Pedro A. Morettin

Wilton de O. Bussab

ESTATÍSTICA BÁSICA

6ª edição
Revista e atualizada

Editora Saraiva

Capítulo 10

Introdução à Inferência Estatística

10.1 Introdução

Vimos, na Parte 1, como resumir descritivamente variáveis associadas a um ou mais conjuntos de dados. Na Parte 2, construímos modelos teóricos (probabilísticos), identificados por parâmetros, capazes de representar adequadamente o comportamento de algumas variáveis. Nesta terceira parte apresentaremos os argumentos estatísticos para fazer afirmações sobre as características de uma população, com base em informações dadas por amostras.

O uso de informações de uma amostra para concluir sobre o todo faz parte da atividade diária da maioria das pessoas. Basta observar como uma cozinheira verifica se o prato que ela está preparando tem ou não a quantidade adequada de sal. Ou, ainda, quando um comprador, após experimentar um pedaço de laranja numa banca de feira, decide se vai comprar ou não as laranjas. Essas são decisões baseadas em procedimentos amostrais.

Nosso objetivo nos capítulos seguintes é procurar dar a conceituação formal a esses princípios intuitivos do dia-a-dia para que possam ser utilizados cientificamente em situações mais complexas.

10.2 População e Amostra

Nos capítulos anteriores, tomamos conhecimento de alguns modelos probabilísticos que procuram medir a variabilidade de fenômenos casuais de acordo com suas ocorrências: as distribuições de probabilidades de variáveis aleatórias (qualitativas ou quantitativas). Na prática, freqüentemente o pesquisador tem alguma idéia sobre a forma da distribuição, mas não dos valores exatos dos parâmetros que a especificam.

Por exemplo, parece razoável supor que a distribuição das alturas dos brasileiros adultos possa ser representada por um modelo normal (embora as alturas não possam assumir valores negativos). Mas essa afirmação não é suficiente para determinar qual a distribuição normal correspondente; precisaríamos conhecer os parâmetros (média e variância) dessa normal para que ela ficasse completamente especificada. O propósito do pesquisador seria, então, descobrir (estimar) os parâmetros da distribuição para sua posterior utilização.

Se pudéssemos medir as alturas de todos os brasileiros adultos, teríamos meios de obter sua distribuição exata e, daí, produzir os correspondentes parâmetros. Mas nessa situação não teríamos necessidade de usar a inferência estatística!

Raramente se consegue obter a distribuição exata de alguma variável, ou porque isso é muito dispendioso, ou muito demorado ou às vezes porque consiste num processo destrutivo. Por exemplo, se estivéssemos observando a durabilidade de lâmpadas e testássemos todas até queimarem, não restaria nenhuma para ser vendida. Assim, a solução é selecionar parte dos elementos (amostra), analisá-la e *inferir* propriedades para o todo (população).

Outras vezes estamos interessados em explorar relações entre variáveis envolvendo experimentos mais complexos, para a obtenção dos dados. Por exemplo, gostaríamos de obter resposta para a seguinte indagação: a altura que um produto é colocado na gôndola de um supermercado afeta a sua venda? Observe que para responder a questão precisamos obter dados de vendas com o produto oferecido em diferentes alturas, e que essas vendas sejam controladas para evitar interferências de outros fatores que não a altura. Nesse caso não existe claramente um conjunto de *todos* os elementos para os quais pudéssemos encontrar os parâmetros populacionais. Recorrer a modelos para descrever o todo (população) facilita a identificação e solução do problema. Nesse exemplo, supondo que as vendas V_h do produto oferecido na altura h ($h = 1$ representando *baixo*, $h = 2$ representando *meio* e $h = 3$ representando *alto*) segue uma distribuição próxima a normal, ou seja, $V_h \sim N(\mu_h, \sigma^2)$, o nosso problema passa a ser o de verificar, por meio de dados coletados do experimento (amostra), se existe evidência de igualdade das médias μ_1, μ_2 e μ_3. Note que, em nossa formulação do problema, supusemos que as três situações de alturas resultam observações com a mesma variância σ^2. Essa suposição poderia ser modificada.

Soluções de questões como as apresentadas acima são o objeto da *inferência estatística*.

Dois conceitos básicos são, portanto, necessários para o desenvolvimento da Inferência Estatística: população e amostra.

Definição. *População* é o conjunto de todos os elementos ou resultados sob investigação. *Amostra* é qualquer subconjunto da população.

Vejamos outros exemplos para melhor entender essas definições.

Exemplo 10.1. Consideremos uma pesquisa para estudar os salários dos 500 funcionários da Companhia MB. Seleciona-se uma amostra de 36 indivíduos, e anotam-se os seus salários. A variável aleatória a ser observada é “salário”. A população é formada pelos 500 funcionários da companhia. A amostra é constituída pelos 36 indivíduos selecionados. Na realidade, estamos interessados nos salários, portanto, para sermos mais precisos, devemos considerar como a população os 500 salários correspondentes aos 500 funcionários. Conseqüentemente, a amostra será formada pelos 36 salários dos indivíduos selecionados. Podemos estudar a distribuição dos salários na amostra, e esperamos que esta reflita a distribuição de todos os salários, desde que a amostra tenha sido escolhida com cuidado.

Exemplo 10.2. Queremos estudar a proporção de indivíduos na cidade A que são favoráveis a certo projeto governamental. Uma amostra de 200 pessoas é sorteada, e a opinião de cada uma é registrada como sendo a favor ou contra o projeto. A população consiste de todos os moradores da cidade, e a amostra é formada pelas 200 pessoas selecionadas. Podemos, como foi visto no Capítulo 5, definir a variável X, que toma o valor 1, se a resposta de um morador for favorável, e o valor 0, se a resposta for contrária ao projeto. Assim, nossa população pode ser reduzida à distribuição de X, e a amostra será constituída de uma seqüência de 200 zeros e uns.

Exemplo 10.3. O interesse é investigar a duração de vida de um novo tipo de lâmpada, pois acreditamos que ela tenha uma duração maior do que as fabricadas atualmente. Então, 100 lâmpadas do novo tipo são deixadas acesas até queimarem. A duração em horas de cada lâmpada é registrada. Aqui, a variável é a duração em horas de cada lâmpada. A população é formada por todas as lâmpadas fabricadas ou que venham a ser fabricadas por essa empresa, com o mesmo processo. A amostra é formada pelas 100 lâmpadas selecionadas. Note-se que nesse caso não podemos observar a população, ou seja, a distribuição da duração de vida das lâmpadas na população, pois isso corresponderia a queimar todas as lâmpadas. Assim, em alguns casos, não podemos observar a população toda, pois isso significaria danificar (ou destruir) todos os elementos da população. Esse problema geralmente é contornado atribuindo-se um modelo teórico para a distribuição da variável populacional.

Exemplo 10.4. Em alguns casos, fazemos suposições mais precisas sobre a população (ou sobre a variável definida para os elementos da população). Digamos que X represente o peso real de pacotes de café, enchidos automaticamente por uma máquina. Sabe-se que a distribuição de X pode ser representada por uma normal, com parâmetros μ e σ^2 desconhecidos. Sorteamos 100 pacotes e medimos seus pesos. A população será o conjunto de todos os pacotes enchidos ou que virão a ser enchidos pela máquina, e que pode ser suposta como normal. A amostra será formada pelas 100 medidas obtidas dos pacotes selecionados, que pode ser pensada como constituída de 100 observações feitas de uma distribuição normal. Veremos mais adiante como tal amostra pode ser obtida.

Exemplo 10.5. Para investigar a "honestidade" de uma moeda, nós a lançamos 50 vezes e contamos o número de caras observadas. A população, como no caso do Exemplo 10.2, pode ser considerada como tendo a distribuição da variável X, assumindo o valor 1, com probabilidade p, se ocorrer cara, e assumindo o valor 0, com probabilidade $1 - p$, se ocorrer coroa. Ou seja, a população pode ser considerada como tendo distribuição de Bernoulli com parâmetro p. A variável ficará completamente especificada quando conhecermos p. A amostra será uma seqüência de 50 números zeros ou uns.

Exemplo 10.6. Há razões para supor que o tempo Y de reação a certo estímulo visual dependa da idade do indivíduo (esse exemplo será usado nos Capítulos 15 e 16). Suponha, ainda, que essa *dependência seja linear*. Para verificarmos se essa suposição é verdadeira, obtiveram-se 20 dados da seguinte maneira: 20 pessoas foram selecionadas, sendo 10 homens e 10 mulheres. Dentro de cada grupo de homens e mulheres foram selecionadas duas pessoas das seguintes faixas de idade: 20, 25, 30, 35 e 40 anos. Cada pessoa foi submetida ao teste e seu tempo de reação y foi medido. A população poderia ser considerada como formada por todas aquelas pessoas que viessem a ser submetidas ao teste, segundo o sexo e a idade. A amostra é formada pelas 20 medidas, que estão apresentadas na Tabela 15.1.

Observações:

(i) Os três últimos exemplos mostram uma ampliação do conceito definido de população, ou seja, designamos agora a população como sendo a função probabilidade ou função densidade de probabilidade de uma v.a. X, modelando a característica de interesse. Esse artifício simplifica substancialmente o problema estatístico, exigindo no entanto uma proposta de modelo para a variável X. Nesses casos simplificaremos a linguagem, dizendo: "seja a população $f(x)$". Por exemplo, "considere a população das alturas $X \sim N(\mu, \sigma^2)$".

(ii) Essa abordagem, por meio da distribuição de probabilidades, utiliza muitas vezes o conceito de população infinita contínua, exigindo um tratamento matemático mais cuidadoso. É mais fácil apresentar os problemas e soluções por meio de populações finitas. É o que faremos muitas vezes. Entretanto, é importante que o estudante aprenda a trabalhar com o conceito de modelo, explorando o caso de "população $f(x)$".

10.3 Problemas de Inferência

Como já dissemos anteriormente, o objetivo da Inferência Estatística é produzir afirmações sobre dada característica da população, na qual estamos interessados, a partir de informações colhidas de uma parte dessa população. Essa característica na população pode ser representada por uma variável aleatória. Se tivéssemos informação completa sobre a função de probabilidade, no caso discreto, ou sobre a função densidade de probabilidade, no caso contínuo, da variável em questão, não teríamos necessidade de escolher uma amostra. Toda a informação desejada seria obtida por meio da distribuição da variável, usando-se a teoria estudada anteriormente.

Mas isso raramente acontece. Ou não temos qualquer informação a respeito da variável, ou ela é apenas parcial. Podemos admitir, como no exemplo das alturas de brasileiros adultos, que ela siga uma distribuição normal, mas desconhecemos os parâmetros que a caracterizam (média, variância). Em outros casos, podemos ter uma idéia desses parâmetros, mas desconhecemos a forma da curva. Ou ainda, o que é muito freqüente, não possuímos informações nem sobre os parâmetros, nem sobre a forma da curva. Em todos os casos, o uso de uma amostra nos ajudaria a formar uma opinião sobre o comportamento da variável (população).

Embora a identificação e a descrição da população sejam fundamentais no processo inferencial, é comum os pesquisadores dedicarem mais atenção em descrever a amostra do que a população para a qual serão feitas as afirmações. É imprescindível que se explicite claramente a população investigada.

Neste livro estaremos mais preocupados em trabalhar com populações descritas por modelos do que com populações finitas identificadas por elementos portadores de uma característica de interesse. Portanto, na maioria das vezes, iremos nos referir à "população X", significando que a variável de interesse X, definida sobre a população-alvo, segue uma distribuição $f(x)$. Nosso problema de interesse passaria a ser o de fazer afirmações sobre a forma da curva e seus parâmetros.

Alguns exemplos simples nos darão uma noção dos tipos de formulações e problemas que a inferência estatística pode nos ajudar a resolver.

Exemplo 10.5. (continuação) Voltemos ao exemplo da moeda. Indicando por X o número de caras obtidas depois de lançar a moeda 50 vezes, sabemos que, se tomados alguns cuidados quando do lançamento, X segue uma distribuição binomial, ou seja, $X \sim b(50, p)$. Esse modelo é válido, admitindo-se ou não a "honestidade" da moeda, isto é, sendo ou não $p = 1/2$. Lançada a moeda, vamos supor que tenham ocorrido 36 caras. Esse resultado traz evidência de que a moeda seja "honesta"? Para tomarmos uma decisão, podemos partir do princípio de que a moeda não favorece nem cara nem coroa, isto é, $p = 1/2$. Com essa informação e com o modelo binomial, podemos encontrar qual a probabilidade de se obterem 36 caras ou mais, e esse resultado nos ajudaria a tomar uma decisão. Suponha que a decisão foi rejeitar a "honestidade" da moeda: qual é a melhor estimativa para p, baseando-se no resultado observado?

Descrevemos aí os dois problemas básicos da Inferência Estatística: o primeiro é chamado *teste de hipóteses*, e o segundo, *estimação*. Nos capítulos seguintes, esses problemas serão abordados com mais detalhes.

Exemplo 10.4. (continuação) Às vezes, o modelo teórico associado ao problema não é tão evidente. No caso da máquina de encher pacotes de café automaticamente, digamos que ela esteja regulada para enchê-los segundo uma distribuição normal com média 500 gramas e desvio padrão de 100 gramas, isto é, $X \sim N(500, 20^2)$. Sabemos também que, às vezes, a máquina desregula-se e, quando isso acontece, o único parâmetro que se altera é a média, permanecendo a mesma variância. Para manter a produção sob controle, iremos colher uma amostra de 100 pacotes e pesá-los. Como essa amostra nos ajudará a tomar uma decisão? Parece razoável, nesse caso, usarmos a média $\bar{x}$ da amostra como informação pertinente para uma decisão. Mesmo que a máquina esteja regulada, dificilmente $\bar{x}$ será igual a 500 gramas, dado que os pacotes apresentam certa variabilidade no peso. Mas se $\bar{x}$ não se afastar muito de 500 gramas, não existirão razões para suspeitarmos da qualidade do procedimento de produção. Só iremos pedir uma revisão se $\bar{x} - 500$, em valor absoluto, for "muito grande".

O problema que se apresenta agora é o de decidir o que é próximo ou distante de 500 gramas. Se o mesmo procedimento de colher a amostra de 100 pacotes fosse repetido um número muito grande de vezes, sob a condição de a máquina estar regulada, teríamos idéia do comportamento da v.a. $\bar{x}$, e saberíamos dizer se aquele valor observado é ou não um evento raro de ocorrer. Caso o seja, é mais fácil suspeitar da regulagem da máquina do que do acaso.

Vemos, então, a importância nesse caso de se conhecer as propriedades da distribuição da variável $\bar{x}$.

Exemplo 10.6. (continuação) A descrição matemática da v.a. Y: tempo de reação ao estímulo é um pouco mais complexa. Podemos supor que esse tempo, para uma dada idade x, seja uma v.a. com distribuição normal, com média dependendo da idade x, ou seja, podemos escrever

$$Y \sim N(\mu(x), \sigma^2).$$

A *linearidade* expressa no problema pode ser incluída na média $\mu(x)$ da seguinte maneira:

$$\mu(x) = \alpha + \beta x.$$

Voltaremos a esse modelo no Capítulo 16. Outra maneira de escrever as duas relações anteriores é

$$Y \mid x \sim N(\alpha + \beta x; \sigma^2).$$

Leia-se "Y dado x".

Podemos, por exemplo, estimar os parâmetros α e β, baseados na amostra de 20 dados. Ou podemos querer investigar a possibilidade de β ser igual a zero, significando que a idade não afeta o tempo de reação. Novamente, os dois principais problemas de inferência aparecem aqui: estimação e teste de uma hipótese. Um outro problema importante em inferência é o de *previsão*. Por exemplo, considerando um grupo de pessoas de 40 anos, poderemos prever com o modelo acima qual será o respectivo tempo de reação.

Repetir um mesmo experimento muitas vezes, sob as mesmas condições, nem sempre é possível, mas em determinadas condições é possível determinar teoricamente o comportamento de algumas medidas feitas na amostra, como por exemplo a média. Mas isso depende, em grande parte, do procedimento (plano) adotado para selecionar a amostra. Assim, em problemas envolvendo amostras, antes de tomarmos uma decisão, teríamos de responder a quatro perguntas:

(a) Qual a população a ser amostrada?

(b) Como obter os dados (a amostra)?

(c) Que informações pertinentes (estatísticas) serão retiradas da amostra?

(d) Como se comporta(m) a(s) estatística(s) quando o mesmo procedimento de escolher a amostra é usado numa população conhecida?

Nas seções e capítulos subseqüentes tentaremos responder a essas perguntas.

10.4 Como Selecionar uma Amostra

As observações contidas em uma amostra são tanto mais informativas sobre a população quanto mais conhecimento explícito ou implícito tivermos dessa mesma população. Por exemplo, a análise da quantidade de glóbulos brancos obtida de algumas gotas de sangue da ponta do dedo de um paciente dará uma idéia geral da quantidade dos glóbulos brancos no corpo todo, pois sabe-se que a distribuição dos glóbulos brancos é homogênea, e de qualquer lugar que se tivesse retirado a amostra ela seria "representativa". Mas nem sempre a escolha de uma amostra adequada é imediata. Por exemplo, voltando ao Exemplo 10.2, para o qual queríamos obter uma amostra de habitantes para saber a opinião sobre um projeto governamental, escolhendo intencionalmente uma amostra de 200 indivíduos moradores de certa região beneficiada pelo projeto, saberemos de antemão que o resultado conterá um *viés de seleção*. Isto é, na amostra, a proporção de pessoas favoráveis ao projeto deverá ser maior do que no todo, donde a importância da adoção de procedimentos científicos que permitam fazer inferências adequadas sobre a população.

