Intervalos de confiança

Teorema do Limite Central, Intervalos de confiança e distribuição t-Student

André Gustavo dos Santos

Departamento de Informática
Universidade Federal de Viçosa

INF222 - 2022/2

– Fonte do material

O conteúdo foi preparado e gentilmente cedido para uso nesta disciplina pela profª Elizabeth Wanner, do PPGMMC, CEFET-MG.

No último tópico foram adicionadas informações e um exemplo do livro complementar da disciplina, seção 9.3.2.

Tópicos da aula

1. Amostragem
2. Estatística Inferencial
3. Teorema do Limite Central
4. Intervalos de confiança
5. Distribuição t-Student

Probabilidade X Estatística

Probabilidade
Dada a piscina de bolas, qual é a probabilidade de se obter uma certa combinação de cores?

Estatítica
Dadas as cores de poucas bolas, o que eu conheço sobre a piscina de bolas?

Inferência Estatística: usando amostras para obter conclusões sobre populações

População, Amostra e Observação

População: coleção completa de todos os possíveis elementos de interesse; pode ser um conjunto real (ex.: bolas na piscina) ou hipotético (ex.: possíveis resultados de um experimento)

Amostra: subconjunto de elementos selecionados da população, conforme o interesse (ex.: para fazer inferência sobre a população examinando ou medindo os elementos da mostra)

Observação: é um único elemento de uma amostra, um dado coletado individualmente (pode ser vista como uma amostra de tamanho 1)

Amostragem

- **Amostragem:** processo de escolha de uma amostra.
- **Erro amostral:** é o erro que ocorre justamente pelo uso da amostra em vez de toda a população.
- **Risco da amostragem:** tirar conclusões e decisões sobre toda a população tendo como base apenas uma parte dela.
- **Teoria de probabilidade:** fornecer uma ideia do risco envolvido, do erro cometido ao utilizar uma amostra no lugar da população.

Observação: em inglês, as palavras amostra (*sample*) e exemplo (*example*) têm origem comum em *essample*, do francês antigo. Embora tenham significados diferentes atualmente, é importante lembrar dessa origem comum, pois amostra é apenas um exemplo. E, assim como um exemplo geralmente não é suficiente para provar uma teoria, uma amostra geralmente não é suficiente para tirar conclusões definitivas da população.

Tipos de Amostragem

- Amostragem aleatória simples
  - cada elemento individual tem chance igual de ser selecionado
  - toda amostra possível de mesmo tamanho tem a mesma chance de ser selecionada
- Amostragem sistemática
  - define uma ordem de escolha
- Amostragem estratificada
  - separa a amostra por classes
- Amostragem por conglomerado
  - separa a amostra por grupos de classes
- Amostragem de conveniência (Evite!)
  - seleciona da forma que for mais fácil

UFV

Estatística Inferencial

- **Estatística Indutiva ou Inferencial:** Coleção de métodos e técnicas utilizados para se estudar uma população baseados em amostras probabilísticas desta população.
- **Parâmetro:** medida numérica que descreve alguma característica única de uma população.
- **Estimador ou Estatística:** medida numérica que descreve alguma característica de uma amostra.
- **Estimativa:** é um valor particular de um estimador.

Estatística Inferencial

Principais parâmetros de uma população:

- média populacional: μ
- variância populacional: σ^2
- desvio-padrão populacional: $\sigma = \sqrt{\sigma^2}$
- proporção (de elementos com uma característica) populacional: π

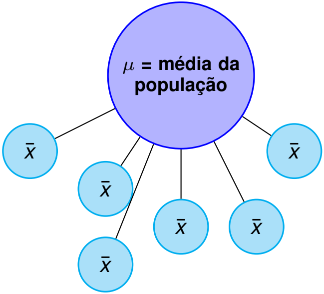
Principais estimadores de uma amostra:

- média amostral: $\bar{x} = \hat{\mu}$
- variância amostral: $s^2 = \hat{\sigma^2}$
- desvio-padrão amostral: $s = \sqrt{s^2} = \hat{\sigma}$
- proporção amostral: $p = \hat{\pi}$

Cada estimador é uma estimativa do respectivo parâmetro populacional.

