CURSO DE ESTATÍSTICA EXPERIMENTAL
PARTE I

ASPECTOS BÁSICOS EM GERAL

Coordenador: JOSÉ RICARDO ESCOBAR

PERÍODO: 08 A 12/09/86
LOCAL: UEPAE DE MANAUS

S U M Á R I O

ESTATÍSTICA EXPERIMENTAL

1. CONCEITOS BÁSICOS

1.1 RACIOCÍNIO DEDUTIVO (Produtor)

Se dispõe de certos princípios e nos interessa saber o que acontecerá sob certas condições específicas. Raciocínio do geral ao particular.
(aplicação de princípios)
(Teste de métodos, técnicas, resultados etc.)

- Temos a fórmula geral do círculo $A = \pi r^2$. Qual é a área de um círculo de raio = 15 cm².

- Temos a seguinte função de resposta a adubos potássicos $Y = 2{,}00024x - 0{,}1392x^2$. Qual é a produção esperada ao aplicar 50 kg do K?

- Dispondo de uma chave e descrições de invasoras no Amazonas. A que espécie e gênero, pertence certa invasora coletada em Manacapuru?

RACIOCÍNIO INDUTIVO (pesquisador)

Se dispõe de vários casos específicos e se deseja chegar a princípios gerais, que serão aplicados a todos os casos. Do particular ao geral.
(geração de princípios, normas, funções. comportamento).

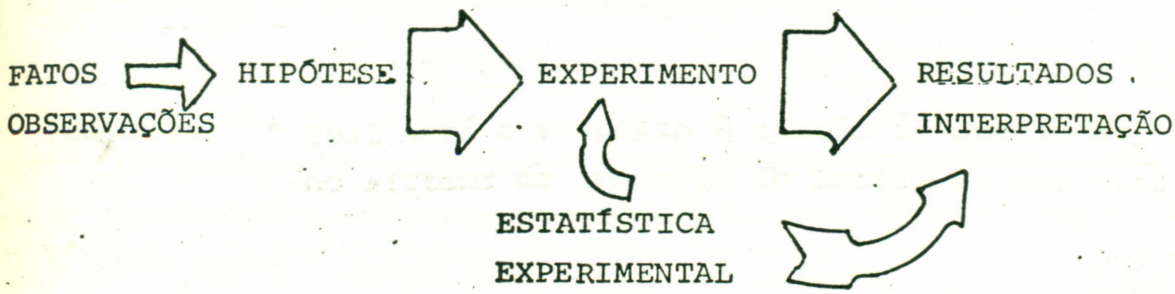
- Com dados das áreas e raios de muitos círculos.
Que fórmula geral podemos gerar, que expresse a relação entre as áreas e os raios de todos os círculos?

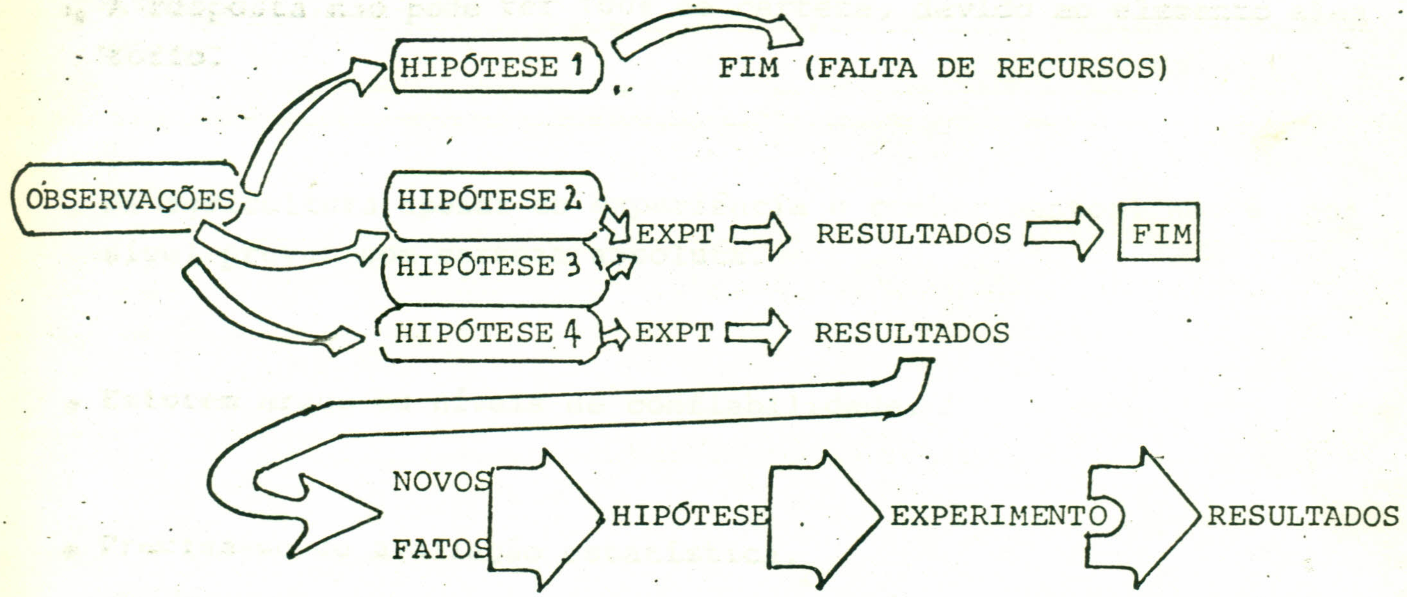
- Com dados das produções obtidas para diferentes níveis de K em vários anos de pesquisa. Qual será a função de resposta para o K para a região X e a cultura Y?

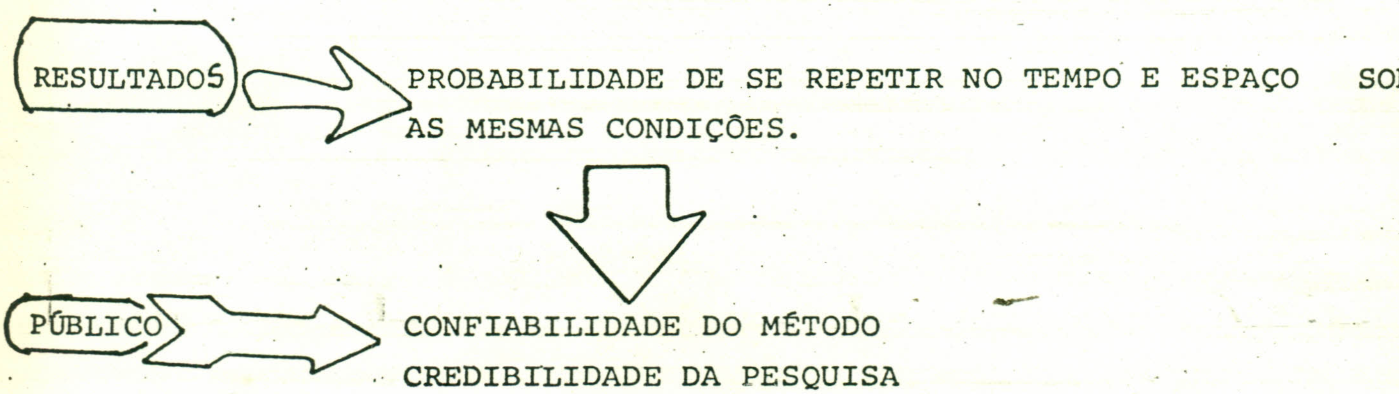
- Se dispõe de numerosas descrições de invasoras no Estado do Amazonas. Qual será uma chave, com descrições detalhadas que permita classificar as invasoras no Estado do Amazonas?

1.2 MÉTODO CIENTÍFICO

Se refere a prática e utilização do método científico.







RACIOCÍNIO DEDUTIVO NA AGRICULTURA

PERGUNTA: " Qual será a resposta à adoção de uma nova tecnologia no sistema de produção do Estado do Amazonas ".

- A resposta não pode ter 100% de certeza, devido ao elemento aleatório.
- Na agricultura apesar da experiência e conhecimentos, não é possível prever com certeza absoluta.
- Existem graus ou níveis de confiabilidade.
- Precisa-se de avaliação estatística.

ESTATÍSTICA EXPERIMENTAL

1.3 DADOS E VARIÁVEIS

CONCEITOS ELEMENTARES

1. DADOS

- 25.000 produtores de guaraná
- 50 vacas leiteiras
- 10 lesões de antracnose

2. SUBSTANTIVOS

Variáveis:
- x = número de cachos/planta
- y = kg adubo/mês
- z = TM de milho/ha

3. ADJETIVOS

Valores das variáveis:

- x_1 = 5 cachos
- x_2 = 15 cachos
- x_3 = 7 cachos
- y_1 = 20 kg adubo/Jan
- y_2 = 50 kg adubo/Mar
- y_3 = 40 kg adubo/Dez
- z_1 = 15 TM/ha
- z_2 = 25 TM/ha
- z_3 = 5 TM/ha

4. VERBOS

Somatória: Σ

Raiz quadradra: $\sqrt{x}$

Potência: X^a

5. ADVÉRBIOS

Modificam à ação do verbo

Exemplo: $\sum_{x=1}^{x=3} = x_1 + x_2 + x_3$

ESTATÍSTICA EXPERIMENTAL

2.1 DISTRIBUIÇÃO DE FREQÜÊNCIAS

NÚMERO DE PLANTAS DE DENDÊ POR HECTARE DE UMA FAZENDA COM PROBLEMAS DE SOBREVIVÊNCIA

● DADOS DE 50 HECTARES

82	115	113	125	93	126	94	117	100	122
100	118	109	119	103	119	108	121	108	118
105	123	106	115	108	116	104	126	111	127
108	130	103	131	110	131	111	142	104	143
112	136	89	139	114	140	114	133	115	133

● DADOS ORGANIZADOS

82	100	105	108	112	115	118	123	130	136
89	103	106	109	113	115	119	125	131	139
93	103	108	110	114	116	119	126	131	140
94	104	108	111	114	117	121	126	133	142
100	104	108	111	115	118	122	127	133	143

● DADOS CLASSIFICADOS

Classes	f	x	fA	fD	FR (%)	fRA	fRD
82– 86	1	84	1	50	2	2	100
87– 91	1	89	2	49	2	4	98
92– 96	2	94	4	48	4	8	96
97–101	2	99	6	46	4	12	92
102–106	6	104	12	44	12	24	88
107–111	8	109	20	38	16	40	76
112–116	8	114	28	30	16	56	60
117–121	6	119	34	22	12	68	44
122–126	5	124	39	16	10	78	32
127–131	4	129	43	11	8	86	22
132–136	3	134	46	7	6	92	14
137–141	2	139	48	4	4	96	8
142–146	2	144	50	2	4	100	4
	50				100		

f = freqüência

x = ponto médio de classe

fA = freqüência ascendente

fD = freqüência descendente

fR = freqüência relativa (%)

fRA= freqüência relativa ascendente

fRD= freqüência relativa descendente

ESTATÍSTICA EXPERIMENTAL

RECURSOS PARA A CLASSIFICAÇÃO DE DADOS

● PRIMEIRO PASSO

Amplitude: valor mais alto - valor mais baixo

: 143 - 82 = 61

●● SEGUNDO PASSO

Determinação do intervalo de classes. Dividir a amplitude pelo número de classes aproximadas.

$\frac{61}{13} = 4.69 = 5$ —— Intervalo de classes

●●● TERCEIRO PASSO

Determinar os limites de classes

INÍCIO	82 + 4 = 86	→ 82-86	: Primeira classe (somar intervalo - 1)
	(+ 5) 87	87-91	: Segunda classe
	(+ 5) 92	:	:
		142-146	: Classe nº 13 - inclue valor mais alto.

●●●● QUARTO PASSO

Classificar a informação (vide exemplo)

DETALHES PARA LEMBRAR

→ PRIMEIRA CLASSE (exemplo) → Limite inferior: 82
Limite superior: 86
Intervalo de classe: 4

→ LIMITES REAIS → Limite inferior: 81.5
Limite superior: 86.5
Intervalo de classe real: 5

ESTATÍSTICA EXPERIMENTAL

Limite real: Limite suoerior + limite inferior próxima classe

$$\frac{86 + 87}{2} = 86.5 = \text{assim sucessivamente} :: :: ::$$

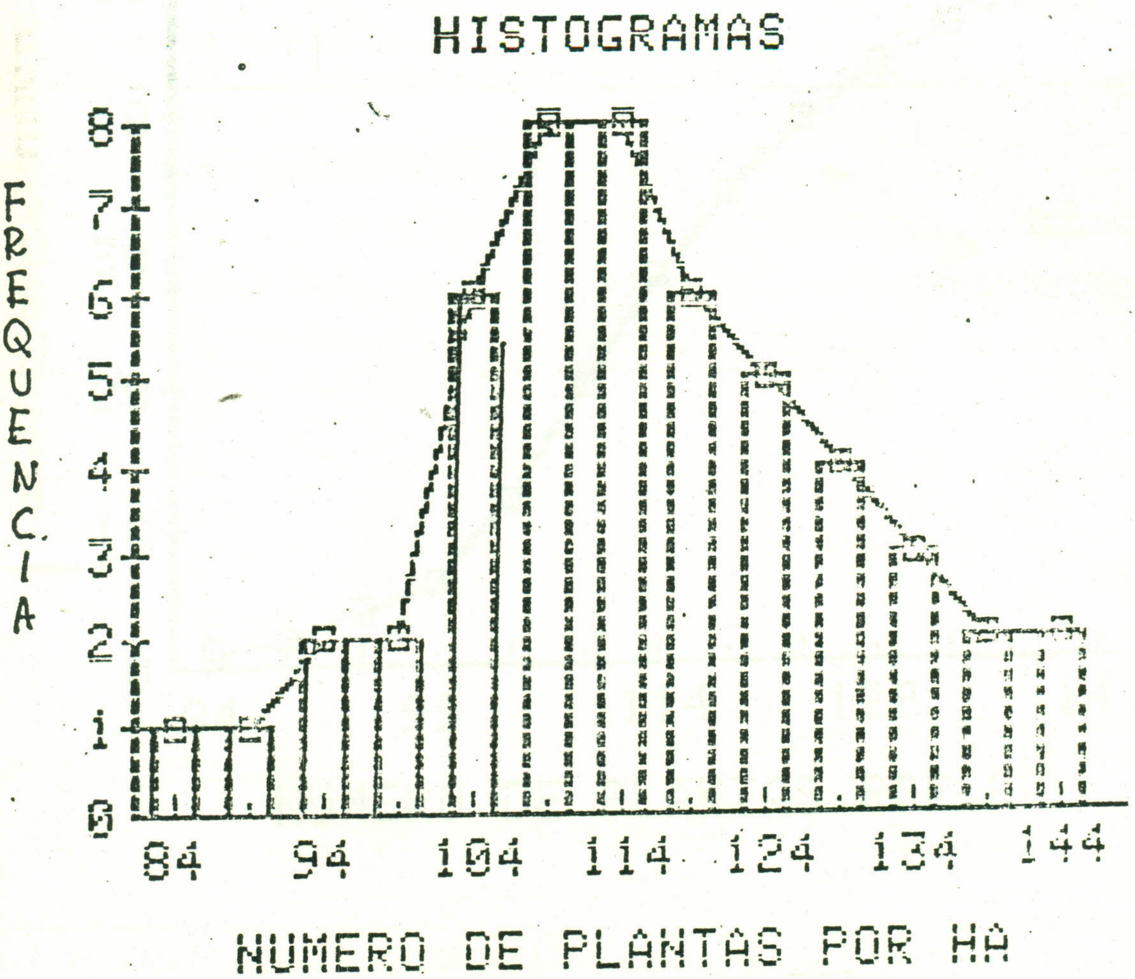
PERGUNTAS (exemplos):

