

MÉTODOS ESTATÍSTICOS

Elementos da Teoria da Estimação

Licenciatura em Engenharia Informática

Departamento de Matemática
Escola Superior de Tecnologia de Setúbal
Instituto Politécnico de Setúbal
2023-2024

Teoria da Estimação

A Estimação é o capítulo da Inferência Estatística que vai permitir avaliar o valor de parâmetros (desconhecidos) da População.

Consiste em utilizar amostras para estimar (ou prever) os valores de parâmetros populacionais desconhecidos.

Teoria da Estimação

Seja X uma característica em estudo numa população e suponha que X tem uma dada distribuição de probabilidade com um parâmetro θ desconhecido.

Estimador

Estimador $\hat{\theta}$ é uma função construída com base numa amostra aleatória sem parâmetros de valor desconhecido e é usado para avaliar ou estimar o parâmetro θ .

Estimativa

Estimativa de um parâmetro de uma população é qualquer valor específico de um estimador desse parâmetro.

Estimação

Estimação é todo o processo que se baseia em utilizar um estimador para produzir uma estimativa de um parâmetro.

Tipo de Estimação

- Estimação pontual,
- Estimação por intervalos.

Teoria da Estimação

Seja X uma característica em estudo numa população e suponha que X tem uma dada distribuição de probabilidade com um parâmetro θ desconhecido.

Estimador

Estimador $\hat{\theta}$ é uma função construída com base numa amostra aleatória sem parâmetros de valor desconhecido e é usado para avaliar ou estimar o parâmetro θ .

Estimativa

Estimativa de um parâmetro de uma população é qualquer valor específico de um estimador desse parâmetro.

Estimação

Estimação é todo o processo que se baseia em utilizar um estimador para produzir uma estimativa de um parâmetro.

Tipo de Estimação

- **Estimação pontual**,
- **Estimação por intervalos**.

Estimação Pontual

A **estimação pontual** consiste em, a partir de uma amostra e usando um estimador, encontrar um valor numérico único (estimativa) que esteja bastante próximo do verdadeiro valor do parâmetro.

Portanto, **estimação pontual** é uma estimativa de um parâmetro populacional obtida por um único número.

Pretende-se que a estimativa esteja bastante próxima do verdadeiro valor do parâmetro, no entanto este procedimento não permite julgar a magnitude do erro que poderá estar a ser cometido.

Ou seja, a **estimação pontual** permite localizar pontualmente cada parâmetro da população, fornece um valor simples que está altamente sujeito a erro, não permite uma avaliação da precisão do estimador.

Observação: Em geral, um estimador de um parâmetro θ representa-se por $\hat{\theta}$ ou por θ^* .

Estimação Pontual

Existem vários métodos para determinar estimadores pontuais:

- método dos Momentos,
- método de máxima verosimilhança,
- método dos mínimos quadrados,
- métodos Bayesianos.

Não vamos falar sobre estes métodos, apenas referir que existem, e vamos utilizar como estimadores pontuais o correspondente amostral do parâmetro que se pretende estimar:

Parâmetros populacionais	Estimador Pontual mais usual
média populacional $\mu = E [X]$	média amostral $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
variância populacional $\sigma^2 = V [X]$	variância amostral $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
proporção populacional $p = \frac{\text{nº casos favoráveis na População}}{\text{nº casos possíveis na População}}$	proporção amostral $\hat{p} = p^* = \frac{\text{nº casos favoráveis na Amostra}}{\text{nº casos possíveis na Amostra}}$

Estimação Pontual

Em relação às distribuições de probabilidade estudadas tem-se:

População / Parâmetros	Estimador Pontual mais usual
$X \sim \text{Binomial}$ $p = \frac{\text{nº casos favoráveis na População}}{\text{nº casos possíveis na População}}$	$\hat{p} = p^* = \frac{\text{nº casos favoráveis na Amostra}}{\text{nº casos possíveis na Amostra}}$
$X \sim P(\lambda)$ $\lambda = E[X]$	$\hat{\lambda} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
$X \sim N(\mu, \sigma)$ $\mu = E[X]$ $\sigma^2 = V[X]$	$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
$X \sim Exp(\theta)$ $\theta = E[X]$	$\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Exemplo 1

Considere uma população Exponencial de média θ desconhecida e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 3$: $(11, 15, 13)$.

- 1 Indique um estimador para a média da população, justificando a sua escolha.

Exemplo 1

Considere uma população Exponencial de média θ desconhecida e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 3$: $(11, 15, 13)$.

- Indique um estimador para a média da população, justificando a sua escolha.

Como temos uma População $X \sim Exp(\theta)$ e sabemos que $E[X] = \theta$, então um possível estimador para θ é a média amostral: $\hat{\theta} = \bar{X}$.

Exemplo 1

Considere uma população Exponencial de média θ desconhecida e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 3$: $(11, 15, 13)$.

- Indique um estimador para a média da população, justificando a sua escolha.

Como temos uma População $X \sim Exp(\theta)$ e sabemos que $E[X] = \theta$, então um possível estimador para θ é a média amostral: $\hat{\theta} = \bar{X}$.

- Indique uma estimativa para a média da população.

Exemplo 1

Considere uma população Exponencial de média θ desconhecida e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 3$: $(11, 15, 13)$.

- Indique um estimador para a média da população, justificando a sua escolha.

Como temos uma População $X \sim Exp(\theta)$ e sabemos que $E[X] = \theta$, então um possível estimador para θ é a média amostral: $\hat{\theta} = \bar{X}$.

- Indique uma estimativa para a média da população.

Uma estimativa para θ (média da população) é $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{11+15+13}{3} = 13$.

Exemplo 1

Considere uma população Exponencial de média θ desconhecida e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 3$: $(11, 15, 13)$.

- Indique um estimador para a média da população, justificando a sua escolha.

Como temos uma População $X \sim Exp(\theta)$ e sabemos que $E[X] = \theta$, então um possível estimador para θ é a média amostral: $\hat{\theta} = \bar{X}$.

- Indique uma estimativa para a média da população.

Uma estimativa para θ (média da população) é $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{11+15+13}{3} = 13$.

- Qual o processo de estimação que usou?

Exemplo 1

Considere uma população Exponencial de média θ desconhecida e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 3$: $(11, 15, 13)$.

- Indique um estimador para a média da população, justificando a sua escolha.

Como temos uma População $X \sim Exp(\theta)$ e sabemos que $E[X] = \theta$, então um possível estimador para θ é a média amostral: $\hat{\theta} = \bar{X}$.

- Indique uma estimativa para a média da população.

Uma estimativa para θ (média da população) é $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{11+15+13}{3} = 13$.

- Qual o processo de estimação que usou?

Estimação pontual.

Exemplo 2

Considere uma população Normal de média μ e desvio padrão σ desconhecidos e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 4$: $(5, 7, 4, 2)$.

- Indique um estimador para a variância da população, justificando a sua escolha.

Exemplo 2

Considere uma população Normal de média μ e desvio padrão σ desconhecidos e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 4$: $(5, 7, 4, 2)$.

- Indique um estimador para a variância da população, justificando a sua escolha.

Como temos uma População $X \sim N(\mu, \sigma^2)$ e sabemos que $V[X] = \sigma^2$, então um possível estimador para σ^2 é a variância amostral: $\hat{\sigma}^2 = S^2$.

Exemplo 2

Considere uma população Normal de média μ e desvio padrão σ desconhecidos e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 4$: $(5, 7, 4, 2)$.

- Indique um estimador para a variância da população, justificando a sua escolha.

Como temos uma População $X \sim N(\mu, \sigma^2)$ e sabemos que $V[X] = \sigma^2$, então um possível estimador para σ^2 é a variância amostral: $\hat{\sigma}^2 = S^2$.

- Indique uma estimativa para a variância da população.

Exemplo 2

Considere uma população Normal de média μ e desvio padrão σ desconhecidos e suponha que foi recolhida a seguinte amostra aleatória de dimensão $n = 4$: $(5, 7, 4, 2)$.

- Indique um estimador para a variância da população, justificando a sua escolha.

Como temos uma População $X \sim N(\mu, \sigma^2)$ e sabemos que $V[X] = \sigma^2$, então um possível estimador para σ^2 é a variância amostral: $\hat{\sigma}^2 = S^2$.

- Indique uma estimativa para a variância da população.

Para calcular a estimativa da variância é necessário calcular a média amostral:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{5 + 7 + 4 + 2}{4} = 4.5$$

Uma estimativa para σ^2 (variância da população) é

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1} = \frac{(5^2 + 7^2 + 4^2 + 2^2) - 4 \times 4.5^2}{4-1} = 4.333$$

Teoria da Estimação

Estimador

Estimador $\hat{\theta}$ é uma função construída com base numa amostra aleatória sem parâmetros de valor desconhecido e é usado para avaliar ou estimar o parâmetro θ .

Estimativa

Estimativa de um parâmetro de uma população é qualquer valor específico de um estimador desse parâmetro.

Estimação

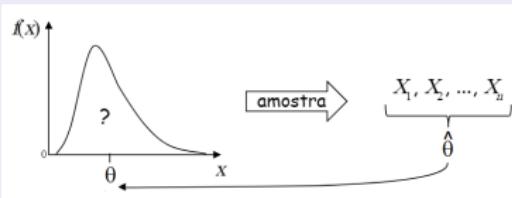
Estimação é todo o processo que se baseia em utilizar um estimador para produzir uma estimativa de um parâmetro.

Tipo de Estimação

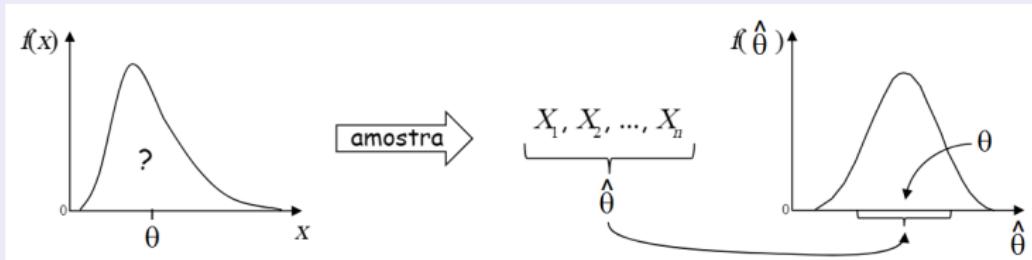
- Estimação pontual,
- Estimação por intervalos.

Tipo de Estimação

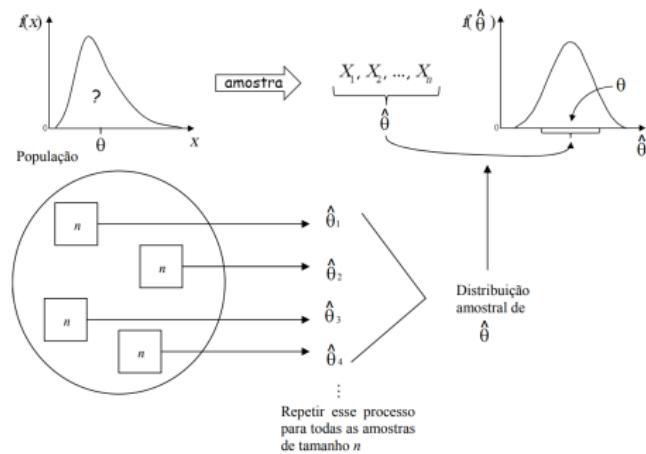
- Estimação pontual** - Permite localizar pontualmente cada parâmetro da população. Fornece um valor simples que está altamente sujeito a erro, não permite uma avaliação da precisão do estimador.



- Estimação por intervalos** - A estimação por intervalos acrescenta à informação pontual uma banda de incerteza que lhe está associada (através de um intervalo). A este processo de estimação chama-se, habitualmente, **Intervalos de confiança**.



Estamos a supor que temos uma dada população X que segue uma dada distribuição de probabilidade $f(x)$ e nessa distribuição de probabilidade existe uma parâmetro θ desconhecido, então vamos recolher uma amostra (X_1, X_2, \dots, X_n) dessa população e com base nessa amostra vamos construir a estatística $\hat{\theta}$, isto é, um estimador para θ . No entanto, como vimos no capítulo 3, os estimadores são variáveis aleatórias (diferentes amostras dão origem a diferentes estimativas) logo têm uma distribuição amostral.