A maneira de se obter a amostra é tão importante, e existem tantos modos de fazê-lo, que esses procedimentos constituem especialidades dentro da Estatística, sendo *Amostragem* e *Planejamento de Experimentos* as duas mais conhecidas. Poderíamos dividir os procedimentos científicos de obtenção de dados amostrais em três grandes grupos:

(a) *Levantamentos Amostrais*, nos quais a amostra é obtida de uma população bem definida, por meio de processos bem protocolados e controlados pelo pesquisador. Podemos, ainda, subdividi-los em dois subgrupos: levantamentos probabilísticos e não-probabilísticos. O primeiro reúne todas aquelas técnicas que usam mecanismos aleatórios de seleção dos elementos de uma amostra, atribuindo a cada um deles uma probabilidade, conhecida *a priori*, de pertencer à amostra. No segundo grupo estão os demais procedimentos, tais como: amostras intencionais, nas quais os elementos são selecionados com o auxílio de especialistas, e amostras de voluntários, como ocorre em alguns testes sobre novos medicamentos e vacinas. Ambos os procedimentos têm suas vantagens e desvantagens. A grande vantagem das amostras probabilísticas é medir a precisão da amostra obtida, baseando-se no resultado contido na própria amostra. Tais medidas já são bem mais difíceis para os procedimentos do segundo grupo.

Estão nessa situação os Exemplos 10.1 (conhecer os salários da Cia. MB), 10.2 (identificar a proporção de indivíduos favoráveis ao projeto), 10.4 (pesos dos pacotes de café) etc.

(b) *Planejamento de Experimentos*, cujo principal objetivo é o de analisar o efeito de uma variável sobre outra. Requer, portanto, interferências do pesquisador sobre o ambiente em estudo (população), bem como o controle de fatores externos, com o intuito de medir o efeito desejado. Podemos citar como exemplos aquele já citado sobre a altura de um produto na gôndola de um supermercado afetar as vendas e o Exemplo 10.6. Em ensaios clínicos em medicina, esse tipo de estudo é bastante usado, como por exemplo para testar se um novo medicamento é eficaz ou não para curar certa doença.

(c) *Levantamentos Observacionais*: aqui, os dados são coletados sem que o pesquisador tenha controle sobre as informações obtidas, exceto eventualmente sobre possíveis erros grosseiros. As séries de dados temporais são exemplos típicos desses levantamentos. Por exemplo, queremos prever as vendas de uma empresa em função de vendas passadas. O pesquisador não pode selecionar dados, esses são as vendas efetivamente ocorridas. Nesses casos, a especificação de um modelo desempenha um papel crucial na ligação entre dados e população.

No caso de uma série temporal, o modelo subjacente é o de *processo estocástico*; podemos pensar que a série efetivamente observada é uma das *infinitas possíveis realizações desse processo*. A população hipotética aqui seria o conjunto de todas essas realizações, e a série observada seria a amostra. Veja Morettin e Toloi (2006) para mais informações.

Neste livro iremos nos concentrar principalmente em levantamentos amostrais e, mais ainda, num caso simples de amostragem probabilística, a *amostragem aleatória simples, com reposição*, a ser designada por AAS. O leitor poderá consultar Bussab e Bolfarine (2005) para obter mais detalhes sobre outros procedimentos amostrais. Um breve resumo sobre alguns planos é dado no Problema 37. Noções sobre planejamento de experimentos podem ser vistas em Peres e Saldiva (1982).

Problemas

1. Dê sua opinião sobre os tipos de problemas que surgiriam nos seguintes planos amostrais:
   (a) Para investigar a proporção dos operários de uma fábrica favoráveis à mudança do início das atividades das 7h para as 7h30, decidiu-se entrevistar os 30 primeiros operários que chegassem à fábrica na quarta-feira.
   (b) Mesmo procedimento, só que o objetivo é estimar a altura média dos operários.
   (c) Para estimar a porcentagem média da receita municipal investida em lazer, enviaram-se questionários a todas as prefeituras, e a amostra foi formada pelas prefeituras que enviaram as respostas.
   (d) Para verificar o fato de oferecer brindes nas vendas de sabão em pó, tomaram-se quatro supermercados na zona sul e quatro na zona norte de uma cidade. Nas quatro lojas da zona sul, o produto era vendido com brinde, enquanto nas outras quatro era vendido sem brinde. No fim do mês, compararam-se as vendas da zona sul com as da zona norte.

2. Refazer o Problema 7 do Capítulo 8.

10.5 Amostragem Aleatória Simples

A amostragem aleatória simples é a maneira mais fácil para selecionarmos uma amostra probabilística de uma população. Além disso, o conhecimento adquirido com esse procedimento servirá de base para o aprendizado e desenvolvimento de outros procedimentos amostrais, planejamento de experimentos, estudos observacionais etc. Comecemos introduzindo o conceito de AAS de uma população finita, para a qual temos uma listagem de todas as N unidades elementares. Podemos obter uma amostra nessas condições, escrevendo cada elemento da população num cartão, misturando-os numa urna e sorteando tantos cartões quantos desejarmos na amostra. Esse procedimento torna-se inviável quando a população é muito grande. Nesse caso, usa-se um processo alternativo, no qual os elementos são numerados e em seguida sorteados por meio de uma tabela de números aleatórios (veja a sua utilização em Problemas e Complementos) ou por meio do uso de computadores, que podem gerar números aleatórios (veja o Capítulo 9).

Utilizando-se um procedimento aleatório, sorteia-se um elemento da população, sendo que todos os elementos têm a mesma probabilidade de ser selecionados. Repete-se o procedimento até que sejam sorteadas as n unidades da amostra.

Podemos ter uma AAS *com reposição*, se for permitido que uma unidade possa ser sorteada mais de uma vez, *e sem reposição*, se a unidade sorteada for removida da população.

Do ponto de vista da quantidade de informação contida na amostra, amostrar sem reposição é mais adequado. Contudo, a amostragem com reposição conduz a um tratamento teórico mais simples, pois ela implica que tenhamos *independência* entre as unidades selecionadas. Essa independência facilita o desenvolvimento das propriedades dos estimadores que serão considerados.

Portanto, para o restante do livro, o plano amostral considerado será o de *amostragem aleatória simples com reposição*, que denotaremos simplesmente por AAS.

Vejamos com algum detalhe o significado mais preciso de uma amostra.

Exemplo 10.7. Considere o Problema 2 acima, em que colhemos todas as amostras possíveis de tamanho 2, com reposição, da população {1, 3, 5, 5, 7}. Defina a variável X: valor assumido pelo elemento na população. Então, a distribuição de X é dada pela Tabela 10.1.

Tabela 10.1: Distribuição da v.a. X para o Problema 2.

x	1	3	5	7
$P(X = x)$	1/5	1/5	2/5	1/5

Indicando por X_1 o número selecionado na primeira extração e por X_2 o número selecionado na segunda extração, vimos que era possível escrever a distribuição conjunta do par (X_1, X_2). Veja também a Tabela 10.2. Além disso, as distribuições marginais de X_1 e X_2 são independentes e iguais à distribuição de X. Desse modo, cada uma das 25 possíveis amostras de tamanho 2 que podemos extrair dessa população corresponde a observar uma particular realização da v.a. (X_1, X_2), com X_1 e X_2 independentes e $P(X_1 = x) = P(X_2 = x) = P(X = x)$, para todo x. Essa é a caracterização de amostra casual simples que iremos usar neste livro.

Definição. Uma amostra aleatória simples de tamanho n de uma variável aleatória X, com dada distribuição, é o conjunto de n variáveis aleatórias independentes $X_1, X_2, ..., X_n$, cada uma com a mesma distribuição de X.

Ou seja, a amostra será a n-upla ordenada $(X_1, X_2, ..., X_n)$, onde X_i indica a observação do i-ésimo elemento sorteado.

Quando a população é caracterizada por uma distribuição de probabilidades, o modo mais simples para sortear uma AAS é usar os procedimentos de simulação estudados no Capítulo 9. O processo de simular uma observação de uma distribuição especificada por seus parâmetros nada mais é do que retirar uma AAS de tamanho um da população. Desse modo, para retirar uma AAS (com reposição) de n indivíduos da população X, basta gerar n números aleatórios independentes dessa distribuição.

Exemplo 10.8. Vamos retirar uma AAS de 5 alturas (em cm) de uma população de mulheres cujas alturas X seguem a distribuição $N(167; 25)$.

Usando-se, por exemplo, o gerador de números aleatórios do Excel, fornecendo os parâmetros $\mu = 167$ e $\sigma = 5$, além do tamanho da amostra $n = 5$, obtemos os valores:

$$x_1 = 165, \qquad x_2 = 161, \qquad x_3 = 168, \qquad x_4 = 173, \qquad x_5 = 173.$$

Note que, se você for gerar uma tal amostra, poderá obter valores diferentes desses. Observe, também, que o primeiro elemento a ser observado pode ser qualquer valor da população simulada $N(167; 25)$. Desse modo, indicando por X_1 o valor observado na primeira extração, concluímos que $X_1 \sim N(167; 25)$. Como a geração do segundo número aleatório é feita independentemente do segundo, resulta que a v.a. X_2, valor observado na segunda extração, também segue uma distribuição $N(167; 25)$, e assim por diante.

Diante do exposto, vemos que continua válida a definição de AAS dada acima, quando a amostra é retirada de uma população referenciada pela sua distribuição de probabilidades.

No caso de uma população X contínua, com f.d.p. $f(x)$, a f.d.p. conjunta da amostra $(X_1, X_2, ..., X_n)$, segundo o que vimos no Capítulo 8, será dada por

$$f(x_1, x_2, ..., x_n) = f_1(x_1)f_2(x_2) \ ... \ f_n(x_n),$$

onde $f_i(x_i)$ denota a distribuição (marginal) de X_i, $i = 1, ..., n$.

Antes de prosseguirmos, seria interessante fazer uma comparação da inferência estatística com o processo de simulação da população.

Podemos imaginar que qualquer característica X de interesse seja produzida por um "programa" (modelo) de gerador de números aleatórios, e que somente o "proprietário" (natureza) desse programa é que conhece a forma da distribuição de X, os valores dos parâmetros etc. relacionados ao programa. Quando "obtemos" a amostra, estamos apenas observando o resultado da simulação, não conhecemos nada do processo gerador dos dados. O objetivo da inferência estatística é fornecer critérios para nos ajudar a descobrir a forma da distribuição e/ou parâmetros usados pelo "proprietário". Bons indicadores desses valores nos ajudam a entender melhor os fenômenos e fazer previsões para futuras observações.

Daqui para frente, a menos que esteja especificada de outra maneira, sempre que mencionarmos a palavra amostra, estaremos entendendo a amostra obtida pelo processo probabilístico AAS, ou seja, o vetor aleatório $(X_1, X_2, ..., X_n)$ definido acima.

Problemas

3. A distribuição do número de filhos, por família, de uma zona rural está no quadro abaixo.

Nº de filhos	Porcentagem
0	10
1	20
2	30
3	25
4	15
Total	100

(a) Sugira um procedimento para sortear uma observação ao acaso dessa população.

(b) Dê, na forma de uma tabela de dupla entrada, as possíveis amostras do número de filhos de duas famílias que podem ser sorteadas e as respectivas probabilidades de ocorrência.

(c) Se fosse escolhida uma amostra de tamanho 4, qual seria a probabilidade de se observar a quádrupla ordenada (2, 3, 3, 1)?

10.6 Estatísticas e Parâmetros

Obtida uma amostra, muitas vezes desejamos usá-la para produzir alguma característica específica. Por exemplo, se quisermos calcular a média da amostra $(X_1, X_2, ..., X_n)$, esta será dada por

$$\bar{X} = \frac{1}{n}\{X_1 + X_2 + ... + X_n\}.$$

É fácil verificar que $\bar{X}$ é também uma variável aleatória. Podemos também estar interessados em qualquer outra característica da amostra, que será sempre uma função do vetor aleatório $(X_1, ..., X_n)$.

Definição. Uma *estatística* é uma característica da amostra, ou seja, uma estatística T é uma função de $X_1, X_2, ..., X_n$.

As estatísticas mais comuns são:

$$\bar{X} = 1/n \sum_{i=1}^{n} X_i \ : \ \text{média da amostra},$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \ : \text{variância da amostra},$$

$$X_{(1)} = \min\,(X_1, X_2, ..., X_n) \ : \text{o menor valor da amostra},$$

$$X_{(n)} = \max(X_1, X_2, \ldots, X_n) \text{ : o maior valor da amostra,}$$

$$W = X_{(n)} - X_{(1)} \text{ : amplitude amostral,}$$

$$X_{(i)} = \text{a } i\text{-ésima maior observação da amostra.}$$

Em geral, como já vimos no Capítulo 3, podemos considerar as *estatísticas de ordem*,

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)},$$

ou seja, os elementos da amostra ordenados.

Outras estatísticas importantes são os quantis (empíricos), $q(p)$, $0 < p < 1$, definidos no Capítulo 3, especialmente os três quartis q_1, q_2 e q_3.

Para facilitar a linguagem usada em Inferência Estatística, iremos diferenciar as características da amostra e da população.

Definição. Um *parâmetro* é uma medida usada para descrever uma característica da população.

Assim, se estivermos colhendo amostras de uma população, identificada pela v.a. X, seriam parâmetros a média $E(X)$ e sua variância $\text{Var}(X)$.

Os símbolos mais comuns são dados na tabela a seguir.

Denominação	População	Amostra
Média	$\mu = E(X)$	$\bar{X} = \sum X_i / n$
Mediana	$\text{Md} = Q_2$	$\text{md} = q_2$
Variância	$\sigma^2 = \text{Var}(X)$	$S^2 = \sum (X_i - \bar{X})^2/(n-1)$
Nº de elementos	N	n
Proporção	p	$\hat{p}$
Quantil	$Q(p)$	$q(p)$
Quartis	Q_1, Q_2, Q_3	q_1, q_2, q_3
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função densidade	$f(x)$	histograma
Função de distribuição	$F(x)$	$F_e(x)$

10.7 Distribuições Amostrais

Vimos na seção 10.3 que o problema da inferência estatística é fazer uma afirmação sobre os parâmetros da população através da amostra. Digamos que nossa afirmação deva ser feita sobre um parâmetro θ da população (por exemplo, a média, a variância ou qualquer outra medida). Decidimos que usaremos uma AAS de n elementos sorteados dessa população. Nossa decisão será baseada na estatística T, que será uma função da amostra $(X_1, X_2, ..., X_n)$, ou seja, $T = f(X_1, ..., X_n)$. Colhida essa amostra, teremos observado um particular valor de T, digamos t_0, e baseados nesse valor é que faremos a afirmação sobre θ, o parâmetro populacional. Veja a Figura 10.1 (a).

A validade da nossa resposta seria melhor compreendida se soubéssemos o que acontece com a estatística T, quando retiramos todas as amostras de uma população conhecida segundo o plano amostral adotado. Isto é, qual a distribuição de T quando $(X_1, ..., X_n)$ assume todos os valores possíveis. Essa distribuição é chamada *distribuição amostral da estatística T* e desempenha papel fundamental na teoria da inferência estatística. Esquematicamente, teríamos o procedimento representado na Figura 10.1, onde temos:

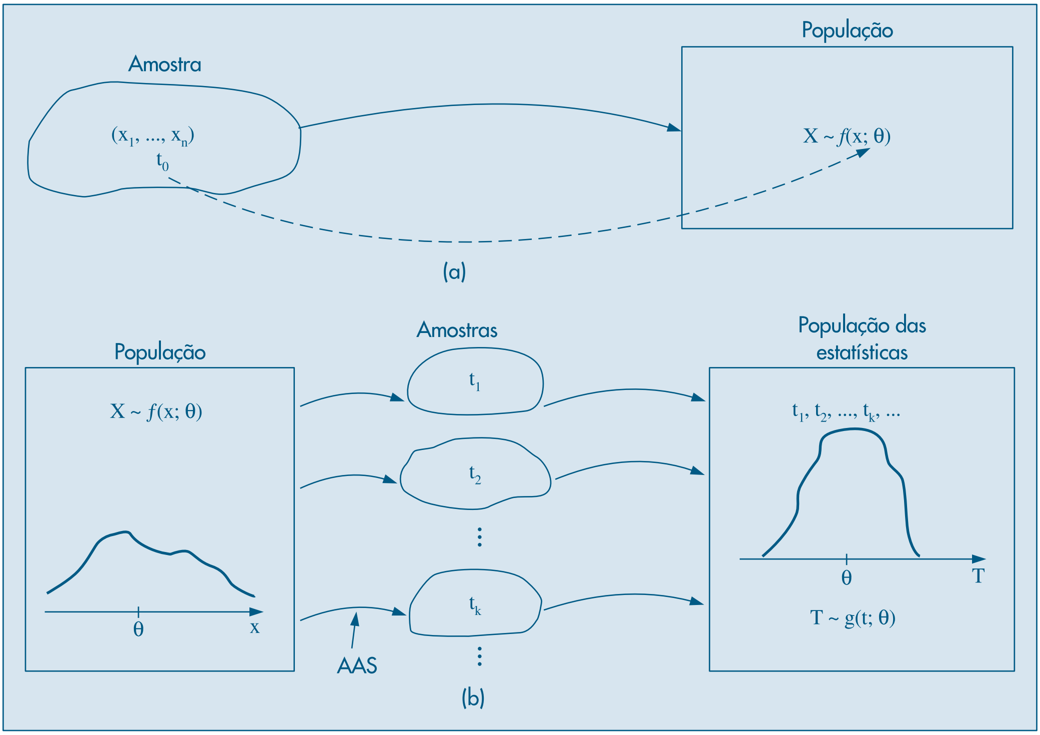
(a) uma população X, com determinado parâmetro de interesse θ;

(b) todas as amostras retiradas da população, de acordo com certo procedimento;

(c) para cada amostra, calculamos o valor t da estatística T; e

(d) os valores t formam uma nova população, cuja distribuição recebe o nome de distribuição amostral de T.

