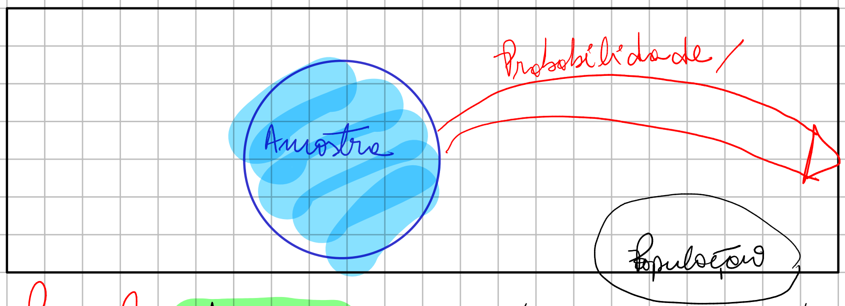
Objetivo:



Inferência: { Estimação Pontual, Estimação Intervalar, Teste de Hipóteses

Estimação Pontual: Aproximar parâmetro por estimativa. Ex.: $\left\{ \begin{array}{l} \hat{\mu} = \bar{x} \approx \mu \\ \hat{\sigma}^2 \approx \sigma^2 \end{array} \right.$

Notação: Colocamos "^" para indicar que a estimativa é uma aproximação do parâmetro!

$\hat{\mu} \approx \bar{x}$ ← média da população ~~tão perto~~ de $\bar{x}$ ~~tão perto~~ O que é perto

Tabela 1: Encontrar o valor do parâmetro dos modelos de probabilidade.
Seja $x_1, \ldots, x_m$ os valores observados de uma variável quantitativa X em uma amostra, então:

Amostra	Distribuição	Parâmetros	Estimador
$X_1, \ldots, X_m$	$U_D[j,k]$ $f(x) = \frac{1}{k-j+1}, \quad x = j, \ldots, k$	j k	$\hat{j} = \min\{x_1, x_2, \ldots, x_m\}$ $\hat{k} = \max\{x_1, x_2, \ldots, x_m\}$
$X_1, \ldots, X_m$	$b(n,p)$ $f(x) = \binom{n}{x} p^x \cdot (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$	p n conhecido	$\hat{p} = \frac{\bar{x}}{n} = \frac{x_1 + x_2 + \cdots + x_m}{n \cdot m}$
$X_1, \ldots, X_m$	Bernoulli(p) $f(x) = p^x \cdot (1-p)^{1-x}, \quad x = 0, 1$	p $p \approx \hat{p}$	$\hat{p} = \bar{x} = \frac{x_1 + x_2 + \cdots + x_m}{m}$
$X_1, \ldots, X_m$	Poison(λ) $f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, 3, \cdots$	λ $\lambda \approx \hat{\lambda}$	$\hat{\lambda} = \bar{x} = \frac{x_1 + x_2 + \cdots + x_m}{m}$
$X_1, \ldots, X_m$	$U[a,b]$ $f(x) = \frac{1}{b-a}, \quad x \in [a,b]$	a b	$\hat{a} = \min\{x_1, x_2, \ldots, x_m\}$ $\hat{b} = \max\{x_1, x_2, \ldots, x_m\}$
$X_1, \ldots, X_m$	Exponencial(α) $f(x) = \alpha \exp(-\alpha x), \quad x \geq 0$	α	$\hat{\alpha} = \frac{1}{\bar{x}} = \frac{m}{x_1 + x_2 + \cdots + x_m}$
$X_1, \ldots, X_m$	Normal(μ, σ^2) $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ, σ^2	$\hat{\mu} = \frac{x_1 + x_2 + \cdots + x_m}{m}$ $\hat{\sigma}^2 = \frac{(x_1 - \hat{\mu})^2 + (x_2 - \hat{\mu})^2 + \cdots + (x_m - \hat{\mu})^2}{m-1}$

→ Dá provar que são estimativas boas!

Intervalo de Confiança: Encontrar $\hat{a}$ e $\hat{b}$ tal que $\left| \hat{a} \leq \mu \leq \hat{b} \right|$ com alguma probabilidade P. (Margem de erro) σ^2 $\hat{\mu}$

Distribuição Normal: $N(\mu, \sigma^2)$ (σ^2 conhecido).

