

ESTIMAÇÃO PARA A PROPORÇÃO POPULACIONAL p

Objetivo

Estimar uma proporção p (desconhecida) de elementos em uma população, apresentando certa característica de interesse, a partir da informação fornecida por uma amostra.

Exemplos:

p : proporção de alunos da USP que foram ao teatro pelo menos uma vez no último mês;

p : proporção de consumidores satisfeitos com os serviços prestados por uma empresa telefônica;

p : proporção de eleitores da cidade de São Paulo que votariam em um determinado candidato, caso a eleição para presidente se realizasse hoje;

p : proporção de crianças de 2 a 6 anos, do estado de São Paulo, que não estão matriculadas em escola de educação infantil.

Dois possíveis procedimentos de estimação:

- **Estimação pontual**

- **Estimação intervalar**

- **Vamos observar n elementos, extraídos ao acaso e com reposição da população;**
- **Para cada elemento selecionado, verificamos a presença (sucesso) ou não (fracasso) da característica de interesse.**

Estimador pontual

O *estimador pontual para p* , também denominado *proporção amostral*, é definido como

$$\hat{p} = \frac{X}{n},$$

sendo que

X denota o número de elementos na amostra que apresentam a característica;

n denota o tamanho da amostra coletada.

Se observamos o valor k da v. a. X , obtemos $\hat{p} = k / n$ que denominamos *estimativa pontual para p* .

Exemplo 1: Sejam,

p : proporção de alunos da USP que foram ao teatro pelo menos uma vez no último mês, e

X : número de estudantes que respondem “sim” em uma pesquisa com n entrevistados.

Suponha que foram entrevistados $n = 500$ estudantes e que, desses, $k = 100$ teriam afirmado que foram ao teatro pelo menos uma vez no último mês.

A estimativa pontual (proporção amostral) para p é dada por:

$$\hat{p} = \frac{k}{n} = \frac{100}{500} = 0,20 ,$$

ou seja, 20% dos estudantes *entrevistados* afirmaram que foram ao teatro pelo menos uma vez no último mês.

Note que, outra amostra de mesmo tamanho pode levar a uma outra estimativa pontual para p .

Estimativa intervalar ou intervalo de confiança

- Para uma amostra observada, os estimadores pontuais fornecem como estimativa um único valor numérico para o parâmetro.
- Os estimadores pontuais são variáveis aleatórias e, portanto, possuem uma distribuição de probabilidade, em geral, denominada *distribuição amostral*.

Idéia: construir intervalos de confiança, que incorporem à estimativa pontual informações a respeito de sua variabilidade (erro amostral).

Intervalos de confiança são obtidos por meio da *distribuição amostral do estimador pontual*.

A **estimativa intervalar** corresponde a um intervalo determinado da seguinte maneira:

$$[\hat{p} - \varepsilon; \hat{p} + \varepsilon],$$

sendo ε o erro amostral ou margem de erro.

Pergunta: *Como encontrar ε ?*

Seja $P(\varepsilon)$ a probabilidade da estimativa pontual estar a uma distância de, no máximo, ε da proporção verdadeira p , ou seja,

$$P(\varepsilon) = P(|\hat{p} - p| \leq \varepsilon).$$

A probabilidade $P(\varepsilon)$ é também denominada **coeficiente de confiança do intervalo**, que denotamos pela letra grega γ (**gama**).

Afirma-se ainda que a estimativa intervalar tem **coeficiente de confiança $\gamma = P(\varepsilon)$** .

