Capítulo 8
Estimação intervalar de proporção e de média

8.1 A parte prática da estimação intervalar de proporção

A ideia é que nesta seção, eu apresente somente o método de solução de problemas do tema "estimação intervalar de proporção", postorgando para a Seção 8.2 o embasamento e a justificativa teóricos do método assim como algumas particularidades que surgem na execução dele. O caminho oposto, isto é, começar com as construções teóricas e deduzir delas a parte instrumental para a estimação intervalar de porporção, possui uma vantagem se for comparado com o caminho aquí tomado. A vantagem é a facilidade e a naturalidade na explicação do coeficiente de confiança de estimação intervalar. Porém, em todos os outros aspectos, o caminho alternativo mostrou ser inferior.

8.1.1 Estimação intervalar de proporção na posse de amostra

I. O problema no título da seção e o procedimento para sua solução, que serão lhe ensinados agora, podem ser apresentados com auxílio de urna e bolas. E é assim que farei.

II. Imagine uma urna com bolas idênticas no tato, mas de duas cores diferentes: preta e branca. Imagine que as proporções de bolas pretas e de bolas brancas da urna são desconhecidas. Nesse ambiente imaginário, porém facilmente imaginávio, estaremos falando sobre a estimação da proporção das bolas pretas com base em resultados de amostra de bolas retiradas da urna.

▷ Observe que a estimativa para a proporção das brancas, caso essa esteja no foco de nosso interesse, vai ser, obviamente, a diferença entre 1 e a estimativa para a proporção das pretas. Isso faz com que toda a atenção pode ser concentrada na proporção de bolas pretas.

Vou denotar por p aquilo que queremos estimar, quer dizer **a proporção populacional** de bolas pretas. O termo "populacional" é cientificamente mais correto de que o termo "da urna", mas você pode usar qualquer um dos dois. Ressalvo que p é só uma notação introduzida para facilitar a escrita da exposição a seguir e que seu valor numérico é desconhecido e deve ser visto como tal.

III. Agora vou esmiuçar o conceito de amostra e de tudo que relaciona-se a esse e está importante para a exposição a seguir. **Amostra** entende-se aqui como o resultado do seguinte procedimento: escolha-se da urna n bolas, em sequencia ao acaso e com reposição, e anotam-se as core dessas. Então, amostra pode e deve ser concebida como uma sequencia

$$x_1, x_2, \ldots, x_n \tag{8.1}$$

onde x_i representa a cor da i-ésima bola retirada. A representação pode usar qualquer codificação cômoda, mas, com a vista nas necessidades e comodidades dos argumentos futuros, será adotada a seguinte codificação específica

$$x_i = \begin{cases} 1, & \text{caso a } i\text{-ísima bola da amostra for preta} \\ 0, & \text{caso a } i\text{-ísima bola da amostra for branca} \end{cases} \tag{8.2}$$

Abaixo está um exemplo da amostra que pode surgir como resultado de retirada de $n = 20$ bolas:

$$\bullet\ \circ\ \bullet\ \bullet\ \bullet\ \circ\ \bullet\ \circ\ \circ\ \bullet\ \circ\ \bullet\ \circ\ \circ\ \bullet\ \bullet\ \bullet\ \circ\ \bullet\ \circ \tag{8.3}$$

Os valores da sequencia x que representa essa amostra são assim:

$$1,\ 0,\ 1,\ 1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 1,\ 1,\ 0,\ 1,\ 0 \tag{8.4}$$

Mais um pouco de notações e nomes: n chamar-se-á **tamanho de amostra**; k denotará a quantidade de bolas pretas da amostra; k/n tem o nome **a proproção amostral** de bolas

pretas. Todos esses termos são autoexplicativos e eu não tenho nada a acrescentar sobre eles no momento. No caso do exemplo (8.3), $k = 11$ (são 11 bolas pretas), e $k/n = 11/20$.

IV. Agora, no ambiente de urna com bolas e de amostra, conforme descritas no texto acima, formula-se a seguinte tarefa:

$$\text{em posse dos valores de } n \text{ e de } k\text{, estime o valor de } p \tag{8.5}$$

Essa colocação, embora suscinta, determina e defina muitas coisas que podem parecer secundárias, mas, na realidade, não são. Então, para que esses detalhes não sejam dispercebidos, vou destacá-los agora:

(a) O tamanho de amostra, n, é um valor conhecido. Pense como se fosse que alguém determinou esse valor e o forneceu para a pessoa que faz a amostragem. Os motivos que guiam a determinação de n serão discutidos na Seção 8.1.2. Mas este será uma outra história, pois relacionar-se-á a um outro ambiente e uma outra tarefa.

(b) O valor numérico de k, que dizer, da quantidade de bolas pretas na amostra, é conhecido. Tal conhecimento só é adquirido após que a amostra foi feita. Então, implicitamente admita-se que a amostragem foi proferida. Observe que isso implica no que sabe-se também toda a sequencia de bolas pretas e brancas na amostra. Entretanto, de acordo com a colocação da tarefa, essa informação não foi convidada a participar na estimação de p. Acerca disso, posso dizer que eu pessoalmente acredito que a sequencia carrega uma certa informação que possa ser útil, mas o método de estimação de p que será apresentado e estudado nesse texto não usa essa informação. É por isso, que já antemão, eu excluo o uso dela.

A tarefa (8.5) é o que eu chamaria de "problema de estimação de proporção populacional na posse de amostra, ou na presença de amostra". Entretanto, esse nome não é tradicional para a literatura estatística. Em primeiro lugar, não é um costume destacar que a amostra já foi feita e já está conhecida, embora tal destaque seja útil já que existe o caso no qual a amostra é vista como algo a ser feito (recordo, esse caso será abordado na Seção 8.1.2). Em segundo lugar, não é costume usar a palavra "populacional". Isso é lógico pois já que a proporção amostral é calculável diretamente do resultado de amostragem então o único desconhecido de peso só pode ser a proporção populacional. Enfim, o que sobrou é **problema de estimação de proproção**. Isso é então o nome para a tarefa (8.5).

V. Se você responder ao problema (8.5) da seguinte maneira:

$$\text{a estimativa para } p \text{ é } \frac{k}{n} \tag{8.6}$$

diz-se que você faz **estimação pontual** e que $\frac{k}{n}$ é a **estimativa pontual** de sua estimação. Não discutimos as vantagens e desvantagens dessa estimativa, assim como não discutimos se há estimativas pontuais diferentes e, talvéz até melhores. O que é relevante agora é avisar que daquí para frente sempre usaremos $\frac{k}{n}$ quando haver necessidade de estimar p por um valor. Acrescento ainda que na Estatística é costume denotar por $\hat{p}$ a estimativa pontual para p. Portanto, no presente texto,

$$\hat{p} \text{ é a notação alternativa para } \frac{k}{n} \tag{8.7}$$

VI. Já se você responder ao problema (8.5) fornecendo dois valores A e B, e alegando que

$$p \text{ encontra-se no intervalo } [A,\, B] \tag{8.8}$$

diz-se então que você fez **estimação por intervalo**; o intervalo $[A, B]$ nesse caso chama-se de **intervalo de estimação**.

Existem diversas maneiras sensatas e eficientes para a determinação de A e B. No presente texto, focamos exclusivamente numa abordagem específica, de acordo com qual o intervalo $[A, B]$ tem o formato

$$\left[\frac{k}{n} - \varepsilon\,,\ \frac{k}{n} + \varepsilon\right] \tag{8.9}$$

para um valor positivo ε que está determinado por método específico a ser descrito abaixo. A interpretação de ε é a distância que há – de acordo com a estimação por intervalo (8.9)– entre o verdadeiro valor de p e sua estimativa pontual $\frac{k}{n}$. Isso é o motivo de ε ser chamado **margem de erro**.

Por exemplo, no caso da amostra (8.3) (recorde, para essa amostra, $n = 20$ e $k = 11$), a escolha $\varepsilon = 0,04$ leva-nos ao intervalo de estimação

$$\left[\frac{11}{20} - 0,04,\ \frac{11}{20} + 0,04\right] \ = \ [0,51,\ 0,59]$$

Observe que se escolher um valor maior para ε, digamos $0,07$, então para mesma amostra, teremos intervalo de estimação maior:

$$\left[\frac{11}{20} - 0,07,\ \frac{11}{20} + 0,07\right] \ = \ [0,48,\ 0,62]$$

É intuitivamente claro que o maior intervalo tem mais "chance" que ele "ache" o verdadeiro valor de p. Essa intuição será formalizada pelo conceito **confiança**.

VII. Nessa parte, vou definir formalmente o procedimento de estimação por intervalo cuja aplicação é o assunto do presente capítulo (Capítulo 8). Recordo que disse que tal procedimento não é o único admissível, mas ele e somente ele será considerado nesse testo. Por isso que o termo genérico **estimação intervalar** será o rótulo para o procedimento aqui definido.

Então, a estimação por intervalo de confiança constitui-se de três peças que estão definidas em (8.10), (8.11) e (8.12) abaixo.

O intervalo de estimação tem o formato $\left[\frac{k}{n} - \varepsilon\,,\ \frac{k}{n} + \varepsilon\right]$ (8.10)

A cada intervalo de estimação atribui-se seu **coeficiente de confiança** que interpreta-se como a probabilidade do intervalo (8.10) captar p; (8.11)

Coeficiente de confiança será denotado por γ.

γ e ε vinculam-se um a outro via a seguinte fórmula (abaixo $Z \sim \mathcal{N}(0,1)$):

$$\underbrace{\varepsilon = z\sqrt{\frac{\frac{k}{n}\left(1 - \frac{k}{n}\right)}{n}}}_{\text{primeira parte}},\ \text{onde}\ \underbrace{z > 0 \text{ é tal que } \gamma = I\!P\left[-z \leq Z \leq z\right]}_{\text{segunda parte}}, \tag{8.12}$$

a fórmula (8.12) vale para valores suficientemente grandes de n, e nós assumiremos implicitamente tal qualidade de n toda vez que usemos a fórmula.

Noto que a introdução de confiança faz com que intervalo de estimação chama-se alternativamente por **intervalo de confiança**. Isso faz com que o método todo é chamado de **estimação por intervalo de confiança**. No presente texto, esse nome é sinônimo de "estimação intervalar".

Gostaria de destacar os aspectos que fazem o método acima definido ser específico. O destaque vai lhe ajudar a imaginar como poderiam ser outros métodos de estimação intervalar, embora tal concepção está fora do arcabouço do presente livro, conforme já avisado acima. A primeira particularidade do método está no formato do intervalo de estimação: ele é centrado no valor de k/n, o que é a estimativa pontual para a proporção estimada. A segunda particularidade está na forma do cálculo do coeficiente de confiança. Esse cálculo será apresentado na Seção 8.2. A lógica por trás dele dá o sentido exato para a interpretação probabilística de γ. Mas como a explicação virá de uma seção futura, então torna-se necessário agora o seguinte aviso. Apesar do valor de p estar desconhecido, ele considera-se fixo, e portanto torna-se indesejada a frase "a probabilidade de p pertencer ao intervalo de estimação construído" pois a mesma insinua que p é uma variável aleatória. O correto é conceber que a aleatoriedade da amostragem faz flutuar o intervalo de estimação, e, consequentemente, ele pode "acertar" no p assim como "não acertar". A probabilidade do acerto é o valor de γ (conforme mostrarão os cálculos da Seção 8.2) e é exatamente isso que está transmitido pela interpretação (8.11).

VIII. Até o momento, eu completei a apresentação do método mais tradicional e (aparentemente) mais usado de estimação por intervalo de confiança (de proporção populacional). Recordo que a justificativa matemática de sua construção está adiada para a Seção 8.2. Meu plano para o resto da presente seção é guiar meu leitor para que ele possa aprender a aplicar o método. Os Exemplos 54 e 56 mostrarão a aplicação na medida adequada para iniciante, enquanto que os Exemplos 55 e 57 mergulharão nums aspectos do método que são secundários mas importantes. Após a leitura desses quatro exemplos, meu leitor pode fazer todos os exercícios da lista, menos aqueles que tratam do dimensiamento de amostra (isto é, os que pedem descobrir o tamanho de amostra a ser feita para que seu uso posterior na estimação intervalar resulte em desejado coeficiente de confiança e desejado margem de erro). Recordo que o dimensiamento de amostratratar-se-á na Seção 8.1.2.

IX. Acredito que a aprendizagem da aplicação do método ensinado seja fácil e mais organizada se meu leitor conceber que na estimação intervalar aplicado as amostras conhecidas há somente dois parâmetros que são ε e γ. A variável z que você vê na fórmula (8.12) não é o terceiro parâmetro; z é só uma variável auxiliar usada para vincular ε a γ. Consequentemente, existem dois tipos de problemas: dado o valor de ε calcular o correspondente valor de γ, e, a recípoca, dado o valor de γ calcular o correspondente valor de ε. Para orientar meu leitor nesse sentido, digo-lo que o problema do primeiro tipo é o problema abordado no Exemplos 54, enquanto que o do segundo tipo está no Exemplo 56. Os cálculos nos dois casos seguem os dois caminhos apresentados abaixo; é claro que ambos foram derivados a partir da definição principal contida em (8.10), (8.11) e (8.12).

$$\text{dado} \quad \varepsilon \quad \text{calcule } z = \frac{\varepsilon}{\sqrt{\left(\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}\right)}} \text{ e obtenha } \gamma = I\!P\left[-z \leq Z \leq z\right] \tag{8.13}$$

$$\text{dado} \quad \gamma \quad \text{ache } z \text{ tal que } \gamma = I\!P\left[-z \leq Z \leq z\right] \text{ e obtenha } \varepsilon = z\sqrt{\left(\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}\right)} \tag{8.14}$$

X. No presente exemplo considera-se o problema de estimação intervalar para proporção do tipo "dada a margem de erro (ε) calcule o coeficiente de confiança (γ)".

↓ **Exemplo 54.** Das 600 pessoas escolhidas ao acaso de uma população, 330 afirmaram apoiar o candidato Zé Fico nas próximas eleições. Calcule o coeficiente de confiança da estimativa que

a proporção do eleitorado do Zé esteja dentro do intervalo de confiança

$$\left[\frac{330}{600} - 0,04;\ \frac{330}{600} - 0,04\right] = [0,51;\ 0,59] \tag{8.15}$$

Vamos concordar que "escolher 600 pessoas ao acaso" refere-se as 600 escolhas sequencias ao acaso com reposição. Na vida real isto não ocorre pois, como bem se sabe, a mesma pessoa nunca é intervistada mais que uma vez. Entretanto, nosso acordo artificial permite garantir que o enredo do exemplo torne-se equivalente à situação de bolas-na-urna.
Vamos também concordar que os votantes não mudarão de suas intenções de voto, isto é, quem disse que apóia Zé agora vai votar nele, e quem disse que não – não vai votar nele nas próximas eleições.