- Qual é o número maior de plantas por ha? R/143 (dos ordenados)
- Quantos hectares têm 116 plantas ou menos? R/ 28
- Quantos hectares têm 143 plantas? R/ 2 ⟶ 1 (dados organizados)
- Que porcentagem da área total não têm a população correta? R/96%

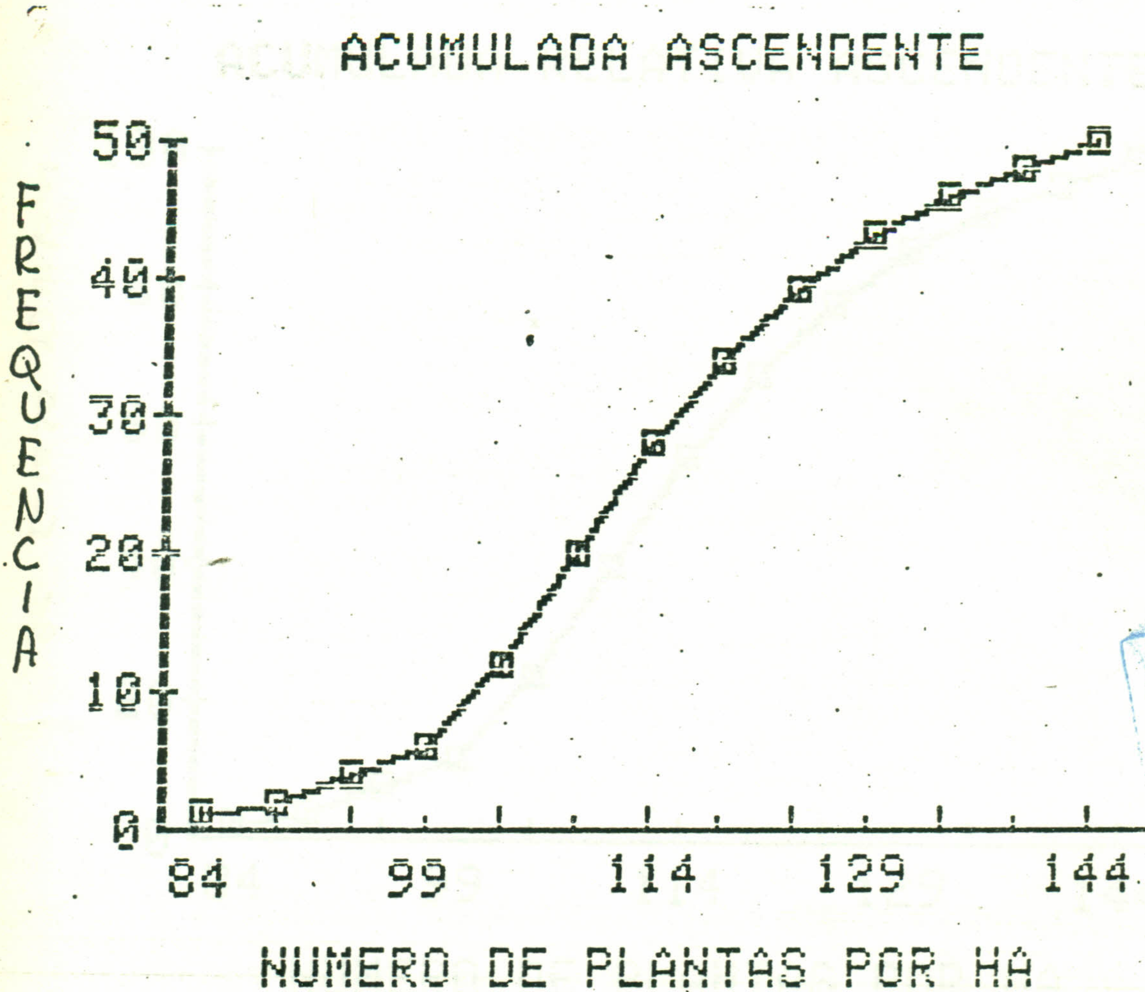
CONCLUSÃO:

- O 16% da área tem entre 107 - 111 plantas/hectare
- 43 hectares tem menos que 132 plantas/hectare

