Estatística Exploratória

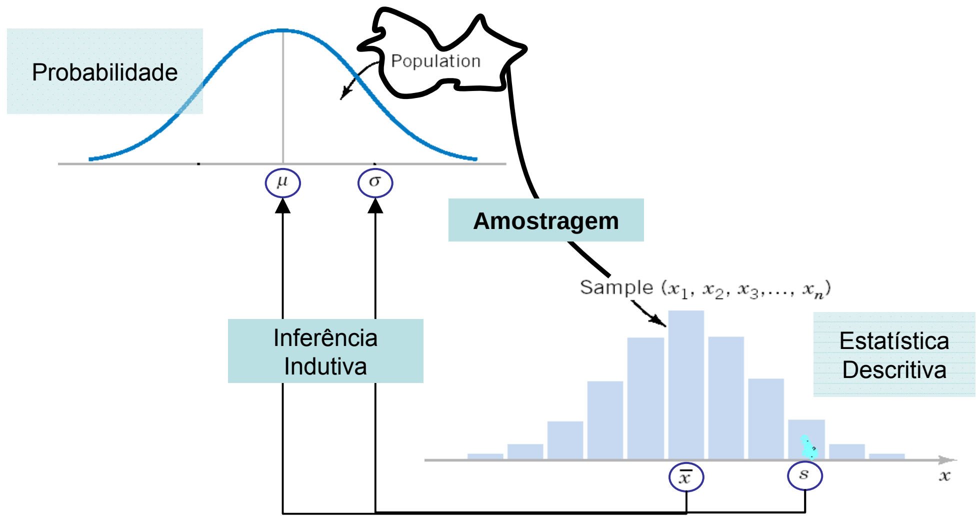
Prof Paulo Renato A. Firmino

praf62@gmailcom

Aulas 21-22

Intervalos de Confiança (IC)

- Temos estudado as nuances de se concluir sobre o todo a partir de amostras
  - Vimos a importância da representatividade das amostras, por exemplo
  - Temos visto como combinar probabilidade e estatística descritica para inferir
  - Aprendemos a recorrer a estimadores pontuais (média, moda, variância,...)
- Aprenderemos agora a recorrer a estimadores intervalares
  - Ao invés da valores (pontuais), teremos intervalos para μ, p, σ, λ, etc



IC para a Média
(Variância Conhecida)

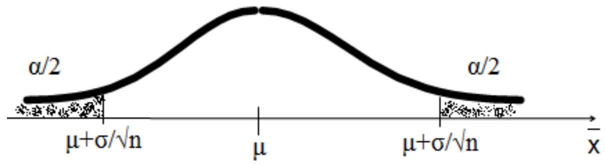
- Qual seria o intervalo [A, B] para o qual $P(A \leq \mu \leq B) = 1-\alpha$?
  - Sabemos que a média amostral é um bom estimador para μ
    - Como utilizar $\overline{X}$ e sua variabilidade para responder a esta questão?
    - Vimos que sob algumas suposições, quando n cresce

$$\overline{X} \sim N\left(\mu, \sigma^2 / n\right) \Rightarrow Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$E(X_i) = \mu$ e $V(X_i) = \sigma^2$

$$P\left(|Z| \leq z_{1-\alpha/2}\right) = 1 - \alpha = P\left(\frac{|\overline{X} - \mu|}{\sigma / \sqrt{n}} \leq z_{1-\alpha/2}\right) = P\left(|\overline{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) =$$

$$= P\left(\mu - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \leq \overline{X} \leq \mu + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) \Rightarrow$$

