

ANÁLISE DE CONTEÚDO CATEGORIAL: MANUAL DE APLICAÇÃO

Rafael Cardoso Sampaio e
Diógenes Lycarião



COLEÇÃO

Metodologias
de pesquisa

Enap Escola Nacional de Administração Pública

Presidente

Diogo Godinho Ramos Costa

Diretora de Altos Estudos

Diana Magalhães de Souza Coutinho

Diretor de Desenvolvimento Profissional

Paulo Marques

Diretora de Inovação

Bruna Silva dos Santos

Diretor de Educação Executiva

Rodrigo Torres de Araujo Lima

Diretora de Gestão Interna

Alana Regina Biagi Silva Lisboa

Revisão

Roberto Araújo e Adriana Braga

Análise de conteúdo categorial: manual de aplicação

*Autores: Rafael Cardoso Sampaio e
Diógenes Lycarião*

Brasília
Enap
2021

- S1921a Sampaio, Rafael Cardoso
Análise de conteúdo categorial: manual de aplicação / Rafael
Cardoso Sampaio, Diógenes Lycarião. -- Brasília: Enap, 2021.
155 p. : il. -- (Coleção Metodologias de Pesquisa)
- Inclui bibliografia
ISBN: 978-65-87791-18-0
1. Pesquisa Científica. 2. Metodologia Científica. 3. Análise
de Conteúdo. 4. Manual. 5. Pesquisa Quantitativa. I. Título. II.
Lycarião, Diógenes.

CDU 001.8:311.2

Bibliotecária: Tatiane de Oliveira Dias – CRB1/2230



Enap, 2021

Este trabalho está sob a Licença Creative Commons – Atribuição: Não Comercial –
Compartilha Igual 4.0 Internacional

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade
do(s) autor(es), não exprimindo, necessariamente, o ponto de vista da Escola
Nacional de Administração Pública (Enap). É permitida a reprodução deste texto e
dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais
são proibidas.

Escola Nacional de Administração Pública (Enap)
Diretoria de Altos Estudos
Coordenação-Geral de Pós-Graduação Stricto Sensu
SAIS – Área 2-A – 70610-900 — Brasília-DF, Brasil

SUMÁRIO

INTRODUÇÃO	6
CAPÍTULO 1: DEFINIÇÃO, APLICAÇÕES E EPISTEMOLOGIA DA ANÁLISE DE CONTEÚDO (AC).....	14
1.1 Definição	14
1.2 Histórico.....	17
1.3 Aplicações da AC	21
1.4 Aplicações em políticas públicas.....	26
1.5 Epistemologia da AC	29
1.6 Fronteiras.....	40
CAPÍTULO 2: MANUAL DA AC.....	45
1. Códigos, codificação e categorias	45
2. Desenho da análise de conteúdo	46
ETAPAS DA ANÁLISE DE CONTEÚDO	49
1. Identificar o problema (revisão de literatura)	49
2. Questões de pesquisa e hipóteses	50
3. Selecionar a(s) unidade(s) e subunidade(s) de análise	51
4. Criar e definir categorias.....	57
5. Amostragem.....	68
5.1 Técnicas probabilísticas (randômicas)	72
5.2 Técnicas não probabilísticas (não randômicas)	75
5.3 Tamanho da amostra	78
6. Pré-teste das categorias e das regras de codificação.....	79
7. Treinamento final e teste de confiabilidade	87
8. Codificação	103
9. Teste de confiabilidade intermediário e final	104
10. Tabulação e aplicação de procedimentos estatísticos.....	107
11. Interpretar e reportar os resultados.....	107
12. Validação e replicabilidade	110

CAPÍTULO 3: ANÁLISE DE RESULTADOS E TESTES ESTATÍSTICOS	113
3.1 Definição de variável e seus tipos.....	113
3.2 Os tipos de variáveis pelo modelo estatístico	118
3.3 Análise descritiva	119
3.4 Estatística inferencial	123
3.5 Testes estatísticos	130
CONCLUSÃO	141
REFERÊNCIAS BIBLIOGRÁFICAS	144

INTRODUÇÃO

Este livro pretende apresentar, de modo simplificado, como realizar uma análise de conteúdo categorial quantitativa¹ com bases científicas. É nosso objetivo que esta obra possa se tornar um manual útil para diferentes pesquisadores que almejam realizar uma aplicação adequada da técnica da análise de conteúdo (AC). Como buscaremos defender, compreendemos que: **análise de conteúdo é uma técnica de pesquisa científica baseada em procedimentos sistemáticos, intersubjetivamente validados e públicos para criar inferências válidas sobre determinados conteúdos verbais, visuais ou escritos, buscando descrever, quantificar ou interpretar certo fenômeno em termos de seus significados, intenções, consequências ou contextos.**

Datam de séculos relatos sobre técnicas de análise de textos, que foram realizadas sob o interesse de se catalogar e classificar textos e materiais sob os mais distintos propósitos, como é o caso de textos religiosos (BARDIN, [1977] 2016; KRIPPENDORFF, 2004; NEUENDORF, 2002; RIFFE; LACY; FICO, 2014). Entretanto, a análise de conteúdo (AC), sob fins científicos, surge para suprir uma necessidade de mensuração dos padrões das mensagens mediáticas, especialmente nos períodos das Grandes Guerras do século 20, como foi o caso, a título de exemplo, do estudo de Harold Lasswell sobre as propagandas de guerra, por meio de análise de conteúdo (LASSWELL, 1927).

Já ao longo da segunda guerra mundial, o uso de mídias massivas (e.g. rádio, cinema) pelos Estados Unidos, pela Alemanha nazista e por outros regimes totalitários emergiu como um fenômeno sobressalente, o que despertou interesse do governo estadunidense em verificar como seus adversários faziam uso dessas mensagens. Harold Lasswell era o coordenador da Divisão Experimental para o Estudo de Comunicações em Tempos de Guerra, criada pelo Congresso americano para tal fim. Para além da avaliação das mensagens mediáticas adversárias, preocupava-o os possíveis efeitos das

¹ Apesar de não ser o objetivo inicial, é bastante provável que boa parte dos passos descritos no livro possa ser útil para análises de conteúdo qualitativas. Em AC qualitativas rigorosas, conforme apresentado por Saldaña (2012) e Mayring (2014), vários passos similares são apresentados, incluindo a apresentação da frequência de códigos.

mensagens dos meios de comunicação de massa nas pessoas. Para analisar essas mensagens, a técnica da análise de conteúdo quantitativa foi, então, largamente utilizada e aperfeiçoada.

Não obstante seu surgimento no campo da comunicação política, a análise de conteúdo² tornou-se uma técnica bastante difundida em toda a área de humanidades, como, por exemplo, nas Ciências Sociais (POHLMANN; BÄR; VALARINI, 2014; TRIVIÑOS, 1987), Ciência da Informação (LIMA; MANINI, 2017; LIMA; MORAES, 2017), Contabilidade (ALVES, 2011), Geografia (PAULA, 2015), História (CONSTANTINO, 2002), Psicologia (GONDIM; BENDASSOLLI, 2014), Serviço Social (LARA, 2011), Turismo (THOMAZ *et al.* 2016), além, claro, de seu vasto uso nas áreas de Administração (BASTOS; OLIVEIRA, 2015; FREITAS, 2011; GOMES *et al.*, 2020; MOZZATO; GRZYBOVSKI; 2011; VERGARA, 2011), Comunicação (HERSCOVITZ, 2007; JORGE, 2015; MARTINEZ; PESSONI, 2015; QUADROS; ASSMANN; LOPEZ, 2014; VIMIEIRO; MAIA, 2011), Educação (FRANCO, 2005; MORO, 1989; OLIVEIRA *et al.* 2003) e de Ciência Política (ALVES; FIGUEIREDO FILHO; HENRIQUE, 2015; CARLOMAGNO; ROCHA, 2016; FERES JR., 2016; FIGUEIREDO *et al.*, 1997; PANKE; CERVI, 2012). Também é visível sua utilização nas ciências da saúde, como Enfermagem e Saúde Coletiva (BELLUCCI JÚNIOR; MATSUDA, 2012; CAMPOS, 2004; MINAYO, 2014; RODRIGUES; LEOPARDI, 1999; ROMEU, 2009; TAQUETTE; MINAYO, 2015).

Diante desse cenário, é seguro afirmar que a análise de conteúdo tem grande capilaridade na ciência brasileira, tendo sido aplicada em um considerável número de estudos em diferentes áreas (Sampaio *et al.*, 2021). Essa aplicação também se mostra acompanhada por debates sobre as benesses e os limites da técnica (CAVALCANTE; CALIXTO; PINHEIRO, 2014; FREITAS, 2011; MORO, 1989; MOZZATO; GRZYBOVSKI, 2011; GONDIM; BENDASSOLLI, 2014; VERGARA, 2011), sobre sua diferenciação e separação das linhas de análise do discurso (CAPPELLE *et al.*, 2003; CAREGNATO; MUTTI, 2006; LIMA, 2003; LIMA; MORAES, 2017; PÁDUA, 2002; ROCHA; DEUSDARÁ, 2005,

² Aqui, estamos nos referindo aqui à análise de conteúdo “clássica” (BAUER, 2007) ou, ainda, humana, baseada na codificação manual por pessoas; portanto, boa parte de nossas reflexões e do passo a passo deste manual não se aplica à AC computadorizada.

2006) e, recentemente, sobre o uso de *softwares* para sua realização (e.g. ALVES; FIGUEIREDO FILHO; HENRIQUE, 2015; LIMA; MANINI, 2017).

Um dos principais motivos que nos incentivou a escrever este livro está justamente na falta de opções de outros manuais no mercado editorial brasileiro. Isso se deveu, em grande medida, pela qualidade do manual de Laurence Bardin, que teve a sua tradução lançada originalmente no Brasil em 1977. Entretanto, enquanto vários outros manuais de alta qualidade foram lançados pelo mundo, após 1977, nenhum recebeu uma tradução para o português brasileiro (e.g. ALONSO *et al.*, 2012; DRISKO; MARCHI, 2016; KRIPPENDORFF [1980] 2004; NEUENDORF, 2002; RIFFE; LACY; FICO, 2014; WEBER, 1990, entre outros). Apesar de a técnica ter sido resumida e explicada em capítulos de coletâneas acadêmicas (BAUER, 2007; FONSECA JÚNIOR, 2005; HERSCOVITZ, 2007; Romeu, 2009), em trechos de manuais de análise qualitativa (FLICK, 2009; MINAYO, 2014; YIN, 2016) ou científica em geral (GIL, 1989; TRIVIÑOS, 1987; VERGARA, 2005), e mesmo em diversos artigos (ALVES, 2011; BELLUCCI JÚNIOR, MATSUDA, 2012; CAMPOS, 2004; CARLOMAGNO; ROCHA, 2016; CAVALCANTE; CALIXTO; PINHEIRO, 2014; GONDIM; BENDASSOLLI, 2014; MORAES, 1999; MOZZATO; GRZYBOVSKI, 2011; OLIVEIRA, 2008; OLIVEIRA; ENS; ANDRADE; MUSSIS, 2003; ROCHA; DEUSDARÁ, 2006; TAQUETTE; MINAYO, 2015; VERGARA, 2011), parecia-nos que faltava um manual atualizado específico exclusivamente dedicado à técnica.

Em nossa revisão do estado da arte da técnica no Brasil (SAMPAIO; LYCARIÃO, 2018; SAMPAIO *et al.*, 2021)³, encontramos poucas obras que buscam ofertar explicações mais detalhadas sobre a aplicação da análise de conteúdo,

³ Neste livro, fazemos referências a várias pesquisas próprias baseadas em análise de conteúdo. Caso haja interesse em receber os livros de códigos, banco de dados e artigos completos, enviar e-mail para os autores: Rafael Cardoso Sampaio (rafael.sampaio@ufpr.br) e Diógenes Lycarião (lycario-d@ufc.br).

para além da tradução do manual de Bardin⁴. Os dois primeiros também são traduções de obras. O primeiro, interessantemente, foi lançado no Brasil antes da primeira tradução do livro de Bardin como *Comunicação de massa: análise de conteúdo*, de Albert Kientz, pela editora Eldorado, em 1973. Trata-se de uma interessante apresentação da técnica, seguindo seu foco inicial na análise de mídias massivas; porém, acabou se tornando um livro com pouca repercussão na ciência brasileira, não tendo novas reedições e se encontrando esgotado. O segundo é a tradução do livro clássico de Harold Lasswell, *A linguagem da política*, que apresenta os primeiros achados da pesquisa de Lasswell e colegas. O livro, original de 1949, ganhou essa tradução no Brasil em 1982, pela Editora Universidade de Brasília. Não obstante, por se tratar dos resultados e das discussões originais do surgimento da técnica, ele tem mais importância por seu caráter histórico do que, de fato, enquanto um manual de aplicação da técnica.

A terceira e quarta obra não são exclusivamente dedicadas à análise de conteúdo, mas acabam sendo, praticamente, as principais referências da área de saúde como um todo, após Bardin, que são dois livros da professora Maria Cecília Minayo da Fiocruz. Na coletânea *Pesquisa Social: teoria, método e criatividade*, Minayo, em conjunto com Suely Ferreira Deslandes e Romeu Gomes, discutem diversos aspectos da pesquisa qualitativa (MINAYO, 2009); nessa coletânea, há um capítulo especificamente sobre a análise de dados qualitativos, que trata de análise de conteúdo (GOMES, 2009). Seja Gomes, seja Minayo, o livro é bastante citado em saúde pública e em enfermagem. Por

⁴ Merecem destaque algumas coletâneas recentes na área. A primeira é organizada por Thais de Mendonça Jorge, *Notícia em Fragmentos: análise de conteúdo no jornalismo*, na qual oito capítulos fazem análises de fenômenos midiáticos pelo uso da AC, enquanto outros oito fazem discussões metodológicas sobre a aplicação da AC em diferentes contextos e fenômenos (JORGE, 2015). Nenhum capítulo, entretanto, busca apresentar um manual passo a passo da aplicação da técnica. Por sua vez, a coletânea *Internet e eleições no Brasil*, organizada por Emerson Cervi, Michele Massuchin e Fernanda Carvalho, apresenta uma miríade de aplicações da técnica em objetos relacionados a ambientes digitais e campanhas eleitorais (CERVI; MASSUCHIN; CARVALHO, 2016). O capítulo inicial de Cervi (2016) apresenta uma boa reflexão sobre a técnica e alguns indicativos de conceitos-chave para sua aplicação. Finalmente, também temos a coletânea *Análise de Conteúdo em pesquisas qualitativas na área de Educação Matemática* de Márcio Urel Rodrigues (RODRIGUES, 2019), que apresenta 14 capítulos exclusivamente baseados em AC para compreender a supracitada área. Os quatro capítulos iniciais são metodológicos e buscam dar uma contribuição ao campo. Coletâneas similares certamente existem.

sua vez, o livro *O Desafio do conhecimento: pesquisa qualitativa em saúde*, de Minayo, é a maior referência em AC em saúde após a própria Bardin. Trata-se de um manual geral sobre a pesquisa qualitativa, mas a análise de conteúdo é considerada uma das formas de se analisar os dados gerados pela pesquisa qualitativa (MINAYO, 2014).

Também da área de saúde, temos o manual brasileiro e exclusivo sobre a técnica *O método de análise de conteúdo: uma versão para enfermeiros*, das professoras Maria Socorro Rodrigues e Maria Tereza Leopardi. Esse manual é uma interessante revisão sobre a técnica, apresentando a lógica da AC de Bardin, Minayo, Kientz e Spradley, fazendo uso de vários autores internacionais especializados sobre o tema e chegando a entrar em questões importantes, como o livro de códigos (RODRIGUES, LEOPARDI, 1999). Surpreendentemente, não se trata de livro muito citado na área.

Outra obra produzida originalmente por uma pesquisadora brasileira que pode ser classificada como um manual específico de aplicação da técnica é o livro *Análise de conteúdo*, de Maria Laura P.B. Franco (2005), dedicado à área de Educação. Trata-se de um manual resumido, ao exemplo deste, para substituir o de Bardin para a aplicação da técnica. Entretanto, em nossa análise exploratória inicial, notamos que, assim como ocorria no livro traduzido de Bardin, o teste de confiabilidade entre codificadores e outras etapas da análise de conteúdo científica estavam ausentes em todos os manuais supracitados. Ademais, todos parecem, em nossa visão, ser mais adequados para análises de conteúdo de cunho qualitativo do que, de fato, para análise de conteúdo categorial quantitativa que aqui desejamos apresentar.

Nesta mesma seara, alguns estudos já indicaram que a técnica pode estar sendo aplicada de forma incompleta ou mesmo incorreta. A pesquisa de Castro, Abs e Sierra (2011) observou 83 artigos de 6 periódicos de Psicologia indexados na base SciELO; segundo os autores, via de regra, os trabalhos não oferecem descrição de nenhuma das operações de AC, tais como critérios de organização, indicadores e procedimentos de inferência. Problemas dessa natureza também foram encontrados por Deslandes e Iriart (2012), que analisaram 113 artigos da área de Ciências Sociais e Humanas em Saúde (entre os anos 2009-2010) usando a “análise de conteúdo”, sobretudo

a análise temática (porém, não desenvolveram, de forma adequada, as categorias de análise).

Por sua vez, Quadros, Assmannz e Lopez (2014, p. 105) chegam à conclusão similar, ao analisarem 57 trabalhos de conclusão de curso de pós-graduação em Comunicação, de que, “apesar de a maior parte das pesquisas basearem-se em Bardin, muitos trabalhos não seguem sistematicamente o método proposto pela autora. Foram poucos os trabalhos que indicaram, por exemplo, as unidades de registro ou contexto utilizadas”.

Já o estudo de Silva e colaboradores (2017, p. 178) realizou uma análise de 31 artigos que fizeram uso de AC na pesquisa em Administração e concluiu que a maioria não esclarece detalhes de sua operacionalização, como a construção das categorias, a forma como se deu a análise e afins e “parecem não [a] ter aplicado na íntegra”. Como já dito, o levantamento realizado por Sampaio e Lycarião (2018) indicou a pouca importância dada à replicabilidade e aos testes de confiabilidade entre codificadores na pesquisa brasileira. É notável que os manuais então disponíveis sobre a técnica não ensinam os procedimentos para realizá-las.

Finalmente, em 40 artigos analisados, Gomes *et al* (2020), numa pesquisa bibliográfica que compreendeu o período de 2016 e 2017 nas bases da SciELO e dos Anais da Associação Nacional dos Cursos de Pós-Graduação em Administração (ANPAD), indicaram que os trabalhos analisados não seguiram adequadamente os preceitos de Bardin em termos do rigor da objetividade e de melhor interpretação dos dados.

Em suma, são indicativos de que há espaços para incrementos na aplicação da AC na pesquisa brasileira. Portanto, mais do que substituir todo esse material já existente, nosso objetivo principal é ofertar uma nova opção complementar⁵. Assim, desejamos especialmente contribuir para estudantes de pós-graduação e graduação sem acesso aos manuais em outras línguas. Algumas pesquisas, inclusive, são indicativas de que o manual de Bardin

⁵ Não é nosso objetivo desconsiderar todos os pontos fortes da obra de Bardin, que é amplamente utilizada em vários pontos deste livro. Não obstante, acreditamos que seu manual apresente uma AC excessivamente simplificada na apresentação de suas fases e com poucas ferramentas para verificar o rigor científico.

é consideravelmente mais utilizado do que outros textos disponíveis em português, ou mesmo do que manuais disponíveis em outras línguas (GOMES *et al.*, 2020; MARTINEZ; PESSONI, 2015; QUADROS; ASSMANN; LOPEZ 2014; SILVA *et al.*, 2017). Sampaio *et al.* (2021), ao revisarem 3.484 referências de AC publicadas na SciELO (2002-2019), alertam que há inclusive um uso categórico de diferentes versões do manual de Bardin tanto no colégio de ciências da vida quanto em humanidades, que, na visão dos autores, denota a necessidade de materiais mais atualizados e que reflitam o debate epistemológico e metodológico dos últimos anos.

É importante também elucidar que este manual é escrito em um momento em que a “crista da onda” da ciência está na utilização e avaliação de *Big Data*, notadamente de grandes quantidades de conteúdos gerados em redes sociais digitais. Portanto, uma parte considerável da pesquisa está justamente preocupada em análises de conteúdo automatizadas (*e.g.* BURSCHER; Vliegenthart; DE VREESE, 2015; HOPKINS; KING, 2010), algo que não estará retratado neste livro. Em outras palavras, é factível questionar a importância da análise de conteúdo humana (ou manual) em um mundo que caminha fortemente para análises automatizadas com base em *machine learning* e afins. Primeiramente, é preciso esclarecer que, mesmo com os avanços alcançados recentemente nessas análises automatizadas, ainda há uma série de restrições a serem consideradas sobre a validade de seus resultados (CARLSON; MONTGOMERY, 2017; GUO *et al.*, 2016).

Em segundo lugar, devemos compreender que boa parte das análises computadorizadas é ou precedida por codificações humanas, ou validada em diversos momentos por codificações humanas (MUDDIMAN; MCGREGOR; STROUD, 2019). Análises de conteúdo manuais são frequentemente a base de análises automatizadas, que buscam aprender (novamente, o *machine learning*) com base na codificação humana (*e.g.* BURSCHER; Vliegenthart; DE VREESE, 2015). Em terceiro lugar, conseqüentemente, a AC manual e a automática podem ser combinadas para se atingir resultados mais válidos e profícuos (ver GANDIN, 2018; CAMARGO; JUSTO, 2013), assim como a AC manual pode ser combinada a outras técnicas, como é o caso de Lycarião e Dos Santos (2017), que combinam AC com análise de redes, e de Panke

e Cervi (2012), que sugerem a sua aplicação em conjunto com a análise de discurso para melhor compreensão de discursos políticos.

Finalmente, no atual estágio da pesquisa e do emprego de Big Data, existem diversos questionamentos sobre o uso de Big Data e sobre a falsa impressão de que eles geram, necessariamente, resultados melhores que as pesquisas de baixo n . Alguns autores têm justamente debatido como o Big Data e os algoritmos de plataformas podem (e, em alguns casos, tendem a) distorcer o debate público (e.g. O'NEIL, 2016). Ao menos no atual momento da tecnologia e da ciência, é inócuo acreditar que toda a AC manual poderá ser substituída por formas computadorizadas de análise. Em verdade, acreditamos que, não importando a evolução de tais técnicas, sempre haverá espaço para análises mais minuciosas e humanas do conteúdo que geramos.

Assim, este livro está organizado da seguinte maneira. No primeiro capítulo, apresentamos diversas definições da análise de conteúdo e oferecemos uma definição própria, tratamos sobre as diferentes aplicações da AC em diversos campos, mais notadamente na área de políticas públicas, e concluímos com uma breve discussão epistemológica sobre a técnica, incluindo um debate sobre algumas fronteiras teóricas e epistemológicas com outras técnicas de pesquisa, como a análise de discurso e a análise léxica.

O segundo capítulo se preocupa com uma apresentação completa da aplicação da análise de conteúdo categorial quantitativa, buscando percorrer um passo a passo completo para o desenho de uma AC baseada em alto rigor científico. O capítulo se preocupa com a identificação dos problemas, a elaboração de perguntas e hipóteses de pesquisa, a criação de categorias, as formas de amostragem do conteúdo, o treinamento de codificadores, o teste de confiabilidade entre codificadores e a codificação final.

Por fim, o terceiro capítulo se ocupará da análise dos resultados da codificação da AC, introduzindo as principais técnicas estatísticas utilizadas para se apresentar descritivamente os resultados e os testes estatísticos de hipótese.

CAPÍTULO 1: DEFINIÇÃO, APLICAÇÕES E EPISTEMOLOGIA DA ANÁLISE DE CONTEÚDO (AC)

Neste capítulo inicial, faremos uma apresentação geral sobre a análise de conteúdo. Inicialmente, trataremos das múltiplas definições da técnica, buscando localizar a definição com a qual iremos trabalhar na obra. Em seguida, apresentaremos um breve histórico da técnica de pesquisa. Depois, faremos um apanhado de aplicações diversas da técnica, com destaque para suas diferentes aplicações nas políticas públicas. Por último, faremos uma breve discussão sobre a epistemologia da técnica, reiterando os três princípios basilares de uma AC científica, nomeadamente a replicabilidade, a confiabilidade e a validade. Essa discussão será concluída com uma reflexão sobre as fronteiras entre AC e outras técnicas de pesquisa.

1.1 Definição

Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens (BARDIN, 2016, p. 48).

Como já abordado, nos poucos estudos que buscam compreender a aplicação da análise de conteúdo em pesquisas brasileiras, Bardin ([1979] 2016) é a autora mais acionada e a definição acima (ou variações dela) corresponde à mais visitada. É, inclusive, a definição adotada por Franco (2005) em seu manual. Em outro momento de sua obra, Bardin busca explicá-la didaticamente: “a técnica consiste em classificar diferentes elementos nas diversas ‘gavetas’ segundo critérios susceptíveis de fazer surgir um sentido dentro de uma ‘confusão’ inicial” (*ibidem*, p. 43).

Para Martin W. Bauer, outra definição bastante usada em textos brasileiros seria a de que a AC “é uma técnica para produzir inferências de um texto focal para seu contexto social de maneira objetivada” (BAUER, 2007,

p. 191). Bauer ainda ressalta que acredita que a AC seja uma técnica híbrida, tendo características tanto quantitativas quanto qualitativas.

Entre várias outras definições existentes e disponíveis, certamente uma a ser destacada é a realizada por Kimberly Neuendorf, que busca não apenas definir a AC, mas também restringi-la a uma certa noção. Para a autora, a AC precisa, necessariamente, seguir seis preceitos, nomeadamente: 1) ser baseada no método científico; 2) possuir uma unidade de análise ou unidade de coleção ou ambos; 3) ser quantitativa; 4) ser condensadora do conteúdo sendo verificado; 5) ser aplicável a todos os contextos; e 6) ter todas as características das mensagens disponíveis de forma que seu conteúdo possa ser analisado (NEUENDORF, 2002). Tendo isso em vista, Neuendorf define a técnica da seguinte maneira:

Análise de conteúdo é uma análise quantitativa condensadora que se baseia no método científico (incluindo atenção a objetividade-inter-subjetividade, design anterior, confiabilidade, validade, generalização, replicabilidade e teste de hipóteses) que não é limitada para os tipos de variáveis que podem ser medidas ou a contexto no qual as mensagens são criadas ou apresentadas (NEUENDORF, 2002, p. 10)⁶.

Apesar de ser certamente uma definição pertinente e completa, Neuendorf assume abertamente que só considera análise de conteúdo aquela que gerar resultados quantitativos, algo seguido no manual de Riffe, Lacy, Fico (2014).

Análise de conteúdo quantitativa é o exame sistemático e replicável de símbolos da comunicação, aos quais são atribuídos valores numéricos de acordo com regras de medida válidas, sendo, também, a análise de relações envolvendo esses valores usando métodos estatísticos, visando descrever a comunicação, gerar inferências sobre seus significados ou inferências sobre a comunicação em termos de seu contexto, produção e consumo (RIFFE; LACY; FICO, 2014, p. 19).

Apesar de apresentarmos justamente a AC quantitativa, não concordamos com os autores que essa é a única forma de se enxergar a

⁶ Este e todos os outros trechos de textos em outras línguas foram traduzidos livremente.

técnica. Na definição de Riffe e colegas, por exemplo, só será AC aquela que for quantitativa e que ainda passar necessariamente por testes estatísticos. Bardin (2016), Bauer (2007), Drisko e Maschi (2016), Franco (2005), Mayring (2014) e Krippendorff (2004) são alguns dos manuais que também não aderem à afirmação, demonstrando haver, sim, análises de conteúdo qualitativas⁷ (ver também ESPÍRITO SANTO; SOARES, 2015)⁸.

Klaus Krippendorff, um dos principais autores sobre análise de conteúdo no mundo, também concorda que a AC não pode ser vista como exclusivamente quantitativa. Então, ele opta por uma definição mais concisa da técnica. A “análise de conteúdo é uma técnica de pesquisa que objetiva criar inferências válidas e replicáveis de textos (ou outro conteúdo significativo) para os contextos de seu uso” (KRIPPENDORFF, 2004, p. 10). Como técnica de pesquisa científica, ela envolve procedimentos especializados e precisa, necessariamente, estar assentada em três princípios básicos da ciência, a saber: validade, confiabilidade e replicabilidade⁹, como vamos discutir no próximo tópico. Enquanto simples e aplicável aos mais diversos contextos, a definição de Krippendorff nos soa como excessivamente aberta.

Uma definição pouco visitada pela literatura, porém mais completa, é apresentada por Downe-Wamboldt (1992), que afirma que:

A análise de conteúdo é um método de pesquisa que providencia meios objetivos e sistemáticos para fazer inferências válidas de dados verbais, visuais ou escritos para descrever e quantificar fenômenos específicos. Infelizmente, para alguns pesquisadores, validade científica é igualada a quantificação [...] Análise de conteúdo é mais que um jogo de soma; ela se preocupa com significados, intenções, consequência e com o contexto (DOWNE-WAMBOLDT, 1992, p.314).

⁷ Notadamente na literatura brasileira há bastante discussão e aplicação de AC qualitativa: Caregnato, Mutti (2006); Gondim, Bendassolli (2014); Lima, Manini (2017); Mozzato, Grzybovski (2011); Oliveira (2008); Taquette, Minayo (2015); Vergara (2011), entre tantos outros.

⁸ O manual de Philipp Mayring (2014) “*Qualitative content analysis: theoretical foundation, basic procedures and software solution*” parece ser um dos mais completos acerca da AC qualitativa. Também podemos citar o “*The coding manual for qualitative researchers*”, de Johnny Saldaña (2012), que apresenta princípios de codificação para análise qualitativa similares à AC.

⁹ Espírito Santo, Soares (2015), por exemplo, apresentam a técnica “análise de conteúdo etnográfica” e realizam pertinentes discussões sobre os três conceitos em sua aplicação.

A definição de Downe-Wamboldt (1992) nos parece, de longe, a mais completa, sem ser totalmente restritiva, como aquela apresentada por Neuendorf e Riffe *et al.* Não obstante, apesar de admitir que não se trata de um jogo de soma, o autor ainda define a AC como necessariamente quantitativa. Assim, adaptamos a sua definição para o que acreditamos que ser uma mais adequada:

Análise de conteúdo é uma técnica de pesquisa científica baseada em procedimentos sistemáticos, intersubjetivamente validados e públicos para criar inferências válidas sobre determinados conteúdos verbais, visuais ou escritos, buscando descrever, quantificar ou interpretar certo fenômeno em termos de seus significados, intenções, consequências ou contextos.

Mesmo nossa apresentação se restringindo à análise de conteúdo categorial quantitativa, acreditamos que a definição acima é capaz de ser ampla o suficiente para diferentes abordagens da AC, inclusive as exclusivamente qualitativas. Em especial, reforçamos que é vital que qualquer AC seja vista como uma técnica de pesquisa científica; portanto, ancorada pelos princípios de replicabilidade, confiabilidade e validade, sobre as quais trataremos na seção “Epistemologia da AC”, neste capítulo.

1.2 Histórico

Como a história da análise de conteúdo está devidamente documentada em outros manuais (BARDIN, 2016; DRISKO; MASHI, 2016; NEUENDORF, 2002; KRIPPENDORFF, 2004), optamos aqui pela apresentação apenas de um condensado de algumas fases. Krippendorff (2004) afirma que o termo *content analysis* não foi utilizado em inglês até 1941, mas que análises sistemáticas de textos podem ser verificadas desde o século 17 na igreja. Há consenso, entretanto, de que as primeiras aproximações de uma técnica científica foram voltadas inicialmente para análises quantitativas dos jornais. Krippendorff cita o estudo de Speed (1893), *Do newspapers now give the news?*, que buscou verificar a diminuição da cobertura de temas literários, científicos e religiosos para privilegiar fofocas, esportes e escândalos. Menciona também a pesquisa de Mathews (1910), que buscou verificar se maior espaço estava sendo dado

a conteúdo desmoralizante, mórbido e trivial em oposição a itens “dignos” de atenção no jornalismo. Após a Crise de 1929, inúmeros problemas políticos e sociais aumentaram nos EUA, e havia uma crença de uma culpa parcial do jornalismo amarelo. Simultaneamente, o rádio ganhava progressivamente mais força e a televisão despontaria em seguida. Surgiam, então, dúvidas sobre a influência das mídias massivas, notadamente no que tange ao uso do rádio para o despertar e para a consolidação do fascismo.

Simultaneamente, segundo o autor, havia a emergência de Ciências Sociais e Comportamentais e um maior aceite de seus métodos e resultados. Havia interesse desses cientistas sociais em estereótipos, estilos, símbolos, valores e mecanismos de propaganda e houve o emprego de ferramentas estatísticas de outras disciplinas, como *surveys* e psicologia experimental. O surgimento do conceito psicológico de “atitude” adicionava dimensões avaliativas à AC, como “pró-contra” e “favorável-desfavorável”, o que não era visto na análise quantitativa de jornais anterior. Logo, era possível verificar a justiça e equilíbrio dos padrões jornalísticos e avaliar algum viés (*bias*). Ademais, dados de análise de conteúdo começaram a fazer parte de esforços mais extensos de pesquisa e a AC não era mais vista como um método à parte (KRIPPENDORFF, 2004).

Nessa fase de consolidação da técnica da AC, Krippendorff destaca dois grupos que se debruçaram sobre a análise de propagandas em mídia, dado seu farto uso na I Guerra Mundial e no período entre guerras, especialmente por regimes fascistas na Europa. O primeiro grupo é liderado por Hans Speier e é um projeto de pesquisa sobre comunicação totalitária em Nova York, financiado pelo *Foreign Broadcast Intelligence Service* do *Federal Communications Commission* (FCC) dos Estados Unidos. O grupo do FCC analisou transmissões de rádio para compreender e mesmo prever eventos da Alemanha Nazista e de outros países do Eixo, avaliando os efeitos das ações militares dos Aliados no sentimento das populações dos países inimigos dos EUA. Segundo Krippendorff (2004), analistas do FCC descobriram que poderiam antever algumas táticas de países do eixo por mensagens que procuravam garantir apoio popular às ações militares em seus países. Foram capazes de prever campanhas políticas e militares, de avaliar as percepções

da elite nazista de sua situação, as mudanças políticas dentro do grupo nazista governante e as mudanças nas relações entre os países do eixo. Ao analisar os discursos de Joseph Goebels, conforme Krippendorff (*ibidem*), eles foram capazes de inferir a data de lançamento do uso de certas armas especiais contra a Grã-Bretanha.

O segundo grupo foi liderado por Harold D. Lasswell, que trabalhou diretamente para a Divisão Experimental com o estudo da comunicação em períodos de guerra da biblioteca do Congresso. Esse grupo se focou em jornais e em transmissões de rádio de outros países e abordou questões básicas de amostra, problemas de medida, confiabilidade e validade das categorias de análise de conteúdo (KRIPPENDORFF, 2004). Assim, tanto Bardin quanto Neuendorf consideram que Lasswell seria o fundador da análise de conteúdo enquanto técnica científica, pois teria sido justamente ele e sua equipe que definiram os métodos para testes-piloto, treinamento de codificadores e avaliações de confiabilidade que serviram como modelos para a prática atual¹⁰.

Segundo Bardin (2016), apesar de os trabalhos iniciais de Lasswell serem da década de 1920, é a década 1940-1950 que sistematiza as primeiras regras da AC quantitativa. A Ciência Política ocupa destaque no desenvolvimento da AC, buscando compreender a cobertura sobre a II Guerra Mundial. Há especialmente uma busca por propaganda subversiva nos meios de comunicação (pró-nazistas ou comunistas) e de um favoritismo/desfavoritismo (temas favoráveis ao inimigo) mediante do emprego de termos-chave. Para além do papel-chave de Lasswell na pesquisa, também realizada na escola de jornalismo de Columbia, Bardin (2016) destaca a importância de Berelson, com ajuda de Lazarsfeld, nas primeiras definições metodológicas da AC. A definição do autor resumiria as preocupações epistemológicas desse período, explicitamente: “A análise de conteúdo é uma técnica de investigação que tem por finalidade a descrição objetiva, sistemática e quantitativa de conteúdo manifesto da comunicação” (BERELSON, 1952, p. 18 *apud* BARDIN,

¹⁰ Note-se que o livro “The language of Politics” (1948), que trata dos primeiros achados da equipe de Lasswell e das discussões metodológicas iniciais da AC, foi traduzido para o português e lançado pela editora da Universidade de Brasília. LASSWELL, Harold. *A linguagem da política*. Brasília: Editora Universidade de Brasília, 1982.

2016, p. 24)¹¹. Para Bardin (2016), a preocupação desse período consistia em trabalhar com amostras reunidas de maneira sistemática, interrogar-se sobre a validade dos procedimentos e dos resultados e verificar a confiabilidade dos codificadores e até medir a produtividade da análise.

Entre o final da década de 1940 e o início dos anos 1950, os autores concordam que há tanto uma expansão da técnica para outros campos de pesquisa quanto certa perda do foco (BARDIN, 2016; KRIPPENDORFF, 2004). Dessa forma, a Etnologia, a História, a Psiquiatria, a Psicanálise, a Linguística acabam por se juntar à Sociologia, Psicologia, Ciência Política e Comunicação no uso da análise de conteúdo. Bardin afirma que, depois dessa expansão, surge um “desencanto” com a técnica, que foi vista como tendo pouco rendimento analítico, enquanto Krippendorff (2004, p. 12) fala de uma falta de foco. Segundo ele, “tudo parecia ser analisável pelo conteúdo e toda análise de fenômenos simbólicos parecia ser uma análise de conteúdo”.

Assim, continuam os autores, buscando resgatar esse foco, o Comitê de Pesquisa em Ciência Social em Linguística e Psicologia realizou uma conferência sobre análise de conteúdo em 1955, tendo participantes de Psicologia, Ciência Política, Literatura, História, Antropologia e Linguística. Dessa conferência, foi publicado um livro chamado *Trends in Content Analysis*, editado por Ithiel de Sola Pool (1959)¹². De acordo com Bardin (2016), os indicativos são de uma diminuição da exigência da objetividade, pois havia uma confusão entre objetividade e cientificidade (vista geralmente enquanto análise de frequência). Portanto, a AC não teria mais um alcance puramente descritivo, e seu grande objetivo passaria a ser a inferência. “Toma-se consciência de que, a partir dos resultados da análise, se pode regressar às causas, ou até descer aos efeitos das características das comunicações” (BARDIN, 2016, p. 27).

As décadas de 1960 e 1970, segundo os autores, trariam as primeiras experiências de análise de conteúdo automatizada (realizadas por computador) e suas questões técnicas, como a dificuldade de verificar linguagem não verbal (visual, vocal e musical), a necessidade de categorias

¹¹ BERELSON, B. *Content analysis in communication research*. Glencoe: The Free Press, 1952.

¹² POOL, I. de S. (Org). *Trends in content analysis*. Urbana: Illinois University Press, 1959.

padronizadas, os problemas envolvendo inferências, os papéis de teorias e construtos analíticos (KRIPPENDORFF, 2004, p. 13). Bardin (2016), tratando do contexto francês, ainda relata uma disputa da AC com a semiologia e linguística e, portanto, com a análise de discurso no geral. Para a autora, havia uma forte necessidade de os pesquisadores da AC se protegerem para não serem “invadidos” pelos opositores.

1.3 Aplicações da AC

A análise de conteúdo apresenta inúmeras aplicações, podendo ser a técnica da pesquisa em si a gerar dados e resultados, ou mesmo como uma técnica intermediária em meio a projetos maiores. Não é raro, por exemplo, o uso da AC apenas para a análise de dados e conteúdo gerado por outras técnicas de pesquisa, como entrevistas em profundidade, grupos focais e diários de campo de etnografia e observações participantes (BARDIN, 2016; FLICK, 2009; HARWOOD; GARRY, 2003; YIN, 2016). Berelson (1952 *apud* KRIPPENDORFF, 2004, p. 45-46) cita 17 diferentes usos:

Descrever tendências no conteúdo da comunicação.

Traçar o desenvolvimento de conhecimento.

Desvelar diferenças no conteúdo de comunicações em contexto internacional.

Comparar as mídias e diferentes níveis de comunicação.

Auditar o conteúdo de comunicação em relação a seus objetivos.

Construir e aplicar padrões de comunicação.

Ajudar em operações técnicas de pesquisa (codificar questões abertas em questionários).

Expor técnicas de propaganda.

Descobrir características estilísticas.

Identificar as intenções e outras características da comunicação.

Determinar o estado psicológico de pessoas ou grupos.

Detectar a existência de propaganda (para termos legais).

Assegurar inteligência militar ou política.

Refletir sobre atitudes, interesses e valores (padrões culturais) de grupos da população.

Revelar o foco da atenção.

Descrever respostas atitudinais e comportamentais para comunicações.

Como é notável, a maior parte dos usos pensados por Berelson ou se aplicam à comunicação, ou à compreensão de comportamentos. Holsti (1969 *apud* KRIPPENDORFF, 2004) também contextualiza, em sua visão, os três propósitos principais da AC no contexto da comunicação:

Descrever características manifestas da comunicação: questionar quem, o quê, como sobre uma mensagem.

