

CAPÍTULO 5

CONHECENDO OS SEUS DADOS: AVALIANDO MENSURAÇÃO E VARIAÇÕES

RESUMO:

Embora cientistas políticos se importem em descobrir relações causais entre conceitos, o que *realmente* examinam é a associação estatística entre variáveis. Portanto, é crítico que tenhamos um entendimento claro dos conceitos que nos importam e que os mensuremos de uma maneira válida e confiável. Neste capítulo, focamos duas tarefas críticas no processo de avaliação de teorias causais: mensuração e estatísticas descritivas. Conforme discutimos a importância da mensuração, usamos diversos exemplos da literatura de ciência política, por exemplo, o conceito de tolerância política. Sabemos que tolerância política e intolerância são uma coisa “real” – que existe em vários graus na mente e no coração das pessoas. Mas como as mensuramos? Quais são as implicações de uma má mensuração? Estatísticas e gráficos descritivos, que são o segundo foco deste capítulo, são o que parecem ser – ferramentas que descrevem variáveis. Essas ferramentas são valiosas porque podem sumarizar de maneira sucinta uma tremenda quantidade de informações. Neste capítulo, discutimos algumas das estatísticas e dos gráficos descritivos mais comumente utilizados, como devemos interpretá-los, como devemos utilizá-los e suas limitações.

Eu reconheço quando vejo. – Potter Stewart, juiz associado da Suprema Corte dos Estados Unidos, em uma tentativa de definir “obscenidade” em um parecer favorável no caso *Jacobellis vs. Ohio* (1964).

Estes vão até onze. – Nigel Tufnel (interpretado por Christopher Guest), ao descrever o botão de volume de seu amplificador no filme *Isto é Spinal Tap*.

5.1 CONHECENDO OS SEUS DADOS

Temos enfatizado o papel da teoria na ciência política. Isto é, que nos preocupamos com as relações causais entre conceitos que nos interessam como cientistas políticos. Neste ponto, esperamos que você esteja começando a desenvolver suas próprias teorias sobre política. Se essas teorias originais estiverem de acordo com as regras que estabelecemos no capítulo 1, elas serão causais, gerais e parcimoniosas. Podem ainda ser elegantes e inteligentes.

Mas, aqui, vale a pena parar e pensar sobre o que uma teoria realmente é e o que *não é*. Para nos ajudar nesse processo, dê uma olhada novamente na Figura 1.2. Uma teoria, como dissemos, é meramente uma conjectura sobre a possível relação causal entre dois ou mais conceitos. Como cientistas, devemos sempre resistir à tentação de ver nossas teorias como suportadas de alguma forma antes que tenhamos avaliado evidências do mundo real e até tenhamos feito tudo que podemos com as evidências empíricas para avaliar o quão bem nossa teoria se sai ao tentar superar os quatro obstáculos causais que identificamos no capítulo 3. Em outras palavras, não podemos avaliar uma teoria até termos passado por todos os processos descritos na Figura 1.2. A primeira parte deste capítulo lida com a operacionalização, ou o movimento das variáveis de um nível conceitual abstrato para um nível mensurável bastante real. Podemos conduzir testes de hipóteses e realizar avaliações razoáveis de nossas teorias somente após termos passado cuidadosamente por esse importante processo com todas as nossas variáveis.

Se nossas teorias são afirmações sobre relações *entre conceitos*, quando procuramos por evidências para testar nossas teorias, somos imediatamente confrontados com a realidade na qual, na verdade, não *observamos* esses conceitos. Muitos dos conceitos pelos quais nos interessamos na ciência política, como brevemente veremos, são inerentemente elusivos e francamente impossíveis de serem observados empiricamente de modo direto e, algumas vezes, incrivelmente difíceis de serem mensurados quantitativamente. Por essa razão, precisamos pensar muito cautelosamente sobre os dados que escolhemos para avaliar nossas teorias.

Até agora, temos visto muitos exemplos de dados, mas não temos discutido os processos de obtenção de dados e os colocados para trabalhar. Se pensarmos novamente na Figura 1.2, estamos agora no estágio no qual nos movemos do nível teórico-conceitual para o nível empírico-mensurado. Para cada conceito teórico, existem múltiplas estratégias de operacionalização ou mensuração. Como discutimos nos capítulos anteriores, uma das primeiras grandes decisões que alguém precisa tomar é se conduzirá um experimento ou alguma forma de teste observacional. Neste capítulo, assumimos que você tenha uma teoria e que conduzirá um teste observacional dela.

Um exercício útil, uma vez que você tenha desenvolvido uma teoria original, é desenhar uma versão da Figura 1.2 e pensar sobre qual seria a configuração ideal para o teste de sua teoria. Qual seria a melhor configuração, um desenho transversal ou de série temporal? Uma vez que você tenha respondido a essa pergunta e que tenha suas dimensões espacial e temporal ideais em mãos, qual seria a medida ideal para suas variáveis dependente e independente?

Tendo passado pelo exercício de pensar sobre os dados ideais, o primeiro instinto da maioria dos estudantes é coletar seus próprios dados e, talvez, até mesmo aplicar um *survey*¹. Na nossa experiência, novos pesquisadores quase sempre subestimam as dificuldades e os custos (em termos de dinheiro e tempo) de coletar seus próprios dados. Assim, recomendamos *fortemente* que você procure por dados que estejam disponíveis para serem utilizados em suas pesquisas.

Para um cientista político, um dos grandes impactos das transformações pelas quais o mundo em que vivemos passou nos últimos tempos é que existe uma quase inesgotável fonte de dados disponível em sítios e em outros lugares de fácil acesso². Um poucas palavras de aviso, no entanto: não é porque os dados são facilmente acessíveis na internet que eles são perfeitamente adequados para as necessidades específicas do seu teste de hipótese.

O que segue no restante deste capítulo é um conjunto de considerações que você deve ter em mente e que o ajudarão a determinar se um conjunto particular de dados que você tenha encontrado é apropriado para o seus propósitos e a conhecer os dados uma vez que você os tenha aberto em um programa estatístico. Começamos com todos os tópicos importantes sobre mensuração de variável. Descrevemos os problemas da mensuração e a importância de mensurar os conceitos em que estamos interessados do modo mais preciso possível. Durante esse processo, você aprenderá algumas habilidades cognitivas para avaliar a estratégia de mensuração de artigos de outros pesquisadores, assim como aprenderá a avaliar a utilidade das medidas que você está considerando usar no teste de suas hipóteses.

Começamos a seção com a mensuração nas ciências sociais em geral. Focamos exemplos da economia e da psicologia, duas ciências sociais que possuem níveis de agregação diferentes para suas principais variáveis – na ciência política, temos uma gama completa de variáveis em termos de como elas devem ser mensuradas. Discutimos alguns conceitos fundamentais da mensuração e apresentamos alguns exemplos de pesquisas em ciência política. Ao longo da discussão desses conceitos fundamentais, focamos a mensuração de variáveis que assumem intervalos numéricos e nos sentimos confortáveis de tratar do modo como normalmente tratamos números. Perto do final do capítulo, quando discutimos o básico sobre como conhecer seus dados utilizando um *software* estatístico, discutiremos isso mais profundamente e focaremos alguns tipos de variáveis que podem assumir diferentes tipos de valores não numéricos.

5.2 A MENSURAÇÃO NAS CIÊNCIAS SOCIAIS: OS VÁRIOS DESAFIOS DE QUANTIFICAR A HUMANIDADE

Mensuração é um “problema” em todas as ciências – das ciências físicas, como a física e a química, às ciências sociais como economia, ciência política, psicologia e

¹ Um *survey* é uma escolha particularmente complicada, porque, ao menos na maioria das universidades, você precisaria ter a aprovação do Comitê de Pesquisa com Seres Humanos.

² Um dos recursos que é frequentemente negligenciado é a biblioteca da sua faculdade. Mesmo que bibliotecas pareçam coisas do passado, a biblioteca da sua faculdade pode ter comprado acesso a fontes de dados, e os bibliotecários muitas vezes são especialistas na localização de dados na internet.

as demais. Mas, nas ciências físicas, o problema da mensuração é frequentemente reduzido ao problema da instrumentalização, no qual cientistas desenvolvem protocolos bastante específicos para mensurar, por exemplo, a quantidade de gás liberado em uma reação química ou a quantidade de luz emitida por uma estrela. As ciências sociais, pelo contrário, são ciências novas e o consenso científico sobre como mensurar importantes conceitos é raro. Talvez mais crucial, porém, é o fato de que as ciências sociais lidam com uma dificuldade inerente ao seu objeto em suas previsões: os seres humanos.

O problema da mensuração existe em todas as ciências sociais. Seria errado, porém, dizer que esse é um problema equânime em todas as disciplinas das ciências sociais. Algumas disciplinas prestam comparativamente menos atenção aos problemas da mensuração, enquanto outras estão envolvidas em constantes controvérsias sobre mensuração e suas dificuldades.

Considere o objeto de muitas pesquisas em economia: dólares (ou euros, ou ienes, ou a moeda que você tiver). Se o conceito de interesse é “produto econômico” (ou “produto interno bruto”), que é comumente definido como o total da soma de todos os bens e serviços produzidos pelo trabalho e pela propriedade em um dado período de tempo, então é relativamente fácil obter uma observação empírica que seja consistente com o conceito de interesse³. Tais medidas não serão controversas entre a maioria dos estudiosos. Contrariamente, uma vez que os economistas concordem sobre como mensurar o produto econômico, eles podem partir para um próximo (e mais interessante) passo do processo científico, a saber, descobrir quais forças *causam* maior ou menor crescimento do produto econômico. (É nesta parte que o acordo entre os economistas termina.)

Contudo, nem todo conceito em economia é mensurado de modo tão fácil. Muitos economistas estão preocupados com a pobreza: por que alguns indivíduos são pobres enquanto outros não? Quais são as forças que fazem a pobreza aumentar ou diminuir ao longo do tempo? A despeito do fato de sabermos que a pobreza é uma coisa real, mensurar quem é e quem não é pobre é um pouco complicado. O governo federal americano define o conceito de pobreza como “um conjunto de linhas de renda ajustado pelo tamanho do domicílio, a idade do chefe do domicílio e o número de crianças com idade abaixo de 18 anos”⁴. A intenção de definir linhas é a de descrever “níveis minimamente decentes de consumo”⁵. Porém, existem dificuldades em obter observações empíricas de pobreza. Entre elas, considere que a maioria das democracias ocidentais (incluindo os Estados Unidos) possui estados de bem-estar social que provêm transfe-

³ Para detalhes sobre como o governo federal dos EUA mede o PIB, ver: <<http://www.bea.gov>>. [Para informações do Brasil, ver: <<http://www.fazenda.gov.br/economia/pib>>. (N.T.)]

⁴ Ver: <www.census.gov/hhes/www/poverty.html>.

⁵ Observe um problema logo de partida: o que é “minimamente decente”? Você suspeita que o que era qualificado como “minimamente decente” em 1950 ou 1985 poderia ser considerado “minimamente decente” hoje? Isso imediatamente leva a problemas de quão sensível é a comparação de taxas de pobreza do passado com as de hoje. Se o que é considerado como o mínimo decente continua a subir, então a comparação é, no pior caso, no mínimo problemática e sem sentido.

rência de renda – em forma de pagamento em dinheiro, vales para trocar por comida ou serviços subsidiados, como saúde – para seus cidadãos que estão abaixo de uma determinada linha de renda. Tais programas, claro, são desenhados para minimizar ou eliminar os problemas que afligem os pobres. Quando economistas buscam mensurar o nível de renda de uma pessoa para definir se ela é ou não pobre, eles devem utilizar a definição “pré-transferência” de renda ou a renda familiar *após* receber alguma das transferências do governo (ou seja uma definição “pós-transferência”)? Qualquer uma das alternativas possui consequências negativas. Escolher a definição pré-transferência de renda nos dá uma medida de como o setor privado da economia está falhando. Por outro lado, uma definição pós-transferência nos dá uma medida de como os estados de bem-estar estão aquém do necessário e como as pessoas realmente vivem. Como a geração do *baby boom*⁶ nos Estados Unidos continua a envelhecer, mais e mais pessoas estão se aposentando. Utilizando uma medida de pobreza pré-transferência significa que os pesquisadores não considerarão os pagamentos da aposentadoria – de longe, a maior fonte de transferência nos EUA – e, portanto, o índice de pobreza (pré-transferência) deve ter um crescimento contínuo nas próximas décadas, a despeito do desempenho geral da economia. Isso pode não ser a representação mais acurada do que chamamos de “pobreza” (Danziger e Gottschalk, 1983).

