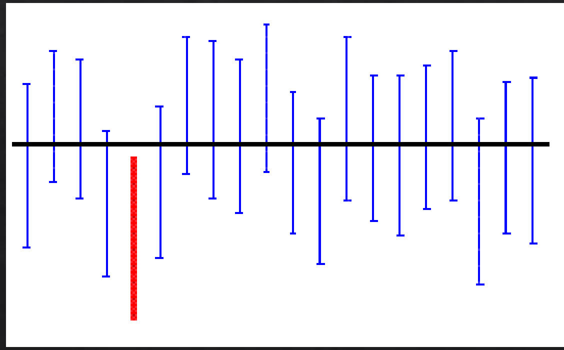
Intervalo de confiança para a média

UESC – Universidade Estadual de Santa Cruz
Probabilidade e Estatística
Hiago Rios Cordeiro
Nathan Rocha

Intervalo de confiança

Um intervalo de confiança (ou estimativa intervalar) é uma faixa de valores usada para estimar o verdadeiro valor de um parâmetro populacional. Um intervalo de confiança é, algumas vezes, abreviado como IC.[1]

O IC é associado a um nível (ou grau, ou coeficiente) de confiança, que é a proporção de vezes que o IC contém o parâmetro populacional, supondo que o processo de estimação seja repetido várias vezes. Essa é sua vantagem em relação à estimativa pontual.

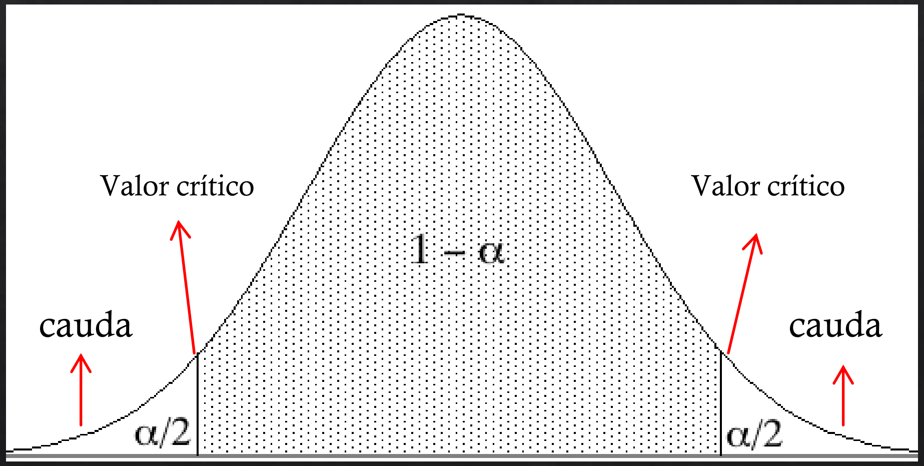


[1] TRIOLA, Mario. Introdução à Estatística. 9ª Ed, pág. 232.

Intervalo de confiança

Um IC é muitas vezes expresso como uma probabilidade ou área $\gamma = 1 - \alpha$, sendo alfa o complemento do nível de confiança gama, chamado nível de significância ou probabilidade de erro.

As escolhas mais comuns para o nível de confiança são 90% (menor largura de intervalo), 95% e 99% (maior largura de intervalo).



Valores críticos são os valores, tabelados para cada distribuição, que separam a área do IC da área de α.

Intervalo de confiança

A margem de erro E (ou erro máximo da estimativa) é a diferença máxima provável (com probabilidade 1 – α) entre o parâmetro amostral e o real valor do parâmetro populacional e é dada por:

$E = V_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$, onde $V_{\alpha/2}$ é o valor crítico da distribuição para α/2.