Fazer inferências dos antecedentes da comunicação: questionar por que algo é dito;

Fazer inferências das consequências da comunicação: questionar efeitos do que é dito (HOLSTI, 1969 *apud* KRIPPENDORFF, 2004, p. 46).

Não obstante, em caminho similar, fazendo uma substituição da ideia de comunicação pela de conteúdo, Bardin (2016, p. 44) afirma que “a intenção da análise de conteúdo é a inferência de conhecimentos relativos às condições de produção (ou, eventualmente, de recepção), inferência essa que recorre a indicadores (quantitativos ou não)”. Portanto, as inferências ou deduções lógicas buscam responder a dois problemas principais:

O que levou a determinado enunciado? Este aspecto diz respeito às causas ou antecedentes da mensagem;

Quais as consequências que determinado enunciado vai provavelmente provocar? Isto se refere aos possíveis efeitos das mensagens (por exemplo: os efeitos de uma campanha publicitária, de propaganda) (BARDIN, 2016, p. 45).

Note-se, portanto, que, nessa definição, buscar compreender a produção ou a consequência do conteúdo nos facilita que a AC possa ser aplicada a contextos fora da comunicação e da política. Na Educação, por exemplo, ACs podem ser utilizadas em livros ou manuais didáticos para

verificar como determinado conceito vem sendo abordado nas escolas. Um pesquisador da história poderia fazer verificação similar para compreender como determinado evento histórico está sendo retratado nas obras educacionais. Em ambos os casos, os pesquisadores poderiam se questionar sobre as explicações que justificam a produção desses conteúdos (Afinal, por que determinado período histórico é retratado desta maneira?); ou mesmo sobre os possíveis efeitos disso (Será que o fato de não estudarmos devidamente a história da África faz os estudantes valorizarem menos as culturas negras?).

Dentro dessa lógica, Krippendorff (2004) propõe inúmeras outras formas de utilizarmos a AC dentro da lógica científica. Ele nos evidencia que a AC pode ser usada para fazer extrapolações, estabelecer tendências, para compreender padrões e avaliar diferenças. Na prática, entre outras possibilidades, o autor ressalta como podemos usar a AC para identificações, fazer avaliações (de várias naturezas), julgamentos (são baseados em padrões, prescritos ou legitimados por instituições), criar índices, análises de conversações e também para a compreensão de processos institucionais.

Para Krippendorff (2004), a AC é mais bem-sucedida quando analistas consideram realidades sociais constituídas linguisticamente que são enraizadas nos tipos de conversações que geram textos a serem analisados. Fenômenos repetitivos, rotineiros, públicos e institucionalizados são mais fáceis de inferir do que aqueles raros e pouco convencionais.

Nesse sentido, essas são as áreas vistas como mais exitosas no uso da AC pelo autor.

- *Atribuições*: conceitos, atitudes, crenças, intenções, emoções e processos cognitivos que se manifestam em atributos verbais de comportamento. Podem ser vistos em conversações, mas também em meios de comunicação.
- *Relações sociais*: foco em como a língua é usada, baseado em gramáticas sociais de discursos gravados ou comunicação escrita de locutores ou escritores, considerando questões como autoridade, poder, acordos contratuais e desigualdades.
- *Comportamentos públicos*: valores individuais, disposições, concepções do mundo e comprometimento ao seu jeito de ser em

conversações envolvem repetições de confirmação. À medida que é um comportamento público, logo observado e julgado pelos outros, ele está no domínio da linguagem.

- *Realidades institucionais*: geralmente ignoramos a natureza institucional da realidade do casamento, dinheiro, governo, história, doenças e mesmo propósitos científicos (KRIPPENDORFF, 2003, p. 76-77, grifos do autor).

Para além dessa avaliação de Krippendorff, Bauer (2007) apresenta vantagens e fraquezas da técnica que nos parecem pertinentes. Como vantagens, ele afirma que a AC: 1) faz uma ponte entre um formalismo estatístico e a/análise qualitativa dos materiais; 2) reduz grandes quantidades de texto em uma descrição curta de algumas de suas características; 3) faz uso de materiais que ocorrem naturalmente; 4) é sistemática e pública (conjunto de procedimentos maduros e bem documentados); e 5) presta-se para dados históricos. Por sua vez, as principais fraquezas da AC seriam: I) introduzir inexatidões de interpretação; II) não reproduzir contexto original na análise; III) tender a se centrar em frequências e descuida do que está ausente; e IV) perder a relação entre códigos e texto (*e.g.* o momento em que algo é dito) (BAUER, 2007)¹³.

Apenas a título de ilustração, para facilitar uma melhor compreensão de possíveis aplicações da AC para o leitor, a pesquisa brasileira já fez uso da técnica para avaliar os mais diferentes tipos de conteúdo, como: artigos científicos em Contabilidade (ALVES, 2011), Administração (CASTRO *et al.*, 2011; GOMES *et al.*, 2020), Medicina (TAQUETTE; MINAYO, 2015); *Papers* apresentados em eventos de Administração (SILVA *et al.*, 2017), Comunicação (MARTINEZ; PESSONI, 2015; QUADROS; ASSMANN; LOPEZ, 2014;), Comunicação e Ciências Sociais (SAMPAIO; BRAGATTO; NICOLÁS, 2016; SARMENTO, 2018); leis (ALVES *et al.*, 2015); memes políticos (CHAGAS *et al.*, 2017); comunicação impressa de instituições de ensino superior (IKEDA; CHANG, 2005); etnografia com professores em formação continuada (URZETTA; CUNHA, 2013); mensagens postadas no Twitter (LYCARIÃO; DOS SANTOS; 2017; RECUERO, 2014; THOMAZ *et al.*, 2016); mensagens postadas no Facebook (ALVES, 2019; CARVALHO; MASSUCHIN; MITOZO, 2018); campanhas on-

¹³ Ver também HARWOOD, GARRY (2003); WHITE, MARSH (2006).

line em sites de redes sociais (AGGIO; REIS, 2013; CERVI; MASSUCHIN; CARVALHO, 2016; ROSSINI *et al.*, 2016); comentários de jornais on-line (BARROS; CARREIRO, 2015; BUENO, 2016); conversações e/ou deliberações on-line (DAHLBERG, 2004; GURGEL; LEITE, 2007; MENDONÇA; SAMPAIO; BARROS, 2016; STEENBERGER *et al.*, 2003); discordâncias e desrespeito on-line (CARREIRO, 2017; PAPACHARISSI, 2004); grupos de conversações sobre questões sensíveis ou polêmicas (MAIA *et al.*, 2017; STEINER, 2012); produção escrita de estudantes em questões discursivas de Matemática (SANTOS; DALTO, 2012); entrevistas sobre condições de trabalho em organizações (CAMARA, 2013); entrevistas com professores de educação física (SOUZA JÚNIOR; MELO; SANTIAGO, 2010); programas de televisão (ROCHA, 2008); telejornais (GOMES, 2009; MIGUEL; BIROLI, 2011); blogs (SANTOS; PENTEADO; ARAÚJO, 2009); pronunciamentos oficiais de presidentes brasileiros para compreender a Política Externa brasileira (SILVA, 2017); pronunciamentos oficiais de Dilma Rousseff para compreender seu mandato (GANDIN, 2018); respostas abertas em *surveys* políticos (HENKEL, 2017); HGPE (CERVI, 2011; FIGUEIREDO *et al.* 1997; PANKE; CERVI, 2012; VEIGA; GONDIM, 2001); campanha eleitoral negativa (BORBA, 2019); anotações de enfermeiros em prontuários (LOURENÇO; ZBOROWSKI; CURY, 2002; MACHADO; FIGUEIREDO, 2010); grupos focais em pesquisas clínico-qualitativas em distúrbios do desenvolvimento (SILVA; ASSIS, 2010); nível de transparência de websites governamentais (AMORIM; ALMADA, 2016; DINIZ; MACHADO; MATOS, 2016; PRADO; RIBEIRO; DINIZ, 2012; RAUPP; PERFOLL, 2017); programas de partidos políticos (TAROUCO; MADEIRA, 2013); websites de partidos políticos (BRAGA; ROCHA; CARLOMAGNO, 2015); notícias de jornais impressos sob a perspectiva do enquadramento (PRUDENCIO; RIZZOTTO; SAMPAIO, 2018); notícias de jornais sob a perspectiva da análise de valência (CERVI; MASSUCHIN, 2013; FERES JÚNIOR, 2016); notícias de jornais on-line (MASSUCHIN; TAVARES; BELEM, 2017); livros e materiais didáticos (RODRIGUES; DELLA JUSTINA; MEGLIORATTI, 2011; VASCONCELOS; SOUTO, 2003); editoriais de jornais impressos (MONT'ALVERNE; MARQUES, 2019); debates eleitorais (VASCONCELLOS, 2018); artigos da Wikipédia (SOUSA; MARQUES, 2016), entre tantos outros possíveis objetos. Para além disso, incorremos em algumas aplicações específicas da AC no campo das políticas públicas na próxima seção.

1.4 Aplicações em políticas públicas

O uso da análise de conteúdo nos estudos de políticas públicas tem se mostrado relevante devido ao poder de a respectiva metodologia gerar dados e interpretações capazes de sintetizar, comparar e testar hipóteses a partir de diferentes tipos de textos. Uma análise sistemática de revisão de literatura de cinco dos principais periódicos¹⁴ da área, de acordo com o Qualis-Capes (2013-2016) e com o *Journal Citation Reports* (ano base 2018 para a América Latina), mostra que, do ano de 2006 a 2018, há 36 estudos que manifestam, no resumo do trabalho, a utilização da análise de conteúdo como, pelo menos, um de seus instrumentos metodológicos.

Os principais tipos de materiais analisados em tais estudos são documentos, tais como estatutos (*e.g.*, LOI; DI GUARDO, 2015, PHILLIPS, 2013), atas de reunião (*e.g.* PAGNEZ, 2006) e até notas de campo (*e.g.*, JACOMETTI *et al.*, 2016). Além de documentos, outros tipos de materiais bastante recorrentes são as entrevistas (*e.g.*, BOZEMAN *et al.*, 2016; FIGUEIREDO, SALLES, 2017) e os formulários (*e.g.*, COUTO; RIBEIRO, 2016; WHITE *et al.*, 2010). Os recortes empíricos e campos de investigação são bastante diversos, os quais abrangem desde a incorporação de diretrizes governamentais na área da educação (SOUZA; MENEZES, 2015), de recursos hídricos (MORAIS; PAIVA; SOUSA; 2017), passando pelas condicionantes da colaboração científica (BOZEMAN *et al.*, 2016) e também por embates entre reforma e cultura organizacional do Poder Judiciário (LIMA; FRAGA; OLIVEIRA, 2016).

Dos 36 estudos em questão, 20 apresentam uma abordagem qualitativa (BLÜMEL; 2017; BOZEMAN *et al.*, 2016; SARAIVA; WAGNER, 2013; COUTO; RIBEIRO, 2016; SOUZA; MENEZES, 2015; LIMA; FRAGA; OLIVEIRA, 2016; FIGUEIREDO; SALLES, 2017; PHILIPPS, 2013; GUERRA; MACHADO, 2011; OLIVEIRA *et al.*, 2010; BIANCHI, DE FIGUEIREDO, 2017; BISPO; GOMES, 2018; CARVALHO; CORREIA, 2018; CASTANHO *et al.*, 2018; RUI *et al.*, 2010; SACRAMENTO; PINHO,

¹⁴ Os cinco periódicos selecionados para a revisão sistemática foram os seguintes: Ensaio – Avaliação e Políticas Públicas em Educação (com 13 artigos revisados), Revista de Políticas Públicas – UFMA (com 6 artigos), Revista de Administração Pública – FGV (com 10 artigos), Poverty and Public Policy (com 1 artigo) e Science and Public Policy (com 6 artigos).

2016; JACOMETTI *et al.*, 2016; CEZAR, 2018; SANTOS *et al.*, 2016; MORAIS; PAIVA; SOUSA, 2017). Ao adotar uma abordagem qualitativa, tais estudos recorrem a categorizações de textos, mas sem estabelecer quantificações dessas categorias; não realizando, portanto, a comparação de unidades semânticas em termos quantitativos. Sua principal contribuição acaba sendo, portanto, a apresentação de uma síntese da diversidade de interpretações e unidades semânticas contidas nos textos analisados.

Já adotando uma abordagem quantitativa, 12 dos 36 estudos revisados apresentam dados de frequência, proporção ou percentagens baseadas em categorias derivadas de análise de conteúdo (JUNGLUT; JUNGLUT, 2016; ARAYA-LEANDRO; CABA-PÉREZ; LÓPEZ-HERNANDEZ, 2016; MOREIRA; IULIANELLI, 2017; CONCEIÇÃO; BROTTTO, 2016; GURGEL; LEITE, 2007; FREIRE; DUARTE, 2016; CRUZ; SZYMANSKI, 2015; MORAES *et al.*, 2018; WHITE *et al.*, 2010; KRIEGER; ANDION, 2014; REIS; GONÇALVES; MESQUITA, 2012; ERLER, 2012). Outro tipo de técnica que implica quantificação se refere à análise de agrupamento (*cluster analysis*), realizada pelos estudos de Abreu, Gomes (2018) e o de Sauerbronn, Sauerbronn (2015, ver p. 729). Esse tipo de análise de agrupamento tem o benefício de tornar possível um contraste visual que evidencia se os atores estudados possuem focos temáticos ou semânticos nas suas falas, relatos ou escritos.

Se a análise de agrupamento implica uma espécie de quantificação devido à capacidade da técnica em identificar desníveis e contrastes nas categorias identificadas, há, também, estudos que, apesar de reivindicarem uma abordagem qualitativa na análise de conteúdo empreendida, acabam realizando técnicas de quantificação ao utilizarem frequência de palavras ou temas. Esse tipo de abordagem quantitativa não reivindicada ocorre nos estudos de Loi, Di Guardo (2015), Pagnez (2006) e Sauerbronn, Sauerbronn (2015).

Uma das razões para não reivindicarem precisamente uma abordagem quantitativa por parte desses estudos reside na perspectiva de que as quantificações são apenas pontos de partida para encontrar temas e unidades semânticas centrais nos textos analisados. Uma vez encontrados tais temas ou unidades, o grande esforço passa a ser destrinchar o emaranhado de interpretações e sentidos acionados por tais temas ou unidades semânticas, recorrendo, para isso, inclusive ao conhecimento de pano de fundo (não

presente no texto, mas no contexto) que permite uma interpretação em profundidade dos textos analisados. Como o esforço da pesquisa acaba se concentrando mais nesse empreendimento interpretativo, explica-se, então, um dos motivos pelos quais tais estudos não reivindicam para si efetivamente uma abordagem quantitativa.

Se há quantificações que funcionam para gerar unidades analíticas ou categorias semânticas a serem aprofundadas, há também a possibilidade de seguir o caminho inverso, ou seja, de se realizar uma análise de conteúdo qualitativa para se gerar dados quantitativos que serão, então, o foco da pesquisa. É o que ocorre quando uma pesquisa realiza entrevistas com perguntas abertas com um número reduzido de participantes para, com base nas respostas desses participantes, identificar, por meio da análise de conteúdo, categorias de respostas e, com elas, gerar questionários com perguntas fechadas. Com tais questionários, faz-se, então, uma aplicação a um volume estaticamente significativo de participantes e, com tais dados, se geram quantificações em relação aos itens mais frequentes nas respostas (ver Capítulo 3).

Com essa técnica, a avaliação de uma política pública (ou preferências) por parte dos usuários pode ser aferida pela aplicação de questionário (no caso, gerando uma *survey*), cujos itens das repostas foram extraídos a partir de uma análise de conteúdo aplicada às repostas na fase preliminar da pesquisa, ou seja, quando os participantes responderam de maneira não estimulada (sem opções de resposta) às perguntas. A vantagem desse tipo de técnica é que as opções dos questionários aplicados são derivadas do próprio público pesquisado e não das expectativas dos pesquisadores. Com isso, evita-se que opções de pergunta que sequer seriam concebíveis pelos participantes da pesquisa não sejam artificialmente emuladas pela pesquisa.

Um olhar panorâmico sobre os 36 estudos da revisão sistemática que realizamos revela, ainda, alguns gargalos metodológicos em relação à aplicação da análise de conteúdo. Isso porque, dos 36 trabalhos revisados, apenas três deles indicam algum elemento capaz de tornar a pesquisa replicável e, assim, escrutável em sua validade e confiabilidade (ver próxima seção). E isso ainda a partir de elementos muito iniciais e incompletos,

sejam eles notas ou esquemas (*memos*) contendo as regras de codificação (BLÜMEL, 2017; BOZEMAN *et al.*, 2016) ou mesmo os códigos de programação (JUNGBLUT; JUNGBLUT, 2016). Em relação à confiabilidade, encontramos uma maior responsividade a esse princípio, pois seis trabalhos (ARAYA-LEANDRO; CABA-PÉREZ; LÓPEZ-HERNANDEZ, 2016; BOZEMAN *et al.*, 2016; FREIRE; DUARTE, 2016; JUNGBLUT; JUNGBLUT, 2016; LOI; DI GUARDO, 2015; WHITE *et al.*, 2010) indicam a existência de testes de confiabilidade, ou mesmo a não necessidade dos mesmos devido à natureza das categorias ou mesmo da codificação. A próxima seção explica as razões pelas quais esses estudos responsivos à replicabilidade e à confiabilidade acabam por apresentar um nível destacado de rigor científico em relação aos demais.

1.5 Epistemologia da AC

As formulações de Karl Popper (1978) estabelecem uma compreensão sobre o desenvolvimento científico que baliza a discussão epistemológica que faremos, neste capítulo, sobre a AC. No caso, o autor compreende o desenvolvimento científico, inclusive nas Ciências Sociais, a partir de um constante exercício de crítica mútua entre os pesquisadores em relação às suas tentativas de solução para os problemas levantados. A importância para o exercício dessa crítica e, portanto, do estabelecimento da revisão entre pares, como princípio constituidor do trabalho científico, é de tal sorte que o autor defende que “se uma solução proposta não está aberta a uma crítica pertinente, então é excluída como não científica” (POPPER, 1978, p. 16).

Sob essa perspectiva, vale esclarecer que a natureza de um trabalho científico não se baseia no quão “verdadeiros” ou “corretos” são seus resultados produzidos, mas na capacidade desses resultados advirem de uma checagem crítica de hipóteses e, oportuno frisar, na possibilidade mesma de tais resultados serem revistos por outros estudos e esforços de pesquisa.

Tal compreensão epistemológica converge com a usual normativa entre especialistas em AC que, para ser balizada cientificamente, ela deve responder a três princípios epistemológicos fundamentais, sendo esses o de

replicabilidade, confiabilidade e validade (KRIPPENDORFF, 2004; MACNAMARA, 2005; MATTHES; KOHRING, 2008; NEUENDORF, 2002; RIFFE; LACY; FICO, 2014).

Todos esses princípios são identificados não pela autoridade ou reputação da fonte, ou seja, dos autores ou instituição de origem do trabalho, mas pela capacidade que a AC tem de ser **revisada** pela comunidade científica em termos de consistência entre conceitos e os instrumentos metodológicos (validade), assim como em termos da precisão e acurácia dos codificadores ao utilizarem tal instrumento (confiabilidade). A própria oferta de condições para realizar essas revisões corresponde precisamente ao princípio de replicabilidade. Sem essa, a revisão não é possível e, portanto, o processo circular do desenvolvimento científico é interrompido.

Nas subseções abaixo, pretendemos apresentar em mais detalhes cada um dos princípios epistemológicos em questão e como cada um deles se relaciona entre si.

Da validade

A validade pode ser definida como sendo “a extensão na qual um processo de medição representa o conceito intencionado e apenas tal conceito. Pensando sobre a validade, perguntamos ‘estamos mensurando o que almejamos mensurar?’” (NEUENDORF, 2002, p.112)¹⁵. De modo sintético, a validade exige que todas as categorias e suas operacionalizações em termos de codificação sejam significativas e úteis em termos do trabalho proposto, sua problemática, seus objetivos e fundamentação teórica (MORAES, 1999, sp.).

Assim, dizer que uma categorização deve ser válida significa dizer que ela deve ser aceita como capaz de traduzir empiricamente conceitos de modo significativo e pertinente. Porém, como toda tradução pressupõe uma correspondência por aproximação e não uma equivalência absoluta,

¹⁵ Os diferentes manuais tendem a apresentar diversas noções de validade. *Grosso modo*, discute-se validade interna ou validade de construto, que é justamente a que desejamos tratar neste livro, e a validade externa. A validade externa tem a ver com a própria validade dos resultados obtidos na análise de conteúdo, ou seja, o quanto eles representam o fenômeno sendo estudado em questão. Não trataremos sobre outras formas de validade, mas há discussões interessantes que podem ser vistas em Alonso e outros. (2012), Neuendorf (2002), Harwood, Garry (2003); Krippendorff (2004), Riffe, Lacy, Rico (2014), Weber (1990).

o nível de qualidade dessa tradução deve ser avaliado em referência aos objetivos da análise, à natureza e constrangimentos do material que está sendo utilizado, ao nível de complexidade dos conceitos e às questões que se pretende responder com a pesquisa.

A validade seria, desse modo, uma adequação epistemológica entre os propósitos da pesquisa e os instrumentos utilizados para identificar o fenômeno sob investigação. Isso, deve-se enfatizar, num nível sensível à especificidade própria que é traduzir em quantificações conceitos abstratos a partir de interpretações de construções linguísticas (de mensagens em forma de sons, texto ou imagem).

Uma especificidade que, apesar de partilhar algumas equivalências do modo como a validade é avaliada nas ciências duras, possui, contudo, nuances no caso da AC, dado que essa pressupõe operações interpretativas cujas decisões podem corresponder a mais de uma alternativa similarmente válida.

Para compreender tais nuances, é oportuno ilustrar como a validade e a confiabilidade podem ser, sem maiores constrangimentos, identificadas em pesquisas ligadas às ciências naturais e exatas. Se a pesquisa está, por exemplo, interessada em descobrir o peso médio dos moradores de uma cidade, serão necessários instrumentos válidos para mensurar esse peso, a exemplo de uma balança. Uma vez que uma amostra representativa (ver seção sobre amostra no próximo capítulo) de indivíduos da cidade tenha o peso de cada um mensurado com essa balança, seria possível, então, chegar ao peso médio.

Entretanto, caso a pesquisa realize essa mensuração com uma fita métrica, mensurando a altura dos indivíduos, isso, de maneira evidente, representaria um sério problema de validade à pesquisa. Se o objetivo fosse mensurar a altura média dos moradores da cidade escolhida, não haveria problemas de tal sorte, mas o objetivo anteriormente anunciado da pesquisa obriga que seus instrumentos metodológicos se mostrem consistentes.

Observe que, em tal situação, não está em questão o quão precisa é a mensuração da fita métrica em relação à balança, pois, *a priori*, uma fita métrica não seria sequer considerada um instrumento válido para o propósito expresso da pesquisa. Em comparação com a balança, a fita métrica poderia ser

um instrumento até mais preciso e, portanto, mais confiável do que a balança, especialmente se a balança estiver desregulada. Entretanto, a altura do indivíduo não seria considerada um indicador válido para se mensurar seu peso.

Esse exemplo almeja demonstrar, assim, que as dimensões de validade e confiabilidade são distintas, não podendo ser confundidas. Enquanto a confiabilidade se refere à estabilidade e precisão do instrumento utilizado pela análise da pesquisa (ou seja, uma balança desregulada ou uma fita métrica sem marcações não são confiáveis), a validade corresponde ao julgamento sobre a pertinência epistemológica do instrumento em relação à pesquisa.

Ainda que tal compreensão se aplique, em linhas gerais, à AC, é importante, contudo, também compreender que, diferentemente das ciências duras, a AC não possui métodos tão estáveis e consensuais para se traduzir um conceito. Assim, enquanto o método para se mensurar “peso” possui uma operacionalização autoevidente e intuitiva (no caso, o uso de balança), o mesmo não se pode imaginar para conceitos como “valência” (ver CERVI; MASSUCHIN, 2013; FERES JÚNIOR, 2016), “deliberatividade” (ver STEENBERGER *et al.*, 2003) ou “temáticas fundamentais” de falas (ver GURGEL; LEITE, 2007). Então, se essas operacionalizações não são tão intuitivas e autoevidentes, isso implica que a validade de uma AC pressupõe que ela seja cuidadosamente elaborada, de modo a fazer com que conceitos possam se converter em operações de codificação de textos. Em tal esforço, cada decisão em relação ao instrumento metodológico irá incidir no nível de aproximação das operações de codificação com o conceito teórico pretendido.

De todo modo, não importa quão cuidadosos e sofisticados se mostrem os critérios e regras que organizam essas operações; ao formar um instrumento metodológico, tais regras e critérios constituirão apenas uma das possibilidades concebíveis de se traduzir um conceito em operações de codificação de textos. Relacionado a isso, Krippendorff (2004) aponta que:

Textos não possuem significados que poderiam ser ‘descobertos’, ‘identificados’ ou ‘descritos’ pelo que são. Assim como textos podem ser lidos por inúmeras perspectivas, então signos podem ter diferentes designações e, assim, dados podem sofrer diversas análises (KRIPPENDORFF, 2004, p. 22).

Vale destacar que a amplitude dessas possibilidades tenderá a se ampliar em função do nível de complexidade teórica dos conceitos operacionalizados. Voltando a um dos exemplos anteriores, o conceito de “deliberatividade”, no caso, carrega consigo um conjunto de pressupostos e elementos teóricos que faz com que sua AC desagregue a codificação em diversas categorias e subcategorias analíticas. Assim, o conceito de deliberatividade acaba sendo desagregado em categorias como “respeito”, “reciprocidade discursiva” e “nível de justificação” etc. Porém, quantos níveis de justificação, por exemplo, devem ser codificados? A esse respeito, sabe-se que podemos codificar um proferimento do jeito mais simples, como “possui justificação” versus “não possui justificação”, até estabelecer uma gradação mais detalhada que parte do nível zero “sem justificação” até o nível 4, de “alto nível de justificação”. Mas, então, quantos níveis deveriam ser utilizados?

Decidir esse tipo de questão é algo que deve levar em conta a importância do “nível de justificação” para a teoria correspondente, o material sob análise e os propósitos da pesquisa. Isso faz com que a avaliação da validade de uma pesquisa de AC deva estar atenta de antemão ao fato de que cada decisão sobre quantas categorias e como (a partir de quais critérios) codificá-las afeta a validade do instrumento, e de que essas decisões dificilmente serão percebidas como inequívocas. Afinal, se há discordâncias até sobre as significações mesmas de conceitos como o de deliberatividade e do peso dos seus conceitos acessórios (STEINER, 2012; ver também discussão sobre a noção de respeito em PAPACHARISSI, 2004), tais discordâncias irão resultar inevitavelmente em diferentes desenhos metodológicos.

Por isso que é possível dizer que a validade de uma AC é dependente do quão convincente ela se mostra à comunidade para a qual ela é dirigida justamente por depender desse poder de demonstração e justificação. O fato de existir diversas possibilidades de desenhar uma AC válida não significa

que ela possua baixo rigor científico. Trata-se do exato oposto. Uma vez que a tradução de conceitos teóricos em uma AC requer diversos critérios de aplicação e complexas decisões teórico-metodológicas, a chance de contestação em relação a alguma, ou algumas, decisões se mostra bastante alta. Isso, de modo algum, seria um problema, mas algo regular e ordinário do trabalho científico. Afinal, como dito anteriormente, é justamente a partir de revisões críticas de trabalhos anteriores que a dinâmica do processo científico se constitui.

Da confiabilidade

Uma vez que percebemos que a validade depende da formulação de regras e critérios de codificação pertinentes e úteis aos objetivos da pesquisa, reside a questão sobre se a aplicação dessas regras se dá ou não de maneira confiável. Voltando ao exemplo anterior, isso implica questionar se a balança está devidamente calibrada ou se as marcações de distância de uma fita métrica estão corretamente impressas. Ou seja, enquanto a validade questiona se as regras formuladas fazem sentido, a confiabilidade questiona se tais regras estão sendo aplicadas de maneira estável e precisa.

A discussão em torno da confiabilidade possui especial relevância na mensuração proporcionada pela análise de conteúdo, pois esse método está sujeito ao erro humano. Por isso, compreende-se que qualquer mensuração derivada de AC apresentará um resultado que, em alguma medida, será influenciado por esse erro (ver NEUENDORF, 2002, p. 111). Um dos desafios da AC é, portanto, justamente reduzir este erro ao máximo, testando a confiabilidade e gerando condições (treinamento e condições de adequadas de trabalho) para a identificação de erros e suas causas.

Entre as causas mais comuns para o erro, pode-se destacar o cansaço do codificador que é exposto a horas contínuas de trabalho sem intervalo, falta de treinamento ou treinamento realizado em espaço de tempo insuficiente, falta de atenção gerada por barulhos e distrações externas do ambiente, assim como prazos exíguos que comprimem a produtividade a

ponto de gerar erros frequentes de digitação dos códigos ou preenchimento das planilhas de dados.

Outra causa para uma baixa confiabilidade de uma AC não reside no erro humano, mas no fato de que as regras estabelecidas para a codificação ainda não estão claras e precisas o suficiente para gerar uma codificação sem maiores dúvidas por parte do analista. Assim, diante de regras pouco claras ou muito vagas, o codificador ou analista acaba produzindo codificações incertas, sem a segurança se a escolha em torno de determinada codificação deveria ser de um código e não de outro. Para evitar isso, os conjuntos de manuais sugerem repetidos pré-testes com amostras do material a ser analisado, para verificar se as regras estão conseguindo ser claras e abrangentes o suficiente, a ponto de conferir segurança aos codificadores em suas decisões.

Alcançar esse nível de segurança é tão importante para a AC que a realização de testes de confiabilidade (ver Capítulo 2) é compreendida como elemento indispensável a uma AC com rigor científico. Por isso, uma pesquisa com AC, para ser confiável, necessita realizar testes de concordância entre diferentes codificadores. Nesses testes, cada codificador analisa o mesmo material, mas separadamente. Ao se comparar as codificações, espera-se que o nível de concordância entre as decisões de cada codificador seja suficiente para aferir a confiabilidade do instrumento e da análise do material.

Isso, contudo, não seria, sob nossa perspectiva (previamente apresentada em SAMPAIO, LYCARIÃO, 2018), suficiente para confirmar efetivamente um padrão de alto rigor no tratamento da confiabilidade. Em tais termos, a confiabilidade de uma pesquisa e de seu instrumento metodológico não é estanque, mas sim passível de constante revisão. Assim, a confiabilidade se confirma toda vez que outros pesquisadores (fora do contexto da pesquisa original), ao utilizarem os mesmos procedimentos e codificado o mesmo material, chegam a resultados iguais ou similares. Tal compreensão se alinha à definição de Neuendorf sobre a confiabilidade, sendo essa, nas palavras da autora:

A extensão com que um procedimento produz os mesmos resultados em diferentes repetições. A noção relevante aqui para

a análise de conteúdo é que uma medida não possui valor se ela puder ser realizada apenas uma vez, por apenas uma pessoa particular (NEUENDORF, 2002, p. 112).

Assim, dados confiáveis seriam aqueles pouco inalterados após sucessivas replicações. Sobre essa questão, deve-se retomar a premissa de que a AC necessariamente envolve uma série de decisões baseadas em interpretações e que, portanto, são sensíveis à subjetividade do analista. A ideia de confiabilidade, deve-se esclarecer, não almeja anular essa subjetividade, mas padronizar, a partir de sucessivos testes e processo de treinamento, as formas com que diferentes codificadores compreendem as mesmas categorias analíticas. Com isso, eleva-se a chance de que a interpretação que esses codificadores fizerem do material da pesquisa seja capaz de revelar um resultado fruto de um instrumento metodológico consistente e compreensível à comunidade acadêmica pertinente.

Dito isso, cabe agora esclarecer algumas relações existentes entre a confiabilidade e a validade. A primeira que destacamos se refere à compreensão de que “confiabilidade não assegura validade, isto é, confiabilidade é necessária, mas não é condição suficiente para a validade” (NEUENDORF, 2002, p. 141). Para compreender isso, basta voltarmos ao exemplo da pesquisa imaginária que se propôs a descobrir o peso médio dos moradores de uma cidade. Não adianta que o instrumento utilizado seja uma balança, se essa estiver desregulada. Ou seja, ainda que seja válida como instrumento, a mensuração que a balança faz do peso dos moradores tem de ser confiável. De outro modo, os resultados encontrados estarão corrompidos.

Contudo, essa não é a única relação relevante entre confiabilidade e validade. Outra relação consiste numa possível proporcionalidade inversa entre uma e outra. Esse tipo de relação tende a ocorrer em pesquisas que trabalham com conceitos de alta complexidade teórica ou com textos especializados (a exemplo de prontuários médicos ou atas de grupos de trabalho de especialistas). Como a codificação desses textos requer um exigente padrão de conhecimento prévio por parte do analista, será relativamente mais custoso treinar diferentes codificadores que consigam ter

o mesmo nível de conhecimento e, assim, chegar aos mesmos resultados em um teste de confiabilidade.

Diante disso, para se alcançar um alto nível de concordância entre codificadores, o desenho da pesquisa exigirá um intenso treinamento desses por um pesquisador mais experiente. Como isso implica altos custos, tanto de ordem financeira como de tempo para a execução da pesquisa, uma forma de viabilizá-la com menos custos pode se dar mediante a redução da complexidade das variáveis ou mesmo do escopo do material a ser analisado. Com isso, os resultados dos testes de confiabilidade tenderão a obter mais concordância, mas podem já não apresentar a mesma fidelidade conceitual (validade) desejável.

Um exemplo disso se dá em relação à codificação do conceito de “reflexividade” nos trabalhos que analisam deliberação em fóruns públicos (ver DAHLBERG, 2004; MENDONÇA, SAMPAIO, BARROS, 2016). No caso, a reflexividade é entendida como o esforço por parte dos participantes em um debate de considerar a perspectiva do interlocutor, o que implica que, ao apresentar suas posições, o participante precisa fazer referência aos argumentos e perspectivas dos outros participantes. A reflexividade se apresentaria, então, no efeito da mudança e revisão dos próprios argumentos a partir da interação com os outros participantes.

Entretanto, identificar a presença desse efeito nas falas de participantes de fóruns públicos é notadamente desafiadora. Isso porque raramente os participantes admitem explicitamente terem mudado seus pontos de vista pela interação discursiva com outros interlocutores. Isso faz com que a incidência de identificação de reflexividade seja muito baixa.

Jensen (2003) propõe resolver esse problema codificando a reflexividade em três possibilidades, que são: (a) persuasão, quando há sinais explícitos de convencimento pela argumentação de outro participante ou pelo debate como um todo; (b) progresso, quando um participante faz referência a outra intervenção e responde com novos argumentos ou informações, ou mesmo quando realiza sínteses do que já foi dito; e (c) radicalização, quando o participante reage a outra intervenção e reforça ainda mais a posição anteriormente assumida, recusando, assim, outras possibilidades de tomada de posição.

Tal esquema de codificação pode trazer ganhos substanciais na confiabilidade, pois oferta baixo nível de dúvida por parte do codificador. Entretanto, tal esquema pode também acarretar perda de validade. Isso porque muitas intervenções podem ser codificadas como “b” (ou seja, progresso) devido à existência de participantes que tentam organizar os debates, produzindo síntese sobre argumentos em disputa. Isso, porém, não significará, efetivamente, que os participantes estão sendo reflexivos no sentido proposto pelo quadro teórico de origem. A partir disso, vale perguntar: o que está sendo mensurado é o que se almeja?

Assim, deve-se ter especial atenção para pesquisas com AC que operacionalizem conceitos teóricos que possuem alto nível de complexidade. Isso porque ganhos rápidos na confiabilidade tendem a sacrificar a validade. Ao mesmo tempo, um alto nível de validade em tais casos tenderá a apresentar um baixo nível de confiabilidade se não houver, previamente à análise, sucessivos pré-testes e intenso treinamento dos codificadores envolvidos na pesquisa.

Da replicabilidade

Como vimos anteriormente, uma AC pode ser considerada válida quando os critérios e regras que normatizam a codificação das categorias traduzem, de maneira razoavelmente aproximada, o conceito teórico subjacente. Já a confiabilidade desta mesma AC será satisfeita caso observadores externos venham a concordar que a aplicação desses critérios e regras se deu em conformidade a essas mesmas regras e critérios. Disso resulta que, quando os resultados de uma AC estiverem publicados sem qualquer especificação de seus critérios ou regras de codificação, tanto a sua validade quanto a confiabilidade não podem ser avaliadas. Daí que tanto a validade quanto a confiabilidade só podem ser concebíveis caso a pesquisa seja replicável, ou seja, caso os elementos que permitem a replicação por observadores externos estejam disponíveis.

Com isso, pode-se dizer que replicabilidade é um princípio epistemológico primordial da AC, pois é ele que permite aferir o nível com

o qual uma pesquisa pode ser replicada por pesquisadores externos aos da pesquisa. Isso implica que a replicabilidade não garante confiabilidade, nem a validade de uma AC, mas é condição de possibilidade das duas. Conforme aponta Krippendorff:

Técnicas devem ser confiáveis. Mais especificamente, técnicas de investigação devem produzir resultados que sejam replicáveis. Isto é, pesquisadores trabalhando em diferentes contextos espaciais e temporais sob diferentes circunstâncias deveriam chegar aos mesmos resultados ao aplicar a mesma técnica aos mesmos dados. A replicabilidade é a forma mais importante de confiabilidade. A pesquisa também deve produzir resultados válidos, no sentido de que os esforços de pesquisa se mostrem abertos ao escrutínio cuidadoso e que as afirmações resultantes possam ser confrontadas diante de evidência disponibilizada independentemente. (KRIPPENDORFF, 2004, p,18).

Para que pesquisadores cheguem a resultados iguais ou similares, considerando os mesmos válidos, são necessárias duas condições básicas: (a) disponibilização de uma descrição detalhada dos critérios e regras utilizadas para a análise; e (b) acessibilidade ao mesmo material analisado, estando o mesmo em condição equivalente ou suficiente para uma nova codificação externa ao contexto da pesquisa original.

Sob essa perspectiva, para fazer com que uma pesquisa atenda a esse princípio fundamental da AC, é importante que a exigência do item (a) seja feita mediante a publicação do máximo de informações sobre os procedimentos e regras utilizadas na análise, o que é estabelecido por meio de um livro de códigos. Oliveira (2008) esclarece a finalidade de um livro de códigos para a replicabilidade, uma vez que:

[...] as unidades decompostas da mensagem, as categorias que servem para classificá-la, devem ser definidas com tal clareza e precisão que outros, a partir dos critérios indicados, possam fazer a mesma decomposição, operar a mesma classificação (OLIVEIRA, 2008, p. 251)

Desse modo, o livro de códigos, além de indicar os códigos alfanuméricos que correspondem a cada variável e categoria, deve discriminar,

detalhadamente, inclusive com exemplos, como a codificação deve ser feita em cada código listado. Já em relação à exigência de se disponibilizar o material da pesquisa de modo a tornar sua replicação possível, é necessário enfatizar que, na atual configuração tecnológica, já não há como alegar falta de condições para tornar o material acessível de maneira pública ou semipública.

Isso porque a emergência de serviços especializados em armazenar e tornar disponível, em modo on-line, uma grande quantidade de dados permite disponibilizar diferentes tipos de linguagem com qualidade suficiente para uma replicação externa. Os detalhes operacionais sobre o uso desse tipo de serviço, assim como sobre a confecção do livro de códigos, podem ser encontrados no segundo capítulo desta obra.

De todo modo, é oportuno repisar que, numa pesquisa, mesmo aquelas que tenham condições plenas de replicabilidade, é possível que os resultados encontrados por outros pesquisadores sejam significativamente distintos dos da pesquisa replicada. Isso seria uma indicação de que, apesar de ser replicável, os resultados de tal pesquisa não teriam se mostrado plenamente confiáveis, uma vez que os resultados encontrados foram bastante distintos. Outra possibilidade diante de uma pesquisa plenamente replicável é que replicações externas podem até chegar a resultados iguais ou muito similares, atestando uma confiabilidade; mas, ao realizar tal replicação, observadores podem avaliar que os critérios e as regras de codificação utilizados não traduzem adequadamente o conceito acionado pela pesquisa, fazendo, então, com que sua validade seja colocada em dúvida. Por isso é que a replicabilidade é uma condição, mas não garantia, da confiabilidade e da validade.