Se, devido à natureza do objeto, economistas raramente (ou ocasionalmente) têm obstáculos de mensuração, no extremo oposto do espectro está a disciplina de psicologia. O objeto da psicologia – o comportamento humano, a cognição e a emoção – é repleto de conceitos que são extremamente difíceis de mensurar. Considere alguns exemplos. Todos sabem que o conceito de “depressão” é uma coisa real; alguns indivíduos são depressivos e outros não. Alguns indivíduos que são depressivos hoje não serão depressivos com o passar do tempo, e alguns que não são depressivos hoje se tornarão depois. No entanto, como é possível avaliar cientificamente se uma pessoa é ou não depressiva?⁷ Por que importa se medimos a depressão acuradamente? Lembre-se dos riscos científicos descritos no início deste capítulo: se não mensurarmos a depressão corretamente, como podemos saber se tratamentos como a terapia clínica ou a utilização de medicamentos antidepressivos são efetivos?⁸ A psicologia lida com uma variedade de outros conceitos que são notoriamente escorregadios, como o foco clínico sobre “ansiedade” ou o foco sociopsicológico sobre conceitos como “estereótipos” ou “preconceito”, que também são preocupações de cientistas políticos.

A ciência política, na nossa visão, está em algum lugar no meio entre os extremos da economia e da psicologia em termos de quão frequentemente encontramos pro-

⁶ Geração nascida logo após o término da Segunda Guerra Mundial. [N.T.]

⁷ Desde 1952, a American Psychiatric Press tem publicado a obra *Diagnostic and Statistical Manual of Mental Disorders*, atualmente em sua quinta edição (chamada DSM 5), que define o diagnóstico de depressão focando quatro conjuntos de sintomas que indicam depressão: humor, síndromes comportamentais (como o afastamento), sintomas cognitivos (como a inabilidade de se concentrar) e sintomas somáticos (como a insônia).

⁸ De fato, a efetividade da “conversa” terapêutica é uma questão de alguma disputa entre psicólogos. Ver *Married with Problems? Therapy May Not Help*. *New York Times*, 19 abr. 2005.

blemas sérios de mensuração. Alguns subcampos da ciência política operam relativamente livres de problemas desse tipo. O estudo da economia política – que examina a relação entre economia e forças políticas, como política governamental, eleições e confiança do consumidor – é bem semelhante à economia, por razões óbvias. Outros subcampos encontram problemas de mensuração regularmente. O subcampo da psicologia política – que estuda o modo como cidadãos individuais interagem com o mundo político – compartilha muitos dos mesmos objetos da psicologia social e, por focar atitudes e sentimentos das pessoas, guarda muita semelhança com os problemas de mensuração da psicologia social.

Considere a seguinte lista de conceitos importantes na disciplina de ciência política e que possuem problemas de mensuração:

- *Ativismo judicial*: nos Estados Unidos, o papel do judiciário no processo de *policy-making* tem sido controverso. Alguns veem as cortes federais como protetoras de importantes liberdades civis, enquanto outros as veem como ameaças à democracia, em razão de os juízes não serem eleitos. Como é possível identificar um “juiz ativista” ou uma “decisão ativista”?
- *Votações motivadas pela ideologia no Congresso*: em cada uma das sucessivas seções do Congresso dos Estados Unidos, comentaristas frequentemente comparam o nível de liberalismo e conservadorismo do atual Congresso com os seus antecessores. Como podemos saber se o Congresso está se tornando mais ou menos liberal ao longo do tempo (Poole e Rosenthal, 1997)?
- *Legitimidade política*: como os analistas podem distinguir um governo “legítimo” de um “ilegítimo”? Um problema conceitual fundamental aqui é “como cidadãos avaliam a autoridade governamental” (Weatherford, 1992). A legitimidade pode ser objetivamente determinada ou é uma propriedade subjetiva inerentemente dos cidadãos? Alguns a veem positivamente, enquanto outros negativamente.
- *Sofisticação política*: alguns cidadãos conhecem mais sobre política e estão mais bem preparados para processar informações políticas do que outros cidadãos que conhecem pouco e se importam pouco com assuntos políticos. Como podemos distinguir cidadãos politicamente sofisticados dos não sofisticados? Ademais, como podemos dizer se o nível de sofisticação política de uma sociedade está aumentando ou diminuindo ao longo do tempo (Luskin, 1987)?
- *Capital social*: algumas sociedades são caracterizadas pelo alto nível relativo de interconectividade, com redes densas de relacionamentos que tornam a população coesa. Outras sociedades, em contraste, são caracterizadas pelo alto nível de isolamento e desconfiança. Como podemos mensurar o que cientistas sociais chamam de *capital social* de modo que nos possibilite comparar o nível de

⁹ Neste caso em particular, poderia ocorrer até um desacordo sobre a definição conceitual de “ativista”. O que um conservador ou um liberal consideraria como “ativista” poderia não gerar acordo algum. Para uma visão jornalística dessa questão, ver: *Activist, Schmactivist*. *New York Times*, 15 ago. 2004.

conectividade de uma sociedade com outra ou como o nível de conectividade varia em diferentes pontos do tempo (Putnam, 2000)?

Nas seções 5.4 e 5.5, descrevemos as controvérsias de mensuração que cercam outros dois importantes conceitos para a ciência política – democracia e tolerância política. Mas antes, na próxima seção, descrevemos alguns pontos fundamentais com os quais cientistas políticos precisam lidar quando mensuram seus conceitos de interesse.

5.3 PROBLEMAS NA MENSURAÇÃO DOS CONCEITOS DE INTERESSE

Podemos sumarizar os problemas da mensuração dos conceitos de interesse na preparação para o teste de hipótese da seguinte maneira: primeiramente, você precisa ter certeza de que tem clareza conceitual. Posteriormente, você precisa decidir um nível razoável de mensuração. Finalmente, assegure-se de que sua medida é válida e confiável. Após repetir esse processo com cada uma das variáveis de sua teoria, você estará pronto para testar suas hipóteses.

Infelizmente, não existe um mapa claro para seguir durante essas etapas com nossas variáveis. Algumas variáveis são bastante fáceis de serem mensuradas, enquanto outras, em razão da natureza do que estamos tentando mensurar, são mais elusivas. Como veremos, debates sobre questões de mensuração estão no centro de muitos dos interessantes campos de estudo da ciência política.

5.3.1 CLAREZA CONCEITUAL

O primeiro passo na mensuração de qualquer fenômeno de interesse de cientistas políticos é ter clareza do conceito que estamos tentando mensurar. Em alguns casos, como os discutidos a seguir, isso é uma tarefa extremamente reveladora e difícil. É preciso considerável disciplina de pensamento para definir precisamente os conceitos sobre os quais estamos teorizando. Até mesmo em alguns exemplos aparentemente fáceis isso é mais difícil do que pode parecer em um primeiro momento.

Considere um *survey* em que precisamos mensurar a *renda* de uma pessoa. Isso parece bastante simples. Uma vez que tenhamos a nossa amostra de adultos, por que não apenas perguntar aos entrevistados “Qual é a sua renda?” e oferecer uma escala de valores na qual os entrevistados possam se alocar? Qual poderia ser o problema com tal medida? Imagine um estudante universitário de 19 anos de idade cujos pais sejam muito ricos, mas que nunca tenha trabalhado, respondendo a tal pergunta. Qual é a renda dessa pessoa no último ano? Zero. Em tal circunstância, essa é a resposta verdadeira para a pergunta. Mas não é uma medida particularmente válida da renda. Provavelmente queremos uma medida de renda que reflita o fato de que os pais do estudante lhe dão uma boa quantidade de dinheiro que o ajuda a manter o luxo de

não trabalhar enquanto estuda, como ocorre com muitos estudantes. Essa medida deveria colocar uma filha (ou filho) de pais ricos à frente de um estudante relativamente pobre, responsável por seu sustento e que trabalha quarenta horas por semana para pagar por seus estudos. Portanto, podemos reconsiderar nossa aparentemente simples pergunta e questionar: “Qual é a quantidade total de renda recebida por você e pelos outros adultos no seu domicílio no último ano contábil, incluindo todas as fontes de renda?”. Essa medida coloca um filho que não trabalha de uma família rica mais bem posicionado do que um estudante de uma família menos rica. Essa é uma medida de “renda” que consideramos teoricamente mais útil para a maioria dos propósitos nas ciências sociais¹⁰.

Neste momento, vale destacar que a *melhor* medida de renda – assim como de outros conceitos – depende de quais são nossos objetivos teóricos. A melhor medida de algo tão simples como a renda de um entrevistado depende do que pretendemos relacionar a essa medida em nosso teste de hipótese.

5.3.2 CONFIABILIDADE

Uma medida operacionalizada de um conceito é considerada confiável na medida em que é replicável e consistente; isto é, quando a aplicação das mesmas regras de mensuração para os mesmos casos ou observações produz resultados idênticos. Uma medida não confiável, pelo contrário, produz resultados inconsistentes para as mesmas observações. Por razões óbvias, todos os cientistas querem que suas medidas sejam confiáveis.

Talvez o exemplo mais simples para ajudar você a entender isso seja sua balança de banheiro. Digamos que você suba na balança uma manhã e que a balança mostre que seu peso é de 68 quilos. Você desce da balança e ela retorna para o zero. Mas você *não* confia na leitura da balança e pensa consigo mesmo: “Talvez se eu subir novamente na balança, ela possa me dar um número melhor”. Esse é um teste de **confiabilidade**. Se você (imediatamente) subir na balança e ela mostrar agora 66 quilos, sua balança *não* é confiável, porque a mensuração do mesmo caso – seu corpo – produz resultados diferentes.

Levando o exemplo da sua balança de banheiro ao extremo, não podemos confundir variabilidade ao longo do tempo com falta de confiabilidade. Se você acordar uma semana depois e pesar 71 em vez de 68 não significa necessariamente que sua balança não é confiável (embora isso possa ser verdade). Talvez você tenha substituído a salada do almoço por batatas fritas nessa semana, e talvez você tenha se exercitado menos vigorosamente ou menos frequentemente.

Confiabilidade é frequentemente uma questão importante quando estudiosos precisam codificar eventos ou textos para análise quantitativa. Por exemplo, se um pesquisador estivesse tentando codificar textos de cobertura da mídia favoráveis ou

¹⁰ Os mesmos problemas podem surgir quando tentamos mensurar a renda de pessoas aposentados que não têm mais participação na força de trabalho.

contrários a um candidato, ele desenvolveria algumas regras específicas de codificação para aplicar ao texto – com efeito, para contar certas referências como “pró” ou “contra” o candidato. Suponha que, para a codificação, o pesquisador empregue um grupo de estudantes para codificar o texto – uma prática bastante comum na pesquisa sobre política. Um conjunto *confiável* de regras de codificação implicaria que, quando um dos estudantes o aplicasse ao texto, os resultados fossem os mesmos que os obtidos por outro estudante que os tenha aplicado no mesmo texto. Um conjunto *não confiável* de regras implicaria o oposto, nominalmente, que, quando dois codificadores diferentes tentassem aplicar as mesmas regras à mesma notícia, eles chegariam a conclusões diferentes¹¹. As mesmas questões surgem quando se codificam coisas como eventos utilizando a cobertura dos jornais¹².

5.3.3 VIÉS DE MENSURAÇÃO E CONFIABILIDADE

Uma das preocupações que acompanham qualquer técnica de mensuração é o **viés de mensuração**, que consiste na subestimação ou na sobre-estimação de modo sistemático dos valores para uma variável. Embora o viés de mensuração seja um problema sério para qualquer um que queira saber o “verdadeiro” valor das variáveis para um caso particular, ele é um problema bem menos sério do que você pode pensar quando o propósito é o teste de teorias. Para melhor entender isso, imagine que você tenha que escolher entre duas diferentes operacionalizações para a mesma variável. A operacionalização *A* é enviesada, mas confiável, e a operacionalização *B* não possui viés, mas não é confiável. Para os propósitos do teste de teoria, preferimos a enviesada mas confiável operacionalização *A*!

Você estará mais bem preparado para entender o porquê dessa escolha uma vez que tenha entendido como o teste estatístico de hipótese funciona no capítulo 7 e nos seguintes. Por agora, porém, tenha em mente que, quando testamos nossas teorias, estamos buscando padrões gerais entre duas variáveis. Por exemplo, com valores *altos* de *X* tendemos a observar valores *altos* de *Y*, ou com valores *altos* de *X* tendemos observar valores *baixos* de *Y*? Se a mensuração de *X* é enviesada para mais, o mesmo padrão geral será visível na associação com *Y*. Mas, se a mensuração de *X* é não confiável, a relação subjacente entre *X* e *Y* se torna obscura.