Estimadores

- O estimador é uma variável aleatória, uma vez que a amostra é probabilística.
- Sendo uma variável aleatória, tem uma função de distribuição (a distribuição amostral, que será comentada mais à frente)
- A figura ilustra um parâmetro (μ, média da população) e as médias amostrais de diferentes amostras ($\bar{x}$, diferentes estimativas do parâmetro)



Estimadores

- Dois conceitos centrais da Estatística Inferencial são os estimadores de ponto e os estimadores de intervalo
- Ambos se referem a usar informação obtida de uma amostra para inferir valores prováveis para parâmetros da população
- Para um dado parâmetro da população temos
  - **Estimador de ponto:** seu valor estimado
  - **Estimador de intervalo:** intervalo estimado de seus possíveis/prováveis valores

Estimadores de ponto

- Um bom estimador deve consistentemente gerar estimativas próximas do valor real do parâmetro
- Um estimador $\hat{\theta}$ é *não-tendencioso* para o parâmetro θ se:

$$\mathbf{E}(\hat{\theta}) = \theta$$

ou seja, se ele varia em torno do parâmetro da população

- O valor $\mathbf{E}(\hat{\theta}) - \theta$ é chamado *tendência* do estimador
- Para um estimador não-tendencioso, $\mathbf{E}(\hat{\theta}) - \theta = 0$
- Os parâmetros usuais (média, variância, ...) são não-tendenciosos

Distribuição amostral

- **Distribuição amostral:** a distribuição de probabilidades de um estimador.
- Em vez de tirarmos uma única amostra de tamanho N, tiramos m amostras de tamanho n. Por exemplo, em vez de 100 observações, tiramos 20 amostras compostas por 5 observações.
- A distribuição amostral de um estimador depende da distribuição da população, do tamanho da amostra e do método de seleção da amostra.
- Se a população for muito grande ou infinita, podemos dispor de modelos probabilísticos para descrever a distribuição amostral do estimador.

Distribuição amostral - exemplo

Melhores atrizes

22	37	28	63	32	26
31	27	27	28	30	26
29	24	38	25	29	41
30	35	35	33	29	38
54	24	25	46	41	28
40	39	29	27	31	38
29	25	35	60	43	35
34	34	17	37	42	41
36	32	41	33	31	74
33	50	38	61	21	41
26	80	42	19	33	35
45	49	39	34	26	25
33	35	35	28		

Melhores atores

44	41	62	52	41	34
34	52	41	37	38	34
32	40	43	56	41	39
49	57	41	38	42	52
51	35	30	39	41	44
49	35	47	31	47	37
57	42	45	42	44	62
43	42	48	49	56	38
60	30	40	42	36	76
39	53	45	36	62	43
51	32	42	54	52	37
38	32	45	60	46	40
36	47	29	43		

	Melhores Atrizes	Melhores Atores
Média Populacional	35.4	43.9

Distribuição amostral - exemplo

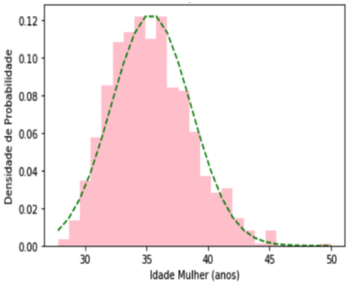
Exercício

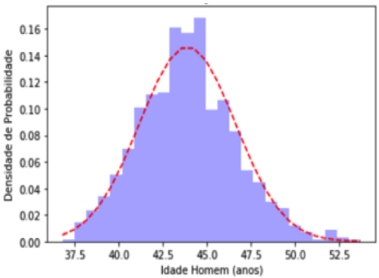
- Escolher uma amostra de tamanho $n = 10$ das idades das atrizes e outra das idades dos atores (sem repetiir escolha na mesma amostra).
- Calcular o valor da média das idades em cada amostra
- Repetir a análise encontrando várias amostras de tamanho 10.
- Estudar a distribuição das médias amostrais.

Distribuição amostral - exemplo

Exercício

- Escolher uma amostra de tamanho $n = 10$ das idades das atrizes e outra das idades dos atores (sem repetiir escolha na mesma amostra).
- Calcular o valor da média das idades em cada amostra
- Repetir a análise encontrando várias amostras de tamanho 10.
- Estudar a distribuição das médias amostrais.





- Note que seguem aproximadamente uma distribuição normal...

