

ME951 - Estatística e Probabilidade I

Parte 15

Notas de aula de ME414 produzidas pelos professores **Samara Kiihl**, **Tatiana Benaglia** e **Benilton Carvalho** modificadas e alteradas pela Profa. **Larissa Avila Matos**

Intervalo de Confiança

Introdução

Vimos que podemos utilizar uma estatística, como \bar{X} , para estimar um parâmetro populacional, como a média populacional μ .

Após coletarmos uma amostra aleatória calculamos \bar{x} , que é a nossa estimativa para μ . Chamamos esta estimativa de **estimativa pontual**.

Uma estimativa pontual fornece apenas um único valor plausível para o parâmetro. E sabemos que ela pode ser diferente para cada amostra obtida: distribuição amostral.

O ideal é que se reporte não só a estimativa, mas também a sua imprecisão.

Duas maneiras: fornecer a estimativa juntamente com o seu **erro-padrão** ou fornecer um intervalo de valores plausíveis para o parâmetro de interesse (**intervalo de confiança**).

Introdução

Suponha que queremos estimar o parâmetro populacional θ através de um intervalo.

Um intervalo de confiança (IC) para θ é sempre da forma:

$$\text{estimativa} \pm \text{margem de erro}$$

$$\hat{\theta} \pm \text{margem de erro}$$

Sendo:

- $\hat{\theta}$ uma estimativa pontual de θ ;
- **margem de erro:** quantidade que depende da distribuição amostral do estimador pontual de θ , do grau de confiança pré-estabelecido e do erro padrão da estimativa.

Intervalo de Confiança como Estimativa de p

Seja $X_i \sim \text{Bernoulli}(p)$, $i = 1, \dots, n$.

Se n suficientemente grande,

$$S_n = \sum_{i=1}^n X_i \approx N(np, np(1-p)).$$

Então, $Z = \frac{S_n - np}{\sqrt{np(1-p)}} \approx N(0, 1)$.

Seja $1 - \alpha$ o grau de confiança do intervalo.

Geralmente usamos $\alpha = 0.05$, então o grau de confiança é 95%.

Queremos encontrar um intervalo tal que a probabilidade do intervalo conter o verdadeiro valor de p seja $(1 - \alpha) \times 100\%$.

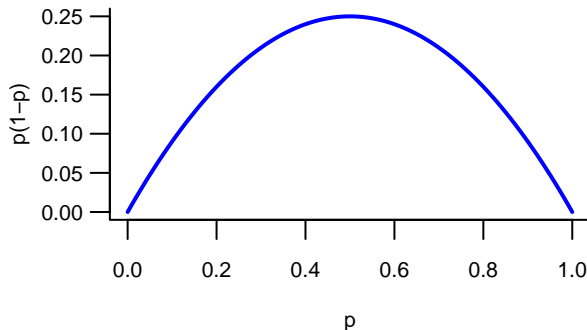
Intervalo de Confiança como Estimativa de p

$$\begin{aligned}0.95 &= P(-1.96 \leq Z \leq 1.96) \\&= P\left(-1.96 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq 1.96\right) \\&= P\left(-1.96\sqrt{np(1-p)} \leq S_n - np \leq 1.96\sqrt{np(1-p)}\right) \\&= P\left(\frac{-1.96\sqrt{np(1-p)}}{n} \leq \frac{S_n - np}{n} \leq \frac{1.96\sqrt{np(1-p)}}{n}\right) \\&= P\left(\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}}\right)\end{aligned}$$

$$\hat{p} = \frac{S_n}{n} = \bar{X}_n$$

Intervalo de Confiança como Estimativa de p

Note que p é desconhecido, mas a variância depende da função de $p(1 - p)$, dada no seguinte gráfico:



A função $p(1 - p)$ atinge o valor máximo quando $p = 1/2$, ou seja, $p(1 - p) \leq \frac{1}{4}$.

Intervalo de Confiança como Estimativa de p

Vimos que $p(1-p) \leq \frac{1}{4}$, então erro-padrão é maximizado por:

$$\sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{1}{4n}} \iff -\sqrt{\frac{p(1-p)}{n}} \geq -\sqrt{\frac{1}{4n}}$$

Portanto, $0.95 \leq P\left(\hat{p} - 1.96\sqrt{\frac{1}{4n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{1}{4n}}\right)$.