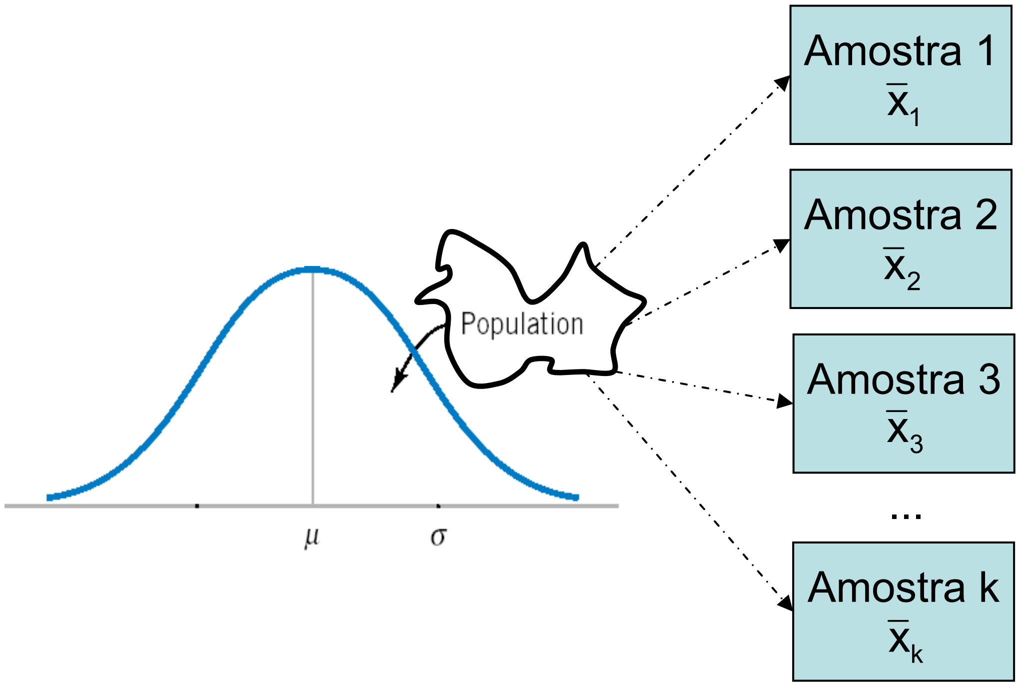


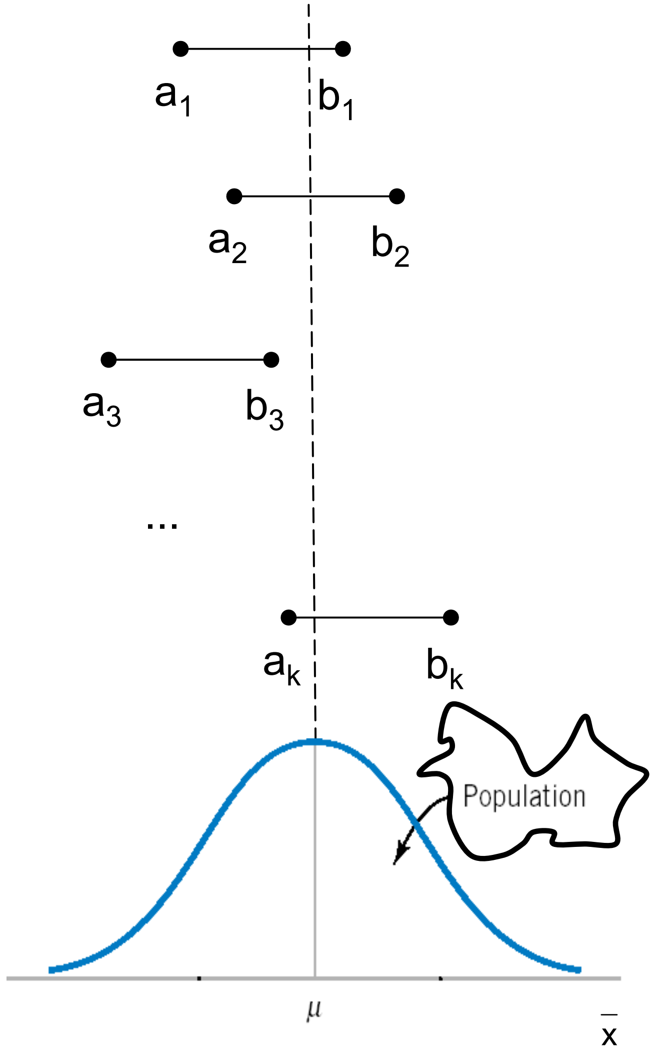
$$= P\left(\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) \Rightarrow [A, B] = \left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right]$$

