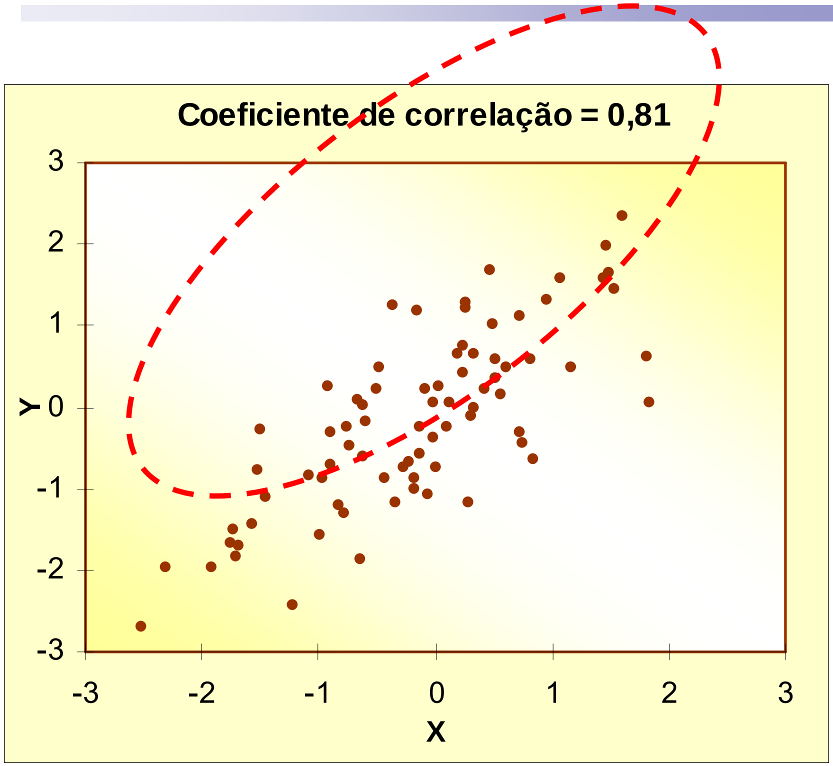
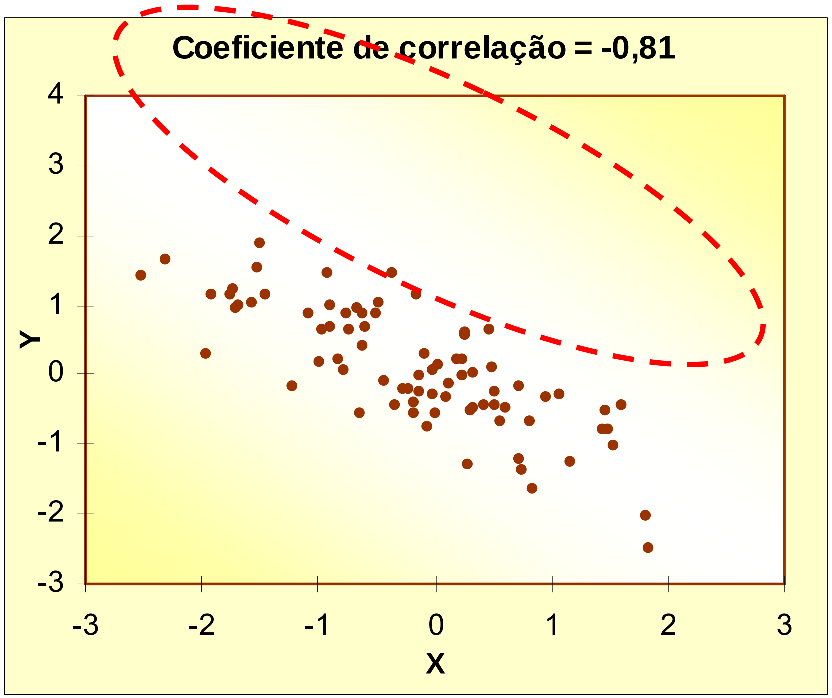
CORRELAÇÃO

- Nas Instituições de Ensino Superior(IES), há uma relação direta entre a qualidade do ensino e a taxa de inadimplência. A taxa de inadimplência das IES que obtiveram conceitos A e B no Provão é 12,1%, nas que obtiveram C é 16% e nas que obtiveram D e E a inadimplência é de 21,9%.
- O frio está para o setor farmacêutico como o Dia das Mães está para o comércio. As vendas de medicamentos não controlados, como analgésicos, antigripais e vitaminas, disparam.
- O faturamento das empresas de energia nos Estados Unidos é diretamente influenciado pela temperatura, especialmente no inverno. Um inverno brando reduz a demanda de energia e pode diminuir drasticamente o lucro.

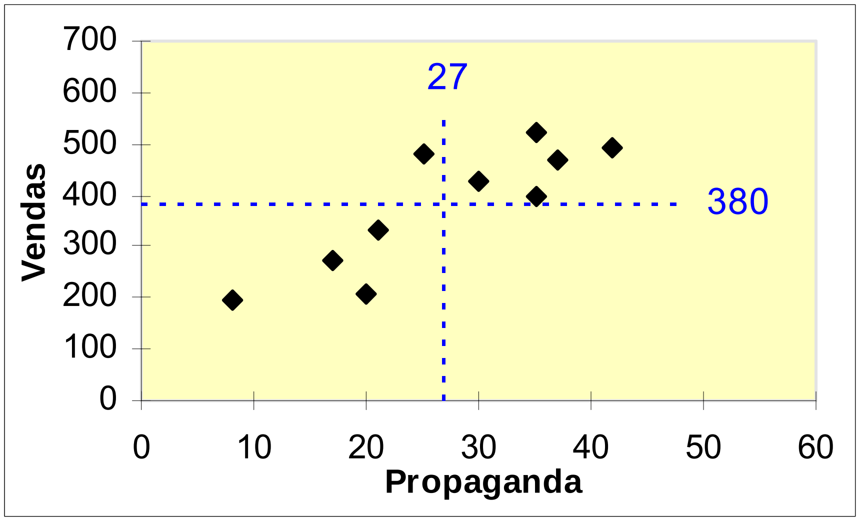


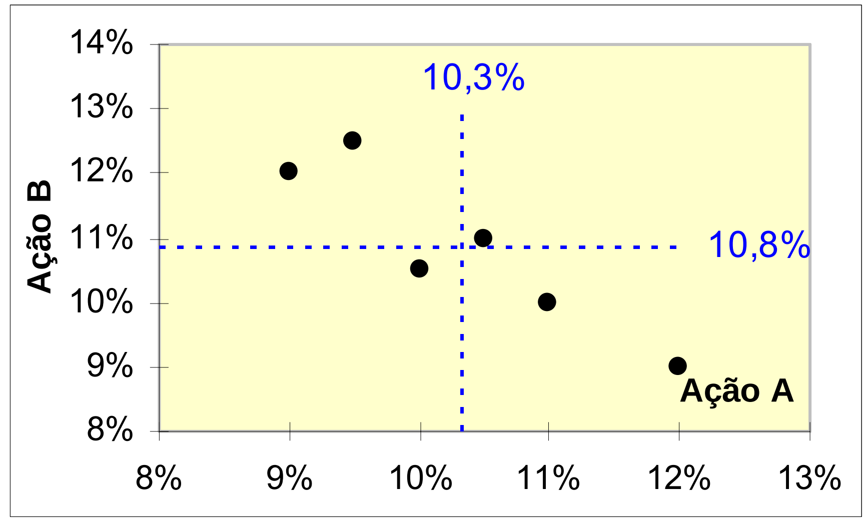


O gráfico de dispersão da esquerda mostra uma relação direta ou positiva entre as variáveis *X* e *Y*, tendência destacada pela declividade positiva da elipse tracejada. Enquanto o gráfico de dispersão da direita mostra uma relação inversa ou negativa, tendência também destacada pela declividade negativa da elipse tracejada.