Caso geral (conservador): Um IC de $100(1-\alpha)\%$ para p é dado por

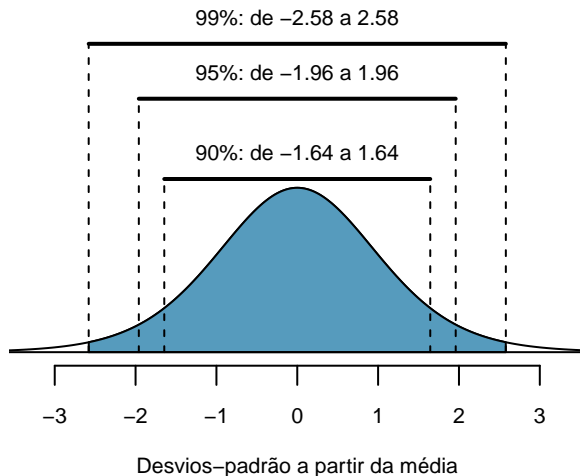
$$IC(p, 1-\alpha) = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{1}{4n}}\right]$$

Veja que tivemos que escolher as quantidades $z_{\alpha/2}$ tal que:

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

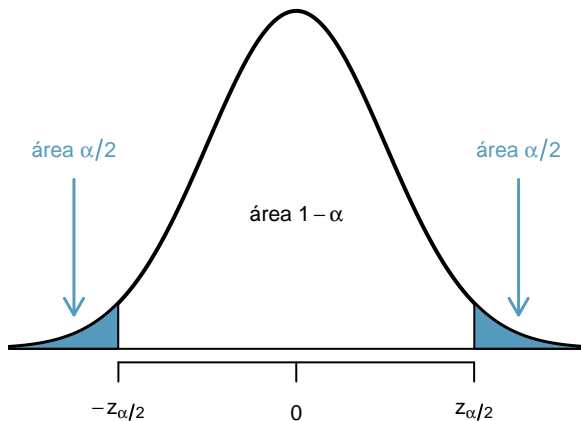
Como encontrar $z_{\alpha/2}$

$$P(|Z| \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



Como encontrar $z_{\alpha/2}$

$$P(|Z| \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



Como encontrar $z_{\alpha/2}$

Seja $Z \sim N(0, 1)$. O percentil $z_{\alpha/2}$ é tal que

$$1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2})$$

Como determinar $z_{\alpha/2}$?

$$\begin{aligned} 1 - \alpha &= P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = P(Z \leq z_{\alpha/2}) - P(Z \leq -z_{\alpha/2}) \\ &= P(Z \leq z_{\alpha/2}) - P(Z \geq z_{\alpha/2}) \\ &= P(Z \leq z_{\alpha/2}) - [1 - P(Z \leq z_{\alpha/2})] \\ &= 2P(Z \leq z_{\alpha/2}) - 1 \\ &= 2\Phi(z_{\alpha/2}) - 1 \end{aligned}$$

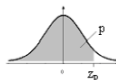
$$\text{Portanto, } 1 - \frac{\alpha}{2} = \Phi(z_{\alpha/2}) \quad \Rightarrow \quad \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = z_{\alpha/2}$$

Procure na tabela o valor de z tal que a probabilidade acumulada até o valor de z , isto é $P(Z \leq z) = \Phi(z)$, seja $1 - \alpha/2$.

Exemplo

Encontrar $z_{0.05}$ tal que $0.90 = P(-z_{0.05} \leq Z \leq z_{0.05})$.

Tabela I: Distribuição Normal Padrão Acumulada



Fornece $\Phi(z) = P(-\infty < Z \leq z)$, para todo z , de 0,01 em 0,01, desde $z = 0,00$ até $z = 3,59$
A distribuição de Z é $\text{Normal}(0,1)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Exemplo

Numa pesquisa de mercado, $n = 400$ pessoas foram entrevistadas (usando amostra aleatória) sobre preferência do produto da marca A, e 60% destas pessoas preferiam a marca A.