5.3.4 VALIDADE

A característica mais importante de uma medida é sua validade. Uma medida válida representa acuradamente o conceito que supostamente mede, enquanto uma medi-

¹¹ Claro, é possível que um *protocolo de codificação* seja perfeitamente confiável, mas os *codificadores em si* não o sejam.

¹² Existe uma variedade de ferramentas para testar a confiabilidade, muitas das quais estão além do escopo dessa discussão.

da inválida mede algo diferente do que originalmente pretendido. Parece que estamos andando em círculos, nós sabemos.

Talvez seja útil pensar sobre alguns importantes conceitos que são exemplos espinhosos de mensuração nas ciências sociais para entender o papel da validade. Tanto na psicologia como na ciência política, o estudo do *preconceito* tem sido particularmente importante. Entre indivíduos, o nível de preconceito pode variar de quase nenhum até altos níveis. Mensurar o preconceito é importante na psicologia social para que possamos tentar determinar quais fatores fazem com que algumas pessoas sejam preconceituosas e outras não. Na ciência política, em particular, frequentemente estamos interessados nas consequências atitudinais e comportamentais do preconceito. Dado que não existe um sorô da verdade, como podemos obter uma medida quantitativa de preconceito capaz de nos dizer quem possui uma grande quantidade, quem possui algum e quem não possui preconceitos? Seria bastante fácil perguntar às pessoas se elas são preconceituosas ou não. Por exemplo, poderíamos perguntar: “Com respeito às pessoas de raças ou etnia diferente da sua, você diria que é extremamente, um pouco ou nem um pouco preconceituoso?”. Mas, com isso, teríamos razões óbvias para duvidar da **validade** das respostas – ou seja, de que essa medida reflete acuradamente o nível real de preconceito.

Existem vários modos para verificar a validade de uma medida, embora seja crítico notar que todas elas são teóricas e sujeitas a altos graus de desacordo. Infelizmente, não existe uma fórmula simples para checar a validade de uma mensuração em uma escala de zero a cem. De fato, dependemos de uma série de meios sobrepostos para determinar a validade da medida. Primeiro, e mais simples, podemos examinar a **validade de face**. Quando examinamos uma estratégia de mensuração, podemos primeiro perguntar se, em sua superfície (ou face), a medida parece ou não mensurar o que se propõe a mensurar. Essa é a validade de face. Segundo, e um pouco mais avançado, podemos escrutinar a **validade de conteúdo** de uma medida. Qual é o conceito a ser medido? Quais são todos os elementos essenciais desse conceito e as suas características que o definem? Você excluiu todas as outras coisas que não pertencem a ele? Por exemplo, o conceito de democracia certamente contém o elemento “eleições”, mas também deve incorporar mais do que meras eleições, porque estas ocorrem em lugares como a Coreia do Norte, que sabemos que não é democrática. O que mais deve conter uma medida válida de democracia? Basicamente, validade de conteúdo é um processo rigoroso que força o pesquisador a ter uma lista de todos os elementos críticos que definem o conceito que desejamos mensurar. Finalmente, podemos examinar a **validade de constructo** de uma medida: o grau em que é relacionada a outras medidas com as quais a teoria requer que ela esteja relacionada. Isto é, se temos uma teoria que conecta democratização e desenvolvimento econômico, então uma medida de democracia que é relacionada a uma medida de desenvolvimento econômico (como nossa teoria propõe) serve, simultaneamente, para confirmar a teoria e para validar nossa medida de democracia. Claro, uma dificuldade dessa abordagem acontece quando a associação esperada não se faz presente. Isso acontece por que nossa medida de democracia é inválida ou por que a teoria está equivocada? Não existe uma resposta conclusiva para essa pergunta.

5.3.5 A RELAÇÃO ENTRE VALIDADE E CONFIABILIDADE

Qual é a conexão entre validade e confiabilidade? É possível ter uma medida válida, mas não confiável? E é possível ter uma medida confiável mas inválida? Com respeito à segunda pergunta, existe algum debate científico; alguns acreditam que é possível ter uma medida confiável, mas inválida. Na nossa visão, isso é possível em termos abstratos. Mas como estamos interessados em mensurar conceitos com o interesse de avaliar teorias causais, acreditamos que, em termos práticos, qualquer medida concebível que seja confiável mas inválida não será útil na avaliação de teorias causais.

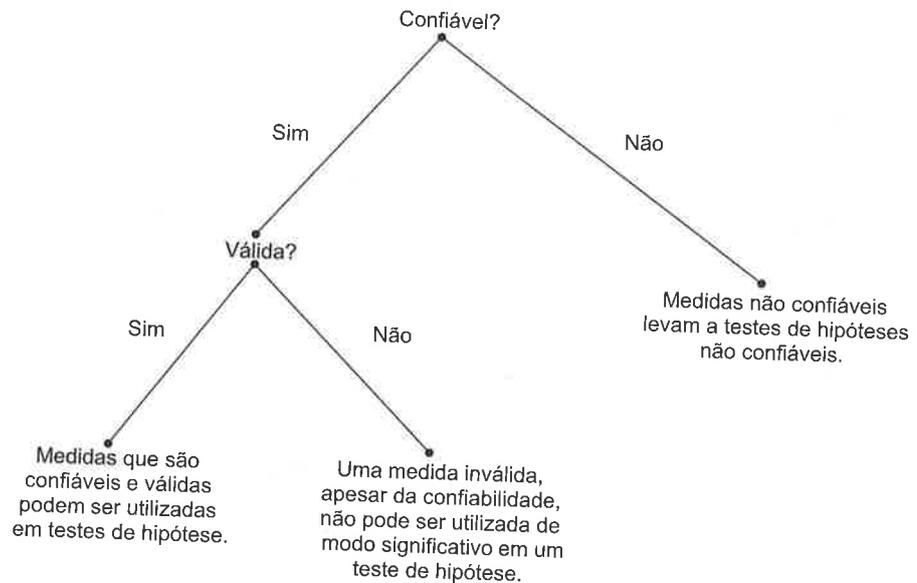


Figura 5.1 – Confiabilidade, validade e teste de hipóteses.

Similarmente, é teoricamente possível ter medidas válidas mas não confiáveis. Mas essas medidas também serão problemáticas para a avaliação de teorias causais, porque não teremos confiança nos testes de hipótese que conduzimos. Na Figura 5.1 apresentamos a relação entre validade e confiabilidade; podemos ver que, se uma medida não é confiável, há pouco sentido em avaliar sua validade. Uma vez que tenhamos estabelecido que uma medida é confiável, podemos avaliar sua validade, e somente medidas válidas e confiáveis são úteis para a avaliação de teorias causais.

5.4 CONTROVÉRSIA 1: MENSURANDO DEMOCRACIA

Embora tenhamos a tendência de pensar que democracia é similar à gravidez – isto é, um país é ou *não é* democrático da mesma forma que uma mulher *está* ou *não está* grávida –, refletindo um pouco mais sobre o assunto, você provavelmente pensará que

a democracia é mais bem entendida como um *contínuo*¹³. Isto é, podem existir vários graus em que um governo é democrático. Adicionalmente, dentro de democracias, alguns países são mais democráticos que outros, e um país pode se tornar mais ou menos democrático com o passar do tempo.

Mas definir um contínuo que varia de democrático a totalitário não é nada fácil. Podemos ser tentados a recorrer à definição de Potter Stewart: “Eu reconheço quando vejo”. Claro que, como cientistas políticos, esta não é uma opção. Temos que começar nos perguntando: o que queremos dizer com democracia? Quais são os elementos centrais que fazem um governo ser mais ou menos democrático? O filósofo político Robert Dahl (1971) persuasivamente argumentou que existem dois conceitos fundamentais atribuídos a democracia: “contestação” e “participação”. Isto é, de acordo com Dahl, democracias têm eleições competitivas para a escolha de líderes e amplas e inclusivas regras para participação.

Diversos grupos de cientistas políticos têm tentado mensurar democracia sistematicamente nas últimas décadas¹⁴. A mais conhecida – embora não signifique que seja universalmente aceita – dessas medidas é a do projeto Polity IV¹⁵. O projeto mensura democracia por meios de *scores* anuais que variam de -10 (fortemente autocrático) a +10 (fortemente democrático) para cada um dos países do mundo de 1800 a 2004¹⁶. Na operacionalização desses pesquisadores, a democracia possui quatro componentes:

1. Regras para o recrutamento do Executivo.
2. Competividade do recrutamento do Executivo.
3. Abertura do recrutamento do Executivo.
4. Limites ao chefe do Executivo.

Para cada uma dessas dimensões, especialistas atribuem aos países uma posição em uma escala. Por exemplo, para o primeiro critério, “regras para o recrutamento do Executivo”, os valores possíveis são os seguintes:

- +3 = Competição regular entre grupos reconhecidos.
- +2 = Competição de transição.
- +1 = Competição entre facções ou restrita.
- 0 = Sem competição.

¹³ Essa posição, porém, é controversa dentro da ciência política. Para uma interessante discussão sobre se pesquisadores devem mensurar democracia como um conceito binário ou contínuo, ver Elkins (2000).

¹⁴ Para uma revisão e comparação útil dessas várias medidas, ver Munck e Verkuilen (2002).

¹⁵ O *site* do projeto, no qual é possível ter acesso a uma vasta quantidade de especificidades dos países ao longo do tempo, é: <<http://www.cidcm.umd.edu/inscr/polity>>.

¹⁶ Eles apresentam os *scores* em duas escalas separadas de 10 pontos, uma para democracias e outra para autocracias. O *score* do Polity para um dado país é dado pelo valor do *score* da escala democrática menos o *score* da escala autocrática; assim, um país que recebesse 10 na escala democrática e 0 na escala autocrática teria um *score* de 10 na escala Polity para um dado ano.

Países que possuem eleições regulares entre grupos que são mais do que rivais étnicos receberão *scores* altos. Por procedimentos similares, estudiosos associados ao projeto definem *scores* para as outras dimensões que compõem a escala de democracia.

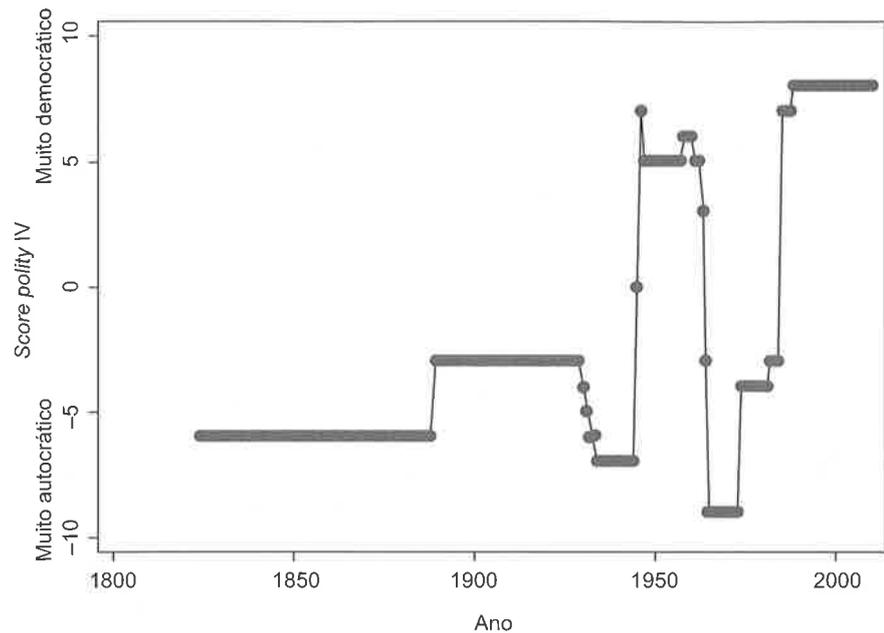


Figura 5.2 – Score do Polity IV para o Brasil.

