

Estimação: (B) Estimação por intervalo

Wagner H. Bonat
Fernando P. Mayer
Elias T. Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

10/05/2018



Sumário

- 1 Introdução
- 2 Intervalos de confiança para a média: σ conhecido
 - IC para a média: σ conhecido
 - Determinação do tamanho amostral
- 3 Intervalos de confiança para a média: σ desconhecido
 - IC para a média: σ desconhecido
 - Determinação do tamanho amostral
- 4 Intervalo de confiança para a proporção p
 - IC para a proporção p
 - Determinação do tamanho amostral
- 5 Exercícios

Estimação

Existem dois tipos de estimativas que podemos obter a partir de uma **amostra aleatória**:

Estimativa pontual

Fornecem como estimativa um único valor numérico para o parâmetro de interesse

Estimativa intervalar

Fornece um intervalo de valores “plausíveis” para o parâmetro de interesse

Estimação

Por serem **variáveis aleatórias**, os estimadores pontuais possuem uma distribuição de probabilidade (distribuições amostrais).

Com isso, podemos apresentar uma estimativa mais informativa para o parâmetro de interesse, que inclua uma medida de **precisão** do valor obtido
→ **estimativa intervalar** ou **intervalo de confiança**.

Os **intervalos de confiança** são obtidos a partir da **distribuição amostral** de seus estimadores.

Sumário

- 1 Introdução
- 2 Intervalos de confiança para a média: σ conhecido
 - IC para a média: σ conhecido
 - Determinação do tamanho amostral
- 3 Intervalos de confiança para a média: σ desconhecido
 - IC para a média: σ desconhecido
 - Determinação do tamanho amostral
- 4 Intervalo de confiança para a proporção p
 - IC para a proporção p
 - Determinação do tamanho amostral
- 5 Exercícios

Suposições necessárias

- A amostra é uma **amostra aleatória simples**. (Todas as amostras de mesmo tamanho tem a mesma probabilidade de serem selecionadas)
- O valor do desvio padrão populacional σ , é conhecido
- Uma ou ambas das seguintes condições são satisfeitas:
 - A população é normalmente distribuída
 - A amostra possui $n > 30$

Erro amostral

Quando coletamos uma **amostra aleatória** e calculamos uma média, sabemos que o valor da média possui um desvio natural, em relação ao verdadeiro valor da média populacional (**erro amostral**), ou seja

$$e = \bar{X} - \mu \quad \Rightarrow \quad \bar{X} = \mu + e$$

Sabemos que a **distribuição amostral da média** é uma distribuição normal, com média μ e variância σ^2/n ,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Margem de erro

Usando a transformação

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{e}{\sigma/\sqrt{n}} \sim N(0, 1)$$

podemos determinar o **erro máximo provável** que assumimos para a média amostral que estamos calculando.

O **erro máximo provável** ou **margem de erro** da média é definido por

$$e = z_{\gamma/2} \cdot \frac{\sigma}{\sqrt{n}}$$

onde $z_{\gamma/2}$ é chamado de **valor crítico**.

Intervalo de confiança

Fixando um valor γ tal que $0 < \gamma < 1$, podemos encontrar um valor $z_{\gamma/2}$ tal que:

$$P[|Z| < z_{\gamma/2}] = \gamma$$

$$P[-z_{\gamma/2} < Z < z_{\gamma/2}] = \gamma$$

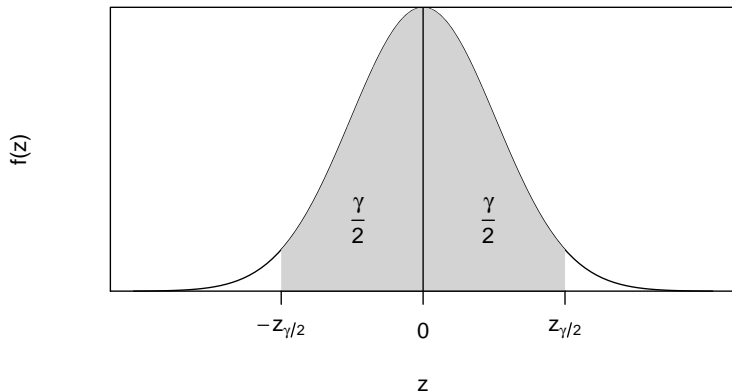
$$P[-z_{\gamma/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\gamma/2}] = \gamma$$

