

Mensuração*

Jorge de Souza

Departamento de Estatística, Universidade de Brasília

Rodrigo Andrés de Souza Peñaloza[†]

Departamento de Economia, Universidade de Brasília

março de 2005

1 Introdução

O ato de medir, tão presente e banal, é objeto aqui de nossas reflexões. É que ele assume conotações que, de um modo geral, transcendem o conhecimento comum das pessoas. Medir é, desse modo, algo que merece de nós uma séria reflexão. Assim, desde logo, ao pensarmos em medir algo cogitamos unicamente de resultados numéricos. Com efeito, o que fazemos quando precisamos medir a largura de uma sala? Usamos uma régua ou uma trena que são instrumentos de mensuração e que são estendidos ao longo dessa dimensão e permitem-nos dizer, por exemplo, que a referida largura vale quatro metros. Muitas vezes não existem instrumentos ou escalas que possam ser aplicados ao que intentamos medir. Do mesmo modo, devemos reconhecer ao longo de nossa leitura que também existem outras expressões de medidas que não sejam números. A questão é por demais complexa realmente e, para abordá-la competentemente, cumpre-nos examinar o problema da mensuração desde de uma perspectiva mais geral, afastando de nossa mente muitas idéias arquetípicas e equivocadas sobre mensuração.

Com efeito, quanto a medir, desde logo, devemos considerar três categorias de elementos essenciais que compõem esse ato: primeiramente, a classe formada pelos objetos que devem ser medidos; em segundo lugar, escolhemos uma propriedade, grandeza ou dimensão inerente a esses mesmos objetos e que constituirão aquilo que deverá ser medido e, finalmente, precisamos de um instrumento de mensuração.

É assim que o físico, por exemplo, aplica um termômetro ao líquido contido em um vaso para medir-lhe a temperatura. Nessa situação o objeto sob análise é o conjunto formado pelo líquido e o vaso; a propriedade ou grandeza que se deseja medir é a temperatura e o instrumento de

*Este texto é parte integrante do livro dos autores Estatística Exploratória, ainda em fase de preparação. Este capítulo, em particular, foi escrito inteiramente pelo primeiro autor, Jorge de Souza, com algumas poucas modificações menores feitas pelo segundo autor. Comentários são bem-vindos.

[†]E-mail: penaloza@unb.br

mensuração para que se alcance esse objetivo é o termômetro. O resultado, leitura ou medida da temperatura do objeto decorre da interação desses três elementos. Nesse caso, a medida ou leitura é obviamente um número real como “dez graus Celsius”.

O exemplo apresentado encontra-se de pleno acordo com a noção usual, corrente ou vulgar do ato de medir, ou seja, o ato de aplicar ao objeto, no processo de mensuração, um instrumento físico com escala numérica de leitura. Entretanto, convém pensarmos em que medir não se apresenta de modo tão simples assim. Ilustremos isso com outro exemplo. Temos um grão de café e desejamos classificá-lo para fins comerciais. O objeto, nesse caso, é obviamente o grão de café. A propriedade ou grandeza mensurável, entretanto, já não assume caráter numérico, mas é meramente tipológico, nominal ou classificatório, ou seja, é a categoria a que pertence o grão de café escolhida num elenco de várias outras. E o instrumento de mensuração, para este exemplo, qual será? Entendemos que, neste caso, pode ser um especialista ou provador de café que vai avaliar o tipo do grão e classificá-lo mediante uma conjectura pessoal inspirada em várias características como seu aspecto visual, sua cor, textura e, talvez, o seu paladar. O resultado dessa operação é também uma medição como, por exemplo, a classificação do grão na categoria conillon, um entre os muitos tipos de café. Neste caso já entendemos por escala de mensuração o conjunto formado pelos diversos tipos de grãos de café existentes.

Na realidade, percebemos que em ambos os exemplos tratados o instrumento de mensuração só pode ser usado mediante o conhecimento de um conjunto de regras de aplicação, de leitura e interpretação do ato de medir. Por outro lado, vemos que desde um ponto de vista mais abstrato o resultado da leitura ou medida tanto pode ser ou não um elemento de um conjunto numérico.

Este último caso nos ensina que uma medida nem sempre se expresa como um número, a não ser que o façamos de um ponto de vista meramente convencional. Se tal convenção for aceita, essa é a razão pela qual podemos conceituar do seguinte modo o ato de medir, tomando-o em sua acepção mais geral: medir é atribuir a certa propriedade de um objeto, dita a grandeza mensurável, mediante a aplicação de um conjunto de regras, dito instrumento de mensuração, um elemento de um conjunto numérico, chamado de conjunto de medidas possíveis ou escala de mensuração.

A idéia tão comum de que uma medida seja, necessariamente, um número é perfeitamente aceitável e pode ser justificada através de um exemplo. Suponhamos, para isso, que desejamos atribuir aos membros de uma comunidade as suas respectivas nacionalidades. Para tanto podemos codificá-los associando-lhes números identificadores. Desta forma, por exemplo, à nacionalidade brasileira atribuímos o número 0 (zero), à argentina o número 1 (um) e assim sucessivamente. Percebemos, por esse procedimento, que essa iniciativa de codificação é sempre possível de ser obtida em todo contexto, mediante a definição de uma correspondência biunívoca, neste caso tomada entre o conjunto de todas as nacionalidades possíveis ou escala de mensuração e um dado subconjunto de números reais, aqui representado pelos códigos 0, 1, 2, 3, ... Ao adotarmos essa convenção, é claro que podemos pensar sempre em que medir signifique a atribuição de um número ao estado inerente a uma propriedade, grandeza ou dimensão de um

objeto. Entretanto, devemos entender ser essa formulação aceita apenas desde um ponto de vista eminentemente formal, já que não podemos admitir, em casos tais como esse que ora ilustramos, operar aritmeticamente com as medidas numéricas daí resultantes, sob pena de cometermos erros imperdoáveis. Com efeito, que sentido teria, por exemplos, acharmos a média aritmética entre as medidas de três pessoas cujas nacionalidades sejam 0, 1 e 5? É claro que a média aritmética é, neste caso, igual a 2. No entanto, do ponto de vista das nacionalidades este valor médio nada representaria relativamente às três medidas efetuadas.

Este exemplo mostra-nos, de forma objetiva, a questão essencial que conduz a uma imprescindível tipificação de uma escala de mensuração segundo a sua admissibilidade quanto às operações aritméticas ou lógicas permitidas sobre seus elementos. Em decorrência disso, concluímos que só poderão ser usados, num dado tipo de escala, os métodos estatísticos que lhes sejam pertinentes, ou seja, ditadas pela natureza das medidas. Essa compreensão torna-se essencial, sobretudo para quem deve lidar competentemente com as ciências comportamentais.

Essas considerações também nos levam à imperiosa necessidade de examinarmos certos modelos padronizados de medidas possíveis ou escalas das propriedades mensuráveis dos objetos. Tais modelos, conforme já foi aqui conceituado, se presididos pelas regras de mensuração, vão determinar as possibilidades de operarmos aritmética e logicamente com os seus elementos ou medidas. Eles são chamados na literatura técnica de escalas de mensuração, escalas de medidas ou mais abreviadamente de escalas.

Resulta desses conceitos que podemos efetuar uma classificação não exaustiva das escalas, classificação ditada pela natureza das operações aritméticas ou lógicas permitidas sobre seus elementos. Já é tradicional, na literatura estatística, nesse sentido, a divisão das escalas de mensuração ou de expressão das propriedades dos objetos em quatro categorias ou modelos, chamados de escalas nominal, ordinal, intervalar e de razão.

Pode parecer, à primeira vista, para os leitores que já detêm algum conhecimento, que essas quatro escalas sejam exaustivas, isto é, que toda escala de mensuração deva necessariamente pertencer a uma dessas quatro categorias. Isso, entretanto, não é exato, pois, por exemplo, podemos defrontar escalas que, apenas parcialmente, satisfaçam às propriedades definidoras de cada um dos quatro tipos anunciados. Ao longo deste capítulo, entretanto, esses temas serão devidamente aclarados. Antes de fazê-lo, porém, cumpre-nos abordar um outro conceito umbelicalmente ligado às escalas de mensuração e freqüentemente usado na pesquisa científica. Trata-se da noção de variável. Para compreendê-la em sua precisa dimensão, devemos recapitular tudo o que examinamos até agora a respeito da mensuração ou ato de medir. Assim é que medimos um objeto relativamente a uma dada propriedade obtendo, por resultado, um número ou medida abstrata sobre sua escala de mensuração. O conceito de variável nasce exatamente do ato de repetirmos essa operação de medir sobre os objetos de um conjunto, fixando-nos sempre na mesma propriedade. Desse modo, uma variável nada mais representa do que uma função que atribui a cada objeto de um conjunto a medida de uma mesma propriedade comum a todos. Variável é, pois nessas circunstâncias, o símbolo ou função que representa as medidas

dos objetos de um conjunto relativamente à propriedade escolhida.

