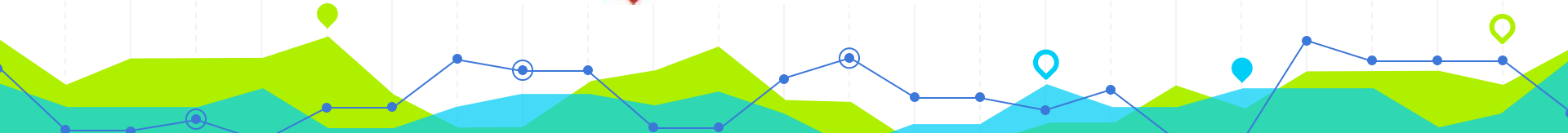




Lisbon School
of Economics
& Management

Universidade de Lisboa



Estatística II

Licenciatura em Gestão do Desporto
2.º Ano/2.º Semestre
2025/2026

Aulas Teórico-Práticas N.ºs 14 e 15 (Semana 8)

Docente: Elisabete Fernandes

E-mail: efernandes@iseg.ulisboa.pt



<https://doity.com.br/estatistica-aplicada-a-nutricao>



<https://basiccode.com.br/produto/informatica-basica/>

Conteúdos Programáticos

Aulas Teórico-Práticas (Semanas 1 a 3)

- **Capítulo 1:** Revisões e Distribuições de Amostragem

Aulas Teórico-Práticas (Semanas 4 a 7)

- **Capítulo 2:** Estimação

Aulas Teórico-Práticas (Semanas 7 a 9)

- **Capítulo 3:** Testes de Hipóteses

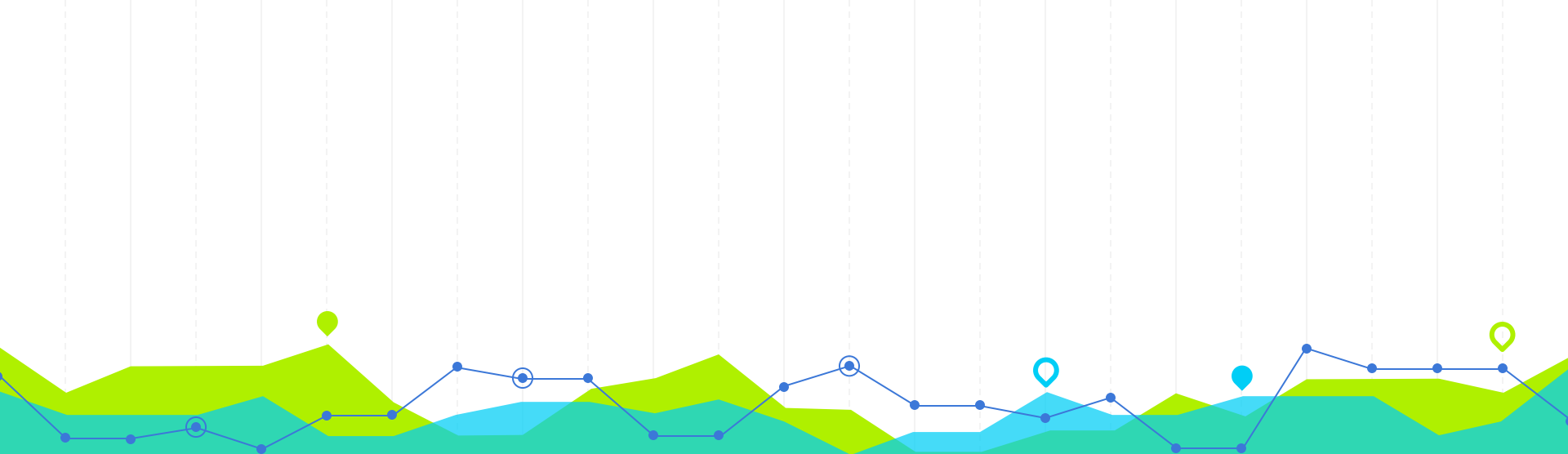
Aulas Teórico-Práticas (Semanas 10 a 13)

- **Capítulo 4:** Modelo de Regressão Linear Múltipla

Material didático: Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

Bibliografia: B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

<https://cas.iseg.ulisboa.pt>



Intervalo de Confiança para a Variância σ^2 : Exercícios

1

7.2 Admita que a densidade de construção, X , num projecto de urbanização tem distribuição normal. Uma amostra aleatória de 50 lotes desse projecto conduziu a

$$\sum_{i=1}^{50} x_i = 227.2; \sum_{i=1}^{50} x_i^2 = 2242.6$$

~~Assumindo que o desvio padrão de X é igual a 4, construa um intervalo de confiança a 95% para a densidade média de construção. Que dimensão deveria ter a amostra para que a amplitude desse intervalo fosse reduzida a metade?~~

Exercício 7.2: Admitindo que se pretende construir um intervalo de confiança a 90% p/ σ^2 .



Exercício 7.2: IC para σ^2

Passo 0 : $X \sim \mathcal{N}(\mu, \sigma^2)$ $\mu, \sigma^2 = ?$
 $I - \alpha_{0.90}(\sigma^2)$

Passo 1 : v. flud
 $\frac{49 s^2}{\sigma^2} \sim \chi^2_{(49)}$

Cálculo dos Quantis da Distribuição Qui-Quadrado de Probabilidade $1-\alpha/2$ e $\alpha/2$ com $n-1$ g.l.'s

Nível de confiança ($1-\alpha=0,90$)

Nível de significância ($\alpha=0,10$)

Área total é igual a 1

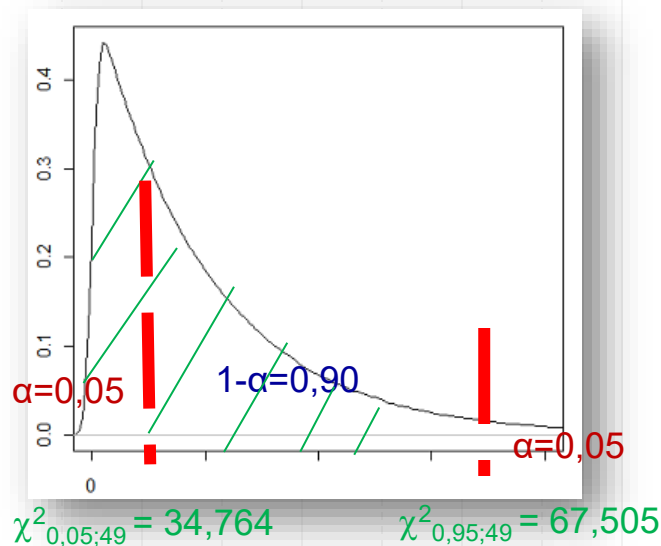
O nível de confiança é $1-\alpha = 0,90 \Leftrightarrow \alpha = 0,10$,
então tem-se $1-\alpha/2 = 0,95$ e $\alpha/2 = 0,05$

Logo, pretende-se calcular o quantil da distribuição
Qui-Quadrado de probabilidade 0,05

$$\chi^2_{0,05;49} = \chi^2_{0,95;49} = 34,764 \text{ (ver tabela a seguir)}$$

Logo, pretende-se calcular o quantil da distribuição
Qui-Quadrado de probabilidade 0,95

$$\chi^2_{0,95;49} = \chi^2_{0,05;49} = 67,505 \text{ (ver tabela a seguir)}$$



Cálculo dos Quantis da Distribuição Qui-Quadrado de Probabilidade $1-\alpha/2$ e $\alpha/2$ com $n-1$ g.l.'s

Tabela do Qui-Quadrado

ϵ	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
n														
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
13	3.565	4.107	5.009	5.892	7.041	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819	34.527
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319	36.124
15	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801	37.698
16	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267	39.252
17	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	30.191	33.409	35.718	40.791
18	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156	42.312
19	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	32.852	36.191	38.582	43.819
20	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997	45.314
21	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	35.479	38.932	41.401	46.796
22	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796	48.268
23	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181	49.728
24	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.588	51.179
25	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928	52.619
26	11.160	12.198	13.844	15.379	17.292	20.843	25.336	30.435	35.563	38.885	41.923	45.642	48.290	54.051
27	11.808	12.878	14.573	16.151	18.114	21.749	26.336	31.528	36.741	40.113	43.195	46.963	49.645	55.475
28	12.461	13.565	15.308	16.928	18.939	22.657	27.336	32.620	37.916	41.337	44.461	48.278	50.994	56.892
29	13.121	14.256	16.047	17.708	19.768	23.567	28.336	33.711	39.087	42.557	45.722	49.588	52.335	58.301
30	13.787	14.953	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	46.979	50.892	53.672	59.702
40	20.707	22.164	24.433	26.560	29.051	33.660	39.335	45.616	51.805	55.758	59.342	63.691	66.766	73.403
50	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	71.420	76.154	79.490	86.660
60	35.534	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	83.298	88.379	91.952	99.608
70	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	95.023	100.425	104.215	112.317
80	51.172	53.540	57.153	60.391	64.278	71.145	79.334	88.130	96.578	101.879	106.629	112.329	116.321	124.839
90	59.196	61.754	65.647	69.126	73.291	80.625	89.334	98.650	107.565	113.145	118.136	124.116	128.299	137.208
100	67.328	70.065	74.222	77.929	82.358	90.133	99.334	109.141	118.498	124.342	129.561	135.807	140.170	149.449

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

$$\chi^2_{0,05;49} = \chi^2_{0,95;49} = 34,764$$

$$\chi^2_{0,95;49} = \chi^2_{0,05;49} = 67,505$$

É a variância amostral corrigida.

NOTA: Saben a variância de amostra não significa se se sabe a variância da população!!

$$S^2 : \frac{1}{50-1} 2242.6 - \frac{50}{50-1} \left(\frac{227.2}{50} \right)^2 = 24.698$$

$$\left[\frac{(n-1) S^2}{\chi^2_{n-1; 1-\frac{\alpha}{2}}}, \frac{(n-1) S^2}{\chi^2_{n-1; \frac{\alpha}{2}}} \right]$$

Limite inferior = $(50-1) \cdot 24,698 / 67,505 = 7,928$
 Limite superior = $(50-1) \cdot 24,698 / 34,764 = 34,811$

$$IC_{90\%}(\sigma^2) : (7,928; 34,811)$$

Exercício 7.2: IC para σ^2

NOTA:

v. flud.: $\frac{49 \cdot 52}{\sigma^2} \sim \chi^2_{(49)}$ ✓

①

v. flud. $\frac{49 \times 24.698}{\sigma^2} \sim \chi^2_{(49)}$ ✗

② Interpretação: $I.C.A._{1-\alpha}(\theta) =]L(x_1, \dots, x_n); U(x_1, \dots, x_n)[$

significa que se recolhamos muitas amostras e calculamos, p/ cada uma delas, o correspondente intervalo então em cerca de $(1-\alpha) 100\%$ os intervalos assim obtidos conterão o verdadeiro valor do parâmetro desconhecido θ .