$$P\left[\bar{x} - z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)\right] = \gamma$$

$$P[\bar{x} - e < \mu < \bar{x} + e] = \gamma$$

Intervalo de confiança

O valor crítico $z_{\gamma/2}$ é o valor de γ dividido por 2, uma vez que a “massa” γ deve ser distribuída igualmente em torno de 0.



Coeficiente de confiança γ

A área γ determina o **coeficiente de confiança** associado ao intervalo de confiança que estamos construindo.

O valor $z_{\gamma/2}$ pode ser obtido da tabela da Normal padrão, localizando o valor de $\gamma/2$ no corpo da tabela e obtendo o valor $z_{\gamma/2}$ nas margens correspondentes.

Exemplo: $\gamma = 0,95$:

- Temos que $\gamma/2 = 0,475$ é a área que devemos procurar no corpo da tabela
- O valor de $z_{\gamma/2}$ será determinado pelos valores correspondentes nas margens da tabela. Nesse caso, $z_{\gamma/2} = 1,96$ é o valor crítico procurado.

Intervalo de confiança

Com estas definições, podemos construir um **intervalo de confiança** para μ , com **coeficiente de confiança** γ :

$$\text{IC}(\mu, \gamma) = \left[\bar{X} - z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}} \right); \bar{X} + z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \right]$$

Outras notações:

$$\bar{x} - e < \mu < \bar{x} + e$$

$$\bar{x} \pm e$$

$$[\bar{x} - e; \bar{x} + e]$$

Procedimentos gerais para a construção de intervalos de confiança

1. Verifique se as suposições necessárias estão satisfeitas
 - Temos uma AAS
 - σ é conhecido
 - A população tem distribuição normal ou $n > 30$
2. Determine o nível de confiança γ , e encontre o valor crítico $z_{\gamma/2}$
3. Calcule a margem de erro $e = z_{\gamma/2} \cdot (\sigma/\sqrt{n})$
4. Calcule $IC(\mu, \gamma)$

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança:

$$IC(\mu, 95\%) = [52; 58]$$

Interpretação 1

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 52 e 58

Interpretação 2

Temos 95% de confiança de que o intervalo entre 52 e 58 realmente contém a verdadeira média populacional μ

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança:

$$IC(\mu, 95\%) = [52; 58]$$

Interpretação 1 — ERRADA

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 52 e 58

Interpretação 2 — CERTA

Temos 95% de confiança de que o intervalo entre 52 e 58 realmente contém a verdadeira média populacional μ

Interpretação de um intervalo de confiança

Como o intervalo de confiança é calculado a partir de uma **amostra aleatória**, este intervalo **também é aleatório!**

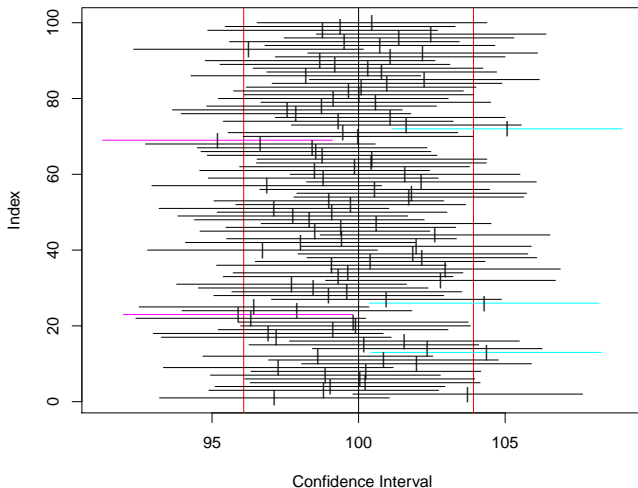
Isso significa que para cada amostra aleatória que tivermos, um intervalo **diferente** será calculado.

