

Metodologia de *Survey* na Pesquisa em Ciências Sociais

Ricardo Ceneviva

ricardo.ceneviva@ufabc.edu.br

Outubro de 2025

Resumo

Este capítulo oferece uma introdução sistemática e aplicada à metodologia de survey na pesquisa sociológica contemporânea, com foco nas possibilidades e limites do método. Partindo da definição de “survey moderno” (coleta padronizada, amostragem definida, inferência populacional e documentação sob o arcabouço do Erro Total de Survey — TSE), organizamos a discussão em torno das duas decisões centrais de desenho: **quem entrevistar** (amostragem, cobertura, não resposta, ponderação/calibração e variância sob desenho) e **o que perguntar** (validade e confiabilidade da mensuração, efeitos de redação, ordem e modo). Como fio condutor didático, utilizamos o caso da **polarização afetiva**, operacionalizada por itens simples (termômetros de sentimento, distância social e traços/estereótipos) e métricas derivadas (IPA e IDS), para ilustrar escolhas de mensuração e seus *trade-offs*. No plano metodológico, integramos o TSE em dois blocos — estrutural/estocástico (cobertura, amostragem, processamento/ajustes) e cognitivo/comportamental (não resposta, desejabilidade social, *satisficing*) — e discutimos como **modos de coleta** (presencial, telefone, web e desenho multimodo) afetam simultaneamente representação e mensuração. Abordamos ainda **comparabilidade** (tradução/retrotradução e invariância de medida) e **governança ética e legal** (LGPD, consentimento, anonimização) como condições para inferência responsável. O capítulo conclui com um checklist de boas práticas de transparência e reprodutibilidade (documentação, dados e códigos), enfatizando que a qualidade inferencial em surveys decorre do acoplamento entre desenho amostral, mensuração válida e gestão explícita do TSE.

Palavras-chave: metodologia de survey; erro total de survey (TSE); amostragem; mensuração; modos de coleta; polarização afetiva; ponderação e calibração; invariância de medida; LGPD; transparência e reprodutibilidade.

Lista de Siglas e Abreviaturas

- **TSE** – Total Survey Error (Erro Total de Survey) – conceito que abrange todas as fontes de erro em uma pesquisa por survey.
- **LGPD** – Lei Geral de Proteção de Dados Pessoais – legislação brasileira de proteção de dados pessoais (Lei n. 13.709/2018).
- **IPA** – Índice de Polarização Afetiva – métrica que quantifica a polarização afetiva, tipicamente como a diferença na avaliação de calor afetivo entre grupos políticos opostos.
- **IDS** – Índice de Distância Social – métrica que quantifica a distância social desejada em relação a membros de um grupo social ou político adversário (e.g., relutância em ter relacionamentos próximos com out-group).

Introdução

No campo da sociologia política contemporânea, os surveys (inquéritos por amostragem) tornaram-se uma ferramenta central para medir opiniões e atitudes em larga escala. A pesquisa por survey envolve coletar sistematicamente informações de uma amostra de respondentes, para inferir características de uma população maior[1].

Este relatório explora em profundidade a metodologia de surveys, cobrindo desde os fundamentos conceituais até as melhores práticas de implementação e análise, empregando o fenômeno da polarização afetiva como caso ilustrativo. A polarização afetiva – entendida como o antagonismo emocional entre grupos políticos rivais – serve aqui apenas como fio condutor metodológico, sem adentrar em discussões substantivas sobre o tema. Serão abordados conceitos chave como a definição de um survey moderno segundo Groves et al. (2009), decisões de desenho amostral e de questionário, o arcabouço do Erro Total de Survey (TSE), modos de aplicação e seus efeitos, técnicas de ponderação e calibração (e implicações na variância), comparabilidade em surveys multiculturais, além de considerações éticas e legais (com ênfase na LGPD para surveys on-line). Por fim, apresentam-se boas práticas de transparência e reprodutibilidade em pesquisa de survey, um checklist de planejamento e uma síntese conclusiva. A linguagem adotada é acessível a leitores iniciantes, sem prejuízo do rigor conceitual e metodológico exigido em contexto acadêmico.

Conceitos Fundamentais e Definição de Survey Moderno

Survey pode ser definido, de forma geral, como um método de pesquisa quantitativa de grande escala, baseado na coleta sistemática de informações de um subconjunto de indivíduos (amostra) que representa uma população de interesse[1]. Groves et al. (2009)

caracterizam o survey moderno como um processo estruturado de levantamento de dados junto a respondentes selecionados, com o objetivo de inferir proporções, médias ou relações presentes no universo populacional[1]. Em outras palavras, trata-se de realizar perguntas padronizadas a uma amostra (geralmente numerosa) de pessoas, obtendo dados que, mediante técnicas estatísticas, permitem descrever ou analisar atributos da população.

Os surveys contemporâneos apoiam-se em princípios científicos de amostragem e mensuração. Eles evoluíram como principal forma de medir opinião pública desde os anos 1930 (era dos institutos de pesquisa de opinião)[2]. Seu predomínio decorre da capacidade de, com base em relativamente poucos casos, “tirar a foto” do que um público maior pensa ou sente sobre determinado assunto. Contudo, como enfatiza Berinsky (2017), ainda que os surveys tenham longa história, não há consenso absoluto entre cientistas sociais sobre a melhor forma de medir opinião pública através de inquéritos, o que exige consciência crítica das escolhas de metodologia feitas em cada pesquisa[3]. Essas escolhas envolvem duas dimensões fundamentais: quem entrevistar e o quê perguntar[3]. Antes de mergulharmos nesses aspectos práticos, é importante delinear a estrutura conceitual que orienta a qualidade de uma pesquisa survey – o arcabouço do Erro Total de Survey – pois ele servirá de referência transversal para as decisões metodológicas discutidas nos próximos tópicos.

Amostragem: Quem Entrevistar?

A definição do público-alvo (população de interesse) e a estratégia de amostragem constituem o alicerce de qualquer survey. A população pode ser ampla – por exemplo, “todos os adultos residentes no Brasil” – ou específica, como “estudantes universitários de 18 a 24 anos na cidade de São Paulo”. Uma vez definida a população, é preciso obter uma lista (frame) de onde selecionar a amostra – por exemplo, cadastro de eleitores, lista de endereços, números de telefone gerados aleatoriamente, etc. Caso não exista uma listagem perfeita dos elementos populacionais, ocorrem erros de cobertura (quando o frame não contempla partes da população)[4][5]. Por exemplo, um survey por telefone fixo pode excluir pessoas que só usam celular; um survey online pode não alcançar quem não tem acesso à Internet, gerando um descompasso entre população-alvo e população acessível (cobertura incompleta)[4].