Figura 10.1: (a) Esquema de inferência sobre θ.
(b) Distribuição amostral da estatística T.



Vejamos alguns exemplos simples para aclarar um pouco mais o conceito de distribuição amostral de uma estatística. Nosso principal objetivo é identificar um modelo que explique bem a distribuição amostral de T. É evidente que a distribuição de T irá depender da distribuição de X e do plano amostral, em nosso caso reduzido a AAS.

Exemplo 10.9. Voltemos ao Exemplo 10.7, no qual selecionamos todas as amostras de tamanho 2, com reposição, da população {1, 3, 5, 5, 7}. A distribuição conjunta da variável bidimensional (X_1, X_2) é dada na Tabela 10.2.

Vejamos qual é a distribuição da estatística

$$\overline{X} = \frac{X_1 + X_2}{2}. \tag{10.1}$$

Essa distribuição é obtida por meio da Tabela 10.2. Por exemplo, quando a amostra selecionada é o par (1, 1), a média será 1; então, temos que $P(\overline{X} = 1) = 1/25$. Obteremos a média igual a 3 quando ocorrer o evento $A = \{(1, 5),(3, 3),(5, 1)\}$, logo

$$P(\overline{X} = 3) = P(A) = \frac{2}{25} + \frac{1}{25} + \frac{2}{25} + \frac{5}{25} = \frac{1}{5}.$$

Tabela 10.2: Distribuição das probabilidades das possíveis amostras de tamanho 2 que podem ser selecionadas com reposição da população {1, 3, 5, 5, 7}.

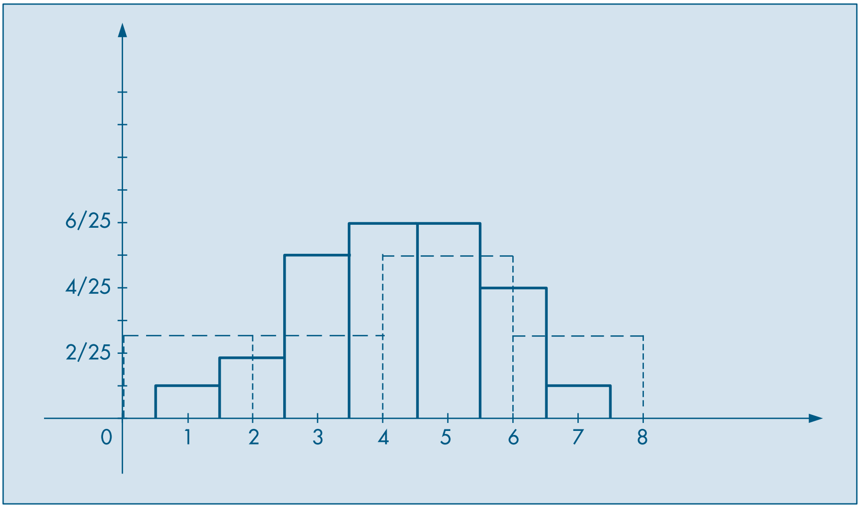
X_2 \ X_1	1	3	5	7	Total
1	1/25	1/25	2/25	1/25	1/5
3	1/25	1/25	2/25	1/25	1/5
5	2/25	2/25	4/25	2/25	2/5
7	1/25	1/25	2/25	1/25	1/5
Total	1/5	1/5	2/5	1/5	1

Procedendo de maneira análoga para os demais valores que $\overline{X}$ pode assumir, obtemos a Tabela 10.3, que dá a distribuição da v.a. $\overline{X}$. Na Figura 10.2 temos as distribuições de X e de $\overline{X}$.

Tabela 10.3: Distribuição amostral da estatística $\overline{X}$.

$\bar{x}$	1	2	3	4	5	6	7	Total
$P(\overline{X} = \bar{x})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25	1,00

Figura 10.2: Distribuição de X(– – – –) e $\overline{X}$(——), obtida de 25 amostras de tamanho 2 de $\{1, 3, 5, 5, 7\}$.



Com um procedimento análogo podemos obter as distribuições amostrais de outras estatísticas de interesse. As Tabelas 10.4 e 10.5 trazem as distribuições amostrais das estatísticas W = amplitude total e $S^2 = \sum(X_i - \overline{X})^2/(n-1)$, respectivamente.

Tabela 10.4: Distribuição amostral de W.

w	0	2	4	6	Total
$P(W = w)$	7/25	10/25	6/25	2/25	1,00

Tabela 10.5: Distribuição amostral de S^2.

s^2	0	2	8	18	Total
$P(S^2 = s^2)$	7/25	10/25	6/25	2/25	1,00

Exemplo 10.5. (continuação) No caso do lançamento de uma moeda 50 vezes, usando como estatística X = número de caras obtidas, a obtenção da distribuição amostral, que já foi vista, é feita por meio do modelo binomial $b(50, p)$, qualquer que seja p = probabilidade de ocorrência de cara num lançamento, $0 < p < 1$. Se estivermos interessados em julgar a "honestidade" da moeda, estaremos verificando se $p = 0{,}5$. Nessas condições, a $P(X \geqslant 36 | n = 50, p = 0{,}5) = 0{,}0013 = 0{,}13\%$.

Portanto, caso a moeda seja honesta, em 50 lançamentos, a probabilidade de se obterem 36 ou mais caras é da ordem de 1 por 1.000. Ou seja, se a moeda fosse honesta, o resultado observado (36 caras) seria muito pouco provável, evidenciando que $p > 0{,}5$.

Comparando os dois últimos exemplos, vemos que nos interessa determinar propriedades das distribuições amostrais que possam ser aplicadas em situações mais gerais (como no caso binomial) e não em situações muito particulares (como no Exemplo 10.7). Iremos, agora, estudar as distribuições amostrais de algumas estatísticas importantes. Nos capítulos seguintes essas distribuições serão usadas para fazer inferências sobre populações.

Quando estivermos trabalhando com populações identificadas pela distribuição de probabilidades, não poderemos gerar *todas as amostras possíveis*. Devemos contentar-nos em simular um número "grande" de amostras e ter uma idéia do que acontece com a estatística de interesse.

Exemplo 10.8. (continuação) Qual seria a distribuição amostral da mediana das alturas de amostras de 5 mulheres retiradas da população $X \sim N(167; 25)$? Como não podemos gerar todas as possíveis amostras de tamanho 5 dessa população, simulamos, via Excel, 200 amostras de tamanho 5 e obtivemos os seguintes resultados:

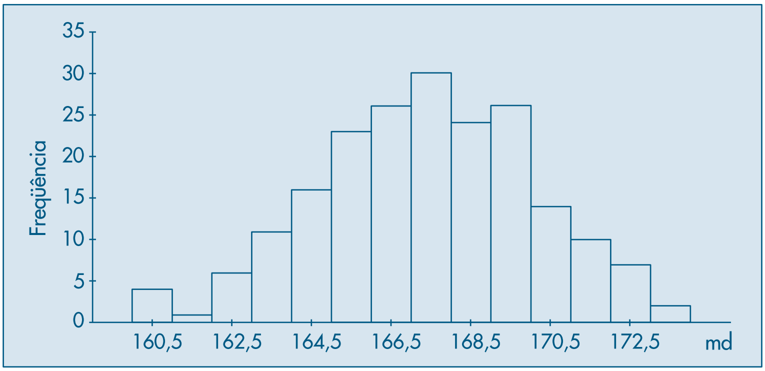
$$E(\text{md}) = 166{,}88, \quad \text{Var(md)} = 7{,}4289, \quad \text{dp(md)} = 2{,}72,$$

$$x_{(1)} = \min(X_1, \ldots, X_{200}) = 160, \quad x_{(200)} = \max(X_1, \ldots, X_{200}) = 173.$$

Observando os resultados somos levados a pensar que a distribuição amostral de md deve ser próxima de uma normal, com média próxima de $\mu = 167$ e desvio padrão menor do que $\sigma = 5$. Veja a Figura 10.3.

Voltaremos a falar na distribuição da mediana amostral em seções futuras.

Figura 10.3: Distribuição amostral da mediana, obtida de 200 amostras de tamanho 5 de $X \sim N(167; 25)$.



Problemas

4. Usando os dados da Tabela 10.2, construa a distribuição amostral da estatística $\hat{\sigma}^2 = \dfrac{\sum (X_i - \overline{X})^2}{n}$.

5. No Problema 3, se X indicar o número de filhos na população, X_1 o número de filhos observados na primeira extração e X_2 na segunda:
   (a) calcule a média e a variância de X;
   (b) calcule $E(X_i)$ e $\text{Var}(X_i)$, $i = 1, 2$;
   (c) construa a distribuição amostral de $\overline{X} = \dfrac{(X_1 + X_2)}{2}$;
   (d) calcule $E(\overline{X})$ e $\text{Var}(\overline{X})$;
   (e) faça num mesmo gráfico os histogramas de X e de $\overline{X}$;
   (f) construa as distribuições amostrais de $S^2 = \sum_{i=1}^{2}(X_i - \overline{X})^2$ e $\hat{\sigma}^2 = \sum_{i=1}^{2}(X_i - \overline{X})^2/2$;
   (g) baseado no resultado de (f), qual dos dois estimadores você usaria para estimar a variância de X? Por quê?
   (h) calcule $P(|\overline{X} - \mu| > 1)$.
6. Ainda com os dados do Problema 3, e para amostras de tamanho 3:
   (a) determine a distribuição amostral de $\overline{X}$ e faça o histograma;
   (b) calcule a média e variância de $\overline{X}$;
   (c) calcule $P(|\overline{X} - \mu| > 1)$.
   (d) se as amostras fossem de tamanho 4, a $P(|\overline{X} - \mu| > 1)$ seria maior ou menor do que a probabilidade encontrada em (c)? Por quê?

10.8 Distribuição Amostral da Média

Vamos estudar agora a distribuição amostral da estatística $\overline{X}$, a média da amostra. Consideremos uma população identificada pela variável X, cujos parâmetros média populacional $\mu = E(X)$ e variância populacional $\sigma^2 = \text{Var}(X)$ são supostos conhecidos. Vamos retirar todas as possíveis AAS de tamanho n dessa população, e para cada uma calcular a média $\overline{X}$. Em seguida, consideremos a distribuição amostral e estudemos suas propriedades. Voltemos a considerar, a título de ilustração, o Exemplo 10.7.

Exemplo 10.10. A população {1, 3, 5, 5, 7} tem média μ = 4,2 e variância σ^2 = 4,16. A distribuição amostral de $\overline{X}$ está na Tabela 10.3, da qual obtemos

$$E(\overline{X}) = \sum_i \bar{x}_i\, p_i = 1 \times \frac{1}{25} + 2 \times \frac{2}{25} + 3 \times \frac{5}{25} + 4 \times \frac{6}{25} + 5 \times \frac{6}{25}$$

$$+ 6 \times \frac{4}{25} + 7 \times \frac{1}{25} = 4{,}2.$$

De modo análogo, encontramos

$$\text{Var}(\overline{X}) = 2{,}08.$$

Verificamos, aqui, dois fatos: primeiro, a média das médias amostrais coincide com a média populacional; segundo, a variância de $\overline{X}$ é igual à variância de X, dividida por $n = 2$. Estes dois fatos não são casos isolados. Na realidade, temos o seguinte resultado.

Teorema 10.1. Seja X uma v.a. com média μ e variância σ^2, e seja $(X_1, \ldots, X_n)$ uma AAS de X. Então,

$$E(\overline{X}) = \mu \quad \text{e} \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}\,.$$

Prova. Pelas propriedades vistas no Capítulo 8, temos:

$$E(\overline{X}) = (1/n)\,\{E(X_1) + \ldots + E(X_n)\}$$
$$= (1/n)\,\{\mu + \mu + \ldots + \mu\} = n\mu/n = \mu.$$

De modo análogo, e pelo fato de $X_1, \ldots, X_n$ serem independentes, temos

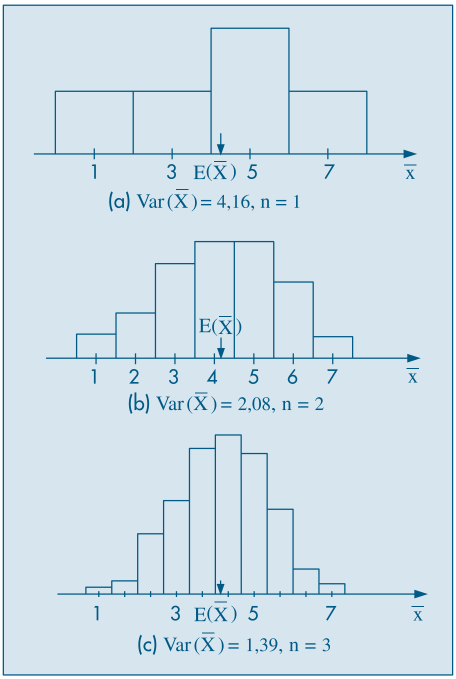
$$\mathrm{Var}(\overline{X}) = (1/n^2)\,\{\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\}$$
$$= (1/n^2)\,\{\sigma^2 + \ldots + \sigma^2\} = n\sigma^2/n^2 = \sigma^2/n.$$

Determinamos, então, a média e a variância da distribuição amostral de $\overline{X}$. Vejamos, agora, como obter informação sobre a forma da distribuição dessa estatística.

Exemplo 10.10. (continuação) Para a população $\{1, 3, 5, 5, 7\}$, vamos construir os histogramas das distribuições de $\overline{X}$ para $n = 1$, 2 e 3.

(i) Para $n = 1$, vemos que a distribuição de $\overline{X}$ coincide com a distribuição de X, com $E(\overline{X}) = E(X) = 4{,}2$ e $\mathrm{Var}(\overline{X}) = \mathrm{Var}(X) = 4{,}16$ (Figura 10.4(a)).

Figura 10.4: Distribuição de $\overline{X}$ para amostras de $\{1, 3, 5, 5, 7\}$.



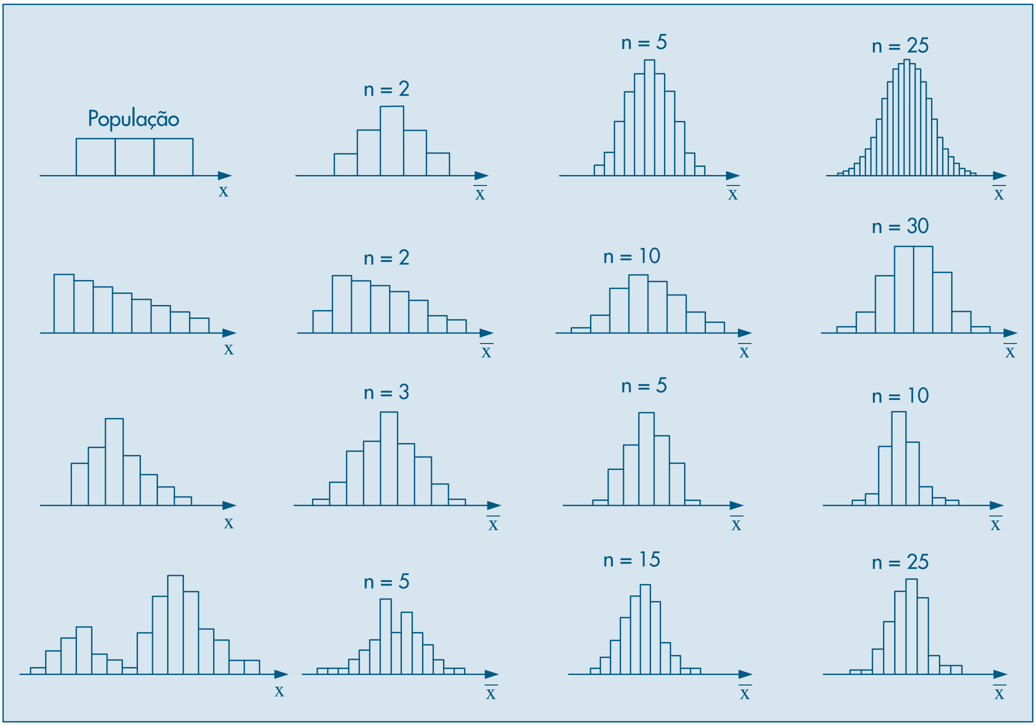
(ii) Para $n = 2$, baseados na Tabela 10.3, temos a distribuição de $\overline{X}$ dada na Figura 10.4(b), com $E(\overline{X}) = 4{,}2$ e $\mathrm{Var}(\overline{X}) = 2{,}08$.