A Figura 5.2 apresenta o *score* de Polity para o Brasil de 1824 até 2010¹⁷. Lembre-se que *scores* mais altos significam que, nesse ponto, o Brasil era mais democrático e que *scores* mais baixos significam que o Brasil era mais autocrático. Como você pode observar, existe uma enorme **variação** na experiência democrática brasileira desde sua declaração de independência de Portugal em 1822. Se fizermos uma comparação grosseira desses *scores* com a história política brasileira, podemos ter uma avaliação da validade de face do Polity como medida de democracia. Após declarar independência de Portugal, o Brasil foi uma monarquia constitucional comandada por um imperador. Após um golpe em 1889, o Brasil se tornou uma república, mas bastante controlada pelas elites de dois estados dominantes. Como podemos observar, essa mudança de regime resultou em um movimento do *score* de Polity de -6 para -3. A partir de 1930, o Brasil passou por uma série de golpes e contragolpes. Estudiosos desse período (como Skidmore, 2009) geralmente concordam que o governo nacional se tornou cada vez mais autocrático durante o período. O *score* de Polity reflete esse movimento. Em 1945, após outro golpe militar, um governo relativamente democrático foi constituído. Esse regime durou até meados da década de 1960, quando outro período de

¹⁷ Fonte: <<http://www.systemicpeace.org/inscr/inscr.htm>>.

instabilidade teve fim com o estabelecimento de uma ditadura militar. Esse período é amplamente reconhecido como o de maior repressão política na história do Brasil independente. Em 1974, notamos uma pequena alteração no *score*. Nesse ano, o governo militar permitiu o uso da televisão durante as eleições e outras atividades de campanha que foram aproveitadas pela oposição para conseguir um resultado expressivo. A eleição de um presidente civil em 1985 marca o começo do atual período democrático brasileiro. Cada um desses grandes movimentos da história política brasileira é refletido no *score* de Polity. Então, por meio dessa avaliação geral, podemos dizer que o *score* de Polity possui validade de face.

A medida de Polity é rica em detalhes históricos, como fica óbvio a partir da Figura 5.2. As regras de codificação são transparentes e claras e a quantidade de informação bruta para compor o *score* de um país para qualquer ano é impressionante. Contudo, ainda é justo criticar o Polity por incluir apenas parte da definição de democracia de Dahl. A medida de Polity contém valiosas informações sobre o que Dahl chama de “contestação” – se um país tem uma contestação ampla para decidir sua liderança. Mas a medida é muito menos rica quando o objetivo é aferir o nível do país no que Dahl chamou de “participação” – o grau em que os cidadãos participam nos processos e atividades políticas. Isso pode ser compreensível, em parte, por causa do impressionante escopo temporal do estudo. Afinal, em 1800 (ano em que a série do Polity começa), poucos países tinham uma participação eleitoral ampla. Desde o final da Segunda Guerra Mundial, a participação democrática ampla tem se espalhado rapidamente pelo globo. Mas, se o mundo tem se tornado um lugar mais democrático, nossas medidas de democracia devem incorporar a realidade. Como a medida de Polity inclui uma parte (“contestação”) do que significa, conceitualmente, ser democrático, mas ignora a outra parte (“participação”), podemos dizer que falta validade de conteúdo à medida. A medida do Polity IV, apesar de suas consideráveis vantagens, não contempla totalmente o que significa, conceitualmente, ser mais ou menos democrático.

Esse problema é bem ilustrado a partir do exame do *score* de Polity para os Estados Unidos. A Figura 5.3 apresenta o *score* de Polity para o período de 1800-2010. O consistente *score* 10 para quase todos os anos após a fundação da república – exceto durante a Guerra Civil, quando o presidente Lincoln suspendeu o direito de *habeas corpus* – ocultar o fato de que os EUA, em muitos aspectos importantes, têm se tornado um país mais democrático ao longo de sua história, particularmente na dimensão não capturada pela medida de Polity. Mesmo considerando algo bastante básico para a participação democrática, como o direito de voto, observamos que os Estados Unidos foram se tornando mais democráticos ao longo de sua história. A escravidão limitava a participação de afro-americanos de muitas formas, incluindo votar, e as leis Jim Crow mantiveram essas proibições nos estados do Sul por quase um século após o final da Guerra Civil. As mulheres também não teriam permissão para votar até que a 19ª emenda à Constituição fosse ratificada em 1920. Seria difícil argumentar que essas mudanças não tornaram os Estados Unidos mais democráticos, mas obviamente elas não estão refletidas na Figura 5.3. Isso não equivale a dizer que a medida de Polity é inútil, mas que lhe falta validade de conteúdo, porque um dos componentes fundamentais da democracia – participação – não está presente nela.

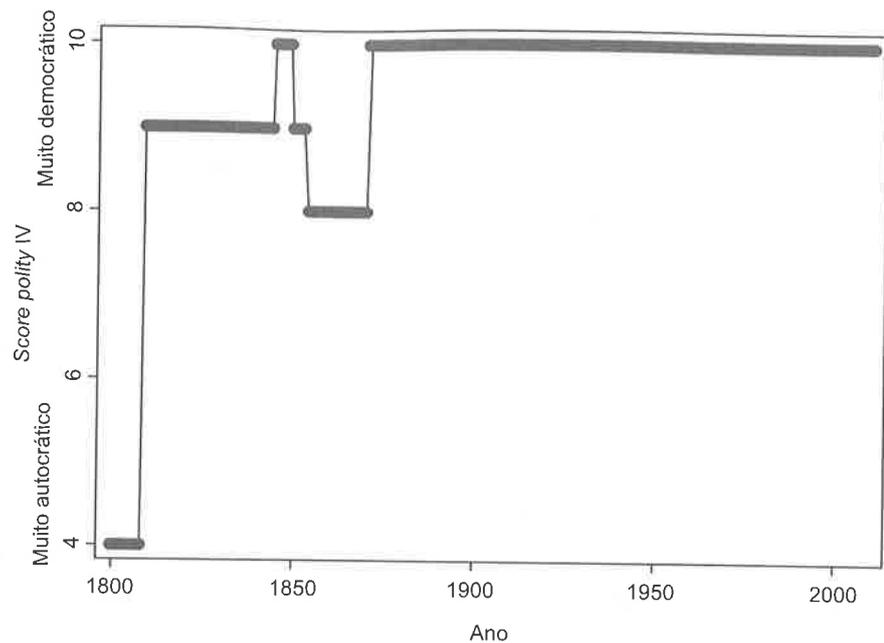


Figura 5.3 – Score do Polity IV para os Estados Unidos.

5.5 CONTROVÉRSIA 2: MENSURANDO TOLERÂNCIA POLÍTICA

Sabemos que existe um contínuo em que, de um lado, estão indivíduos extremamente “tolerantes” e, de outro, estão indivíduos extremamente “intolerantes”. Em outras palavras, tolerância política e intolerância, no nível conceitual, são coisas reais. Alguns indivíduos são mais tolerantes e outros são menos. É fácil imaginar por que cientistas políticos se interessam por tolerância política e intolerância. Existem fatores sistemáticos que causam que algumas pessoas sejam tolerantes e outras intolerantes?

Mensurar tolerância política, por outro lado, está longe de ser fácil. Tolerância não é como colesterol, que depende de um simples teste de sangue para sabermos o quanto é bom e o quanto é ruim. A abordagem inocente para mensurar tolerância política – conduzir um *survey* e perguntar diretamente para as pessoas: “Você é tolerante ou intolerante?” – parece boba logo de partida. Qualquer pergunta de *survey* certamente produziria altas taxas de “tolerância”, porque presumidamente poucas pessoas – mesmo pessoas intolerantes – se acham intolerantes. Mesmo no caso de pessoas que são conscientes da sua intolerância, é improvável que admitam esse fato a um entrevistador. Dada essa situação, como cientistas políticos lidam com esse problema?

Durante a década de 1950, quando o crescimento do socialismo soviético representava a maior ameaça à América, Samuel Stouffer (1955) conduziu uma série de *surveys* de opinião para mensurar como pessoas reagiam à “ameaça vermelha”. Ele perguntou

a amostras nacionais nos Estados Unidos se elas estariam dispostas a estender certas garantias civis – como poder lecionar em uma escola pública, garantir que os telefonemas não fossem grampeados, e outras – a certos grupos impopulares como comunistas, socialistas e ateus. Ele verificou que uma grande parte das pessoas, a partir dessas medidas, eram intolerantes; elas não estavam dispostas a garantir essas liberdades civis a membros desses grupos. A quantidade precisa de intolerância variava, dependendo do grupo-alvo e da atividade mencionada nos cenários, mas a intolerância era substancial – ao menos 70% dos entrevistados deram respostas intolerantes. Stouffer encontrou que o melhor preditor do nível de intolerância de um indivíduo era o nível de educação formal dele ou dela; pessoas com maior grau educacional emergiam como mais tolerantes, e pessoas com menor nível educacional, como menos tolerantes. Na década de 1970, quando a ameaça vermelha tinha diminuído um pouco, um novo grupo de pesquisadores fez as mesmas perguntas a uma nova amostra de americanos. Eles encontraram que os níveis de intolerância tinham diminuído consideravelmente no passar dos vinte anos – de um cenário em que a intolerância excedia 60% para a maioria das questões perguntadas, para um de 50% –, levando alguns a especular que a intolerância política minguara.

Contudo, no final da década de 1970, um grupo diferente de pesquisadores liderado pelo cientista político John Sullivan questionou a validade das medidas de Stouffer e, portanto, suas conclusões. O conceito de tolerância política, escreveram Sullivan, Pierson e Marcus (1979), “pressupõe oposição”. Isto é, a não ser que o respondente do *survey* ativamente se oponha a comunistas, socialistas e ateus, a questão de intolerância ou tolerância simplesmente não surge. Por exemplo, considere perguntar tais questões sobre os ateus. Um ateu que concorda que se permita a ateus ensinar em escolas públicas é politicamente tolerante? Sullivan e seus colegas pensam que não.

Os autores propuseram um novo conjunto de questões para *surveys* que eram, em sua visão, mais consistentes com um entendimento conceitual de tolerância. Se, como eles definiram, tolerância pressupõe oposição, então pesquisadores precisam *descobrir* a quem o respondente do *survey* se opõe; *presumir* que os respondentes se opõem a um grupo em particular não é uma boa ideia. Eles identificaram uma variedade de grupos politicamente ativos no momento – incluindo grupos racistas, pró e antiaborto e mesmo o Exército Simbionês de Libertação – e perguntaram aos entrevistados de qual deles eles mais desgostavam. Fizeram essa pergunta juntamente com outras muito semelhantes às de Souffer, mas somente para os grupos dos quais os próprios respondentes disseram não gostar, em vez dos que Stouffer tinha listado outrora.

Entre outros achados, dois chamaram atenção. Primeiro, os níveis de intolerância foram bastante altos: 66% dos americanos estavam dispostos a proibir que membros dos grupos que eles mais desgostavam se reunissem, e 71% estavam dispostos a ter um governo que banisse totalmente esse grupo. Segundo, sob essa nova conceituação e mensuração de tolerância, os autores encontraram que a percepção individual da natureza da ameaça do grupo-alvo, e não o nível de educação, era o preditor primário da intolerância. Em outras palavras, indivíduos que consideram seu grupo-alvo como particularmente ameaçador foram os mais intolerantes, enquanto os que achavam o

grupo pouco ameaçador foram mais tolerantes. A conclusão dos autores, portanto, foi que a educação não tinha nenhum efeito direto na tolerância. Nesse sentido, mensurar um importante conceito de um modo diferente produziu achados substantivos bastante diferentes sobre causas e efeitos¹⁸.

É importante que você observe a conexão com a validade da mensuração. Sullivan e seus colegas argumentaram que as perguntas do *survey* de Stouffer não eram medidas válidas para tolerância porque não eram acuradas para capturar o que significa, em termos abstratos, ser intolerante (especificamente, faltava o componente da oposição). A criação de medidas de tolerância e intolerância que melhor espelham o conceito de interesse produziu achados significativamente diferentes sobre a persistência da intolerância, assim como sobre os fatores que fazem um indivíduo ser tolerante ou intolerante.

5.6 AS MENSURAÇÕES RUINS TÊM CONSEQUÊNCIAS?

O que acontece quando falhamos em mensurar conceitos-chave de nossas teorias de modo que eles sejam tanto válidos como confiáveis? Retorne à Figura 1.2, que destaca a distinção entre conceitos abstratos de interesse teórico e as variáveis que observamos no mundo real. Se as variáveis que observamos no mundo real não fazem um bom trabalho em espelhar os conceitos abstratos, então isso afeta nossa habilidade de avaliar conclusivamente o suporte empírico de uma teoria. Isto é, como sabemos se nossa teoria tem respaldo se fizemos um trabalho ruim ao mensurar os conceitos-chave que observamos? Se nossa análise empírica é baseada em medidas que não capturam a essência dos conceitos abstratos de nossas teorias, então é improvável que tenhamos alguma confiança nos nossos achados.

5.7 CONHECENDO SEUS DADOS ESTATISTICAMENTE

Até o momento, discutimos detalhes sobre a mensuração de variáveis. A mensuração de variáveis envolve muito esforço e reflexão. Mas, uma vez que o pesquisador tenha coletado os dados e ficado familiarizado e satisfeito com a forma pela qual as variáveis foram mensuradas, é importante que ele tenha uma boa ideia dos tipos de valores que elas assumem antes de passar para o teste de conexões causais entre duas ou mais variáveis. Quais são os valores “típicos” de uma variável? Quão agrupados (ou dispersos) são os valores da variável?