Com o frame em mãos, seleciona-se a amostra. O método mais recomendado, do ponto de vista inferencial, é a amostragem probabilística, na qual cada indivíduo tem chance conhecida de ser selecionado (ex.: sorteio aleatório simples, amostragem estratificada ou por conglomerados). Esse procedimento permite calcular a margem de erro estatística devida à amostragem – o erro amostral –, que deriva do uso de um subconjunto e não de um censo[6]. Erro amostral é puramente aleatório (resulta do acaso na seleção) e desaparece em um censo completo[7]. Em contrapartida, em amostras não probabilísticas (por conveniência, voluntários de painel online, “bola de neve” etc.), não se pode assegurar representatividade

estatística nem calcular a probabilidade de erro, o que pode introduzir vieses sistemáticos. Uma decisão importante é o tamanho da amostra, que afeta a precisão das estimativas: amostras maiores tendem a reduzir o erro amostral (a margem de erro diminui), embora com custos crescentes. No entanto, mais respondentes não corrigem erros de viés; por exemplo, se a amostra estiver enviesada por subrepresentar certo grupo, aumentar o n não elimina esse viés – a solução, nesse caso, é aprimorar o plano amostral ou aplicar ponderações.

Quem entrevistar, no caso da polarização afetiva? Se quisermos medir a polarização afetiva na sociedade, a população de interesse provavelmente seriam eleitores (ou cidadãos em geral) de um país. Para captar adequadamente diferenças afetivas entre grupos políticos, é essencial que a amostra inclua proporções realistas de adeptos de diferentes correntes ou partidos. Uma amostra enviesada que contenha poucos simpatizantes de um dos lados polarizados resultaria em métricas de polarização sub ou superestimadas. Por exemplo, se em uma pesquisa hipotética somente 10% da amostra se identifica com o partido X quando, na população real, esse grupo corresponde a 30%, as medidas de antipatia entre grupos podem não refletir a realidade. Portanto, é recomendável estratificar a amostragem pela delcação de voto nas últimas eleições ou pela preferência partidária, ou ao menos acompanhar quotas de variáveis correlacionadas (região, idade, escolaridade) para garantir presença suficiente de todos os segmentos politicamente relevantes.

Outra preocupação é a taxa de resposta. Mesmo com amostragem probabilística, nem todos os selecionados respondem à pesquisa, gerando erro de não-resposta – diferenças potenciais entre quem respondeu e quem não respondeu (no nível do indivíduo inteiro, chamado *unit nonresponse*, ou de perguntas específicas, *item nonresponse*)[8]. A não-resposta pode introduzir viés se certos grupos têm menor propensão a participar. Por exemplo, em um survey político, simpatizantes desiludidos ou antipartidários extremos poderiam se recusar a responder por desconfiança, o que enviesaria os resultados sobre polarização afetiva. Para mitigar isso, usam-se técnicas de acompanhamento (*follow-up*) de não respondentes, incentivos para resposta, e posteriormente ponderação de não-resposta, ajustando pesos de acordo com perfis demográficos para alinhar a amostra respondente à população (ver seção de Ponderação). Berinsky (2017) destaca que escolher quem entrevistar envolve lidar com desafios de cobertura e participação, e que tais decisões impactam diretamente a validade das inferências sobre a opinião pública obtidas pelo survey[3].

Em suma, amostragem rigorosa é condição necessária para um survey de qualidade. Definir claramente a população, obter (ou construir) um frame abrangente, selecionar uma amostra probabilística adequada e maximizar a resposta são etapas fundamentais para garantir representatividade – isto é, que os entrevistados efetivamente “representem” o todo. Esse cuidado estruturante reduz erros de representação, um dos dois blocos principais do Erro Total de Survey, conforme veremos adiante.

Mensuração: O Que Perguntar?

Tão importante quanto escolher os entrevistados é definir o que será perguntado e como. A fase de mensuração abrange a elaboração do questionário e das variáveis que medirão os conceitos de interesse. Uma boa mensuração começa pela operacionalização: definir em termos observáveis aquilo que se deseja investigar. No caso da polarização afetiva, que é um conceito abstrato (antagonismo emocional entre adversários políticos), é preciso traduzi-lo em perguntas concretas que capturem esse sentimento. Poderíamos operacionalizá-la, por exemplo, pedindo aos respondentes que avaliem em uma escala sua simpatia ou antipatia por determinados grupos ou líderes políticos adversários, ou que indiquem quão dispostos estariam a interações sociais próximas com pessoas do outro campo político.

As perguntas devem ter validade de conteúdo, ou seja, realmente refletir o construto teórico. Groves et al. (2009) apontam que problemas de validade ocorrem quando há descompasso entre o tema de pesquisa e as perguntas usadas para mensurá-lo[9]. Se as questões não capturam bem o fenômeno desejado, teremos erro de validade. Por exemplo, medir polarização afetiva apenas perguntando sobre posicionamento ideológico (esquerda-direita) incorre em baixa validade, pois polarização afetiva diz respeito a sentimentos por grupos, não a ideologia em si. Assim, deve-se formular itens que captem efetivamente os afetos (positivos ou negativos) em relação a grupos políticos.

Alguns exemplos de itens utilizados para aferir polarização afetiva incluem:

- **“Termômetros” de sentimento:** o respondente atribui uma nota (ex.: de 0 a 100, ou 0 a 10) indicando o quão frio/negativo ou quente/positivo se sente em relação a determinado grupo, partido ou figura pública. Por exemplo: “Em uma escala de 0 (desgostar totalmente) a 10 (gostar muito), que nota você dá para o grupo/partido X? E para o grupo/partido Y?”. A polarização afetiva pode então ser mensurada pela diferença entre as notas dadas ao grupo político preferido vs. ao adversário. Essa diferença constitui um Índice de Polarização Afetiva (IPA) individual – quanto maior a discrepância, maior a polarização afetiva daquele indivíduo. Se calcularmos o valor médio dessa diferença absoluta na amostra, obtemos um indicador agregado de polarização afetiva na população[10][11]. Por exemplo, suponha que um eleitor deu nota 8 (em 0–10) ao político de sua preferência e nota 2 ao oponente; o IPA individual seria 6 pontos. Em média, se a amostra inteira apresenta IPA médio de ~ 7 pontos, indica alto nível de polarização afetiva agregada. Importante notar que o IPA pode ser considerado com ou sem sinal – se quisermos saber não apenas o grau, mas também o lado da polarização, poderíamos calcular uma média não-absoluta (onde sinais positivos ou negativos indicam direção do viés afetivo). No entanto, usualmente reporta-se o valor absoluto médio $|IPA|$ para quantificar a intensidade da polarização afetiva independentemente de lado.
- **Itens de distância social:** perguntas inspiradas na tradição da escala de distância

social de Bogardus, adaptadas para contexto político. Por exemplo: “Quão confortável você se sentiria se seu filho(a) casasse com uma pessoa que apoia o partido adversário ao seu?” com opções de resposta variando de “sem nenhuma objeção” a “com certeza eu ficaria infeliz”. Respostas de desconforto ou infelicidade com tal cenário indicam alta polarização afetiva, pois revelam recusa a vínculos próximos com membros do out-group[12]. Podem-se incluir outras situações: trabalhar com alguém do outro partido, ter como vizinho, amigo próximo, etc. A partir dessas respostas, pode-se construir um Índice de Distância Social (IDS) – por exemplo, a porcentagem de respondentes que se dizem “infelizes” com um relacionamento familiar envolvendo um opositor político[12], ou uma escala somando pontos conforme o nível de rejeição a contatos sociais com o out-group. Esse índice reflete a profundidade da clivagem afetiva na vida social.

