

# Intervalos de Confiança - Amostras Grandes

Prof. Eduardo Bezerra

CEFET/RJ

14 de Setembro de 2018

- 1 Intervalo de confiança para médias
- 2 Intervalo de confiança para proporções

- 1 Intervalo de confiança para médias
- 2 Intervalo de confiança para proporções

# Intervalo de confiança para médias

Vamos estudar um procedimento para calcular intervalos de confiança para a média amostral  $\bar{X}$ .

# Intervalo de confiança para médias

Vamos estudar um procedimento para calcular intervalos de confiança para a média amostral  $\bar{X}$ .

Aqui, consideramos que o tamanho da amostra é *grande* (i.e.,  $n > 30$ ).

# Intervalo de confiança para médias

Vamos estudar um procedimento para calcular intervalos de confiança para a média amostral  $\bar{X}$ .

Aqui, consideramos que o tamanho da amostra é *grande* (i.e.,  $n > 30$ ).

Nesse caso, podemos utilizar o TLC.

# Intervalo de confiança para médias

Considere uma amostra de 100 alturas retirada a partir de alunos de uma universidade.

- São calculados a média e o desvio padrão dessa amostra:  $\bar{x} = 1,58$  e  $s = 0,2$ .
- O valor  $\bar{x} = 1,58$  é uma estimativa pontual de  $\mu$ .
- Para ter uma ideia do quanto  $\mu$  difere de 1,58, construímos um **intervalo de confiança** (em torno de 1,58), que *provavelmente* contenha  $\mu$ .

# Intervalo de confiança para médias

Para construir esse intervalo de confiança, considere que

- a) a distribuição da população tem média  $\mu$  e variância  $\sigma^2$ ;
- b) os valores de  $\mu$  e  $\sigma$  são desconhecidos;
- c)  $(X_1, X_2, \dots, X_{100})$  são as 100 alturas amostradas.

Visto que  $\bar{X}$  é a média de uma *amostra grande*, o TLC nos diz que esse valor vem de uma distribuição normal:



# Intervalo de confiança para médias

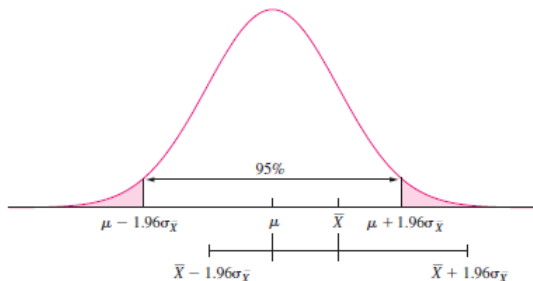
Para construir esse intervalo de confiança, considere que

- a) a distribuição da população tem média  $\mu$  e variância  $\sigma^2$ ;
- b) os valores de  $\mu$  e  $\sigma$  são desconhecidos;
- c)  $(X_1, X_2, \dots, X_{100})$  são as 100 alturas amostradas.

Visto que  $\bar{X}$  é a média de uma *amostra grande*, o TLC nos diz que esse valor vem de uma distribuição normal:

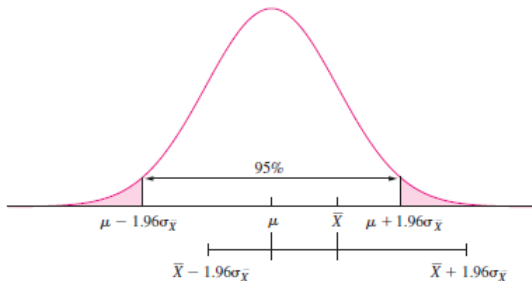
$$\bar{X} \sim N(1,58; \frac{(0,2)^2}{100})$$

# Intervalo de confiança para médias



A figura apresenta a distribuição (normal) da v.a.  $\bar{X}$ . Apresenta também um cenário possível em que a estimativa pontual  $\bar{X}$  é maior do que  $\mu$ .

# Intervalo de confiança para médias



A figura apresenta a distribuição (normal) da v.a.  $\bar{X}$ . Apresenta também um cenário possível em que a estimativa pontual  $\bar{X}$  é maior do que  $\mu$ .

De onde vêm os valores **1,96** ou **95%** nessa figura?!

