Aula 4
Estatística II

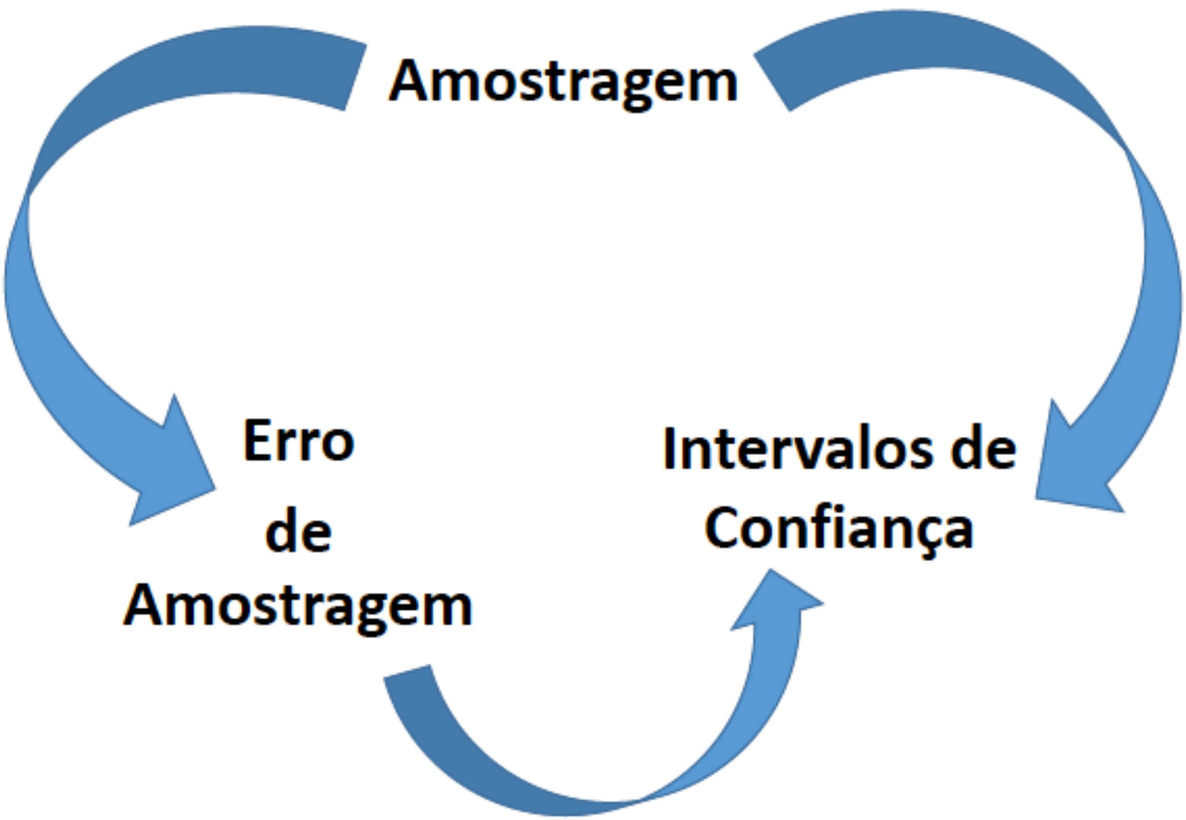
Estimação por intervalos

(Ilustrações e exemplos)

Estimação para obter a confiança

ESTIMAÇÃO POR INTERVALOS

Estimação por intervalos privilegia a confiança



A ter em conta: a variabilidade da amostra

ESTIMAÇÃO POR INTERVALOS

VARIABILIDADE DA AMOSTRA

Para diferentes amostras da mesma população as estimativas assumem diferentes valores

ERRO DE AMOSTRAGEM (ε)

