

13 Inferência Estatística – Teoria da Estimação

13.1 Introdução

A partir deste capítulo abordaremos situações em que o interesse está em obter informações da população a partir dos resultados de uma amostra. Como exemplo, consideremos que uma empresa comercializa um programa educacional para escolas e afirma que o percentual de aceitação é de 85%. Para testar a afirmação da empresa, realizou-se uma experiência com uma amostra de 50 escolas, verificando-se a aceitação em 42 delas.

Outro exemplo trata-se de um psiquiatra interessado em determinar o tempo de reação de um medicamento anti-depressivo em crianças. Uma amostra aleatória de crianças que utilizaram o medicamento é obtida e analisamos o tempo médio de reação. Nestes dois exemplos, deseja-se avaliar uma afirmação sobre o parâmetro ou determinar o valor de um parâmetro por meio de uma amostra.

A seguir vamos definir alguns conceitos básicos de inferência estatística.

- **Parâmetro:** é uma medida numérica, em geral desconhecida, que descreve uma característica de interesse da população. São representados, geralmente, por letras gregas tais como, μ (média populacional) e σ (desvio-padrão populacional), entre outros. Neste texto, usaremos a letra p para representar a proporção populacional.
- **Estatística:** é qualquer valor calculado a partir dos dados amostrais. Por exemplo, $\bar{x}$ (média amostral), s (desvio-padrão amostral), $\hat{p}$ (proporção amostral), etc. A estatística é uma variável aleatória, no sentido que (a) é uma quantidade incerta (antes de obter a amostra não sabemos seu valor) e (b) seu valor varia de amostra para amostra. É claro que, quando uma amostra é selecionada, e uma estatística é calculada, torna-se então uma constante, ou seja, é o resultado de uma variável aleatória.
- **Estimador e Estimativa:** uma estatística destinada a estimar um parâmetro populacional é chamada estimador. Dada uma amostra, o valor assumido pelo estimador é chamado de estimativa ou valor estimado do parâmetro. As estimativas obtidas por meio da estatística variam de acordo com a amostra selecionada.

Os procedimentos de inferência estatística compreendem duas metodologias. Uma é chamada de estimação, na qual nós usamos o resultado amostral para estimar o valor desconhecido do parâmetro; a outra, é conhecida como teste de hipóteses, em que nós usamos o resultado amostral para avaliar se uma afirmação sobre o parâmetro (uma hipótese) é sustentável ou não. Teoria de estimação é o assunto principal deste capítulo e teste de hipóteses será retomado no próximo capítulo.

É importante estudar as propriedades do estimador, para avaliá-lo, ou seja, para poder responder a pergunta: Será que é um bom estimador para o parâmetro? Essas propriedades estão baseadas na distribuição de probabilidades do estimador, chamada de distribuição amostral.

13.2 Distribuições Amostrais

Na seção anterior, vimos que as estatísticas e, portanto os estimadores são variáveis aleatórias. A distribuição de probabilidades de uma estatística é conhecida como distribuição amostral e seu desvio-padrão é referido como erro padrão.

Uma forma de obter a distribuição amostral de um estimador é pensarmos em todas as amostras possíveis de tamanho n que podem ser retiradas da população, usando, por exemplo, amostragem aleatória simples com reposição.

Quando obtemos todas as amostras de tamanho n da população encontramos estimativas diferentes para o mesmo parâmetro, porém, em média, são iguais ao parâmetro; e tendem a ser mais homogêneas com o aumento do tamanho da amostra. Este resultado é tão surpreendente, que torna possível o uso de uma amostra para estimar os parâmetros da população. Se esta fosse a única forma de obter a distribuição amostral, o processo de inferência ficaria inviável para populações reais, pois é necessário obter todas as possíveis amostras de tamanho n para construir a distribuição amostral. Felizmente, pela teoria de probabilidades podemos mostrar que se X tem distribuição Normal, a média amostral, $\bar{x}$ também tem distribuição Normal.