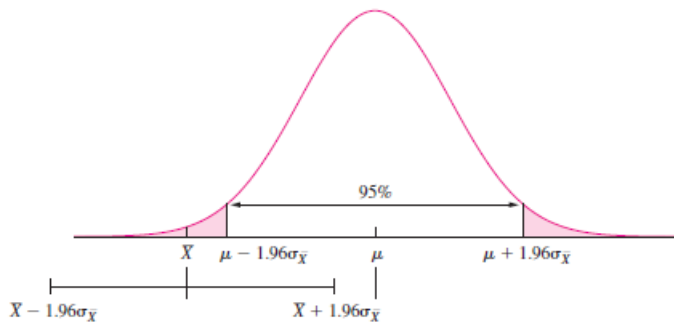
# Intervalo de confiança para médias

Repare que

- para essa amostra em particular, a estimativa para  $\bar{X}$  “caiu dentro” do intervalo correspondente à região que cobre 95% da área central da curva normal.
- para 95% de todas as amostras que poderiam ser colhidas da população e para as quais o intervalo correspondente fosse calculado,  $\mu$  estaria contido nesse intervalo.

Portanto, dizemos que o intervalo de confiança para  $\mu$  no nível de 95% é  $\bar{X} \pm 1,96\sigma_{\bar{X}}$ .

# Intervalo de confiança para médias



Outro cenário possível de amostragem: dessa vez, a estimativa pontual  $\bar{X}$  “caiu fora” do intervalo correspondente a 95% da área central.

# Intervalo de confiança para médias

Repare que

- para essa amostra em particular,  $\bar{X}$  “caiu fora” do intervalo correspondente à região que cobre 95% da área central da distribuição.
- 5% de todas as amostras que poderiam ser colhidas da população apresentariam esse comportamento.

# Intervalo de confiança para médias

Repare que

- para essa amostra em particular,  $\bar{X}$  “caiu fora” do intervalo correspondente à região que cobre 95% da área central da distribuição.
- 5% de todas as amostras que poderiam ser colhidas da população apresentariam esse comportamento.

Para as amostras não usuais (i.e., aquelas que compõem esses 5%), o intervalo de confiança não contém a média populacional  $\mu$ .

# Intervalo de confiança para médias

Seja calcular o intervalo de confiança no nível de 95% para a média das alturas dos 100 alunos da amostra.

- $\bar{x} = 1,58$
- O desvio padrão populacional  $\sigma$  e conseqüentemente o desvio padrão da média amostral  $\sigma_{\bar{x}} = \sigma/\sqrt{100}$  são desconhecidos.
- Entretanto, dado que  $n > 30$ , podemos aproximar  $\sigma$  pela estimativa pontual do desvio padrão da amostra:  $\sigma \approx s = 0,2$ .



# Intervalo de confiança para médias

Seja calcular o intervalo de confiança no nível de 95% para a média das alturas dos 100 alunos da amostra.

- $\bar{x} = 1,58$
- O desvio padrão populacional  $\sigma$  e conseqüentemente o desvio padrão da média amostral  $\sigma_{\bar{x}} = \sigma/\sqrt{100}$  são desconhecidos.
- Entretanto, dado que  $n > 30$ , podemos aproximar  $\sigma$  pela estimativa pontual do desvio padrão da amostra:  $\sigma \approx s = 0,2$ .

Portanto, podemos calcular o intervalo de confiança da média populacional das alturas como

$$1,58 \pm (1,96)\left(\frac{0,2}{\sqrt{100}}\right) \rightarrow 1,58 \pm (1,96)(0,02) \rightarrow (1,54; 1,62)$$

# Intervalo de confiança para médias

A interpretação para o intervalo obtido é a seguinte: **há 95% de confiança de que a média populacional esteja acima de 1,54m e abaixo de 1,62m.**

# Intervalo de confiança para médias

A interpretação para o intervalo obtido é a seguinte: **há 95% de confiança de que a média populacional esteja acima de 1,54m e abaixo de 1,62m.**

Podemos afirmar com certeza que a média populacional esteja realmente dentro desse intervalo?