Portanto, com base no estimador definido, vamos calcular uma estimativa pontual para o parâmetro θ e com base na distribuição amostral vamos construir um intervalo em torno dessa estimativa pontual: $[\hat{\theta} - d, \hat{\theta} + d]$. Como este intervalo vai ser calculado com base na distribuição amostral do estimador, então vai ter uma probabilidade associada e d representa o erro associado ao intervalo. Agora de uma forma mais formal:

Estimação por Intervalos ou Intervalos de Confiança

Seja X uma variável aleatória que representa uma população e θ um parâmetro desconhecido dessa população. Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão n dessa população e $\hat{\theta}$ um estimador de θ . Partindo da distribuição amostral de $\hat{\theta}$ calcula-se

$$P(a < \hat{\theta} < b) = 1 - \alpha$$

de modo a construir dois estimadores, L_1 e L_2 , para θ tais que:

$$L_1 \leq L_2 \quad \text{e} \quad P(L_1 < \theta < L_2) = 1 - \alpha$$

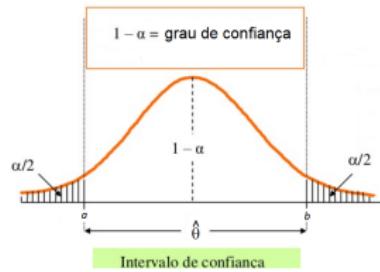
então

- ao intervalo $]l_1, l_2[$ (l_1 e l_2 são estimativas de L_1 e L_2) chama-se **intervalo de confiança a $(1 - \alpha) \times 100\%$ para θ** ;
- $(1 - \alpha)$ é o coeficiente ou **grau de confiança** do intervalo;
- α é o coeficiente ou grau de erro inerente ao intervalo e chama-se **nível de significância**;
- aos extremos do intervalo, l_1 e l_2 , chamam-se **limites de confiança inferior e superior**, respetivamente;
- a **amplitude** do intervalo de confiança $= l_2 - l_1$;
- a **margem de erro** do intervalo de confiança $= \frac{l_2 - l_1}{2}$, ou seja, metade da amplitude do intervalo de confiança.

Para chegar aos estimadores L_1 e L_2 é necessário começar por definir quem serão a e b tais que

$$P(a < \hat{\theta} < b) = 1 - \alpha.$$

Como já foi referido, é a distribuição amostral do estimador $\hat{\theta}$ que irá permitir calcular a probabilidade pretendida e, como vimos no capítulo 3, só vamos trabalhar com distribuições de variáveis aleatórias contínuas. Portanto, pretende-se apenas determinar a e b tal que a área dentro do intervalo $[a, b]$ seja $1 - \alpha$. Existe uma infinidade de intervalos nessas condições, mas pretende-se o intervalo que tenha a **menor amplitude**. O cálculo do intervalo de confiança foi "pensado" para a distribuição Normal, portanto o intervalo com a menor amplitude e que tenha associado uma probabilidade de $1 - \alpha$ é o intervalo que coloca a parte mais alta da distribuição no centro do intervalo.



Todos os intervalos de confiança serão definidos desta forma, o grau de confiança no centro da distribuição e metade do nível de significância nas caudas da distribuição, logo $a = \text{quantil de probabilidade } \frac{\alpha}{2}$ e $b = \text{quantil de probabilidade } 1 - \frac{\alpha}{2}$.

Em resumo, para determinar um **intervalo de confiança a $(1 - \alpha) \times 100\%$ para θ** basta seguir os seguintes passos:

Metodologia para a Construção de um Intervalo de Confiança para um parâmetro θ desconhecido

- ① Escolher um estimador pontual para o parâmetro θ : $\hat{\theta}$.
- ② Determinar a distribuição amostral desse estimador.
- ③ Fixar o nível de significância (α) e deduzir o intervalo de confiança:

$$\textcircled{1} \quad P(a < \hat{\theta} < b) = 1 - \alpha$$

com $a = \text{quantil de probabilidade } \frac{\alpha}{2}$ e $b = \text{quantil de probabilidade } 1 - \frac{\alpha}{2}$.

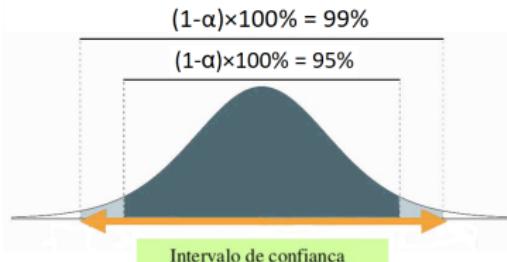
$$\textcircled{2} \quad P(a < \hat{\theta} < b) = 1 - \alpha \Leftrightarrow \dots \Leftrightarrow P(L_1 < \theta < L_2) = 1 - \alpha$$

- ④ Com base na amostra recolhida, um intervalo de confiança a $(1 - \alpha) \times 100\%$ para θ é $]l_1, l_2[$, onde l_1 e l_2 são estimativas de L_1 e L_2 .

Estimação por Intervalos ou Intervalos de Confiança

Como é óbvio, a estimação de parâmetros através de um intervalo de confiança tem a vantagem de associar uma probabilidade mas tem a desvantagem de dar um conjunto de valores como hipótese para o parâmetro. O ideal é ter um intervalo de confiança com a maior confiança possível e ao mesmo tempo muito preciso (no sentido de ser pouco vago, com poucos valores), mas conjugar os dois não é fácil.

- Um intervalo de confiança tem uma amplitude inversamente proporcional à dimensão da amostra, ou seja, **quanto maior a dimensão da amostra menor a amplitude do intervalo**. No limite, quando n coincide com a dimensão da população, o intervalo reduz-se a um único ponto, isto é, o parâmetro é conhecido com exatidão.
- Se a dimensão da amostra é fixa, a amplitude do intervalo é proporcional ao grau de confiança atribuído, ou seja, **quanto maior o grau de confiança do intervalo maior a amplitude logo mais vago é o intervalo**.



Observação: O valor mais usual para o grau de confiança é 95%.

Estimação por Intervalos ou Intervalos de Confiança

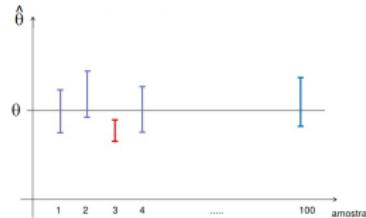
Um intervalo de confiança a $(1 - \alpha) \times 100\%$ para θ é

$$]l_1, l_2[$$

onde l_1 e l_2 são estimativas de L_1 e L_2 .

Outro aspecto é a interpretação de um intervalo de confiança. Não esquecer que, no final um intervalo de confiança é uma estimativa, ou seja, é a concretização para uma única amostra. Com outra amostra teríamos outro intervalo de confiança. Portanto, dizer que a um intervalo está associado $(1 - \alpha) \times 100\%$ de confiança significa que tem-se $(1 - \alpha) \times 100\%$ de hipóteses desse ser o intervalo que contém o verdadeiro valor do parâmetro desconhecido.

Por exemplo, suponha que vai recolher 100 amostras de dimensão n e vai construir 100 intervalos de confiança a 99% para θ . Atribuir uma confiança de 99% significa que espera ter 99 intervalos que contêm o verdadeiro valor de θ e apenas 1 que não contém.



Como na prática não recolhe 100 amostras mas apenas recolhe 1, não sabe qual dos intervalos construiu, logo tem 99% de hipóteses do intervalo que construiu ser um dos que contém θ .

Estimação por Intervalos ou Intervalos de Confiança

De acordo com as distribuições amostrais estudadas, podem-se definir os seguintes intervalos de confiança:

- Intervalo de Confiança para a média μ ;
- Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$;
- Intervalo de Confiança para a proporção p ;
- Intervalo de Confiança para a diferença de proporções $p_1 - p_2$;
- Intervalo de Confiança para a variância σ^2 ;
- Intervalo de Confiança para o quociente de variâncias $\frac{\sigma_1^2}{\sigma_2^2}$.

Intervalo de Confiança para a média μ (com σ conhecido)

Seja X uma **população Normal** com média μ e desvio padrão σ que é **conhecido** e (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão n dessa população.

Estimador pontual para μ : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Distribuição Amostral

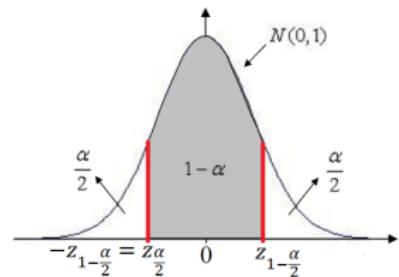
ou $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

Variável fulcral

Intervalo de Confiança para a média μ (com σ conhecido)

Dedução do intervalo de confiança:

$$\begin{aligned} P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha &\Leftrightarrow \\ \Leftrightarrow P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha &\Leftrightarrow \\ \Leftrightarrow P\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha & \end{aligned}$$



Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Intervalo de Confiança para a média μ (com σ conhecido)

Seja X uma **população Qualquer** (pode não ser Normal ou ser desconhecida) com média μ e desvio padrão σ que é **conhecido** e (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão $n \geq 30$ dessa população.

Estimador pontual para μ : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Como $n \geq 30$, então pelo Teorema do Limite Central tem-se

Distribuição Amostral

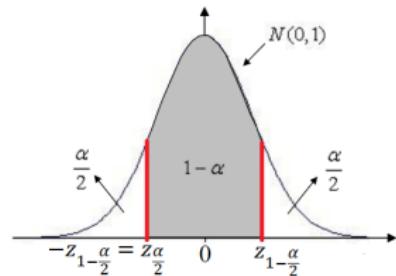
ou $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \stackrel{d}{\sim} N(0, 1)$

Variável fulcral

Intervalo de Confiança para a média μ (com σ conhecido)

Dedução do intervalo de confiança:

$$\begin{aligned} P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha &\Leftrightarrow \\ \Leftrightarrow P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha &\Leftrightarrow \\ \Leftrightarrow P\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha & \end{aligned}$$



Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Intervalo de Confiança para a média μ (com σ desconhecido)

Seja X uma **população Normal** com média μ e desvio padrão σ que é **desconhecido** e (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão n dessa população.

Estimador pontual para μ : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Estimador pontual para σ^2 : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Distribuição Amostral

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

Variável fulcral

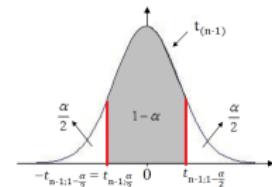
Intervalo de Confiança para a média μ (com σ desconhecido)

Dedução do intervalo de confiança:

$$P\left(-t_{1-\frac{\alpha}{2};n-1} < T < t_{1-\frac{\alpha}{2};n-1}\right) = 1 - \alpha \Leftrightarrow$$

$$\Leftrightarrow P\left(-t_{1-\frac{\alpha}{2};n-1} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{1-\frac{\alpha}{2};n-1}\right) = 1 - \alpha \Leftrightarrow$$

$$\Leftrightarrow P\left(\bar{X} - t_{1-\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2};n-1} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$



Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ

$$\left[\bar{x} - t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right]$$

Intervalo de Confiança para a média μ (com σ desconhecido)

Seja X uma **população Qualquer** (pode não ser Normal ou ser desconhecida) com média μ e desvio padrão σ que é **desconhecido** e (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão $n \geq 30$ dessa população.

Estimador pontual para μ : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Estimador pontual para σ^2 : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Como $n \geq 30$, então pelo Teorema do Limite Central tem-se

Distribuição Amostral

ou

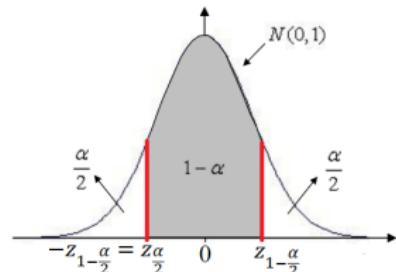
$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \stackrel{d}{\sim} N(0, 1)$$

Variável fulcral

Intervalo de Confiança para a média μ (com σ desconhecido)

Dedução do intervalo de confiança:

$$\begin{aligned} P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha &\Leftrightarrow \\ \Leftrightarrow P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha &\Leftrightarrow \\ \Leftrightarrow P\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = 1 - \alpha & \end{aligned}$$



Intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

Em resumo ([Formulário](#)):

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para a média μ

Condições		
População	Parâmetros e Amostra	Intervalo de Confiança
Normal	σ conhecido	$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$
	σ desconhecido	$\left[\bar{x} - t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right]$
Qualquer	σ conhecido e $n \geq 30$	$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$
	σ desconhecido e $n \geq 30$	$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$

Observação:

Quando a População é Normal e $n \geq 30$, em vez $\left[\bar{x} - t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right]$ pode ser usado $\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$

Intervalo de Confiança para a diferença de médias

$$\mu_1 - \mu_2$$

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 conhecidos)

Sejam X_1 e X_2 variáveis aleatórias que representam **populações Normais**, respetivamente, com médias μ_1 e μ_2 e com desvios padrão σ_1 e σ_2 que são **conhecidos**. Selecionaram-se duas **amostras aleatórias independentes** de dimensões n_1 (da população X_1) e n_2 (da população X_2).