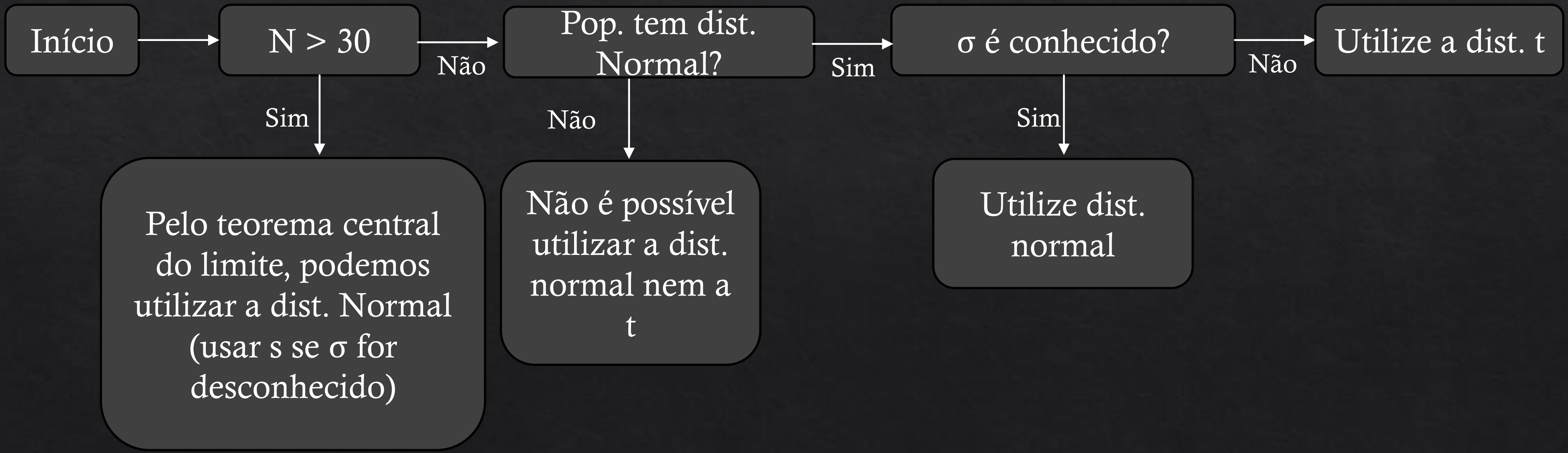
Seja θ um parâmetro populacional e θ' uma estimativa pontual desse parâmetro.

O intervalo de confiança é, em geral, escrito nos seguintes formatos:

$\theta' \pm E$ ou $\theta' - E < \theta < \theta' + E$ ou $(\theta' - E, \theta' + E)$

Intervalo de confiança para a média

- Para se fazer inferências estatísticas sobre uma população, geralmente, são utilizadas as distribuições Normal Padrão e T. Um método de escolha é:



Intervalo de confiança para a média: σ conhecido

Suposições:

1. A amostra é aleatória simples
2. O valor do desvio padrão populacional é conhecido σ
3. Pelo menos uma das condições seguintes é satisfeita: a distribuição é aproximadamente normal (simétrica, uma moda e nenhum outlier) ou $n > 30$.

Se as suposições forem satisfeitas, usamos a distribuição normal para construir o IC.

Intervalo de confiança para a média: σ conhecido

A média amostral $\bar{x}$ é a melhor estimativa pontual da média populacional. Apesar disso, ela não fornece uma indicação do quão boa ela é, por isso usamos IC.

Margem de erro para IC com σ conhecido:

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

Onde Z é o valor crítico tabelado da distribuição normal para α/2.

A estimativa intervalar será então: $\bar{x} - E < \mu < \bar{x} + E$

Exemplo – Intervalo de confiança para a média: σ conhecido

- A duração de vida de uma peça de equipamento é tal que o desvio padrão é de 5 horas. Foram amostradas 100 dessas peças obtendo-se a média de 500 horas. Deseja – se construir um intervalo de confiança para a verdadeira duração média da peça com um nível de 95% de confiança.