# Intervalo de confiança para médias

A interpretação para o intervalo obtido é a seguinte: **há 95% de confiança de que a média populacional esteja acima de 1,54m e abaixo de 1,62m.**

Podemos afirmar com certeza que a média populacional esteja realmente dentro desse intervalo?

Não! Em vez disso, o que podemos afirmar é que, se repetíssemos o processo de amostragem uma grande quantidade de vezes e calculássemos o intervalo de confiança correspondente, em aproximadamente 95% dos casos, a amostra conteria a média populacional.

# Intervalo de confiança para médias

Note que não há nada de especial com o valor 95% do nível de confiança usado no exemplo anterior.

- Em geral, qualquer nível de confiança pode ser usado.
- Por exemplo, se tivéssemos usado o valor 68%, teríamos obtido o intervalo de confiança (1,56; 1,60).

Repare também que o intervalo para o nível 95% é *maior* do que o intervalo para 68%.

- Esse é um resultado plausível?
- O que significa um intervalo de confiança no nível 100%?

# Intervalo de confiança para médias

Podemos agora definir formalmente o termo **nível de confiança**.

# Intervalo de confiança para médias

Podemos agora definir formalmente o termo **nível de confiança**.

## Nível de confiança

Seja  $\alpha$  um número tal que  $0 \leq \alpha \leq 1$ . O nível de confiança  $(1 - \alpha)$  é a probabilidade de que o intervalo construído contenha o verdadeiro valor da média populacional.

# Intervalo de confiança para médias

Podemos agora definir formalmente o termo **nível de confiança**.

## Nível de confiança

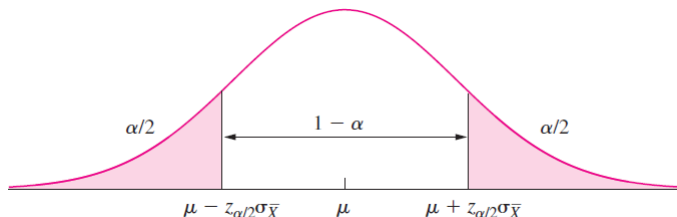
Seja  $\alpha$  um número tal que  $0 \leq \alpha \leq 1$ . O nível de confiança  $(1 - \alpha)$  é a probabilidade de que o intervalo construído contenha o verdadeiro valor da média populacional.

No exemplo anterior,  $1 - \alpha = 0,95$  é o nível de confiança. Em geral, é um valor definido pelo projetista.



# Intervalo de confiança para médias

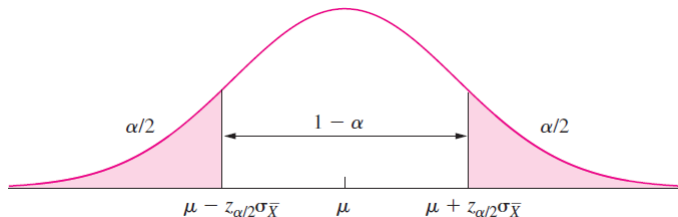
Considere que  $z_{\alpha/2}$  é o escore na distribuição normal padrão que delimita uma área de  $\alpha/2$  no lado direito da distribuição:



Por exemplo  $z_{0,025} = 1,96$ , pois 2,5% da área sob a curva normal padrão está à direita de 1,96.

# Intervalo de confiança para médias

Note que a região central da curva que possui área  $(1 - \alpha)$  corresponde ao intervalo  $\mu \pm z_{\alpha/2}\sigma_{\bar{X}}$ .



Em geral, para  $(1 - \alpha)\%$  de todas as amostras possíveis, o intervalo  $\bar{X} \pm z_{\alpha/2}\sigma_{\bar{X}}$  irá conter a média populacional  $\mu$ .

# Procedimento geral

O procedimento geral para a determinação de intervalos de confiança para  $\bar{X}$ , para amostras grandes, é resumido a seguir.

## Procedimento geral

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$  ( $n > 30$ ) de uma população cuja média é  $\mu$  e o desvio padrão é  $\sigma$ . Então um intervalo de confiança no nível  $100(1 - \alpha)\%$  para  $\mu$  é

$$\bar{X} \pm z_{\alpha/2} \sigma_{\bar{X}}$$

em que  $\sigma_{\bar{X}} = \sigma / \sqrt{n}$ .