Formalmente,

$$\begin{aligned} P(\varepsilon) &= P(|\hat{p} - p| \leq \varepsilon) = P\left(\left|\frac{X}{n} - p\right| \leq \varepsilon\right) \\ &= P\left(p - \varepsilon \leq \frac{X}{n} \leq p + \varepsilon\right) = P(np - n\varepsilon \leq X \leq np + n\varepsilon) \\ &= P\left\{-\frac{n\varepsilon}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{n\varepsilon}{\sqrt{np(1-p)}}\right\}. \end{aligned}$$

Como $X \sim b(n,p)$ temos que, para n grande, a variável aleatória

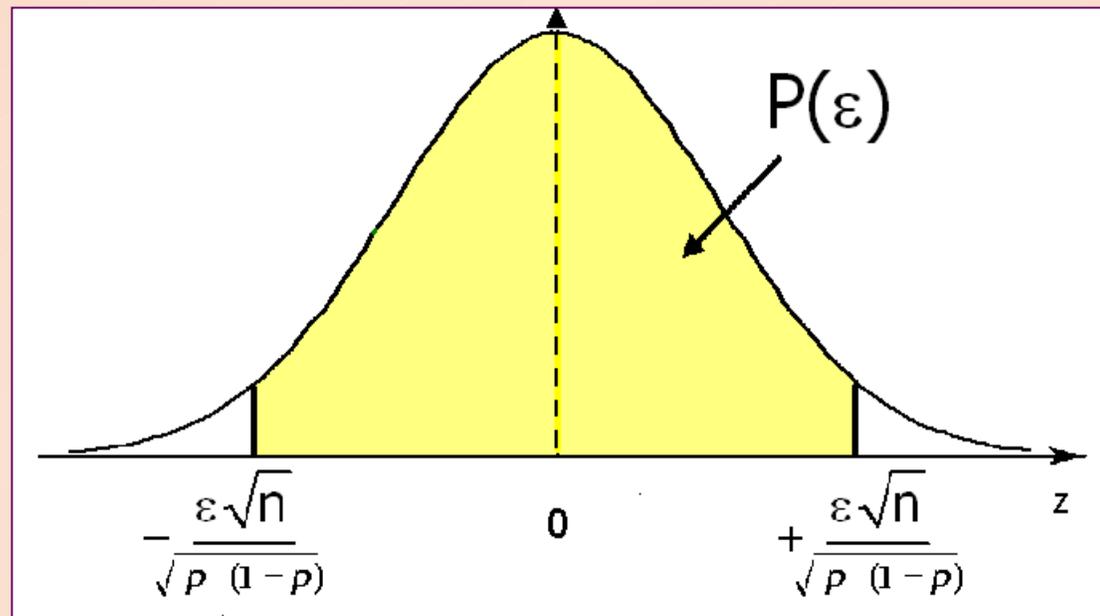
$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

tem distribuição $N(0,1)$.

Deste modo, para n grande,

$$P(\varepsilon) \cong P \left\{ -\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \leq \mathbf{Z} \leq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \right\},$$

onde $Z \sim N(0,1)$.

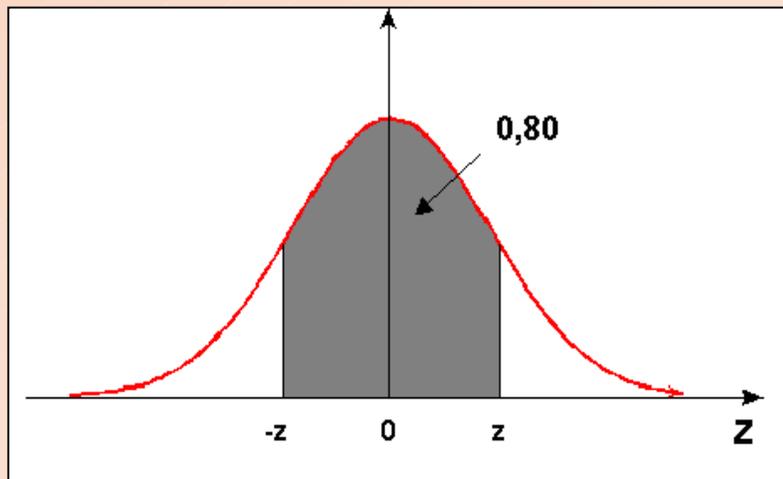


Denotando $\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} = z$, temos que

$$P(\varepsilon) = \gamma = P(-z \leq Z \leq z).$$

Assim, podemos obter z conhecendo-se γ (ou $P(\varepsilon)$).