Vamos á solução. O primeiro passo da solução é estabelecer a equivalência da situação do enunciado com a de "bolas-na-urna". Eis esta:

a população do distrito eleitoral do Zé	-	bolas em urna;
qualquer pessoa que votará no Zé	-	bola preta;
qualquer pessoa que não votará nele	-	bola branca;
pessoas intervistadas	-	amostra, isto é, bolas retiradas da urna ao acaso e com reposição;
600	-	o tamanho da amostra (recordo: a notação genérica é n);
330	-	a quantidade de bolas pretas na amostra (recordo: a notação ganérica é k);
0,04	-	ε (essa equivalência estabelese-se ao comparar a forma genérica (8.10) do intervalo de confiança com a forma (8.15) que ele adquire no exemplo);

Agora, que a analogia com as bolas-na-urna foi estabelecida, tornou-se claro que o coeficiente de confiança solicitado no exemplo é o valor do parâmetro γ, e que esse está fornecido pela fórmula (8.12). Ao colocar na fórmula, em sua primeira parte, $0,04$ no lugar de ε, e 600 e 330 nos lugares de n e k respectivamente, temos que

$$z = \frac{\varepsilon}{\sqrt{\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}}} = \frac{0,04}{\sqrt{\frac{0,55(1-0,55)}{600}}} \approx \frac{0,04}{0,02} = 2,00 \tag{8.16}$$

O valor achado do parâmetro auxiliar z permite-nos calcular, com emprego da segunda parte da fórmula (8.12), o valor de γ. O cálculo usa exclusivamente as habilidades que você deve ter adquirido na capítulo sobre as variáveis aleatórias normais. Eis a conta; nela, Z significa a variável aleatória normal padrão:

$$\begin{aligned} \gamma &= \mathbb{P}[-2,00 \le Z \le 2,00] \\ &= 2\mathbb{P}[0 \le Z \le 2,00] \\ &= 2\left(\mathbb{P}[Z \le 2,00] - 0,5\right) = (0,9772 - 0,5) = 0,9544, \\ &\quad \text{o que é a resposta final.} \end{aligned} \tag{8.17}$$

XI. O exemplo agora tratado nos fornece uma situação na qual há um embasamento prático e lógico para escolha do valor para a margem de erro. Em outras palavras, há uma razão para o surgimento de "$0,04$" na fórmula (8.15). Tal razão será explicada em seguida. Mas a exposição

é optativa, isto é, você não precisa entendé-la para poder fazer os exercícios do tema e questões da prova.

Recorde que p, o objeto de estitiva no exemplo, é a proporção da população que votará no Zé nas próximas eleições. Portanto, a probabilidade com a qual possamos garantir que p é de no mínimo $0,5$ é a probabilidade de Zé ganhar as próximas eleições. Tenha isso em sua mente quendo for ler o parágrafo abaixo que fornece a prometida explicação das razões que nos fizeram a escolher $\varepsilon = 0,04$.

Imagine então que Zé realmente deseja estimar a probabilidade de sua vitória nas próximas eleições. Imagine que para tal estimação, ele encomendou a pesquisa de intenção de voto que mostrou os dados que eu apresentei-lhe no enunciado do exemplo: dos 600 intervistados, 330 disseram que voltariam no Zé. Imagine por fim, que esses dados forma levados a um estatístico e ele/a decidiu abordar o problema do Zé com auxílio de intervalo de confiança. Então, o estatístico sabe que com os dados trazidos pelo Zé, o centro do intervalo de confiança está inevitavelmente no $\frac{330}{600} = 0,55$. Por outro lado, o argumento do parágrafo anterior elucida para o estatístico que para que possa-se dizer algo sobre a vitória do Zé nas próximas eleições, todo o intervalo de confiança deve estár à direita de $0,5$. Já que o centro do intervalo está à direita de $0,5$, então para que todo ele esteja à direita de $0,5$ é suficiente que seu extremo esquerdo esteja maior que $0,5$. Sendo que $0,55 - 0,5 = 0,05$, o estatístico conclui então que a margem do erro do intervalo a ser construido deve ser menor que $0,05$. Ele informa isso para Zé, e esse sugere o valor $0,04$ (se Zé fosse um matemático, ele poderia sugerir também $0,049999\ldots$). Com $0,04$ como o valor da margem de erro, o estatístico calcula o correspondente o coeficiente de confiança; esse é $0,9544$. Esse resultado leva – de acordo com o argumento apresentado no parágrafo anterior – a seguinte resposta ao Zé:

$$\text{com a probabilidade de no mínimo 0,9544, Zé vai ganhar as eleições} \tag{8.18}$$

▷ Vale comentar sobre o por quê a resposta diz "no mínimo 0,9544" e não "exatamente 0,9544". Para tal, observo que a abordagem empregada deduziu que a proporção populacional dos apoiadores de Zé está no intervalo $[0,51;\ 0,59]$ com o coeficiente de confiança 0,9544. Portanto, esta proporção está fora do intervalo com a "restante probabilidade" $0,0456$. Nesse valor, inclui-se a possibilidade da proporção populacional estar tanto à direita de $0,59$ quanto entre $0,5$ e $0,51$. Essas possibilidades, se ocorrerem, acarretam a vitória de Zé nas eleições, mas suas ocorrências não foram "contabilisadas" na confiança $0,9544$. É por isso que a probabilidade da vitória do Zé é "no mínimo $0,9544$".

Fim do Exemplo 54↑

↓ **Exemplo 55.** Com os dados do Exemplo 54, vou agroa calcular os coeficientes de confiança para dois intervalos de confiança: o primeiro corresponde ao valor $0,06$ para ε e sua cara portanto é

$$[0,55 - 0,06\,;\ 0,55 + 0,06] = [0,49,\ 0,61] \tag{8.19}$$

o o segundo corresponde ao valor $0,02$ para ε e sua cara portanto é

$$[0,55 - 0,02\,;\ 0,55 + 0,02] = [0,53,\ 0,57] \tag{8.20}$$

O objetivo desses cálculos é revelar – via exemplo numérico – a maneira com a qual a mudança do valor de ε afeta o valor de γ. O resultado estará na tabela colocada no final do presente exemplo. Posteriormente, ele servirá como motivação e como ilustração para a proposição 21. Devo esclarecer que minhas escolhas $0,06$ e $0,02$ não têm nenhuma motivação prática diferentemente daquilo que houve com a escolha $0,04$ tratada no Exemplo 54.

Usando a fórmula $z = \varepsilon/\sqrt{\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}}$ (que você conhece desde (8.16) e sabe ela que veio da primeira parte da fórmula principal (8.12)), tem-se que

$$\begin{array}{l} \varepsilon = 0,06 \;\; \text{acarreta} \;\; z = \frac{0,06}{\sqrt{\frac{0,55(1-0,55)}{600}}} \approx \frac{0,06}{0,02} = 3,00 \\ \varepsilon = 0,02 \;\; \text{acarreta} \;\; z = \frac{0,02}{\sqrt{\frac{0,55(1-0,55)}{600}}} \approx \frac{0,02}{0,02} = 1,00 \end{array} \tag{8.21}$$

Em seguida, usando a fórmula $\gamma = I\!P[-z \leq Z \leq z] = 2\,(I\!P[Z \leq z] - 0,5)$ (que foi deduzida em (8.17) a partir da segunda parte da fórmula principal (8.12)) e a Tabela da Distribuição Normal Padrão, tem-se que

$$\begin{array}{l} \varepsilon = 0,06 \;\; \text{acarreta} \;\; \gamma = I\!P[-3,00 \leq Z \leq 3,00] = 2(0,9987 - 0,5) = 0,9974 \\ \varepsilon = 0,02 \;\; \text{acarreta} \;\; \gamma = I\!P[-1,00 \leq Z \leq 1,00] = 2(0,8413 - 0,5) = 0,6826 \end{array} \tag{8.22}$$

isso quer dizer que os coeficientes de confiança dos intervalos (8.19) e (8.20) são, respetivamente, $0,9974$ e $0,6826$.

Gostaria de juntar os resultados desse exemplo e o do anterior numa tabela. Eis esta:

ε	0,02	0,04	0,06
z	1,00	2,00	3,00
γ	0,6826	0,9544	0,9974

Essa tabela mostra os valores do coeficiente de confiança (a última linha da tabela) correspondentes aos valores da margem de erro $0,02$, $0,04$ e $0,06$ (a primeira linha); na linha do meio da tabela estão os correspondentes valores do parámetro auxiliar z. Todos os cálculos foram feitos para a mesma amostra; nessa amostra, $n = 600$ e $k = 330$. Os detalhes dos cálculos encontram-se nos Exemplos 54 e 55. A tabela ilustra que o aumento do valor de margem de erro acarreta o aumento do valor do coeficiente de confiança quando a amostra fica inalterada. Esse fato será formulado em forma genérica na Proposição 21.

Fim do Exemplo 55↑

↓ **Exemplo 56.** No presente exemplo considera-se o problema de estimação intervalar para proporção do tipo "dado o coeficiente de confiança (γ) calcule a margem de erro (ε)". O exemplo conta uma história verdadeira com nomes alterados.

Até um certo momento, os pneus Durodondo eram comercializados por revendedores de pneus de todas as marcas do mercado. Aí, então, os fabricantes de Durodondo foram sondar a possibilidade da abertura de rede própria e exclusiva de comercialização do seu produto. Os donos da fábrica aceitam o risco de 10% no investimento na construção da rede, para qualquer que seja o valor do investimento. Tal valor, entretanto, depende da quantidade de lojas da rede, e esta quantidade, depende - por sua vez - da proporção dos donos de automóveis que usem Durodondo. Acredita-se que dono de automóvel quelquer troca sua preferência pela marca de pneu poucas vezes. Isto faz-se acreditar que os quem já usa Durodondo, continuará a usar, e os que não usa, não tornará a usar. Para estimar a proporção populacional dos quem usa Durodondo, foi encaminhada uma pesquisa que revelou que dentro das 600 pessoas escolhidas ao acaso, 80 usam Durodondo. A tarefa então é usar a proporção amostral revelada pela pesquisa para construir a estimativa intervalar da proporção populacional.

Do enunciado, deduz-se diretamente que os parâmetros do método de estimação intervalar adquiram os seguintes valores:

$$\begin{array}{rcl} n & - & 600 \\ k & - & 80 \\ \gamma & - & 90\% \end{array}$$

As primeiras duas linhas são óbvias. Já a terceira deduz-se da interpretação do coeficiente de confiança, e tal dedução será lhe apresentada lá pela frente, perto do fechamento do presente exemplo.

Para a construção do intervalo de estimação solicitado, falta só o valor de ε, já que a posição de centro está em

$$\frac{k}{n} = \frac{80}{600}$$

Buscaremos o valor de ε com auxílio da fórmula (8.12). Primeiramente, usaremos sua segunda parte para calcular o valor de z a partir do valor desejado para γ:

$$0,90 = I\!P[-z \leq Z \leq z] \implies 0,95 = I\!P[Z \leq z] \implies z = 1,65$$

No segundo e último passo, implantamos o valor achado na primeira parte de fórmula. Isto dos dá:

$$\varepsilon = 1,65\sqrt{\frac{\frac{80}{600}\left(1-\frac{80}{600}\right)}{600}} = 0,023$$

Então, temos a resposta à tarefa formulada no exemplo: o intervalo de confiança, com o solicitado coeficiente de confiança de 90%, para a verdadeira proporção dos usuários de pneu da marca Durodondo é

$$\left[\frac{80}{600} - 0,023\,;\ \frac{80}{600} + 0,023\right] = [0,11\,;\ 0,153]$$

Agora imagine que essa estimativa para a proporção populacional dos compradores de pneus da marca Dorodondo foi levada as economistas que availiaram que nesse caso o investimento de 1.150 milhoes na construção da rede própria seja lucrativo, nem mais nem menos. É possível entender a conclusão da avaliação: é o valor necessário para abrir a quantidade certa de lojas que comercializarão somente Duradondo. Se a quantidade de lojas for menor, os frequeses potenciais comprariam pneus de outras marcas porque pode não haver loja de Durodondo por perto. Por outro lado, se criar lojas demais, então a maioria delas ficarão osciosa pois lhes faltará clientela.

Então, imagina que o estatístico que fez as contas chegue ao conselho dos diretores e apresenta: "A pesquisa mostrou que com 1.150 milhoes de investimento em rede própria vocês terão lucro." Aí, obviamente, o conselho indiga: "Qual é o risco de investir essa quantia e não haver lucro? Recorde que aceitamos o risco de 10%!" A resposta do estatísco soa assim: "De acordo com o método aplicado, há 90% de certeza que a verdadeira proporção dos compradores de pneus da marca de vocês esteja entre 0,11 e 0,153. Para tais limites, o investimento lucrativo é 1.150 milhoes, conforme a conta de economistas. Tal investimento pode deixar de ser lucrativo se a verdadeira proporção estiver fora desses limites, o que acontece com os restantes 10%. Essa é o valor do risco que você pediram, e é para poder chegar nele, eu tomei 90% para o valor do coeficiente de confiança na minha construção de estimativa intervalar."

Espero que você, meu leitor, entendeu – por vias do presente exemplo – que existem situações práticas cujo enredo fixa o coeficiente de confiança, caso sua solução der-se por método de intervalo de confiança. Se numa de tais situações ainda a amostra for fornecida, então o que falta para construir o desejado coeficiente de confiança é o valor da margem de erro (ε). Na falta de amostra, a situação é pouco mais sutil e essa será discutida na Seção 8.1.2.

Fim do Exemplo 56↑

↓ **Exemplo 57.** Com os dados do Exemplo 56, vou agora calcular as margens de erro e construir os correspondentes intervalos de confiança para dois valores de coeficiente de confiança: $0,9974$ e $0,6826$. O objetivo desses cálculos é revelar – via exemplo numérico – a maneira com a qual a mudança do valor de γ afeta o valor de ε. O resultado estará na tabela colocada no final do presente exemplo. Posteriormente, ele servirá como motivação e como ilustração para a proposição 21. Devo esclarecer que a minhas escolhas $0,9974$ e $0,6826$ não têm nenhuma relação específica com o enredo descrito no Exemplo 56. Naquele exemplo, o valor de γ era de $0,9$ e isso foi justificado pelas preferências de investimento da empresa que encaminhou pesquisa de mercado. Já no momento, os valores $0,9974$ e $0,6826$ são simplesmente os que agradaram meu gosto.

Usando a fórmula $\gamma = I\!P[-z \leq Z \leq z]$ (apresentada a você em (8.12)), calculo que

$$\begin{array}{l} \gamma = 0,9974 \text{ acarreta } z = 3,00 \\ \gamma = 0,6826 \text{ acarreta } z = 1,00 \end{array} \tag{8.23}$$

(Não supreenda-se que essa relação entre valores de γ e de z você já viu em (8.22). É natural que a relação (8.22) repita-se em (8.23) pois a fórmula usada para derivar tanto (8.22) quanto (8.23) é a mesma e ela não depende nem de n nem de k. A única diferença é que em (8.22) ela foi usada para achar γ dado ε, enqaunto que em (8.23) ela foi usada para achar z dado γ.)

Em seguida, usando a fórmula $\varepsilon = z\left(\sqrt{\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}}\right)$ (apresentada a você em (8.12)), concluo que

$$\begin{array}{ll} z = 3,00 & \text{acarreta } \varepsilon = 3,00\sqrt{\frac{\frac{80}{600}\left(1-\frac{80}{600}\right)}{600}} \approx 0,042 \text{ e consequentemente} \\ & I.C.(p, 0,9974) = [0,133 - 0,042;\ 0,133 + 0,042] = [0,091;\ 0,175] \\ z = 1,00 & \text{acarreta } \varepsilon = 0,01\sqrt{\frac{\frac{80}{600}\left(1-\frac{80}{600}\right)}{600}} \approx 0,014 \text{ e consequentemente} \\ & I.C.(p, 0,6826) = [0,133 - 0,042;\ 0,133 + 0,042] = \end{array} \tag{8.24}$$

Gostaria de juntar os resultados desse exemplo e o do anterior numa tabela. Eis esta:

γ	0,9974	0,9000	0,6826
z	3,00	1,65	1,00
ε	0,042	0,023	0,014

Essa tabela mostra os valores da margem de erro (a última linha da tabela) correspondentes aos valores do coeficiente de confiança $0,9974$, $0,9000$ e $0,6826$ (a primeira linha); na linha do meio da tabela estão os correspondentes valores do parámetro auxiliar z. Todos os cálculos foram feitos para a mesma amostra; nessa amostra, $n = 600$ e $k = 80$. Os detalhes dos cálculos encontram-se nos Exemplos 56 e 57. A tabela ilustra que o aumento do valor do coeficiente de confiança acarreta o aumento do valor da margem de erro quando a amostra fica inalterada. Esse fato será formulado em forma genérica na Proposição 21.