# Procedimento geral

O procedimento geral para a determinação de intervalos de confiança para  $\bar{X}$ , para amostras grandes, é resumido a seguir.

## Procedimento geral

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$  ( $n > 30$ ) de uma população cuja média é  $\mu$  e o desvio padrão é  $\sigma$ . Então um intervalo de confiança no nível  $100(1 - \alpha)\%$  para  $\mu$  é

$$\bar{X} \pm z_{\alpha/2} \sigma_{\bar{X}}$$

em que  $\sigma_{\bar{X}} = \sigma / \sqrt{n}$ .

Quando o valor de  $\sigma$  é desconhecido, ele pode ser substituído pelo valor de  $s$ , a estimativa pontual do desvio padrão populacional.

# Exercício 01

Calculam-se a média e o desvio padrão para os pesos das caixas em uma amostra formada por 100 unidades, o que resulta em

- $\bar{x} = 12,05$ ,
- $s = 0,1$ .

Encontre um intervalo de confiança ao nível de 85% para a média dos pesos das caixas.

## Exercício 01 - solução

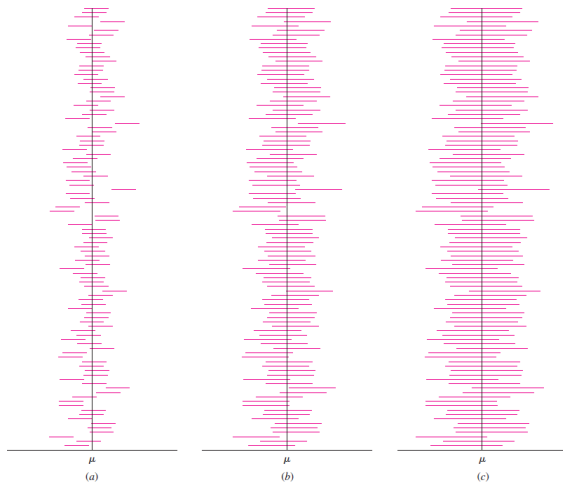
- O nível de confiança desejado é 85%, portanto  $1 - \alpha = 0,85$ , o que resulta em  $\alpha = 0,15$  e  $\alpha/2 = 0,075$ .
- Agora, procuramos pelo valor  $z_{0,075}$ , i.e., o valor que delimita 7,5% da área à direita da curva. Ao fazer isso, encontramos o valor  $z_{0,075} = 1,44$ .
- Vamos aproximar  $\sigma_{\bar{X}} \approx s/\sqrt{n} = 0,01$ .
- Portanto o intervalo de confiança ao nível de 85% é  $12,05 \pm (1,44)(0,01)$ .
- Esse intervalo também pode ser representado como  $12,05 \pm 0,0144$ , ou como  $(12,0356, 12,0644)$ .

# Confiabilidade x Precisão

Na ilustração seguinte, cada figura apresenta 100 intervalos de confiança (cada qual para uma amostra distinta) para a média populacional no nível de:

- (a) **68%**, que, embora precisos, não contêm a média da população em 32% das vezes, o que faz com que esse nível seja inaceitável para fins práticos.
- (b) **95%**, o que representa um bom compromisso entre precisão e confiabilidade para muitas finalidades práticas.
- (c) **99,7%**, o que significa que esses intervalos não cobrem a população em apenas três vezes em 1000. Esses intervalos são extremamente confiáveis, mas imprecisos.

# Confiabilidade x Precisão



Fonte da ilustração: Navidi, pág. 329



# Confiabilidade x Precisão

- O nível de confiança mais frequentemente utilizado na prática é de 95%.
- Para muitas aplicações, este nível oferece um bom compromisso entre confiabilidade e precisão.
- Níveis de confiança abaixo de 90% são raramente utilizados.
- Para algumas aplicações de controle de qualidade, nas quais a confiabilidade do produto é extremamente importante, são usados intervalos de confiança de níveis mais elevados, tais como 99,7%.

## Exercício 02

Um intervalo de confiança de 90% para o diâmetro médio (em centímetros) de hastes de aço fabricadas em uma certa máquina de extrusão é calculado como  $(14,73, 14,91)$ .

