Universidade Federal de Lavras

Departamento de Ciências Exatas

Prof. Daniel Furtado Ferreira

$9^{\underline{a}}$ Lista de Exercícios — Teoria da Estimação

1) O que são estimadores, estimativas e parâmetros? Dê exemplos.

2) Discuta as propriedades de um estimador e dê exemplos.

3) Uma amostra do teor de Zn em ppm no dedo médio de $n = 100$ aves de postura apresentou média de $\bar{X} = 400$ e variância $S^2 = 420$. Esta amostra foi obtida em um plantel em que não foi realizada nenhum tratamento especial com fonte de Zn na ração. Obtenha o intervalo de 95% de confiança para a média populacional μ.

   Dado $t_{0,025;\nu=99} = 1{,}984$

4) Após a aplicação de Zn na ração, outra amostra de tamanho $n = 100$, apresentou média $\bar{X} = 700$ ppm e variância de $S^2 = 630$ ppm^2. Obtenha o intervalo de 95% de confiança para a média populacional μ após a dieta com Zn na ração. Compare as médias populacionais antes e após os tratamentos a base de Zn com base nos intervalos de confiança e tire suas conclusões.

5) Obtenha o intervalo de confiança para média de uma população normal μ a partir da afirmativa probabilística dada por:

$$P\left(-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

6) O erro de estimação e, também conhecido como semi-amplitude do intervalo de confiança (IC) dado por

$$e = t_{\alpha/2}\frac{S}{\sqrt{n}}$$

   pode ser manipulado, aumentando ou diminuindo o comprimento do intervalo de confiança, de que forma?

7) Desejando saber a incidência de uma doença em uma região foram amostrados $n = 600$ animais. Destes $y = 200$ estavam contaminados por esta doença. Determine o intervalo de 95% de confiança para a verdadeira proporção de doença da região, utilizando a aproximação normal. Tire todas as conclusões de interesse.

Resolução

1) Estimadores são variáveis aleatórias que são funções dos elementos amostrais. A média amostral é um exemplo de um estimador e é dada por

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Estimativas são os valores dos estimadores em uma amostra específica realizada (ex. $\bar{X} = 10$, $S^2 = 9$, etc.).

Parâmetros são constantes inerentes as populações, como sua média μ e a sua variância σ^2, entre muitos outros exemplos.

2) As principais propriedades dos estimadores são:

- **Viés ou Vício**: um estimador é não viciado se a média das estimativas obtidas em todas as amostras possíveis de tamanho n extraídas de uma população for igual ao parâmetro que está sendo estimado. Essa média é conhecida por *esperança matemática* do estimador. São exemplos de estimadores não-viciados a média $\bar{X}$ e a variância S^2 amostrais. É um exemplo de estimador viciado o desvios padrão amostral S. Isso porque a média de todas as possíveis estimativas de desvios padrões em amostras de tamanho n resulta em um valor diferente do parâmetro σ, o que não ocorre para a média e para a variância.
- **Eficiência**: um estimador é dito eficiente se entre todos os estimadores não viciados ele possuir a menor variância. Por exemplo, em populações simétricas podemos utilizar a média ($\bar{X}$) ou a mediana (m_d) amostrais para estimar a média da população μ. A média e a mediana neste caso de populações simétricas são não viesadas e as variâncias de cada uma são dadas por

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \qquad \text{e} \qquad \sigma^2_{m_d} \cong \frac{\sigma^2}{n} \times \frac{\pi}{2}.$$

  Assim, a média é um estimador mais eficiente do que a mediana, pois possui menor variância.
- **Consistência**: um estimador é consistente se além de não-viciado, possuir a propriedade de ter sua variância tendendo para 0 quando o tamanho da amostra n tende para ∞. Assim, $\bar{X}$, por exemplo, é um estimador consistente, pois é não viciado e

$$\lim_{n\to\infty} \sigma^2_{\bar{X}} = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

3) O intervalo de confiança para μ é dado por:

$$\begin{aligned} IC_{1-\alpha}(\mu) :& \bar{X} \pm t_{\alpha/2;\nu} \frac{S}{\sqrt{n}} \\ &= 400 \pm 1{,}984 \times \sqrt{\frac{420}{100}} \\ &= 400 \pm 4{,}07 \\ &= [395{,}93; 404{,}07]. \end{aligned}$$

Logo, podemos afirmar, com 95% de confiança, que a verdadeira média μ do teor de zinco na constituição *química* das aves que não receberam dieta a base de zinco (suplemento mineral na ração) deve ser um valor entre 395,93 ppm e 404,07 ppm.