Antes de prosseguir para o teste das relações teorizadas entre duas ou mais variáveis, é essencial entender as propriedades e características de cada uma das variáveis. Para colocar isso de modo diferente, queremos aprender algo sobre como os valores de cada variável “se parecem”. Como realizamos isso? Uma possibilidade é listar todos os valores observados de uma variável mensurada. Por exemplo, a seguir são expostos os percentuais de votos para os candidatos do partido que ocupava a presidência durante

¹⁸ Ver Gibson (1992).

as eleições presidenciais americanas de 1880 a 2008¹⁹: 50,22; 49,846; 50,414; 48,268; 47,76; 53,171; 60,006; 54,483; 54,708; 51,682; 36,119; 58,244; 58,82; 40,841; 62,458; 54,999; 53,774; 52,37; 44,595; 57,764; 49,913; 61,344; 49,596; 61,789; 48,948; 44,697; 59,17; 53,902; 46,545; 54,736; 50,265; 51,2; 46,311. Podemos observar a partir desse exemplo que, ao menos que tenhamos um número baixo de observações, a listagem de observações é difícil de ser interpretada – se não temos a menor ideia do padrão geral da floresta, nos perdemos no meio das árvores. Por essa razão, utilizamos estatísticas e gráficos descritivos para pegar uma grande quantidade de informações e reduzi-la a pequenos pedaços que sumarizam a informação.

Estatísticas e gráficos descritivos são ferramentas úteis para ajudar pesquisadores a conhecer seus dados antes de passar a testes de hipóteses. Eles também são úteis, algumas vezes, quando escrevemos sobre a pesquisa de outra pessoa. A apresentação de estatísticas e gráficos descritivos na versão final do seu trabalho é uma decisão que deve ser tomada caso a caso. É, contudo, cientificamente importante que essa informação esteja disponível aos consumidores da sua pesquisa de alguma forma²⁰.

Uma das principais formas de distinguir variáveis é pela **métrica de mensuração**. A métrica de mensuração de uma variável é o tipo de valores que essa variável assume. Discutiremos isso em detalhe na próxima seção, quando descreveremos os três diferentes tipos de variáveis. Então explicaremos que, apesar da natureza imperfeita das distinções entre esses três tipos de variáveis, somos forçados a escolher entre duas grandes classificações de variáveis – categórica e contínua – quando as classificamos. O restante deste capítulo discute estratégias para descrever **variáveis contínuas** e categóricas.

5.8 O QUE É A MÉTRICA DE MENSURAÇÃO DE UMA VARIÁVEL?

Não existem regras rígidas e rápidas para descrever variáveis, mas um ponto de partida importante envolve a métrica de mensuração de cada variável. Lembre-se do capítulo 1, em que pensamos em cada variável em termos dos seus valores e seus nomes. O nome é uma descrição da variável – por exemplo, “gênero do respondente” – e seu valor são as denominações nos quais as variáveis ocorrem – por exemplo, “homem” ou “mulher”. Para o tratamento na maioria das análises estatísticas, somos forçados a dividir nossas variáveis em dois tipos, de acordo com a métrica em que seus valores ocorrem: categóricas ou contínuas. Na verdade, variáveis podem ter ao menos três tipos diferentes de métricas e existem muitas variáveis que não se encaixam ordenadamente em apenas uma dessas classificações. Para ajudá-lo a entender melhor cada

¹⁹ Esta medida é construída de modo que seja comparável ao longo do tempo. Embora partidos menores e candidatos independentes tenham ocasionalmente participado de eleições, focamos apenas os votos dos dois maiores partidos. E, como queremos testar a teoria do voto econômico, queremos ter uma medida de apoio ao incumbente. Em eleições em que o presidente atual não está concorrendo à reeleição, ainda existe um partido que será avaliado por ser responsável pelo desempenho econômico.

²⁰ Muitos pesquisadores apresentarão essas informações em um apêndice, a não ser que exista algo particularmente digno de nota sobre as características de uma ou mais variáveis.

um desses tipos de variáveis, apresentaremos em cada um deles um exemplo. Todos os exemplos que utilizamos nesta descrição inicial são provenientes de pesquisas de *survey*, mas os princípios básicos das métricas de mensuração são as mesmas independentemente do tipo de dado analisado.

5.8.1 VARIÁVEIS CATEGÓRICAS

Variáveis categóricas são variáveis em que os casos assumem valores que são iguais ou diferentes aos valores de outros casos, mas dos quais não podemos fazer qualquer distinção ordinal universal entre eles. Se considerássemos uma variável que tenha o nome “identificação religiosa”, alguns valores dessa variável seriam “católico”, “muçulmano”, “não religioso”, entre outros. Embora esses valores sejam claramente diferentes uns dos outros, não podemos fazer uma ordenação universal com eles. Em outras palavras, com variáveis categóricas como essa, não é possível ranquear as categorias do menor valor para o maior: por exemplo, o valor “muçulmano” não é nem maior, nem menor do que o “não religioso” (e assim por diante). Em vez disso, sabemos que os casos que possuem o mesmo valor são iguais e os que possuem valores diferentes são diferentes. O termo “categórica” sintetiza a essência desse tipo de variável; podemos colocar casos individuais em categorias baseadas em seus valores, mas não podemos ranquear ou ordenar esses valores de outra forma.

5.8.2 VARIÁVEIS ORDINAIS

Como as variáveis categóricas, as **variáveis ordinais** também são variáveis para as quais os casos assumem valores que são iguais ou diferentes aos de outros casos. A distinção entre variáveis categóricas e ordinais é que *podemos* fazer uma ordenação universal que distinga os casos quando a variável é ordinal. Por exemplo, considere a variável nomeada “situação financeira familiar no passado”, que tem sido comumente utilizada como variável independente em estudos sobre o voto econômico no nível do indivíduo. No National Election Study (NES) de 2004, pesquisadores criaram essa variável perguntando, primeiramente, aos entrevistados a seguinte questão: “Estamos interessados em saber como as pessoas estão financeiramente hoje em dia. Você diria que você (e sua família que vive com você) está melhor ou pior do que estava um ano atrás?”. Pesquisadores então perguntam aos entrevistados que responderam “melhor” ou “pior”: “Muito [melhor/pior] ou um pouco [melhor/pior]?”. A variável era então codificada como:

1. Muito melhor
2. Um pouco melhor
3. A mesma coisa
4. Um pouco pior
5. Muito pior

É bastante claro que essa variável é ordinal porque, à medida que vamos do começo ao final da lista, nos movemos de avaliações mais positivas a mais negativas de como os indivíduos (e as famílias com quem eles vivem) estavam financeiramente no último ano.

Outro exemplo é a variável denominada “identificação partidária”. No NES de 2004, pesquisadores criaram essa variável a partir das respostas dos entrevistados para a seguinte pergunta: “De um modo geral, você usualmente se considera republicano, democrata, independente ou outro?”²¹.

1. Republicano
2. Independente
3. Democrata

Se todos os casos que assumem o valor “independente” representarem indivíduos cuja visão política esteja em algum ponto entre “republicano” e “democrata”, podemos chamar a variável “identificação partidária” de ordinal. Se esse não for o caso, então ela é uma variável categórica.

5.8.3 VARIÁVEIS CONTÍNUAS

Uma importante característica que as variáveis ordinais *não* possuem é uma **diferença de unidade constante**. Uma variável possui diferença de unidade constante se o aumento de uma unidade no valor da variável *sempre* significar a mesma coisa. Se retornarmos para os exemplos da seção anterior, podemos ranquear as cinco categorias da variável “situação financeira familiar no passado” de 1 para a melhor situação até 5 para a pior situação. Mas não nos sentimos muito confiantes em trabalhar com esses valores como normalmente trabalhamos com números. Em outras palavras, podemos dizer que a diferença entre “um pouco pior” e “a mesma coisa” (4–3) é a mesma que entre “muito pior” e “um pouco pior” (5–4)? E que a diferença entre “muito pior” e “a mesma coisa” (5–3) é o dobro da entre “um pouco pior” e “muito melhor” (2–1)? Se a resposta para essas duas perguntas for “sim”, então a variável “situação financeira familiar no passado” é contínua.

Se fizermos a mesma pergunta sobre a “identificação partidária”, devemos ser um pouco céticos. Podemos ranquear as três categorias de “identificação partidária”, mas

²¹ Quase todos os entrevistados se colocam em uma das três primeiras categorias. Por exemplo, em 2004, 1.128 dos 1.212 respondentes (93,1%) do NES realizado pós-eleição afirmaram ser republicanos, democratas ou independentes. Para os nossos propósitos, ignoraremos as respostas “outro”. Note que pesquisadores usualmente apresentam a identificação partidária a partir de uma escala de sete pontos que varia de “republicano convicto” a “democrata convicto” baseada em uma série de perguntas feitas aos entrevistados para melhor caracterizar suas posições.

não podemos, com grande confiança, afirmar que “republicano” tem valor igual a 1, “independente” tem valor igual a 2 e “democrata” tem valor igual a 3 e trabalhar com esses valores do mesmo modo como tipicamente trabalhamos com números. Não podemos dizer que a diferença entre um “Independente” e um “Republicano” ($2 - 1$) é a mesma que a diferença entre um “democrata” e um “independente” ($3 - 2$) – apesar do fato de que $3 - 2 = 2 - 1 = 1$. Certamente, não podemos dizer que a diferença entre um “democrata” e um “republicano” ($3 - 1$) é o dobro da diferença entre um “independente” e um “republicano” ($2 - 1$), apesar do fato de 2 ser o dobro de 1.

Podemos dizer que a métrica na qual mensuramos uma variável tem diferença de unidade constante se um aumento de uma unidade no valor de uma variável indica a mesma quantidade de mudança entre *todos os valores* da variável. Variáveis contínuas são variáveis que possuem diferença de unidade constante²². Imagine, por exemplo, uma variável com o nome “idade em anos”. O aumento de uma unidade nessa variável *sempre* indica um indivíduo um ano mais velho; isso é verdade quando estamos falando tanto sobre um caso com valor igual a 21 quanto sobre um caso com valor igual a 55.

5.8.4 TIPOS DE VARIÁVEIS E ANÁLISE ESTATÍSTICA

Como vimos nas seções anteriores, nem sempre variáveis se encaixam perfeitamente em uma das três categorias. Quando passamos para a análise estatística, devemos decidir entre tratar nossa variável como uma variável categórica ou como uma variável contínua. Para algumas variáveis, essa é uma escolha bastante simples. Contudo, para outras, é uma escolha muito difícil. Se tratarmos uma variável ordinal como se fosse categórica, estamos agindo como se soubéssemos menos sobre os valores dessa variável do que realmente sabemos. De outro lado, tratar uma variável ordinal como se fosse uma variável contínua significa que estamos assumindo que ela possui diferença de unidade constante. De qualquer forma, é crítico que tenhamos consciência das nossas decisões. Podemos sempre repetir nossa análise utilizando um diferente pressuposto e observar quão robustos são nossos achados.

Com tudo isso em mente, apresentamos discussões separadas do processo de descrição da variação de uma variável categórica e uma variável contínua. Uma variação de uma variável é a distribuição de valores que a variável assume entre os casos para os quais foi mensurada. É importante que tenhamos um forte conhecimento da variação de cada uma de nossas variáveis antes de traduzir nossa teoria em hipóteses, verificar se existe covariação entre duas variáveis (terceiro obstáculo causal do capítulo 3) e pensar se pode existir uma terceira variável que possa fazer que a covariação observada entre nossa variável independente e dependente seja espúria (quarto obstáculo).

²² Algumas vezes podemos chamar esse tipo de variável de “variável intervalar”. Uma distinção adicional que você encontrará para variáveis contínuas é se elas tiverem um valor zero substancialmente significativo. Usualmente descrevemos variáveis que possuem essa característica como “variáveis-razão” (ou “proporção”).

Como acabamos de delinear, estatísticas e gráficos descritivos são úteis para sumarizar a variação para variáveis individuais. Outro modo pelo qual descrevemos distribuições de variáveis é por meio de medidas de **tendência central**. Medidas de tendência central nos dizem os valores típicos encontrados no centro da distribuição de uma variável.

5.9 DESCREVENDO VARIÁVEIS CATEGÓRICAS

Com variáveis categóricas, podemos entender a frequência com que cada um dos valores da variável ocorre em nossos dados. O modo mais simplista para fazermos isso é produzir uma tabela de frequência na qual os valores da nossa variável são expostos em uma coluna e a frequência em que eles ocorrem (em número absoluto de casos e/ou em termos percentuais) em outra(s) coluna(s). A Tabela 5.1 mostra um exemplo desse tipo de tabela para a variável “identificação religiosa” do *survey* NES conduzido durante as eleições nacionais nos Estados Unidos em 2004.