Por exemplo, considere $\gamma = 0,80$.



z é tal que $A(z) = 0,90$.

Pela tabela, temos $z = 1,28$.

Erro da estimativa intervalar

Da igualdade $z = \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}$,

é imediato mostrar que o erro amostral ε é dado por

$$\varepsilon = z\sqrt{\frac{p(1-p)}{n}},$$

onde z é tal que $\gamma = P(-z \leq Z \leq z)$, com $Z \sim N(0,1)$.

Dimensionamento da amostra

Da relação $\varepsilon = z \sqrt{\frac{p(1-p)}{n}}$,

segue que o **tamanho amostral** n , dados γ e a margem de erro ε , tem a forma

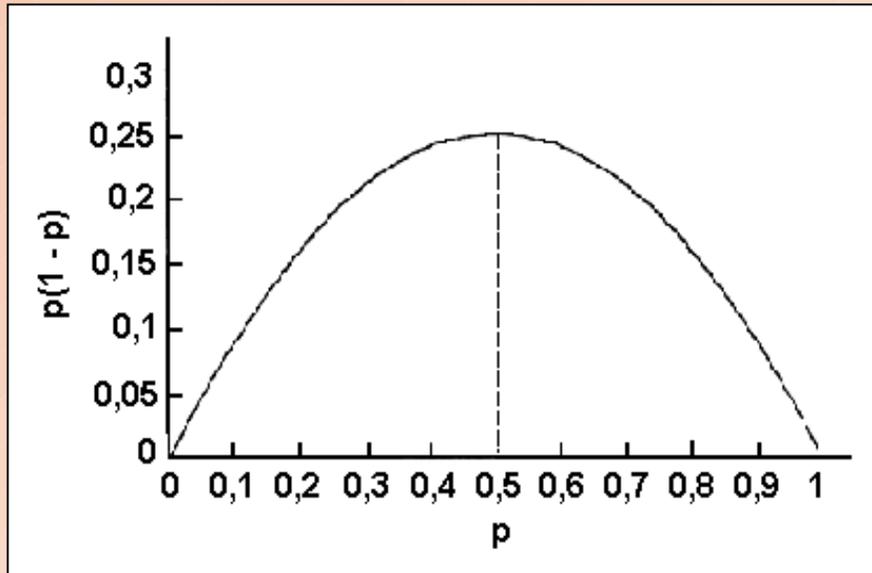
$$n = \left(\frac{z}{\varepsilon} \right)^2 p(1-p),$$

onde z é tal que $\gamma = P(-z \leq Z \leq z)$ e $Z \sim N(0,1)$.

Entretanto, nesta expressão, n depende de $p(1-p)$, que é desconhecido.

- **Como calcular o valor de n ?**

Gráfico da função $p(1-p)$, para $0 \leq p \leq 1$.



Pela figura observamos que:

- a função $p(1-p)$ é uma parábola simétrica em torno de $p = 0,5$;
- o máximo de $p(1-p)$ é 0,25, alcançado quando $p = 0,5$.

Assim, na prática, substituímos $p(1-p)$ por seu valor máximo, obtendo

$$n = \left(\frac{z}{\varepsilon} \right)^2 0,25 ,$$

que pode fornecer um valor de n maior do que o necessário.

Exemplo 2:

No exemplo da USP (Exemplo 1) suponha que nenhuma amostra foi coletada. Quantos estudantes precisamos consultar de modo que a estimativa pontual esteja, no máximo, a 0,02 da proporção verdadeira p , com uma probabilidade de 0,95?