Estimador pontual para μ_1 : \bar{X}_1

Estimador pontual para μ_2 : \bar{X}_2

Distribuição Amostral

ou

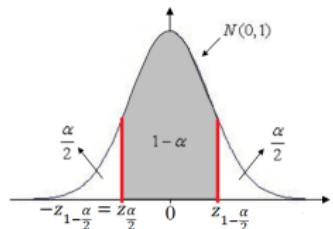
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Variável fulcral

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 conhecidos)

Dedução do intervalo de confiança:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(\left(\bar{X}_1 - \bar{X}_2\right) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{X}_1 - \bar{X}_2\right) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$

$$\left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 conhecidos)

Sejam X_1 e X_2 variáveis aleatórias que representam **populações Quaisquer** (podem não ser Normais ou serem desconhecidas), respetivamente, com médias μ_1 e μ_2 e com desvios padrão σ_1 e σ_2 que são **conhecidos**. Selecionaram-se duas **amostras aleatórias independentes** de dimensões $n_1 \geq 30$ (da população X_1) e $n_2 \geq 30$ (da população X_2).

Estimador pontual para μ_1 : \bar{X}_1

Estimador pontual para μ_2 : \bar{X}_2

Como $n_1 \geq 30$ e $n_2 \geq 30$, então pelo Teorema do Limite Central tem-se

Distribuição Amostral

ou

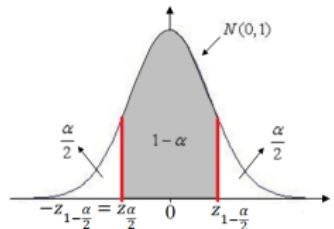
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Variável fulcral

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 conhecidos)

Dedução do intervalo de confiança:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(\left(\bar{X}_1 - \bar{X}_2\right) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{X}_1 - \bar{X}_2\right) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$

$$\left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 desconhecidos)

Sejam X_1 e X_2 variáveis aleatórias que representam **populações Normais**, respectivamente, com médias μ_1 e μ_2 e com desvios padrão σ_1 e σ_2 que são **desconhecidos**. Selecionaram-se duas **amostras aleatórias independentes** de dimensões n_1 (da população X_1) e n_2 (da população X_2).

Estimador pontual para μ_1 : \bar{X}_1
Estimador pontual para σ_1^2 : S_1^2

Estimador pontual para μ_2 : \bar{X}_2
Estimador pontual para σ_2^2 : S_2^2

Se $\sigma_1^2 = \sigma_2^2$

Distribuição Amostral

ou

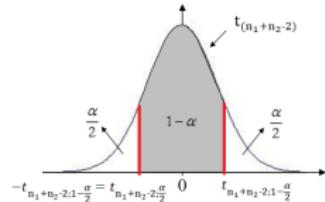
$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}} \sim t_{(n_1+n_2-2)}$$

Variável fulcral

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 desconhecidos)

Dedução do intervalo de confiança:

$$P\left(-t_{1-\frac{\alpha}{2};n_1+n_2-2} < T < t_{1-\frac{\alpha}{2};n_1+n_2-2}\right) = 1-\alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(\overline{X}_1 - \overline{X}_2 - t_{1-\frac{\alpha}{2};n_1+n_2-2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} < \mu_1 - \mu_2 < \overline{X}_1 - \overline{X}_2 + t_{1-\frac{\alpha}{2};n_1+n_2-2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$

$$\left[(\overline{X}_1 - \overline{X}_2) - t_{1-\frac{\alpha}{2};n_1+n_2-2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}, (\overline{X}_1 - \overline{X}_2) + t_{1-\frac{\alpha}{2};n_1+n_2-2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \right]$$

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 desconhecidos)

Sejam X_1 e X_2 variáveis aleatórias que representam **populações Normais**, respetivamente, com médias μ_1 e μ_2 e com desvios padrão σ_1 e σ_2 que são **desconhecidos**. Selecionaram-se duas **amostras aleatórias independentes** de dimensões n_1 (da população X_1) e n_2 (da população X_2).

Estimador pontual para μ_1 : \bar{X}_1
Estimador pontual para σ_1^2 : S_1^2

Estimador pontual para μ_2 : \bar{X}_2
Estimador pontual para σ_2^2 : S_2^2

Se $\sigma_1^2 \neq \sigma_2^2$

Distribuição Amostral

ou

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(gl_2)}$$

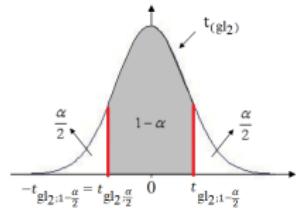
Variável fulcral

$$\text{com } gl_2 \approx \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}, \text{ habitualmente considera-se o inteiro mais próximo ou faz-se a correção de Welch-Satterthwaite}$$

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 desconhecidos)

Dedução do intervalo de confiança:

$$P\left(-t_{1-\frac{\alpha}{2};gl_2} < T < t_{1-\frac{\alpha}{2};gl_2}\right) = 1-\alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(\left(\bar{X}_1 - \bar{X}_2\right) - t_{1-\frac{\alpha}{2};gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{X}_1 - \bar{X}_2\right) + t_{1-\frac{\alpha}{2};gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$

$$\left[(\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2};gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2};gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

$$\text{com } gl_2 \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}, \text{ habitualmente considera-se o inteiro mais próximo ou faz-se a correção de Welch-Satterthwaite}$$

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 desconhecidos)

Sejam X_1 e X_2 variáveis aleatórias que representam **populações Quaisquer** (podem não ser Normais ou serem desconhecidas), respetivamente, com médias μ_1 e μ_2 e com desvios padrão σ_1 e σ_2 que são **desconhecidos**. Selecionaram-se duas **amostras aleatórias independentes** de dimensões $n_1 \geq 30$ (da população X_1) e $n_2 \geq 30$ (da população X_2

Estimador pontual para μ_1 : \bar{X}_1
Estimador pontual para σ_1^2 : S_1^2

Estimador pontual para μ_2 : \bar{X}_2
Estimador pontual para σ_2^2 : S_2^2

Como $n_1 \geq 30$ e $n_2 \geq 30$, então pelo teorema do Limite Central tem-se

Distribuição Amostral

ou

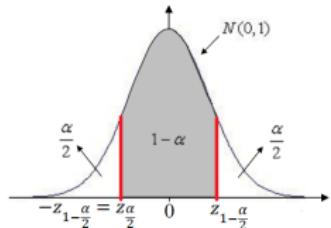
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

Variável fulcral

Intervalo de Confiança para a diferença de médias $\mu_1 - \mu_2$ (com σ_1 e σ_2 desconhecidos)

Dedução do intervalo de confiança:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(\left(\bar{X}_1 - \bar{X}_2\right) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{X}_1 - \bar{X}_2\right) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\mu_1 - \mu_2$

$$\left[\left(\bar{x}_1 - \bar{x}_2\right) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \left(\bar{x}_1 - \bar{x}_2\right) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

Em resumo (**Formulário**):

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para a diferença de médias $\mu_1 - \mu_2$

Condições	Intervalo de confiança
Populações Normais σ_1 e σ_2 conhecidos Amostras independentes	$\left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$
Populações Normais σ_1 e σ_2 desconhecidos $\sigma_1 = \sigma_2$ Amostras independentes	$\left[(\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}; n_1+n_2-2} \sqrt{(\frac{1}{n_1} + \frac{1}{n_2}) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}; n_1+n_2-2} \sqrt{(\frac{1}{n_1} + \frac{1}{n_2}) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \right]$
Populações Normais σ_1 e σ_2 desconhecidos $\sigma_1 \neq \sigma_2$ Amostras independentes	$\left[(\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}; gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}; gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$
Populações Quaisquer σ_1 e σ_2 conhecidos Amostras independentes $n_1 \geq 30$ e $n_2 \geq 30$	$\left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$
Populações Quaisquer σ_1 e σ_2 desconhecidos Amostras independentes $n_1 \geq 30$ e $n_2 \geq 30$	$\left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$

Intervalo de Confiança para a variância σ^2

Intervalo de Confiança para a variância σ^2

Seja X uma **população Normal** com média μ (desconhecida) e desvio padrão σ e (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão n dessa população.

Estimador pontual para σ^2 : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Distribuição Amostral

ou $X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

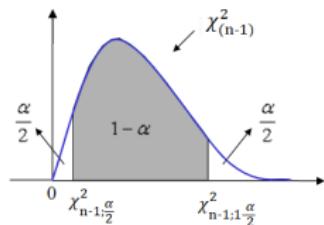
Variável fulcral

Intervalo de Confiança para a variância σ^2

Dedução do intervalo de confiança:

$$P\left(x_{\frac{\alpha}{2};n-1}^2 < X^2 < x_{1-\frac{\alpha}{2};n-1}^2\right) = 1 - \alpha \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{(n-1)S^2}{x_{1-\frac{\alpha}{2};n-1}^2} < \sigma^2 < \frac{(n-1)S^2}{x_{\frac{\alpha}{2};n-1}^2}\right) = 1 - \alpha$$



Intervalo de confiança a $(1 - \alpha) \times 100\%$ para σ^2

$$\left[\frac{(n-1)s^2}{x_{1-\frac{\alpha}{2};n-1}^2}, \frac{(n-1)s^2}{x_{\frac{\alpha}{2};n-1}^2} \right]$$

Intervalo de Confiança para a variância σ^2

Se pretender um intervalo de confiança para o **desvio padrão**, então basta ir ao intervalo anterior e calcular a raiz quadrada, isto é

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para σ

$$\left[\sqrt{\frac{(n-1)s^2}{x_{1-\frac{\alpha}{2};n-1}^2}}, \sqrt{\frac{(n-1)s^2}{x_{\frac{\alpha}{2};n-1}^2}} \right]$$

Em resumo ([Formulário](#)):

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para a variância σ^2

Condição	Intervalo de Confiança
População Normal	$\left[\frac{(n-1)s^2}{x_{1-\frac{\alpha}{2}; n-1}^2}, \frac{(n-1)s^2}{x_{\frac{\alpha}{2}; n-1}^2} \right]$

Intervalo de Confiança para o quociente de variâncias $\frac{\sigma_1^2}{\sigma_2^2}$

Intervalo de Confiança para o quociente de variâncias $\frac{\sigma_1^2}{\sigma_2^2}$

Sejam X_1 e X_2 variáveis aleatórias que representam **populações Normais**, respectivamente, com médias μ_1 e μ_2 (desconhecidas) e com desvios padrão σ_1 e σ_2 . Selecionaram-se duas **amostras aleatórias independentes** de dimensões n_1 (da população X_1) e n_2 (da população X_2).