Como o valor de μ é fixo, é o intervalo que deve conter o valor de μ , e não o contrário.

Isso significa que se pudessemos obter 100 amostras diferentes, e calcularmos um intervalo de confiança de 95% para cada uma das 100 amostras, esperaríamos que 5 destes intervalos **não** contenham o verdadeiro valor da média populacional μ .

Interpretação de um intervalo de confiança

Confidence intervals based on z distribution



Exemplo

Uma empresa de computadores deseja estimar o tempo médio de horas semanais que as pessoas utilizam o computador.

Uma amostra aleatória de 25 pessoas apresentou um tempo médio de uso de 22,4 horas. Com base em estudos anteriores, a empresa assume que $\sigma = 5,2$ horas, e que os tempos são normalmente distribuídos.

Construa um intervalo de confiança para a média μ com coeficiente de confiança de 95%.

Amplitude de um intervalo

A **amplitude** de um intervalo de confiança é dada pela diferença entre o limite superior e inferior, ou seja,

$$\begin{aligned} \text{AMP}_{IC} &= \left[\bar{x} + z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \right] - \left[\bar{x} - z_{\gamma/2} \cdot \left(\frac{\sigma}{\sqrt{n}} \right) \right] \\ &= 2 \times z_{\gamma/2} \cdot (\sigma / \sqrt{n}) \end{aligned}$$

- Note que, claramente, um intervalo de confiança depende conjuntamente de três componentes:
 - Coeficiente de confiança γ , expresso pelo valor crítico $z_{\gamma/2}$
 - Desvio-padrão populacional σ
 - Tamanho da amostra n

Amplitude de um intervalo

$z_{\gamma/2} \rightarrow$ Cada vez que aumentamos a confiança γ , o valor de $z_{\gamma/2}$ fica maior, e consequentemente a amplitude do intervalo aumenta.

$\sigma \rightarrow$ Um grande desvio padrão indica a possibilidade de um considerável distanciamento dos valores amostrais em relação à média populacional

$n \rightarrow$ Quanto maior for o tamanho da amostra, maior será a quantidade de informação disponível. Com isso, valores maiores de n produzem intervalos mais informativos

Exemplo

Seja $X \sim N(\mu, 36)$

- a) Para uma amostra de tamanho 50, obtivemos média amostral 18,5. Construa intervalos de confiança de
 - (i) 90% (ii) 95% (iii) 99%
- b) Calcule as amplitudes dos intervalos acima e explique a diferença.
- c) Para um nível de confiança de 95%, construa intervalos de confiança (admita a mesma média amostral 18,5) supondo tamanhos de amostra
 - (i) $n = 15$ (ii) $n = 100$
- d) Calcule as amplitudes dos intervalos acima e explique a diferença.

Determinação do tamanho amostral

Nosso objetivo é coletar dados para estimar a **média populacional** μ .

A questão é:

Quantos elementos (itens, objetos, pessoas, ...) devemos amostrar?

Já vimos que, de maneira (bem) geral, $n > 30$ é um tamanho de amostra mínimo para a maioria dos casos.

Será que podemos ter uma estimativa melhor de quantos elementos devem ser amostrados para estimarmos a média populacional com uma precisão conhecida?