Teorema do Limite Central

Teorema do Limite Central (TLC)

Se $x_1, \ldots, x_n$ for uma amostra aleatória de tamanho n, retirada de uma população (finita ou infinita) com média μ e variância σ^2 e se $\bar{X}$ for a média da amostra, então a forma limite da distribuição de

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

quando $n \to \infty$, é a distribuição normal padrão.

UFV

Teorema do Limite Central

Considere uma variável aleatória X com média μ e variância σ^2 e m amostras de X de tamanho n: $x_1, \ldots, x_n$:

Ao analisarmos a distribuição das médias das amostras, temos que:

1. a média das médias amostrais tenderá para: $\mu_{\bar{x}} = \mu$
2. a variância das médias amostrais tenderá para $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$.
3. o desvio padrão das médias amostrais tenderá para $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
4. a distribuição das médias amostrais tenderá para uma distribuição normal quando n aumenta: $\bar{X} \sim N(\mu, \sigma^2/n)$.

Teorema do Limite Central

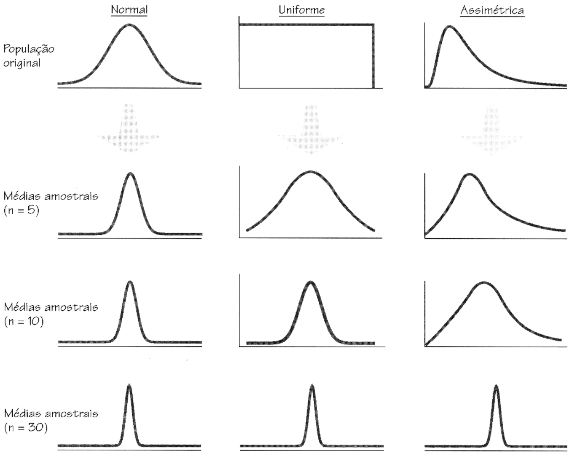
O que há de extraordinário no TLC??

Ele nos diz que, qualquer que seja a forma da distribuição original, suas médias amostrais tendem para uma distribuição normal quando o tamanho da amostra cresce.

Para encontrarmos a distribuição das médias amostrais, basta conhecermos a média da população, a variância (ou o desvio padrão) e o tamanho da amostra.

Por outro lado, permite o uso de técnicas baseadas da distribuição normal, mesmo que a população não siga distribuição normal.

Teorema do Limite Central



Teorema do Limite Central

Observações:

- Para uma população com distribuição qualquer, quanto maior o tamanho das amostras, mais próxima será a distribuição das médias amostrais de uma distribuição normal.
- Regra prática (muito conservadora): para $n \geq 30$, a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma distribuição normal.
- Se a distribuição for originalmente uma distribuição normal, então a distribuição das médias amostrais terá distribuição normal para qualquer tamanho amostral n.

Teorema do Limite Central

Exemplo

A renda de um conjunto de pessoas de uma região tem média $\mu = 6$ e desvio-padrão $\sigma = 2,5$ salários mínimos. Se desta população for extraída uma amostra ($m = 1$) de 100 pessoas ($n = 100$), qual é a probabilidade da média desta amostra acusar um valor superior a 6,3 salários mínimos?

Solução:

- $n = 100 > 30 \Rightarrow$ TLC Ok!

$$\bar{X} \sim N\left(6, \frac{2.5}{\sqrt{100}}\right) \Rightarrow Z = \frac{\bar{X} - 6}{2.5/\sqrt{100}}$$

$$\mathbf{P}\{\bar{X} > 6.3\} = \mathbf{P}\{Z > 1.2\} = \mathbf{P}\{Z < -1.2\} = 0.1151$$

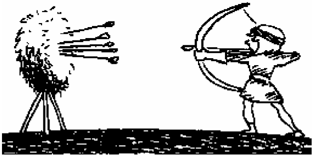
- A probabilidade da média ser superior a 6,3 salários é 11,5% (evento raro).

Estimadores de Intervalos

- A média amostral do conjunto de dados, $\bar{X}$, é a melhor estimativa para a média populacional μ.
- Entretanto, o que nos garante que a amostra usada compõe uma boa estimativa da população?
- A média amostral, $\bar{X}$, é um estimador pontual.
- Podemos associá-la a uma outra estimativa - intervalo de confiança (estimativa de intervalo).