Estimador pontual para σ_1^2 : S_1^2

Estimador pontual para σ_2^2 : S_2^2

Distribuição Amostral

ou

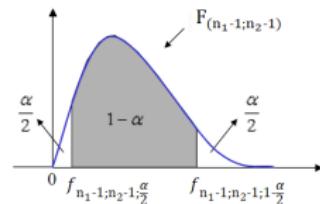
$$F = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n_1 - 1, n_2 - 1)}$$

Variável fulcral

Intervalo de Confiança para o quociente de variâncias $\frac{\sigma_1^2}{\sigma_2^2}$

Dedução do intervalo de confiança:

$$P(f_{\frac{\alpha}{2};n_1-1;n_2-1} < F < f_{1-\frac{\alpha}{2};n_1-1;n_2-1}) = 1 - \alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(\frac{1}{f_{1-\frac{\alpha}{2};n_1-1;n_2-1}} \times \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{\frac{\alpha}{2};n_1-1;n_2-1}} \times \frac{S_1^2}{S_2^2}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$

$$\left[\frac{1}{f_{1-\frac{\alpha}{2};n_1-1;n_2-1}} \times \frac{s_1^2}{s_2^2}, \frac{1}{f_{\frac{\alpha}{2};n_1-1;n_2-1}} \times \frac{s_1^2}{s_2^2} \right]$$

Intervalo de Confiança para o quociente de variâncias $\frac{\sigma_1^2}{\sigma_2^2}$

Se pretender um intervalo de confiança para o quociente dos **desvios padrão**, então basta ir ao intervalo anterior e calcular a raiz quadrada, isto é

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $\frac{\sigma_1}{\sigma_2}$

$$\left[\sqrt{\frac{1}{f_{1-\frac{\alpha}{2};n_1-1;n_2-1}} \times \frac{s_1^2}{s_2^2}}, \sqrt{\frac{1}{f_{\frac{\alpha}{2};n_1-1;n_2-1}} \times \frac{s_1^2}{s_2^2}} \right]$$

Em resumo ([Formulário](#)):

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para o quociente de variâncias $\frac{\sigma_1^2}{\sigma_2^2}$

Condições	Intervalo de Confiança
Populações Normais Amostras Independentes	$\left[\frac{1}{f_{1-\frac{\alpha}{2}; n_1-1; n_2-1}} \times \frac{s_1^2}{s_2^2}, \frac{1}{f_{\frac{\alpha}{2}; n_1-1; n_2-1}} \times \frac{s_1^2}{s_2^2} \right]$

Intervalo de Confiança para a proporção p

Intervalo de Confiança para a proporção p

Seja X uma população Binomial em que p é desconhecido e (X_1, X_2, \dots, X_n) uma amostra aleatória de dimensão $n \geq 30$ dessa população.

Estimador pontual para p : $p^* = \frac{n^q \text{ de casos favoráveis na Amostra}}{n^q \text{ de casos possíveis na Amostra}}$

Como $n \geq 30$, então pelo Teorema do Limite Central tem-se

Distribuição Amostral

ou

$$Z = \frac{p^* - p}{\sqrt{\frac{pq}{n}}} \cong \frac{p^* - p}{\sqrt{\frac{p^* q^*}{n}}} \sim N(0, 1)$$

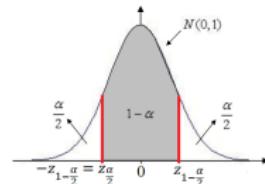
Variável fulcral

Observação: Para utilizar a distribuição amostral no cálculo do intervalo de confiança é necessário efetuar uma aproximação no denominador, em vez de colocar o verdadeiro desvio padrão é colocada uma aproximação.

Intervalo de Confiança para a proporção p

Dedução do intervalo de confiança:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \Leftrightarrow$$



$$\Leftrightarrow P\left(p^* - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^*q^*}{n}} < p < p^* + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^*q^*}{n}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para p

$$\left[p^* - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^*q^*}{n}}, p^* + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^*q^*}{n}} \right]$$

Em resumo ([Formulário](#)):

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para a proporção p

Condições	Intervalo de Confiança
$n \geq 30$	$\left[p^* - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}}; p^* + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}} \right]$

Intervalo de Confiança para a diferença de proporções $p_1 - p_2$

Intervalo de Confiança para a diferença de proporções $p_1 - p_2$

Sejam X_1 uma população Binomial em que p_1 é desconhecido e X_2 uma população Binomial em que p_2 é desconhecido. Selecionaram-se duas **amostras aleatórias independentes** de dimensões $n_1 \geq 30$ (da população X_1) e $n_2 \geq 30$ (da população X_2).

Estimador pontual para p_1 : p_1^*

Estimador pontual para p_2 : p_2^*

Como $n_1 \geq 30$ e $n_2 \geq 30$, então pelo Teorema do Limite Central tem-se

Distribuição Amostral

ou

$$Z = \frac{(p_1^* - p_2^*) - (p_1 - p_2)}{\sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}} \cong \frac{(p_1^* - p_2^*) - (p_1 - p_2)}{\sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}} \sim N(0, 1)$$

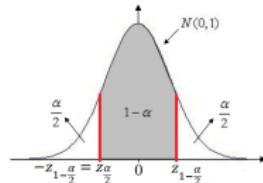
Variável fulcral

Observação: Para utilizar a distribuição amostral no cálculo do intervalo de confiança é necessário efetuar uma aproximação no denominador, em vez de colocar o verdadeiro desvio padrão é colocada uma aproximação.

Intervalo de Confiança para a diferença de proporções $p_1 - p_2$

Dedução do intervalo de confiança:

$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$



$$\Leftrightarrow P\left((p_1^* - p_2^*) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}} < p_1 - p_2 < (p_1^* - p_2^*) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}\right) = 1 - \alpha$$

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para $p_1 - p_2$

$$\left[(p_1^* - p_2^*) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}, (p_1^* - p_2^*) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}} \right]$$

Em resumo ([Formulário](#)):

Intervalo de confiança a $(1 - \alpha) \times 100\%$ para a diferença de proporções
 $p_1 - p_2$

Condições	Intervalo de Confiança
Amostras independentes $n_1 \geq 30$ e $n_2 \geq 30$	$\left[(p_1^* - p_2^*) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}, (p_1^* - p_2^*) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}} \right]$

Intervalos de Confiança no R

- Para as médias (uma média ou diferença de médias):
 - ▶ `z.test()` (é necessário instalar o package "BSDA")
 - ▶ `t.test()`
- Para variâncias:
 - ▶ uma variância: `varTest()` (é necessário instalar o package "EnvStats")
 - ▶ quociente de variâncias: `var.test()`
- Para as proporções (uma proporção ou diferença de proporções):
 - ▶ `z.test()` (é necessário instalar o package "BSDA")

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- Indique uma estimativa do valor médio do pH da solução.

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- Indique uma estimativa do valor médio do pH da solução.

Como vimos, um possível estimador para a média populacional é a média amostral:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

então uma estimativa do valor médio do pH da solução (média da população) é

$$\bar{x} = \frac{8.18 + 8.16 + 8.17 + 8.22 + 8.19}{5} = 8.184$$

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- ② Calcule um intervalo de confiança a 90% para o valor médio do pH da solução.

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- ② Calcule um intervalo de confiança a 90% para o valor médio do pH da solução.

- **População**

X = valor do pH da solução

$$X \sim N(\mu, 0.08)$$

média populacional = μ

desvio padrão populacional = $\sigma = 0.08$

- **Amostra Aleatória**

dimensão = $n = 5$

estimativa:

média amostral = $\bar{x} = 8.184$

Pretende-se um intervalo de confiança a 90% para μ .

- grau de confiança = $1 - \alpha = 0.90$
- nível de significância = $\alpha = 0.10$

A População tem distribuição Normal, $\sigma = 0.08$ conhecido, então o intervalo de confiança para a média é:

Distribuição Amostral	Intervalo de Confiança	R
$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$	$[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$	<code>z.test()</code> <code>library(BSDA)</code>

então um intervalo de confiança a 90% para o valor médio do pH da solução é

$$[8.125, 8.243]$$

Com 90% de confiança, o valor médio do pH da solução está entre 8.125 e 8.243.

Observação:

Neste caso como interpretação do intervalo de confiança calculado foi escrito:

“Com 90% de confiança, o valor médio do pH da solução está entre 8.125 e 8.243.”

mas há outra forma de dizer o mesmo:

“Com 90% de confiança, o valor médio do pH da solução é 8.184 com uma margem de erro de 0.059.”

ou seja, referir qual a estimativa pontual obtida e a quantidade que se está a subtrair e a somar a essa estimativa pontual:

$$]8.125, 8.243[=]8.184 - 0.059, 8.184 + 0.059[$$

A margem de erro é metade da amplitude do intervalo de confiança, que neste caso é

$$\text{margem de erro} = \frac{\text{amplitude}}{2} = \frac{8.243 - 8.125}{2} = 0.059$$

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- 3 Para um certo processo químico é importante que uma dada solução tenha um pH de 8.20. Com base no resultado da alínea anterior, o que pode concluir relativamente à utilização desta solução no referido processo químico?

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- 3 Para um certo processo químico é importante que uma dada solução tenha um pH de 8.20. Com base no resultado da alínea anterior, o que pode concluir relativamente à utilização desta solução no referido processo químico?

Com 90% de confiança, é possível utilizar esta solução no processo químico, pois o intervalo de confiança a 90% para o valor médio do pH da solução inclui o 8.20, $8.20 \in [8.125, 8.243]$.

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- 4 Qual deveria ser a dimensão da amostra se pretendesse obter um intervalo de confiança a 90% para o valor médio do pH da solução com amplitude inferior a 0.05?

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.08. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- 4 Qual deveria ser a dimensão da amostra se pretendesse obter um intervalo de confiança a 90% para o valor médio do pH da solução com amplitude inferior a 0.05?

Pretende-se determinar a dimensão da amostra, n , tal que amplitude do intervalo de confiança a 90% para μ seja inferior a 0.05.

Na alínea (2) calculou-se um intervalo de confiança a 90% para μ , $[8.125, 8.243]$, com

$$\text{amplitude} = 8.243 - 8.125 = 0.118$$

Como $0.118 > 0.05$, pretende-se diminuir a amplitude do intervalo, sem alterar o grau de confiança, então é necessário aumentar a dimensão da amostra, ou seja, terá de ser recolhida uma amostra de dimensão $n > 5$.

Como vimos, o intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ selecionado foi

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

logo

$$\text{amplitude} = \left(\bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

portanto pretende-se

$$\begin{aligned} \text{amplitude} < 0.05 &\Leftrightarrow 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < 0.05 \Leftrightarrow 2z_{0.95} \frac{0.08}{\sqrt{n}} < 0.05 \Leftrightarrow \\ &\Leftrightarrow 2 \times 1.645 \times \frac{0.08}{\sqrt{n}} < 0.05 \Leftrightarrow n > \left(\frac{2 \times 1.645 \times 0.08}{0.05} \right)^2 \Leftrightarrow n > 27.71 \Rightarrow n \geq 28 \end{aligned}$$

É necessário recolher uma amostra de dimensão $n \geq 28$ para obter um intervalo de confiança a 90% para o valor médio do pH da solução com amplitude inferior a 0.05.

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.02. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- 5 Qual deveria ser o grau de confiança se pretendesse obter um intervalo de confiança para o valor médio do pH da solução com margem de erro inferior a 0.03?

Exemplo 3

Um método utilizado na determinação do pH de uma dada solução fornece medições que se admite terem distribuição normal com desvio padrão de 0.02. Para avaliar o pH de uma solução, efetuaram-se 5 medições independentes tendo-se obtido os seguintes valores: 8.18; 8.16; 8.17; 8.22; 8.19.

- 5 Qual deveria ser o grau de confiança se pretendesse obter um intervalo de confiança para o valor médio do pH da solução com margem de erro inferior a 0.03?

Pretende-se determinar o grau de confiança, $1 - \alpha$, tal que a margem de erro do intervalo de confiança para μ seja inferior a 0.03.

Na alínea (2) calculou-se um intervalo de confiança a 90% para μ , $[8.125, 8.243]$, com

$$\text{margem de erro} = \frac{8.243 - 8.125}{2} = 0.059$$

Como $0.059 > 0.03$, pretende-se diminuir a margem de erro do intervalo, sem alterar a amostra recolhida, então é necessário diminuir o grau de confiança, ou seja, terá de ser considerado um grau de confiança $1 - \alpha < 0.90$.

Como vimos, o intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ selecionado tem

$$\text{amplitude} = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

portanto

$$\text{margem de erro} = \frac{2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}{2} = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

pretende-se

$$\begin{aligned} \text{margem de erro} < 0.03 &\Leftrightarrow z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < 0.03 \Leftrightarrow z_{1-\frac{\alpha}{2}} \frac{0.08}{\sqrt{5}} < 0.03 \Leftrightarrow \\ &\Leftrightarrow z_{1-\frac{\alpha}{2}} < 0.84 \Leftrightarrow 1 - \frac{\alpha}{2} < \Phi(0.84) \Leftrightarrow 1 - \frac{\alpha}{2} < 0.7995 \Leftrightarrow \alpha > 0.401 \Leftrightarrow \\ &\Leftrightarrow 1 - \alpha < 0.599 \end{aligned}$$

É necessário considerar um grau de confiança inferior a 59.9% ($1 - \alpha < 0.599$) para obter um intervalo de confiança para o valor médio do pH da solução com margem de erro inferior a 0.03.