Determinação do tamanho amostral

A partir da equação do **erro máximo provável**

$$e = z_{\gamma/2} \cdot \frac{\sigma}{\sqrt{n}}$$

podemos isolar n e chegar na seguinte equação para a determinação do tamanho amostral

$$n = \left[\frac{z_{\alpha/2} \cdot \sigma}{e} \right]^2$$

Determinação do tamanho amostral

Note que, em

$$n = \left[\frac{z_{\alpha/2} \cdot \sigma}{e} \right]^2$$

- O tamanho amostral n **não** depende do tamanho populacional N
- O tamanho amostral depende:
 - do nível de confiança desejado (expresso pelo valor crítico $z_{\alpha/2}$)
 - do erro máximo *desejado*
 - do desvio-padrão σ (embora veremos que não é estritamente necessário)
- Como o tamanho amostral precisa ser um número inteiro, arredondamos sempre o valor para o **maior** número inteiro mais próximo

Exemplo

Seja $X \sim N(\mu, 36)$

- a) Calcule o tamanho da amostra, para que com 95% de probabilidade, a média amostral não difira da média populacional por mais de
 - (i) 0,5 unidades (ii) 2 unidades
- b) Qual o impacto do erro máximo assumido para o tamanho da amostra?
- c) Calcule o tamanho da amostra, para que a diferença da média amostral para a média populacional (em valor absoluto) seja menor ou igual a 2 unidades, com níveis de confiança de
 - (i) 90% (ii) 95%
- d) Compare as estimativas do item anterior e analise o impacto do nível de confiança para a determinação do tamanho amostral.

Sumário

- 1 Introdução
- 2 Intervalos de confiança para a média: σ conhecido
 - IC para a média: σ conhecido
 - Determinação do tamanho amostral
- 3 Intervalos de confiança para a média: σ desconhecido
 - IC para a média: σ desconhecido
 - Determinação do tamanho amostral
- 4 Intervalo de confiança para a proporção p
 - IC para a proporção p
 - Determinação do tamanho amostral
- 5 Exercícios

Estimativa da variância amostral

Na maioria das situações práticas, não sabemos o verdadeiro valor do desvio padrão populacional σ .

Se o desvio padrão é desconhecido, ele precisa ser estimado.

Sendo (X_1, \dots, X_n) VAs onde $X \sim N(\mu, \sigma^2)$, vimos que o “melhor” estimador para σ^2 é a variância amostral

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

que é não viciada e consistente para σ^2 .

A distribuição t de Student

Definindo a variável padronizada

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

o denominador S^2 fará com que a função densidade de T seja diferente da Normal.

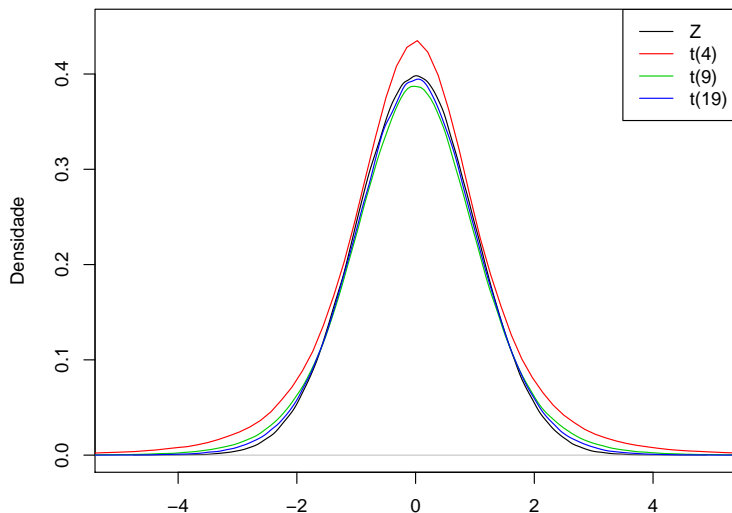
Essa nova densidade é denominada **t de Student**, e seu parâmetro é denominado **graus de liberdade**, que nesse caso é $n - 1$. Assim:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$$