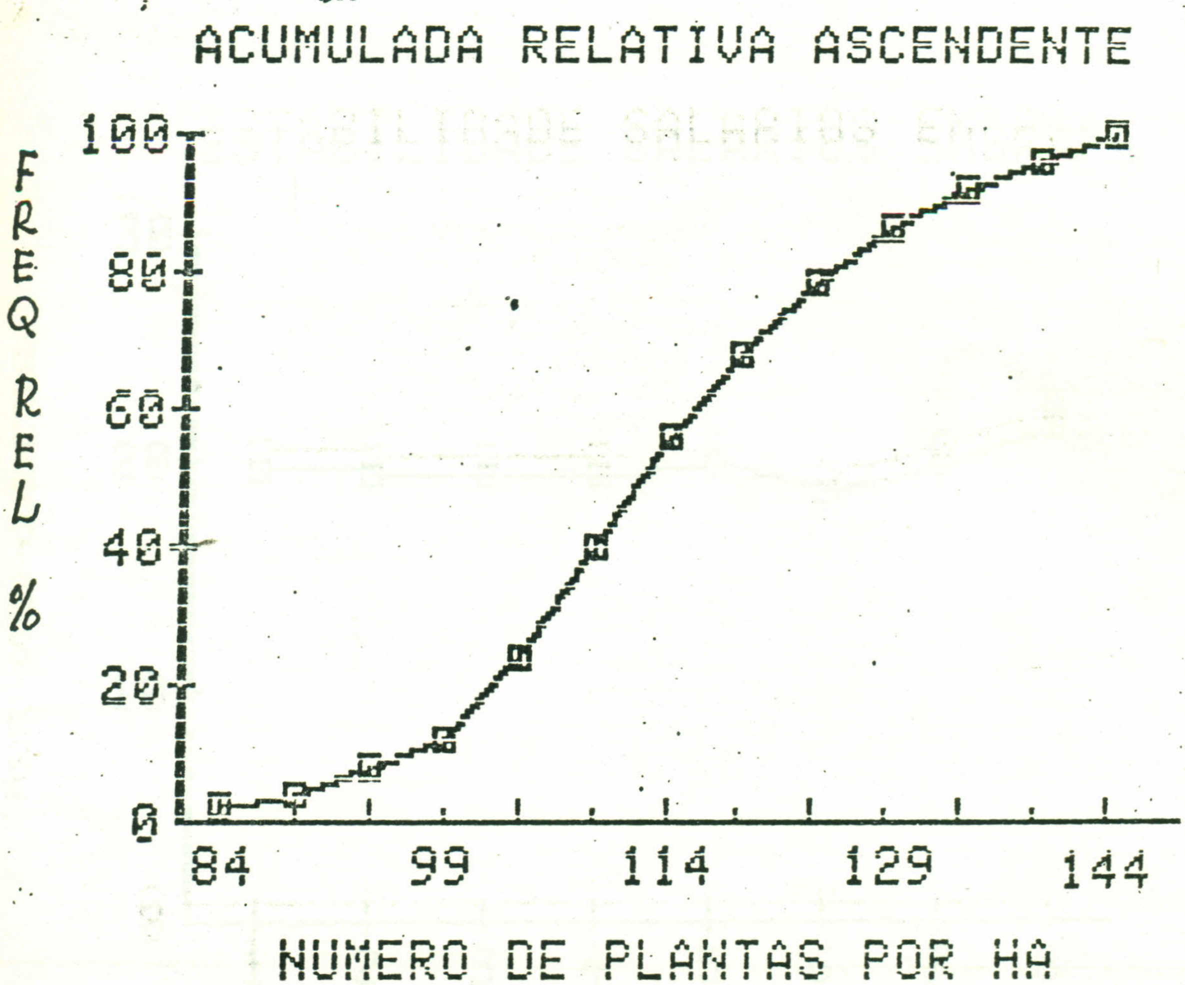
REPRESENTAÇÕES GRÁFICAS



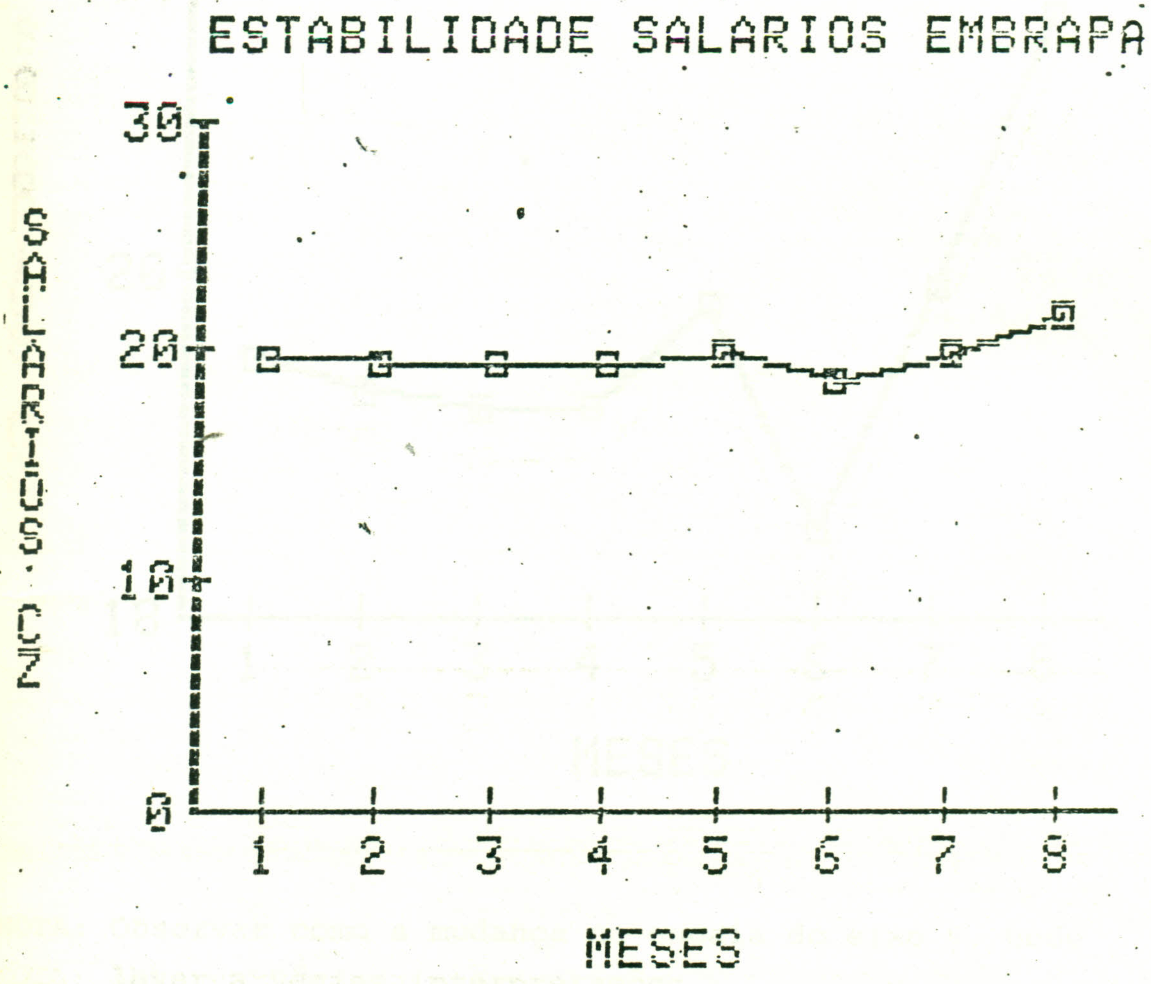
REPRESENTAÇÕES GRÁFICAS



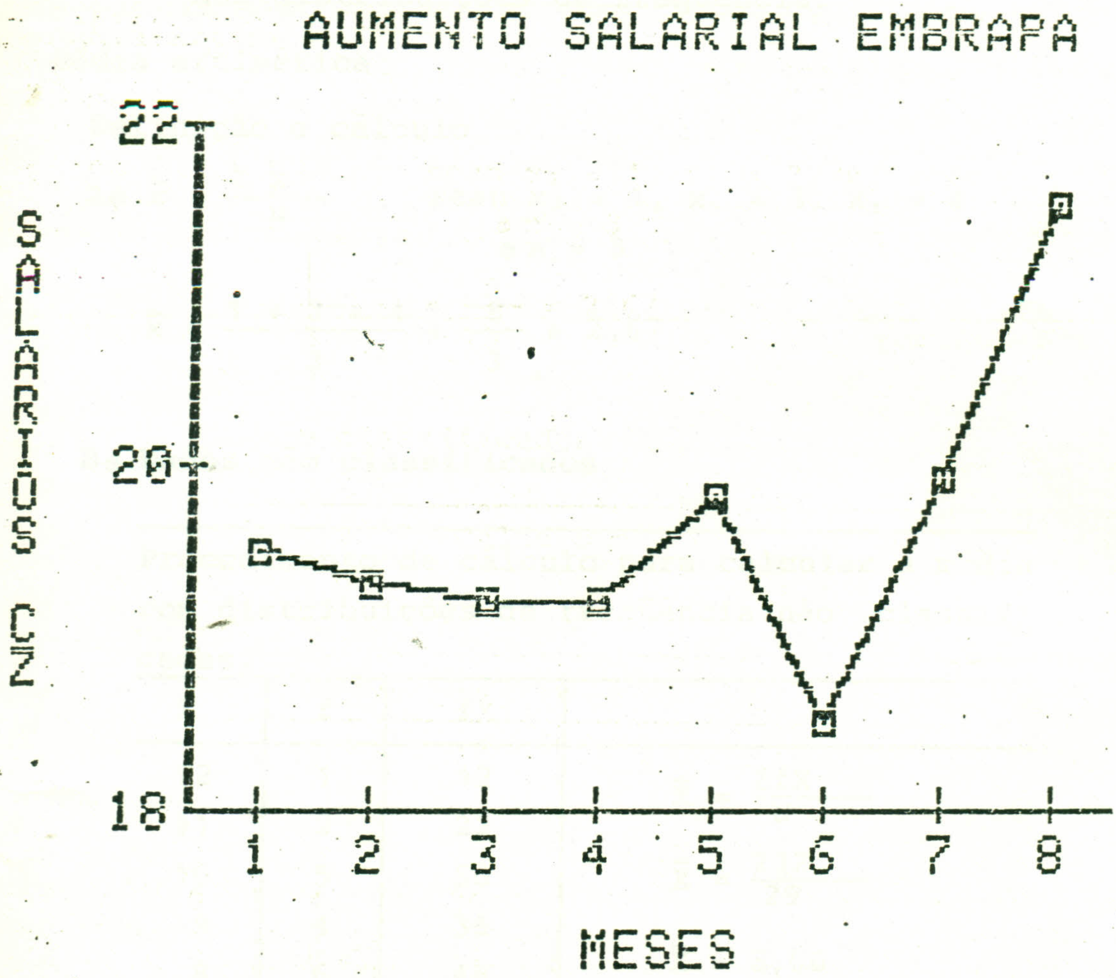
REPRESENTAÇÕES GRÁFICAS



REPRESENTAÇÕES GRÁFICAS



REPRESENTAÇÕES GRÁFICAS



NOTA: Observar como a mudança na escala do eixo y, pode levar a várias interpretações.

2.3 Medidas de tendência central

CONCEITO: Indice de localização central empregado na descrição das distribuições de freqüência.

Média artimética

Definição e cálculo

A. $\bar{x} = \frac{\Sigma x}{n}$ sean $x_1 = 1,\ x_2 = 3,\ x_3 = 4$

$n = 3$

$$\bar{x} = \frac{1 + 3 + 4}{3} = \frac{8}{3} = 2.67$$

B. Dados não classificados

Procedimento de cálculo para calcular a média com distribuições de freqüência não classificadas.

X	f	fX	
12	1	12	$\bar{X} = \frac{\Sigma fX}{N}$
11	2	22	
10	5	50	$\bar{X} = \frac{232}{29}$
9	4	36	
8	6	48	$\bar{X} = 8,00$
7	4	28	
6	3	18	
5	2	10	
4	2	8	
	$N = 29$	$\Sigma f X = 232$	

2.3 MEDIDAS DE TENDÊNCIA CENTRAL

C. Dados Classificados

Procedimentos para calcular a média a partir de uma distribuição de freqüências classificadas. Se usa o ponto médio de classe como valor da variável.

1 Intervalo de classe	2 Freqüência (f)	3 Ponto Médio (X)	4 fX	
125 - 129	2	127	254	• $\bar{X} = \frac{fX}{N}$
120 - 124	5	122	610	
115 - 119	8	117	936	
110 - 114	10	112	1.120	• $\bar{X} = \frac{10.195}{100}$
105 - 109	15	107	1.605	
100 - 104	20	102	2.040	
95 - 99	15	97	1.455	• $\bar{X} = 101,95$
90 - 94	10	92	920	
85 - 89	8	87	696	
80 - 84	4	82	328	
75 - 79	3	77	231	
	N = 100		Σ fX = 10.195	

PROPRIEDADES DA MÉDIA

A. A soma dos desvios com relação a média é zero

$\Sigma (x - \bar{x}) = 0$ • Sejam $x_1 = 1$, $x_2 = 3$, $x_3 = 4$

$$\bar{x} = \Sigma \frac{(1 + 3 + 4)}{3} = 2.66$$

x_i	$(x_i - \bar{x})$
1	-1.66
3	0.33
4	1.33
	0

$\Sigma (x - \bar{x}) = -1.66 + 0.33 + 1.33 = 0$

ESTATÍSTICA EXPERIMENTAL

Medidas de tendência central

B. A soma dos desvios ao quadrado com relação a média e um mínimo.

1	2	3	4	5	6
x_i	$(x_i - \bar{x})^2$	$(x_i - 2)^2$	$(x_i - 3)^2$	$(x_i - 5)^2$	$(x_i - 6)^2$
2	4	0	1	9	16
3	1	1	0	4	9
4	0	4	1	1	4
5	1	9	4	0	1
6	4	16	9	1	0
6	10	30	15	15	30

$N = 5$

$\bar{x} = 4$

2. Mediana

E o valor da variável que divide em duas uma serie de dados ordenados de menor a maior ou vice-versa.

- Se é impar é o valor central

  1, 2, 3, 4, 5, 9, 11, 12, 14

  $M = 5$

- Se é par e a média dos valores centrais

  1, 2, 3, 4, 5, 9, 11, 12,

  $$M = \frac{4 + 5}{2} = 4.5$$

ESTATÍSTICA EXPERIMENTAL

MEDIANA

Para dados classificados se usa a seguinte relação:

$$M = + \frac{N/2 - fAA}{fA} \times I$$

L = Limite inferior da classe que contém a mediana.

Do nosso exemplo de 50 ha de palma nós temos.

N = Soma de freqüências

fAA = freqüências acumuladas até a anterior classe.

$$M = 111.5 + \frac{25 - 20}{8} \times 5$$

fA = frequência absoluta da que contém a mediana.

$$M = 114.625 = 115$$

I = Intervalo de classe

:: Notar que se usou o intervalo real de classe I = 5 e não I = 4; e o limite real de classe.

Também se pode determinar diretamente dos dados ordenados.

$$M = \frac{115 + 115}{2} = 115$$

3. Moda

É o valor da variável que ocorre com mais freqüência

$$Mo = Li + \frac{fP}{fA + fP} \times I$$

fP = freqüência da classe posterior a que contém à máxima freqüência.

De nosso exemplo

fA = freqüência da classe anterior a que contém a máxima freqüência

$$Mo = 106.5 + \frac{8}{8 + 6} x = 109.3$$

O que não é correto devido que o valor 108 é o mais freqüênte, no entanto 109.3 indica que a classe 107-111 contém a moda.

2.3 - Medidas de Tendência Central

PARA LEMBRAR: Quando a informação é influenciada por valores extremos, a média se torna uma medida imperfeita.

x_1	x_2
2	2
3	3
5	5
7	7
8	33
$\Sigma x_1 = 25$	$\Sigma x_2 = 50$
$\bar{x}_1 = 5{,}00$	$\bar{x}_2 = 10{,}00$

→ dobrou o valor

O valor da média por si mesma, não indica nada acerca da variabilidade da informação:

$$\bar{x}_1 = \frac{5 + 5}{2} = 5$$

$$\bar{x}_2 = \frac{6 + 4}{2} = 5$$

$$\bar{x}_3 = \frac{7 + 3}{2} = 5$$

$$\bar{x}_4 = \frac{8 + 3}{2} = 5$$

$$\bar{x}_5 = \frac{9 + 1}{2} = 5$$

+ consistente

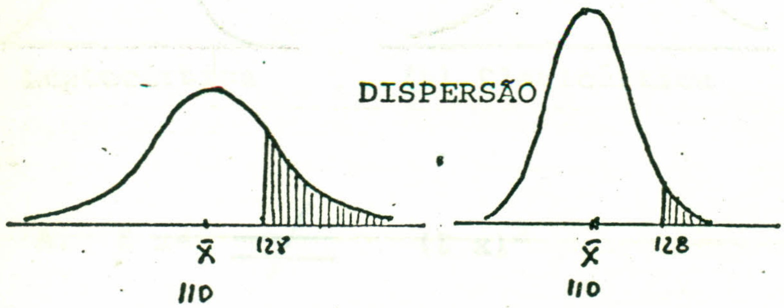
↓

- consistente

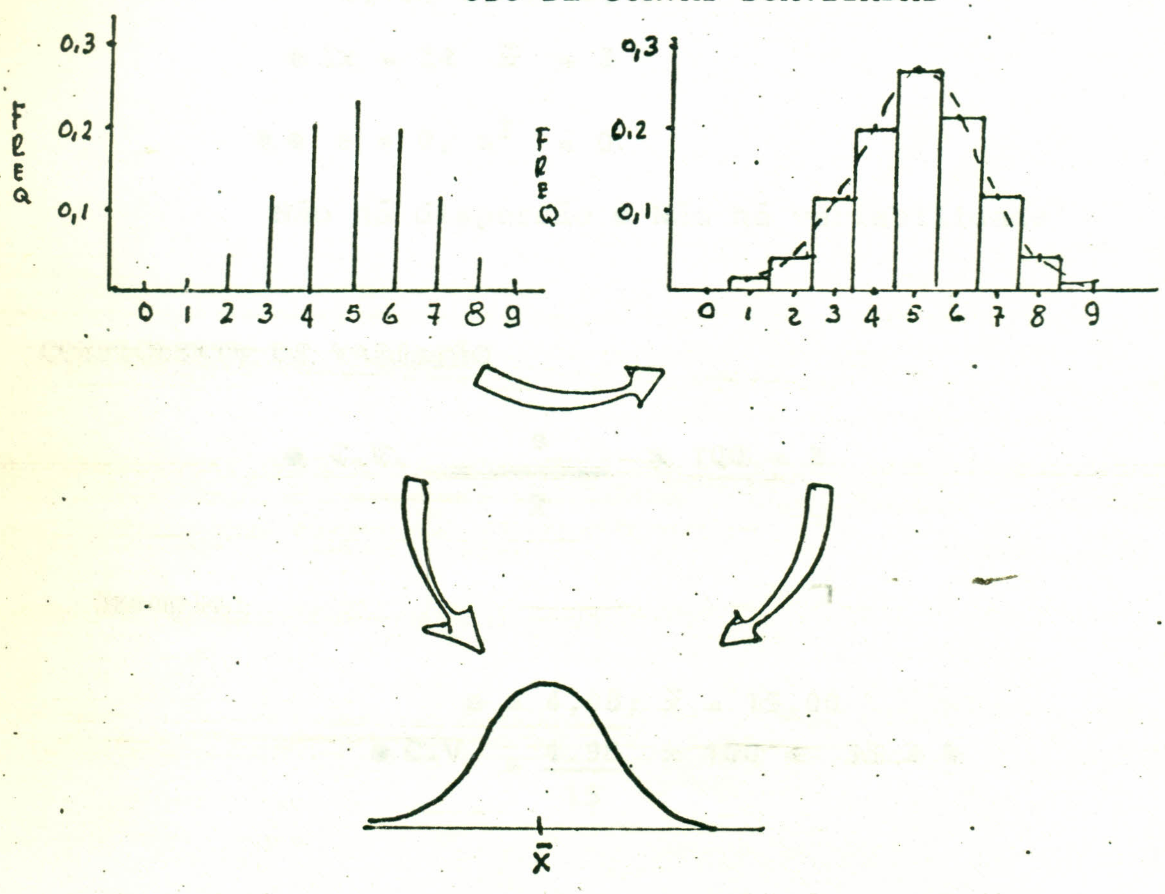
2.4 Medidas de dispersão

A. CONCEITOS

Dois grupos de dados com igual média e diferente dispersão.



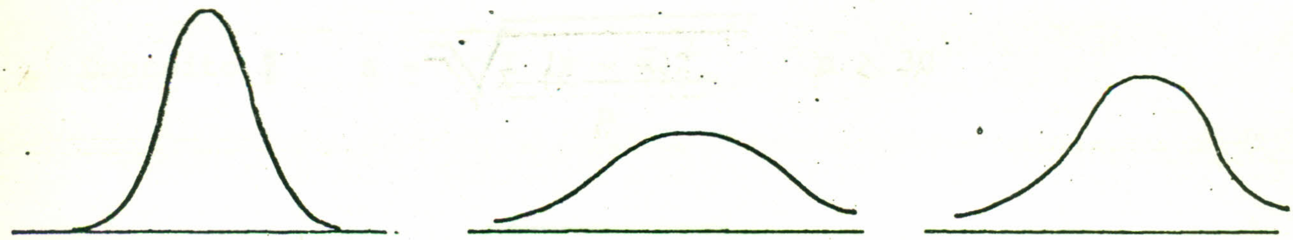
USO DE CURVAS SUAVIZADAS



ESTATÍSTICA EXPERIMENTAL

MEDIDAS DE DISPERSÃO

FORMAS DA DISPERSÃO:



(a) Leptocúrtica (b) Platicúrtica (c) Mesocúrtica

LEMBRAR:

A. $\Sigma x^2 \neq (\Sigma x)^2$

$\Sigma f x^2 \neq (\Sigma f x)^2$

B. Se todos os valores são idênticos

3, 3, 3, 3, 3, 3

● $\Sigma x = 24$ $\bar{x} = 3$

●● $s = 0$, $s^2 = 0$

Não há dispersão ● Não há variabilidade

COEFICIENTE DE VARIAÇÃO

$$\bullet\ C.V. = \frac{s}{\bar{x}} \times 100 = \%$$

Exemplo:

$s = 4{,}98$; $\bar{x} = 15{,}00$

$$\bullet\ C.V. = \frac{4.98}{15} \times 100 = 33.2\ \%$$

ESTATÍSTICA EXPERIMENTAL

MEDIDAS DE DISPERSÃO

B. DESVIO PADRÃO

Conceito :

$$s = \sqrt{\frac{\Sigma\,(x - \bar{x})^2}{n}} \qquad n > 30$$

$$s = \sqrt{\frac{\Sigma\,(x - \bar{x})^2}{n-1}} \qquad n < 30$$

Trabalho :

$$s = \sqrt{\frac{\Sigma\, x^2 - \frac{(\Sigma x)^2}{n}}{n}} \qquad n > 30$$

$$s = \sqrt{\frac{\Sigma\, x^2 - \frac{(\Sigma x^2)}{n}}{n-1}} \qquad n < 30$$

O desvio padrão e a variância representam a dispersão do conjunto de informação. A variabilidade da informação pode comparar-se em termos de desvio padrão:

x_1	x_2
3	4
4	9
5	2
6	1
7	10 ←
8	7

• $\Sigma\, x_1 = 33$ $\quad\Sigma\, x_2 = 33$

•• $\bar{x}_1 = 5.5$ $\quad\bar{x}_2 = 5.5$

••• $s = 1.87$ $\quad s = 3.72$

MEDIDAS DE DISPERSÃO

C. Variância

É o quadrado do desvio padrão

- Conceito = $s^2 = \dfrac{\Sigma\,(x - \bar{x})^2}{n - 1}$ $n < 30$

- Trabalho = $s^2 = \dfrac{\Sigma x^2 - \dfrac{(\Sigma\, x)^2}{n}}{n - 1}$ $n < 30$

D. Cálculos vários

Procedimento para calcular s, usando o método da desviação média.