③

A amplitude do intervalo é tanto menor, em geral, qto maior for a dimensão da amostra.

30. Com base numa amostra casual com 16 observações, retirada de uma população normal, construiu-se, segundo o processo habitual, o seguinte intervalo de confiança para o valor esperado: (7.398, 12.602).
- a) Sabendo que, com a informação da amostra, se obteve $s = 3.872$, qual o grau de confiança que pode atribuir ao intervalo atrás referido?
 - b) Com base na mesma amostra, construa um intervalo de confiança a 95% para a variância da população.
 - c) Suponha que a verdadeira variância é 36. Se pretender construir um intervalo de confiança a 95% para a média, de modo que a respectiva amplitude não exceda 6.5, qual é a dimensão mínima da amostra a recolher?



Nesta alínea não se considerou a variância amostral, mas neste caso é igual como se verifica por esta igualdade:

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2}$$

Exercício 30 b)

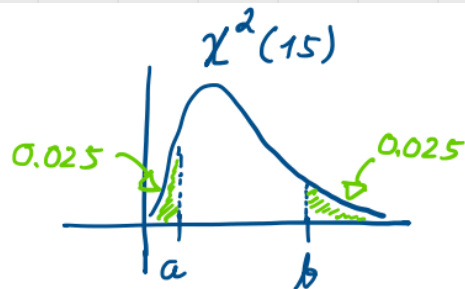
$$b) \quad \frac{nS^2}{\sigma^2} \sim \chi^2(n-1) = \chi^2(15)$$

$$P\left(a < \frac{nS^2}{\sigma^2} < b\right) = 0.95 \quad (\Rightarrow)$$

$$(\Rightarrow) P\left(\frac{nS^2}{a} > \sigma^2 > \frac{nS^2}{b}\right) = 0.95 \quad (\Rightarrow) a = \chi_{15, 0.975}^2 = 6.262$$

$$(\Rightarrow) P\left(\frac{nS^2}{b} < \sigma^2 < \frac{nS^2}{a}\right) = 0.95$$

$$b = \chi_{15, 0.025}^2 = 27.488$$



Exercício 30 b)

$$\left[\frac{(n-1) s^2}{\chi^2_{n-1; 1-\frac{\alpha}{2}}}, \frac{(n-1) s^2}{\chi^2_{n-1; \frac{\alpha}{2}}} \right]$$

$$IC_{95\%}(\sigma^2) = \left(\frac{n s^2}{b}; \frac{n s^2}{a} \right) =$$

$$= \left(\frac{16 \times 3.872^2}{27.488}; \frac{16 \times 3.872^2}{6.262} \right) =$$

$$= (8.72665; 38.30695)$$

Nota: Podemos considerar n *variância não corrigida ou $(n-1)$ *variância corrigida... O resultado é o mesmo...

31. Considere uma população com distribuição normal de parâmetros desconhecidos. Dessa população foi retirada uma amostra casual de dimensão 25. Suponha que a amostra forneceu os seguintes resultados:

$$\sum_{i=1}^{25} x_i = 75 \text{ e } \sum_{i=1}^{25} x_i^2 = 321.$$

- a) Construa um intervalo de confiança a 95% para a média.
- b) Construa um intervalo de confiança a 95% para o desvio padrão.

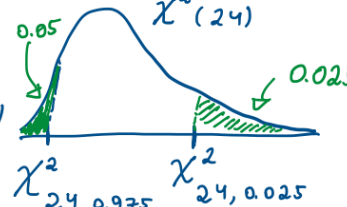


Exercício 31 b)

b) I.C. a 95% para σ^2

$$1 - \alpha = 0.95 \quad \alpha = 0.05 \quad \alpha/2 = 0.025$$

Variaável fucional: $Q = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{(n-1)} = \chi^2_{(24)}$

$$P\left(\chi^2_{24, 0.975} < Q < \chi^2_{24, 0.025} \right) = 0.95 \Leftrightarrow$$


$$\Leftrightarrow P(12.401 < Q < 39.364) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow P\left(12.401 < \frac{(n-1)S'^2}{\sigma^2} < 39.364\right) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{1}{12.401} > \frac{\sigma^2}{(n-1)S'^2} > \frac{1}{39.364}\right) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{(n-1)S'^2}{39.364} < \sigma^2 < \frac{(n-1)S'^2}{12.401}\right) = 0.95$$

Exercício 31 b)

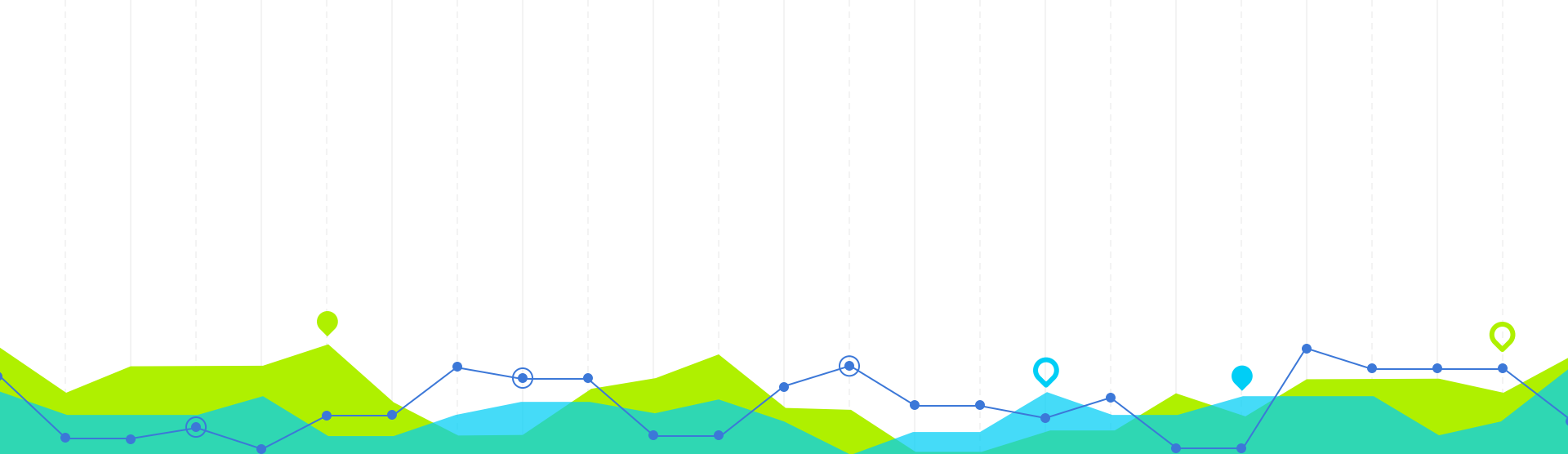
$$I.A. : \left[\frac{24 S'^1}{39.364} ; \frac{24 S'^2}{12.401} \right]$$

$$I.C. : \left[\frac{24 S'^2}{39.364} ; \frac{24 S'^2}{12.401} \right] = (S'^2 = 4)$$

$$= \left[\frac{24 \times 4}{39.364} ; \frac{24 \times 4}{12.401} \right] = [2.439, 7.741]$$

Nota: exercício pede I.C. para σ mas as soluções apresentam o resultado aqui obtido, ou seja, o I.C. para σ^2 .

Se quisermos o IC para σ em vez de σ^2 basta fazer $IC = (\sqrt{2.439}, \sqrt{7.741})$.



Intervalo de Confiança para a Relação de Variâncias σ_1^2/σ_2^2

2

Intervalo de Confiança para σ_1^2/σ_2^2

Quando as populações são Normais, o I. C. para $\frac{\sigma_1^2}{\sigma_2^2}$ com $100(1 - \alpha)\%$ de confiança é dado por:

$$\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}}; \frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1, \frac{\alpha}{2}}} \right] \cdot \frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$$

Variável Fulcral

Como $F_{n_1-1, n_2-1; \frac{\alpha}{2}} = \frac{1}{F_{n_2-1, n_1-1; 1-\frac{\alpha}{2}}}$ este intervalo pode ser escrito da seguinte forma:

$$\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1; 1-\frac{\alpha}{2}}}; \frac{S_1^2}{S_2^2} F_{n_2-1, n_1-1; 1-\frac{\alpha}{2}} \right]$$

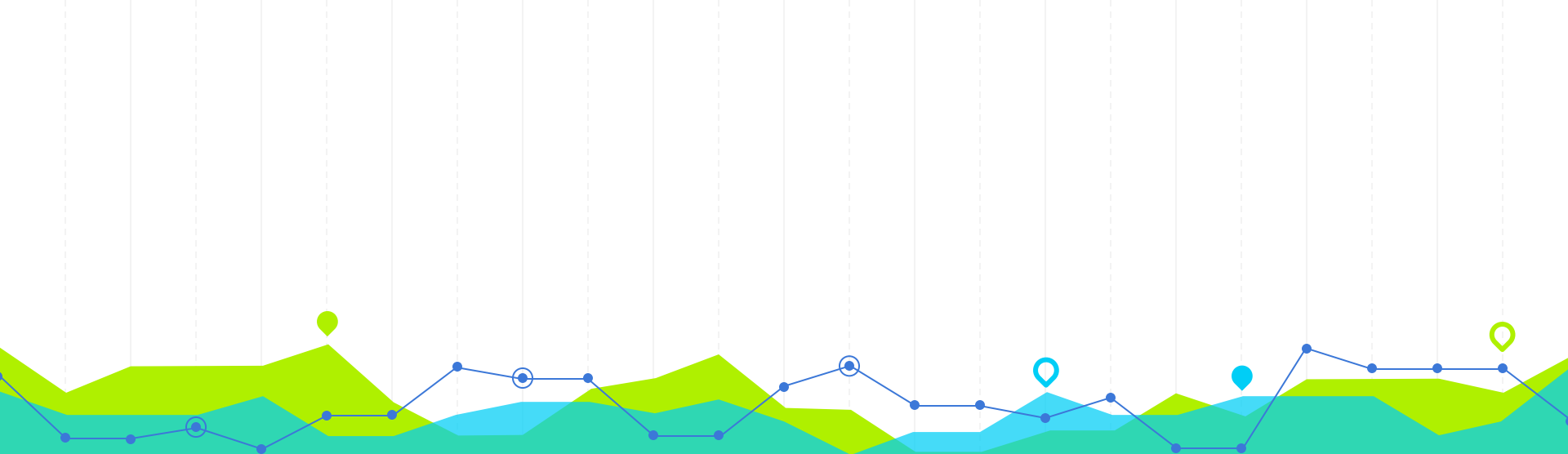
Variâncias corrigidas

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

IC para σ_1^2/σ_2^2 : Formulário

• POPULAÇÕES NORMAIS

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(\nu)$
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	onde ν é o maior inteiro contido em r, $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{S_2'^2}{n}\right)^2}$
Variância	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	Variâncias corrigidas
Relação de variâncias	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$



Intervalo de Confiança para a Relação de Variâncias σ_1^2/σ_2^2 : Exercícios

3

Para substituir uma máquina antiquada encaram-se 2 alternativas: o equipamento A ou equipamento B. Dado tratar-se de uma decisão que envolve custos consideráveis uma vez que os equipamentos são bastante dispendiosos, resolveu-se testar os dois equipamentos durante um período experimental.

