto, uma decisão tomada em determinada etapa produz efeitos sobre as etapas seguintes. Um instrumento inadequado, uma amostra incompatível com os objetivos ou uma base tratada sem critérios explícitos, por exemplo, não podem ser compensados apenas pela aplicação posterior de técnicas estatísticas.

Nesse percurso, também se mostrou que os dados não falam por si mesmos. Tabelas, gráficos e medidas estatísticas selecionam e evidenciam determinados aspectos do fenômeno, mas precisam ser

interpretados em relação ao contexto em que os dados foram produzidos. Um valor incomum pode representar tanto um erro quanto uma ocorrência relevante; uma associação entre variáveis não comprova necessariamente uma relação causal; e uma diferença numérica pode não possuir o mesmo significado em grupos, períodos ou instituições distintas. Por isso, os procedimentos adotados, as transformações realizadas e as limitações encontradas devem ser documentados e considerados na apresentação dos resultados.

Na Ciência da Informação, essas orientações podem ser aplicadas a diferentes objetos, como usuários, acervos, serviços, documentos, sistemas de informação, produção científica e rastros digitais. Embora cada investigação apresente características próprias, permanece a necessidade de estabelecer correspondência entre aquilo que se pretende conhecer e os dados utilizados para representar o fenômeno. Nesse sentido, a Estatística não substitui o conhecimento teórico e contextual da área, mas oferece recursos para organizar evidências, estabelecer comparações e sustentar interpretações.

Por fim, este guia não pretendeu esgotar as técnicas estatísticas nem substituir obras especializadas. Buscou-se apresentar conceitos e procedimentos introdutórios de forma integrada e aplicável às situações encontradas por estudantes e pesquisadores da Ciência da Informação. Mais do que selecionar fórmulas, gráficos ou programas, realizar um estudo quantitativo requer compreender o que os dados representam, explicitar as decisões tomadas e reconhecer os limites das conclusões produzidas. É essa articulação entre rigor metodológico, clareza dos procedimentos e interpretação responsável que permite utilizar a Estatística de forma consistente na produção do conhecimento científico.

REFERÊNCIAS

AGUILLO, Isidro F. **Cibermetría**: introducción teórico-práctica a una disciplina emergente. [S.l.]: [S.n.], jan. 2007. Disponível em: https://digital.csic.es/handle/10261/4143. Acesso em: 28 set. 2026.

Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação - ANCIB. **Programas de Pós-Graduação em Ciência da Informação no Brasil**. 2026. Disponível em: https://ancib.org/programas-de-pos-graduacao-em-ciencia-da-informacao/. Acesso em: 23 set. 2026.

BARCELOS, Janinne; MARICATO, João de Melo. Altmetria: uma análise de seus termos, expressões, conceitos e definições. **Em Questão**, [*s.l.*], v. 29, 24 nov. 2023. DOI: https://doi.org/10.1590/1808-5245.29.129518. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/129518. Acesso em: 28 set. 2026.

BAYER, Arno; BITTENCOURT, Hélio; ROCHA, Josy; ECHEVESTE, Simone. A estatística e sua história. In: XII SIMPÓSIO SULBRASILEIRO DE ENSINO DE CIÊNCIAS. 2004. **Anais**. Universidade Luterana do Brasil (ULBRA), 2004. Disponível em: https://notasdeaula.wordpress.com/wp-content/uploads/2009/08/estatistica-e-sua-historia.pdf. Acesso em: 28 set. 2026.

BJÖRNEBORN, Lennart. **Small-world link structures across an academic web space**: a library and information science approach. 2004. 399 f. Tese (Doutorado) – Royal School of Library and Information Science, Dinamarca, 2004.

BRASIL. [Lei Geral de Proteção de Dados Pessoais (LGPD)] **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília: Presidência da República. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 28 set. 2026.

BRASIL. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. **Diário Oficial da União**: seção 1, Brasília, DF, ano 153, n. 98, p. 44–46, 24 maio 2016. Disponível em: https://www.in.gov.br/web/dou/-/resolucao-n-510-de-7-de-abril-de-2016-22917558. Acesso em: 28 set. 2026.

DESSAUNE FILHO, Nilton. **Noções básicas de estatística experimental**. Vitória, ES: EMCAP, 1987. 59 p. (EMCAPA - Documentos, 32). Disponível em: https://biblioteca.incaper.es.gov.br/digital/handle/item/1576. Acesso em: 28 set. 2026.

FOWLER JR., Floyd J. **Survey research methods**. 5. ed. Los Angeles: SAGE, 2014. 171 p.