Conforme o nome da escala em que expressem os valores de uma variável, ela pode ser chamada de variável nominal, ordinal, intervalar ou, ainda, variável de razão. Entretanto, muitas outras formas de classificação das variáveis podem ser usadas ao longo deste livro. Assim, podemos falar de variáveis contínuas em contraposição a variáveis discretas. Outras vezes distinguimos as variáveis pela dicotomia que as classifica como quantitativas ou qualitativas. Ambas as classificações aqui lembradas são de uso corrente na literatura estatística e devemos estar atentos a suas propriedades que, por sua vez, vão requerer um tratamento estatístico próprio, sempre de acordo com a natureza da escala. Deste modo, por exemplo, não tem sentido o cálculo da média aritmética ou da variância das medidas efetuadas numa escala nominal ou ordinal. Contrariamente, esses mesmos sumários estatísticos têm sentido quando lidamos com as escalas de medidas intervalares e de razão.

2 Construtos e instrumentos conceituais de mensuração

Para outros propósitos devemos considerar, agora sob novo ponto de vista, as variáveis ou as correspondentes grandezas. Com efeito, queremos fazer a importante distinção entre as variáveis que são diretamente observáveis daquelas outras, de natureza eminentemente conceitual e que só indiretamente podem ser conhecidas. São exemplos de variáveis diretamente observáveis o comprimento e a temperatura. Já a ideologia ou a inteligência de uma pessoa são exemplos de variáveis que só podemos medir de forma indireta. Estas são chamadas de construtos, palavra que nos remete a outra de mesma raiz – construção. Um construto é, assim, uma variável abstrata concebida, construída ou definida no plano mental e estudadas pelas manifestações concretas que lhes atribuímos.

As variáveis indiretamente observáveis ou construtos necessitam, no entanto, de uma conceituação que lhes dê operacionalidade, isto é, elas exigem uma definição que, ao mesmo tempo que as descreva, vincule-as a algo observável. Assim, por exemplo, a inteligência pode ser definida operacionalmente ou semanticamente como a capacidade de resolver problemas. Uma definição como essa é chamada de definição operacional ou definição semântica. É através dessa definição operacional que poderemos medir a inteligência de uma pessoa determinada por seu desempenho em uma prova ou teste de inteligência, isto é, submetendo-a a um conjunto de problemas cujas soluções lhes são solicitadas e que supomos sejam capazes de refletir ou representar o conceito de inteligência adotado.

Chamamos a atenção do leitor, aqui, para a natureza conceitual do instrumento de mensuração utilizado para medir as variáveis abstratas – também chamado de teste, questionário ou prova – em contraposição à natureza eminentemente física do instrumento de mensuração das variáveis diretamente observáveis.

Surgem, assim, em decorrência dessa distinção, duas novas questões que são relevantemente cruciais para o trabalho do estatístico. Referimo-nos aos conceitos de validade e de confiabilidade

dos instrumentos de mensuração. Antes de abordá-los, porém, devemos aclarar alguns outros pontos referentes à nomenclatura utilizada para designar as variáveis indiretamente observáveis. Elas também são chamadas na literatura técnica de variáveis latentes, de traços latentes ou ainda de fatores, todos esses termos de uso dos psicometristas, e de construtos, segundo a preferência expressada pelos sociólogos. A diferenciação dessa nomenclatura reflete os esforços autônomos que os dois tipos de profissionais despenderam ao tratarem do mesmo assunto sem o conhecimento mútuo. Cada um criando os imprescindíveis métodos estatísticos mais apropriados às suas ciências. Esta certamente não é uma postura eficaz. No entanto, foi assim que as coisas ocorreram.

De um modo geral, a mensuração de um construto ou, se quisermos agora, de uma variável latente, dá-se através de manifestações diretamente medidas e indicadas pela definição operacional do mesmo construto. Ocorre, assim, que quase sempre, essas variáveis representantes ou vicárias – também chamadas de indicadores pelos sociólogos, de testes pelos psicometristas e de proxies pelos econométricos – exibem aspectos parciais ou as muitas faces que compõem o construto em exame. Em razão disso, concluímos que o construto pode ser medido de várias e diferentes maneiras alternativas, conforme as faces escolhidas para representá-lo, e formam o conjunto mais amplo de todas as faces possíveis. Se este é o caso, falamos em indicadores ou fatores múltiplos para nos referirmos às variadas faces do mesmo construto por elas representado.

Ao considerarmos, agora, um instrumento físico de mensuração de uma variável diretamente observável ninguém pode pôr em dúvida que esse instrumento possa servir exclusivamente ao objetivo para o qual ele foi constituído. Desse modo, por exemplo, podemos dizer que um tensiômetro é um instrumento certo para medir a pressão sanguínea, enquanto uma trena só tem por propósito medir um comprimento. Ninguém ousaria mudar-lhes os usos para os quais cada um deles foi destinado. Entretanto, quando o instrumento de mensuração é conceitual, como nos casos das variáveis indiretamente observáveis ou latentes, já não podemos garantir que ele meça somente aquilo para o qual ele foi construído e não uma outra grandeza diversa daquela. Por exemplo, um teste de inteligência pode não medir, exclusivamente, esse construto pois ele é também capaz de medir outros conceitos abstratos ou construtos como o aprendizado ou a cultura. Com efeito, muitas vezes a capacidade de resolver problemas pode não somente refletir a inteligência, mas também o aprendizado ou um treinamento anteriormente adquirido. Esta multiplicidade de fins é que caracteriza o problema da validade ou da validez do instrumento de mensuração. Este, em resumo, consiste em responder à seguinte pergunta: “um certo instrumento é apropriado para medir exclusivamente uma dada variável?” Esta é uma questão que só tem sido superficialmente abordada e, apenas, nos textos de psicometria, mas que, de um modo geral, é desconhecida de outros profissionais, como os economistas, por exemplo.

Um outro conceito importante que se liga ao instrumento de mensuração é o da sua confiabilidade ou precisão. Nesse novo contexto o que examinamos é a fidedignidade das medidas por ele efetuadas sobre o mesmo objeto e relativamente à mesma propriedade. A confiabilidade de um instrumento de mensuração representa, assim, a sua capacidade de medir o estado de uma

grandeza com erro preferencialmente desprezível. Desse modo, quando confrontamos validade e confiabilidade devemos admitir que um instrumento de mensuração pode ser muito preciso ou confiável em suas medidas e pode, ao mesmo tempo, ser pouco válido para o fim a que se propõe. Contrariamente, ele também pode ser muito válido e pouco preciso. Em geral, os instrumentos físicos de mensuração são bem mais confiáveis que os seus co-irmãos, os instrumentos conceituais, isto é, as medidas proporcionadas pelos instrumentos físicos são mais fiéis na previsão dos verdadeiros valores que tentam medir do que as medidas obtidas por instrumentos conceituais de mensuração. Esta observação tem um significado estatístico prático muito importante, por causa dos erros em que incorremos ao medir os construtos. Assim, muitas vezes, equivocamo-nos por imaginar que o uso de métodos estatísticos analiticamente sofisticados possa superar essa dificuldade própria dos dados obtidos pela mensuração efetuada por instrumentos conceituais. Assim, tratamos do uso de instrumentos analíticos sofisticados sobre uma base de dados de má qualidade, resultante tanto da imprecisão com que foram obtidos quanto pela baixa validade do instrumento de mensuração que os produziu. Isto, evidentemente, não nos conduz a bons resultados e por todos os modos deve ser evitado.

3 Mensuração de variáveis abstratas

Medir, nas ciências humanas, pode ser entendido já agora como o problema de determinar o grau de manifestação de um atributo mental, de uma faculdade perceptiva ou de uma face do comportamento social, psicológico ou político de um indivíduo. Todas essas manifestações, já sabemos, são construtos. O processo de mensuração nas ciências psicossociais é realizado, como vimos, através de um instrumento de mensuração ou medição representado por um teste, uma escala ou questionário. Neste livro, muitas vezes, usaremos o termo simplificado teste para designar o instrumento conceitual de mensuração psicossocial em lugar das expressões equivalentes ou alternativas como teste subjetivo, teste mental, escala, questionário, etc. Esta é, assim, uma mera escolha de termos convencionais, uma opção que fazemos pela palavra teste.

Um teste, por sua vez, é formado por questões ou itens mais simples e que vão representar as diversas faces ou dimensões com que o construto pode apresentar-se ao nosso entendimento. A existência dessas faces, conforme já foi dito, decorre da conexão usada na definição semântica ou operacional desse mesmo construto. O caso da inteligência humana ilustra muito isto que acabamos de afirmar. Com efeito, se definirmos inteligência pelo conceito operacional que a traduz como a capacidade de resolver problemas, somos capazes de identificar as várias faces ou dimensões da inteligência. Com efeito, evidentemente ela possui uma face lógica, uma face de percepção espacial, uma face de memória e assim sucessivamente. Desta forma, um teste de inteligência deve conter itens ou quesitos pertinentes a todas as faces ou dimensões que, a priori, admitimos como suas componentes. Esses mesmos itens, por sua vez, quando aplicamos o teste, funcionam como estímulos psicológicos que vão produzir respostas concretas manifestadas pelo indivíduo, respondente ou examinando que a ele se submete. As questões ou itens componentes

de um teste, por outro lado, devem ser estatisticamente associadas ao atributo mental, à faculdade perceptiva ou à atitude manifestada pelo respondente. Não cremos ser desprezível dizer que, por exemplo, se desejamos realizar um estudo sociológico de comportamentos e atitudes, as predisposições mentais para agir constituem as faces que devem representar o comportamento examinado.