Exemplo 4

Para avaliar o peso médio das maçãs produzidas por um determinado agricultor analisaram-se 20 maçãs selecionadas ao acaso da produção. Com base nessa amostra obteve-se o seguinte intervalo de confiança a 95% para o peso médio das maçãs:

$$]311.59, 328.41[$$

Sabe-se que os pesos das maçãs (em gramas) seguem uma distribuição Normal.

- 1 Qual a média e o desvio padrão do peso das maçãs da amostra?

Exemplo 4

Para avaliar o peso médio das maçãs produzidas por um determinado agricultor analisaram-se 20 maçãs selecionadas ao acaso da produção. Com base nessa amostra obteve-se o seguinte intervalo de confiança a 95% para o peso médio das maçãs:

$$]311.59, 328.41[$$

Sabe-se que os pesos das maçãs (em gramas) seguem uma distribuição Normal.

- Qual a média e o desvio padrão do peso das maçãs da amostra?

• População

X = peso das maçãs, em gramas

$$X \sim N(\mu, \sigma)$$

média populacional = μ

desvio padrão populacional = σ

• Amostra Aleatória

dimensão = $n = 20$

estimativas:

média amostral = \bar{x}

desvio padrão amostral = s

Calculou-se um intervalo de confiança a 95% para μ , então a média amostral está no meio do intervalo:

$$\bar{x} = \frac{311.59 + 328.41}{2} = 320 \text{ gramas}$$

A População tem distribuição Normal e σ desconhecido, então o intervalo de confiança para a média considerado foi:

Parâmetros da População	Intervalo de confiança
σ conhecido	$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$
σ desconhecido	$\left[\bar{x} - t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right]$

tem-se

- grau de confiança = $1 - \alpha = 0.95$
- nível de significância = $\alpha = 0.05$
- $t_{1-\frac{\alpha}{2};n-1} = t_{1-\frac{0.05}{2};20-1} = t_{0.975;19} = 2.09$

então um intervalo de confiança a 95% para o peso médio das maçãs é

$$\left[320 - 2.09 \times \frac{s}{\sqrt{20}}, 320 + 2.09 \times \frac{s}{\sqrt{20}} \right] = [311.59, 328.41]$$

Considerando, por exemplo, o limite inferior tem-se

$$320 - 2.09 \times \frac{s}{\sqrt{20}} = 311.59 \Leftrightarrow s = 18 \text{ gramas}$$

Exemplo 4

Para avaliar o peso médio das maçãs produzidas por um determinado agricultor analisaram-se 20 maçãs selecionadas ao acaso da produção. Com base nessa amostra obteve-se o seguinte intervalo de confiança a 95% para o peso médio das maçãs:

$$]311.59, 328.41[$$

Sabe-se que os pesos das maçãs (em gramas) seguem uma distribuição Normal.

- 2 Considerando que não há alteração nas estimativas observadas, qual deve ser o tamanho da amostra de forma a que a amplitude do intervalo de confiança a 95% para o peso médio das maçãs seja de 5 gramas?

Exemplo 4

Para avaliar o peso médio das maçãs produzidas por um determinado agricultor analisaram-se 20 maçãs selecionadas ao acaso da produção. Com base nessa amostra obteve-se o seguinte intervalo de confiança a 95% para o peso médio das maçãs:

$$]311.59, 328.41[$$

Sabe-se que os pesos das maçãs (em gramas) seguem uma distribuição Normal.

- ② Considerando que não há alteração nas estimativas observadas, qual deve ser o tamanho da amostra de forma a que a amplitude do intervalo de confiança a 95% para o peso médio das maçãs seja de 5 gramas?

Pretende-se determinar a dimensão da amostra, n , tal que amplitude do intervalo de confiança a 95% para μ seja de 5 gramas.

O intervalo de confiança a 95% para μ calculado com $n = 20$ tem

$$\text{amplitude} = 328.41 - 311.59 = 16.82 \text{ gramas}$$

Como $16.82 > 5$, pretende-se diminuir a amplitude do intervalo, sem alterar o grau de confiança, então é necessário aumentar a dimensão da amostra, ou seja, terá de ser recolhida uma amostra de dimensão $n > 20$.

Como vimos, o intervalo de confiança a $(1 - \alpha) \times 100\%$ para μ selecionado foi

$$\left[\bar{x} - t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right]$$

logo

$$\text{amplitude} = \left(\bar{x} + t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right) - \left(\bar{x} - t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \right) = 2t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}}$$

portanto pretende-se

$$\text{amplitude} = 5 \Leftrightarrow 2t_{1-\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} = 5 \Leftrightarrow 2t_{0.975;n-1} \frac{18}{\sqrt{n}} = 5$$

Como o quantil depende de n , $t_{0.95;n-1}$, seria necessário fazer as contas uma a uma para cada possível valor de n até encontrar o menor valor que respeita a condição pretendida. No entanto sabemos que para $n \geq 30$ o intervalo de confiança utilizado seria

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

logo

$$\text{amplitude} = \left(\bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right) = 2z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

cujo quantil, $z_{1-\frac{\alpha}{2}}$ não depende de n , então vamos usar este resultado para determinar n (se não resultar então é necessário ir calcular o quantil da t de Student para $n \in [20, 30]$).

Para $n \geq 30$ tem-se

$$\begin{aligned} \text{amplitude} = 5 &\Leftrightarrow 2z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 5 \Leftrightarrow 2z_{1-\frac{0.05}{2}} \frac{18}{\sqrt{n}} = 5 \Leftrightarrow 2z_{0.975} \frac{18}{\sqrt{n}} = 5 \Leftrightarrow \\ &\Leftrightarrow 2 \times 1.96 \times \frac{18}{\sqrt{n}} = 5 \Leftrightarrow n = \left(\frac{2 \times 1.96 \times 18}{5} \right)^2 \Leftrightarrow n = 199.15 \Rightarrow n = 200 \end{aligned}$$

Como o valor obtido é $n = 200 \geq 30$, então já se encontrou o valor pretendido.

É necessário recolher uma amostra de dimensão $n = 200$ para obter um intervalo de confiança a 95% para μ com amplitude de 5 gramas (supondo que a estimativa do desvio padrão se mantinha em 18).

Exemplo 5

Pretende-se construir um intervalo de confiança a 95% para a diferença das médias de pontos obtidos por dois golfistas em determinado torneio. Sabe-se que as pontuações de ambos os golfistas são normalmente distribuídas e que as variâncias podem ser consideradas iguais. Seleccionaram-se aleatoriamente 10 jogos do golfista A e 16 jogos do golfista B e obtiveram-se os seguintes resultados:

golf. A:	36	34	34	36	35	34	33	35	33	35
golf. B:	26	37	28	26	34	28	32	27	23	28

- Construa o referido intervalo de confiança. Comente.

Exemplo 5

Pretende-se construir um intervalo de confiança a 95% para a diferença das médias de pontos obtidos por dois golfistas em determinado torneio. Sabe-se que as pontuações de ambos os golfistas são normalmente distribuídas e que as variâncias podem ser consideradas iguais. Seleccionaram-se aleatoriamente 10 jogos do golfista A e 16 jogos do golfista B e obtiveram-se os seguintes resultados:

golf. A:	36	34	34	36	35	34	33	35	33	35						
golf. B:	26	37	28	26	34	28	32	27	23	28	26	36	33	30	31	27

- Construa o referido intervalo de confiança. Comente.

Atenção:

Como as amostras são amostras aleatórias **independentes**, então o intervalo de confiança a calcular deve ser um intervalo de confiança para $\mu_1 - \mu_2$ usando como estimador pontual $\bar{X}_1 - \bar{X}_2 \rightarrow$ estamos interessados na diferença das médias.

• População A

X_1 = pontuação do golfista A

$$X_1 \sim N(\mu_1, \sigma_1)$$

média populacional = μ_1

desvio padrão populacional = σ_1

• População B

X_2 = pontuação do golfista B

$$X_2 \sim N(\mu_2, \sigma_2)$$

média populacional = μ_2

desvio padrão populacional = σ_2

- $\sigma_1^2 = \sigma_2^2 \Leftrightarrow \sigma_1 = \sigma_2$

• Amostra Aleatória de A

dimensão = $n_1 = 10$

estimativas:

média amostral = $\bar{x}_1 = 34.5$

desvio padrão amostral = $s_1 = 1.0801$

• Amostra Aleatória de B

dimensão = $n_2 = 16$

estimativas:

média amostral = $\bar{x}_2 = 29.5$

desvio padrão amostral = $s_2 = 3.9833$

Pretende-se um intervalo de confiança a 95% para $\mu_1 - \mu_2$.

- grau de confiança = $1 - \alpha = 0.95$
- nível de significância = $\alpha = 0.05$

As Populações têm distribuição Normal, σ_1 e σ_2 desconhecidos mas $\sigma_1 = \sigma_2$, as amostras são independentes e as amostras têm dimensão $n_1 = 10 < 30$ e $n_2 = 16 < 30$, então o intervalo de confiança para a diferença de médias é:

	R
D. A.	$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}} \sim t_{(n_1+n_2-2)}$
I. C.	$\left[(\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}; n_1+n_2-2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}; \right.$ $\left. ; (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}; n_1+n_2-2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \right]$

então um intervalo de confiança a 95% para a diferença entre as médias das pontuações dos golfistas é

$$]2.3229, 7.6771[$$

Com 95% de confiança, as pontuações médias dos golfistas não podem ser consideradas iguais pois o intervalo de confiança não inclui o zero, $0 \notin]2.3229, 7.6771[$.
(Observação: Se $\mu_1 - \mu_2 = 0$, tem-se $\mu_1 = \mu_2$)

Com 95% de confiança, a pontuação média do golfista A deve ser superior à do golfista B pois o intervalo de confiança só tem valores positivos.

(Observação: Se $\mu_1 - \mu_2 > 0$, tem-se $\mu_1 > \mu_2$)

Ou seja, com 95% de confiança, o que foi observado na amostra recolhida ($\bar{x}_1 > \bar{x}_2$) parece manter-se na população ($\mu_1 > \mu_2$).

Exemplo 5

Pretende-se construir um intervalo de confiança a 95% para a diferença das médias de pontos obtidos por dois golfistas em determinado torneio. Sabe-se que as pontuações de ambos os golfistas são normalmente distribuídas e que as variâncias podem ser consideradas iguais. Selecionaram-se aleatoriamente 10 jogos do golfista A e 16 jogos do golfista B e obtiveram-se os seguintes resultados:

golf. A:	36	34	34	36	35	34	33	35	33	35						
golf. B:	26	37	28	26	34	28	32	27	23	28	26	36	33	30	31	27

- 2 Com recurso a um intervalo de confiança a 95% verifique se a suposição das variâncias poderem ser consideradas iguais é válida.

● **População A**

X_1 = pontuação do golfista A

$$X_1 \sim N(\mu_1, \sigma_1)$$

média populacional = μ_1

desvio padrão populacional = σ_1

● **População B**

X_2 = pontuação do golfista B

$$X_2 \sim N(\mu_2, \sigma_2)$$

média populacional = μ_2

desvio padrão populacional = σ_2

● $\sigma_1^2 = \sigma_2^2 \Leftrightarrow \sigma_1 = \sigma_2$

● **Amostra Aleatória de A**

dimensão = $n_1 = 10$

estimativas:

média amostral = $\bar{x}_1 = 34.5$

desvio padrão amostral = $s_1 = 1.0801$

● **Amostra Aleatória de B**

dimensão = $n_2 = 16$

estimativas:

média amostral = $\bar{x}_2 = 29.5$

desvio padrão amostral = $s_2 = 3.9833$

Pretende-se um intervalo de confiança a 95% para $\frac{\sigma_1^2}{\sigma_2^2}$.