- **Itens de traços e estereótipos:** questões que avaliam atributos percebidos dos partidários do outro grupo – por exemplo, pedir para classificar se os membros do partido rival são “honestos ou desonestos”, “inteligentes ou ignorantes”, “pacíficos ou agressivos”, etc. A atribuição sistemática de traços negativos ao grupo opositor (e positivos ao próprio grupo) é sintoma de polarização afetiva. Métricas podem ser derivadas da diferença nas avaliações de traços concedidas ao in-group vs. out-group, similar ao termômetro.

Na construção dessas perguntas, a redação importa imensamente. Deve-se usar linguagem clara, evitar termos técnicos ou ambíguos, e atentar para vieses de formulação. Perguntas enviesadas ou sugestivas acarretam erro de mensuração, que ocorre quando há discrepância entre o que o pesquisador pretendeu perguntar e como o respondente de fato entendeu e respondeu[9]. Por exemplo, perguntar “Você odeia completamente o partido X?” é tendencioso e extremo; pode induzir respostas defensivas. Melhor seria usar termos neutros: “Qual seu sentimento em relação ao partido X?” com opções que permitam gradientes (como o termômetro ou escalas de gostar/desgostar). Ainda, a ordem das perguntas e das alternativas deve ser considerada – efeitos de contexto podem influenciar respostas subsequentes. Em surveys políticos, costuma-se embaralhar a ordem de apresentação de nomes de candidatos ou partidos para evitar viés de primazia ou novidade.

Confiabilidade também é crucial: refere-se à consistência das medições. Se repetíssemos a pergunta, obteríamos resposta semelhante? Garantir confiabilidade envolve padronização na aplicação (todos os entrevistados recebem as mesmas perguntas, nas mesmas condições), perguntas específicas e não vagas, e, quando possível, uso de múltiplos itens para o mesmo construto (por exemplo, vários itens de traços que em conjunto indiquem hostilidade afetiva, permitindo depois calcular um alpha de Cronbach).

O processo de resposta a uma pergunta de survey tem um componente cognitivo-comportamental complexo. Tourangeau, Rips e Rasinski (2000) propuseram um modelo de quatro etapas

mentais pelo qual o respondente passa ao responder: compreensão da pergunta, recuperação da informação da memória, julgamento/estimativa com base nas memórias recuperadas, e formatação da resposta adequando-a às opções oferecidas[9]. Problemas em qualquer estágio geram erros: se a pergunta não for bem compreendida, ou se o respondente não lembrar direito do fato perguntado, ou ainda se tiver de chutar uma resposta por dificuldade de julgamento, o dado obtido pode não refletir a “verdade” que se quer medir. Por isso, testar cognitivamente as perguntas antes do campo – por meio de entrevistas cognitivas ou testes piloto – é uma boa prática para identificar mal-entendidos, interpretações dúbias ou opções confusas. Essa etapa preemptiva consome tempo, mas aumenta a qualidade da mensuração, reduzindo erros de medida[13].

No exemplo da polarização afetiva, testes prévios poderiam revelar, por exemplo, se todos os respondentes entendem da mesma forma o conceito de “grupo político contrário” (alguns podem pensar em partidários do partido X, outros em adeptos de certa ideologia). Ou se a escala 0–10 de sentimento é usada de forma consistente (há culturas em que evitar extremos é comum, o que afetaria médias). Esses refinamentos garantem que os dados de opinião coletados correspondam fielmente aos conceitos pesquisados, assegurando validade e confiabilidade.

Resumindo, o quê perguntar deve ser guiado pela teoria (boa operacionalização do conceito) e pelas melhores práticas de redação e teste de perguntas. A mensuração cuidadosa é o segundo pilar – juntamente com a amostragem – que sustenta um survey robusto. Uma vez definidos “quem” e “o quê”, podemos examinar de forma integrada o conjunto de possíveis erros que ameaçam a pesquisa e como controlá-los – isto é, o Erro Total de Survey.

Erro Total de Survey (TSE): Arcabouço Estrutural vs. Comportamental

O Erro Total de Survey (TSE) é um arcabouço conceitual que engloba todas as fontes de erro que podem ocorrer em um survey, desde o planejamento até a coleta e análise dos dados[14]. Em vez de focalizar apenas o erro amostral (margem de erro), o paradigma TSE alerta que há múltiplos erros potenciais – alguns de natureza mais “estrutural ou estocástica” (ligados ao plano amostral e procedimentos) e outros de natureza “cognitivo-comportamental” (ligados ao comportamento dos respondentes e entrevistadores e aos processos mentais de resposta). A literatura frequentemente organiza esses erros em dois grandes blocos: erros de representação (ou de representatividade) e erros de mensuração[15][9]. Podemos correlacionar esses blocos com a classificação pedida de estrutural/estocástico vs. cognitivo/comportamental, como detalhado a seguir:

Erros de Representação (Estruturais/Estocásticos): são aqueles que afetam quem de fato entra na amostra e no conjunto de respostas, ou seja, dizem respeito à relação

entre a amostra obtida e a população pretendida. Incluem:

- **Erro de Cobertura:** ocorre quando a lista (frame) usada para selecionar a amostra não cobre perfeitamente a população-alvo. Há um mismatch entre população e frame[6]. Exemplo: realizar um survey via internet sobre população geral gera erro de cobertura, pois exclui quem não tem acesso online. É um erro estrutural de desenho, que existe a priori da coleta. Para mitigá-lo, busca-se frames mais abrangentes (e.g., combinar múltiplas fontes)[16] ou ajustar o modo de coleta (usar múltiplos modos para alcançar quem ficaria de fora em apenas um, ver seção de Modos).
- **Erro de Amostragem:** deriva do fato de se ouvir apenas uma parte (amostra) e não todo o universo. Mesmo sem nenhum viés, amostrar aleatoriamente implica incerteza estatística – a cada subamostra os resultados variam um pouco ao redor do verdadeiro valor populacional. É o erro aleatório (estocástico) medido pela margem de erro. Embora inevitável em amostras, pode-se reduzi-lo aumentando o tamanho amostral ou otimizando o desenho (e.g., estratificar para garantir representação de subgrupos), dentro do limite de recursos disponíveis[16]. Diferentemente de outros erros, o amostral tende a zero se a amostra for infinitamente grande ou um censo (portanto não é viés, e sim imprecisão).
- **Erro de Não-Resposta:** surge quando nem todos os selecionados respondem (unit nonresponse) ou deixam perguntas sem resposta (item nonresponse). Se os não-respondentes diferem sistematicamente dos respondentes quanto ao tema pesquisado, introduz-se viés. Por exemplo, se em um survey político os eleitores mais desinteressados ou mais radicalizados se recusam a responder, as estimativas de polarização afetiva podem ficar subestimadas (faltando justamente os mais polarizados)[8]. Esse erro tem componente comportamental (decisão do indivíduo de participar ou não), mas é tratado no bloco de representação porque afeta quem compõe o conjunto final de dados. Estratégias para lidar com não-resposta incluem design de coleta com múltiplas tentativas de contato, uso de diferentes modos, e ajustes pós-coleta via ponderação para alinhar a amostra respondente à composição conhecida da população (por exemplo, calibrar para margens demográficas oficiais)[17]. Contudo, ponderação só corrige completamente se as diferenças forem explicáveis pelas variáveis usadas no ajuste; diferenças inobserváveis podem persistir.
- **Erro de Ajuste (Pós-Coleta):** refere-se a erros introduzidos nas etapas de ajuste estatístico dos dados, como ponderação, pós-estratificação, imputação de respostas faltantes, calibração etc. Embora esses procedimentos sejam necessários para corrigir outros problemas (não-resposta, cobertura), eles próprios podem inserir vieses se mal executados ou baseados em informações imprecisas[18]. Por exemplo, se ao ponderar atribuímos peso excessivo a poucos respondentes de um grupo minoritário, qualquer erro ou acaso nas respostas desses indivíduos fica exagerado, podendo distorcer os

resultados (um efeito colateral conhecido como ampliação da variância – ver discussão de ponderação mais adiante). Assim, o erro de ajuste é um risco estrutural decorrente das decisões de processamento dos dados brutos.

Em síntese, os erros de representação fazem com que os respondentes obtidos não correspondam perfeitamente à população que se quer estudar – seja porque alguns nunca tiveram chance de ser selecionados (cobertura), seja porque o acaso e a participação limitada intervieram (amostragem e não-resposta), seja por ajustes imperfeitos. O resultado típico de erros de representação são estimativas enviesadas (p. ex., superestimar uma opinião presente em excesso na amostra e subestimar outra ausente).

Erros de Mensuração (Cognitivos/Comportamentais): ocorrem quando, dada uma amostra de respondentes, as respostas fornecidas não refletem com exatidão os valores verdadeiros ou o significado pretendido. Estão ligados ao processo de resposta e à interação pesquisador-respondente. Incluem:

- **Erro de Validade (construct validity):** ocorre se as perguntas ou indicadores escolhidos não medem de fato o conceito teórico alvo[9]. É um problema conceitual: a pesquisa pode estar perguntando algo diferente do que imagina. Por exemplo, medir “polarização afetiva” perguntando só sobre discordância em políticas públicas confunde afetividade com ideologia. Esse erro se previne na fase de desenho, escolhendo boas definições operacionais e, se possível, utilizando baterias de itens já validadas na literatura. Exemplo: utilizar termômetros de sentimento é validado em inúmeros estudos como medida apropriada de afetividade partidária, ao passo que perguntar “Você concorda com a ideologia do partido oposto?” não capta diretamente afeto e misturaria componente racional/ideológico. Colaborar com especialistas em psicometria ou testar a estrutura fatorial das respostas (ver invariância de medida adiante) ajuda a garantir validade de construto.
- **Erro de Mensuração (resposta):** trata de discrepâncias entre a resposta dada e o valor “verdadeiro” que se gostaria de obter[9]. Aqui estão vários fenômenos: vieses de resposta (como desejabilidade social, aquiescência, tendenciosidade do entrevistador), compreensões errôneas da pergunta, erro de recordação, respostas aleatórias por desinteresse ou falta de conhecimento (satisficing). Por exemplo, por desejabilidade social, um respondente pode atenuar declarações de ódio ao grupo adversário numa entrevista cara a cara para não parecer intolerante, embora pessoalmente ele desaprove fortemente o outro grupo. Isso resulta em sub-relato do verdadeiro antagonismo (erro de mensuração por viés de resposta). A literatura mostra que esse viés é menor em modos autoaplicados anônimos (como web) comparado a entrevistas presenciais ou telefônicas[19] – retornaremos a isso em Modos de Coleta. Outro componente é o erro introduzido pelo próprio entrevistador humano, caso haja: entonações ou explicações não padronizadas podem influenciar respostas (por isso

treinamentos e scripts rígidos são importantes). Em surveys autoaplicados, erros de mensuração podem vir de interpretação ambígua de alguma palavra ou da tendência de o respondente marcar opções sem muita reflexão quando cansado (por exemplo, escolhendo sempre “neutro” ou repetindo um padrão nas escalas). Minimizar erros de mensuração requer questionário bem formulado, pré-testado e, idealmente, simples e curto o suficiente para manter engajamento, além de técnicas como randomizar ordem de itens para diluir efeitos de posição.

- **Erro de Processamento:** engloba problemas ocorridos após a coleta, na etapa de manipulação dos dados[9]. Exemplos: erros de codificação (quando respostas abertas são categorizadas), erros de digitação, falhas ao transcrever resultados, ou bugs em programas de análise. Por exemplo, se por engano os valores da escala 0–10 forem recodificados invertidos para metade da amostra, as estatísticas calculadas estarão incorretas. Esses erros são comportamentais no sentido de serem cometidos por humanos ou sistemas durante o processamento. Para mitigá-los, adota-se dupla checagem, procedimentos automatizados de consistência, e documentação cuidadosa de todas as transformações realizadas nos dados.

O arcabouço TSE enfatiza que todo survey possui erros – eliminar completamente cada fonte é impraticável –, mas o objetivo do pesquisador é controlar e reduzir o erro total dentro dos recursos disponíveis[14]. Importante, há trade-offs: esforços para diminuir uma fonte de erro podem aumentar outra. Groves et al. (2009) exemplificam: ampliar o frame para cobrir melhor a população pode ser custoso ou inviável; aumentar a amostra reduz erro amostral mas pode intensificar problemas de não-resposta ou elevar custo; muita insistência para obter resposta pode introduzir vieses de resposta satisfeito (respondente apressado)[16][17]; aplicação de ponderação corrige vieses de representação mas pode aumentar variância e até induzir novos erros se baseada em suposições erradas[18]; perguntas altamente validadas e detalhadas consomem tempo e podem levar ao abandono da entrevista (aumentando não-resposta). Assim, o pesquisador deve buscar um equilíbrio ótimo entre as dimensões de erro, considerando os objetivos do estudo e os recursos (tempo, orçamento). A chave do TSE é fornecer um mapa para identificar onde a pesquisa pode falhar em qualidade e permitir decisões informadas de desenho para maximizar a qualidade dos dados dentro das limitações práticas[20][16].