1.6 Fronteiras

Como denotado por Bardin (2016), em certas áreas das humanidades, a análise de conteúdo esteve em rota de colisão com a análise do discurso (AD) e há, frequentemente, dúvidas sobre as diferenças entre as duas técnicas. Essa discussão, entretanto, já foi devidamente realizada pela literatura brasileira,

buscando tanto diferenciar as técnicas quanto a trajetória de cada uma (CAPPELLE *et al.*, 2003; CAREGNATO; MUTTI, 2006; LIMA, 2003; LIMA; MORAES, 2017; PÁDUA, 2002; ROCHA; DESDARA, 2005, 2006). Como já sabemos, a análise de discurso tem uma trajetória mais ligada à semiologia e à linguística, sendo assim mais preocupada com as condições de produção do discurso; porém, não somente no sentido do meio ambiente material e institucional do discurso, mas também no que tange as representações imaginárias dos agentes interagentes com sua própria identidade, ou ainda, “a análise do discurso se interessa pela linguagem tomada como prática: mediação, trabalho simbólico, e não instrumento de comunicação” (LIMA, 2003, p. 80). Esses imaginários sociais configuram um universo de significações que funda a identidade de um grupo, pessoa ou organização.

Em outras palavras, compreendemos que a análise de discurso se aproxima mais de um método de pesquisa fortemente ancorado em certas teorias sobre o discurso e a língua, enquanto a análise de conteúdo pode ser tratada exclusivamente como uma técnica de pesquisa que pode partir dos mais diferentes pressupostos. Portanto, análise do discurso parte de uma série de pressupostos que geralmente não existem (ou, ao menos, não precisam existir) na análise de conteúdo. Importa dizer que, apesar de poder trabalhar com o conteúdo latente ou mesmo buscar identificar as lacunas e conteúdo não presentes, a análise de conteúdo geralmente se baseia na análise do conteúdo manifesto, enquanto a análise de discurso naturalmente irá verificar de maneira mais enfática o contexto da produção do conteúdo. Isso não quer dizer, evidentemente, que toda AC é quantitativa e toda AD é qualitativa.

Importa ainda dizer que, uma vez que consideramos a AD como um método¹⁶, isso significa que ela pode ser conjugada ou mesmo baseada na AC como técnica de pesquisa¹⁷; afinal, “embora a diferença entre os dois conceitos seja porosa, o método é muito mais abrangente e se aproxima da epistemologia, contemplando estratégias gerais, enquanto a técnica é

¹⁶ “Métodos seriam estratégias de produção de conhecimento científico, incluindo a geração e a validação de teorias” (CANO, 2012, p.107).

¹⁷ “Técnicas seriam formas padronizadas de coleta e análise de dados, com a mesma finalidade, a de produzir conhecimento válido” (CANO, 2012, p.107).

específica e concreta” (CANO, 2012, p. 107). A possibilidade da conjugação das duas já foi discutida nessa mesma literatura que buscou compreender o lugar de cada. Apesar de variados, há exemplos tangíveis dessa aproximação, como é a sugestão de Panke e Cervi (2012) de se analisar o Horário Gratuito Político Eleitoral (HGPE) pela conjugação das suas técnicas e como foi a tese de Lucas Gandin (2018), que analisou todos os discursos proferidos pela Presidente Dilma Rousseff em seu mandato. Gandin faz uma interessante conjugação de análise léxica em conjunto com análise de conteúdo categorial quantitativa, mas sob pressupostos da análise de discurso.

Por sinal, uma segunda importante fronteira a ser traçada é justamente a diferença entre a análise de conteúdo e a análise léxica. É natural que haja uma confusão entre as duas, uma vez que, de fato, algumas análises de conteúdo são inerentemente léxicas, como denota Bardin (2016). Para Freitas e Janissek (2000, p. 33), “a análise léxica é o estudo científico do vocabulário, com aplicações de métodos estatísticos para a descrição do vocabulário” e sempre começa pela contagem de palavras. Portanto, a análise léxica é consideravelmente diferente da análise de conteúdo categorial que vamos apresentar neste manual. Isso não significa, evidentemente, que elas não possam ser utilizadas em conjunto. No momento atual, muitas pesquisas têm feito uso de *softwares* para análise léxica automatizada, como o Iramuteq (CAMARGO; JUSTO, 2013), e combinando os resultados com análises de conteúdo manuais, como é o supracitado caso de Gandin (2018), que inicialmente fez uma análise de conteúdo categorial dos discursos da presidente e depois os tratou no Iramuteq para compreender as conexões entre suas categorias manuais e as categorias léxicas entregues pelo *software*. Para uma discussão sobre análise lexical em conjunção com outros métodos de pesquisa, ver também Justo e Camargo (2014).

Por fim, importa dizer que, apesar de similares, há diferenças entre a análise de conteúdo e a análise documental. Esta seria um método para representar o conteúdo de um ou mais documentos, de modo a facilitar sua consulta e referência posteriormente (BARDIN, 2016). Novamente, é interessante compreender a análise documental como um método de pesquisa; portanto, é possível que a análise de conteúdo seja uma das

técnicas a serem empregadas para se retirar informações úteis aos objetivos da pesquisa nos documentos a serem analisados. Por outro lado, é possível que a análise documental não careça, frequentemente, de todos os critérios e passos de uma análise de conteúdo categorial quantitativa, como a aqui apresentada, e que possa ser em certas horas mais heurística e qualitativa. Ademais, enquanto a análise de conteúdo é diretamente conectada às inferências que podem ser produzidas após o conteúdo ser avaliado, a análise documental geralmente deseja produzir dados descritivos.

Portanto, enquanto resguardam todas essas similaridades, essas fronteiras da análise de conteúdo com outras técnicas e métodos de pesquisa precisam ser definidas. Igualmente, é importante notar como certas áreas de pesquisa trabalham com metodologias tão próximas à AC que haveria ganhos em efetivamente passar a considerar seus procedimentos. Apenas a título de exemplo, consideremos duas áreas.

A primeira é a área de análise de transparência digital, que não apenas se preocupa com a acessibilidade e “a visibilidade de dados e informações técnicas, contábeis, estatísticas e econômicas sobre representantes e negócios públicos, preocupando-se, sobretudo, com a inteligibilidade desses dados” (AMORIM; ALMADA, 2016, p. 55), como também afeta, como os meios digitais, notadamente a internet, esses processos de transparência. *Grosso modo*, uma considerável parte da pesquisa sobre transparência digital se concentra na análise de websites ou perfis de redes sociais digitais de órgãos governamentais. Os sites são analisados seguindo um livro de códigos definido pelo pesquisador, buscando identificar quais são os quesitos mais importantes em termos de transparência existentes ou não na plataforma digital em questão, assim como fazer uma avaliação sobre a qualidade dessas informações e/ou das ferramentas digitais existentes. Entre outros elementos, Amorim e Almada (2016) sugerem verificar a existência de: informações financeiro-orçamentárias, informações administrativas (processos internos), instrumento de acompanhamento e interação. E todos os conteúdos e ferramentas são avaliados em termos de usabilidade e acessibilidade. Para a geração de índices que permitam testes estatísticos mais refinados posteriormente, as autoras sugerem que certos conteúdos ou ferramentas

devem apresentar pesos maiores que outros, ou seja, serão considerados em maior proporção no índice proposto pelas autoras. Fica notável que boa parte do artigo das autoras se trata basicamente da apresentação e defesa de um livro de códigos, que será utilizado para avaliar conteúdos, no sentido mais amplo, dos websites governamentais. Ao verificar a literatura, é notável que vários estudos sigam a mesma lógica de análise (MACHADO; MATOS, 2016; PRADO; RIBEIRO; DINIZ, 2012; RAUPP; PERFOLL, 2017).

De maneira similar, cresceu, nos últimos anos, o interesse da pesquisa brasileira sobre a própria pesquisa. Estudos bibliográficos ou mesmo cientométricos crescem em abundância, buscando compreender a organização dos campos de estudos das mais diferentes áreas. Da mesma maneira, há um movimento surgido das ciências médicas e que ganha força em outras áreas para abandonarmos revisão de literaturas narrativas e partirmos para revisões sistemáticas da literatura. Em todos os casos elencados, a pesquisa, frequentemente, baseia-se na análise de conteúdo de materiais acadêmicos (artigos, *papers*, dissertações, teses etc.), observando-se autores, instituições, abordagens teóricas e técnicas de pesquisa e autores mais citados. Considerável parte dessas análises advém de interesses específicos dos pesquisadores e são materializados em livros de códigos (ALVES, 2011; CASTRO *et al.*, 2011; TAQUETTE; MINAYO, 2015; SILVA *et al.*, 2017; MARTINEZ; PESSONI, 2015; QUADROS; ASSMANN; LOPEZ, 2014; SAMPAIO; BRAGATTO; NICOLÁS, 2016; SARMENTO, 2018).

Em ambas as áreas, acreditamos que ou se trate de análises de conteúdo no sentido estrito, ou, no mínimo, baseiem-se em técnicas de pesquisa suficientemente similares para poderem se beneficiar das discussões e dos métodos apresentados neste livro. Como discutimos anteriormente, são pesquisas que poderiam se valer das discussões e procedimentos da AC em termos de replicabilidade e confiabilidade. É notável, por exemplo, que diversas pesquisas nessas áreas não tratam sobre o treinamento ou mesmo sobre a confiabilidade entre codificadores. Em diversos casos, falta transparência na falta de disponibilização dos respectivos livros de códigos e bancos de dados, o que evidentemente diminui a possibilidade de uma replicação de seus resultados.

CAPÍTULO 2: MANUAL DA AC

Este capítulo busca explicar, de forma detalhada, a aplicação da análise de conteúdo categorial quantitativa. Faremos isso em três partes. A primeira envolve as definições básicas das diferenças entre códigos, codificação e categorias, elementos usualmente tomados como sinônimos na literatura. A segunda, o desenho adequado de uma AC, apresentando um guia ou manual a ser seguido. A terceira, cada etapa da AC para facilitar sua aplicação.

1. Códigos, codificação e categorias

A análise de conteúdo categorial, como já dito, é uma técnica de pesquisa que busca permitir a criação de inferências sobre determinado conteúdo. Para tanto, os pesquisadores realizam a **codificação** do conteúdo, fazendo a aplicação de **códigos**, que vão formar **categorias**. Apesar de, frequentemente, serem vistos como sinônimos, cada um desses termos é importante para uma aplicação adequada da técnica.

Iniciamos pelo código, que é o elemento essencial da AC. Um código é uma palavra ou frase curta que confere um atributo saliente, essencial, evocativo e/ou que resume um dado baseado em texto ou mesmo visual (SALDAÑA, 2012, p. 3). Um código geralmente irá resumir, filtrar ou condensar dados de acordo com os interesses de pesquisa. Em outras palavras, código é um rótulo ou uma etiqueta que usamos para classificar, qualificar, registrar partes do conteúdo de acordo com os objetivos da pesquisa.

Segundo Holsti (1969), “a codificação é o processo pelo qual os dados brutos são transformados sistematicamente e agregados em unidades, as quais permitem uma descrição exata das características pertinentes do conteúdo” (*apud* BARDIN, 2016, p. 133). Todavia, a ideia de exatidão é, em si, questionável. Para Saldaña (2012), a **codificação** significa arranjar as coisas de forma sistemática, de modo a criar algum tipo de sistema, classificação ou categorização. Quando os códigos são aplicados ao conteúdo analisado, o pesquisador está codificando – um processo que permite aos dados serem segregados, reagrupados e reconectados para consolidar sentido e gerar

explicação. Codificar, então, é um método que permite ao pesquisador organizar e agrupar dados codificados em categorias ou famílias pelo compartilhamento de suas características. Não se trataria, então, apenas de etiquetar o conteúdo, mas de conectar o pesquisador dos dados às ideias e das ideias a todos os dados pertencentes a essa ideia (SALDAÑA, 2012, p. 3).

Por sua vez, categorias são construtos analíticos derivados de teorias ou práticas existentes; experiência ou conhecimento de *experts* ou pesquisa (KRIPPENDORFF, 2004, p. 173). Categorias são elementos que nos dão meios para descrever o fenômeno sobre investigação, aumentando o conhecimento e gerando conhecimento. Essencialmente, os códigos são agrupados em categorias quando são relacionados em termos de conteúdo ou contexto. Ou ainda, “códigos são organizados numa categoria quando eles estão descrevendo diferentes aspectos, similaridades ou diferenças do conteúdo do texto que devem permanecer juntos” (ERLINGSSON; BRYSEWICZ, 2017). As categorias são geralmente baseadas na pergunta de pesquisa, na unidade de análise selecionada, em teorias relevantes, em pesquisa prévia e mesmo com base nos próprios dados (CAVANAGH, 1997; DOWNE-WAMBOLDT, 1992). Portanto, em resumo, na análise de conteúdo categorial aqui descrita, a codificação se dará pela aplicação de códigos contidos em categorias.

Essas são definições que já nos ajudam a compreender o que é de fato uma AC. Todavia, esses códigos não podem ser aplicados indiscriminadamente a diferentes porções de conteúdo ou mesmo ser aplicados sob lógicas e olhares diferentes. A codificação precisa ser rigorosa e, para tanto, precisa seguir instruções e ser aplicada da mesma maneira. Então, iniciemos pela montagem de um esquema de codificação.

2. Desenho da análise de conteúdo

Uma das maiores lacunas do manual de Laurence Bardin (2016) está justamente na excessiva simplificação do desenho e do processo da análise de conteúdo. Como massivamente reproduzido em textos baseados em seu manual no Brasil (FRANCO, 2005; OLIVEIRA, 2008; MORAES, 1999), a autora resume a AC em apenas três fases, nomeadamente:

Primeira Etapa: pré-análise

Nesta etapa, são desenvolvidas as operações preparatórias para a análise propriamente dita. Consiste num processo de escolha dos documentos ou definição do corpus de análise; formulação das hipóteses e dos objetivos da análise; elaboração dos indicadores que fundamentam a interpretação final.

Segunda Etapa: exploração do material ou codificação

Consiste no processo pelo qual os dados brutos são transformados sistematicamente e agregados em unidades, as quais permitem uma descrição exata das características pertinentes ao conteúdo expresso no texto.

Terceira Etapa: tratamento dos resultados - inferência e interpretação

Busca-se, nesta etapa, colocar em relevo as informações fornecidas pela análise, por meio de quantificação simples (frequência) ou mais complexa como a análise fatorial, permitindo apresentar os dados em diagramas, figuras, modelos etc. (BARDIN, 1977 *apud* OLIVEIRA, 2008, p. 572).

Notadamente, todo o processo de desenho da pesquisa em AC estaria contido na segunda etapa do modelo proposto por Bardin, e toda a análise contida na terceira etapa. Apenas a título de exemplo, Bauer (2007) evidencia que tais etapas poderiam ser, facilmente, subdivididas em oito diferentes fases:

1. Seleção de textos específicos
2. Amostra
3. Construir referencial de codificação
4. Teste-piloto e revisar referencial
5. Teste de fidedignidade dos códigos
6. Codificar materiais e estabelecer fidedignidade do processo
7. Construir arquivo de dados para análise estatística
8. Folheto com:
 - a) racional para referencial de codificação

- b) distribuições de frequência de todos os códigos;
- c) fidedignidade do processo de codificação (BAUER, 2007, p. 215)

Da fase 1 até a 6, estaríamos na exploração do material, enquanto da fase 7 em diante estaríamos no tratamento dos resultados. Então, o modelo sugerido por Bauer é um avanço àquele de Bardin, por tornar mais explícito os inúmeros processos existentes na análise de conteúdo. Ainda assim, Bauer dá pouca atenção à primeira etapa proposta por Bardin, que é onde justamente surgem as inquietações acadêmicas que vão guiar o desenho de uma pesquisa científica. Nesse caso, optamos por apresentar um passo a passo da condução de uma análise de conteúdo categorial baseada na sugestão de Riffe, Lacy, Fico (2014) da divisão em três grandes etapas: 1) **Conceituação e propósito**; 2) **Desenho** e 3) **Análise**. Entretanto, na especificação das subetapas, fizemos um amálgama entre as sugestões de outros autores (DOWNE-WAMBOLDT, 1992; WEBER, 1990; NEUENDORF, 2002). Dessa maneira, sugere-se o esquema abaixo:

ETAPAS DA ANÁLISE DE CONTEÚDO	
Conceituação	
1.	Identificar o problema (revisão de literatura)
2.	Questões de pesquisa e hipóteses
Desenho	
3.	Selecionar a(s) unidade(s) e subunidade(s) de análise
4.	Criar e definir categorias
	a. elaboração do livro de códigos
	b. elaborar a planilha de codificação
5.	Amostragem
6.	Pré-teste das categorias e das regras de codificação
	a. treinamento
	b. revisão do livro de códigos
	c. teste de confiabilidade-piloto
7.	Treinamento final e teste de confiabilidade das categorias
8.	Codificação
9.	Testes de confiabilidade intermediário e final
Análise	
10.	Tabulação e aplicação de procedimentos estatísticos
11.	Interpretar e reportar os resultados
12.	Validação e replicabilidade

1. Identificar o problema (revisão de literatura)

Como já discutido anteriormente, trabalhamos com a análise de conteúdo enquanto uma técnica de pesquisa científica. Como tal, ela deve ser aplicada para responder a problemas e, especialmente, a questões de pesquisa. Em outras palavras, a própria maneira como uma análise de conteúdo será configurada depende diretamente do tipo de questões a que ela deseja dar resposta.

Enquanto não se deseja aprofundar a questão sobre como elaborar um problema de pesquisa, definitivamente se recomenda que o mesmo seja

oriundo de uma revisão de literatura adequada, que busque efetivamente levantar um estado da arte do objeto a ser investigado. Uma revisão de literatura adequada deve responder à questão “quanto já sabemos sobre o fenômeno a ser estudado”?

Em especial, é muito comum que as revisões de literatura foquem excessivamente no estado da arte da teoria ou do conceito a ser utilizado, enquanto defende-se que será mais profícua a concentração no estado da arte das pesquisas relacionadas ao fenômeno que se pretende estudar. Notadamente, diversas áreas começam a sugerir o uso da revisão de literatura sistemática (PEREIRA; GALVÃO, 2014; FIGUEIREDO FILHO *et al.*, 2014), que busca justamente apresentar uma revisão da literatura menos narrativa e mais centrada em resultados de pesquisas científicas. Uma revisão de literatura adequada produzirá um estado da arte das pesquisas sobre o fenômeno, que apresentará, entre outras coisas, 1) métodos e técnicas utilizadas em estudos anteriores; 2) corpus utilizado para a pesquisa; e 3) resultados encontrados. Dessa maneira, tende a ser mais viável a identificação das lacunas presentes em estudos da área e dos pontos que justamente merecem receber mais atenção, ou seja, do problema de pesquisa.

2. Questões de pesquisa e hipóteses

A revisão de literatura adequada dará ao pesquisador uma noção do estado da arte da pesquisa sobre o fenômeno a ser investigado. Em especial, uma revisão de literatura adequada que possa ser útil para a elaboração de uma análise de conteúdo terá justamente verificado estudos anteriores que aplicaram a técnica, ressaltando a amostra ou corpus, a tipologia ou o mesmo o livro de códigos utilizado e os resultados apontados. As amostras e corpus serão úteis para o delineamento do próprio corpus da pesquisa em tela, enquanto o acesso a livros de códigos e tipologias anteriores permite tanto a replicação de resultados anteriores quanto a construção de verdadeiras linhas de pesquisa que se apoiam em bases similares.

A apresentação adequada de resultados anteriores é, definitivamente, um dos pontos mais importantes. Com base nesse estado da arte do que já foi

identificado nos estudos já realizados, será mais fácil a elaboração de questões de pesquisa e hipóteses pertinentes do ponto de vista científico. Em suma, acredita-se que uma revisão de literatura adequada e a criação de questões de pesquisa e hipóteses baseadas em resultados prévios contribuem para o acúmulo de conhecimento na ciência.

3. Selecionar a(s) unidade(s) e subunidade(s) de análise

Uma das decisões mais importantes para o desenho da pesquisa da análise de conteúdo é a definição da unidade de análise e de possíveis subunidades de análise. Em outras palavras, qual é a unidade de conteúdo sendo investigada (são *posts* de uma rede social? Editoriais de jornais? Prontuários de pacientes? Projetos de lei? A transcrição de falas de discursos de parlamentares?) e exatamente como o conteúdo está sendo analisado, ou seja, cada palavra está sendo avaliada, ou são frases, ou ainda parágrafos? Ou a unidade de análise é o texto como um todo? São questões que precisam de respostas inequívocas, inclusive para fins de transparência, confiabilidade e replicabilidade da pesquisa. Trata-se inclusive de erro comum a existência de ACs que não deixam claras ou explícitas as unidades de análise sendo verificadas.

O problema, provavelmente, surge da própria taxonomia empregada na análise de conteúdo, na qual diferentes autores fazem referência a diferentes nomes. A título de exemplo, Neuendorf (2002) fala de unidade de amostras, unidades de coleção e unidade de análise; Krippendorff (2004) as divide em unidade amostral, unidade de codificação e unidade de contexto, enquanto Bardin (2016) trata de unidades de registro e unidades de contexto. Outros autores, tendo análises léxicas, temáticas e semânticas em vista, chegam a tratar de unidades físicas, unidades de análise, unidades semânticas, unidades sintáticas, entre outras (*e.g.* BAUER, 2007; CERVI, 2016).

Como estamos tratando exclusivamente da análise de conteúdo categorial, é possível simplificar a questão, tendo apenas três divisões em vista. Conforme sugerido por Neuendorf (2002) e Krippendorff (2004), a primeira definição está na **unidade amostral**, afinal, quais serão as “porções” de texto ou de conteúdo a serem analisadas? Como dito anteriormente, aqui

estamos falando das unidades a serem amostradas (as regras de amostragem serão mais detalhadas na quinta seção deste capítulo) para a análise de conteúdo. Algumas ocorrem naturalmente, como *posts* de uma rede social digital, prontuários de pacientes, projetos de lei, leis, livros didáticos, artigos acadêmicos, trabalhos de conclusão de curso, teses, dissertações, programas de TV ou rádio, revistas semanais, fotografias, programas do horário eleitoral gratuito, editoriais de um jornal, notícias de um jornal, colunas de um jornal, diários, documentos, cartas, contratos, relatórios, filmes, pinturas, desenhos, ordens régias, leis, ofícios, relatórios, correspondências, anuários, alvarás, transcrição de discursos políticos ou de outras naturezas, entre tantos outros; enquanto outros serão conteúdos gerados pelos próprios pesquisadores, como entrevistas, grupos focais, diários de campo etc.

Naturalmente, a unidade amostral já definirá a unidade física (CERVI, 2016), ou seja, qual o meio físico do qual o conteúdo se originou, se era jornal impresso, rádio, TV, blogs, sites de redes sociais on-line, ou documentos impressos, por exemplo; não sendo, portanto, útil a separação taxonômica.

Note-se que determinadas pesquisas podem trabalhar com mais de uma unidade amostral, de acordo com seu interesse. Em pesquisas de jornalismo e de comunicação política, por exemplo, é muito comum a análise de conteúdo em jornais impressos, nas quais há, ao menos, quatro unidades amostrais distintas, nomeadamente capas, editoriais, colunas de opinião e notícias. Em certas pesquisas, o interesse se concentra exclusivamente na opinião do jornal, da empresa, então a unidade amostral é apenas o editorial (MONT'ALVERNE; MARQUES, 2019); em outras, a curiosidade já se localiza na cobertura diária jornalística, então a unidade amostral é a própria notícia, enquanto é possível ainda um foco no destaque dado em capas de jornal (FERES JÚNIOR, 2016), algo também bastante estudado nas revistas semanais. Não obstante, há trabalhos que comparam a cobertura jornalística em editoriais e em notícias; portanto, são duas unidades amostrais diferentes.

Por exemplo, vamos supor que um pesquisador está interessado em verificar se houve notável imparcialidade jornalística em um evento político de grande natureza, como é o caso do impeachment da Presidente Dilma Rousseff, em 2015. Essa pesquisa poderia incluir unidades amostrais de

capa, editoriais e notícias. Fazendo-se um recorte temporal, por exemplo, é natural que haveria mais notícias que editoriais, ou mesmo chamadas de capa a respeito do evento; portanto, são necessárias amostras diferentes. Em outro exemplo, estudos bibliométricos frequentemente lidam apenas com artigos publicados em periódicos científicos; mas há casos em que as pesquisas recolhem também *papers* apresentados em eventos, trabalhos de conclusão de curso, dissertações, teses e relatórios. Cada um desses é uma unidade amostral distinta. Cabe ao pesquisador, posteriormente, avaliar se as naturezas de cada unidade amostral são comparáveis entre si e são passíveis de serem subsumidas numa só.

Definida a unidade amostral, é preciso determinar a **unidade de análise** (também chamada na literatura de unidade de codificação), o elemento unitário de conteúdo a ser classificado (MORAES, 1999). Muitas vezes, a unidade de análise será idêntica à unidade amostral, ou seja, a análise de conteúdo se dará no texto/documento como um todo. Em análises bibliométricas, por exemplo, a unidade de análise frequentemente é o artigo acadêmico como um todo. Na comunicação política, como já citado, pode ser o editorial ou a notícia como um todo (CERVI; MASSUCHIN, 2013; FERES JÚNIOR, 2016; MONT'ALVERNE; MARQUES, 2019; PRUDENCIO; RIZZOTTO; SAMPAIO, 2018). Em análises de conteúdo em *posts* de redes sociais digitais, também o mais visto é a análise de cada postagem (AGGIO; REIS, 2013; ALVES, 2019; CARVALHO; MASSUCHIN; MITOZO, 2018; LYCARIÃO; DOS SANTOS, 2017; ROSSINI *et al.*, 2016).

Contudo, em determinados casos, pode ser que a análise do conteúdo como um todo possa não ser o mais recomendado, por ser um conteúdo complexo e/ou extenso. Imaginemos o caso de uma análise de conteúdo aplicada a filmes de longa-metragem. Ora, se um filme vai ser analisado como um todo, será bastante difícil se alcançar bons valores nos testes de confiabilidade, e a replicação do estudo será muito complexa. Se o desejo é analisar tal conteúdo como um todo, análises semióticas ou heurísticas de forma geral seriam mais adequadas. Em casos como esse, na análise de conteúdo, o mais indicado é a escolha por unidades de análise menores e bem definidas.

A identificação correta da unidade de análise é vital para replicabilidade e confiabilidade da análise de conteúdo. Os codificadores devem estar cientes da unidade a ser codificada ou, então, poderemos ter resultados distintos, porque cada codificador pode estar simplesmente codificando com base em trechos ou mesmo premissas diferentes.

Para facilitar essa compreensão, peguemos as pesquisas sobre horários gratuitos políticos eleitorais (HGPE). Normalmente, a unidade amostral é um programa inteiro, enquanto a unidade de análise utilizada por boa parte da literatura é o “segmento”, que é uma definição análoga a uma cena de programas audiovisuais. O segmento irá se alterar quando: a) o cenário mudar; b) os personagens mudarem e/ou c) a narração mudar. Então, em vez de analisar o programa televisivo como um todo, os pesquisadores aplicam as categorias de análise de conteúdo a cada segmento de cada programa de HGPE (CERVI, 2011; FIGUEIREDO *et al.* 1997; PANKE; CERVI 2012). Tal unidade de análise também já foi aplicada com sucesso em séries televisivas (SOUZA; ALMEIDA; DRUMMOND, 2017). Neuendorf (2002), por sua vez, sugere que a divisão pode ser simplesmente por tempo. Por exemplo, a codificação se dará em cima de cinco minutos, logo, cada cinco minutos do produto audiovisual será uma unidade de análise distinta.

Em determinados casos, a unidade de análise poderá ser definida pela própria organização do conteúdo original. Um bom exemplo é o caso de grupos focais. Ora, cada fala de cada participante de um grupo focal poderá funcionar adequadamente como a unidade de análise, o que, inclusive, pode gerar conhecimento por si (por exemplo, quantas vezes cada pessoa se manifestou no grupo focal e quão longas foram tais falas). Entretanto, pode ser que algumas falas sejam consideravelmente maiores que outras, o que pode dificultar a análise. Se uma pesquisa se detivesse sobre discursos de parlamentares, então estaríamos tratando de falas com vários minutos de duração, o que também seria um empecilho para codificações confiáveis e replicáveis.

De forma análoga, cada postagem em uma rede social on-line tende a funcionar como uma boa unidade de análise e é o mais visto na literatura; porém, pode não ser o suficiente quando há imagens e textos na mesma

postagem. Se o codificador está analisando a postagem como um todo, ele considera o texto e a imagem igualmente? E no caso dos conteúdos presentes no texto e na imagem serem contraditórios? Como é tomada a decisão? Haverá uma hierarquia?

Em todos esses exemplos, o mais adequado seria a opção por uma unidade de análise ainda menor ou mais específica. No caso da análise de textos, pode-se optar por definir o parágrafo ou mesmo a sentença como a unidade de análise, o que gerará a certeza de que os codificadores estão considerando os mesmos trechos do conteúdo para a codificação. No caso de leis, relatórios, livros ou textos igualmente extensos, essa pode ser uma saída apropriada.

Como no caso das unidades amostrais, também é possível se definir mais de uma unidade de análise em uma mesma pesquisa. No exemplo da postagem das redes sociais digitais, o texto da postagem pode ser uma unidade de análise e a imagem poderá ser outra. Consoante Bardin (2016), isso tenderá a permitir análises e comparações mais adequadas. O estudo de Prudencio, Rizzotto e Sampaio (2016), ao analisar notícias de jornais impressos, nessa linha, optou por considerar o texto da matéria como uma unidade de análise e as fotografias como outra unidade de análise. Ao final, eles possuem mais textos que imagens como unidades de análise, uma vez que nem toda notícia apresenta uma fotografia ou ilustração.

Uma alternativa pode se dar por meio da definição de *unidade de contexto*, ou seja, quanto um codificador leva em consideração do conteúdo original para a aplicação dos códigos. Para Krippendorff (2004, p. 103), as “unidades de contexto são unidades que delineiam o escopo de informação que os codificadores precisam consultar para caracterizar as unidades” de análise (*recording units*, no original). Em certos objetos, a unidade de contexto será idêntica à unidade de análise, como é o caso do estudo sobre notícias, ou um programa do HGPE. Enquanto é possível que haja referências a casos anteriores, não é recomendável que a unidade de contexto seja superior à unidade de análise. As inferências sobre essa possível complementaridade ou serialidade desses objetos seriam passíveis de ser verificadas após a finalização da AC.

Em outra pesquisa, a unidade de análise pode ser o parágrafo de livros didáticos destinados a crianças no ensino fundamental. Em variados momentos, é possível que a análise exclusiva do parágrafo não seja suficiente para a codificação, porque há referências a parágrafos ou mesmo capítulos anteriores. Todavia, o uso do livro como um todo como unidade de contexto pode, novamente, ter um impacto negativo na confiabilidade da codificação. Então, uma opção seria considerar apenas os últimos parágrafos como unidade de contexto, ou, no limiar, declarar o capítulo do livro como unidade de contexto. Já no caso da análise de uma conversação (digital ou presencial) ou de um grupo focal, talvez o mais indicado seja considerar o debate como um todo como unidade de contexto. Portanto, não há uma regra específica para essa definição, que deve se adequar aos objetivos da pesquisa. Há, todavia, uma dimensão ótima desejável:

É evidente que uma unidade de contexto alargado exige uma releitura do meio, mais vasta. Por outro lado, existe uma dimensão ótima, ao nível do sentido: se a unidade de contexto for muito pequena ou muito grande, já não se encontra adaptada; também aqui são determinantes quer o tipo de material, quer o quadro teórico (BARDIN, 2016, p. 134).

Dito de outra forma, no geral, unidades de contexto maiores permitem análises mais adequadas (ganho em validade); porém, exigem mais esforço e treinamento por parte dos codificadores, podendo, conforme vimos no 1º capítulo, gerar perda em confiabilidade (KRIPPENDORFF, 2004).

Box 1: exemplo de pesquisa com suas unidades definidas

A pesquisa de Felipe Borba sobre o uso de propaganda negativa durante a campanha presidencial de 2014 nos soa como um excelente exemplo de AC com diferentes unidades de análise. É sua intenção verificar, de modo amplo e complexo, como a propaganda negativa se deu nos mais diferentes meios midiáticos; logo, o “estudo incluiu todos os *spots* televisivos e radiofônicos veiculados pelos três candidatos, todas as postagens do Facebook durante o período de campanha, os oito debates organizados pelas emissoras Bandeirantes, Record, SBT e TV Globo, além de

todas as falas publicadas no jornal impresso ‘O Globo’ e no jornal televisivo ‘Jornal Nacional’, ambos de propriedade das Organizações Globo [...] A coleta dos dados ocorreu entre os dias 19 de agosto e 2 de outubro, no primeiro turno, e entre os dias 9 de outubro e 24 de outubro, no segundo turno, em período correspondente ao início e fim da propaganda no rádio e na televisão. No total, foram codificadas o total de 4.098 mensagens” (BORBA, 2019, p. 45). Assim, o autor faz a seguinte divisão para meio midiático e unidade de análise para AC:

- Spots TV – O Componente textual (imagens foram ignoradas), n=1033
- Spot Rádio – O componente textual, n=1297
- Debates – A fala do candidato em cada intervenção (pergunta, resposta, réplica e tréplica), n=450
- O Globo – A fala do candidato transcrita no jornal, n= 626
- Jornal Nacional – A fala do candidato exibida no telejornal, n= 133-
Facebook – O texto escrito na postagem. Não foi considerado o conteúdo compartilhado, n= 543 (BORBA, 2019, p. 45-46, adaptado do Quadro 1 original)

Portanto, mesmo sendo uma única pesquisa, Borba trabalha com diferentes meios, com diferentes unidades de análise, que, por sua vez, possuem diferentes tamanhos de população. Ele posteriormente realiza uma classificação das falas em a) aclamação, b) ataque ou c) defesa; e os tipos de ataques realizados, nomeadamente i) político, ii) pessoal e iii) misto, e observa as diferenças das estratégias das campanhas e dos candidatos para cada meio.

4. Criar e definir categorias

- A. elaboração do livro de códigos
- B. elaborar a planilha de codificação

A tarefa provavelmente mais importante na AC (categorial) está justamente na definição de um referencial de codificação (uma tradução para *coding scheme* ou, literalmente, esquema de codificação). “Ele é um conjunto de questões (códigos) com o qual o codificador trata os materiais, e do qual o codificador consegue respostas, dentro de um conjunto predefinido de

alternativas (valores de codificação)” (BAUER, 2007, p. 199). Os construtos analíticos podem ser derivados de: 1) teorias ou práticas existentes; 2) experiência ou conhecimento de peritos e pesquisa anterior (WHITE; MARSH, 2006).

Como já definido anteriormente no capítulo sobre epistemologia da AC, um texto não tem um significado único, estando aberto a inúmeras possíveis questões, embora “a AC interpreta o texto apenas à luz do referencial de codificação, que constitui uma seleção teórica que incorpora o objetivo da pesquisa” (*ibidem*, p. 199).

Então, o referencial de codificação vai determinar quais são as categorias e códigos a serem aplicados, assim como as regras para a codificação que deverão ser devidamente seguidas pelos codificadores. Portanto, esse referencial de codificação é materializado numa espécie de manual de codificação (*codebook*) ou simplesmente **livro de códigos**¹⁸. Falemos inicialmente sobre as várias regras para a criação de categorias e, posteriormente, sobre o livro de códigos.

Como já explicado, o código é a unidade elementar da AC. Ele irá resumir, filtrar ou condensar dados de acordo com os objetivos e com os interesses da pesquisa. Grupos de códigos, por sua vez, são agrupados em categorias, ou seja, unidades analíticas que materializam as questões a serem verificadas. Segundo Berelson (1952, p. 147), uma análise de conteúdo será produtiva na medida em que suas categorias forem claramente formuladas e bem adaptadas ao problema de pesquisa e ao conteúdo sendo estudado. Um exemplo bastante simples é o estudo das valências, muito visto na Ciência Política (CERVI; MASSUCHIN, 2013; FERES JÚNIOR, 2016), que busca verificar se a cobertura midiática é enviesada a favor ou contra determinado governo, tema ou ator político. A categoria é Valência e os possíveis códigos a serem aplicados são: 1. Negativo, 2. Positivo e 3. Neutro. Portanto, cada notícia,

¹⁸ Outra lacuna presente nos principais manuais disponíveis em português sobre a AC (BARDIN, 2016; FRANCO, 2005) está na não especificação clara sobre o que é um livro de códigos e sua importância. Eles apresentam regras e dicas para a criação de categorias e para a codificação, porém não elucidam a importância de que essas regras estejam formuladas e devidamente apresentadas em um livro de códigos.

editorial, telejornal etc. será definido como negativo, positivo ou neutro. Neste exemplo, uma tentação muito grande seria marcar dois códigos simultaneamente; afinal, uma matéria poderia se iniciar negativa para um candidato (realçado no título e no *lead*), mas fazer um contraponto positivo ao final, talvez não podendo ser classificada como neutra. A aplicação de dois códigos numa mesma categoria ou mesmo de categorias sobrepostas da mesma unidade de análise, entretanto, é algo que deve ser evitado.

Conforme bem resumido por Bardin (2016), as categorias de uma análise de conteúdo de boa qualidade devem ser: a) homogêneas, b) exaustivas, c) exclusivas, d) objetivas e e) adequadas ou pertinentes; algo que é bastante reproduzido pela literatura brasileira de AC. Todavia, as categorias devem ser objetivas (codificadores diferentes devem chegar a resultados iguais) e adequadas ou pertinentes (adaptadas ao conteúdo e ao objetivo) por questão de confiabilidade e validade, respectivamente; portanto, podem e devem ser avaliadas de outras maneiras. No que tange explicitamente às categorias, são apenas os primeiros itens que devem ser observados. Logo, argumentamos que as categorias de uma análise de conteúdo devem ser: 1) exclusivas, 2) exaustivas e 3) homogêneas, nesta ordem de importância.

Essas três características (denotadas não apenas por Bardin, mas basicamente por boa parte da literatura especializada, consoante BAUER, 2007; DOWNE-WAMBOLDT, 1992; NEUENDORF, 2002; KRIPPENDORFF, 2004; PRASAD, 2008; WEBER, 1990; WHITE, MARSH, 2006) se aplicam tanto às categorias no geral quanto aos códigos que as formam. Em outras palavras, também os códigos que formam uma categoria precisam ser exclusivos, exaustivos e homogêneos.

Em primeiro lugar, as categorias **precisam** ser mutuamente exclusivas. Isso significa que o mesmo conteúdo não é classificado de forma similar ou idêntica em diferentes categorias ou, mais frequentemente, que ele só pode ser classificado em um único código no interior de uma categoria. No caso da comparação entre categorias, se a mesma unidade de análise está sendo verificada de forma similar ou idêntica em diferentes categorias, isso tende a significar que a mesma coisa está sendo medida duas ou mais vezes sem necessidade. Porém, mais importante é ressaltar que as categorias não podem

ter elementos que se sobreponham ou sejam redundantes (CARLOMAGNO; ROCHA, 2016). Isso significa dizer que cada **código contido em uma categoria também é mutuamente exclusivo**. Portanto, no exemplo, uma mesma matéria não poderia ser classificada como tendo, simultaneamente, uma valência positiva e neutra. Caso o material em questão apresente característica dos dois códigos, o codificador deve tomar uma decisão, marcando o código preponderante (ou segundo instruções contidas no respectivo livro de códigos).

Stemler (2001) apresenta outro exemplo simples que evidencia a necessidade de categorias bem desenhadas. Ele apresenta um referencial de codificação que busca compreender cursos oferecidos numa escola divididos em cinco grupos: matemática, ciência, literatura, biologia e cálculo. Ora, um curso de biologia já naturalmente pode ser classificado tanto dentro de ciência quanto biologia, enquanto um curso de cálculo também poderia ser classificado em cálculo e também em matemática. Nesse exemplo, o único jeito é revisar o próprio referencial de codificação.

Por conseguinte, categorias e códigos exclusivos tendem a facilitar tanto melhores índices no teste de confiabilidade quanto uma maior chance de replicabilidade do estudo. Afinal, se o codificador pode decidir incluir o conteúdo em várias categorias e/ou códigos, isso tende a diminuir a necessidade de maior precisão na codificação e tende a aumentar a disparidade com outros codificadores (sejam aqueles presentes no mesmo estudo ou outros que desejam replicá-lo). A exclusividade, na visão de Krippendorff (2004), tende a expressar uma clareza em distinguir características do fenômeno sendo estudado. Logo, a criação de códigos como “ambíguo”, “híbrido”, “aplicável a mais de um código” tende a gerar categorias menos confiáveis, enviesar os resultados de pesquisa e revelar a falta de conceitos adequados para avaliar o conteúdo. Para além disso, a regra da exclusividade é vital para a realização posterior de testes estatísticos, uma vez que tanto as categorias quanto os códigos agora poderão, potencialmente, superar os 100%, o que acaba não se adequando aos testes mais tradicionais usados na AC (CAVANAGH, 1997; DOWNE-WAMBOLDT, 1992; RIFFE; LACY; FICO, 2014; STEMLER, 2001).

Em segundo lugar, as categorias e seus códigos precisam ser exaustivos. Em outras palavras, todo o conteúdo analisado precisa ser passível de codificação, de acordo com os interesses da pesquisa. Caso exista conteúdo pertinente que não possa ser devidamente codificado, isso significa que o referencial de codificação está inadequado ou incompleto e deveria ser revisado antes da codificação completa. Segundo Bardin (2016), é preciso:

esgotar a totalidade do 'texto'; é preciso ter-se em conta todos os elementos desse corpus. Em outras palavras, não se pode deixar de fora qualquer um dos elementos por esta ou aquela razão (dificuldade de acesso, impressão de não interesse), que não possa ser justificável no plano do rigor. (BARDIN, 2016, p. 126-127).