- grau de confiança = $1 - \alpha = 0.95$
- nível de significância = $\alpha = 0.05$

As Populações têm distribuição Normal, e as amostras são independentes, então o intervalo de confiança para o quociente de variâncias é:

Distribuição Amostral	Intervalo de Confiança	R
$F = \frac{s_1^2}{s_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n_1 - 1, n_2 - 1)}$	$\left[\frac{1}{F_{1 - \frac{\alpha}{2}; n_1 - 1; n_2 - 1}} \times \frac{s_1^2}{s_2^2}, \frac{1}{F_{\frac{\alpha}{2}; n_1 - 1; n_2 - 1}} \times \frac{s_1^2}{s_2^2} \right]$	var.test()

então um intervalo de confiança a 95% para o quociente das variâncias das pontuações dos golfistas é

$$]0.0235, 0.2772[$$

Com 95% de confiança, as variâncias das pontuações dos golfistas não podem ser consideradas iguais pois o intervalo de confiança não inclui o um, $1 \notin]0.0235, 0.2772[$.

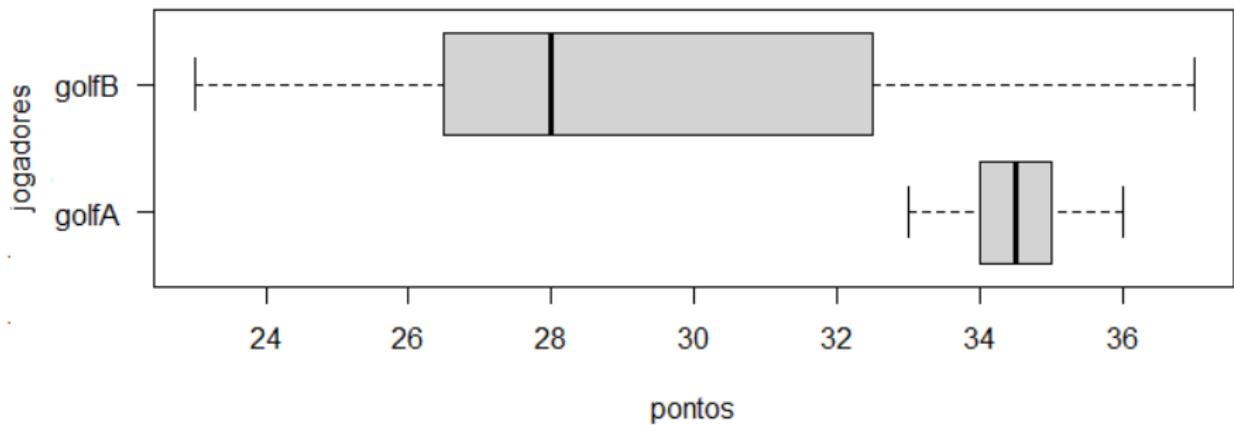
(Observação: Se $\frac{\sigma_1^2}{\sigma_2^2} = 1$, tem-se $\sigma_1^2 = \sigma_2^2$)

Com 95% de confiança, a variância da pontuação do golfista A deve ser inferior à variância da pontuação do golfista B pois o intervalo de confiança só tem valores inferiores a 1.

(Observação: Se $\frac{\sigma_1^2}{\sigma_2^2} < 1$, tem-se $\sigma_1^2 < \sigma_2^2$)

Ou seja, com 95% de confiança, o que foi observado na amostra recolhida ($s_1^2 < s_2^2$) parece manter-se na população ($\sigma_1^2 < \sigma_2^2$).

Se tivesse feito o diagrama de extremos e quartis para os dois golfistas, podíamos ver que de facto os dados têm comportamentos muito diferentes:



Exemplo 5

Pretende-se construir um intervalo de confiança a 95% para a diferença das médias de pontos obtidos por dois golfistas em determinado torneio. Sabe-se que as pontuações de ambos os golfistas são normalmente distribuídas e que as variâncias podem ser consideradas iguais. Selecionearam-se aleatoriamente 10 jogos do golfista A e 16 jogos do golfista B e obtiveram-se os seguintes resultados:

golf. A:	36	34	34	36	35	34	33	35	33	35						
golf. B:	26	37	28	26	34	28	32	27	23	28	26	36	33	30	31	27

- ③ Com base no resultado na alínea 2. que comentário pode fazer sobre a resolução da alínea 1.?

Exemplo 5

Pretende-se construir um intervalo de confiança a 95% para a diferença das médias de pontos obtidos por dois golfistas em determinado torneio. Sabe-se que as pontuações de ambos os golfistas são normalmente distribuídas e que as variâncias podem ser consideradas iguais. Selecionearam-se aleatoriamente 10 jogos do golfista A e 16 jogos do golfista B e obtiveram-se os seguintes resultados:

golf. A:	36	34	34	36	35	34	33	35	33	35						
golf. B:	26	37	28	26	34	28	32	27	23	28	26	36	33	30	31	27

- ③ Com base no resultado na alínea 2. que comentário pode fazer sobre a resolução da alínea 1.?

Com 95% de confiança, na alínea 1. usámos uma condição falsa pois não devíamos ter considerado as variâncias populacionais iguais.

O intervalo de confiança calculado devia ter sido:

● **População A**

X_1 = pontuação do golfista A

$$X_1 \sim N(\mu_1, \sigma_1)$$

média populacional = μ_1

desvio padrão populacional = σ_1

● **População B**

X_2 = pontuação do golfista B

$$X_2 \sim N(\mu_2, \sigma_2)$$

média populacional = μ_2

desvio padrão populacional = σ_2

● $\sigma_1^2 = \sigma_2^2 \Leftrightarrow \sigma_1 = \sigma_2$

● **Amostra Aleatória de A**

dimensão = $n_1 = 10$

estimativas:

média amostral = $\bar{x}_1 = 34.5$

desvio padrão amostral = $s_1 = 1.0801$

● **Amostra Aleatória de B**

dimensão = $n_2 = 16$

estimativas:

média amostral = $\bar{x}_2 = 29.5$

desvio padrão amostral = $s_2 = 3.9833$

Pretende-se um intervalo de confiança a 95% para $\mu_1 - \mu_2$.

- grau de confiança = $1 - \alpha = 0.95$
- nível de significância = $\alpha = 0.05$

As Populações têm distribuição Normal, σ_1 e σ_2 desconhecidos mas $\sigma_1 \neq \sigma_2$, as amostras são independentes e as amostras têm dimensão $n_1 = 10 < 30$ e $n_2 = 16 < 30$, então o intervalo de confiança para a diferença de médias é:

R

D. A.	$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{(gl_2)}$	t.test()
I. C.	$\left[(\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}; gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}; gl_2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$	

então um intervalo de confiança a 95% para a diferença entre as médias das pontuações dos golfistas é

$$]2.7909, 7.2091[$$

Como o intervalo de confiança obtido não é muito diferente do anterior, a conclusão é idêntica.

Com 95% de confiança, as pontuações médias dos golfistas não podem ser consideradas iguais pois o intervalo de confiança não inclui o zero, $0 \notin]2.7909, 7.2091[$.

(Observação: Se $\mu_1 - \mu_2 = 0$, tem-se $\mu_1 = \mu_2$)

Com 95% de confiança, a pontuação média do golfista A deve ser superior à do golfista B pois o intervalo de confiança só tem valores positivos.

(Observação: Se $\mu_1 - \mu_2 > 0$, tem-se $\mu_1 > \mu_2$)

Ou seja, com 95% de confiança, o que foi observado na amostra recolhida ($\bar{x}_1 > \bar{x}_2$) parece manter-se na população ($\mu_1 > \mu_2$).

Exemplo 6

Uma empresa deseja estudar o efeito de uma pausa de dez minutos para um cafezinho sobre a produtividade dos trabalhadores. Para isso, sorteou seis trabalhadores, e contou o número de peças produzidas durante uma semana sem intervalo e uma semana com intervalo:

trabalhador	1	2	3	4	5	6
sem intervalo	23	35	29	33	43	32
com intervalo	28	38	30	37	42	30

Suponha que o número de peças produzidas pode ser modelado segundo uma distribuição normal. Recorrendo a um intervalo de confiança a 95%, os resultados sugerem que há melhoria na produtividade?

Exemplo 6

Uma empresa deseja estudar o efeito de uma pausa de dez minutos para um cafezinho sobre a produtividade dos trabalhadores. Para isso, sorteou seis trabalhadores, e contou o número de peças produzidas durante uma semana sem intervalo e uma semana com intervalo:

trabalhador	1	2	3	4	5	6
sem intervalo	23	35	29	33	43	32
com intervalo	28	38	30	37	42	30

Suponha que o número de peças produzidas pode ser modelado segundo uma distribuição normal. Recorrendo a um intervalo de confiança a 95%, os resultados sugerem que há melhoria na produtividade?

Atenção:

Neste caso as amostras são amostras aleatórias emparelhadas, então é necessário construir a amostra das diferenças, $d_i = x_{1i} - x_{2i}$ e o intervalo de confiança a calcular deve ser um intervalo de confiança para μ usando como estimador pontual \bar{X}_D (média das diferenças) → estamos interessados na diferença média.

Neste caso temos uma única população onde as amostras foram obtidas com os mesmos trabalhadores em períodos diferentes, ou seja, temos amostras aleatórias emparelhadas. Como as amostras aleatórias são emparelhadas, não é possível construir intervalos de confiança para $\mu_1 - \mu_2$ pois uma das hipóteses na construção desses intervalos impõe que as amostras aleatórias sejam independentes.

Como as amostras aleatórias são emparelhadas, vamos construir uma única amostra, a amostra das diferenças:

trabalhador	1	2	3	4	5	6
sem intervalo	23	35	29	33	43	32
com intervalo	28	38	30	37	42	30
$D = \text{com intervalo} - \text{sem intervalo}$	5	3	1	4	-1	-2

Pretende-se recorrer a um intervalo de confiança a 95% para verificar se há melhoria na produtividade, então vamos construir um intervalo de confiança a 95% para a média das diferenças (μ_D) e se o intervalo só tiver valores positivos, então com 95% confiança o intervalo para o cafezinho trouxe melhoria na produtividade.

A População tem distribuição Normal, σ desconhecido e Amostras Emparelhadas então o intervalo de confiança para a média é:

Distribuição Amostral	Intervalo de Confiança	R
$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$	$\left[\bar{x} - t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}; n-1} \frac{s}{\sqrt{n}} \right]$	t.test()

Pretende-se um intervalo de confiança a 95% para μ_D

- amostra: $(5, 3, 1, 4, -1, -2)$
- grau de confiança $= 1 - \alpha = 0.95$
- nível de significância $= \alpha = 0.05$

portanto um intervalo de confiança a 95% para μ_D é

$$]-1.277, 4.610[$$

Com 95% de confiança os resultados não sugerem que haja melhoria na produtividade, pois o intervalo de confiança inclui o zero, $0 \in]-1.277, 4.610[$.

Ou seja, com 95% de confiança, não é possível dizer que o que foi observado na amostra ($\bar{x}_D > 0$) se mantém na população.

Portanto com 95% de confiança não há evidência estatística que a pausa para o cafezinho melhore a produtividade.

Exemplo 7

Um fabricante de máquinas para cortar relva pretende determinar o desvio padrão do tempo de vida de um de seus modelos, mas apenas sabe que o tempo de vida desses modelos segue uma distribuição Normal.

- Para tal selecionou aleatoriamente 12 dessas máquinas que tinham sido vendidas há vários anos e verificou que o desvio padrão dessa amostra é de 3.25 anos. Construa um intervalo de confiança a 90% para o desvio padrão pretendido.

Exemplo 7

Um fabricante de máquinas para cortar relva pretende determinar o desvio padrão do tempo de vida de um de seus modelos, mas apenas sabe que o tempo de vida desses modelos segue uma distribuição Normal.

- Para tal selecionou aleatoriamente 12 dessas máquinas que tinham sido vendidas há vários anos e verificou que o desvio padrão dessa amostra é de 3.25 anos. Construa um intervalo de confiança a 90% para o desvio padrão pretendido.

- População**

X = tempo de vida de um modelo de máquina de cortar relva

$$X \sim N(\mu, \sigma)$$

média populacional = μ

desvio padrão populacional = σ

- Amostra Aleatória**

dimensão = $n = 12$

estimativa:

desvio padrão amostral = $s = 3.25$

Pretende-se um intervalo de confiança a 90% para σ .