Entendemos que a qualidade das respostas ou das observações estatísticas produzidas por um instrumento de mensuração, em qualquer domínio do conhecimento, tanto pode ser julgada sob a ótica do uso a que elas se destinam quanto, ainda, por considerarmos a fidelidade do instrumento de mensuração. Segue-se, como decorrência do que já é sabido, que os conceitos de validade e confiabilidade podem ser descritos, sucintamente, da seguinte maneira:

- (i) a validade é a adequação do instrumento ou, em última instância, a pertinência dos resultados ou medidas por ele produzidos para medir aquilo que se pretende;
- (ii) a confiabilidade, por sua vez, é a consistência ou a estabilidade estatística das medidas quando estas são obtidas por repetição e mantidas imutáveis as condições sob as quais são geradas.

A validade de um instrumento de mensuração tanto pode ser apreciada em seu aspecto teórico quanto desde um ponto de vista eminentemente prático. Assim, caso as medidas produzidas pelo ato de medir, através de um teste, sejam apreciadas segundo um contexto teórico, para o exame de um certo construto, dizemos que o teste tem validade formal, teórica, lógica ou, ainda, que ele tem validade de conteúdo. De outra feita, se essas mesmas medidas são agora apreciadas mediante um critério objetivo que afirma sua adequação para o exame dos propósitos para os quais o teste foi produzido, dizemos que o instrumento de mensuração tem validade experimental. Decorre dessas considerações que a validade formal de um teste é uma validade apriorística, é teórica, de opinião ou de juízo, enquanto a validade experimental é de natureza empírica, afirmada no uso que se dá às medidas produzidas pelo teste.

A validade prática ou experimental também exige, para o seu pleno entendimento, a definição de um critério ou padrão externo de avaliação, ou seja, a escolha de uma variável ou um outro teste que possa manifestar uma associação estatística com as medidas originais examinadas: o grau de associação estatística assim determinado configurará o nível de validade empírica do instrumento de mensuração.

Esses mesmos conceitos, agora reportando-nos à confiabilidade, autorizam-nos a afirmar que ela diz respeito à precisão ou fidedignidade das medidas produzidas pelo instrumento de mensuração, enquanto a validade refere-se ao uso que se pretende dar às mesmas medidas. Por exemplo, a trena utilizada por um carpinteiro é um instrumento menos confiável que o paquímetro utilizado por um físico, se o problema comum é medir o diâmetro de um parafuso. A confiabilidade é, desse modo, restrita ao instrumento de mensuração, supondo mantidas as condições exteriores nas várias medições efetuadas. A validade, entretanto, é de outra natureza, pois seu problema é mais inerente às ciências psicossociais sem que essa restrição, evidentemente, implique em

exclusividade. Com efeito, o problema da validade está presente sempre que desejamos medir uma dimensão latente da realidade. Por exemplo, o economista que objetiva medir a riqueza de uma nação, que é um conceito abstrato ou construto econômico, usa como instrumento de mensuração a renda nacional, que é um conceito empírico, ou a própria definição semântica ou operacional da riqueza, segundo a teoria econômica vigente. Esse instrumento de mensuração só é válido para o propósito estipulado desde que ele consiga medir, convenientemente, aquilo que se entende por riqueza. Neste exemplo, se o instrumento para medir a riqueza fosse o número de automóveis produzidos na economia, ele seria menos válido que a renda nacional porque, desde logo, há nações notoriamente mais ricas que produzem menos automóveis que outras nações evidentemente mais pobres. Essa contradição labora a desfavor do instrumento representado pela produção de automóveis e ela já não se dá com a renda nacional.

Freqüentemente, ao pensarmos em instrumentos de mensuração, por um mecanismo inconsciente de condicionamento cultural, imaginamos tais instrumentos como objetos concretos como um metro ou uma balança. Essa idéia, como já sabemos, é insubsistente nas ciências psicossociais porque, por exemplo, não se pode construir – e até agora não o fizemos – um inteligentômetro, isto é, um aparelho físico que meça diretamente a inteligência das pessoas. É claro que, mesmo nas chamadas ciências naturais, os instrumentos de mensuração por elas utilizados podem ser semelhantes à maioria dos instrumentos de medição encontrados nas ciências psicossociais. Com efeito, admitamos, ainda, que desejamos medir o comprimento de uma mesa. Em lugar de usarmos uma trena como um instrumento natural e mais apropriado de medição, poderíamos optar por ouvir a opinião de várias pessoas sobre suas avaliações do comprimento da mesa tomando por base a um referencial ou padrão e, a partir daí, definindo a medida do comprimento da mesa pela média aritmética das avaliações individuais. Neste caso tomamos como referência, por exemplo, uma vareta e pedimos às pessoas que conjecturem ou opinem sobre o valor do comprimento da mesma dizendo quantas vezes, aí incluída uma fração, o comprimento da mesa expressa-se pela vareta. Assim procedendo teríamos resultados do tipo “4,00”, “4,20”, “4,80”, etc. É claro que esse procedimento, de caráter eminentemente opinativo, é trabalhoso e pouco confiável porque, ao repeti-lo em circunstâncias idênticas, a medida obtida será sempre outra, pois diversas deverão ser, também, as medidas das avaliações individuais. O comprimento da mesa, além do mais, é uma dimensão palpável da realidade, está ao alcance dos sentidos e, por isso mesmo, torna-se dispensável, e até mesmo inconseqüente ousadia, a mensuração mediante o uso de um instrumento baseado no subjetivismo das opiniões. Quem o faria, se dispomos de trenas?

Assim, concluímos ser bastante complexo o problema de construção de instrumentos de mensuração no domínio das ciências psicossociais. Ainda em apoio às novas idéias, suponhamos agora que queremos medir o grau de liberalidade política dos membros do Congresso Nacional. O conceito de liberalidade política é obviamente abstrato ou latente e dele só conhecemos certas manifestações que, mediante consenso, podem ser atribuídas à liberalidade política, ou seja, mediante a escolha de uma definição operacional que registre as ações políticas tidas consen-

sualmente como liberais. É por isso que, por exemplo, um parlamentar favorável ao voto dos analfabetos deve ser tido como politicamente mais liberal que outro que seja contrário a essa proposição. De maneira semelhante, o que se opusesse à legalização de um partido extremista seria politicamente menos liberal que aquele que não admitisse a atividade normal desse partido, supondo respeitador da ordem pública, evidentemente.

O leitor, com certeza, deve estar bastante inquieto com essa argumentação. É que, de um lado, ela lhe abre uma perspectiva imensa de pesquisa nos domínios da política e, de outro, porque as questões apresentadas como faces de liberalidade política também podem, em realidade, caracterizar uma outra variável abstrata ou latente, aqui ironicamente chamada de irresponsabilidade política. Desse modo é que surge ante nossos olhos a questão já conceituada da validade do instrumento de mensuração.

Podemos ainda, no exemplo em tela, muito bem admitir este último ponto de vista, da mesma maneira que o primeiro. O leitor não seria volúvel por isso, mas compreenderia que é muito complexa a escolha das questões que devem compor o instrumento de mensuração da variável latente ou, o que é de certo modo equivalente, a escolha da definição operacional. Existem algumas técnicas que facilitam esse estudo, como a análise de conteúdo do discurso, e é esse um dos assuntos deste palpitante livro.

Após tantas novidades, antes impensáveis em um curso de estatística, confessemos já se compreende que em um teste podem ser identificados dois elementos básicos: um, o seu antecedente ou estímulo; outro, o conseqüente, resposta ou comportamento. Decorre disso que um teste é, deste novo ponto de vista, um proceso de mensuração que, a partir de um estímulo, induz uma resposta por parte do indivíduo, respondente ou examinando que a ele se submete. Face a essa nova forma de vê-lo entendemos, perfeitamente, que as condições psicológicas ou ambientais sob as quais um indivíduo responde a um teste são temporariamente mutáveis e, de igual modo, variam de uma pessoa a outra. E, mais que tudo, essas variações influenciam os resultados ou medidas.

No que diz respeito à variabilidade temporal das medidas devemos admitir que os resultados de um teste a que responde um indivíduo, em vários instantes, não devem ser muito diferentes sob certas condições de estabilidade. As flutuações possíveis nesses resultados deverão ser atribuídas, preferencialmente, a fatores ocasionais, de natureza aleatória e, por essa razão mesma, são incontroláveis para o analista. Decorre daí que de um teste devemos exigir uma baixa variabilidade em suas medidas obtidas do mesmo indivíduo, em ocasiões diferentes, supondo estáveis as condições internas ou externas.