- Distribuição amostral das médias

Considere todas as amostras possíveis de tamanho n retiradas de uma população e, para cada amostra, determine, por exemplo, a sua média e então reponha as observações na população. O resultado é uma série de médias de amostras de tamanho n . Se cada média da série for agora tratada como uma observação individual e colocadas em uma distribuição de frequências, então, obtém-se uma distribuição amostral das médias.

A média da distribuição amostral das médias é igual a média populacional, μ e o desvio-padrão da distribuição amostral das médias é $\sigma / \sqrt{n}$.

De maneira análoga, podemos determinar a distribuição amostral de diversas características tais como: a distribuição amostral das proporções que estuda a forma como uma característica aparece nas amostras; distribuição amostral da diferença entre as médias etc.

13.3 Teorema Central do Limite

Uma razão para a distribuição Normal ser considerada tão importante é porque qualquer que seja a distribuição da variável de interesse para grandes amostras ($n \geq 30$), a distribuição da média amostral $\bar{x}$ é aproximadamente Normal com média μ e desvio padrão $\sigma / \sqrt{n}$, e tenderão a uma distribuição normal à medida que o tamanho de amostra aumenta. Então podemos ter uma variável original com uma distribuição muito diferente da Normal, mas se tomarmos várias amostras grandes desta distribuição, então fizermos um histograma das médias amostrais, a forma se parecerá como uma curva Normal.

A aproximação para a Normal melhora à medida que o tamanho amostral aumenta. Este resultado é conhecido como o Teorema Central do Limite e é notável porque permite-nos conduzir alguns procedimentos de inferência sem qualquer conhecimento da distribuição da população.

13.4 Estimação de Parâmetros

Quando estamos interessados em determinado parâmetro de uma população, retiramos uma amostra dessa população, estudamos seus elementos e procuramos por meio dessa amostra estimar o parâmetro populacional.

Exemplo: Um candidato a deputado estadual pode querer avaliar a proporção de eleitores de seu estado que o favorecem, consultando uma amostra de 100 eleitores. A proporção de eleitores da amostra favoráveis a ele servirá como estimativa da correspondente proporção populacional, que só será conhecida após as eleições.

Vamos distinguir dois casos de estimação de parâmetros: por ponto e por intervalo.

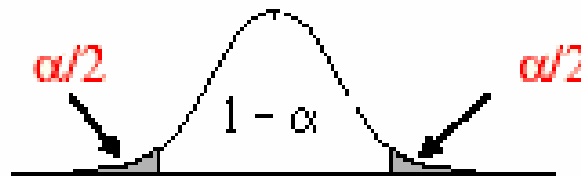
- **Estimação por ponto (pontual)**

Quando a partir de uma amostra representativa da população, o pesquisador procura obter um único valor numérico para o parâmetro de interesse.

- **Estimação intervalar (intervalo de confiança)**

Ao invés de estimar o parâmetro de interesse por um único valor, é muito mais informativo estimá-lo por um intervalo de valores que considere a variação presente na amostra e que contenha o seu verdadeiro valor com determinada confiança. Este intervalo é chamado de intervalo de confiança. Para construir um intervalo de confiança precisamos conhecer a distribuição de probabilidade do estimador. Lembre que um estimador é uma variável aleatória e que uma variável aleatória é completamente caracterizada por sua distribuição de probabilidade.

Um intervalo de confiança está associado a um grau de confiança que é uma medida da nossa certeza de que o intervalo contém o parâmetro populacional. A definição de grau de confiança utiliza α para descrever uma probabilidade que corresponde a uma área. Veja a figura abaixo, onde a probabilidade α está dividida igualmente entre duas regiões extremas sombreadas (geralmente chamadas caudas) na distribuição normal padronizada.



Grau de confiança
É a probabilidade $1 - \alpha$ (comumente expressa como valor percentual equivalente) de o intervalo de confiança conter o verdadeiro valor do parâmetro populacional.