Desvio Padrão
Proporção

Vamos começar com distribuição amostral.

População $N(\mu, \sigma^2)$, σ^2 conhecido

Amostras de tamanho n

Amostra 1 → $y_1, y_2, \ldots, y_n : \bar{y} = \frac{y_1 + \ldots + y_n}{n}$

Amostra 2 → $x_1, x_2, \ldots, x_n : \bar{x} = \frac{x_1 + \ldots + x_n}{n}$

⋮

Amostra K → $z_1, z_2, \ldots, z_n : \bar{z} = \frac{z_1 + \ldots + z_n}{n}$

→ K Médias diferentes

Abstração: Essas K médias são realizações de uma variável aleatória, e representamos essa variável aleatória por $\bar{X}$.

Teorema do Limite Central: $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$.

Exemplo:

Tabela 1: Nota dos 30 alunos na primeira prova.

5.80	7.27	7.04	6.83	7.64	7.11	7.51	7.30	6.89	6.30
5.60	8.80	7.05	6.13	5.42	6.24	8.54	5.74	5.43	5.95
6.82	5.73	7.84	5.90	7.80	5.48	8.59	5.76	3.90	7.10

Teorema do Limite Central: O modelo de probabilidade adequado para médias é Normal!

$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$ $(n=5)$ Teorema! Teorema do Limite Central

Tabela 2: Dez amostras com cinco alunos provenientes dos 30 alunos da Tabela 1.

Amostras	Notas dos cinco indivíduos da amostra de Heloísa					Média
Amostra 1	6,83	7,30	6,89	7,51	3,90	6,49
Amostra 2	5,76	6,13	8,59	6,89	6,82	6,84
Amostra 3	5,80	5,48	6,13	6,82	6,24	6,09
Amostra 4	8,59	5,73	5,80	7,10	6,83	6,81
Amostra 5	6,82	6,30	6,13	7,10	8,80	7,03
Amostra 6	7,51	8,59	5,73	3,90	6,82	6,51
Amostra 7	5,95	6,89	5,73	3,90	7,30	5,95
Amostra 8	8,59	5,74	6,30	7,10	6,24	6,79
Amostra 9	6,83	7,80	7,51	6,89	5,95	7,00
Amostra 10	5,43	7,80	6,89	5,80	7,30	6,64

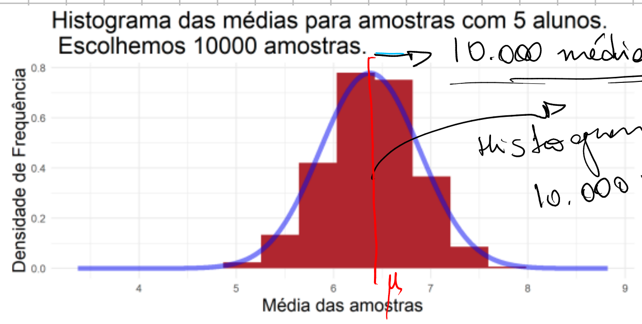


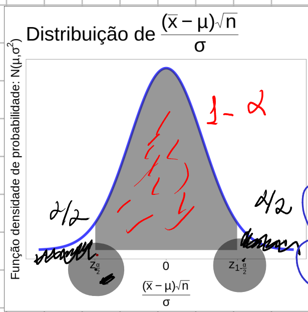
Figura 3: Histograma das médias de amostras com cinco alunos dos alunos da tabela 1, e o modelo de probabilidade segundo o Teorema Central do Limite (curva azul).

Amágico!