Características da Covariância

- A covariância mede a tendência e a força da relação linear entre duas variáveis.





A covariância pode ser nula, negativa ou positiva. A covariância é a medida do afastamento simultâneo das respectivas médias. Se as ambas variáveis aleatórias tendem a estar simultaneamente acima, ou abaixo, de suas respectivas médias, então a covariância tenderá a ser positiva e nos outros casos poderá ser negativa, como mostram os gráficos acima.

- A covariância de uma variável e ela mesma é a própria variância da variável, seja no caso de população ou amostra. Como *Y* = *X*,

$$S_{XX} = \frac{\sum_{i=1}^{n}(X_i - \bar{x}) \times (X_i - \bar{x})}{n} = \frac{\sum_{i=1}^{N}(X_i - \bar{x})^2}{n} = S_X^2$$

- A permutação das variáveis não altera o resultado da covariância, se os pares de valores não forem alterados:

$$S_{XY} = S_{YX}$$

Coeficiente de Correlação

- Para facilitar a relação entre duas variáveis e evitar a unidade de medida da covariância, foi definido o coeficiente de correlação $r_{XY.}$
- Os valores de r_{XY} estão limitados entre os valores -1 e $+1$, e sem nenhuma unidade de medida
- É um valor único para população ou amostra, tomando o cuidado de utilizar dados coerentes no procedimento de cálculo.
- Da fórmula do coeficiente de correlação pode-se obter também a covariância das mesmas variáveis quando conhecidos os desvios padrões correspondentes:

$$S_{XY} = r_{XY} \times S_X \times S_Y$$

Características de *r*

- Se a variável *Y* é a mesma variável *X*, então o coeficiente de correlação é igual a 1:

$$r_{XX} = \frac{\sigma_{XX}}{\sigma_X \times \sigma_X} = \frac{\sigma_X^2}{\sigma_X^2} = 1$$

- A permutação das variáveis não altera o resultado do coeficiente de correlação, se os mesmos pares de valores forem mantidos.

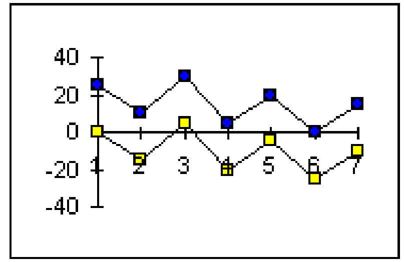
$$r_{XY} = r_{YX}$$

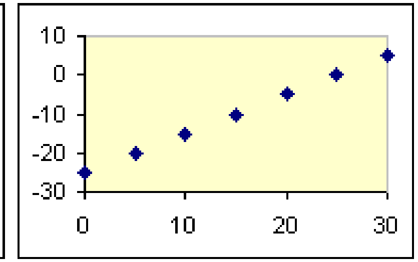
Exemplos

Amostras perfeitamente correlacionadas em sentido positivo

X	Y
25	0
10	-15
30	5
5	-20
20	-5
0	-25
15	-10

r = 1,00

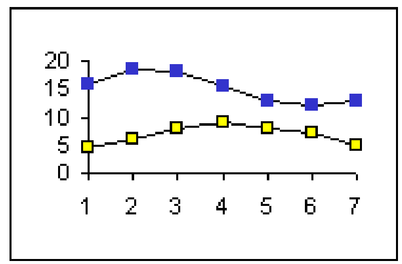


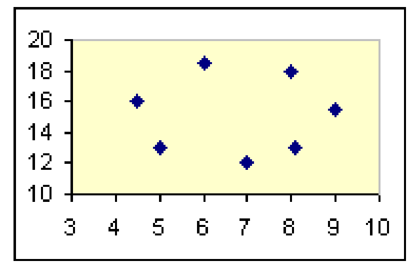


Amostras não correlacionadas

X	Y
4,5	16
6	18,5
8	18
9	15,5
8,1	13
7	12
5	13

r = 0,00

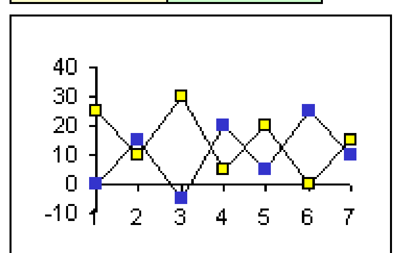


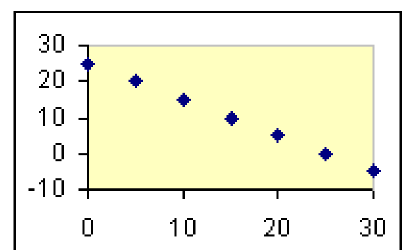


Amostras perfeitamente correlacionadas em sentido negativo

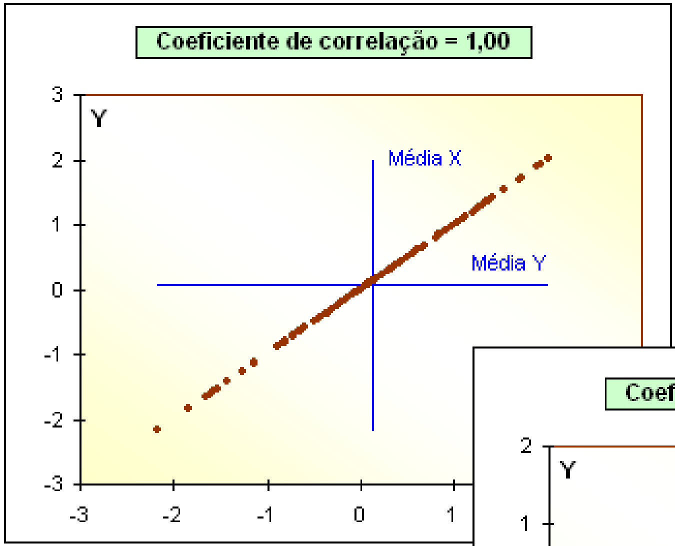
X	Y
25	0
10	15
30	-5
5	20
20	5
0	25
15	10