É evidente, também, sobre isso, que existe uma gradação notória e, até certo ponto previsível e intuitiva, no que diz respeito à confiabilidade ou precisão inerentes a certas classes de testes. Nessa circunstância, é fácil concluir que um teste de inteligência é mais confiável que um teste de conhecimentos, sendo este, por sua vez, mais confiável que um teste de personalidade. Isto quer dizer, exatamente, que as variações possíveis nos resultados de múltiplas aplicações de um teste de inteligência têm, em princípio, menor variabilidade ou mais estabilidade estatística que

as variações dos resultados de um teste de conhecimentos.

É fácil perceber, também, referindo-nos agora à validade, que pode haver um divórcio notável entre o conteúdo do teste e o objetivo para o qual ele foi proposto. Nesse caso, dizemos que sua validade de conteúdo é baixa e, contrariamente, havendo grande sintonia entre o teste e o fim a que ele se destina, afirmamos a sua alta validade teórica. Depreendemos daí, por conseguinte, que a validade de um teste não pode somente ser apreciada no teste em si mesmo mas, também, na sua aplicação, isto é, nas inter-relações entre o teste e o seu fim último.

Como o leitor já deve ter percebido, a validade é uma questão ligada essencialmente aos instrumentos conceituais de mensuração ou testes, já que os instrumentos físicos, por sua própria natureza, não requerem o exame da finalidade para a qual eles foram concebidos.

A questão da validade de um teste, como já vimos anteriormente, pode, em princípio, ser apreciada de dois modos diversos: os modos de validade objetiva e de validade conceitual. A primeira forma diz respeito à capacidade do teste para o propósito de previsão relativamente a algum estado mental, comportamento ou desempenho. Dizemos, nesse caso, que a validade é preditiva ou corrente, conforme ela se refira, respectivamente, ao futuro ou ao presente. Nessas circunstâncias, de um teste vocacional exigimos que tenha validade preditiva e, em contrapartida, um teste de personalidade deve possuir validade corrente. Da mesma maneira, concebemos que uma escala de atitudes deve possuir mais validade corrente do que preditiva, isto é, ela deve ser capaz de prever comportamentos atuais mais do que futuros.

Quanto à conceituação de categoria de validade conceitual, havemos de entendê-la através da capacidade do teste para medir o construto ou traço latente objeto de sua concepção, ou ainda, de qualquer outro traço que possa, porventura, manifestar-se de seus itens. Dessa maneira, concluímos haver duas formas diferentes de conceber a validade conceitual:

- (i) a da validade de conteúdo, que responde à questão referente à representatividade dos itens do teste enquanto uma amostra das faces do construto que lhe inspira a construção;
- (ii) a da validade de construto, que responde à questão referente à diversidade ou espectro de construtos que o teste pode medir e que caracteriza a sua heterogeneidade, multifatorialidade ou multidimensionalidade.

A primeira dessas formas de validade conceitual, ou seja, a da validade de conteúdo, pode ser entendida considerando o domínio de conteúdo do construto que o teste intenta medir. Assim, a validade de conteúdo implica na análise e na medida da pertinência dos itens para representar o construto em tela por suas diversas faces. A outra forma de validade conceitual, a de construto, expressa a natural fragilidade do instrumento conceitual de mensuração por constatarmos que uma amostra de itens do domínio de conteúdo de um construto também pode ser uma amostra do domínio de conteúdo de outro construto, distinto do primeiro, como ocorre no caso já examinado da inteligência.

A necessidade de examinarmos a validade conceitual de um teste é decorrência da natural inacessibilidade direta do construto. Entretanto, através de uma teoria apropriada, podemos

saber alguma coisa relativa à observação de suas manifestações. Desse modo é possível ter uma noção do domínio de conteúdo do construto, de como os examinandos o percebem e das possíveis influências que podem alterar essa percepção pelos mesmos respondentes.

A validade objetiva, por outro lado, em qualquer de suas manifestações, atual ou futura, necessita de um ou vários critérios objetivos Y_1, Y_2, \dots, Y_k para referendar a medida X produzida pelo teste. Decorre desse fato a imprescindibilidade de conhecermos, de antemão, os seguintes elementos necessários para caracterizar a validade objetiva ou empírica:

- a identificação dos critérios empíricos de avaliação Y_1, Y_2, \dots, Y_k ;
- a seleção de uma amostra de examinandos sobre a qual vamos observar a medida X do teste e as correspondentes medidas Y_1, Y_2, \dots, Y_k dos diversos critérios empíricos;
- a escolha de uma medida de associação estatística $a(X, Y_1, Y_2, \dots, Y_k)$ que meça o grau de previsibilidade de X a partir dos critérios exteriores Y_1, Y_2, \dots, Y_k .

O valor absoluto $|a(X, Y_1, Y_2, \dots, Y_k)|$ dessa medida de associação estatística é chamado de coeficiente de validade empírica ou simplesmente coeficiente de validade do teste. Frequentemente usamos um único critério Y predictor e, no caso onde as medidas X e Y expressam-se em escala intervalar ou de razão, o coeficiente de validade $|a(X, Y)|$ é tomado como o valor absoluto do coeficiente de correlação linear de Pearson. Em outras situações, também muito comuns, o critério de previsão Y é dicotômico ou mesmo classificatório com mais de duas modalidades. Nesses casos, usamos para calcular o coeficiente de validade um dos múltiplos coeficientes de associação estatística disponíveis na literatura técnica.

Também é necessário termos em conta, no concernente à determinação do coeficiente de validade, a existência de certas escolhas que podem mascarar os resultados quando não levadas em consideração. Nesse contexto podem ser revelados os seguintes problemas, todos eles afetando ponderavelmente o valor do coeficiente de validade:

- a escolha inadequada da amostra de examinandos, que introduz vieses nas medidas do teste e dos critérios empíricos usados;
- as baixas confiabilidades do teste e dos critérios empíricos escolhidos.

No caso de um único critério Y e desde que X e Y estejam expressados em escala intervalar ou de razão, por exemplo, podemos usar a conhecida fórmula de atenuação com vistas à eliminação do viés exercido sobre o coeficiente de correlação de Pearson pelas inconfiabilidades do teste e do critério.

Abordando um teste T desde um outro ponto de vista e abstraindo-nos daquilo que foi comentado anteriormente, um teste é visto como um conjunto finito $T = \{Q_1, Q_2, \dots, Q_n\}$ constituído por n questões ou itens $Q_i, i = 1, 2, \dots, n$, onde supomos que cada uma dessas questões Q_i possa ter uma resposta no conjunto $R = \{R_1, R_2, \dots, R_m\}$. O número n de questões

que compõem o teste T é o seu comprimento. O conjunto de respostas a tais itens pode tanto ser formado por categorias ou modalidades quanto por respostas ordenadas ou por números reais que expressem notas ou scores, isto é, as respostas podem ser variáveis representadas em qualquer escala de mensuração.

Dado um teste T e um indivíduo ou examinando i , escolhido imparcialmente numa população P que a ele se submete, a seqüência de variáveis aleatórias (X_1, X_2, \dots, X_n) de respostas dadas por i às questões de T é chamada de padrão de respostas de i no teste T e a soma $X = X_1 + X_2 + \dots + X_n$, quando possível calcular-se, é chamada de nota ou score bruto do mesmo indivíduo no teste T .

São muitas as medidas que podem ser associadas a um teste quando aplicado a um indivíduo. Podemos medir, por exemplo, o tempo que ele leva para responder ao teste; podemos contar, quando for o caso, o número de respostas corretas por ele dadas e assim sucessivamente. Essas diferentes medidas, pelas razões expostas anteriormente, são consideradas variáveis aleatórias, pois o examinado foi escolhido segundo uma dada lei de probabilidade.

Consideremos, agora, um conjunto P de pessoas ou respondentes que se submetem a um teste T e representemos por X o score bruto de um membro de P , escolhido imparcialmente entre as N pessoas que compõem a população P .

Em tais circunstâncias, o score bruto X do teste T é uma variável aleatória real que se sujeita a influências incontroláveis, de caráter interno e externo ao indivíduo escolhido.

A teoria estatística clássica dos testes, formulada por Charles Spearman, admite que o score bruto X seja o resultado da adição de duas parcelas: uma, V , designada como score verdadeiro; outra, ε , chamada de erro. Desse modo, o score bruto X , obtido no teste T , é representado pela identidade $X = V + \varepsilon$.

O erro aleatório ε reflete todas as possíveis perturbações estocásticas que podem acentuar a diferença entre o score verdadeiro V e o score bruto X que o indivíduo obteria no teste T , caso esses mesmos fatores deixassem de agir. Tais fatores aleatórios originam-se nas flutuações ambientais e resultam, igualmente, da própria instabilidade das condições subjetivas naturais do indivíduo. Por exemplo, ao aplicar-se um teste de inteligência a uma pessoa febril, este fator ocasional há de afetar, de algum modo, o seu score final no teste. Esta é uma causa subjetiva. De igual maneira, o ruído ambiental é uma causa exterior que contribui para ressaltar a diferenciação entre os scores brutos e verdadeiro.