IC: $\bar{x} - E < \mu < \bar{x} + E$

$\bar{x}$ = 500 horas
n = 100 peças
σ = 5 horas
n >= 30
α = 0,05

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = Z_{0,025} \cdot \frac{5}{\sqrt{100}}$$

$$E = 1,96 \cdot \frac{1}{2}$$

$$E = 0,98$$

IC: $499,02 < \mu < 500,98$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	**0,4495**	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,4750**	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993

Intervalo de confiança para a média: σ desconhecido

- Em muitas situações da vida real, o desvio padrão populacional é desconhecido. Além disso, em função de fatores como tempo e custo, frequentemente não é prático colher amostras de tamanho 30 ou mais.
- Assim, como demonstrar intervalos de confiança para a média populacional nessas condições? Se a variável aleatória tem distribuição aproximadamente normal, ou seja, simétrica, com uma moda e nenhum outlier, utilizamos a distribuição t.

Distribuição T de Student

- Trata-se de um modelo de distribuição contínua de densidade de probabilidade que se assemelha a distribuição normal padrão, ou seja, com formato de sino e com área total sob a curva igual a 1, mas reflete uma maior variabilidade (com curvas mais alargadas).
- A distribuição possui um parâmetro φ (Phi) = n - 1 denominado graus de liberdade que, conforme aumenta (a partir de $n = 30$), torna a curva mais semelhante à da normal padrão.
- A distribuição T encontra-se tabelada para diferentes combinações de probabilidade e graus de liberdade.
- É utilizada para inferências estatísticas, particularmente, quando se tem amostras pequenas ($n < 30$).

Distribuição T de Student

É simétrica à sua média 0 e sua variância é dada pela expressão a seguir, implicando que σ^2 vai se reduzindo e σ se aproximando de 1 com o aumento de φ:

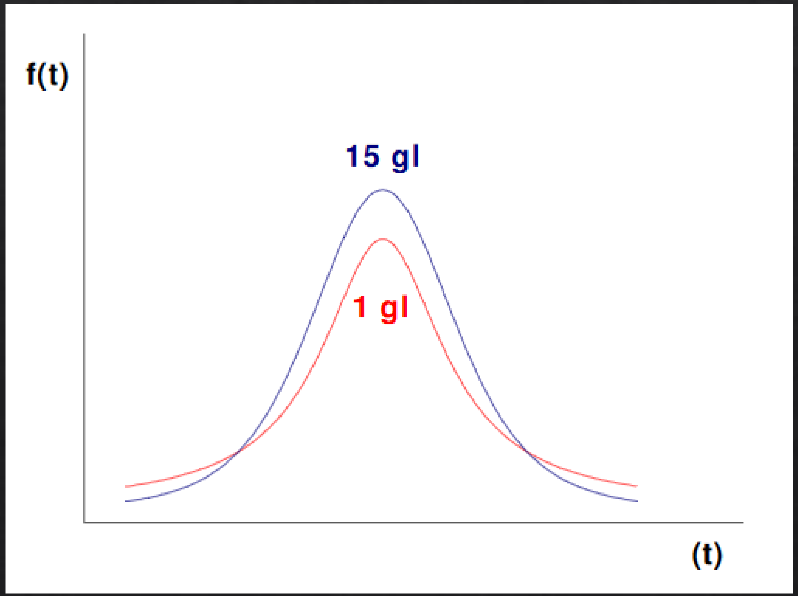
$$Var(t_\varphi) = \sigma^2\,(t_\varphi) = \frac{\varphi}{\varphi-2}, (\varphi > 2)$$

Ex:

$$\sigma(t_4) = \sqrt{\frac{4}{4-2}} = 1{,}41$$

$$\sigma(t_{35}) = \sqrt{\frac{35}{35-2}} = 1{,}03$$

$$\sigma(t_{60}) = \sqrt{\frac{60}{60-2}} = 1{,}02$$



Distribuição T de Student

- Seja Y uma variável aleatória proveniente de uma população normal, Y~N(μ, σ^2).
- Seja uma amostra aleatória simples extraída desta população.
- Sejam m e s, respectivamente, a média e o desvio padrão obtidos a partir da amostra.

❖ A variável $t = \frac{m - \mu}{s_m}$ tem distribuição t de Student com n-1 graus de liberdade.