No final do período experimental seleccionaram-se 31 e 61 peças da produção dos equipamentos A e B, respetivamente, tendo-se registado os seguintes valores relativamente à característica de interesse na avaliação da qualidade do trabalho das máquinas:

$$\sum_{i=1}^{31} x_{iA} = 43,4; \quad \sum_{i=1}^{31} x_{iA}^2 = 123,76; \quad \sum_{i=1}^{61} x_{iB} = 91,5; \quad \sum_{i=1}^{61} x_{iB}^2 = 269,25$$

Utilizando um intervalo de confiança a 95% diga se há razões para crer que com a máquina A se consegue uma menor variabilidade da característica de avaliação do que com a máquina B. Admita a normalidade das distribuições.



Exercício: IC para σ_1^2/σ_2^2

Sejam:

- X_1 a v.a. que representa o valor da característica de interesse no equipamento A,
- X_2 a v.a. que representa o valor da característica de interesse no equipamento B,

Com $X_1 \sim N(\mu_1 = ?; \sigma_1 = ?)$ e $X_2 \sim N(\mu_2 = ?; \sigma_2 = ?)$.

$n_1 = 31$ e $n_2 = 61$.

Aqui são variâncias amostrais corrigidas:

$$\bar{x}_1 = \sum_{i=1}^{31} \frac{x_{1i}}{31} = \frac{43,4}{31} = 1,4;$$

$$s_1^2 = \frac{1}{30} \left(\sum_{i=1}^{31} x_{1i}^2 - 31 \times \bar{x}_1^2 \right) = \frac{123,76 - 31 \times 1,4^2}{30} = 2,1;$$

$$\bar{x}_2 = \sum_{i=1}^{61} \frac{x_{2i}}{61} = \frac{91,5}{61} = 1,5;$$

$$s_2^2 = \frac{1}{60} \left(\sum_{i=1}^{61} x_{2i}^2 - 61 \times \bar{x}_2^2 \right) = \frac{269,25 - 61 \times 1,5^2}{60} = 2,2.$$

O I. C. a 95% para $\frac{\sigma_1^2}{\sigma_2^2}$ é dado por:

$$\left[\frac{S_1^2}{S_2^2} \frac{1}{f_{n_1-1; n_2-1; 1-\frac{\alpha}{2}}}, \frac{S_1^2}{S_2^2} f_{n_2-1; n_1-1; 1-\frac{\alpha}{2}} \right].$$

$$f_{0,025;30,60} = f^*_{0,975;30,60} = 1 / f^*_{0,025;60,30} = 1 / 1,94 = 0,5154$$

$$f_{0,975;30,60} = f^*_{0,025;30,60} = 1,82$$

Exercício: IC para σ_1^2/σ_2^2

Tabela da F-Snedcor

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

		m – graus de liberdade do numerador																			
	ε	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
n – graus de liberdade do denominador	20	.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
		.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.87
		.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.10
		.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.43
	22	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
		.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.79
		.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
		.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
	24	.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.54
		.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
		.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
		.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.22
	26	.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
		.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
		.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
		.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.14
	28	.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
		.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
		.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
		.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
30	.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.63	
	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.00	
40	.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	
	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.53	
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	
60	.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	
	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.49	
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.61	

Exercício: IC para σ_1^2/σ_2^2

$$P(0,5154 < s_1'^2/s_2'^2 * \sigma_2^2/\sigma_1^2 < 1,82) = 0,95$$

A probabilidade da variável
fulcral “estar entre” os quantis
0,5154 e 1,82 é igual a 0,95.
Posteriormente, desenvolve-
se essas desigualdades e
obtem-se os limites:

Aqui são variâncias amostrais corrigidas:

$$\text{Limite inferior} = (2,1/2,2) * (1/1,82) = 0,526$$

$$\text{Limite superior} = (2,1/2,2) * (1/0,5154) = 1,852$$

$$\text{IC}_{95\%}(\sigma_1^2/\sigma_2^2) = (0,526; 1,852)$$

Com 95% confiança não há razões para crer que exista diferença na variabilidade da característica de avaliação obtida com as duas máquinas (a igualdade das variâncias), uma vez que o valor 1 está presente no intervalo.

37. Para comparar dois métodos de ensino, dividiu-se aleatoriamente uma turma de 22 alunos em dois grupos iguais. Cada grupo foi ensinado com um método diferente (método A e método B) e, no fim da formação, foram sujeitos à mesma prova de avaliação. Os resultados da avaliação, numa escala de 0 a 100, foram os seguintes:

$$\text{Método } A: \bar{x}_A = 74.8; s_A'^2 = 81.5.$$

$$\text{Método } B: \bar{x}_B = 72.1; s_B'^2 = 110.5.$$

Admita que as classificações obtidas em cada grupo têm distribuição normal.

- Construa um intervalo de confiança a 95% para o quociente das variâncias dos dois grupos.
- Tendo em conta o resultado da alínea anterior, construa um intervalo de confiança a 99% para a diferença de médias. Interprete o resultado obtido.



Exercício 37 a)

$X_A \equiv$ Classificação de um aluno (0 a 100) ensinado pelo método A.

$X_B \equiv$ " " " " " " " " " " B.

$$n_A = n_B = 11$$

$$X_A \sim N(\mu_A, \sigma_A^2) \quad \bar{x}_A = 74.8 \quad s_A'^2 = 81.5$$

$$X_B \sim N(\mu_B, \sigma_B^2) \quad \bar{x}_B = 72.1 \quad s_B'^2 = 110.5$$

Exercício 37 a)

a) I.C. 95% para $\frac{\sigma_B^2}{\sigma_A^2}$ $\alpha = 0.05$ $\frac{\alpha}{2} = 0.025$

Variável aleatória: $\frac{S_A'^2}{S_B'^2} \cdot \frac{\sigma_B^2}{\sigma_A^2} \sim F(m_A-1, m_B-1)$
 $F \sim F(10, 10)$

$$P(F_{10,10,0.975} < F < F_{10,10,0.025}) = 0.95 \Leftrightarrow$$

$$P\left(F_{10,10,0.975} < \frac{S_A'^2}{S_B'^2} \cdot \frac{\sigma_B^2}{\sigma_A^2} < F_{10,10,0.025}\right) = 0.95$$

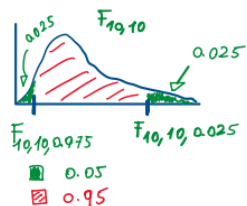


Tabela 8: $F_{10,10,0.025} = 3.72$

$$P(F < F_{10,10,0.975}) = 0.025 \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{1}{F} > \frac{1}{F_{10,10,0.975}}\right) = 0.025 \Leftrightarrow$$

$$\Leftrightarrow P\left(F > \frac{1}{F_{10,10,0.975}}\right) = 0.025$$

$\hookrightarrow F_{10,10,0.025} = 3.72$

Porque $F \sim (10, 10)$ e $\frac{1}{F} \sim F(10, 10)$

$$\log_{10} F_{10,10,0.975} = \frac{1}{F_{10,10,0.025}} = \frac{1}{3.72}$$

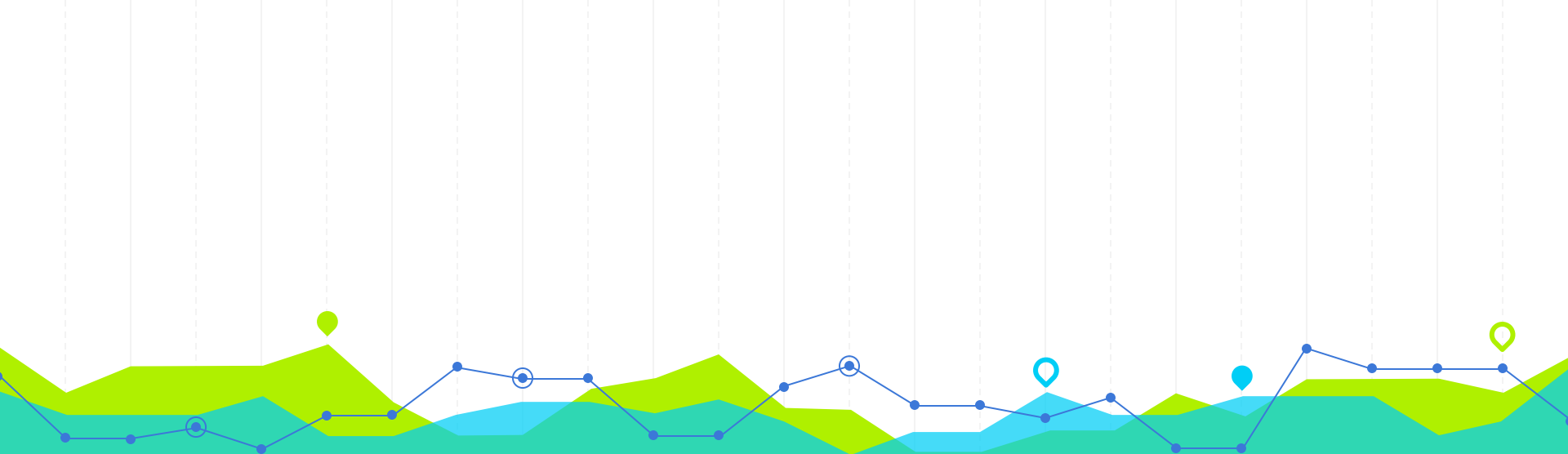
$$\Leftrightarrow P\left(\frac{1}{3.72} < \frac{S_A'^2}{S_B'^2} \cdot \frac{\sigma_B^2}{\sigma_A^2} < 3.72\right) = 0.95 \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{1}{3.72} \frac{S_B'^2}{S_A'^2} < \frac{\sigma_B^2}{\sigma_A^2} < 3.72 \frac{S_B'^2}{S_A'^2}\right) = 0.95$$

Exercício 37 a)

$$\begin{aligned} \text{I.C.}_{95\%} \text{ para } \frac{\sigma_B^2}{\sigma_A^2} &= \left(\frac{1}{3.72} \frac{s_B^2}{s_A^2}; 3.72 \frac{s_B^2}{s_A^2} \right) = \\ &= \left(\frac{1}{3.72} \times \frac{110.5}{81.5}; 3.72 \times \frac{110.5}{81.5} \right) = \\ &= (0.36; 5.04) \end{aligned}$$

Conclusão: Se $\sigma_A^2 = \sigma_B^2$ então $\sigma_B^2 / \sigma_A^2 = 1$. Um possível teste para a hipótese de igualdade de variâncias consiste em construir um I.C. para σ_B^2 / σ_A^2 e verificar se "1" pertence ao intervalo. Como $\sigma_A^2 / \sigma_B^2 = 1 \in \text{I.C.}_{95\%}(\sigma_A^2 / \sigma_B^2) = (0.36, 5.04)$ não rejeitamos a hipótese de igualdade de variâncias.