Encontre um IC de 95% para a proporção de pessoas que preferem a marca A.

Pelo resultado da pesquisa, $\hat{p} = 0.6$.

Logo, o IC com grau de confiança $1 - \alpha = 0.95$ é dado por:

$$\begin{aligned} IC(p, 0.95) &= \left[0.6 - 1.96 \frac{1}{\sqrt{1600}}; 0.6 + 1.96 \frac{1}{\sqrt{1600}} \right] \\ &= [0.551; 0.649] \end{aligned}$$

Exemplo

Suponha que em $n = 400$ entrevistados, tivéssemos obtido $k = 80$ respostas de pessoas que preferem a marca A.

Vamos obter um intervalo de confiança para p , com grau de confiança de 90%:

- $\hat{p} = \frac{80}{400} = 0.2$

- $1 - \alpha = 0.90$. Então $\alpha = 0.10$ e $\alpha/2 = 0.05$

- $z_{\alpha/2} = z_{0.05} = 1.64$

$$\begin{aligned} IC_1(p, 0.90) &= \left[0.2 - 1.64 \frac{1}{\sqrt{1600}}; 0.2 + 1.64 \frac{1}{\sqrt{1600}} \right] \\ &= [0.159; 0.241] \end{aligned}$$

Exemplo

E se usarmos a estimativa \hat{p} para também estimar o erro-padrão

$$\sqrt{\frac{p(1-p)}{n}}?$$

Podemos construir o seguinte IC de $100(1 - \alpha)\%$

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

■ No exemplo anterior,

$$\begin{aligned} IC_2(p, 0.90) &= \left[0.2 - 1.64 \sqrt{\frac{(0.2)(0.8)}{400}}; 0.2 + 1.64 \sqrt{\frac{(0.2)(0.8)}{400}} \right] \\ &= [0.167; 0.233] \end{aligned}$$

Exemplo

O intervalo que utiliza \hat{p} também para estimar o erro-padrão tem menor amplitude do que o intervalo que utiliza o fato de $p(1-p) \leq \frac{1}{4}$. Por isso esse último é chamado de **conservador**.

Veja as amplitudes dos IC 's que encontramos no exemplo anterior:

$$\blacksquare IC_1(p, 0.90) = [0.159; 0.241] \Rightarrow A_1 = 0.241 - 0.159 = 0.082$$

$$\blacksquare IC_2(p, 0.90) = [0.167; 0.233] \Rightarrow A_2 = 0.233 - 0.167 = 0.066$$

Intervalo de Confiança para p

Em resumo, os intervalos de $100(1 - \alpha)\%$ de confiança para p podem então ser de duas formas:

1 Método Conservador

$$IC_1(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{1}{4n}} \right]$$

2 Usando \hat{p} para estimar o erro-padrão

$$IC_2(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

Veja que nos dois casos, os IC 's são da forma $\hat{p} \pm$ margem de erro.

Resumo

Coletamos uma amostra aleatória X_1, X_2, \dots, X_n de uma população com distribuição de Bernoulli com probabilidade de sucesso igual a p , portanto com média p e a variância $p(1 - p)$ e usamos $\bar{X}_n = \hat{p}$ para estimar p .

Pelo TCL:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Propriedade da Normal:

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0, 1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Intervalo de Confiança para p

Então, um intervalo de $100(1 - \alpha)\%$ de confiança para p :

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} ; \hat{p} + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Problema: não conhecemos p . Portanto, usamos:

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

ou, pelo método conservador,

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{1}{4n}} ; \hat{p} + z_{\alpha/2} \sqrt{\frac{1}{4n}} \right]$$

Exemplo: Universitários Não Fumantes

De uma amostra aleatória de 100 alunos de uma universidade, 82 afirmaram ser não fumantes.