Fim do Exemplo 57↑

XII. Nesta parte da apresentação discute-se o melhoramento de estimação intervalar no sentido de diminuição da margem de erro ou no aumento do coeficiente de confiança. Mostrar-se-á que é impossível melhorar os dois ao mesmo tempo. Esse fato está formulado precisamente na Proposição 21. Se você consegue entede-lá, você então não precisa ler o texto que precede à formulação da proposição. Aviso-lhe ainda que a proposição é importante para a prática da estimação intervalar e, além disso, prepara meu leitor para a compreenção do conteúdo da Seção 8.1.2 que dicutirá as ações que permitiriam melhorar a estimação intervalar tanto no

sentido de diminuição de sua margem de erro quanto no sentido de aumento de seu coeficiente de confiança.

Recordo-lhe que no Exemplo 55 eu construi diversos intervalos de confiança para p com base na amostra que apresentou $k = 330$ bolas pretas entre $n = 600$ bolas retiradas, e descobri que a diminuição da margem de erro do intervalo de confiança acarretava a diminuição do correspondente coeficiente de confiança. Essa relação está resumida na tabela no final do Exemplo 55. Os cálculos do exemplo mostram que a relação está "programada" nas fórmulas que formam a base matemática para a construção do intervalo de estimação. Entretanto, vale comentar que nossa intuição é capaz de conceber tal relação sem precisar fazer cálculos. De fato, é intuitivamente óbvio que quanto menor o tamanho de intervalo, mais difícil para ele acertar no p, e consequentemente, aos menores valores da margem de erro (ε) devem corresponder menores valores do coeficiente de confiamça (γ).

Recorde-lhe também que no Exemplo 57 eu fui no sentido contrário do raciocínio do Exemplo 55: eu alterei o valor do coeficiente de confiança e analisei a consequencia disso no valor da margem de erro; tudo foi feito para a amostra que apresentou $k = 80$ bolas pretas entre $n = 800$ bolas retiradas. A conclusão foi que a diminuição do coeficiente de confiança faz diminuir a margem de erro. Assim como no caso do Exemplo 55, a conclusão deduz-se a partir do embasamento matemático para a construção de intervalo de estimação e está de acordo com nossa concepção intuitiva do funcionamento de estimação por intervalo.

XIII. Os dois exemplos supramencionados foram incluídos no texto com o intuito de desencadear a discussão, que você vé nos dois parágrafos acima, e que destina-se à implantação na sua mente do seguinte fato: Na situação quando a amostra já foi feita, os valores da margem de erro e do coeficiente de confiança estão rigidamente amarados um no outro. Se você alterar um deles, o outro responderá imediatamente. E tem mais: a amarração não lhe permite melhorar ao mesmo tempo a precisão numérica da estimativa de p (expressa via a margem de erro ε) e a probabilidade de que p de fato foi emglobado pelo intervalo (expressa via o coeficiente de confiança γ); acontece que se você melhorar a precisão (isto é, diminuia ε) a probabilidade de acerto vai piorar (vai diminuir γ), e acontece que a recípoca também vale: se você melhorar a probabilidade de acerto (aumentar γ) então a precisão numérica vai piorar (vai aumentar ε).

Tudo o dito acima pode ser demonstrado rigorosamente (na verdade, a demostração já está nos cálculos dos Exemplos 57 e Exemplo 55 só que escrita para os casos particulares desses exemplos). Isso me motiva a formular o resultado em formato de um teorema:

Proposição 21 *(sobre a dependência entre a margem de erro e o coeficiente de confiança na estimação intervalar quando o tamanho de amostra e a proporção amostral ficam inalteradas).*

(a) Se os valores de n (o tamanho de amostra) e $\frac{k}{n}$ (a proporção amostral) fiquem inalterados enquanto que o valor de ε (a margem de erro) aumenta-se, então, como a consequencia, o valor de γ (o coeficiente de confiança) aumentar-se-á.

(b) Se os valores de n (o tamanho de amostra) e o valor de $\frac{k}{n}$ (a proporção amostral) fiquem inalterados enquanto que o valor de γ (o coeficiente de confiança) aumentar-se, então como a consequencia, o valor de ε (a margem de erro) aumentar-se-á.

8.1.2 Planejamento de amostra no problema de estimação intervalar

Imagine que puseram na sua frente uma urna com bolas pretas e brancas e lhe pediram que determine o tamanho mínimo de amostra que seja suficiente para que seu resultado permita

construir estimação intevalar para bolas pretas na urna com o coeficiente de confiança $0,9$ e com a margem de erro $0,04$. Observe o requerimento "mínimo" dessa tarefa. Ele é importante na prática pois cada retirada de bola no nosso modelo de bolas em urna corresponde na prática a um experimento que tem custo, o qual, as vezes, pode ser bem elevado.

Á luz da Proposição 21 e a exposição que a precede e que foi marcada por XIII, você sabe que a margem de erro e o coeficiente de confiança estão amarrados entre si,e que a amarração depende de n e k. Exatamente falando, se exigir que γ seja $0,9$, então os valores numéricos de n e k definirão unicamente o valor de ε e nada garante que n e k sejam tais que garantam para ε o valor $0,04$. Portanto, a tarefa formulada não faz muito sentido.

Entretanto, para nos é óbvio que a pessoa que colocou a tarefa não estaria insatisfeita se ε fosse menor do que o desejado $0,04$, desde que, claro, o valor de γ seja mantido em $0,9$. Essa observação sugere que a tarefa original seja executável se for corrigida da seguinte maneira: determinar o tamanho mínimo de amostra $n_{\min}$ da maneira tal que após que essa for feita, para qualquer seu resultado k, seja possível estimar a proporção (de bolas pretas da urna) com coeficiente de confiança $0,9$ e a margem de erro não maior que $0,04$.

Acontece que a tarefa corrigida realmente tem solução. Essa está dada pela Proposição 22 abaixo. Observe que a proposição cuida para recolocar a tarefa em seu formato correto que está pouco diferente daquilo formulado no parágrafo acima. O emprego dessa proposição na prática está mostrado pelo Exemplo 58. Vale você lé-lo, pois ele apresenta as idéias básicas das quais você precisará para fazer exercícios. Alguns detalhes secundários, porém não dispresíveis, estão unidos no Exemplo 59.

Proposição 22 *(sobre a determinação do tamanho de amostra no procedimento de estimação intervalar para proporção).*
O tamanho mínimo de amostra ($n_{\min}$) a ser feita para garantir que para qualquer resultado (k), a desconhecida proporção populacional (p) possa ser estimada com a margem de eror de no máximo ε e o coeficiente de confiança de no mínimo γ fixados antemão, determina-se pelas seguintes fórmulas:

$$n_{\min} = M_{\mathcal{D}} \left(\frac{z}{\varepsilon}\right)^2 \qquad \text{caso for conhecido apriori que } p \tag{8.25}$$

pode estar somente no conjunto $\mathcal{D}$; neste caso

$$M_{\mathcal{D}} = \max_{x\in\mathcal{D}} x(1-x); \tag{8.26}$$

em particular, se nenhuma informação acerca de p estar disponível, então $\mathcal{D} = [0,1]$, $\max_{x\in[0,1]} x(1-x) = 0,5(1-0,5)$ (Figura 8 explica esse valor máximo), e consequentemente

$$n_{\min} = 0,5(1-0,5)\left(\frac{z}{\varepsilon}\right)^2 = 0,25\left(\frac{z}{\varepsilon}\right)^2 \tag{8.27}$$

Em todas as fórmulas, z é um número positivo tal que $\gamma = I\!P[-z \le Z \le z]$, com $Z \sim \mathcal{N}(0,1)$.

↓ **Exemplo 58** (novamente, uma história que aconteceu de verdade). Um $\mathcal{F}$undo de $\mathcal{P}$ensão resolveu comprar de um $\mathcal{B}$anco uma carteira da pagamentos de uma parte dos clientes do $\mathcal{B}$anco.[1] Especificamente falando, a história era assim. O $\mathcal{B}$anco possui uma quantidade enorme de clientes que tomaram emprestimo e pagavam parcelas devolvendo o mesmo (com juros e tudo

[1] Com o intuito de não revelar os nome verdadeiros, vou me referir às instituições envolvidas nesta história por $\mathcal{F}$undo de $\mathcal{P}$ensão e $\mathcal{B}$anco.

aquilo que faz o banco ser um banco). O $\mathcal{B}$anco concordou em passar os futuros pagamentos de uma parte de seus clientes para o $\mathcal{F}$undo de $\mathcal{P}$ensão; em troca, o $\mathcal{F}$undo de $\mathcal{P}$ensão paga para o $\mathcal{B}$anco, no momento da assinatura do contrato, uma vez só uma certa quantia; denotamos essa por x.

É uma operação que interessa o $\mathcal{B}$anco, pois este pode usar o recebido x para oferecer novos emprestimos.

Para o $\mathcal{F}$undo de $\mathcal{P}$ensão,a operação também é boa, pois ele desembolsa x agora, mas vai receber, nos meses subsequentes, um fluxo de valores que em somatória dá mais que x (por causa dos juros embutidos nas parcelas dos tomadores de emprestimo, a soma das parecelas a vencer é maior – as vezes, muito maior – que x). Como o $\mathcal{F}$undo de $\mathcal{P}$ensão paga pensões mensalmente, então este "alongamento" dos recebimentos não é prejudicial.

Para dar o preço, o $\mathcal{B}$anco fez conta tal simples qual o pure de batata: "A soma que emprestei para meus clientes é z; estes já me devolveram y; então se receber agora x, fico satisfeito." (Os valores numericos de x, y e z não nos interessam neste exemplo.)

Para o $\mathcal{F}$undo de $\mathcal{P}$ensão, entretanto, o preço justo a pagar agora depende daquilo que será recebido no futuro. Esse, por sua vez, depende da proporção de pessoas que iam deixar de pagar suas parcelas; isto chama-se *taxa de inadimplência*. Para saber, se uma dada pessoa tornar-se-á inadimplete, é precisa analizar seus cadastro e histórico de pagamentos e aplicar alguns critérios que darão resposta binária: ou "sim, este será um bom pagador e pagará todas as parecelas" ou "não, este é um pagador ruim, num momento ele deixará de pagar suas parcelas". Só que aqui começam os problemas: a quantidade dos clientes na carteira a ser comprada é tal emensa que analizar todos é um trabalho muito custoso.

▷ Devo realçar que no que foi descrito até o momento, não há nada que pode ser interpretado ou modelado por conceitos relacionados à probabilidade; em particular, aos quem entende do tema Risco de Crédito, peço não relacionar nenhum aspecto do presente problema ao conceito "probabilidade de default". A probabilidade que surgir-se-á nesta história é a probabilidade embutida na construção do método de estimação intervalar, cujo papel no desenrolar do problema está lhe contado em seguida.

Podemos dizer que o $\mathcal{F}$undo de $\mathcal{P}$ensão está com uma urna que contém bolas pretas e brancas na quantidade N igual á quantidade de clientes da carteira do $\mathcal{B}$anco que será comprada pelo $\mathcal{F}$undo de $\mathcal{P}$ensão; bolas brancas são os clientes que pagarão direitinho todas as futuras parcelas do emprestimos que tomaram do $\mathcal{B}$anco, enquanto que as bolas pretas correspondem àqueles clientes que deixarão de pagar num momento futuro. A proporção de bolas pretas (p) está desconhecida e deseja-se estimá-la.

Então, o $\mathcal{F}$undo de $\mathcal{P}$ensão contratou uma $\mathcal{E}$mpresa de $\mathcal{C}$onsultoria que sugeriu fazer a estimativa intervalar por amostragem – exatamente aquilo que você aprendeu nas sub-seções anteriores, e aquilo que a empresa de consultoria explicou para os diretores do fundo. "Otimo", - disseram os diretores após ter ouvido a explicação e ter entendido as interpretações de γ e de ε: "o coeficiente de confiança desta estimativa deve ser 99%. Tal valor é a exigência do Banco Central que fiscalisa transações do tipo que pretendemos fazer com o banco. Se for menor, o Banco Central pode punir o fundo com base na estratégia arriscada de investimento."' Então, a $\mathcal{E}$mpresa de $\mathcal{C}$onsultoria deduziu que

$$0,99 \text{ é o valor desejado do coeficiente de confiança} \tag{8.28}$$

"E pedimos,- continuaram os diretores, –"Que a margem do erro seja $0,1\%$, o que é $0,001$ em valores absolutos; pois o grupo de clientes que pretendemos adquirir contem uns dezenas de milhar de clientes, fato que faz com que $0,1\%$ desse grupo já é valor alto para a contabilidade de nosso fundo." Portanto,

$$0,001 \text{ é o valor desejado da margem de erro} \tag{8.29}$$

Para comprir o prometido, a $\mathcal{E}$mpresa de $\mathcal{C}$onsultoria vai executar o método de estimação intervalar descrito no meu texto acima. Como o objetivo é alcançar certo coeficiente de confiança com certo margem de erro, a execução compõe-se da solução de duas tarefas:

Tarefa 1: Achar o tamanho de amostra a ser feita (n) que garante que o intervalo de estimação construido com base nos resultados da amostra tenha os valores desejados (ou melhores que os desejados) de seu coeficiente de confiança e de sua margem de erro.

Tarefa 2: Proferir a amostra e construir o intervalo.

A Tarefa 1 resolve-se seguindo a Proposição 22. No primeiro passo, acha-se z a partir de γ desejado:

$$\gamma = 0,99 \implies z \text{ satisfaz } \mathbb{P}[-z \leq Z \leq z] = 0,99 \implies z = 2,58$$

Depois, é a vez da fórmula (8.27), que dá a resposta final:

$$n = 0,25 \left(\frac{2,58}{0,001} \right)^2 \times 0,25 = 1\,664\,100$$

A magnitude do valor da resposta fez com que dois representantes da $\mathcal{E}$mpresa de $\mathcal{C}$onsultoria contratada apareceram na minha sala no Instituto de Matemática e Estatística. Suspeito alias, que os dois eram toda a empresa. Veio um senhor de gravata colorida com pasta de couro alemã, e uma moça de sapato Prado, coleção verão 2012 que adequava-se bem ao inverno brasileiro 2013. Não me entenda mal: os dois vestiam outra roupa além da gravata e sapatos supramencionados, mas só estes dois itens fincaram-se na minha memória. Então o engravatado e a calçada explicaram para mim que o banco de dados com as caraterísticas dos clientes é muito mal estruturado e que a procura por informação de um cliente demora ums minutos, e que, consequentemente, a análise de amostra de tamanho 1 664 100 demoraria 3 anos.

Sugeri então usar o método de redução, quer dizer, sugeri seguir as fórmulas (8.25), (8.26) da Proposição 22. Para isso precisava achar um subconjunto de $[0, 1]$ (denotado por $\mathcal{D}$ no enunciado da proposição) de tal sorte que possa ser garantido que p está nele. Alcançamos isso da seguinte maneira. Consultamos os dados do Banco Central e vimos que para os grupos de clientes parecidos com os do grupo em questão, a proporção de inadimpletes nunca ultrapassou 6%. Então, aceitamos que

$$p \in (0,\ 0,06],\ \text{ou, em termos da proposição, que } p \in \mathcal{D} \text{ com } \mathcal{D} = (0,\ 0,06] \tag{8.30}$$

▷ Como afirmações do tipo (8.30) causam frequentemente confusão nos mentes de meus leitores e alunos, gostaria de colocar explicitamente que (8.30) não é nenhuma estimativa intervalar para o desconhecido p. É uma informação que adveio de uma fonte externa.

Para prosseguir, precisamos achar o máximo que a função $x(1-x)$ pode assumir enquanto x percorre por $\mathcal{D}$ (a Proposição 22 usa a notação $M_{\mathcal{D}}$ para o máximo agora procurado). Farei isto aqui com auxílio de um desenho que foi colocado na Figura 8. Este desenho junto com a argumento relacionado requerem sua atenção redobrada pois apresentam a única ferramenta necessária para que você possa resolver os exercícios sobre a estimação do tamanho de amostra similhantes ao que estamos tratando agora.