X	X - $\bar{X}$	$(X - \bar{X})^2$
9	+4	16
8	+2	9
7	+2	4
7	+2	4
7	+2	4
5	0	0
5	0	0
5	0	0
5	0	0
4	-1	1
4	-1	1
3	-2	4
3	-2	4
2	-3	9
1	-4	16
$\Sigma x = 75$	$\Sigma(x - \bar{x}) = 0$	$\Sigma(x - \bar{x})^2 = 72$

N = 15, $\bar{x} = 5$

$$S = \sqrt{\frac{\Sigma\,(x - \bar{x})^2}{n - 1}} \quad *$$

$$S = \sqrt{72/14}$$

$$S = \sqrt{5,14}$$

$$S = 2,27$$

*Para dados apresentados em forma de distribuição, a fórmula do desvio padrão é:

$$S = \sqrt{\frac{f\,(x - \bar{x})^2}{n - 1}} \qquad N \leqslant 30$$

Observe que a freqüência que aparece na fórmula, é para recordar que cada $(x - \bar{x})^2$ se deve multiplicar por sua correspondente freqüência antes de ser somada. Ainda que utilizemos uma série de dados não classificados.

Outra forma equivalente de cálculo, só quando n > 30 é:

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} \quad (1)*$$

(1) esta forma é equivalente a as outras 2, quando n>30. n>30

DADOS CLASSIFICADOS

Procedimentos para calcular desvio padrão a partir das distribuições de freqüência, usando o método de dados originais.

1 Intervalo de classe	2 f	3 X	4 fX	5 fX^2
25 - 28	1	27	27	729
23 - 25	4	24	96	2304
20 - 22	7	21	147	3087
17 - 19	12	18	216	3888
14 - 16	18	15	270	4050
11 - 13	11	12	132	1584
8 - 10	9	9	81	729
5 - 7	3	6	18	108
2 - 4	1	3	3	9

N = 66 $\quad \Sigma fX = 990 \quad \Sigma fX^2 = 16488$

$$\bar{X} = \frac{990}{66} = 15{,}00$$

$$s = \sqrt{\frac{\Sigma fx^2}{N} - \bar{X}^2}$$

$$s = \sqrt{\frac{16488}{66} - 15^2}$$

$$s = \sqrt{249{,}82 - 225{,}00}$$

$$s = \sqrt{24{,}82}$$

$$s = 4{,}98$$

Resumindo o cálculo do desvio padrão a partir de uma distribuição de freqüência:

Passo 1. Seguir todos os passos necessários para calcular a média a partir de dados agrupados.

Passo 2. Agregar uma coluna adicional, fX^2.

Passo 3. Multiplicar os valores da coluna fX por os correspondentes valores da coluna X e por os resultados na coluna fX^2.

Passo 4. Somar os valores da coluna fX^2 para obter ΣfX^2.

Passo 5. Substituir os valores ΣfX^2, N, $\bar{X}$, na fórmula e resolver as operações indicadas para encontrar o desvio padrão.

2.5 - Análise de Correlação

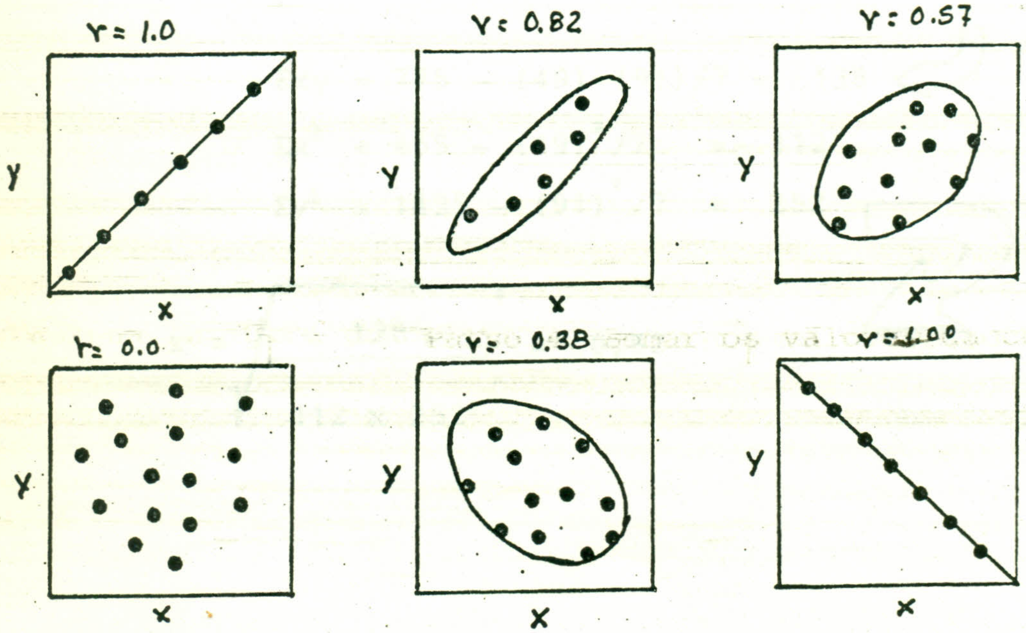
A. Conceito

- Freqüentemente é necessário determinar as relações entre dois ou mais variáveis independentes.
- A expressão quantitativa do grau de relação entre as variáveis é o coeficiente de correlação.

CÁLCULO

$$r = \frac{\Sigma\, xy - (\Sigma\, x)\ (\Sigma\, y)/n}{\sqrt{[\Sigma\, x^2 - (\Sigma\, x)^2/n\ ] \bullet [\Sigma\, x^2 - (\Sigma\, y)^2\ /n]}}$$

$$r = \pm \sqrt{\frac{\text{Variação explicada}}{\text{Variação total}}}$$



Procedimento de cálculo para o coeficiente r utilizando o método de soma de quadrados

INDIVÍDUO	X	Y	X^2	Y^2	XY
A	1	7	1	49	7
B	3	4	9	16	12
C	5	13	25	169	65
D	7	16	49	256	112
E	9	10	81	100	90
F	11	22	121	484	212
G	13	19	169	361	247
	$\Sigma x = 49$	$\Sigma y = 91$	$\Sigma x^2 = 455$	$\Sigma y^2 = 1435$	$\Sigma xy = 775$

$n = 7$

$$r = \sqrt{\frac{\Sigma xy - \Sigma x \Sigma y / n}{(\Sigma x^2 - (\Sigma x)^2/4) \bullet (\Sigma y^2 - (\Sigma y)^2/4}} \approx \sqrt{\frac{\Sigma xy}{(\Sigma x^2) \bullet (\Sigma y^2)}}$$

$$\Sigma xy = 775 - (49)\ (91)/7 = 138$$

$$\Sigma x^2 = 455 - (49)^2/7 = 112$$

$$\Sigma y^2 = 1435 - (91)^2/7 = 252$$

$$r = \sqrt{\frac{138}{112 \times 252}} = 0{,}82$$

ESTATÍSTICA EXPERIMENTAL

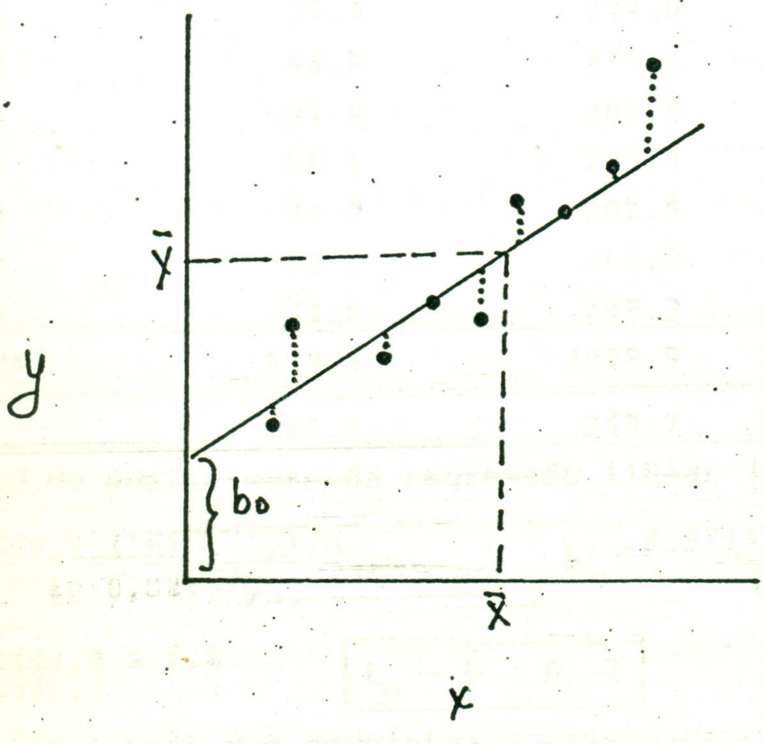
2.6 - Regressão Linear

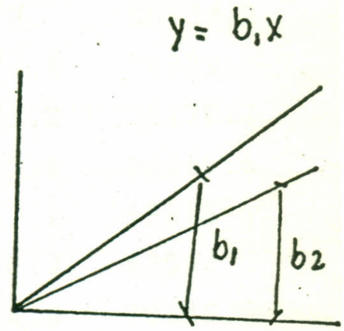
A. Cálculo

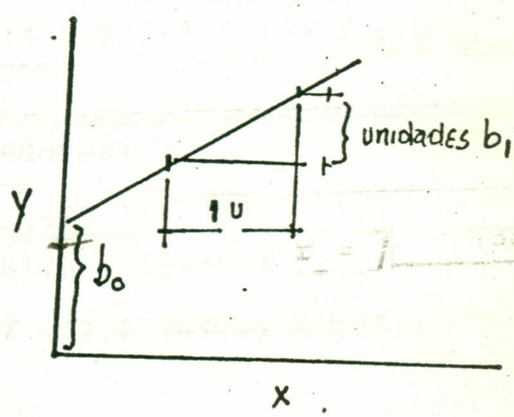
$y = b_o + b_1 x$ • equação da reta

$$b_1 = \frac{\Sigma\, xy - \Sigma x\, \Sigma y/n}{\Sigma\, x^2 - (\Sigma\, x)^2/n}$$

• $b_o = \bar{y} - b_1\, \bar{x}$







Exemplo: Cálculo regressão linear

Em um experimento de colheita foi determinado o número de frutos soltos antes da colheita e o número de frutos soltos total. Se deseja saber si o número total de frutos desprendidos do cacho na colheita (y) depende do número de frutos soltos antes da colheita (x). (Cultura de Dendê).

Nº de cachos observados	Média de frutos soltos antes da colheita (x)	Frutos soltos total (y)	xy	x^2	y^2
40	57.0	230.9	13161.3	3249	53314.8
40	45.4	208.6	9470.4	2061.2	43514.0
50	60.7	233.9	14197.7	3684.5	54709.2
60	59.2	256.7	15196.6	3504.6	65894.9
35	45.9	197.5	9065.3	2106.8	39006.3
42	61.2	209.6	12827.5	3745.4	43932.2
40	62.1	236.4	14680.4	3856.4	55885
45	69.7	259.8	18108.1	4858.1	67496
46	73.0	294.6	21505.8	5329	86789.2
50	63.4	277.0	17561.8	4079.6	76729
55	64.6	279.6	18062	4273.2	78176.2
56	91.9	300.8	27643.5	8445.6	90480.6
40	56.1	236.4	13262.0	3147.2	55885
42	42.9	208.6	8948.9	1840.4	43514
40	52.2	265.8	13874.8	2724.8	70649.6
50	74.0	299.3	22148.2	5476	89580.5
SOMA	979.3	3995.5	249714.5	62281.8	1015556.5
$\bar{X}$	61.2	249.7			

Cálculo do coeficiente da regressão linear (esimadores).