Variação devido à amostragem

$\bar{x} = 130grs$

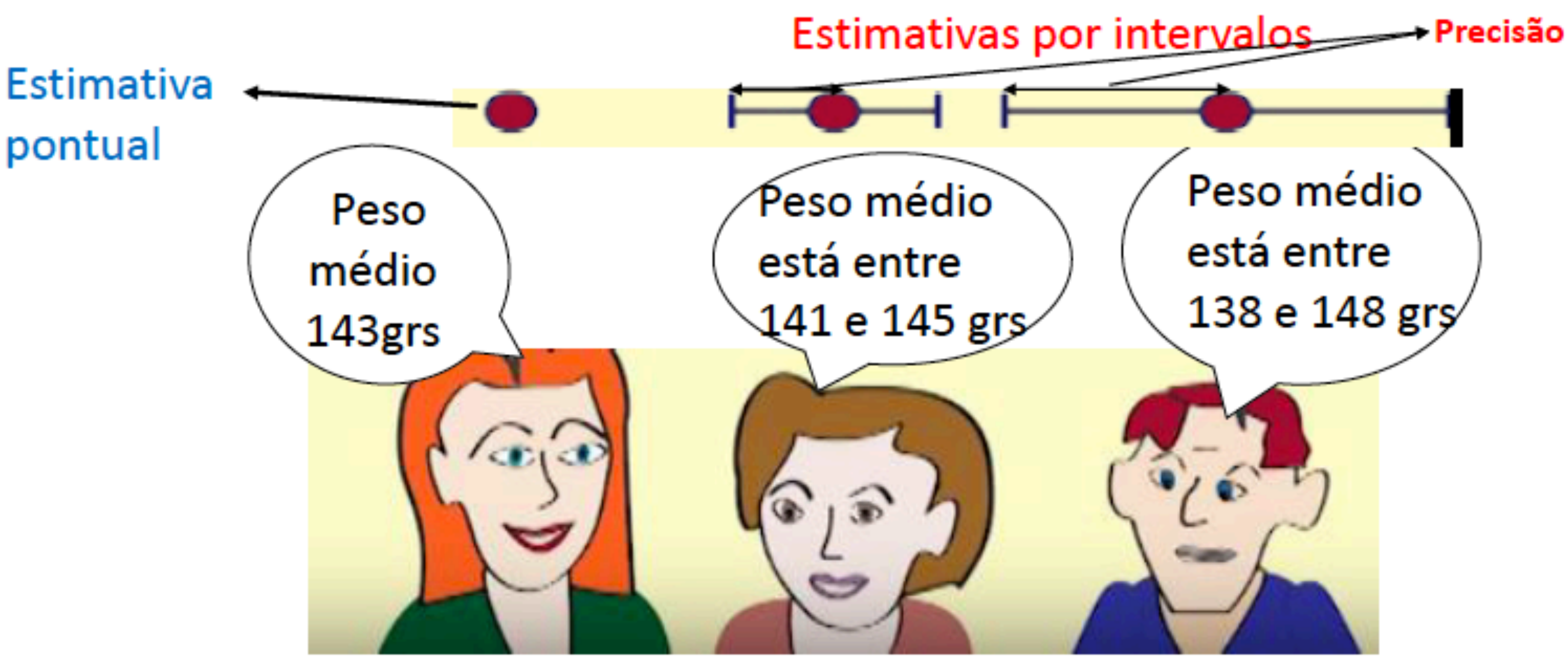
$\bar{x} = 149grs$

$\bar{x} = 153grs$

Qual a precisão associada a um grau de confiança?

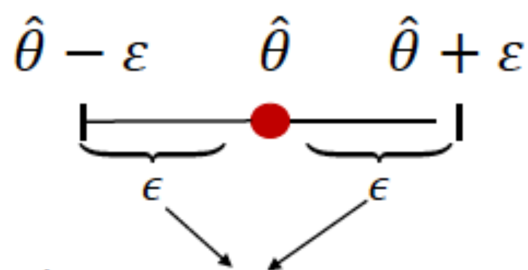
ESTIMAÇÃO POR INTERVALOS

Quando estimamos um parâmetro de uma população é prática comum exprimir a estimativa sob a forma de um intervalo. O intervalo de confiança indica a **precisão** da estimativa associada a um **grau de confiança** fixado.



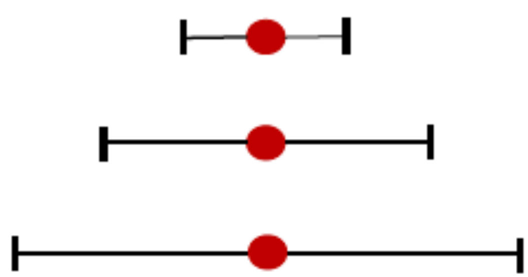
Exprimir a precisão: Intervalo de confiança (IC)

O intervalo de confiança $(\hat{\theta}-\varepsilon, \hat{\theta}+\varepsilon)$ indica a **precisão** da estimativa associada a um **grau de confiança** fixado.



Erro de amostragem\margem de erro – medida de precisão

O que afecta a amplitude do Intervalo de confiança?