Como verificado na literatura, para garantir a exaustividade, é possível incluir o código “Outros” ou códigos de maneira similar. Afinal, “por mais precisas que sejam suas categorias, sempre há algum conteúdo que não se encaixa em nenhuma” outra (CARLOMAGNO; ROCHA, 2016, p. 180-181); ou mesmo para conteúdos raros, residuais que não ocorrem com frequência o suficiente para merecer um código respectivo (PRASAD, 2008), contudo, os autores advertem que a opção “outros” deve ser sempre residual, caso contrário é um indicativo de “problemas na criação e existência das suas categorias de classificação” (CARLOMAGNO; ROCHA, 2016, p. 181).

A esse respeito, voltemos ao exemplo de Stemler (2001) sobre a classificação de cursos oferecidos numa escola. As únicas opções apresentadas, inicialmente, são matemática, ciência, literatura, biologia e cálculo. Para além do fato de não serem mutuamente exclusivos, como já denotado, os códigos também não são exaustivos. Note-se que um curso de língua estrangeira, conforme o autor, não se encaixaria em qualquer das opções apresentadas, demonstrando que elas não são exaustivas. Pensemos no caso de um pesquisador da área de enfermagem que está avaliando os diferentes relatórios, protocolos, prontuários e documentos produzidos no interior de um posto de saúde, e que seja seu desejo avaliar qual a ocupação de cada indivíduo, produzindo um documento. Intuitivamente, esperaríamos médicos, enfermeiros, técnicos de enfermagem e agentes de saúde. Todavia, eventualmente, o pesquisador em questão poderia se deparar com

documentos assinados por profissionais de segurança, de limpeza ou mesmo de outros cargos administrativos da prefeitura em questão. Uma categoria exaustiva teria códigos que dessem conta de todos os tipos de profissionais que assinam documentos no posto em tela.

Finalmente, as categorias precisam ser homogêneas. Isso quer dizer que tanto as categorias quanto os códigos que as formam devem buscar realizar classificações homogêneas entre si. Em outras palavras, não é adequado que tenhamos categorias ou códigos com “tamanhos” muito heterogêneos entre si, apresentando lado a lado códigos muito abrangentes e muito restritos. Para Carlomagno e Rocha (2016, p. 180), as categorias “não devem ser tão amplas ao ponto de serem capazes de abarcar coisas muito diferentes em uma mesma categoria, sob pena de não ter significado prático para o estudo”. Os autores fazem uma citação de uma recomendação de Kenneth Bailey sobre a criação de tipologias, cuja reprodução nos parece adequada:

Nós geralmente procuramos minimizar a variância dentro do grupo, enquanto maximizamos a variância entre grupos. Isto significa que nós organizamos um conjunto de características em grupos, para que cada grupo seja tão diferente quanto possível de todos os outros grupos, mas cada grupo seja internamente homogêneo quanto possível. Ao maximizar ambos, homogeneidade intra-grupo e heterogeneidade entre-grupos, nós criamos grupos que são tão distintos (não sobrepostos) quanto possível, com todos os membros dentro de um grupo sendo tão iguais quanto possível (BAILEY, 1994, p. 1 *apud* CARLOMAGNO; ROCHA, 2016, p. 181)¹⁹.

Os autores ainda dão um exemplo que nos parece bem adequado. Eles apresentam o exemplo de uma pesquisa baseada na análise de conteúdo de uma campanha eleitoral, que busca identificar os temas mais tratados pelos candidatos. Nessa pesquisa, inicialmente, seriam apresentadas quatro opções de códigos: propostas de políticas públicas, construção da própria imagem, campanha negativa (ataques a adversários) e agenda (divulgação de eventos etc.). O primeiro código, segundo Carlomagno e Rocha (2016), estaria pouco homogêneo. Inicialmente, porque a grande maioria das

¹⁹ BAILEY, K. *Typologies and taxonomies: an introduction to classification techniques*. Sage Series on Quantitative Applications in the Social Sciences. California: Sage, 1994.

mensagens poderia ser classificada como tratando de políticas públicas e, segundo, porque existem diferentes tipos de políticas públicas (*e.g.* sociais, econômicas, de saúde) que poderiam ser mais exploradas em uma categoria com códigos mais homogêneos.

Bauer (2007) também salienta a importância de que as categorias sejam homogêneas internamente, no sentido de não verificar aspectos diferentes entre si. Então, ele exemplifica, não poderíamos ter um código para “vermelho”, ao lado de “pequeno”. Da mesma maneira, não deveríamos verificar na mesma categoria elementos diferentes como notícia de televisão e editorial; afinal, uma trata do meio de comunicação enquanto a outra lida com o tipo de material jornalístico a ser analisado. Seria mais correto, portanto, dividir as classificações em duas categorias diferentes: meio de comunicação (televisão, rádio, impresso, internet etc.), e formato das notícias (reportagem, entrevista, editorial etc.).

Observadas essas três regras principais, a tipologia adequada será materializada na criação de categorias e códigos de acordo com os interesses da pesquisa, com a pergunta de pesquisa e mesmo com possíveis hipóteses a serem testadas. Então, tal esquema ou referencial de codificação precisa ser transformado em um livro de códigos. Conforme seu sentido original em inglês *codebook*, trata-se de uma espécie de manual de codificação. Portanto, o livro de códigos (LdC) deve descrever as categorias e seus códigos detalhadamente, assim como as regras para a codificação, que devem ser seguidas explicitamente pelos codificadores²⁰.

Adaptando a sugestão de Macqueen *et al.* (1998, p. 32), um livro de códigos deve incluir, ao menos, seis componentes básicos: 1) categorias e seus códigos; 2) breve descrição; 3) definição completa; 4) regras para quando aplicar os códigos; 5) regras para quando não aplicar os códigos; 6)

²⁰ A título de exemplo, *vide* o LdC da pesquisa de Prudencio, Rizzotto, Sampaio (2018), que é bastante detalhado em suas explicações sobre a aplicação de cada código. Disponível em: <https://bdc.c3sl.ufpr.br/handle/123456789/27>.

exemplos²¹. Em resumo, deve conter definições claras, instruções fáceis de serem seguidas e exemplos que não deixem ambiguidade (WHITE; MARSH, 2006)²². As instruções devem ser escritas detalhada e cuidadosamente, e mesmo detalhes mundanos de codificação precisam ser especificados (NEUENDORF, 2002)²³; e os exemplos, por sua vez, ilustram a complexidade e a profundidade de pensamento que será necessário na codificação (WHITE; MARSH, 2006).

Box 2: Livro de códigos adaptado de Lycarião e Dos Santos (2017)

Livro de Códigos para a Codificação de Tweets com a hashtag #precisamosfalarsobreaborto

V. 01 - Identificação do Tweet [Tweet_ID]:

- Por favor, escrever o número de identificação do tweet de acordo com a ordem de coleta do mesmo.

V. 02 - Link do Tweet [Tweet_Link]:

- Por favor, inserir (copiar e colar) o endereço do *tweet*.

V. 03 - Texto do tweet [Tweet_Texto]:

- Por favor, copiar e colar o texto da mensagem do tweet.

V. 04 - Posicionamento da mensagem em relação à campanha [Posicionamento]

- Por favor, ler a mensagem contida no tweet e codificá-la em uma das categorias abaixo.

²¹ Outro exemplo é o da pesquisa de Lombard que analisa vídeos, que dá várias explicações detalhadas do que deve e do que não deve ser feito durante a codificação. Há especificações inclusive sobre o tempo que se deve tomar para analisar cada imagem/vídeo e sobre o que fazer em caso de dúvidas. Disponível em: https://academic.csuohio.edu/neuendorf_ka/content/codebooks/lombard_codebook.pdf.

²² Outro excelente modelo de LdC é o de Andreas M. Wüst e Andrea Volkens para analisar programas de governo de partidos durante as eleições. O LdC é tornado disponível online, mas também é oferecido um histórico da pesquisa, a forma como partidos e programas são selecionados e, claro, os procedimentos de codificação. Há um bom uso dos exemplos e de regras explícitas para ajudar na codificação. Disponível em: <http://www.mzes.uni-mannheim.de/projekte/manifestos/EMCodingInstructions.pdf>.

²³ Vide o livro de códigos de Neuendorf (2002) sobre personagens de filmes e TV, que chega a denotar: sotaque, cor do cabelo, pelo facial, altura, peso, uso de óculos, e mesmo nível de calvície. Disponível em: https://academic.csuohio.edu/neuendorf_ka/content/codebooks/neuendorf_codebook1.pdf

1 – Convergente

Por favor, codificar como ‘1’ todos os *tweets* que apoiam ou endossam a hashtag #precisamosfalarsobreaborto (criada pela Revista TPM como parte de uma campanha pela descriminalização do aborto). Por exemplo, mensagem contendo apenas a hashtag #precisamosfalarsobreaborto devem ser codificadas nesta categoria. Ademais, *tweets* que disseminam informação que ratifica os enquadramentos e opiniões relacionadas à descriminalização do aborto (o número de mulheres que morrem em clínicas ilegais, por exemplo) devem ser codificadas como ‘1’, após se verificar o contexto da mensagem do *tweet*. Ver exemplo abaixo:

“Fiz um aborto, não sou criminosa. Jandira poderia ser eu, poderia ser você. <http://t.co/EP2MTuPQda> #precisamosfalarsobreaborto”

Atenção: Ver *Figura 6* deste livro para verificar como o exemplo acima aparece na planilha de dados preenchida.

2 – Divergente

Por favor, codificar como ‘2’ os *tweets* que contenham opiniões e enquadramentos divergentes daqueles apresentados pela Revista TPM. Por exemplo, mensagens que contenham a hashtag #precisamosfalarsobrevida ou #precisamosfalarsobreassassinatodebebês devem ser codificadas nesta categoria, após se verificar o contexto da mensagem. Ver exemplo a seguir:

O Ex-BBB @jeanwyllys_real publicou uma foto com
#PrecisamosFalarSobreAborto - que vergonha Deputado!
#JeanWyllysNaoMeRepre...

Atenção: Ver *Figura 6* deste livro para verificar como o exemplo em questão aparece na planilha de dados preenchida.

3 – Ambíguo

Por favor, codificar como ‘3’ todos os *tweets* que não apresentam opiniões discerníveis ou que apresentam enquadramentos ambíguos em relação àqueles apresentados pela Revista TPM em sua campanha pela

descriminalização do aborto. Ademais, *tweets* que não apresentam o contexto da mensagem para verificar a relação com o perfil do criador do *tweet*, devido às configurações de privacidade, também devem ser codificadas nessa categoria. Ver exemplo a seguir:

“E você é contra ou a favor do aborto? #precisamosfalarsobreaborto
<http://t.co/al0DyPLGDg>”

Atenção: Ver Figura 6 deste livro para verificar como o exemplo em questão aparece na planilha de dados preenchida.

Conforme evidenciado no livro de códigos acima, o mais adequado é numerar as suas categorias e seus respectivos códigos. Isso facilita que a codificação possa ser realizada sem o auxílio de *softwares* complexos, podendo ser realizada em uma planilha de Excel, Calc ou *software* semelhante. O uso de uma planilha acaba sendo mais adequado que um editor de texto, como Word ou Writer, para facilitar os testes estatísticos posteriores (que serão mais detidamente abordados no Capítulo 3). Portanto, após a elaboração do livro de códigos, é igualmente importante que seja elaborado um formulário de codificação (FdC), que é a planilha de dados que será preenchida pelos diferentes codificadores, seja nos treinamentos, seja no teste de confiabilidade, seja na codificação final. Conseqüentemente, ele deve ser formatado para apresentar os espaços apropriados para a aplicação de todos os códigos possíveis nas devidas categorias, conforme o exemplo²⁴.

²⁴ Esta é uma pesquisa que busca estudar artigos sobre Democracia Digital publicados em periódicos. Na primeira coluna, deve ser preenchido a URL do artigo a ser analisado. Na segunda coluna, o nome do codificador (isso é importante para possíveis revisões e dúvidas ao longo da pesquisa). Em seguida, são retiradas várias informações do artigo, como o ano de publicação, seu título, palavras-chave, resumo e quantidade de autores/as. A partir daí, são extraídas informações de cada autor/a, como nome, instituição, escolaridade etc. O processo é repetido para cada coautor/a existente.

Figura 1 – Exemplo de FdC não preenchido de pesquisa (adaptado de LYCARIÃO e DOS SANTOS, 2017)

	A	B	C	D
1	TWEET_ID	Tweet_Link	Tweet_Texto	Posicionamento
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

Fonte: elaboração própria.

Atenção: Ver Figura 6 deste livro para visualizar a planilha de dados acima preenchida na fase final da codificação.

Segundo Neuendorf (2002), o objetivo de criar LdC e FdC é dar informação mais completa e menos ambígua possível, ao ponto de quase diminuir ao mínimo as diferenças individuais entre os codificadores. Portanto, LdC e FdC devem funcionar sozinhos como o protocolo de análise de conteúdo. Tal detalhamento também é importante para facilitar a validação dos dados da pesquisa em um primeiro momento, e mesmo para fins de replicabilidade posteriormente, como alertado por Macqueen e colegas:

Ao definir os códigos, não assuma que qualquer coisa é óbvia; sempre explicita especificamente o que código deveria ou não capturar. [...] Inclua todas as informações na definição completa do código e nas explicações para seu uso. Coisas que podem parecer óbvias aos codificadores em um certo lugar e momento podem ser totalmente obscuras em outras circunstâncias (MACQUEEN *et al.*, 1998, p. 35).

Naturalmente, é bastante raro se obter um livro de códigos adequado de imediato. Tanto o LdC quanto o FdC tenderão a sofrer inúmeras revisões até serem considerados adequados; porém, isso será abordado ainda neste capítulo.

5. Amostragem

A ideia de amostra ou amostragem é bastante presente não apenas nos estudos que envolvem análise de conteúdo, mas nos diversos tipos de pesquisas que se veem diante do desafio de produzir resultados significativos ou representativos a partir de uma seleção da população a ser pesquisada; ou seja, sem analisar a totalidade de unidades ou indivíduos dessa população. Nas palavras de Neuendorf (2002, p. 83), “amostragem é o processo de se selecionar um subconjunto de unidades para o estudo da população como um todo”¹⁰.

No caso, deve-se esclarecer que o termo população é utilizado, na literatura estatística, para se referir não apenas a conjuntos de indivíduos (humanos), mas a qualquer ser ou objeto que se almeja investigar. Ou seja, uma população pode ser constituída por instituições (*e.g.*, tribunais, prefeituras, secretarias etc.), documentos (*e.g.*, estatutos, legislações, regimentos etc.) e, claro, também por pessoas (usuários de postos de saúde, servidores públicos federais, habitantes de um país etc.).

Por isso, a ideia de amostra só se faz concebível e ponto de atenção quando os recursos disponíveis para realizar a pesquisa não são suficientes para se analisar todas as unidades da população. Quando isso é possível, ou seja, quando a população pode ser devidamente pesquisada em sua totalidade, no lugar de uma amostragem, realiza-se, então, o que se denomina de *censo*. Por isso, uma pesquisa *censitária* é aquela que coleta e analisa todas as unidades de uma população. Os censos demográficos do IBGE, por exemplo, propõem-se a realizar exatamente uma contagem e levantamento de toda a população brasileira.

Quando um censo não é possível devido a recursos insuficientes, o desenho de uma amostragem se faz necessário. Daí se questiona: como fazer com que os dados de apenas uma parte (os dados da amostra) possam falar sobre o conjunto de uma população?

Esse tipo de questionamento, de acordo com Krippendorff (2004, p. 112), ganhou proeminência quando pesquisadores precisaram gerar técnicas que pudessem mostrar as preferências da opinião pública de países com milhões de

habitantes, entrevistando apenas alguns milhares desses cidadãos. Tal tipo de necessidade se fez e se faz especialmente forte em períodos eleitorais, quando é preciso construir sondagens semanais sobre as preferências do eleitorado em relação aos candidatos. No caso, seria absolutamente inviável entrevistar todos os eleitores de um país a cada semana.

Essa necessidade impulsionou, então, todo um ramo de pesquisas de sondagens de opinião pública, as denominadas *surveys*, a se utilizarem de princípios da estatística inferencial (ver Capítulo 3, Seção 4) para selecionar uma pequena parte da população que pudesse ser **representativa** do conjunto total dessa mesma população. O acúmulo gerado por esse ramo de pesquisas tem produzido técnicas cada vez mais sofisticadas de amostragem para se reduzir os vieses da seleção e, portanto, diminuir o erro amostral, ou seja, a diferença entre os resultados provenientes da amostra e os da população como um todo (ver FOWLER, 2013; MOSER, KALTON, 2017).

Nosso objetivo aqui, naturalmente, não é discorrer sobre o desenvolvimento, nem mesmo sobre o estado atual dessas técnicas de amostragem para pesquisas de opinião pública, mas apenas ilustrar que, a depender da natureza do fenômeno pesquisado, desenhos complexos de amostragem são necessários para se chegar a uma base de dados com alto nível de confiança e um baixo erro amostral²⁵. Assim, o conjunto desses elementos (quantas unidades serão e como serão selecionadas) visa a garantir a representatividade da amostra. A esse respeito, pode-se dizer que uma amostra é representativa no momento em que “ela leva a conclusões que são aproximadamente as mesmas que aquelas que se alcançaria caso toda a população fosse pesquisada” (KRIPPENDORFF, 2004, p. 112).

²⁵ O nível de confiança se refere a uma medida estatística diretamente dependente da margem de erro amostral. Isso porque margens extensas permitem que o nível de confiança da amostra seja elevado. Uma amostragem com 95% de confiança e 3% de margem de erro, por exemplo, implica que em 100 repetições do mesmo método de amostragem, em 95 delas, os resultados estariam dentro da margem de erro, ou seja, 3% acima ou abaixo do encontrado. Se aumentarmos a margem, o nível de confiança irá subir. Contudo, a mensuração da amostragem será menos precisa. Para se elevar o intervalo de confiança alto (acima de 95%) e se garantir uma pequena margem de erro (ou seja, uma amostragem relativamente mais confiável e precisa), faz-se necessário aumentar o tamanho da amostra (ver CERVI, 2017, p.135-137).

A atribuição de representatividade de uma amostra requer, em linhas gerais (a aplicação efetiva terá grande variação), uma seleção probabilística (com algum teor de aleatoriedade) das unidades de uma população. Contudo, essa seleção probabilística, conforme esclarece Krippendorff (*idem*), embute uma série de premissas de teoria estatística que dificilmente se aplicam plenamente à análise de conteúdo.

Isso implica que, por mais que a amostra de uma análise de conteúdo siga rigorosamente as técnicas de seleção probabilística mais adequadas para a população selecionada e aplique procedimentos que visem a reduzir vieses, ainda assim, é muito provável que a representatividade não possa ser assumida. Conforme Krippendorff (2004, p. 110-113), isso ocorre pelas seguintes constatações:

1ª: Na análise de conteúdo, não há necessariamente uma indivisibilidade das unidades coletadas, pois a maior parte dos textos e materiais analisáveis pode ser dividida em diversas unidades e subunidades analíticas. Um filme, por exemplo, pode ter seu conteúdo codificado tendo como unidade de codificação a sequência, a cena ou até mesmo sua menor unidade possível, no caso, o quadro (*frame*).

2ª: Como consequência da primeira constatação, ocorre que raramente as unidades da amostra coletada são as unidades quantificadas, o que viola justamente a expectativa da teoria da representação estatística tradicional, a qual requer que a unidade coletada seja a unidade quantificada. Assim, uma pesquisa que se propõe a coletar e analisar o conteúdo de filmes ou documentos (unidade de coleta) terá como unidade de codificação algo menor, ou seja, sequências (dos filmes) ou sentenças (dos documentos). Como essa seleção pode se dar em diferentes níveis de grandeza, os resultados contabilizados tendem a apresentar essa seleção e grandezas. Não irão, portanto, necessariamente representar estatisticamente as unidades coletadas (ou seja, os filmes, os documentos) em sua totalidade.

3ª: Nas sondagens de opinião (*surveys*), cada indivíduo possui a mesma relevância e deve ser selecionado com algum nível de aleatoriedade. Uma vez selecionados, cada indivíduo possui o mesmo valor informativo. Na análise de conteúdo, por outro lado,

não é incomum que certas unidades possuam mais relevância (valor informativo) para a pesquisa do que outras unidades. Assim, por exemplo, uma notícia de um portal jornalístico que se encontra entre as mais lidas pode ter mais valor informativo do que as notícias sem visualizações. Essa relevância, claro, será qualificada pelos interesses da pesquisa. Caso ela almeje identificar as preferências dos leitores de um portal, a categoria “mais lidas” apresenta mais valor informativo do que notícias sem quantidade relevante de visualizações. Porém, caso a pesquisa almeje mensurar as preferências temáticas do veículo (ignorando, portanto, as preferências das audiências), então todas as notícias tendem, a princípio, a ter a mesma relevância. Contudo, a temática de uma notícia colocada na primeira página do portal, com vídeo e com grande quantidade de palavras, possui a mesma relevância do que as temáticas de notícias que estão na parte inferior do portal, contendo apenas alguns parágrafos, sem qualquer elemento visual? Tais diferenças não deveriam ser levadas em consideração como indício de “temática preferencial” dada pelo veículo jornalístico?

4ª: Como os exemplos acima ilustram os critérios de relevância, ao serem algo esperado e legítimos numa análise de conteúdo, acabam, contudo, violando mais um dos pressupostos da teoria da representação estatística tradicional. Com isso, não se pode atribuir à amostragem um poder de representação (no sentido de espelhamento) de uma população, pois as características da amostra não irão necessariamente espelhar, em proporções muito similares, um conjunto de características que estão presentes na população como um todo.

As constatações de Krippendorff, acima, ajudam, em seu conjunto, a lidar com mais realismo em relação às três questões que, segundo Bauer (2002, p. 196), qualquer pesquisa que utiliza análise de conteúdo precisa responder; essas três questões são: (a) a da representatividade, (b) a do tamanho da amostra e, por fim, (c) a da diferença entre unidade de coleta (amostragem) e de codificação. No caso, as constatações de Krippendorff (2004, p. 113) demonstram que o problema da representatividade de uma análise de conteúdo deve ser colocado em perspectiva do interesse da pesquisa e da natureza do fenômeno analisado (a produção de sentido) e que, mais relevante do que se tentar criar uma amostra representativa, é construí-

la de modo a tornar possível a existência de resultados significativos para se responder ao interesse da pesquisa. Isso reforça, como vimos anteriormente (no primeiro capítulo, Seção 1.5), a importância da replicação e da revisão dos resultados por outras pesquisas, pois as escolhas que constituem um método de codificação são passíveis de críticas e reformulações.

Por isso, a utilização de técnicas de amostragem por parte da análise de conteúdo, mais do que representatividade, é necessária para se chegar a uma metodologia que dê a mesma chance para que resultados positivos e negativos ao teste das hipóteses possam aparecer. Em outras palavras, deve-se realizar um desenho de amostragem que não privilegie um resultado ou conclusão de antemão. Do contrário, quando a chance, pelo próprio método de seleção das unidades, descarta previamente aquelas unidades que geram resultados contrários aos de determinada perspectiva teórica, pode-se dizer que há um problema de viés nos critérios de composição da amostra.

Para evitar isso, é necessário que o desenho da pesquisa apresente e justifique os critérios de seleção das unidades de acordo com uma ou mais técnicas de amostragem. No caso, iremos apresentar aquelas mais utilizadas. Tais técnicas são divididas por Neuendorf (2002, p. 83-88) em dois grandes grupos, um de natureza probabilística (randômica) e outro de natureza não probabilística (não randômico). Iremos, a partir de uma síntese dos manuais de Neuendorf (2002) e Krippendorff (2004), apresentar as principais técnicas de cada grupo. Ressaltamos que, comumente, essas técnicas são usadas em combinação; a apresentação em separado de cada técnica se deveu a meros fins didáticos.

5.1 Técnicas probabilísticas (randômicas)

Amostragem aleatória simples

Nesse tipo de amostragem, as unidades devem ser sorteadas a partir de uma lista contendo todas as unidades. Para tornar esse sorteio mais viável, pode-se tabular a lista em algum programa ou pacote de programação para

processamento e análise de dados (*e.g.*, Calc, Excel, Projeto R²⁶ etc.) e utilizar o comando para se gerar um ordenamento aleatório da lista ou uma amostra sob mesmo critério.

Amostragem sistemática

A amostragem sistemática consiste em selecionar as unidades de uma população a partir de determinado intervalo. Por exemplo, a cada três pessoas de uma fila de posto de saúde, uma é selecionada para entrevista. Tal tipo de técnica pode não ser desejável para fenômenos que sejam de natureza cíclica. Caso um intervalo de quatro anos de notícias de jornais seja selecionado para uma AC, deve-se levar em conta se o intervalo selecionado não possui, ele mesmo, um viés de sub ou sobreamostragem. Se a AC de conteúdo quiser, por exemplo, analisar notícias políticas, caso o ano inicial da amostragem seja um ano eleitoral, haverá uma tendência natural de que mais notícias sobre política apareçam, pois o intervalo selecionado só irá selecionar anos de disputa eleitoral (assumindo, nesse caso, que as eleições ocorrem a cada quatro anos). Assim, a amostragem terá esse viés, sendo esperado que os outros anos do intervalo possuam uma quantidade e perfil de notícias diferente do selecionado.

Amostragem por agrupamento

A seleção por agrupamento ocorre geralmente por questões logísticas, especialmente quando não é possível obter uma lista com todas as unidades da população para sorteá-las ou selecionar unidades sistematicamente. Em tais circunstâncias, faz-se oportuno selecionar grupos de unidades que já estão agregadas previamente à coleta. Exemplos comuns são arquivos de documentos (não digitalizados) ordenados alfabeticamente, sendo possível o sorteio de uma ou mais letras, cada letra correspondendo, portanto, a um agrupamento. No caso de jornais e outros documentos ordenados cronologicamente, uma semana também se constitui como um agrupamento usual.

²⁶ Para este pacote, recomendamos o manual de Aquino (2014), “R para cientistas sociais”, disponível em: http://www.uesc.br/editora/livrosdigitais_20140513/r_cientistas.pdf.

Amostragem estratificada

A estratificação de uma amostragem ocorre quando suas unidades são coletadas ou manipuladas (após a coleta) com base em algum critério que subdivida a amostra em diferentes estratos. Assim, estatutos de instituições podem, por exemplo, ser divididos por região geográfica do país. Se certa região possuir três vezes mais o número de instituições do que outra, a amostragem pode ser estratificada de modo a espelhar essa diferença, ou seja, selecionando-se aleatoriamente três vezes mais unidades da região correspondente. Tal tipo de estratificação também pode ser realizado após a coleta, quando parte das unidades é aleatoriamente descartada para gerar proporções correspondentes às características da população como um todo. Esse tipo de procedimento é comum em sondagens (*surveys*), pois frequentemente a proporção de mulheres respondentes acaba se mostrando significativamente maior do que a de homens; maior ainda, no caso, do que a existente na população como um todo. Para tornar a amostra mais próxima da proporcionalidade existente na população (conhecida por meio dos censos do IBGE), faz-se uma correção da amostra, eliminando-se aleatoriamente respostas de mulheres, até que a amostra reflita as proporções identificadas na pesquisa censitária. Tal tipo de procedimento também é utilizado para outras características, como renda, faixa etária, religião etc.

Amostragem em múltiplos estágios

Esse tipo de amostragem ocorre quando a coleta se dá em diferentes estágios, ou seja, quando se realiza uma coleta de uma coleta já realizada. Um exemplo é ofertado por Lycarião e Dos Santos (2017), em que inicialmente uma coleta de *tweets* de uma campanha foi realizada pelo aplicativo NodeXL, resultando em 1010 *tweets*. Desses, uma nova amostragem de 376 *tweets* foi realizada para uma análise de conteúdo. Há pelo menos duas situações em que realizar uma coleta dentro de (ou após) uma coleta é oportuna. Uma é quando se deseja estratificar a amostra para um conjunto específico de unidades (ver amostragem estratificada acima). Outra ocorre quando o número de unidades coletadas inicialmente for superior à capacidade de

análise da equipe de codificadores, sendo necessário, portanto, proceder a novas amostragens, até que se chegue a uma quantidade significativa e viável de análise (ver seção abaixo sobre tamanho da amostra).

Amostragem em bola de neve

Uma terceira situação em que uma amostragem em múltiplos estágios se mostra oportuna é quando se deseja identificar fluxos precedentes de influência ou de ideias, conceitos etc. Para isso, a amostragem bola de neve é comumente utilizada. Ela consiste em, a partir de uma unidade, coletar e analisar outras unidades que foram citadas ou indicadas pela primeira unidade. No caso, isso ocorre geralmente quando se pretende identificar, coletar e analisar as fontes bibliográficas de documentos (artigos, livros) citados em um artigo, livro etc. Ou seja, acaba sendo também um método de revisão bibliográfica adequado para saber as bases e o desenvolvimento de construtos que precederam e influenciaram certa descoberta ou inovação científica. Outro uso da técnica bola de neve se refere a identificar líderes de opinião ou fluxos de influência sobre alguma decisão cotidiana, tal como sobre qual filme assistir no cinema ou qual candidato votar na eleição. No caso, ao se entrevistar participantes de uma pesquisa, pode-se perguntar sobre alguém que influencia esse tipo de decisão e, caso um conjunto de pessoas sejam citadas por várias outras, chega-se, então, a líderes de opinião. Esses, por sua vez, são entrevistados sobre quem os influencia, construindo-se, com isso, uma cadeia hierárquica de líderes e opinião (ver KATZ; LAZARFELD, 1955).

5.2 Técnicas não probabilísticas (não randômicas)

Enquanto os métodos probabilísticos atribuem, em alguma medida, ao acaso (à aleatoriedade) a escolha das unidades que irão compor a amostragem, a escolha não probabilística realiza essa seleção tendo por base outros critérios. Veremos a seguir quais são eles e como eles determinam diferentes tipos de amostragem não probabilística.

Amostragem por conveniência

Esse tipo de abordagem se justifica fundamentalmente pelas limitações de recursos da pesquisa ou da própria impossibilidade de acesso aos dados de uma maneira mais ampla e completa. Tal tipo de escolha acaba sendo necessária quando se pretende analisar, por exemplo, políticas públicas ou fenômenos nacionais, mas os pesquisadores só possuem acesso à realidade (pessoas e documentos) local dessas políticas ou fenômenos. Daí que a amostragem por conveniência acaba sendo necessária, pois seria inconveniente, ou até mesmo inviável, ter acesso aos usuários de uma política pública ou a seus documentos em todo o território nacional. Porém, novamente, pode-se combinar esse método de amostragem não probabilístico com algum de tipo probabilístico, em que a escolha das unidades, ainda que restrita ao âmbito local por conveniência, pode ser selecionada a partir de uma seleção aleatória simples, sistemática, por bola de neve etc.

Outra situação comum em que a conveniência se estabelece é a forma com que participantes são entrevistados. Uma fila de ônibus ou de qualquer outro tipo de serviço (banco, de posto de saúde, do INSS) oferece uma oportunidade não apenas conveniente, mas também, em certa medida, bastante oportuna para se pesquisar as perspectivas dos usuários de uma política pública, justamente no momento da utilização desses serviços.

Por propósito ou relevância

Como visto no começo deste capítulo, materiais textuais não são unidades indivisíveis como um indivíduo. Daí que a análise de conteúdo, muitas vezes precisa, pelo propósito da pesquisa e pela natureza mesma do tipo de população estudada, descartar boa parte do material do texto inicialmente coletado, justamente por ele não se constituir como parte do objeto da pesquisa e, portanto, por não possuir relevância dentro do escopo de interesse previamente definido. Desse modo, análises de conteúdo sobre discursos de autoridades políticas sobre determinado assunto, por exemplo, tenderão a analisar apenas os trechos dos discursos que se referem ao tema de interesse da pesquisa, descartando os demais. Assim, percebe-se que esse

tipo de amostragem é bastante comum à análise de conteúdo, mesmo quando combinado com outras técnicas aqui discutidas.

Por cotas

Esse tipo de amostragem é oportuno quando se almeja construir um questionário para ser aplicado numa sondagem (usando, em tal momento, de técnicas probabilísticas). Assim, seleciona-se uma quantidade de participantes com características da população a ser estudada. Dessa forma, busca-se, por exemplo, ter uma cota de pessoas com renda média, alta e baixa, de pertencentes às religiões mais populares do país etc. Com essas cotas, as entrevistas podem ser realizadas com questões abertas e cujas respostas, após serem submetidas a uma análise de conteúdo, e ajudarão a compor os itens do questionário. Essa composição irá privilegiar, no caso, as respostas mais frequentes contidas nas respostas dos participantes selecionados por cotas.

Outra situação em que as cotas são oportunas se trata de grupos focais, especialmente quando se almeja fazer com que os grupos apresentem diferentes pontos de vista decorrentes da classe social, religião, gênero, orientação sexual etc.

Censo

O censo é uma forma não probabilística de análise de uma população, pois ele estabelece que todas as suas unidades serão coletadas e analisadas. A rigor, nem faz muito sentido tratar “censo” como um tipo de amostragem, pois a ideia de amostra pressupõe um recorte, uma exemplificação, o que não é o caso de um censo. Esse tipo de análise requer grandes somas de recursos humanos e de tempo para ser realizada quando envolve grandes populações. Entretanto, há situações em que censos são plenamente possíveis, mesmo com poucos recursos. Imagine que uma universidade elabore um projeto de extensão numa escola para testar um novo método de alfabetização, até, então, não experimentado. Caso todas as provas aplicadas aos estudantes para testar a eficiência do método sejam submetidas a uma

análise de conteúdo, pode-se dizer que essa análise foi censitária, pois todas as unidades da população definida pela pesquisa foram analisadas.

5.3 Tamanho da amostra

Um dos desafios mais recorrentes no planejamento de uma análise de conteúdo se refere à definição do tamanho da amostra. O primeiro aspecto que configura esse desafio deriva da própria dificuldade de se aplicar os pressupostos da estatística inferencial que permitem atribuir a uma amostra a capacidade de representar os parâmetros, com certo nível de confiança e dentro de um erro amostral pré-determinado, da população estudada. Trata-se, no caso, de atribuir aos resultados da amostra a capacidade de apresentar características a todo o conjunto de unidades que não foram efetivamente analisadas.

Entretanto, tal desafio (o da representação) não deve ser necessariamente uma aspiração epistemológica de qualquer análise de conteúdo. Como vimos anteriormente, o fato de uma amostra não poder representar uma população mais ampla não retira dos resultados alcançados significância ou validade para testar constructos teóricos. Por isso que, mesmo quando não é possível atribuir a qualidade de representação a uma amostra, é preciso avaliar se o seu tamanho é suficiente para permitir que os testes estatísticos a serem aplicados ao final da codificação (a exemplo do qui-quadrado, ver terceiro capítulo desta obra) alcancem alguma chance de apresentar resultados significativos do ponto de vista estatístico. Isso precisa ainda levar em conta que, mesmo quando significativos, tais resultados podem advir de amostras que, por contarem com poucas unidades, diminuem a chance de esses resultados serem encontrados por outros estudos (BUTTON *et al.*, 2013). Por isso que, sempre que possível do ponto de vista da viabilidade da pesquisa, deve-se privilegiar amostras com mais unidades, mesmo que seja necessário, posteriormente, descartar parte delas por motivo de correção (ver seção deste capítulo sobre **amostragem estratificada**).

Contudo, a depender da natureza do material analisado, o estabelecimento de uma grande quantidade de codificação pode implicar

desperdício significativo de tempo e de recursos. Para evitar isso, recomenda-se a utilização de calculadoras (algumas disponíveis on-line) que determinam o intervalo de confiança e a margem de erro para cada tamanho de amostra, ou estabelecem o tamanho da amostra a partir da indicação dessas medidas (para melhor compreender tais noções, ver Capítulo 3, Seção 4.1). Deve-se, ao tomar a decisão sobre o tamanho, buscar a combinação mais próxima possível com um alto intervalo de confiança (95% ou acima) e uma margem de erro geralmente aceitável (igual ou inferior a 3%).

Contudo, a definição do tamanho da amostra, além da combinação entre intervalo de confiança e margem de erro, dependerá de outros aspectos, tais como o tipo de teste estatístico a ser aplicado, o nível de homogeneidade da amostra (do seu desvio-padrão) e os propósitos da pesquisa (ver CERVI, 2017, p. 134-137; KRIPPENDORFF, 2004, p. 122). Por isso, recomenda-se que, sempre que possível, tanto o desenho da amostra (incluindo seu tamanho) quanto a aplicação e a interpretação de testes estatísticos sejam assessorados e supervisionados por especialistas em estatística.

6. Pré-teste das categorias e das regras de codificação

- a. treinamento
- b. revisão do livro de códigos
- c. teste de confiabilidade-piloto

A seção anterior se preocupou em apresentar as principais regras a serem observadas na criação de categorias e de seus códigos; entretanto, nenhum livro de códigos tende a ser confiável e válido sem a devida avaliação. Um livro de códigos precisa ser testado antes de a codificação final ser realizada, e o mais recomendado é que seja revisado algumas vezes antes dela ocorrer.

Em um momento inicial, há inúmeras maneiras de se revisar e incrementar um referencial de codificação. O pesquisador pode apresentá-lo a *experts*, parceiros de pesquisa ou mesmo aos pares para receber críticas qualificadas. É importante que a própria equipe de pesquisa e a equipe de codificação (caso sejam diferentes) estejam diretamente envolvidas nesse

processo de revisão do livro de códigos. Esse é o momento para tornar conceitos e explicações mais claros e explícitos. Também é o momento para se avaliar a qualidade e utilidade das categorias e códigos, que podem eventualmente ser descartados ou mesmo refeitos. Por exemplo, uma categoria pode ser vista como excessivamente ampla e ser dividida em duas ou mais derivações. Consoante Macqueen e colegas (1998, p. 36), “jogue fora os códigos que não funcionem e trabalhe em redefinições para códigos problemáticos. Alguns códigos podem capturar a especificidade de uma resposta única em vez de padrões e temas gerais”.

Os codificadores devem ser incentivados a apresentarem suas dúvidas e questionamentos sobre as categorias e regras de codificação. Caso haja a confirmação de se tratar de um problema, o mesmo já pode ser corrigido imediatamente no livro de códigos. Não obstante, acreditamos que a revisão do livro de códigos deva acontecer em paralelo ao treinamento dos codificadores. Neuendorf faz uma apresentação detalhada das diversas etapas nas quais o livro de códigos pode ser revisado e como o treinamento dos codificadores está intrinsecamente ligado a isso:

- Escrever um livro de códigos com variáveis e categorias válidas para o fenômeno de interesse da pesquisa.
- Treinamento dos codificadores com discussão.
- Codificadores praticam em conjunto, engajando em discussões que buscam construir consensos.
- Possíveis revisões do livro de código.
- Treinamento dos codificadores nas revisões,
- Codificadores praticam a codificação de modo independente em um número de unidades que represente a variedade da população estudada.
- Codificadores discutem os resultados da prática de codificação.
- Possíveis revisões do livro de códigos.
- Treinamento dos codificadores nas revisões

- Codificadores fazem uma codificação-piloto em uma amostra para testar a confiabilidade.
- Pesquisador verifica confiabilidade mediante uma operação matemática que pondere a chance aleatória de concordância.
- Possíveis revisões ao livro de códigos.
- Treinamento dos codificadores nas revisões
- Codificação independente final (incluindo checagens de confiabilidade). Nessa checagem, o cálculo do nível de confiabilidade deve ponderar a chance aleatória de concordância.
- Relato dos codificadores de suas experiências (NEUENDORF, 2002, p. 134).

Em outras palavras, toda vez que o livro de códigos é revisado, um novo treinamento com os codificadores deve ser realizado. Esse processo é repetido tantas vezes quantas forem necessárias²⁷. Como parte do processo de treinamento, o livro de códigos pode ser revisado até que os codificadores estejam confortáveis com o esquema de codificação. As revisões podem ser realizadas até que a codificação final comece.

De forma resumida, por mais detalhado, bem explicado e bem definido que seja o livro de códigos, não se deve assumir que as categorias e que a categorização soarão como óbvios aos codificadores e que a codificação será fácil (WHITE; MARSH, 2006). Então, o treinamento dos codificadores busca justamente permitir a melhor compreensão possível das categorias e das regras de codificação por parte dos codificadores. O treinamento, então, tem quatro funções principais: 1) dirimir dúvidas sobre códigos; 2) aumentar as perícias dos codificadores em fazer a aplicação correta do LdC; 3) nivelar o nível de codificadores entre si; 4) se necessário, ajudar na revisão do livro de códigos. Para Neuendorf (2002), o treinamento envolve padronizar as técnicas de codificação dos codificadores. Seus métodos de codificação precisam ser

²⁷ Isso, claro, tendo uma pesquisa ideal como parâmetro. Evidentemente, no contexto brasileiro, haverá constrangimentos de tempo e demanda para a entrega de resultados de pesquisa. Em diversos casos, será impossível se repetir o processo até o nível adequado. Caberá ao pesquisador ou à pesquisadora-líder avaliar qual o nível mínimo aceitável para se prosseguir em cada etapa.

calibrados para que eles vejam o conteúdo da mesma forma, para poderem codificar independentemente, sem discussão ou colaboração depois.