r = -1,00

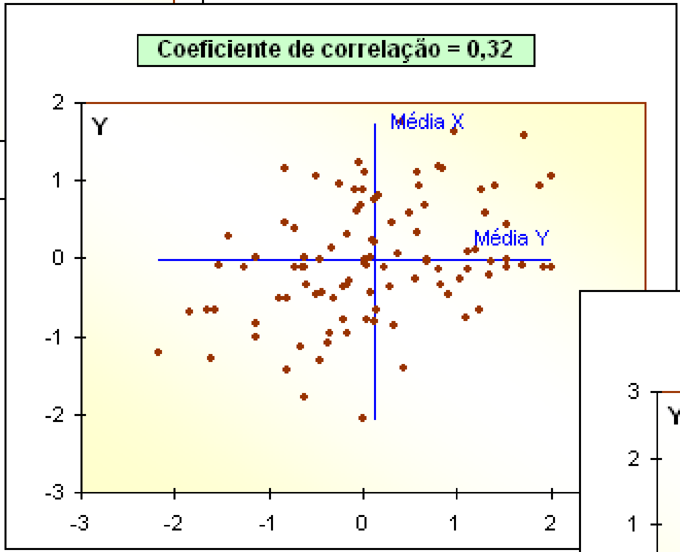




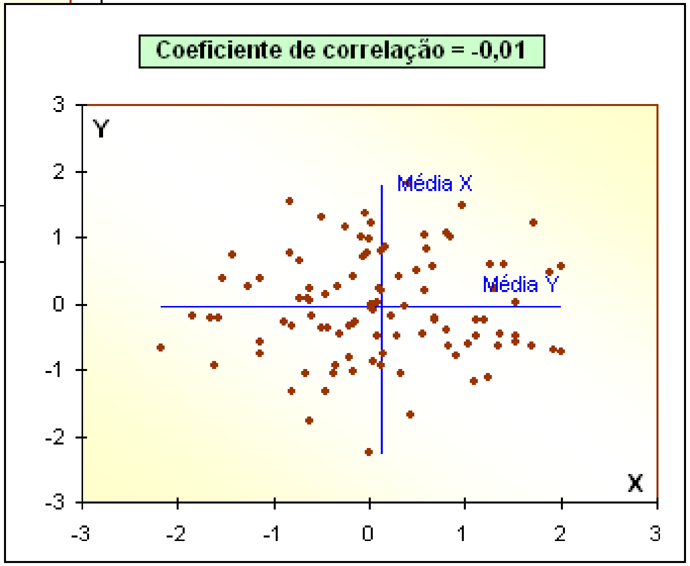
Exemplos



Coeficiente de correlação = 1,00
Y
Média X
Média Y



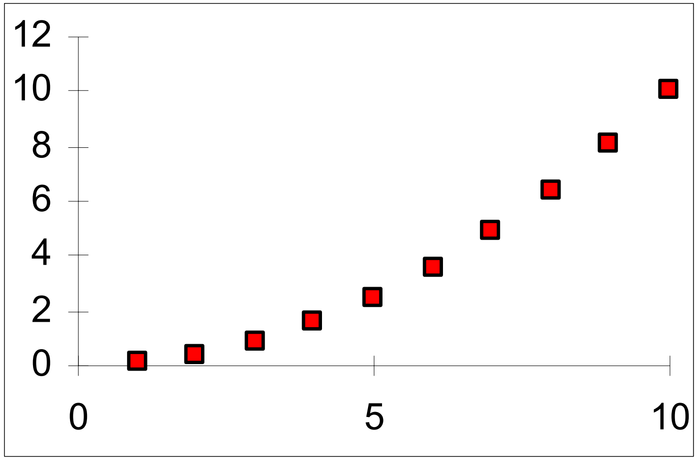
Coeficiente de correlação = 0,32
Y
Média X
Média Y

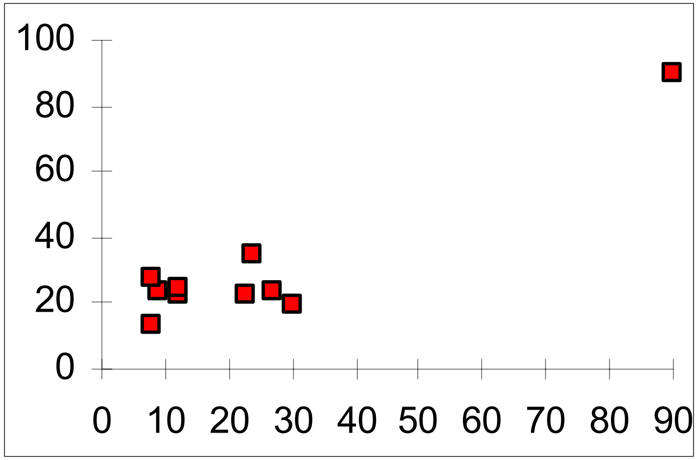


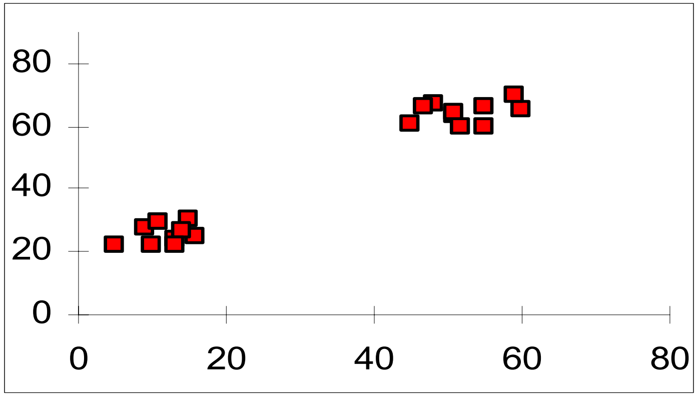
Coeficiente de correlação = -0,01
Y
Média X
Média Y
X

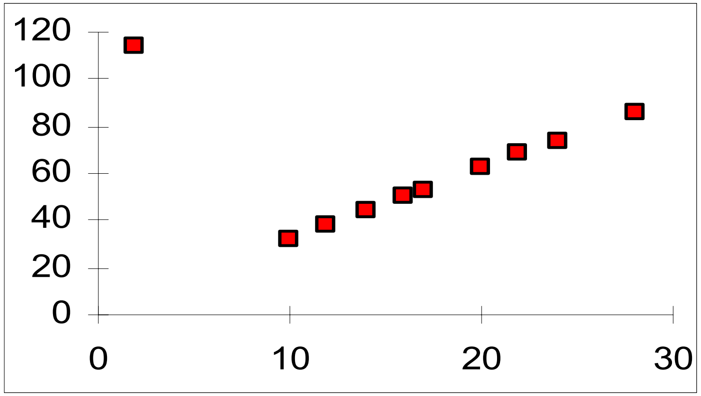
- Em alguns casos, a relação causa-efeito pode ser provocada por um ou mais fatores ocultos, uma variável não considerada na análise.
  - Por exemplo, suponha que o número de vendas diárias de um jornal e a produção diária de ovos tenham uma forte correlação positiva.
  - Não se pode afirmar que o aumento da produção de ovos seja a causa do aumento do número de jornais vendidos, nem que o aumento do número de jornais vendidos resulte no aumento da produção de ovos!
  - Para compreender a forte e positiva correlação, devem procurar fatores ocultos, por exemplo, o aumento de riqueza da população que resulta em aumento de demanda dos dois produtos ao mesmo tempo, jornais e ovos.

Anomalias com *r* próximo de +1







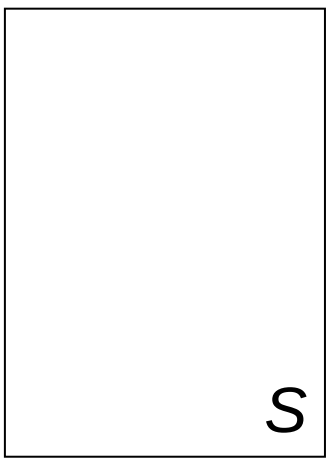


Índices de preços (%)

Mês	IBGE INPC	FIPE IPC	DIEESE ICV	IBGE IPCA	FGV IGP-DI	FGV IGPM	FGV IPA-M	IBGE IPCA-E
jun/03	-0,06	-0,16	-0,26	-0,15	-0,70	-1,00	-1,62	0,22
jul/03	0,04	-0,08	0,35	0,20	-0,20	-0,42	-0,75	-0,18
ago/03	0,18	0,63	-0,15	0,34	0,62	0,38	0,20	0,27
set/03	0,82	0,84	1,26	0,78	1,05	1,18	1,54	0,57
out/03	0,39	0,63	0,47	0,29	0,44	0,38	0,36	0,66
nov/03	0,37	0,27	0,26	0,34	0,48	0,49	0,57	0,17
dez/03	0,54	0,42	0,32	0,52	0,60	0,61	0,64	0,46
jan/04	0,83	0,65	1,46	0,76	0,80	0,88	0,98	0,68
fev/04	0,39	0,19	-0,18	0,61	1,08	0,69	0,79	0,90
mar/04	0,57	0,12	0,47	0,47	0,93	1,13	1,33	0,40