Construa um intervalo de confiança de 99% para a proporção de não fumantes entre todos os alunos da universidade.

$$\hat{p} = 0.82, n = 100, \alpha = 0.01, \text{ e } z_{0.005} = 2.58$$

$$\begin{aligned} IC_1(p, 0.99) &= \left[\hat{p} - z_{0.005} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}; \hat{p} + z_{0.005} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right] \\ &= \left[0.82 - 2.58 \sqrt{\frac{(0.82)(0.18)}{100}}; 0.82 + 2.58 \sqrt{\frac{(0.82)(0.18)}{100}} \right] \\ &= [0.72; 0.92] \end{aligned}$$

Exemplo: Universitários Não Fumantes

Podemos também calcular o IC de 99% pelo método conversador:

$$\begin{aligned} IC_2(p, 0.99) &= \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{1}{4n}} \right] \\ &= \left[0.82 - 2.58 \sqrt{\frac{1}{400}}; 0.82 + 2.58 \sqrt{\frac{1}{400}} \right] \\ &= [0.69; 0.95] \end{aligned}$$

Interpretação: Com um grau de confiança de 99%, estimamos que a proporção de não fumantes entre os alunos está entre 72% e 92% (resultado do slide anterior).

E pelo método conservador, com um grau de confiança de 99%, estimamos que a proporção de não fumantes entre os alunos está entre 69% e 95%.

Exemplo: A esposa deve sacrificar a carreira?

Pesquisa do GSS. Você concorda ou não com a seguinte frase: “é mais importante para um esposa ajudar a carreira do marido do que ter uma carreira própria.”

A última vez que esta pergunta foi incluída no GSS foi em 1998 onde 1823 pessoas responderam e 19% concordaram.

- Calcule e interprete o IC de 95% para a proporção na população que concorda com a frase.
- Encontre e interprete a margem de erro do IC de 95%.

Exemplo: A esposa deve sacrificar a carreira?

Calcule e interprete o IC de 95% para a proporção na população que concorda com a frase.

$$\hat{p} = 0.19, n = 1823, \alpha = 0.05, \text{ e } z_{0.025} = 1.96$$

Então,

$$\begin{aligned} IC(p, 0.95) &= \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \\ &= \left[0.19 - 1.96\sqrt{\frac{0.19(1-0.19)}{1823}}; 0.19 + 1.96\sqrt{\frac{0.19(1-0.19)}{1823}} \right] \\ &= [0.17; 0.21] \end{aligned}$$