4) O intervalo de confiança para μ é dado por:

$$\begin{aligned} IC_{1-\alpha}(\mu) :& \bar{X} \pm t_{\alpha/2;\nu} \frac{S}{\sqrt{n}} \\ &= 700 \pm 1{,}984 \times \sqrt{\frac{630}{100}} \\ &= 700 \pm 4{,}98 \\ &= [695{,}02; 704{,}98]. \end{aligned}$$

Logo, podemos afirmar, com 95% de confiança, que a verdadeira média μ do teor de zinco na constituição *química* das aves que receberam dieta a base de zinco (suplemento mineral na ração) deve ser um valor entre 695,02 ppm e 704,98 ppm.

Para compararmos as médias populacionais de antes (exercício anterior) e após (exercício atual) a dieta a base de zinco, podemos utilizar um critério que não é formal e nem é ótimo. Este critério baseia-se na sobreposição dos intervalos de confiança. Desta forma podemos contemplar os erros da estimação. Se os intervalos não se sobrepuserem, então consideraremos as médias populacionais diferentes e teremos uma confiança de 95% para afirmarmos isso. Por outro lado, se os intervalos se sobrepuserem, então consideraremos as médias populacionais não diferentes entre si, com esta mesma confiança nominal.

No presente exemplo, os intervalos não se sobrepõem e, portanto, concluímos que as aves que receberam a dieta com suplementação mineral possuem média de zinco maior na sua constituição química do que a daqueles que não receberam tal dieta. Este exemplo isso era um tanto quanto óbvio, pois as estimativas (400 e 700) são muito diferentes entre si. Mas realizar inferência, sem contemplar a margem de erro não é ciência.

Obs. Sugiro que este exemplo seja resolvido também, considerando que as duas amostras sejam independentes (como são de fato), estimando por intervalo a diferença entre médias $\mu_1 - \mu_2$, em que os índices 1 e 2, indicam população 1 (não tratada) e 2 (tratada), respectivamente.

5) Desenvolvendo a expressão temos:

$$\begin{aligned}
P\left(-t_{\alpha/2} \leq \frac{\bar{X}-\mu}{\dfrac{S}{\sqrt{n}}} \leq t_{\alpha/2}\right) &= 1-\alpha\\
P\left(-t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \bar{X}-\mu \leq t_{\alpha/2}\frac{S}{\sqrt{n}}\right) &= 1-\alpha\\
P\left(-\bar{X}-t_{\alpha/2}\frac{S}{\sqrt{n}} \leq -\mu \leq -\bar{X}+t_{\alpha/2}\frac{S}{\sqrt{n}}\right) &= 1-\alpha\\
P\left(\bar{X}+t_{\alpha/2}\frac{S}{\sqrt{n}} \geq \mu \geq \bar{X}-t_{\alpha/2}\frac{S}{\sqrt{n}}\right) &= 1-\alpha\\
P\left(\bar{X}-t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}+t_{\alpha/2}\frac{S}{\sqrt{n}}\right) &= 1-\alpha.
\end{aligned}$$

De onde obtemos o IC para μ por:

$$IC_{1-\alpha}(\mu) : \bar{X} \pm t_{\alpha/2;\nu}\frac{S}{\sqrt{n}}.$$

6) Basicamente podemos alterar o valor de e por meio de duas formas: a) aumentando o tamanho da amostra. Se fizermos isso, fixados as demais quantidades, iremos diminuir a semi-amplitude do intervalo de confiança, ou seja, o erro da estimação. O preço é o maior custo; b) alterando o valor de α, o que altera o coeficiente de confiança $1-\alpha$. Se diminuirmos α, o quantil superior $t_{\alpha/2}$ irá aumentar e o erro de estimação será maior.Desta forma,para termos uma confiança maior, fixadas as demais quantidades, pagamos o preço de lidarmos com intervalos de confiança mais amplos. Essa é razão pela qual não aumentamos a confiança para 100%, pois o intervalo é inútil. Veja o exemplo: afirmamos com 100% de confiança que a altura média dos alunos da UFLA está entre 0 e ∞ m. Essa afirmativa não nos ajuda em nada.

7) O intervalo de confiança aproximado para o parâmetro p pode ser obtido por:

$$\begin{aligned}
IC_{1-\alpha}(p) : \hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= \frac{y}{n} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\\
&= \frac{200}{600} \pm 1{,}96\sqrt{\frac{0{,}3333 \times 0{,}6667}{600}} = 0{,}3333 \pm 0{,}0377\\
&= [0{,}2956; 0{,}3710].
\end{aligned}$$

Assim, podemos afirmar com 95% que a verdadeira proporção de animais doentes na região deve estar entre 29,56% e 37,10%. O erro para menos e para mais foi de 3,77 pontos percentuais e a estimativa pontual foi de 33,33%.