	INPC	*IPC*	*ICV*	*IPCA*	*IGP-DI*	*IGPM*	*IPA-M*	*IPCA-E*
INPC	1							
IPC	0,70	1						
ICV	0,82	0,60	1					
IPCA	0,91	0,68	0,69	1				
IGP-DI	0,80	0,65	0,42	0,90	1			
IGPM	0,89	0,67	0,57	0,90	0,97	1		
IPA-M	0,90	0,66	0,59	0,91	0,96	1,00	1	
IPCA-E	0,62	0,52	0,26	0,62	0,68	0,59	0,59	1

Inferência Estatística



distribuição desconhecida
e/ou
parâmetros desconhecidos

Exemplo:
Numa imagem, um *pixel* é selecionado ao acaso. Define se uma v.a. *X* cujo valor representa seu valor digital. Qual a probabilidade deste *pixel* possuir valor entre 100 e 150?

Quais os valores possíveis de *X*?
X: {0, 1, ..., 255} (considerando uma imagem 8 *bits*)

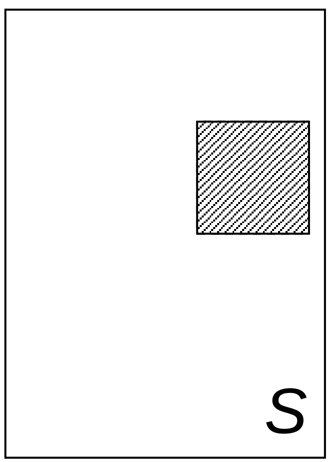
Qual a distribuição de probabilidade de *X*?
Desconhecida (discreta)

Que parâmetros são necessários para definir esta distribuição?

???????

Inferência Estatística

inferir certas características da população



distribuição desconhecida
e/ou
parâmetros desconhecidos

amostra

- n indivíduos (ou objetos) da população
  ex: sortear n *pixels* de uma imagem (com ou sem reposição)
- n realizações (sequencia de observações) de uma v.a.
  ex: medir a reflectância de um objeto n vezes

$\Downarrow$

Uma amostra constitui um conjunto de n v.a. $X_1, X_2, \ldots, X_n$ com mesma distribuição (desconhecida). Se as variáveis são independentes, elas são chamadas de iiid (independentes e identicamente distribuídas)

$\Downarrow$

$\boxed{\text{Amostra Aleatória}}$

Estimação de Parâmetros

População ⟶ Amostra

Distribuição de Probabilidade (ou FDP)   Distribuição Amostral

Parâmetros (valor fixo) $\xleftarrow{\text{estimar}}$ Estatísticas (variável aleatória)

Estimação
- pontual (estatísticas)
- por intervalo (intervalos de confiança)

OBS: estatística: é a v.a. que estima (pontualmente) um parâmetro (populacional)
as vezes é chamada simplesmente de estimador

estimativa: é o valor do estimador obtido para uma amostra específica

Estimação Pontual

Seja *X* uma v.a. normalmente distribuída com a média (μ) e a variância (σ^2) desconhecidas. Retira-se uma amostra de tamanho *n* com a finalidade de se estimar μ e σ^2.

- média populacional μ

De que maneira os valores da amostra podem ser combinados a fim de se produzir uma "boa" estimativa de μ?

$\hat{\mu}_k$ é o k-ésimo estimador de μ

$$\hat{\mu}_1 = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{\mu}_2 = x_i$$

Mas qual é o melhor estimador pontual?

- não tendencioso

$$E(\hat{\mu}_k) = \mu$$

- variância mínima

$$Var(\hat{\mu}_k) < Var(\hat{\mu}_j) \quad \forall k \neq j$$

Estimação Pontual

Seja X uma v.a. normalmente distribuída com a média (μ) e a variância (σ^2) desconhecidas. Retira-se uma amostra de tamanho n com a finalidade de se estimar μ e σ^2.

- variância populacional σ^2

De que maneira os valores da amostra podem ser combinados a fim de se produzir uma "boa" estimativa de σ^2?

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}$$

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i^2\right) - E\left(\bar{X}^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{E\left(X_i^2\right)}_{\sigma^2+\mu^2} - \underbrace{E\left(\bar{X}^2\right)}_{\frac{\sigma^2}{n}+\mu^2}$$

$$= \sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2} = \frac{n\sigma^2 - \sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

estimador tendencioso!

Estimação Pontual

Seja *X* uma v.a. normalmente distribuída com a média (μ) e a variância (σ^2) desconhecidas. Retira-se uma amostra de tamanho *n* com a finalidade de se estimar μ e σ^2.

- variância populacional σ^2

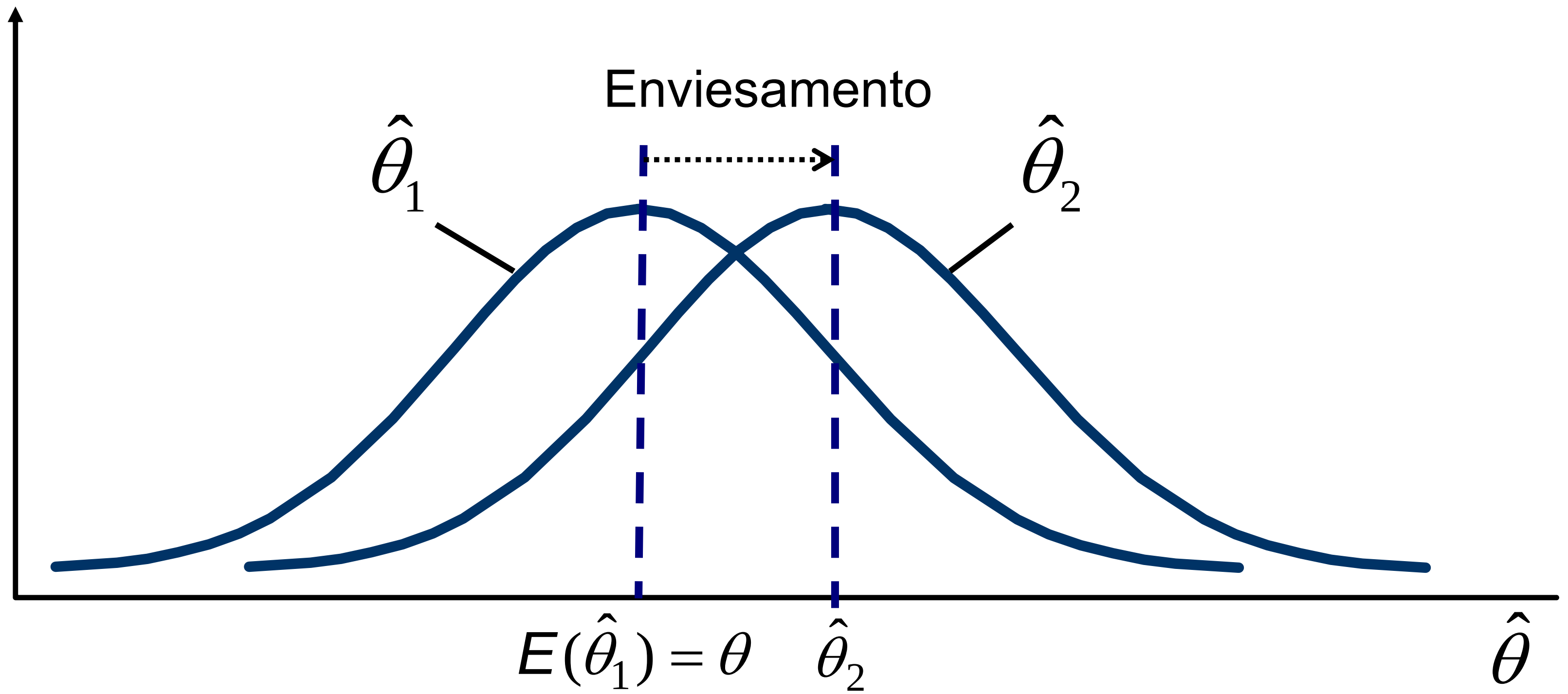
De que maneira os valores da amostra podem ser combinados a fim de se produzir uma “boa” estimativa de σ^2?