Observações:

- Nível de significância (α) do intervalo é igual ao complemento de seu grau de confiança; portanto, mede a probabilidade de estarmos errando a estimativa.
- A escolha do nível ou grau de confiança ($1 - \alpha$) depende da precisão com que desejamos estimar o parâmetro. Evidentemente que o aumento de confiança no intervalo, implica no aumento de sua amplitude, que pode muitas vezes não interessar em termos práticos.

13.5 Estimação da Proporção Populacional

Nesta seção, as variáveis são referentes a contagens, como o número de pessoas que se sentem mal após tomar uma vacina, número de unidades defeituosas em um lote de determinado produto manufaturado, número de usuários de um site, e assim por diante. Abordamos a estimativa pontual do parâmetro p e em seguida, construiremos estimativas intervalares.

Da mesma forma que um conjunto de médias amostrais são distribuídas nas proximidades da média populacional, as proporções amostrais $\hat{p}$ são distribuídas ao redor da verdadeira proporção populacional p . Devido ao Teorema Central do Limite, para n grande e p não muito próximo de 0 ou 1, a distribuição de $\hat{p}$ será aproximadamente Normal com média p e um desvio-padrão dado

por $\sqrt{\frac{pq}{n}}$ que chamamos de erro padrão da proporção amostral.

Estimativa Pontual

Consideremos o caso em que o parâmetro a ser estimado é a proporção p de indivíduos em uma população, que apresentam uma certa característica. Retira-se da população uma amostra de tamanho n . X será o número de elementos da amostra que apresentam a característica em estudo.

Um estimador da proporção populacional p será a proporção amostral $\hat{p}$.

$$\hat{p} = X/n$$

Exemplo 1: A prefeitura de uma cidade pretende avaliar a aceitação de certo projeto educacional. Depois de apresentá-lo aos moradores do município, os responsáveis por sua execução desejam avaliar o valor aproximado do parâmetro p = proporção de favoráveis ao projeto, dentre os indivíduos residentes no município. Para estimar este parâmetro, a prefeitura planeja observar uma amostra aleatória simples de $n = 400$ moradores e calcular o valor da estatística $\hat{p}$ = proporção de favoráveis ao projeto na amostra. Por exemplo, se na amostra 240 favoráveis, temos a seguinte estimativa pontual para o parâmetro p :

$$\hat{p} = X/n = 240/400 = 0,60$$

Contudo, não devemos esperar que este valor coincida com o valor do parâmetro p , pois haverá uma variação devido ao que chamamos de erro amostral.

☞ Intervalo de Confiança

O desvio-padrão da distribuição amostral das proporções, σ_p , também conhecido como erro padrão, pode ser estimado pelos dados da amostra.

$$s_p = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

A Tabela 13.1 foi construída a partir da Tabela Normal Padrão, que associa os níveis de confiança mais usados, com valores de $z_{\alpha/2}$.

Tabela 13.1- Parte de uma tabela da Normal Padrão

Área	0,800	0,850	0,900	0,950	0,980	0,990	0,995	0,998
$z_{\alpha/2}$	1,28	1,44	1,65	1,96	2,33	2,58	2,81	3,09

Fixado o nível de confiança, podemos obter o correspondente valor de $z_{\alpha/2}$. E, a partir daí, calcular erro máximo da estimativa.

Definição

Quando utilizamos dados amostrais para estimar uma proporção populacional p , a **margem de erro**, denotado por **E**, é a diferença máxima provável (com probabilidade $1 - \alpha$) entre a proporção amostral observada e a verdadeira proporção populacional. A margem de erro **E** é chamada também **erro máximo da estimativa** e pode ser obtida multiplicando-se o valor crítico pelo desvio-padrão da distribuição amostral das proporções, conforme a fórmula abaixo:

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

O intervalo de confiança para p :