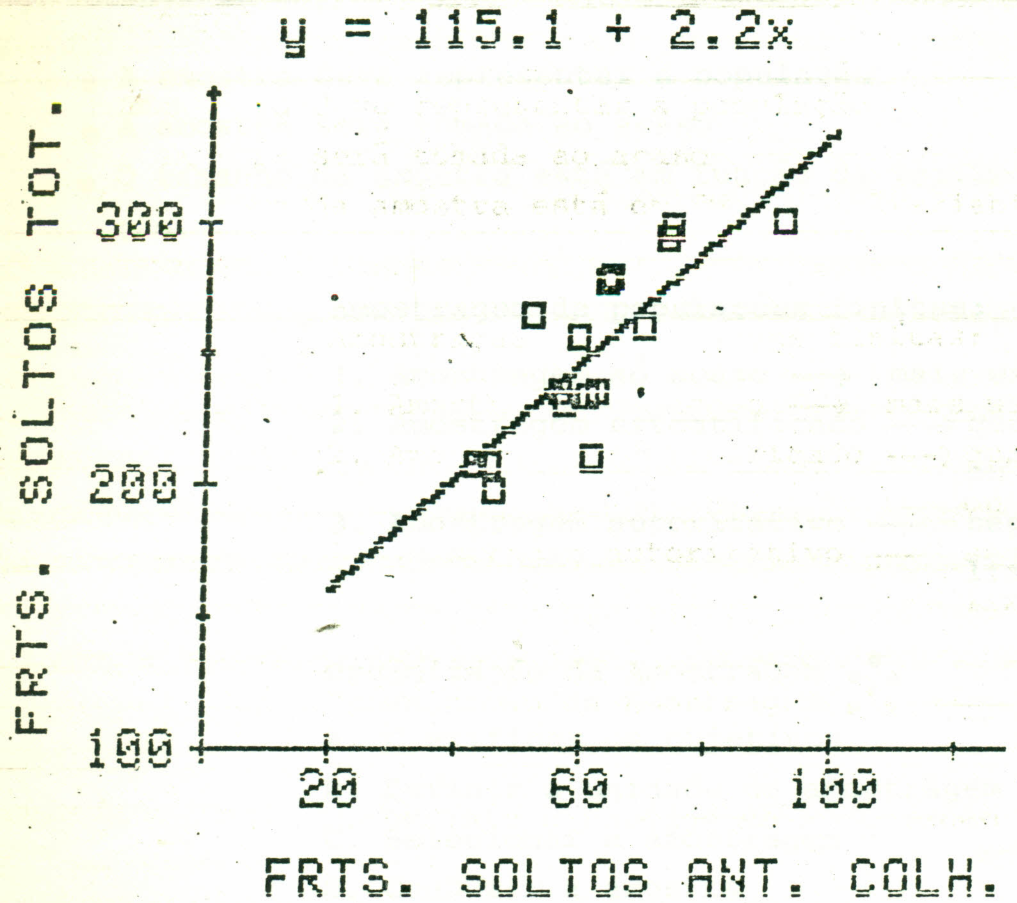
$$b_1 = \frac{\Sigma xy - (\Sigma x)\quad(\Sigma y)/n}{\Sigma x^2 - (\Sigma x^2)/n}, \qquad b_1 = \frac{249714.5 - (979.3)\quad(3995.5)/16}{62281.8 - 59939.3}$$

$$b_1 = \frac{5164.8}{2342.5} = 2.2 \qquad \boxed{b_o = \bar{y} - b_1\,\bar{x}} \qquad 249.7 - 2.2\ (61.2) = 115.1$$

A equação linear é a seguinte: $\boxed{y = 115.1 + 2.2x}$

Com esta equação podemos, calcular qualquer valor de "y" (frutos soltos total) para qualquer valor de "x", frutos soltos antes da colheita. (ver gráfico).

EXEMPLO REGRESSÃO



Nota: Existem testes de hipóteses para verificar se o coeficiente b, é significativo ou não.

ESTATÍSTICA EXPERIMENTAL

AMOSTRAGEM

3.1 - CONCEITOS

- A amostra deve representar a população
- A amostra será tomada ao acaso
- O tamanho da amostra está em função da variabilidade da informação.

Amostragem de populações finitas:

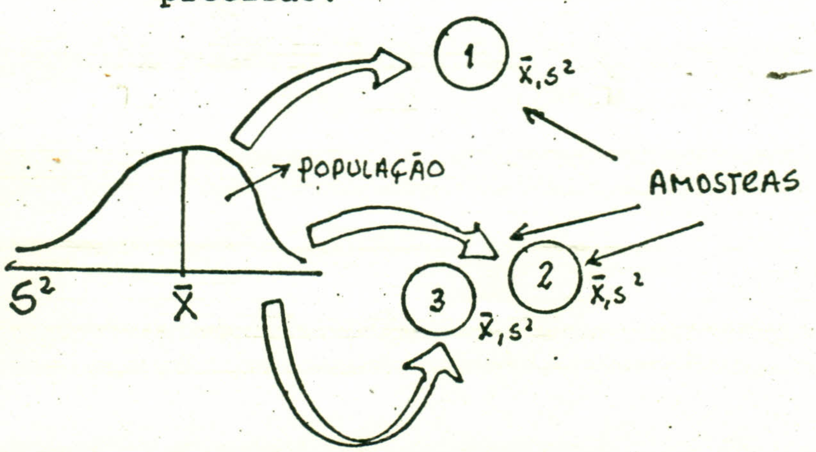
1. Amostragem ao acaso ⟶ mais usado
2. Amostragem estratificado ⟶ quando existam grupos ou estratos.
3. Amostragem autoritativo — segurança de quem realiza a amostragem.

Organização da Amostragem ∴

A. Clarificar os objetivos
B. Definir a unidade de amostragem e a população
C. Selecionar a amostragem
D. Conduzir a amostra
E. Analizar a informação

Tamanho da amostra ::: depende em parte de:

- Fundos ou financiamento. Quanto maior o tamanho da amostra o custo é mais alto. Menor tamanho ⟶ menor precisão.



ESTATÍSTICA EXPERIMENTAL

DEFINIÇÕES:

- Marco: É a lista organizada de todos os componentes da população; e a apresentação organizada do conjunto universal definido pela variável Y_i

$$y_i = y_1, y_2, y_3, \ldots\ldots\ldots\ldots y_N$$

- Amostra : É a coleção de n unidades obtidas do marco anterior.

$$y_i = y_1, y_2, y_3, \ldots\ldots\ldots y_n$$

ANÁLISE DA INFORMAÇÃO:

- Estimadores:

  - Média: $\bar{y} = \frac{1}{n} \Sigma y_i$

  - Variância: $S^2 = \frac{\Sigma y_i^2 - (\Sigma y_i)^2/n}{n-1}$

  - Variância das Médias: $S^2_{\bar{y}} = \frac{S^2}{n} \cdot \frac{N-n}{N-1}$ N = Tamanho população n = Tamanho amostra

  - Desvio padrão da Média: $S^2_{\bar{y}} = S^2$

ESTATÍSTICA EXPERIMENTAL

ESTIMADOS

Total da população $\hat{y} = N\bar{y}$ N = Tamanho da população finita

Média população $\hat{\bar{y}} = \hat{y}$

Variância população $s^2\hat{y} = S^2\bar{y}\, N^2$

Erro padrão população $s\bar{y} = S^2\bar{y}\quad N$

PARÂMETROS

Total $y = \Sigma yi$

Média $\bar{y} = \dfrac{1}{N}\ \Sigma yi$

Variância $\delta^2 = \dfrac{\Sigma yi^2 - (\Sigma yi)^2/N}{N}$

CONFIABILIDADE

• Intervalo de confiança para estimar a média a partir de uma amostra.

Conceito: a partir da amostra

Probabilidade: $\left\{ \bar{y} - t_\alpha s_{\bar{y}} \leqslant \bar{y} \leqslant \bar{y} + t_\alpha s_{\bar{y}} \right\} = 1 - \alpha$

• $\bar{y} \pm t_\alpha s_{\bar{y}}$

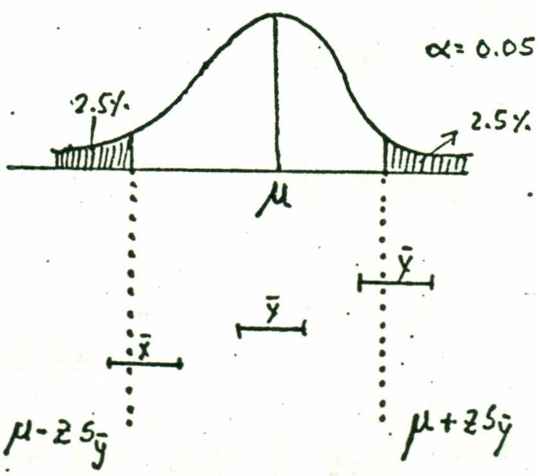
$s_{\bar{y}}$ = Variabilidade da amostra
s = desvio padrão
$s_{\bar{y}}$ = erro padrão

$$s_{\bar{y}} = \frac{s}{\sqrt{n}}$$

t_α ou z = Valor de probabilidade a confiança para n - 1 observações ma amostra.

• Quando se conhece a média da população (μ) y δ^2 o intervalo de confiança é:

• $\bar{y} - zs_{\bar{y}} \leqslant \mu \leqslant \bar{y} + zs_{\bar{y}}$



AMOSTRAGEM

Exemplo:

Seja a seguinte amostra:

$y_i = 4 + 6 + 8 + 7 + 3$

$\bar{y} = 5.6 \qquad s = 2.07$

Quais são os intervalos de confiança desta média para P = 0.05 ?

$\bar{y} \pm tS_{\bar{x}}$

$\bar{y} = 5.6$

$t_{0.05}\ (GL = 4) = 2.77$

$S_{\bar{y}} = \frac{s}{n} = \frac{2.07}{2.24} = 0.92$

$\bar{y} \pm 2.776 = 0.92$

$\bar{y} \pm 2.55 \longrightarrow \bar{x} = 5.6 \pm 2.55 \longrightarrow$ amplitude: 3.1 a 8.1

ESTATÍSTICA EXPERIMENTAL

AMOSTRAGEM

Tamanho da amostra e precisão

Conceito::

• $d^2 = t^2 \dfrac{s^2}{n}$

t = 95% = n − 1 G.L.
99% = n − 1 G.L.

d = precisão
t = probabilidade
s^2 = variância

Tamanho da amostra (população infinita)

• $n = \dfrac{t^2 \; s^2}{d^2}$ vem de:: $d^2 = t^2 \dfrac{s^2}{n}$

$$\frac{d^2}{t^2} = \frac{s^2}{n}$$

$$\sqrt{\frac{d^2}{t^2}} = \sqrt{\frac{s^2}{n}}$$

$$\frac{d}{t} = \frac{s}{\sqrt{n}}; \quad \sqrt{n} = \frac{st}{d}; \quad n = \frac{(st)^2}{d^2}$$

PARA POPULAÇÕES FINITAS

$$n = \frac{1}{\dfrac{d^2}{t^2 s^2} + \dfrac{1}{N}}$$

N: Tamanho população

ESTATÍSTICA EXPERIMENTAL

AMOSTRAGEM

3.2- Amostragem inteiramente ao acaso

USOS:
- Nos casos de relativa homogeneidade.
- Não existe outra fonte de variabilidade além da atribuível ao erro de amostragem.

EXEMPLO

- Vamos supor que desejamos avaliar uma floresta de *Pinus* sp., que se estende em forma homogênea em 2.000 ha.
- O objetivo da pesquisa é estimar o volume médio de madeira por ha e o volume total de madeira nessa floresta.
- A unidade de amostragem é um hectare, sobre a qual se avalia o volume de madeira.

- MÉTODO:
  - O número de ha possíveis de serem amostrados é N = 2.000
  - Em forma aleatória amostramos N= 20 ha ao longo da floresta
  - Os resultados da avaliação se apresentam à seguir:

Volume Madeira em m^3/unidade amostrada (ha)

Unidade	Volume	Unidade	Volume
1	8	11	9
2	7	12	8
3	6	13	7
4	9	14	6
5	10	15	9
6	7	16	10
7	8	17	6
8	9	18	7
9	6	19	8
10	8	20	8

N = 2.000 ; n = 20

- ESTIMADORES:

Média $\bar{y} = 7,8 \ m^3/ha$

Variância $s^2 = 1.6421$

Variância das médias $= \dfrac{1.6421}{20} \cdot \dfrac{2.000 - 20}{2.000 - 1} = 0.0213$

Erro padrão da média $= 0.0813 = 0.285$

ESTIMADOS:

Média $\bar{y} = 7.8\ m^3/ha$

Total $y = (7.8)\ (2000) = 15.600\ m^3$

Erro padrão $s_{\bar{y}} = (0,285)\ (2000) = 570$

LIMITES DE CONFIANÇA

- Para os parâmetros total (y) e a média $(\bar{y})$

  - $y \pm s_{\bar{y}} t$ $\left\{ \begin{array}{l} \hat{y} = 15.600\ m^2,\ \hat{\bar{y}} = 7,8 \\ s_{\bar{y}} = 570 \quad ,\ s_{\bar{y}} = 0,285 \\ t\ (20-1 = 19GL)\ e\ P = 0,05 = 2.093 \end{array} \right.$

  - $y \pm (570)\ (2.093) \longrightarrow y \pm 1193.01$

A um nível de 95% de probabilidade o volume total de madeira estará na seguinte amplitude:

$$14406\ m^3 \leqslant y \leqslant 16793\ m^3$$

- Enquanto que os limites de confiança da média são:

$$\bar{y} \pm (0.285)\ (2.093) \longrightarrow \bar{y} \pm 0.5965$$

Logo, a média estará na amplitude:

$$7.20\ m^3 \leqslant \bar{y} \leqslant 8.40\ m^3$$

- Para outras florestas de condições similares, podemos estimar o número de ha a serem avaliadas, considerando o nível de variabilidade encontrada de $s_{\bar{y}} = 0.285$ a um nível de probabilidade $P = 0,05$ e para uma margem de erro das estimativas de $d = \pm 0,2 m^3$

$$n = \frac{1}{\dfrac{d^2}{t^2\ s^2_{\bar{y}}} + \dfrac{1}{N}} \qquad \left\{ \begin{array}{l} d^2 = 0,2^2 = 0,04 \\ t^2 = 2.093^2 = 4.38 \\ s^2_{\bar{y}} = 0,285^2 = 0.0812 \end{array} \right.$$

$$n = 8.85 \cong 9$$

Se a floresta mostrar uma variabilidade semelhante a encontrada, será necessário avaliar apenas 9 ha, ou seja menos da metade do número de ha da amostra inicial.

ESTATÍSTICA EXPERIMENTAL

AMOSTRAGEM

EXEMPLOS

• Se um número grande de frangos (N = grande). A variância estimada do peso médio (ℓb) é: $S^2 = 600$ em base a uma amostra de 15 frangos (N=15).