Intervalo de Confiança para uma proporção p

4

Intervalo de Confiança para p

Estimador de p

$$\bar{P} = \hat{P}$$

Variável Fulcral

$$Z = \frac{\bar{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0; 1).$$

Portanto, quando a amostra é grande, o I. C. para p com $100(1 - \alpha)\%$ de confiança é:

$$\left[\bar{P} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}; \bar{P} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{P}(1-\bar{P})}{n}} \right].$$

[ProbabilidadesEstatistica_2019 \(uevora.pt\)](#)

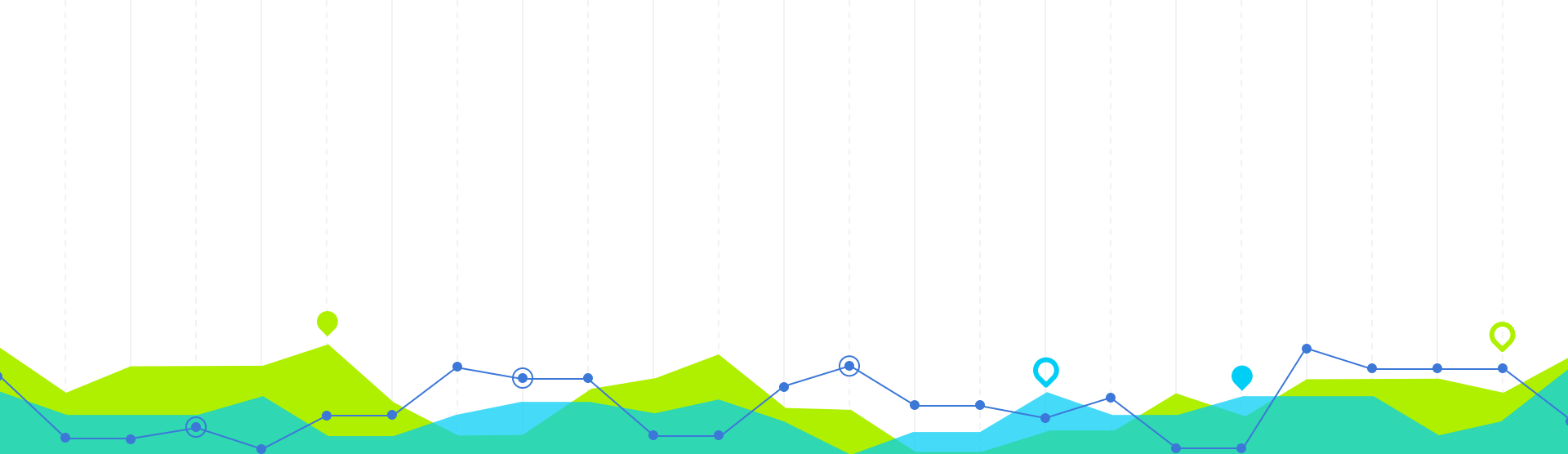
IC para p: Formulário

Considerar em Testes de Hipóteses

Considerar em Intervalos de Confiança

População de Bernoulli

Proporção	$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$
Diferença de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$
Igualdade de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\hat{\theta}(1-\hat{\theta})}} \stackrel{a}{\sim} N(0,1) \text{ onde } \hat{\theta} = \frac{m\bar{X}_1 + n\bar{X}_2}{m+n}$	



Intervalo de Confiança para uma proporção p : Exercícios

5

Numa certa cidade A recolheu-se uma amostra aleatória de 150 homens tendo 54 afirmado que viam o telejornal todos os dias.

- a) Com 90% de confiança, será que se pode considerar que a proporção de homens, daquela cidade, que veem o telejornal todos os dias é de 40%.
- b) Mantendo-se o resto constante, qual deveria ser a dimensão da amostra de forma a que o erro de estimativa do intervalo de confiança não ultrapasse 5%?



Exercício (a): IC para p

Sejam:

- X_i a v. a. que designa se o i -ésimo homem afirmou ver o telejornal,
- $\bar{P}$ a v. a. que representa a proporção de homens que afirmaram ver o telejornal, em n homens.

$$n = 150 \text{ e } \bar{p} = \frac{54}{150} = 0,36.$$

a) Afirmação: $p = 0,4$.

I. C. a 90% para p é dado por:

$$\left[\bar{p} - z_{0,95} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}; \bar{p} + z_{0,95} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \right].$$

Substituindo pelos valores conhecidos, com $z_{0,95} = 1,645$, obtém-se

$$\left[0,36 - 1,645 \sqrt{\frac{0,36(1 - 0,36)}{150}}; 0,36 + 1,645 \sqrt{\frac{0,36(1 - 0,36)}{150}} \right] =]0,3464; 0,5077[.$$

Deste modo, face aos resultados obtidos (0,4 está contido do I. C. a 90%) não é de rejeitar a hipótese de que a proporção de homens que vê o telejornal todos os dias é de 40%, pois com 90% de confiança a percentagem de homens que vê diariamente o telejornal situa-se entre 34,64% e 50,77%.

Exercício (b): IC para p

b) Erro de estimativa $\leq 0,05$, então $n = ?$

O erro de estimativa associado ao I. C. a 90% para p é:

Margem de erro ou Erro de estimativa é metade da amplitude do IC

$$z_{0,95} \sqrt{\frac{\bar{P}(1 - \bar{P})}{n}},$$

que se pretende que seja inferior ou igual a 0,5. Portanto,

$$z_{0,95} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \leq 0,05 \Leftrightarrow 1,645 \sqrt{\frac{0,36(1 - 0,36)}{n}} \leq 0,05$$

$$\Leftrightarrow \sqrt{n} \geq \frac{1,645 \sqrt{0,36(1 - 0,36)}}{0,05} \Leftrightarrow \sqrt{n} \geq 15,792$$

$$\Rightarrow n \geq 15,792^2 \Leftrightarrow n \geq 249,4 \Rightarrow n \geq 250$$

Desta forma, a dimensão mínima da amostra que garante que o erro de estimativa do I. C. a 90% é no máximo de 5% é de 250 homens.

40. Uma cadeia de televisão fixou como objectivo ter uma audiência, para a telenovela XYZ, de pelo menos 55% dos telespectadores. Consultada uma empresa de audio-

metria, ficou estabelecido que semanalmente a empresa interrogaria por telefone, dentro do horário da novela, uma amostra casual de 200 telespectadores e registaria quantos estavam a ver a referida novela. Em determinada semana obtiveram-se 90 respostas positivas. Construa um intervalo de confiança a 95% para a percentagem de telespectadores da telenovela XYZ nessa semana. Com base neste intervalo, diga o que pode concluir sobre o cumprimento do objectivo fixado.



Exercício 40

Pop. de Bernoulli: $X \sim \text{Ber}(\theta)$

$$X = \begin{cases} 1 & (\text{se } \theta \text{ telespectador assiste à novela}) \\ 0 & (\text{caso contrário}) \end{cases}$$

$$X \sim \text{Ber}(\theta) \quad \text{objetivo: } \theta \geq 0.55$$

Amostra: $n = 200$ telespectadores

$$\sum_{i=1}^{200} x_i = 90 \quad \bar{x} = \frac{90}{200} = 0.45$$

$$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \approx N(0,1) \quad z_{0.025} = 1.96$$

$$P\left(-z_{0.025} < \frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} < z_{0.025}\right) = 0.95$$

Exercício 40

$$\begin{aligned} IC_{95\%}(\theta) &= \bar{x} \pm z_{0.025} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = \\ &= 0.45 \pm 1.96 \sqrt{\frac{0.45(1-0.45)}{200}} = \\ &= (0.381, 0.519) \end{aligned}$$

$$\theta = 0.55 \notin IC_{95\%}(\theta) \text{ e } \forall \theta \in IC_{95\%}(\theta) \text{ temos } \theta < 0.55.$$

O intervalo de confiança obtido sugere que o objetivo fixado não foi cumprido.

44. Pretende saber-se qual a proporção de consumidores do produto A . Para tal foram recolhidas amostras de 200 pessoas nas cidades F e G , tendo-se observado, respectivamente, 140 e 160 consumidores desse produto.
- a) Determine, com um grau de confiança de 95%, a percentagem de consumidores do produto A na cidade F .
 - b) Pode afirmar-se, com um grau de confiança de 99%, que o consumo do produto A é superior na cidade F ?



Exercício 44 a)

Dois Populações de Bernoulli independentes

$$X_F \sim B(1, \theta_F) \quad X_G \sim B(1, \theta_G)$$

θ_F = Proporção de consumidores do produto A na cidade F.

θ_G = Proporção de consumidores do produto A na cidade G.

$$n_F = n_G = 200 \quad (\text{amostras grandes})$$

$$\sum_{i=1}^{200} x_{Fi} = 140$$

$$\sum_{i=1}^{200} x_{Gi} = 160$$

$$\bar{x}_F = \frac{140}{200} = 0.7$$

$$\bar{x}_G = \frac{160}{200} = 0.8$$

Exercício 44 a)

a) I.C. para θ_F com $1 - \alpha = 0.95$
 $\alpha = 0.05 \quad \frac{\alpha}{2} = 0.025$

Variável fatorial: $Z = \frac{\bar{X}_F - \theta_F}{\sqrt{\frac{\bar{X}_F(1-\bar{X}_F)}{n_F}}} \approx N(0,1)$

$$P(-z_{0.025} < Z < z_{0.025}) = 0.95 (=)$$

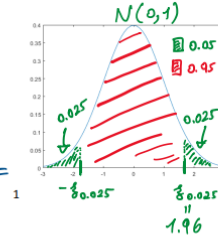
$$= P\left(-z_{0.025} < \frac{\bar{X}_F - \theta_F}{\sqrt{\frac{\bar{X}_F(1-\bar{X}_F)}{n_F}}} < z_{0.025}\right) =$$