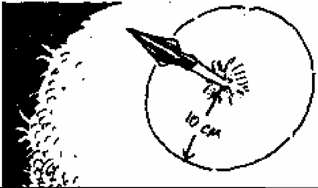
Motivação

Considere uma arqueira atirando em um alvo. Suponha que ela acerta o alvo, uma circunferência de raio 10 cm, em 95% das vezes.

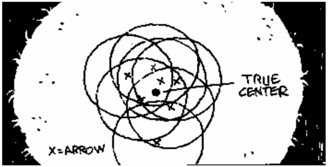


A arqueira atira a primeira flecha. Sentado atrás do alvo está um detetive, que não vê onde está o alvo.

Conhecendo o nível da habilidade da arqueira, o detetive desenha um círculo com 10 cm de raio ao redor da flecha. Ele tem 95% de confiança de que o seu círculo inclua o centro do alvo.



Ele raciocinou que se desenhasse círculos com 10 cm de raio ao redor de muitas flechas, seus círculos incluiriam o centro do alvo em 95% dos casos.



Intervalo de Confiança

Ideia: construir um intervalo de confiança (IC) para o parâmetro com uma probabilidade de $1 - \alpha$ de que o intervalo contenha o verdadeiro parâmetro:

$$\mathbf{P}\{L_I < \theta < L_S\} = 1 - \alpha.$$

Neste caso, o nível de significância α é uma estimativa para o erro cometido, mesmo sem sabermos se o intervalo contém o verdadeiro valor do parâmetro populacional.

Desejável: obter um intervalo que seja curto o suficiente para que seja possível tomar uma decisão e que, ao mesmo tempo, apresente um grau de confiança adequado, de acordo com o tamanho da amostra.

Intervalo de Confiança para a Média Populacional

Intervalo de confiança para a média populacional μ desconhecida e uma distribuição normal com a variância populacional σ^2 conhecida:

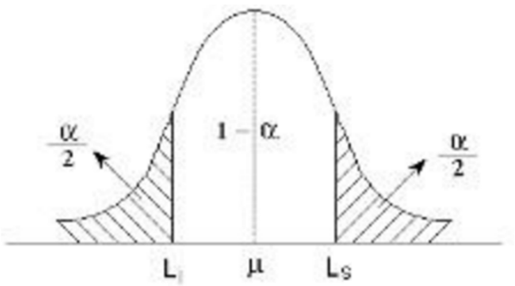
- O intervalo de confiança para μ é construído em torno da estimativa pontual da média amostral $\bar{x}$.
- Deve-se fixar uma probabilidade de $1-\alpha$ de que o intervalo encontrado contenha o verdadeiro parâmetro populacional.
- A média amostral segue uma distribuição normal: $\bar{X} \sim N(\mu, \sigma^2/n)$.
- Podemos utilizar a distribuição normal padrão Z para estabelecer os limites do intervalo de confiança:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Intervalo de Confiança para a Média Populacional

Intervalo de confiança para μ para distribuição Normal com σ^2 conhecida:

- nível de confiança de $1-\alpha$: remove os piores $\alpha/2$ dos valores em cada extremidade da distribuição
- descobrir os limites L_I e L_S para incluir $1-\alpha$ das observações, usando a distribuição $Z \sim N(0,1)$ inversamente
- ou seja, é necessário encontrar $z_{\alpha/2}$ tal que $\alpha/2$ das observações encontrem-se acima e abaixo desse valor.



Intervalo de Confiança para a Média Populacional

Intervalo de confiança desejado	valor de z usado
95%	$z_{\alpha/2} = 1.96$
99%	$z_{\alpha/2} = 2.58$
99.8%	$z_{\alpha/2} = 3.10$

Intervalo de Confiança para a Média Populacional

CASO 1

Intervalo de confiança para a média populacional μ desconhecida e uma distribuição normal com a variância populacional σ^2 conhecida:

- amostra $x_1, x_2, \cdots, x_n$, em que n é grande
- calcular a média amostral: $\bar{x}$
- usar o desvio padrão amostral populacional: σ

$$\mathbf{P}\left\{-z_{\alpha/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right\} = 1-\alpha$$

Então, o intervalo de confiança é:

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

Intervalo de Confiança para a Média Populacional

Exemplo

Uma população tem desvio-padrão igual a 10 e média desconhecida. Uma amostra de tamanho $n = 100$ é retirada e fornece uma média amostral de $\bar{x} = 50$. Encontre o intervalo de confiança de 95% para a média desta população.