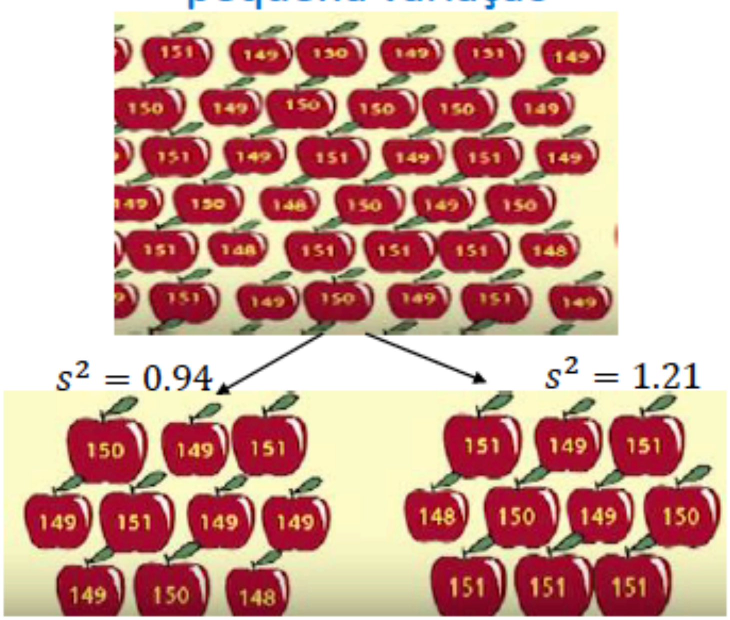
O que afecta o intervalo da confiança?