- grau de confiança = $1 - \alpha = 0.90$
- nível de significância = $\alpha = 0.10$

A População tem de ter obrigatoriamente distribuição Normal, como é o caso, e tem-se μ desconhecido, então o intervalo de confiança para a variância é:

Distribuição Amostral	Intervalo de Confiança	R
$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$	$\left[\frac{(n-1)s^2}{x^2_{1-\frac{\alpha}{2};n-1}}, \frac{(n-1)s^2}{x^2_{\frac{\alpha}{2};n-1}} \right]$	varTest() library(EnvStats)

então

- $x^2_{\frac{\alpha}{2};n-1} = x^2_{\frac{0.10}{2};12-1} = x^2_{0.05;11} = 4.57$
- $x^2_{1-\frac{\alpha}{2};n-1} = x^2_{1-\frac{0.10}{2};12-1} = x^2_{0.95;11} = 19.7$

portanto um intervalo de confiança a 90% para a **variância** do tempo de vida de um modelo de máquina de cortar relva é

$$\left[\frac{(12 - 1) \times 3.25^2}{19.7}, \frac{(12 - 1) \times 3.25^2}{4.57} \right]$$

e um intervalo de confiança a 90% para o **desvio padrão** do tempo de vida de um modelo de máquina de cortar relva é

$$\left[\sqrt{\frac{(12 - 1) \times 3.25^2}{19.7}}, \sqrt{\frac{(12 - 1) \times 3.25^2}{4.57}} \right] = [2.429, 5.042]$$

Com 90% de confiança, o desvio padrão do tempo de vida de um modelo de máquina de cortar relva está entre 2.429 e 5.042 anos.

Exemplo 7

Um fabricante de máquinas para cortar relva pretende determinar o desvio padrão do tempo de vida de um de seus modelos, mas apenas sabe que o tempo de vida desses modelos segue uma distribuição Normal.

- ② Para tal selecionou aleatoriamente 12 dessas máquinas que tinham sido vendidas há vários anos e registou os seus tempos de vida:

(14.06, 6.08, 9.76, 9.75, 15.45, 6.77, 13.81, 6.28, 9.56, 10.75, 15.16, 5.57)

Construa um intervalo de confiança a 90% para o desvio padrão pretendido.

Exemplo 7

Um fabricante de máquinas para cortar relva pretende determinar o desvio padrão do tempo de vida de um de seus modelos, mas apenas sabe que o tempo de vida desses modelos segue uma distribuição Normal.

- ② Para tal selecionou aleatoriamente 12 dessas máquinas que tinham sido vendidas há vários anos e registou os seus tempos de vida:

(14.06, 6.08, 9.76, 9.75, 15.45, 6.77, 13.81, 6.28, 9.56, 10.75, 15.16, 5.57)

Construa um intervalo de confiança a 90% para o desvio padrão pretendido.

• População

X = tempo de vida de um modelo de máquina de cortar relva

$$X \sim N(\mu, \sigma)$$

média populacional = μ

desvio padrão populacional = σ

• Amostra Aleatória

dimensão = $n = 12$

estimativa:

desvio padrão amostral = $s = 3.65$

Pretende-se um intervalo de confiança a 90% para σ .

- grau de confiança = $1 - \alpha = 0.90$
- nível de significância = $\alpha = 0.10$

A População tem de ter obrigatoriamente distribuição Normal, como é o caso, e tem-se μ desconhecido, então o intervalo de confiança para a variânciá é:

Distribuição Amostral	Intervalo de Confiança	R
$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$	$\left[\frac{(n-1)s^2}{x^2_{1-\frac{\alpha}{2}; n-1}}, \frac{(n-1)s^2}{x^2_{\frac{\alpha}{2}; n-1}} \right]$	<pre>varTest() library(EnvStats)</pre>

então, recorrendo ao R tem-se

um intervalo de confiança a 90% para a **variância** do tempo de vida de um modelo de máquina de cortar relva é

$$]7.4570, 32.0705[$$

e um intervalo de confiança a 90% para o **desvio padrão** do tempo de vida de um modelo de máquina de cortar relva é

$$\left] \sqrt{7.4570}, \sqrt{32.0705} \right[=]2.7307, 5.6631[$$

Com 90% de confiança, o desvio padrão do tempo de vida de um modelo de máquina de cortar relva está entre 2.7307 e 5.6631 anos.

Exemplo 8

Pretende-se verificar se duas máquinas (A e B) produzem peças com a mesma homogeneidade quanto à resistência à tensão. Para tal, selecionaram-se duas amostras, sete peças da máquina A e seis peças da máquina B, e obtivemos as seguintes resultados sobre as resistências: $s_A^2 = 40$ e $s_B^2 = 36$. Sabe-se que a resistência à tensão é normalmente distribuída. Com base num intervalo de confiança de 99% para o quociente de variâncias, verifique se as máquinas produzem peças com a mesma homogeneidade.

Exemplo 8

Pretende-se verificar se duas máquinas (A e B) produzem peças com a mesma homogeneidade quanto à resistência à tensão. Para tal, selecionaram-se duas amostras, sete peças da máquina A e seis peças da máquina B, e obtivemos as seguintes resultados sobre as resistências: $s_A^2 = 40$ e $s_B^2 = 36$. Sabe-se que a resistência à tensão é normalmente distribuída. Com base num intervalo de confiança de 99% para o quociente de variâncias, verifique se as máquinas produzem peças com a mesma homogeneidade.

• População A

X_A = resistência à tensão da máquina A

$$X_A \sim N(\mu_A, \sigma_A)$$

média populacional = μ_A

desvio padrão populacional = σ_A

• População B

X_B = resistência à tensão da máquina B

$$X_B \sim N(\mu_B, \sigma_B)$$

média populacional = μ_B

desvio padrão populacional = σ_B

• Amostra Aleatória de A

dimensão = $n_A = 7$

estimativa:

variância amostral = $s_A^2 = 40$

• Amostra Aleatória de B

dimensão = $n_B = 6$

estimativa:

variância amostral = $s_B^2 = 36$

Pretende-se um intervalo de confiança a 99% para $\frac{\sigma_A^2}{\sigma_B^2}$.

- grau de confiança = $1 - \alpha = 0.99$
- nível de significância = $\alpha = 0.01$

As Populações têm de ter obrigatoriamente distribuição Normal, como é o caso, tem-se μ_A e μ_B desconhecidos e as amostras são independentes, então o intervalo de confiança para o quociente de variâncias é:

Distribuição Amostral	Intervalo de Confiança	R
$F = \frac{s_1^2}{s_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n_1-1, n_2-1)}$	$\left[\frac{1}{f_{1-\frac{\alpha}{2}; n_1-1; n_2-1}} \times \frac{s_1^2}{s_2^2}, \frac{1}{f_{\frac{\alpha}{2}; n_1-1; n_2-1}} \times \frac{s_1^2}{s_2^2} \right]$	var.test()

então

- $f_{\frac{\alpha}{2}; n_1-1; n_2-1} = f_{\frac{\alpha}{2}; n_A-1; n_B-1} = f_{0.01; 7-1; 6-1} = f_{0.005; 6; 5} = 0.087$
- $f_{1-\frac{\alpha}{2}; n_1-1; n_2-1} = f_{1-\frac{\alpha}{2}; n_A-1; n_B-1} = f_{1-\frac{0.01}{2}; 7-1; 6-1} = f_{0.995; 6; 5} = 14.5$

portanto um intervalo de confiança a 99% para o quociente entre as variâncias das resistências à tensão é

$$\left[\frac{1}{14.5} \times \frac{40}{36}, \frac{1}{0.087} \times \frac{40}{36} \right] = [0.077, 12.771]$$

Com 99% de confiança, as variâncias das resistências à tensão não podem ser consideradas diferentes pois o intervalo de confiança inclui o um, $1 \in [0.077, 12.771]$.

(Observação: Se $\frac{\sigma_A^2}{\sigma_B^2} = 1$, tem-se $\sigma_A^2 = \sigma_B^2$)

Ou seja, com 99% de confiança, não é possível dizer que o que foi observado nas amostras ($s_A^2 > s_B^2$) mantém-se na população.

Portanto, com 99% de confiança, não se pode concluir que as máquinas produzem peças com homogeneidade diferente quanto à resistência à tensão.

Exemplo 9

Em 60 lançamentos de uma moeda, foram obtidas 24 caras. Com base num intervalo de confiança de 96%, pode-se concluir que a moeda é viciada?

Exemplo 9

Em 60 lançamentos de uma moeda, foram obtidas 24 caras. Com base num intervalo de confiança de 96%, pode-se concluir que a moeda é viciada?

- **População**

X = número de caras, em N lançamentos

$$X \sim \text{Binomial}$$

proporção populacional = p

- **Amostra Aleatória**

dimensão = $n = 60$

estimativa:

$$\text{proporção amostral} = p^* = \frac{24}{60} = 0.4$$

Pretende-se um intervalo de confiança a 96% para p .

- grau de confiança = $1 - \alpha = 0.96$
- nível de significância = $\alpha = 0.04$

Como a População é Binomial, então é obrigatório recolher uma amostra de dimensão $n \geq 30$ para se poder recorrer ao Teorema do Limite Central. Como $n = 60 \geq 30$, então o intervalo de confiança para a proporção é:

Distribuição Amostral	Intervalo de Confiança	R
$Z = \frac{p^* - p}{\sqrt{\frac{p^* q^*}{n}}} \sim N(0, 1)$	$\left[p^* - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}}, p^* + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}} \right]$	<code>z.test()</code> <code>library(BSDA)</code>

atenção que este intervalo de confiança é uma adaptação do intervalo de confiança para a média:

$$\left[p^* - z_{1-\frac{\alpha}{2}} \frac{\sqrt{p^* q^*}}{\sqrt{n}}, p^* + z_{1-\frac{\alpha}{2}} \frac{\sqrt{p^* q^*}}{\sqrt{n}} \right] = \left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

onde

- $\bar{x} = p^*$
- $s = \sqrt{p^* q^*}$

Se pretender usar a função z.test do R, é necessário:

- considerar o sucesso = 1 e o insucesso = 0 e construir o vetor da amostra com 24 uns e $(60 - 24) = 36$ zeros
- considerar $s = \sqrt{p^*q^*} = \sqrt{p^* \times (1 - p^*)}$

portanto um intervalo de confiança a 96% para a proporção de caras é

$$]0.2701, 0.5299[$$

Com 96% de confiança, não se pode concluir que a moeda seja viciada pois 0.5 pertence ao intervalo de confiança, $0.5 \in]0.2701, 0.5299[$.

Como é óbvio, é sempre possível calcular o intervalo de confiança pretendido sem recorrer à função z.test. Basta fazer:

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.04}{2}} = z_{0.98} = 2.0537$$

portanto um intervalo de confiança a 96% para a proporção de caras é

$$\left[0.4 - 2.0537 \times \sqrt{\frac{0.4 \times (1 - 0.4)}{60}}, 0.4 + 2.0537 \times \sqrt{\frac{0.4 \times (1 - 0.4)}{60}} \right] = \\ =]0.2701, 0.5299[$$

Com 96% de confiança, não se pode concluir que a moeda seja viciada pois 0.5 pertence ao intervalo de confiança, $0.5 \in]0.2701, 0.5299[$.

Exemplo 10

Suponha que pretende estimar a proporção de uma população Binomial. Qual a dimensão da amostra a recolher para estimar essa proporção com uma margem de erro de 0.05 e um grau de confiança de 95%?

Exemplo 10

Suponha que pretende estimar a proporção de uma população Binomial. Qual a dimensão da amostra a recolher para estimar essa proporção com uma margem de erro de 0.05 e um grau de confiança de 95%?