Note que $\dfrac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$, ou seja,

Para olhar na Tabela

$\boxed{\bar{X} \sim N(\mu;\ \sigma^2/n)}$ $\qquad \dfrac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} = \dfrac{(\bar{X}-\mu)\sqrt{n}}{\sigma}$



Notação: $\begin{cases} P(Z \leq z_{\alpha/2}) = \alpha/2 \\ P(Z \leq z_{1-\alpha/2}) = 1-\alpha/2 \end{cases}$

$z_{\alpha/2}$ | $P(Z \leq z_{\alpha/2}) = \alpha/2$

$z_{1-\alpha/2}$ | $P(Z \leq z_{1-\alpha/2}) = 1-\alpha/2$

$z_{\alpha/2} = -z_{1-\alpha/2}$

Então $P\left(z_{\alpha/2} \leq \dfrac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \leq z_{1-\alpha/2}\right) = 1-\alpha$

$$P\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}-\bar{X} \leq -\mu \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}-\bar{X}\right) = P\Big(\underbrace{z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}+\bar{X}}_{\hat{a}} \leq \mu \leq \underbrace{z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}+\bar{X}}_{\hat{b}}\Big) = 1-\alpha$$

Encontramos nosso $\hat{a}$ e $\hat{b}$!

$1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05$

$\alpha/2 = 0{,}025 \Rightarrow 1-\alpha/2 = 0{,}975$

Notação: $IC(\mu;\ 1-\alpha) = \left(-z_{1-\alpha/2}\cdot\dfrac{\sigma}{\sqrt{n}}+\bar{X};\ z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}+\bar{X}\right)$.

ATENÇÃO À INTERPRETAÇÃO

Tabela 3: Intervalos de confiança e amostras de uma população com distribuição normal com média populacional $\mu = 1,75$ e desvio padrão $\sigma = 0,1$.

μ		Amostra					a	b	$a < \mu < b$?
1,75	Amostra 1	2,050	1,909	1,893	1,858	1,651	1,785	1,960	Não
	Amostra 2	1,667	1,909	1,958	1,771	2,028	1,779	1,954	Nao
	Amostra 3	1,835	1,905	1,995	1,805	1,820	1,784	1,960	Não
σ	Amostra 4	1,824	1,870	1,965	1,637	1,711	1,714	1,889	Sim
0,1	Amostra 5	1,773	1,796	1,895	1,872	1,812	1,742	1,917	Sim
	Amostra 6	1,741	1,885	1,896	1,629	1,664	1,675	1,851	Sim

$N(1{,}75;\ 0{,}1^2)$

$1{,}785 \leq \cancel{1{,}75} \leq 1{,}960$

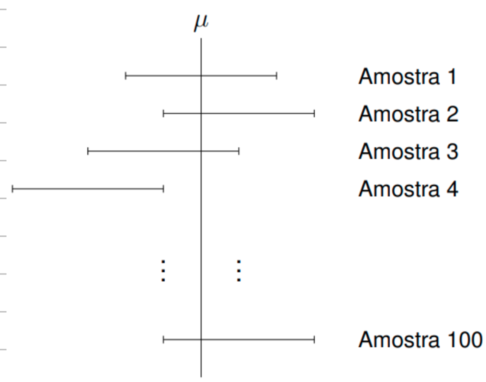
μ pode ou não satisfazer $-z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}+\bar{x} \leq \mu \leq z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}+\bar{x}$!

O importante é que $(1-\alpha)\%$ dos intervalos de confiança vão conter a média populacional.

Desvio Padrão

proporção

Figura 1: Interpretação do coeficiente de confiança.



4. Por analogia com produtos similares, o tempo de reação de um novo medicamento tem distribuição normal com desvio padrão igual a 2 minutos (a média é desconhecida). Vinte pacientes foram sorteados, receberam o medicamento e tiveram seu tempo de reação anotado. Os dados foram os seguintes (em minutos): 2,9; 3,4; 3,5; 4,1; 4,6; 4,7; 4,5; 3,8; 5,3; 4,9; 4,8; 5,7; 5,8; 5,0; 3,4; 5,9; 6,3; 4,6; 5,5 e 6,2. Obtenha um intervalo de confiança para o tempo médio de reação. Use como coeficiente de confiança $\gamma = 96\%$.

Solução: $\sigma = 2$ min; $n = 20$; $1-\alpha = 0,96 \Rightarrow 1 - \frac{\alpha}{2} = \frac{0,96+1}{2} = 0,98$

$z_{0,98} = 2,05$; $\bar{x} = \frac{2,9+3,4+3,5+\dots+5,5+6,2}{20} = 4,745$

$$IC(\mu; 96\%) = \left(-z_{0,98}\cdot\frac{\sigma}{\sqrt{n}} + \bar{x}\,;\, z_{0,98}\frac{\sigma}{\sqrt{n}} + \bar{x}\right) = \left(-2,05\cdot\frac{2}{\sqrt{20}} + 4,745\,;\, 2,05\cdot\frac{2}{\sqrt{20}} + 4,745\right)$$

$IC(\mu; 96\%) = (3,83; 5,66)$.