Estimador Intervalar

IC para a Média (Variância Conhecida)

- Uma vez calculada a estimativa da média da iª amostra de tamanho n …





Temos 100(1-α)% de **confiança** de que μ pertenca ao iº intervalo

$$\left[\bar{x}_i - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}, \bar{x}_i + \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}\right]$$

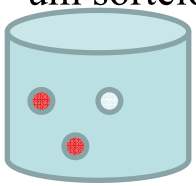
Estimativa Intervalar

IC para a Média (Variância Conhecida)

- Observações importantes
  - (1-α) é chamado de nível (ou coeficiente) de confiança do intervalo
  - Diz-se que $\left[\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}\right]$ é um intervalo de

    100(1-α)% de confiança para μ
  - Uma boa analogia para entender o significado de um IC ...
    - Ele pode envolver μ ou não
      - Assim como uma bola sorteada aleatoriamente de uma urna pode ser branca ou não
        - » Qual é a probabilidade de uma bola branca ser selecionada de uma urna que envolve 100 bolas (das quais 100(1-α) são brancas)?
      - Um IC está para uma bola assim como um IC que envolve μ está para uma bola branca
        - » Da analogia, **confiamos,** a um nível de 1-α, que a "bola selecionada é branca"
      - Confiança é, aqui, um sinônimo cauteloso para Probabilidade
        - » Probabilidade é uma função cujo argumento é um evento
        - » Daí, para que $P(a \leq \mu \leq b) = 1-\alpha$ tenha sentido devemos lembrar que μ, embora constante, é desconhecida, dando margem à incerteza

um sorteio



IC: Exercício 1

1. A variabilidade acerca do valor da cesta básica de uma região é dada por um desvio-padrão de R$50.00. Qual é a estimativa intervalar para o valor médio da cesta basica na região a um nível de confiança de 90%, dado que o custo médio da cesta de uma amostra com 30 cidades foi de R$250,00? E para um nível de confiança de 99%?
2. A variabilidade acerca do tempo de vida de lâmpadas é dada por um desvio-padrão de 1000 horas. Qual é a estimativa intervalar para o tempo médio de vida do modelo de lâmpadas sob um nível de confiança de 95%, dado que o tempo médio de vida de uma amostra com 100 lâmpadas é de 4000h? E para um nível de confiança de 99.5%?
3. A previsão para a temperatura média em dado dia é baseada em uma amostra envolvendo dias semelhantes. A variabilidade da temperatura média ao longo de dias semelhantes pode ser representada por um desvio-padrão de σ = 4ºC. De uma amostra envolvendo 64 dias semelhantes, obtém-se uma média de 28ºC. Qual é a estimativa intervalar para a temperatura média do dia sob estudo a um nível de confiança de: (a) 96%, (b) 99.9%?
4. Quais suposições embasam as análises acima?

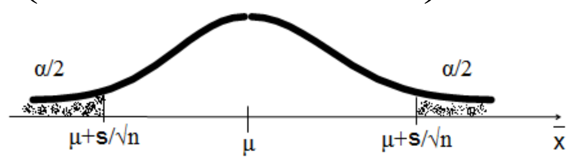
IC para a Média (Variância Desconhecida)

- Qual seria o IC para μ se σ fosse desconhecido?
  - Vimos que sob algumas suposições, quando n cresce
    - Utilizando o estimador da variância amostral (S^2) de σ^2

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \qquad E(X_i) = \mu \text{ e } V(X_i) = \sigma^2$$

$$P\left(|T| \le t_{1-\alpha/2}\right) = 1 - \alpha = P\left(\frac{|\overline{X} - \mu|}{s/\sqrt{n}} \le t_{1-\alpha/2}\right) = P\left(|\overline{X} - \mu| \le \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}\right) =$$

$$= P\left(\mu - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2} \le \overline{X} \le \mu + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}\right) \Rightarrow$$



$$= P\left(\overline{X} - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2} \le \mu \le \overline{X} + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}\right) \Rightarrow IC = \left[\overline{X} - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}, \overline{X} + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}\right]$$