No contexto do nosso caso-guia, pensar em TSE auxilia, por exemplo, a identificar riscos: na medição da polarização afetiva, erros de representação poderiam surgir se certos subgrupos politicamente relevantes não estiverem bem representados (cobertura/não-resposta) – quem sabe os mais radicalizados evitam responder pesquisas, ou jovens engajados só estejam online. Já erros de mensuração poderiam ocorrer se os respondentes moderassem suas expressões de antipatia dependendo do modo (temas sensíveis em público) ou interpretassem diferentemente escalas de sentimento. Estar ciente dessas possibilidades permite adotar

contramedidas: reforçar coleta junto a segmentos de difícil acesso; garantir anonimato e modo apropriado para honestidade; e testar se as medidas de afeto coletadas realmente refletem o antagonismo (validade). A aplicação prática do TSE, portanto, permeia todas as próximas etapas – modo de coleta, ponderação, comparabilidade e ética –, guiando-nos na condução de um survey confiável.

Modos de Coleta e Efeitos de Modo

O modo de coleta de um survey refere-se à forma como o questionário é administrado aos respondentes. Os principais modos clássicos incluem: entrevista face a face (presencial, em domicílio ou local público), entrevista por telefone (tradicionalmente fixo, mas hoje também celular), questionário autopreenchido em papel (enviado por correio ou entregue) e survey via Internet (questionários online, seja por email, link web ou painel). Há ainda modalidades mistas ou intermediárias, como entrevistas assistidas por computador (CATI para telefone, CAPI para presencial, ou CASI no computador pelo próprio respondente), e novas plataformas como aplicativos móveis e mensagens instantâneas.

Cada modo apresenta vantagens e limitações, e a escolha impacta tanto erros de representação quanto de mensuração. Do lado da representação, o modo condiciona quem pode ser alcançado: surveys pela internet, por exemplo, embora rápidos e de baixo custo, excluem pessoas sem acesso ou habilidade digital (gerando erro de cobertura, principalmente entre idosos ou populações rurais desconectadas). Surveys por telefone enfrentam problemas semelhantes hoje, pois linhas fixas caíram em desuso entre jovens, enquanto listas de celulares podem não estar disponíveis e sofrem alta taxa de não-resposta (muitas pessoas não atendem números desconhecidos). Já entrevistas presenciais, apesar de oferecerem melhor controle amostral (podendo visitar domicílios selecionados aleatoriamente), têm custo altíssimo e logísticas complexas, o que frequentemente limita o tamanho amostral ou escopo geográfico. Surveys postais podem atingir quase todos com endereço, mas sofrem de retorno baixo e demora. Assim, muitas pesquisas optam por modos mistos, combinando estratégias (ex.: primeiro enviar carta com link online, depois ligar para quem não respondeu, etc.) para melhorar cobertura e resposta.

Do lado da mensuração, o modo influencia como o respondente interage com o questionário e como isso afeta as respostas. Um aspecto notório é o já citado viés de desejabilidade social: em presença de um entrevistador (face a face ou telefone), as pessoas tendem a dar respostas mais socialmente aceitas e menos embaraçosas, comparado a respostas anônimas em papel ou web[19]. Assim, para temas sensíveis (preconceitos, comportamentos não aceitáveis, intensidades de antipatia), modos autoaplicados tendem a colher respostas mais sinceras. Pesquisa de Tourangeau et al. (2013) e outros demonstram que, por exemplo, respondentes relatam maior preconceito ou comportamentos estigmatizados em questionários online do que em entrevistas telefônicas, indicando que o anonimato reduz a

inibição da desejabilidade social[19][21]. No nosso exemplo, alguém altamente polarizado afetivamente pode relutar em expressar ódio pelo grupo oposto a um entrevistador humano (por receio de julgamento), mas se estiver sozinho diante de uma tela pode atribuir nota 0 ao oponente sem constrangimento. Portanto, para medir verdadeiramente sentimentos negativos intensos, um modo auto administrado poderia captar melhor a polarização latente.

Por outro lado, entrevistas com entrevistador permitem esclarecer dúvidas do respondente e assegurar que ele siga as instruções corretamente, reduzindo certos erros de mensuração como má interpretação ou perguntas ignoradas. Já nos modos autoaplicados, um respondente pode pular itens difíceis ou interpretar errado sem ninguém para auxiliar. Além disso, aspectos de formato influenciam: numa entrevista oral (telefone), listas longas de opções são mais difíceis de recordar e podem levar a satisficing (escolher algo “bom o bastante” rapidamente); já num questionário visual (web/papel), o respondente vê todas as opções e pode compará-las antes de marcar, o que pode reduzir alguns erros mas introduzir outros (como tendência a escolher opções do meio da lista em visuais, ou efeitos de layout). Há também diferenças de taxa de respostas faltantes: pesquisas postais e web às vezes têm item nonresponse maior porque é fácil deixar em branco, enquanto entrevistadores normalmente forçam resposta (perguntando de novo se a pessoa diz “não sei” ou pulou sem querer).

Efeito de modo é o termo para diferenças sistemáticas nas respostas atribuíveis ao modo de coleta. Por exemplo, o Pew Research Center encontrou que, embora muitas estimativas não mudem drasticamente, há casos onde as respostas variam ~5–10 pontos percentuais entre um survey telefônico e online, especialmente em perguntas sensíveis ou de formatação diversa (questões de satisfação com a vida tendem a ser mais positivas por telefone, possivelmente por desejabilidade; perguntas de conhecimento tendem a ter mais “não sei” no autoaplicado onde não há pressão para responder)[22].

Para acomodar efeitos de modo, pesquisadores podem: (a) calibrar questionários para cada modo, garantindo que visualmente ou textualmente as versões sejam equivalentes; (b) se necessário, ajustar estatisticamente as diferenças ou ao menos reconhecer a possibilidade delas na análise; (c) escolher um modo alinhado ao tema (por exemplo, usar web para obter maior privacidade em temas delicados, ou usar face a face para aplicar testes cognitivos complexos que requerem monitoramento). Dillman, Smyth e Christian (2014) enfatizam o *Tailored Design Method*, isto é, adaptar o design do survey (inclusive modo e contatos) para maximizar participação e qualidade dos dados em cada contexto específico[23]. Isso inclui, por exemplo, enviar pré-aviso por carta mesmo que a resposta seja online, usar layouts de questionário web amigáveis (com barras de progresso, divisão em páginas curtas), ou no caso de telefone, treinar entrevistadores para ler perguntas exatamente como escritas, em tom neutro.