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

$$\mathbf{P}\{z > z_{\alpha/2}\} = 0.025 \Rightarrow z_{\alpha/2} = 1.96$$

$$-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96$$

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

$$48.04 < \mu < 51.96$$

Intervalo de Confiança para a Média Populacional

Exemplo (cont...)

O que acontece se escolhemos um nível de confiança maior, de 99%?

$$\alpha/2 = 0.005 \Rightarrow \mathbf{P}\{z > z_{\alpha/2}\} = 0.005 \Rightarrow z_{\alpha/2} = 2.58$$

$$-2.58 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 2.58$$

$$\bar{X} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2.58\frac{\sigma}{\sqrt{n}}$$

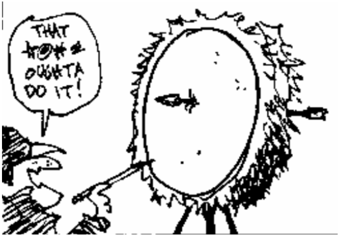
$$47.42 < \mu < 52.58$$

** Note que, com o aumento do nível de confiança, o intervalo de confiança aumentou*

Como melhorar a confiança?

Como ter mais confiança que o centro real está dentro do círculo desenhado?

Aumentando o tamanho do círculo...



ou melhorando a mira da arqueira!

Tamanho da amostra

- A precisão é inversamente proporcional ao nível de confiança.
- Se o nível de confiança aumenta, o tamanho do intervalo de confiança também aumenta.
- Para obter um intervalo de confiança que seja curto o suficiente com a confiança adequada, deve-se escolher bem o tamanho da amostra.
- Se $\bar{x}$ for escolhido como estimador de μ, podemos estar $100(1-\alpha)\%$ seguros de que o erro $E = |\bar{X} - \mu|$ não excederá um valor especificado quando o tamanho da amostra for:

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

Intervalo de Confiança

A margem de erro, o nível de confiança e o tamanho da amostra caminham lado a lado!

Modificar qualquer um altera os restantes:

- Para uma amostra fixa, reduzir a margem de erro obriga a reduzir o nível de confiança e aumentar o nível de confiança obriga a aumentar a margem de erro.
- Para manter ou reduzir a margem de erro e manter ou aumentar o nível de confiança, obriga-se a aumentar o tamanho da amostra.
- Se aumentar o tamanho da amostra, é possível reduzir a margem de erro e/ou aumentar o nível de confiança.

Idade das Melhores Atrizes e dos Melhores Atores

Melhores atrizes

22	37	28	63	32	26
31	27	27	28	30	26
29	24	38	25	29	41
30	35	35	33	29	38
54	24	25	46	41	28
40	39	29	27	31	38
29	25	35	60	43	35
34	34	17	37	42	41
36	32	41	33	31	74
33	50	38	61	21	41
26	80	42	19	33	35
45	49	39	34	26	25
33	35	35	28		

Melhores atores

44	41	62	52	41	34
34	52	41	37	38	34
32	40	43	56	41	39
49	57	41	38	42	52
51	35	30	39	41	44
49	35	47	31	47	37
57	42	45	42	44	62
43	42	48	49	56	38
60	30	40	42	36	76
39	53	45	36	62	43
51	32	42	54	52	37
38	32	45	60	46	40
36	47	29	43		

Idade das Melhores Atrizes e dos Melhores Atores

Calcule os IC de 95% para as idades com amostras de tamanho $n = 5$:

	Melhores Atrizes	Melhores Atores
Média	35.4	43.9
Desvio Padrão	11.36	9.06

Melhores atrizes

$$35.4 - 1.96 \times \frac{11.36}{\sqrt{5}} < \mu < 35.4 + 1.96 \times \frac{11.36}{\sqrt{5}}$$

$$25.44 < \mu < 45.36$$

Melhores atores

$$43.9 - 1.96 \times \frac{9.06}{\sqrt{5}} < \mu < 43.9 + 1.96 \times \frac{9.06}{\sqrt{5}}$$

$$35.96 < \mu < 51.84$$

Idade das Melhores Atrizes e dos Melhores Atores

Calcule os IC de 95% para as idades com amostras de tamanho $n = 10$:

	Melhores Atrizes	Melhores Atores
Média	35.4	43.9
Desvio Padrão	11.36	9.06

Melhores atrizes

$$35.4 - 1.96 \times \frac{11.36}{\sqrt{10}} < \mu < 35.4 + 1.96 \times \frac{11.36}{\sqrt{10}}$$

$$28.36 < \mu < 42.44$$

Melhores atores

$$43.9 - 1.96 \times \frac{9.06}{\sqrt{10}} < \mu < 43.9 + 1.96 \times \frac{9.06}{\sqrt{10}}$$

$$38.28 < \mu < 49.52$$

Idade das Melhores Atrizes e dos Melhores Atores

Encontre o tamanho da amostra para um intervalo de confiança de 95% considerando o erro $E = |\bar{X} - \mu|$ menor que 5 anos na idade das atrizes.

$$\text{No exemplo anterior, } E = |\mu - 35.4| < 1.96 \times \frac{11.36}{\sqrt{10}} = 7.04 \text{ anos}$$

$$\text{Precisaria de uma amostra de } n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 11.36}{7}\right)^2 = 10.12$$

$$\text{Para um erro menor que 5: } n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 11.36}{5}\right)^2 = 19.83 \rightarrow 20$$

$$\text{O novo intervalo é: } 30.42 < \mu < 35.4 + 1.96 \times \frac{11.36}{\sqrt{20}} = 40.38$$

Intervalo de Confiança para a Média Populacional

Intervalo de confiança para a média populacional μ de uma distribuição qualquer para amostra grande:

- Se o tamanho da amostra é grande, então o Teorema do Limite Central implica que a média amostral tende para uma distribuição Normal, mesmo em caso de variância desconhecida, uma vez que a troca da variância populacional pela amostral tem pouco efeito na distribuição de Z.
- Considere amostra grande como $n \geq 30$ se a distribuição for Normal e $n \geq 40$ em caso de distribuição qualquer.
- Desta forma, podemos utilizar a distribuição normal padrão para estabelecer os limites do intervalo de confiança.

Intervalo de Confiança para a Média Populacional

Pelo Teorema do Limite Central:

- Quando a população tiver uma distribuição desconhecida e a variância conhecida, mas o **tamanho da amostra for grande**, a média amostral tende para uma distribuição normal:
$$\bar{X} \sim N(\mu, \sigma^2/n).$$
- Quando a população tiver uma distribuição desconhecida e a variância desconhecida, mas o **tamanho da amostra for grande**, a média amostral tende para uma distribuição normal:
$$\bar{X} \sim N(\mu, s^2/n).$$

Intervalo de Confiança para a Média Populacional

CASO 2

Intervalo de confiança para a média populacional μ desconhecida de uma distribuição qualquer, com variância populacional σ^2 conhecida e amostra grande:

- amostra $x_1, x_2, \cdots, x_n$, em que n é grande
- calcular a média amostral: $\bar{x}$
- usar o desvio padrão populacional: σ

$$P\left\{-z_{\alpha/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right\} = 1-\alpha$$

Então, o intervalo de confiança é:

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

Intervalo de Confiança para a Média Populacional

CASO 3

Intervalo de confiança para a média populacional μ desconhecida de uma distribuição qualquer, com variância populacional σ^2 desconhecida e amostra grande:

- amostra $x_1, x_2, \cdots, x_n$, em que n é grande
- calcular a média amostral: $\bar{x}$
- calcular o desvio padrão amostral: s

$$P\left\{-z_{\alpha/2} < \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} < z_{\alpha/2}\right\} = 1-\alpha$$

Então, o intervalo de confiança é:

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

Intervalo de Confiança para a Média Populacional

Exemplo

Uma amostra de $n = 53$ peixes foi selecionada para investigar a contaminação por mercúrio em lagos da Flórida, em que mediu-se a concentração de mercúrio no tecido muscular. Os valores de concentração de mercúrio foram: média amostral $\bar{x} = 0.5250$ e desvio padrão amostral $s = 0.3486$. Queremos achar um intervalo de confiança aproximado de 95% para a média populacional μ.