Contraditoriamente, a maior parte dos manuais afirma que o treinamento dos codificadores é uma etapa muito importante da AC, mas é igualmente evasiva em explicar como deve ser esse treinamento dos codificadores. Não há uma “receita de bolo” aceita e as descrições sobre o treinamento são, frequentemente, genéricas. Não obstante, apresentamos aqui algumas boas práticas e dicas que conseguimos extrair de tais manuais e também de nossas próprias pesquisas.

Krippendorff (2004), por exemplo, sugere o método de autoaprendizado. Programadores treinam a codificação e depois recebem um gabarito com as respostas “certas” (feitas por uma equipe de *experts*). A comparação dá um *feedback* imediato à *performance* do codificador e permite que ele se adapte ao padrão de interpretação das instruções. Também permite que os codificadores mais adequados sejam escolhidos. Entretanto, isso pode tornar o processo extremamente lento, uma vez que nem sempre há *experts* disponíveis para avaliar uma codificação, mostrando-se, portanto, como uma opção irrealista.

O método sugerido por Hak e Bernst (1996) no artigo *Coder Training: theoretical training or practical socialization?* parece-nos mais próximo do que é feito na práxis da AC em diferentes lugares. No caso, eles deram aos codificadores as instruções sem explicação formal. Foram codificadas 40 unidades de análise e obtidas várias respostas diferentes. O pesquisador marcou uma reunião para discutir as diferenças encontradas. Uma segunda codificação do mesmo material foi realizada e ainda encontrou resultados diferentes. Uma segunda reunião foi realizada com instruções ainda mais detalhadas. Alcançou-se um resultado de confiabilidade alto o suficiente, mas ainda houve uma discussão aprofundada sobre os detalhes do processo, algo pouco realizado em outros lugares.

Riffe, Lacy, Fico (2014) sugerem que a melhor ideia seria, inicialmente, familiarizar os codificadores com o conteúdo a ser codificado (ver também KRIPPENDORFF, 2004), o que deve ser feito com conteúdo similar, mas não idêntico ao que será codificado na pesquisa. Depois, o livro de códigos deve

conter instruções explícitas do máximo de conteúdo a ser codificado por sessão de codificação, e mesmo regras para a releitura de todas as regras de codificação antes de sua realização. Em seguida, os codificadores precisam se familiarizar com o LdC e discuti-lo tanto com o supervisor quanto com outros codificadores nas sessões de treinamento e lidar com os problemas vistos na codificação. É importante verificar se os codificadores não estão codificando a partir de perspectivas diferentes. Para os autores, é importante, nessa etapa, compreender se há discordância entre os codificadores porque as categorias são ambíguas, mal desenhadas, ou simplesmente porque um ou outro codificador não consegue compreender o conceito para operacionalizá-las. Eles dizem que, se vários codificadores discordam de certa categoria, a maior chance é que ela seja o problema.

Para facilitar, com base em nossas próprias experiências, apresentamos uma sugestão de como realizar o treinamento dos codificadores.

I. Explicações teóricas e metodológicas detalhadas

O pesquisador sênior e/ou líder da pesquisa apresenta o livro de códigos em detalhes. Essa apresentação tende a funcionar melhor com a leitura coletiva do livro de códigos. Os codificadores devem retirar todas as dúvidas neste momento. Idealmente, essa reunião acontece com o líder da pesquisa e com todos os codificadores presentes. No caso de alguém participar por ferramenta de comunicação on-line, é preciso cuidado extra para assegurar que houve uma compreensão geral adequada do livro de códigos. O formulário de codificação já deve ser apresentado e enviado aos codificadores. Macqueen e equipe (1998) acreditam que codificadores só podem lidar razoavelmente com 30-40 categorias por vez. Portanto, se o livro de códigos for superior a isso, o ideal é que a codificação seja realizada por estágios.

II. Familiarização com o conteúdo

O pesquisador-líder deve providenciar excertos do conteúdo para familiarizar os codificadores com o tipo de conteúdo a ser analisado. Isso já pode ocorrer na primeira sessão de treinamento, junto com o passo 1. De toda forma, os codificadores tenderão a precisar de tempo extra para ler e se familiarizar com o conteúdo providenciado.

III. Codificação inicial independente

Os codificadores fazem uma primeira codificação inicial independente em um excerto de material (no mínimo, 10 unidades), que deve, a cada treinamento, ser diferente e o mais representativo possível da diversidade do conteúdo a ser analisado. Idealmente, cada codificador já trabalhará no local escolhido para a codificação (*e.g.* escritório, sala do grupo de pesquisa, ou casa). Os codificadores não devem se comunicar entre si e nem com o líder da pesquisa, buscando realizar a codificação com base no livro de códigos e nas explicações iniciais. Os codificadores devem anotar a) todas as dúvidas existentes na codificação e b) tempo de codificação de cada unidade de análise.

IV. Comparação e verificação das divergências

Em uma nova sessão presencial, o pesquisador-líder ou algum pesquisador encarregado do treinamento dos codificadores irá comparar a codificação inicial e verificar as possíveis divergências. Os codificadores devem apresentar as dificuldades já anotadas no passo 3 e deve haver uma ampla discussão dos codificadores com o pesquisador-líder e entre os codificadores. É o momento para compartilhar dificuldades, opiniões, experiências, exemplos e afins. Todos os codificadores devem sair o mais ciente possível de erros na compreensão e na aplicação dos códigos. Uma forma de se verificar isso produtivamente é a comparação entre pares. Suponhamos que sejam quatro codificadores, então, o responsável pelo treinamento deve comparar os resultados entre o codificador 1 e o codificador 2, entre 1 e 3, entre 1 e 4, entre 2 e 3, entre 2 e 4 e entre 3 e 4. A porcentagem de concordância entre os

pares poderá ser útil para se verificar possíveis problemas de compreensão. Tal procedimento também facilitará encontrar, caso exista, o codificador que mais discorda dos colegas (se houver 3 ou mais codificadores). O tempo de codificação também deve ser informado ao pesquisador-líder para facilitar a organização do tempo na codificação final.

É importante compreender que as divergências também podem advir de vieses dos codificadores (bagagem cultural e opinião sobre determinado assunto), de diferenças de nível de treinamento em AC, ou simplesmente de erros de codificação, que podem ser causados pela excessiva repetição ou mesmo pelo cansaço por excessiva codificação. Logo, algumas divergências podem ocorrer simplesmente por erros aleatórios de preenchimento.

V. Mudanças/melhorias no livro de códigos

Exemplos de dúvidas e dificuldades podem (e devem!) ser inseridos no LdC (caso não seja a versão final). Todos os consensos alcançados na discussão devem ser devidamente incorporados ao LdC. O formulário de codificação também pode ser revisado, caso isso se mostre útil para facilitar ou acelerar a codificação.

VI. Codificação independente

Uma nova amostra de excertos dos materiais é tornada disponível aos codificadores, que devem novamente codificá-los de forma independente, sem se comunicar entre si ou com o pesquisador-líder (ou com o pesquisador que treina os codificadores). O tempo de codificação por unidade de análise deve ser registrado novamente para verificar se, dirimidas as dúvidas, houve alguma redução.

VII. Repetição dos passos 4-6

Os passos 4-6 devem ser repetidos tantas vezes quantas necessárias²⁸.

²⁸ Novamente, se as condições de pesquisa permitirem.

VIII. Teste de confiabilidade-piloto

Uma vez que haja consenso o suficiente, recomendamos a realização de um teste de confiabilidade-piloto. Esse teste pode ser realizado em amostras menores e menos representativas do conteúdo total. Aqui, o objetivo é apenas verificar as categorias que geram mais divergência entre os codificadores. O teste irá simplesmente prover números que facilitem essa compreensão. Se os resultados do teste forem insuficientes (ver Seção 7.2 para mais informações sobre o teste de confiabilidade), voltar para o passo 4. Se o teste alcançar bons resultados, é o momento para um último treinamento de fixação da codificação.

Box 3: O que fazer com codificadores que nunca concordam entre si?

Alguns manuais e a maioria dos artigos são bastante esquivos com o fato de que, em certos casos, simplesmente os codificadores nunca concordam entre si. O mais indicado é que seja verificado qual o codificador que tende a discordar mais vezes dos colegas. Se isso acontece de maneira sistemática, há uma boa chance de que esse codificador não seja apropriado para o estudo em questão (o que, evidentemente, não significa nenhum demérito intelectual). Segundo Riffe e colegas (2014), é preciso verificar se a) o codificador está tendo o tempo satisfatório para codificar e b) o referencial de codificação foi devidamente revisado em termos de seus procedimentos de codificação. Se a resposta for positiva, eles sugerem que o codificador seja retirado da pesquisa. Neuendorf (2002) também menciona que, em certos casos, pode haver esse *Rogue Coder* que, por algum motivo, simplesmente não consegue aplicar corretamente os códigos e, depois de alguma insistência, deve ser retirado da pesquisa. Portanto, em uma situação ideal de pesquisa (*i.e.*, recursos e tempo), recomendamos que o treinamento se inicie com ao menos três codificadores distintos.

7. Treinamento final e teste de confiabilidade

7.1 Treinamento

As divergências verificadas no teste-piloto serão a baliza para esse treinamento final. É o último momento para se retirar as últimas dúvidas. Neste momento, devem permanecer apenas os codificadores escolhidos para o teste de confiabilidade e não há mais espaço para revisões no livro de códigos. Portanto, deve ser o líder da pesquisa ou pesquisador encarregado do treinamento dos codificadores a decidir quando entrar nessa fase.

Neuendorf (2002) e Krippendorff (2004) recomendam o uso de codificadores “cegos”, ou seja, que não saibam exatamente o propósito do estudo para diminuir a inclusão de vieses na codificação; pois, de outra maneira, haveria uma tendência de os codificadores darem ao pesquisador o que ele deseja. Isso, claro, acontece em um contexto de pesquisa em que codificadores são contratados com fundos da pesquisa, podendo inclusive ser codificadores profissionais. Krippendorff (2004), em específico, chega a recomendar que os codificadores possuam um histórico similar em termos educacionais, profissionais e culturais e que não sejam usados os mesmos codificadores que ajudaram a construir o livro de códigos, uma vez que eles podem possuir consensos internos que não são desejáveis.

No contexto brasileiro, não raro os codificadores tenderão a ser alunos de graduação (geralmente, bolsistas de iniciação) e bolsistas de pós-graduação (frequentemente, participantes do grupo do pesquisador ou mesmo seus orientandos), e em diversos casos é possível que os codificadores trabalhem de maneira voluntária e/ou dividam seu tempo com o trabalho e mesmo com outras demandas da universidade. Em outras palavras, o mais corriqueiro em nosso contexto de pesquisa será trabalhar com codificadores que: 1) sabem o objetivo da pesquisa, 2) possuem históricos, conhecimento e treinamento em níveis bem diferentes entre si e 3) não estão exclusivamente disponíveis para a codificação. Todas essas características, em nossa visão, apontam para uma ênfase ainda maior na importância do treinamento.

Box 4: O treinamento aumenta a qualidade da codificação?

O treinamento dos codificadores é um assunto relativamente controverso por si. Enquanto todos os principais manuais admitem a importância em si do treinamento dos codificadores, frequentemente alguns autores assumem que o referencial de codificação e notadamente o LdC detalhados, ajustados e exequíveis deveriam permitir uma boa aplicação geral, como evidenciado nos trechos abaixo:

Idealmente, **as instruções deveriam ser compreendidas como estão escritas**, mas é normal que os codificadores recebam treinamento adicional para aplicá-las (KRIPPENDORFF, 2004, p. 129, grifos nossos).

Quando um conjunto de categorias atende ao critério da objetividade, **a classificação não será afetada pela subjetividade dos codificadores**. Nestas condições, diferentes pesquisadores deverão chegar a resultados **semelhantes** quando categorizando as mesmas unidades de conteúdo, a partir das mesmas regras de classificação. (MORAES, 1999, sp., grifos nossos).

As diferentes partes de um mesmo material [...] devem ser codificadas da mesma maneira, mesmo quando submetidas a várias análises. As distorções devidas à subjetividade dos codificadores e à variação dos juízos **não ocorrem** se a escolha e a **definição das categorias forem bem** estabelecidas. (BARDIN, 2016, p. 150, grifos nossos).

Nossa defesa aqui é de que se trata de posições epistemologicamente inadequadas, pois mesmo um livro de códigos exemplarmente elaborado pode (e tende a) gerar dúvidas entre os codificadores. Na prática, a preocupação de Krippendorff e colegas (*e.g.* MACQUEEN *et al.*, 1998) é que o treinamento pode gerar consensos *ad hoc* que podem não estar devidamente registrados no LdC, e que eles não são reproduzidos em outros contextos, prejudicando a replicabilidade do estudo. Então, enquanto todos recomendam ao menos um treinamento mínimo para a compreensão dos códigos e das regras de codificação, são poucos os estudos que conseguem provar a relação causal entre mais treinamento e melhor codificação. Ver o estudo de Milne e Adler (1999) para um exemplo de indicativo de melhora após o treinamento.

7.2 Teste de confiabilidade entre codificadores

Enquanto há boas descrições em português acerca das fases da AC, que inclusive apresentam a necessidade do teste de confiabilidade (e.g. BAUER, 2007; HERSCOVITZ, 2007; VIMIEIRO; MAIA 2011)²⁹, não encontramos descrições mais extensas sobre os passos de um teste de confiabilidade entre codificadores³⁰ na literatura brasileira³¹.

Os testes de confiabilidade se dão a partir da comparação entre as codificações de dois ou mais codificadores sobre um mesmo excerto de material. Em outras palavras, nos testes de confiabilidade, **todos os codificadores codificam exatamente o mesmo material, mas de forma independente**. Isso implica que eles não podem conversar ou trocar qualquer tipo de informação entre si durante a codificação. Para isso, é recomendável que os codificadores realizem a codificação do teste em locais e/ou momentos diferentes.

Ademais, todos os testes de confiabilidade (com exceção dos pilotos)³² devem ocorrer em uma amostra que seja aleatória e representativa da população estudada (ver seção anterior sobre amostra). Se, por exemplo, um estudo decide analisar os prontuários de um posto médico no período de dois anos, isso significa dizer que os prontuários escolhidos para o teste devem compreender todos os períodos dentro do recorte temporal. Afinal, se os prontuários do exemplo forem excessivamente concentrados em certo período, é possível que determinadas condições dos pacientes sejam

²⁹ Bardin (2016), inclusive, afirma que “para um maior rigor, esses resultados [da AC] são submetidos a provas estatísticas, assim como a testes de validação” (BARDIN, 2016, p. 131).

³⁰ Para além dos manuais especializados já elencados, o pesquisador Matthew Lombard mantém um site dedicado ao tema que resume bem todas as questões envolvidas no teste. Disponível em: <http://matthewlombard.com/reliability>.

³¹ Esta seção e a seção seguinte reproduzem trechos do texto “Eu quero acreditar! Da importância, formas de uso e limites dos testes de confiabilidade na Análise de Conteúdo” publicado na Revista de Sociologia e Política, v. 26, n. 66, 2018. Buscamos, entretanto, fazer algumas melhorias e reformulações do texto original (SAMPAIO; LYCARIÃO, 2018).

³² A exceção são os testes-piloto (ou pré-testes), quando o objetivo principal é afinar a compreensão dos codificadores, criar consensos e fazer as modificações ao livro de códigos. Nesse caso, a amostra não precisa ser representativa do total, mas o ideal é que ainda seja escolhida de forma aleatória.

superestimadas, como seria o caso se houvesse um surto de dengue ou mesmo em um período de muitas viroses. A amostra concentrada também poderia corromper o teste, se fosse concentrada no período de férias de determinados profissionais. Esses são apenas exemplos genéricos de como é importante a representatividade temporal do período estudado.

Inicialmente, o resultado do teste de confiabilidade era calculado apenas pela concordância absoluta, ou seja, comparando-se as codificações de modo a obter a percentagem dos casos em que houve concordância. Entretanto, atualmente, esse é considerado um procedimento bastante limitado. Isso porque há variáveis (especialmente as dicotômicas) em que há uma chance alta de concordância aleatória entre os codificadores (HAYES; KRIPPENDORFF, 2007).

Essa situação pode ser ilustrada a partir de uma simplificação da variável denominada valência. Ela se encontra em diversos estudos da Comunicação e da Ciência Política e visa a identificar como matérias jornalísticas constroem a imagem pública de atores políticos (*e.g.* FERES JÚNIOR 2016; FIGUEIREDO *et al.* 1997), ou seja, se a matéria apresenta uma imagem positiva ou negativa de um determinado ator político. Como há apenas duas possibilidades de codificação (ignoremos, para fins didáticos, a valência neutra), isso significa que há, em cada decisão, 50% de chance de os codificadores concordarem de maneira aleatória. Isso quer dizer que, mesmo sem ler o material, e apenas preenchendo aleatoriamente uma das duas opções, os codificadores ainda terão uma alta chance de concordar entre si em cada decisão.

Diante disso, a literatura especializada aponta que a simples percentagem nominal da concordância entre os codificadores não é suficiente para garantir a confiabilidade da codificação (HAYES; KRIPPENDORFF, 2007; LOMBARD; SNYDER-DUCH; BRACKEN, 2002; MACNAMARA, 2005; MATTHES; KOHRING, 2008). Consoante essa literatura, é, então, necessário aplicar algum índice de teste de confiabilidade entre codificadores que pondere a chance aleatória de concordância³³. Tais testes se dão a partir de equações matemáticas que incorporam, além da concordância absoluta entre os

³³ Não é nosso objetivo entrar nos meandros das diferentes fórmulas, uma vez que isto já foi realizado com propriedade anteriormente (FENG, 2014; HAYES; KRIPPENDORFF, 2007).

codificadores, a covariação, a particular raridade ou excessiva aparição de certas categorias, além, claro, da chance de os codificadores concordarem aleatoriamente entre si³⁴.

Os valores dos testes de confiabilidade, geralmente, variam de -1 a 1, sendo que 1 indicaria uma concordância perfeita, zero uma falta de confiabilidade por serem pareamentos aleatórios e abaixo de zero significa que há uma discordância não aleatória ocorrendo. Ou seja, quando temos um índice negativo, isso implica que os codificadores estão codificando o mesmo material a partir de compreensão significativamente e sistematicamente distinta uns dos outros. Qualquer valor acima de 0.9 é, em geral, considerado muito confiável, e acima de 0.8, suficientemente confiável. Já valores entre 0.667 e 0.8 são considerados suficientes para variáveis experimentais (em aperfeiçoamento) no caso do alpha de Krippendorff (*vide* abaixo), mas serão passíveis de diferentes interpretações, a depender do índice utilizado. Por sua vez, valores abaixo de 0.667 tendem a ser aceitos apenas para estudos em fase de teste (NEUENDORF, 2002).

Qual índice utilizar e como?

Apesar de existirem mais de 20 índices diferentes para se realizar um teste de confiabilidade, três deles se destacam pelo seu uso mais disseminado: o kappa de Cohen, o pi de Scott, e o alpha de Krippendorff³⁵ (LACY *et al.*, 2015). Desses, o alpha de Krippendorff, além de ser altamente exigente, mostra-se bastante prático e versátil, pois não tem restrição em termos de número de codificadores (já que alguns índices só funcionam para dois codificadores, como o kappa de Cohen³⁶) nem em termos de variáveis; ou seja, se são qualitativas (subtipo ordinal ou nominal) ou quantitativas

³⁴ Este é o motivo de a maioria não recomendar o uso de técnicas estatísticas simples (alpha de Cronbach, r de Pearson ou qui-quadrado), uma vez que esses testes não medem a concordância entre os codificadores ou mesmo a chance da concordância aleatória, mas apenas a covariação entre os resultados encontrados.

³⁵ Para mais sobre as diferenças entre os índices, ver Hayes e Krippendorff (2007) e Feng (2014).

³⁶ O kappa de Fleiss é uma adaptação do kappa de Cohen para mais de dois codificadores.

(subtipo discreta ou contínua)?³⁷ (HAYES; KRIPPENDORFF, 2007). Entretanto, para pesquisas que trabalham com um número alto de categorias pouco presentes, a escolha de outro índice pode ser mais adequada (ver WOZNIAK, LÜCK, WESSLER, 2015).

A maior parte dos *software* de AC, como Nvivo³⁸, e *software* de análise estatística, como SPSS³⁹, Stata⁴⁰ e R⁴¹, possui pacotes ou adaptações para a realização dos testes de confiabilidade, conforme os *links* disponíveis nas notas de rodapé. Há também *software* específicos para o cálculo, como o Simstate⁴² e o *Program for Reliability Assessment with Multiple Coders* (PRAM)⁴³.

Apresentando uma curva de aprendizado mais baixa, há páginas *on-line* que calculam os índices de forma simples, sendo alimentadas diretamente⁴⁴ ou por meio de arquivos⁴⁵. Um notório exemplo é a página desenvolvida pelo pesquisador Dean Freelon (<http://dfreelon.org/utills/recalfront/>). Ela faz o cálculo do pi de Scott, do kappa de Cohen, do kappa de Fleiss e do alpha de Krippendorff, utilizando a ferramenta *ReCal*⁴⁶ (FREELON, 2010, 2013); e nos parece a melhor opção para pesquisadores iniciantes na AC. O site apresenta duas formas de funcionar.

A primeira é desenhada exclusivamente para o caso de termos apenas dois codificadores e é chamada de ReCal2⁴⁷. A vantagem dessa ferramenta é

³⁷ Para mais sobre os tipos de variáveis, ver Capítulo 3 e também Cervi (2014) e Krippendorff (2004).

³⁸ Ver mais no *link*:http://help-nv11.qsrinternational.com/desktop/procedures/run_a_coding_comparison_query.htm. Apenas em inglês.

³⁹ Ver mais no *link* <http://afhayes.com/spss-sas-and-r-macros-and-code.html> para a instalação. Informações mais detalhadas disponíveis neste *link* (<http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf>). Apenas em inglês.

⁴⁰ Ver mais sobre em: <https://ideas.repec.org/p/boc/dsug17/07.html>

⁴¹ No caso do R, como é um *software* aberto, há um enorme leque de opções de pacotes a serem instalados para isso. Por exemplo: http://www.cookbook-r.com/Statistical_analysis/Inter-rater_reliability/.

⁴² Disponível em: <https://provalisresearch.com/downloads/trial-versions/>.

⁴³ Disponível em: <http://academic.csuohio.edu/kneuendorf/c63311/PRAMS/>.

⁴⁴ No site <http://justusrandolph.net/kappa/>, o pesquisador preenche diretamente *on-line* as tabelas, evidenciando onde houve concordância entre os codificadores.

⁴⁵ Ver <https://nlp-ml.io/jg/software/ira/>.

⁴⁶ Para mais explicações sobre o site, ver Freelon (2010, 2013).

⁴⁷ <http://dfreelon.org/utills/recalfront/recal2/>.

que é possível testar todas as variáveis de uma única vez. Basta considerar as linhas da planilha (.csv) como os casos (unidades de análise) e as colunas como as categorias codificadas, e colocar lado a lado as codificações de dois codificadores. Neste exemplo, temos 10 casos, ou unidades de análise, e duas variáveis, avaliadas por dois codificadores.

Tabela 1 – Exemplo de teste de confiabilidade entre dois codificadores

	Codificador 1	Codificador 2	Codificador 1	Codificador 2
	Variável I		Variável II	
Caso 1	1	1	0	1
Caso 2	2	1	2	2
Caso 3	1	2	3	2
Caso 4	3	3	3	3
Caso 5	2	2	2	2
Caso 6	2	2	3	2
Caso 7	2	1	2	2
Caso 8	3	3	3	3
Caso 9	1	1	1	1
Caso 10	2	3	2	2

Fonte: elaboração própria.

A coluna A será a codificação do codificador 1 para a variável I, a coluna B será a codificação do codificador 2 para a variável I, a coluna C será a codificação do codificador 1 para a variável II, a coluna D será a codificação do codificador 2 para a variável II e assim por diante. Logo, para a Recal2, é sempre necessário que as colunas sejam pares e o *software* considera cada par como uma variável diferente. Para funcionar adequadamente, o pesquisador deve colocar apenas as codificações em uma planilha (*software* ou *on-line*), retirando todas as letras e símbolos diferentes dos números. Deve-se salvar em formato.csv (o site não funciona com outros formatos de arquivo). Nesse exemplo, teríamos uma planilha similar a esta:

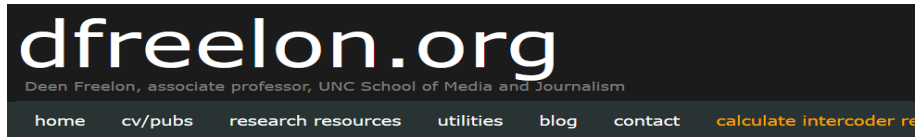
Figura 2 – Exemplo de resultados numéricos em planilha

	A	B	C	D
1	1	1	0	1
2	2	1	2	2
3	1	2	3	2
4	3	3	3	3
5	2	2	2	2
6	2	2	3	2
7	2	1	2	2
8	3	3	3	3
9	1	1	1	1
10	2	3	2	2

Fonte: elaboração própria.

Este arquivo.csv alimentará diretamente a ferramenta Recal2, conforme a Figura 2. No exemplo acima, colocamos as discordâncias dos codificadores em negrito apenas para termos didáticos, o que não precisa ser feito ao alimentar a ferramenta. Clique no botão “browse” e selecione o arquivo.csv em seu computador. Em seguida, clique no botão “Calculate Reliability”.

Figura 3 – Recal2



ReCal2: Reliability for 2 Coders

ReCal2 ("Reliability Calculator for 2 coders") is an online utility that computes intercoder/interrater reliability coefficients for **nominal data coded by two coders**. (Versions for 3 or more coders working on nominal data and for any number of coders working on ordinal, interval, and ratio data are also available.) Here is a brief feature list:

- ▶ Calculates four of the most popular reliability coefficients for nominal data: percent agreement, Scott's Pi, Cohen's Kappa, and Krippendorff's Alpha.
- ▶ Can calculate reliability for multiple variables at a time
- ▶ Accepts any range of possible variable values
- ▶ Results should be valid for **nominal data coded by two coders** (other uses are not endorsed, and accurate results are not guaranteed in any case — trust but verify!)

If you have used ReCal2 before, you may submit your data file for calculation via the form below. If you are a first-time user, please read [the documentation](#) first. (Note: failure to format data files properly may produce incorrect results!) You should also read ReCal's [very short license agreement](#) before use.

Browse... No file selected. Calculate Reliability

Fonte: www.dfreelon.org

Assim, teríamos a seguinte página como resultado do teste do exemplo anterior:

Figura 4 – Resultado do teste do Recal2

Congratulations! Your file has passed a basic error-check and is probably OK. But please doublecheck if the output below seems off.

ReCal 0.1 Alpha for 2 Coders

results for file "recal2.csv"

File size: 90 bytes
 N columns: 4
 N variables: 2
 N coders per variable: 2

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha (nominal)	N Agreements	N Disagreements	N Cases	N Decisions
Variable 1 (cols 1 & 2)	60%	0.389	0.403	0.42	6	4	10	20
Variable 2 (cols 3 & 4)	70%	0.528	0.545	0.551	7	3	10	20

Export Results to CSV [\(what's this?\)](#)

Select another CSV file for reliability calculation below.

Browse... No file selected. Calculate Reliability

Save results history [\(what's this?\)](#)

Fonte: www.dfreelon.org

Conforme vemos na figura 4, os dois codificadores concordaram em 60% do tempo na variável 1 e 70% na variável 2. Ainda assim, os resultados do alpha de Krippendorff estão abaixo do mínimo aceitável (que seria entre 0.667 e 0.8); portanto, o teste falhou e precisaria ser repetido (veja mais sobre isso abaixo).

Por sua vez, a Recal3⁴⁸ é a ferramenta do site Dfreelon para o cálculo do índice do teste de confiabilidade entre três ou mais codificadores. Diferentemente da ferramenta anterior, a Recal3 só calcula uma variável por vez, sendo que o site considera cada coluna como sendo a codificação de um codificador diferente; portanto, é necessário preparar um arquivo.csv para cada variável a ser testada. A título de exemplo, se houver quatro colunas, ele não considerará que são dois codificadores analisando duas variáveis diferentes (como é o caso da Recal2), mas, sim, que quatro codificadores diferentes estão analisando uma única variável. Tentamos exprimir isso no exemplo abaixo. Os negritos representam as discordâncias.

Tabela 2 – Exemplo de teste de confiabilidade com quatro codificadores

	Codificador 1	Codificador 2	Codificador 3	Codificador 4
Caso 1	1	1	1	1
Caso 2	2	2	2	2
Caso 3	1	1	2	1
Caso 4	3	3	3	1
Caso 5	2	2	3	3
Caso 6	3	2	2	2
Caso 7	2	2	2	2
Caso 8	3	3	3	3
Caso 9	1	1	1	1
Caso 10	2	2	2	2

Fonte: elaboração própria.

Ao utilizar a ferramenta Recal, o site já retorna com o resultado completo, incluindo o número de casos e o número de tomada de decisões entre codificadores, além da concordância média (porcentagem) entre todos os codificadores e, também, entre cada par de codificadores (o que pode ser particularmente útil para compreender se algum codificador está com mais

⁴⁸ <http://dfreelon.org/utills/recalfront/recal3>.

dificuldade que o grupo). Como o Kappa de Cohen é pensado apenas para dois codificadores, o seu resultado também é apresentado para cada par.

Figura 5 – Resultados da ferramenta ReCal3

ReCal 0.1 Alpha for 3+ Coders results for file "recal3.csv"

File size: 90 bytes
N coders: 4
N cases: 10
N decisions: 40

Average Pairwise Percent Agreement

Average pairwise percent agr.	Pairwise pct. agr. cols 1 & 4	Pairwise pct. agr. cols 1 & 3	Pairwise pct. agr. cols 1 & 2	Pairwise pct. agr. cols 2 & 4	Pairwise pct. agr. cols 2 & 3	Pairwise pct. agr. cols 3 & 4
78.333%	70%	70%	90%	80%	80%	80%

Fleiss' Kappa

Fleiss' Kappa	Observed Agreement	Expected Agreement
0.664	0.783	0.355

Average Pairwise Cohen's Kappa

Average pairwise CK	Pairwise CK cols 1 & 4	Pairwise CK cols 1 & 3	Pairwise CK cols 1 & 2	Pairwise CK cols 2 & 4	Pairwise CK cols 2 & 3	Pairwise CK cols 3 & 4
0.666	0.545	0.538	0.846	0.688	0.683	0.697

Krippendorff's Alpha (nominal)

Krippendorff's Alpha	N Decisions	$\sum_c o_{cc}^{***}$	$\sum_c n_c(n_c - 1)^{***}$
0.672	40	31.3333333333	528

***These figures are drawn from [Krippendorff \(2007, case C.\)](#)

[\(what's this?\)](#)

Select another CSV file for reliability calculation below:

No file selected.

Save results history [\(what's this?\)](#)

Fonte: www.dfreelon.org.

Nesse exemplo, o teste foi bem-sucedido, mas dentro da margem mínima aceitável para qualquer avaliação (0,672 no índice do alpha de

Krippendorff). Um periódico (ou parecerista) realmente exigente poderia questionar o porquê de o resultado não ser acima de 0.8. Ressalta-se que, no caso de três ou mais codificadores, o teste acima precisa ser realizado separadamente para cada categoria do livro de códigos.

Deve-se destacar que o procedimento de um teste rigorosamente executado não deve fazer a média entre os valores obtidos em cada variável ou qualquer outra adaptação dos resultados. Assim, é recomendável que seja exibido o valor do índice para cada variável do livro de códigos da pesquisa⁴⁹. Se o menor valor for abaixo do índice aceitável, o teste falhou e deve ser repetido, após nova rodada de treinamentos.

Segundo os manuais especializados (KRIPPENDORFF, 2004; NEUENDORF, 2002; RIFFE, LACY; FICO, 2014), se uma variável não passou, o teste falhou e deve ser repetido para todas as variáveis. Particularmente, parece-nos aceitável que os testes sejam repetidos apenas para as variáveis que não passaram, em vez de todas, especialmente quando o livro de códigos for extenso e quando houver fortes constrangimentos de pesquisa. Todavia, essa prática deve ser utilizada com cuidado. Um tempo muito excessivo entre o treinamento e a efetiva codificação final tende a diminuir a memória dos codificadores sobre o treinamento e diminuir a confiabilidade da codificação.

Box 5: E se o teste de confiabilidade continua falhando?

Caso a complexidade de uma variável seja elevada, o esperado é que sejam necessários vários testes para se alcançar a confiabilidade desejável (LACY *et al.*, 2015). Nessa situação, mais testes-piloto e maior tempo de treinamento serão necessários em relação a variáveis que requerem alta densidade interpretativa. Por exemplo, identificar **a editoria** de uma matéria de jornal impresso requer menor densidade interpretativa do que identificar a **valência** de um ator político nessa mesma matéria (LACY *et al.*, 2015). De toda forma, os novos treinamentos deveriam se focar, particularmente,

⁴⁹ Caso a pesquisa apresente um único valor, infere-se que ele se refere ao menor valor obtido no teste de confiabilidade.

nas divergências entre os codificadores e na compreensão das razões pelas quais eles discordaram durante as codificações (ver subseção anterior sobre treinamento).

Deve-se observar que os testes de confiabilidade tendem a falhar quando os codificadores alcançam próximo a 100% de concordância, uma vez que a maior parte dos índices espera alguma discordância entre sujeitos (cf. HAYES; KRIPPENDORFF, 2007). Logo, os testes de confiabilidade entre codificadores são pensados para variáveis que exigem decisões dos codificadores. Se são totalmente objetivas (os codificadores só discordam em caso de erro, por exemplo), as variáveis não necessitam de teste de confiabilidade (e.g. preencher a data do material coletado). Também é possível que o material escolhido para os testes não seja suficientemente representativo, notadamente quando há códigos que aparecem mais raramente. Nesses casos, uma sugestão é realizar outro teste com excertos que amostram esses itens mais raros, conforme sugerido por Hayes e Krippendorff (2007) e Lacy e colegas (2015).

Caso os testes continuem falhando, Lima (2003, p. 22) sugere a reformulação do livro de códigos. Para ele, três alternativas para aperfeiçoar o livro de códigos e, assim, permitir um maior nível de concordância entre os codificadores seriam: a) proceder à fusão de categorias; b) alterar os descritivos das categorias para torná-las mais claras; c) inserir exemplos típicos mais adequados para ilustrar o conteúdo indicativo dessas categorias.

Veja o exemplo da supracitada categoria nível de justificação dos argumentos (STEINER, 2012):

1. O orador (*speaker*) não apresenta nenhum argumento (pergunta, por exemplo, meramente por informação adicional).
2. O orador apenas diz que X deveria ou não ser feito, que isso é uma ideia maravilhosa ou terrível etc.; porém, nenhuma razão é apresentada sobre porque X deveria ou não ser feito.
3. O orador justifica apenas com ilustrações de porque X deveria ou não ser feito.
4. O orador dá uma razão Y sobre por que X deveria ou não ser feito, mas não há nenhuma conexão sobre como Y irá contribuir para X.

5. O orador dá uma razão Y sobre por que X deveria ou não ser feito e há uma conexão sobre como Y irá contribuir para X.
6. O orador dá ao menos duas razões por que X deveria ser feito e também ao menos duas conexões sobre como as razões contribuirão para X.

Trata-se claramente de uma categoria complexa e de difícil aplicação. Notadamente, compreender o que o autor define como “conexão” (*linkage* no original) é tarefa complexa; portanto, que tende a ser difícil para se atingir aceitáveis índices de confiabilidade. Maia e equipe (2017) optaram por uma simplificação da categoria acima, reduzindo de seis para apenas quatro códigos, buscando atingir um índice de confiabilidade aceitável, como a seguir: 1) sem argumentos, 2) opinião, 3) argumentos simples e 4) argumentos complexos. O principal problema em uma adaptação dessa natureza é a perda de uma comparabilidade plena de dados; todavia, como em qualquer pesquisa, escolhas precisam ser feitas.

Dados os limites da pesquisa brasileira, em última instância, a alternativa final seria enviar o artigo para evento ou periódico com os resultados abaixo dos indicados como ideais e realizar a justificativa pública adequada. Cabe aos pares avaliar se ainda assim os resultados da pesquisa podem ser considerados como suficientemente confiáveis. Não obstante, essa alternativa não deve ser utilizada com frequência.

7.3 Procedimentos para pesquisas individuais

Enquanto os procedimentos acima correspondem a uma realidade mais profissionalizada de pesquisa, aquela que conta com ao menos dois codificadores, as indicações desta seção se voltam a organizar procedimentos para pesquisadores individuais e que não têm condições de contar com a colaboração de assistentes de pesquisa.

Esse é o caso de pesquisas produzidas em trabalhos de conclusão de curso (TCC), especializações, mestrado e doutorado⁵⁰. Tais procedimentos

⁵⁰ Contudo, no nível da pós-graduação, reforça-se a importância do uso de dois ou mais codificadores, como explicamos anteriormente.

respondem à demanda por confiabilidade não mediante a realização de um teste com outros codificadores, mas com o(a) próprio(a) pesquisador(a). Apesar de não recomendar seu uso, Krippendorff (2004, p. 214) reconhece o método, classificando-o como “estabilidade”. Em suma, esse método tende a indicar baixa presunção de confiabilidade, mas ao menos demonstra a estabilidade da classificação do próprio pesquisador. Visando a aumentar a confiabilidade ao máximo, tornamos a exigência maior do que aquela retratada por Krippendorff. Tais procedimentos se encontram explicitados e sua formulação se volta a um uso crítico e flexível por parte da comunidade científica pertinente:

- a) Escreva um livro de códigos especificando, a partir de códigos alfanuméricos, além das variáveis e categorias, todas as regras de codificação para cada categoria⁵¹.
- b) Construa uma planilha de dados em que, ao lado de cada variável, conste uma outra coluna para se inserir o código da regra utilizada em cada codificação.
- c) Quando não for possível identificar uma regra no livro de códigos que seja capaz de subsidiar a codificação, reformule o livro de códigos, incluindo a regra necessária à codificação.
- d) Quando mais de uma regra for aplicável, reformule o livro, reduzindo a ambiguidade entre as regras ou criando uma adicional que “desempate” as demais.
- e) Repetir o procedimento até que 10% do material da pesquisa (estabelecendo o mínimo de 50 unidades) tenha sido codificado⁵².
- f) Após finalizada a codificação desse material, realize, após o intervalo de, ao menos, uma semana, uma nova codificação das mesmas unidades, mas sem consultar a primeira codificação.
- g) De posse da nova codificação, realize testes de confiabilidade contrastando a primeira com a segunda.

⁵¹ Tal procedimento também pode ser realizado em pesquisas com mais de um codificador. Trata-se de garantir, além de um nível maior de verificabilidade externa da codificação, um instrumento por meio do qual o pesquisador poderá ter clareza da precisão de suas variáveis e categorias.

⁵² O ideal, no entanto, é utilizar material que não faça parte do banco de dados da pesquisa a ser reportada no trabalho final.

- h) Caso os testes não tenham atingido níveis *ótimos* de confiabilidade, deve-se revisar os casos incongruentes e refinar o livro de códigos.
- i) Repetir os procedimentos acima com outras unidades (evitando, portanto, codificar uma mesma unidade mais de duas vezes), até que se alcance níveis ótimos de confiabilidade.
- j) Uma vez alcançado níveis ótimos de confiabilidade (acima de 0.9), codificar todas as unidades da pesquisa.
- k) Reportar os níveis alcançados de confiabilidade para cada variável.
- l) Disponibilizar, na internet, o livro de códigos, a planilha de dados codificada e, quando possível, o material da pesquisa.

No caso das instruções acima, deve-se esclarecer que a exigência em torno de um resultado acima de 0,9 no teste realizado é recomendável, pois se trata de apenas um único codificador. Se o próprio pesquisador não consegue replicar seus próprios resultados, isso significa que as regras e as distinções não estão claras. Portanto, em tal caso, é ainda mais improvável que outros pesquisadores venham a conseguir replicar os resultados.

Dito isso, concluímos esta seção reforçando a necessidade de, em qualquer pesquisa com AC quantitativa, realizar-se testes de confiabilidade. Isso porque eles são ferramentas cruciais para se garantir que uma pesquisa seja replicável. Uma vez que a replicabilidade é condição de possibilidade da confiabilidade, uma pesquisa sem um teste e procedimento similar ao aqui descrito deixa de garantir tal condição.