No contexto de surveys sobre polarização afetiva, a escolha do modo pode afetar direta-

mente o resultado. Pesquisas recentes no Brasil, por exemplo, usaram termos fortes de desaprovação ao outro grupo (como perguntar se ficaria “infeliz” com um familiar do grupo oposto) e aplicaram via websurvey para garantir anonimato[12]. Isso provavelmente captou níveis altos de polarização (mais pessoas admitindo infelicidade) comparado a se a mesma pergunta fosse feita ao vivo. Entretanto, deve-se cuidar com possíveis vieses de cobertura ao optar por web: se grupos importantes (ex: eleitores mais idosos ou de baixa renda) ficarem de fora on-line, a medição global da polarização pode ficar distorcida. Uma solução é o modo misto: por exemplo, combinar coleta web (para quem tem acesso) com entrevistas telefônicas (para quem não respondeu on-line), aplicando o mesmo questionário. Assim se tenta equilibrar cobertura e efeitos de modo. Nesse caso, é prudente testar se há diferença sistemática nas respostas entre modos – se houver, pode-se inserir uma variável de controle no modelo ou analisar separadamente. No ideal, todos os respondentes estariam no mesmo modo para comparabilidade interna, mas a realidade muitas vezes impõe combinações. Concluindo, a decisão do modo de coleta envolve pesar considerações de custo, cobertura, conveniência e qualidade da resposta. Modos autoaplicados podem melhorar a franqueza das respostas em tópicos sensíveis e reduzir custo por entrevista, mas demandam atenção para inclusão digital e clareza do questionário. Modos com entrevistador podem atingir amostras mais representativas (em especial face a face com sorteio domiciliar) e garantir preenchimento completo, porém custam caro e podem introduzir certos vieses (desejabilidade, efeito entrevistador). Conhecer esses trade-offs permite ao pesquisador projetar surveys mais eficazes. Independentemente do modo escolhido, práticas como treinamento de entrevistadores, teste piloto do instrumento e monitoramento de dados em tempo real (para identificar inconsistências ou padrões suspeitos de respostas) contribuem para mitigar problemas de modo e melhorar a confiabilidade dos dados coletados.

Ponderação, Calibração e Variância sob Desenho

Após coletar os dados, é comum aplicar ponderação (weighting) para ajustar a amostra obtida de forma que ela melhor represente a população-alvo. Cada respondente recebe um peso que indica quantas pessoas daquele perfil ele representa na população. Na prática, os pesos compensam desproporções amostrais e efeitos de não-resposta. Por exemplo, se mulheres com ensino superior ficaram sobre-representadas na amostra (digamos 60% da amostra contra 30% na população), seus pesos serão ajustados para valer menos ($\text{peso} < 1$), enquanto homens jovens eventualmente sub-representados terão $\text{peso} > 1$ para compensar. O objetivo é minimizar vieses de representação pós-coleta, aliviando erros de cobertura e não-resposta remanescentes[18].

Um tipo específico de ajuste de pesos é a calibração (ou pós-estratificação), na qual se define controle para certas margens populacionais conhecidas (por exemplo, distribuir pesos de modo que a amostra ponderada tenha exatamente 50% mulheres e 50% homens,

ou reflita a distribuição oficial por faixa etária, região, nível educacional, etc.). Métodos comuns de calibração incluem *raking* (iterative proportional fitting) e uso de algoritmos de minimização de distância que modificam pesos originais para atender múltiplos totais simultaneamente. Em surveys eleitorais ou de opinião, costuma-se calibrar pelos dados censitários e, às vezes, por variáveis como voto anterior (quando conhecidas). No nosso caso ilustrativo, poderíamos calibrar para que a proporção de simpatizantes de diferentes partidos na amostra reflita alguma medida externa confiável (como resultado eleitoral ou pesquisas prévias), embora nem sempre haja esse dado – mais frequentemente calibra-se por demografia e assume-se que dentro de estratos demográficos a variância política está alinhada.

É importante destacar que a ponderação é uma faca de dois gumes: corrige vieses mas introduz incremento de variância. Intuitivamente, ao dar pesos muito altos a poucos casos (para compensar minorias sub-representadas), aumenta-se a influência de casos que, sendo poucos, podem se desviar do verdadeiro médio populacional por acaso. Assim, as estimativas ponderadas podem ter maior volatilidade (maior erro-padrão) que as não ponderadas. Mede-se isso através do efeito do desenho (design effect), especialmente o componente de efeito de peso. Se todos os pesos fossem iguais, o design effect devido a pesos seria 1; quando variam, o design effect > 1 indica inflação da variância. Por exemplo, se alguns pesos são 5 vezes maiores que outros, a variância das estimativas pode aumentar consideravelmente (o efeito de desenho pode elevar a margem de erro efetiva em, digamos, 20%). Em surveys, frequentemente se reporta o n efetivo (effective sample size), que é $n_{\text{eff}} = n/\text{deff}$, indicando o tamanho amostral equivalente após considerar a penalização do efeito de desenho (devido a pesos desiguais ou outro aspecto do plano amostral).

Além dos pesos, o desenho amostral em si (estratificação, conglomerados) afeta a variância das estimativas. Estratificação homogênea tende a reduzir variância (dado que dentro de cada estrato a variabilidade é menor, aumentando precisão se estimadores estratificados são combinados corretamente). Amostragem por conglomerados (clusters) tende a aumentar variância em relação à amostragem aleatória simples, pois indivíduos do mesmo cluster (ex.: moradores do mesmo bairro) costumam ser mais semelhantes entre si do que dois indivíduos aleatórios da população – ou seja, a informação coletada é um pouco redundante dentro do cluster. Isso é quantificado pelo *intra-cluster correlation* e reflete-se também no design effect.

Portanto, ao analisar os dados de um survey, é imprescindível usar técnicas de estimação de variância complexa, que considerem pesos e estrutura do desenho, em vez de aplicar fórmulas simples que assumem amostragem aleatória simples. Softwares de análise survey (como Stata, R `survey` package, SPSS Complex Samples) permitem especificar pesos, estratos e clusters, calculando erros-padrão ajustados por design. Por exemplo, para estimar o intervalo de confiança do IPA médio na população com nossos dados, deveríamos computá-lo respeitando os pesos (cada respondente contribuindo proporcionalmente) e

usando formulação de variância que incorpore possíveis clusters (se, por exemplo, a amostra foi feita por municípios). Caso contrário, poderíamos reportar intervalos muito estreitos (otimistas) se ignorássemos o efeito de desenho, ou mal posicionados se ignorássemos os pesos.

