Estatística: Aplicação ao Sensoriamento Remoto

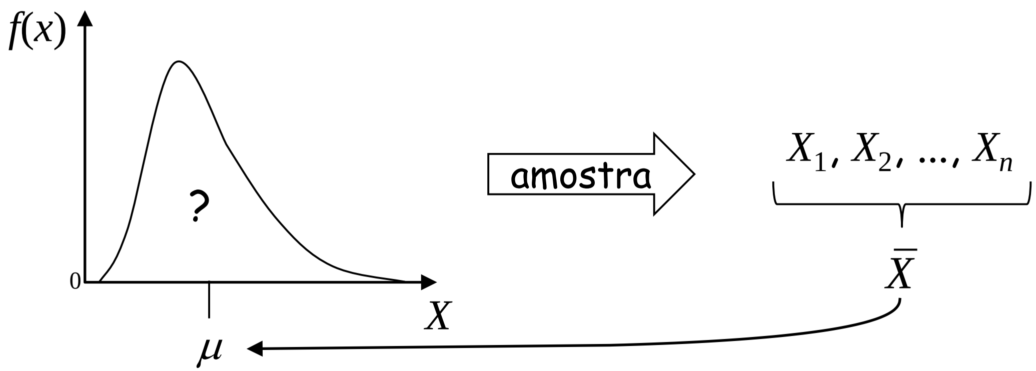
SER 204 - ANO 2024

Intervalo de Confiança

Camilo Daleles Rennó
camilo.renno@inpe.br
http://www.dpi.inpe.br/~camilo/estatistica/

Intervalo de Confiança

Um parâmetro pode ser estimado através de um único valor (estimador pontual)



Qual a probabilidade de que $\bar{X}$ tenha exatamente o valor de μ?

$P(\bar{X} = \mu) = 0$ (improvável)

Uma alternativa é definir um intervalo de estimativas mais prováveis de acordo com a distribuição teórica da estatística (estimador), que é uma v.a.

Para isso, é necessário conhecer esta distribuição

Intervalo de Confiança para μ

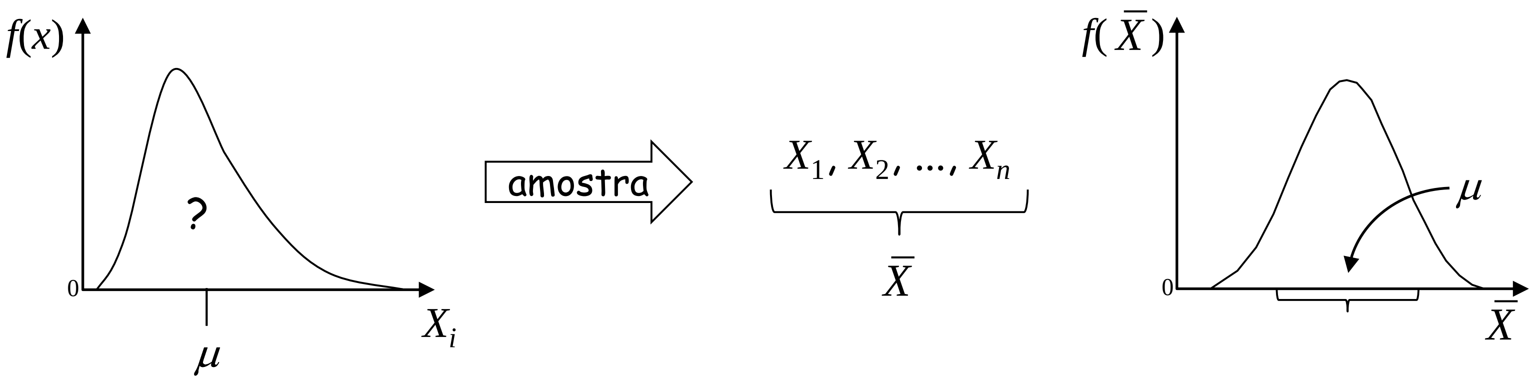
$X = \{X_1, X_2, \cdots, X_n\}$ amostra aleatória

$X_i \sim ?(\mu, \sigma^2)$ distribuição desconhecida, μ desconhecido, mas σ^2 conhecido

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Se $X_i \sim N(\mu, \sigma^2)$ ou se n for grande (ou seja, adotando-se o TLC):

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \qquad E\left(\bar{X}\right) = \mu \qquad Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