(iii) Finalmente, para $n = 3$, com os dados da Tabela 10.6, temos a distribuição de $\overline{X}$ na Figura 10.4 (c), com $E(\overline{X}) = 4{,}2$ e $\mathrm{Var}(\overline{X}) = 1{,}39$.

Observe que, conforme n vai aumentando, o histograma tende a se concentrar cada vez mais em torno de $E(\overline{X}) = E(X) = 4{,}2$, já que a variância vai diminuindo. Os casos extremos passam a ter pequena probabilidade de ocorrência. Quando n for suficientemente grande, o histograma alisado aproxima-se de uma distribuição normal. Essa aproximação pode ser verificada analisando-se os gráficos da Figura 10.5, que mostram o comportamento do histograma de $\overline{X}$ para várias formas da distribuição da população e vários valores do tamanho da amostra n.

Esses exemplos sugerem que, quando o tamanho da amostra aumenta, independentemente da forma da distribuição da população, a distribuição amostral de $\overline{X}$ aproxima-se cada vez mais de uma distribuição normal. Esse resultado, fundamental na teoria da Inferência Estatística, é conhecido como *Teorema Limite Central* (TLC).

Figura 10.5: Histogramas correspondentes às distribuições amostrais de $\overline{X}$ para amostras extraídas de algumas populações.



Teorema 10.2. (TLC) Para amostras aleatórias simples $(X_1, \ldots, X_n)$, retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral da média $\overline{X}$ aproxima-se, para n grande, de uma distribuição normal, com média μ e variância σ^2/n.

A demonstração completa desse teorema exigiria recursos dos quais não dispomos, portanto não será dada, mas o importante é sabermos como esse resultado pode ser usado.

Observemos que, se a população for normal, então $\overline{X}$ terá distribuição *exata* normal. Esse resultado segue do fato de que a distribuição de uma combinação linear de v.a.'s normais independentes tem ainda distribuição normal. No caso da $\overline{X}$, a média e variância dessa normal serão dadas pelo Teorema 10.1. A prova dessa propriedade depende do conceito de função geradora de momentos, que não será objeto deste livro. O leitor interessado pode consultar Meyer (1965), por exemplo.

Exemplo 10.11. Voltemos ao Exemplo 10.4, onde uma máquina enchia pacotes cujos pesos seguiam uma distribuição $N(500, 100)$. Colhendo-se um amostra de $n = 100$ pacotes e pesando-os, pelo que foi dito acima, $\overline{X}$ terá uma distribuição normal com média 500 e variância 100/100 = 1. Logo, se a máquina estiver regulada, a probabilidade de encontrarmos a média de 100 pacotes diferindo de 500 g de menos de 2 gramas será

$$P(|\overline{X} - 500| < 2) = P(498 < \overline{X} < 502) = P(-2 < Z < 2) \approx 95\%.$$

Ou seja, dificilmente 100 pacotes terão uma média fora do intervalo (498, 502). Caso 100 pacotes apresentem uma média fora desse intervalo, podemos considerar como um evento raro, e será razoável supor que a máquina esteja desregulada.

Outra maneira de apresentar o TLC é por meio do

Corolário 10.1. Se $(X_1, ..., X_n)$ for uma amostra aleatória simples da população X, com média μ e variância σ^2 finita, e $\overline{X} = (X_1 + ... + X_n)/n$, então

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1). \tag{10.2}$$

Basta notar que se usou a transformação usual de reduzir a distribuição de $\overline{X}$ a uma normal padrão. Observe, também, que (10.2) pode ser escrita como

$$Z = \frac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma} \sim N(0, 1). \tag{10.3}$$

Chamemos de e a v.a. que mede a diferença entre a estatística $\overline{X}$ e o parâmetro μ, isto é, $e = \overline{X} - \mu$; e é chamado o *erro amostral da média*. Então, temos o

Corolário 10.2. A distribuição de e aproxima-se de uma distribuição normal com média 0 e variância σ^2/n, isto é,

$$\frac{\sqrt{n}\, e}{\sigma} \sim N(0, 1). \tag{10.4}$$

O TLC afirma que $\overline{X}$ aproxima-se de uma normal quando n tende para o infinito, e a rapidez dessa convergência (veja a Figura 10.5) depende da distribuição da população da qual a amostra é retirada. Se a população original tem uma distribuição próxima da normal, a convergência é rápida; se a população original se afasta muito de uma normal, a convergência é mais lenta, ou seja, necessitamos de uma amostra maior para que $\bar{X}$ tenha uma distribuição aproximadamente normal. Para amostras da ordem de 30 ou 50 elementos, a aproximação pode ser considerada boa.

Problemas

7. Uma v.a. X tem distribuição normal, com média 100 e desvio padrão 10.
   (a) Qual a $P(90 < X < 110)$?
   (b) Se $\bar{X}$ for a média de uma amostra de 16 elementos retirados dessa população, calcule $P(90 < \bar{X} < 110)$.
   (c) Represente, num único gráfico, as distribuições de X e $\bar{X}$.
   (d) Que tamanho deveria ter a amostra para que $P(90 < \bar{X} < 110) = 0{,}95$?

8. A máquina de empacotar um determinado produto o faz segundo uma distribuição normal, com média μ e desvio padrão 10 g.
   (a) Em quanto deve ser regulado o peso médio μ para que apenas 10% dos pacotes tenham menos do que 500 g?
   (b) Com a máquina assim regulada, qual a probabilidade de que o peso total de 4 pacotes escolhidos ao acaso seja inferior a 2 kg?

9. No exemplo anterior, e após a máquina estar regulada, programou-se uma carta de controle de qualidade. De hora em hora, será retirada uma amostra de quatro pacotes e esses serão pesados. Se a média da amostra for inferior a 495 g ou superior a 520 g, encerra-se a produção para reajustar a máquina, isto é, reajustar o peso médio.
   (a) Qual é a probabilidade de ser feita uma parada desnecessária?
   (b) Se o peso médio da máquina desregulou-se para 500 g, qual é a probabilidade de continuar a produção fora dos padrões desejados?

10. A capacidade máxima de um elevador é de 500 kg. Se a distribuição X dos pesos dos usuários for suposta $N(70, 100)$:
    (a) Qual é a probabilidade de sete passageiros ultrapassarem esse limite?
    (b) E seis passageiros?

10.9 Distribuição Amostral de uma Proporção

Vamos considerar uma população em que a proporção de elementos portadores de certa característica é p. Logo, podemos definir uma v.a. X, da seguinte maneira:

$$X = \begin{cases} 1, & \text{se o indivíduo for portador da característica} \\ 0, & \text{se o indivíduo não for portador da característica,} \end{cases}$$

logo,

$$\mu = E(X) = p, \quad \sigma^2 = \text{Var}(X) = p(1-p).$$

Retirada uma AAS dessa população, e indicando por Y_n o total de indivíduos portadores da característica na amostra, já vimos que

$$Y_n \sim b(n, p).$$

Vamos definir por $\hat{p}$ a proporção de indivíduos portadores da característica na amostra, isto é,

$$\hat{p} = \frac{Y_n}{n}.$$

Então,

$$P(Y_n = k) = P(Y_n/n = k/n) = P(\hat{p} = k/n),$$

ou seja, a distribuição amostral de $\hat{p}$ é obtida da distribuição de Y_n.

Vimos na seção 7.5 que a distribuição binomial pode ser aproximada pela distribuição normal. Vamos mostrar que a justificativa desse fato está no TLC. Inicialmente, observe que

$$Y_n = X_1 + X_2 + \ldots + X_n,$$

onde cada X_i tem distribuição de Bernoulli, com média $\mu = p$ e variância $\sigma^2 = p(1 - p)$, e são duas a duas independentes. Podemos escrever que

$$Y_n = n\overline{X},$$

mas pelo TLC, $\overline{X}$ terá distribuição aproximadamente normal, com média p e variância $\frac{p(1-p)}{n}$, ou seja,

$$\overline{X} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Logo, a transformação $Y_n = n\overline{X}$ terá a distribuição

$$Y_n \sim N(np, np(1 - p)),$$

que foi a aproximação adotada na seção 7.5.

Observe que $\overline{X}$, na expressão acima, é a própria variável $\hat{p}$ e, desse modo, para n grande podemos considerar a distribuição amostral de p como aproximadamente normal:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Exemplo 10.12. Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de $n = 10$ estudantes e calculamos $\hat{p}$ = proporção de mulheres na amostra. Qual a probabilidade de que $\hat{p}$ difira de p em menos de 0,01? Temos que essa probabilidade é dada por

$$P(|\hat{p} - p| < 0{,}01) = P(-0{,}01 < \hat{p} - p < 0{,}01).$$

Mas, $\hat{p} - p \sim N\left(0, \dfrac{p(1-p)}{n}\right)$, e como $p = 0{,}3$, temos que

$$\mathrm{Var}(\hat{p}) = (0,3)(0,7)/10 = 0{,}021,$$

e, portanto, a probabilidade pedida é igual a

$$P\left(\frac{-0{,}01}{\sqrt{0{,}021}} < Z < \frac{0{,}01}{\sqrt{0{,}021}}\right) = P(-0{,}07 < Z < 0{,}07) = 0{,}056.$$

Problemas

11. Sabe-se que 20% das peças de um lote são defeituosas. Sorteiam-se oito peças, com reposição, e calcula-se a proporção $\hat{p}$ de peças defeituosas na amostra.
    (a) Construa a distribuição exata de $\hat{p}$ (use a tábua da distribuição binomial).
    (b) Construa a aproximação normal à binomial.
    (c) Você pensa que a segunda distribuição é uma boa aproximação da primeira?
    (d) Já sabemos que, para dado p fixo, a aproximação melhora à medida que n aumenta. Agora, se n for fixo, para qual valor de p a aproximação é melhor?

12. Um procedimento de contrôle de qualidade foi planejado para garantir um máximo de 10% de itens defeituosos na produção. A cada 6 horas sorteia-se uma amostra de 20 peças e, havendo mais de 15% de defeituosas, encerra-se a produção para verificação do processo. Qual a probabilidade de uma parada desnecessária?

13. Supondo que a produção do exemplo anterior esteja sob controle, isto é, $p = 10\%$, e que os itens sejam vendidos em caixas com 100 unidades, qual a probabilidade de que uma caixa:
    (a) tenha mais do que 10% de defeituosos?
    (b) não tenha itens defeituosos?

10.10 Outras Distribuições Amostrais

Do mesmo modo que estudamos a distribuição amostral de $\overline{X}$, podemos, em princípio, estudar a distribuição amostral de qualquer estatística $T = f(X_1, \ldots, X_n)$. Mas, quanto mais complexa for essa relação f, mais difícil será a derivação matemática das propriedades dessa estatística. Vejamos alguns exemplos.

Exemplo 10.13. Na Tabela 10.6 apresentamos a distribuição de três outras estatísticas; a variância da amostra,

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

a mediana amostral, md, e o estimador

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

que difere de S^2 apenas no denominador, e que foi estudado no Capítulo 3. Desta tabela, obtemos as distribuições amostrais apresentadas nas Tabelas 10.7, 10.8 e 10.9.

Tabela 10.6: Distribuição amostral de algumas estatísticas obtidas de amostra de tamanho $n = 3$, retiradas da população $\{1, 3, 5, 5, 7\}$ ($\mu = 4{,}2$, $\sigma^2 = 4{,}16$ e $Md = 5$).

Tipo de amostra	Freqüência (prob. × 125)	Soma	Soma dos quadrados	Média $\bar{x}$	Mediana md	Variância	
						s^2	$\hat{\sigma}^2$
111	1	3	3	1,00	1	0	0
113	3	5	11	1,67	1	4/3	8/9
115	6	7	27	2,33	1	16/3	32/9
117	3	9	51	3,00	1	12	8
133	3	7	19	2,33	3	4/3	8/9
135	12	9	35	3,00	3	4	8/3
137	6	11	59	3,67	3	28/3	56/9
155	12	11	51	3,67	5	16/3	32/9
157	12	13	75	4,33	5	28/3	56/9
177	3	15	99	5,00	7	12	8
333	1	9	27	3,00	3	0	0
335	6	11	43	3,67	3	4/3	8/9
337	3	13	67	4,33	3	16/3	32/9
355	12	13	59	4,33	5	4/3	8/9
357	12	15	83	5,00	5	4	8/3
377	3	17	107	5,67	7	16/3	32/9
555	8	15	75	5,00	5	0	0
557	12	17	99	5,67	5	4/3	8/9
577	6	19	123	6,33	7	4/3	8/9
777	1	21	147	7,00	7	0	0
Total	125						

Tabela 10.7: Distribuição amostral da variância S^2, para amostras de tamanho 3, retiradas da população $\{1, 3, 5, 5, 7\}$.

s^2	0,00	1,33	4,00	5,33	9,33	12,00
$P(S^2 = s^2)$	11/125	42/125	24/125	24/125	18/125	6/125

$$E(S^2) = 4{,}16, \quad \mathrm{Var}(S^2) = 11{,}28.$$

Tabela 10.8: Distribuição amostral da mediana da amostra md para amostras de tamanho 3, retiradas da população $\{1, 3, 5, 5, 7\}$.

md	1	3	5	7
Prob.	13/125	31/125	68/125	13/125

$$E(md) = 4{,}30, \quad \mathrm{Var}(md) = 2{,}54.$$

Tabela 10.9: Distribuição amostral da variância $\hat{\sigma}^2$, para amostras de tamanho 3, retiradas da população $\{1, 3, 5, 5, 7\}$.

$\hat{\sigma}^2$	0,00	0,89	2,67	3,56	6,22	8,00
Prob.	11/125	42/125	24/125	24/125	18/125	6/125

$$E(\hat{\sigma}^2) = 2{,}77, \quad \mathrm{Var}(\hat{\sigma}^2) = 5{,}04.$$

Os gráficos das funções de probabilidade estão nas Figuras 10.6, 10.7 e 10.8. A obtenção das propriedades dessas estatísticas, de modo geral, não é uma tarefa fácil, e os modelos de probabilidade resultantes correspondem a distribuições mais complexas.

Figura 10.6: Distribuição amostral de S^2 para amostras de tamanho $n = 3$ extraídas de $\{1, 3, 5, 5, 7\}$.

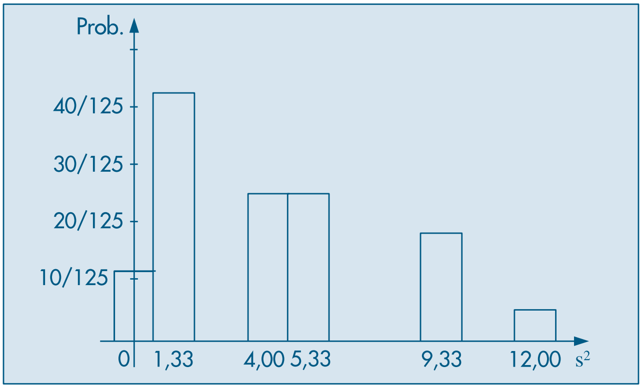


Figura 10.7: Distribuição amostral de md para amostras de tamanho $n = 3$ de $\{1, 3, 5, 5, 7\}$.

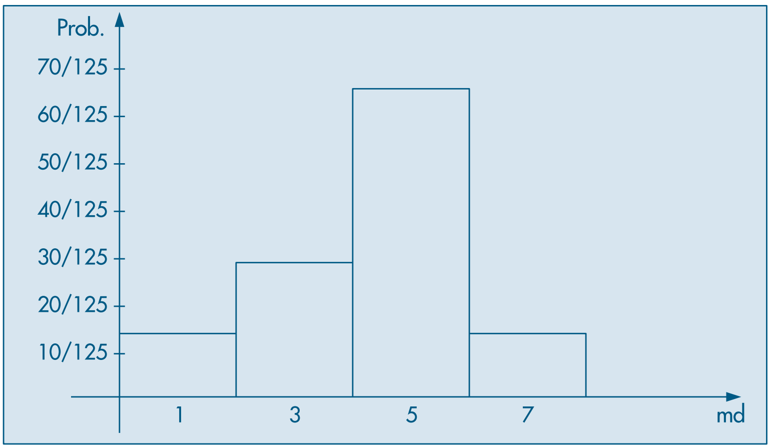
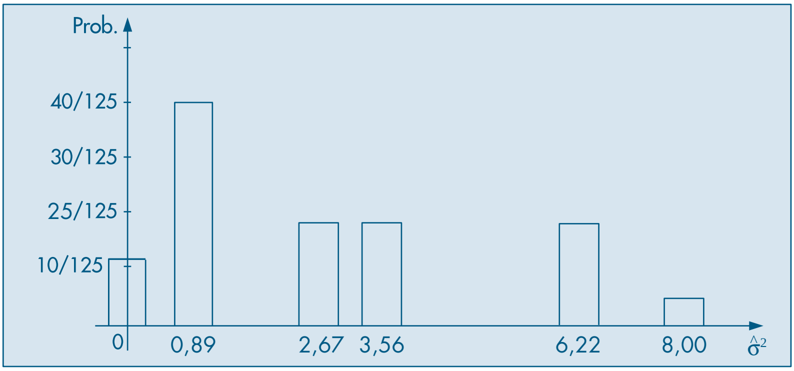


Figura 10.8: Distribuição amostral de $\hat{\sigma}^2$ para amostras de tamanho $n = 3$ extraídas de $\{1, 3, 5, 5, 7\}$.