Do lado esquerdo do desenho, há ao grafo da função $x(1-x)$ para os valores de x entre 0 e 1. Estes dois valores são os limites para a posição de p no caso quando não se sabe nada sobre p. No caso estudado agora, assumimos que p não pode ultrapassar $0,06$, quer dizer, os limites para p tornaram a ser 0 e $0,06$. Como a Proposição 22 manda a gente procurar pelo máximo da função $x(1-x)$ no intervalo $(0,\ 0,06]$, então eu recortei do desenho à esquerda somente a parte que interessa, ampliei-la e coloquei à direita. O intervalo $(0,\ 0,06]$ está destacado em preto. Os valores da função em cima deste intervalo também estão destacados em preto. É fácil ver que o máximo da parte destacada da função assume-se para $x = 0,06$ (quer dizer, assume-se

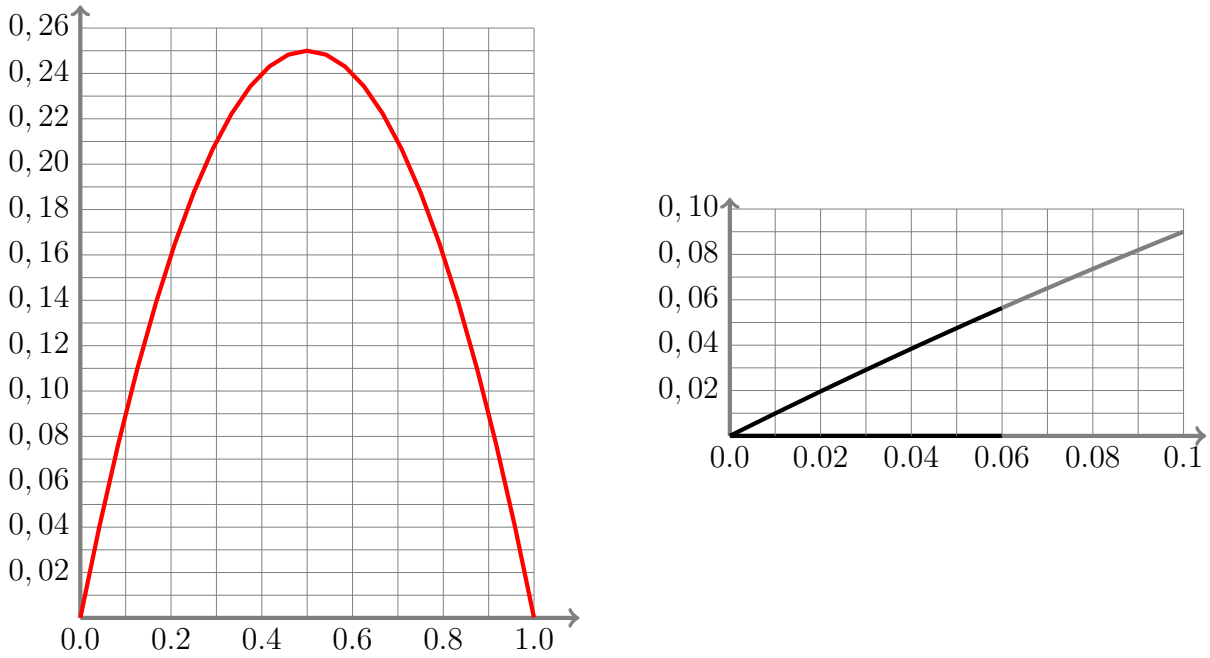


Figura 8.1: Os gráficos que são usados na aplicação da Proposição 22 para o Exemplo 8.1.2. O gráfico à direita é a parte inicial da parábola desenhada no gráfico à esquerda. O gráfico à direita parece ser uma reta (embora não é) por duas razões: por ser um trexo muito curto de parábola, e, em segundo lugar, pelo fao do desenho ter escalas diferentes em seus eixos.

8

no limite direito do intervalo) e é igual a $0,06(1-0,06)$. Em notações da proposição, tem-se: $M_{\mathcal{D}} = 0,06(1-0,06)$.

Então, segundo a proposição, ao saber por certo que $p \in (0,\ 0,06]$, o tamanho de amostra

$$n = 0,06(1-0,06)\left(\frac{2,58}{0,001}\right)^2 \times 0,06(1-0,06) = 375\,420,96 \text{ arred. de por cima pelo } 375\,421$$

torna-se o suficiente para garantir que após a amostragem e indpendetemente do seu resultado (k), seja possível estimar p com margem de erro de no máximo $0,001$ e coeficiente de confiança de no mínimo $0,9$.

Poderia agora fechar a exposição do exemplo, pois alcancei meu objetivo: mostrei um caso real que precisava da técnica de redução. Mas eu gostaria de contar a história até o fim para lhe mostrar a sensibilidade da fórmula à mudança de valores de sesu parâmetros.

Então, continuo. Como 375 421 foi julgado ser número ainda grande demais para o tamanho de amostra, fomos procurar os meios da sua diminuição. Acreditamos que o Banco Central não reclamaria se tomassemos confiança a ser 96% em vez de 99%. Menos que 96% não dava, pois nossa experiência nos dizia que valores menores que 96% nunca passaram pela auditoria do Banco Central. Então, tercendo que o Banco Central aciete o valor 96% para o coeficiente de confiança, refizemos a conta:

$$\gamma = 0,96 \implies z \text{ satisfaz } \mathbb{P}[-z \leq Z \leq z] = 0,96 \implies z = 2,06$$

$$n = 0,06(1-0,06)\left(\frac{2,06}{0,001}\right)^2 = 239\,339,04 \text{ arred. de por cima pelo } 239\,340$$

mas achamos que este ainda esteja alto demais. Resolvemos então sugerir que a precisão da

estimação seja duas vezes pior: $\varepsilon = 0,002$. Refizemos a conta de novo:

$$n = 0,06(1-0,06)\left(\frac{2,06}{0,002}\right)^2 = 59\,834,76 \text{ arred. de por cima pelo } 59\,835$$

e por aqui paramos. Os representantes da $\mathcal{E}$mpresa de $\mathcal{C}$onsultoria, satisfeitos, se despediram de mim. Eles nunca me contaram como foi a intervalo de confiança que saiu após a amostragem (quer dizer, não fui informado sobre a execução daquilo que chamamos acima por "Tarefa 2"), e talpouco sei como tudo isso foi recebido pela direitoria do $\mathcal{F}$undo de $\mathcal{P}$ensão. Mas ganhei uma experiência interessante, e estou contente por poder compatilhar esta com meus alunos.

Fim do Exemplo ↑

8.1.3 Um exemplo que toca nos aspectos pouco discutidos até o momento

Há alguns aspectos do tema Estimação Intervalar de Proporção que ficaram fora do foco principal das seções anteriores. Esses "despercebidos e despresados" são os itens do exemplo da presente seção. Em cada item, o correspondente comentário revela o que tem de particular em cada especto.

↓ **Exemplo 59.** **(a)** Necessita-se estimar, por intervalo de confiança cuja margem de erro seja de no máximo $0,05$ e cujo coeficiente de confiança seja de no mínimo $0,9$, a proporção de pessoas contaminadas pelo virus de um gripe numa certa região. Qual é o tamanho mínimo da amostra que deve ser feita para atender tais necessidades?

A solução: $\gamma = 0,9 \Rightarrow z = 1,65 \Rightarrow n_{\min} = 0,5(1-0,5)\left(\frac{1,65}{0,05}\right)^2 = 272,25 \rightarrow n_{\min} = 273$
Comentário: O aredondamento aqui em todos as outras aplicações semelhantes da Proposição 22, é para cima, pois a proposição disse que n não pode ser menor que $272,25$ e, portanto, a resposta 272 não serve.

(b) As autoridades de uma região informaram que 25% de toda a população foram vacinados antes do surto. Essa informação permite diminuir o tamanho de amostra calculado no item (a) (desde que os requesitos sobre o coeficiente de confiança e a margem de erro sejam mantidas)? Entende-se que pessoa vacinada não contamina-se pelo virus.

A solução: Se 25% da população não podem ser contaminados, então p, a proporção de contaminados, pode assumir qualquer valor entre $0,25$ e 1. Em termos da Proposição 22, tem-se:

$$p \in \mathcal{D}, \text{ onde } \mathcal{D} = (0,25;\ 1]$$

O máximo da função $x(1-x)$ no domínio $\mathcal{D}$ alcança-se no ponto $x = 0,5$ e é igual $0,5(1-0,5)$ (para descobrir esse fato, você não precisa analisar a Figura 8 e não mais que isso). Aplicando a Proposição 22, conclui-se que a resposta é $n_{\min} = 273$. Eis o cálculo que leva a tal resposta:

$$\gamma = 0,9 \Rightarrow z = 1,65 \Rightarrow n_{\min} = 0,5(1-0,5)\left(\frac{1,65}{0,05}\right)^2 = 272,25 \rightarrow n = 273$$

Comentário: Ao comparar o resultado presente com o obtido no item (a), vê-se que são iguais, ou que, em outras palavras, a informação acrescida ao item (b) não ajudou reduzir o valor de $n_{\min}$ em comparação com aquele valor que foi obtido no caso quando nenhuma informação sobre p era disponível. Se mergulharmos nos cálculos proferidos em ambos os casos, veremos a razão desse não melhoramento: o máximo da função $x(1-x)$ no intervalo $[0,1]$ coincidiu com o máximo dessa função no domínio $\mathcal{D} = (0,25;\ 1]$.

(c) As autoridades de uma outra região informaram que 75% de toda a população foram vacinados antes do surto. Essa informação permite diminuir o tamanho de amostra calculado no item (a) (desde que os requesitos sobre o coeficiente de confiança e a margem de erro sejam mantidas)?
Entende-se que pessoa vacinada não contamina-se pelo virus.

A solução: Se 75% da população não podem ser contaminados, então p, a proporção de contaminados, pode assumir qualquer valor entre $0,75$ e 1. Em termos da Proposição 22, tem-se:

$$p \in \mathcal{D}, \text{ onde } \mathcal{D} = (0,75;\ 1]$$

O máximo da função $x(1-x)$ no domínio $\mathcal{D}$ alcança-se no ponto $x = 0,75$ e é igual $0,75(1-0,75)$. Aplicando a Proposição 22, conclui-se que a resposta é $n_{\min} = 205$. Eis o cálculo que leva a tal resposta:

$$\gamma = 0,9 \ \Rightarrow\ z = 1,65 \ \Rightarrow\ n_{\min} = 0,75(1-0,75)\left(\frac{1,65}{0,05}\right)^2 = 204,1875 \ \rightarrow\ n = 205$$

Comentário: O máximo da função $x(1-x)$ no intervalo [0, 1] é maior do que seu máximo no domínio $\mathcal{D} = (0,75;\ 1]$. É por isso que $n_{\min} = 273$ obtido no item (a) reduziu-se para 205 com a ajuda da informação de que $p \in (0,75;\ 1]$.

(d) Uma equipe de médicos foi estimar a proporção de pessoas contaminadas pelo virus de um gripe numa região do país. Desejava-se obter estimativa intervalar para a proporção populacional cuja margem de erro seja de no máximo $0,05$ e cujo coeficiente de confiança seja de no mínimo $0,9$. Como não se sabia nada apriori sobre a proporção a ser estimada, aplicou-se o resultado do item (a) acima, isto é foi tomada uma amostra aleatória de 273 pessoas da região. Foi constatado que 82 delas eram contaminadas pelo vírus. Com esse resultado, qual é a estimativa da proporção populacional dos contaminados para a margem de erro $0,05$. Qual é o valor de seu coeficiente de confiança?

A solução segue-se pelo caminho das fórmulas (8.10)–(8.12). Primeiramente, apresenta-se a estimativa intervalar:

$$\left[\tfrac{k}{n} - \varepsilon,\ \tfrac{k}{n} + \varepsilon\right] = \left[\tfrac{82}{273} - 0,05,\ \tfrac{82}{273} + 0,05\right]$$
$$\text{aproximadamente } [0,3 - 0,05;\ 0,3 + 0,05] = [0,25;\ 0,35]$$

Agora, vamos calcular seu coeficiente de confiança:

$$\tfrac{k}{n} = \tfrac{82}{273} \approx 0,3 \ \Rightarrow\ z = \varepsilon/\sqrt{\tfrac{k/n(1-k/n)}{n}} = 0,05/\sqrt{\tfrac{0,3(1-0,3)}{273}} \approx 1,80 \Rightarrow$$
$$\Rightarrow\ \gamma = 2(0,9641 - 0,5) = 0,9282$$

Comentário: O coeficiente de confiança saiu maior que $0,9$. Isso não disqualifica a resposta, pois a tarefa dizia "no mínimo $0,9$". E meu presente comentário é justamente sobre a extrema inteligência da tarefa que não institiu erroneamente que γ deva ser $0,9$ e nada diferente disso. Tal inteligência adveio da demonstração da Proposição 22: a demonstração sabia que n, ε e γ estão amarrados pela relação

$$n = p(1-p)\left(\frac{z_\gamma}{\varepsilon}\right)^2 \tag{8.31}$$

(onde eu escrevi z_γ no lugar de z para assinalar que z e só z representa γ nessa relação). A relação revela que uma vez que fixamos ε e γ que desejamos ter após a amostragem, tem-se que n depende de p via a expressão $p(1-p)$. Para cada p seria "seu" n, mas como p está desconhecido (e só pode ser desconhecido, já que toda nossa conversa é sobre a estimação de

p) então escolha-se seu maior valor possível, que é o valor correspondente ao $0,5(1-0,5)$ no lugar de $p(1-p)$ na expressão (8.31) (já que $0,5(1-0,5)$ é o máximo de $p(1-p)$ no intervalo de valores de p entre 0 e 1). Então, temos:

$$n_{\min} = 0,5(1-0,5)\left(\frac{z_\gamma}{\varepsilon}\right)^2 \tag{8.32}$$

Imagine agora que formos colher amostra de tamanho $n_{\min}$ e, a partir dos resultados da amostra, calculamos $\hat{p} = k/n_{\min}$. Por um lado, como a consequencia das fórmulas (8.10)–(8.12), esperamos que deve valer a igualdade

$$n_{\min} = \hat{p}(1-\hat{p})\left(\frac{z_\gamma}{\varepsilon}\right)^2 \tag{8.33}$$

Por outro lado, se $\hat{p}$ estar diferente de $0,5$, então a igualdade (8.33) não é possível já que, no caso $\hat{p} \neq 0,5$, ela contradiz à relação (8.32). Então, no caso $\hat{p} \neq 0,5$, para forçar a igualdade (8.33), ou aumenta-se γ ou diminui-se ε. No presente item, γ subiu de $0,9$ para $0,9282$. O item seguinte mostra o caso quando diminui-se ε.

(e) Nas condições do item (d) foi decidio construir o intervalo de confiança com o coeficiente de confiança igual a $0,9$. Qual é a margem de erro e qual é o intervalo?

A solução: $\gamma = 0,9 \Rightarrow z = 1,65 \Rightarrow \varepsilon = 1,65\sqrt{\frac{0,3(1-0,3)}{273}} = 0,0457 \approx 0,046$.

$$I.C.(p;0,9) = [0,3-0,046\,;\ 0,3+0,046] = [0,234\,;\ 0,346]$$

Comntário: veja o Comentário ao item (d).

Fim do Exemplo 59↑

8.2 Estimação intervalar de proporção em detalhes

De acordo com meu plano de ensino, no presente momento, você, meu leitor, deve estar com conhecimento da execução do método de estimação intervalar de proporção e sua experiência com esse deve ter gerado na sua cabeça a pergunta sobre o sentido do elemento de método chamado "coeficiente de confiança". Recordo-lhe que nos exemplos da Seção 8.1 tal coeficiente foi interpretado por algo vinculado a probabilidade, mas tal interpretação nunca adquiriu explicação rigorosa. Ainda mais: foi sempre lhe dito que a interpretação probabilística não é bem óbvia pois no ambiente para o qual a estimação intervalar de proporção foi lhe apresentada não há experimento aleatório a acontecer, e, portanto, não há probabilidade em sua essência pura como a inserteza vinculada a algo que acontecerá.