Em termos práticos, muitos relatórios apresentam tanto os percentuais brutos quanto os ponderados. A diferença entre eles indica se a ponderação alterou substancialmente os resultados – se sim, possivelmente a amostra tinha viés que foi corrigido. Por exemplo, imagine que sem peso o IPA médio fosse 5 pontos, mas com peso tornou-se 6; isso sugeriria que grupos muito polarizados estavam sub-representados originalmente e foram corrigidos via peso. Entretanto, se as diferenças forem grandes, isso também significa maior incerteza (pois depende de poucos indivíduos com peso alto, por exemplo).

O processo de ponderação e calibração deve ser transparente e documentado, incluindo quais variáveis de ajuste foram usadas e quais fontes populacionais serviram de referência. Biemer (2010) ressalta que ajustes feitos de maneira descuidada ou com bases imprecisas podem introduzir novos erros em vez de reduzir o TSE[18]. Um exemplo seria calibrar atitudes políticas a partir de pesquisas antigas não mais válidas – isso poderia enviesar resultados atuais. Assim, a máxima é: ponderar quando necessário, mas com critério. Quando a amostra já é de alta qualidade e balanceada, pode-se optar por mínima ponderação (p. ex., peso proporcional à probabilidade inversa de seleção, se conhecida, e só). Em contrapartida, quando há suspeita de vieses residuais, a ponderação é fundamental para alinhamento.

Em surveys on-line modernos, onde muitas vezes se usam painéis de respondentes não probabilísticos, a ponderação/calibração robusta é essencial para conseguir qualquer tipo de inferência para a população. Métodos como pós-estratificação por múltiplas variáveis e até ajustes por *propensity score* de participação são empregados para tentar aproximar a amostra voluntária de algo representativo. Esses métodos entram no erro de ajuste mencionado em TSE – um mal necessário, mas que deve ser bem conduzido para realmente melhorar a acurácia.

Por fim, cabe mencionar o conceito de precisão e testes estatísticos no contexto de surveys. Quando comparamos subgrupos (e.g., nível de polarização afetiva entre jovens vs idosos), devemos levar em conta o plano amostral para testar se diferenças são significativas. O uso de modelos (regressões, por exemplo) também requer *design-based standard errors*. Em amostras complexas, analistas muitas vezes utilizam reamostragem (bootstrap, Jackknife) para estimar variâncias de estimadores complexos (como medianas, ou índices não lineares como o % de respondentes “muito infelizes” com casamento interpartidário). Em suma, toda a inferência estatística no pós-coleta precisa estar alinhada às características do desenho e ponderação para evitar conclusões incorretas.

Resumindo esta seção: ponderação e calibração são ferramentas poderosas para melhorar representatividade de resultados de survey, corrigindo desequilíbrios da amostra obtida, mas vêm acompanhadas de aumento da variabilidade das estimativas. O pesquisador deve

aplicar essas técnicas cuidadosamente e analisar os dados usando métodos apropriados que reflitam o desenho amostral. Transparência nos procedimentos e comunicação franca da margem de incerteza (que pode ser maior do que a do simples acaso) são partes integrantes de uma boa prática em surveys.

Comparabilidade em Surveys Multiculturais: Tradução, Retrotradução e Invariância de Medida

Ao conduzir surveys em múltiplos países, línguas ou grupos culturais – ou mesmo ao comparar resultados de um survey atual com estudos anteriores –, surge o desafio da comparabilidade. Uma pesquisa só é comparável entre contextos distintos se as perguntas tiverem o mesmo significado e interpretação para todos os respondentes e medirem o mesmo construto subjacente com a mesma estrutura. Isso envolve duas grandes questões: (1) a tradução e adaptação linguística do questionário, e (2) a invariância de medida dos constructos medidos.

Tradução e Retrotradução: A tradução de questionários de survey não é um simples exercício literal, mas uma tarefa de equivalência conceitual. As melhores práticas recomendam o método TRAPD – Translation, Review, Adjudication, Pretesting, and Documentation – conforme as diretrizes de Harkness et al. (2010) para surveys multilíngues. Em suma, isso envolve: traduzir cada item por pelo menos dois tradutores independentes; depois realizar uma retradução (back-translation) da versão traduzida de volta ao idioma original por um terceiro tradutor que não viu o original, para checar se o conteúdo se mantém[24]. Equipes de revisão com linguistas e especialistas no tema então comparam as versões, discutem discrepâncias e refinam a formulação até atingir equivalência semântica e funcional. Back-translation serve como diagnóstico – se a retradução diverge muito do original, é sinal de que a tradução pode ter se desviado do significado pretendido.

Além da tradução literal, muitas vezes é necessária adaptação cultural. Certos termos ou exemplos podem não ter correspondência direta em outra cultura. Por exemplo, perguntar “Em uma escala de 0 a 100, qual o seu sentimento em relação ao Partido X?” pressupõe que o respondente entenda a ideia de usar uma escala termômetro. Em algumas culturas, escalas numéricas extensas podem confundir; pode ser preferível uma escala verbal de 0 a 10, ou de “muito negativo” a “muito positivo” com passos claros. O importante é que, independentemente do formato, a intenção da pergunta e a dimensão psicológica medida sejam as mesmas. Às vezes, nomes de pessoas ou grupos usados em exemplos precisam ser adaptados (ex.: um nome popular num país pode soar estranho noutro; em vez de “Fulano e Sicrano”, usar equivalentes locais). Essa adaptação deve ser feita sem alterar a dificuldade ou conotação da pergunta.

No caso de polarização afetiva, imagine estender um survey dos EUA para o Brasil: a ideia

de “Democrats vs Republicans” é substituída por “partidos ou grupos políticos relevantes no Brasil (por exemplo, lulistas vs bolsonaristas em 2022)”. Mas isso requer cautela – se no contexto original bipartidário a pergunta supunha apenas dois grupos, num contexto multipartidário teria que se adaptar possivelmente para “grupo político que você mais gosta vs. grupo que você menos gosta”. Adaptações assim devem ser deliberadas e testadas. Pré-testes locais são cruciais para garantir que os respondentes compreendam as perguntas como esperado.

Invariância de Medida: Além de traduzir, é necessário verificar se as escalas e conceitos medem a mesma coisa nos diferentes grupos. O conceito de invariância de medida (proveniente da psicometria) refere-se a termos técnicos como invariância configural, métrica e escalar. Em termos simples:

- *Invariância configural* significa que o construto tem a mesma estrutura de fatores nos grupos comparados.
- *Invariância métrica* implica que as unidades de medida são equivalentes – ou seja, uma diferença de 1 ponto na escala representa a mesma diferença de atitude em todos os grupos.
- *Invariância escalar* significa que os pontos de origem das escalas são equivalentes – isto é, na ausência do traço, os grupos responderiam de forma igual.