Impõem-se adicionalmente à definição anterior algumas condições extras de profundo significado empírico e necessárias à formulação teórica subsequente. Essas propriedades, de natureza estatística, são as seguintes:

- (a) $E[\varepsilon] = 0$, ou seja, o erro no teste T tem esperança matemática nula;
- (b) $\text{corr}[V, \varepsilon] = 0$, ou seja, é nula a correlação entre o score verdadeiro e o erro de um teste.

O primeiro desses axiomas apenas revela que o erro não contém qualquer componente sistemática porque, se por absurdo fosse $E[\varepsilon] = a \neq 0$, então a parcela a seria parte integrante do

score verdadeiro V , que, nessas circunstâncias, seria igual a $a + V$.

Ao quociente:

$$\rho^2 = \frac{Var(V)}{Var(X)}$$

entre as variâncias do score verdadeiro V e do score bruto X damos o nome de coeficiente de confiabilidade do teste T . Assim, compreendemos que quanto maior for o valor do coeficiente ρ^2 da confiabilidade de um teste, maior será a precisão das medidas por ele produzidas.

4 Escalas de mensuração

4.1 Escala nominal

Uma escala de mensuração é chamada de nominal, categórica ou classificatória – e seus elementos são chamados de categorias ou modalidades – quando as medidas dos objetos que lhes correspondem cumprem, necessariamente, as duas propriedades seguintes:

- (i) a cada objeto atribui-se uma única modalidade;
- (ii) dois objetos com a mesma modalidade são tidos como equivalentes ou pertencentes ao mesmo agrupamento.

Ilustremos com um exemplo. O conjunto de modalidades {masculino, feminino} é uma escala nominal para o variável sexo das pessoas porque:

- (i) a cada pessoa atribui-se um sexo ou uma categoria ou modalidade nessa escala;
- (ii) se duas pessoas são masculinas (femininas) elas têm o mesmo sexo, ou seja, são equivalentes por pertencerem à mesma categoria sexual ou agrupamento.

As modalidades dessa escala poderiam ser numeradas como 0 (zero), para a categoria masculino e 1 (um), para a categoria feminino. Esses números obviamente têm, na escala nominal, apenas propriedades classificatórias para o efeito de considerar duas pessoas como pertencentes ao mesmo sexo ou tendo sexos diferentes. Nenhuma operação aritmética tem sentido com os elementos dessa escala. Por exemplo, se três pessoas têm modalidades 1, 0 e 1 nessa escala, ou seja, se elas são mulher, homem e mulher, respectivamente, nenhum sentido existe em obter-se a sua média aritmética, por ser ela totalmente desprovida de significado interpretativo. Para que o leitor compreenda melhor o que se afirma, suponhamos que no lugar do sexo se atribua às pessoas as suas rendas em dado mês do ano. Neste caso, sendo x_1 , x_2 e x_3 , respectivamente, os valores de suas rendas, a soma $x_1 + x_2 + x_3$ tem o sentido óbvio de significar a renda grupal ou conjunta das três pessoas consideradas.

Ainda para afirmarmos a impossibilidade de quaisquer outros significados das modalidades 1 (um) e 0 (zero) relativas à determinação do sexo das pessoas e que não sejam as da classificação

nessas duas categorias, frisamos a falta de sentido de outras relações formais como, por exemplo, a relação menor do que: $0 < 1$ (0 (zero) menor que 1 (um)). Com efeito, que significado teria a afirmação correspondente de que o masculino é inferior ao feminino? Pois não é esse, no caso, o significado aritmético da relação $0 < 1$?

A construção de uma escala nominal requer, na prática estatística, que escolhamos antecipadamente um determinado número de categorias para expressar todas as possibilidades de medidas e, é claro, exige a nossa competência para identificar as medidas dos objetos sobre a escala definida pelas diversas modalidades adotadas. Por exemplo, a maturidade das pessoas, vista aqui como a expressão de suas experiências biopsicológicas pode ser medida sobre a escala nominal definida pelas modalidades criança, jovem, adulto e idoso. Desse modo, o conjunto formado por essas categorias em que se pode expressar, por convenção, a maturidade humana {criança, jovem, adulto, idoso}, constitui uma escala nominal para essa propriedade ou grandeza. O estatístico deve ser capaz, por outro lado, de classificar cada pessoa em uma única dessas categorias, de modo inequívoco, através do enunciado de critérios precisos de classificação ou, ainda, em outras palavras, escolhendo uma definição operacional apropriada para a variável conceitual maturidade. Poderia ser, neste caso, uma classificação expressada por faixas etárias. Compreendemos, no entanto, que qualquer que seja a escolha efetuada sempre, sobre ela, teremos argumentos favoráveis ou contrários.

4.2 Escala ordinal

São estas escalas ordinais uma forma mais rica de mensuração de uma grandeza do que aquela proporcionada pela escala nominal. Esta escala ordinal caracteriza-se pelas três propriedades seguintes:

1. A cada objeto atribui-se um único valor na escala.
2. Existe, na escala, uma relação de ordem simbolizada por \prec (menor do que, inferior a, precede a, etc) de modo que sendo A e B dois objetos quaisquer, com medidas respectivamente iguais a x_A e x_B e sendo $x_A \neq x_B$, só vale uma das seguintes afirmações:
 - (a) $x_A \prec x_B$, equivalente a dizer que o objeto A precede o objeto B ou que o objeto A mede menos que o objeto B ;
 - (b) $x_B \prec x_A$ significando que o objeto B mede menos que o objeto A .

Esta propriedade é dita ser antissimétrica para a relação de ordem \prec ;

3. Sendo A , B e C três objetos quaisquer e se suas respectivas medidas x_A , x_B e x_C satisfazem às duas relações $x_A \prec x_B$ e $x_B \prec x_C$, então, necessariamente, vale a relação transitiva $x_A \prec x_C$. Esta propriedade é chamada de transitiva para a relação \prec .

Ao leitor pode parecer mero preciosismo a necessidade de impormos a propriedade transitiva para a relação de ordem inerente a uma escala ordinal. A esse respeito, no entanto, nunca é demais lembrar a existência de relações binárias nas ciências comportamentais (isto é, de relações que comparam os elementos constitutivos de pares de elementos de um conjunto) e que não cumprem, necessariamente, a propriedade citada. Com efeito, admitamos que numa eleição com três candidatos A , B e C um eleitor, solicitando a estabelecer as suas preferências entre os três pares de candidatos (A, B) , (A, C) e (B, C) , manifeste as seguintes opiniões:

$$\left\{ \begin{array}{l} \text{Prefere } A \text{ a } B \\ \text{Prefere } B \text{ a } C \\ \text{Prefere } C \text{ a } A \end{array} \right.$$

Representando pelo símbolo $X \succ Y$ a preferência por X em detrimento de Y , as manifestações desse eleitor podem ser simbolizadas, respectivamente, como $A \succ B$, $B \succ C$ e $C \succ A$. Podemos ver, facilmente, a partir deste caso, que a propriedade transitiva, caracterizada aqui pela preferência $A \succ C$, não foi nele respeitada.

Não é difícil notar, também, que toda escala ordinal é, necessariamente, e antes de mais nada, uma escala nominal. Com efeito, na escala ordinal todo objeto tem por medida uma única modalidade nominal e, além disso, dois objetos com a mesma modalidade ordinal são equivalentes por pertencerem à mesma categoria ou classe.

Resulta dessas considerações que a diferença de uma escala ordinal para uma escala nominal está em que, na primeira, os objetos são ordenáveis segundo a intensidade com que eles manifestam a propriedade medida, enquanto na segunda eles são vistos, apenas, segundo a ótica mais pobre de serem meramente iguais ou diferentes. Desse modo, a relação de ordem (menor do que, precede a, mais jovem que, mais grave do que, etc) possibilita uma comparação elementar dos graus ou de conteúdo da propriedade ou grandeza, construto ou traço em exame. Isto já não ocorre, conforme vimos, na escala nominal.

Dispomos de uma escala ordinal, por exemplo, quando desejamos medir o grau de apoio a um governo. Nesse caso podemos solicitar aos membros de uma amostra escolhida que optem por uma das seguintes categorias que denotam o seu grau de avaliação do governo: CONTRÁRIO, INDIFERENTE e A FAVOR. O grau de apoio de quem é “contrário” ao governo é, evidentemente, inferior ao de quem é “indiferente” que, por sua vez, é menor do de quem é “a favor”.

Ocorre, muitas vezes, nesse contexto de pesquisa política, a necessidade de definir a nova categoria NÃO TENHO OPINIÃO FORMADA OU NÃO SEI para que todas as pessoas, sem distinção, possam expressar suas medidas quanto ao possível apoio. Nesse caso, enquanto a primeira escala tricategórica {contrário, indiferente, a favor} é obviamente ordinal, a última escala {contrário, indiferente, a favor, não tenho opinião formada} já não o é, pela impossibilidade de compararmos, com as demais, a nova categoria introduzida. Desse modo, para sermos rigorosos, deveríamos vê-la como uma escala nominal ou, quando muito, como a uma escala intermediária entre a nominal e a ordinal.