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1} \qquad E\left(s^2\right) = \sigma^2$$

estimador
não tendencioso

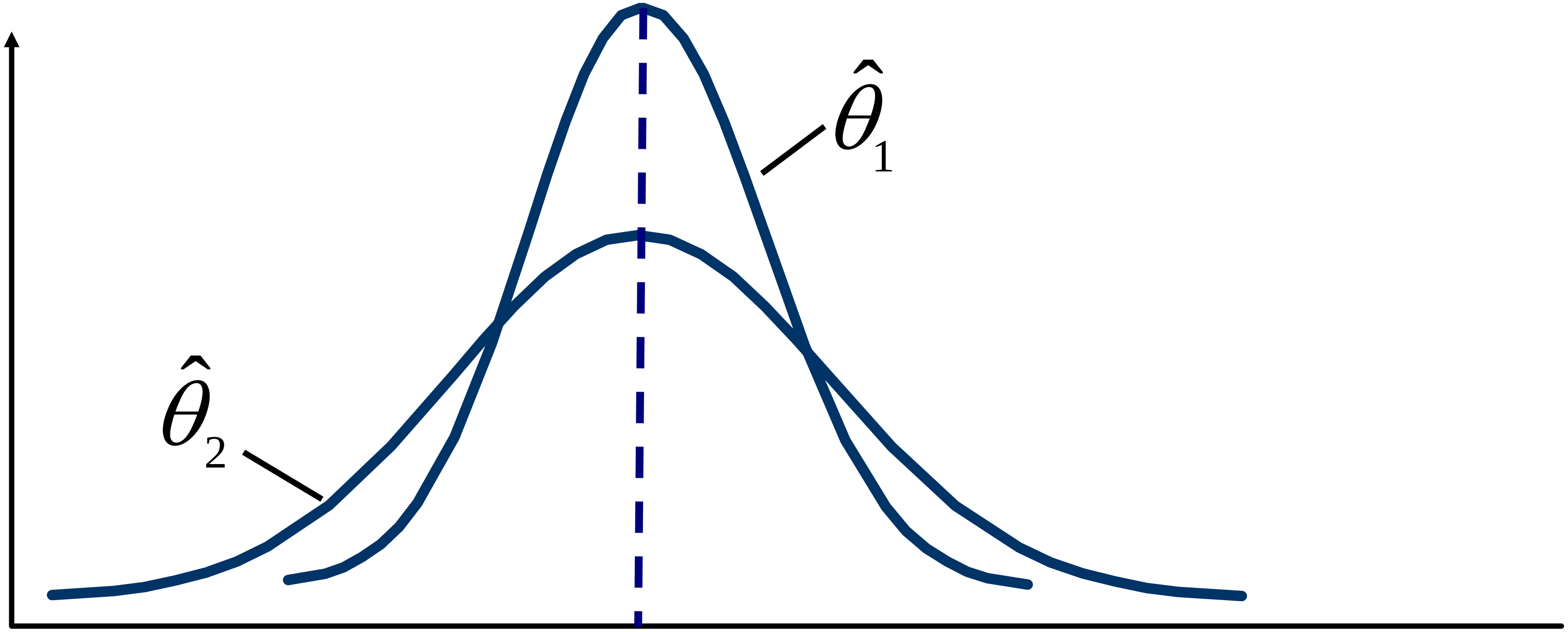
Propriedades dos estimadores pontuais: Não enviesado

$\hat{\theta}_1$ - não enviesado e $\hat{\theta}_2$ - enviesado



Propriedades dos estimadores pontuais: Eficiente

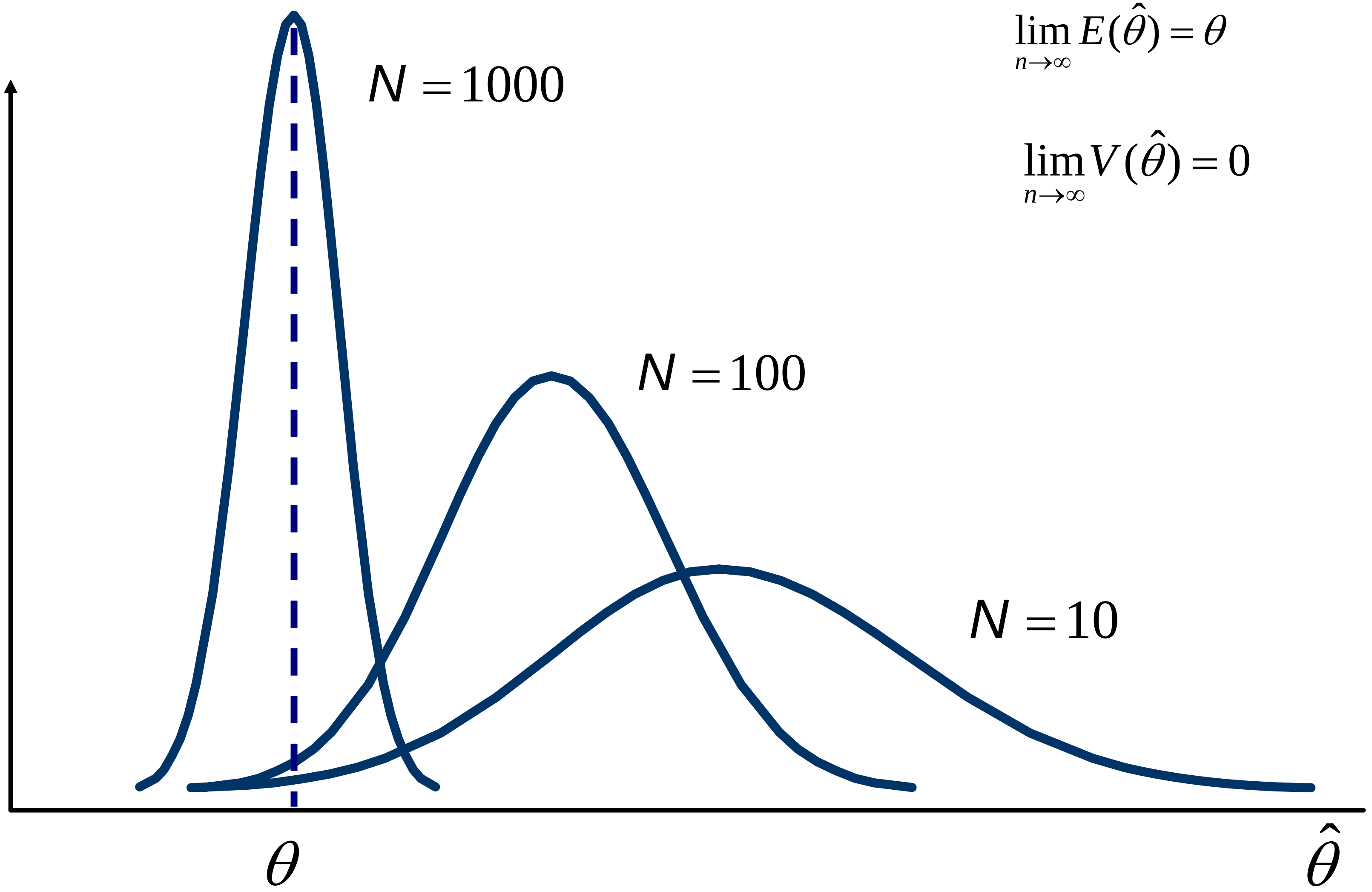
$\hat{\theta}_1$ mais eficiente do que $\hat{\theta}_2$