Na prática, usam-se análises fatoriais confirmatórias multi-grupo e testes de DIF para identificar itens que se comportam diferentemente entre grupos.

Pesquisas como WVS, ESS e outras investem na padronização de tradução e em testes de invariância. Se itens não forem invariantes, pode ser necessário excluí-los ou tratá-los separadamente.

No tocante à polarização afetiva, ao comparar IPA médio entre países, é crucial garantir equivalência de procedimento e de uso de escala. Diferenças culturais no uso de extremos, por exemplo, podem afetar médias. Técnicas como *vignette anchoring* às vezes ajudam a corrigir estilos de resposta.

Resumindo: traduzir cuidadosamente (com retrotradução e revisão), adaptar referências culturais sem perder o sentido, documentar alterações, pré-testar localmente e testar invariância são passos-chave para comparabilidade.

Ética e LGPD em Surveys On-line

Conduzir pesquisas de opinião envolve diretamente seres humanos como participantes, o que demanda cuidadosa consideração ética em todas as etapas. Princípios centrais incluem consentimento informado, privacidade e justiça.

No Brasil, essas obrigações alinham-se à LGPD (Lei 13.709/2018). A LGPD classifica dados sobre opinião política como dados pessoais sensíveis. Para pesquisa, a lei permite tratamento

com salvaguardas: idealmente, anonimização, uso exclusivo para fins acadêmicos, e sem repasse a terceiros[25][26]. Mesmo quando a base legal dispense consentimento formal, boas práticas recomendam **TCLE** claro e acessível. Em surveys on-line, o TCLE aparece na primeira página (explicando objetivos, procedimentos, confidencialidade e direito de desistir).

Medidas técnicas: plataformas seguras (HTTPS), armazenamento protegido, minimização de dados (coletar apenas o necessário), separação entre dados identificáveis e analíticos, retenção limitada e descarte/anonimização após a finalidade.

Conteúdo sensível pede linguagem neutra, evitando incitar ofensas. Questionários devem ser concisos para reduzir abandono. Pesquisas com seres humanos passam por CEP (Res. CNS 466/12 e 510/16). A ANPD pode supervisionar cumprimento da LGPD em caso de incidentes.

Em surveys on-line sobre polarização afetiva, anonimato completo é imprescindível, dado o risco de exposição de opiniões potencialmente estigmatizantes.

Boas Práticas de Transparência e Reprodutibilidade

Adoção de princípios de ciência aberta eleva a confiança e o valor cumulativo dos estudos:

- **Documentação completa** do questionário, plano amostral, taxas de resposta, modos, datas, ponderações (AAPOR Transparency Initiative)[27].
- **Dados anonimizados** em repositórios (Dataverse, Zenodo), quando possível, com revisão de risco de reidentificação (Salganik, 2019)[28].
- **Código de análise** (R, Stata etc.) para reprodutibilidade exata.
- **Pré-registro** de hipóteses e planos de análise (OSF), distinguindo o que é confirmatório do que é exploratório.
- **Relatos completos e claros**, com ICs, tamanhos amostrais válidos e discussão franca de limitações.

Checklist de Planejamento e Execução de um Survey de Qualidade

- Definir objetivos e conceitos (ex.: medir polarização afetiva entre eleitores).
- Delimitar população-alvo e critérios de elegibilidade.
- Selecionar método de amostragem e tamanho amostral; definir frame.
- Desenhar questionário alinhado à teoria; incluir itens para IPA/IDS.
- Pré-testar (cognitivo/piloto); ajustar linguagem e layout.
- Planejar modo(s) de coleta e treinamento; programar instrumento (web/telefone/presencial).
- Monitorar campo e qualidade; protocolos para não-resposta e verificação.

- Garantir ética/LGPD (TCLE, anonimização, minimização de dados).
- Ponderar e calibrar; inspecionar pesos e *design effect*.
- Estimar variância sob desenho; usar erros-padrão apropriados.
- Testar comparabilidade/invariância quando aplicável.
- Documentar e divulgar metodologia, dados (anonimizados) e código.

Síntese e Considerações Finais

A metodologia de survey, conforme explorada neste relatório, revela-se um campo rico e multidimensional, integrando conceitos estatísticos, comportamentais e éticos. Partindo da definição de Groves et al. (2009) de um survey moderno – um método sistemático de coletar informação de uma amostra para inferir sobre uma população[1] –, percorremos as etapas essenciais de concepção e implementação. Destacamos decisões de amostragem (“quem entrevistar”) e de mensuração (“o que perguntar”), estruturadas sob o TSE[29][9]. Modos de coleta afetam representação e mensuração; para temas sensíveis, autoaplicados tendem a reduzir desejabilidade social[19]. Ponderação e calibração corrigem vieses de representação, mas elevam variância; a análise deve respeitar pesos e desenho. Comparabilidade demanda tradução rigorosa e testes de invariância. Ética e LGPD exigem consentimento, anonimização e governança de dados. Transparência e reprodutibilidade consolidam a confiabilidade dos achados.

Usando a polarização afetiva como fio condutor, mostramos como índices simples (termômetros, distância social) tornam mensurável um construto complexo, desde que apoiados por desenho amostral e mensuração rigorosos. Em síntese: a qualidade do survey depende de escolhas sólidas sobre **quem entrevistar** e **o que perguntar**, e da gestão consciente do **Erro Total de Survey**.

Referências Bibliográficas

- ANPD (Autoridade Nacional de Proteção de Dados). *A LGPD e o tratamento de dados pessoais para fins acadêmicos e para a realização de estudos por órgão de pesquisa*. Texto para Discussão n° 1/2022. Brasília: ANPD, 2022. Disponível em: <https://www.gov.br/anpd/>.
- BERINSKY, Adam J. Measuring Public Opinion with Surveys. *Annual Review of Political Science*, v. 20, p. 309–329, 2017.
- BIEMER, Paul P. Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, v. 74, n. 5, p. 817–848, 2010.
- DILLMAN, Don A.; SMYTH, Jolene D.; CHRISTIAN, Leah M. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4. ed. Hoboken, NJ: John Wiley & Sons, 2014.

- GROVES, Robert M.; FOWLER Jr., Floyd J.; COUPER, Mick P.; LEPKOWSKI, James M.; SINGER, Eleanor; TOURANGEAU, Roger. *Survey Methodology*. 2. ed. Hoboken, NJ: John Wiley & Sons, 2009.
- SALGANIK, Matthew J. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press, 2019.
- TOURANGEAU, Roger; RIPS, Lance J.; RASINSKI, Kenneth. *The Psychology of Survey Response*. Cambridge: Cambridge University Press, 2000.
- TOURANGEAU, Roger; CONRAD, Frederick J.; COUPER, Mick P. *The Science of Web Surveys*. New York: Oxford University Press, 2013.