MAT02025 - Amostragem 1

O erro quadrático médio

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2022

O erro quadrático médio

O erro quadrático médio

- Para comparar um estimador enviesado com um estimador imparcial, ou dois estimadores com diferentes valores de enviesamento, um critério útil é o **erro quadrático médio (EQM)** da estimativa, medido a partir do valor da população que está sendo estimado.
- Formalmente

$$
\begin{aligned}
EQM(\hat{\theta}) &= \mathrm{E}\,(\hat{\theta}-\theta)^2\\
&= \mathrm{E}\,\{[\hat{\theta}-\mathrm{E}\,(\hat{\theta})]+[\mathrm{E}\,(\hat{\theta})-\theta]\}^2\\
&= \mathrm{E}\,\{[\hat{\theta}-\mathrm{E}\,(\hat{\theta})]^2+2[\hat{\theta}-\mathrm{E}\,(\hat{\theta})][\mathrm{E}\,(\hat{\theta})-\theta]+[\mathrm{E}\,(\hat{\theta})-\theta]^2\}\\
&= \mathrm{E}\,\{[\hat{\theta}-\mathrm{E}\,(\hat{\theta})]^2\}+2[\mathrm{E}\,(\hat{\theta})-\theta]\mathrm{E}\,[\hat{\theta}-\mathrm{E}\,(\hat{\theta})]+\mathrm{E}\,[\mathrm{E}\,(\hat{\theta})-\theta]^2\\
&= \mathrm{Var}\,(\hat{\theta})+2[\mathrm{E}\,(\hat{\theta})-\theta][\mathrm{E}\,(\hat{\theta})-\mathrm{E}\,(\hat{\theta})]+[B(\hat{\theta})]^2\\
&= \mathrm{Var}\,(\hat{\theta})+[B(\hat{\theta})]^2.
\end{aligned}
$$

O erro quadrático médio

- Note que se um estimador $\hat{\theta}$ é não enviesado para θ, então

$$EQM(\hat{\theta}) = \text{Var}\,(\hat{\theta}) + [0]^2 = \text{Var}\,(\hat{\theta}) = \sigma^2_{\hat{\theta}}.$$

- No exemplo da aula passada, considerando o parâmetro populacional de interesse como a média, μ, temos $EQM(\hat{\mu}) = \sigma^2_{\hat{\mu}} + B^2$, em que $B = m - \mu$.

O erro quadrático médio

- O uso do EQM como critério de precisão de um estimador equivale a considerar equivalentes duas estimativas que têm o mesmo erro quadrático médio.
- Isso não é inteiramente verdadeiro, porque a distribuição de frequência de erros $(\hat{\mu} - \mu)$ não será a mesma para dois estimadores, caso eles apresentem viéses de valores diferentes.
- Entretanto, **Hansen, Hurwitz e Madow (1953)**[1] mostraram que se B/σ for menor que cerca de $0,5$, as duas distribuições de frequência são quase idênticas em relação aos erros **absolutos** $|\hat{\mu} - \mu|$ de tamanhos diferentes.

[1] Hansen, M. H., Hurwitz, W. N. e Madow, W. G. (1953) **Sample Survey methods and theory**, John Wiley & Sons, Nova York, Vol. I, pg. 58.

O erro quadrático médio

- Mais uma vez suponha que $\hat{\mu}$ tem uma distribuição aproximadamente normal com média $m = \mathrm{E}(\hat{\mu})$ e desvio padrão $\sigma = \sigma_{\hat{\mu}}$. Ainda, denote $EQM = EQM(\hat{\mu}) = \sigma^2 + B^2$.
- Então