1. Variação na população
2. A dimensão da amostra
3. O grau de confiança pretendido

Maior a variação na população, maior a amplitude do intervalo da confiança

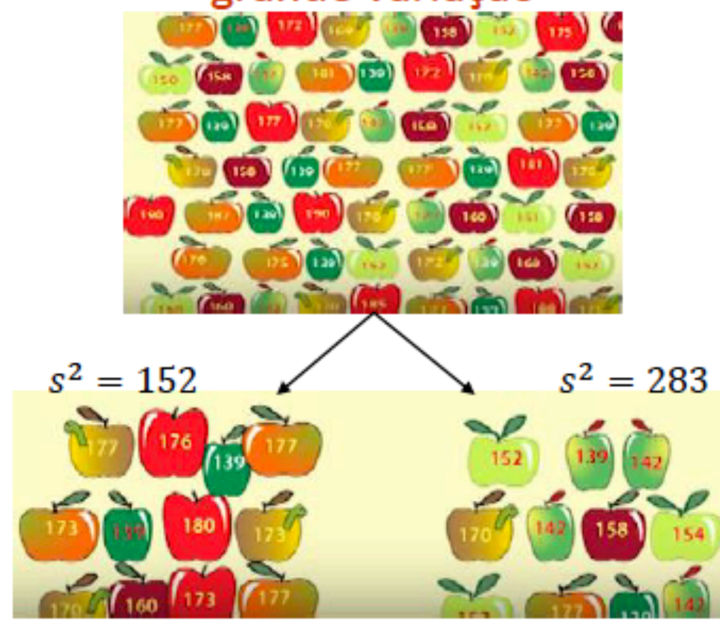
1. Variação na população

População com
pequena variação



Intervalos de confiança com
pequena amplitude ⇔ ε **pequenas**

População com
grande variação



Intervalos de confiança com
grande amplitude ⇔ ε **grandes**

Maior a amostra, menor a amplitude do intervalo

2. Variação na dimensão da amostra

$\bar{x} = 160$
$s^2 = 336.6$

$\bar{x} = 165.6$
$s^2 = 380$

Amostras pequenas contém pouca informação e apresentam maior variabilidade

Intervalos com maior amplitude

$\bar{x} = 160.6$
$s^2 = 15.8$

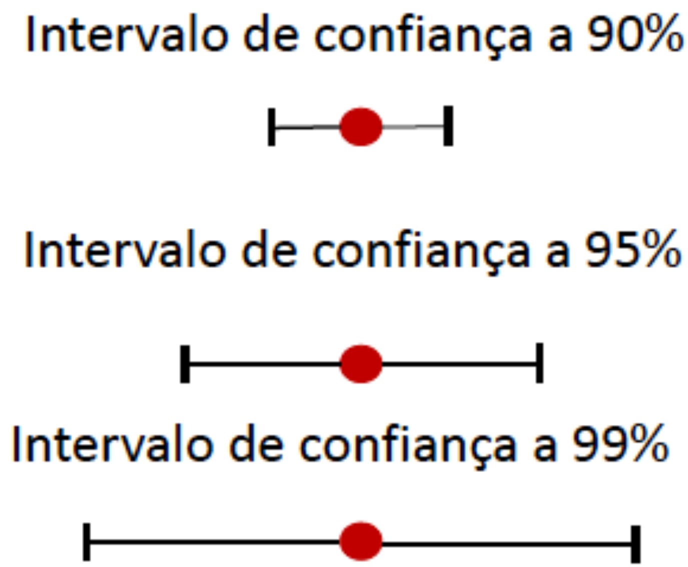
$\bar{x} = 159.4$
$s^2 = 20.3$

Intervalos com menor amplitude

Amostras grandes contém mais informação e apresentam menor variabilidade

Maior o grau de confiança, maior a amplitude do intervalo

3. Variação no grau de confiança



Quanto maior o grau de confiança maior a amplitude do intervalo

menor a precisão

IC pelo método da variável fulcral

- **Obtenção de um intervalo de confiança:**

1. Encontrar uma variável fulcral adequada ao problema em estudo

2. Fixado o grau de confiança, $1-\alpha$, procurar dois números no domínio de Z, $z_1(\alpha)$ e $z_2(\alpha)$ tais que $P\big(z_1(\alpha) < Z < z_2(\alpha)\big) = 1-\alpha \quad \forall\theta \in \boldsymbol{\Theta}$

   Existem muitos (infinitos) pares de valores que satisfazem esta equação. A ideia é, para o mesmo grau de confiança, procurar o par de valores que minimiza a amplitude (ou o seu valor esperado se esta for aleatória) para o intervalo resultante.

   Geralmente faz-se: $P\big(Z < z_1(\alpha)\big) = P\big(Z > z_2(\alpha)\big) = \alpha$.

3. Se a partir de $z_1(\alpha) < Z < z_2(\alpha) \longrightarrow T_1(X_1, X_2, \cdots, X_n) < \theta < T_2(X_1, X_2, \cdots, X_n)$

Então, $P\big(z_1(\alpha) < Z < z_2(\alpha)\big) = P\big(T_1(X_1, X_2, \cdots, X_n) < \theta < T_2(X_1, X_2, \cdots, X_n)\big) = 1-\alpha \quad \forall\theta \in \boldsymbol{\Theta}$

Exemplo 1

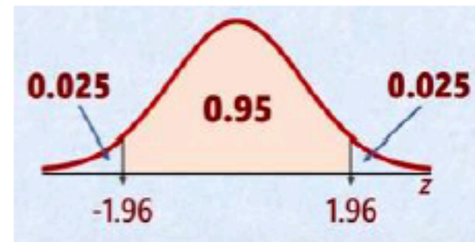
Estimação por intervalos
(variância **conhecida**)

Intervalo de confiança para a média (variância conhecida) – Populações normais

Grau\nível de confiança$= 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

Variável Fulcral $T = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$z_{\alpha/2} : P\left(-z_{\alpha/2} < T < z_{\alpha/2}\right) = 1 - \alpha$



$-z_{\alpha/2}$ $\quad z_{\alpha/2}$

$$-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

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

Intervalo aleatório para $\mu = \left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

Margem de erro

Intervalo confiança para μ a $(1-\alpha) * 100\% = \left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

continuação..

Intervalo de confiança para a média (variância conhecida)

Exemplo: População $X \sim N(\mu, \sigma^2 = \boxed{605})$

Seja amostra casual $(807.7, 790.7, 818.8, 853.4, 858.6) \Rightarrow n = 5,\ \bar{x} = 825{,}84$

Grau\nível de confiança$= 0.95 \Leftrightarrow 1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

Variável fulcral – $T = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

$\Rightarrow z_{0.025} = \mathrm{invnorm}(0.025, 0, 1) = 1.96$ ou lê-se na tabela 5

Margem de erro - $\varepsilon = 1.96 \frac{\boxed{24{,}597}}{\sqrt{5}} = 21{,}56 \ \Rightarrow \ \bar{X} - 21{,}56 < \mu < \bar{X} + 21{,}56$

Intervalo aleatório para $\mu = (\bar{X} - 21{,}56 < \mu < \bar{X} + 21{,}56)$

Intervalo confiança para $\mu = (825.84 - 21.56,\ 825.84 + 21.56) = (804.28,\ \ 847.4)$

Inverse normal distribution

TABELA 5 – DISTRIBUIÇÃO NORMAL: $\Phi^{-1}(z)$

ε	.0005	.0010	.0050	.0100	.0200	.0250	.0500	.1000	.2000	.3000	.4000
z_{ε}	3.290	3.090	2.576	2.326	2.054	1.960	1.645	1.282	.842	.524	.253
$z_{\varepsilon/2}$	3.481	3.290	2.807	2.576	2.326	2.241	1.960	1.645	1.282	1.036	.842

$$z_{\varepsilon} : P(Z > z_{\varepsilon}) = \varepsilon \, ; \quad z_{\varepsilon/2} : P(|Z| > z_{\varepsilon/2}) = \varepsilon \, .$$

Efeito da amostra (**n**) na estimação por intervalos

Exemplo (continuação): População $X \sim N(\mu, \sigma^2 = 605)$

Grau\nível de confiança$= 0.95 \Leftrightarrow 1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05$