Em 96% das amostras, a média populacional satisfaz $3,83 \leq \mu \leq 5,66$.

No jargão de estatística: com coeficiente de confiança 96%, o tempo médio de reação está entre 3,83 min e 5,66 min. ■

Distribuição Normal: $N(\mu, \sigma^2)$ − não conhecemos σ^2

População: $N(\mu, \sigma^2)$

σ^2 desconhecido

Amostras

Amostra 1 → $x_1, \dots, x_n$: $t_1 = \frac{(\bar{x}-\mu)\sqrt{n}}{s}$

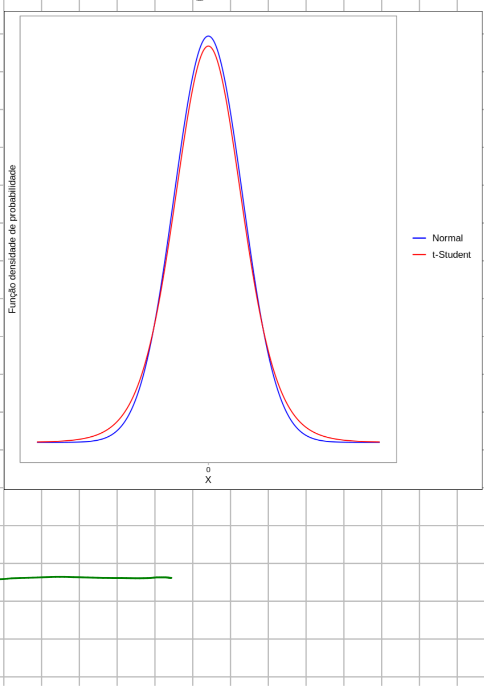
Amostra 2 → $y_1, \dots, y_n$: $t_2 = \frac{(\bar{y}-\mu)\sqrt{n}}{s}$

Amostra k → $z_1, \dots, z_n$: $t_k = \frac{(\bar{z}-\mu)\sqrt{n}}{s}$

→ k valores diferentes!

Abstração: Esses k valores são observados uma variável aleatória t_{n-1}.

Notação: $\frac{(\bar{X}-\mu)\sqrt{n}}{S} \sim t_{n-1}$. (Nova distribuição)



- Valores concentrados em torno do zero
- Intervalos longe do zero são pouco prováveis
- Intervalos longe de zero p/ t-Student são mais prováveis do que N(0,1).

Função Densidade de Probabilidade: $f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$, $\nu = n-1$.

NÃO TEM PRIMITIVA!!! → Métodos Numéricos; Tabela t

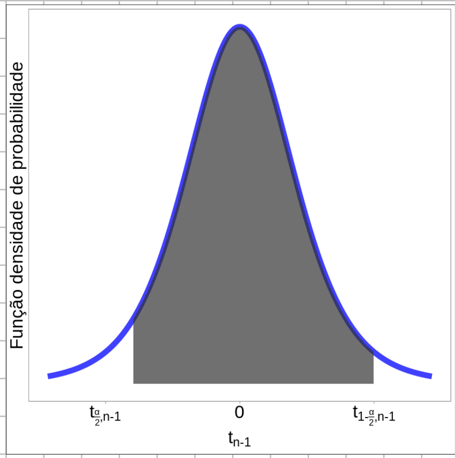
A mágica!

Note que $t_{\frac{\alpha}{2};n-1} = -t_{1-\alpha/2;n-1}$

$$P\left(t_{\alpha/2;n-1} \leq \frac{(\overline{X}-\mu)\sqrt{n}}{S} \leq t_{1-\alpha/2;n-1}\right) =$$

$$P\left(-t_{1-\alpha/2;n-1} \cdot \frac{S}{\sqrt{n}} + \overline{X} \leq \mu \leq t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}} + \overline{X}\right) = 1-\alpha$$

$$IC(\mu; 1-\alpha) = \left[\overline{X} - t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}} ; \overline{X} + t_{1-\alpha/2;n-1} \cdot \frac{S}{\sqrt{n}}\right]$$



18. Uma máquina produz hastes de metal usadas em sistema de suspensão de automóveis. Uma amostra aleatória de 15 rodas foi coletada, e o diâmetro é mensurado. Os dados (em milímetros) estão na Tabela 2. Assuma a normalidade do diâmetros das hastes de metal. Construa um intervalo de confiança com coeficiente de confiança $\gamma = 99\%$ para o diâmetro das hastes de metal.