Intervalo de Confiança para a Proporção Populacional p

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

Exemplo 1: (continuação) - Admita que na amostra de $n = 400$ elementos, encontramos 60% de favoráveis. Temos, então, $\hat{p} = 0,60$ e o erro padrão dado por:

$$s_p = \sqrt{\frac{(0,60).(0,40)}{400}} = 0,024$$

Usando nível de confiança de 95%, temos um erro da estimativa de:

$$1,96 \cdot s_p = 1,96.(0,024) = 0,048.$$

O intervalo de 95% de confiança para p :

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$0,60 - 0,048 \leq p \leq 0,60 + 0,048$$

$$0,552 \leq p \leq 0,648$$

Desta forma, podemos dizer que o intervalo 0,552 a 0,648 contém, com 95% de confiança, o parâmetro p = proporção de favoráveis em toda a população de moradores do município.

Exemplo 2: Uma amostra aleatória de 400 clientes de um supermercado inclui 288 que utilizam regularmente cupons de desconto. Construa um intervalo de confiança de 98% para a proporção dos clientes do supermercado que utilizam os cupons.

Exemplo 3: Uma amostra aleatória de 500 consumidores revelou que 100 consumidores tinham planos de adquirir um novo carro durante o próximo ano. Construa um intervalo de confiança de 90% para a proporção de consumidores com os mesmos planos.

13.6 Estimação da Média Populacional (μ)

Consideremos uma população em que há interesse em estimar, μ , a média de X , uma variável aleatória. Suponhamos que uma amostra aleatória de tamanho n foi selecionada da população para obter uma estimativa de μ .

O estimador $\bar{x}$ é aquele que apresenta as melhores propriedades para estimar média populacional, μ .

O erro padrão da média amostral pode ser estimado, a partir do desvio-padrão amostral:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- **Amostras Grandes**

Pelo Teorema Central do Limite vimos que para uma amostra suficientemente grande a distribuição das médias amostrais em torno da média populacional é Normal com desvio-padrão $\sigma / \sqrt{n}$. Chamamos de $\sigma / \sqrt{n}$ o erro padrão da média, uma vez que quanto menor seu valor tanto mais próximas estarão as médias amostrais da média populacional μ .

Quando temos uma amostra grande, podemos avaliar o erro máximo ou margem de erro da estimativa por:

$$E = z_{\alpha/2} \frac{s}{\sqrt{n}}$$

em que $z_{\alpha/2}$ pode ser obtido da Tabela 13.1, em função do nível de confiança desejado.

Intervalo de Confiança para a Média Populacional μ
$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$

Este tipo de intervalo de confiança para a média pode ser usado para grandes amostras, independentemente da distribuição da variável original.

Exemplo 1: Para estudar o efeito da merenda escolar, introduzida nas escolas de um grande município, planeja-se acompanhar uma amostra de $n = 100$ crianças, que estão entrando na rede municipal de ensino. Dentre diversas características de interesse, pretende-se avaliar o parâmetro μ = ganho médio de peso, dentre todas as crianças da rede municipal de ensino, durante o primeiro ano letivo. Observando a amostra aleatória simples de $n = 100$ crianças, encontraram-se as seguintes estatísticas relativas à variável ganho de peso ao longo do ano.

- Ganho médio de peso das crianças da amostra: $\bar{x} = 6$ kg;
- Desvio-padrão dos pesos das crianças na amostra $s = 2$ kg.

Com o objetivo de estimar o parâmetro μ = ganho médio de peso da população, podemos calcular uma estimativa para o erro padrão da média amostral

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$s_{\bar{x}} = \frac{2}{\sqrt{100}} = 0,2$$

e o erro máximo da estimativa usando 95% de confiança é:

$$1,96 (0,2) = 0,392$$

O intervalo de 95% de confiança para μ :

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$6 - 0,392 \leq \mu \leq 6 + 0,392$$

Ou seja, a partir do acompanhamento da amostra das cem crianças, chegamos a conclusão que o intervalo é

$$5,608 \leq \mu \leq 6,392$$

Podemos dizer então que o intervalo entre 5,608 e 6,392 contém o ganho médio de peso, com 95% de confiança.