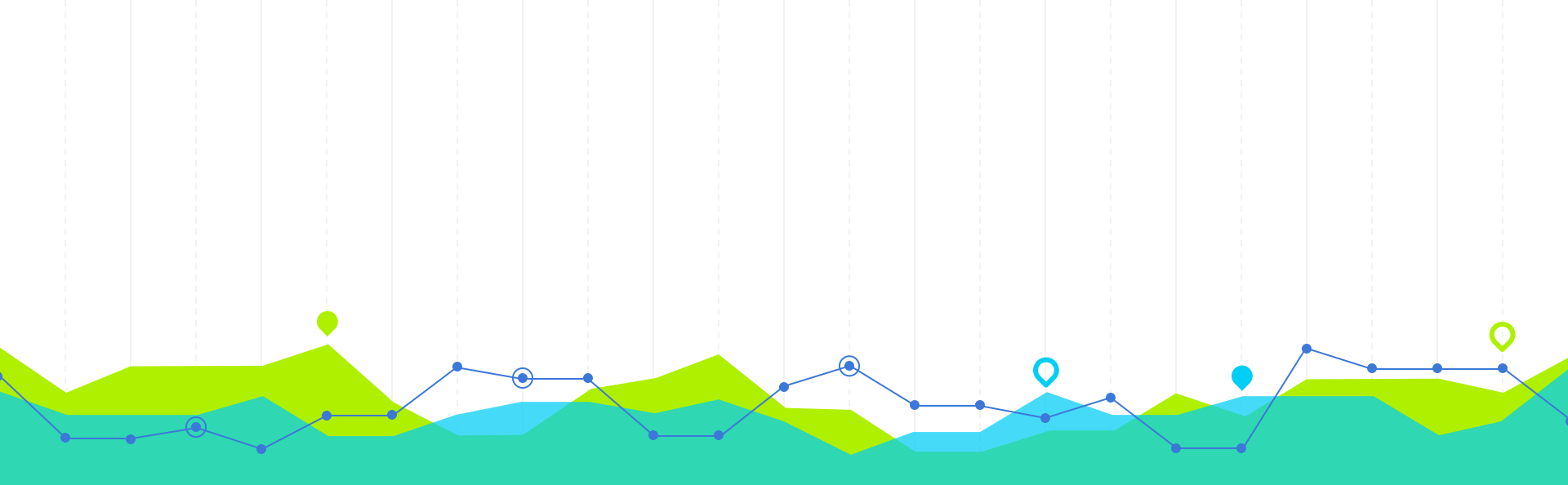
$$= P\left(\bar{X}_F - z_{0.025} \sqrt{\frac{\bar{X}_F(1-\bar{X}_F)}{n_F}} < \theta_F < \bar{X}_F + z_{0.025} \sqrt{\frac{\bar{X}_F(1-\bar{X}_F)}{n_F}}\right) = 0.95$$

I.A. 95% $\bar{X}_F \pm z_{0.025} \sqrt{\frac{\bar{X}_F(1-\bar{X}_F)}{n_F}}$

I.C. 95% : $\bar{x}_F \pm z_{0.025} \cdot \sqrt{\frac{\bar{x}_F(1-\bar{x}_F)}{n_F}} =$
Tabela 5
 $= 0.7 \pm 1.96 \sqrt{\frac{0.7(1-0.7)}{200}} =$
 $= (0.6365, 0.7635)$

ou I.C. 95% = (63.65%, 76.35%)





Intervalo de Confiança para a Diferença de Proporções $p_1 - p_2$

6

Intervalo de Confiança para $p_1 - p_2$

Estimadores de p_1 e p_2

$$\overline{P_1} = \hat{P}_1$$

$$\overline{P_2} = \hat{P}_2$$

Variável Fulcral

$$Z = \frac{(\overline{P_1} - \overline{P_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \stackrel{\circ}{\sim} N(0; 1).$$

Portanto, quando **as amostras são grandes**, o **I. C. para $p_1 - p_2$ com $100(1 - \alpha)\%$ de confiança** é dado por:

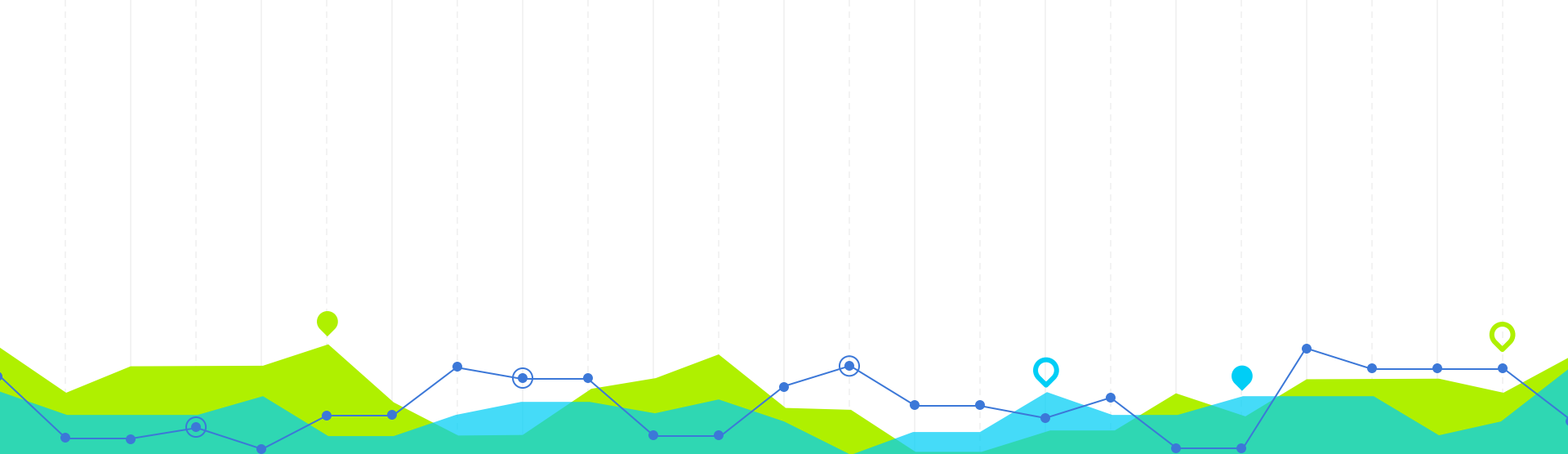
$$\left[\overline{P_1} - \overline{P_2} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{P_1}(1-\overline{P_1})}{n_1} + \frac{\overline{P_2}(1-\overline{P_2})}{n_2}}; \overline{P_1} - \overline{P_2} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{P_1}(1-\overline{P_1})}{n_1} + \frac{\overline{P_2}(1-\overline{P_2})}{n_2}} \right].$$

ProbabilidadesEstatistica_2019 (uevora.pt)

IC para p_1-p_2 : Formulário

População de Bernoulli

Proporção	$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$
Diferença de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$
Igualdade de proporções	Considerar em Testes de Hipóteses $\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\theta(1-\theta)}$	Considerar em Intervalos de Confiança $\hat{\theta} = \frac{m\bar{X}_1}{m+n}$



Intervalo de Confiança para a Diferença de Proporções $p_1 - p_2$: Exercícios

7

Exercício Suplementar que não consta
do livro Murteira et al (2015)

Numa certa cidade A recolheu-se uma amostra aleatória de 150 homens tendo 54 afirmado que viam o telejornal todos os dias. Numa outra cidade do país, cidade B, 80 dos 200 homens seleccionados aleatoriamente responderam afirmativamente.

Com 95% de confiança, será de admitir que a proporção de homens que vê o telejornal todos os dias é igual nas duas cidades?

ProbabilidadesEstatistica_2019 (uevora.pt)



44. Pretende saber-se qual a proporção de consumidores do produto A . Para tal foram recolhidas amostras de 200 pessoas nas cidades F e G , tendo-se observado, respectivamente, 140 e 160 consumidores desse produto.
- a) Determine, com um grau de confiança de 95%, a percentagem de consumidores do produto A na cidade F .
 - b) Pode afirmar-se, com um grau de confiança de 99%, que o consumo do produto A é superior na cidade F ?

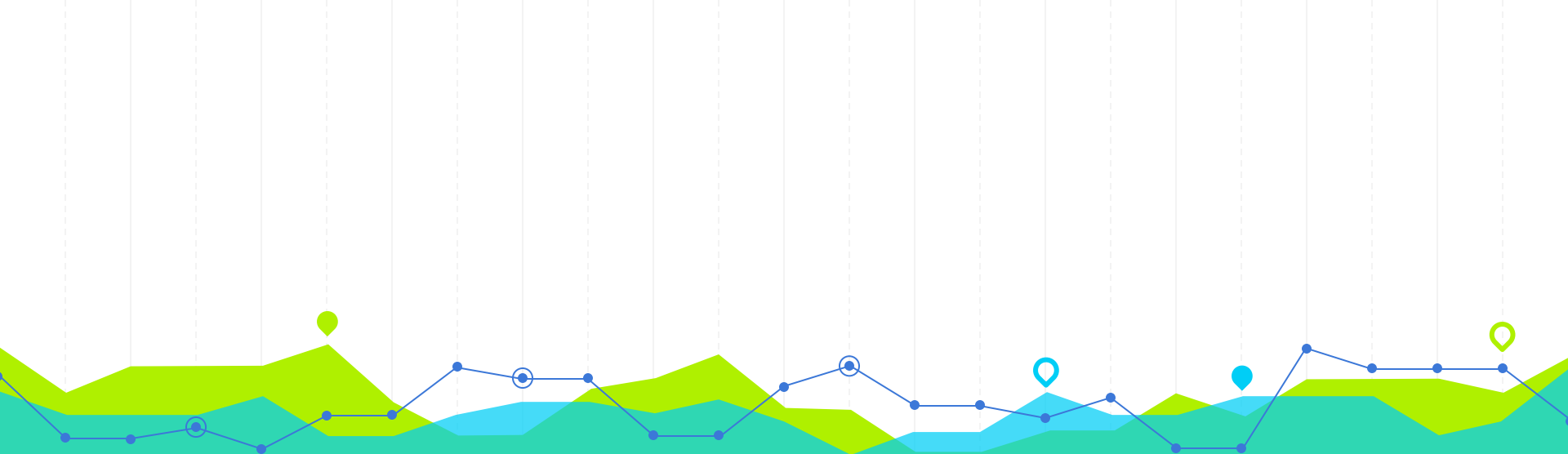


49. Com o objectivo de avaliar o efeito da dimensão da turma sobre o aproveitamento dos alunos recolheu-se a seguinte informação (amostras casuais de alunos da mesma disciplina):

Classificação (0-20)	< 8	[8, 10)	[10, 12)	[12, 14)	≥ 14	Total
Turmas grandes	12	42	30	11	5	100
Turmas pequenas	4	30	40	21	5	100

- Construa um intervalo de confiança a 90% para a nota média dos alunos de turmas pequenas.
- Construa um intervalo de confiança a 90% para a diferença na classificação média das duas populações e comente o resultado obtido.
- Com base num intervalo de confiança a 90%, diga se pode afirmar que a proporção de alunos com nota positiva é maior nas turmas pequenas.