Propriedades dos estimadores pontuais: Consistente

$$\lim_{n\to\infty} E(\hat{\theta}) = \theta$$

$$\lim_{n\to\infty} V(\hat{\theta}) = 0$$



Alguns métodos de estimacão pontual:

- Método de Máxima Verosimilhança ("Maximum Likelihood" – ML);
  - Verossímil: semelhante à verdade.
  - Amostra verossímil: amostra que fornece a melhor informação possível sobre um paramêtro que desejamos estimar.
  - Princípio da verossimilhança: devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter uma amostra observada, ou seja, o valor que torna aquela amostra mais provável.
  - Suponha que temos n provas de Bernoulli com sucesso *p*. Qual o valor de *p* pelo método de máxima verossimilhança?

$$L(p)=\prod_{i=1}^{n} f_{dp}(x,p)=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

  - O maximo da função acima também ocorre no logaritmo neperiano dessa função

$$\ln L(p)=\sum_{i=1}^{n} x_i \log p+(1-x_i)\log(1-p)$$

  - Derivando e igualando a zero

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

Alguns métodos de estimacão pontual:

- Momentos:
- Seja X uma variável aleatória o k-ésimo momento de X em torno de zero é

$$E(X^k)$$

- Sejam $X_1,.., X_k$ uma amostra e $\beta_1,.., \beta_k$ parâmetros a estimar. Esses parâmetros são estimadores pelo método dos momentos se eles forem soluções das equações.

$$E(X^k) = m_k \qquad\qquad m_k = \frac{1}{n}\sum_{i-1}^{n} X_i^k$$

  - Suponha que temos n provas de Bernoulli com sucesso *p* e X=numero de sucessos. Qual o valor de *p* pelo método de momentos?
  - Como existe um parâmetro, o princípio é igualar o primeiro momento populacional com o primeiro momento amostral.

$$p = \frac{x}{n}$$

Portanto $\hat{p} = x / n$

Alguns métodos de estimacão pontual:

- Mínimos quadrados:
- Um engenheiro está estudando a resistência Y de uma fibra em função do seu diâmetro X e notou a seguinte relação

$$Y = \beta X$$

- Considere a amostra:
- X: 1,2 1,5 1,7 2,0 2,6 $\overline{X} = 1{,}8$
- Y: 3,9 4,7 5,6 5,8 7,0 $\overline{Y} = 5{,}4$
- O estimador de β é obtido encontrando o mínimo da função

$$S = \sum_{i=1}^{5} (Y_i - \beta X_i)^2$$

- Assim, derivando e igualando a zero

$$\hat{\beta} = \frac{\sum_{i=1}^{5} X_i Y_i}{\sum_{i=1}^{5} X_i^2} = 2{,}94$$

Intervalo de Confiança para μ

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

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

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

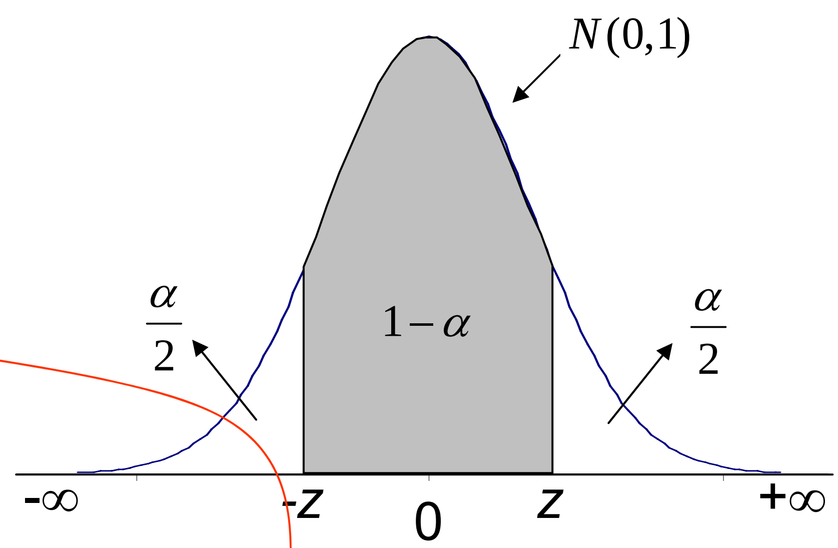
z

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

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

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

IC para *m*



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

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

Intervalo de Confiança para μ

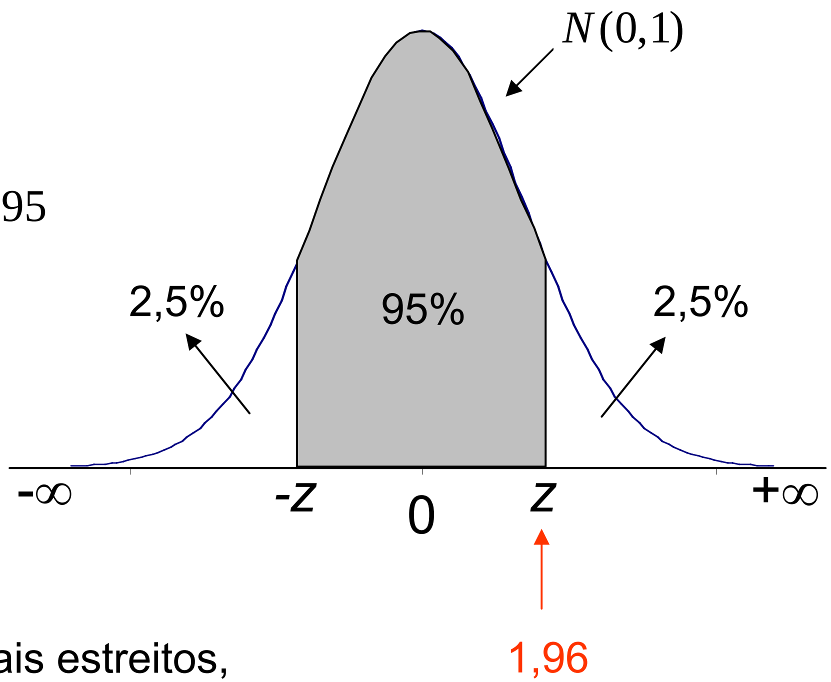
Exemplo: uma v.a. qualquer tem uma distribuição desconhecida com média *m* também desconhecida e variância σ^2 = 16. Retira-se uma amostra de 25 valores e calcula-se a média amostral. Construa um IC de 95% para *m* supondo que $\bar{X} = 12,7$.