Amostra anterior: $n = 5$, $\bar{x} = 825{,}84$

$I.C.$ para $\mu = (804.28, \;\; 847.4)$

Amplitude do $I.C. = \boxed{43.12}$

Amostra:

$$\begin{Bmatrix} 807.7, 790.7, 818.8, 853.4, 858.6, \\ 798.6, 812.2, 813.1, 839.4, 812.8 \end{Bmatrix}$$

$n = 10$, $\bar{x} = 820.53$

$I.C.$ para $\mu = (805.28, \;\; 835.78)$

Amplitude do $I.C. = \boxed{30.49}$

Amplitude do intervalo de confiança reduz-se quando aumenta a dimensão da amostra

Exemplo 2
Estimação por intervalos (variância **desconhecida**)

Exemplo: População $X \sim N(\mu, \sigma^2)$, μ, σ^2 desconhecidos

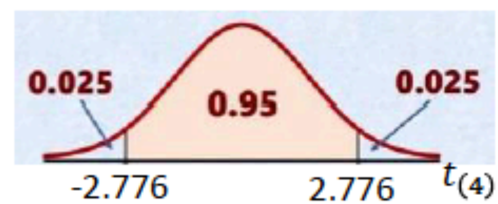
Grau\nível de confiança$= 0.95 \Leftrightarrow 1-\alpha = 0.95 \Leftrightarrow \alpha = 0.05$

Distribuição T-student

Seja amostra casual $(807.7, 790.7, 818.8, 853.4, 858.6) \Rightarrow n = 5,\ \bar{x} = 825{,}84, s' = 29.352$

Variável fulcral $-T = \frac{\bar{X}-\mu}{s'/\sqrt{n}} \sim t\left(\underbrace{5-1}_{4}\right)$

$\Rightarrow t_{(4)}^{0.025} = \text{invt}(0.025, 4) = 2.776$

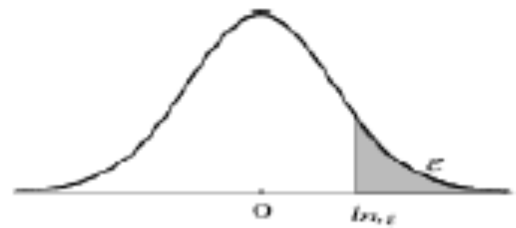


Margem de erro - $\varepsilon = 2.776\frac{29.352}{\sqrt{5}} = 36.44 \Rightarrow \bar{X} - 36.44 < \mu < \bar{X} + 36.44$

Intervalo aleatório para $\mu = (\bar{X} - 36.44 < \mu < \bar{X} + 36.44)$

Intervalo confiança para $\mu = (825{,}84 - 36.44,\ 825.84 - 36.44) = (789.40,\ \ 862.28)$

$$t_{n,\varepsilon} : P(X > t_{n,\varepsilon}) = \varepsilon$$



n \ ε	.400	.250	.100	.050	.025	.010	.005	.001
1	.325	1,000	3.078	6.314	12.706	31.821	63.656	318.289
2	.289	.816	1.886	2.920	4.303	6.965	9.925	22.328
3	.277	.765	1.638	2.353	3.182	4.541	5.841	10.214
4	.271	.741	1.533	2.132	2.776	3.747	4.604	7.173
5	.267	.727	1.476	2.015	2.571	3.365	4.032	5.894
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.208
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.785
8	.262	.706	1.397	1.860	2.306	2.896	3.355	4.501
9	.261	.703	1.383	1.833	2.262	2.821	3.250	4.297
10	.260	.700	1.372	1.812	2.228	2.764	3.169	4.144

Exemplo 2 continuação:

grau de confiança 95% → 99%

Efeito da variação do grau de confiança na amplitude do Intervalo de confiança para a média (variância desconhecida)

