ACH3584 - Estatística II

Aula 06: Estimação

Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br

Organização

1. Estimação
   - Primeiras Ideias

2. Intervalos de confiança

Organização

1. Estimação
   - Primeiras Ideias

2. Intervalos de confiança

Organização

1. Estimação
   - Primeiras Ideias
2. Intervalos de confiança

Primeiras Ideias

- Inferência tem por objetivo fazer generalizações sobre uma população com base nos dados de uma amostra
- Dois problemas básicos desse processo:

  a) estimação e parâmetros; e
  b) testes de hipóteses sobre parâmetros

  Obs. *Parâmetros* são funções de valores populacionais, enquanto *estatísticas* são funções de valores amostrais
- Vamos discutir ideias básicas sobre estimação.

Exemplo

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

Solução Se 300 pessoas responderam **Sim**, então uma estimativa natural para essa proporção seria:

$$\frac{300}{500} = 0,6 = 60\%$$

Observações

- Nossa resposta é baseada na suposição de que a amostra é *representativa* da população;
- Sabemos que outra amostra poderia resultar em outra estimativa;
- Conhecer as propriedades desses *estimadores* é um dos propósitos mais importantes da Inferência Estatística.

Exemplo

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

Solução Se 300 pessoas responderam **Sim**, então uma estimativa natural para essa proporção seria:

$$\frac{300}{500} = 0,6 = 60\%$$

Observações

- Nossa resposta é baseada na suposição de que a amostra é *representativa* da população;
- Sabemos que outra amostra poderia resultar em outra estimativa;
- Conhecer as propriedades desses *estimadores* é um dos propósitos mais importantes da Inferência Estatística.

Exemplo

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

Solução Se 300 pessoas responderam **Sim**, então uma estimativa natural para essa proporção seria:

$$\frac{300}{500} = 0,6 = 60\%$$

Observações

- Nossa resposta é baseada na suposição de que a amostra é *representativa* da população;
- Sabemos que outra amostra poderia resultar em outra estimativa;
- Conhecer as propriedades desses *estimadores* é um dos propósitos mais importantes da Inferência Estatística.

Vamos definir as v.a.'s $X_1, X_2, \dots, X_n$ tais que:

$$X_i = \begin{cases} 1, & \text{se a } i\text{-ésima pessoa na amostra responde } \textbf{Sim} \\ 0, & \text{se a } i\text{-ésima pessoa na amostra responde } \textbf{Não} \end{cases}$$

Seja $p = \mathbf{P}(\text{Sucesso})$, em que *sucesso* significa resposta **Sim**.

$$Y_n = \sum_{i=1}^{n} X_i = \text{número de pessoas na amostra que responderam } \textbf{Sim}$$

$$Y_n \sim \text{Binomial}(n, p)$$

O problema consiste em encontrar p.

Um possível *estimador* de p é:

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\text{número de } \textbf{Sim's}}{\text{número de indivíduos}}$$

Se $Y_n = k$, uma estimativa para p é:

$$\hat{p} = \frac{k}{n}$$

Estudamos anteriormente a distribuição de $\hat{p}$:

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

Esse resultado nos ajuda avaliar as qualidades desse estimador. Ele diz que:

i) Em média $\hat{p}$ acerta p

- Dizemos que $\hat{p}$ é um estimador não viciado de p

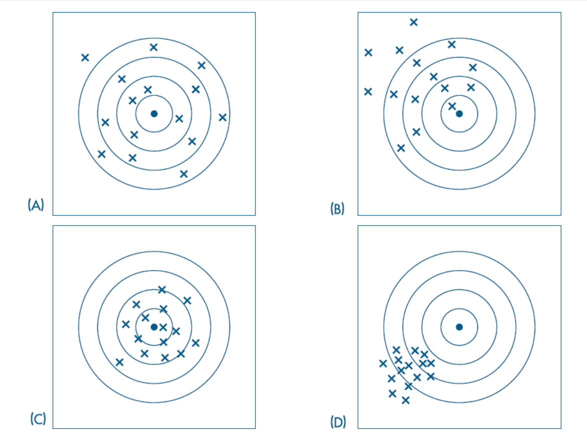
ii) Para amostras grandes, a diferença entre $\hat{p}$ e p tende a ser pequena pois:

$$n \to \infty \Rightarrow \mathrm{Var}(\hat{p}) \to 0$$

Em algumas situações, podemos ter mais de um estimador para um mesmo parâmetro.

Exemplo

Desejamos comprar um rifle e, após algumas seleções, restaram 4 alternativas: A, B, C e D. Foi feito um teste com cada rifle, que consistiu em fixá-lo num cavalete, mirar o centro e disparar 15 tiros. Resultado:



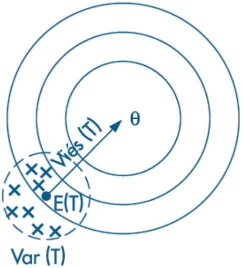
Avaliação

A: não viesada e baixa precisão.