Dados do problema:

$\varepsilon = 0,02$ (erro da estimativa);

$P(\varepsilon) = \gamma = 0,95 \Rightarrow z = 1,96.$

$$n = \left(\frac{1,96}{0,02} \right)^2 p(1 - p) \leq \left(\frac{1,96}{0,02} \right)^2 0,25 = 2401 \text{ estudantes .}$$

Pergunta: *É possível reduzir o tamanho da amostra quando temos alguma informação a respeito de p ?*

Por exemplo, sabemos que:

- p não é superior a 0,30, ou
- p é pelo menos 0,80, ou
- p está entre 0,30 e 0,60.

Resposta: *Depende do tipo de informação sobre p .*

Em alguns casos, podemos substituir a informação $p(1-p)$, que aparece na expressão de n , por um valor menor que 0,25.

Redução do tamanho da amostra

Vimos que, se nada sabemos sobre o valor de p , no cálculo de n , substituímos $p(1-p)$ por seu valor máximo, e calculamos

$$n = \left(\frac{z}{\varepsilon} \right)^2 0,25.$$

Se temos a informação de que p é *no máximo 0,30* ($p \leq 0,30$), então o valor máximo de $p(1-p)$ será dado por $0,3 \times 0,7 = 0,21$.

Logo, reduzimos o valor de n para

$$n = \left(\frac{z}{\varepsilon} \right)^2 0,21.$$

Agora, se p é pelo menos $0,80$ ($p \geq 0,80$), então o máximo de $p(1-p)$ é $0,8 \times 0,2 = 0,16$ e temos

$$n = \left(\frac{z}{\varepsilon} \right)^2 0,16 .$$

Mas, se $0,30 \leq p \leq 0,60$, o máximo de $p(1-p)$ é $0,5 \times 0,5 = 0,25$ e, neste caso, não há redução, ou seja,

$$n = \left(\frac{z}{\varepsilon} \right)^2 0,25 .$$

Exemplo 3:

No Exemplo 2, suponha que temos a informação de que no máximo 30% dos alunos da USP foram ao teatro no último mês. Portanto, temos que $p \leq 0,30$ e, como vimos, o máximo de $p(1-p)$ neste caso é 0,21.

Assim, precisamos amostrar

$$n = \left(\frac{z}{\varepsilon} \right)^2 0,21 = \left(\frac{1,96}{0,02} \right)^2 0,21 = 2017 \text{ estudantes ,}$$

conseguindo uma redução de $2401 - 2017 = 384$ estudantes.

Intervalo de confiança para p

Vimos que a estimativa intervalar para p tem a forma: $\left[\hat{p} - \varepsilon ; \hat{p} + \varepsilon \right]$,

com $\varepsilon = z \sqrt{\frac{p(1-p)}{n}}$ e z tal que $\gamma = P(-z \leq Z \leq z)$ na $N(0,1)$.

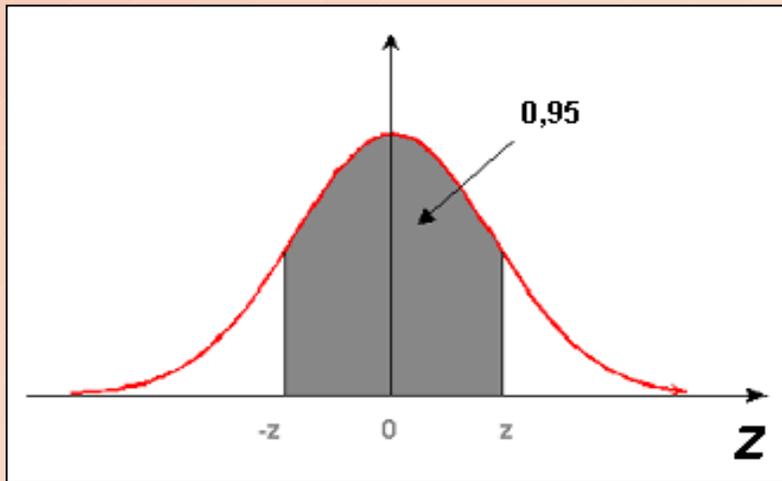
Na prática, substituímos a proporção desconhecida p pela proporção amostral \hat{p} , obtendo o seguinte intervalo de confiança com coeficiente de confiança γ :

$$\text{IC}(p; \gamma) = \left[\hat{p} - z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Exemplo 4:

No exemplo da USP, temos $n = 500$ e $\hat{p} = 0,20$.