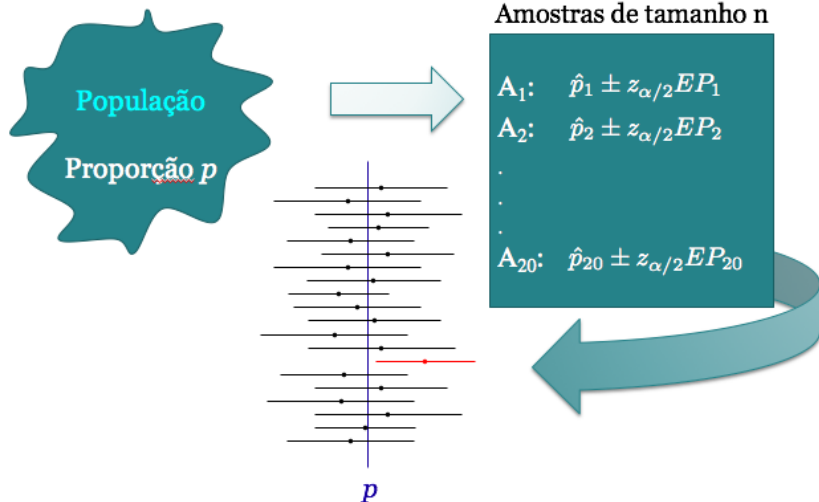
Interpretação do Intervalo de Confiança

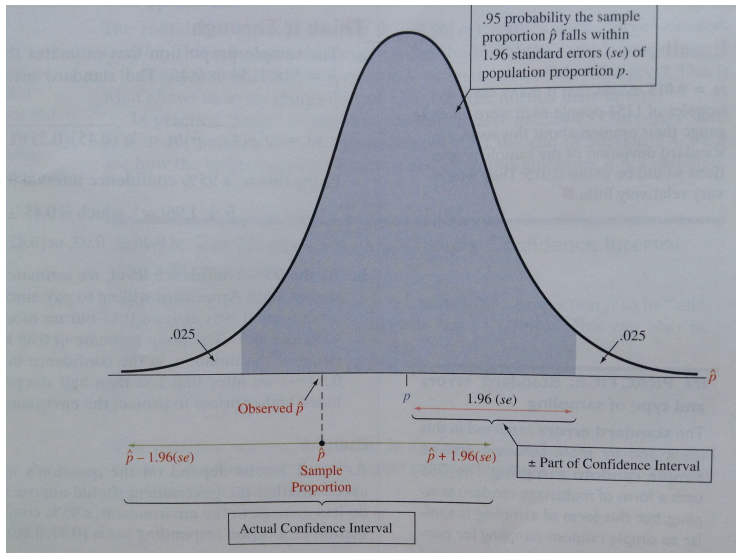
Interpretação: Se várias amostras forem retiradas da população e calcularmos um *IC* de 95% para cada amostra, cerca de 95% desses intervalos irão conter a verdadeira proporção na população, p .

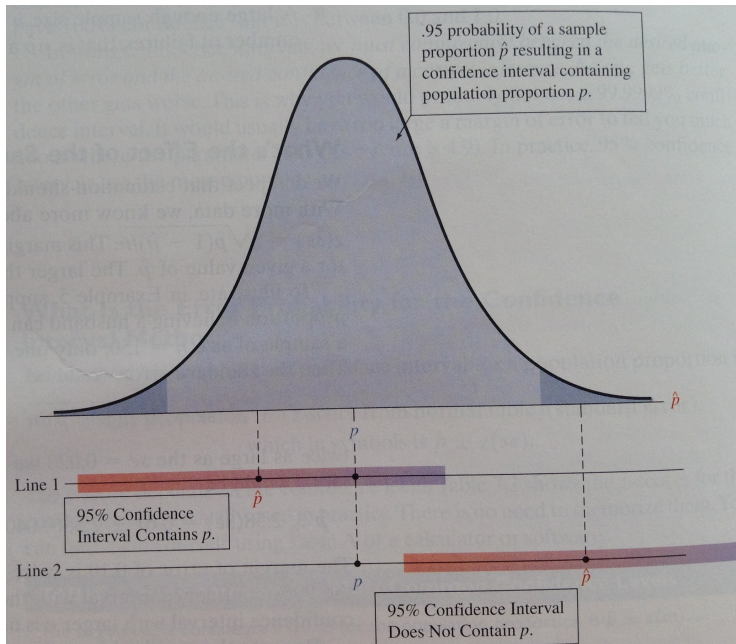
INCORRETO: Dizer que “a probabilidade de que p esteja dentro do intervalo é 95%”.

Por que incorreto? p é uma constante, não é variável aleatória. Ou p está no intervalo ou não está. O intervalo é que é aleatório.

Interpretação do Intervalo de Confiança







Exemplo (continuação)

Um IC de 95% para p é: $[0.17; 0.21]$

A margem de erro (metade do comprimento do IC) é:

$$ME = 1.96 \sqrt{\frac{0.19(1 - 0.19)}{1823}} = 0.02$$

$$P(|\hat{p} - p| < 0.02) = 0.95$$

Interpretação: Com probabilidade 0.95, o erro ao usar a proporção amostral para estimar a proporção populacional não excede 0.02.

Curiosidade: em 1977 a pergunta foi feita pela primeira vez no GSS. $\hat{p} = 0.57$ e IC de 95% foi $[0.55; 0.59]$.

Você pagaria mais por um produto em favor ao meio-ambiente?

Exemplo: Na teoria, muita gente se considera “*eco-friendly*”. Mas e na prática?

Em 2000, GSS perguntou: “Você estaria disposto a pagar mais pela gasolina para proteger o ambiente?”

Entre $n = 1154$ participantes, 518 responderam que sim.

- Encontre IC 95% para a proporção da população que concorda.
- Interprete.