8,24	8,21	8,23
8,25	8,26	8,23
8,20	8,26	8,19
8,23	8,20	8,28
8,24	8,25	8,24

Tabela 2: Hastes de metal usadas em sistema de suspensão de automóveis.

Solução: $\bar{x} = 8.234$; $s = 0.03$; $n = 15$; $\gamma = 0,99$; $1 - \frac{\alpha}{2} = \frac{1+\gamma}{2} = 0,995$.

$z_{0,995} = 2,58$.

$$IC(\mu; 95\%) = \left(-2,58 \cdot \frac{S}{\sqrt{n}} + \bar{x} ; 2,58 \cdot \frac{S}{\sqrt{n}} + \bar{x}\right)$$

$$= \left(-2,58 \cdot \frac{0,03}{\sqrt{15}} + 8,234 ; 2,58 \cdot \frac{0,03}{\sqrt{15}} + 8,234\right)$$

$$= (8,21 ; 8,25).$$

Com coeficiente de confiança 99%, a média do diâmetro das hastes de metal está entre 8,21 e 8,25 em milímetros. ■

Distribuição Normal: $N(\mu, \sigma^2)$ (σ^2 desconhecido)

População: $N(\mu, \sigma^2)$
μ – desconhecido
σ^2 – desconhecido

Amostras

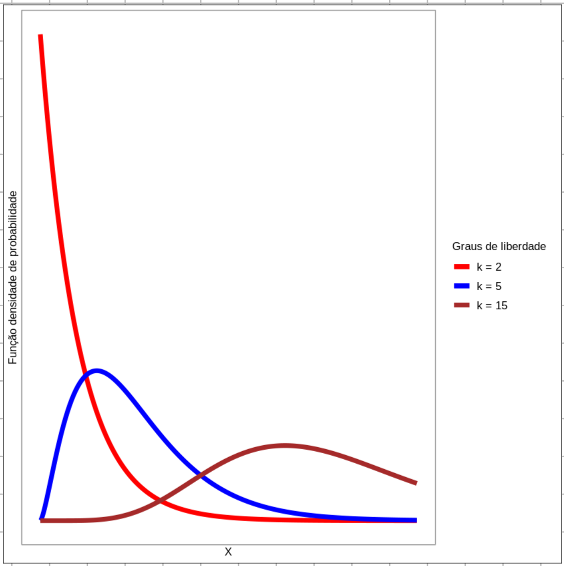
Amostra 1 → $x_1, \ldots, x_n$: $\chi^2 = \frac{(x_1-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{\sigma^2(n-1)}$

Amostra 2 → $y_1, \ldots, y_n$: $\chi^2 = \frac{(y_1-\bar{y})^2 + \cdots + (y_n-\bar{y})^2}{\sigma^2(n-1)}$

Amostra k → $z_1, \ldots, z_n$: $\chi^2 = \frac{(z_1-\bar{z})^2 + \cdots + (z_n-\bar{z})^2}{\sigma^2(n-1)}$

→ k valores diferentes.

Abstração: Esses k valores são observados de uma variável aleatória χ^2_{n-1}.



$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad k = n-1.$$

- Intervalos perto do zero têm maiores chances!