Exemplo: População $X \sim N(\mu, \sigma^2)$, μ, σ^2 desconhecidos

Amostra anterior: $n = 5$, $\bar{x} = 825{,}84$, $s' = 29.352$

Grau de confiança $= 0.95$

$$\Leftrightarrow 1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05$$

$$t_{(4)}^{0.025} = 2.776$$

$I.C.$ para $\mu = (789.40,\ 862.28)$

Amplitude do $I.C. = 72.88$

Grau de confiança$= 0.99$

$$\Leftrightarrow 1 - \alpha = 0.99 \Leftrightarrow \alpha = 0.01$$

$$t_{(4)}^{0.005} = 4.604$$

$I.C.$ para $\mu = (765.40,\ \ 886.28)$

Amplitude do $I.C. = 120.87$

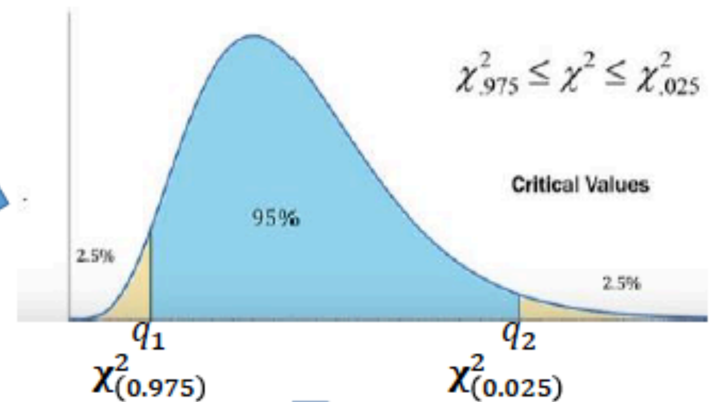
A amplitude do intervalo de confiança aumenta quando se aumenta o grau de confiança

Exemplo 3: IC para a variância

Intervalo de confiança para a variância – Populações normais

Grau\nível de confiança$= 1-\alpha = 0.95 \Rightarrow \alpha = 0.05$

Variável Fulcral $T = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{(n-1)}$



$$q_1 < \frac{(n-1)S'^2}{\sigma^2} < q_2$$

$$\frac{(n-1)S'^2}{q_2} < \sigma^2 < \frac{(n-1)S'^2}{q_1}$$

Intervalo aleatório para $\sigma^2 = \left(\frac{(n-1)S'^2}{q_2}, \frac{(n-1)S'^2}{q_1}\right)$

Intervalo confiança para $\sigma^2 = \left(\frac{(n-1)s'^2}{q_2}, \frac{(n-1)s'^2}{q_1}\right)$

Aplicação concreta: IC para a variância

Intervalo de confiança para a variância – Populações normais

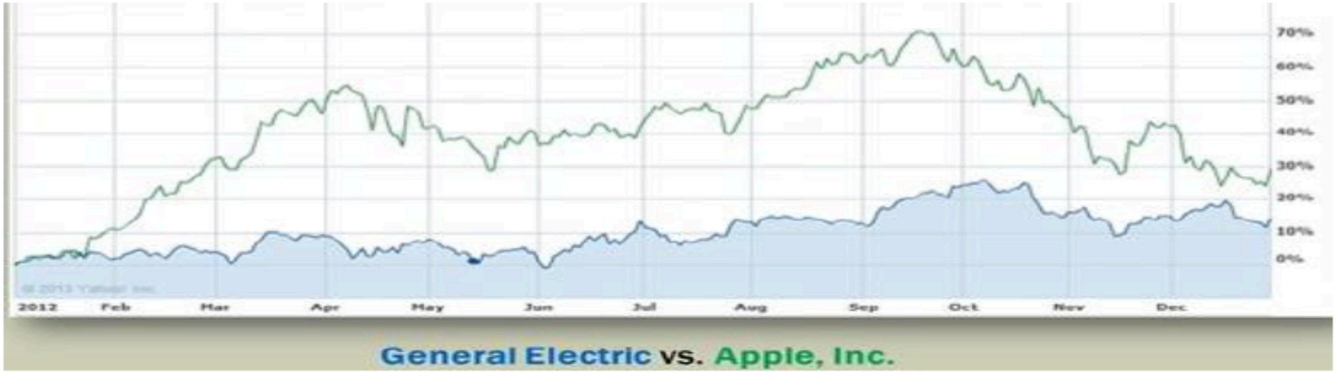
Exemplo: Ao investir em acções, geralmente há um trade-off: risco vs rendimento.

Em termos financeiros, "risco" é sinónimo de variância no valor das acções.

Algumas acções são estáveis (baixo risco), mas oferecem baixos retornos potenciais (GE), outras variam descontroladamente (maior risco), mas oferecem retornos potenciais mais elevados (Apple).

Suponham que compramos acções da GE e da Apple e as mantemos por um ano.

Rendimentos das acções em percentagem



Os dados

Exemplo (continuação):

Date	GE	APPL	GE%	AAPL%
January 2012	0.043869	0.119655	4.39%	11.97%
February 2012	0.026947	0.172533	2.69%	17.25%
March 2012	0.051809	0.100103	5.18%	10.01%
April 2012	-0.02451	-0.026306	-2.45%	-2.63%
May 2012	-0.02567	-0.010762	-2.57%	-1.08%
June 2012	0.096491	0.010796	9.65%	1.08%
July 2012	-0.00444	0.044801	-0.44%	4.48%
August 2012	-0.00198	0.089724	-0.20%	8.97%
September 2012	0.099801	0.002791	9.98%	0.28%
October 2012	-0.07535	-0.113841	-7.53%	-11.38%
November 2012	0.003376	-0.012450	0.34%	-1.25%
December 2012	0.002404	-0.095123	0.24%	-9.51%
Mean	0.016063	0.023493	1.61%	2.35%
Variance	0.002590	0.007330	25.89	73.30
Standard Dev.	0.050890	0.085618	5.09%	8.56%