Exemplo 2: A média de uma amostra de 43 medidas é 26,3 e o desvio-padrão, 1,9. Calcule um intervalo de 98% de confiança para a média dessa população.

- **Amostras Pequenas**

Quando dispomos de uma amostra pequena ($n < 30$), não temos a garantia da aplicação do TCL, portanto a distribuição amostral da média pode não estar próxima da distribuição normal. A teoria estatística mostra que é possível construir estimativas intervalares para média populacional, μ , utilizando uma certa distribuição, denominada de **t de Student**, que também é tabelada.

A distribuição **t**, tem forma parecida com a Normal Padrão, sendo um pouco mais dispersa. Esta dispersão varia com o tamanho da amostra, sendo bastante dispersa para amostras pequenas, mas se aproximando da Normal Padrão para amostras grandes. Em geral, a distribuição é apresentada em função de um parâmetro, denominado graus de liberdade, gl , definido, no caso de estimação de uma média, por $gl = n - 1$.

Para obter o valor **t** da distribuição t de Student, basta calcular os graus de liberdade: $gl = n - 1$, fixar o nível de confiança desejado e usar a tabela.

Intervalo de Confiança para a Média Populacional μ (Com base em uma amostra pequena e σ desconhecido)
$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}$

Condição: Para que a distribuição t de Student seja aplicável, a distribuição da população básica deve ser essencialmente normal; não precisa ser exatamente normal, mas se tem apenas uma moda e é basicamente simétrica, obtemos em geral, bons resultados com intervalos de confiança precisos. Se há forte evidência de que a população tem distribuição bastante assimétrica, então uma alternativa é utilizar métodos não-paramétricos.

Pode parecer um pouco estranho que, com uma população distribuída normalmente, venhamos eventualmente a utilizar a distribuição t para achar os valores associados ao nível de confiança; mas quando σ não é conhecido, a utilização de s de uma amostra pequena incorpora outra fonte de erro. Para manter o grau desejado de confiança compensamos a variabilidade adicional ampliando o intervalo de confiança por um processo que substitui o valor de $z_{\alpha/2}$ por um valor maior $t_{\alpha/2}$.

Exemplo 1: Para verificar a eficácia de um programa de prevenção de acidentes de trabalho, fez-se um estudo experimental, implementando este programa em dez empresas da construção civil, escolhidas ao acaso numa certa região. Os dados abaixo referem-se aos percentuais de redução de acidentes de trabalho, na amostra destas dez empresas em que o programa foi implementado.

Amostra	Estatísticas
20 15 23 11 29 5 20 22 18 17	Média: $\bar{x} = 18$
	Desvio-padrão: $s = 6,65$

O objetivo é estimar o parâmetro μ = média da redução percentual de acidentes de trabalho, devido ao programa preventivo, em todas as empresas da construção civil da região. Podemos obter uma estimativa para o erro padrão da média, como segue:

$$s_{\bar{x}} = \frac{6,65}{\sqrt{10}} = 2,103$$

Usando nível de 95% de confiança, graus de liberdade $gl = 9$ (pois, $n = 10$ e $gl = n - 1$), obtemos na tabela o valor $t_{\alpha/2} = 2,26$, donde podemos calcular o erro máximo da estimativa:

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = 2,26 \frac{6,65}{\sqrt{10}} = 4,753$$

Então, temos o seguinte intervalo de 95% de confiança para o parâmetro μ :

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$18 - 4,753 \leq \mu \leq 18 + 4,753$$

$$13,247 \leq \mu \leq 22,753.$$

Exemplo 2: A média e o desvio-padrão de uma amostra com 19 medidas foram determinados como sendo 24,7 e 1,8, respectivamente. Determine um intervalo de 98% de confiança para a média populacional correspondente.