Tabela 5.1 – Frequência para a identificação religiosa segundo o NES de 2004.

Categoria	Número de casos	Percentual
Protestante	672	56,14
Católico	292	24,39
Judeu	35	2,92
Outro	17	1,42
Nenhum	181	15,12
Total	1197	99,9

A única medida de tendência central apropriada para dados categóricos é a **moda**, definida como o valor que ocorre com maior frequência. Na Tabela 5.1, a moda da distribuição é o valor “protestante”, porque existem mais protestantes do que membros de qualquer outra categoria.

Um modo comum com que pessoas não habituadas com estatística apresentam dados de frequência é por meio de um gráfico de setores, tal qual na Figura 5.4. Gráficos de setores são um dos modos que temos para visualizar o percentual de casos de uma determinada categoria. Muitos estatísticos são contrários ao uso desse tipo de gráfico e recomendam fortemente o uso do gráfico de barras. O gráfico de barras, como o da Figura 5.5, é outro modo que temos para ilustrar a frequência de variáveis categóricas. Vale ressaltar, contudo, que a informação que podemos extrair desses dois tipos de gráfico é apresentada de maneira clara e precisa nas colunas de frequência e percentuais da Tabela 5.1.

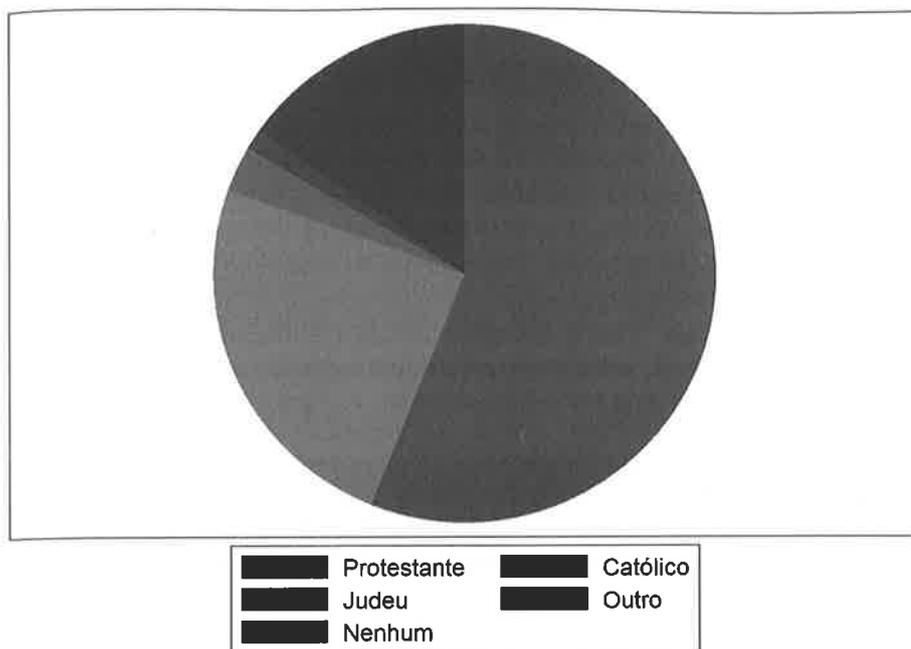


Figura 5.4 – Gráfico de setores da identificação religiosa, NES 2004.

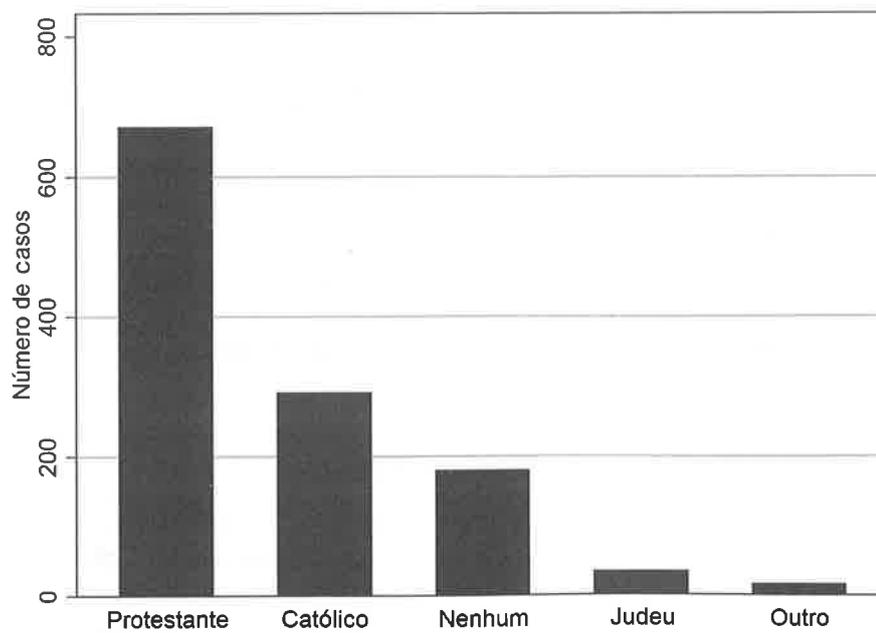


Figura 5.5 – Gráfico de barras da identificação religiosa, NES 2004.

5.10 DESCREVENDO VARIÁVEIS CONTÍNUAS

As estatísticas e os gráficos para descrever variáveis contínuas são consideravelmente mais complicados do que os para variáveis categóricas. Isso ocorre porque variáveis contínuas são matematicamente mais complexas do que variáveis categóricas. Com variáveis contínuas, queremos saber a tendência central e a amplitude ou variação dos valores ao redor da tendência central. Com variáveis contínuas, também queremos estar conscientes dos **valores discrepantes**²³. Valores discrepantes são casos para os quais os valores de uma variável são extremamente altos ou baixos em comparação com o resto dos valores da variável. Quando encontramos um valor discrepante, queremos ter certeza de que o caso não é produto de algum tipo de erro.

A maioria dos *softwares* estatísticos possui um comando que calcula uma bateria de estatísticas descritivas para variáveis contínuas. A Figura 5.6 mostra o produto do comando *summarize* do Stata com a opção *detail* para a variável “percentual de votos recebidos pelo partido do incumbente nas eleições presidenciais americanas” entre os anos de 1876 e 2006. Nas estatísticas do lado esquerdo (primeiras três colunas da esquerda), estão o que chamamos de **estatísticas de ordenamento** e, no lado direito, (duas colunas do lado direito) as conhecidas como **estatísticas de momento**. Embora ambos os tipos sejam utilizados para descrever a variação de variáveis contínuas, eles fazem isso de modos ligeiramente diferentes e, portanto, juntos, são bastante úteis para fornecer um quadro completo da variação de uma única variável.

```
. summarize inc_vote, det
```

inc_vote				
Percentiles		Smallest		
1%	36.148	36.148		
5%	40.851	40.851		
10%	44.842	44.71	Obs	34
25%	48.516	44.842	Sum of Wgt.	34
50%		51.4575	Mean	51.94718
			Std. Dev.	5.956539
		Largest		
75%	54.983	60.006	Variance	35.48036
90%	60.006	61.203	Skewness	-.3065283
95%	61.791	61.791	Kurtosis	3.100499
99%	62.226	62.226		

Figura 5.6 – Exemplo do produto do comando *summarize* do Stata com a opção *detail*.

5.10.1 ESTATÍSTICAS DE ORDENAMENTO

O cálculo das estatísticas de ordenamento começa com a ordenação dos valores de uma variável contínua do menor para o maior, seguida pela identificação de conjunturas cruciais ao longo do caminho. Uma vez que temos os casos ordenados, o ponto do meio da nossa contagem de casos é conhecido como caso mediano. Lembre-se

²³ NT: No original, outliers.

que anteriormente, no capítulo, definimos a variável na Figura 5.6 como o percentual de votos populares recebidos pelo candidato do partido que ocupava a presidência durante uma eleição presidencial nos Estados Unidos entre os anos de 1876 e 2006. Para simplificar a exposição, chamaremos essa variável de “voto do incumbente” a partir deste momento. Para calcular as estatísticas de ordenamento para essa variável, precisamos primeiro ordenar os casos do menor valor observado para o maior. Essa ordem é exposta na Tabela 5.2. Com estatísticas de ordenamento, mensuramos a tendência central como o **valor mediano** da variável. O valor mediano é o valor do caso que está exatamente no centro dos nossos casos quando os ranqueamos do menor valor observado ao maior. Quando temos um número par de casos, como temos na Tabela 5.2, calculamos a média utilizando os dois valores mais próximos do centro para obter o valor mediano (em nosso exemplo, calculamos a mediana como $\frac{51,233 + 51,682}{2} = 51,4575$). Esse valor também é conhecido como o valor da variável

no ponto de 50% da ordenação. De modo similar, podemos obter o valor da variável em qualquer outro percentual do ranking no qual temos interesse. Outros pontos do *ranking* que frequentemente são de interesse são os de 25% e de 75%, que também são conhecidos como primeiro e terceiro quartis da distribuição. A diferença entre o valor da variável no ponto de 25% (primeiro quartil) e 75% (terceiro quartil) é conhecida por “intervalo interquartis” ou “IIQ” da variável. Em nosso exemplo, o valor no ponto de 25% é 48,516 e o valor no ponto de 75% é 54,983. Isso faz o IIQ = 54,983 – 48,516 = 6,467. Em termos de estatística de ordenamento, o valor mediano para uma variável é uma medida da sua tendência central, enquanto o IIQ é uma medida de **dispersão**, ou amplitude, dos valores.

Com as estatísticas de ordenamento, também queremos observar os valores maiores e menores para identificar valores discrepantes. Lembre-se que definimos valor discrepante no começo desta seção como “casos para os quais os valores de uma variável são extremamente altos ou baixos em comparação com o resto dos valores da variável”. Se observarmos os maiores valores na Tabela 5.2, podemos ver que não existem casos que realmente se adequam a essa descrição. Embora certamente existam alguns valores que são maiores que a mediana e o valor na posição de 75%, eles não são “extremamente” maiores que o restante dos valores. Pelo contrário, parece haver um crescimento progressivo do valor da posição de 75% até o valor mais elevado. A história no outro extremo dos valores da Tabela 5.2 é um pouco diferente. O valor de 36,148 em 1920 parece se encaixar na nossa definição de um valor discrepante. O valor 40,851 em 1932 também é um caso que fica na fronteira. Sempre que observarmos valores discrepantes, devemos checar se mensuramos os valores desses casos de modo acurado. Algumas vezes, verificamos que os valores discrepantes são resultado de erros que cometemos ao coletar os dados. No nosso exemplo, uma checagem do nosso conjunto de dados revela que o caso discrepante acontece em 1920, quando o partido do candidato incumbente recebeu apenas 36,148% dos votos. Uma checagem adicional dos nossos dados indica que esse é, de fato, o valor correto para a variável em 1920²⁴.

²⁴ Uma questão óbvia é: “Por que o valor de 1920 é tão baixo?”. Essa foi a primeira eleição presidencial após o fim da Primeira Guerra Mundial, durante um período em que ocorreu muita perturbação econômica e política. Por sua vez, a eleição em 1932 ocorreu bem no início de uma grande desaceleração econômica conhecida como a “grande depressão”, então faz sentido que o partido do presidente em exercício tenha recebido uma porcentagem de votos relativamente mais baixa nessa eleição.

Tabela 5.2 – Valores da votação do incumbente ordenados do menor para o maior.