GLAZIER, Jack D. Structured observation: how it works. **College & Research Libraries News**, [s.l.], v. 46, p. 105–108, 1985. Disponível em: https://api.semanticscholar.org/CorpusID:113675986. Acesso em: 28 set. 2026.

GOUVEIA, Fábio Castro; ARAÚJO, Ronaldo Ferreira de. Webmetria: origens e usos contemporâneos. In: GRÁCIO, Maria Cláudia Cabrini; MARTÍNEZ-ÁVILA, Daniel; OLIVEIRA, Ely Francina Tannuri de; ROSAS, Fábio Sampaio. **Tópicos da bibliometria para bibliotecas universitárias**. Marília: Oficina Universitária, 2020. p. 208–228. DOI: https://doi.org/10.36311/2020.978-65-86546-91-0.p208-228. Disponível em: https://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/book/187. Acesso em: 28 set. 2026.

MACIAS-CHAPULA, Cesar A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, 1998. DOI: https://doi.org/10.18225/ci.inf.v27i2.794. Disponível em: https://revista.ibict.br/ciinf/article/view/794. Acesso em: 28 set. 2026.

MEMÓRIA, José Maria Pompeu. **Breve história da estatística**. Brasília: Embrapa Informação Tecnológica, 2004. 111 p. (Texto para Discussão, 21). Disponível em: https://www.ime.usp.br/~rvicente/JMPMemoria_Historia_Estatistica.pdf. Acesso em: 28 set. 2026.

MOMESSO, Ana Carolina; NORONHA, Daisy Pires. Bibliométrie ou Bibliometrics: o que há por trás de um termo? **Perspectivas em Ciência da Informação**, [s.l.], v. 22, n. 2, p. 118–124, 29 jun. 2017. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/22506. Acesso em: 25 set. 2026.

NARIN, Francis. Patent bibliometrics. **Scientometrics**, [s.l.], v. 30, n. 1, p. 147–155, 1 maio 1994. DOI: https://doi.org/10.1007/BF02017219. Disponível em: https://doi.org/10.1007/BF02017219. Acesso em: 28 set. 2026.

NEUENDORF, Kimberly A. **The content analysis guidebook**. 2. ed. Los Angeles: SAGE, 2017. 438 p.

NOGUEIRA, Clélia Maria Ignatius. Pesquisas atuais sobre a construção do conceito de número: para além de Piaget? **Educar em Revista**, [s.l.], p. 109–124, 18 ago. 2011. Disponível em: https://revistas.ufpr.br/educar/article/view/22611. Acesso em: 25 set. 2026.

OTLET, Paul. **Traité de Documentation**: le livre sur le livre. Bruxelles: Mundaneum, 1934. 452 p. Disponível em: https://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf. Acesso em: 14 abr. 2021.

PINTO, Adilson Luiz. Arquivometria. **ÁGORA: Arquivologia em debate**, [*s.l.*], v. 21, n. 42, p. 59–69, 26 set. 2011. Disponível em: https://agora.emnuvens.com.br/ra/article/view/263. Acesso em: 28 set. 2026.

PRITCHARD, Allan. Statistical bibliography or bibliometrics. **Journal of Documentation**, [*s.l.*], v. 25, n. 4, p. 348–349, 1969. Disponível em: https://api.semanticscholar.org/CorpusID:62432515. Acesso em: 28 set. 2026.

SANTOS, Carla Maria Lopes da Silva Afonso dos. **Estatística descritiva: manual de auto-aprendizagem**. 3. ed. Lisboa: Edições Sílabo, 2018. 297 p.

SILVA, José Aparecido da; BIANCHI, Maria de Lourdes Pires. Cientometria: a métrica da ciência. **Paidéia (Ribeirão Preto)**, Ribeirão Preto, v. 11, n. 21, p. 5–10, 1 jan. 2001. DOI: https://doi.org/10.1590/S0103-863X2001000200002. Disponível em: https://revistas.usp.br/paideia/article/view/46613. Acesso em: 28 set. 2026.

SPEZIALI, Marcelo; NASCIMENTO, Raphael. Patentometria: uma ferramenta indispensável no estudo de desenvolvimento de tecnologias para a indústria química. **Química Nova**, [s.l.], v. 43, n. 10, p. 1538–1548, 2020. DOI: https://doi.org/10.21577/0100-4042.20170620. Disponível em: http://quimicanova.sbq.org.br/audiencia_pdf.asp?aid2=9173&nomeArquivo=AG2020-0224.pdf. Acesso em: 28 set. 2026.

STERRANTINO, Anna Freni. Observational studies: practical tips for avoiding common statistical pitfalls. **The Lancet Regional Health – Southeast Asia**, [s.l.], v. 25, 1 jun. 2024. DOI: https://doi.org/10.1016/j.lansea.2024.100415. Disponível em: https://www.thelancet.com/journals/lansea/article/PIIS2772-3682(24)00065-9/fulltext. Acesso em: 25 set. 2026.