Devemos ter em conta, ainda, no concernente a uma escala ordinal, que não podem ser comparados os intervalos ou variações entre os diversos níveis de medidas proporcionados por essa escala. Assim é que, referindo-nos ainda ao último exemplo, nada podemos afirmar sobre a variação da grandeza do grau de apoio ao governo referente às passagens da categoria “contrário” a “indiferente” e relativamente à passagem desta última para a categoria “a favor”.

O que sabemos, unicamente, nesse caso, é que o conteúdo de apoio é gradativo e crescente com passagem da categoria “contrário” a “indiferente” e desta para a “favorabilidade”. Do mesmo modo, nessa escala, não conhecemos o seu ponto zero, no sentido de significar a ausência de qualquer conteúdo do traço ou construto “apoio ao governo”. Com efeito, ponhamos o caso do cidadão que não é totalmente contrário ao governo e que tampouco se encontre na situação mais amena da indiferença. Nessas circunstâncias, seria mais óbvio que ele fosse classificado entre os contrários ou na categoria dos indecisos, apesar de possuir algum grau de favorabilidade. Essa é a mesma situação que notamos, em outro exemplo, no concernente à medição da temperatura de um corpo na escala Celsius. Não passa por nossa mente que um corpo a zero grau Celsius não possua calor. É que, nesse contexto, o nível zero da escala é apenas convencional, não implicando tal fato, de nenhum modo, a ausência de calor. Esse zero é tão só uma marca convencional a partir da qual temos temperaturas positivas, de um lado e, negativas, do outro.

Decorre, assim, dessas considerações, que na escala ordinal só valem as comparações de ordem entre os números que sempre podem, por codificação meramente, frisemos, expressar as suas categorias. Nenhuma operação aritmética (adição, subtração, multiplicação ou divisão) com as medidas obtidas faz sentido nessa escala mas valem, nesse caso, as relações de ordem.

Algumas vezes a categorização de uma variável nominal apresenta-se de modo intuitivo e quase natural. É o caso, por exemplo, da já conhecida variável sexo. Diante desse exemplo iremos categorizá-la, de um modo simples, nas tradicionais modalidades masculino e feminino. Se bem que para esse exemplo ainda possamos questionar a dificuldade de classificação binária de todas as pessoas de acordo com as teorias psicológicas de gênero, podemos admitir que a definição do construto sexo seja a de caráter morfológico e não a de cunho funcional. Esse esclarecimento da definição semântica da variável sexo se melhora, não resolve de todo a dificuldade com que algumas vezes nos defrontamos para medi-la. Uma outra solução, capaz de superar esse impasse consiste em, comodisticamente, incorporar a categoria “outros” às demais categorias da escala. Ela, evidentemente, revela informações relevantes concernentes à sombra definicional que paira sobre a tradicional categorização dicotômica masculino versus feminino. Entretanto, em outros contextos como o das variáveis ordinais, essa solução leva, conforme já vimos, a uma descontinuidade na natureza algébrica da escala ordinal do tipo {ótimo, bom, regular, ruim e péssimo}. Por outro lado, no que diz respeito a essa mensuração, é comum encontrarmos alguém que não tenha opinião formada e, por isso mesmo, não a manifeste ou, provavelmente, alguém que, simplesmente, recuse-se a responder ao inquérito por um motivo qualquer. Ao tentar solucionar a dificuldade por acrescentarmos à escala ordinal a categoria “outras” ou “outras avaliações”, perdemos a natureza ordinal da escala.

Todas essas observações chamam a atenção do leitor para a complexidade do problema, aparentemente simplório, de tratar as variáveis nominais e ordinais. Antes de tecer outras considerações referentes ao assunto, entretanto, importa enunciarmos alguns princípios que devem ser seguidos quando da definição de uma variável categórica ou ordinal.

Princípio I: As categorias da variável devem ser bem definidas, exaustivas e mutuamente exclusivas, isto é, todo os objetos submetidos à avaliação devem pertencer a uma e única categoria e nenhum objeto deve deixar de ser medido.

No que respeita à “boa definição”, queremos dizer que não podem pairar dúvidas quanto à definição semântica ou operacional do construto medido ou da variável observada. Assim, por exemplo, referindo-nos ao sexo devemos tornar claro se o construto sexo diz respeito às suas manifestações morfológica, funcional, psicológica ou social. Aclarado esse aspecto, dizemos que a categorização está bem definida.

Ao afirmarmos, por outro lado, que as categorias escolhidas são exaustivas e mutuamente exclusivas precisamos, no que concerne a esta última propriedade, que as diversas categorias não se superponham umas às outras e, também, no que tange à exaustão, queremos dizer que todo objeto investigado há de, imprescindivelmente, pertencer a uma das diversas categorias adotadas.

Princípio II: Para obtermos um conteúdo informativo mais rico do construto, traço ou grandeza estudado, o número de categorias em que ele se expressa deve ser o maior possível porque enriquecemos com isso a compreensão quanto à natureza do construto.

Realmente, não é difícil percebermos que o construto apoio ao governo fica mais aclarado em sua natureza quando a ele acrescentamos outras categorias além da simples dicotomia contra e a favor. Tratando do construto profissão, por exemplo, ponhamos o leitor diante de duas escalas possíveis: uma, mais agregada, dada pela listagem {industrial, comercial, rural, liberal, servidor público, outros} e, outra, mais desagregada, definida pelo rol {conferente, engenheiro civil, estatístico, cantor, ...}. É claro, nesse último caso, que ao concebermos uma lista imensa de profissões temos algumas vantagens imediatas sobre a classificação mais agregada da primeira escala. De um lado, porque reduzimos os casos duvidosos de mensuração ou classificação que porventura existam na escala mais agregada e, de outro, porque obtemos uma informação mais pormenorizada sobre a distribuição de frequências das profissões com a segunda escala.

O problema da categorização ou de escolha de uma escala nominal assume, muitas vezes, um aspecto teórico bastante instigante e muito enriquecedor através do que denominamos continuum latente da variável nominal. Para compreendermos este último conceito suponhamos, ainda, o mesmo caso da escala nominal referente ao sexo. Seguindo as concepções da escola psicofísica alemã, o grande estatístico inglês Karl Pearson admitia que a categorização de uma variável desse tipo expressava, de forma resumida, uma avaliação mental latente que se efetuava no chamado

continuum mental subjacente a essa categorização. Assim, por exemplo, diante da dicotomia belo versus feio, Pearson supunha a existência de um continuum mental onde se expressaria a variável conceitual grau de beleza. Nesse caso, se a avaliação – latente, é bom frisar – do grau de beleza fosse inferior a um certo número, por ele chamado de limiar, a classificação anunciada seria na categoria feio; contrariamente, se essa medida fosse superior ao mesmo limiar a classificação dar-se-ia na categoria belo. Essa hipótese, ou teoria do continuum mental de Pearson, tornou-se não só a base de posteriores formulações da psicometria no domínio da denominada escalagem psicológica como, igualmente, funda-se nela a interpretabilidade das interações estatísticas examinadas por uma técnica estatística muito importante nos dias presentes, os chamados modelos loglineares e a análise de correspondências de tábuas de contingências. Sem discutir a plausibilidade científica ou filosófica, nela inserida, a idéia de continuum subjacente a uma categorização pode, apenas, aclarar o seu conteúdo mas não supera, em absoluto, as dificuldades que lhe são inerentes. Esse é o caso, por exemplo, das categorias contíguas ou vizinhas em uma escala ordinal ou, ainda, no concernente a uma escala nominal definida por classes ou intervalos contínuos, onde sobrepaira o grave problema de definição dos seus limites. Em ambas as situações incorremos em erros incontroláveis e inevitáveis de classificação.

Já com a escala ordinal, como vimos, torna-se possível não só uma ordenação das suas categorias – e essa é a sua essência – como, de igual modo, podem ser ordenados, por indução natural, os objetos segundo ela avaliados. Com efeito, os objetos podem ser ordenados segundo as suas determinações ou medidas obtidas na escala ordinal e assim, se x_i e x_j são as modalidades dos objetos i e j e se $x_i < x_j$, então dizemos que $i < j$, ou seja, o objeto i antecede ao objeto j , porque a modalidade correspondente x_i precede a outra, x_j .

O problema maior da escala ordinal, como já observamos, está na sua impossibilidade de comparar as variações de intensidade entre as suas diversas modalidades e essa limitação implica em uma outra, de muito interesse prático, e que pode ser assim descrita:

“Sendo x , y e z três categorias contíguas e sucessivas da escala ordinal, consideremos três objetos i , j e k que, respectivamente, manifestam-nas. Como as modalidades da escala significam o quantum da grandeza medida, perguntamos se são iguais as suas variações? Quando a resposta é negativa dizemos, nesse caso, que os intervalos sucessivos de mensuração são desiguais ou de comprimentos desiguais. Em decorrência do que já expusemos anteriormente, podemos dizer que uma escala ordinal não admite uma unidade de medida no sentido de que a modalidade de qualquer objeto valha tantas vezes essa mesma unidade”.