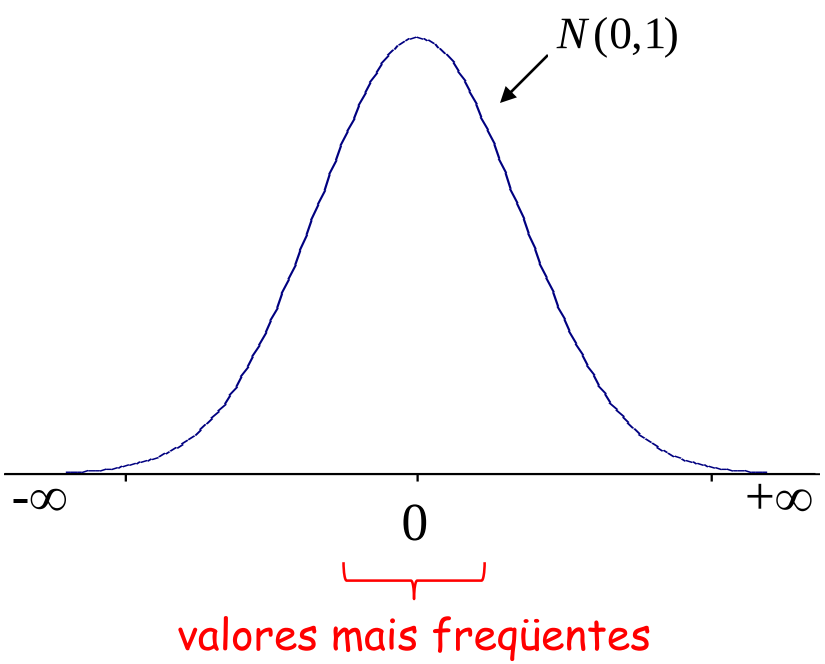


Intervalo de Confiança para μ

$X \sim ?(\mu, \sigma^2)$ distribuição desconhecida, μ desconhecido, mas σ^2 conhecido

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ se X tiver distribuição normal ou n for grande (TLC)

$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ (Normal Padrão)



Intervalo de Confiança para μ

$X \sim ?(\mu, \sigma^2)$ distribuição desconhecida, μ desconhecido, mas σ^2 conhecido

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ se X tiver distribuição normal ou n for grande (TLC)

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$ (Normal Padrão)

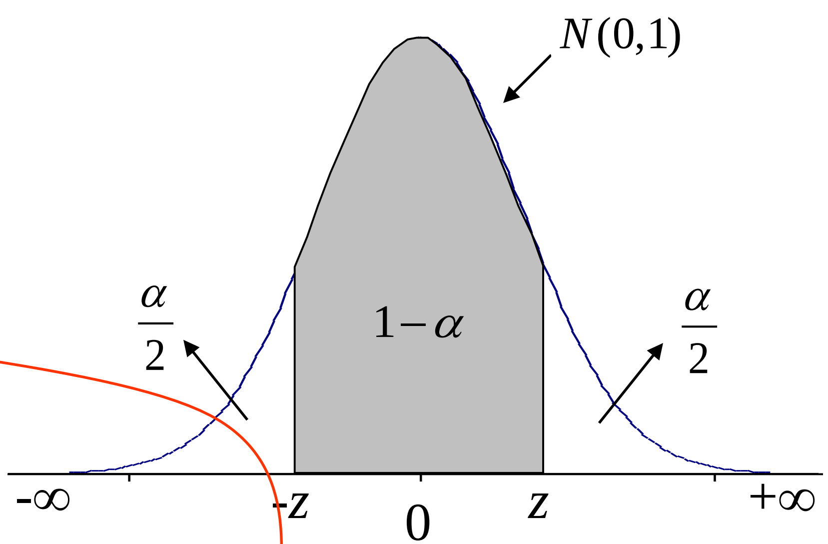
Z

$$P(-z < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z) = 1 - \alpha$$

$$P(-z\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\boxed{P(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha}$$

IC para μ



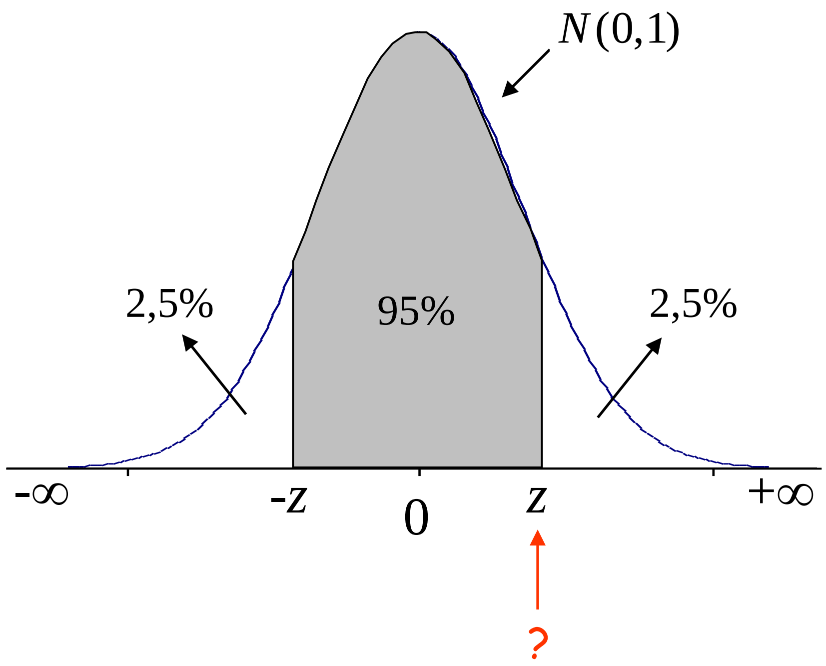
$P(|Z| > z) = \alpha$ nível de significância