Inferência Estatística

Definição e Construção de Testes de Hipóteses

8

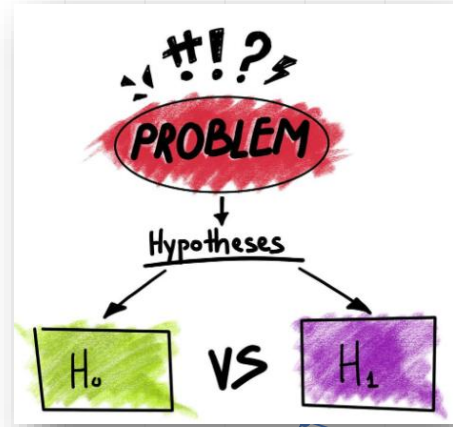
Teste de Hipóteses

- ✓ Procedimento que conduz a uma tomada de decisão, com base na informação fornecida pelos dados recolhidos, sobre a aceitação ou a não aceitação de determinada **hipótese estatística** que se coloca sobre uma ou mais populações.

“A **Statistical Test** is a way to evaluate the evidence the data provides against a hypothesis. This hypothesis is called **the null hypothesis (H_0)**. H_0 is usually opposed to a hypothesis called the **alternative hypothesis (H_1)**.

If the data does not provide enough evidence against H_0 , H_0 is not rejected. If instead, the data shows strong evidence against H_0 , H_0 is rejected and H_1 is considered as true with a quantified (low) risk of being wrong.

A **Statistical Test** allows to reject/not to reject H_0 .”



Hipótese é uma afirmação

Em uma certa população as médias de tempo de recuperação dos indivíduos que tomam um certo remédio e daqueles que não tomam são iguais

O teste de hipótese é uma pergunta

Será que em uma certa população as médias de tempo de recuperação dos indivíduos que tomam um certo remédio e daqueles que não tomam são iguais?

O resultado do teste é uma resposta

Com base na amostra pode-se dizer que

- Não há indícios de diferença estatisticamente significativa no tempo médio de recuperação entre os que tomam e os que não tomam o remédio, ou
- Há indícios de diferença estatisticamente significativa no tempo médio de recuperação entre os que tomam e os que não tomam o remédio

Qual o Teste Estatístico a Usar?

Depende de vários factores:

- Tipo de efeito que se pretende testar
- N° de variáveis envolvidas
- Nível de medição das variáveis
- Independência das observações
- Outras características dos dados (tipo de distribuição de frequências, igualdade de variâncias, etc)



Testes Paramétricos vs Testes Não-Paramétricos

Testes Paramétricos:

- **Fazem pressupostos sobre a distribuição** dos dados:
 - Exigem a normalidade em amostras pequenas ($n < 30$).
 - Nas amostras grandes ($n \geq 30$) como a distribuição da variável média amostral é assintoticamente normal pelo **Teorema do Limite central (TLC)** podem ser aplicados os Testes Paramétricos.
- Dados contínuos.

Testes Não-Paramétricos:

- **Não fazem pressupostos sobre a distribuição** dos dados:
 - São usados quando a(s) variável(is) não tem(têm) distribuição normal ou quando, apesar da(s) amostra(s) ser(em) grande(s), se opta por conclusões mais conservadoras.
- Dados nominais, ordinais, discretos ou contínuos.
- Em geral, envolvem cálculos mais simples.
- Pouco influenciados por valores extremos.
- A desvantagem destes testes é que não são tão potentes quanto os Testes Paramétricos, ou seja, podem levar a uma maior probabilidade do erro tipo II (definido num slide mais à frente).

Paramétrico

Distribuição da variável na população conhecida (ex: Normal)

Estudos anteriores com amostras grandes revelam normalidade

Amostras grandes
 $n \geq 30$

Comparo médias e/ou variâncias

Mais poderoso (*)

Não Paramétrico

Não conheço a distribuição da variável na população

Amostras pequenas

Comparo medianas e/ou distribuições

Utiliza postos (ranks)

Menos poderoso (*)

*) Poder: Habilidade do teste de detectar um efeito dado que ele realmente exista

Testes de Hipóteses em Estudo...

Testes Paramétricos

- Testes de hipóteses para um valor médio μ
- Testes de hipóteses para uma variância σ^2
- Testes de hipóteses para uma proporção populacional p
- Testes de Hipóteses para a igualdade de médias (amostras independentes)
- Testes de Hipóteses para a razão de variâncias
- Testes de Hipóteses para a igualdade de proporções
- Testes de Hipóteses para a igualdade de médias (amostras emparelhadas)

Testes Não Paramétricos

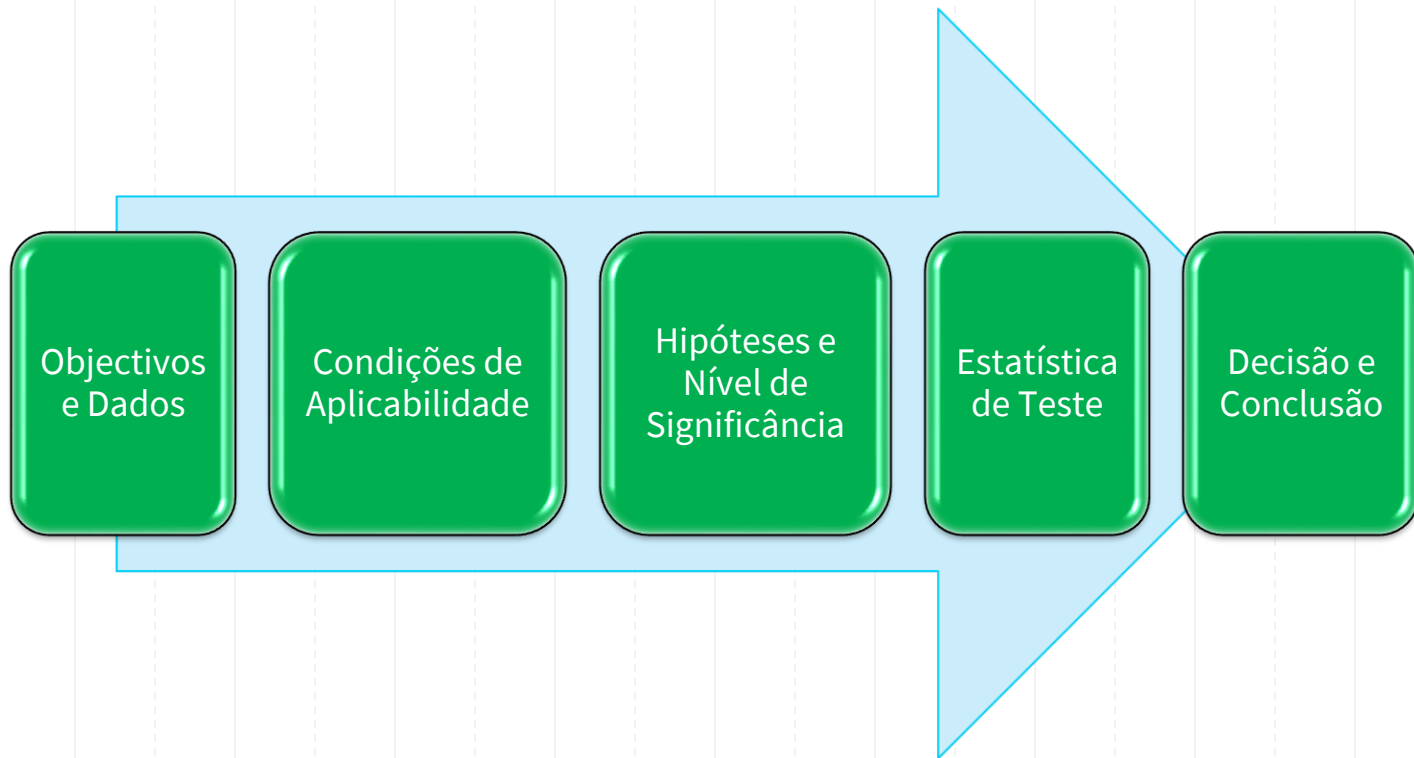
- Teste de Independência do Qui-Quadrado

Condições de Aplicabilidade de Testes Paramétricos

Normalidade

Homogenidade de
variâncias

Construção de um Teste de Hipóteses



Tipo de Hipóteses

1) $H_0: \mu = 1,70$ simples	vs. $H_1: \mu = 1,80$ simples
2) $H_0: \mu = 1,70$ simples	vs. $H_1: \mu > 1,70$ composta
3) $H_0: \mu \leq 1,70$ comp.	vs. $H_1: \mu > 1,70$ comp.
4) $H_0: \mu = 1,70$ simples	vs. $H_1: \mu \neq 1,70$ composta

Tipos de Erros vs Nível de Significância

Tipos de erros

- Erro tipo I (1ª espécie): rejeitar H_0 quando H_0 é verdadeira
- Erro tipo II (2ª espécie): não rejeitar H_0 quando H_0 é falsa

Decisão	H_0 Verdadeira	H_0 Falsa
Não rejeitar H_0	Correto $1-\alpha$	Erro tipo II β
Rejeitar H_0	Erro tipo I α	Correto $1-\beta$

Poder ou potência do teste

Definição: Chama-se potência do teste à probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira ($= 1 - \beta$).

Nível de significância ($0 \leq \alpha \leq 1$)

- Probabilidade que o investigador estabelece à priori como limite para decidir se rejeita H_0
- **Níveis usuais de significância:** 0,5%, 1%, 5% e 10%

$$\begin{aligned}\alpha &= P(\text{erro do tipo I}) = \\ &= P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira})\end{aligned}$$

A α chama-se nível de significância.

$$\begin{aligned}\beta &= P(\text{erro do tipo II}) = \\ &= P(\text{"Aceitar" } H_0 | H_0 \text{ é falsa})\end{aligned}$$

Decisão: Região de Rejeição vs Valor-p

Região de rejeição (RR) ou Região crítica (RC):
Conjunto para o qual H_0 é rejeitada

- Teste unilateral à esquerda: $RR =]-\infty; z_\alpha]$
- Teste unilateral à direita: $RR = [z_{1-\alpha}; +\infty[$
- Teste bilateral: $RR =]-\infty; -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}; +\infty[$



Nota: Supondo que a estatística de teste tem distribuição normal.