Além dessa restrição, por si mesma tão grave, existe uma outra que também vai tipificar uma escala ordinal. Trata-se da ausência, nela, da modalidade chamada de zero absoluto, ou seja, a modalidade que denota a ausência de todo o conteúdo inerente à grandeza medida pela escala.

Ponhamos, para aclarar este assunto, o exemplo da escala ordinal que expressa as avaliações do rendimento escolar dos alunos da Universidade de Brasília. Essa escala é formada pelos

seguintes graus ou modalidades: *SR* = sem rendimento, *MI* = média insuficiente, *MM* = grau médio, *MS* = grau superior, *SS* = grau excelente. A partir da referida categorização ordinal é fácil perceber que:

- (i) o estado *SR* não significa, necessariamente, que o aluno não possua nenhum conhecimento da disciplina avaliada;
- (ii) mesmo que um aluno não respondesse a nenhum dos quesitos formulados no exame, ou seja, caso ele obtivesse uma nota zero, ainda assim esse grau não seria interpretado como um zero absoluto, ou seja, ela não significaria, necessariamente, uma completa ausência de conhecimentos;
- (iii) ainda que a escala contivesse um zero, o seu significado relativamente ao conteúdo de conhecimento seria sempre relativo;
- (iv) não se pode dizer que a passagem do grau *MM* para *MS* indique o mesmo quantum de acréscimo de conhecimentos que o da correspondente passagem de *MS* para *SS* ou, em outras palavras, a escala de graus da universidade não admite uma unidade de mensuração por ter intervalos sucessivos desiguais.

Admitamos, agora, dando continuidade ao exemplo com que estamos aclarando o entendimento de uma escala ordinal, que a avaliação do desempenho escolar se dê através das respostas a um conjunto de questões previamente valorizadas por notas. Suponhamos, além disso, que as referidas notas ou scores tenham um mínimo igual a zero e uma soma máxima igual a cem. Queremos saber, diante dessas condições, se a escala de graus assim produzida é ordinal. Esta questão, tal como foi formulada, suscita uma resposta inadequada e, desgraçadamente, muito corriqueira e que decorre da confusão que se estabelece entre as escalas ordinal e intervalar, tema da próxima seção. Antes de examinarmos essa nova categoria de escala de mensuração, cumpre dizermos que tanto uma escala ordinal quanto uma escala nominal podem ser contínuas, pelo menos conceitualmente contínuas. No caso em tela, podemos imaginar, teoricamente, que a nota ou o grau de qualquer aluno seja um número do intervalo $[0, 100]$ de números reais. Entretanto não podemos afirmar, com rigor, que essa escala não seja ordinal. Para compreender essa afirmação cumpre-nos conceber a noção intuitiva de dificuldade, conceito de cunho psicométrico. A dificuldade de uma questão ou item componente de um teste vem a ser a probabilidade de que sua resposta por um examinando escolhido imparcialmente seja correta em um dado grupo de respondentes. Nesse sentido, se todas as questões que compõem o teste têm dificuldades diferentes e se essas questões medem um mesmo e único traço, a escala resultante será ordinal, isto é, os graus ou notas por ela revelados só servirão aos propósitos de dizer quem teve menor ou maior aproveitamento. Outra seria a resposta, entretanto, se todas as dificuldades fossem iguais na mesma situação anterior. Nesse caso a escala de graus do rendimento escolar já seria mais que ordinal, pois uma escala intervalar, tema da próxima seção.

4.3 Escalas intervalar e de razão

Uma escala intervalar é, também, uma escala ordinal à qual lhe acrescentamos a seguinte condição: ela possui uma unidade de mensuração μ , de tal modo que toda modalidade da escala é um múltiplo dessa unidade.

De acordo com essa conceituação, podemos concluir, imediatamente, que numa escala intervalar os intervalos de mesma amplitude denotam que o correspondente quantum da grandeza têm, neles, a mesma variação matemática. Essa propriedade significa que esse quantum do intervalo categórico $[x_1, x_2]$ pode ser medido por sua amplitude $m[x_1, x_2]$. A amplitude $m[x_1, x_2]$ é uma função não negativa dos intervalos categóricos $[x_1, x_2]$ definidos na escala intervalar. Assim é que quando, por exemplo, a escala se expressa no conjunto dos números reais – e esse é o caso mais freqüente, mas não exclusivo – podemos considerar como amplitude o comprimento $x_2 - x_1$ do intervalo $[x_1, x_2]$. Assim, numa escala intervalar, vista dessa última forma, têm sentido as operações de adição e subtração entre suas determinações além, é claro, da multiplicação por um escalar ou número real. Desde o ponto de vista estatístico essas propriedades da escala intervalar é que vão permitir, dentre outras, a aplicação dos clássicos sumários das médias, medianas, percentis e das diversas maneiras de medir a variabilidade das séries estatísticas envolvendo as medidas e as suas correspondentes freqüências, simultaneamente.

Por outro lado quando a escala ordinal admite, simultaneamente, o zero absoluto e a unidade de mensuração, temos o que se chama de escala de razão. Esta é a escala mais aperfeiçoada no sentido de que, sobre ela, faz sentido a divisão entre as suas determinações. Daí, em decorrência dessa propriedade, é que se justifica o seu nome, derivado da palavra latina ratio, isto é, razão ou quociente. Nesta escala de razão podemos afirmar que dados dois objetos com medidas y e x a razão $y/x = k$ expressa, exatamente, que o primeiro objeto tem um quantum y do construto igual a k vezes o quantum x medido pelo outro. Um outro modo de dizer isso é que entre as escalas intervalar e de razão podemos afirmar que o quantum do conteúdo da variável está plenamente representado em suas determinações. Contrariamente, a escala intervalar é interdita a essas comparações pois, nela, só podemos ter medidas as variações dos quanta entre dois estados ou modalidades. Assim, por exemplo, a escala Kelvin de temperatura é uma escala de razão por possuir o zero absoluto, além da unidade de medida, o grau kelvin. Esse grau zero é o estado correspondente à ausência de calor pela ausência de movimento nas partículas íntimas da matéria. Em contrapartida, a escala Celsius é, apenas, uma escala intervalar. Nesse caso, se dois corpos possuem temperaturas medidas na escala Kelvin e iguais a $10^0 K$ e $20^0 K$, podemos dizer que esse último possui o dobro do calor do primeiro, pois $20^0 K = 2 \times 10^0 K$. Entretanto, se expressadas essas temperaturas na escala Celsius, já não se pode afirmar o mesmo. Com efeito, essas temperaturas, agora, são iguais respectivamente a $-263^0 C$ e $-253^0 C$ que, como se vê, já não têm a mesma razão de duplicidade antes expressada na escala Kelvin. Elas apenas mantêm, neste caso, a mesma diferença de antes igual a dez unidades.

Cumpramos ainda sabermos, no concernente às escalas intervalares, que suas determinações não são essencialmente numéricas e que, além disso, podem ser concebidas escalas intervalares

finitas, isto é, escalas com um número finito de modalidades ou determinações. Dá-se esse caso, por exemplo, na situação de um teste formado por n itens dicotômicos do tipo certo-errado positivamente associados, com os mesmos índices de dificuldades e medindo o mesmo traço ou variável abstrata. Nessas circunstâncias, se o 0 (zero) é a nota da resposta errada a um item qualquer e se 1 (um) é a correspondente nota da resposta correta, a escala intervalar resultante se expressa no conjunto $\{0, 1, 2, \dots, n\}$ dos inteiros não negativos menores ou iguais ao número de itens que compõem o teste.

Uma outra questão importante no tocante às escalas intervalares de mensuração reside no problema da incomensurabilidade da unidade. Comensuralidade significa que a unidade pode ser subdividida em quaisquer números de subunidades que desejemos usar. Quando a escala intervalar é incomensurável não existe uma unidade final, por assim dizer, e, em tal circunstância, ela pode expressar-se no conjunto dos números reais. É o caso, por exemplo, da escala Celsius de temperatura. A sua unidade, o grau Celsius, pode ser subdividida em quantas subunidades se deseje pois, para isso, basta subdividir o intervalo da coluna de mercúrio entre dois graus inteiros sucessivos no correspondente número de subintervalos de mesma amplitude.

Devemos chamar a atenção do leitor para o fato de que dizer que uma medida empírica pode ser um número real é, apenas, uma força de expressão, porque, do modo como se dá o processo de mensuração, a medida, se numérica, deve ser sempre expressada por um número racional ou fracionário. Não podemos, por exemplo, com rigor matemático, dizer que um corpo tem a temperatura igual a $\pi^{\circ}C$, (π graus Celsius) porque este é um número irracional. Entretanto, pela incomensurabilidade da unidade da temperatura medida em graus Celsius, podemos sempre acercarmo-nos desse valor por um número racional tão próximo dele quanto desejemos.