## Exercício 02

Um intervalo de confiança de 90% para o diâmetro médio (em centímetros) de hastes de aço fabricadas em uma certa máquina de extrusão é calculado como  $(14,73, 14,91)$ .

Verdadeiro ou falso: A probabilidade de que o diâmetro médio das hastes fabricadas por este processo está entre 14,73 e 14,91 é de 90%.

## Exercício 02 - solução

**Falso.** Um intervalo de confiança *para uma amostra específica* é dado. A média ou está nesse intervalo, ou não está. Estamos 90% confiantes de que a média da população está entre 14,73 e 14,91. O termo *probabilidade* é inadequado nesse caso.

## Exercício 03

No exemplo do peso de enchimento discutido anteriormente, o desvio padrão da amostra de pesos de 100 caixas foi  $s = 0,1$  onças.

## Exercício 03

No exemplo do peso de enchimento discutido anteriormente, o desvio padrão da amostra de pesos de 100 caixas foi  $s = 0,1$  onças.

Quantas caixas devem ser amostradas para se obter um intervalo de confiança de 99% da largura  $\pm 0,012$  onças?

## Exercício 03 - solução

O nível é de 99%, de modo  $1 - \alpha = 0,99$ . Portanto  $\alpha = 0,01$  e  $z_{\alpha/2} = 2,58$ . o valor de  $s$  é estimado com  $s = 0,1$ . O tamanho da amostra necessária é encontrada resolvendo a equação

$$(2,58)(0,1)/\sqrt{n} = 0,012.$$

Ao resolver essa equação, obtemos  $n \approx 463$ .

## Exercício 03 - solução

O nível é de 99%, de modo  $1 - \alpha = 0,99$ . Portanto  $\alpha = 0,01$  e  $z_{\alpha/2} = 2,58$ . o valor de  $s$  é estimado com  $s = 0,1$ . O tamanho da amostra necessária é encontrada resolvendo a equação

$$(2,58)(0,1)/\sqrt{n} = 0,012.$$

Ao resolver essa equação, obtemos  $n \approx 463$ .

### Situação geral

Se o nível de confiança  $100(1 - \alpha)\%$  é especificado, podemos procurar pelo valor  $z_{\alpha/2}$  na tabela para  $N(0, 1)$ . Se o desvio padrão da população  $\sigma$  também é especificado, então podemos calcular o valor de  $n$  necessário para produzir uma largura especificada.

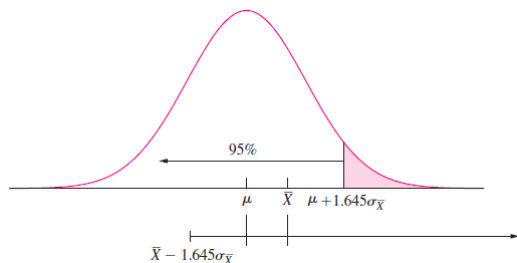


# Intervalos de confiança unilaterais

- Os intervalos de confiança discutidos até agora foram *bilaterais*, posto que especificam dois limites de confiança, superior e outro inferior. Ocasionalmente, estamos interessados em apenas um destes limites.
- Por exemplo, suponha que um engenheiro quer estimar a média de força de esmagamento de um tipo de bloco de concreto.
- O engenheiro estará provavelmente interessado em apenas um **limite inferior** para a força, uma vez que as especificações para várias aplicações irão geralmente especificar apenas uma resistência mínima.
- Nestes casos, os **intervalos de confiança unilaterais** são adequados.

# Intervalos de confiança unilaterais

A média amostral  $\bar{X}$  é extraída de uma distribuição normal com média  $\mu$  e desvio padrão  $\sigma$ . Para essa amostra,  $\bar{X}$  vem da área 95% inferior da distribuição, de modo que o intervalo de confiança unilateral de 95%  $(\bar{X} - 1,645\sigma_{\bar{X}}, \infty)$  contém a média populacional  $\mu$ .



# Procedimento geral

O procedimento geral para a determinação de intervalos unilaterais de confiança para  $\bar{X}$ , para amostras grandes, é resumido a seguir.