B: viesada e baixa precisão.

C: não viesada e boa precisão.

D: viesada e alta precisão.



Organização

1. Estimação
   - Primeiras Ideias

2. Intervalos de confiança

Intervalos de confiança

- Até agora falamos de estimação pontual
- Queremos acoplar medidas de precisão a essas estimativas

Exemplo

- Suponha que queiramos estimar a média μ de uma população qualquer
- Para tanto usamos a media $\bar{X}$ de uma amostra de tamanho n
- Do TLC:

$$e = (\bar{X} - \mu) \sim N\left(0, \frac{\sigma^2}{n}\right)$$

Com a constatação anterior, podemos calcular probabilidade de ocorrência de erros:

$$\mathbf{P}\left(|\bar{X} - \mu| < 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Equivalente a:

$$\mathbf{P}\left(-1,96\frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95$$

e finalmente,

$$\mathbf{P}\left(\bar{X} - 1,96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95 \tag{2}$$

Interpretação:

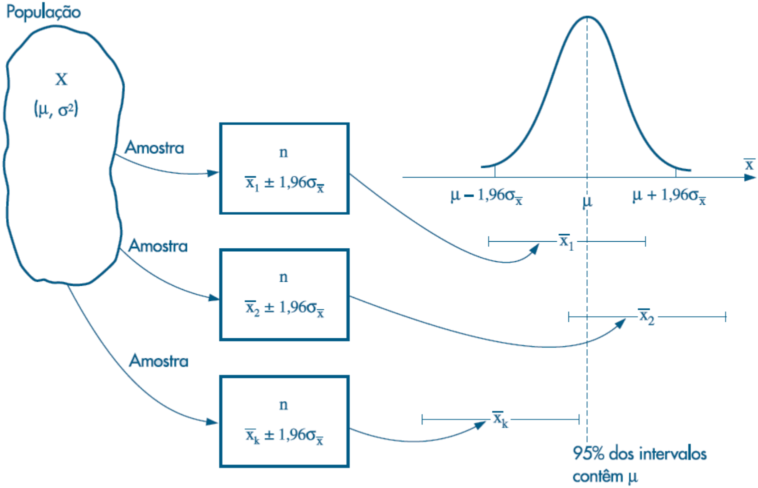
Se pudéssemos construir vários intervalos da forma:

$$\left[\bar{X} - 1,96\frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

baseados em amostras de tamanho *n*, 95% deles conteriam o parâmetro μ.

Obs. Dizemos que $\gamma = 0,95$ é o coeficiente de confiança.

Denotando $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, a figura abaixo ilustra o significado de um IC para μ com $\gamma = 0,95$.



Escolhida uma amostra e encontrada sua média $\bar{x}_0$ e admitindo-se σ conhecido, construímos o intervalo:

$$\left[\bar{x}_0 - 1,96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}_0 + 1,96\frac{\sigma}{\sqrt{n}}\right] \qquad (3)$$

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

Exemplo

Uma máquina enche pacotes de café com uma variância igual a 100 g^2. Ela estava regulada para encher os pacotes com 500 g, em média. Agora ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 25 pacotes apresentou média igual a 485 g. Vamos construir um intervalo de confiança com 95% de confiança para μ.

Solução O intervalo procurado tem a forma:

$$IC(\mu; 0,95) = \left[\bar{x}_0 - 1,96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}_0 + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

- $\bar{x}_0$ é o valor observado de $\bar{X}$. Logo, $\bar{x}_0 = 485$
- Temos $n = 25$ e $\sigma = 10$, logo:

Logo:

$$\begin{aligned} IC(\mu; 0,95) &= \left[485 - 1,96\frac{10}{\sqrt{25}}\ ,\ 485 + 1,96\frac{10}{\sqrt{25}}\right] \\ &= [485 - 1,96 \times 2\ ,\ 485 + 1,96 \times 2] \\ &= [481, 489] \end{aligned}$$

Exercício

De 50.000 válvulas fabricadas para uma determinada companhia petroleira, retira-se uma amostra de 400 válvulas, e observa-se na amostra tempo de duração médio de 700 horas. Sabe-se que o tempo de duração desse tipo de válvula tem desvio padrão de 100 horas.

a) Determine um intervalo de confiança de 95% de confiança para a média dos tempos de duração das válvulas.
b) O contrato com o fornecedor estabelece que as peças tenham duração mínima de 800 horas, em média. Que evidências o IC calculado no item a) traz sobre isso?