$P(-z < Z < z) = 1 - \alpha$ nível de confiança

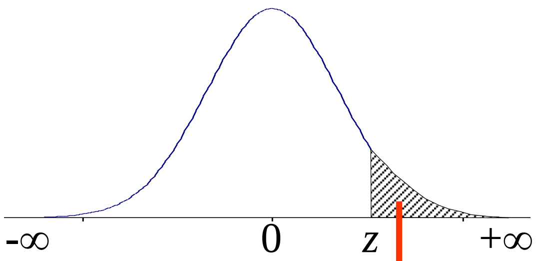
Intervalo de Confiança para μ

Exemplo: uma v.a. qualquer tem uma distribuição desconhecida com média μ também desconhecida e variância σ^2 = 16. Retira-se uma amostra de 36 valores e calcula-se a média amostral. Construa um IC de 95% para μ supondo que $\bar{X} = 12,7$.

$$P(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}) = 0,95$$



Intervalo de Confiança para μ



$$P(Z > 1{,}96) = 0{,}025$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010

Intervalo de Confiança para μ

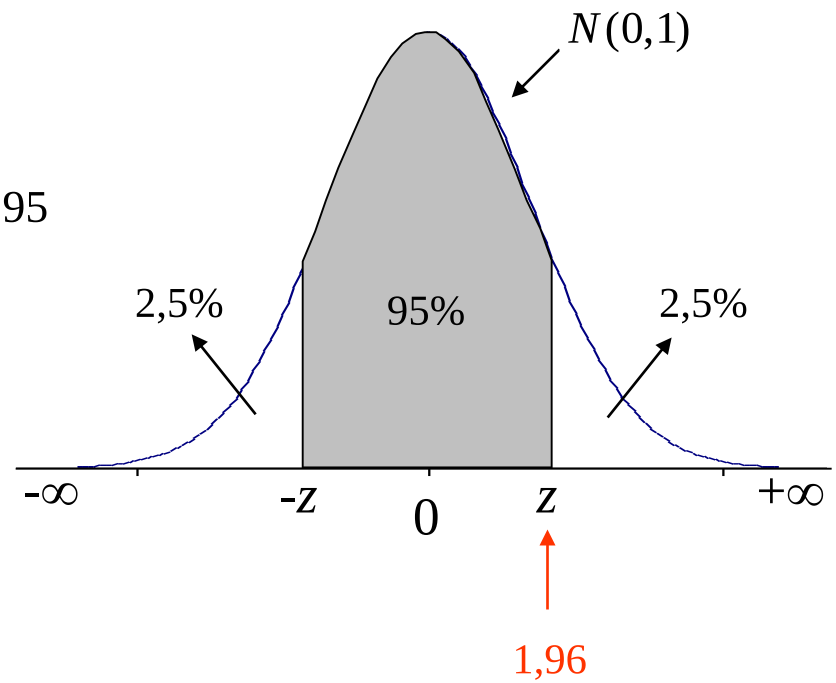
Exemplo: uma v.a. qualquer tem uma distribuição desconhecida com média μ também desconhecida e variância σ^2 = 16. Retira-se uma amostra de 36 valores e calcula-se a média amostral. Construa um IC de 95% para μ supondo que $\bar{X} = 12,7$.

$$P(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}) = 0,95$$

$$P(12,7 - 1,96\frac{4}{\sqrt{36}} < \mu < 12,7 + 1,96\frac{4}{\sqrt{36}}) = 0,95$$

$$P(12,7 - 1,307 < \mu < 12,7 + 1,307) = 0,95$$

$$\boxed{P(11,393 < \mu < 14,007) = 0,95}$$



Mas o que significa realmente este IC?

Como Interpretar o IC para μ?

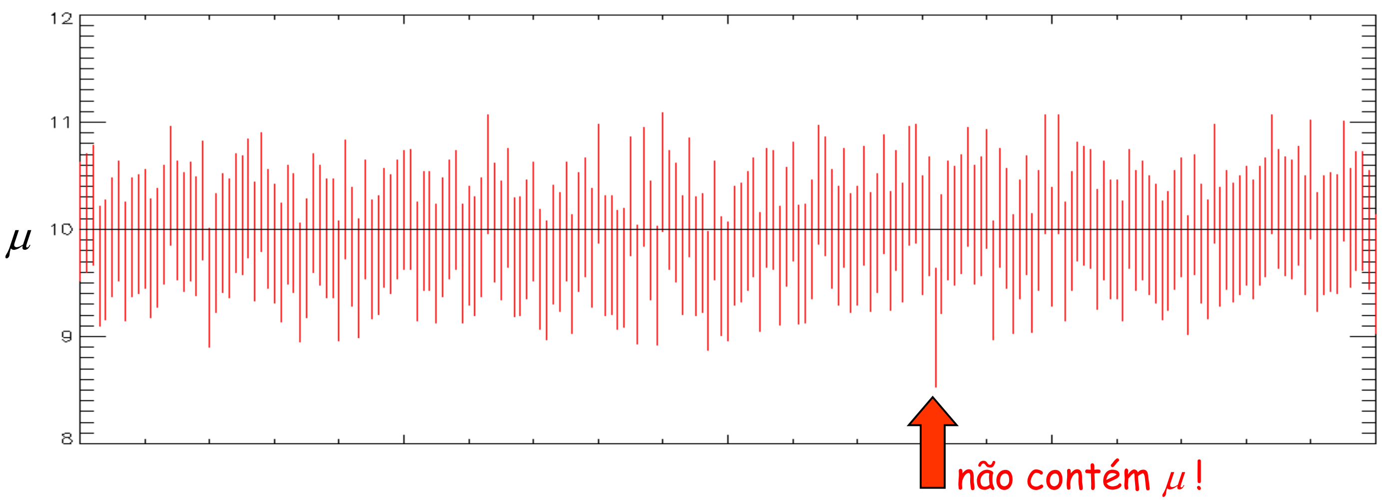
Suponha uma v.a. X normalmente distribuída com $\mu = 10$ e $\sigma^2 = 4 \quad \rightarrow \quad X \sim N(10,4)$