•• Se deseja estimar o peso médio por frango dentro $\pm$ 5 lb com uma confiança de P =, 0.05

Valor probabilidade: $t_{0.05}$ para 14 G.L. = 2.145 (da tabela)

••• Qual tamanho de amostra será necessária para uma precisão de $\pm$ 5ℓb

$$n = \frac{(t\ 0.05\ s^2)}{d^2}$$

$t_{0.05} = 2.145$ (da tabela)

$s^2 = 600$

$d^2 = (5)^2 = 25$

$$n = \frac{(2.145) \times 600}{25} = \frac{4.6 \times 600}{25} = 110 \text{ frangos}$$

•• Se si trata de uma população de 200 frangos, se tem portanto, •• uma população finita.

n = 200 $\longrightarrow$ população finita

$$n = \frac{1}{\frac{d^2}{t^2 s^2} + \frac{1}{N}} \quad ; \quad N = \frac{1}{\frac{2.5}{4.6 \times 600} + \frac{1}{200}} = 71$$

Notar a diferença com •••

$$N = 2000,\ n = \frac{1}{\dfrac{25}{4.6 \times 600} + \dfrac{1}{2000}} = 105$$

$$N = 20{,}000\ n = \frac{1}{\dfrac{25}{4.6 \times 600} + \dfrac{}{20{,}000}} = 109$$

Nota: Com o aumento do tamanho da população finita, esta se acerca a uma população grande (definita).

ERROS DE AMOSTRA ::::

$\theta - \hat{\theta}$ = ERRO = precisão = d

θ = valor real

$\hat{\theta}$ = valor estimado

LEMBRAR:

A amostragem nem sempre é de conceito, as vezes pode-se realizar uma análise lógica com outras técnicas. Vide o exemplo a seguir.

ESTATÍSTICA EXPERIMENTAL

Controle de qualidade de cachos de dendê em carretas que entram na Usina.

Dados Observados (cachos)

CARRETA	VERDES P	VERDES M	MADUROS P	MADUROS M	PODRES P	PODRES M	TOTAL P	TOTAL M
30	13	4	103	21	0	0	116	25
40	26	6	81	19	2	0	109	25
78	20	5	95	25	4	0	119	30
46	12	5	87	25	4	2	103	32
12	21	6	108	24	1	0	130	26
70	6	1	99	29	4	1	109	31
55	13	3	86	22	15	5	114	30
16	9	1	56	32	7	1	112	34
10	9	5	117	20	4	1	130	26
92	11	3	109	29	3	1	123	33
52	15	6	100	25	0	0	115	31

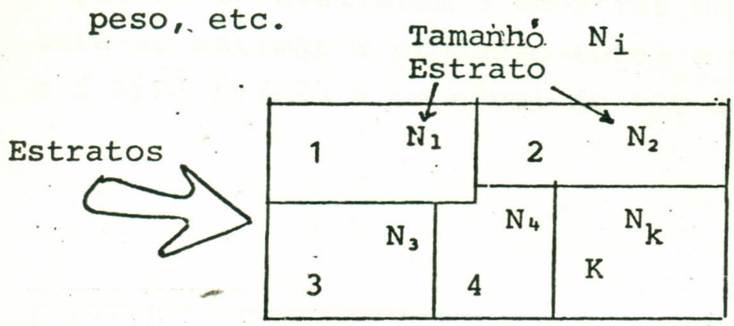
Dados em porcentagem

CARRETA	VERDES P	VERDES M	VERDES E	MADUROS P	MADUROS M	MADUROS E	PODRES P	PODRES M	PODRES E
30	11.2	16.0	+ 4.8	88.8	84.0	- 4.8	0	0	0
40	23.9	24.0	+ 0.1	74.3	76.0	+ 1.7	1.8	0	-1.8
78	16.8	16.7	- 0.1	79.8	83.3	+ 3.5	3.4	0	-3.4
46	11.6	15.6	+ 4.0	84.5	78.1	- 6.4	3.9	6.3	+2.4
12	16.1	20.0	+ 3.9	83.1	80.0	- 3.1	0.8	0	-0.8
70	5.5	3.2	- 2.3	90.8	93.5	+ 2.7	3.7	3.3	-0.4
55	11.4	10.0	- 1.4	75.4	73.3	- 2.1	13.2	16.7	+3.5
16	8.0	2.9	- 5.1	85.7	94.1	+ 8.4	6.3	3.0	-3.3
10	6.9	19.2	+12.3	90.0	76.9	-13.1	3.1	3.9	+0.8
92	8.9	9.1	+ 0.2	88.6	87.9	- 0.7	2.5	3.0	+0.5
52	13.0	19.3	+ 6.3	87.0	80.7	- 6.3	0	0	0
	P = população	M = amostra	E = erro da amostra						
Correlação	r = 0.723*			r = 0.577			r=0.918**		*(P=0.05)
		Combinado	r = 0.992**						**(P=0.01)

3.3- Amostragem Estratificada

Estratificação: É o processo de dividir a população de acordo com:

- Características Naturais: Clima, solo, topografia, diferenças entre plantas quanto ao crescimento, etc.
- Divisões político-administrativas: Municípios, Comunidades, Fazendas, etc.
- Ou de maneira artifical em: Classes por tamanho, idade, peso, etc.



N = Tamanho população

Em forma separada para cada Estrato, se estimam os parâmetros da mesma forma que para a amostragem inteiramente ao acaso, ou seja:

$J = \{ 1,2..., k \}$ = Estratos

$\bar{Y}_j = \frac{1}{nj} \Sigma Y_i$ Média/Estrato

- $S_j^2 = \frac{\Sigma Y_j^2 - (\Sigma Y_j^2/_{nj})}{n - 1}$ Variáveis/Estrato

- $S^2_{\bar{y}_j} = \frac{S_j^2}{n_j} \cdot \frac{(N_j - n_j)}{(N_j - 1)}$ Variância das médias/Estratos.

N população $= \Sigma N_j$

ESTATÍSTICA EXPERIMENTAL

ESTIMADOS:

- Total $\hat{Y} = \Sigma \bar{Y}_j N_j$

Erro Padrão: $S\hat{y} = \Sigma S\bar{y}\, Nj$

Exemplo:

Vamos supor uma população de 200 árvores jovens de pupunha. Nota-se que a área apresenta três estratos bem diferenciados quanto a topografia do terreno: levemente inclinado; inclinado e plano. Se avaliaram 3 amostras de tamanho Nj = {5,4,6}. Procura-se estimar a produção total e média de palmito por planta e o erro padrão a um nível de 90% de probabilidade:

Estrato	Nj	nj	Amostras (kg/planta/pl)	Nj/N	$\bar{y}_i$
1	40	5	{1.6, 1.5, 1.4, 1.5, 1.7}	0,20	1,54
2	50	4	{1.2, 1.1, 1.0, 0.8}	0,25	1,02
3	110	6	{2.2, 1.9, 2.3, 1.8, 2.4, 2.0}	0,55	2,10
	N = 200	n= 15		Σ = 1,0	

Continuação ...

Sj	$S\bar{y}j$	$\bar{Y}j$ (Nj/N)	$S\bar{y}_j$ (Nj/N)	$\bar{Y}j$ (Nj)	$S\bar{y}_j$ (Nj)
0,114	0,048	0,31	0,0096	61,6	0,384
0,171	0,082	0,25	0,0207	51,0	1,035
0,236	0,094	1,15	0,0517	231,0	5,687
	Soma:	1,71	0,082	343,6	6,760

De acordo com os resultados da nossa amostragem temos:

Média de palmito/planta: 1.71 kg

Total de palmito na população: 343,6 kg

Desejamos conhecer a confiabilidade destas estimações a um nível de 90% de probabilidade (P = 0,010).

$\bar{y} \pm s_{\bar{y}} t \longrightarrow$ Média: $1.71 \pm (0.082) \cdot (1.782)$

$1.71 \pm 0.15 \longrightarrow 1.56 \text{ kg} \leqslant \bar{y} \leqslant 1.86 \text{ kg}$

$\hat{y} \pm s_{\hat{y}} t \longrightarrow$ Total: $343.6 \pm (6.760)\ (1.782)$

$343.6 \pm 12.0 \longrightarrow 331.6 \text{ kg} \leqslant \hat{y} \leqslant 355.6 \text{ kg}$

NOTA: (Para o valor de t da tabela, para n-3 G.L, um por estrato): n = 15; GL: 15-3 = 13

TAMANHO DA AMOSTRA DENTRO DE CADA ESTRATO

Antes de amostrar uma população seguindo o método de amostragem estratificada; temos que decidir quantas amostras serão coletadas em cada estrato:

- Quando não se dispõe de nenhuma informação prévia: O tamanho da amostra n_j é proporcional ao tamanho do estrato, ou seja:

$$n_j = n(N_j/N) \quad \begin{cases} n_j = \text{Tamanho amostra/estrato} \\ n = \text{Tamanho amostra total/população} \\ N_j = \text{Tamanho estrato} \\ N = \text{Tamanho população} \end{cases}$$

Considerando todos os estratos temos:

Estrato	N_j	Fração $N_j N$	Tamanho n_j
I	N_1	N_1/N	$n_1 = n(N_1/N)$
II	N_2	N_2/N	$n_2 = n(N_2/N)$
:	:	:	: :
K	N_k	N_k/N	$n_k = n(N_k/N)$
	$\Sigma N_j = N$		$\Sigma n_j = n$

Por exemplo temos uma população estratificada em 4 estratos. Se dispõe de recursos para coletar apenas 30 amostras. Quantas amostras coletaremos por estrato.

Estrato	Tamanho (N_j)	Fração	Número amostras/Estrato	
1	50	0.13	3.9 $\longrightarrow$	4
2	140	0.38	11.4	11
3	80	0.22	6.6	7
4	100	0.27	8.1	8
	N = 370	1.0	n = 30	

As vezes se dispõe de informação acerca da variabilidade dentro dos estratos por amostragens ou estimativas prévias. Nesta situação o tamanho da amostra e proporcional a variância (variabilidade) e ao tamanho do estrato:

$$n_j = n(S_j^2)\ (N_j/N)$$

Ou seja:

Estrato	Nj	S^2_j	$(N_j/N)\ S_j^2 = f_j$	n_j	
I	N_1	S_1^2	$(N_1/N)\ S_1^2 = f_1$	n_1	$= nf_1$
II	N_2	S_2^2	$(N_2/N)\ S_2^2 = f_2$	n_2	nf_2
:	:	:	:	:	:
K	N_k	S_k^2	$(N_k/N)\ S_k^2 = f_3$	n_k	nf_k
$\Sigma N_j = N$				$\Sigma nj = n$	

Com as informações sobre os estratos desejamos coletar 15 amostras:

Estrato	Tamanho	Variância	Fração	Tamanho amostra	
1	40	1.29	0.258	3.87 →	4
2	50	1.66	0.415	6.225	6
3	110	1.84	1.01	15.15	15
	N = 200				n = 26

Notar que o tamanho da amostra total subiu de 15 para 25, isto em função da variabilidade observada e o tamanho do estrato.

TAMANHO AMOSTRA TOTAL (n)

Podemos também estimar o tamanho da amostra total n de forma proporcional a variânci e introduzindo um fator de precisão D, que é determinado pelo pesquisador:

$$n = \frac{N\, \Sigma N_j\ S_j^2}{N^2 D^2 + \ N_j\ S_j^2}$$

N = tamanho população
N_j = tamanho estrato
S_j^2 = variância/estrato
D = precisão ±

Considerando nosso exemplo de pupunha temos:

Estrato	Tamanho (N_j)	Variância (S^2_j)	Proporção ($N_j\ S^2_j$)
1	40	0.012	0.48
2	50	0.029	1.45
3	110	0.055	6.05
	N = 200		7.98

Desejamos estimar os parâmetros a uma precisão de ± 0,05kg(50g). Portanto o tamanho de amostra total a ser coletada nessa população estratificada será:

$$n = \frac{200\ (7.98)}{(200^2)(0.05)^2 + 7.98} = 14.78 \longrightarrow 15 \text{ árvores}$$

Se aumentarmos a precisão D para ± 0.025 kg (25g) teremos:

$$n = \frac{200\ (7.98)}{(200^2)(0.025)^2 + 7.97} = 48 \text{ árvores}$$

ESTATÍSTICA EXPERIMENTAL

ELEMENTOS DE ESTATÍSTICA (1981)

4.1 - Análises de Contagem

Basicamente se refere a análise de variáveis discretas como: Número de plantas de vacas, de casas, etc.

Temos unicamente a aplicação do teste de X^2, para determinar o grau de associação entre duas variáveis, que podem ser um tratamento e sua resposta. (Tabelas contigência 2 x 2).

Hipóteses Ø nula: unicamente levanta: que as variáveis são independentes.

Hipóteses alterna: que as variáveis estão associadas ou relacionadas.

Conceito $X^2 = \dfrac{\Sigma\ (o - e)^2}{e}$ $\begin{cases} o = \text{dados observados} \\ e = \text{dados esperados} \end{cases}$

A doença que produz secamento das ponteiras do Freijó (essência florestal) em Manaus, não é letal.

Se costuma realizar tratamentos com fungicidas para ajudar na recuperação das plantas. O pesquisador deseja saber si a aplicação do fungicida está associada com a recuperação das árvores, ou as árvores se recuperam expontaneamente?

Para testar esta hipótese, se trataram algumas árvores, deixando outros sem tratar (testemunhas).

Os resultados foram:

Tratamento	Recuperadas	Mortas	Total
Fungicida	32	2	34
(Esperado)	(29)	(5)	
Não tratadas	36	9	45
(Esperado)	(39)	(6)	
Total	68	11	79

Se as duas variáveis são independentes, a proporção esperada de plantas recuperadas respeito ao total será igual a proporção de plantas tratadas com fungicida; respeito ao total:

$\left(\frac{68}{79} \times \frac{34}{79}\right)79 = 29$ plantas esperamos que se recuperem com o tratamento.

∴ Com este mesmo raciocínio .. continua:

A proporção esperada de plantas recuperadas respeito ao total será igual a`proporção de plantas não tratadas respeito ao total:

$$\left(\frac{68}{79} \times \frac{45}{79}\right)79 = 39$$

O restante do quadro se completa por diferença.

34 - 29 = 5 ::::

45 - 39 = 6 ::::

Aplicando o conceito:

$$x^2 = \frac{\Sigma(o - e)^2}{e} \qquad (r - 1) \times (c - 1) \qquad \begin{matrix} r = \text{Fileiras} \\ c = \text{Colunas} \end{matrix}$$

$$x^2 = \frac{(32 - 29)^2}{29} + \frac{(2 - 5)^2}{5} + \frac{(36 - 39)^2}{39} + \frac{(9 - 6)^2}{6}$$

$$x^2 = 3.841$$

x^2_t = das tabelas (p = 0.05) (2-1) = 1 G.L.