Por exemplo, note que $E(S^2) = 4{,}16 = \sigma^2$, logo S^2 satisfaz uma propriedade análoga a $E(\overline{X}) = \mu$; dizemos que $\overline{X}$ e S^2 são estimadores *não-viesados* dos respectivos parâmetros μ e σ^2. Esta propriedade já não vale para md e $\hat{\sigma}^2$, pois $E(md) = 4{,}3$, enquanto $Md = 5{,}0$ e $E(\hat{\sigma}^2) = 2{,}77$ e não 4,16. Vemos que $\hat{\sigma}^2$ sub-estima a verdadeira variância.

Também pode-se demonstrar que S^2 segue uma distribuição que é um múltiplo de uma distribuição qui-quadrado (χ^2), quando a população tem distribuição normal. Ver a seção 11.9. Já a mediana *md*, obtida de amostras de uma população simétrica, com média μ e variância σ^2, segue aproximadamente uma distribuição normal, com média $E(md) = \mu$ e $\text{Var}(md) = (\pi\sigma^2)/(2n)$. Note que se exigem mais suposições do que aquelas mencionada no TLC. Nos Capítulos 11 e 12 voltaremos a discutir algumas distribuições amostrais e suas aplicações.

Problemas

14. Usando os dados da Tabela 10.2:
    - (a) construa a distribuição amostral de $\hat{\sigma}^2$ e compare com a distribuição amostral de S^2 (Tabela 10.5). Você notou alguma propriedade de S^2 que seja "melhor" do que de $\hat{\sigma}^2$?
    - (b) seja U a média de elementos distintos de amostras de tamanho $n = 3$. Por exemplo, se a amostra observada for (1, 1, 3), então $u = (1 + 3)/2 = 2$. Construa a distribuição amostral de U;
    - (c) compare as distribuições amostrais de U e $\overline{X}$.

15. Na tabela abaixo tem-se a distribuição dos salários da Secretaria A.

Classes de salários	Freqüência relativa
4,5 ⊢ 7,5	0,10
7,5 ⊢ 10,5	0,20
10,5 ⊢ 13,5	0,40
13,5 ⊢ 16,5	0,20
16,5 ⊢ 19,5	0,10

(a) Calcule a média, a variância e a mediana dos salários nessa população.
(b) Construa a distribuição amostral da média e da mediana para amostras de tamanho 2, retiradas dessa população.
(c) Mostre que a média $\overline{X}$ e a mediana md da amostra são estimadores não-viesados da mediana Md da população, no sentido que $E(\overline{X}) = E(\text{md}) = \text{Md}$.
(d) Qual dos dois estimadores não-viesados você usaria para estimar Md nesse caso? Por quê?
(e) Baseado na distribuição amostral da média, encontre a distribuição amostral da estatística

$$Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n},$$

para $n = 2$.
(f) Quais são os valores de $E(Z)$ e $\text{Var}(Z)$?
(g) Construa a distribuição amostral da estatística

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

e faça o seu histograma.
(h) Calcule a média e variância de S^2.
(i) Baseando-se nas distribuições amostrais anteriores, determine a distribuição amostral da estatística

$$t = \frac{\overline{X} - \mu}{S}\sqrt{n},$$

e construa seu histograma. Qual o problema encontrado?
(j) Calcule a média e variância de t, quando possível.
(k) Calcule a $P(|t| < 2)$ e $P(|t| < 4{,}30)$.

16. Tente esboçar como ficariam os histogramas das estatísticas abaixo, para amostras de tamanho grande.
(a) S^2 (faça o histograma da distribuição da Tabela 10.5)
(b) $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$ (Veja o Teorema Limite Central)
(c) $t = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$, definida no problema anterior (compare com a expressão e o resultado obtido em (b)).

10.11 Determinação do Tamanho de uma Amostra

Em nossas considerações anteriores fizemos a suposição que o tamanho da amostra, n, era conhecido e fixo. Podemos, em certas ocasiões, querer determinar o tamanho da amostra a ser escolhida de uma população, de modo a obter um erro de estimação previamente estipulado, com determinado grau de confiança.

Por exemplo, suponha que estejamos estimando a média μ populacional e para tanto usaremos a média amostral, $\overline{X}$, baseada numa amostra de tamanho n. Suponha que se queira determinar o valor de n de modo que

$$P(|\overline{X} - \mu| \leqslant \varepsilon) \geqslant \gamma, \tag{10.5}$$

com $0 < \gamma < 1$ e ε é o *erro amostral* máximo que podemos suportar, ambos valores fixados.

Sabemos que $\overline{X} \sim N(\mu, \sigma^2/n)$, logo $\overline{X} - \mu \sim N(0, \sigma^2/n)$ e portanto (10.5) pode ser escrita

$$P(-\varepsilon \leqslant \overline{X} - \mu \leqslant \varepsilon) = P\left(\frac{-\sqrt{n}\,\varepsilon}{\sigma} \leqslant Z \leqslant \frac{\sqrt{n}\,\varepsilon}{\sigma}\right) \approx \gamma,$$

com $Z = (\overline{X} - \mu)\sqrt{n}/\sigma$. Dado γ, podemos obter z_γ da $N(0,1)$, tal que $P(-z_\gamma < Z < z_\gamma) = \gamma$, de modo que

$$\frac{\sqrt{n}\,\varepsilon}{\sigma} = z_\gamma,$$

do que obtemos finalmente

$$n = \frac{\sigma^2 z_\gamma^2}{\varepsilon^2}. \tag{10.6}$$

Note que em (10.6) conhecemos z_γ e ε, mas σ^2 é a variância desconhecida da população. Para podermos ter uma idéia sobre n devemos ter alguma informação prévia sobre σ^2 ou, então, usar uma pequena amostra piloto para estimar σ^2.

Exemplo 10.13. (continuação) Suponha que uma pequena amostra piloto de $n = 10$, extraída de uma população, forneceu os valores $\overline{X} = 15$ e $S^2 = 16$. Fixando-se $\varepsilon = 0{,}5$ e $\gamma = 0{,}95$, temos

$$n = \frac{16 \times (1{,}96)^2}{(0{,}5)^2} \approx 245.$$

No caso de proporções, usando a aproximação normal da seção 10.9 para $\hat{p}$, é fácil ver que (10.6) resulta

$$n = \frac{z_\gamma^2\, p(1-p)}{\varepsilon^2}. \tag{10.7}$$

Como não conhecemos p, a verdadeira proporção populacional, podemos usar o fato de que $p(1-p) \leqslant 1/4$, para todo p, e (10.7) fica

$$n \approx \frac{z_\gamma^2}{4\varepsilon^2}. \tag{10.8}$$

Por outro lado, se tivermos alguma informação sobre p ou pudermos estimá-lo usando uma amostra piloto, basta substituir esse valor estimado em (10.7).

Exemplo 10.14. Suponha que numa pesquisa de mercado estima-se que no mínimo 60% das pessoas entrevistadas preferirão a marca A de um produto. Essa informação é baseada em pesquisas anteriores. Se quisermos que o erro amostral de $\hat{p}$ seja menor do que $\varepsilon = 0{,}03$, com probabilidade $\gamma = 0{,}95$, teremos

$$n \approx \frac{(1{,}96)^2(0{,}6)(0{,}4)}{(0{,}03)^2} = 1.024,$$

na qual usamos o fato de que $p \geqslant 0{,}60$. Veja também os Problemas 19, 20 e 41.

Problemas

17. Suponha que uma indústria farmacêutica deseja saber a quantos voluntários se deva aplicar uma vacina, de modo que a proporção de indivíduos imunizados na amostra difira de menos de 2% da proporção verdadeira de imunizados na população, com probabilidade 90%. Qual o tamanho da amostra a escolher? Use (10.8).

18. No problema anterior, suponha que a indústria tenha a informação de que a proporção de imunizados pela vacina seja $p \geqslant 0{,}80$. Qual o novo tamanho de amostra a escolher? Houve redução?

19. Seja o tamanho de amostra dado por (10.7) e n_0 dado por (10.8). Prove que, para todo p, temos $n \leqslant n_0$. (Use a função $f(p) = p(1-p)$ para sua resposta.)

20. Suponha que haja a informação $p \leqslant p_0 < 0{,}5$, com p_0 conhecida. Se $n_1 = z_\gamma^2 p_0(1-p_0)/\varepsilon^2$, mostre que $n \leqslant n_1 < n_0$. Mostre que essa mesma relação vale se soubermos que $p \geqslant p_0 > 0{,}5$.
    [Sugestão: note que $f(p) = p(1-p)$ é crescente em $[0; 0{,}5]$, atinge o máximo em 0,5 e depois é decrescente em $[0{,}5; 1]$.]

10.12 Exemplos Computacionais

Vimos, no Exemplo 10.7, como escolher todas as possíveis amostras de tamanho $n = 2$, com reposição, da população {1, 3, 5, 5, 7}. Obtemos $5^2 = 25$ amostras. Como já salientamos em seções anteriores, ao escolher uma amostra de uma população, estamos na realidade gerando valores de uma v.a. com determinada distribuição de probabilidades, supostamente conhecida. No exemplo, podemos pensar na v.a. X, assumindo os valores $x_1 = 1$, $x_2 = 3$, $x_3 = 5$, $x_4 = 5$, $x_5 = 7$, com probabilidades todas iguais a 0,2. Portanto, para escolher uma amostra de tamanho $n = 2$, basta gerar dois valores dessa distribuição, como aprendemos no Capítulo 9.

Os programas Excel, SPlus e Minitab têm comandos apropriados para gerar amostras de uma população especificada.

Exemplo 10.15. O Excel usa a opção *Amostragem*, dentro de "Análise de Dados" do menu "Ferramentas". Na coluna G do quadro do Exemplo 9.5, temos uma amostra aleatória simples (com reposição), de tamanho $n = 5$ da população $\mathcal{P} = \{1, 2, ..., 10\}$, que está na coluna F.

Capítulo 11

Estimação

11.1 Primeiras Idéias

Vimos que a Inferência Estatística tem por objetivo fazer generalizações sobre uma população, com base nos dados de uma amostra. Salientamos que dois problemas básicos nesse processo são:

(a) estimação de parâmetros; e

(b) teste de hipóteses sobre parâmetros.

Lembremos que *parâmetros* são funções de valores populacionais, enquanto *estatísticas* são funções de valores amostrais.

O problema do teste de hipóteses sobre parâmetros de uma população será tratado no Capítulo 12. Neste capítulo iremos discutir as idéias básicas sobre estimação. Para ilustrar, consideremos o exemplo seguinte.

Exemplo 11.1. Uma amostra de n = 500 pessoas de uma cidade é escolhida, e a cada pessoa da amostra é feita uma pergunta a respeito de um problema municipal, para o qual foi apresentada uma solução pela prefeitura. A resposta à pergunta poderá ser SIM (favorável à solução) ou NÃO (contrária à solução). Deseja-se estimar a proporção de pessoas na cidade favoráveis à solução apresentada.

Se 300 pessoas responderam SIM à pergunta, então uma estimativa natural para essa proporção seria 300/500 ou 60%. Nossa resposta é baseada na suposição de que a amostra é representativa da população. Sabemos, também, que outra amostra poderia levar a outra estimativa. Conhecer as propriedades desses *estimadores* é um dos propósitos mais importantes da Inferência Estatística. Vejamos o que pode ser feito nesse caso particular.

Definamos as v.a. $X_1, \ldots, X_n$, tais que:

$$X_i = \begin{cases} 1, & \text{se a i-ésima pessoa na amostra responder SIM,} \\ 0, & \text{se a i-ésima pessoa na amostra responder NÃO,} \end{cases}$$

e seja $p = P$ (sucesso), onde aqui *sucesso* significa resposta SIM à questão formulada.

Portanto, se $Y_n = \sum_{i=1}^{n} X_i$, sabemos que Y_n tem distribuição binomial com parâmetros n e p, e o problema consiste em estimar p. É claro que Y_n representa o número de pessoas na amostra que responderam SIM; portanto, um possível *estimador* de p é

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\text{número de SIM}}{\text{número de indivíduos}}. \tag{11.1}$$

Então, se $Y_n = k$, isto é, observarmos o valor k da variável Y_n, obteremos $\hat{p} = k/n$ como uma *estimativa* de p. Observe que $\hat{p}$, dado por (11.1), é uma v.a., ao passo que k/n é um número, ou seja, um valor da v.a. No exemplo acima, uma estimativa é 0,6 ou 60%.

O estimador $\hat{p}$ teve sua distribuição amostral estudada na seção 10.9. De lá podemos concluir que $\hat{p}$ tem distribuição aproximadamente normal, com parâmetros:

$$E(\hat{p}) = p, \tag{11.2}$$

$$\text{Var}(\hat{p}) = p(1-p)/n. \tag{11.3}$$

Esses resultados nos ajudam a avaliar as qualidades desse estimador. Por exemplo, o resultado (11.2) indica que o estimador $\hat{p}$, em média, "acerta" p. Dizemos que $\hat{p}$ é um estimador *não-viesado* (ou não-viciado) de p. Ou ainda, o resultado (11.3) indica que para amostras grandes, a diferença entre $\hat{p}$ e p tende a ser pequena, pois para $n \to \infty$, $\text{Var}(\hat{p}) \to 0$. Nesse caso, dizemos que $\hat{p}$ é um estimador *consistente* de p. Observe que essas propriedades são válidas para o estimador no conjunto de todas as amostras que poderiam ser extraídas da população. Para uma particular amostra, $\hat{p}$ pode estar distante de p.

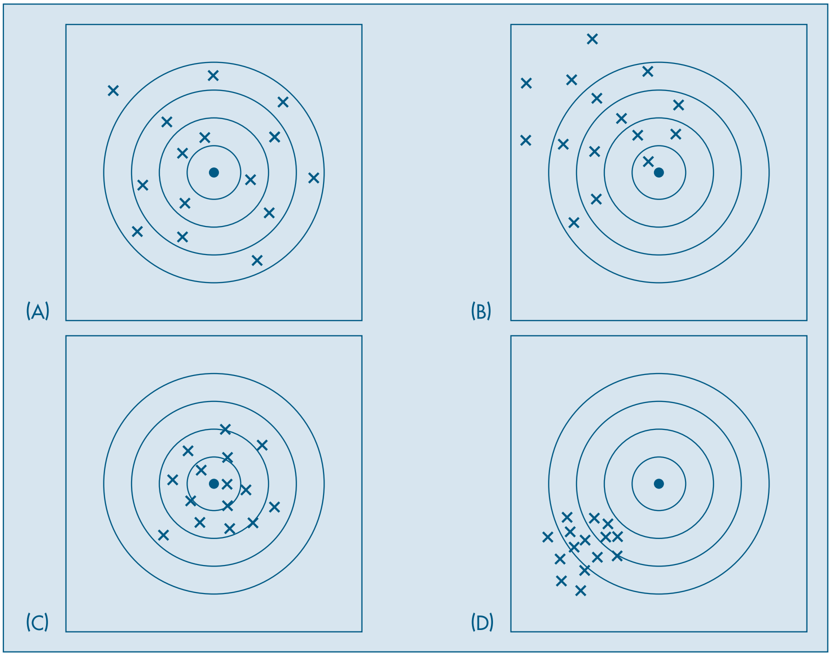
Em algumas situações, podemos ter mais de um estimador para um mesmo parâmetro, e desejamos saber qual deles é "melhor". O julgamento pode ser feito analisando as propriedades desses estimadores. Vejamos um exemplo.

Exemplo 11.2. Desejamos comprar um rifle e, após algumas seleções, restaram quatro alternativas, que chamaremos de rifles A, B, C e D. Foi feito um teste com cada rifle, que consistiu em fixá-lo num cavalete, mirar o centro de um alvo e disparar 15 tiros. Os resultados estão ilustrados na Figura 11.1.

Para analisar qual a melhor arma, podemos fixar critérios. Por exemplo, segundo o critério de "em média acertar o alvo", escolheríamos as armas A e C. Segundo o critério de "não ser muito dispersivo" (variância pequena), a escolha recairia nas armas C e D. A arma C é aquela que reúne as duas propriedades e, segundo esses critérios, seria a melhor arma. Mas, se outro critério fosse introduzido (por exemplo, menor preço), talvez não fosse a arma escolhida. Muitas vezes, a solução deve ser um compromisso entre as propriedades.

Esse exemplo também nos permite introduzir os conceitos de acurácia e precisão.A *acurácia* mede a proximidade de cada observação do valor alvo que se procura atingir. A *precisão* mede a proximidade de cada observação da média de todas as observações.

Figura 11.1: Resultados de 15 tiros dados por 4 rifles.



Desse modo, podemos descrever cada arma da seguinte maneira:

Arma A: não-viesada, pouco acurada e baixa precisão.
Arma B: viesada, pouco acurada e baixa precisão.
Arma C: não-viesada, muito acurada e boa precisão.
Arma D: viesada, pouco acurada e alta precisão.

Do exposto acima, notamos a importância de se definir propriedades desejáveis para estimadores. Trataremos desse assunto na próxima seção. Outro problema que aparece em inferência é como obter um estimador de determinado parâmetro. Nem sempre temos uma sugestão para um estimador, como no caso da proporção, no Exemplo 11.1. Nas seções 11.3, 11.4 e 11.5 trataremos de três desses métodos.