Sorteia-se 50 valores aleatoriamente e calcula-se $\bar{X}$. Em seguida determina-se o IC para μ com 95% de confiança, ou seja

$$P(\bar{X} - 1,96\frac{2}{\sqrt{50}} < \mu < \bar{X} + 1,96\frac{2}{\sqrt{50}}) = 0,95$$

$$P(\bar{X} - 0,5544 < \mu < \bar{X} + 0,5544) = 0,95$$ (O IC varia para cada amostra!!!)

Através de simulações, foram gerados inúmeros IC, um para cada amostra...

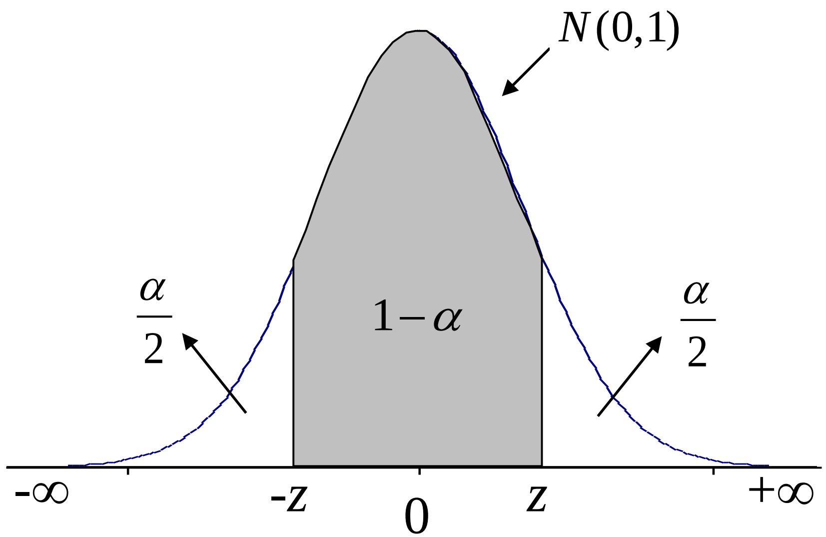


Interpretação: 95% dos possíveis IC obtidos a partir de uma amostra de tamanho 50, conterão de fato a verdadeira média μ

(ver IC.xls)

Intervalo de Confiança para μ

$$P(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$



Como poderíamos obter intervalos de confiança mais estreitos, ou seja, com limites mais próximos da média verdadeira μ?

- diminuindo-se o nível de confiança
  quanto menor $1-\alpha$, menor será z
  mas maior será a probabilidade do IC
  não conter a verdadeira μ!

$P(|Z| > z) = \alpha$ nível de significância

$P(-z < Z < z) = 1 - \alpha$ nível de confiança

- aumentando-se o tamanho da amostra ← melhor opção
  quanto maior n menor será a variância
  de $\bar{X}$

É possível obter ICs com 100% de confiança? Não $P(-\infty < \mu < +\infty) = 1$

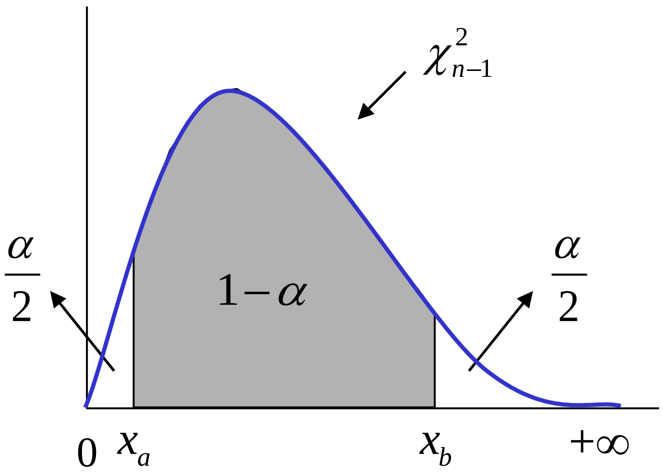
Intervalo de Confiança para σ^2

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

χ^2

$$P\left(x_a < \frac{(n-1)s^2}{\sigma^2} < x_b\right) = 1-\alpha$$

$$P\left(\frac{1}{x_b} < \frac{\sigma^2}{(n-1)s^2} < \frac{1}{x_a}\right) = 1-\alpha$$