Probabilidade

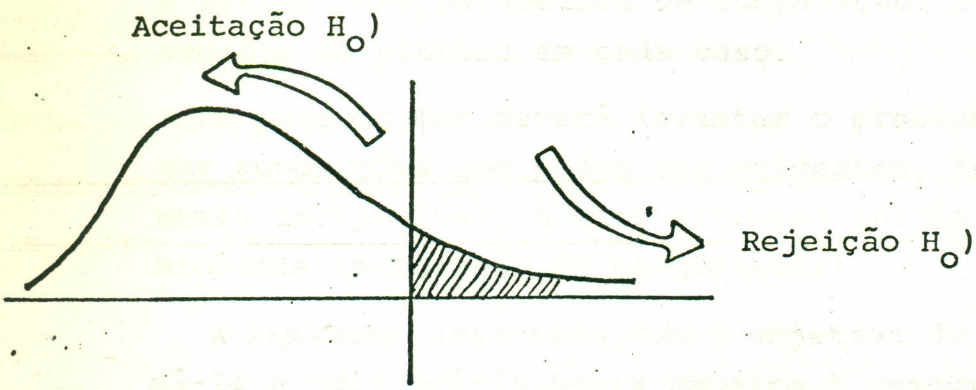
0.05

Graus de liberdade

$$1 \longrightarrow 3.841$$

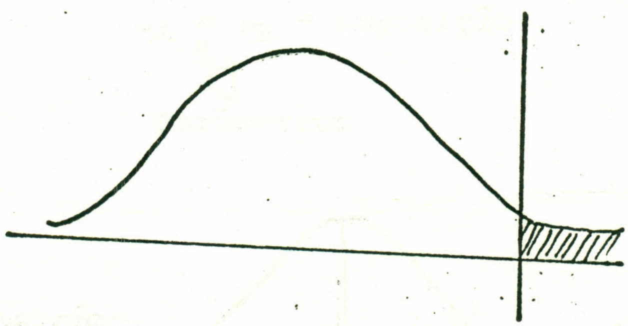
Decisão: O valor calculado $x^2 = 3.841$ é igual ao valor da tabela, por tanto se aceita H_o).

Conclusão: As variáveis são independentes.



$x^2_t = 3.841$ (1GL) 0.05 • tabelas

$x^2_c = 3.841$ • calculado



$x^2_t = 6.635$ (1GL) 0.01

4.2 - TESTE DE HIPÓTESE (CONCEITOS)

É muito comum a necessidade de tomar alguma decisão a partir de uma informação obtida por amostragem. Por exemplo: desejamos comparar o crescimento de plantas de Seringueira com cobertura de puerária com plantas com cobertura natural nas suas entrelinhas, através da medição do diâmetro do caule e a altura como parâmetros de comparação. Suponhamos que foram medidas 25 plantas em cada caso.

O pesquisador deverá levantar o problema com base a algumas suposições que podem ser evidentes, falsas ou aparentemente inobjetivas. A esta situação inicial se conhece como hipótese (suposição do pesquisador).

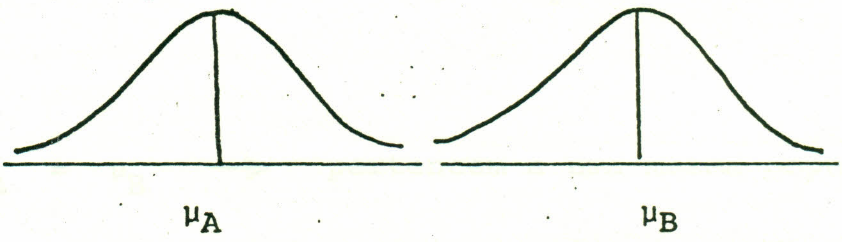
A hipóteses levantada com o objetivo de rejeitá-la, eliminá-la e nulificá-la.Desta maneira o pesquisador pode fazer pressuposição acerca dos parâmetros. Por exemplo, supor que a média de certo tratamento o seu efeito é igual ou diferente a outro tratamento.

Hipóteses:

$\mu_A = \mu_B \rightarrow$ suposição .

Parâmetros

Suposição:



$\mu_A - \mu_B = 0 \rightarrow$ hipótese nula

ESTATÍSTICA EXPERIMENTAL

Teste de hipótese

Para testes reais:

$H_O = \mu_A = \mu_B$

$H_A = \mu_A \neq \mu_B \rightarrow$ hipótese alterna (provas de 2 caudas)

Outro tipo de hipótese alterna:

$H_A)$ $\mu_A > \mu_B$

implica compromisso

H_A $\mu_A < \mu_B$

Recomenda-se não ter compromisso, como $\mu_A > \mu_B$, que denota parcialização. (provas de 1 cauda).

•• Também pode-se levantar uma hipótese, acerca de outros parâmetros, por exemplo, acerca da variabilidade:

$H_O)$ $\delta_A^2 = \delta_B^2$

$H_A)$ $\delta_A^2 \neq \delta_B^2$ ó $H_A)$ $\delta_A^2 > \delta_B^2$; $\delta_A^2 < \delta_B^2$

Entre a $H_O)$ e $H_A)$ uma delas é evidente, não podem ser ambas.

Por exemplo:

Se $\mu_A = \mu_B \rightarrow$ pertencem a uma mesma população

Se $\mu_A \neq \mu_B \rightarrow$ pertencem a diferentes populações.

ESTATÍSTICA EXPERIMENTAL

Teste de Hipóteses

LEMBRAR:

- Os testes de hipóteses se realizam com base as amostras.
- Uma decisão de êxito, depende de uma eleição acertada da distribuição a usar. (critério de prova).

Tabela A. Alguns critérios para eleger distribuições.

1. Para comparar duas amostras: Z ou "t"
2. Para comparar duas amostras: F ou X^2.

   F = número fixo de tratamentos

   X^2 = depende dos eventos do experimento.
3. Para amostras de tamanho n >30: Z
4. Para amostras de tamanho n <30: t
5. Se se conhece a variância da população: Z
6. Se não se conhece a variância da população: t

X^2 = ⟶ variáveis discretas

t, x, F ⟶ variáveis continuas

IMPORTANTE: A decisão está também em função de probabilidade. Em cada distribuição, a diferente nível de probabilidade e graus de liberdade, se tem valores críticos de significancia estadística calculados, que são a referência para a decisão final. (vide tabelas de Z,t,X^2, y F).

ESTATÍSTICA EXPERIMENTAL

Teste de Hipótese

Metodologia para o teste de hipótese

Se aplica o método científico e se deve definir claramente o seguinte:

1. População e parâmetros objetos da pesquisa
2. Hipótese estatística a ser testada
3. Distribuição estatística aplicável ao experimento
4. Obtenção de dados
5. Nível de significância estatística

Cálculos Estimadores

Teste "Z" para amostras maiores que 30 ou quando se conhece a variância da população.

$$Z_o = \frac{\bar{X}_A - \bar{X}_B}{\delta_{\bar{X}_A - \bar{X}_B}}$$

$\bar{X}_A$, $\bar{X}_B$ = Médias das amostras A e B

$\delta_{\bar{X}_A - \bar{X}_B}$ = Desvio padrão das diferenças entremédias.

Onde:

$$\delta_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\delta_A^2}{n_A} + \frac{\delta_B^2}{n_B}}$$

δ_A^2, δ_B^2 = Variâncias das amostras A e B

n_A; n_B = Tamanho das amostras A e B.

4.3 - Teste de "t"

Apresenta variações de acordo com os tamanhos das amóstras e a homogeneidade das variâncias:

1. Amostras de diferente tamanho e variâncias homogêneas

$$\boxed{n_a \neq n_b \;;\; s_a^2 \doteq s_b^2}$$

$$\therefore s_c^2 = \frac{(n_a - 1)s_a^2 + (n_b - 1)\,s_b^2}{(n_a - 1) + (n_b - 1)}$$

$$\therefore s_{\bar{x}_a} \doteq \bar{x}_b = \sqrt{\frac{s_c^2}{n_a} + \frac{s_c^2}{n_b}}$$

$$\boxed{t_o = \frac{\bar{X}_a - \bar{X}_b}{s_{\bar{x}_a} - \bar{x}_b}}$$

Graus de Liberdade G.L para t_t

$$t_t = (n_a - 1) + (n_b - 1) = G.L$$

s_c^2 = Variância combinada
s_a^2, s_b^2 = Variâncias amostras a, b
n_a, n_b = Tamanho amostras a, b
$s_{\bar{x}a-\bar{x}b}$ = Desvio padrão das diferenças
$\bar{X}_a$, $\bar{X}_b$ = Médias das amostras a, b
t_o = Valor de t calculado
t_t = Valor de t das tabelas.

2. Amostras de igual tamanho, variâncias homogêneas.

$$\boxed{n_a = n_b \;,\; s_a^2 = s_b^2}$$

$$s_{\bar{x}_a} - \bar{x}_b = \sqrt{\frac{s_a^2 + s_b^2}{n}}$$

$$\boxed{t_o = \frac{\bar{x}_a - \bar{x}_b}{s_{\bar{x}a} - \bar{x}_b}}$$

$$t_t = (n - 1) = G.L.$$

3. Amostras igual tamanho, e variâncias não homogêneas

$$\boxed{n_a = n_b,\ s_a^2 \neq s_b^2}$$

$$s_c^2 = \frac{s_a^2 + s_b^2}{2}$$

$$s_{\bar{x}_a - \bar{x}_b} = \sqrt{2\,\frac{s^2c}{n}}$$

$$\boxed{t_o = \frac{\bar{x}_a - \bar{x}_b}{s_{\bar{x}_a - \bar{x}_b}}}$$

$$t_t = 2\,(n-1)$$

4. Amostras de diferente tamanho e variâncias não homogêneas.

$$\boxed{n_a \neq n_b\ ,\quad s_a^2 \neq s_b^2}$$

$$t_g = \frac{w_a t_a + w_b t_b}{w_a + w_b}$$

$$w_a = \frac{s_a^2}{n_a}\ ,\quad w_b = \frac{s_b^2}{n_b}$$

t_g = valor de t gerado

$w_a w_b$ = ponderações para as amostras a,b

t_a, t_b = Valores de t da tabela das amostras a,b

Graus de liberdade

$$\boxed{\begin{array}{l} t_a\ (n_a - 1)\ \text{da tabela} \\ t_b\ (n_b - 1)\ \text{da tabela} \end{array}}$$

$$s_{\bar{x}_a - \bar{x}_b} = \sqrt{\frac{s_a^2}{n} + \frac{s_b^2}{n_b}}$$

$$\boxed{t_o = \frac{\bar{x}_a - \bar{x}_b}{s_{\bar{x}_a} - \bar{x}_b}}$$

si $t_o > t_g$ há diferença

$\boxed{\text{Comparar } t_g \text{ com } t_o}$

5. Teste de t quando os dados são em pares:

$$\boxed{t_o = \frac{\bar{D}}{s_{\bar{d}}}}$$

s_d = Desvio padrão da diferença entre pares

$s_{\bar{d}}$ = Desvio padrão das diferenças

D_i = Diferença entre cada par

$\bar{D}$ = Média da diferença entre pares

$$\bar{D} = \bar{X}_a - \bar{X}_b$$

n = Número de pares

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

$$s_d = \sqrt{\frac{\Sigma D_i^2 - (\Sigma D_i)^2/n}{(n-1)}}$$

t_t = (n pares) − 1 = G.L.

Antes de aplicar o teste de t para a comparação das médias de duas amostras de igual ou diferente tamanho, se verifica se as variâncias são homogêneas ou não, através do teste de F.

$$F_o = \frac{\text{Variância maior}}{\text{Variância menor}} : \frac{\text{GL (n - 1 numerador)}}{\text{GL (n - 1 denominador)}}$$

ESTATÍSTICA EXPERIMENTAL

Ilustração

Sejam duas fazendas com diferente número de quadras de dendê em produção, pertencentes a uma empresa produtora de óleo. O gerente deseja saber, se existe diferença entre as duas fazendas quanto a produção de fruta fresca, devido a elevação dos custos de produção numa delas.

Quadra	TM de fruta fresca/ha/ano Fazendas A	B
1	8.3	4.2
2	7.4	3.3
3	6.5	5.6
4	7.3	4.8
5	8.2	4.5
6	6.4	
7	7.6	

Metodologia

- Populações e parâmetros
  - Palmeiras de 3 anos de idade de duas fazendas dividas em: (quadras de colheita).
  - Parâmetros: produção de fruta fresca em TM/ha/ano.

- Hipótese a testar

$H_o)\ \mu_A = \mu_B \Longrightarrow$ (Média de produção de fruta fresca em TM/ha/ano da fazenda A igual a fazenda B)

$H_A)\ \mu_A \neq \mu_B$

ESTATÍSTICA EXPERIMENTAL

Teste

∴ <u>Critério de teste:</u>

Precisamos decidir que distribuição usar (Vide tabela A)

- No nosso caso usaremos t ou Z, por que estamos comparando 2 amostras (Fazenda A e B).