Exemplo (continuação)

Estimativa: $\hat{p} = 518/1154 = 0.45$

Desvio padrão da estimativa (erro padrão): $\sqrt{\frac{0.45(1-0.45)}{1154}} = 0.015$

$$\begin{aligned} IC(p, 0.95) &= \left[0.45 - 1.96\sqrt{\frac{(0.45)(0.55)}{1154}}; 0.45 + 1.96\sqrt{\frac{(0.45)(0.55)}{1154}} \right] \\ &= [0.45 - 1.96 \times 0.015; 0.45 + 1.96 \times 0.015] \\ &= [0.42; 0.48] \end{aligned}$$

Interpretação: Com grau de confiança de 95%, estimamos que a proporção populacional que concorda em pagar mais está entre 0.42 e 0.48. A estimativa pontual, 0.45, tem margem de erro de 3%.

Exemplo (continuação)

E se estivéssemos interessados na proporção que não pagaria mais?

Estimativa: $\hat{p} = 1 - 518/1154 = 0.55$

Desvio padrão da estimativa (erro padrão): $\sqrt{\frac{0.55(1-0.55)}{1154}} = 0.015$

$$\begin{aligned} IC(p, 0.95) &= \left[0.55 - 1.96\sqrt{\frac{(0.55)(0.45)}{1154}}; 0.55 + 1.96\sqrt{\frac{(0.55)(0.45)}{1154}} \right] \\ &= [0.55 - 1.96 \times 0.015; 0.55 + 1.96 \times 0.015] \\ &= [0.52; 0.58] \end{aligned}$$

Interpretação: Com grau de confiança de 95%, estimamos que a proporção populacional que não pagaria mais está entre 0.52 e 0.58. A estimativa pontual, 0.55, tem margem de erro de 3%.

Exemplo: Se a esposa quer ter um filho, mas o marido não, é justo que ele se recuse a ter um filho?

GSS: 598 responderam, 366 acham justo. Encontre um IC de 99%.

Estimativa: $\hat{p} = 366/598 = 0.61$

Desvio padrão da estimativa (erro padrão): $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.02$

$$IC(p, 0.99) = [0.61 - 2.58 \times 0.02; 0.61 + 2.58 \times 0.02] = [0.56; 0.66]$$

Com grau de confiança igual a 99%, estimamos que a proporção populacional que concorda está entre 0.56 e 0.66. A estimativa pontual, 0.61, tem margem de erro de 5%.

Exemplo (continuação)

E o IC de 95%?

$$\begin{aligned} IC(p, 0.95) &= [0.61 - 1.96 \times 0.02; 0.61 + 1.96 \times 0.02] \\ &= [0.57; 0.65] \end{aligned}$$

Com grau de confiança igual a 95%, estimamos que a proporção populacional que concorda está entre 0.57 e 0.65. A estimativa pontual, 0.61, tem margem de erro de 4%.

Com maior grau de confiança, temos uma margem de erro um pouco maior.

Tamanho da amostra para estimar p

Exemplo

A Datafolha quer fazer uma pesquisa de boca-de-urna para prever o resultado de uma eleição com apenas dois candidatos.

Seleciona então uma a.a. de eleitores e pergunta em quem cada um votou. Para esta pesquisa, o Datafolha quer uma margem de erro de 4%. Qual o tamanho de amostra necessário?

- O grau de confiança é 95% e $IC(p, 0.95) = \hat{p} \pm 1.96 \times EP(\hat{p})$
- Erro padrão de \hat{p} é $EP(\hat{p}) = \sqrt{p(1-p)/n}$
- Margem de erro: $1.96 \times EP(\hat{p}) = 1.96\sqrt{p(1-p)/n}$
- Margem de erro desejada é 0.04. Então, o tamanho amostral necessário n é:

$$1.96\sqrt{\frac{p(1-p)}{n}} = 0.04 \quad \Rightarrow \quad n = \frac{1.96^2 p(1-p)}{0.04^2}$$

Exemplo

Problema é que não conhecemos p .

Assim como para encontrar os IC 's, podemos usar o método conservador ou então usar informações obtidas em pesquisas anteriores (caso existam).

Método Conservador:

Lembre que $p(1 - p)/n$ é a variância da estimativa \hat{p} e já vimos anteriormente que $p(1 - p) \leq 1/4$.