Estimador Intervalar

IC: Exercício 2

1. De uma amostra com 30 cidades, estima-se que o desvio-padrão e a média do valor da cesta básica de uma região é de R$50.00 e R$250,00, respectivamente. Qual é a estimativa intervalar para o valor médio da cesta basica na região a um nível de confiança de 90%? E para um nível de confiança de 99%? E se o desvio-padrão, ao invés de uma estimativa, fosse populacional?
2. De uma amostra com 100 lâmpadas, estima-se que o desvio-padrão e a média do tempo de vida de um modelo de lâmpadas é de 1000 horas e 4000h, respectivamente. Qual é a estimativa intervalar para o tempo médio de vida do modelo de lâmpadas sob um nível de confiança de 95%? E para um nível de confiança de 99.9%? E se o desvio-padrão, ao invés de uma estimativa, fosse populacional?
3. A previsão para a temperatura média em dado dia é baseada em uma amostra envolvendo dias semelhantes. De uma amostra envolvendo 64 dias semelhantes, obtém-se uma média de 28ºC e um desvio-padrão de 4ºC para a temperatura. Qual é a estimativa intervalar para a temperatura média do dia sob estudo a um nível de confiança de: (a) 95%, (b) 99.9%?
4. Quais suposições embasam as análises acima?

IC para a Proporção

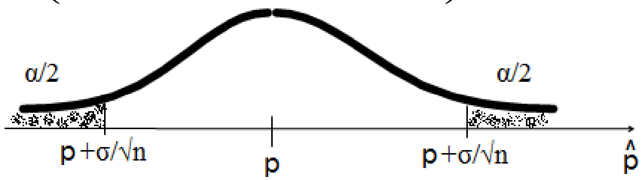
- Qual seria o intervalo [A, B] para o qual $P(A \leq p \leq B) = 1\text{-}\alpha$?
  - Sabemos que a proporção amostral $\hat{p}$ é um bom estimador para p
    - Como utilizar $\hat{p}$ e sua variabilidade para responder a esta questão?
    - Vimos que sob algumas suposições, quando n cresce

$$\hat{p} \sim N\left(p, \sigma^2/n\right) \Rightarrow Z = \frac{\hat{p} - p}{\sigma/\sqrt{n}} \sim N(0,1)$$

$p=E(X_i)$ e $\sigma^2=V(X_i)=p(1-p)$

$$P\left(|Z| \leq z_{1-\alpha/2}\right) = 1-\alpha = P\left(\frac{|\hat{p}-p|}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) = P\left(|\hat{p}-p| \leq \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) =$$

$$= P\left(p - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \leq \hat{p} \leq p + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) \Rightarrow$$



$$= P\left(\hat{p} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \leq p \leq \hat{p} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right) \Rightarrow [A, B] = \left[\hat{p} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, \hat{p} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}\right]$$

Estimador Intervalar

Supõe-se que σ é o maior possível (σ=0.5)

IC: Exercício 3

1. De uma amostra com 3000 eleitores de um país, a proporção de votantes em B foi de 46%. Qual é a estimativa intervalar para a proporção de votantes em B no país a um nível de confiança de 94%? E para um nível de confiança de 98%?
2. De uma amostra com 100 lâmpadas, 6 falharam antes de completar 8000 horas de operação. Qual é a estimativa intervalar para a probabilidade de uma lâmpada falhar antes de completar 8000 horas de funcionamento, sob um nível de confiança de 96%? E para um nível de confiança de 97%?
3. Deseja-se calcular uma estimativa intervalar para a probabilidade de que uma especulação financeira configure-se concretamente. De um banco de dados históricos envolvendo 64 especulações, ele verificou que 28 se tornaram verdade. Qual nível de confiança você adotaria? Por que? Qual é a estimativa intervalar associada?
4. Quais suposições embasam as análises acima?