## Procedimento geral

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$  ( $n > 30$ ) de uma população cuja média é  $\mu$  e o desvio padrão é  $\sigma$ . Então o limite de confiança inferior no nível  $100(1 - \alpha)\%$  para  $\mu$  é

$$\bar{X} - z_{\alpha}\sigma_{\bar{X}}$$

e o limite de confiança superior no nível  $100(1 - \alpha)\%$  para  $\mu$  é

$$\bar{X} + z_{\alpha}\sigma_{\bar{X}}$$

em que  $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ . Quando  $\sigma$  é desconhecido, ele pode ser substituído pelo valor de  $s$ , a estimativa pontual do desvio padrão.

## Exercício 05

Em uma amostra de 50 microfuradeiras, o tempo de vida útil médio (expresso como o número de orifícios perfurados antes da primeira falha) foi 12,68 com um desvio padrão de 6,83. Determine

- um limite de confiança inferior de 95% para o tempo de vida útil média das microfuradeiras;
- um limite de confiança superior de 99% para o tempo de vida útil média das microfuradeiras.

## Exercício 05 - solução

A média e o desvio padrão da amostra são  $\bar{X} = 12,68$  e  $s = 6,83$ , respectivamente. O tamanho da amostra é  $n = 50$ . Estimamos  $\sigma_{\bar{X}} \approx s/\sqrt{n} = 0,9659$ .

- a) O limite de confiança inferior de 95% é  $\bar{X} - 1,645\sigma_{\bar{X}} = 11,09$ .
- b) O limite de confiança superior de 99% é  $\bar{X} + 2,33\sigma_{\bar{X}} = 14,93$ .

- 1 Intervalo de confiança para médias
- 2 Intervalo de confiança para proporções

# Intervalo de confiança para proporções

De forma análoga ao que foi feito para médias, há também procedimento geral para determinar intervalos de confiança para proporções.

# Intervalo de confiança para proporções

De forma análoga ao que foi feito para médias, há também procedimento geral para determinar intervalos de confiança para proporções.

Regra prática: o procedimento a seguir deve ser usado apenas quando a amostra contém pelo menos 10 sucessos e 10 falhas.



# Intervalo de confiança para proporções

Seja  $X$  a variável aleatória que conta a quantidade de sucessos em uma amostra de tamanho  $n$ . Sendo assim,  $X \sim \text{Bin}(n, p)$ , e a estimativa para  $p$  (proporção populacional) é  $\hat{p} = X/n$ .

# Intervalo de confiança para proporções

Seja  $X$  a variável aleatória que conta a quantidade de sucessos em uma amostra de tamanho  $n$ . Sendo assim,  $X \sim \text{Bin}(n, p)$ , e a estimativa para  $p$  (proporção populacional) é  $\hat{p} = X/n$ .

O desvio padrão da variável  $\hat{p}$  é  $\sqrt{p(1-p)/n}$ .

# Intervalo de confiança para proporções

Seja  $X$  a variável aleatória que conta a quantidade de sucessos em uma amostra de tamanho  $n$ . Sendo assim,  $X \sim \text{Bin}(n, p)$ , e a estimativa para  $p$  (proporção populacional) é  $\hat{p} = X/n$ .

O desvio padrão da variável  $\hat{p}$  é  $\sqrt{p(1-p)/n}$ .

Considerando que o tamanho da amostra é grande, pelo TLC, temos que:

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

# Intervalo de confiança para proporções

Se considerarmos o nível de confiança de 95%, então a proporção populacional satisfaz a seguinte desigualdade:

$$\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + 1,96\sqrt{\frac{p(1-p)}{n}}$$

# Intervalo de confiança para proporções

Se considerarmos o nível de confiança de 95%, então a proporção populacional satisfaz a seguinte desigualdade:

$$\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + 1,96\sqrt{\frac{p(1-p)}{n}}$$

Para obter o intervalo de confiança, a abordagem tradicional recomenda substituir  $p$  por  $\hat{p}$ . Sendo assim, obtemos o intervalo de confiança a seguir:

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

# Intervalo de confiança para proporções

## Procedimento geral

Seja  $\hat{p}$  a proporção de sucessos em um grande número  $n$  de ensaios independentes de Bernoulli com probabilidade de sucesso  $p$ .

Então, o intervalo de confiança no nível  $100(1 - \alpha)\%$  para  $p$  é

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$