Ademais, a relevância e a necessidade de testes de confiabilidade se tornaram tão fortes que a prática-padrão dos periódicos de alto impacto é publicar apenas trabalhos que apresentam algum teste com alto nível de concordância⁵³. Entretanto, compreendemos que tal procedimento é apenas um dos que podem sustentar a presunção de confiabilidade, sendo necessárias outras ações ainda mais garantidoras. A seção seguinte esclarece quais ações

⁵³ Apesar de, como verificado por Lovejoy *et al.* (2014), ainda existem artigos publicados em periódicos de alto impacto que não retratam adequadamente como foi testada a confiabilidade.

seriam essas, assim como apresenta os argumentos que indicam os limites que (apenas) um índice de confiabilidade “alto” apresenta para a satisfação de seu princípio epistemológico correspondente.

8. Codificação

A codificação final, de fato, só começa nesta etapa. Esse é um dos motivos para acreditarmos que o modelo proposto por Bardin tende a simplificar excessivamente todas as etapas necessárias para uma análise de conteúdo com base científica. Assim, não seria surpresa verificar que certas pesquisas baseadas em AC suprimiram uma ou mais etapas anteriores, o que provavelmente teve efeitos na qualidade geral dos dados.

Em resumo, a codificação se inicia quando cinco requisitos foram suficientemente atendidos: 1) a unidade de análise foi definida; 2) a amostra foi definida e realizada; 3) o referencial de codificação, notadamente o livro de códigos, foi devidamente revisado para aumentar a validade e a confiabilidade da pesquisa; 4) houve treinamento suficiente e adequado para os codificadores participantes; e 5) há um resultado considerado suficiente no teste de confiabilidade entre codificadores inicial. Assim, se todos os requisitos foram cumpridos, é hora de a codificação final acontecer.

Seja na população geral de textos/conteúdos, seja na amostra, é o momento de dividir o número total de unidade de análises entre os codificadores aprovados no teste de confiabilidade. Exatamente como nos testes de confiabilidade, cada um codifica de modo independente, não havendo consulta entre si ou com o líder da pesquisa. Em alguns casos, a divisão será simples, até mesmo óbvia. Se estão sendo analisadas 400 matérias jornalísticas em uma AC e há quatro codificadores, basta cada um codificar 100 matérias distintas. Em outros exemplos, a divisão poderá ser mais complexa, notadamente quando a unidade de análise for menor. Peguemos o exemplo de análises de conteúdo de materiais audiovisuais materializado em análises do horário gratuito político eleitoral (HGPE). Talvez a separação entre codificadores por programa de campanha não seja adequada; afinal, certas campanhas possuem mais tempo de HGPE do

que outras. Como o HGPE tem a mesma duração diária, uma possibilidade seria atribuir diferentes dias do programa a cada codificador por exemplo. Ou, alternativamente, verificar o número de minutos de programa por dia e dividir de maneira similar entre os codificadores. Independente do método de decisão, é importante que os codificadores recebam um material diverso para codificar. No exemplo do HGPE, é mais provável que o codificador inclua um viés pessoal ao codificar todas as campanhas de um único candidato do que ao receber programas de diferentes candidatos, assim como já ocorreu nos diferentes testes de confiabilidade anteriores.

Com base nas medições que já ocorreram ao longo dos diversos treinamentos, cabe ao líder da pesquisa tornar um tempo razoável disponível para que a codificação seja realizada com a devida atenção. Então, deve-se considerar o tempo médio de análise de cada unidade de análise e o tempo disponível para o codificador diariamente. Além disso, é pertinente notar que pouco adianta seguir todas as etapas anteriores se, ao final, exigem-se metas impossíveis dos codificadores. Isso pode ter um efeito negativo para a codificação final, o que, idealmente, ainda precisa ser verificado mais uma vez.

Figura 6 – Exemplo de codificação concluída

TWEET_ID				
	A	B	C	D
TWEET_ID	Tweet_Link	Tweet_Texto		Posicionamento
1	https://twitter.com/Byanka	RT @Byanka: @ninalemos fiz um aborto, não sou criminosa. sandra poderia ser eu, poderia ser você. http://t.co/EP2MtuPQda #precisamosfal...		1
2	https://twitter.com/jeanywlyp_real	RT @jeanywlyp_real: O Estado brasileiro é Laico e a Mulher é livre! http://t.co/QRWib3iX #Precisamosfalarsobreaborto @revista_tpm http://t.co/300yP1G3g		1
3	https://twitter.com/fabiocara	RT @fabiocara: é você é contra ou a favor do aborto? #precisamosfalarsobreaborto http://t.co/300yP1G3g		3
4	https://twitter.com/jeanywlyp_real	RT @jeanywlyp_real: O Ex-BBB @jeanywlyp_real publicou uma foto com #PrecisamosFalarsobreAborto - que vergonha! Deputado! @jeanywlypKaoMeRepre...		2
5	https://twitter.com/jeanywlyp_real	RT @jeanywlyp_real: O Estado brasileiro é Laico e a Mulher é livre! http://t.co/QRWib3iX #precisamosfalarsobreaborto @revista_tpm http://t.co/300yP1G3g		1
6	https://twitter.com/jeanywlyp_real	RT @jeanywlyp_real: O Estado brasileiro é Laico e a Mulher é livre! http://t.co/QRWib3iX #precisamosfalarsobreaborto @revista_tpm http://t.co/300yP1G3g		1
7	https://twitter.com/millyacombe	#precisamosfalarsobreaborto 🙏🙏 @revistatpm @gduvillier @revistatpm #magazine http://t.co/1EwOpWW09		1
8	https://twitter.com/millyacombe	RT @millyacombe: Quem é contra o aborto deveria estar na rua exigindo educação sexual, camisinha, contraceptivo... (C. Sagan) #precisamosfa...		1
9	https://twitter.com/millyacombe	Leandra Leal <3		
10	https://twitter.com/millyacombe	#precisamosfalarsobreaborto http://t.co/UVV1785tC7		1
11	https://twitter.com/jeanywlyp_real	RT @jeanywlyp_real: O Estado brasileiro é Laico e a Mulher é livre! http://t.co/QRWib3iX #Precisamosfalarsobreaborto @revista_tpm http://t.co/300yP1G3g		1
12	https://twitter.com/jeanywlyp_real	RT @leandraleal: Abortos inseguros e mal sucedidos são a 5a causa de morte materna		1
13	https://twitter.com/jeanywlyp_real	http://t.co/eqp4Kou7u8 #precisamosfalarsobreaborto		1
14	https://twitter.com/jeanywlyp_real	RT @Be_reviani: @revista_tpm @NinaLemos também está na campanha. #precisamosfalarsobreaborto http://t.co/VXnu3BOHVT		1
15	https://twitter.com/jeanywlyp_real	RT @leandraleal: A campanha #precisamosfalarsobreaborto propõe uma discussão profunda, livre de preconceitos, sobre um... http://t.co/yDDQDh...		1
16	https://twitter.com/jeanywlyp_real	RT @marciaburti: #precisamosfalarsobreaborto CAMPANHA DA @revistatpm http://t.co/a5emlp36C		1
17	https://twitter.com/jeanywlyp_real	#precisamosfalarsobreaborto na sociedade brasileira.		1
18	https://twitter.com/jeanywlyp_real	É uma questão de saúde pública, não de ética... http://t.co/LA3Mq1AD2P		1

Fonte: Adaptado de Lycarião e Dos Santos, 2017.

9. Teste de confiabilidade intermediário e final

Convém ressaltar que os testes de confiabilidade, segundo as etapas acima, podem ocorrer em diferentes momentos da pesquisa conteúdo

(SAMPAIO; LYCARIÃO, 2018). Nas pesquisas especializadas mais criteriosas, repetidos testes de confiabilidade são realizados durante todo o processo de pesquisa, nomeadamente: 1) em testes-piloto, 2) no teste de confiabilidade oficial antes da codificação, 3) no meio da codificação e 4) após a codificação.

No teste inicial, que ocorre antes da codificação, busca-se averiguar se as categorias são confiáveis e se os codificadores participantes foram capazes de assimilá-las dentro do padrão científico esperado dentro da AC (LACY *et al.*, 2015). O teste intermediário, por sua vez, busca verificar se a codificação não está caindo em sua qualidade. É um teste que pode ser bastante útil para se verificar possíveis erros antes da conclusão da codificação completa. Idealmente, ele deve ocorrer em torno de 40% a 50% da codificação final. Finalmente, o teste de confiabilidade final ocorre após a conclusão da codificação e busca justamente demonstrar que a codificação final manteve a qualidade atestada pelo teste de confiabilidade inicial (LOMBARD; SNYDER-DUCH; BRACKEN, 2002)⁵⁴.

Então, se pudéssemos resumir o passo a passo sugerido por Neuendorf (2002), sugeriríamos que um teste de confiabilidade fosse composto das seguintes etapas:

- I) Sucessivas codificações-piloto em uma amostra heterogênea (mínimo de 10 unidades) levando a revisões do livro de códigos, até que a confiabilidade se mostre aceitável (não aleatória).
- II) Codificação independente final em amostra aleatória representativa (sugerimos 10% da amostra, sendo o mínimo de 50 unidades).
- III) Pesquisador verifica resultado do índice de confiabilidade.
- VI) Se passar, relato dos codificadores de suas experiências. Em caso de falha, retornar ao passo I ou excluir a variável da análise do material completo.
- V) Disponibilizar, na internet, o livro de códigos, a planilha de dados codificada e o material da pesquisa.
- VI) Apresentar um relato detalhado do processo no texto científico.

⁵⁴ Entretanto, boa parte da literatura reporta apenas o resultado do teste inicial, aquele realizado antes da codificação em si, o que é considerado por alguns autores como incorreto (e.g. LOMBARD; SNYDER-DUCH, BRACKEN, 2002).

Portanto, por último, o pesquisador precisa explicar como as divergências finais (ou seja, aquelas existentes no teste de confiabilidade) foram tratadas. Dito de outra forma, mesmo após o teste de confiabilidade ser bem-sucedido, existirão algumas discordâncias de codificação. Os textos especializados recomendam que essa divergência final seja discutida entre os codificadores e o pesquisador responsável (LOMBARD; SNYDER-DUCH; BRACKEN, 2002; NEUENDORF, 2002). Afinal, a codificação do material em si será iniciada na sequência. Contudo, frisamos que o livro de códigos não pode ser mais alterado a essa altura em termos gerais⁵⁵.

Porém, o que acontece se o teste de confiabilidade final estiver consideravelmente pior do que o inicial/intermediário? Esse é justamente o motivo para se indicar a realização de, ao menos, um teste de confiabilidade intermediário, uma vez que, aí, correções podem ser realizadas durante a codificação. Em suma, se o teste de confiabilidade final for consideravelmente inferior e se estiver abaixo dos índices mínimos para garantia da confiabilidade, a melhor ideia é descartar a codificação e refazê-la. Aqui, cabe a avaliação e palavra-final do pesquisador-líder. A depender, ele poderia voltar até a fase 3 do treinamento dos codificadores para verificar o que houve de errado. Em casos extremos, uma solução poderia ser a troca completa dos codificadores (ou, ao menos, daqueles que mostraram mais divergência na codificação final) e o reinício da AC pela fase 6. Ressaltamos, entretanto, que se todos os passos foram seguidos adequadamente e se houve um tempo apropriado para a realização da codificação final, esse cenário será totalmente excepcional e não a regra.

Não obstante, reconhecemos que a realização de múltiplos testes de confiabilidade (já excluídos os realizados em caráter de treinamento) tende a aumentar consideravelmente os custos da pesquisa, aqui se incluindo o tempo para sua realização. Ademais, boa parte da literatura não reporta a realização de vários testes de confiabilidade, geralmente apresentando apenas um resultado, já excluindo os diversos trabalhos que nem buscam

⁵⁵ Ou seja, não é possível mais alterar a descrição ou escopo das categorias ou variáveis. Por outro lado, pequenos refinamentos, como a melhoria de uma descrição, ou o acréscimo de alguns exemplos não tende a macular a confiabilidade já atingida.

apresentar tais resultados (cf. KRIPPENDORFF, 2004; LOMBARD; SNYDER-DUCH; BRACKEN, 2002; LOVEJOY *et al.*, 2014; MACNAMARA, 2005). Em outro momento, afirmamos acreditar que o teste reportado deve ser relativo ao teste de confiabilidade realizado antes da codificação, uma vez que ele é o primordial para o início da codificação final em si (SAMPAIO; LYCARIÃO, 2018). Portanto, no mínimo, um teste de confiabilidade entre codificadores deve ser realizado com sucesso e seus valores devem ser devidamente apresentados na comunicação científica apropriada.

Análise

Estando a codificação e os testes de confiabilidade concluídos, a análise de conteúdo em si estará concluída; todavia, ainda há passos importantes para que a AC seja analisada e reportada de maneira pertinente.

10. Tabulação e aplicação de procedimentos estatísticos

Como supracitado, este livro trata da análise de conteúdo categorial quantitativa. Uma das principais vantagens desse tipo específico de AC é justamente a geração de frequências simples da presença ou ausência das categorias e códigos criados anteriormente. Ao fazer uso de uma planilha e manter apenas os números referentes às categorias e seus códigos, a primeira vantagem possível é a aplicação de procedimentos estatísticos que possam melhor responder às perguntas e hipóteses de pesquisa. Dessa forma, todo o Capítulo 3 é dedicado ao tema, dada a sua complexidade.

11. Interpretar e reportar os resultados

A interpretação dos resultados pode ser uma tarefa complexa na AC. Bardin (2016) afirma que a técnica chegou a ser quase abandonada na década de 1950 por uma impressão de seu baixo rendimento; afinal, a AC era excessivamente descritiva, o que teria mudado em sua adaptação como técnica que permite fazer inferências sobre as causas ou sobre os efeitos das características das comunicações a partir dos resultados das análises. A noção de inferência é importante na análise de conteúdo; afinal, o pesquisador usa

construtos analíticos para se mover do texto para as respostas de questões de pesquisa (WHITE; MARSH, 2006). Ou, como já tratamos, traduzir em quantificações conceitos abstratos a partir de interpretações de construções linguísticas.

Consoante Krippendorff (2004, p. 25), textos, mensagens e símbolos nunca falam por si. A AC “infere respostas para questões específicas de pesquisa de seus textos. As inferências são meramente mais sistemáticas, explicitamente informadas e (idealmente) verificáveis do que a interpretação de leitores comuns”. Exatamente por isso, o autor defende a necessidade de se informar o contexto (teórico, conceitual, histórico etc.) que guia a análise.

De acordo com Bardin (2016), a AC permite fazer inferências sobre: 1) o emissor ou produtor de mensagem⁵⁶, 2) o receptor⁵⁷, 3) a mensagem e 4) o *medium* (meio, canal, instrumento, objeto técnico). Boa parte da literatura trata justamente dessa ideia de causas e efeitos da comunicação (CAVANAGH, 1997; PRASAD, 2008; RIFFE; LACY; RICO, 2014; WEBER, 1990). Aqui, comunicação pode ser compreendida como quaisquer conteúdos verbais, visuais ou escritos, como já definimos. Ademais, é preciso não ficar preso à ideia de comunicação. Apenas a título de exemplo, Krippendorff (2004) sugere que é possível fazer inferências em termos de: extrapolações (tendências, padrões, diferenças); identificação, avaliações e julgamentos de padrões; índices; representações linguísticas; conversações; e processos institucionais.

A ideia de inferências (interpretações, conexões, deduções etc.) é largamente aceita pelos especialistas em AC; porém, alguns, a exemplo de Neuendorf (2002), tendem a considerar o conceito pouco produtivo, especialmente em se tratando de uma técnica quantitativa; Neuendorf (2002) defende que todas as interpretações devem seguir os princípios de métodos quantitativos e estatísticos para as respostas das questões de

⁵⁶ “Neste caso, insiste-se na função expressiva ou representativa da comunicação. Com efeito, pode-se seguir com a hipótese de que a mensagem exprime e representa o emissor” (BARDIN, 2016, p. 165).

⁵⁷ “Nesta ótica, insiste-se no fato de a mensagem se dirigir a este indivíduo (ou conjunto de indivíduos) com a finalidade de agir (função instrumental da comunicação) ou de se adaptar a ele (ou a eles). Por consequência, o estudo da mensagem poderá fornecer informações relativas ao receptor ou ao público” (BARDIN, 2016, p. 166).

pesquisa. Apesar de mencionarem a ideia de inferência, uma ideia similar é vista em Riffe, Lacy, Rico (2014).

Neste livro, entendemos que tanto a noção de inferências quanto de interpretações por via estatísticas são importantes. Na prática, as inferências, interpretações e conclusões devem responder às perguntas de pesquisa. Todo o Capítulo 3 deste livro é justamente voltado para uma interpretação à luz das técnicas já desenvolvidas e testadas em pesquisas quantitativas. Uma das vantagens da abordagem quantitativa é justamente a verificação da frequência, indicando a existência, intensidade e relativa importância de certo fenômeno sendo avaliado. Se a amostragem e as variáveis forem adequadas, os dados vão permitir o teste estatístico de hipóteses. Contudo, qualquer generalização para a população de textos/conteúdo fora daqueles que foram efetivamente codificados deve ser realizada com precaução e plena consciência de que tal movimento analítico é passível de diversas contestações (ver seção sobre amostragem deste capítulo). De todo modo, essas inferências quantitativas provavelmente não poderiam ser feitas com rigor científico de outro modo⁵⁸ (NEUENDORF, 2002; RIFFE; LACY; RICO, 2014; WEBER, 1990).

Consoante Krippendorff (2004), concluídas a análise de conteúdo, as inferências e as interpretações, é hora de narrar cientificamente os resultados; ou seja, de reportá-los adequadamente em produto científico (relatórios, *papers* de eventos, artigos, capítulos e livros, monografias, dissertações e teses). Com base no ressaltado pelo autor, achamos importante que qualquer produto científico que tenha feito uso da AC reporte:

- Revisão de literatura que contextualize a pesquisa em termos teóricos.
- Apresentação da abordagem epistemológica ou metodológica, incluindo as questões e hipóteses de pesquisa.

⁵⁸ Por outro lado, há evidentemente perdas semânticas na avaliação do conteúdo, que tendem a ser mitigadas em técnicas qualitativas.

- Apresentação do desenho de pesquisa, incluindo os preparativos, complicações e problemas (e como eles foram resolvidos), e também definições em termos de:
 - amostragem dos textos (estratégias e justificativas de escolha);
 - unidades de análise (definições operacionais, como foram usadas no processo);
 - processo de codificação (explicação dos construtos teóricos, disponibilização pública do livro de códigos, uso de codificadores e testes de confiabilidade).
- Os resultados da pesquisa (respostas às perguntas e hipóteses de pesquisa).
- Limites da pesquisa e autoavaliação: quais os limites dos resultados? Há sugestões metodológicas? Os esforços foram válidos? Trouxeram resultados relevantes?
- Conclusão: onde a pesquisa se encontra? Estes resultados vão ao encontro ou de encontro àqueles da literatura? O que podemos generalizar e o que é especificidade do estudo? Quais são os mais promissores indicativos de futuras pesquisas?

12. Validação e replicabilidade

Como já dito, compreendemos validade enquanto adequação epistemológica entre os propósitos da pesquisa e os instrumentos utilizados para identificar o fenômeno sob investigação. Refere-se, assim, à capacidade dos construtos analíticos criados e/ou empregados em traduzirem o conceito, a teoria ou simplesmente o fenômeno que se deseja apreender. Reforçando as palavras de Neuendorf (2002, p. 112): “estamos mensurando o que almejamos mensurar?”.

Como já discutimos em outro texto (SAMPAIO; LYCARIÃO, 2019), a validade tende a ser inversamente proporcional à confiabilidade, afinal, quanto mais simples for o parâmetro (*i.e.*, código), mais fácil de ser aplicado

e reproduzido (confiabilidade); porém, menos representa o fenômeno ou conceito que se deseja (validade). Enquanto alguns manuais tendem a apresentar formas complexas (e muitas vezes questionáveis) para averiguar a validade do construto teórico e mesmo dos resultados, acreditamos que a melhor avaliação da validade ocorre de forma transacional. Como já dito, a validade de uma AC é dependente do quão convincente ela se mostra à comunidade para a qual ela é dirigida. Como aqui estamos pensando em uma AC enquanto técnica científica, o mais recomendável é que o próprio campo científico no qual a pesquisa se insere seja justamente o melhor árbitro para avaliar a validade do estudo em tela.

Porém, a simples apresentação dos resultados pode ser pouco para uma avaliação adequada. Como já insistimos em outros pontos, o ideal é que os pares possuam mais material para analisar a qualidade da AC em questão. Nomeadamente, é preciso tornar o livro de códigos completo disponível. No caso de trabalhos de conclusão, como monografias, teses e dissertações, ele deve vir como anexo, enquanto no caso de *papers* e artigos, ele deve ser tornado disponível *online* e o *link* deve ser indicado no texto⁵⁹. O LdC, em conjunto com a revisão teórica e a pergunta de pesquisa, são premissas básicas para a avaliação da validade de qualquer AC. Então, durante os eventos científicos e/ou durante a realização dos pareceres cegos para os periódicos, os pares poderão avaliar se a pesquisa em questão apresenta um aceitável nível de validade para ser publicada ou não.

Tornar disponível o LdC também é essencial para a replicabilidade do estudo. Entretanto, no caso desta, também se faz necessária a acessibilidade ao mesmo material analisado, estando o mesmo em condição equivalente ou suficiente para uma nova codificação externa ao contexto da pesquisa original. Em outras palavras, é necessário disponibilizar o material e/ou banco de dados que serviu de base para a pesquisa. Atualmente, alguns dos periódicos mais qualificados já solicitam que o LdC e que o banco de dados também sejam enviados em conjunto com o artigo a ser considerado. Também

⁵⁹ Como é o caso da pesquisa sobre populismo em *releases* para imprensa na Alemanha. Disponível em: https://populism.byu.edu/App_Data/DataSetFiles/codebook_franzmann2016_afd_oct2016.pdf.

já temos disponíveis repositórios de bancos de dados, que geralmente são liberados após a publicação original da pesquisa⁶⁰. Também é usual (como parte da ética científica) enviar o banco de dados quando solicitado por um par acadêmico, pois a condição de replicabilidade é uma base da AC e também da ciência.

Portanto, sugerimos que: 1) o LdC já seja disponibilizado assim que a pesquisa for apresentada e que 2) o banco de dados seja disponibilizado, caso solicitado (*e.g.* por um periódico ou por partes) ou depois da publicação do artigo científico (assim, a originalidade da pesquisa já estará garantida)⁶¹. Em ambos os casos, o ideal é a publicação em repositórios institucionais e fixos e não em páginas aleatórias da internet, que poderão deixar de existir no futuro breve.

⁶⁰ Um bom exemplo é o Consórcio de Informações Sociais (CIS) criado pela Anpocs para ser repositório de pesquisas, LdC e banco de dados de todo tipo de pesquisa em Ciências Sociais. Disponível em: <http://www.nadd.prp.usp.br/cis/index.aspx>.

⁶¹ O *Mannheimer Zentrum für Europäische Sozialforschung* (MZES) torna disponível os bancos de dados de suas pesquisas: <https://www.mzes.uni-mannheim.de/d7/en/datasets-list>.

CAPÍTULO 3: ANÁLISE DE RESULTADOS E TESTES ESTATÍSTICOS

As análises estatísticas permitem estabelecer a análise quantitativa de eventos e as características de uma população. Isso pode se dar de maneira mais elementar por intermédio da estatística descritiva, o que compreende técnicas e maneiras mais usuais, como frequências e gráficos (de pizza, barras, por exemplo.), medidas de tendência central (média, mediana, por exemplo) e dispersão (desvio padrão, por exemplo). Porém, técnicas estatísticas mais complexas, geralmente associadas à estatística inferencial, também são recursos úteis para algumas pesquisas.

Contudo, alertamos que a utilização de recursos estatísticos mais complexos, tais como análise de correlação, regressão e modelos estatísticos multivariados deveria ser supervisionada ou mesmo realizada por profissionais especializados no campo. De outro modo, há considerável chance de se empregar técnicas inadequadas ao tipo de análise que se pretende, ou mesmo de se inferir conclusões que os resultados estatísticos não permitem com segurança razoável.

Feita essa observação, esclarecemos que o intuito deste capítulo fundamentalmente consiste em trazer as noções básicas necessárias para que pesquisadores iniciantes na AC e também em estatística possam formular o desenho básico de produção dos resultados. De todo modo, recomendamos, além de assistência especializada, a utilização do material bibliográfico aqui referenciado para melhor iniciação na análise estatística de dados provenientes de AC.

3.1 Definição de variável e seus tipos

Até agora, este livro tem se referido ao conceito de categorias para designar, na linguagem da análise de conteúdo, aquilo que, na linguagem estatística, corresponde às chamadas **variáveis**. Uma variável, como o próprio nome sugere, refere-se a uma característica da população que é passível de apresentar diferentes resultados (portanto, uma variação) ao

longo das observações dos casos. Assim, unidades textuais podem apresentar diferentes quantidades de palavras, temas, valências etc.

Ao se observar o leque bastante amplo de técnicas estatísticas disponíveis, imediatamente surge a questão: qual técnica utilizar para analisar meus dados? Para saber qual técnica utilizar, seja ela simples ou complexa, é necessário ter muito claro como os componentes empíricos (os materiais semânticos) são resultado de uma tradução teórica que estrutura a relação ou investigação desses mesmos componentes. Em suma, deve-se ter clareza absoluta da pergunta de pesquisa ou das hipóteses que estão conduzindo o esforço da investigação empírica. Uma vez tendo isso claro, é preciso identificar os componentes empíricos de investigação em termos de correspondência a certos tipos de variável.

Diferentes tipos de variáveis aparecem de maneira evidente nos exemplos anteriormente referidos, quando se leva em conta que “número de palavras” é um tipo de variável que apresenta um valor quantitativo em cada observação, ou seja, um número. Já a variável “tema”, por outro lado, apresenta, no lugar de um número, um conjunto de códigos, tais como “política”, “crime”, “esportes” etc., sendo, por isso, chamada de variável qualitativa nominal. Já “número de palavras” seria uma variável quantitativa discreta, pois trabalha apenas com números inteiros, recusando frações.

No caso de variáveis qualitativas nominais, apenas pela soma de todas as observações da amostra ou da população é que é possível chegar a um indicador quantitativo, tais como a percentagem de notícias com o tema crime em contraste com o tema “política”, por exemplo. Se de 100 notícias identificadas, 80 tratam do tema crime e apenas 20 de política, então se pode deduzir que 80% da amostra estudada possui notícias voltadas ao tema crime, enquanto 20% voltadas ao tema política. Esse tipo de variável é

chamado de qualitativa nominal porque, em cada observação (codificação), o que se atribui à unidade é uma categoria e não um número⁶².

Tal diferença na denominação dos tipos de variáveis já indica por qual razão as variáveis na análise de conteúdo são chamadas de categorias. Isso porque uma análise de conteúdo, a rigor, está, em um primeiro momento, voltada a produzir análises que identifiquem conceitos e significados dos textos em forma de categorias (as subdivisões/classificações de uma variável qualitativa). Apenas posteriormente (após a codificação) é que as variáveis qualitativas e suas categorias podem se tornar objeto de quantificações como frequências, proporções, taxas e percentagens. Feitos esses esclarecimentos, cabe agora compreender os diferentes tipos de variáveis:

Variáveis qualitativas

- **Nominal:** São variáveis qualitativas que classificam as unidades semânticas em categorias, mas que não estabelecem nenhum tipo de ordem entre as categorias, tais como “tema”, “enquadramento” e “metodologia de avaliação” (ver GURGEL; LEITE, 2007).
- **Ordinal:** São variáveis qualitativas que classificam as unidades semânticas em categorias que se organizam a partir de uma ordem, ou “ordenam as unidades observadas segundo processo de comparação” (CERVI, 2017, p. 42). Isso faz com que as unidades analisadas possuam valor hierárquico distintamente entre si, a exemplo do “nível de justificação”, “êxito” e “posicionamento” (ver Figura 1 e Figura 6 deste livro). Por ser de tipo ordinal, isso implica que as categorias dessas variáveis possuem uma ordem. Assim, para a variável “nível de justificação”, por exemplo, ela deve comportar o menor nível de justificação (nível 0), até o maior (nível 3).

⁶² Note que, conforme explicado no capítulo anterior, o formulário de codificação será usualmente preenchido com números; porém, o fato de se atribuir esse número (apenas em termos de preenchimento) não torna a variável quantitativa. Por isso, é preciso informar ao software se aquele número representa uma categoria (subtipo nominal ou ordinal) ou uma quantidade (subtipo contínua ou discreta).

Variáveis quantitativas

- **Discreta:** Uma variável quantitativa é considerada discreta quando ela assume apenas números inteiros, recusando, portanto, a possibilidade de frações. Derivam geralmente de contagens de unidades de interesse. No caso da análise de conteúdo, usa-se esse tipo de variável para determinar os elementos mais estruturais de um *corpus* semântico. No caso de filmes, o número de cenas; no caso de um livro, o número de páginas, capítulos, parágrafos, palavras etc.
- **Contínua:** Uma variável contínua, diferentemente da discreta, assume frações e, portanto, expressa intervalo ou composição de números reais. Na análise de conteúdo, esse tipo de variável tende a surgir a partir da composição de variáveis mais simples, em que dados primários são compostos de modo a gerar uma nova medida e, portanto, uma nova variável. Tal tipo de operação é bastante útil quando se deseja comparar dois conjuntos semânticos (duas *corpora*) com tamanhos bastante distintos.

A título de exemplo, imagine que um pesquisador deseja comparar dois programas policiais quanto ao uso do termo “bandido” para se referir a indivíduos acusados de algum crime. Após uma análise de conteúdo, observou-se que os dois programas possuíam as mesmas frequências para o termo. Em cada um dos programas, observou-se que o termo “bandido” foi usado 40 vezes para se referir a um acusado de cometer crime. No entanto, também se observou que os programas possuíam diferentes tipos de duração, um com 15 minutos e o outro com 30 minutos.

Tal situação faz com que a simples contagem da frequência da palavra de interesse não seja adequada para comparar os programas entre si e, portanto, inferir que os dois programas recorrem ao termo “bandido” com a mesma grandeza quantitativa. Isso porque a incidência do termo no programa de maior duração, ainda que na mesma quantidade do outro programa (o de menor duração), acaba revelando que há uma menor incidência ou ênfase no uso do termo. Enquanto, no programa de maior duração, as 40 aparições

do termo ficaram distribuídas ao longo de 30 minutos; no outro, as mesmas aparições ficam distribuídas num intervalo de apenas 15 minutos. Tal tipo de situação gera a necessidade de compor uma nova variável que seja capaz de gerar uma comparação adequada que efetivamente compare a incidência e ênfase de aparição do termo, levando em consideração a extensão do *corpus*, o que, no caso hipotético aqui, trata-se da duração em minutos de um programa.

Para isso, é preciso formar uma variável de tipo contínua a partir da composição de duas variáveis discretas com dados primários. A primeira variável discreta se refere à contagem ou número de vezes em que a identificação do acusado pelo termo “bandido” aparece no programa, o que podemos denominar de variável “Id”. A segunda variável discreta se refere ao tempo de duração do programa em minutos, o que podemos denominar de variável “Tmin”. Como a variável a ser gerada pela composição dessas variáveis discretas prevê que os valores de “Id” sejam relacionados a uma distribuição no tempo (expresso pela variável “Tmin”), trata-se de uma taxa. Assim, podemos denominar a variável resultante de Tx. Isso pode ser traduzido a partir da seguinte equação:

$$Tx = Id/Tmin$$

Se convencionarmos denominar o programa de 15 minutos como sendo o caso “1” e o programa de 30 minutos como sendo caso “2”, a equação para o cálculo da taxa do uso termo “bandido” pode ser escrita do seguinte modo (para cada caso):

$$\text{Caso 1: } Tx_1 = 40/30 = 1.3$$

$$\text{Caso 2: } Tx_2 = 40/15 = 2.6$$

Assim, o programa de 30 minutos terá a taxa de 1,3 de uso do termo “bandido” por minuto, enquanto o programa de 15 minutos terá uma taxa de 2,6 por minuto do uso do termo. Esse exemplo mostra que, a partir de dados primários da análise de conteúdo provenientes de variáveis discretas, é possível compor uma variável contínua capaz de gerar o tipo de medida estatística mais apropriada para os objetivos de uma pesquisa.

3.2 Os tipos de variáveis pelo modelo estatístico

As variáveis se diferenciam em tipos, não apenas em correspondência ao tipo de informação que geram (qualitativa x quantitativa) e seus subtipos (nominal, ordinal x discreta, contínua), mas também em função do papel que a variável ocupa no modelo explicativo adotado ou elaborado pela pesquisa.

Quando se fala em modelo explicativo, estamos nos referindo às relações entre os componentes (ou variáveis para a estatística) que estão sendo analisados pela pesquisa. No caso, quanto mais elementos são investigados, mais variáveis surgem e, com isso, o nível de complexidade do modelo explicativo tende a se elevar. Para organizar as relações entre as variáveis e tornar essa complexidade inteligível, a estatística prevê três tipos fundamentais de variáveis: dependente, independentes e de controle.

Dependente

A variável dependente, também chamada de variável resposta, é aquela que está sendo explicada, gerando assim o ponto de partida e a busca por outras variáveis que possam explicar o comportamento da variável dependente, ou de interesse. No caso das pesquisas sobre políticas públicas, a variável dependente pode ser o problema social que se pretende compreender, a exemplo da criminalidade, do nível de educação da população, de saúde etc. No exemplo dado acima sobre os programas policiais, a variável de interesse ou a ser explicada seria a incidência do uso do termo “bandido” para se referir a acusados de cometer crimes.

Independente

A variável independente, também chamada de explicativa, será usada e testada para explicar o comportamento da variável dependente. No caso, as políticas públicas e a qualidade das mesmas tendem a ser assumidas como variáveis que incidem sobre problemas sociais e que, portanto, ajudam a explicar o comportamento de variáveis dependentes citadas acima, a exemplo da criminalidade, do nível de educação da população, de saúde etc.

No exemplo dos programas policiais (exemplo acima), vimos que um programa apresentou uma incidência do uso do termo “bandido”, para se referir a acusados de cometer crime, maior do que o outro programa. Mas o que explica essa diferença? A política editorial do programa? A política editorial da emissora? A profissão e o nível de formação do apresentador? O perfil da audiência? A partir do momento em que essas variáveis são investigadas como sendo explicativas e de modo quantitativo, as análises e os dados devem mostrar até que ponto essa relação, de fato, se confirma ou se, ao contrário, tais variáveis não conseguem explicar de modo suficiente o comportamento da variável dependente.

De controle

É o tipo de variável que consegue interferir “de maneira indireta na variável dependente” (CERVI, 2017, p. 45). Nesse caso, se a adoção de uma política pública se mostrar eficaz no combate à criminalidade, é plausível que existam variáveis que afetem a eficácia da implantação da política e que não concernem à própria política pública, a exemplo do nível de experiência dos gestores e executores da política, a estabilidade do financiamento etc.

Uma vez compreendidas as noções básicas dos principais tipos de variáveis que ajudam a traduzir estatisticamente os componentes envolvidos numa pesquisa de AC, passaremos a delimitar as principais opções de análise estatística passíveis de serem realizadas após a codificação do material de pesquisa. Opções essas que, como veremos adiante, são moldadas e refinadas por meio das premissas e hipóteses da pesquisa desde seu início.

3.3 Análise descritiva

A análise estatística de tipo descritiva é aquela que organiza os dados em termos de medidas descritivas, tais como frequência, contagem e medidas de tendência central e dispersão, a exemplo da média, da mediana e do desvio-padrão. Medidas essas que, para efeito de inteligibilidade mais ampla de apresentação dos resultados, são usualmente apresentadas em tabelas e gráficos. Justamente por sua inteligibilidade e pelo seu caráter de

constituição básica das análises estatísticas, tais tipos de medidas estão entre as mais recorrentes na apresentação dos resultados.

Contagem, frequência e percentagem

A contagem de um evento diante da população pesquisada (ver definição de população na Seção 5 do Capítulo 2) é que permite extrair sua frequência, ou, mais precisamente, sua frequência relativa. Isso porque a própria contagem de um evento é considerada uma frequência, mas, no caso, de tipo absoluto. Para se chegar à frequência relativa, deve-se contrastar a contagem da população (todo o material coletado e analisado) com a contagem de um determinado componente ou evento (que se traduz em uma variável). Ao realizar esse contraste, chega-se então a uma medida de frequência relativa. Para se obter, então, uma percentagem, basta multiplicar o valor da frequência relativa por 100.

Gurgel e Leite (2007), por exemplo, recorrem a esse recurso. Ao entrevistar 40 docentes do ensino superior, os autores relatam que a análise de conteúdo das entrevistas revelou que 60% dos entrevistados disseram que se dedicam mais ao ensino do que à pesquisa. Para chegar a essa percentagem, foi preciso contar a frequência absoluta dos docentes que deram essa informação, o que corresponde a 24 professores, e contrastá-la com a frequência absoluta da população estudada, no caso, 40 docentes.

Ainda que o recurso da percentagem seja bastante comum, deve-se atentar que, para amostras ou populações com um número pequeno de unidades (n baixo), o uso de tal recurso faz pouco sentido. Isso porque o próprio termo “percentagem” sugere uma escala-padrão de 100 que é bastante útil para tornar inteligível os resultados de amostras ou populações com muitas unidades (próxima ou superior a 100). Porém, para amostras ou populações pequenas, o uso desse tipo de medida pode levar a crer que a amostra ou população seria maior do que efetivamente é. Se apenas três professores fossem entrevistados, por exemplo, dizer que um deles afirmou se dedicar mais à pesquisa do que ao ensino se mostra mais adequado do que relatar a percentagem de 33,3%. Em suma, para se usar percentagem no

lugar da frequência absoluta ou relativa, deve-se observar se o tamanho da amostra ou da população é adequado para esse tipo de recurso.

Medidas de tendência central

Medidas de tendência central indicam os valores típicos, mais frequentes ou mais representativos de uma amostragem ou população analisada. Nas palavras de Cervi (2017, p. 105), tais medidas formam “um tipo de sumarização de dados, pois descreve o meio ou ponto central da distribuição, representando-a”.

As medidas mais comuns desse tipo são a média, a moda e a mediana. Abaixo, uma breve descrição de cada uma e o que torna cada uma útil na apresentação dos resultados de uma análise de conteúdo.

- **Média:** Pode ser aritmética ou ponderada. A aritmética é realizada mediante a soma dos valores de cada unidade, dividindo-se o resultado da soma pela quantidade total de unidades observadas. Já a média ponderada estratifica a média em subgrupos da amostra ou da população analisada, tendo a possibilidade de ponderar a medida em questão em função do tamanho de cada subgrupo ou mesmo em função de pesos diferentes atribuíveis a eles. Para exemplos e uma explicação mais detalhada, ver Cervi (2017, p. 107-109).

A média é bastante útil na análise de conteúdo, sendo a mais comum das medidas centrais, mas deve ser usada para variáveis quantitativas e evitadas para variáveis categóricas ordinais (ver RIFFE; LACY; FICO, 2014, p. 142).

- **Moda:** Trata-se do valor ou classificação mais recorrente das unidades analisadas. É aplicável para variáveis tanto quantitativas quanto qualitativas. Contudo, seu uso tende a se mostrar mais útil em variáveis categóricas, pois, no caso de variáveis quantitativas, a variedade de valores tende a ser maior, fazendo com que o valor informativo da moda seja pouco informativo em comparação a variáveis categóricas (qualitativas).

- **Mediana:** A mediana estabelece o valor central entre os dois extremos de uma população cujas observações estejam ordenadas em valores crescentes. Portanto, o valor da mediana possui equidistância entre a unidade com menor e a com maior valor. Não por acaso, esse é um tipo de medida recomendável “quando a distribuição apresenta valores extremos” (CERVI, 2017, p. 110). Importante notar que, em distribuições simétricas como a normal, a média, a moda e a mediana apresentam o mesmo valor.

Medidas de dispersão

Como vimos anteriormente, as medidas de tendência central são bastante úteis para sintetizar uma grande quantidade de dados em um único valor ou em alguns deles. Contudo, ao realizar essa síntese, algumas informações importantes podem ficar ausentes. Uma das mais relevantes, nesse caso, corresponde saber o quão a média, a moda ou a mediana, de fato, representam a amostra ou população estudada. Essa questão fica em aberto quando não se sabe o quanto os valores dessa mesma amostra ou população variam. Isso é importante, pois caso os valores sejam muito distantes entre si, as medidas de tendência central acabam tendo pouco valor informativo e compreensivo sobre a amostra ou população em análise.

Tal questão fica bastante nítida quando pensamos em medidas como renda *per capita*, por exemplo. Esse tipo de medida indica o quanto há de recursos financeiros disponíveis para cada indivíduo de uma população de um país, pois a renda *per capita* é calculada dividindo-se a renda nacional pelo tamanho da população. Tal tipo de medida fornece, assim, uma medida capaz de dizer quão rica é a população de um país, mas apenas de maneira superficial. Isso porque, em países muito desiguais (como é o caso do Brasil), alguns poucos indivíduos concentram grande parte dos recursos financeiros. Isso faz, portanto, com que a maior parte das pessoas possua uma renda bem abaixo do que a renda *per capita* do país.