A presente seção destina-se à explicação rigorosa do sentido de coeficiente de confiança no método de estimação intervalar de proporção. Preste a atenção ao que a explicação seguir-se-á da visão bocado artificial em tudo que acontece no método. Essa visão é comum na Teoria Estataística. Diria que ela é o princípio central daquilo que os cientistas chamam por **inferência frequentista**. Ela aprecerá toda vez que tentarei apresenta-lhe os motivos matemáticos para a construção de métodos estatísticos estudados no nosso curso. Agora pouco chamei-a "artificial" pois estou convencido que nenhum de meus leitores-de-primeira-viajem-em-estatística conseguiria sozinho elaborar tal visão. Acerca desse "fracasso", gostaria de lhe contar que alguns de meus alunos-iniciantes-em-estatística conseguiram elaborar uma base probabilística para métodos estatísticos para eles ensinados, mas todos os esses "sucessos" levaram à construção de visão **bayesiana** a qual é uma alternativa e um concorrente principal para a visão frequentista. Tal "acerto" involutário mostra que **inferência bayesiana** é algo mais natural para seres humanos com mentes intocadas por formalismo matemático. Essa naturalidade deve ser reconhecida e, como o reconhecimento, planejo acrescentar a meu livro um capítulo ou mais sobre essa tal de inferência bayesiana.

Diria que há dois pilares que sustentam toda a explicação.

O primeiro pilar é que há um experimento aleatório no ambiente para o qual desenvolvemos a estimação intervalar de proporção. Tal experimento aleatório sempre pode ser representado em termos de bolas-na-urna e a representação é assim: Há ma urna com bolas idênticas no tato, mas pintadas de duas cores diferentes: preta e branca. A proporção das bolas pretas denota-se por p. Retira-se da urna um certo número de bolas, ao acaso e com reposição, e observam-se as cores das bolas retiradas.

O segundo pilar é que há duas visões nesse experimento aleatório. Uma delas será chamada **visão pós-experimental** pois corresponde exatamente aquilo que se vé após ter executado o experimento aleatório. Desse ponto de vista, o experimento aleatório resulta numa sequencia de n números composta de 1's e 0's, a qual está denotada por

$$x_1, x_2, \ldots, x_n \tag{8.34}$$

e na qual o número na i-ésima posição é 0 se a i-ésima bola foi branca, e é 1 caso ela foi preta. Vale ainda dizer explicitamente que no ambiente agora descrito n considera-se conhecido. Já a visão chamada **pré-experimental** é a de observador colocado no eixo de tempo antes do experimento aleatório acontecer, e portanto, a cor de cada bola pode ser tanto branca quanto preta. É importantissimo que essa inserteza pode ser expessada por variável aleatória. Nessa expressão, a aparecer agora, eu já vou dar o nome cómodo para a variável aleatória. Então, a cor a ser vista da i-ésima a ser retirada denota-se agora pela X_i sobre a qual sabe-se a cara de sua distribuição:

valor	0	1
a probabilidade de X_i assumir o valor	$1-p$	p

Na construção da variável aleatória X_i, codificamos as cores "branca" e "preta" por 0 e 1 respectivamente. Então, podemos agora "anotar" tudo que será observado por

$$X_1, X_2, \ldots, X_n \tag{8.35}$$

Tal escrita naturalmente levanta a pergunta sobre a relação entre as suas variáveis aleatórias . A resposta é que elas são independentes em conjunto. (ACRESCENTAR O ARGUMENTO QUE JUSTIFICA A INDEPENDÊNCIA.)

Observe que até o momento não houve mencionada nenhum problema ou objetivo por cima do experimento aleatório. Isso será feito agora.

Então suponha que veio uma pessoa (chamaremos essa por Iris), que na realidade pós-experimental faz o seguinte:

cores na urna são desconhecidos, e a tarefa é estimar a proporção de bolas pretas usando amostra. A qualidade e a forma de estimação ainda serão discutidas e especificadas no que se segue. No momento, vamos lembrar as notações e os conceitos que serão usados na exposição e discussão a seguir.

Denoto por p a proporção de bolas azuis. p é desconhecido.

Vou estimar p da seguinte maneira.

Vou retirar n bolas ao acaso e com reposição. Aqui n não é um valor desconhecido. É uma notação genérica. Vou contar o número de bolas azuis nesta amostra. A notação para o resultado da contagem é X. Observe: considero o resultado como variável aleatória pois o experimento (retirada aleatória com reposição) ainda não aconteceu. Sabemos que

$$X \sim Bin(n, p) \tag{8.36}$$

(embora não sabemos o valor de p, podemos alegar (8.36)).

A princípio, X poderá ser qualquer valor inteiro entre 0 e n. Sua média é np e sabe-se que a probabilidade de X assumir valor próximo a sua média é maior que a de um valor afastado da média. Isto sugere que

$$\frac{X}{n} \tag{8.37}$$

seja um bom estimador para o desconhecido p. (Se X tem tendência estar próximo a np, então X/n tem tendência estar próximo a p.)

Mas como não há garantia nenhuma que X assuma valor np e portanto, X/n assuma valor exatamente p, então sugiro estimar p não pontualmente mas por meio de um intervalo. Sugiro que este seja simêtrico e com centro em X/n (sugestão razoável):

$$[X/n - \varepsilon, X/n + \varepsilon] \tag{8.38}$$

Aqui, ε não é incognito. É uma notação genérica

Pergunto: Com qual probabilidade meu intervalo "acerta" em p? Eis a expressão formal para a pergunta: Achar $I\!P\{p \in [X/n - \varepsilon, X/n + \varepsilon]\}$.

Na descrição acima, muios aspectos ficaram obscuros; por exemplo, o significado da "confiança", ou a omissão da informação acerca da determinação dos valores de ε e de γ: "É o interessado na solução quem deve sugerir os valores ao estatístico, ou é este quem determiná-os?" A falta de clareza é onus do objetivo pretencioso do conteudo deste seção: introduzir um procedimento sem explicar os motivos e os detalhes de sua construção. Entretanto, creio que podemos continuar andando neste caminho reto, pois lá, no final, uma boa parte da dúvidas cairão. É calro que para o esclarecimento completo, o caminho de explicação deve ser outro. Este será exhibido na próxima seção para os quem quizer ou precisar aprender todos os "por quês".

Eis a resposta: (abaixo uso o seguinte fato válido para quaisquer números reais:

$$a - \varepsilon \leq b \leq a + \varepsilon \text{ se e somente se } b - \varepsilon \leq a \leq b + \varepsilon$$

$$\begin{aligned} &I\!P\left\{X/n - \varepsilon \leq p \leq X/n + \varepsilon\right\} \\ &= I\!P\left\{p - \varepsilon \leq X/n \leq p + \varepsilon\right\} \\ &I\!P\left\{n(p-\varepsilon) \leq X \leq n(p+\varepsilon)\right\} \end{aligned}$$

Substituo X por $Y \sim \mathcal{N}\left(np, np(1-p)\right)$ (é aqui que usamos fato que $X \sim Bin(n,p)$), e faço a padronização:

$$\begin{aligned} &I\!P\left\{n(p-\varepsilon) \leq Y \leq n(p+\varepsilon)\right\} \\ &= I\!P\left\{-\frac{n\varepsilon}{\sqrt{np(1-p)}} \leq \frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{np}{\sqrt{np(1-p)}}\right\} \\ &= I\!P\left\{-\frac{n\varepsilon}{\sqrt{np(1-p)}} \leq Z \leq \frac{np}{\sqrt{np(1-p)}}\right\} \end{aligned}$$

Interpretação: Se for fazer etimações de p por intervalos

$$[X/n - \varepsilon, X/n + \varepsilon]$$

então a proporção dos intervalos que acertarão o verdadeiro valor de p será

$$I\!P\left\{-\frac{n\varepsilon}{\sqrt{np(1-p)}} \leq Z \leq \frac{np}{\sqrt{np(1-p)}}\right\}$$

Em segundo lugar, o ambiente do presente exemplo nos permite a interpretar a confiança da estimativa intervalar. Recorde, nossas contas concluiram que esta é 95,44%. Então, se Zé perguntar:
– Está garantido que a proporção populacional de meus eleitores esteja entre 51% e 59%?
Ai, responderemos: – Não, só podemos dar a confinaça de 95,44% disso.
Se Zé tivesse noção básica de probabilidade, ele intepretaria a resposta como a chance dele ganhar a eleição. A intepretação está correta, e, em geral, está correta a ponte que Zé fez ligando a interpretação profana mas prático com nossa resposta que carrega o termo científico "confiança". As pontes do tipo do Zé você vai erguer ao solucionar exercícios deste capítulo, e isto lhe dará a prática para entender a lógica da interpretação.

Já a interpretação rigorosa exige um mergulho na construção do intervalo de estimação, e isto será feito na seção seguinte. Para os quem preferir a dispensar a leitura daquela seção, aviso que o assunto não é trivial. De fato, suponha que Zé estudou a Teoria de Probabilidades e portanto sabe que a probabilidade surge só onde e quando há um experimento aleatório. Então, suponha que ele pergunta:
– Onde está o experimento aleatório que resultará na minha eleição com aquela prometida probabilidade $0,9544$ (e, alternativamente, na eleição de meu concorrente com a probabilidade $1 - 0,9544$)?
Você, meu leitor, ficou agora confuso? Se sim, não é à toa, pois a aleatoriedade está ausente. De fato, cada pessoa já escolheu em quem irá votar, e a proporção daqueles que votarão no Zé está lá, na população; somos nós quem desconhece o valor desta proporção, mas ela existe e é nada aleatória.

Bom, e agora! Vai fazer o que perante este paradoxo? Acalme-se, pois os quem inventou o método pensou em tudo. A ausência do experimento aleatório, que é um fato óbvio para os profissionais, fez com que a palavra "probabilidade" foi banida da resposta final. No seu lugar entrou **o nível de confiança** (chamado alternativamente **grau de confiança** e **coeficiente de confiança**). Deste jeito, os guardiões do rigor não reclamam, e, por outro lado, os quem tiram conclusões práticas da estimação por intervalo, interpretam "confiança" como "probabilidade" e tal interpretação é precisa e correta o suficiente para que suas conclusões aproveitem dos resultados da estimação.

Infelizmente, a resposta depende do valor desconhecido de p e, ao rigor, não pode ser empregada. Felizmente, ao substituir o descolhecido p por conhecido (após ter feito o experimento!!!) valor X/n, não cometeremos um grande erro no cálculo desta probabilidade (as contas que justificam isto são chatas).

Vejamos como isto funciona na prática.

Suponha que decidi por $n = 100$. Ao fazer meu experimento, observei 20 bolas azuis. Suponha que decidi por $\varepsilon = 0,03$. Então, o *intervalo de confiança* resultante é

$$[0,2 - 0,03\,;\, 0,2 + 0,03]$$

e a correspondente probabilidade de acerto *confinaça* é

$$\begin{aligned} &I\!P\left\{-\frac{100\cdot 0,03}{\sqrt{100\cdot 0,2(1-0,2)}} \leq Z \leq \frac{100\cdot 0,03}{\sqrt{100\cdot 0,2(1-0,2)}}\right\} \\ &= I\!P\left\{-0,75 \leq Z \leq 0,75\right\} = 2\left(A(0,75) - 0,5\right) = 2(0,7734 - 0,5) = 0,5468 \approx 0,55 \end{aligned}$$

A interpretação científica é: o intervalo de confiança para o desconhecido p é

$$[0,2 - 0,03\,;\, 0,2 + 0,03]$$

com coeficiente de confiança $\gamma = 0,55$.

A interpretação profana desta interpretação é: com confiança de 55% podemos alegar captaram p por $[0,2 - 0,03\,;\, 0,2 + 0,03]$.

Esta última pode ser corrigida assim: Se fizessemos 1.000.000 (muitas quer dizer) repetições deste procedimento, teriamos muitos intervalos diferentes (pois nada garante que ao repetir 100 retiradas teriamos de novo 20 bolas azuis). 55% destes captariam (cobririam) o verdadeiro valor de p. Aquele único intervalo que veio de nosso único experimento, pode estar em 55% dos casos felizes, assim como em 45% dos casos infelizes. Então damos 55% de confiança para ele.

Prosseguimos então. A primiera coisa a fazer é olhar o desenho abaixo que exemplifica a situação

Neste desenho, $n = 10$ e $k = 7$. Observe que o valor de N naõ foi especificado, e observe que, de fato, N não interessa pois retiramos bolas com reposição.

Algumas das caraterísticas do exemplo acima apresentado enquadram-se dentro da nomenclatura introduzida nos capítulas anteriores; observe a relação pois precisaremos dela no futuro: as bolas na urna – população N – o tamanho da população $p, 1-p$ a distribuição populacional de frequencia sobre as duas cores (preta e branca) da população n tamanha da amostra $\frac{k}{n}$, $1-\frac{k}{n}$ a distribuição amostral de frequenca sobre as duas cores (preta e branca) da amostra

Vamos introduzir mais dois termos: p a proporção populacional $\frac{k}{n}$ a proporção amostral

A fórmula é

$$\left\{\begin{array}{l} \varepsilon \;\approx\; z\sqrt{\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}} \\ \text{onde } z>0 \text{ satisfaz a equação } \gamma = I\!P\left[-z \leq Z \leq z\right] \text{ com } Z \sim \mathcal{N}(0,1), \\ \text{e onde a aproximação segue-se da aplicação do Teorema Central} \\ \quad \text{de Limite, para cuja validade no caso é necessário que } np \text{ seja} \\ \qquad\qquad \text{suficientemente grande} \end{array}\right. \tag{8.39}$$

A ressalva é levantada pelo fato da fórmula ter sido derivada com emprego do Teorema Central de Limite, ou, mais especificamente, daquela consequencia do teorema que garante a aproximação da distribuição binomial pela distribuição normal. Na apresentação desta aproximação que fizemos no Capítulo ??, avisamos que a mesma realmente aproxima bem uma distribuição pela outra só quando o produto np assume valor suficientemente grande (aquí, n e p são os parâmetros da distribuição binomial). .... Então, vamos "fechar os olhos" para este detalhe.

Na equação, k e n são valores conhecidos, uma vez que a amostra é conhecida. Portanto, a equação nos mostra que os parâmetros da estimação ε e γ são vinculados via uma equação. Isto significa que

(a) uma vez que o valor de ε foi estabelecido (oriundo de algumas considerações da natureza prática, sobre as quais ainda versaremos no futuro), o correspondente valor de γ segue-se da Eq.(8.12);

(b) uma vez que o valor de γ foi estabelecido (oriundo de algumas considerações da natureza prática, sobre as quais ainda versaremos no futuro), o correspondente valor de ε segue-se da Eq.(8.12);

As propriedades (a) e (b) agora descobertas, sendo transcritas na âmbito da relação professor-aluno, soam assim:

em qualquer exercício sobre o tema "estimação intervalar de proporção a partir de amostra já feita" a parte computacional está vinculada completamente e exclusivamente à relação (8.39), e ocorre uma das duas:

(a) ou professor fornece o valor do parâmetro ε e pede calcular o valor de γ;

(b) ou professor fornece o valor do parâmetro γ e pede calcular o valor de ε.

Isto foi sobre a parte computacional. A outra parte, a conceitual, está ligada à interpretação dos parâmetros ε e γ. Esta será introduzida nos exemplos abaixo. Adiantando-me, gostaria de lhe dizer que a interpretação da parâmetro γ pode ser entendida por completo só em conjunto com a consideração do método que deduz a relação (8.12). Portanto, espere até Seção ??.

8.3 "De no mínimo" e "de no máximo" nas respostas e arredondamento por de cima

Parece ser simples. Eis a solução:

$$n = \left(\frac{z}{\varepsilon}\right)^2 \hat{p}(1-\hat{p}) \tag{8.40}$$

entretanto, a coisa só parece ser simples, e a simplicedade surgiu oriundo de um pequeno engano: escrevi na fórmula $\hat{p}$ em vez de k/n. Se o último estivesse aparecido na fórmula, você reconheceria de imediato que seu valor numérico só será conhecido após a amostragem e concluiria, com toda a razão, de que a estratégia (8.40) é inexecutável.