$$P(x_a < \chi^2_{n-1} < x_b) = 1-\alpha$$

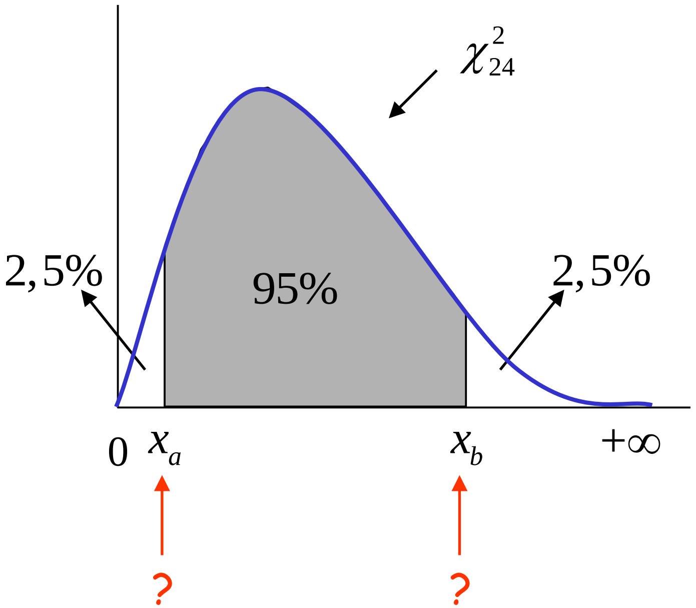
$$\boxed{P\left(\frac{(n-1)s^2}{x_b} < \sigma^2 < \frac{(n-1)s^2}{x_a}\right) = 1-\alpha}$$

IC para σ^2

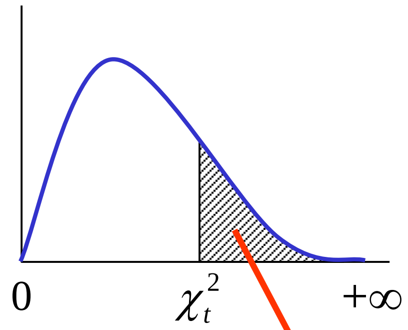
Intervalo de Confiança para σ^2

Exemplo: uma v.a. qualquer tem uma distribuição normal com média μ e variância σ^2 desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a variância amostral. Construa um IC de 95% para σ^2 supondo que $s^2 = 2{,}34$.

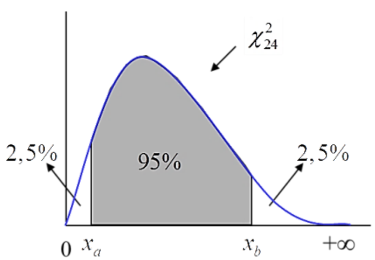
$$P\left(\frac{(n-1)s^2}{x_b} < \sigma^2 < \frac{(n-1)s^2}{x_a}\right) = 0{,}95$$



Distribuição χ^2



$$P(\chi_g^2 > \chi_t^2)$$



$$P(\chi_{24}^2 > x_a) = 0{,}975$$

$$P(\chi_{24}^2 > x_b) = 0{,}025$$

g	**0,005**	**0,010**	**0,025**	**0,050**	**0,100**	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**
1	7,88	6,63	5,02	3,84	2,71	0,016	0,0039	0,0010	0,00016	0,00004
2	10,60	9,21	7,38	5,99	4,61	0,21	0,10	0,051	0,020	0,010
3	12,84	11,34	9,35	7,81	6,25	0,58	0,35	0,22	0,11	0,072
4	14,86	13,28	11,14	9,49	7,78	1,06	0,71	0,48	0,30	0,21
5	16,75	15,09	12,83	11,07	9,24	1,61	1,15	0,83	0,55	0,41
6	18,55	16,81	[illegible]	12,59	10,64	2,20	1,64	[illegible]	0,87	0,68
7	20,28	18,48	[illegible]	14,07	12,02	2,83	2,17	[illegible]	1,24	0,99
8	21,95	20,09	[illegible]	15,51	13,36	3,49	2,73	[illegible]	1,65	1,34
9	23,59	21,67	[illegible]	16,92	14,68	4,17	3,33	[illegible]	2,09	1,73
10	25,19	23,21	[illegible]	18,31	15,99	4,87	3,94	[illegible]	2,56	2,16
11	26,76	24,72	21,92	19,68	17,28	5,58	4,57	3,82	3,05	2,60
12	28,30	26,22	23,34	21,03	18,55	6,30	5,23	4,40	3,57	3,07
13	29,82	27,69	24,74	22,36	19,81	7,04	5,89	5,01	4,11	3,57
14	31,32	29,14	26,12	23,68	21,06	7,79	6,57	5,63	4,66	4,07
15	32,80	30,58	27,49	25,00	22,31	8,55	7,26	6,26	5,23	4,60
16	34,27	32,00	28,85	26,30	23,54	9,31	7,96	6,91	5,81	5,14
17	35,72	33,41	30,19	27,59	24,77	10,09	8,67	7,56	6,41	5,70
18	37,16	34,81	31,53	28,87	25,99	10,86	9,39	8,23	7,01	6,26
19	38,58	36,19	32,85	30,14	27,20	11,65	10,12	8,91	7,63	6,84
20	40,00	37,57	34,17	31,41	28,41	12,44	10,85	9,59	8,26	7,43
21	41,40	38,93	35,48	32,67	29,62	13,24	11,59	10,28	8,90	8,03
22	42,80	40,29	36,78	33,92	30,81	14,04	12,34	10,98	9,54	8,64
23	44,18	41,64	38,08	35,17	32,01	14,85	13,09	11,69	10,20	9,26
24	45,56	42,98	39,36	36,42	33,20	15,66	13,85	12,40	10,86	9,89
25	46,93	44,31	40,65	37,65	34,38	16,47	14,61	13,12	11,52	10,52
26	48,29	45,64	41,92	38,89	35,56	17,29	15,38	13,84	12,20	11,16
27	49,64	46,96	43,19	40,11	36,74	18,11	16,15	14,57	12,88	11,81
28	50,99	48,28	44,46	41,34	37,92	18,94	16,93	15,31	13,56	12,46
29	52,34	49,59	45,72	42,56	39,09	19,77	17,71	16,05	14,26	13,12
30	53,67	50,89	46,98	43,77	40,26	20,60	18,49	16,79	14,95	13,79
40	66,77	63,69	59,34	55,76	51,81	29,05	26,51	24,43	22,16	20,71
50	79,49	76,15	71,42	67,50	63,17	37,69	34,76	32,36	29,71	27,99
60	91,95	88,38	83,30	79,08	74,40	46,46	43,19	40,48	37,48	35,53
70	104,21	100,43	95,02	90,53	85,53	55,33	51,74	48,76	45,44	43,28
80	116,32	112,33	106,63	101,88	96,58	64,28	60,39	57,15	53,54	51,17
90	128,30	124,12	118,14	113,15	107,57	73,29	69,13	65,65	61,75	59,20
100	140,17	135,81	129,56	124,34	118,50	82,36	77,93	74,22	70,06	67,33