Então, para saber o quão uma população possui certa característica (a exemplo da renda) distribuída, as medidas de dispersão são bastante úteis.

A mais comum delas é o desvio-padrão, que, segundo Bussab e Morettin (2013, p. 41), “indica, em média, qual será o “erro” (desvio) cometido ao tentar substituir cada observação pela medida resumo do conjunto de dados (no caso, a média)”. O desvio padrão é calculado, desse modo, por meio da raiz quadrada da média quadrática das diferenças de cada unidade analisada em relação à média. Para detalhes técnicos sobre como calcular essa e outras medidas de dispersão, recomendamos o manual de Cervi (2017, p. 114-121).

3.4 Estatística inferencial

O uso de técnicas derivadas da estatística inferencial não é tão comum na análise de conteúdo quanto o das derivadas da estatística descritiva. Isso acontece provavelmente por uma série de razões. A primeira e mais intuitiva delas deriva da constatação realizada no capítulo anterior (Seção 5) de que a análise de conteúdo, pelas suas próprias características, tende a violar algumas premissas da representação estatística (base da inferência estatística), o que torna o uso de técnicas de estatística inferencial passível de questionamentos estruturais.

Ademais, há razões bem mais simples. Isso porque as técnicas de estatística inferencial pressupõem técnicas da própria estatística descritiva, que abriga um conjunto básico de instrumentos que são necessários para compor as técnicas mais avançadas de estatística inferencial. Ou seja, mesmo quando se utiliza a estatística inferencial, também se recorre a técnicas da estatística descritiva, que são, portanto, de uso mais universal do que aquelas. Também se deve lembrar que as técnicas de estatística descritiva são bem mais acessíveis e compreensíveis ao público em geral, o que as tornam mais eficazes na divulgação científica. Portanto, faz sentido, sob esse ponto de vista, que elas sejam mais recorrentes.

Contudo, o elemento mais determinante que estabelece a necessidade ou não do uso de técnicas de estatística inferencial corresponde à ambição epistêmica do desenho da pesquisa. Uma ambição precisamente em relação ao grau de representatividade e precisão que o recorte empírico analisado (a amostra) busca obter em relação a uma dada população (para a diferença entre o conceito de população e amostra na estatística, ver Capítulo 2, Seção 5).

Para ilustrar isso, voltemos ao caso da pesquisa de Gurgel e Leite (2007) que, ao realizarem uma AC de entrevistas com 40 docentes da UFPI (Universidade Federal do Piauí), chegaram ao dado de que 60% dos entrevistados afirmaram se dedicar mais ao ensino do que à pesquisa. De posse desse dado, o quão representativo e preciso ele seria para descrever a realidade de todas e todos os docentes da UFPI? Ou sendo ainda mais ambicioso, o quão esse mesmo dado seria representativo de todos os docentes do Piauí ou mesmo do Brasil?

As técnicas desenvolvidas pela estatística inferencial buscam justamente fazer com que esses tipos de pergunta possam gerar dados capazes de informar o nível de precisão e probabilidade com que eles representariam informações de uma população. Porém, para isso, a estatística inferencial estabelece técnicas que envolvem não apenas a análise dos dados já previamente obtidos de uma amostra, mas fundamentalmente o próprio método de amostragem, seu tamanho, formas de extração dos dados etc. Ou seja, para fazer com que obtivéssemos, por meio de uma amostragem, a percentagem de docentes da UFPI, do Piauí ou do Brasil que se dedicam mais ao ensino do que à pesquisa, seria necessário planejar, com base nas premissas da estatística inferencial, a quantidade de docentes a serem entrevistados, o método da entrevista, de abordagem para obter os participantes da pesquisa etc. Isso tudo antes mesmo de se ir a campo. Ou seja, a ambição epistêmica de uma pesquisa, ao recorrer à estatística inferencial, gera consequências não apenas sobre como os dados obtidos serão analisados, mas também sobre todas as etapas de produção desses dados. Por isso, antes de utilizar técnicas de estatística inferencial para se analisar dados da sua pesquisa de análise de conteúdo, deve-se saber se seu uso:

(a) é possível ou recomendável (não o é, por exemplo, para amostras pouco representativas probabilisticamente);

(b) se é necessária (não o é, para parâmetros da população⁶³, em caso de censos); e

⁶³ A exemplo da média, desvio-padrão, moda etc.

(c) se as técnicas de análises de dados são adequadas às características das variáveis e às ambições epistemológicas previamente estabelecidas pela pesquisa.

Esta seção se volta a prover assistência didática inicial em relação ao item “c”, o que, como já indicado, pressupõe que os dados tenham sido coletados já em acordo com os princípios e técnicas da estatística inferencial (para isso, ver Capítulo 2, Seção 5). Ao prover essa assistência didática, insiste-se em relação a seu caráter inicial e que, portanto, tem como principal utilidade ser um primeiro passo no aprendizado dessas técnicas.

Por ser um primeiro passo, serão apresentados apenas os princípios e as técnicas mais úteis e simples para a análise de conteúdo. Para a utilização autônoma dessas técnicas em nível apropriado e para o reconhecimento mais amplo das demais técnicas e princípios da estatística inferencial, será necessário, portanto, prosseguir e avançar no aprendizado se utilizando da literatura mais especializada referenciada aqui e, também, dos diversos vídeos, cursos, programas e pacotes estatísticos (disponíveis em *R*, *Python*, *Stata*, *SPSS* etc.) existentes.

Inferência estatística como algo improvável:

Na seção sobre estatística descritiva deste capítulo, passamos por algumas medidas de tendência central e de dispersão, sendo a média e o desvio-padrão as medidas correspondentes mais utilizadas de cada tipo de parâmetro populacional em questão. Tais medidas foram apresentadas partindo do pressuposto de que toda a população foi analisada, ou seja, um censo. Porém, caso uma pesquisa ambicione, a partir de uma amostra, chegar à média ou a outro parâmetro dessa população como um todo, como isso seria possível?

A resposta a essa questão é dada justamente pelos fundamentos da inferência estatística, a qual, partindo de amostras aleatórias da população, consegue estimar qual seria o valor aproximado dos parâmetros e características dessa população. No caso, a natureza aleatória (probabilística) da amostra é fundamental para realizar essa aproximação. Isso porque:

A noção fundamental de uma amostragem aleatória é que cada indivíduo da população tem a mesma chance de ser selecionado na amostra. Se isso ocorre, as características encontradas com mais frequência na população – seja em novelas, em notícias de jornal ou em poemas – também se mostrarão mais frequentes na amostra, assim como as características menos frequentes na população se mostrarão menos frequentes na amostra. Isso ocorre pelas leis da probabilidade (RIFFE; LACY; FICO, 2014, p. 77).

São justamente essas leis que permitem predizer que, se selecionarmos aleatoriamente um bom número de pessoas em uma população que possui a mesma proporção de homens e mulheres, há uma boa probabilidade de que a proporção de homens e mulheres dessa seleção (amostra) seja próxima da existente na população como um todo. Tal fenômeno é descrito pelo teorema central do limite, uma lei probabilística que diz que a distribuição das somas padronizadas de variáveis aleatórias de uma amostra probabilística da população tende, com o aumento do tamanho da amostra, a aproximar-se de uma distribuição normal⁶⁴ (ver LUDWIG-MAYERHOFER, LIBESKIND, GEISSLER, 2014, p. 114).

Em termos práticos, isso significa que, **quanto maior for o tamanho da amostra, maior é a probabilidade de que a características dessa amostragem sejam próximas àquelas encontradas na população como um todo**. Quando se fala em probabilidade maior ou menor, isso implica que a estatística inferencial não estabelece qualquer certeza absoluta de que as estimativas de uma amostra, de fato, correspondam aos parâmetros da população. O melhor que uma estatística inferencial exigente consegue fazer é indicar que há uma probabilidade muito pequena de que uma estimativa obtida da amostra (a média, por exemplo) não esteja muito próximo ao parâmetro da população que seria encontrado caso realizássemos um censo.

Mas qual seria o limiar para essa probabilidade muito pequena? A determinação desse limiar é dada *a priori* pelo desenho da pesquisa e é

⁶⁴ “Uma distribuição normal segue o formato de um sino, simetricamente distribuído ao redor da média, ou seja, o mesmo número de caso à esquerda e à direita da média. Quando existem mais casos de um dos lados da média, diz-se que a curva não segue uma distribuição normal” (CERVI, 2017, p. 94).

indicada pela probabilidade de significância estatística, o que é chamado de valor p . Segundo Riffe, Lacy e Fico (2014, p. 143), as pesquisas sociais convencionam os valores p como sendo $p < 0,05$ ou $p < 0,01$, o que indica, respectivamente, de 95% e 99% de probabilidade (ou, no jargão estatístico, **nível de confiança**) de que os valores encontrados se encontram dentro de uma margem de erro previamente estabelecida. Atentar, no caso, de que quanto menor for o valor de p , maior é o nível de confiança.

O valor - p e os 2 tipos de erros nos testes de hipótese:

O valor p é usado não apenas para se chegar ao nível de confiança de uma estimativa de uma amostra dentro de uma margem de erro (como falamos anteriormente e com mais detalhes no tópico abaixo), mas também para indicar, no contexto de um teste de hipótese (ver próxima seção deste capítulo), o nível de confiança em se rejeitar a hipótese nula. No caso, a hipótese nula estipula que os resultados do teste derivam do mero acaso, não representando relações ou parâmetros previstos pela hipótese de pesquisa. Assim, o valor $-p$ também é utilizado nos testes estatísticos de hipótese.

Nos testes de hipótese, há dois tipos de erro. O primeiro, denominado de alfa (ou α), consiste em rejeitar a hipótese nula, ainda que verdadeira. Já o segundo, denominado de beta (ou β), consiste em não se rejeitar a hipótese nula, ainda que essa seja falsa. Segundo Cervi (2017, p. 178), é possível calcular apenas a probabilidade do erro tipo 1, pois “do ponto de vista estatístico, nunca se tem certeza de estar acertando ou errando ao rejeitar a hipótese nula”. Por isso, o nível de confiança de uma estatística inferencial é algo que, nos padrões mais exigentes, indica um alto nível de probabilidade de que a conta não seja fruto de um evento altamente improvável (como o é, por exemplo, o de ser sorteado na Mega-Sena). A possibilidade de erro (assim como de ser sorteado e se tornar, da noite para o dia, milionário) sempre existe. Por isso, podemos dizer que uma estatística inferencial exigente e metodologicamente recomendável é aquela que faz o erro tipo 1 se aproximar do altamente improvável.

Uma vez compreendidos os fundamentos teóricos básicos da estatística inferencial, na próxima seção, iremos compreender o significado e as principais técnicas na inferência sobre os parâmetros de uma população, assim como planejar um tamanho amostral em que se saiba de antemão a probabilidade de a conta não estar errada dentro de uma certa margem de erro.

Inferência sobre os parâmetros de uma população

No começo deste capítulo, deixamos claro que o uso da estatística inferencial está condicionado a uma série de pressupostos. Um dos mais relevantes se refere à ambição epistemológica do estudo em conseguir, a partir de uma seleção limitada de casos, chegar a informações (parâmetros) sobre uma dada população. Como vimos acima, a obtenção de tais informações é possível através de uma amostra probabilística (em que a seleção aleatória dos casos esteja em alguma medida presente), mas, ainda assim, restrita a um nível de confiança. Porém, quando se almeja obter informações de uma população por intermédio desse tipo de procedimento, o nível de confiança não é a única restrição em jogo, mas também a margem de erro. No caso, ela corresponde ao intervalo de variação que é tolerado pelo próprio desenho da pesquisa, em relação ao ponto de estimação encontrado após a análise da amostra.

Para ilustrar isso com o caso mencionado no começo deste tópico sobre estatística inferencial, imagine que uma pesquisa se propusesse a produzir uma amostra estatisticamente representativa dos docentes das universidades federais do Brasil, para saber qual seria a percentagem (aproximada) dos que afirmam se dedicar mais ao ensino do que à pesquisa. Para fazer com que a percentagem encontrada nessa amostra possa ser estatisticamente representativa da população total dos docentes, é preciso, além de estabelecer de antemão o nível de confiança (vamos convencionar que ela seja de 95%, ou seja o valor $p < 0,05$), também estabelecer qual é a margem de erro tolerada. No caso, se a pesquisa estabelecer três pontos percentuais como margem de erro e encontrar que 60% dos docentes

afirmam se dedicar mais ao ensino, isso significa que a percentagem se situa entre 57% e 63%, a um nível de confiança de 95%. Em outras palavras, há 95% de chance de que o valor real dos docentes que dizem se dedicar mais ao ensino do que à pesquisa se situar nesse intervalo. E lembrar que há também 5% de chance de não estar em tal margem de erro.

Algo importante a se observar é que a margem de erro, quando se torna muito tolerante (acima de 5 pontos percentuais), passa a ter menor precisão e, portanto, menor valor informativo. Do mesmo modo que, quanto maior for o limite crítico do teste de significância estatística (o valor- p), menor será o nível de confiança dos resultados. Para que possa ter um alto nível de precisão (uma margem de erro pouco tolerante) e um alto nível de confiança, a pesquisa precisa de um tamanho de amostra (designado como n) também alto.

Para ilustrar isso, voltemos ao nosso exemplo. Quantos docentes seriam necessários para obtermos estimativas dos parâmetros da sua respectiva população a um nível de 95% de confiança e com uma margem de erro de três pontos percentuais? Para realizar esse tipo de cálculo, há diferentes aplicativos disponíveis *on-line* que demandam, para fins do cálculo, além da margem de erro e do intervalo de confiança, o tamanho total da população. Ou seja, o ideal é que saibamos o número total da população.

Contudo, vamos imaginar, para fins de um procedimento mais próximo de uma análise de conteúdo (em que o tamanho total da população muitas vezes é desconhecido), que não saibamos o número total de docentes. Para, então, saber a quantidade deles a se entrevistar, podemos calcular o tamanho necessário da amostra utilizando um valor certamente mais alto do que o valor real (no caso, 50.000.000, cinquenta milhões). Ao realizar o cálculo⁶⁵ para o tamanho da amostra, chegamos a 1.067. Ou seja, para obtermos estimativas de parâmetros de uma população, cujo tamanho total é desconhecido, a um nível de 95% de confiança, com três pontos de margem de erro, necessitaríamos selecionar 1.067 docentes para a pesquisa.

⁶⁵ Utilizamos a calculadora da página *The Survey System*, disponível em: <https://www.surveysystem.com/sscalc.htm>.

Caso almejemos tornar a pesquisa ainda mais confiável (no caso, com 99% de nível de confiança) e precisa, diminuindo a margem de erro de três para dois pontos percentuais, precisaríamos entrevistar 4.160 docentes. Ou seja, o aumento da precisão e do nível de confiança de uma amostra probabilística faz com que n também aumente. Por isso que amostras pequenas para populações de tamanho desconhecido tendem a obter um baixo nível de confiança e/ou baixa precisão (margem de erro muito ampla).

3.5 Testes estatísticos

Na estruturação de pesquisas que envolvem análise de conteúdo, o teste de hipótese tem se mostrado bastante produtivo e eficiente na organização das análises dos seus resultados. No caso, Neuendorf (2002, p. 168) explica que a hipótese é “uma asserção preditiva sobre a relação entre duas ou mais variáveis”. A autora reforça que a proposição de uma hipótese deve estar justificada pela presença de evidências preliminares e de uma fundamentação teórica.

Ao se realizar o teste de uma hipótese, há consequências analíticas de natureza estatística que permitem estabelecer se há ou não o tipo de relação entre variáveis que foi previsto. Nesse caso, Riffe, Lacy e Fico (2014, p. 148) mencionam dois tipos básicos de resultados que podem aparecer no teste de hipótese:

(a) hipótese nula (estabelece que não há qualquer relação entre as variáveis); e

(b) hipótese alternativa (estabelece que há uma relação entre as variáveis no sentido proposto pela pesquisa).

Para verificar qual tipo de hipótese tem mais força estatística, dois tipos de teste se mostram bastante comuns: associações e correlações. Krippendorff (2004) explica a diferença entre eles da seguinte maneira:

Quando as variáveis são nominais (um conjunto de categorias sem ordem hierárquica), estamos falando de associações [...], mas quando consistem em valores ordenados numericamente, estamos falando de correlações. (KRIPPENDORFF, 2004, p. 196).

Abaixo, iremos apresentar brevemente as principais premissas teóricas e as técnicas envolvidas em cada tipo.

Teste qui-quadrado (χ^2):

Como observado anteriormente, o tipo de teste estatístico que corresponde a um teste de hipótese envolvendo relações entre variáveis nominais se denomina por associações. Tal tipo de teste é bastante comum na análise de conteúdo, pois, como anteriormente abordado (ver Seção 3.1.1), esse tipo de variável estabelece categorias conceituais como base para sua codificação, que é justamente uma das operações mais recorrentes da análise de conteúdo.

Para testar se esse tipo de variável consegue explicar outra variável de interesse (ver tipos de variáveis em Seção 3.1), o teste qui-quadrado χ^2 é o tipo de técnica estatística mais comum que testa a associação entre esses tipos de variáveis. Sua realização é viabilizada por aquilo que se denomina de **tabela de contingência**, que sumariza as observações das variáveis.

Para ilustrar, imagine uma pesquisa que deseja testar a hipótese de que partidos de direita seriam mais favoráveis a políticas de transferência de renda voltada aos mais pobres, enquanto os partidos de esquerda seriam mais favoráveis à renda universal (em que todos os cidadãos têm a prerrogativa de receber certa renda mínima do Estado). Com o objetivo de testar tal hipótese, a pesquisa se propôs a realizar uma análise de conteúdo sobre a posição manifesta em programas partidários de diferentes democracias do mundo em relação a propostas de transferência de renda. Para isso, a pesquisa codificou os programas em relação à transferência de renda em três categorias (propostas sobre o tema) : 1- renda básica universal , 2- renda básica aos mais pobres e 3- sem proposta encontrada. Ao mesmo tempo, os partidos foram categorizados a partir da autodeclaração em relação a serem de “direita” ou de “esquerda”. O resultado da pesquisa hipotética foi organizado na seguinte tabela de contingência:

Tabela 3 – Posicionamento dos partidos em relação a políticas de transferência de renda

	Renda básica universal	Renda para os mais pobres	Sem Proposta	TOTAIS
<i>Direita</i>	5	32	15	52
<i>Esquerda</i>	16	7	25	48
TOTAIS	21	39	40	100

Fonte: elaboração própria.

A Tabela 3 mostra que, dos 100 programas partidários que foram analisados, 52 eram de partidos que se autodeclaravam de direita e 48, de esquerda. Dos 52 programas dos partidos de direita, cinco se manifestaram a favor da renda mínima universal, 32 a favor de renda básica aos mais pobres e 15 não manifestaram posição. Já dos programas dos partidos de esquerda, 16 se mostraram favoráveis à renda básica universal, sete a favor da renda básica apenas aos mais pobres e 25 não manifestaram qualquer posição sobre o tema. Juntando os dois tipos de partidos, 21 se manifestaram a favor da renda mínima universal, 39 da renda mínima aos mais pobres e 40 não manifestaram qualquer posição sobre o tema.

Para se testar a hipótese da pesquisa, o teste qui-quadrado se faz oportuno, pois a hipótese associa duas variáveis nominais, nomeadamente:

(orientação ideológica do partido)

x

(a posição de seu programa partidário em relação a políticas de transferência de renda)

Além da natureza das variáveis, o teste qui-quadrado se faz aqui oportuno, pois a ideia básica do teste é “comparar as frequências efetivamente observadas com as frequências esperadas por aquelas providas por independência estatística” (LUDWIG-MAYERHOFER; LIEBESKIND; GEISLER, 2014, p. 212). Ou seja, trata-se de verificar se os resultados obtidos em uma ou mais células da Tabela 3 se mostram significativamente acima ou abaixo daquilo que se esperaria no caso da hipótese nula, que, como vimos anteriormente, estabelece que não há relação entre as variáveis. Como a hipótese a ser testada no exemplo aqui em questão propõe que haveria uma associação entre as variáveis (a orientação

ideológica do partido influenciando a posição do mesmo em relação a uma política pública), os valores do teste precisam indicar se uma ou mais células apresentam valores significativamente acima ou abaixo dos valores esperados pela independência estatística (hipótese nula).

Para se realizar o teste qui-quadrado de maneira prática e rápida, há diferentes ferramentas disponíveis *on-line* que solicitam o preenchimento da tabela de contingência e, após o preenchimento dos dados (e observados os pressupostos do teste), é possível obter resultados. No caso do nosso exemplo, a tabela de contingência fica do seguinte modo após a realização do teste⁶⁶:

Tabela 4 – Tabela 3 com os valores do teste qui-quadrado

	Universal	Mais pobres	Sem Posição	TOTAIS
<i>Direita</i>	5 (10.92) [3.21]	32 (20.28) [6.77]	15 (20.80) [1.62]	52
<i>Esquerda</i>	16 (10.08) [3.48]	7 (18.72) [7.34]	25 (19.20) [1.75]	48
TOTAIS	21	39	40	100

Fonte: elaboração própria.

Os valores entre parênteses representam os valores esperados por independência estatística e os valores entre colchetes correspondem ao resultado do teste qui-quadrado para cada célula (quanto maior, mais destoante do valor esperado). Ao somar os valores da estatística do teste para todas as células, o valor total para o teste qui-quadrado é de 24.1662 com um valor $p < 0.00001$. Como resultado, o teste qui-quadrado se mostrou significativo a um $p < 0.01$. Para compreender tais noções, ver tabela de escala de significância do p -valor atribuída a Fischer em Bussab e Morettin (2013, p. 356).

Observe ainda que os maiores valores para o teste qui-quadrado foram justamente nas células da coluna correspondente à hipótese da pesquisa de nosso exemplo, ou seja, na coluna que identifica a preferência à proposta de renda mínima voltada aos mais pobres. Em tal coluna, os valores foram 6.77 e 7.34 para os partidos de direita e de esquerda, respectivamente. Tais resultados devem ser lidos em referência ao valor esperado (entre parênteses). O valor esperado para os partidos de esquerda era de 18.27, mas o valor observado foi bem abaixo disso (no caso, 7). Já o número esperado de

⁶⁶ No caso, utilizamos a calculadora da página “Social Science Statistics”, disponível em: <https://www.socscistatistics.com/tests/chisquare2/default2.aspx>.

partidos de direita que seriam a favor da renda mínima aos mais pobres era de 20.28; mas o valor observado foi bem acima disso, no caso, 32. O exato oposto ocorre em relação à categoria “renda universal”, em que os partidos de esquerda apresentam uma frequência superior à esperada, enquanto os partidos de direita uma frequência abaixo da esperada. Como resultado, o teste qui-quadrado foi mais forte para a hipótese da pesquisa (a hipótese alternativa) do que a hipótese nula.

No caso, a hipótese alternativa postulou que os partidos de direita seriam mais favoráveis do que os partidos de esquerda à proposta de renda mínima voltada aos mais pobres, enquanto estes seriam mais favoráveis à renda mínima universal do que aqueles. Em termos estatísticos, isso implica que a hipótese nula foi rejeitada; portanto, a hipótese alternativa do nosso exemplo resistiu à refutação, ficando por hora (até novos estudos mostrarem algo diverso) como válidos.

Contudo, antes de realizar um teste qui-quadrado, deve-se observar se a tabela de contingência satisfaz os pressupostos do teste, os quais foram esquematizados a partir de Cervi (2017, p. 207) e Giolo (2017, p. 43-46) do seguinte modo:

- não pode existir nenhuma casa com ausência total de casos (zero na casa);
- todas as frequências esperadas devem ser ao menos iguais a 5;
- deve haver independência entre as observações das variáveis testadas;
- o χ^2 não pode ser usado para testar dados correlacionados como pesquisas de painel, pois isso quebraria o pressuposto anterior.

Quando os dados não atendem aos pressupostos do teste qui-quadrado listados acima, uma alternativa se encontra no teste exato de Fisher. Para mais informações sobre, ver Giolo (2017, p.47).

O teste qui-quadrado, como já indicado, é oportuno para testar relações entre variáveis nominais. A seguir, veremos os principais testes de diferença entre médias e percentagens, os quais são adequados para verificar se as diferenças das médias ou proporções entre variáveis quantitativas são estaticamente significativas.

ANOVA

Anova (análise de variância) é um dos tipos de teste mais comuns para se verificar se as médias de um grupo de amostras são estatisticamente significativas. Esse tipo de teste é oportuno quando se deseja medir a “independência de variações entre uma variável contínua e uma categórica com mais de duas categorias” (CERVI, 2017, p. 194). Riffe, Lacy e Fico (2014) explicam a natureza inferencial desse tipo de teste:

Um teste de diferença de médias ou de proporções calcula o quão provável é que a diferença amostral entre dois grupos de uma amostra probabilística possa ter ocorrido por acaso. Se a diferença amostral for muito alta a ponto de ser muito improvável que ela decorra desse acaso, então a hipótese nula é rejeitada em favor da hipótese de que os dois grupos de fato pertencem a duas populações diferentes (RIFFE; LACY; FICO, 2014, p. 145).

Para ilustrar o uso desse teste, vamos imaginar uma pesquisa que deseja comparar a grade curricular de escolas estaduais e municipais brasileiras para verificar, através de uma análise de conteúdo das grades curriculares, se elas possuem médias diferentes de componentes curriculares voltados à história da América Latina (AL). A amostra contou com uma análise de conteúdo da grade curricular de 15 escolas municipais e 15 estaduais, os quais foram sumarizados na Tabela 5.

Tabela 5 – Quantidade de componentes curriculares sobre história da AL

Esc. Municipais	Esc. Estaduais
5	6
6	8
4	6
5	6
5	7
4	8
3	6
5	8
4	7
6	8
4	6
2	4
6	7
4	6
2	4

Fonte: elaboração própria

A Tabela 5 mostra a quantidade de componentes curriculares que cada uma das 15 escolas municipais e estaduais pesquisadas dedicou à história da América Latina. A média das escolas municipais foi de 4.3 componentes curriculares sobre o assunto, enquanto a média das escolas estaduais foi de 6.4. Seriam essas médias de 15 escolas selecionadas aleatoriamente fruto do mero acaso ou estariam elas indicando uma diferença real entre os dois tipos de escola existentes na população como um todo? O teste Anova permite responder a essa questão, recorrendo a medidas de variação como aquela existente dentro dos grupos e entre os grupos.

Ao realizar o teste⁶⁷, o resultado foi um valor $F = 20.30595$, com valor $p = 000107$, o que implica um valor inferior a um $p < 0.01$. Assim, a probabilidade de que a diferença encontrada não decorra do mero acaso é de 99%. Riffe, Lacy e Fico (2014) explicam da seguinte maneira os valores resultantes de um teste Anova:

⁶⁷ No caso, utilizamos a calculadora da página “Social Science Statistics”, disponível em <https://www.socscistatistics.com/tests/anova/default2.aspx>.

Quanto maior o valor F , maior é a chance de que a correlação obtida reflita algo real na população no lugar de um evento estatístico gerado pela amostragem aleatória. O processo computacional que produz o valor F é também acompanhado por um valor de probabilidade que informa a probabilidade de que os dados foram produzidos pelo acaso. (RIFFE; LACY; FICO, 2014, p. 156).

Do mesmo modo que o teste qui-quadrado, a execução do teste Anova também só deve ser realizada quando os seguintes pressupostos são satisfeitos:

- as observações da variável resposta devem ser independentes (Pimentel Gomes, 1985, p.12);
- a variável dependente é contínua (intervalar ou de razão);
- a variável dependente apresenta distribuição normal;
- existe homogeneidade na variância (CERVI, 2017, p.195).

Se algum dos pressupostos acima listados não puder ser satisfeito, há testes alternativos como o teste “ t ”. Para mais informações sobre esse tipo de teste, ver Pimentel Gomes (1985, p. 20).

Uma vez compreendidos os principais testes estatísticos de associações entre variáveis, assim como testado se há diferença significativa entre os parâmetros de dois grupos de uma amostra, agora vamos passar a tratar das correlações, as quais envolvem variáveis de natureza quantitativa e permitem testar o pressuposto de que uma variável independente influencia outra dependente.

Correlação de Pearson

As análises de correlação são testes estatísticos que permitem verificar se uma variável quantitativa está correlacionada a outra, seja de maneira positiva (quando uma aumenta, a outra também) ou negativa (quando uma aumenta, a outra diminui). Além disso, esse tipo de teste informa a intensidade dessa correlação através do valor r , assim como a probabilidade de significância estatística (ou seja, a chance do resultado de r ser fruto do mero acaso) por meio do valor p .

O valor r pode variar de -1.0 a 1.0 . Quando a correlação entre duas variáveis é perfeitamente negativa, seu valor é -1.0 ; quando a relação é perfeitamente aleatória, seu valor corresponde a 0 ; e quando a correlação é perfeitamente positiva, o valor de r é igual a 1.0 . Porém, os resultados reais obtidos nas pesquisas sociais raramente apresentam correlações perfeitas e, portanto, é preciso interpretar as faixas de valores que estão entre os extremos. Sobre isso, Riffe, Lacy e Fico (2014) apontam que:

Os livros de estatística geralmente consideram correlações de 0.7 ou acima disso como correlações fortes, já correlações entre 0.4 e 0.7 seriam moderadas e correlações entre 0.2 e 0.4 seriam fracas ou moderadas. (RIFFE; LACY; FICO, 2014, p. 155).

Assim como as análises de associações, as análises de correlações permitem testar a possibilidade de que uma variável influencie o comportamento da outra, desde que tal relação possa ser considerada como epistemologicamente válida e não resultado de uma correlação espúria (ver discussão sobre o conceito mais adiante).

Para ilustrar isso com mais um exemplo fictício e similar ao dado anteriormente⁶⁸, imaginemos que uma pesquisa se propôs a verificar se a quantidade de componentes curriculares sobre Ciências Humanas de escolas influenciaria no desempenho dos alunos no Enem (Exame Nacional do Ensino Médio), especificamente o desempenho nas questões do exame que versam sobre Ciências Humanas e suas tecnologias. Para isso, a pesquisa fez uma análise de conteúdo da grade curricular de 15 escolas selecionadas aleatoriamente para identificar a quantidade desses componentes e, em seguida, calculou a percentagem média de acerto dos alunos de cada escola no Enem relativo às questões sobre Ciências Humanas (CH). Os resultados da

⁶⁸ No exemplo do teste de variância Anova, todos os dados eram resultantes de uma análise de conteúdo. Já nesse exemplo da correlação, os dados referentes ao Enem seriam retirados do Inep (Instituto responsável pela realização do Enem). Ou seja, cada exemplo ilustraria cada um dos dois tipos de associações ou correlações que Krippendorff (2004, p.196) aponta como possível fazer com a análise de conteúdo: 1- com dados provenientes apenas da análise conteúdo; 2- dados obtidos por meio da análise de conteúdo com dados obtidos independentemente.

pesquisa hipotética foram catalogados na Tabela abaixo, em que cada linha representa uma escola:

Tabela 6 – N° de componentes curriculares da escola X Percentagem média de acerto dos alunos da respectiva escola

No de Componentes curriculares em CH	Percentagem média de acerto no ENEM
4	52%
6	77%
8	84%
2	11%
5	65%
6	86%
7	92%
8	94%
2	32%
1	13%
5	66%
3	51%
6	83%
7	87%
8	93%

Fonte: elaboração própria.

Na Tabela 6, a coluna da esquerda representa a variável independente, e a coluna da direita, a variável dependente. A primeira das 15 escolas analisadas teve quatro componentes curriculares identificados na análise de conteúdo e a percentagem média de acertos dos alunos da respectiva escola nas questões do Enem sobre Ciências Humanas foi de 52%. Ao se observar os demais 14 casos de escolas analisadas, percebe-se que, em geral, quando o número de componentes curriculares aumenta, a percentagem média de acertos dos alunos da escola também aumenta. Ao realizar o teste de correlação de Pearson (o mais comum entre os testes desse tipo) para verificar a força dessa correlação linear, obtemos os seguintes resultados⁶⁹:

⁶⁹ No caso, utilizamos a calculadora da página “Social Science Statistics”, disponível em <https://www.socscistatistics.com/tests/pearson/default2.aspx>.

- Valor de $r = 0.9576$
- O valor de $p < .00001$, sendo significativo, portanto, no nível mais exigente $p < 0.01$.

O valor de r apresenta um valor muito próximo ao nível máximo de correlação positiva (1.0), ou seja, trata-se de uma correlação linear muito alta (algo raro nas pesquisas reais). Aqui, assim como nos testes de associações (a exemplo do qui-quadrado visto neste livro), há também um valor p que indica a significância estatística do resultado encontrado, o que, no caso acima, é um valor que indica baixa probabilidade de que o resultado encontrado seja fruto do mero acaso.

A propósito do mero acaso, devemos aqui retomar o conceito de correlação espúria para uma melhor compreensão de que qualquer teste estatístico deve fazer sentido epistemológico e não atestar uma linha de influência pela mera correlação estatística entre duas (ou mais) variáveis. Para ilustrar isso, retomemos nosso exemplo das escolas e imaginemos que, no lugar de número de componentes curriculares, a pesquisa tivesse quantificado o tamanho médio do cabelo dos alunos da escola e tivesse feito o mesmo cruzamento com o desempenho desses alunos no Enem. Caso o resultado da correlação tivesse sido o mesmo do exemplo anterior (ou seja, um valor $r = 0.95$ e um valor $p < 0,01$), o tamanho dos cabelos influenciaria no desempenho dos estudantes? Como tal postulada não faz sentido algum, pois não é concebível tal linha de causalidade, um eventual resultado deveria ser considerado como fruto de uma correlação espúria.

Talvez, o tamanho dos cabelos dos alunos esteja associado ao gênero dos discentes ou aspectos culturais envolvidos nas escolas que, então, talvez expliquem a correlação espúria. De todo modo, o esforço de um desenho de pesquisa é tentar encontrar correlações significativas do ponto de vista conceitual e, portanto, de compreensão do fenômeno social em questão e não confundir isso com meros artefatos estatísticos.

CONCLUSÃO

O objetivo principal deste livro foi oferecer um guia sobre análise de conteúdo em língua portuguesa que refletisse as técnicas e procedimentos mais rigorosos do método, especificamente em sua dimensão quantitativa. Para isso, buscamos oferecer um material alinhado ao estado da arte na literatura internacional, mas adaptado à realidade predominante das pesquisas desenvolvidas no Brasil. Com isso, almejamos contribuir para a conquista, a médio e longo prazos, de um padrão médio mais exigente tanto no uso da AC quanto na recepção e avaliação de trabalhos que dizem se utilizar do método.

Tal elevação se mostra particularmente necessária quando as revisões bibliográficas mais atuais mostram que as pesquisas brasileiras dificilmente reportam como (e se) realizaram os procedimentos que constituem a análise de conteúdo. A maioria das pesquisas ainda se apoia excessiva ou exclusivamente no manual de Bardin (cf. SAMPAIO *et al.*, 2021). Portanto, os déficits metodológicos são graves e preocupantes, uma vez que nem mesmo o relativamente pouco exigente manual de Bardin é efetivamente aplicado (GOMES *et al.*, 2020; MARTINEZ; PESSONI, 2015; QUADROS *et al.*, 2015; SAMPAIO; LYCARIÃO, 2018; SILVA *et al.*, 2017).

Certamente, isso não é exclusividade da pesquisa brasileira, pois até mesmo periódicos internacionais de alto impacto têm apenas recentemente – e de maneira ainda limitada a um subconjunto restrito de estudos – demonstrado elementos básicos de replicabilidade e confiabilidade (ver LOVEJOY *et al.*, 2016). Ademais, conforme o levantamento de Lovejoy e colegas (2014) em torno de periódicos de alto impacto, dos 581 artigos analisados (publicados entre 1985 e 2010), “a maioria dos artigos não usou censo nem amostra probabilística e nem foram transparentes sobre como o processo de construção de amostra foi realizado” (LOVEJOY *et al.*, 2014, p. 220).

De todo modo, por mais que até mesmo os periódicos de maior impacto ainda demonstrem uma quantidade relevante de trabalhos com déficits metodológicos preocupantes na operacionalização da análise de conteúdo, é seguro afirmar que tais déficits se mostram ainda mais graves na esfera dos

periódicos nacionais⁷⁰. Isso porque a própria literatura predominantemente utilizada em tal esfera recorre a manuais pouco exigentes e detalhados em relação às diversas fases da análise de conteúdo. Diante disso, esta obra se propõe a oferecer uma fonte mais acessível em torno desse detalhamento, mas também adaptado à realidade de nossa estrutura de pesquisa.

No caso, devemos ser enfáticos no reconhecimento de que tal estrutura torna a utilização das práticas mais exigentes da análise de conteúdo absolutamente inviável. Isso porque as melhores práticas da análise de conteúdo requerem prazos razoáveis de execução, recursos humanos qualificados (e, portanto, de elevado custo financeiro) e equipes que contem com diversos pesquisadores envolvidos. Contudo, tais condições se mostram impraticáveis quando o financiamento devotado à pesquisa acadêmica se mostra insuficiente, instável e sujeito a bruscas alterações de acordo com a conjuntura política e econômica do país.

Como resultado, as condições objetivas em que normalmente as análises de conteúdo são realizadas no país são muito distantes das necessárias para que as práticas mais rigorosas sejam implementadas. Cientes disso, confeccionamos este livro de modo a permitir uma análise com um nível razoável de rigor metodológico, mas sem tornar inviável a execução da pesquisa. Daí que oferecemos, pela primeira vez na literatura de manuais de que temos notícias, um conjunto de procedimentos voltados a pesquisas individuais de graduandos, pós-graduandos e até de docentes que não estão em condições de contar com a participação de outros colegas em sua pesquisa.

Outro aspecto deste manual voltado a elevar o rigor metodológico, mas sem elevar o custo financeiro da pesquisa, foi a prática de utilizar exemplos e aplicações provenientes de calculadoras, aplicativos e programas disponíveis gratuitamente na internet, que prescindem o uso de *software* proprietários.

⁷⁰ Apesar dos esforços iniciais já supracitados, um caminho importante a ser trilhado para a pesquisa brasileira é a metapesquisa ou meta-análise das pesquisas que utilizam (ou dizem utilizar) a análise de conteúdo. A verificação do preenchimento de certas fases vitais, como amostragem, teste de confiabilidade entre codificadores e disponibilização do livro de códigos seria pertinente para avaliarmos o atual estado da pesquisa brasileira e pontos a serem avançados.

Ademais, tendem a ser ferramentas com baixas curvas de aprendizado, facilitando seu uso sem consideráveis conhecimentos prévios. Por mais que algumas dessas ferramentas possam ser descontinuadas nos próximos anos, é muito provável que outras surjam ou que alguma(s) das suas alternativas atuais permaneçam *on-line*. Daí, recomendamos a busca ativa por parte dos leitores desta obra por esses tipos de ferramenta.

De todo modo, indicamos, para fins de qualificação continuada, o aprendizado de linguagens de programação (a exemplo dos requeridos para o uso do projeto *R* ou *Python*), pois tal conhecimento confere autonomia e ampliação da capacidade tanto de coleta (especialmente se forem de dados disponíveis na internet) quanto de análise estatística dos resultados. Sobre isso, também reforçamos que, mais importante do que saber escrever uma linha de código para se rodar um teste estatístico, é reconhecer a adequabilidade de tal teste ao contexto da pesquisa e ter a capacidade de interpretar tais resultados de modo consequente e significativo. Para isso, a literatura presente em nossas referências bibliográficas busca abrir caminho para essa formação continuada e consciente dos limites e dilemas que envolvem cada decisão metodológica de uma pesquisa.⁷¹

⁷¹ Agradecemos aos colegas Rousiley Maia, Pedro Mundim, Patrícia Rossini, Emerson Cervi, Antal Wozniak, Hartmut Wessler pelas indicações e ensinamentos na técnica. Agradecemos ainda à equipe da ENAP que se esforçou efusivamente pela publicação desta obra em um cenário complexo, aqui representadas nas figuras de Fernando Filgueiras, Regina Luna Santos de Souza, José Guilherme de Lara Resende, Kelly Sartório e Amanda Soares.