Mean Monthly Return

GE = 1.61%   AAPL = 2.35%

Monthly Return Variance, $\boxed{s'^2}$

GE = 25.89   AAPL = 73.30

Obter o intervalo de confiança para a variância (grau 95%)

$$\frac{(n-1)\boxed{s'^2}}{\chi^2_{.025}} \le \sigma^2 \le \frac{(n-1)\boxed{s'^2}}{\chi^2_{.975}}$$

n = sample size

$\boxed{s'^2}$ = sample variance

$$\chi^2_{.975} \le \chi^2 \le \chi^2_{.025}$$

$$3.82 \le \chi^2 \le 21.92$$

Critical Values

Monthly Return Variance, s^2

GE = 25.89

$$\frac{(12-1)25.89}{21.92} \le \sigma^2 \le \frac{(12-1)25.89}{3.82}$$

$$12.99 \le \sigma^2 \le 74.55$$

$$3.60\% \le \sigma \le 8.63\%$$

AAPL = 73.30

$$\frac{(12-1)73.30}{21.92} \le \sigma^2 \le \frac{(12-1)73.30}{3.82}$$

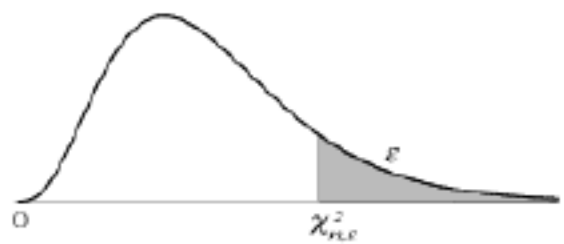
$$36.78 \le \sigma^2 \le 211.07$$

$$6.06\% \le \sigma \le 14.53\%$$

Valores q1, q2 da Tabela 6

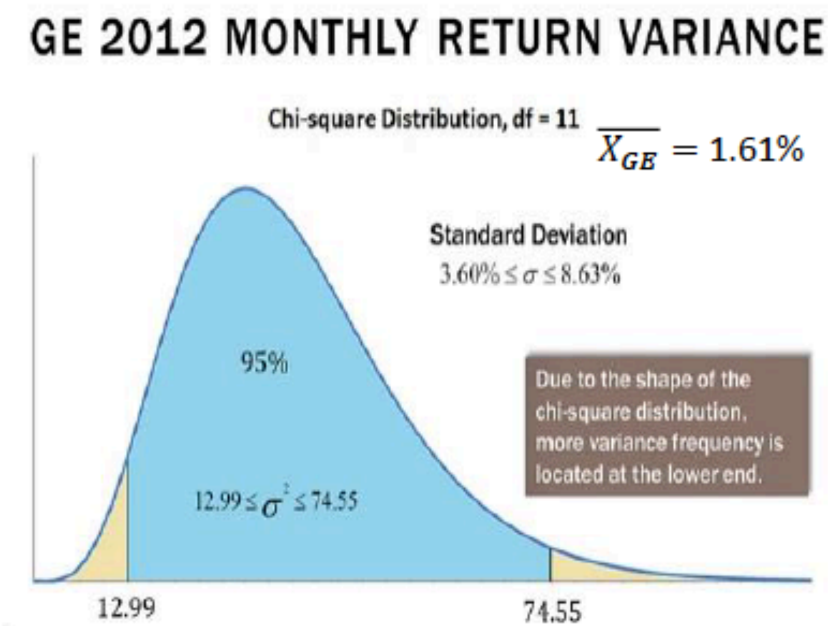
TABELA 6 – DISTRIBUIÇÃO DO QUI-QUADRADO

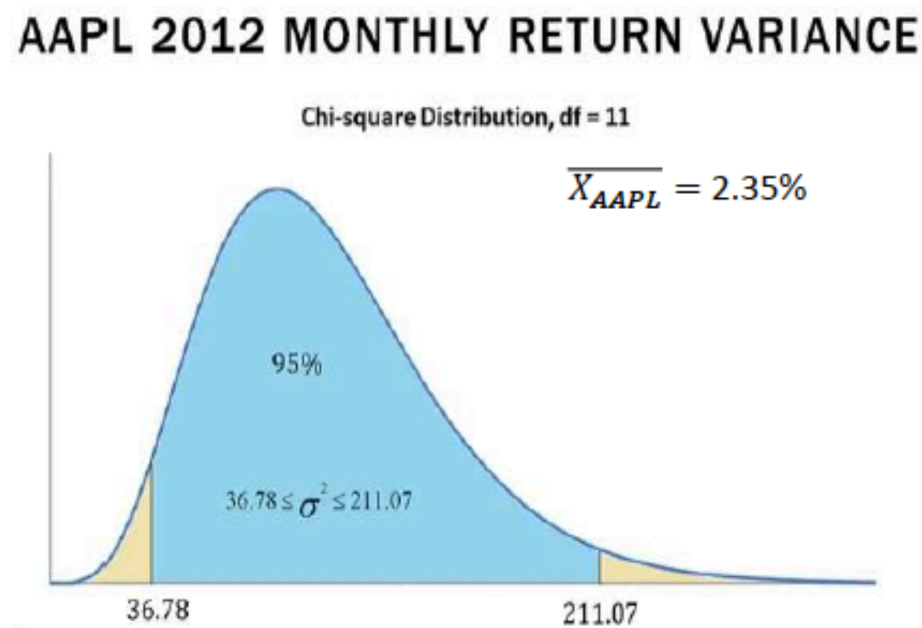
$\chi^2_{n,\varepsilon} : P(X > \chi^2_{n,\varepsilon}) = \varepsilon$



ε / n	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757	31.264
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300	32.909

Para a **variância**: O uso da distribuição Qui-quadrado (**Chi-square)**





Amplitude do Intervalo = 61.56

Amplitude do Intervalo = 174.29

A amplitude do intervalo de confiança aumenta quando aumenta a variação na população

Exemplo 4:
IC para a diferença de medias

Intervalo de confiança para a diferença de médias – Populações normais

Variâncias conhecidas e iguais

População X - $X \sim N(\mu_X, \sigma_X^2)$

População Y - $Y \sim N(\mu_Y, \sigma_Y^2)$

$\sigma_X^2 = \sigma_Y^2$

$\overbrace{(X_1, X_2, \cdots, X_m)}^{amostra\ da\ população\ X}$ → $\bar{X}$

$\overbrace{(Y_1, Y_2, \cdots, Y_n)}^{amostra\ da\ população\ Y}$ → $\bar{Y}$