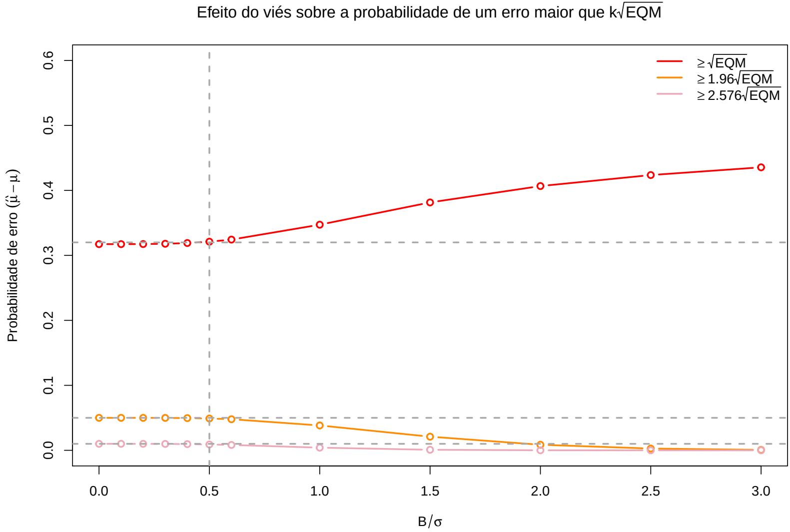
$$
\begin{aligned}
\Pr\left(|\hat{\mu} - \mu| \geq k\sqrt{EQM}\right) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu + k\sqrt{EQM}}^{\infty} e^{-(\hat{\mu}-m)^2/2\sigma^2} d\hat{\mu} \\
&\quad + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu - k\sqrt{EQM}} e^{-(\hat{\mu}-m)^2/2\sigma^2} d\hat{\mu}.
\end{aligned}
$$

O erro quadrático médio

Table 1: Proporção de casos em que o valor verdadeiro, μ, não está incluído no intervalo $\hat{\mu} \pm k\sqrt{EQM}$, para diferentes níveis de viés em $\hat{\mu}$

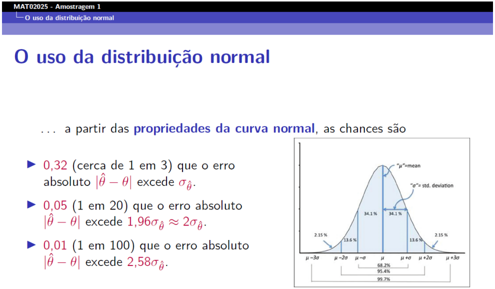
B/σ	$\geq \sqrt{EQM}$	$\geq 1,96\sqrt{EQM}$	$\geq 2,576\sqrt{EQM}$
0.0	0.32	0.050	0.0100
0.1	0.32	0.050	0.0100
0.2	0.32	0.050	0.0100
0.3	0.32	0.050	0.0098
0.4	0.32	0.050	0.0095
0.5	0.32	0.049	0.0090
0.6	0.32	0.048	0.0083
1.0	0.35	0.038	0.0041
1.5	0.38	0.021	0.0008
2.0	0.41	0.009	0.0001
2.5	0.42	0.003	-
3.0	0.44	0.001	-

O erro quadrático médio



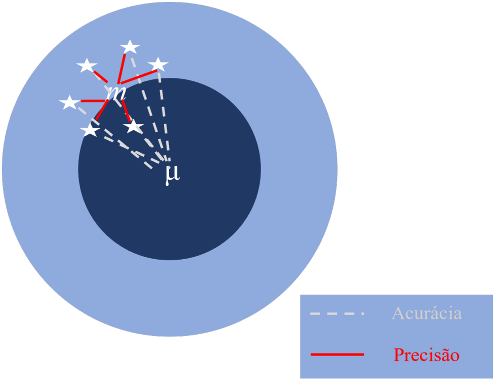
Comentários

- Esses resultados, para muitos propósitos práticos, concordam com as interpretações baseadas nos múltiplos correspondentes do desvio padrão quando uma estimativa não enviesada é usada.
  - Ou seja, quando $\sqrt{EQM} = \sqrt{\sigma^2 + B^2} = \sqrt{\sigma^2 + 0} = \sigma$.
- Da aula passada, temos

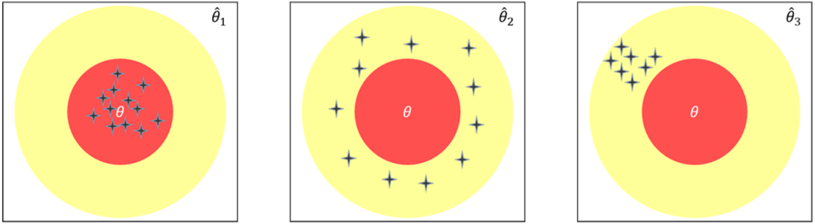


Comentários

- Devido à dificuldade de garantir que nenhum viés insuspeitado entre nas estimativas, geralmente falaremos da **precisão** de uma estimativa em vez de sua **acurácia (exatidão)**.
- A **acurácia** se refere ao tamanho dos desvios da verdadeira média (populacional) μ, enquanto a **precisão** se refere ao tamanho (médio) dos desvios da média m obtida pela aplicação repetida do procedimento de amostragem.