Como $n = 53 > 40$, podemos considerar a suposição de normalidade pelo TLC. Então, o intervalo de confiança aproximado é:

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

$$0.4311 < \mu < 0.6189$$

Intervalo de Confiança para a Média Populacional

Intervalo de confiança para a média populacional μ de uma distribuição normal com a variância populacional σ^2 desconhecida e amostra pequena:

- O intervalo de confiança para μ é construído em torno da estimativa pontual $\bar{x}$ e do desvio-padrão amostral s.
- Se o tamanho da amostra é pequeno, então a variável aleatória

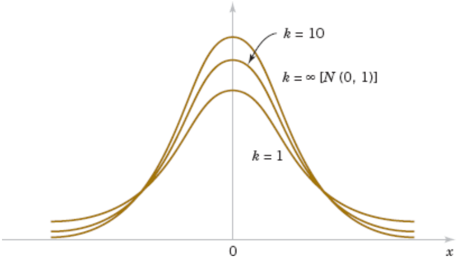
$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

tem uma distribuição t com $n - 1$ graus de liberdade.

- O número de graus de liberdade para t é o número de graus de liberdade associado ao desvio-padrão estimado: $n - 1$.

Distribuição *t*

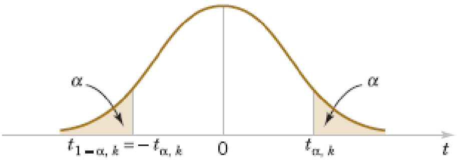
- A aparência geral da distribuição *t* é similar à da distribuição normal padrão.
- Ela é simétrica, unimodal e tem o valor máximo na média $\mu = 0$.
- Entretanto, tem extremidades (caudas) mais espessas; ou seja, ela tem mais probabilidade nas extremidades (caudas) do que a distribuição normal.
- Ela se aproxima da normal à medida que o número de graus de liberdade cresce



- No limite, a distribuição *t* é a distribuição normal padrão.

Distribuição *t*

- Pontos percentuais da distribuição *t*:



- Como a distribuição *t* tem um pico mais baixo (e caudas mais largas) que a distribuição normal, o corte de área α do lado direito acontece num valor maior
- isto é, $t_\alpha > z_\alpha$
- Assim, o intervalo de confiança $100(1-\alpha)\%$ é maior na *t* que na normal
- É o preço que se paga por não se conhecer o desvio σ da população...

Distribuição *t* - curiosidade

- A distribuição *t* foi desenvolvida pelo químico e estatístico inglês William Gosse
- Trabalhando na cervejaria Guinness, ficou interessado em estudos com amostra pequena no problema de controle de qualidade da fermentação
- Por não poder publicar com seu nome, usava o pseudônimo “Student”
- Por isso é conhecida como distribuição *t* de Student

Intervalo de Confiança para a Média Populacional

CASO 4

Intervalo de confiança para a média populacional μ de uma distribuição normal com a variância populacional σ^2 desconhecida e amostra pequena:

- amostra $x_1, x_2, \cdots, x_n$
- calcular a média amostral $\bar{X}$
- calcular o desvio padrão amostral s

$$P\left\{-t_{\alpha/2;n-1} < \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} < t_{\alpha/2;n-1}\right\} = 1-\alpha$$

Então, o intervalo de confiança é:

$$\bar{X} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}$$

Exemplo

Exemplo

Se uma pessoa não autorizada acessa um computador com o nome de usuário e a senha correta (roubada ou craqueada), é possível detectar que é um intruso?

Um método é comparar seu comportamento com os do usuário verdadeiro, por exemplo, sua velocidade de digitação e o intervalo de tempo ao pressionar as teclas.

Um intruso é detecado se houver diferença significativa.

Um usuário digitou o nome e a senha com os seguintes tempos entre as teclas:

$$.24, .22, .26, .34, .35, .32, .33, .29, .19, .36, .30, .15, .17, .28, .38, .40, .37, .27 \text{ segundos}$$

Como primeiro passo para detectar um intruso, calcule o intervalo de confianca de 99% para a média do tempo entre as teclas, assumindo uma distribuição normal.

Solução:

- A amostra tem tamanho $n = 18$
- A média amostral é $\bar{x} = 0.29$ s e o desvio padrão amostral é $s = 0.074$ s.
- Para $n - 1 = 17$ graus de liberdade temos $t_{\alpha/2;n-1} = t_{0.005;17} = 2.898$
- Temos então $\bar{X} \pm t_{\alpha/2;n-1} \frac{s}{\sqrt{n}} = 0.29 \pm (2.898) \frac{0.074}{\sqrt{18}} = 0.29 \pm 0.05 \Rightarrow [0.24, 0.34]$