Entretanto, há um argumento que revitaliza a fórmula (8.40). Ele é muito simples, fato que faz todo o assunto aparecer nos temas ensinados neste curso. Suponha, só para o titulo de um exemplo passageiro, que você colocou, sem nenhuma razão sólida, $\hat{p} = 0,3$ na equação, e suponha que você desja que γ seja $95,44\%$ e que ε seja $0,02$. As contas do exemplo anterior mostra que ao $\gamma = 0,9544$ corresponde $z = 2$. Coloque os valores na equação. Esta vai nos dar:

$$n = \left(\frac{2}{0,02}\right)^2 \times 0,3(1-0,3) = 10.000 \cdot 0,3 \cdot 0,7 = 2100$$

quer dizer

$$0,02 = 2\sqrt{\frac{0,3(1-0,3)}{2100}} \tag{8.41}$$

Agora supoonha que você foi e retirou uma amostra e tamanho 2100, mas ela acusou $\hat{p} = 0,4$. Como

$$0,4(1-0,4) > 0,3(1-0,3)$$

então (8.41) não procede, e no seu lugar ocorre

$$0,02138 = 2\sqrt{\frac{0,3(1-0,3)}{2100}}$$

o que significa que você não consegue que a margem de erro seja o desejado valor 0,02; a margem ficou maior.

Isto mostra que a flutuação do valor de $\hat{p}(1-\hat{p})$ que você não controla antes de ir fazer amostra pode estragar sua estimativa, se ele acontece a ser maior que você usou. Esta conclusão motiva a pensaqr sobre qual é o maior valor que a expressão $\hat{p}(1-\hat{p})$ pode ter. Se tal valor existir, vale coloca-lo nas contas pois pior que ele não vai acontecer. Felizmente, há uma resposta simples nesta pergunta: o maior valor que $\hat{p}(1-\hat{p})$ pode assumir quando $\hat{p}$ estiver entre 0 e 1 é $0,5(1-0,5) = 0,25$. O desenho abaixo comprova este resultado.

8.3.1 Sem amostra

O assunto principal desta sub-seção exige alguns preparativos, com os quais começaremos nossa exposição.

Vamos itroduzir dois termos que não são tradicionais para a linguagem da Teoria de Estatística, mas que são apelativos à intuição. A margem de erro (denotado por ε) vamos chamar de **precisão numérica** de intervalo de estimação, e o coeficiente de confiança (denotado por γ) vamos chamar de **precisão probabilística**.

quanto maior o valor de ε (quer dizer, maior o valor da margem de erro), pior é a precisão numérica

quanto maior o valor de γ (quer dizer, maior o valor do coeficiente de confiança), melhor é a precisão probabilística

Na posse de uma amostra, as precisões numérica e probabilística estão amarradas: para um valor de uma delas corresponde um e único valor da outra. Suponha então que desejamos a estimativa intervalar cujas precisões não estejam na referida correspondência. Esta tarefa está discutida na presente seção. Acerca desta, nossa intuição sugere que, talvéz, seja necessário fazer uma nova amostra. Veremos já-já que a intuição sugere caminho correto. Mas antes disto, observamos que se a amostra a fazer deva ser maior que a que já temos, então a tentação é acrescentar novas observações às que já existem. Tal acrescimo exige um cuidado, que, por sua vez, exige uma explicação. Também, cabe a explicação a respeito se seja possível aproveitar uma parte da amostra existente, caso a nova amostra for menor desta. Para eliminar as explicações, encurtando e simplificando assim a exposição, optamos por assumir que não haja amostra alguma no momento de colocação da tarefa a discutir. Começamos formulando a resposta.

que o intervalo de estimação a ser construido com base nesta tenha e suponha que desejamos que o intervalo de estimação tenha o que estou perante aquela minha urna já desenhada na sub-seção anterior, e suponha que assim como antes, desejo estimar a proporção populacional das bolas pretas da urna, só que agora antes de tirar amostra, vou encomendar (determinar,

em outras palavras) os parâmteros da estimação, tipo:
–Quero que a margem de erro seja $0,03$ e o coeficiente de confiança seja $0,96$.

Porquê pedir os valores de parâmetros antes da amostragem e porquê tal pedido está sendo tratado numa sub-seção separada segue-se daquela fórmula central do método de estimação por nos discutido. De fato, olhe novamente nela apresentada da seguinte maneira:

$$\begin{array}{l}\varepsilon = z\left(\gamma\right)\sqrt{\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n}},\\ \text{onde } z(\gamma) \text{ significa que } z \text{ depende exclusivamente de } \gamma,\\ \text{a dependência sendo aquela que você já conhece: } \gamma = I\!P\left[-z \leq Z \leq z\right]\end{array} \tag{8.42}$$

Nesta apresentação, os destaque em vermelho esclarecem a seguinte

Propriedade: no método de estimação intervalar aqui apresentado, seus parâmetros ε, margem de erro, e γ, coeficiente de confiança, estão vinculados por uma equação, e esta é tal que com valores de n e $\hat{p}$ fixos, a cada valor de ε corresponde um e único valor de γ, e, *vice versa*, a cada valor de γ corrreponde um e único valor de ε.

Segue-se disto, que com a amostra na mão, há pares fixos de valores de ε e γ, e se você desejar construir intervalo de confiança com um par diferente, pode não conseguir.

Mas e se há razões fortes para ter estimativa com os valores desejados que não concordam entre si pela equação? A própria equação dá-nos

A abordagem: coloque os valores desejados dos parâmetros ε e γ na equação, e tira dela o valor de n. Este é o tamanho de amostra da qual precisa para satisfazer suas exigências.

Entretanto, vé-se claramente que na fórmula aparece mais um parâmetro, que é k, número de "bolas pretas na amostra"(falando em termos de bolas-na-urna). Este obviamente, só será revelado após a amsotragem. Portanto, a abordagem acima sugerida não é implementável. Não obstante, o argumento apresentado abaixo corrige esta falha.

Está certo que o verdadeiro valor que $\hat{p} = \frac{k}{n}$ assumirá só será conhecido após a amostragem. Entretanto, ele nunca será menor que 0 e nunca maior que 1. No intervalo de todos os valores possíveis, quer dizer no intervalo $[0,1]$, a expressão $\frac{k}{n}\left(1-\frac{k}{n}\right)$ não fica maior que

$$0,5(1-0,5) = 0,25 \tag{8.43}$$

Este fato fica óbvio, se você substituir $\frac{k}{n}$ por x na experessão, para que esta fique com a cara das funções cujos máximos e mínimos você analisava na época do colégio. Então, em termos daquela época, você precisa achar o máximo da função $f(x) = x(1-x)$ no intervalo de valores $x \in [0,1]$. Ao escrever $f(x)$ como $x - x^2$ você vé que esta função é uma parábola invertida e deslocada. É facil ver que sua cimeira esta no ponto $x = 0,5$, e, portanto, a altura da cimeira é $0,25$.

Com esta substituição, você Esta fica então com

O argumento apresnetado acima aplica-se no caso quando nada sobre $\hat{p}$ é conhecido apriori. Existem situações nas quais há uma informação adicional (que não tem nada aver com a obtenção da amostra) que garante que $\hat{p}$ pode esta só num domínio $\mathcal{D}$ (no Exemplo 3 abaixo, este domínio é o intervalo $(0;0,06]$)

Esta considera a ser uma boa **estimativa pontual** para a proproção populacional (aquela que foi denotada por nós pelo p). A bondade de $\frac{k}{n}$ como estimativa não será dicutida aqui e agora. Chamo sua atenção ao termo "pontual"; ele quer dizer que estima-se p por um número. Tal estimação apresenta alguns desvantagens que podem ser eliminadas esta possui uma notação alternativa que é $\hat{p}$. Cada uma das duas notações, tem suas vantagens. Por exemplo, $\hat{p}$ indica para leitor que trata-se de estimativa pontual para p, e isto justifica que se desejar-se construir

uma estimativa de p por intervalo, então este deve ser centrado em $\hat{p}$, quer dizer, tenha a cara

$$\text{proporção amostral } - \text{tiquinho, proporção amostral } + \text{tiquinho} \tag{8.44}$$

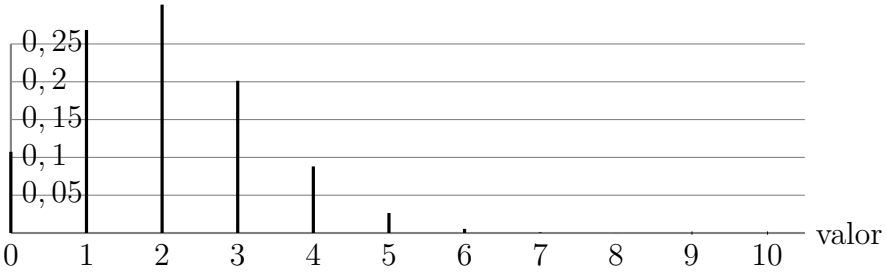
quer dizer, a cara apresentada em (8.10), se formos usar notação formal.[2] Mas o uso da notação $\frac{k}{n}$ traz mais vantagens, e é por isto que usarei esta predominantemente.

Dois desenhos que mostram a flutuação probabilística do centro de IC

0.9672065 é a probabilidade de "cobertura" de $p = 0,2$ por

$$\left[\frac{X_1 + \ldots + X_{10}}{10} - 0,25\,;\ \frac{X_1 + \ldots + X_{10}}{10} - 0,25\right]$$

FUNÇÃO DE PROBABILIDADE de $B \sim Bin(10;\, 0,2)$

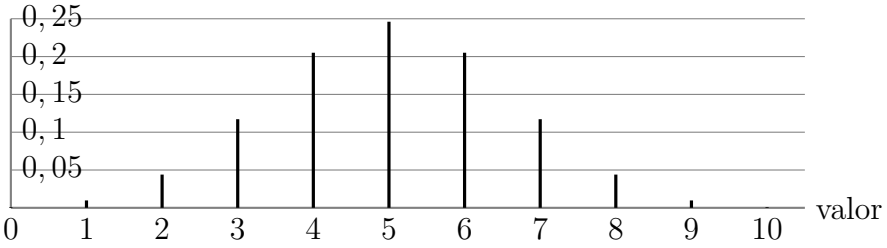


0.1073741824 0.2684354560 0.3019898880 0.2013265920
0.0880803840 0.0264241152 0.0055050240 0.0007864320
0.0000737280 0.0000040960 0.0000001024 (valores)

0,890625 é a probabilidade de "cobertura" de $p = 0,5$ por

$$\left[\frac{X_1 + \ldots + X_{10}}{10} - 0,25\,;\ \frac{X_1 + \ldots + X_{10}}{10} - 0,25\right]$$

FUNÇÃO DE PROBABILIDADE de $B \sim Bin(10;\, 0,5)$



0.0009765625 0.0097656250 0.0439453125 0.1171875000
0.2050781250 0.2460937500 0.2050781250 0.1171875000
0.0439453125 0.0097656250 0.0009765625 (valores)

[2]Recorde que num momento de minha apresentação, eu mencionei a regra de notação aceita em Estatística: um estimador para qualquer parâmetro u denota-se por $\hat{u}$. Preste a atenção: eu não provo aqui, assim como não provei em lugar algum, que $\frac{k}{n}$ é uma boa estimativa pontual para p. Só estou dizendo que a notação $\hat{p}$ arpou leitor pela intuição e faz este acreditar que entendeu as razões que fazem $\hat{p}$ estar no centro do intervalo.

8.4 Exercícios sobre a estimação intervalar de proporção

Exc. 130. 400 pessoas da população da Região Metropolitana (RM) da cidade de São Paulo com idade entre 30 e 60 anos, foram escolhidas ao acaso. Foi constatado que 35 são naturais do Minas Gerais. Pede-se estimar, por intervalo de confiança com a margem de erro de $0,005$, a proporção dos mineiros na população da RM de São Paulo que estejam na faxa etária 30 – 60 anos; pede-se também calcular o coeficiente de confiança dessa estimativa.

Exc. 131. 800 pessoas da população paulistana foram escolhidas ao acaso. Foi constatado que 420 delas são mulheres. Pede-se construir intervalo de confiança, com coeficiente de confiança 92%, para a proporção de mulheres na população paulistana.

Exc. 132. Ao intervistar 200 fregueses de um shoppin center, constatou-se que 25% deles residem longe (a mais de 10 km) do shopping. Construa intervalos de confiança para a verdadeira proporção de fregueses que moram longe do shopping center aos coeficientes de confiança de 85%, 90% e 95% (são solicitados três intervalos de confiança, um para cada valor do coeficiente de confiança). Compare os intervalos e comente.

Exc. 133. Suponha que ao intervistar 150 fregueses de um shoppin center, constatou-se que 25% deles residem longe (a mais de 10 km) do shopping. Suponha que no dia seguinte, foram intervistados 200 fregueses, e, novamente, descobriu-se que 25% deles residem longe. Suponha ainda que no terceiro dia, a amostra foi feita com 250 pessoas, e novamente, 25% deles revelaram que residem longe do shopping.

Construa três intervalos de confiança, um para cada amostra, com o mesmo coeficiente de confiança que é 88%. Compare os intervalos. Comente o resultado da comparação.

Exc. 134. (a) Pretendemos tomar uma amostra aleatória de n alunos da população de alunos da USP que estudaram a disciplina "Estatística Básica" no 1-o semestre de 2015, e intervistar os alunos escolhidos acerca de sua aprovação do método empregado no ensino da disciplina. Com base nesta amostra, pretendemos construir um intervalo de confiança para a proporção populacional dos quem aprova o método de ensino dentro dos alunos que estudaram a disciplina "Estatística Básica" no 1-o semestre de 2015. Qual deve ser o tamanho da amostra que garanta que o intervalo de confiança esteja com a margem de erro de no máximo $0,1$ e o valor de seu coeficiente de confiança esteja de no mínimo 85%?
(b) O enredo e a pergunta são os mesmos que no item (a), mas agora, para dar a resposta, use o fato que 74% é a porcentagem dos aprovados dentre os alunos da USP que estudaram a disciplina "Estatística Básica" no 1-o semestre de 2015, e junte esse fato com o fato que um aluno aprovado sempre responde que gostou do método de ensino.

Exc. 135. (a) Um cartaz no metrô de São Paulo proclama: 87% da população usa metrô pelo menos 5 vezes por semana.

Resolvi testar esta afirmação por amostragem e posterior construção de intervalo de confiança.

Desejo que o intervalo de confiança a ser construído tenha as seguintes caraterísticas: que sua margem de erro seja não maior que 0,01, e que seu coeficiente de confinaça seja não menor que 0,95. Qual é o número de cidadãos de São Paulo que devem ser intervistadas acerca dos costumes do uso de metrô para que o intervalo de confiança construido com base nos resutlados da amostra atenda aos requisitos desejados?

Comentário: Muitos alunos interpretaram esse enunciado da maneira muito mais complexa que era pretendido por mim quando compús-o. Isso requer de mim alguns esclarecimentos

que descomplicam a historinha aqui contada. Em primeiro lugar, é precisa imaginar todos os maradores de São Paulo como bolas, sendo que se morador usa metrô pelo menos 5 vezes por semana então ele representa-se por bola preta, e senão, então representa-se por bola branca. O cartaz supracitado alega que p, a proporção de bolas pretas, é 87%. Eu não acreditei nesse fato e quero testá-lo. O procedimento de "teste" é assim: vou tomar uma amostra de n bolas e com uso dela construir intervalo de confiança para p. Após a construção, verificarei se 87% esteja dentro do intervalo. Se estiver, vou acreditar na afirmação do cartaz, senão – não. Alias, vale notar que poderei também dizer que esse teste da afirmação de cartaz possui "probabilidade de acerto" que é igual ao valor do coeficiente de confiança do intervalo de estimação para p. Mas tudo isso será feito e dito somente após ter conseguido amostra. Já a questão do exercício é sobre o tamanho da amostra a ser feita. Com isso sendo esclarecido, fica evidente que "87%" não desempenhará nenhum papael na solução da questão. Em outras palavras, a questão é um puro e limpo exemplo de questão sobre o dimencionamente de amostra em função dos desejados limites para a margem de erro e o coeficiente de confiança.