Comentários



- **Pergunta:** o que podemos concluir sobre $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\theta}_3$?

Levantamentos por amostragem: mais alguns conceitos

Relembrando

- **"Levantamento por amostragem"** é a tradução ao português da expressão em língua inglesa *survey sampling* ou *sample surveys*.
- Com a finalidade de produzir instantâneos das realidades estudadas, os levantamentos por amostragem reúnem as seguintes características:
  1. Aplicam-se a conjuntos reais e finitos (na maioria dos casos), compostos de elementos, denominados **população de estudo**.
  2. Os elementos podem ser seres humanos, animais, árvores, fichas, prontuários, domicílios, áreas ou objetos.
  3. As características ou atributos são observados em cada elemento e, posteriormente, agregados por meio de medidas estatísticas chamadas **parâmetros** ou **valores populacionais**.
  4. Os dados são coletados em amostras das populações de estudo, e as medidas calculadas (estimativas) passam a ser a informação disponível para os valores populacionais desconhecidos.

Unidade elementar, amostral e resposta

- A **unidade elementar**, ou simplesmente **elemento** de uma população é o objeto ou entidade portadora das informações que pretende-se coletar.
  - Pode ser uma pessoa, família, domicílio, loja, empresa, estabelecimento, classe de alunos, escola, etc.
- Qualquer **plano amostral**[2] (ou **plano de amostragem**) fará recomendações para selecionar elementos da população por meio das **unidades amostrais**.
  - Pode ser formada por uma única unidade elementar ou por várias.

[2] A cada amostra possível, S_j, é atribuído uma probabilidade conhecida de seleção, π_j. A função π_j é também conhecida de **plano amostral**.

Unidade elementar, amostral e resposta

Unidade amostral

Uma pesquisa eleitoral usa eleitores como sendo a unidade elementar.

1. Um levantamento pode escolher um ponto da cidade e entrevistar os cem primeiros **eleitores** que passam por lá. Usou-se a **unidade elementar** como **unidade amostral**.
2. Um plano alternativo decidiu selecionar **domicílios** e entrevistar todos os eleitores residentes nos domicílios escolhidos. A **unidade elementar** continua sendo *eleitor* mas agora a **unidade amostral** passou a ser *domicílio*, um **conjunto de unidades elementares**.

Unidade elementar, amostral e resposta

- Às vezes é conveniente ressaltar quem é a **unidade respondente** ou a **unidade de resposta**.
  - O censo demográfico tem uma primeira parte com questõees simples sobre cada morador do domicílio, tais como sexo, idade, grau de instrução, etc. Um ´unico morador pode responder por todos os outros; este morador é considerado a **unidade de resposta**.

Exercícios

Exercícios

- Exercício ConVid (Aula 03)
- Lista de Exercícios 01 (Aula 04)
- Para casa (Aula 05)
- Exercício 1.11 (Bolfarine e Bussab)
- Ao estimar por amostragem a idade média dos alunos do Bacharelado em Estatística da UFRGS, um questionário foi empregado e a idade (em anos) de cada aluno na amostra foi registrada. Para a amostra, a idade média foi 24,89, sendo esta estimativa aproximadamente normalmente distribuída com um erro padrão de 3,51. Calcule limites de confiança de 95% para a idade média populacional. Interprete este resultado.

Próxima aula

- **Atividade de avaliação I.**
  - Dia: 14/07
  - Horário: 8hs 30min
  - Local: Sala de aula
  - Modalidade: presencial e individual
  - **Conteúdo:** Área I (aulas 01-07)
  - **Não esqueça de trazer:** Lápis, Borracha, Caneta e Calculadora

Por hoje é só!

Bons estudos!