VAN DEN BROECK, Jan; ARGESEANU CUNNINGHAM, Solveig; EECKELS, Roger; HERBST, Kobus. Data cleaning: detecting, diagnosing, and editing data abnormalities. **PLoS Medicine**, [s.l.], v. 2, n. 10, p. e267, 6 set. 2005. DOI: https://doi.org/10.1371/journal.pmed.0020267. Disponível em: https://dx.plos.org/10.1371/journal.pmed.0020267. Acesso em: 28 set. 2026.

WICKHAM, Hadley. Tidy Data. **Journal of Statistical Software**, [s.l.], v. 59, n. 10, 2014. DOI: https://doi.org/10.18637/jss.v059.i10. Disponível em: http://www.jstatsoft.org/v59/i10/. Acesso em: 28 set. 2026.

SOBRE OS AUTORES



Bianca Savegnago de Mira

Lattes: http://lattes.cnpq.br/7638396535395640

Orcid: https://orcid.org/0000-0001-7913-4084

Mestre e Doutora em Ciência da Informação pela Universidade Estadual Paulista. Especialista em Formação Didático-Pedagógica para Cursos na Modalidade à Distância pela Universidade Virtual do Estado de São Paulo. Bacharel em Ciências Econômicas pela Universidade Estadual Paulista. Atua como docente na Universidade de Marília para os cursos do Departamento de Tecnologia e como membro do Grupo de Pesquisa de Estudos Métricos em Informação da Universidade Estadual Paulista.



Marcelo Schiessl

Lattes: https://lattes.cnpq.br/4402964117780625

Orcid: https://orcid.org/0000-0002-3895-5844

Doutor em Ciência da Informação pela Universidade de Brasília, com estágio de doutoramento na Universidade de Koblenz-Landau, Koblenz, Alemanha. Mestre em Ciência da Informação pela Universidade de Brasília e especialista em Inteligência Organizacional e Competitiva na Sociedade da Informação pela Universidade de Brasília e Bacharel em Estatística pela Universidade de Brasília. Atua em pesquisas sobre descoberta de Conhecimento em Texto/Mineração de Textos/Dados e Processamento de Linguagem Natural com foco na lexicalização de ontologias.



Milton Shintaku

Lattes: http://lattes.cnpq.br/8605833104600600

Orcid: https://orcid.org/0000-0002-6476-4953

Mestre e Doutor em Ciência da Informação pela Universidade de Brasília (UnB). Coordenador de Tecnologia para Informação (Cotec) do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), credenciado no Programa de Pós-Graduação em Gestão da Informação (PPGGI) da Universidade Federal do Paraná (UFPR) e no Programa de Pós-Graduação em Ciência da Informação (PPGCI) do Ibict.



Editora
Ibict

Brasília
2026



A obra Pesquisa Quantitativa na Ciência da Informação, estruturada em cinco capítulos, reúne conceitos teóricos basilares para a compreensão da pesquisa e de seus procedimentos, com ênfase em dados quantitativos. Parte de uma problemática concreta sobre o planejamento e a execução de pesquisas, utilizando fundamentos da estatística descritiva, além de ilustrá-los com exemplos de peculiaridades metodológicas frequentes e considerações sobre a aplicação desses métodos. Ao refletir sobre objetos e situações comuns, adota distanciamento suficiente para desnaturalizar ou desconstruir noções estereotipadas presentes no cotidiano do pesquisador, discutindo procedimentos de coleta, organização, tratamento, análise e representação de dados em tabelas, gráficos e medidas estatísticas. Observa como a qualidade dos resultados depende tanto da vigilância em relação aos cálculos quanto da coerência entre problema, objetivos e procedimentos metodológicos. Assim, serve de guia relevante ao desenvolvimento de pesquisas, contribuindo para o fortalecimento da produção científica em Ciência da Informação. Com linguagem clara e didática, oferece subsídios a pesquisadores de diferentes níveis de familiaridade com os Estudos Métricos da Informação. O conteúdo apresentado enfatiza a importância do planejamento metodológico, da delimitação de problemas e objetivos, da definição de populações e amostras e da escolha adequada de variáveis, métodos e instrumentos de coleta de dados. Por óbvio, o texto não substitui obras especializadas em Estatística; entretanto, seu potencial reside justamente em privilegiar o espaço de intersecção entre a Estatística e a Ciência da Informação.

Leilah Santiago Bufrem

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

ibict Instituto Brasileiro de Informação em Ciência e Tecnologia

Editora Ibict

978-65-82912-62-7

9 78658 2912627