Construir um intervalo de confiança para p com coeficiente de confiança $\gamma = 0,95$.



Como $\gamma = 0,95$ fornece $z = 1,96$, o intervalo é dado por:

$$\left[\hat{p} - z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

$$\begin{aligned} &= \left[0,20 - 1,96 \sqrt{\frac{0,20 \times 0,80}{500}} ; 0,20 + 1,96 \sqrt{\frac{0,20 \times 0,80}{500}} \right] \\ &= [0,20 - 0,035 ; 0,20 + 0,035] = [0,165 ; 0,235]. \end{aligned}$$

Nesse intervalo ($\gamma=0,95$), a estimativa pontual para p é 0,20, com um erro amostral ε igual a 0,035.

Interpretação do IC com $\gamma = 95\%$:

Se sortearmos 100 amostras de tamanho $n=500$ e construirmos os respectivos 100 intervalos de confiança, com coeficiente de confiança de 95%, esperamos que, aproximadamente, 95 destes intervalos contenham o verdadeiro valor de p .

Comentários:

Da expressão $\varepsilon = z \sqrt{\frac{p(1-p)}{n}}$, é possível concluir que:

- para γ fixado, o erro diminui com o aumento de n .
- para n fixado, o erro aumenta com o aumento de γ .

Exemplo 5:

Ainda no exemplo da USP, temos $k = 100$ e $n = 500$.

Qual é a probabilidade da estimativa pontual estar a uma distância de, no máximo, 0,03 da proporção verdadeira?

Dados do problema:

$$n=500, \hat{p}=0,20 \text{ e } \varepsilon=0,03 \quad \Rightarrow \quad P(\varepsilon) = \gamma = ??$$

Como a proporção verdadeira p é desconhecida, utilizamos a estimativa pontual \hat{p} para calcular z e, assim, obter γ (ou $P(\varepsilon)$).

Cálculo de z:

$$z = \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \cong \frac{0,03\sqrt{500}}{\sqrt{0,2 \times 0,8}} = 1,68 .$$

Logo, obtemos

$$\begin{aligned} P(\varepsilon) &\cong 2 A(z) - 1 \\ &= 2 A(1,68) - 1 \\ &= 2 \times 0,953 - 1 \\ &= 0,906 \quad (90,6\%). \end{aligned}$$

Exemplo 6: Suponha que estamos interessados em estimar a proporção p de pacientes com menos de 40 anos diagnosticados com câncer nos pulmões que sobrevivem pelo menos 5 anos.

Em uma amostra aleatoriamente selecionada de 52 pacientes, somente seis sobreviveram mais de 5 anos.

- Estimativa por ponto para p : $\hat{p} = \frac{6}{52} = 0,115$

- Intervalo de confiança aproximado de 95% para p :

$$\left(0,115 - 1,96 \sqrt{\frac{0,115(1 - 0,115)}{52}} ; 0,115 + 1,96 \sqrt{\frac{0,115(1 - 0,115)}{52}} \right)$$
$$= (0,028, 0,202)$$

Comentário:

Embora esse intervalo tenha sido construído usando a aproximação normal para a distribuição binomial, poderíamos ter gerado um intervalo de confiança *exato* para p usando a própria distribuição binomial.

Um intervalo exato é particularmente útil para pequenas amostras, em que o uso da aproximação normal não pode ser justificada.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.