Referências bibliográficas

- ABREU, M.; GOMES, R. Medindo o nível da influência colaborativa dos stakeholders e das ações de orçamento aberto. *Revista de Administração Pública*, v. 52, n. 4, p.593-609, 2018.
- AGGIO, C.; REIS, L. Campanha eleitoral no Facebook: usos, configurações e o papel atribuído a esse site por três candidatos eleitos nas eleições municipais de 2012. *Revista Compolitica*, v. 3, p. 155-188, 2013.
- ALONSO, S.; VOLKENS, A.; GÓMES, B. *Análisis de contenido de textos políticos: un enfoque cuantitativo*. Centro de Investigaciones Sociológicas: Madrid, 2012.
- ALVES, D.; FIGUEIREDO, D.; HENRIQUE, A. O poderoso NVivo: uma introdução a partir da análise de conteúdo. *Revista Política Hoje*, v. 24, n. 2, p. 119-134, 2015.
- ALVES, M. Análise de conteúdo: sua aplicação nas publicações de contabilidade. *Revista Universo Contábil*, v. 7, n. 3, p. 146-166, 2011.
- ALVES, M. *VaipraCuba? A gênese das redes de direita no Facebook*. Curitiba: Appris, 2019.
- AMORIM, P.; ALMADA, M. E-transparência: proposta de modelo metodológico para avaliação de portais de executivos nacionais. *Logos*, v. 23, n. 2, 2016.
- AQUINO, J. *R para cientistas sociais*. Ilhéus, BA: Editus, 2014.
- ARAYA-LEANDRO, C.; CABA-PÉREZ, M.; LÓPEZ-HERNANDEZ, A. A convergência dos países da América Central para as Normas Internacionais de Contabilidade. *Revista de Administração Pública*, v. 50, n. 2, p. 265-283, 2016.
- BARDIN, L. *Análise de conteúdo*. Edição revista e ampliada. São Paulo: Edições 70 Brasil, [1977] 2016.
- BARROS, S.; CARREIRO, R. A discussão pública e as redes sociais online: o comentário de notícias no Facebook. *Revista Fronteiras*, v. 17, p. 174-185, 2015.
- BASTOS, M.; OLIVEIRA, U. Análise de discurso e análise de conteúdo: um breve levantamento bibliométrico de suas aplicações nas ciências sociais aplicadas da Administração. *XII SEGeT-Simpósio de Excelência em Gestão e Tecnologia*, 2015.
- BAUER, M. Análise de conteúdo clássica: uma revisão. In: BAUER, M.; GASKELL, G. (Orgs.). *Pesquisa qualitativa com texto, imagem e som: um manual prático*. Petrópolis, Editora Vozes, p.189-217, 2007.
- BELLUCCI JÚNIOR, J.; MATSUDA, L. Construção e validação de instrumento para avaliação do Acolhimento com Classificação de Risco. *Revista brasileira de enfermagem*, v.65, n. 5, p.751-757, 2012.
- BERELSON, B. *Content analysis in communications research*. New York: Free Press, 1952.
- BIANCHI, C.; FIGUEIREDO, J. Construção de políticas para melhoria do processo de internacionalização da arquitetura no Brasil: caso Apex-Brasil. *Revista de Políticas Públicas*, v. 21, n.1, p.177-198, 2017.
- BISPO, F.; GOMES, R. Os papéis dos stakeholders na formulação do Pronatec. *Revista de Administração Pública*, v. 52, n. 6, p.1258-1269, 2018.
- BLÜMEL, C. Translational research in the science policy debate: a comparative analysis of documents. *Science and Public Policy*, v. 45, n.1, p. 24-35, 2017.
- BORBA, F. Medindo a propaganda negativa na TV, rádio, debates, imprensa e Facebook: o caso

- das eleições presidenciais de 2014. *Intercom*, v. 42, p. 37-56, 2019.
- BOZEMAN, B.; GAUGHAN, M.; YOUTIE, J.; SLADE, C.; RIMES, H. Research collaboration experiences, good and bad: Dispatches from the front lines. *Science and Public Policy*, v. 43, n. 2, p. 226-244, 2016.
- BRAGA, S.; ROCHA, L.; CARLOMAGNO, M. A internet e os partidos brasileiros. *Cadernos Adenauer*, v. 16, p. 47-74, 2015.
- BUENO, T. Como pesquisadores e veículos têm tratado o tema comentário de leitores na internet. *Liinc em Revista*, v. 12, p. 180-194, 2016.
- BURSCHE, B.; Vliegenthart, R.; DE VREESE, C. Using supervised machine learning to code policy issues. *The Annals of the American Academy of Political and Social Science*, v. 659, n. 1, p.122–131, 2015.
- BUSSAB, W. O.; MORETTIN, P. A. Estatística Básica. São Paulo: Editora Saraiva, 2013.
- BUTTON, K.; IOANNIDIS, J.; MOKRYSZ, C.; NOSEK, B.; FLINT, J.; ROBINSON, E.; MUNAFÒ, M. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, v.14, n. 5, p.365–376, p. 2013.
- CÂMARA, R. Análise de conteúdo: da teoria à prática em pesquisas sociais aplicadas às organizações. *Gerais: Revista Interinstitucional de Psicologia*, v. 6, n. 2, p. 179-191, 2013.
- CAMARGO, B.; JUSTO, A. IRAMUTEQ: um software gratuito para análise de dados textuais. *Temas em psicologia*, v. 21, n. 2, p. 513-518, 2013.
- CAMPOS, C. Método de análise de conteúdo: ferramenta para a análise de dados qualitativos no campo da saúde. *Revista Brasileira de Enfermagem*, v. 57, n. 5, p. 611-614, 2004.
- CANO, I. Nas trincheiras do método: o ensino da metodologia das ciências sociais no Brasil. *Sociologias*, v. 14, n. 31, p. 94-119, 2012.
- CAPPELE, M.; MELO, M.; GONÇALVES, C. Análise de conteúdo e análise de discurso nas ciências sociais. *Organizações rurais & agroindustriais*, v. 5, n. 1, 2003.
- CAREGNATO, R.; MUTTI, R. Pesquisa qualitativa: análise de discurso versus análise de conteúdo. *Texto & Contexto-Enfermagem*, v. 15, n. 4, p. 679-684, 2006.
- CARLOMAGNO, M.; ROCHA, L. Como criar e classificar categorias para fazer análise de conteúdo: uma questão metodológica. *Revista Eletrônica de Ciência Política*, v.7, n.1, 2016.
- CARLSON, D.; MONTGOMERY, J. A pairwise comparison framework for fast, flexible, and reliable human coding of political texts. *American Political Science Review*, v.111, p. 835 – 843, 2017.
- CARREIRO, R. *A discussão política em rede: um estudo sobre a divergência política no Facebook*. Tese (Doutorado em Comunicação e Cultura Contemporâneas), 244p. Universidade Federal da Bahia, Salvador, 2017.
- CARVALHO, F.; MASSUCHIN, M.; MITOZO, I. Radicalização nas redes sociais: comentários no Facebook durante a disputa presidencial em 2014 no Brasil. *Análise Social (Lisboa)*, v. 13, p. 898-926, 2018.
- CARVALHO, M.; CORREIA, H. Autoavaliação das escolas: o caso do grupo disciplinar de Biologia/ Geologia. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 26, n. 101, p. 1262-1282, 2018.
- CASTANHO, M.; SILVA, M.; CASARIN, S.; DORIGHELLO, L. Concepções e ações de professoras sobre propostas de um programa de educação integral. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 26, n. 100, p. 780-800, 2018.

- CASTRO, T.; ABS, D.; SARRIERA, J. Análise de conteúdo em pesquisas de Psicologia. *Psicologia: ciência e profissão*, v. 31, n. 4, p. 814-825, 2011.
- CAVALCANTE, R., CALIXTO, P., PINHEIRO, M. Análise de Conteúdo: considerações gerais, relações com a pergunta de pesquisa, possibilidades e limitações do método. *Informação & Sociedade: Estudos*, v. 24, n. 1, p. 13-18, 2014.
- CAVANAGH, S. Content analysis: concepts, methods and applications. *Nurse researcher*, v. 4, n. 3, p. 5-16, 1997.
- CERVI, E. *Análise de Dados Categóricos em Ciência Política*. Curitiba: PPGCP, 2014.
- CERVI, E. Campanhas eleitorais em redes sociais: transparência ou pornografia. In: CERVI, E.; MASSUCHIN, M.; CARVALHO, F. (Org.) *Internet e eleições no Brasil*. Curitiba: CPOP, 2016, p. 17-35.
- CERVI, E. O uso do HGPE como recurso partidário em eleições proporcionais no Brasil: um instrumento de análise de conteúdo. *Opinião Pública*, v. 17, n.1, 106-136, 2011.
- CERVI, E. *Manual de métodos quantitativos para iniciantes em Ciência Política*. Curitiba: CPOP, 2017
- CERVI, E.; MASSUCHIN, M. Metodologia quantitativa em pesquisas sobre cobertura jornalística. *Revista FAMECOS*, v. 20, p. 840-865, 2013.
- CERVI, E.; MASSUCHIN, M.; CARVALHO, F. (Org.) *Internet e eleições no Brasil*. Curitiba: CPOP, 2016.
- CEZAR, L. Reflexões sobre a comunicação em políticas públicas: proposta de um modelo de avaliação da comunicação governamental. *Revista de Administração Pública*, v. 52, n.1, 52-70, 2018.
- CHAGAS, V.; FREIRE, F.; RIOS, D.; MAGALHÃES, D. A política dos memes e os memes da política: proposta metodológica de análise de conteúdo de memes dos debates eleitorais de 2014. *Intexto*, n. 38, p. 173-196, 2017.
- CONCEIÇÃO, L.; BROTTTO, M. Residência Multiprofissional em saúde: o olhar dos preceptores sobre o processo de ensino-aprendizagem. *Revista de Políticas Públicas*, v. 20, n. 2, p. 885-901, 2016.
- CONSTANTINO, N. Pesquisa histórica e análise de conteúdo: pertinência e possibilidades. *Estudos Ibero-Americanos*, v. 28, n. 1, p. 183-194, 2002.
- COUTO, H.; RIBEIRO, F. Objetivos e desafios da política de compras públicas sustentáveis no Brasil: a opinião dos especialistas. *Revista de Administração Pública*, v. 50, n. 2, p. 331-343, 2016.
- CRUZ, J.; SZYMANSKI, M. Materialização das políticas de educação do campo: a relação entre professores e Diretrizes da Educação do Campo em Cascavel. *Revista de Política Públicas da UFMA*, v. 19, p. 1-20, 2015.
- DAHLBERG, L. Net-public sphere research: beyond the "first phase". *Javnost-The Public*, v. 11, n. 1, p. 27-43, 2004.
- DESLANDES, S.; IRIART, J. Usos teórico-metodológicos das pesquisas na área de Ciências Sociais e Humanas em Saúde. *Cadernos de Saúde Pública*, v. 28, n. 12, p. 2380-2386, 2012.
- DINIZ, G.; MACHADO, D; MATOS, F. Transparência digital na gestão pública: uma análise de conglomerados nos municípios cearenses. *Revista Controle: Doutrinas e artigos*, v. 14, n. 2, p. 197-217, 2016.
- DOS SANTOS, M. A rede de oposição radical no facebook: cartografia e apontamentos sobre os haters políticos. *Revista de Estudos da Comunicação*, v. 15, p. 309-324, 2014

- DOWNE-WAMBOLDT, B. Content analysis: method, applications, and issues. *Health care for women international*, v. 13, n. 3, p. 313-321, 1992.
- DRISKO, J. W.; MASCHI, T. *Content Analysis*. Pocket Guides to Social Work Research Methods. Oxford: Oxford University Press, 2016.
- ERLER, A. A new face of poverty? Economic crises and poverty discourses. *Poverty & Public Policy*, v. 4, n. 4, p. 183-204, 2012.
- ERLINGSSON, C.; BRYSEWICZ, P. A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, v. 7, n. 3, p. 93-99, 2017.
- ESPÍRITO SANTO, P.; SOARES, I. Unindo as pontas soltas com a análise de conteúdo etnográfica. *Revista Debates*, v. 9, n. 1, p. 25-48, 2015.
- FENG, G. Intercoder reliability indices: disuse, misuse, and abuse. *Quality & Quantity*, v. 48, n. 3, p. 1803-1815, 2014.
- FERES JÚNIOR, J. Análise de valências, debate acadêmico e contenda política. *Revista Brasileira de Ciência Política*, v. 20, p. 313-322, 2016.
- FIGUEIREDO FILHO, D.; PARANHOS, R.; SILVA JÚNIOR, J.; ROCHA, E.; ALVES, D. O que é, para que serve e como se faz uma meta-análise?. *Teoria & Pesquisa: Revista de Ciência Política*, v. 23, n. 2, 2014.
- FIGUEIREDO, M.; ALDÉ, A.; DIAS, H., JORGE, V. Estratégias de persuasão eleitoral: uma proposta metodológica para o estudo da propaganda eleitoral. *Opinião Pública*, v. 4, n. 3, p. 109-120, 1997.
- FIGUEIREDO, N.; SALLES, D. Educação Profissional e evasão escolar em contexto: motivos e reflexões. *Revista Ensaio: Avaliação e Políticas Públicas em Educação*, v. 25, n. 9, p. 356-392, 2017.
- FLICK, U. *Introdução à pesquisa qualitativa*. São Paulo: Artmed editora, 2009.
- FONSECA JÚNIOR, W. Análise de conteúdo. In: DUARTE, J. (Org.). *Métodos e técnicas de pesquisa em comunicação*. São Paulo: Atlas, 2005.
- FOWLER JR, F. *Survey research methods*. Sage publications, 2013.
- FRANCO, M. *Análise de conteúdo*. Brasília: Liber Livro Editora, 2005.
- FRELON, D. ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, v. 5, n. 1, p. 20-33, 2010.
- FRELON, D. ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science*, v. 8, n.1, p. 10-16, 2013.
- FREIRE, L.; DUARTE, A. Concepções de estudantes universitários brasileiros sobre os fatores e as funções da aprendizagem. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 24, n. 91, p. 380-394, 2016.
- FREITAS, H. Réplica 1—Análise de conteúdo: faça perguntas às respostas obtidas com sua ‘pergunta’!. *RAC-Revista de Administração Contemporânea*, v. 15, n. 4, p. 748-760, 2011.
- FREITAS, H.; JANISSEK, R. *Análise léxica e análise de conteúdo: técnicas complementares, seqüenciais e recorrentes para exploração de dados qualitativos*. Porto Alegre: Sagra Luzzatto, 2000.
- GANDIN, L. *Eu, nós e eles: a luta política no discurso presidencial de Dilma Rousseff*. 2018. 238f. Tese (Doutorado em Ciência Política) – Universidade Federal do Paraná, Curitiba, 2018.
- GIL, A. *Métodos e técnicas de pesquisa social*. São Paulo: Editora Atlas, 1989.

- GIOLO, S. Introdução à análise de dados categóricos com aplicações. São Paulo: Blucher, 2017.
- GOMES, A.; BOCKORNI, B.; SANTOS, A.; NASCIMENTO, K. As contribuições da Análise de Conteúdo e do Discurso para os estudos em Administração. *Revista Foco*, v. 13, n. 1, p. 146-170, 2020.
- GOMES R. Análise e interpretação de dados de pesquisa qualitativa. In: MINAYO, M. (org.) *Pesquisa social: teoria, método e criatividade*. 28a edição. Petrópolis: Vozes, 2009, p. 79-108
- GOMES, W. *Audiosfera política e visibilidade pública: os atores políticos no Jornal Nacional. Televisão e realidade*. Salvador: EDUFBA, p. 175-222, 2009.
- GONDIM, S.; BENDASSOLLI, P. Uma crítica da utilização da análise de conteúdo qualitativa em psicologia. *Psicologia em Estudo*, v. 19, n. 2, p. 191-199, 2014.
- GUERRA, G.; MACHADO, L. Representações sociais de avaliação processual construídas por professoras. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 19, n. 71, 2011.
- GUO, L. *et al.* Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, v. 93, n. 2, p. 332-359, 2016.
- GURGEL, C.; LEITE, R. Avaliar aprendizagem: uma questão de formação docente. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 15, n. 54, p. 145-168, 2007.
- HAK, T.; BERNTS, T. Coder training: Theoretical training or practical socialization? *Qualitative Sociology*, v. 19, n. 2, p. 235-257, 1996.
- HARWOOD, T.; GARRY, T. An overview of content analysis. *The marketing review*, v. 3, n. 4, p. 479-498, 2003.
- HAYES, A. F.; KRIPPENDORFF, K. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, v. 1, n. 1, p. 77-89, 2007.
- HENKEL, K. A categorização e a validação das respostas abertas em surveys políticos. *Opinião Pública*, v. 23, n. 3, p. 786-808, 2017.
- HERSCOVITZ, H. Análise de conteúdo em jornalismo. In: LAGO, C.; BENETTI, M. (Org.). *Metodologia de pesquisa em jornalismo*. Petrópolis: Vozes, 2007, p. 123-142.
- HOLSTI, O. *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley, 1969.
- IKEDA, A.; CHANG, S. Análise de Conteúdo - uma experiência de aplicação na pesquisa em comunicação social. *Comunicação & Inovação*, v. 6, n. 11, 2005.
- HOPKINS, D.; KING, G. A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, v. 54, n. 1, p. 229-247, 2010.
- JACOMETTI, M., CASTRO, M.; GONÇALVES, S.; COSTA, M. Análise de efetividade das políticas públicas de Arranjo Produtivo Local para o desenvolvimento local a partir da teoria institucional. *Revista de Administração Pública*, v. 50, n. 3, p. 425-454, 2016.
- JENSEN, J. L. Public Spheres on the Internet: anarchic or government-sponsored – a comparison. *Scandinavian Political Studies*, v. 26, n. 4, 2003, p. 349-374.
- JORGE, T. (org.). *Notícia em fragmentos: Análise de conteúdo no jornalismo*. Florianópolis: Editora Insular, 2015.
- JUNGBLUT, J.; JUNGBLUT, M. All different? All equal? differentiation of universities' mission statements and excellence initiatives in Germany. *Science and Public Policy*, v. 44, n. 4, p. 535-545, 2017.

- JUSTO, A.; CAMARGO, B. Estudos qualitativos e o uso de softwares para análises lexicais. In: NOVIKOFF, C.; SANTOS, S.; MITHIDIERI, O. (Org.). *Caderno de artigos: X SIAT & II Serpro*, Duque de Caxias, RJ, p. 37-54, 2014. Disponível em: <<https://lageres.wordpress.com/>>
- KATZ, E.; LAZARSFELD, P. *Personal influence: The part played by people in the flow of communications*. Nova York: New York Free Press, 1955.
- KIENTZ, A. *Comunicação de massa: análise de conteúdo*. Rio de Janeiro: Eldorado, 1973.
- KRIEGER, M.; ANDION, C. Legitimidade das organizações da sociedade civil: análise de conteúdo à luz da teoria da capacidade crítica. *Revista de Administração Pública*, v. 48, n. 1, p. 83-110, 2014.
- KRIPPENDORFF, K. *Content analysis: an introduction to its methodology*. Londres: Sage, [1980] 2004.
- LACY, S.; WATSON, B.; RIFFE, D.; LOVEJOY, J. Issues and best practices in content analysis. *Journalism & Mass Communication Quarterly*, v. 92, n. 4, p. 791-811, 2015.
- LARA, R. *A Produção de Conhecimento no Serviço Social: o mundo do trabalho em debate*. São Paulo: Unesp, 2011.
- LASSWELL, H. (org.). *A linguagem da política*. Brasília: Universidade de Brasília, 1982.
- LASSWELL, H. The theory of political propaganda. *The American Political Science Review*, v. 21, n. 3, p. 627-631, 1927.
- LIMA, D.; FRAGA, V.; OLIVEIRA, F. O paradoxo da reforma do Judiciário: embates entre a nova gestão pública e a cultura organizacional do jeitinho. *Revista de Administração Pública*, v. 50, n. 6, p. 893-912, 2016.
- LIMA, J. Por uma análise de conteúdo mais fiável. *Revista portuguesa de pedagogia*, p. 7-29, 2013.
- LIMA, J.; MANINI, M. Metodologia para análise de conteúdo qualitativa integrada à técnica de mapas mentais com o uso dos softwares Nvivo e Freemind. *Informação & Informação*, v. 21, n.3, p. 63-100, 2017.
- LIMA, L.; MORAES, J. A legitimação dos elementos teórico-metodológicos da análise do discurso na ciência da informação brasileira: um aporte da análise de conteúdo. *Brazilian Journal of Information Science: Research Trends*, v. 11, n. 2, 2017.
- LIMA, M. Análise do discurso e/ou análise de conteúdo. *Psicologia em revista*, v. 9, n. 13, p. 76-88, 2003.
- LOI, M.; DI GUARDO, M. The third mission of universities: an investigation of the espoused values. *Science and Public Policy*, v. 42 n. 6, p. 855-870, 2015.
- LOMBARD, M.; SNYDER-DUCH, J.; BRACKEN, C. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research*, v. 28, n. 4, p. 587-604, 2002.
- LOURENÇO, M.; ZBOROWSKI, I.; CURY, A. Análise de conteúdo dos registros da equipe de enfermagem. In: *Proceedings of the 8. Brazilian Nursing Communication Symposium*, 2002.
- LOVEJOY, J.; WATSON, B.; LACY, S.; RIFFE, D. Assessing the reporting of reliability in published content analyses: 1985–2010. *Communication Methods and Measures*, v. 8, n. 3, p. 207-221, 2014.

LOVEJOY, J.; WATSON, B.; LACY, S.; RIFFE, D. Three Decades of Reliability in Communication Content Analyses: Reporting of Reliability Statistics and Coefficient Levels in Three Top Journals. *Journalism & Mass Communication Quarterly*, v. 93, n. 4, p. 1135–1159, 2016.

LUDWIG-MAYERHOFER, W.; LIEBESKIND, U.; GEISSLER, F. *Statistik Eine Einführung für Sozialwissenschaftler*. Beltz Juventa, 2014.

LYCARIÃO, D.; DOS SANTOS, M. Bridging semantic and social network analyses: the case of the hashtag# precisamosfalarsobreaborto (we need to talk about abortion) on Twitter. *Information, Communication & Society*, v. 20, n. 3, p. 368-385, 2017.

LYCARIÃO, D. *Mudanças climáticas sob o prisma da esfera pública: a mediação jornalística como fator de legitimação democrática no caso da COP-15*. Tese de Doutorado. Universidade Federal de Minas Gerais, 333p, 2014.

MACHADO, D.; FIGUEIREDO, N. Registros de enfermagem: a mensagem sobre o cuidado contida na linguagem escrita. *Revista de Pesquisa Cuidado é Fundamental Online*, v. 2, p. 293-297, 2010.

MACNAMARA, J. Media content analysis: Its uses, benefits and best practice methodology. *Asia-Pacific Public Relations Journal*, v. 6, n. 1, 2005.

MACQUEEN, K. M.; MCLELLAN, E.; KAY, K.; MILSTEIN, B. Codebook development for team-based qualitative analysis. *Cam Journal*, v. 10, n. 2, p. 31-36, 1998.

MAIA, R.; CAL, D.; HAUBER, G.; OLIVEIRA, V.; ROSSINI, P.; SAMPAIO, R.; GARCÊS, R. Conversação e deliberação sobre questões sensíveis: um estudo sobre o uso das razões que circulam nos media. *Galáxia*, n. 34, p. 55-72, 2017.

MARTINEZ, M.; PESSONI, A. Intercom: pesquisas feitas com o método (1996 a 2012). In: JORGE, T. (Org.). *Notícia em fragmentos: análise de conteúdo no jornalismo*. Florianópolis: Editora Insular, 2015, p. 299-319.

MASSUCHIN, M.; TAVARES, C.; BELEM, V. Entretenimento no conteúdo jornalístico: um estudo das notícias nos portais informativos brasileiros. *RAZÓN Y PALABRA*, v. 21, p. 273-298, 2017.

MATTHES, J.; KOHRING, M. The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, v. 58, n. 2, 258-279, 2008.

MAYRING, P. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt, 2014. Disponível em: <<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173>>.

MENDONÇA, R.; SAMPAIO, R.; BARROS, S. (Org.). *Deliberação online no Brasil: entre iniciativas de democracia digital e redes sociais de conversação*. Salvador: Edufba, 2016. Disponível em: <<https://repositorio.ufba.br/ri/handle/ri/19267>>.

MIGUEL, L.; BIROLI, F. *Caleidoscópio convexo: mulheres, política e mídia*. Editora Unesp, 2011.

MILNE, M.; ADLER, R. Exploring the reliability of social and environmental disclosures content analysis. *Accounting, Auditing & Accountability Journal*, 1999.

MINAYO, M. (org.) *Pesquisa social: teoria, método e criatividade*. 28ª edição. Petrópolis: Vozes, 2009.

MINAYO, M. *O desafio do conhecimento: pesquisa qualitativa em saúde*. 14ª edição. São Paulo: Hucitec editora, 2014.

MONT'ALVERNE, C.; MARQUES, F. Conflitos, barganhas e desprezo pelo interesse público: enquadramentos do Congresso Nacional em editoriais da Folha de São Paulo. *Revista Famecos*, v. 26, p. 30158, 2019.

- MORAES, M., GALASSO, B.; JANOARIO, R.; ESDRAS, D. Graduates: an overview of the trajectory of graduates in education from the National Institute for the Deaf. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 26, n.100, p. 1084-1107, 2018.
- MORAES, R. Análise de conteúdo. *Revista Educação*, Porto Alegre, v. 22, n. 37, p. 7-32, 1999.
- MORAIS, H.; Paiva, J.; Sousa, W. Avaliação do programa um milhão de cisternas rurais (P1MC): eficácia, eficiência e efetividade nos territórios do Rio Grande do Norte (2003/2015). *Revista de Políticas Públicas*, v. 21, n. 1, p. 133-158, 2017.
- MOREIRA, L.; IULIANELLI, J. Formação docente e ensino de política educacional em instituições de educação superior do Rio de Janeiro. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 25, n. 97, p. 854-873, 2017.
- MORO, M. O problema da validade na pesquisa sobre a alfabetização. *Educar em Revista*, v. 8, p. 157-181, 1989.
- MOSER, C.; KALTON, G. *Survey methods in social investigation*. Londres: Routledge, 2017.
- MOZZATO, A.; GRZYBOVSKI, D. Tréplica-Análise de Conteúdo: ampliando e aprofundando a reflexão sobre a técnica de análise de dados qualitativos no campo da administração. *Revista de Administração Contemporânea*, v. 15, n. 4, 766-775, 2011.
- MUDDIMAN, A.; MCGREGOR, S.; STROUD, N. (Re) Claiming Our Expertise: parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, v. 36, n. 2, p. 214-226, 2019.
- NEUENDORF, K. *The content analysis guidebook*. Londres: Sage, 2002.
- OLIVEIRA, A.; TINÔCO, D.; ALLOUFA, J.; ARAÚJO, R. Uma análise do processo gestor do Programa de Aquisição de Alimentos (PAA): o caso Rio Grande do Norte. *Revista de Políticas Públicas*, v. 14, n. 1, 2010.
- OLIVEIRA, D. Análise de conteúdo temático-categorial: uma proposta de sistematização. *Revista enfermagem UERJ*, v. 16, n. 4, p. 569-576, 2008.
- OLIVEIRA, E.; ENS, R.; ANDRADE, D.; MUSSIS, C. Análise de conteúdo e pesquisa na área da Educação. *Revista Diálogo Educacional*, v. 4, n. 9, p. 1-17, 2003.
- O'NEIL, C. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Portland: Broadway Books, 2016.
- PÁDUA, E. Análise de conteúdo, análise de discurso: questões teórico-metodológicas. *Revista de Educação PUC-Campinas*, n. 13, p. 21-30, 2002.
- PAGNEZ, K. Projeto Eureka: a implantação da informática educativa na rede municipal de campinas no período de 1989-1997. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 14, n. 51, p. 249-260, 2006.
- PANKE, L.; CERVI, E. Análise da comunicação eleitoral: uma proposta metodológica para os estudos do HGPE. *Contemporanea-Revista de Comunicação e Cultura*, v. 9, n. 3, p. 390-404, 2012.
- PAPACHARISSI, Z. Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New media & society*, v. 6, n. 2, p. 259-283, 2004.
- PAULA, C. Análise dialógica de conteúdo e diálogos de saberes. *Boletim Gaúcho de Geografia*, v. 42, n. 1, 2015.
- PEREIRA, M.; GALVÃO, T. Etapas de busca e seleção de artigos em revisões sistemáticas da literatura. *Epidemiologia e Serviços de Saúde*, v. 23, p. 369-371, 2014.

- PHILIPPS, A. Mission statements and self-descriptions of German extra-university research institutes: a qualitative content analysis. *Science and Public Policy*, v. 40, n. 5, p. 686-697, 2013.
- PIMENTEL, F. Curso de Estatística Experimental. São Paulo, Nobel, 1985.
- POHLMANN, M.; BÄR, S.; VALARINI, E. The analysis of collective mindsets: introducing a new method of institutional analysis in comparative research. *Revista de sociologia e política*, v. 22, n. 52, p. 7-25, 2014.
- POOL, I. (Org.). *Trends in content analysis*. Urbana: University of Illinois Press, 1959.
- POPPER, K. *Lógica das ciências sociais*. Rio de Janeiro: Tempo Brasileiro, 1978.
- PRADO, O.; RIBEIRO, M.; DINIZ, E. Governo eletrônico e transparência: olhar crítico sobre os portais do governo federal brasileiro. In: PINHO, A. (Org.). Estado, sociedade e interações digitais: expectativas democráticas. Salvador: EDUFBA, 2012, v. 1, p. 13-40.
- PRASAD, B. Content analysis. In: LAL DAS, D.; BHASKARAN, V. (org.). *Research methods for Social Work*. New Delhi: Rawat, 2008, p.173-193.
- PRUDENCIO, K.; RIZZOTTO, C.; SAMPAIO, R. A normalização do golpe: o esvaziamento da política na cobertura jornalística do impeachment de Dilma Rousseff. *Revista Contracampo*, v. 37, p. 1-15, 2018.
- QUADROS, M.; ASSMANN, G.; LOPEZ, D. A análise de conteúdo nas pesquisas brasileiras em comunicação: aplicações e derivações do método. In: BARICHELLO, E.; RUBLESCKI, A. (org.). *Pesquisa em comunicação: olhares e abordagens*. Santa Maria: Facos – UFSM, 2014, p. 89-108.
- RAUPP, F.; PERFOLL, A. Construindo a transparência em assembleias legislativas dos estados brasileiros: uma investigação nos portais eletrônicos. *Revista de Administração e Contabilidade da FAT*, v. 5, n. 2, p. 04-20, 2017.
- RECUERO, R. Discutindo análise de conteúdo como método: o# DiadaConsciênciaNegra no Twitter. *Cadernos de Estudos Linguísticos*, v. 56, n. 2, p. 289-309, 2014.
- REIS, P.; GONÇALVES, T.; MESQUITA, L. A avaliação do período probatório de professores em Portugal: perspectivas de professores, mentores e gestores. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 20, n. 76, p. 525-546, 2012.
- RIFFE, D.; LACY, S.; FICO, F. *Analyzing media messages: using quantitative content analysis in research*. Londres: Routledge, 2014.
- ROCHA, D.; DEUSDARÁ, B. Análise de Conteúdo e Análise do Discurso: aproximações e afastamentos na (re)construção de uma trajetória. *Alea: estudos neolatinos*, v. 7, n. 2, p. 305-322, 2005.
- ROCHA, D.; DEUSDARÁ, B. Análise de conteúdo e análise do discurso: o linguístico e seu entorno. *Delta*, v. 22, n. 1, p. 29-52, 2006.
- ROCHA, S. Análise de conteúdo articulada à análise de gênero televisivo: proposta metodológica para interpretação das representações nas narrativas midiáticas. *Revista Fronteiras*, p. 121-134, 2008.
- RODRIGUES, M. (org.). *Análise de conteúdo em pesquisas qualitativas na área de educação matemática*. Curitiba: Editora CRV, 2019.
- RODRIGUES, M.; DELLA JUSTINA, L.; MEGLIORATTI, F. O conteúdo de sistemática e filogenética em livros didáticos do ensino médio. *Ensaio Pesquisa em Educação em Ciências*, v. 13, n. 2, p. 65-84, 2011.
- RODRIGUES, M.; LEOPARDI M. *O método de análise de conteúdo: uma versão para enfermeiros*. Fortaleza: Fundação cearense de pesquisa e cultura, 1999.

- ROSSINI, P.; BAPTISTA, A.; OLIVEIRA, V.; SAMPAIO, R. O uso do Facebook nas eleições presidenciais brasileiras de 2014: a influência das pesquisas eleitorais nas estratégias das campanhas digitais. *Fronteiras-estudos midiáticos*, v. 18, n. 2, p. 145-157, 2016.
- RUI, M.; GONÇALVES, N.; ALVES, P.; GARCIA, M. Entre o discurso e o desejo: a intersectorialidade nas ações dos técnicos do CRAS e da ESF de Planalto Serrano, Serra/ES. *Revista de Políticas Públicas*, v. 14, n. 2, p. 245-253, 2010.
- SACRAMENTO, A.; PINHO, J. O processo de implementação da answerability no Brasil contemporâneo. *Revista de Administração Pública*, v. 50, n. 2, p. 193-213, 2016.
- SALDAÑA, J. *The coding manual for qualitative researchers*. Londres: Sage, 2012.
- SAMPAIO, R. C.; BRAGATTO, R. C.; NICOLÁS, M. A. A construção do campo de internet e política: análise dos artigos brasileiros apresentados entre 2000 e 2014. *Revista Brasileira de Ciência Política*, v. 4, p. 285-320, 2016.
- SAMPAIO, R.; LYCARIÃO, D. Eu quero acreditar! Da importância, formas de uso e limites dos testes de confiabilidade na Análise de Conteúdo. *Revista de Sociologia e Política*, v. 26, n. 66, p. 31-47, 2018.
- SAMPAIO, R.; LYCARIÃO, D.; CODATO, A.; MARIOTO, D.; BITTENCOURT, M.; NICHOLS, B. Uma técnica parada no tempo? Mapeamento da produção científica baseada em análise de conteúdo na SciELO Brasil (2002-19). *Preprint Scielo*. Disponível em: <https://doi.org/10.1590/SciELOPreprints.1913>
- SANTOS, M.; PENTEADO, C.; ARAÚJO, R. Metodologia de pesquisa de blogs de política: análise das eleições presidenciais de 2006 e do movimento "cansei". *Revista de Sociologia e Política*, v. 17, n. 34, p. 159-181, 2009.
- SANTOS, A.; BRAGA, I.; GUIMARÃES-IOSIF, R.; ZARDO, S. Governance: concepts and emergence in Brazilian education. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 24, n. 93, p. 939-967, 2016.
- SANTOS, J.; DALTO, J. Sobre a análise de conteúdo, análise textual discursiva e análise narrativa: investigando produções escritas em matemática. *Anais do V seminário internacional de pesquisa em Educação Matemática*. Rio de Janeiro, 2012.
- SARAIVA, L.; WAGNER, A. A relação família-escola sob a ótica de professores e pais de crianças que frequentam o ensino fundamental. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 21, n. 81, p. 739-772, 2013.
- SARMENTO, R. Estudos feministas de mídia e política: uma visão geral. *BIB-Revista brasileira de informação bibliográfica em Ciências Sociais*, v. 87, p. 181-202, 2018.
- SAUERBRONN, F.; SAUERBRONN, J. Representações sociais da reforma do Judiciário - um estudo baseado nas perspectivas dos servidores do Tribunal de Justiça do Estado da Bahia. *Revista de Administração Pública*, v. 49, n.3, 719-737, 2015.
- SILVA, D. A análise de conteúdo como método analítico no campo da política externa. In: CONGRESSO LATINO-AMERICANO DE CIÊNCIA POLÍTICA, 2017.
- SILVA, A.; CUNHA, D.; GASPARY, E.; MOURA, G.; FIGUEIRA, K.; HÖRBE, T. Análise de conteúdo: fazemos o que dizemos? Um levantamento de estudos que dizem adotar a técnica. *Conhecimento Interativo*, v. 11, p. 168-184, 2017.
- SILVA, J.; ASSIS, S. Grupo focal e análise de conteúdo como estratégia metodológica clínica-qualitativa em pesquisas nos distúrbios do desenvolvimento. *Cadernos de Pós-Graduação em Distúrbios do Desenvolvimento*, v. 10, n. 1, 2010.

- SOUSA, C.; MARQUES, F. Deliberação online em contextos de disputa eleitoral: um estudo sobre as edições da Wikipédia durante as eleições de 2012 em São Paulo. *E-Legis*, v. 21, p. 144-165, 2016.
- SOUZA, D.; MENEZES, J. Elaboração e aprovação de planos de educação no Brasil: do nacional ao local. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 23, n. 89, p. 901-936, 2015.
- SOUZA, N.; ALMEIDA, V.; DRUMMOND, D. Poder, resistência e interseccionalidade: as disputas discursivas por identidade no seriado brasileiro *Sexo e as Negras*. *EX AEQUO (OEIRAS)*, v. 35, p. 65-82, 2017.
- SOUZA JÚNIOR, M.; MELO, M.; SANTIAGO, M. A análise de conteúdo como forma de tratamento dos dados numa pesquisa qualitativa em Educação Física escolar. *Movimento (ESEFID/UFRGS)*, v. 16, n. 3, p. 29-47, 2010.
- STEENBERGEN, M.; BÄCHTIGER, A.; SPÖRNDLI, M.; STEINER, J. Measuring political deliberation: A discourse quality index. *Comparative European Politics*, v. 1, n. 1, p. 21-48, 2003.
- STEINER, J. *The foundations of deliberative democracy: empirical research and normative implications*. Cambridge: Cambridge University Press, 2012.
- STEMLER, S. An overview of content analysis. *Practical assessment, research, and evaluation*, v. 7, n. 1, p. 17, 2000.
- TAQUETTE, S.; MINAYO, M. Características de estudos qualitativos conduzidos por médicos: revisão da literatura. *Ciência & Saúde Coletiva*, v. 20, n. 8, p. 2423-2430, 2015.
- TAROUCO, G.; MADEIRA, R. Partidos, programas e o debate sobre esquerda e direita no Brasil. *Revista de Sociologia e política*, v. 21, n. 45, p. 149-165, 2013.
- THOMAZ, G.; BIZ, A.; BETTONI, E.; MENDES FILHO, L. Mineração de Conteúdo em Mídias Sociais: análise de conteúdos publicados por usuários sobre atrativos turísticos de Curitiba-PR. *Marketing & Tourism Review*, v. 1, n. 2, 2016.
- TRIVIÑOS, A. *Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em Educação*. São Paulo: Atlas, 1987.
- URZETTA, F.; CUNHA, A. Análise de uma proposta colaborativa de formação continuada de professores de ciências na perspectiva do desenvolvimento profissional docente. *Ciência & Educação*, v. 19, n. 4, p. 841-858, 2013.
- VASCONCELLOS, F. Debates eleitorais na TV como eventos de campanha. In: FIGUEIREDO, A.; BORBA, F. (Org.). *25 Anos de eleições presidenciais no Brasil*. Curitiba: Appris, 2018. p. 225-238.
- VASCONCELOS, S.; SOUTO, E. O livro didático de ciências no ensino fundamental proposta de critérios para análise do conteúdo zoológico. *Ciência & Educação*, v. 9, n. 1, p. 93-104, 2003.
- VEIGA, L.; GONDIM, S. A utilização de métodos qualitativos na ciência política e no marketing político. *Opinião pública*, v. 7, n. 1, p. 1-15, 2001.
- VERGARA, S. Réplica 2-análise de conteúdo como técnica de análise de dados qualitativos no campo da administração: potencial e desafios. *Revista de Administração Contemporânea*, v. 15, n. 4, p. 761-765, 2011.
- VERGARA, S. *Métodos de pesquisa em administração*. São Paulo: Atlas, 2005.
- VIMIEIRO, A., MAIA, R. Análise indireta de enquadramentos da mídia: uma alternativa metodológica para a identificação de frames culturais. *Revista FAMECOS*, v. 18, n. 1, p. 235, 2011.
- WEBER, R. B. *Basic content analysis*. Londres: Sage, 1990.

WHITE, D.; WUTICH, A.; LARSON, K.; GOBER, P.; LANT, T.; SENNEVILLE, C. Credibility, salience, and legitimacy of boundary objects: water managers' assessment of a simulation model in an immersive decision theater. *Science and Public Policy*, v. 37, n. 3, p. 219-232, 2010.

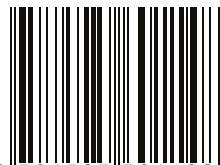
WHITE, M.; MARSH, E. Content analysis: A flexible methodology. *Library trends*, v. 55, n. 1, p. 22-45, 2006.

WOZNIAK, A., LÜCK, J., WESSLER, H. Frames, stories, and images: the advantages of a multimodal approach in comparative media content research on climate change. *Environmental Communication*, v. 9, n. 4, p. 469-490, 2015.

YIN, R. *Pesquisa qualitativa do início ao fim*. São Paulo: Penso Editora, 2016.

Este livro pretende apresentar, de modo simplificado, como realizar uma análise de conteúdo (AC) categorial quantitativa com bases científicas. Não obstante seu surgimento no campo da comunicação política, a análise de conteúdo tornou-se uma técnica bastante difundida em toda a área de ciências sociais aplicadas e ciências humanas como, por exemplo, Administração, Comunicação, Educação, Ciência Política, Políticas Públicas e também nas ciências da saúde, como Enfermagem e Saúde Coletiva. Um dos principais motivos que nos incentivou a escrever este livro está justamente na falta de opções de outros manuais especializados no Brasil. Assim, desejamos especialmente contribuir para o debate sobre os avanços da aplicação da AC e oferecer um tutorial simplificado para estudantes de pós-graduação e graduação sem acesso a manuais em outras línguas.

ISBN 978-65-87791-18-0



9 786587 791180