Regra (considerando os valores críticos):

- $z_0 \leq z_\alpha \Rightarrow$ Rejeita-se H_0
- $z_0 \geq z_{1-\alpha} \Rightarrow$ Rejeita-se H_0
- $|z_0| \geq z_{1-\alpha/2} \Rightarrow$ Rejeita-se H_0

Regra: $z_0 \in RR \Rightarrow$ Rejeita-se H_0

Dimensão do teste

Valor-p ou P-value: Probabilidade sob H_0 de a estatística de teste tomar valores tão ou mais desfavoráveis a H_0 do que o seu valor observado

- Teste unilateral à esquerda: $\text{valor-p} = P(Z \leq z_0)$
- Teste unilateral à direita: $\text{valor-p} = P(Z \geq z_0)$
- Teste bilateral: $\text{valor-p} = P(Z \leq -z_0 \text{ ou } Z \geq z_0) = 2 \times P(Z \geq |z_0|)$

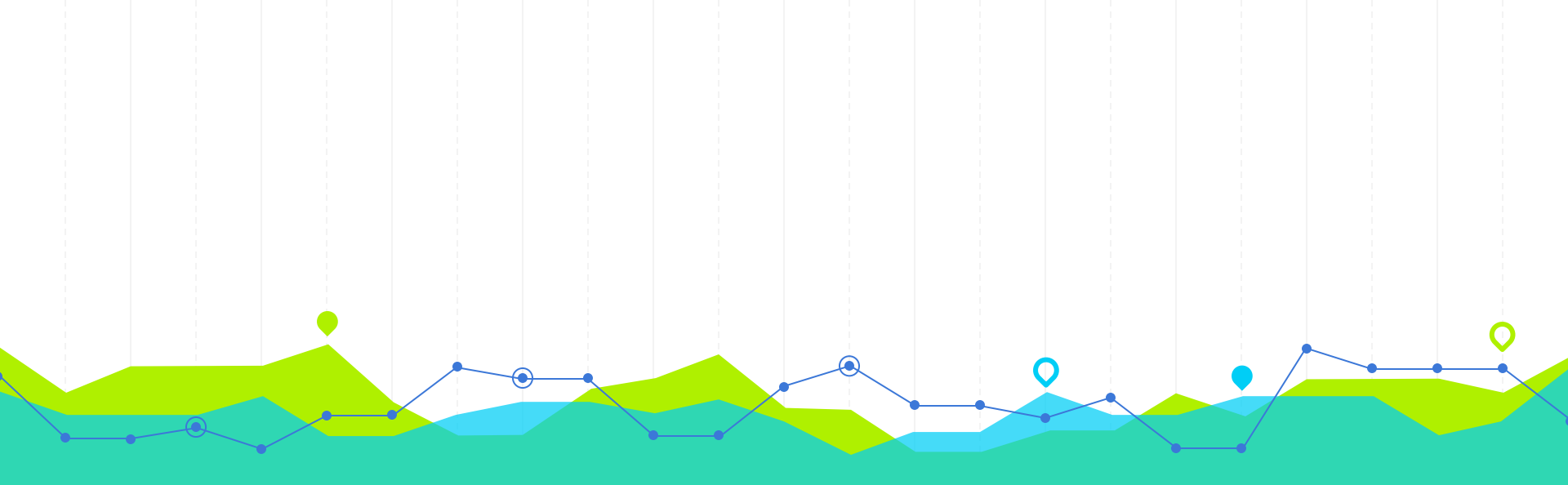
Regra: $\text{Valor-p} < \alpha \Rightarrow$ Rejeita-se H_0

O nível de significância α é a probabilidade de rejeitar a **hipótese nula** quando ela é verdadeira (conhecido como **erro do tipo I**).^[5] Em **testes de hipóteses** estatísticas, diz-se que há significância estatística ou que o resultado é estatisticamente significativo quando o p -valor observado é menor que o nível de significância α definido para o estudo.^{[2][3]} O nível de significância é geralmente determinado pelo pesquisador *antes* da coleta dos dados e é tradicionalmente fixado em 0,05 ou menos, dependendo da área de estudo.^{[6][7][8]} Em muitas áreas de estudo, resultados com nível de significância de 0,05 (probabilidade de erro de 5%) são considerados estatisticamente relevantes.^{[9][10][11]}

O p -valor (nível descritivo ou probabilidade de significância) é a probabilidade de se obter uma estatística de teste igual ou mais extrema que a estatística observada a partir de uma amostra aleatória de uma população quando a **hipótese nula** é verdadeira. Em outras palavras, o p -valor é o menor nível de significância para o qual se rejeita a hipótese nula. Por exemplo, a hipótese nula é rejeitada a 5% quando o p -valor é menor que 5%.^[12]

https://pt.wikipedia.org/wiki/Signific%C3%A2ncia_estat%C3%ADstica

Definição: valor- p é o menor nível de significância a partir do qual H_0 é rejeitado.



Testes de Hipóteses para μ (σ^2 Conhecida)

Hipóteses, Estatística de Teste e Decisão

9

8.4 Da produção diária de determinado fertilizante tiraram-se seis pequenas porções que se analisaram para calcular a percentagem de nitrogénio. Os resultados foram os seguintes:

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \bar{x} = 5.917 \\ 6.2 & 5.7 & 5.8 & 5.8 & 6.1 & 5.9 & \end{array}$$

Sabe-se, por experiência, que o processo de análise fornece valores com distribuição que se pode considerar normal com $\sigma^2 = 0.25$.

- (a) Suportam as observações a garantia de que a percentagem esperada de nitrogénio, μ , é igual a 6% ao nível de significância de 10%?
- (b) Responda à alínea anterior usando o valor- p .



Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Passo 0 [descrição da situação]

$X = \%$ de nitrogénio num fertilizante $\sim N(\mu, 0.25)$
(população normal, de valor esperado desconhecido
e de variância conhecida)

amostra de dimensão 6; $\bar{x} = 5.917$

A não esquecer: $\bar{X}$ é a mesma coisa que $E(X)$

S^2 " " " " $Var(X)$

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Hipóteses:

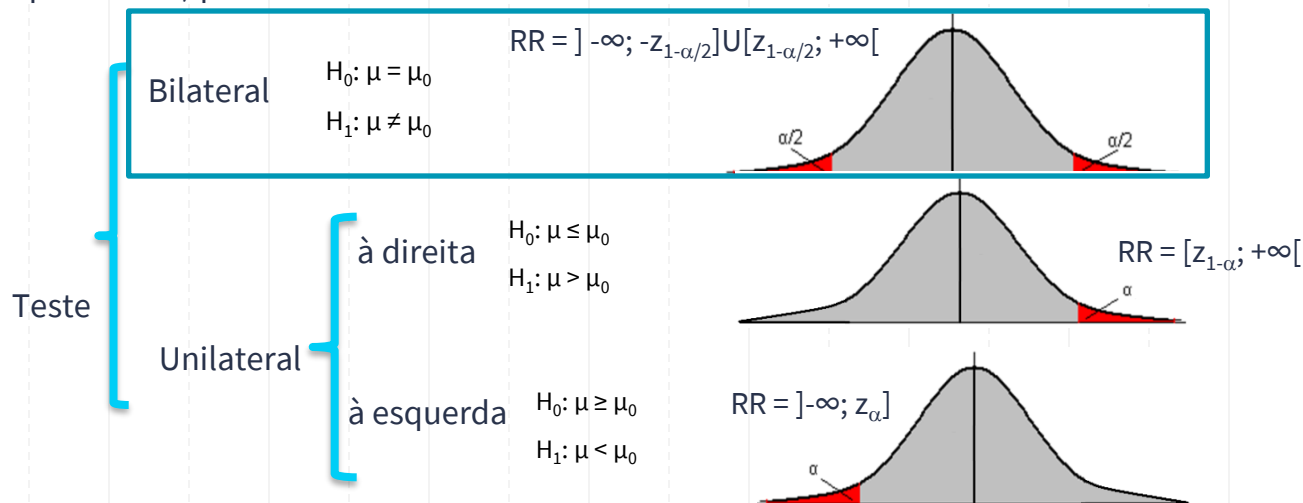
Passo 1 : Indicar quais as hipóteses que vamos testar e qual a significância

[nula] $H_0: \mu = 6$ [i.e.: vamos testar a hipótese H_0 i.e., se $\mu = 6$]

[alternativa] $H_A: \mu \neq 6$ [i.e.: em alternativa, vamos decidir se $\mu \neq 6$]

Tipos de Testes de Hipóteses para μ (σ^2 Conhecida)

Um **teste de hipóteses paramétrico** para o parâmetro μ (valor médio ou média populacional) pode ser:



onde μ_0 é o valor numérico específico considerado em H_0 e H_1 .

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Estatística de Teste:

Passo 2: Escolhe de v. fndal e consequente estatística de teste.

Regras para escolha de v. fndal no cap. 7:

I.C. p/ μ, σ
conhecido,
 $X \sim N(\mu, \sigma^2)$

I.C. p/ μ, σ
desconhecido,
 $X \sim N(\mu, \sigma^2)$

idên, n grande

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$	$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$ ①	$\frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} N(0, 1)$	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ ②		$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim} \chi^2_{(k-m-1)}$

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

I.C. p/ σ , com
 μ desconhecido
 $X \sim N(\mu, \sigma^2)$

IC para μ : Formulário

• POPULAÇÕES NORMAIS

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(\nu)$
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	onde ν é o maior inteiro contido em r, $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{S_2'^2}{n}\right)^2}$
Variância	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Estatística de Teste:

No presente exemplo, estamos a construir um teste de hipótese para o valor esperado de uma população normal de variância conhecida.

$$V.F. \equiv Z \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\text{Est. Teste} \equiv Z_0 \quad \frac{\bar{X} - 6}{\sigma/\sqrt{n}} \stackrel{H_0}{\sim} N(0, 1)$$

(ie, a estatística de teste é a variável aleatória,

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Estatística de Teste:

depois de substituído o parâmetro em Teste pelo valor que consta em H_0)

$$\text{valor observado: } z_0 = \frac{5.917 - 6}{\sqrt{0.25}/\sqrt{6}} = -0.41$$

$H_0: \mu = 6$

n (dimensão da amostra)

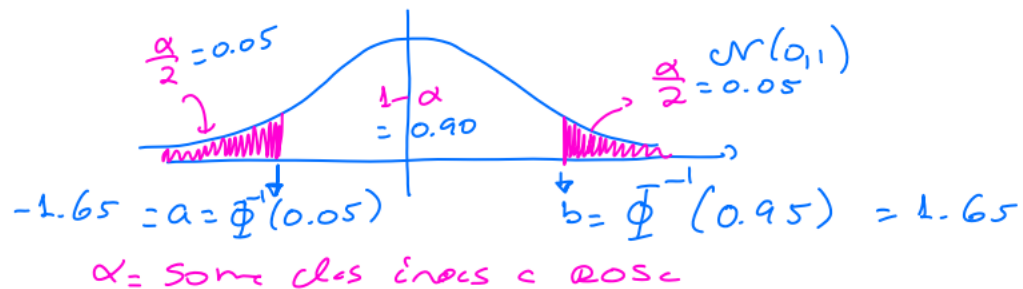
~~$N(0,1)$~~

[Passo 2: dist. de $x + H_0$ + saber quais os parâmetros conhecidos]

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Decisão (pela região de rejeição):

PASSO 3 : Construção da Região de Rejeição



[Passo 3 = distribuição de v. fluee + α + forme? do H_1]

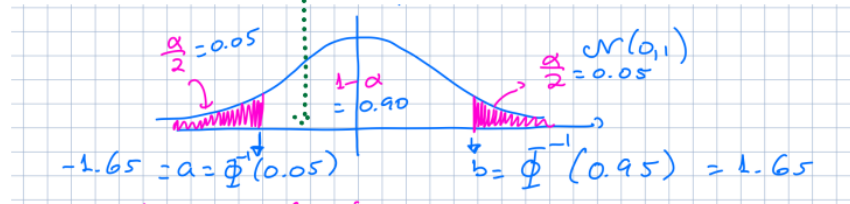
Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Decisão (pela região de rejeição):

Passo 4 : Decisão sobre aceitação ou rejeição de H_0

$$Z_0 = -0.41$$

Região de rejeição $RR =]-\infty; -1,65] \cup [1,65; +\infty[$



[Mariana: está bonita a aula de hoje?
aguardo resposta!]