11.2 Propriedades de Estimadores

Inicialmente vejamos a questão da estimação de um modo mais geral. Consideremos uma amostra $(X_1, X_2, ..., X_n)$ de uma v.a. que descreve uma característica de interesse de uma população. Seja θ um parâmetro que desejamos estimar, como por exemplo a média $\mu = E(X)$ ou a variância $\sigma^2 = \mathrm{Var}(X)$.

Definição. Um *estimador* T do parâmetro θ é qualquer função das observações da amostra, ou seja, $T = g(X_1, ..., X_n)$.

Notemos que, segundo essa definição, um estimador é o que chamamos antes de *estatística*, porém associando-o a um parâmetro populacional.

O problema da estimação é, então, determinar uma função $T = g(X_1, X_2, ..., X_n)$ que seja "próxima" de θ, segundo algum critério. O primeiro critério que iremos abordar é dado a seguir.

Definição. O estimador T é *não-viesado* para θ se

$$E(T) = \theta, \tag{11.4}$$

para todo θ.

Se (11.4) não valer T diz-se *viesado* e a diferença V(T) = E(T) – θ é chamado o *viés* de T.

Notemos que a esperança de T em (11.4) é calculada sobre a distribuição amostral de T, como tratada no capítulo anterior.

Definição. *Estimativa* é o valor assumido pelo estimador em uma particular amostra.

Assim, no Exemplo 11.1, $\hat{p}$ é um estimador de p, enquanto 60% é uma estimativa de p.

Exemplo 11.3. Vimos que a média amostral $\overline{X}$ é um estimador não-viesado de $\mu = E(X)$, colhida uma amostra $(X_1, \ldots, X_n)$ da v.a. X. Do mesmo modo, como vimos na seção 10.9, a proporção amostral $\hat{p}$ é um estimador não-viesado da proporção p de indivíduos de uma população que tem certa característica comum.

Exemplo 11.4. Considere uma população com N elementos e a variância populacional

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2, \tag{11.5}$$

onde $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$ é a média populacional. Um possível estimador para σ^2, baseado numa AAS de tamanho n extraída dessa população, é

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{11.6}$$

Mostremos que esse estimador é viesado. Pela fórmula (3.11), temos que

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2,$$

logo

$$E(\hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}^2).$$

Mas, pela definição de AAS e definição de variância de uma v.a., $E(X_i^2) = \text{Var}(X_i) + [E(X_i)]^2 = \sigma^2 + \mu^2$. Também, usando o Teorema 10.1, temos que $E(\overline{X}^2) = \text{Var}(\overline{X}) + [E(\overline{X})]^2 = \frac{\sigma^2}{n} + \mu^2$.

Seque-se que

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^{n} (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right),$$

ou seja,

$$E(\hat{\sigma}^2) = \frac{1}{n}(n(\sigma^2 + \mu^2)) - \frac{\sigma^2}{n} - \mu^2 = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\left(1 - \frac{1}{n}\right).$$

Finalmente,

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2. \tag{11.7}$$

De (11.7) vemos que $\hat{\sigma}^2$ é viesado para σ^2 e o viés é dado por

$$V = V(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = -\frac{\sigma^2}{n}. \tag{11.8}$$

Como esse viés é negativo, o estimador $\hat{\sigma}^2$ em geral subestima o verdadeiro parâmetro σ^2. Por outro lado, por (11.8), o viés diminui com n, ou seja, formalmente, para $n \to \infty$, o viés de $\hat{\sigma}^2$ tende a zero. Note também que o viés de $\hat{\sigma}^2$ é uma função de σ^2. Uma estimativa do viés seria dada por

$$\hat{V} = -\frac{\hat{\sigma}^2}{n},$$

ou seja, substituímos o valor desconhecido de σ^2 por uma estimativa, como por exemplo $\hat{\sigma}^2$.

É fácil ver que para obter um estimador não-viesado de σ^2 basta considerar $(n/(n-1))\hat{\sigma}^2$, pois de (11.7) segue-se que

$$E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2.$$

Logo, se definirmos

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2, \tag{11.9}$$

então $E(S^2) = \sigma^2$ e S^2 é um estimador não-viesado para σ^2. Essa é a razão para se usar $n - 1$, em vez de n, como denominador da variância da amostra. No Capítulo 3 usamos sempre n como denominador, porque não havia preocupação em saber se estávamos trabalhando com uma população ou uma amostra. Daqui por diante, será feita essa distinção.

Vimos que o estimador $\hat{p}$ é não-viesado e tem variância que tende a zero, quando $n \to \infty$. Ver (11.2) e (11.3). Dizemos que $\hat{p}$ é consistente. Esse conceito de consistência é um pouco mais difícil de se definir. Vejamos um exemplo para motivar a definição que será dada.

Considere a média $\overline{X}$ calculada para diversos tamanhos de amostras; obtemos, na realidade, uma seqüência de estimadores $\{\overline{X}_n, n = 1, 2, ...\}$. À medida que n cresce, a distribuição de $\overline{X}_n$ torna-se mais concentrada ao redor da verdadeira média μ. Veja, por exemplo, a Figura 10.4 do Capítulo 10. Dizemos que $\{\overline{X}_n\}$ é uma seqüência *consistente* de estimadores de μ.

Definição. Uma seqüência $\{T_n\}$ de estimadores de um parâmetro θ é *consistente* se, para todo $\varepsilon > 0$,

$$P\{|T_n - \theta| > \varepsilon\} \to 0, \quad n \to \infty. \tag{11.10}$$

Não é muito difícil ver que essa condição está satisfeita para $\{\overline{X}_n\}$. Veja o Problema 34.

Em vez de usar (11.10) para verificar se uma seqüência de estimadores é consistente, podemos usar o seguinte resultado.

Proposição. Uma seqüência $\{T_n\}$ de estimadores de θ é consistente se

$$\lim_{n \to \infty} E(T_n) = \theta \tag{11.11}$$

e

$$\lim_{n \to \infty} \operatorname{Var}(T_n) = 0. \tag{11.12}$$

Se T_n for não-viesado, a primeira condição estará, obviamente, satisfeita. Usando esse resultado, vemos que $\hat{p}$ e $\overline{X}_n$ são estimadores consistentes de p e μ, respectivamente, nos Exemplos 11.1 e 11.3.

Exemplo 11.5. Vimos que S^2, dado por (11.9), é não-viesado para σ^2. É possível demonstrar, no caso que $X_1, \ldots, X_n$ são observações de uma distribuição $N(\mu, \sigma^2)$, que

$$\operatorname{Var}(S^2) = \frac{2\sigma^4}{n-1}. \tag{11.13}$$

Como $E(S^2) = \sigma^2$, e $\lim_{n \to \infty} \operatorname{Var}(S^2) = 0$, segue-se que S^2 é um estimador consistente para σ^2. Dado o que foi dito acima, talvez fosse melhor escrever S_n^2.

Exemplo 11.6. Vimos que $E(\hat{\sigma}^2) = \sigma^2(1 - 1/n)$, de modo que $\lim_{n \to \infty} E(\hat{\sigma}^2) = \sigma^2$. Também, de (11.6) e (11.13) e supondo que as observações são de uma distribuição normal $N(\mu, \sigma^2)$, temos que

$$\operatorname{Var}(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 \operatorname{Var}(S^2) = \frac{n-1}{n^2}(2\sigma^4), \tag{11.14}$$

o que mostra que $\operatorname{Var}(\hat{\sigma}^2) \to 0$, quando $n \to \infty$, logo $\hat{\sigma}^2 = \hat{\sigma}_n^2$ também é consistente para σ^2.

De (11.14) obtemos, também, que

$$\operatorname{Var}(\hat{\sigma}^2) < \frac{2\sigma^4}{n-1} = \operatorname{Var}(S^2). \tag{11.15}$$

Portanto, usando-se somente o critério de "ter menor variância", $\hat{\sigma}^2$ seria um "melhor" estimador de σ^2. Mas observe que estamos nos referindo a amostras de uma distribuição normal.

Vejamos agora um critério que nos permite escolher entre dois estimadores do mesmo parâmetro.

Definição. Se T e T' são dois estimadores não-viesados de um mesmo parâmetro θ, e ainda

$$\mathrm{Var}(T) < \mathrm{Var}(T'), \tag{11.16}$$

então T diz-se mais *eficiente* do que T'.

Exemplo 11.7. Consideremos uma população normal X, com parâmetros μ e σ^2. Queremos estimar a mediana dessa população. Por ser uma distribuição simétrica, sabemos que $\mu = Md(X)$. Definindo como $\overline{X}$ a média e como md a mediana de uma amostra de tamanho n dessa população, qual dos dois estimadores é o melhor para estimar a mediana populacional?

Pelo que vimos no capítulo anterior,

$$\overline{X} \sim N(\mu, \sigma^2/n). \tag{11.17}$$

Pode-se demonstrar que a distribuição da mediana amostral pode ser aproximada por uma normal, especificamente,

$$\mathrm{md} \sim N(Md(X), \pi\sigma^2/2n). \tag{11.18}$$

Vemos, portanto, que os dois estimadores são não-viesados, mas $\overline{X}$ é mais eficiente, pois

$$\mathrm{Var}(\mathrm{md})/\mathrm{Var}(\overline{X}) = \pi/2 > 1.$$

Conclui-se que, para estimar a mediana dessa população, é preferível usar a média da amostra como estimador, o que contraria um pouco a nossa intuição.

Para precisar o conceito de estimador acurado, discutido na seção anterior, vamos agora introduzir o conceito de erro quadrático médio.

Chamemos de

$$e = T - \theta,$$

o *erro amostral* que cometemos ao estimar o parâmetro θ da distribuição da v.a. X pelo estimador $T = g(X_1, \dots, X_n)$, baseado na amostra $(X_1, \dots, X_n)$.

Definição. Chama-se *erro quadrático médio* (EQM) do estimador T ao valor

$$EQM(T; \theta) = E(e^2) = E(T - \theta)^2. \tag{11.19}$$

De (11.19) temos

$$\begin{aligned} EQM(T; \theta) &= E(T - E(T) + E(T) - \theta)^2 \\ &= E(T - E(T))^2 + 2E[(T - E(T))(E(T) - \theta)] + E(E(T) - \theta)^2 \\ &= E(T - E(T))^2 + E(E(T) - \theta)^2, \end{aligned}$$

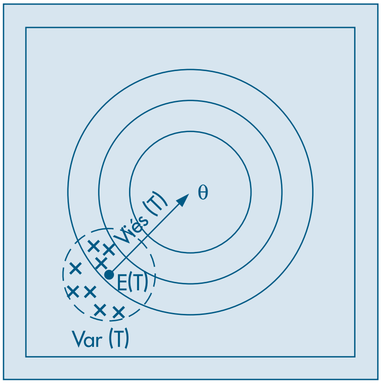
já que $E(T) - \theta$ é uma constante e $E(T - E(T)) = 0$. Podemos, pois, escrever,

$$EQM(T; \theta) = \mathrm{Var}(T) + V^2, \tag{11.20}$$

onde $V = V(T) = E(T) - \theta$ indica, como vimos, o viés de T. A Figura 11.2 ilustra essas duas medidas, usando o caso das armas discutido no Exemplo 11.2.

Vemos, portanto, que um estimador preciso tem variância pequena, mas pode ter EQM grande.

Figura 11.2: Representação gráfica para o EQM.



Problemas

1. Obtenha a distribuição de $\hat{p}$ quando $p = 0{,}2$ e $n = 5$. Depois calcule $E(\hat{p})$ e $\mathrm{Var}(\hat{p})$.

2. Encontre um limite superior para $\mathrm{Var}(\hat{p})$ quando $n = 10$, 25, 100 e 400. Faça o gráfico em cada caso.

3. Suponha um experimento consistindo de n provas de Bernoulli, com probabilidade de sucesso p. Seja X o número de sucessos, e considere os estimadores

   (a) $\hat{p}_1 = X/n$; (b) $\hat{p}_2 = \begin{cases} 1, & \text{se a primeira prova resultar sucesso,} \\ 0, & \text{caso contrário.} \end{cases}$

   Determine a esperança e a variância de cada estimador. Por que $\hat{p}_2$ não é um "bom" estimador?

4. Verifique se $\hat{p}_1$ e $\hat{p}_2$ do Problema 3 são consistentes.

5. Tem-se duas fórmulas distintas para estimar um parâmetro populacional θ. Para ajudar a escolher o melhor, simulou-se uma situação onde $\theta = 100$. Dessa população retiraram-se 1.000 amostras de dez unidades cada uma, e aplicaram-se ambas as fórmulas às dez unidades de cada amostra. Desse modo obtêm-se 1.000 valores para a primeira fórmula t_1 e outros 1.000 valores para a segunda fórmula t_2, cujos estudos descritivos estão resumidos abaixo. Qual das duas fórmulas você acha mais conveniente para estimar θ. Por quê?

Fórmula 1	Fórmula 2
$\bar{t}_1 = 102$ $\mathrm{Var}(t_1) = 5$ Mediana = 100 Moda = 98	$\bar{t}_2 = 100$ $\mathrm{Var}(t_2) = 10$ Mediana = 100 Moda = 100

11.3 Estimadores de Momentos

Neste capítulo e em anteriores, temos usado certos estimadores de parâmetros populacionais, como a média e a variância, simplesmente tentando "imitar" na amostra o que acontece na população. Foi assim que construímos $\overline{X}$, por exemplo.

A média populacional é um caso particular daquilo que chamamos de momento. Na realidade, ela é o *primeiro momento*. Se X for uma v.a. contínua, com densidade $f(x;\ \theta_1, \ldots, \theta_r)$, dependendo de r parâmetros, então

$$\mu_1 = E(X) = \int_{-\infty}^{\infty} x f(x;\ \theta_1, \ldots, \theta_r) dx. \tag{11.21}$$

Essa média dependerá, genericamente, dos parâmetros desconhecidos $\theta_1, \ldots, \theta_r$. Por exemplo, suponha que X tenha distribuição normal, com parâmetros μ e σ^2. Aqui, $\theta_1 = \mu$, $\theta_2 = \sigma^2$ e $r = 2$. Temos, nesse caso, que $E(X) = \mu$.

Podemos, em geral, definir o *k-ésimo momento* de X por

$$\mu_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x;\ \theta_1, \ldots, \theta_r) dx, \quad k = 1, 2, \ldots \tag{11.22}$$

Assim, para $k = 2$, obtemos o segundo momento

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x;\ \theta_1, \ldots, \theta_r) dx.$$

No caso acima da normal, temos que $E(X^2) = \text{Var}(X) + [E(X)]^2 = \sigma^2 + \mu^2$. Suponha, agora, que colhemos uma amostra de tamanho n da população $(X_1, \ldots, X_n)$. Definimos o chamado *k-ésimo momento amostral* por

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k, \quad k = 1, 2, \ldots \tag{11.23}$$

Temos, portanto, que $m_1 = \overline{X}$ e $m_2 = \sum_{i=1}^{m} X_i^2 / n$.

Definição. Dizemos que $\hat{\theta}_1, \ldots, \hat{\theta}_r$ são estimadores obtidos pelo método dos momentos se eles forem soluções das equações

$$m_k = \mu_k, \quad k = 1, 2, \ldots, r. \tag{11.24}$$

O procedimento consiste em substituir os momentos teóricos pelos respectivos momentos amostrais.

Exemplo 11.8. Se $X \sim N(\mu, \sigma^2)$, teremos as seguintes relações válidas para os dois primeiros momentos populacionais:

$$E(X) = \mu, \quad E(X^2) = \sigma^2 + \mu^2,$$

do que obtemos

$$\mu = E(X), \quad \sigma^2 = E(X^2) - E^2(X).$$

Temos, também, os dois primeiros momentos amostrais:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Os estimadores obtidos pelo método dos momentos serão

$$\hat{\mu}_M = m_1 = \overline{X},$$

$$\hat{\sigma}_M^2 = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \hat{\sigma}^2.$$

Ou seja, obtemos os já mencionados estimadores $\overline{X}$ e $\hat{\sigma}^2$.

Na realidade, podemos ter, às vezes, mais de um estimador de momentos. Suponha, por exemplo, que a v.a. Y tenha uma distribuição de Poisson com parâmetro $\lambda > 0$. Vimos que $E(Y) = \text{Var}(Y) = \lambda$, de modo que λ pode ser estimado por $\overline{Y}$ ou por $\sum_{i=1}^{n}(Y_i - \overline{Y})^2/n$, ou seja, $\hat{\lambda}_M = \overline{X}$ ou $\hat{\lambda}_M = \hat{\sigma}^2$, que podem resultar em valores muito diferentes. Veja o Problema 46.

11.4 Estimadores de Mínimos Quadrados

Um dos procedimentos mais usados para obter estimadores é aquele que se baseia no princípio dos mínimos quadrados, introduzido por Gauss em 1794, mas que primeiro apareceu com esse nome no apêndice do tratado de Legendre, *Nouvelles Méthodes pour la Determination des Orbites des Comètes*, publicado em Paris em 1806. Gauss somente viria a publicar seus resultados em 1809, em Hamburgo. Ambos utilizaram o princípio em conexão com problemas de Astronomia e Física.

Vejamos o procedimento por meio de um exemplo simples.

