Distribuições Amostrais - Tamanho da Amostra

Prof. Eduardo Bezerra

Inferência Estatística

21 de Setembro de 2018

Motivação

Suponha que queremos estimar o CR médio de estudantes de graduação de uma universidade. Historicamente, o desvio padrão do CR é conhecido: $\sigma = .30$. Se uma amostra aleatória de tamanho n = 25 produzir $\overline{X} = 3,05$, então a média populacional μ pode ser estimada como estando dentro do intervalo

$$\overline{X} \pm 2(\sigma/\sqrt{n}) = 3.05 \pm 2(.30/\sqrt{25}) = 3.05 \pm .12$$

com 95% de confiança.

Motivação (cont.)

A quantidade mais ou menos 0,12 é chamada margem de erro da média da amostra associada a um nível de confiança de 95%. Também é correto dizer "estamos 95% confiantes de que μ está dentro de 0,12 da média da amostra 3,05". No entanto, .12 é um grande número. O valor $3,05 \pm .12$ pode ser tão baixo quanto 2,93 ou tão alto quanto 3,17.
Existe alguma maneira de reduzir a margem de erro? Em particular, o que fazer se quisermos diminuir a margem de erro de .12 para .05?

Determinação do tamanho da amostra

- Até aqui, consideramos que o tamanho da amostra, n, é conhecido.
- Entretanto, podemos em certas ocasiões querer determinar o *tamanho mínimo da amostra* a ser colhida de uma população, de modo a obter um erro de estimação previamente estipulado, com determinado grau de confiança.

Determinação do tamanho da amostra (cont.)

Por exemplo, considere que estejamos estimando a média populacional μ e para tanto usaremos a média amostral, $\overline{X}$, baseada em uma amostra de tamanho n. Suponha que se queira determinar o valor de n de modo que

$$\Pr(\left|\overline{X} - \mu\right| \leq \epsilon) \geq \gamma$$

onde:

- γ é o grau de confiança $(0 < \gamma < 1)$;
- ϵ é o erro amostral máximo (ou erro de estimação) que podemos suportar.

Ambos os valores, γ e ϵ, são fixados durante o planejamento da pesquisa estatística.

Determinação do tamanho da amostra (cont.)

Sabemos que $\overline{X} \sim N(\mu, \sigma^2/n)$, logo $\overline{X} - \mu \sim N(0, \sigma^2/n)$ e portanto podemos reescrever a equação anterior como

$$\Pr(-\epsilon \leq \overline{X} - \mu \leq \epsilon) = \Pr\left(\frac{-\sqrt{n}\epsilon}{\sigma} \leq Z \leq \frac{\sqrt{n}\epsilon}{\sigma}\right) \geq \gamma$$

com $Z = (\overline{X} - \mu)\sqrt{n}/\sigma$. Dado γ, podemos obter z_γ da $N(0, 1)$, tal que $\Pr(-z_\gamma < Z < z_\gamma) = \gamma$, de modo que

$$\frac{\sqrt{n}\epsilon}{\sigma} = z_\gamma$$

do que obtemos

$$n = \frac{\sigma^2 z_\gamma^2}{\epsilon^2}$$

Determinação do tamanho da amostra (cont.)

Portanto, o tamanho mínimo de uma amostra pode ser obtido por

$$n = \frac{\sigma^2 z_\gamma^2}{\epsilon^2}$$

Note que nessa expressão derivada para n, são conhecidos z_γ e ϵ, mas σ^2 é a variância populacional desconhecida. Portanto, para calcularmos n,

- devemos ter alguma informação prévia sobre σ^2 (obtida a partir de estudos anteriores ou pela experiência do projetista do estudo estatístico),
- ou então usar uma pequena amostra piloto para estimar σ^2.

Determinação do tamanho da amostra - exemplo 01

Suponha que uma pequena amostra piloto de $n = 10$, extraída de uma população, forneceu valores $\overline{X} = 15$ e $S^2 = 16$. Com valores fixos de $\epsilon = 0,5$ e $\gamma = 0,95$, temos

$$n = \frac{16 \times (1,96)^2}{(0,5)^2} = 245.$$

Na expressão acima, 1,96 é o valor na normal padrão necessário para que se obtenha 95% da área sob a curva.

Determinação do tamanho da amostra

Quando o parâmetro que se deseja estimar é p, a proporção populacional, podemos usar a aproximação normal para $\widehat{p}$, o que resulta em

$$n = \frac{z_\gamma^2 p(1-p)}{\epsilon^2}$$

Quando não conhecemos o valor de p, podemos usar o fato de que

$$\forall p : \; p(1-p) \leq \frac{1}{4},$$

de modo que

$$n \approx \frac{z_\gamma^2}{4\epsilon^2}$$

Determinação do tamanho da amostra - exemplo 02

Em uma pesquisa de mercado, estima-se que aproximadamente 60% das pessoas entrevistadas preferirão a marca A de um produto. Essa informação é baseada em pesquisas anteriores. Considere que, no planejamento da pesquisa, definiu-se que o erro amostral máximo de $\widehat{p}$ deve ser $0,03$ (i.e., $\epsilon = 0,03$), com probabilidade $\gamma = 0,95$. Com os valores fornecidos para ϵ e γ, temos

$$n = \frac{(1,96)^2(0,6)(0,4)}{(0,03)^2} = 1.024.$$

em que usamos o fato de que $p \approx 0,60$.