Notação: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

- $P\left(\chi^2_{n-1} \leq \chi^2_{\alpha/2; n-1}\right) = \alpha/2$
- $P\left(\chi^2_{n-1} \leq \chi^2_{1-\alpha/2; n-1}\right) = 1 - \alpha/2$

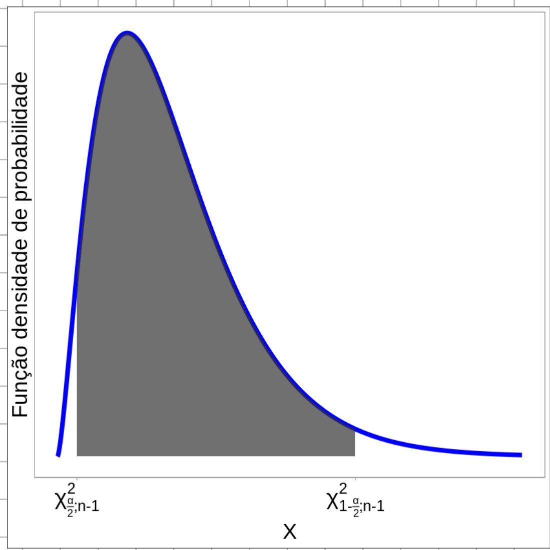
- Não tem primitiva $\begin{cases} \text{Tabela} \\ \text{métodos numéricos} \end{cases}$

A mágica!

$$P\left(\chi^2_{\alpha/2; n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2; n-1}\right) =$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2; n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2; n-1}}\right) = 1-\alpha$$

$$Ic(\sigma^2; \gamma) = \left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2; n-1}} ; \frac{(n-1)S^2}{\chi^2_{\alpha/2; n-1}}\right)$$



21. Um estudo com o objetivo de estudar o nível de composição de aminoácido essencial (Lysine) de farejo de soja está na Tabela 4 (g/kg). Assuma que o nível de composição de aminoácido essencial (Lysine) de farejo de soja tem distribuição normal. Construa um intervalo de confiança com coeficiente de confiança $\gamma = 99\%$ para σ^2.

22,20	20,90	27,00	26,50	25,60
24,70	26,00	24,80	23,80	23,90

Tabela 4: Nível de aminoácido (Lysine) de farejo de soja.

Solução: $S^2 = 3,66$, $n = 10$; $\begin{cases} \frac{1+\gamma}{2} = 1 - \frac{\alpha}{2} = 0,995 \\ \frac{1-\gamma}{2} = \frac{\alpha}{2} = 0,005 \end{cases}$ $\quad \chi^2_{0,995} = 23,59$, $\chi^2_{0,005} = 1,73$

$$Ic(\sigma^2, 99\%) = \left(\frac{(n-1) \cdot S^2}{\chi^2_{1-\alpha/2; n-1}} ; \frac{(n-1)S^2}{\chi^2_{\alpha/2; n-1}}\right) = \left(\frac{9 \cdot 3,66}{23,59} ; \frac{9 \cdot 3,66}{1,73}\right)$$
$$= (1,40; 19,04)$$

Com coeficiente de confiança 99%, o desvio padrão do nível de postássio está entre 1,40 e 19,04.

Distribuição Bernoulli: $X \sim Bernoulli(p)$.

População: Bernoulli(p)

p: Probabilidade de Sucesso

Amostras

Amostra 1 → $x_1, \dots, x_n$: $\frac{(\bar{x}-p)\sqrt{n}}{\sqrt{p(1-p)}}$

Amostra 2 → $y_1, \dots, y_n$: $\frac{(\bar{y}-p)\sqrt{n}}{\sqrt{p(1-p)}}$

Amostra K → $z_1, \dots, z_n$: $\frac{(\bar{z}-p)\sqrt{n}}{\sqrt{p(1-p)}}$

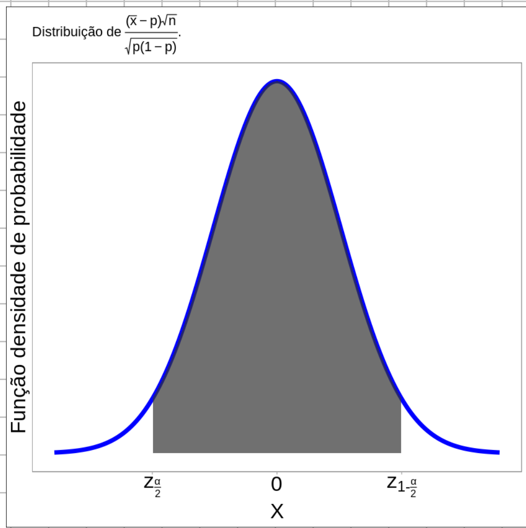
→ K valores diferentes

Abstração: Esses K valores são observações de uma variável aleatória $N(0,1)$

A mágica!