(b) Na prefeitura me disseram que no mínimo 30% da população de São Paulo trabalha e anda de metrô no percurso de seu domicílio ao trabalho; estas pessoas usam o metrô no mínimo duas vezes por dia, portanto, no mínimo 10 vezes por semana. Esta informação ajuda a diminuir o tamanho da amostra a ser feita?

(c) Uma fonte de informação revelou, que a porcentagem fornecida pela prefeitura está superada e que há novas pesquisas que mostram que esta porcentagem é 80% e não 30%, quer dizer, no mínimo 80% da população usa metrô 5 vezes ou mais por semana. Assumindo isto como fato verdadeiro, você consegue diminuir o tamanho da amostra?

8.5 Soluções dos exercícios sobre a estimação intervalar de proporção

Solução do Exc. 132. Em todos os três casos, o valor da proporção amostral foi $0,25$. Nas notações da teoria apresentada nas aulas, isso significa que $\frac{k}{n} = 0,25$. Na solução apresentada abaixo, usa-se a notação $\hat{p}$ para $\frac{k}{n} = 0,25$. O outro símbolo usado na solução é $I.C.(p;\gamma)$; esse leia-se assim: "o intervalo de confiança para p com o coeficiente de confiança γ". Em todos os três casos, usa-se a fórmula

$$I.C.(p;\gamma) = \left[\hat{p} - z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\ \hat{p} + z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

onde z determina-se pelo γ de acordo com a relação $\gamma = I\!P\left[-z \leq Z \leq z\right]$.

Ao $\gamma = 0,85$ corresponde $z = 1,44$, assim:

$$\begin{aligned} I.C.(p;0,85) &= \left[0,25 - 1,44 \times \sqrt{\tfrac{0,25\times 0,75}{200}};\ 0,25 + 1,44 \times \sqrt{\tfrac{0,25\times 0,75}{200}}\right] \\ &\approx [0,206;\ 0,294] \end{aligned}$$

Ao $\gamma = 0,90$ corresponde $z = 1,64$, assim:

$$\begin{aligned} I.C.(p;0,90) &= \left[0,25 - 1,64 \times \sqrt{\tfrac{0,25\times 0,75}{200}};\ 0,25 + 1,64 \times \sqrt{\tfrac{0,25\times 0,75}{200}}\right] \\ &\approx [0,20;\ 0,30] \end{aligned}$$

Ao $\gamma = 0,95$ corresponde $z = 1,96$, assim:

$$\begin{aligned} I.C.(p;0,95) &= \left[0,25 - 1,96 \times \sqrt{\tfrac{0,25\times 0,75}{200}};\ 0,25 + 1,96 \times \sqrt{\tfrac{0,25\times 0,75}{200}}\right] \\ &\approx [0,19;\ 0,31] \end{aligned}$$

Podemos notar que se usermos a mesma amostra e aumentemos o coeficiente de confiança γ, então a margem de erro aumenta, quer dizer, aumento o tamanho do correspondente intervalço de confiança, o que signiica, em termos práticos, a piora da precisão da estimativa.

Solução do Exc. 133. Para $\gamma = 0,88$, temos que $z = 1,55$. Em todos os três casos, $\frac{k}{n} = \hat{p} = 0,25$. O que muda é o valor de n.

Para $n = 150$,

$$\begin{aligned} I.C.(p;0,88) &= \left[0,25 - 1,55 \times \sqrt{\tfrac{0,25\times 0,75}{150}};\ 0,25 + 1,55 \times \sqrt{\tfrac{0,25\times 0,75}{150}}\right] \\ &\approx [0,195;\ 0,305] \end{aligned}$$

Para $n = 200$,

$$\begin{aligned} I.C.(p;0,88) &= \left[0,25 - 1,55 \times \sqrt{\tfrac{0,25\times 0,75}{200}};\ 0,25 + 1,55 \times \sqrt{\tfrac{0,25\times 0,75}{200}}\right] \\ &\approx [0,203;\ 0,297] \end{aligned}$$

Para $n = 250$,

$$\begin{aligned} I.C.(p;0,88) &= \left[0,25 - 1,55 \times \sqrt{\tfrac{0,25\times 0,75}{250}};\ 0,25 + 1,55 \times \sqrt{\tfrac{0,25\times 0,75}{250}}\right] \\ &\approx [0,208;\ 0,292] \end{aligned}$$

Podemos notar que, caso $\frac{k}{n}$ e γ permaneçam fixos, mas n cresce, a margem de erro descresce, fazendo o intervalo de confiança diminuir; em termos práticos isso significa o aumento da precisão.

8.6 Estimação da média de uma distribuição normal; o resumo

Toda a teoria, que está apresentada abaixo e usada em sequencia para solução de exemplos e exercícios, funciona desde que a população tratada nesses exemplos e exercícios tenha distribuição normal. Acerca dessa condição, vale notar que quando o enunciado de exemplo ou de exercício diz que "uma população tem distribuição aproximadamente normal", você então, ao elaborar a solução, tratará essa população como se ela fosse perfeitamente normal.

Em todos os exemplos e exercícios do tema tratado na presente seção, o objetivo é estimar a média da população. Mas como a população segue uma distribuição normal – conforme o pressuposto primordial explicado no parágrafo acima – então o objetivo supraformulado chama-se também de estimação da média de distribuição normal. Essa média, que é o objeto de estimação e que sempre está considerada como desconhecida, denota-se por μ.

Em tudo o dito até o momento e no tudo a ser dito para frente, há uma imprecisão na terminologia. A saber: "a população tem distribuição normal" é a forma errada de dizer que "a distribuição da frequencia relativa por atributo de interesse na população é normal". Mas a segunda frase é mais comprida e menos cómoda que a primeira, fato que faz com que essa está mais usada que aquela.

A fonte principal para a estimação é amostra simples aleatória denotada por

$$x_1, x_2, \ldots, x_n \tag{8.45}$$

Especificamente falando, x_i denota o valor do atributo (a média de cuja distribuição é o obejto de estimação) medido para i-ésimo individuo retirado aleatoreamente da população. Assume-se que cada indivíduo retirado devolve-se à população após a medição. Assume-se também que na retirada, cada individuo tem a mesma chance de ser escolhido.

Na presente seção, estaremos discutindo a estimativa intervalar para média que tem o seguinte formato:

$$[\bar{x} - \varepsilon,\, \bar{x} + \varepsilon]\,, \tag{8.46}$$

Outras maneira de estimação intervalar existem, mas todas elas estão fora do escopo da apresentação. Na fórmula (8.46), $\bar{x}$ chama-se **média amostral** e defina-se pelo sequinte:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n} \tag{8.47}$$

Quanto ao significado de ε usado em (8.46), esse será apresentado em seguida. No momento, é importante lhe informar que cada intervalo de estimação (8.46) possui seu **coeficiente de confiança** denotado por γ. O significado desse também será apresentado em seguida. No momento, só observo que a qualidade do intervalo representado pelo esse coeficiente faz com que seu nome é **intervalo de confiança**.

Agora vamos falar de γ e de ε. O parâmetro γ é interpretado como a probabilidade do que o intervalo de confiança construído de fato contem o valor verdadeiro da média estimada. O parâmetro ε, que você vê em (8.46), chama-se por **margem de erro** e é a distânica máxima com a qual a média estimada desvia-se da sua estimativa pontual caso o intervalo de estimação de fato captou o verdadeiro valor da média estimada. Vale enfatizar que a **estimativa pontual** agora mencionada é – por decfreto – $\bar{x}$ (é por causa do seu uso como estimador pontual para μ que $\bar{x}$ denota-se as vezes por $\hat{\mu}$).

Observe que a palavra "média" está usada em dois conceitos: um deles é "média populacional" (o que é o objeto de estimação), e o outro é "média amostral" (que é o estimador pontual para a média populacional e também toma-se como o centro do intervalo de estimação para a média populacional). As vezes, no calor de discussão, os adjetivos "amostral" e "populacional" são omitidos, mas sempre podem ser reconstruidos a partir de contexto.

Entre as questões que surgem no ambiente de estimação intervalar acima formulado há três que serão tradadas no presente texto. São essas:

(i) Dado o valor numérico da margem de erro (valor de ε, quer dizer) achar o valor do coeficiente de confiança (isto é, o valor de γ) para o intervalo de confiança $[\bar{x} - \varepsilon,\, \bar{x} + \varepsilon]$.

(ii) Dado o valor do coeficiente de confiança (γ), achar o valor numérico da margem de erro (ε) que faz com que o intervalo $[\bar{x} - \varepsilon,\, \bar{x} + \varepsilon]$ estime a média populacional com a desejada confiança γ.

(iii) Dados os valores do coeficiente de confiança (γ) e da margem de erro (ε) estabelecer o tamanho de amostra mínimo $n_{\min}$ tal que se formos fazer amostra de tamanho $n \geq n_{\min}$ e, com base nos resultados da amostra e do valor de ε dado, construirmos o intervalo $[\bar{x} - \varepsilon,\, \bar{x} + \varepsilon]$, então o coeficiente de confiança desse será no mínimo o valor dado de γ.

As soluções das questões (i)-(iii) dependem de se

(a) a variância da distribuição normal (cuja média é o objeto de estimação) é conhecida (a variância denotar-se-á por σ^2), ou

(b) tal variância é desconhecida.

Antes de prosseguir, vale notar que a situação (a) é possível embora é muito estranha, a estranhesa sendo o fato que nessa situação sabe-se o valor da variância mas não o da média de distribuição normal.

Vamos agora à descrição dos métodos de solução das questões (i)-(iii). No caso (a), cada uma das soluções deduz-se da seguinte relação entre n, ε e γ:

$$\varepsilon = z\sqrt{\frac{\sigma^2}{n}},\;\; \gamma = I\!P\left[-z \leq Z \leq z\right], \text{ onde } Z \sim \mathcal{N}(0, 1^2) \tag{8.48}$$

No caso da questão (i), a sequencia que leva de γ (dado) para ε (procurado) é assim: primeiramente acha-se z que satisfaz a segunda fórmuala em (8.48), e coloca-se o valor achado na primeira fórmula para que essa forneça o valor de ε. No caso da questão (ii), a sequencia de uso das fórmulas se inverte: primeiramente acha-se z segundo a fórmula $z = \varepsilon\sqrt{\frac{n}{\sigma^2}}$, e depois determina-se γ a partir de z. Finalmente, no caso (iii), primeiramente usa-se a segunda fórmula de (8.48) para achar z, e depois usa-se a primeira das fórmulas para calcular $n_{\min} = \frac{z^2\sigma^2}{\varepsilon^2}$.

Preste a atenção que as fórmulas (8.48) que vinculam ε (a margem de erro) e γ (o coeficiente de confiança), determima a relação entre γ e z

$$\text{via } \gamma = I\!P\left[-z \leq Z \leq z\right] \text{ e } \underline{\text{não}} \text{ via } \gamma = I\!P\left[Z \leq z\right]$$

Essas duas fórmulas são diferentes, fato que está ilustrado na Figura 8.2, onde o desenho á esquerda corresponde à primeira fórmula e o à direita a segunda; ambos os desenhos ilustram suas respectivas fórmulas para o caso $z = 1,65$. Como se pode ver, no caso da relação correta, a resposta não adveio diretamente da Tabela da Distribuição Acumulada para a Normal

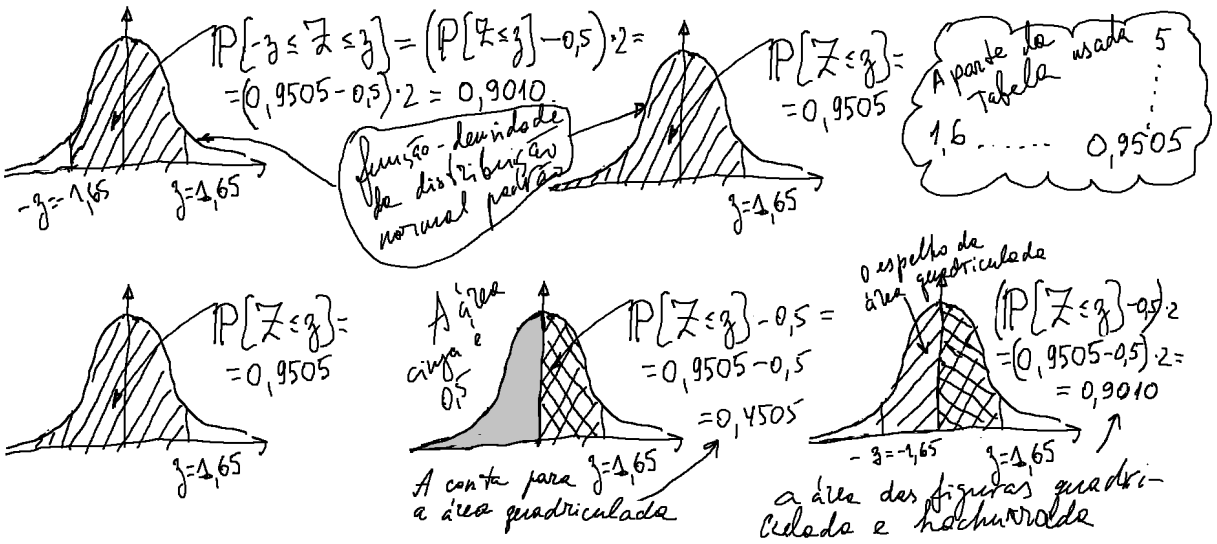


Figura 8.2: A ilustração do funcionamento da fórmula $\mathbb{P}[-z \leq Z \leq z]$ que faz parte do vinculo entre ε (a margem de erro) e γ (o coeficiente de confiança) (veja (8.48)). A figura mostra também a diferença desse funcionamento com o da fórmula $\mathbb{P}[Z \leq z]$.

Padrão, diferentemente do outro caso, no qual a resposta foi dada diretamente pela Tabela. O procedimento adicional necessário está ilustrado nos desenhos da fileira inferior da Figura 8.2.

No caso (b), as fórmulas das quais deduzam-se todas as soluções são assim:

$$\varepsilon = z\sqrt{\frac{s^2}{n}}, \quad \gamma = \mathbb{P}\left[-z \leq T_{n-1} \leq z\right] \tag{8.49}$$

onde s^2 (denotada também por s_x^2) é a estimativa pontual para σ^2 feita com base na amostra (8.45) de acordo com uma das três seguintes fórmulas (naturalmente, equivalentes entre si):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \;=\; \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 \;- n(\bar{x})^2\right) \;=\; \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 \;- \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right) \tag{8.50}$$

Na fórmula (8.49), T_{n-1} é uma variável aleatória com a distribuição específica chamada "t de Student com $n-1$ graus de liberdade" (lembro que n usado na notação dessa variável aleatória significa o tamanho de amostra). Os valores dessa distribuição são tabeladas. A tabela aparece em literatura estatística em lugares diferentes com nomes diferentes, mas em cada caso, o nome contêm alguma referência ao nome completo: "Tabela da Distribuição t de Student". A tabela apresentada na Figura 8.3, por exemplo, chama-se "t-Table".

No caso (b), as sequencias de ações que que resolvem questões (i) e (ii) são as mesmas que as já apresentadas para o caso (a); a diferença é que usam-se as fórmulas (8.49) em vez das (8.48). A questão (iii) não faz sentido no caso (b), pois sua solução estaria dada pela fórmula $n_{\min} = \frac{z^2 s_x^2}{\varepsilon^2}$ a qual envolve s_x^2 cujo valor será conhecido somente após ter feito amostra.