Variável fulcral: $T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim N(0,1)$

Intervalo Confiança para $\mu_X - \mu_Y$ a $100 * (1-\alpha)\%$:

$$\left((\bar{x} - \bar{y}) - z_{\alpha/2} * \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}, (\bar{x} - \bar{y}) + z_{\alpha/2} * \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \right)$$

IC para a diferença de medias (variâncias **desconhecidas**)

Intervalo de confiança para a diferença de médias – Populações normais

Variâncias desconhecidas

População X - $X \sim N(\mu_X, \sigma_X^2)$ População Y - $Y \sim N(\mu_Y, \sigma_Y^2)$

$$\underbrace{\sigma_X^2 = \sigma_Y^2}_{supõe-se}$$

$$\overbrace{(X_1, X_2, \cdots, X_m)}^{amostra\ casual\ da\ população\ X} \qquad \overbrace{(Y_1, Y_2, \cdots, Y_n)}^{amostra\ casual\ da\ população\ Y}$$

$\bar{X}$ e $S_X'^2$ $\bar{Y}$ e $S_Y'^2$

Variável fulcral: $T = \dfrac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\left(\frac{1}{m}+\frac{1}{n}\right)\sqrt{\frac{(m-1)S_X'^2 + (n-1)S_Y'^2}{m+n-2}}} \sim t(m+n-2)$

Intervalo Confiança para $\mu_X - \mu_Y$ a $100 * (1-\alpha)\%$:

$$\left((\bar{x} - \bar{y}) - t_{\alpha/2} * \left(\frac{1}{m} + \frac{1}{n}\right) \sqrt{\frac{(m-1)s_X'^2 + (n-1)s_Y'^2}{m+n-2}}, (\bar{x} - \bar{y}) - t_{\alpha/2} * \left(\frac{1}{m} + \frac{1}{n}\right) \sqrt{\frac{(m-1)s_X'^2 + (n-1)s_Y'^2}{m+n-2}} \right)$$

Ver exemplo 7.34 no livro (slide 46-47 na Estimação)

Conclusão

Estimação por intervalos (Pop. Normais)

- IC para a media (variância conhecida)→Distribuição normal Z ~N(0,1)
- IC para a media (variância desconhecida)→Distribuição t-student
- IC para a variância → Distribuição Qui-quadrado
- IC para a diferença de medias (variâncias conhecidas e iguais)→ Distribuição normal Z ~N(0,1)
- IC para a diferença de medias (variâncias desconhecidas e iguais)→Distribuição t-student

Material de estudo: SLIDES 34-47 na *Estimação.pdf*