(Script em R)

Intervalo de confiança para a média: σ desconhecido

$$P(-t_0 \le t \le t_0) = 1-\alpha$$

$$P\left(-t_0 \le \frac{m-\mu}{s_m} \le t_0\right) = 1-\alpha \quad (s_m)$$

$$P(-t_0 \cdot s_m \le m-\mu \le t_0 \cdot s_m) = 1-\alpha \quad (-1)$$

$$P(t_0 \cdot s_m \ge -m+\mu \ge -t_0 \cdot s_m) = 1-\alpha \quad (somando\ m)$$

$$P(m+t_0 \cdot s_m \ge \mu \ge m-t_0 \cdot s_m) = 1-\alpha$$

$$P(m-t_0 \cdot s_m \le \mu \le m+t_0 \cdot s_m) = 1-\alpha \qquad \therefore \qquad Como\ \ s_m = \frac{s}{\sqrt{n}}$$

$$P\left(m-t_0 \cdot \frac{s}{\sqrt{n}} \le \mu \le m+t_0 \cdot \frac{s}{\sqrt{n}}\right) = 1-\alpha$$

Temos que a área 1 - α = m ± $t_0 \cdot \frac{s}{\sqrt{n}}$, ou seja:

$$IC(\mu) = m \pm t_{\alpha\%} \cdot \frac{s}{\sqrt{n}}$$

Ou $IC(\mu) = m \pm E$

Onde:
E = margem de erro
m = média amostral
s = desvio padrão amostral
n = nº de elementos da amostral
$t_{\alpha\%}$ = valor crítico tabelado em função de gl (graus de liberdade = n-1)

Exemplo – Intervalo de confiança para a média: σ desconhecido

- Uma empresa estuda o atraso na entrega dos pedidos recebidos. Supõe-se que o atraso nas entregas possui distribuição normal e os tempos de atraso (em dias) dos últimos 20 pedidos tem média de 3,3 dias e desvio padrão de 3,0105. Qual é o IC para o atraso médio nas entregas com α = 10%?
- Como a amostra é pequena e não conhecemos o desvio-padrão populacional, utilizamos a distribuição t-student. Devemos encontrar o valor crítico t para α/2 = 0,05 e graus de liberdade (n - 1) = 19:

IC: $\bar{x}$ - E < μ < $\bar{x}$ + E

$\bar{x}$ = 3,3 dias
n = 20 pedidos
s = 3,0105 dias
n < 30
α = 0,1

$E = t_{\alpha/2,\ gl} \cdot \frac{s}{\sqrt{n}}$

$E = t_{0{,}05,\ 19} \cdot \frac{3{,}0105}{\sqrt{20}}$

E = 1,729 . 0,6731

E = 1,1638

IC: 2,1362 < μ < 4,4638

gl	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,660
30	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
35	0,682	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
40	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
45	0,680	1,301	1,679	2,014	2,412	2,690	2,952	3,281	3,520
50	0,679	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496

Determinação do tamanho amostral

Quando trabalhamos com amostras maiores que 30 elementos, é importante determinar o tamanho necessário para estimar o parâmetro, já que amostras desnecessariamente grandes gastam mais tempo e dinheiro.

Resolvendo a equação da margem de erro em relação a n, obtemos:

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2,$$

Onde E é a margem de erro desejada e α o nível de significância desejado.

Se o uso da expressão acima não resultar um número inteiro, arredonde-o para o número inteiro maior mais próximo.

Se σ for desconhecido, usar um método para estima-lo, como por exemplo a regra intuitiva da amplitude, onde σ ≅ (amplitude amostral) / 4.

Determinação da Estimativa pontual e de E

- Com um intervalo de confiança, podemos determinar a média amostral e a margem de erro usadas em sua construção:

- $\overline{x} = \frac{limite\ superior + limite\ inferior}{2}$ e $E = \frac{limite\ superior - limite\ inferior}{2}$

Referências

- Icm_apresentação, José Carlos da Silva Adão
- https://support.minitab.com/pt-br/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/what-is-a-confidence-interval/
- http://www.portalaction.com.br/inferencia/41-intervalo-de-confianca-para-media
- Notas de aulas expandidas. Faria, José Claudio.
- TRIOLA, M.F. Introdução à estatística. 9 ed. 2005.