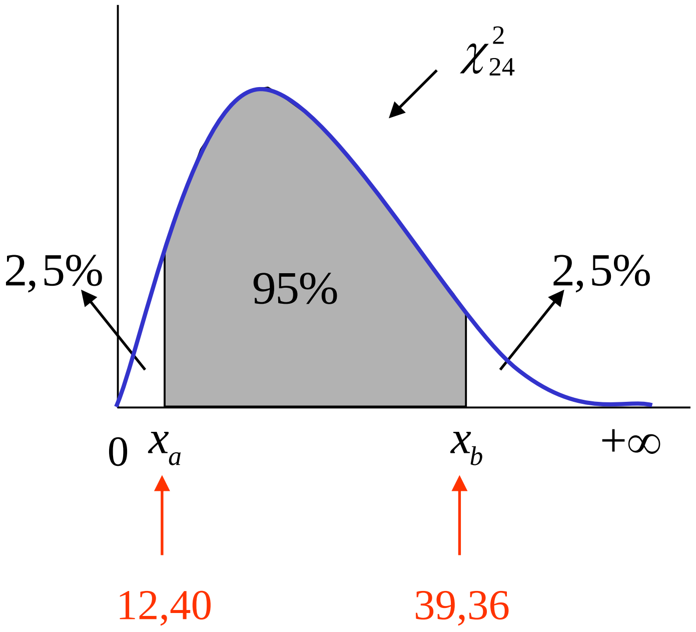
Intervalo de Confiança para σ^2

Exemplo: uma v.a. qualquer tem uma distribuição normal com média μ e variância σ^2 desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a variância amostral. Construa um IC de 95% para σ^2 supondo que $s^2 = 2{,}34$.

$$P\left(\frac{(n-1)s^2}{x_b} < \sigma^2 < \frac{(n-1)s^2}{x_a}\right) = 0{,}95$$

$$P\left(\frac{24\,.\,2{,}34}{39{,}36} < \sigma^2 < \frac{24\,.\,2{,}34}{12{,}40}\right) = 0{,}95$$

$$\boxed{P\left(1{,}43 < \sigma^2 < 4{,}53\right) = 0{,}95}$$



Intervalo de Confiança para μ com σ^2 desconhecida

$X \sim N(\mu, \sigma^2)$ μ e σ^2 desconhecidos

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

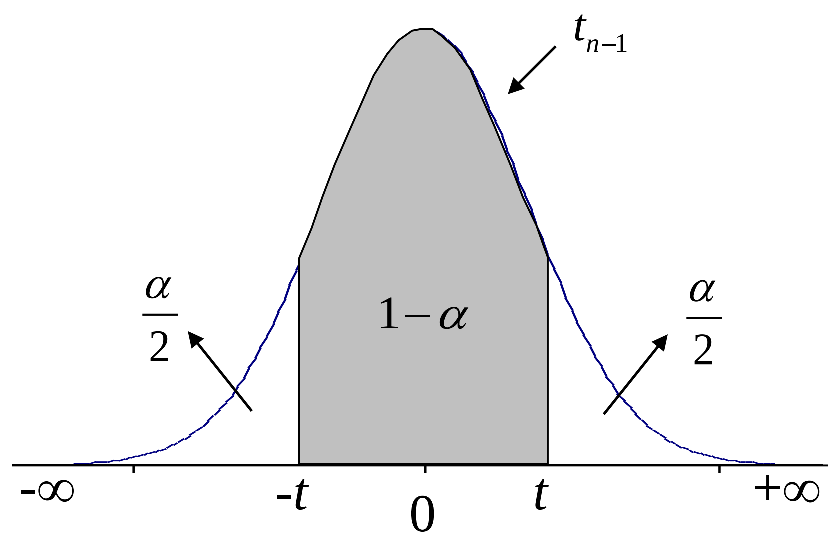
T

$$P(-t < \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} < t) = 1 - \alpha$$

$$P(-t\frac{s}{\sqrt{n}} < \bar{X} - \mu < t\frac{s}{\sqrt{n}}) = 1 - \alpha$$

$$\boxed{P(\bar{X} - t\frac{s}{\sqrt{n}} < \mu < \bar{X} + t\frac{s}{\sqrt{n}}) = 1 - \alpha}$$