Exemplo 11.9. Um engenheiro está estudando a resistência Y de uma fibra em função de seu diâmetro X e notou que as variáveis são aproximadamente proporcionais, isto é, elas obedecem à relação

$$Y \approx \theta X, \tag{11.25}$$

onde θ é o coeficiente de proporcionalidade. Agora ele deseja estimar o parâmetro θ, baseado numa amostra de cinco unidades, que, submetidas a mensuração e testes, produziram os resultados:

$$X: \quad 1{,}2 \quad 1{,}5 \quad 1{,}7 \quad 2{,}0 \quad 2{,}6, \quad \overline{X} = 1{,}8;$$

$$Y: \quad 3{,}9 \quad 4{,}7 \quad 5{,}6 \quad 5{,}8 \quad 7{,}0, \quad \overline{Y} = 5{,}4.$$

Inspecionando os resultados, conclui-se que $\hat{\theta} = 3$ parece ser um valor razoável. Como verificar a qualidade dessa estimativa? Podemos utilizar o modelo $\hat{Y} = 3X$ e ver como esse prevê os valores de Y, para os dados valores de X, e como são as discrepâncias entre os valores observados e os estimados pelo modelo. Essa análise está resumida na Tabela 11.1.

Os valores da coluna ($Y - 3X$) medem a inadequação do modelo para cada observação da amostra, enquanto o valor $\sum_{i=1}^{5}(Y_i - 3X_i)^2 = 1{,}06$ é uma tentativa de medir "o erro quadrático total da amostra". Como em situações anteriores, elevou-se ao quadrado para evitar o problema do sinal. Quanto menor for o erro quadrático total, melhor será a estimativa. Isso nos sugere procurar a estimativa que torne mínima essa soma de quadrados. Matematicamente, o problema passa a ser o de encontrar o valor de θ que minimize a função

$$S(\theta) = \sum_{i=1}^{5}(Y_i - \theta X_i)^2. \tag{11.26}$$

Tabela 11.1: Análise do modelo $\hat{Y} = 3X$.

X	Y	$3X$	$Y - 3X$	$(Y - 3X)^2$
1,2	3,9	3,6	0,3	0,09
1,5	4,7	4,5	0,2	0,04
1,7	5,6	5,1	0,5	0,25
2,0	5,8	6,0	–0,2	0,04
2,6	7,0	7,8	0,8	0,64
		Total	0	1,06

O mínimo da função é obtido derivando-a em relação a θ, e igualando o resultado a zero (Ver Morettin et al., 2005), o que resulta

$$\frac{dS(\theta)}{d\theta} = \sum_{i=1}^{5}(Y_i - \hat{\theta}X_i)(-2X_i) = 0.$$

Resolvendo essa equação, obtemos

$$\hat{\theta}_{MQ} = \frac{\sum_{i=1}^{5} X_i Y_i}{\sum_{i=1}^{5} X_i^2}.$$

Usando os dados acima encontramos $\hat{\theta}_{MQ} = 2{,}94$, que conduz a um valor mínimo para $S(\theta)$ de 0,94. Observe que esse valor é realmente menor do que o observado para $\theta = 3$, ou seja, 1,06.

Como foi dito, não esperávamos uma relação perfeita entre as duas variáveis, já que o diâmetro da fibra não é o único responsável pela resistência; outros fatores não controlados afetam o resultado. Desse modo, duas amostras obtidas do mesmo diâmetro X não teriam obrigatoriamente que apresentar o mesmo resultado Y, mas valores em torno de um valor esperado θX.

Em outras palavras, estamos supondo que, para um dado valor da variável explicativa X, os valores da variável resposta Y seguem uma distribuição de probabilidade $f_Y(y)$, centrada em θX. Isso equivale a afirmar que, para cada X, o desvio $\varepsilon = Y - \theta X$ segue uma distribuição centrada no zero. Para melhor entendimento dessa proposição, veja o Capítulo 16. Podemos, então, escrever

$$E(Y \mid x) = \theta x, \quad \text{para todo valor } x.$$

É comum supor que ε tem a mesma distribuição, para todo valor x da variável explicativa X. Desse modo, é comum escrever

$$Y = \theta x + \varepsilon,$$

com ε seguindo a distribuição $f_\varepsilon(.)$, com média zero. Como ilustração, poderíamos supor que $\varepsilon \sim N(0, \sigma^2)$, para todo x. Quanto menor for a variância σ^2, melhor será a "previsão" de Y como função de x. Assim, parece razoável escolher θ que torna mínima a soma dos quadrados do erros:

$$\sum_{i=1}^{5} \varepsilon_i^2 = \sum_{i=1}^{5} (Y_i - \theta X_i)^2.$$

O modelo acima pode ser generalizado, de modo a envolver outras funções do parâmetro θ, resultando no modelo

$$Y = g(X;\, \theta) + \varepsilon, \tag{11.27}$$

e devemos procurar o valor de θ que minimize a função

$$S(\theta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - g(X_i;\, \theta))^2, \tag{11.28}$$

para uma amostra $(X_1, Y_1), \ldots, (X_n, Y_n)$ das variáveis X e Y. A solução $\hat{\theta}_{MQ}$ é chamada de estimador de mínimos quadrados (EMQ) de θ.

Nos Capítulos 15 e 16 voltaremos a esse tópico e trataremos com mais detalhes os chamados modelos lineares.

Problemas

6. Estamos estudando o modelo $y_t = \mu + \varepsilon_t$, para o qual uma amostra de cinco elementos produziu os seguintes valores para y_t: 3, 5, 6, 8, 16.
   (a) Calcule os valores de $S(\mu) = \sum_t (y_t - \mu)^2$, para $\mu = 6, 7, 8, 9, 10$, e faça o gráfico de $S(\mu)$ em relação a μ. Qual o valor de μ que parece tornar mínimo $S(\mu)$?
   (b) Derivando $S(\mu)$ em relação a μ, e igualando o resultado a zero, você encontrará o EMQ de μ. Usando os dados acima, encontre a estimativa para μ e compare com o resultado do item anterior.

7. Os dados abaixo referem-se ao índice de inflação (y_t) de 1967 a 1979.

Ano (t)	1967	1969	1971	1973	1975	1977	1979
Inflação (y_t)	128	192	277	373	613	1.236	2.639

   (a) Faça o gráfico de y_t contra t.
   (b) Considere ajustar o modelo $y_t = \alpha + \beta t + \varepsilon_t$ aos dados. Encontre as estimativas de mínimos quadrados de α e β.
   (c) Qual seria a inflação em 1981?
   (d) Você teria alguma restrição em adotar o modelo linear nesse caso?

8. No Problema 7, determinamos os estimadores de mínimos quadrados para o modelo $y_t = f(t) + \epsilon_t$, no qual $f(t) = \alpha + \beta t$. Suponha agora que

$$f(t) = \alpha + \beta x_t, \quad t = 1, ..., n,$$

ou seja, temos n valores fixos $x_1, ..., x_n$ de uma variável fixa (não-aleatória) x. Obtenha os EMQ de α e β para esse modelo.

9. Aplique os resultados do Problema 8 para os dados a seguir:

t	1	2	3	4	5	6	7	8	9	10
x_t	1,5	1,8	1,6	2,5	4,0	3,8	4,5	5,1	6,5	6,0
y_t	66,8	67,0	66,9	67,6	68,9	68,7	69,3	69,8	71,0	70,6

11.5 Estimadores de Máxima Verossimilhança

O Novo Dicionário Aurélio da Língua Portuguesa (2ª edição, 1986) define *verossímil* (ou verossimilhante) aquilo que é semelhante à verdade, provável, e *verossimilhança* (ou verossimilidade, ou ainda verossimilitude), à qualidade ou caráter de verossímil. O que seria uma amostra verossímil? Seria uma amostra que fornecesse a melhor informação possível sobre um parâmetro de interesse da população, desconhecido, e que desejamos estimar.

O princípio da verossimilhança afirma que devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a "mais provável". O uso desse princípio conduz a um método de estimação pelo qual se obtêm os chamados estimadores de máxima verossimilhança que, em geral, têm propriedades muito boas. Esse princípio foi enunciado por Fisher pela primeira vez em 1912 e, em 1922, deu-lhe forma mais completa, introduzindo a expressão "likelihood" (verossimilhança). Veja Fisher (1935) para mais detalhes. Vamos começar com um exemplo.

Exemplo 11.10. Suponha que temos n provas de Bernoulli com P (sucesso) = p, $0 < p < 1$ e X = número de sucessos. Devemos tomar como estimador aquele valor de p que torna a amostra observada a mais provável de ocorrer.

Suponha, por exemplo, que $n = 3$ e obtemos dois sucessos e um fracasso. A função de verossimilhança é

$$L(p) = P(2 \text{ sucessos e } 1 \text{ fracasso}) = p^2(1 - p).$$

Maximizando essa função em relação a p, obtemos

$$L'(p) = 2p(1 - p) - p^2 = 0 \Rightarrow p(2 - 3p) = 0,$$

do que seguem $p = 0$ ou $p = 2/3$. É fácil ver que o ponto máximo é $\hat{p} = 2/3$, que é o estimador de máxima verossimilhança (EMV) de p.

De modo geral, o EMV do parâmetro p de uma distribuição binomial é

$$\hat{p}_{MV} = \frac{X}{n}, \tag{11.29}$$

que é o estimador usado anteriormente no Exemplo 11.1.

Para chegar a (11.29), observe que a função de verossimilhança nesse caso é

$$L(p) = p^x (1 - p)^{n-x},$$

que é a probabilidade de se obter x sucessos e $n - x$ fracassos. O máximo dessa função ocorre no mesmo ponto que $\ell(p) = \log_e L(p)$. Denotando o logaritmo natural simplesmente por log, temos

$$\ell(p) = x \log p + (n - x) \log(1 - p).$$

Derivando e igualando a zero obtemos $\hat{p}_{MV} = x/n$.

O procedimento, pois, é obter a função de verossimilhança, que depende dos parâmetros desconhecidos e dos valores amostrais, e depois maximizar essa função ou o logaritmo dela, o que pode ser mais conveniente em determinadas situações. Chamando de $L(\theta; X_1, \ldots, X_n)$ a função de verossimilhança, a log-verossimilhança será $\ell(\theta; X_1, \ldots, X_n) = \log_e L(\theta; X_1, \ldots, X_n)$.

No caso de variáveis contínuas, a função de verossimilhança é definida da seguinte maneira. Suponha que a v.a. X tenha densidade $f(x; \theta)$, onde destacamos a dependência do parâmetro θ desconhecido. Retiramos uma amostra de X, de tamanho n, $(X_1, \ldots, X_n)$, e sejam $(x_1, \ldots, x_n)$ os valores efetivamente observados.

Definição. A função de verossimilhança é definida por

$$L(\theta; x_1, \ldots, x_n) = f(x_1; \theta) \ldots f(x_n; \theta), \tag{11.30}$$

que deve ser encarada como uma função de θ. O estimador de máxima verossimilhança de θ é o valor $\hat{\theta}_{MV}$ que maximiza $L(\theta; x_1, \ldots, x_n)$.

Se indicarmos por $\boldsymbol{x} = (x_1, \ldots, x_n)'$ o vetor contendo a amostra, é costume denotar a verossimilhança por $L(\theta|\boldsymbol{x})$ e a log-verossimilhança por $\ell(\theta|\boldsymbol{x})$. O parâmetro θ pode ser um vetor, como no caso de querermos estimar a média μ e a variância σ^2 de uma normal. Nesse caso, $\theta = (\mu, \sigma^2)'$.

Exemplo 11.11. Suponha que a v.a. X tenha distribuição exponencial, com parâmetro $\alpha > 0$, desconhecido, e queremos obter o EMV desse parâmetro. A densidade de X é dada por (7.26):

$$f(x; \alpha) = \begin{cases} \dfrac{1}{\alpha} e^{-x/\alpha}, & \text{se } x \geqslant 0 \\ 0, & \text{se } x < 0. \end{cases}$$

Então, a verossimilhança é dada por

$$L(\alpha|x) = (1/\alpha)^n e^{-\Sigma x_i/\alpha}$$

e a log-verossimilhança fica

$$\ell(\alpha|x) = -n \log \alpha - \sum_{i=1}^{n} x_i/\alpha.$$

Derivando e igualando a zero obtemos que o EMV de α é

$$\hat{\alpha}_{MV} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{11.31}$$

que nada mais é do que a média amostral. Lembremos que na distribuição exponencial E(X) = α, e portanto o estimador obtido é o esperado pelo senso comum.

No caso discreto, a função de verossimilhança pode ser escrita na forma

$$L(\theta; x_1, ..., x_n) = P(X_1 = x_1|\theta) \ldots P(X_n = x_n|\theta).$$

Veja o Problema 37 para o caso de termos mais de um parâmetro.

Problemas

10. Na função de verossimilhança $L(p)$ da binomial, suponha que $n = 5$ e $x = 3$. Construa o gráfico da função para os possíveis valores de p = 1/5, 2/5, 3/5, 4/5, e verifique que o máximo ocorre realmente para $p = 3/5$.

11. Observa-se uma seqüência de ensaios de Bernoulli, independentes, com parâmetro p, até a ocorrência do primeiro sucesso. Se X indicar o número de ensaios necessários:
    (a) Mostre que $P(X = x) = (1 - p)^{x-1}p$ (distribuição geométrica).
    (b) Repetiu-se esse experimento n vezes, e em cada um deles o número de ensaios necessários foram $x_1, x_2, ..., x_n$. Encontre o EMV para p.
    (c) Usando uma moeda, repetiu-se esse experimento 5 vezes, e o número de ensaios necessários até a ocorrência da primeira coroa foi 2, 3, 1, 4, 1, respectivamente. Qual a estimativa de MV para p = probabilidade de ocorrência de coroa nessa moeda? Existiria outra maneira de estimar p?

12. Suponha que X seja uma v.a. com distribuição normal, com média μ e variância 1. Obtenha o EMV de μ, para uma amostra de tamanho n, $(x_1, ..., x_n)$.

13. Considere Y uma v.a. com distribuição de Poisson, com parâmetro $\lambda > 0$. Obtenha a EMV de λ, baseado numa amostra de tamanho n.

11.6 Intervalos de Confiança

Até agora, todos os estimadores apresentados foram pontuais, isto é, especificam um único valor para o estimador. Esse procedimento não permite julgar qual a possível magnitude do erro que estamos cometendo. Daí, surge a idéia de construir os intervalos de confiança, que são baseados na distribuição amostral do estimador pontual.

Exemplo 11.12. Suponha que queiramos estimar a média μ de uma população qualquer, e para tanto usamos a média $\overline{X}$ de uma amostra de tamanho n. Do TLC,

$$e = (\overline{X} - \mu) \sim N(0, \sigma^2_{\bar{x}}), \tag{11.32}$$

com $\text{Var}(\overline{X}) = \sigma^2_{\bar{x}} = \sigma^2/n$. Daqui podemos determinar qual a probabilidade de cometermos erros de determinadas magnitudes. Por exemplo,

$$P(|e| < 1{,}96\sigma_{\bar{x}}) = 0{,}95$$

ou

$$P(|\overline{X} - \mu| < 1{,}96\sigma_{\bar{x}}) = 0{,}95,$$

que é equivalente a

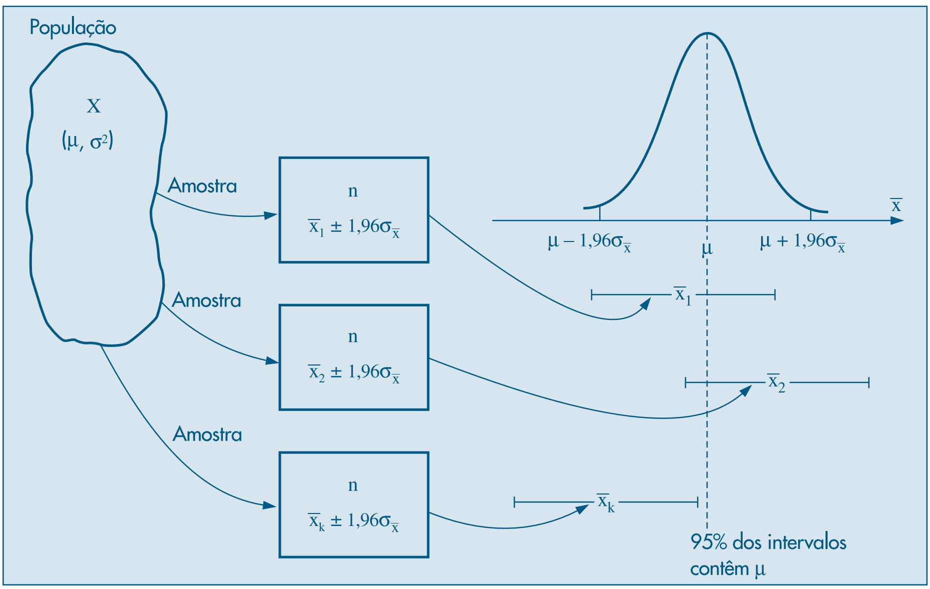
$$P(-1{,}96\sigma_{\bar{x}} < \overline{X} - \mu < 1{,}96\sigma_{\bar{x}}) = 0{,}95,$$

e, finalmente,

$$P(\overline{X} - 1{,}96\sigma_{\bar{x}} < \mu < \overline{X} + 1{,}96\sigma_{\bar{x}}) = 0{,}95. \tag{11.33}$$

Convém lembrar que μ não é uma variável aleatória e sim um parâmetro, e a expressão (11.33) deve ser interpretada da seguinte maneira: se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma $]\overline{X} - 1{,}96\sigma_{\bar{x}},\ \overline{X} + 1{,}96\sigma_{\bar{x}}[$, todos baseados em amostras de tamanho n, 95% deles conteriam o parâmetro μ. Veja a Figura 11.3. Dizemos que $\gamma = 0{,}95$ é o *coeficiente de confiança*. Nessa figura estão esquematizados o funcionamento e o significado de um intervalo de confiança (IC) para μ, com $\gamma = 0{,}95$ e σ^2 conhecido.