Então, como $-1.65 < Z_0 < 1.65$, significa que para
10% = 5% não há evidências estatísticas para rejeitar H_0

[Passo 4 = Região rejeição + t_0]

Decisão pela região de rejeição:

Não se rejeita H_0 para $\alpha = 10\%$, pois $-1,65 < -0,41 < 1,65$. Não existe evidência estatística para afirmar que o valor médio é diferente de 6 para $\alpha = 10\%$.

a) Sim ($-1.645 < -0.408 < 1.645$)

b) Valor-p = $0.6818 > 0.1$

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Realidade vs H ₀	decisão	
	Aceitar H ₀	Rejeitar H ₀
Verdadeira H ₀	O.K.	ERRO Tipo 1
Falsa H ₀	ERRO Tipo 2	O.K.

$$\alpha = P(\text{erro tipo 1}) = P(\text{Rejeitar } H_0 \mid H_0 \text{ é verdadeira})$$

(nível de significância)

(por diversas razões, damos mais relevância ao erro tipo 1)

No exemplo: $\alpha = 0.10$

$$[\text{Passo 1} = H_0 + H_1 + \alpha]$$

Exercício 8.4 (a): Teste de Hipóteses para μ (σ^2 Conhecida)

Comentário: A forma do H_1 é essencial no passo 3

$$H_0: \mu = 6$$

$$H_1: \mu \neq 6$$

$$H_0: \mu = 6$$

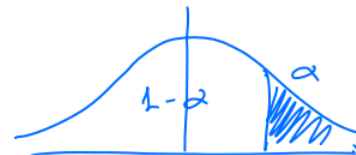
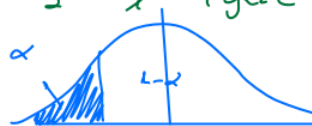
$$H_1: \mu < 6$$

$$H_0: \mu = 6$$

$$H_1: \mu > 6$$

Passo 0 é $\neq$ igual

Passo 2 é igual



$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

α é a área de zona de rejeição!

Exercício 8.4 (a): Teste t para o Valor Médio (σ^2 Conhecida)

Hipóteses

Teste Bilateral

$H_0: \mu = 6$ versus $H_1: \mu \neq 6$

Nota: A variável média amostral tem distribuição normal, logo este teste de hipóteses é válido.

Estatística de Teste

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

Valor Observado da Estatística de Teste (VOE)

$z_0 = -0,41$

Regra: $z_0 \in RR \Rightarrow$ Rejeita-se H_0

Decisão

Pela região de rejeição: $z_0 = -0,41$ não pertence à região de rejeição $RR =]-\infty; -1,645] \cup [1,645; +\infty[$

Pelo valor-p: Valor-p = 0,6818 > 0,10 (ver slide a seguir)

Regra: Valor-p < $\alpha \Rightarrow$ Rejeita-se H_0

Não se rejeita-se H_0 para $\alpha = 10\%$. Não existe evidência estatística para afirmar que a percentagem esperada de nitrogénio é diferente de 6% para $\alpha = 10\%$.

Dados:

$N = 6$

Média amostral = 5,917

$\sigma^2 = 0,25$

$\mu_0 = 6$

$\alpha = 0,10$

$z_{1-\alpha/2} = z_{0,95} = 1,645$

$$RR =]-\infty; -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}; +\infty[$$

Cálculo do Quantil da Distribuição Normal de Probabilidade $1-\alpha/2$

Nível de significância ($\alpha=0,1$)

Nível de confiança ($1-\alpha=0,90$)

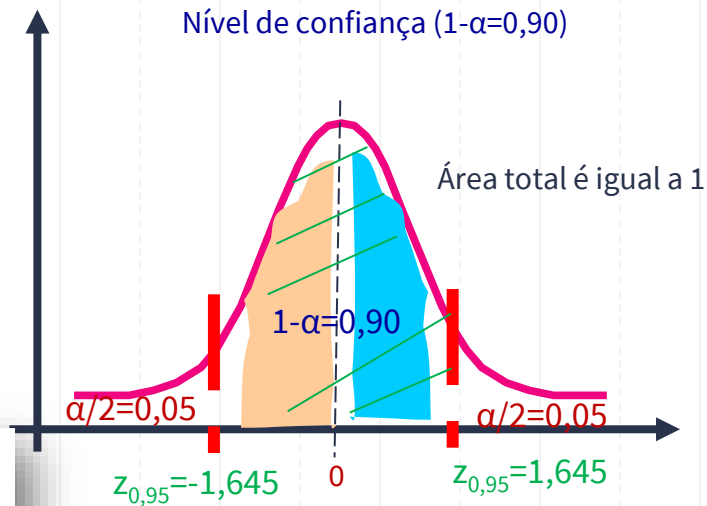


TABELA 5 – DISTRIBUIÇÃO NORMAL: $\Phi^{-1}(z)$

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

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

O nível de significância é igual a $\alpha = 0,10$, então tem-se $z_{1-\alpha/2} = z_{0,95} = 1,645$

Teste bilateral: valor-p = $P(Z \leq -z_0 \text{ ou } Z \geq z_0) = P(Z \leq -z_0) + P(Z \geq z_0) = 2 \times P(Z \geq |z_0|)$

Exercício 8.4 (b): Valor-p quando a Estatística de Teste tem Distribuição Normal Padrão

Decisão (pelo p-value):

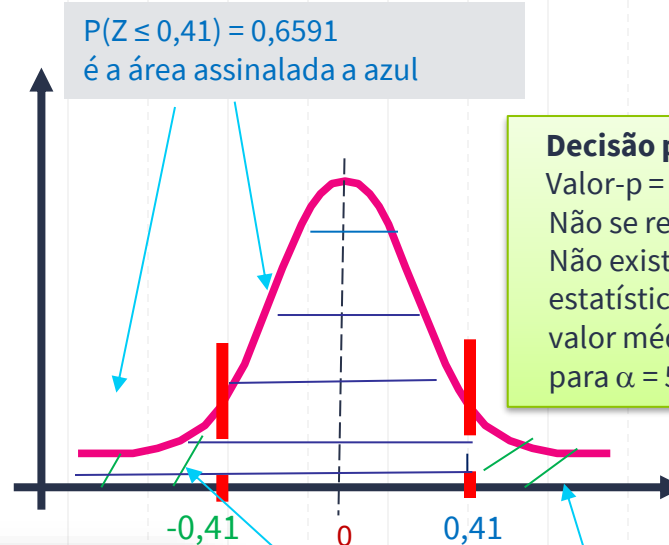
$$\begin{aligned} \text{valor-p} &= P(Z \leq -0,41 \text{ ou } Z \geq 0,41) \\ &= 2 \times P(Z \geq 0,41) = 2 \times [1 - P(Z < 0,41)] \end{aligned}$$

A tabela geral só permite obter probabilidades de quantis positivos e do tipo $P(Z \leq z)$.

Então, tem-se
 $P(Z \leq 0,41) = 0,6591$, logo

$$\text{valor-p} = 2 \times [1 - P(Z \leq 0,41)] \sim 2 \times [1 - 0,6591] = 0,6818$$

Área total é igual a 1



Decisão pelo valor-p:

Valor-p = 0,6818 > 0,10
 Não se rejeita H_0 para $\alpha = 10\%$.
 Não existe evidência estatística para afirmar que o valor médio é diferente de 6 para $\alpha = 5\%$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879

$P(Z \leq -0,41) = 1 - P(Z \leq 0,41) = P(Z \geq 0,41) = 0,3409$
 é a área assinalada a verde

a) Sim ($-1.645 < -0.408 < 1.645$)

b) Valor-p = 0.6818 > 0.1

1. Para cada uma das proposições seguintes refira se se trata ou não de uma hipótese estatística:
- a) $\mu = 3$;
 - b) $\bar{x} = 4$;
 - c) $P(X < 2.5) = 0.4$;
 - d) $2 < \sigma < 3$;
 - e) $\bar{X} < 3$.



14. Um comerciante recebe ovos de um determinado aviário, onde os ovos são classificados, consoante o peso, em duas classes, A e B . O peso dos ovos da classe A tem distribuição normal de média 50 gramas e desvio padrão 6 gramas, enquanto o peso dos ovos da classe B tem distribuição normal de média 55 gramas e desvio padrão idêntico ao da classe A . O comerciante acaba de receber uma remessa de ovos com garantia de serem da classe B e tem um prazo de 2 dias para reclamar, caso considere ter havido engano da parte do aviário. Considere $H_0 : \mu = 55$ contra $H_1 : \mu = 50$.
- a) Tal como o teste foi definido, pode afirmar-se que o aviário tem o benefício da dúvida?
 - b) Para tomar uma decisão, o comerciante analisou uma amostra de uma dúzia de ovos cujo peso total foi de 630 gramas. Qual a atitude que o comerciante deve tomar? (utilize a dimensão de 0.05 como referência, e de 0.10 para comparação).
 - c) Determine a probabilidade do erro de 2.^a espécie. Qual o seu significado?
 - d) Se pretender que a probabilidade do erro de 2.^a espécie seja idêntica à do de 1.^a espécie (0.05), qual deve ser a dimensão da amostra a considerar?

Nota: Nesta questão, pretende-se determinar o p-value.



16. Seja X uma variável aleatória que representa a quantidade de vinho existente numa garrafa de 75 centilitros. Admita que X tem distribuição normal com desvio padrão igual a 2. Para testar $H_0 : \mu = 75$ contra $H_1 : \mu < 75$ seleccionou-se uma amostra casual de 10 garrafas, rejeitando-se a hipótese nula se $\bar{x} < 74.1$, onde $\bar{x}$ é a quantidade média de vinho, por garrafa, na amostra observada.
- a) Calcule a dimensão deste teste.
 - b) Determine a função potência, e calcule o seu valor quando $\mu = 74$ e $\mu = 72.5$.

Nota: Nesta questão, pretende-se determinar o p-value.



Obrigada!

Questões?