IC para μ

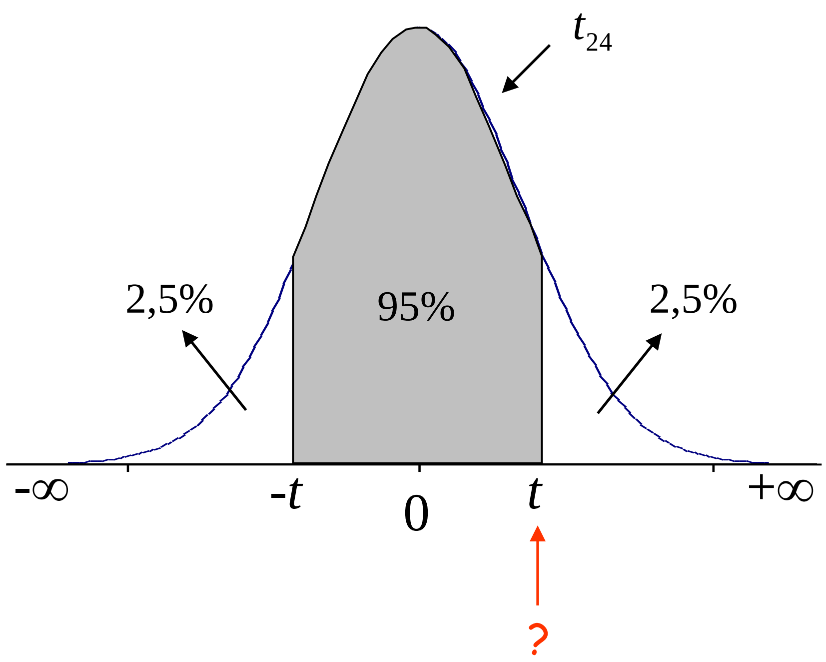


$$P(-t < T < t) = 1 - \alpha$$

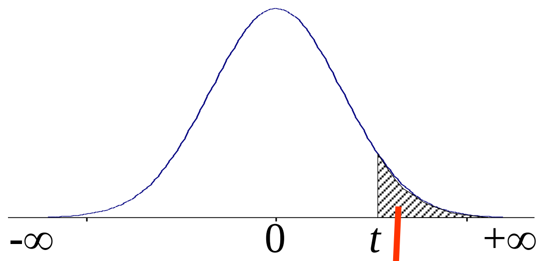
Intervalo de Confiança para μ com σ^2 desconhecida

Exemplo: uma v.a. qualquer tem uma distribuição normal com média μ e variância σ^2 desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a média amostral e a variância amostral. Construa um IC de 95% para μ supondo que $\bar{X} = 12,7$ e $s^2 = 16$.

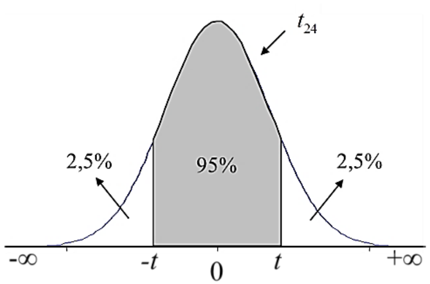
$$P(\bar{X} - t\frac{s}{\sqrt{n}} < \mu < \bar{X} + t\frac{s}{\sqrt{n}}) = 0,95$$



Distribuição *t de Student*



$$P(T_g > t)$$



$$P(T_{24} > t) = 0,025$$

$$P(T_{24} > 2,064) = 0,025$$

g	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**
1	3,078	6,314	12,706	31,821	63,656
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	$\boxed{2,064}$	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

Intervalo de Confiança para μ com σ^2 desconhecida

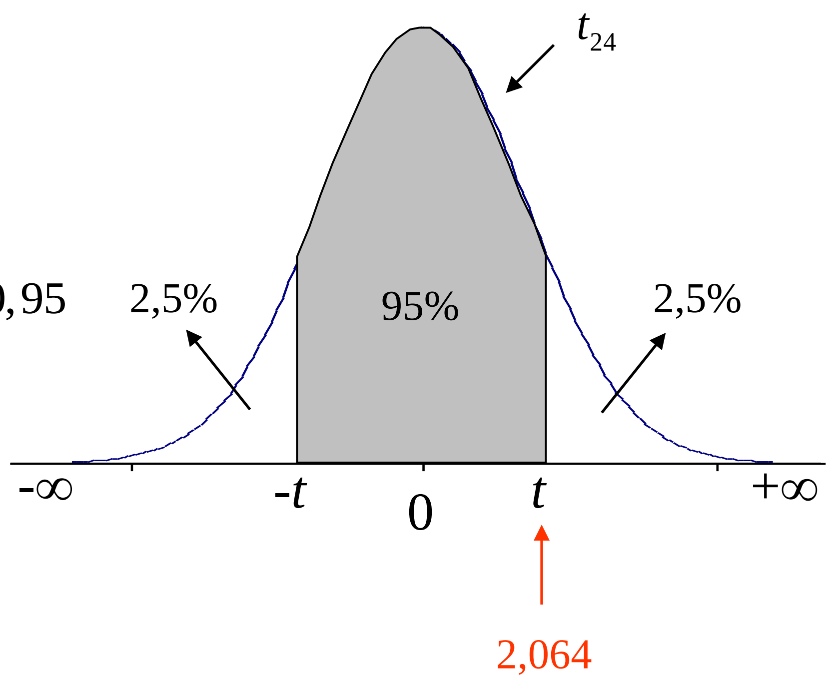
Exemplo: uma v.a. qualquer tem uma distribuição normal com média μ e variância σ^2 desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a média amostral e a variância amostral. Construa um IC de 95% para μ supondo que $\bar{X} = 12,7$ e $s^2 = 16$.