Características da distribuição t

- É simétrica com média $t = 0$ (assim como $z = 0$)
- É diferente para tamanhos de amostra diferentes
- Possui maior área nas caudas e menor área no centro (quando comparada com a distribuição normal) \rightarrow para incorporar a incerteza
- O desvio padrão da distribuição t varia com o tamanho da amostra (ao contrário da distribuição z onde $\sigma = 1$)
 - $n \downarrow \quad \sigma \uparrow$
 - $n \uparrow \quad \sigma \downarrow$
- A medida que o n amostral aumenta, a distribuição t se aproxima cada vez mais de uma distribuição normal padrão Z
 - Por isso, para amostras grandes ($n > 30$) o resultado das duas é similar

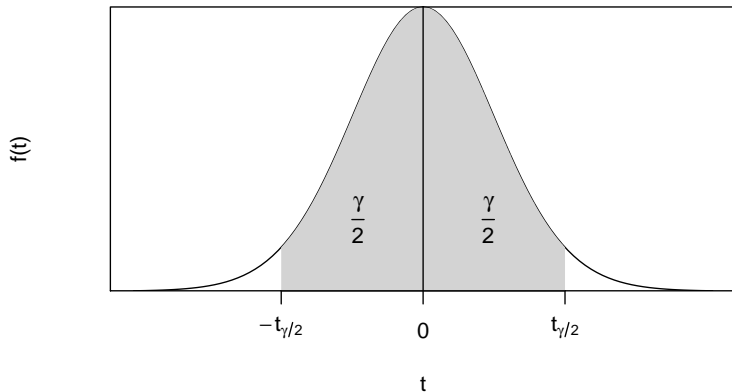
Características da distribuição t



Encontrando valores críticos de t

Com a definição do **nível de confiança** e sabendo o tamanho da amostra n , sabemos então o valor de γ e dos gl, e devemos encontrar o **valor crítico** de $t_{\gamma/2}$. Usando como exemplo $\gamma = 0,95$ e uma amostra de $n = 7$

- Temos que $n = 7 \Rightarrow gl = n - 1 = 6$
- Na tabela da distribuição t de Student procure a linha correspondente aos gl, e coluna correspondente ao valor de $1 - \gamma = 1 - 0,95 = 0,05 = 5\%$
- O valor de $t_{\gamma/2}$ será determinado pelos valores correspondentes **no corpo da tabela**. Nesse caso, $t_{\gamma/2} = 2,447$ é o valor crítico procurado.

Encontrando valores críticos de t 

Intervalo de confiança

Com estas definições, podemos construir um **intervalo de confiança** para μ , com **coeficiente de confiança** γ , e σ desconhecido:

$$\text{IC}(\mu, \gamma) = \left[\bar{X} - t_{\gamma/2} \cdot \left(\frac{S}{\sqrt{n}} \right); \bar{X} + t_{\gamma/2} \cdot \left(\frac{S}{\sqrt{n}} \right) \right]$$

Outras notações:

$$\bar{x} - e < \mu < \bar{x} + e$$

$$\bar{x} \pm e$$

$$[\bar{x} - e; \bar{x} + e]$$

Procedimentos gerais para a construção de intervalos de confiança

1. Verifique se as suposições necessárias estão satisfeitas
 - Temos uma AAS
 - Temos uma estimativa de s
 - A população tem distribuição normal ou $n > 30$
2. Determine o nível de confiança γ , e encontre o valor crítico $t_{\gamma/2}$
3. Calcule a margem de erro $e = t_{\gamma/2} \cdot (S/\sqrt{n})$
4. Calcule $IC(\mu, \gamma)$

Exemplo

Em um teste da eficácia do alho na dieta para a redução do colesterol, 49 pessoas foram avaliadas e seus níveis de colesterol foram medidos antes e depois do tratamento. As **mudanças** nos níveis de colesterol apresentaram média de 0,4 e desvio-padrão de 21.