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{(\bar{X}-p)\sqrt{n}}{\sqrt{p(1-p)}} \leq z_{1-\alpha/2}\right)$$

$$= P\left(-z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} + \bar{X} \leq p \leq z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} + \bar{X}\right) = 1-\alpha$$

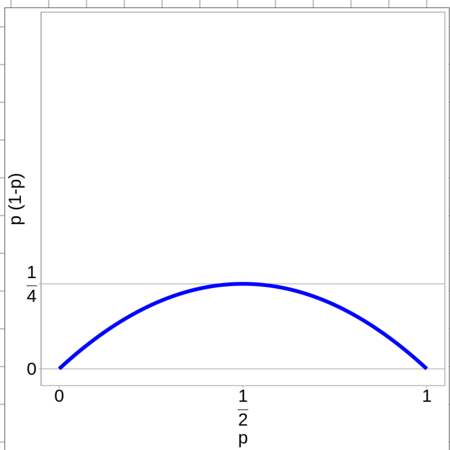


$\bar{x} = \hat{p}$

$$IC(p; 1-\alpha) = \left(-z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} + \bar{X}; z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} + \bar{X}\right)$$

Não conhecemos p!

Solução:



• $p(1-p) \leq \frac{1}{4}$

• $\frac{z_{1-\alpha/2}}{\sqrt{n}} \cdot \sqrt{p(1-p)} \leq \frac{z_{1-\alpha/2}}{\sqrt{n}} \cdot \frac{1}{2}$ e $\frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{p(1-p)} + \bar{X} \leq \frac{z_{1-\alpha/2}}{\sqrt{n}}\frac{1}{2} + \bar{X}$

• $-\frac{z_{1-\alpha/2}}{\sqrt{n}} \cdot \frac{1}{2} \leq -\frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{p(1-p)}$ e $-\frac{z_{1-\alpha/2}}{2\sqrt{n}} + \bar{X} \leq -\frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{p(1-p)} + \bar{X}$

Logo, $IC(p; 1-\alpha) = \left[-z_{1-\alpha/2} \cdot \frac{1}{2\sqrt{n}} + \bar{X}; z_{1-\alpha/2} \cdot \frac{1}{2\sqrt{n}} + \bar{X}\right]$.

22. A fração de circuitos integrados defeituosos produzidos em um processo de fotolitografia está sob análise. Uma amostra aleatória de 300 circuitos foram testadas e descobrimos que 13 circuitos estavam defeituosos. Construa um intervalo de confiança a fração de circuitos defeituosos com coeficiente de confiança $\gamma = 95\%$.

Solução: $n = 300$; $p = \frac{13}{300} = 0{,}04$; $1-\alpha = 0{,}95$; $1 - \frac{\alpha}{2} = 0{,}975$.

$z_{0,975} = 1{,}96$.

$$IC(p; 95\%) = \left[-\frac{1{,}96}{2\sqrt{300}} + 0{,}04; \frac{1{,}96}{2\sqrt{300}} + 0{,}04\right] = [-0{,}02; 0{,}1] = [0; 0{,}1]$$

Com coeficiente de confiança 95%, a proporção de circuitos defeituosos é no máximo 10%.

$X : \Omega \longrightarrow \mathbb{R}$

↳ Fenômeno Aleatório e Probabilidade

- ω_1 é resultado do Fenômeno Aleatório $\Rightarrow X(\omega_1) = x_1$

x_1 é uma realização ou observação de X.

$\bar{x}, \bar{y}, \dots, \bar{z}$: são realizações de $\bar{X} : \Omega^* \longrightarrow \mathbb{R}$

↳ p/ algum espaço amostral e p/ algum Fenômeno Aleatório

- ω_1 resultado do F.A. tal que $\bar{X}(\omega_1) = \bar{x}$
- ω_2 resultado do F.A. tal que $\bar{X}(\omega_2) = \bar{y}$

$\bar{X}$ é uma variável aleatória contínua E $\bar{X} \sim N(\mu ; \sigma^2/n)$