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

$$P(12,7 - 1,96\frac{4}{\sqrt{25}} < \mu < 12,7 + 1,96\frac{4}{\sqrt{25}}) = 0,95$$

$$P(12,7 - 1,568 < \mu < 12,7 + 1,568) = 0,95$$

$$\boxed{P(11,132 < \mu < 14,268) = 0,95}$$



Como poderia obter intervalos de confiança mais estreitos, ou seja, com limites mais próximos a média verdadeira?

- diminuindo-se o nível de confiança
- aumentando-se o tamanho da amostra

Como Interpretar o IC para μ?

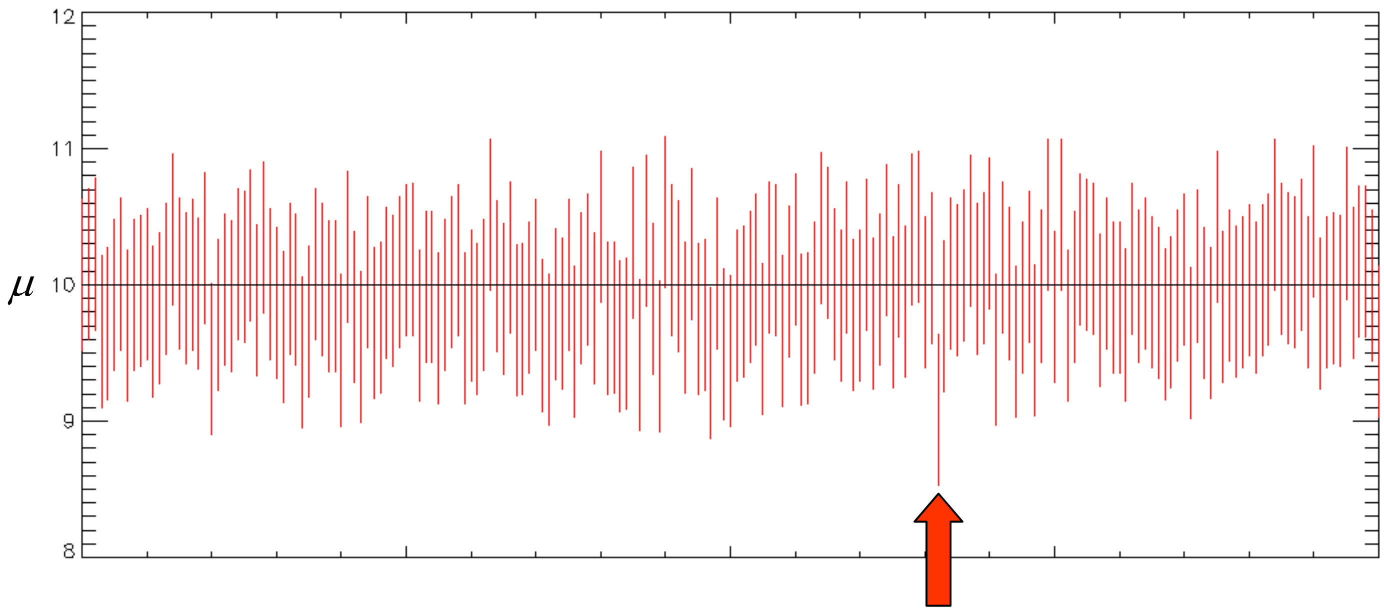
Suponha uma v.a. X normalmente distribuída com μ = 10 e σ^2 = 4 $\rightarrow$ $X \sim N(10,4)$

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

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

$$P(\bar{X} - 0{,}5544 < \mu < \bar{X} + 0{,}5544) = 95\%$$

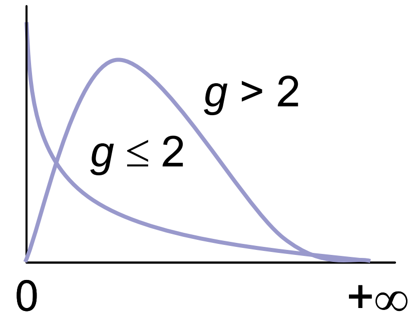
Interpretação: 95% dos possíveis ICs obtidos a partir de uma amostra de tamanho 50, conterão de fato a verdadeira média μ



Distribuição χ^2

(lê-se qui-quadrado)

$$f(x) = \frac{1}{2^{g/2}\Gamma(g/2)} x^{g/2-1} e^{-x/2} \qquad x \geq 0$$



$$\left.\begin{array}{l} E(X) = g \\ Var(X) = 2g \end{array}\right\} \quad X \sim \chi^2_g$$

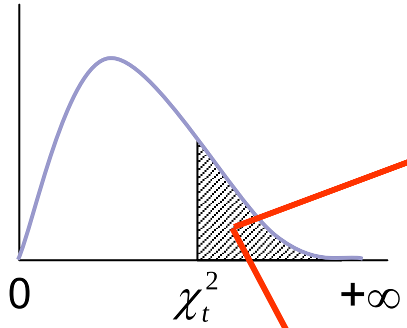
(lê-se: *X* tem distribuição qui-quadrado com *g* graus de liberdade)

Propriedades:

a) se $Z \sim N(0,1)$, então $Z^2 \sim \chi^2_1$

b) se $X_i \sim \chi^2_1$, então $\sum_{i=1}^{n} X_i \sim \chi^2_n$

Distribuição χ^2



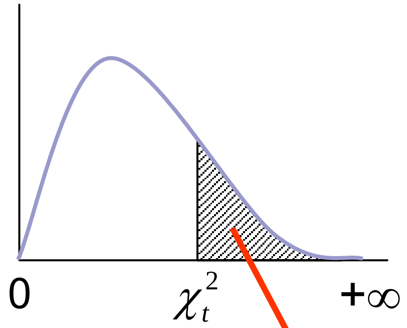
$$P(\chi_g^2 > \chi_t^2)$$

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	[illegible]	0,30	0,21
5	16,75	15,09	12,83	11,07	9,24	1,61	1,15	[illegible]	0,55	0,41
6	18,55	16,81	14,45	12,59	10,64	2,20	1,64	[illegible]	0,87	0,68
7	20,28	18,48	16,01	14,07	12,02	2,83	2,17	[illegible]	1,24	0,99
8	21,95	20,09	17,53	15,51	13,36	3,49	2,73	2,18	1,65	1,34
9	23,59	21,67	19,02	16,92	14,68	4,17	3,33	2,70	2,09	1,73
10	25,19	23,21	20,48	18,31	15,99	4,87	3,94	3,25	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

$$P(\chi_{10}^2 > 3,25) = ?$$

$$P(\chi_{10}^2 > 3,25) = 0,975$$

Distribuição χ^2



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

$$P(\chi_{10}^2 > 3,25) = ?$$

$$P(\chi_{10}^2 > 3,25) = 0,975$$

$$P(\chi_{15}^2 > ?) = 0,9$$

$$P(\chi_{15}^2 > 8,55) = 0,9$$

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	14,45	12,59	10,64	[illegible]	1,64	1,24	0,87	0,68
7	20,28	18,48	16,01	14,07	12,02	[illegible]	2,17	1,69	1,24	0,99
8	21,95	20,09	17,53	15,51	13,36	[illegible]	2,73	2,18	1,65	1,34
9	23,59	21,67	19,02	16,92	14,68	[illegible]	3,33	2,70	2,09	1,73
10	25,19	23,21	20,48	18,31	15,99	[illegible]	3,94	3,25	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	$\boxed{8,55}$	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