Posição	Ano	Valor
1	1920	36,148
2	1932	40,851
3	1952	44,71
4	1980	44,842
5	2008	46,311
6	1992	46,379
7	1896	47,76
8	1892	48,268
9	1876	48,516
10	1976	48,951
11	1968	49,425
12	1884	49,846
13	1960	49,913
14	1880	50,22
15	2000	50,262
16	1888	50,414
17	2004	51,233
18	1916	51,682
19	1948	52,319
20	1900	53,171
21	1944	53,778
22	1988	53,832
23	1908	54,483
24	1912	54,708
25	1996	54,737
26	1940	54,983
27	1956	57,094
28	1924	58,263
29	1928	58,756
30	1984	59,123
31	1904	60,006
32	1964	61,203
33	1972	61,791
34	1936	62,226

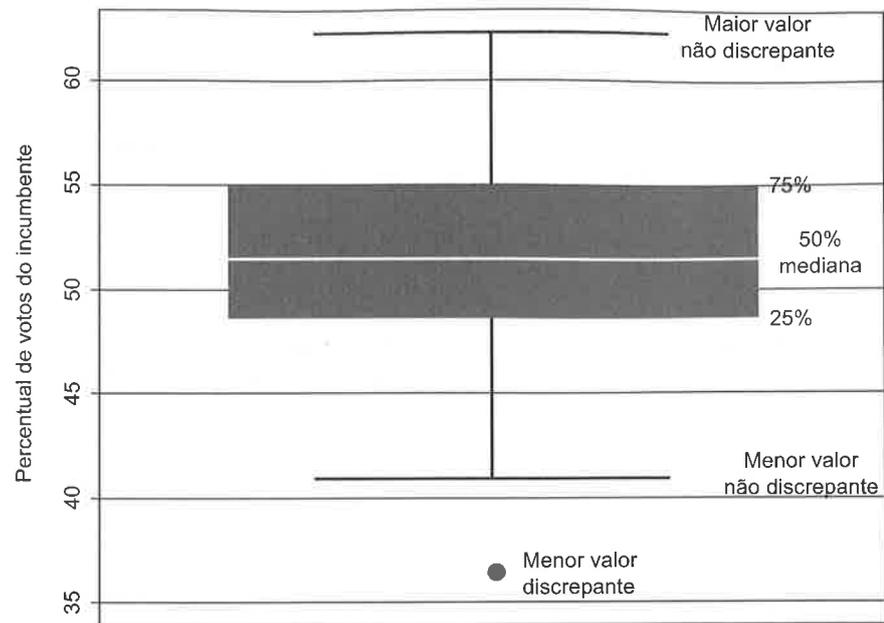


Figura 5.7 – Box-plot do percentual de votação do partido incumbente para a presidência, 1876-2008.

A Figura 5.7 apresenta um gráfico *box-plot*²⁵ das estatísticas de *ranking* para nossa variável de voto presidencial. Esse gráfico mostra a distribuição da variável ao longo da dimensão vertical. Se começarmos pelo centro da caixa na Figura 5.7, temos o valor mediano (ou o ponto de 50%) da nossa variável representado como a linha no centro da caixa. Os dois extremos da caixa mostram os valores do primeiro percentil (posição de 25%) e do terceiro percentil (posição de 75%) da nossa variável. As linhas nos extremos do lado externo da caixa representam os valores mais alto e mais baixo de nossa variável, excetuando valores discrepantes. Cada programa de estatística tem suas próprias regras para lidar com valores discrepantes, portanto é importante saber se nosso *box-plot* é ou não configurado para exibir os valores discrepantes. Essas configurações usualmente são ajustáveis no *software* estatístico. O cálculo para definir se um caso é considerado como um valor discrepante em um *box-plot* é bastante direto. O cálculo começa com o valor do IIQ para a variável. Definimos um caso como um valor discrepante se o valor dele for superior ao valor do terceiro percentil em uma vez e meia o valor do IIQ ou menor que o primeiro percentil em uma vez e meia o valor do IIQ. Para a Figura 5.7, fizemos um gráfico que apresenta os valores discrepantes, e podemos observar um valor na parte inferior do nosso gráfico. Como sabemos pela Tabela 5.2, esse é o valor da eleição de 1920, em que o candidato do partido do presidente em exercício recebeu 36,148% dos votos.

²⁵ *Box-plot* em português é também conhecido pela expressão “caixa de bigode”. [N.T.]

5.10.2 ESTATÍSTICA DE MOMENTO

Estatísticas de momento de uma variável são um conjunto de estatísticas que descrevem a tendência central para uma variável e a distribuição dos valores ao redor dela. A mais familiar dessas estatísticas é conhecida como o **valor médio** para a variável. Para uma variável Y , a média é descrita e calculada por

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n},$$

em que \bar{Y} , conhecido como “Y-barra”, indica a média de Y , que é a soma de todos os valores de Y para todos os casos de Y , Y_p , dividido pelo número total dos casos, n ²⁶. Embora a média ou o valor médio sejam familiares para todos, nem todos estão familiarizados com duas características da média que tornam essa medida particularmente atrativa. A primeira é conhecida como “**propriedade da soma zero**”:

$$\sum_{i=1}^n (Y_i - \bar{Y}) = 0,$$

que significa que a soma das diferenças de cada valor Y , Y_p , e o valor médio de Y , \bar{Y} , é igual a zero. A segunda característica desejável do valor médio é conhecida como “**propriedade dos mínimos quadrados**”:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 < \sum_{i=1}^n (Y_i - c)^2 \forall c \neq \bar{Y},$$

que significa que a soma dos quadrados da diferença entre cada valor Y , Y_p , e o valor da média de Y , \bar{Y} , é menor que a soma dos quadrados da diferença entre cada valor Y , Y_p , e algum valor c , para todo (\forall) c 's não igual a (\neq) \bar{Y} . Por causa dessas duas propriedades, a média também é conhecida como o **valor esperado** da variável. Pense nisso do seguinte modo: se alguém lhe pedisse um palpite para o valor de um caso individual oferecendo como informação apenas o valor médio, baseado nessas duas propriedades da média, o valor médio seria o melhor palpite.

A próxima estatística de momento para uma variável é a **variância**. Representamos a variância e a calculamos por:

$$\text{var}(Y) = \text{var}_Y = s_Y^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1},$$

que significa que a variância de Y é igual à soma dos quadrados das diferenças de cada valor Y , Y_p , e sua média dividida pelo número de casos menos um²⁷. Observando

²⁶ Para entender uma fórmula como essa, é útil ler cada um dos termos da fórmula e traduzi-los em palavras, como fizemos aqui.

²⁷ O “menos um” nessa equação é um ajuste feito para considerar o número de “graus de liberdade” com o qual esse cálculo foi feito. Discutiremos o conceito de graus de liberdade no capítulo 7.

essa fórmula, o que aconteceria se não tivéssemos variação alguma em Y ($Y_i = \bar{Y} \forall i$)? Nesse caso, a variância seria igual a zero. Quando casos individuais estão mais e mais distantes em relação à média, o resultado desse cálculo aumenta. A lógica da variância é, portanto, que ela transmite a amplitude dos dados ao redor da média. Uma medida mais intuitiva da variância é o **desvio-padrão**:

$$sd(Y) = sd_Y = s_Y = \sqrt{\text{var}(Y)} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

Grosseiramente falando, o desvio-padrão é a média das diferenças entre os valores de Y (Y) e a média de Y (\bar{Y}). À primeira vista, isso pode não ser aparente. Mas o ponto importante para entender sobre essa fórmula é que elevamos ao quadrado cada diferença da média e então tiramos a raiz quadrada do resultado dos desvios quadrados para evitar que os valores positivos e negativos dos desvios cancelem uns aos outros²⁸.

A variância e o desvio-padrão nos dão um resumo numérico da distribuição dos casos ao redor do valor médio de uma variável²⁹. Também podemos descrever as distribuições visualmente. A ideia de descrever visualmente distribuições consiste em produzir um gráfico com duas dimensões, no qual o eixo horizontal (eixo x) apresenta os valores da variável, e a dimensão vertical (eixo y), a frequência relativa dos casos. Um dos modos mais populares para apresentar a distribuição de uma variável é o **histograma**, como o da Figura 5.8.

Um problema com histogramas é que nós (ou o programa de computador que estamos utilizando) devemos escolher quantos blocos são expostos no nosso histograma. Mudar o número de blocos em um histograma pode mudar nossa impressão da distribuição da nossa variável. A Figura 5.9 mostra a mesma variável que apresentamos na Figura 5.8 com dois e, então, dez blocos. Embora tenham sido gerados com os mesmos dados, os gráficos da Figura 5.9 são bastante diferentes um do outro.

Outra opção é o **gráfico kernel de densidade**, como o da Figura 5.10, que é baseado no cálculo da densidade dos casos ao longo dos valores.

²⁸ Um método alternativo que produziria resultados bastante similares a esse seria calcular o valor médio do valor absoluto de cada diferença da média: $\left(\frac{\sum_{i=1}^n |Y_i - \bar{Y}|}{n} \right)$.

²⁹ A **obliquidade** (*skewness*) e a **curtose** (*kurtosis*) de uma variável transmitem aspectos adicionais da distribuição de uma variável. O cálculo da obliquidade indica a assimetria da distribuição ao redor da média. Se os dados são simetricamente distribuídos ao redor da média, então essa estatística será igual a zero. Se a obliquidade é negativa, isso indica que existem mais valores abaixo da média do que acima; se a obliquidade é positiva, isso indica que existem mais valores acima da média do que abaixo. A curtose indica o achatamento da distribuição estatística. Uma curtose positiva indica uma distribuição bastante afunilada ou uma concentração de valores próximos à média; enquanto valores negativos da curtose indicam uma distribuição achatada ou mais casos distantes do valor médio. Como discutiremos no capítulo 6, a obliquidade e a curtose são medidas que possuem valor zero para a distribuição normal.

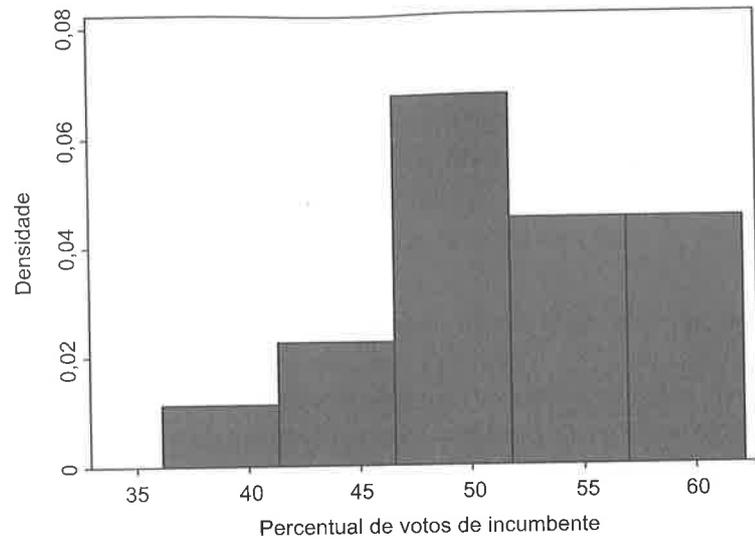


Figura 5.8 – Histograma do percentual de votação do partido incumbente para a presidência, 1876-2008.

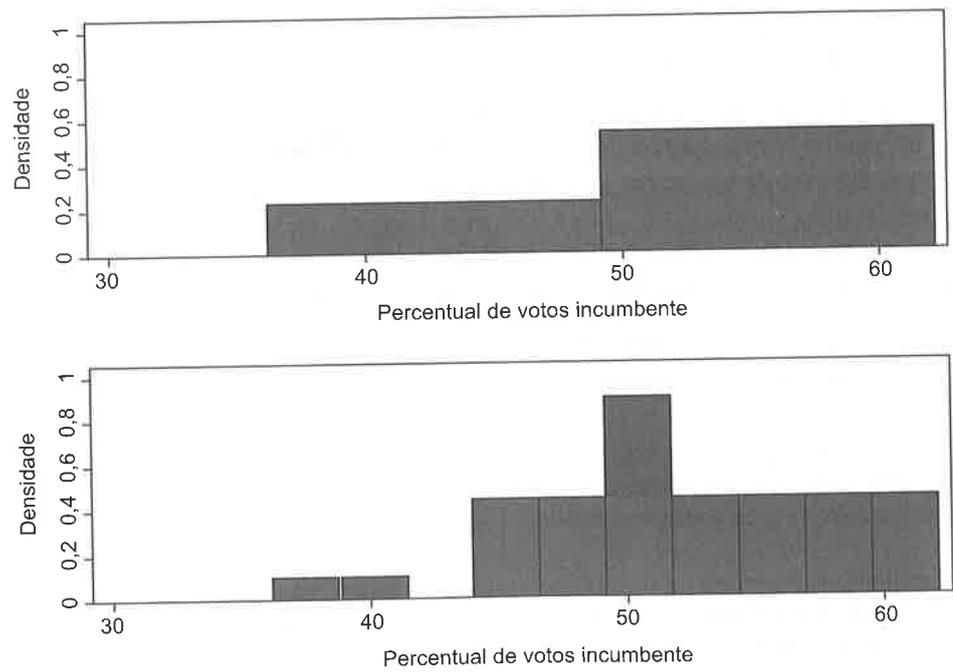


Figura 5.9 – Histogramas do percentual de votação do partido incumbente para a presidência, 1876-2008, descritos com dois e dez blocos.