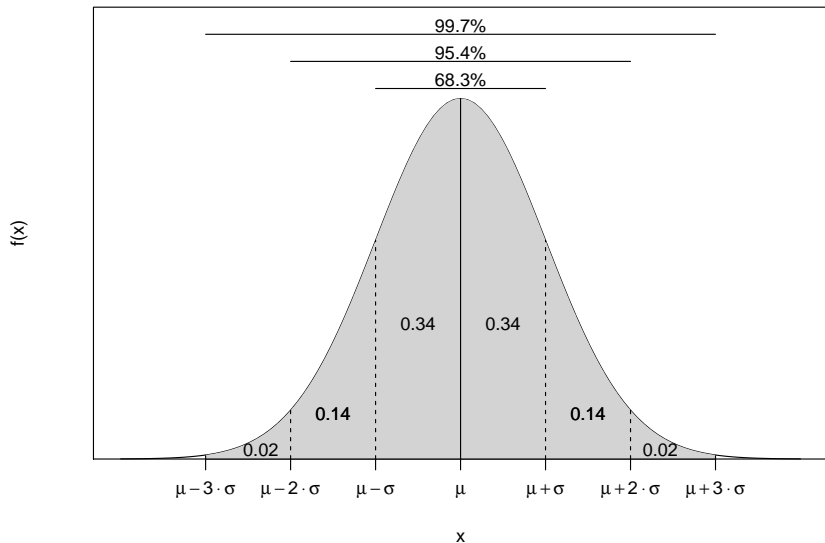
- a) Para um nível de confiança de 95%, calcule o intervalo para a verdadeira média das mudanças no nível de colesterol
- b) O que o intervalo de confiança sugere sobre a eficácia do uso do alho na dieta para a redução do colesterol?
- c) Resolva o mesmo exemplo supondo que o $\sigma = s$ é conhecido (ou seja, usando a distribuição Z). Compare os dois métodos.

Determinação do tamanho amostral

Se σ for desconhecido?

- Estime o valor de σ com base em algum estudo feito anteriormente
- Faça uma amostra piloto e estime o desvio padrão amostral s , e use-o como uma aproximação para o desvio-padrão populacional σ
- Use a **regra empírica** da amplitude para dados com distribuição (aproximadamente) normal

Regra empírica para uma distribuição normal



Regra empírica para uma distribuição normal

Define-se **valores usuais** aqueles que são típicos e não muito extremos.

Como sabemos que em uma distribuição (aproximadamente) normal, aproximadamente 95% dos dados encontram-se a 2 desvios-padrões acima e abaixo da média, temos que

$$4\sigma = (\max - \min)$$
$$\sigma = \frac{(\max - \min)}{4}$$

pode ser utilizado como uma estimativa para σ .

Exemplo

Um professor deseja estimar o salário médio de professores do ensino médio de uma cidade. Quantos professores devem ser selecionados para termos 90% de confiança que a média amostral esteja a menos de R\$30,00 da média populacional? Sabe-se apenas que os salários variam entre R\$800,00 e R\$1.200,00.

Sumário

- 1 Introdução
- 2 Intervalos de confiança para a média: σ conhecido
 - IC para a média: σ conhecido
 - Determinação do tamanho amostral
- 3 Intervalos de confiança para a média: σ desconhecido
 - IC para a média: σ desconhecido
 - Determinação do tamanho amostral
- 4 Intervalo de confiança para a proporção p
 - IC para a proporção p
 - Determinação do tamanho amostral
- 5 Exercícios

Proporção amostral

A proporção amostral

$$\hat{p} = \frac{x}{n} = \frac{\text{número de sucessos}}{\text{total de tentativas}}$$

é a “melhor estimativa” para a proporção populacional p

Através do estudo da distribuição amostral da proporção, chegamos aos seguintes resultados:

- A proporção amostral \hat{p} **tende** para o valor da proporção populacional p
- A distribuição das proporções amostrais tende a ser uma **distribuição normal**
- $E(\hat{p}) = \mu_{\hat{p}} = p$
- $\text{Var}(\hat{p}) = \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$