4.4 Transformação de escalas

Muitas vezes ao tratarmos com problemas multidimensionais, isto é, com a observação simultânea ou conjunta de várias variáveis defrontamo-nos com a dificuldade inerente à diversidade de suas escalas de expressão. Assim, por exemplo, um politicólogo examina uma série estatística multidimensional quando suas observações referem-se, simultaneamente, ao voto dado pelo eleitor a um candidato ou partido político, à sua renda, ao seu nível de instrução e à sua idade. O leitor reconhece, desde logo, nesse exemplo, a mistura das escalas nominal e de razão no seio dessa variável quadridimensional. Essa mistura de escalas é um grave empecilho ao uso dos métodos estatísticos porque estes, em geral, exigem a fidelidade das variáveis a um mesmo tipo de escala de mensuração.

Um tratamento inicial, capaz de contornar dificuldade levantada, consiste em tentar a expressão de todas as variáveis numa mesma escala. Assim, neste mesmo exemplo, uma primeira solução do problema consistiria em degradar o nível da variável renda – expressada, originalmente, numa escala de razão – para uma escala mais pobre, a nominal, onde são medidas as três outras variáveis. Isso se faz, por exemplo, definindo-se para a renda uma categorização do tipo renda baixa, média e elevada, segundo os seus valores originais. Desde logo, seguindo essa

metodologia, configuram-se dois problemas novos no exame da questão em tela:

- (i) a arbitrariedade da escolha da escala degradada da renda;
- (ii) a notória perda de informações estatísticas inerente ao abandono da escala original da rendas para uma escala mais pobre como a nominal.

Considerando que as escalas padrões formam uma graduação ajustada no sentido da riqueza interpretativa das medidas por elas produzidas inferimos, desde logo, que a interpretação das medidas efetuadas por uma certa escala através de uma escala sucedânea mais pobre configura uma degradação ou perda de informações reveladas pela primeira. Contrariamente a isso, a leitura numa escala superior das medidas efetuadas numa escala mais pobre constitui uma elevação que só pode ser justificada mediante hipóteses adicionais ou exógenas ao processo original de mensuração porque, nada no processo inicial de mensuração, em si mesmo, autoriza-nos a efetuarmos essa elevação. Dessa maneira, necessitamos de uma informação extra que nos autorize a promoção em tela.

A seguir, neste capítulo, serão examinadas algumas dessas técnicas de degradação ou de elevação das escalas designadas aqui, genericamente, sob a denominação de conversão de escalas.

A conversão da escala intervalar ou de razão numa escala ordinal requer a partição do campo de variação da variável X em m intervalos ou classes contíguas I_1, I_2, \dots, I_m de tal sorte que esses m intervalos sejam disjuntos e exaustivos. Desse modo, estabelecemos uma ordenação natural ($I_1 \prec I_2 \prec \dots \prec I_m$) entre essas m classes e, assim, quando uma observação pertence a uma das classes I_i ela recebe o valor correspondente na escala ordinal. Esse valor é chamado de valor de escala e, em princípio, pode ser adotado o próprio índice i para representá-lo.

Esse procedimento, obviamente, implica em perda de informações contidas na escala original e, portanto, ele constitui um método de degradação. São duas as principais fontes responsáveis pela perda de informações agora constatada:

- (a) duas medições distintas x_a e x_b , quando pertencentes à mesma classe I_i , transformar-se-ão, na escala ordinal, num mesmo valor i , tornando-as indistinguíveis;
- (b) duas medições diferentes x_a e x_b serão distintas na escala ordinal apenas por pertencerem a classes diferentes, independentemente do valor original de sua diferença $x_a - x_b$.

O problema da construção da nova escala ordinal divide-se, por sua vez, em dois outros, a saber:

- a escolha do número m de categorias ou de intervalos da partição;
- a determinação dos limites ou pontos de corte de cada um dos intervalos ou classes.

É evidente, do exposto, que qualquer critério de resolução do problema proposto deve minimizar a perda de informações resultante da conversão. Entretanto, como não podemos explicitar

uma função que represente essa perda de informações, propugnamos pelo uso de métodos heurísticos, de fácil entendimento. Duas soluções, muito usadas, consistem em definir as classes da partição em intervalos de mesmo comprimento ou em m classes de mesma frequência absoluta n/m .

Uma outra forma de degradação de escalas resulta de converter as medidas efetuadas em uma escala ordinal para uma escala nominal. A solução, neste caso, é totalmente arbitrária e consiste, basicamente, em escolher quais as modalidades da escala ordinal vão ser reunidas para comporem uma modalidade de escala nominal. O caso particular mais simples desse método consiste em usar as mesmas categorias da escala ordinal como as correspondentes categorias da escala nominal. Nesse caso, é óbvio, abandonamos ou perdemos a informação pertinente à ordenação das categorias.

Quanto à elevação de escala, a sua forma mais elementar é resultado de transformar as medidas de uma escala nominal em ordinal. Uma solução simples desse problema exige o conhecimento de uma nova variável ordinal fortemente associada ou correlacionada com a variável nominal. Nesse caso as categorias da escala ordinal, obtida exogenamente, é que determinarão a ordenação buscada. Assim, a técnica proposta usa uma informação exterior ao processo de mensuração para a elevação da escala.

Devemos considerar, por outro lado, que uma maneira mais sistemática de resolver esse problema seria atribuir valores ou scores às diversas categorias da escala nominal de tal modo que, com esses valores, seja máxima a correlação da variável nominal com uma variável intervalar ou de ordem já fortemente associada à variável nominal. Esse método é chamado de escalagem matemática e constitui a base das técnicas de análise de correspondências.

5 Anatomia do construto

Já compreendemos que um construto torna-se conhecido por suas faces ou traços que constituem um rol de suas manifestações concretas no plano das observações. No entanto, como ele é uma abstração, justo é nos inquirirmos quanto a sua existência real. Assim é que, por exemplo, podemos interrogar-nos quanto às realidades do conceito de inteligência ou da tolerância racial. Em verdade é pouco relevante se um construto tem ou não existência real. O fato mais importante é que ele constitui uma espécie de sumário ou resumo mental de um complexo de manifestações concretas vistas como suas componentes. Pensemos, sobre isso, no construto sociológico que chamamos de tolerância racial. Ora, a natureza dessa idéia reside exatamente naquelas atitudes mentais ou nos comportamento efetivos que entendemos serem as encarnações precisas dessa postura ante a convivência com as diversas raças. Assim, se alguém ante uma seqüência de questões que o interrogam sobre o modo pelo qual reage em uma situação desse tipo de acordo com as respostas obtidas poderemos não só caracterizá-lo como medir-lhe o grau de tolerância racial. Suponhamos, assim, que estamos diante das três seguintes interrogações:

- (i) Você aceita ter de trabalhar com alguém de outra raça?

(ii) Você aceita ter um amigo de outra raça?

(iii) Você aceita que sua filha ou seu filho namore alguém de outra raça?

Ora, o que entendemos sobre esse questionário nos parece muito óbvio. Primeiramente compreendemos que ele contém graus crescentes de conteúdo tolerância racial no sentido de que alguém que só responde positivamente à primeira pergunta é menos tolerante do que outro que responde positivamente às duas primeiras interrogações. Diante desse questionário, também, defrontamo-nos com diversas faces da tolerância racial representadas por uma espécie de aceitação que só se estende a um certo ponto. Com efeito, as questões propostas nos interrogam de modo a poder verificar até que ponto de tolerância nós somos capazes de ir. Apesar de incompleta quanto a uma compreensão mais reveladora do construto examinado entendemos que, antes de mais nada, necessitamos identificar as diversas faces, traços ou manifestações capazes de nos elucidar quanto aos porquês da tolerância ou de sua opositora, a discriminação racial. Neste caso, entre muitas outras questões referentes a sentimentos de superioridade racial e cultural deverão ser abordadas no exame de construto. Um meio de fazê-lo consiste em consultar especialistas e pesquisar, na literatura sociológica, os comportamentos que manifestem a tolerância racial. Esta pesquisa constitui o que entendemos por análise de conteúdo de um construto. Analisemos agora, mais concretamente, um construto político como o terrorismo. Este é um conceito que pode ser entendido como a violência ou a ameaça, com fins políticos, praticadas por um grupo organizado de pessoas que se opõem à autoridade constituída. Entendemos, também, que o terrorismo visa à destituição dessas autoridade e, para isso, entende que atos intimidatórios ou violentos são justificáveis para alcançarem esse fim.

Ora, o que podemos fazer a partir dessa conceituação do construto terrorismo é construir um instrumento conceitual capaz de medir a sua existência e a própria intensidade. Para isso deveremos, a partir da definição proposta, identificar-lhe, os vários tipos de manifestações e posturas para, através de um questionário guiado por tais elementos de identificação dessas faces avaliarmos, em cada país, o nível de manifestação do terrorismo.