Exemplo 3: O lucro por carro novo vendido por uma agência de automóveis varia de carro para carro, devido à variação dos abatimentos dados ao comprador. A tabela a seguir registra o lucro com a venda da semana passada. Determine um intervalo de 90% de confiança para o lucro médio por venda de carros.

Lucro nas vendas
210
300
120
620
450
510

Exercícios:

1. Uma pesquisa efetuada com 130 funcionários selecionados ao acaso de uma empresa revelou que 52 deles não mantinham convênio com alguma empresa particular de assistência médica. Construa um intervalo de confiança de 90% para a proporção de funcionários que mantém convênio com alguma empresa particular de assistência médica.

2. Entrevistaram em uma cidade 1500 pessoas em idade de trabalho e constata-se que 145 estão desempregadas. Construir um intervalo de 95% de confiança para a taxa de desemprego.

3. Uma máquina automática de refrigerantes foi projetada para descarregar, em média, 7 onças de bebida em cada copo. Para testar essa máquina, retiram-se 10 copos de refrigerante, medindo o conteúdo de cada um. A média e o desvio-padrão obtidos foram, respectivamente, 7,1 e 0,12 onças de bebida. Determine um intervalo de 90% de confiança para a descarga média da máquina.

4. Calcule o intervalo de confiança para a média em cada um dos casos abaixo:

	Média Amostral	Tamanho da amostra	Desvio-padrão	Grau de confiança
a.	170	100	15	95%
b.	165	184	30	85%
c.	180	225	30	80%
d.	3,2	18	0,447	95%

5. Uma amostra aleatória de 625 donas de casa revela que 70% delas preferem a marca K de detergente. Construir um intervalo de 90% de confiança para p = proporção das donas de casa que preferem K.

6. Testou-se uma amostra de 36 cigarros de uma certa marca, a fim de verificar o teor de nicotina presente. Essa amostra indicou uma média de 22 e um desvio-padrão de 4. Determine um intervalo de 98% de confiança para μ , a média efetiva do teor de nicotina dessa marca de cigarro.

7. Dada uma amostra de tamanho 150 e uma proporção amostral de 0,32.

a. Calcule o erro-padrão da proporção

b. Construa um I.C. de 95%

c. Qual é a estimativa pontual?

8. Calcule o erro-padrão da média com a seguinte amostra composta de 32 escores:

4	2	5	6	1	1	7	8	7	8	8	2	6	5	6	4
4	3	6	6	7	1	5	7	8	8	4	5	3	2	6	5

9. Com a média amostral do problema anterior, calcule:

a. Intervalo de Confiança de 95%

b. Intervalo de Confiança de 99%

10. Calcule o erro-padrão da média com a seguinte amostra composta de 30 escores:

3	3	2	1	5	4	5	1	6	3	2	1	1	2	3
5	3	3	2	2	3	2	4	6	1	1	3	4	3	4

11. Com a média amostral do problema anterior, calcule:

a. Intervalo de confiança de 95%

b. Intervalo de confiança de 99%

12. Suponha que utilizemos uma amostra para estimar uma proporção populacional p . Determine a margem de erro que corresponde aos valores dados de n , x e ao grau de confiança:

a. $n = 800$, $x = 200$ e $1 - \alpha = 95\%$

b. $n = 4275$, $x = 2576$ e $1 - \alpha = 98\%$

13.7 Intervalo de Confiança para a Diferença entre Duas Médias Populacionais μ_1 e μ_2

Em muitas situações há necessidade de comparar duas populações diferentes. Como exemplos, podemos ter interesse em saber: se idosos que praticam exercícios físicos diariamente apresentam nível de colesterol menor do que idosos, com as mesmas condições, mas que não praticam exercícios físicos diariamente; se um tipo de equipamento tem maior durabilidade do que outro; e assim por diante. A seguir vamos utilizar um método para construir um intervalo de confiança para a diferença entre duas médias, com duas amostras independentes.

Para tanto, uma amostra aleatória é selecionada independentemente de cada população.