Distribuição amostral da proporção \hat{p}

Assim, sabemos que

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

Ainda podemos mostrar que a quantidade

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

Quando não conhecemos p , usamos $\hat{p} = x/n$ como estimativa

Erro amostral da proporção

O **erro amostral** da proporção pode ser definido por

$$e = \hat{p} - p \Rightarrow p = \hat{p} + e$$

então

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

Portanto o **erro máximo provável** ou **margem de erro** da proporção é

$$e = z_{\gamma/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

Intervalo de confiança

Com estas definições, podemos construir um **intervalo de confiança** para p , com **coeficiente de confiança** γ

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

Outras notações:

$$\hat{p} - e < p < \hat{p} + e$$

$$\hat{p} \pm e$$

$$[\hat{p} - e; \hat{p} + e]$$

Procedimentos gerais para a construção de intervalos de confiança

1. Verifique se as suposições necessárias estão satisfeitas
 - Temos uma AAS
 - As condições para a distribuição binomial são satisfeitas
 - as tentativas são independentes
 - há duas categorias de resultado (“sucesso”, “fracasso”)
 - a probabilidade de sucesso p permanece constante
 - A distribuição normal pode ser usada como aproximação para a distribuição binomial, ou seja, $np \geq 5$ e $np(1 - p) \geq 5$
2. Determine o nível de confiança γ , e encontre o valor crítico $z_{\gamma/2}$
3. Calcule a margem de erro $e = z_{\gamma/2} \cdot \sqrt{\frac{p(1-p)}{n}}$
4. Calcule $IC(p, \gamma)$

Exemplo

Em uma pesquisa realizada por um instituto de pesquisa Norte-Americano, 1500 adultos foram selecionados aleatoriamente para responder à pergunta se acreditam ou não no aquecimento global. 1050 entrevistados responderam que sim. Com isso:

- a) Para um nível de confiança de 95%, calcule o intervalo de confiança para a verdadeira proporção de pessoas que acreditam no aquecimento global
- b) Com base nesse resultado, podemos concluir que a maioria dos adultos acredita no aquecimento global?

Determinação do tamanho amostral

A partir da equação do erro máximo provável

$$e = z_{\gamma/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

podemos isolar n e chegar na seguinte equação para a determinação do tamanho amostral para uma proporção populacional

$$n = \left(\frac{z_{\gamma/2}}{e} \right)^2 \cdot p(1-p)$$

Determinação do tamanho amostral

Como **não conhecemos** a proporção populacional p , temos duas alternativas:

1. Usar \hat{p} no lugar de p (estimativa otimista)
2. Usar $p = 0,5$ (estimativa conservadora). Porque quando $p = 0,5$ teremos o maior tamanho de amostra possível.

p	$(1 - p)$	$p(1 - p)$
0,1	0,9	0,09
0,3	0,7	0,21
0,5	0,5	0,25
0,6	0,4	0,24
0,8	0,2	0,16

Exemplo

Um fabricante de peças deseja estimar a verdadeira proporção de peças defeituosas no processo de fabricação, com um erro máximo de 3% e nível de confiança de 99%. Calcule o tamanho da amostra necessário para se estimar esta proporção se:

- a) O fabricante tem uma estimativa de que em uma amostra anterior, aproximadamente 10% das peças eram defeituosas.
- b) O fabricante não tem nenhuma informação prévia sobre a proporção de peças defeituosas.

Sumário

- 1 Introdução
- 2 Intervalos de confiança para a média: σ conhecido
 - IC para a média: σ conhecido
 - Determinação do tamanho amostral
- 3 Intervalos de confiança para a média: σ desconhecido
 - IC para a média: σ desconhecido
 - Determinação do tamanho amostral
- 4 Intervalo de confiança para a proporção p
 - IC para a proporção p
 - Determinação do tamanho amostral
- 5 Exercícios

Exercícios recomendados

- Seção 7.4 - 1 a 5
- Seção 7.5 - 17, 19, 21, 23, 25, 29
- **Seção 8.3 (distribuição t) - 1**