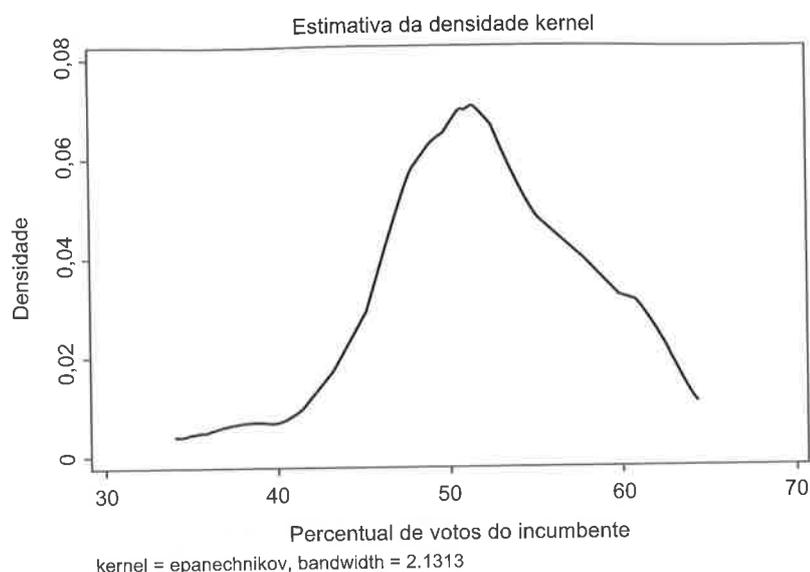


Figura 5.10 – Gráfico de densidade *kernel* do percentual de votação do partido incumbente para a presidência, 1876-2008.

5.11 LIMITAÇÕES DAS ESTATÍSTICAS DESCRITIVAS E DOS GRÁFICOS

As ferramentas que apresentamos nas últimas três seções deste capítulo são úteis para propiciar uma primeira visão dos dados para uma variável. Observar os dados com essas ferramentas ajudará você a conhecer melhor seus dados e cometer poucos erros no longo prazo. É importante, contudo, notar que não podemos testar teorias causais com uma única variável. Afinal, como temos notado, uma teoria é uma tentativa de afirmação sobre uma possível relação causal entre duas variáveis. Não é porque discutimos como descrever uma única variável que começamos a discutir como fazer testes apropriados de teorias causais.

5.12 CONCLUSÕES

O modo como mensuramos nossos conceitos importa. Como foi possível observar nos exemplos precedentes, diferentes estratégias de mensuração podem produzir (e, algumas vezes, de fato produzem) diferentes conclusões sobre relações causais.

Um dos pontos que deve ser guardado deste capítulo é que a mensuração não pode ocorrer em um vácuo teórico. O *propósito teórico* do empreendimento acadêmico deve informar o processo de como mensuramos o que mensuramos. Por exemplo, lembre-se da nossa discussão anterior sobre os vários modos de medir pobreza. O modo como mensuramos esse conceito depende de qual é o nosso objetivo. Em um processo de mensuração da pobreza, se nosso objetivo teórico é avaliar a efetividade de diferen-

tes políticas de combate à pobreza, teríamos uma forma de mensuração diferente do que estudiosos cujo objetivo teórico é estudar como ser pobre influencia as atitudes políticas de uma pessoa. No primeiro caso, daríamos ênfase em uma medida de pobreza pré-transferências, enquanto, no último caso, uma medida pós-transferências seria mais apropriada.

Outros pontos importantes que discutimos ao longo do capítulo ressaltam a necessidade de se ter clareza conceitual, confiabilidade e validade das medidas. Verificamos que alguns conceitos, como democracia e tolerância política, necessitam ser cuidadosamente mensurados para que tenhamos maior confiança em nossos resultados e afirmações. Adicionalmente, apresentamos como estatísticas descritivas podem ser utilizadas para conhecermos as variáveis que utilizamos em nossas pesquisas.

CONCEITOS INTRODUZIDOS NESTE CAPÍTULO

- Confiabilidade – a extensão na qual a aplicação das mesmas regras de mensuração para o mesmo caso ou observação produzirá resultados idênticos.
- Curtose – uma medida estatística que indica o achatamento da distribuição estatística para uma única variável.
- Desvio-padrão – uma medida estatística da dispersão de uma variável ao redor da sua média.
- Diferença de unidade constante – uma variável tem diferenças de unidade igual se o aumento de uma unidade no valor dessa variável significa sempre a mesma coisa.
- Dispersão – a extensão ou alcance de valores de uma variável.
- Estatísticas de momento – um grupo de estatísticas utilizado para descrever a variação de uma variável contínua com base em cálculos numéricos.
- Estatísticas de ordenamento – um grupo de estatísticas utilizado para descrever a variação de uma variável contínua com base na ordenação das observações do menor para o maior valor.
- Gráfico *kernel* de densidade – uma descrição visual da distribuição de uma única variável baseada no cálculo da densidade dos casos ao longo da variação dos valores.
- Histograma – uma descrição visual da distribuição de uma única variável que é representada por gráfico de duas dimensões em que a dimensão horizontal (eixo x) exibe os valores da variável, e a dimensão vertical (eixo y), a frequência relativa dos casos.
- Métrica de mensuração – tipo de valor que uma variável assume.
- Moda – o valor mais frequente de uma variável.
- Obliquidade – uma medida estatística que indica a simetria da distribuição ao redor da média.

- Propriedade dos mínimos quadrados – uma propriedade do valor médio de uma única variável Y , na qual a soma dos quadrados da diferença entre cada valor Y , Y_p e o valor médio de Y , \bar{Y} , é menor que a soma dos quadrados das diferenças entre cada valor Y , Y_p e algum valor c , para todo (\forall) c 's não igual a (\neq) \bar{Y} .
- Propriedade da soma zero – uma propriedade do valor médio de variável Y , que significa que a soma das diferenças entre cada valor Y , Y_p e o valor da média de Y , \bar{Y} , é igual a zero.
- Tendência central – valores típicos para uma variável no centro de sua distribuição.
- Validade – o grau com o qual uma medida representa acuradamente o conceito que supostamente mede.
- Validade de constructo – o grau no qual a medida é relacionada a outra medida com a qual a teoria afirma que ela é relacionada.
- Validade de conteúdo – o grau no qual uma medida contém todos os elementos críticos que, como grupo, definem o conceito que queremos medir.
- Validade de face – se uma medida parece (ou não), em sua face, estar mensurando o que se propõe a mensurar.
- Valor discrepante – um caso para o qual o valor da variável é extremamente alto ou baixo em relação ao resto dos valores da variável.
- Valor esperado – um sinônimo para média.
- Valor mediano – o valor do caso que está exatamente no centro da distribuição dos nossos casos quando ordenamos os valores de uma única variável do seu menor para o maior valor observado.
- Valor médio – média aritmética de uma variável que é igual à soma de todos os valores Y para todos os casos individuais de Y , Y_p dividida pelo número total de casos.
- Variação – a distribuição que os valores de uma variável assumem ao longo dos casos para os quais ela foi mensurada.
- Variância – uma medida estatística de dispersão de uma variável ao redor de sua média.
- Variáveis categóricas – variáveis para as quais os casos possuem valores que são diferentes ou iguais aos valores para os outros casos, mas para as quais não é possível fazer uma ordenação universal que distinga os valores.
- Variável contínua – uma variável cuja métrica tem diferença de unidade constante, em que o aumento de uma unidade em um valor de uma variável indica a mesma quantidade de mudança em todo o espectro de valores da variável.
- Variável ordinal – uma variável para a qual podemos fazer um ordenamento universal para seus valores, mas cuja métrica não possui diferenças de unidades iguais.

- Viés de mensuração – a sobrenotificação ou subnotificação sistemática dos valores para uma variável.

EXERCÍCIOS

1. Suponha que um pesquisador desejasse mensurar os esforços do governo federal para tornar a educação dos cidadãos uma prioridade. O pesquisador propôs calcular o orçamento do governo para a educação como percentual do PIB total e utilizar essa medida para verificar o comprometimento do governo com a educação. Em termos de validade, quais são os pontos fortes e fracos dessa medida?
2. Suponha que um pesquisador desejasse mensurar a cobertura na mídia na campanha de um candidato a um cargo político e, portanto, criasse um conjunto de regras para codificar palavras que apareceram em artigos de jornais como “a favor” ou “contra” o candidato. Em vez de contratar estudantes para implementar as regras de codificação, o pesquisador utilizou um computador para codificar o texto, por meio da contagem da frequência de certas palavras em uma série de artigos. Uma estratégia de mensuração como essa seria confiável? Por quê?
3. Para cada um dos seguintes conceitos, identifique se eles sofreriam de prováveis problemas de viés de mensuração, invalidez, não confiabilidade ou nenhum dos mencionados. Explique suas respostas.
 - a) A mensuração do conceito de aprovação do presidente a partir do uso de resultados das respostas de entrevistados em um *survey* sobre se eles aprovam ou desaprovam o desempenho do presidente.
 - b) A mensuração do conceito de corrupção política como o percentual de políticos em um país em um determinado ano que foram condenados por práticas corruptas.
 - c) A mensuração do conceito de democracia em cada país do mundo a partir da leitura da constituição, observando se nela há a afirmação de que o país é “democrático”.

Tabela 5.3 – Renda média dos cinquenta estados americanos, 2004-2005.

Estado	Valor	Posição	Valor
Alabama	37,502	Montana	36,202
Alaska	56,398	Nebraska	46,587
Arizona	45,279	Nevada	48,496
Arkansas	36,406	New Hampshire	57,85
California	51,312	New Jersey	60,246
Colorado	51,518	New Mexico	39,916
Connecticut	56,889	New York	46,659

Estado	Valor	Posição	Valor
Delaware	50,445	North Carolina	41,82
Florida	42,44	North Dakota	41,362
Georgia	44,14	Ohio	44,349
Hawaii	58,854	Oklahoma	39,292
Idaho	45,009	Oregon	43,262
Illinois	48,008	Pennsylvania	45,941
Indiana	43,091	Rhode Island	49,511
Iowa	45,671	South Carolina	40,107
Kansas	42,233	South Dakota	42,816
Kentucky	36,75	Tennessee	39,376
Louisiana	37,442	Texas	42,102
Maine	43,317	Utah	53,693
Maryland	59,762	Vermont	49,808
Massachusetts	54,888	Virginia	52,383
Michigan	44,801	Washington	51,119
Minnesota	56,098	West Virginia	35,467
Mississippi	34,396	Wisconsin	45,956
Missouri	43,266	Wyoming	45,817

Fonte: <<http://www.census.gov/hhes/www/income/income05/stateinhi2.html>>. Acesso em: 11 jan. 2007.

4. Faça o *download* de um livro de códigos de um banco de dados de ciência política pelo qual você tenha interesse.
 - a) Descreva o banco de dados e o propósito para o qual ele foi montado.
 - b) Quais são as dimensões física e temporal do banco de dados?

Leia os detalhes sobre como uma das variáveis que interessaram a você foi codificada e responda às seguintes perguntas:
 - c) Esse parece um método confiável de operacionalização dessa variável? Como a confiabilidade dessa operacionalização pode ser melhorada?
 - d) Avalie os vários elementos de validade para essa variável. Como a validade dessa operacionalização pode ser melhorada?
5. Se você ainda não fez o exercício 5 do capítulo 3, faça-o agora. Para a teoria que você desenvolveu, avalie o processo de mensuração das variáveis independente e dependente. Escreva sobre a confiabilidade e os vários aspectos da validade para cada uma das medidas. Você pode pensar em um melhor modo para operacionalizar essas variáveis para o teste de sua teoria?
6. *Coletando e descrevendo uma variável categórica.* Encontre dados para uma variável categórica pela qual você tenha interesse. Prepare um arquivo com os

dados de modo que possa ser lido por um programa de estatística. Produza uma tabela de frequência e descreva o que você observa.

7. *Coletando e descrevendo uma variável contínua.* Encontre dados para uma variável contínua pela qual você tenha interesse. Prepare um arquivo com os dados de modo que possa ser lido em um *software* estatístico. Produza uma tabela com estatísticas descritivas e um histograma ou um gráfico *kernel* de densidade. Descreva o que você encontrou após realizar esse procedimento.
8. Na Tabela 5.1, por que seria problemático calcular o valor médio da variável “identificação religiosa”?
9. *Da formulação matemática à afirmação textual.* Escreva uma sentença que sintetize o que cada uma das seguintes equações quer dizer:
 - a) $Y = 3 \forall X_i = 2$;
 - b) $Y_{\text{total}} = \sum_{i=1}^n Y_i = n\bar{Y}$
10. *Calculando médias e desvios-padrão.* A Tabela 5.3 contém a mediana da renda para cada um dos cinquenta estados americanos para os anos de 2004 e 2005. Qual é a média dessa distribuição e qual é o seu desvio-padrão? Escreva todos os cálculos.