Seja X a variável aleatória a ser comparada. Suponhamos que X tenha média μ_1 e desvio-padrão σ_1 na População 1 e tenha média μ_2 e desvio-padrão σ_2 na População 2. Suponhamos também que as amostras tenham tamanhos n_1 e n_2 para as Populações 1 e 2, respectivamente.

☞ **1º caso:** quando os desvios- padrão populacionais σ_1 e σ_2 são conhecidos ou (se n_1 e $n_2 \geq 30$).

Portanto, o intervalo de confiança para $\mu_1 - \mu_2$ será dado por:

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

em que o valor de $z_{\alpha/2}$ é encontrado na tabela da distribuição Normal Padrão, conforme o nível de confiança do intervalo.

Quando os parâmetros, σ_1 e σ_2 , são desconhecidos podem ser substituídos por s_1 e s_2 , os desvios-padrão amostrais.

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Exemplo 1: Duas populações normais tem desvios-padrão de 1,7 e 1,3, respectivamente. Da população 1 foi extraída uma amostra de 25 elementos obtendo-se $\bar{x}_1 = 42$ e da população 2 foi extraída uma amostra de 16 elementos obtendo-se $\bar{x}_2 = 35$. Construir o I.C. de 90 % para a diferença de médias.

Observação: Quando o zero pertence ao intervalo de confiança, há forte evidência de que não há diferença significativa entre as duas médias populacionais.

☞ **2º caso:** quando os desvios-padrão populacionais σ_1 e σ_2 são desconhecidos e (se n_1 e $n_2 < 30$).

Assumimos normalidade para as distribuições populacionais, como no caso de uma amostra. Além disso, devemos assumir aqui que $\sigma_1 = \sigma_2$, isso quer dizer que as duas populações tem o mesmo desvio-padrão, digamos, σ . A variância populacional, pode ser estimada por:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

que representa uma variância combinada, pois é uma média ponderada das duas variâncias amostrais, s_1^2 e s_2^2 .

O intervalo de confiança para $\mu_1 - \mu_2$ será dado por:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Exemplo 2: Um experimento realizado com funcionários de uma empresa consistia em distribuir porções de café solúvel entre consumidores de café. Essas porções deveriam ser misturadas com água quente e tomadas antes de dormir. Os batimentos cardíacos deveriam ser medidos antes e depois de tomar o café e a diferença registrada. Um grupo recebeu porções de café com cafeína e outro, de um placebo. Os dados da diferença entre os batimentos cardíacos estão apresentados abaixo. Construir o intervalo de confiança de 90% para a diferença entre as médias.

Café com cafeína	Placebo
2	1
5	2
9	3
8	6
5	5
6	6
4	4
7	3
8	6
9	1
5	4
3	
2	

Exercícios:

1. Os preços que duas ações atingiram ao final do dia foram anotados durante um período de 14 dias. As médias e variâncias obtidas foram:

Ação A	Ação B
$\bar{x}_A = 45,33$ e $s^2_A = 1,54$	$\bar{x}_B = 43,54$ e $s^2_B = 2,96$

Determine um intervalo de confiança de 99% para a diferença entre os preços médios desses dois conjuntos de ações.

2. Realizou-se uma experiência a fim de comparar dois tipos de dieta, A e B, destinados à redução de peso. Dois grupos, cada um com 30 pessoas gordas, foram tratadas, cada um usando uma dessas dietas, sendo cada pessoa associada a uma dieta aleatoriamente. Antes e depois de um período de 30 dias de dieta, foram anotados os pesos de todas as pessoas, obtendo-se os resultados a seguir, expressos em termo da redução de peso verificada.

Dieta A	Dieta B
$\bar{x}_A = 21,3$	$\bar{x}_B = 13,4$
$s_A = 2,6$	$s_B = 1,9$

Determine um intervalo de confiança de 95% para a diferença entre a média de perda de peso das pessoas que fizeram os dois tipos de dieta.