Então,

$$n = \frac{1.96^2 \times (1/4)}{0.04^2} = 600$$

Exemplo

Outra alternativa

O Datafolha fez uma pesquisa na semana passada e o resultado foi 58% votariam no candidato A e 42% no B .

Podemos usar então estas estimativas:

$$n = \frac{1.96^2 \hat{p}(1 - \hat{p})}{0.04^2} = \frac{1.96^2 (0.58)(0.42)}{0.04^2} = 585$$

Uma a.a. de tamanho 585 deverá resultar numa margem de erro de 4% para um IC de 95% para a proporção da população que vota no candidato A .

Exemplo

Uma firma de propaganda está interessada em estimar a proporção de domicílios que estão assistindo a final do campeonato brasileiro de futebol. Para isso, está planejando ligar para os domicílios selecionados aleatoriamente a partir de uma lista. Qual o tamanho da amostra necessário se a firma quer 90% de confiança de que a estimativa obtida tenha uma margem de erro igual a 0.02?

Exemplo - Solução

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{1}{4n}} \right]$$

Margem de erro 0.02: $z_{\alpha/2} \sqrt{\frac{1}{4n}} = 0.02$

Como eles querem 90% de confiança, $\alpha = 0.10$ e $z_{0.05} = 1.645$

$$1.645 \sqrt{1/4n} = 0.02 \quad \Longleftrightarrow \quad 1/4n = (0.02/1.645)^2 \quad \Rightarrow \quad n = 1691.3$$

Tamanho amostral: 1692.

Em geral, para uma margem de erro m :

$$n = \left(\frac{z_{\alpha/2}}{2m} \right)^2$$

Exemplo

Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres.

Coletamos uma amostra aleatória simples de $n = 10$ estudantes e calculamos a proporção de mulheres na amostra, ou seja, \hat{p} .

Qual a probabilidade de que \hat{p} difira de p em menos de 0.01? E se $n = 50$?

Adaptado de: Morettin & Bussab, Estatística Básica 5ª edição, pág 276.

Solução: Temos que a probabilidade que desejamos encontrar é dada por

$$P(|\hat{p} - p| < 0.01) = P(-0.01 < \hat{p} - p < 0.01)$$

onde p é o valor verdadeiro da proporção de mulheres, e \hat{p} a proporção observada na amostra.

Exemplo - Solução

Seja X_i a v.a. indicando se a pessoa i é mulher ou não. Temos que $X_i \sim \text{Bernoulli}(p)$, com $p = 0.3$.

Então sabemos que $\mathbb{E}(X_i) = p$ e $\text{Var}(X_i) = p(1 - p)$.

Coletamos uma amostra de tamanho n : X_1, \dots, X_n . Calculamos a proporção de mulheres na amostra:

$$\bar{X}_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$$

Exemplo - Solução

Sabemos que $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_i) = p$ e $Var(\bar{X}_n) = \frac{Var(X_i)}{n} = \frac{p(1-p)}{n}$.

Sabemos também, pelo TCL, que se n é grande, a distribuição de $\bar{X}_n = \hat{p}$ pode ser aproximada por uma normal $N(p, p(1-p)/n)$.

Como $p = 0.3$ e $n = 10$, temos que:

$$Var(\hat{p}) = \frac{0.3 \times 0.7}{10} = 0.021$$

Exemplo - Solução

$$P(|\hat{p} - p| < 0.01) = P(-0.01 < \hat{p} - p < 0.01)$$

$$P\left(-\frac{0.01}{\sqrt{\text{Var}(\hat{p})}} < \frac{\hat{p} - p}{\sqrt{\text{Var}(\hat{p})}} < \frac{0.01}{\sqrt{\text{Var}(\hat{p})}}\right)$$

$$P\left(\frac{-0.01}{\sqrt{0.021}} < Z < \frac{0.01}{\sqrt{0.021}}\right) = P(-0.07 < Z < 0.07) = 0.056.$$

Mas $n = 10$ é grande? Podemos comparar essa probabilidade com o resultado exato.