- **População**

$$X \sim \text{Binomial}$$

proporção populacional = p

- **Amostra Aleatória**

dimensão = n

estimativa:

proporção amostral = p^*

Pretende-se um intervalo de confiança a 95% para p :

- grau de confiança = $1 - \alpha = 0.95$
- nível de significância = $\alpha = 0.05$

com uma margem de erro igual a 0.05:

$$\text{margem de erro} = \frac{\text{amplitude}}{2}$$

Como a População é Binomial, então é obrigatório recolher uma amostra de dimensão $n \geq 30$ para se poder recorrer ao Teorema do Limite Central, e, nesse caso, o intervalo de confiança para a proporção é:

Distribuição Amostral	Intervalo de Confiança	R
$Z = \frac{p^* - p}{\sqrt{\frac{p^* q^*}{n}}} \sim N(0, 1)$	$\left[p^* - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}}; p^* + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}} \right]$	<code>z.test()</code> <code>library(BSDA)</code>

logo

$$\text{amplitude} = \left(p^* + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}} \right) - \left(p^* - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}} \right) = 2z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}}$$

portanto

$$\text{margem de erro} = \frac{2z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}}}{2} = z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^* q^*}{n}}$$

pretende-se

$$\text{margem de erro} = 0.05 \Leftrightarrow z_{1-\frac{\alpha}{2}} \sqrt{\frac{p^*q^*}{n}} = 0.05$$

como

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$$

fica

$$\text{margem de erro} = 0.05 \Leftrightarrow 1.96 \times \sqrt{\frac{p^*q^*}{n}} = 0.05$$

Pretende-se determinar n mas é necessário conhecer p^* , só que p^* depende de n . Então para resolver este problema considera-se $p^* = 0.5$ (valor que irá dar origem ao maior n , ou seja, a "pior situação") e fica

$$\begin{aligned} \text{margem de erro} = 0.05 &\Leftrightarrow 1.96 \times \sqrt{\frac{0.5 \times (1 - 0.5)}{n}} = 0.05 \Leftrightarrow \\ &\Leftrightarrow n = \left(\frac{1.96 \times \sqrt{0.5 \times (1 - 0.5)}}{0.05} \right)^2 \Leftrightarrow n = 384.16 \Rightarrow n = 385 \end{aligned}$$

É necessário recolher uma amostra de dimensão $n = 385$ para obter um intervalo de confiança a 95% para a proporção com uma margem de erro de 0.05.



Exemplo 11

Em virtude dos protestos quanto à degradação do meio ambiente, algumas empresas têm adotado materiais totalmente recicláveis nos seus produtos. Numa pesquisa foram inquiridos 270 consumidores do distrito de Setúbal e 290 consumidores do distrito de Faro, aleatoriamente selecionados, e foi-lhes perguntado se eles comprariam um produto que não tivesse etiqueta informativa de que os materiais utilizados na embalagem são totalmente recicláveis. Dos entrevistados, 81 consumidores de Setúbal e 116 de Faro responderam que não comprariam o produto. Com base num intervalo de confiança a 99% verifique se existem diferenças na proporção de consumidores dos dois distritos que não comprariam o produto.

Exemplo 11

Em virtude dos protestos quanto à degradação do meio ambiente, algumas empresas têm adotado materiais totalmente recicláveis nos seus produtos. Numa pesquisa foram inquiridos 270 consumidores do distrito de Setúbal e 290 consumidores do distrito de Faro, aleatoriamente selecionados, e foi-lhes perguntado se eles comprariam um produto que não tivesse etiqueta informativa de que os materiais utilizados na embalagem são totalmente recicláveis. Dos entrevistados, 81 consumidores de Setúbal e 116 de Faro responderam que não comprariam o produto. Com base num intervalo de confiança a 99% verifique se existem diferenças na proporção de consumidores dos dois distritos que não comprariam o produto.

• População 1

X_1 = número de consumidores que não comprariam o produto, em N_1 consumidores de Setúbal

$$X_1 \sim \text{Binomial}$$

proporção populacional = p_1

• População 2

X_2 = número de consumidores que não comprariam o produto, em N_2 consumidores de Faro

$$X_2 \sim \text{Binomial}$$

proporção populacional = p_2

• Amostra Aleatória de 1

dimensão = $n_1 = 270$

estimativa:

$$\text{proporção amostral} = p_1^* = \frac{81}{270} = 0.3$$

• Amostra Aleatória de 2

dimensão = $n_2 = 290$

estimativas:

$$\text{proporção amostral} = p_2^* = \frac{116}{290} = 0.4$$

Pretende-se um intervalo de confiança a 99% para $p_1 - p_2$.

- grau de confiança = $1 - \alpha = 0.99$
- nível de significância = $\alpha = 0.01$

Como as Populações são Binomiais, então é obrigatório recolher amostras de dimensão $n \geq 30$ para se poder recorrer ao Teorema do Limite Central. Como $n_1 = 270 \geq 30$, $n_2 = 290 \geq 30$ e as amostras são independentes, então o intervalo de confiança para a diferença de proporções é:

	R
D. A.	$Z = \frac{(p_1^* - p_2^*) - (p_1 - p_2)}{\sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}} \sim N(0, 1)$
I. C.	$\left[(p_1^* - p_2^*) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}}; (p_1^* - p_2^*) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}} \right]$

Atenção que este intervalo de confiança é uma adaptação do intervalo de confiança para a diferença de médias:

$$\left[(p_1^* - p_2^*) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}}; (p_1^* - p_2^*) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1^* q_1^*}{n_1} + \frac{p_2^* q_2^*}{n_2}} \right] = \\ = \left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}; (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

onde

- $\bar{x}_1 - \bar{x}_2 = p_1^* - p_2^*$
- $s_1^2 = p_1^* q_1^*$
- $s_2^2 = p_2^* q_2^*$

Se pretender usar a função `z.test` do R, é necessário:

- considerar o sucesso = 1 e o insucesso = 0 e construir os vetores das amostras: um com 81 uns e $(270 - 81) = 186$ zeros e o outro com 116 uns e $(290 - 116) = 174$ zeros
- considerar $s_1 = \sqrt{p_1^*q_1^*} = \sqrt{p_1^* \times (1 - p_1^*)}$ e $s_2 = \sqrt{p_2^*q_2^*} = \sqrt{p_2^* \times (1 - p_2^*)}$

portanto um intervalo de confiança a 99% para a diferença entre as proporções de consumidores de Setúbal e Faro que não comprariam os produtos é

$$]-0.2032, 0.0032[$$

Com 99% de confiança, não é possível afirmar que existem diferenças na proporção de consumidores dos dois distritos que não comprariam o produto, pois o intervalo de confiança inclui o zero, $0 \in]-0.2032, 0.0032[$.

(Observação: Se $p_1 - p_2 = 0$, tem-se $p_1 = p_2$)

Ou seja, com 99% de confiança, não é possível dizer que o que foi observado nas amostras ($p_1^* < p_2^*$) mantém-se na população.

Como é óbvio, é sempre possível calcular o intervalo de confiança pretendido sem recorrer à função z.test. Basta fazer:

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.01}{2}} = z_{0.995} = 2.576$$

portanto um intervalo de confiança a 99% para a diferença entre as proporções de consumidores de Setúbal e Faro que não comprariam os produtos é

$$\left[(0.3 - 0.4) - 2.576 \times \sqrt{\frac{0.3 \times (1-0.3)}{270} + \frac{0.4 \times (1-0.4)}{290}}, (0.3 - 0.4) + 2.576 \times \sqrt{\frac{0.3 \times (1-0.3)}{270} + \frac{0.4 \times (1-0.4)}{290}} \right] = \\ = [-0.2032, 0.0032]$$

Com 99% de confiança, não é possível afirmar que existem diferenças na proporção de consumidores dos dois distritos que não comprariam o produto, pois o intervalo de confiança inclui o zero, $0 \in [-0.2032, 0.0032]$.

(Observação: Se $p_1 - p_2 = 0$, tem-se $p_1 = p_2$)

Ou seja, com 99% de confiança, não é possível dizer que o que foi observado nas amostras ($p_1^* < p_2^*$) mantém-se na população.

Exemplo 12

Considere a base de dados "iris" disponível no R. Esta base de dados foi construída por Edgar Anderson com o objetivo de quantificar a variação morfológica das flores da íris de três espécies: setosa, versicolor e virginica. A tabela contém cinco colunas: comprimento da sépala, largura da sépala, comprimento da pétala, largura da pétala e espécie.

Base de dados das Flores de Íris

Iris flower dataset

Setosa



Tra Monk, CC BY-SA 4.0, via Wikimedia Commons

Versicolor



Charles de Mille-Falets from McIvor-Ides, Canada, CC BY 2.0, via Wikimedia Commons

Virginica



Rober H. Miklosbrek, Courtesy of USDA NRCS, Public domain, via Wikimedia Commons

- Calcule um intervalo de confiança a 90% para o comprimento médio da sépala supondo que a população é Normal.

- **População**

X = comprimento da sépala

$$X \sim N(\mu, \sigma)$$

parâmetros:

média populacional = μ

desvio padrão populacional = σ

- **Amostra Aleatória**

dimensão = $n = 150$

estimativas:

média amostral = $\bar{x} = 5.8433$

desvio padrão amostral = $s = 0.8281$

Pretende-se um intervalo de confiança a 90% para μ :

- grau de confiança = $1 - \alpha = 0.90$
- nível de significância = $\alpha = 0.10$

Como temos População Normal, σ desconhecido e dimensão da amostra $n = 150 \geq 30$, há duas possibilidades para calcular o intervalo de confiança pretendido:

Condições	Distribuição Amostral	Intervalo de Confiança	R
P. Normal σ desconhecido	$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$	$\left[\bar{x} - t_{1-\frac{\alpha}{2}}; n-1 \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}}; n-1 \frac{s}{\sqrt{n}} \right]$	<code>t.test()</code>
P. Qualquer σ desconhecido $n \geq 30$	$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim N(0, 1)$	$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$	<code>z.test()</code> <code>library(BSDA)</code>

O intervalo de confiança a 90% para o comprimento médio da sépala é

- usando o teste Z:

$$]5.7321, 5.9545[$$

- usando o teste T:

$$]5.7314, 5.9552[$$

Exemplo 12

Considere a base de dados "iris" disponível no R. Esta base de dados foi construída por Edgar Anderson com o objetivo de quantificar a variação morfológica das flores da íris de três espécies: setosa, versicolor e virginica. A tabela contém cinco colunas: comprimento da sépala, largura da sépala, comprimento da pétala, largura da pétala e espécie.



- 2 Recorrendo a um intervalo de confiança a 90% verifique se há diferenças no comprimento médio da sépala das íris setosa e das íris versicolor.

● **População 1**

X_1 = comprimento das sépalas das íris setosa

média populacional = μ_1

desvio padrão populacional = σ_1

● **População 2**

X_2 = comprimentos das sépala das íris versicolor

média populacional = μ_2

desvio padrão populacional = σ_2

● **Amostra Aleatória da População 1**

dimensão = $n_1 = 50$

estimativas:

média amostral = $\bar{x}_1 = 5.006$

desvio padrão amostral = $s_1 = 0.3525$

● **Amostra Aleatória da População 2**

dimensão = $n_2 = 50$

estimativas:

média amostral = $\bar{x}_2 = 5.936$

desvio padrão amostral = $s_2 = 0.5162$

Pretende-se um intervalo de confiança a 90% para $\mu_1 - \mu_2$:

- grau de confiança = $1 - \alpha = 0.90$
- nível de significância = $\alpha = 0.10$

Populações Quaisquer, σ_1 e σ_2 desconhecidos, amostras independentes e $n_1 = 50 \geq 30$ e $n_2 = 50 \geq 30$, agora só há uma possibilidades para calcular o intervalo de confiança pretendido:

		R
D. A.	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$	z.test()
I. C.	$\left[(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$	library(BSDA)

Observação: Se no enunciado dissesse que as populações podiam ser consideradas normais, então voltávamos a ter 2 possibilidades: teste T e teste Z.

O intervalo de confiança a 90% para a diferença dos comprimentos médios das sépalas das íris setosa e versicolor é

$$]-1.0754, -0.7846[$$

Com 90% de confiança, os comprimentos médios das sépalas das íris setosa e versicolor não podem ser considerados iguais pois o intervalo de confiança não inclui o zero, $0 \notin]-1.0754, -0.7846[$.

(Observação: Se $\mu_1 - \mu_2 = 0$, tem-se $\mu_1 = \mu_2$)

Com 90% de confiança, o comprimento médio da sépala da íris setosa deve ser inferior ao comprimento médio da sépala da íris versicolor pois o intervalo de confiança só tem valores negativos.

(Observação: Se $\mu_1 - \mu_2 < 0$, tem-se $\mu_1 < \mu_2$)

Ou seja, com 90% de confiança, o que foi observado na amostra recolhida ($\bar{x}_1 < \bar{x}_2$) parece manter-se na população ($\mu_1 < \mu_2$).