Exemplo 60. Podemos assumir que o tempo de rodagem de pneus de uma certa marca tenha distribuição aproximadamente normal. Estime, por intervalo de confiança com o coeficiente de confiança $0,8$, a média dessa distribuição usando a seguinte amostra que apresenta o tempo de rodagem (em anos) de de $n = 13$ pneus da marca escolhidos aleatoreamente:

1, 1.4, 2, 2.4, 2.7, 2.9, 3.1, 3.5, 3.9, 4, 4.6, 5.2, 6.1

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figura 8.3: Essa é a tabela de limiares para as distribuições t de Student com diversos valores do parâmentro "graus de liberdade" (denotado na tabela por "**df**"). A linha **z** da tabela assinala que para qualquer valor dos graus de liberdade maior que 1000, a correspondente distribuição t de Student está próxima à distribuição Normal Padrão, e, por causa dessa proximidade, os limiares podem ser obtidos consultando a Tabela da Distribuição Normal Padrão. Tal tabela dá os valores de limiares apresentados pela tabela presente na linha dela marcada por "**z**". Aviso ainda que para o uso na construção de intervalos de confiança é cómodo usar a marcação de colunas apresentada na última linha da tabela, pois nessa linha marca-se diretamente o coeficiente de confiança (*confidence level*, γ) da correspondente coluna. E por fim, apresento aqui explicitamente a interpretação do valor em cada célula do corpo da tabela: na célula encontrada na intersecção da linha k com a coluna marcada por $A\%$ encontra-se o valor de z tal que $A\% = \mathbb{P}[-z \leq T_k \leq z]$ onde T é a variável aleatória t de Student com k graus de liberdade.

Solução. A média dessa amostra é

$$\bar{x}_{13} \equiv \hat{\mu} = \frac{1 + 1.4 + \cdots + 6.1}{13} = 3,292308 \approx 3,3$$

Ela toma-se como a estimativa pontual de μ, a média populacional.

A variância dessa amostra é

$$s_x^2 = \frac{1}{13-1}\left(\sum_{i=1}^{13} x_i^2 \; - \; n(\bar{x})^2\right) = 2,165769 \approx 2,17$$

e, portanto, o desvio padrão é $s_x = \sqrt{2,165769} \approx 1,47$. Esses tomam-se como as estimativas pontuais para a variância (σ^2) e o desvio padrão (σ) populacionais.

O intervalo de confiança tem o seguinte formato (os dois abaixo são, obviamente, equivalentes entre si):

$$\left[\hat{\mu} \; - \; z\sqrt{\frac{s_x^2}{n}};\; \hat{\mu} \; + \; z\sqrt{\frac{s_x^2}{n}};\right], \text{ ou } \left[\hat{\mu} \; - \; z\frac{s_x}{\sqrt{n}};\; \hat{\mu} \; + \; z\frac{s_x}{\sqrt{n}};\right]$$

O que nos falta para dar o resultado em termos de intervalo numérico é achar o valor numérico da expressão $z\frac{s_x}{\sqrt{n}}$. Nessa expressão, só o valor de z não foi estabelicido até o momento. Ele virá do valor desejado do coeficiente de confiança (que é $0,8$) com auxílio de t-Table.

Eis os passos da "consulta" da tabela. Toma-se a linha marcada à esquerda por "12" pois no nosso caso, o número de graus de liberdade é $13-1$ ("graus de liberdade" é igual, por definição, ao "tamanho de amostra" -1). Depois, toma-se a coluna marcada "80%" nas marcações entituladas "Confidence level". Na intersecção entre a linha e a coluna escolhidas, há o valor que nos interessa: $z = 1,356$.

Então, podemos concluir a solução:

$$\varepsilon = 1,356 \times \frac{1,47}{\sqrt{13}} = 1,356 \times \frac{1,47}{3,6055} \approx 0,5528$$

e, consequenteente, o intervalo de confiança é:

$$[3,3 - 0,5528;\; 3,3 + 0,5528]$$

8.7 Estimação de média; apresentação detalhada

Aqui vai ser colocado o texto de aula que até o momento só existe na versão escrita à mão.

8.8 Exercícios sobre a estimação intervalar de média

Ex. 136. Uma máquina de encher pacotes de café pode ser regulada para qualquer peso de pacotes. Porém, devido ao processo tecnológico, sempre haverá pequenos desvios do peso regulado, para mais ou para menos. É conhecido que estes desvios têm distribuição normal com média 0 e o desvio padrão 0,01 Kg (10g). Um supermercado recebeu um lote de café em pacotes de 1 kg. **(a)** Deseja-se verificar se a máquina que os encheu de fato foi regulada para 1kg. Deseja-se que a estimativa tenha a precisão de 0,005 (5 g) e que o coeficiente de confiança desta seja 90%. Quantos pacotes de café devem ser tomados na amostra para que estas exigências sejam satisfeitas?
(b) Deseja-se verificar se a máquina que os encheu de fato foi regulada para 1kg. Deseja-se que a estimativa tenha a precisão de 0,005 (5 g) e que o coeficiente de confiança desta seja 95%. Quantos pacotes de café devem ser tomados na amostra para que estas exigências sejam satisfeitas? **(c)** Para uma amostra de tamanho 100, qual é a confiança que o intervalo centrado na média amostral cujo tamnho é 0,003 ($\varepsilon = 0,003/2 = 0,0015$) está contendo o verdadeiro valor do peso médio das pacotes do lote recebido?
Observação: É óbvio que o tamanho do intervalo de confiança $[\bar{x}_n - \varepsilon\,;\bar{x}_n + \varepsilon]$ é 2ε, mas que a precisão da estimativa é ε e isto ocorre pois por "precisão" entende-se a diferença entre o valor estimado μ por sua estimativa $\bar{x}_n$. Tendo em mente esta diferença, é natural perguntar se o valor de ε desejado seria $0,003/2$ em vez de $0,003$. É a questão de interpretação. Na minha interpretação, o quem compus o exercício equivocou-se e escreveu "tamanho" onde quiz "precisão".

Exc. 137. Sabe-se por certo que a altura das pessoas de qualquer população, que têm os mesmos hábitos alimentícios e são expostas às mesmas condições climáticas, segue a distribuição normal com desvio padrão de 0,1 m.
(a) Para estimar a altura média da população de uma ilha, medimos as alturas de 30 pessoas escolhidas ao acaso da população dos moradores da ilha. A média amostral foi de 1,45 m. Com base nestes valores, qual é o intervalo de confiança da altura média da população com coeficiente de confiança de 95%.
(b) Qual seria o tamanho da amostra a ser tomado para que o erro cometido seja menor que 0,05 metros com confiança de 99%.

Exc. 138. O objetivo de uma pesquisa é estimar o tempo médio que um certo analgésico demora a fazer efeito. Este foi aplicado aos 100 pacientes e foram observados os tempos da reação para cada um deles. A média das observações foi de 20 minutos e o desvio padrão delas foi de 6 minutos. Com base nestes dados, afirmamos que o tempo médio da reação está entre 19 e 21 minutos.
(a) Qual é o valor do coeficiente de confiança desta estimativa?
(b) Qual o tamanho da amostra a ser tomada para obter um intervalo de tamanho 1 minuto com confiança 95%?

Exc. 139. Em períodos de pico, os clientes de um banco são obrigados a enfrentar longas filas para sacar dinheiro nos caixas eletrônicos. Dados históricos de vários anos de operação indicam que o tempo de transação nesses caixas tem distribuição normal com a média igual a 270 segundos. Para aliviar esta situação , o banco resolve instalar, em caráter experimental, alguns caixas eletrônicos de concepção mais avançada. Os tempos de 24 transações escolhidos ao acaso nesses caixas são
240 245 286 288 238 239 278 287 291 248 257 225 257 264 282 252 243 260 248 259 262 271 234 250
O banco deve trocas os caixas antigos por caixas testados?

Exc. 140. (Fonte: Bussab & Morettin "Estatística Básica" 5-a edição, pag. 308.) Calcule o intervalode confinaça para a média de uma distribuição normal em cada um dos casos abaixo.

Média amostral	Tamanho da amostra	Desvio Padrão da amostra	Coeficiente de confiança
170 cm	100	15 cm	95%
165 cm	184	30 cm	80%
180 cm	225	30 cm	70%

Exc. 141. (Fonte: Bussab & Morettin "Estatística Básica" 5-a edição, pag. 308.) De 50.000 válvulas fabricadas por uma companhia retira-se uma amostra de 400 válvulas e obtém-se a média das vidas úteis delas de 800 horas e o desvio padrão das vidas 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 da população encontra-se no intervalo $800 \pm 0,98$?

8.9 Soluções dos exercicios sobre estimação de média

Solução do Exc. 136. Que seja claro: "1kg" é o que está escrito nos pacotes de cafe recebidos por supermercado, mas o supermecado não tem certeza que isto é verdade, quer dizer, que a máquina que encha pacotes foi de fato regulada para 1kg.

Por não saber a regulagem da máquina, o peso média de pacotes de café, μ, é um incognito. Porém, devido a processo de enchimento, sabemos que para qualquer valor de μ, o desvio padrão do peso de pacotes de café é $\sigma = 0,01$Kg. Em outras palavras, sabemos o desvio padrão populacional mas não sabemos a média populacional e queremos estimar esta por intervalo de confiança.

No item **(a)** do exercício, o coeficiente de confiança da estimativa deve ser de 90%, e a precisão da estimativa deve ser $0,005$Kg. Recorde que a estimativa intervalar terá a "cara":

$$[\bar{x}_n - \varepsilon\,;\bar{x}_n + \varepsilon]$$

onde $\bar{x}_n$ denota a média amostral a ser construida com base na amosta de tamanho n, e

$$\varepsilon = \frac{z\sigma}{\sqrt{n}}, \text{ sendo que } I\!P\,[-z \leq Z \leq z] = \gamma.$$

No nosso caso, $\gamma = 0,9$, de onde (usando a tabela da distribuição normal padrão) $z = 1,65$, e $\varepsilon = 0,005$. Portanto, o desconhecido n deve satisfazer a relação

$$0,005 = \frac{1,65 \cdot 0,01}{\sqrt{n}} \;\Leftrightarrow n = \frac{0,0165^2}{0,005^2} = (3,3)^2 \approx 11.$$

(b) A diferença do item (a) está somente no valor de γ, que era de 90% e ficou agora em 95%. Isto afeta o valor de z: no item (a) era $z_{\gamma=0,9} = 1,65$, enquanto que agora $z_{\gamma=0,95} = 1,96$. Repetindo os argumentos do item (a), temos que

$$n = \left(\frac{z_\gamma \sigma}{\varepsilon}\right)^2 = \left(\frac{1,96 \cdot 0,01}{0,005}\right)^2 \approx 16.$$

(c) Aquí, a situação é pouco diferente daquela que era em comum nos itens (a) e (b). Agora a amostragem foi feita: $n = 100$ pacotes foram pesados. Deseja-se então saber a confiança (γ) do intervalo de confiança para a média populacional cuja precisão é de 0,003 (isto é: $\varepsilon = 0,003$). Observe que o resultado da amostragem, apropreadamente dito, não foi revelado pelo enunciado, isto é, não foi dado para nos o valor que $\bar{x}_{n=100}$ assumiu. Mas tambem, não precisamos deste valor para responder na pergunta colocada, pois da fórmula $\varepsilon = \frac{z\sigma}{\sqrt{n}}$ deduz-se que

$$z = \frac{\varepsilon\sqrt{n}}{\sigma} = \frac{0,003\sqrt{100}}{0,01} = 3.$$

De onde, com o uso da tabela da distribuição Normal Padrão, segue-se que

$$\gamma = I\!P\,[-3 \leq Z \leq 3] = 0,9973.$$

Solução do Exc 137 (a). Pelo enunciado, $\sigma = 0,1$ m, quer dizer, o desvio padrão populacional da altura de toda a população é 0,1 m, valendo para quelquer que seja a altura média populacional.

Neste item, sabemos que $n = 30$ e que $\bar{x}_{n=30}$ assumiu valor 1,45 m, e queremos contruir o intervalo de confiança com $\gamma = 0,95$ (coeficiente de confiança igual a 95%, quer dizer). Já fizemos as contas que mostram que $z_{\gamma=0,95} = 1,96$. Portanto,

$$\varepsilon = \frac{z_{\gamma=0,95}\sigma}{\sqrt{n}} = \frac{1,96 \cdot 0,1}{\sqrt{30}} = 0,035.$$

Portanto, o intervalo de confiança para a média tem a seguinte forma:

$$[1,45 - 0,035\,;\, 1,45 + 0,035]\,.$$

(b) Neste item queremos estimar a média populacional por intervalo de confiança com $\varepsilon = 0,05$ m e $\gamma = 99\%$. A pergunta é qual deve ser o tamanho de amostra para que a precisão e a confiança desejadas sejam garantidas. Já tivemos um problema parecido e sabemos que para sua solução é precisa achar z_γ e usar a fórmula $n = \left(\frac{z_\gamma \sigma}{\varepsilon}\right)^2$. Então:

$$n = \left(\frac{z_{\gamma=0,99}\sigma}{\varepsilon}\right)^2 = \left(\frac{2,58 \cdot 0,1}{0,05}\right)^2 = 26,6256 \approx 27.$$

Solução do Exc. 138(a).

Do enunciado, temos que $n = 100$, que o valor média da amostra (valor de $\bar{x}_{n=100}$) é 20, que a desvio padrão da amostra é 6 (isto é, $s = 6$), e que $\varepsilon = 1$ (pois 2ε corresponde ao tamanho do intervalo de confiança, e sendo que o tamanho foi de $21 - 19 = 2$ minutos, concluimos que $\varepsilon = 1$).

Para a solução, vamos assumir, em ambos os itens, que $n = 100$ é grande o suficiente para que a distribuição t-Student seja aproximada pela distribuição de Normal Padrão com a precisão aceitável.

A fórmula que vincula z_γ a todos estes valores (menos o de $\bar{x}_{n=100}$) é:

$$z_\gamma = \frac{\varepsilon\sqrt{n}}{s}$$

Logo,

$$z_\gamma = \frac{1 \cdot \sqrt{100}}{6} = 1,66$$

o que dá, via a tabela da distribuição Normal Padrão, que

$$\gamma = 0,90308.$$

(b) Já que este item pergunta sobre o tamanho de amostra, entendemos que tal amostra não foi feita, e portanto, não tem como saber nem a média amostral nem o desvio padrão amostral (s). O desvio padrão populacional (σ) tambem não foi dado. Sem saber s ou σ, não há como usar nossas fórmulas. Portanto, a única maneira de achar alguma solução do problema, é aproveitar do resultado da amostragem feita com 100 pacientes para usar s como estimativa para o desvio pasdrão amostral. Então, assumiremos que $\sigma = 6$. Temos do enunciado que $\varepsilon = 1$ e que $\gamma = 95\%$. Para que γ seja $0,95$ é precisa que z for $1,96$ (pois, pela tabela da Distribuição Acumulada da Normal Padrão, $0,975 = I\!P(Z \leq 1,96)$, e, portanto, $0,95 = I\!P[-1,96 \leq Z \leq 1.96]$). Portanto,

$$n = \left(\frac{z_{\gamma=0,95} \cdot \sigma}{\varepsilon}\right)^2 = \left(\frac{1,96 \cdot 6}{1}\right)^2 \approx 138,30 \to 139.$$

Solução do Exc. 139 A media amostral é 258,5, é com o intuito de sigerir a troca de caixas é precisa ter uma confiança de que a média real (populacional, quer dizer) não esteja igual ou menor que a das caixas que já estão em uso. Como a média populacional dos caixas antigas é 280, então tomando e=280-258,5, precisamos calcular a confiança de que a média populacionl de caixas novas esteja no intervalo [258,5-e, 258,5+e].Se tal confiança for grande, conclui-se-á que vale a pena substituir as caixas. Portanto, a "comparação"das médias sobre a qual você está falando, torna-se numa estimativa intervalar.

Solução do Exc. 140 a fazer.

Solução do Exc. 141 a fazer.