Exemplo - Solução

Não sabemos a distribuição de \hat{p} , mas o evento $\hat{p} = \gamma$, onde $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$, é igual ao evento $\sum_{i=1}^n X_i = n\gamma$, onde X_i são v.a. independentes e identicamente distribuídas Bernoulli(0.3).

A soma é portanto Binomial(10, 0.3).

O evento $\{|\hat{p} - p| < 0.01\}$ é igual ao evento $\{|\sum X_i - 10 \cdot 0.3| < 0.1\}$.

Como $\sum X_i$ assume somente valores inteiros, temos que

$$\left\{ \left| \sum_{i=1}^{10} X_i - 10 \cdot 0.3 \right| < 0.1 \right\} = \left\{ \sum_{i=1}^{10} X_i = 3 \right\}.$$

Exemplo - Solução

Portanto,

$$P\left(\left\{\sum_{i=1}^{10} X_i = 3\right\}\right) = \binom{10}{3} 0.3^3 0.7^7 = 0.267.$$

Temos uma probabilidade que é 5 vezes maior que a aproximação.

Exemplo - Solução

Tome $n = 50$, agora. Podemos modificar rapidamente as contas da aproximação normal. A variância agora é $\frac{p(1-p)}{n} = 0.0042$, e portanto a probabilidade aproximada é:

$$P\left(\frac{-0.01}{\sqrt{0.0042}} < Z < \frac{0.01}{\sqrt{0.0042}}\right) = P(-0.154 < Z < 0.154) = 0.12239$$

A probabilidade exata agora é dada pelo evento $|\sum X_i - 50 \cdot 0.3| < 0.5$, ou simplesmente $\sum_{i=1}^{50} X_i = 15$.

Intervalo de Confiança para proporções

Observe agora que

$$P\left(\sum_{i=1}^{50} X_i = 15\right) = \binom{50}{15} 0.3^{15} 0.7^{50-15} = 0.12237$$

A diferença agora é muito menor e, à medida que $n \rightarrow \infty$ ela tende a 0, pelo TCL.

É preciso contudo ter em mente que a aproximação só é válida para grandes tamanhos de amostra.

Intervalo de Confiança para proporções

Exercício

Suponha que estejamos interessados em estimar a porcentagem de consumidores de um certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:

- 1 O intervalo de confiança de p , com i.c. de 95%; interprete o resultado.
- 2 O tamanho da amostra para que o erro da estimativa não exceda 0.02 unidades com probabilidade de 95%; interprete o resultado.

Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 309.

Intervalo de Confiança para proporções

1 O intervalo de confiança de 95% é dado por:

$$IC(p; 0.95) = 0.333 \pm 1.96 \sqrt{\frac{0.333 \cdot 0.667}{300}} = 0.333 \pm 0.053$$

Ou simplesmente (0.280; 0.387).

Interpretação: Se pudéssemos construir um grande número de intervalos aleatórios para p , todos baseados em amostras de tamanho n , 95% deles conteriam o parâmetro p .

Intervalo de Confiança para proporções

- 2** Utilizando a estimativa da amostra observada ($\hat{p} = 0.333$), temos que n é dado por

$$n = \left(\frac{1.96}{0.02} \right)^2 \times 0.333 \times 0.667 \cong 2134.$$

Contudo, frequentemente devemos determinar o tamanho da amostra antes de realizar qualquer experimento, isto é, sem nenhuma informação prévia de p . Se esse for o caso, devemos considerar o caso em que a variância da amostra é a pior possível.

Intervalo de Confiança para proporções

2 Utilizando o valor máximo de $p(1 - p)$, isto é, $1/4$, obtemos

$$n = \left(\frac{1.96}{0.02} \right)^2 \times \frac{1}{4} \cong 2401$$

Interpretação: Utilizando o tamanho amostral encontrado, teremos uma probabilidade de 95% de que a proporção amostral não difira do verdadeiro valor de p em menos que 2%.

Note que a prática de obter amostras pequenas para examinar p , e aí determinar o tamanho amostral sem utilizar o “pior caso”, é no que consiste a idéia de *amostras piloto*.



- [Ross](#): capítulo 8.
- [OpenIntro](#): seção 4.2.
- [Magalhães](#): seção 7.4.