Distribuição χ^2

Se $X_i \sim N(\mu, \sigma^2)$ ⇨ $\dfrac{X_i - \mu}{\sigma} \sim N(0,1)$

$$\frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_1^2$$

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

Substituindo-se μ por $\bar{X}$ tem-se que

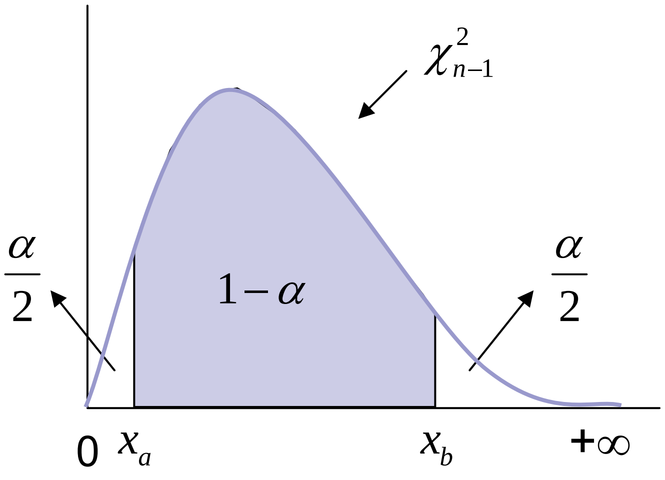
$$\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$ (perde-se 1 grau de liberdade)

mas $s^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \Rightarrow \sum_{i=1}^{n} (X_i - \bar{X})^2 = (n-1)s^2$

logo $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$

Intervalo de Confiança para σ^2

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



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

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

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

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

IC para σ^2

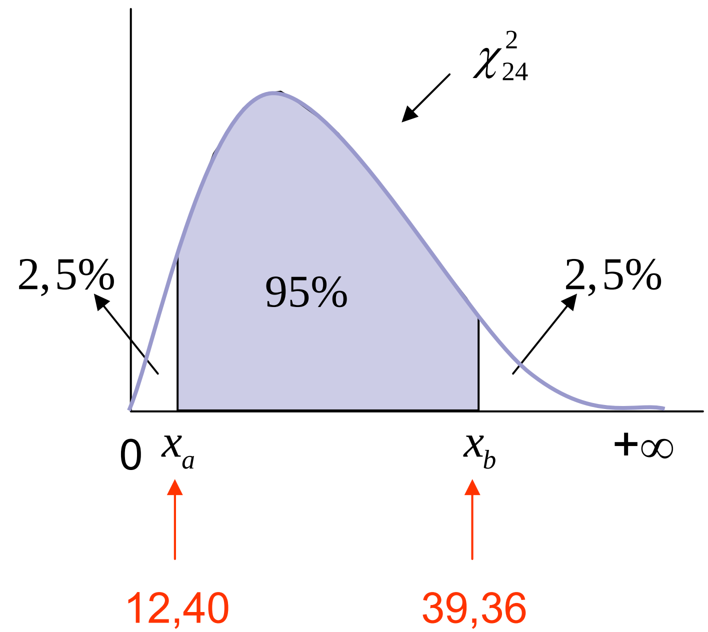
Intervalo de Confiança para σ^2

Exemplo: uma v.a. qualquer tem uma distribuição desconhecida 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.

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

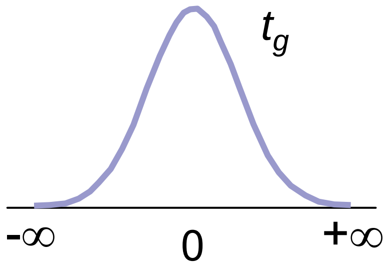
$$P\left(\frac{24\;2,34}{39,36} < \sigma^2 < \frac{24\;2,34}{12,40}\right) = 0,95$$

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



Distribuição *t de student*

$$f(x) = \frac{\Gamma[(g+1)/2]}{\Gamma(g/2)\sqrt{\pi g}}\left(1+\frac{x^2}{g}\right)^{-(g+1)/2} \qquad -\infty < x < \infty$$



$$\left.\begin{array}{l} E(X)=0 \\ Var(X)=\dfrac{g}{g-2} \end{array}\right\} \quad X \sim t_g$$

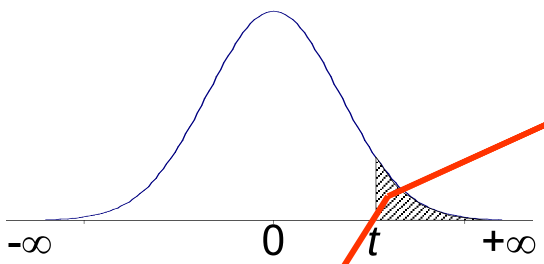
(lê-se: X tem distribuição *t de student* com g graus de liberdade)

Propriedades:

a) se $Z \sim N(0,1)$ e $W \sim \chi_g^2$ então $\dfrac{Z}{\sqrt{\dfrac{W}{g}}} \sim t_g$

b) se $n \to \infty$ então $t_g \to N(0,1)$

Distribuição *t de student*



$$P(0 < T_g < t)$$

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	[illegible]	4,604
5	1,476	2,015	2,571	[illegible]	4,032
6	1,440	1,943	2,447	[illegible]	3,707
7	1,415	1,895	2,365	[illegible]	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	$\boxed{2{,}764}$	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	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

$$P(T_{10} > 2{,}764) = ?$$

$$P(T_{10} > 2{,}764) = 0{,}01$$

Distribuição *t de student*

Se $X_i \sim N(\mu, \sigma^2)$

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}}} = \frac{\frac{\bar{X} - \mu}{\frac{\cancel{\sigma}}{\sqrt{n}}}}{\frac{s}{\cancel{\sigma}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

Intervalo de Confiança para μ

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

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

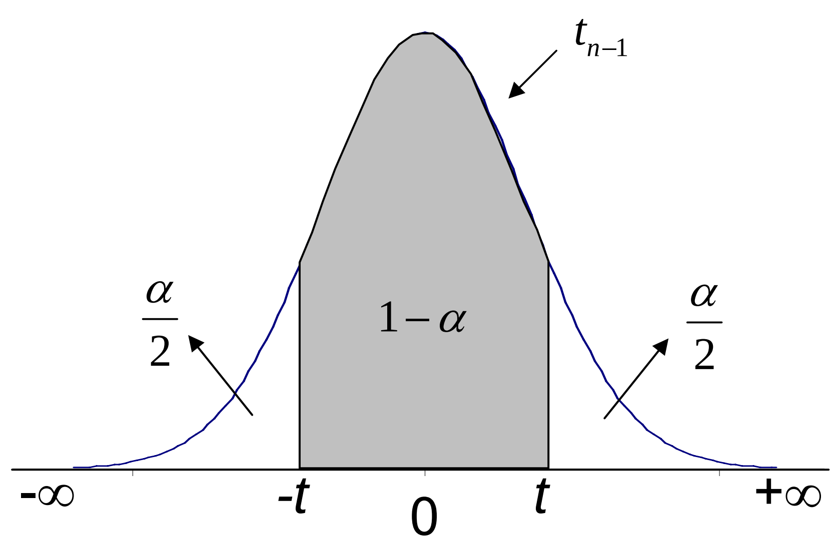
T

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

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

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

IC para *m*



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

Intervalo de Confiança para μ

Exemplo: uma v.a. qualquer tem uma distribuição desconhecida com média μ e variância σ^2 também 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 e $s^2 = 16$. $\bar{X} = 12,7$

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

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

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

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