Figura 11.3: Significado de um IC para μ, com $\gamma = 0{,}95$ e σ^2 conhecido.



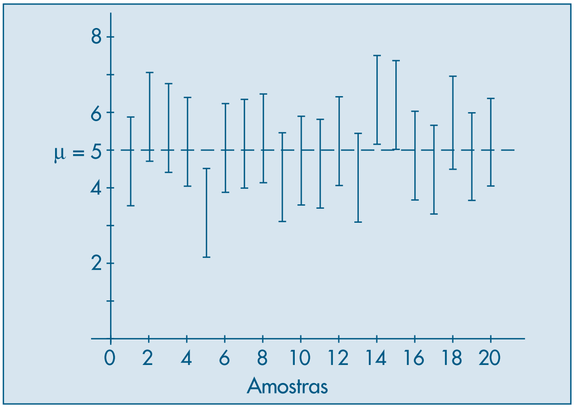
Escolhida uma amostra e encontrada sua média $\bar{x}_0$, e admitindo-se $\sigma_{\bar{x}}$ conhecido, podemos construir o intervalo

$$]\bar{x}_0 - 1{,}96\sigma_{\bar{x}}, \bar{x}_0 + 1{,}96\sigma_{\bar{x}}[. \qquad (11.34)$$

Esse intervalo pode ou não conter o parâmetro μ, mas pelo exposto acima temos 95% de confiança de que contenha.

Para ilustrar o que foi dito acima, consideremos o seguinte experimento de simulação. Geramos 20 amostras de tamanho n = 25 de uma distribuição normal de média μ = 5 e desvio padrão σ = 3. Para cada amostra construímos o intervalo de confiança para μ, com coeficiente de confiança γ = 0,95, que é da forma $\bar{X} \pm 1{,}176$, usando (11.34). Na Figura 11.4, temos esses intervalos representados e notamos que três deles (amostras de números 5, 14 e 15) não contêm a média μ = 5.

Figura 11.4: Intervalos de confiança para a média de uma N(5, 9), para 20 amostras de tamanho n = 25.



Exemplo 11.13. Uma máquina enche pacotes de café com uma variância igual a 100 g². Ela estava regulada para encher os pacotes com 500 g, em média. Agora, ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 25 pacotes apresentou uma média igual a 485 g. Vamos construir um intervalo de confiança com 95% de confiança para μ. De (11.34), teremos

$$IC(\mu; 0{,}95) = 485 \pm 1{,}96 \times 2,$$

ou seja,

$$IC(\mu; 0{,}95) =]481, 489[,$$

pois $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 10/5 = 2$g.

Se T for um estimador do parâmetro θ, e conhecida a distribuição amostral de T, sempre será possível achar dois valores t_1 e t_2, tais que

$$P(t_1 < \theta < t_2) = \gamma, \qquad (11.35)$$

a probabilidade interpretada como em (11.33), e γ um valor fixo, $0 < \gamma < 1$. Para uma dada amostra, teremos dois valores fixos para t_1 e t_2, e o intervalo de confiança para θ, com coeficiente de confiança γ, será indicado do seguinte modo:

$$\mathrm{IC}(\theta;\ \gamma) =]t_1,\ t_2[. \tag{11.36}$$

Se a variância populacional σ^2 não for conhecida, podemos substituir em (11.34) $\sigma_{\bar{x}}$ por $S/\sqrt{n}$, onde S^2 é a variância amostral dada em (11.9). Para n grande, da ordem de 100, o intervalo (11.34), com essa modificação, pode ainda ser usado. Para n não muito grande, a distribuição normal não pode mais ser usada e terá de ser substituída pela distribuição t de Student, que estudamos no Capítulo 7. Esse assunto voltará a ser abordado no Capítulo 12.

Para um coeficiente de confiança qualquer γ, teremos de usar o valor $z(\gamma)$ tal que $P(-z(\gamma) < Z < z(\gamma)) = \gamma$, com $Z \sim N(0,\ 1)$. O intervalo fica

$$\mathrm{IC}(\mu;\ \gamma) =]\bar{X} - z(\gamma)\sigma_{\bar{x}};\ \bar{X} + z(\gamma)\sigma_{\bar{x}}[. \tag{11.37}$$

Observe, também, que a amplitude do intervalo (11.37) é $L = 2z(\gamma)\sigma/\sqrt{n}$, que é uma constante, independente de $\bar{X}$. Se construirmos vários intervalos de confiança para o mesmo valor de n, σ e γ, estes terão extremos aleatórios, mas todos terão a mesma amplitude L.

Exemplo 11.14. Vamos obter um intervalo de confiança para o parâmetro p de uma distribuição $b(n,\ p)$. Sabemos que se X = número de sucessos nas n provas, então X tem distribuição aproximadamente normal, com média $\mu = np$ e variância $\sigma^2 = npq$, com $q = 1 - p$. Logo,

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0,\ 1),$$

ou ainda,

$$Z = \frac{X/n - p}{\sqrt{pq/n}} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{pq}} \sim N(0,\ 1). \tag{11.38}$$

Assim, se $\gamma = 0{,}95$, temos, consultando a Tabela III, que

$$P(-1{,}96 \leqslant Z \leqslant 1{,}96) = 0{,}95,$$

ou seja,

$$P\left\{-1{,}96 \leqslant \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{pq}} \leqslant 1{,}96\right\} = 0{,}95.$$

Portanto, com probabilidade 0,95, temos que

$$-1{,}96\sqrt{pq/n} \leqslant \hat{p} - p \leqslant 1{,}96\sqrt{pq/n},$$

do que segue

$$\hat{p} - 1{,}96\sqrt{pq/n} \leqslant p \leqslant \hat{p} + 1{,}96\sqrt{pq/n}.$$

Como não conhecemos p, podemos proceder de duas maneiras. Uma é usar o fato que $pq \leqslant 1/4$, de modo que $\sqrt{pq/n} \leqslant 1/\sqrt{4n}$, obtendo

$$\hat{p} - \frac{1{,}96}{\sqrt{4n}} \leqslant p \leqslant \hat{p} + \frac{1{,}96}{\sqrt{4n}}. \tag{11.39}$$

Temos, então, que $]\hat{p} - 1{,}96/\sqrt{4n};\ \hat{p} + 1{,}96/\sqrt{4n}[$ é um intervalo de confiança para p, com coeficiente de confiança de 95%.

Para um γ qualquer, $0 < \gamma < 1$, (11.39) fica

$$\hat{p} - \frac{z(\gamma)}{\sqrt{4n}} \leqslant p \leqslant \hat{p} + \frac{z(\gamma)}{\sqrt{4n}}, \tag{11.40}$$

onde $z(\gamma)$ é definido como em (11.37).

Exemplo 11.15. Numa pesquisa de mercado, $n = 400$ pessoas foram entrevistadas sobre determinado produto, e 60% delas preferiram a marca A. Aqui, $\hat{p} = 0{,}6$ e um intervalo de confiança para p com coeficiente de confiança $\gamma = 0{,}95$ será

$$0{,}6 \pm (1{,}96)\ 1/\sqrt{1600} = 0{,}6 \pm 0{,}049,$$

ou seja

$$IC(p;\ 0{,}95) =]0{,}551;\ 0{,}649[.$$

O intervalo (11.40) é chamado *conservador*, pois se p não for igual a 1/2 e estiver próximo de zero ou de um, então ele fornece um intervalo desnecessariamente maior, porque substituímos pq pelo seu valor máximo, 1/4. Uma outra maneira de proceder é substituir pq por $\hat{p}\hat{q}$, com $\hat{q} = 1 - \hat{p}$, sendo $\hat{p}$ o estimador de máxima verossimilhança de p, por exemplo. O intervalo obtido fica

$$\hat{p} - z(\gamma)\sqrt{\hat{p}\hat{q}/n} \leqslant p \leqslant \hat{p} + z(\gamma)\sqrt{\hat{p}\hat{q}/n}, \tag{11.41}$$

com $z(\gamma)$ definido como em (11.40).

Exemplo 11.16. Suponha que em $n = 400$ provas obtemos $k = 80$ sucessos. Vamos obter um intervalo de confiança para p com $\gamma = 0{,}90$. Como $\hat{p} = 80/400 = 0{,}2$ e $\hat{q} = 1 - \hat{p} = 0{,}8$, então (11.41) fica

$$0{,}2 \pm (1{,}645)\sqrt{(0{,}2)(0{,}8)/400} = 0{,}2 \pm 0{,}033,$$

ou seja,

$$IC(p;\ 0{,}90) =]0{,}167;\ 0{,}233[.$$

Usando (11.40) o intervalo conservador é

$$IC(p;\ 0{,}90) =]0{,}159;\ 0{,}241[.$$

Observe que o primeiro intervalo tem amplitude menor que o segundo. Outra observação importante é que por (11.40) e um γ fixo, os intervalos que podemos obter para amostras diferentes (mas de mesmo tamanho n) terão a mesma amplitude, dada por $2z(\gamma)/\sqrt{4n}$. Por outro lado, usando (11.41), a amplitude do intervalo será $2z(\gamma)\dfrac{\sqrt{\hat{p}\hat{q}}}{n}$, que é variável de amostra para amostra, pois $\hat{p}$ (e, conseqüentemente, $\hat{q}$) variará de amostra para amostra.

Problemas

14. Calcule o intervalo de confiança para a média de uma $N(\mu, \sigma^2)$ em cada um dos casos abaixo.

Média Amostral	Tamanho da Amostra	Desvio Padrão da População	Coeficiente de Confiança
170 cm	100	15 cm	95%
165 cm	184	30 cm	85%
180 cm	225	30 cm	70%

15. De 50.000 válvulas fabricadas por uma companhia retira-se uma amostra de 400 válvulas, e obtém-se a vida média de 800 horas e o desvio padrão de 100 horas.
    (a) Qual o intervalo de confiança de 99% para a vida média da população?
    (b) Com que confiança dir-se-ia que a vida média é $800 \pm 0{,}98$?
    (c) Que tamanho deve ter a amostra para que seja de 95% a confiança na estimativa $800 \pm 7{,}84$?
    (Que suposições você fez para responder às questões acima?)

16. Qual deve ser o tamanho de uma amostra cujo desvio padrão é 10 para que a diferença da média amostral para a média da população, em valor absoluto, seja menor que 1, com coeficiente de confiança igual a:
    (a) 95% *(b)* 99%

17. Uma população tem desvio padrão igual a 10.
    (a) Que tamanho deve ter uma amostra para que, com probabilidade 8%, o erro em estimar a média seja superior a uma unidade?
    (b) Supondo-se colhida a amostra no caso anterior, qual o intervalo de confiança, se $\overline{x} = 50$?

18. Uma amostra aleatória de 625 donas de casa revela que 70% delas preferem a marca A de detergente. Construir um intervalo de confiança para p = proporção das donas de casa que preferem A com c.c. $\gamma = 90\%$.

19. Encontre os intervalos de confiança para p se $k/n = 0{,}3$, com c.c. $\gamma = 0{,}95$. Utilize os dois enfoques apontados na seção 11.6, com $n = 400$.

20. Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato em questão.

(a) Determine o tamanho da amostra necessário para que o erro cometido na estimação seja de, no máximo, 0,01 com probabilidade de 80%.

(b) Se na amostra final, com tamanho igual ao obtido em (a), observou-se que 55% dos eleitores eram favoráveis ao candidato em questão, construa um intervalo de confiança para a proporção p. Utilize $\gamma = 0{,}95$.

21. Suponha que estejamos interessados em estimar a proporção de consumidores de um certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:

(a) o intervalo de confiança para p, com coeficiente de confiança de 95% (interprete o resultado);

(b) o tamanho da amostra para que o erro da estimativa não exceda a 0,02 unidades com probabilidade de 95% (interprete o resultado).

11.7 Erro Padrão de um Estimador

Vimos que, obtida a distribuição amostral de um estimador, podíamos calcular a sua variância. Se não pudermos obter a distribuição exata, usamos uma aproximação, se essa estiver disponível, como no caso de $\overline{X}$, e a variância do estimador será a variância dessa aproximação. Por exemplo, para a média amostral $\overline{X}$, obtida de uma amostra de tamanho n, temos que

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n},$$

na qual σ^2 é a variância da v.a. X definida sobre a população.

À raiz quadrada dessa variância chamaremos de erro padrão de $\overline{X}$ e o denotaremos por

$$\mathrm{EP}(\overline{X}) = \frac{\sigma}{\sqrt{n}}. \qquad (11.42)$$

Definição. Se T for um estimador do parâmetro θ, chamaremos de *erro padrão de* T a quantidade

$$\mathrm{EP}(T) = \sqrt{\mathrm{Var}(T)}. \qquad (11.43)$$

A variância de T dependerá dos parâmetros da distribuição de X, o mesmo acontecendo com o erro padrão. Por exemplo, em (11.42), $\mathrm{EP}(\overline{X})$ depende de σ, que em geral é desconhecida. Podemos, então, obter o erro padrão estimado de $\overline{X}$, dado por

$$\mathrm{ep}(\overline{X}) = \widehat{\mathrm{EP}}(\overline{X}) = S/\sqrt{n}, \qquad (11.44)$$

na qual S^2 é a variância amostral. Genericamente, o *erro padrão estimado* de T é dado por

$$\widehat{\mathrm{EP}}(T) = \sqrt{\widehat{\mathrm{Var}}(T)}. \qquad (11.45)$$

Muitas vezes a quantidade (11.45) é chamada de erro amostral. Mas preferimos chamar de erro amostral à diferença $e = T - \theta$.

Exemplo 11.17. Para o Exemplo 11.15, $\hat{p}$ = 0,6, e o erro padrão de $\hat{p}$ será dado por

$$\text{EP}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}. \tag{11.46}$$

Como não conhecemos p usamos no seu lugar o estimador $\hat{p}$, obtendo-se

$$\widehat{\text{EP}}(\hat{p}) = \sqrt{(0{,}6)(0{,}4)/400} = 0{,}025.$$

Observe que o intervalo de confiança (11.41) pode ser escrito

$$\hat{p} \pm z(\gamma)(\widehat{\text{EP}}(\hat{p})),$$

ao passo que o intervalo para μ dado por (11.37) pode ser escrito

$$\overline{X} \pm (1{,}96)(\text{EP}(\overline{X})).$$

11.8 Inferência Bayesiana

O estabelecimento de uma ponte entre os valores observados na amostra e os modelos postulados para a população, objeto da inferência estatística, exige a adoção de princípios teóricos muito bem especificados. Neste livro usaremos a chamada *teoria freqüentista* (às vezes também chamada de clássica). Seus fundamentos encontram-se em trabalhos de J. Neyman, E. Pearson, R. Fisher e outros.

Consideremos um exemplo para ilustrar esse enfoque. Suponha que tenhamos uma amostra observada $(x_1, ..., x_n)$ de uma população normal, $N(\mu, \sigma^2)$, e queremos fazer inferências sobre os valores de μ e σ^2, baseados nas n observações.

Por meio de algum procedimento estudado neste capítulo, selecionamos estimadores $\hat{\mu}(x)$ e $\hat{\sigma}^2(x)$ que sejam funções do vetor de observações $\boldsymbol{x} = (x_1, ..., x_n)'$. Considere dados hipotéticos $\boldsymbol{x}_1, \boldsymbol{x}_2, ...$, todos amostras de tamanho n, que poderiam ter sido gerados da população em questão. Obtemos, então, as distribuições amostrais de $\hat{\mu}(\boldsymbol{x})$ e $\hat{\sigma}^2(\boldsymbol{x})$, como na seção 10.7. Podemos também obter intervalos de confiança para os parâmetros desconhecidos μ e σ^2, bem como testar hipóteses sobre esses parâmetros, assunto a ser discutido no Capítulo 12.

Para construir intervalos de confiança e testar hipóteses será necessário conhecer a distribuição amostral dos estimadores. Como só temos um conjunto de dados e não dados hipotéticos, estas distribuições amostrais terão de ser obtidas de outra maneira, e não como no Exemplo 10.7. Usualmente isso é feito usando teoremas como o Teorema Limite Central, discutido na seção 10.8, obtendo-se uma distribuição aproximada para os estimadores, que vale para tamanhos de amostras grandes.

A crítica que se faz à teoria freqüentista é a possibilidade de "replicar dados", bem como o recurso à teoria assintótica. Uma teoria que não faz uso de tais argumentos é a inferência bayesiana, cujos fundamentos foram estabelecidos por T. Bayes em 1763. Outros expoentes dessa corrente foram Bernoulli (1713), Laplace (1812) e Jeffreys (1939). Aqui, o Teorema de Bayes, estudado no Capítulo 5, tem papel fundamental. A noção de probabilidade prevalente aqui é a subjetiva, discutida brevemente no mesmo capítulo.