$$P(\bar{X} - t\frac{s}{\sqrt{n}} < \mu < \bar{X} + t\frac{s}{\sqrt{n}}) = 0,95$$

$$P(12,7 - 2,064\frac{4}{\sqrt{25}} < \mu < 12,7 + 2,064\frac{4}{\sqrt{25}}) = 0,95$$

$$P(12,7 - 1,6512 < \mu < 12,7 + 1,6512) = 0,95$$

$$\boxed{P(11,0488 < \mu < 14,3512) = 0,95}$$



Intervalo de Confiança para proporção p

$$\hat{p} \sim N(p, \frac{pq}{n})$$

$$\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

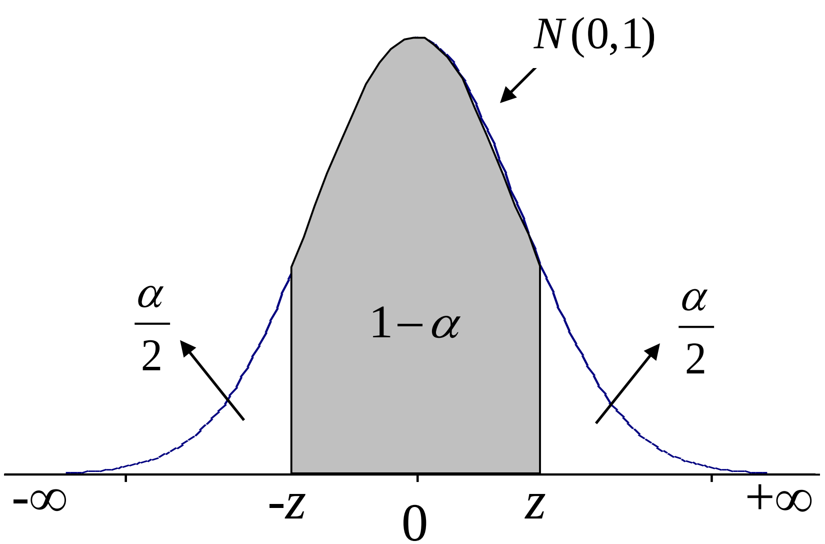
Z

$$P(-z < \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} < z) = 1-\alpha$$

$$P(\hat{p} - z\sqrt{\frac{pq}{n}} < p < \hat{p} + z\sqrt{\frac{pq}{n}}) = 1-\alpha$$

$$\boxed{P(\hat{p} - z\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z\sqrt{\frac{\hat{p}\hat{q}}{n}}) = 1-\alpha}$$

IC para p



$$P(-z < Z < z) = 1-\alpha$$

Intervalos de Confiança (Resumo)

$$\text{para } \mu \begin{cases} N(0,1) & \text{se } \sigma^2 \text{ é conhecida} \\ t_{n-1} & \text{se } \sigma^2 \text{ é desconhecida} \end{cases}$$

$$\text{para } \sigma^2 \left\{ \chi^2_{n-1} \right.$$

$$\text{para } p \left\{ N(0,1) \right.$$

É possível também construir IC de modo a comparar parâmetros de 2 populações (ver o material extra desse tema)

No entanto, esse propósito será melhor abordado em Testes de Hipótese

Intervalos de Confiança (Resumo)

Observações importantes:

- Os ICs são construídos a partir de uma distribuição que relaciona o estimador pontual ao seu parâmetro;
- Para se conseguir ICs mais estreitos, conservando-se o mesmo nível de confiança, deve-se aumentar o tamanho da amostra;
- Caso o IC seja utilizado para verificar se o parâmetro para o qual o IC foi construído tem um determinado valor, deve-se aceitar qualquer valor presente dentro do intervalo, considerando o nível de confiança adotado;

  Ex: se o IC para μ for $P(20{,}3 < \mu < 43{,}8) = 0{,}95$

  μ pode ser 30? SIM

  μ pode ser 21? SIM

  μ pode ser 45? NÃO

  - "não se pode negar que ela seja 30" ou
    "não há razões para discordar que a verdadeira média μ seja de fato 30"

  considerando 95% de confiança

  (se mudar o nível de confiança, pode-se mudar a conclusão)

- Intervalo de Confiança (estimação de parâmetro)

  $\neq$

  Intervalo de Credibilidade (ocorrência de valores simulados)