4.8. Estimação de parâmetros

Como foi dito na seção 4.1, a *Inferência Estatística* reúne técnicas que nos permitem tirar conclusões sobre populações a partir do que é encontrado nas amostras (lembre-se de que 'população', em Estatística, não é necessariamente um conjunto de pessoas, e sim um conjunto de pessoas, animais ou objetos que nos interessam, dos quais a amostra é um subconjunto).

Estas conclusões podem ser de dois tipos. Primeiro, se temos alguma hipótese sobre o valor de um parâmetro (ou sobre a diferença entre dois parâmetros), podemos testar esta hipótese a partir de amostras, e decidir se ela deve ou não ser aceita como verdadeira; para isso, usamos os *Testes de Hipótese*, vistos nos capítulos anteriores. Segundo, podemos calcular *estimativas* do valor do parâmetro desconhecido (ou da diferença entre parâmetros desconhecidos) a partir do que é observado em amostras; para isto, usamos as técnicas de *Estimação de Parâmetros*, que serão vistas neste capítulo.

É importante aqui atentar para os detalhes da terminologia. *Estimação* é o nome desta área da Estatística, e a *Teoria da Estimação* é a parte da Matemática que desenvolve a teoria necessária. O *estimador* é uma variável, definida como função dos dados da amostra, que serve de base para a *estimativa*, o resultado a que chegamos depois de fazer os cálculos.

Por exemplo, se queremos estimar a média μ de uma população desconhecida, provavelmente usaremos como *estimador* a média $\overline{X}$ calculada na amostra. Da mesma forma, se queremos estimar a proporção π de sucessos em uma população, provavelmente usaremos como *estimador* a proporção P de sucessos encontrada na amostra. Isto parece bastante razoável, e de fato a teoria demonstra que $\overline{X}$ e P são os melhores estimadores possíveis para μ. e π. (No entanto, esta lógica, que parece muito natural, nem sempre funciona; a teoria mostra por exemplo que a variância s de uma amostra não é um bom estimador da variância σ da população, pois leva a estimativas que tendem a ser *menores* do que o valor real de σ; veja a seção 4.7.1.1).

Escolhido o *estimador*, calculamos a *estimativa*. Há dois tipos de estimativa, a *pontual* e a *intervalar*. A estimava é *pontual* quando dada simplesmente pelo valor do *estimador* encontrado na amostra. Já usamos este tipo de estimativa nos capítulos anteriores. Por exemplo, para testar uma hipótese sobre a média de uma população precisamos do desvio-padrão σ desta população; como em geral não o conhecemos, usamos o desvio-padrão s da amostra como estimativa pontual de σ (seção 4.5.3.1). Da mesma forma, para testar hipóteses sobre a proporção π de sucessos de uma população, precisamos do valor deste mesmo π para calcular o desvio-padrão da variável de teste P. A saída é fazer uma estimativa pontual de π a partir de P (seção 4.4.4). Além dessas aplicações técnicas dentro da Estatística, estimativas pontuais também são muito comuns na mídia. Se o IBGE diz por exemplo que o salário médio do trabalhador brasileiro é de R$ 2.261,00 (IBGE, 4º trimestre de 2019), esta é uma estimativa *pontual*, igual à média $\overline{X}$ encontrada numa amostra.

A estimava é *intervalar* quando expressa por meio de um intervalo centrado no valor do estimador. Um exemplo deste tipo de estimação é o das pesquisas que procuram estimar a intenção de voto dos eleitores antes de eleições. Os resultados são geralmente publicados na forma de um intervalo cujo centro é a proporção P encontrada na amostra, cercada por uma margem de erro que depende do tamanho da amostra. Por exemplo, se uma pesquisa diz que o candidato conta com 34 % dos votos, com margem de erro ±2 %, ele provavelmente deve contar com algo entre 32 a 36 % dos votos. Este intervalo é chamado de *intervalo de confiança.*

Intervalos de confiança podem ser obtidos para qualquer parâmetro (ou diferença entre parâmetros). Veremos a seguir as técnicas básicas para calcular intervalos em alguns casos mais simples, como a estimação de médias, proporções e diferenças de médias e de proporções, para amostras grandes ou pequenas. Como foi feito anteriormente nas seções sobre testes de hipóteses, usaremos apenas a teoria aplicável quando as amostras são aleatórias simples.