- Usaremos t, devido a que tamanho de n e menor que 30 (7 e 5 quadras de colheita)

<u>Cálculo de Estimadores:</u>

• No nosso caso temos $n_A \neq n_B$, no entanto, não sabemos nada acerca das variâncias portanto uma hipótese prévia:

$H_o : \delta^2_A = \delta^2_B$ variância A = variância B

$H_A : \delta^2_A \neq \delta^2_B$ variância A ≠ variância B

Critério de teste ∴ teste de hipótese acerca das variâncias. Usaremos sempre para este tipo de hipóteses a distribuição ou critério "F" de Fisher.

$$F_o = \frac{\delta^2 >}{\delta^2 <}$$

Aritmética ::: com os dados do exemplo temos:

Lotes colheitas	Fazendas A (TM/ha/ano)	Fazendas B (TM/ha/ano)
1.	8.3	4.2
2.	7.4	3.3
3.	6.5	5.6
4.	7.3	4.8
5.	8.2	4.5
6.	6.4	
7.	7.6	

$\Sigma = 51.7$ $\Sigma = 22.2$

$n_A = 7$ $n_B = 5$

$\bar{x}_A = 7.4$ $\bar{x}_B = 4.5$

$s^2_A = 0.548$ $s^2_B = 0.706$

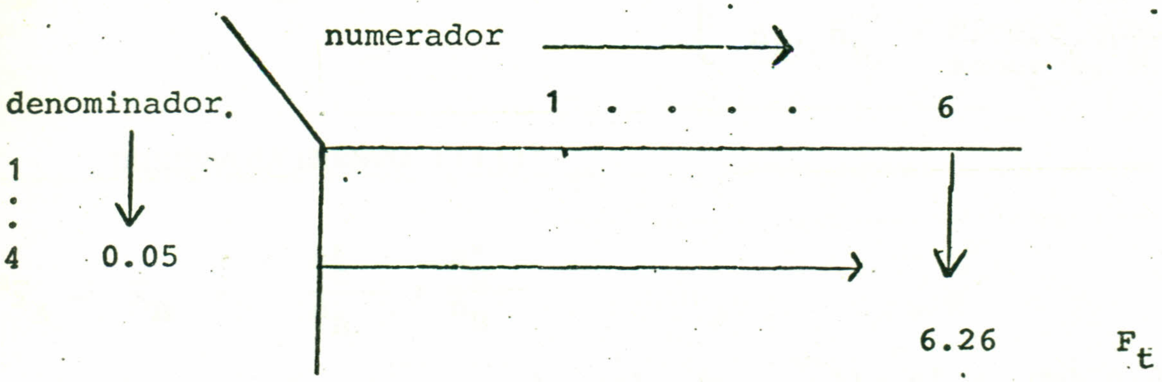
$s_A = 0.74$ $s_B = 0.84$

Teste de Hipótese

$$F_O = \frac{S_B^2}{S_A^2} \quad \frac{0.706}{0.548} = 1.288$$

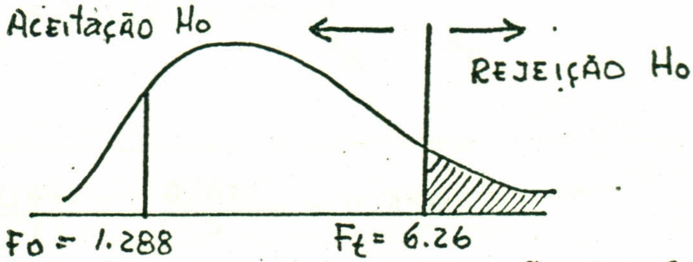
Ver tabela de "F" para $\dfrac{n\ (\text{numerador}) - 1}{n\ (\text{denominador}) - 1}$

Neste caso: $\dfrac{4}{6}$ para $P = 0.05$



Decisão: Nosso valor F_O = 1,288 cai dentro da zona de aceptação da hipótese (H_O).

Portanto: $\delta_A^2 = \delta_B^2 \implies$ homogêneas.



Com este critério voltamos a nossa situação de cálculo do estimador, para testar a hipótese inicial.

- Se decide pelo teste de t (duas amostras, n < 30), no nosso caso:

ESTATÍSTICA EXPERIMENTAL

$$\boxed{n_A \neq n_B \; ; \; s_A^2 = s_B^2}$$

DADOS ::::::

$\bar{x}_A = 7.4 \; ; \; s_A^2 = 0.548 \; ; \; n_A = 7$

$\bar{x}_B = 4.5 \; ; \; s_B^2 = 0.706 \; ; \; n_B = 5$

$\bar{x}_A, \bar{x}_B$ = Médias fazendas A,B

s_A^2, s_B^2 = Variâncias fazendas A, B

n_A, n_B = Número quadras fazenda A, B

DESENVOLVIMENTO :::::

$$s_{\bar{x}_A - \bar{x}_B} = \frac{s_c^2}{n_A} + \frac{s_c^2}{n_B}$$

$$s_c^2 = \frac{s_A^2 \; (n_a - 1) \; + \; s_B^2 \; (s_B^2 \; (n_B - 1)}{(n_A - 1) \; + \; (n_B - 1)}$$

s_c^2 = combinada (ponderada)

$$s_c^2 = \frac{0.548 \; (6) + 0.706 \; (4)}{6 + 4}$$

$s_c^2 = 0.611$: Substituir em $s^2_{\bar{x}_A - \bar{x}_B}$

$$s^2_{\bar{x}_A - \bar{x}_B} = \frac{0.611}{7} + \frac{0.611}{5} = 0.458$$

$$t_o = \frac{\bar{x}_A - \bar{x}_B}{s_{\bar{x}_A - \bar{x}_B}} , \quad t_o = \frac{7.4 - 4.5}{0.458} = 6.33$$

ESTATÍSTICA EXPERIMENTAL

Testes de Hipótese

∷ Nível de significância:

$P = 0.05$ para $(n_A - 1) + (n_B - 1)$ graus de liberdade.

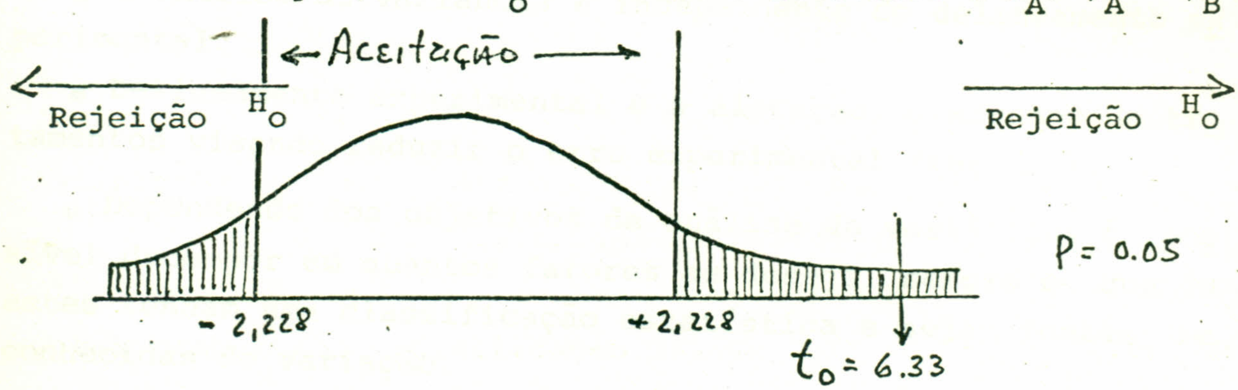
$6 + 4 = 10$ graus de liberdade

• Ver tabela de "t".

df (graus de liberdade) $\xrightarrow{\text{significância}}$ 0.05

1
⋮
10 ⟶ $2.228 = t_t$

Decisão : Nosso valor calculado $t_o = 6.33$ cai dentro de uma zona de rejeição da H_o . Portanto aceitamos a H_A) $\mu_A \neq \mu_B$

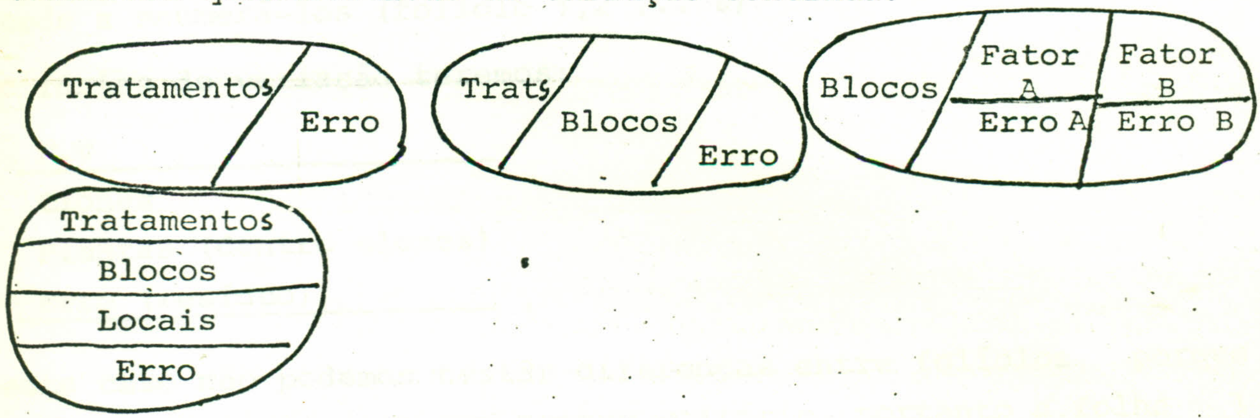


Conclusão: A diferença de produção observada entre as fazendas A e B, e estatisticamente diferente.

IMPORTANTE ::::::::: Outros problemas, similares serão meras aplicações de critérios de teste adequados.

4.4 - Conceitos Gerais para a Análise de Variância

A análise de variância é essencialmente um processo aritmético para a partição da soma total de quadrados em componentes associados com fontes de variação. É particularmente útil para o teste de hipóteses quando a informação corresponde a variáveis quantitativas de variação contínua.



- A análise de variância é independente do delineamento experimental.

- Delineamento experimental e a alocação, o arranjo de tratamentos visando reduzir o erro experimental (resíduo).

- Dependendo dos objetivos da análise de variância, é possível decompor em quantos fatores se deseja, sempre e quando estes tenham uma classificação estatística e sejam fontes reconhecidas de variação.

A análise de variância e um teste de hipótese.

- Exemplo que ilustra quando uma fonte de variação produz ou não variação. Em outras palavras, o fator é ou não uma classificação estatística.

Certo pesquisador deseja saber se há diferenças entre clones de guaraná quanto a alometria das suas folhas. Para isto:

a) Seleciona 5 clones e 6 plantas de cada clone.

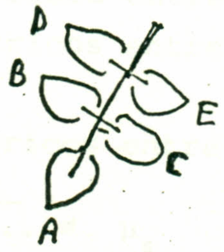
b) Em cada planta seleciona cinco folíolos normais ao acaso e procede a enumerá-los (folíolo 1,2 ... 6)

Como fontes de variação teremos:

F.V
Clones
Plantas (dentro clones)
Erro (resíduo)

- Neste caso não podemos testar diferenças entre folíolos, porque foram tomados ao acaso sem nenhum critério, portanto a folha 1 da planta 2 por exemplo, não tem relação com a folha 1 da planta 6. Isto devido que a enumeração das folhas foi apenas para identificá-las quando as medidas a forem tomadas, e não seguiu um critério de classificação estatística.

- No entanto, os folíolos podem ter um critério de classificação: Nós sabemos que as folhas do guaranazeiro possuem 5 folíolos que podem ser identificados segundo a sua posição:



- Se medirmos agora 5 folíolos A, 5 folíolos B,e...5 folíolos D, podemos agora realizar a comparação entre folíolos, já que o folíolo A da planta 1 terá a mesma posição ou qualidade dos demais folíolos a das plantas restantes. Portanto, a partição da variação total poderá ser:

F. V.
Clones
Plantas

- Se teríamos interesse em conhecer a interação entre fatores, as fontes de variação serão:

F.V.
Clones
Plantas
Folíolos
Clones x Plantas
Clones x Folíolos
Plantas x Folíolos
Plantas x Plantas x Folíolos
Erro

Para ilustrar o método geral para o teste de hipótese e cálculo da análise de variância, usaremos o mesmo exemplo do guaraná.

LEMBRAR: Quando se deseja fazer inferença sobre mais de duas amostras, devemos usar a análise de variância ou distribuição de F.

- Seguindo com a nossa ilustração, temos as seguintes hipóteses. Com relação ao índice alométrico (L/C) = largura do folíolo/comprimento do folíolo.

$$H_o = \mu_1 = \mu_2 = \ldots\ldots = \mu_6$$

- Os índices alométricos entre clones são iguais
- Os índices alométricos entre plantas dentro clones dos são iguais.
- Os índices alométricos entre folíolos são iguais

$$H_A: \mu_1 \neq \mu_2 \neq \ldots\ldots \neq \mu_6$$

- Os índices alométricos entre clones não são iguais.

Para outros fatores da mesma maneira ....

Os dados podem ser apresentados da seguinte maneira para 3 clones, 3 plantas por clone e 4 folíolos ao acaso por clone:

Clones	Folíolos	Plantas 1	Plantas 2	Plantas 3	Somas (clones)
222	1	0,42	0,46	0,41	5,28
	2	0,46	0,49	0,40	
	3	0,43	0,50	0,39	
	4	0,48	0,41	0,43	
	Soma(I)	1,79	1,86	1,63	
274	1	0,54	0,56	0,58	7,00
	2	0,58	0,59	0,59	
	3	0,56	0,61	0,60	
	4	0,57	0,60	0,62	
	Soma(I)	2,25	2,36	2,39	
276	1	0,78	0,69	0,78	8,85
	2	0,74	0,71	0,77	
	3	0,72	0,77	0,73	
	4	0,70	0,72	0,74	
	Soma(I)	2,94	2,89	3,02	
Somas (plantas)		6,98	7,11	7,04	21,13

(I) = Soma interação

A tabela de resultados de análise de variância clássica, é apresentada da seguinte forma:

Fonte Variação (F.V.)	Graus liberdade (G.L.)	Soma Quadrados (S.Q)	Quadra do Médio (Variância) Q.M	Valor de F Calculado F	Tabela $F_{\alpha .05}$
Clones (c)	2	0.533	0.2665	360.13 **	3.35
Plantas (p)	2	0.003	0.000/5	0.203 ns	3.35
Clones x Plantas	4	0.009	0.00225	3.040 *	2.73
Erro	27	0.02	0.00074		
Total:	35	0.565			

1. Cálculo Termo Correção (TC):

Lembrar: $S^2 = \dfrac{\Sigma x^2 - (\Sigma x)^2/n}{n-1}$ (1) $\left\{\right.$ O termo $(\Sigma x)^2/n$ corresponde ao termo de Correção na análise de variância.

$$TC = \frac{(y..)^2}{n}$$

$\left\{\right.$ y.. = Soma total
n = Número de dados total

$$TC = \frac{(21.13)^2}{36} = 12,4$$

2. Graus Liberdade

Se refere ao número de dados disponíveis para cada fator. Seria o denominador (n-1) da nossa geral (1).

G.L. Clones (3-1) = 2 (C-1)
G.L. Plantas (3-1) = 2 (P-1)
G.L. Clones Plantas (3-1) (2-1) = 4 (C-1) (P-1)
G.L. Total: (36-1) = 35 (CP-1)
G.L. Erro (por diferença): 35-(2+2+4) = 27

- Na literatura a determinação dos G.L. esta bem clara para cada caso em particular.

3. **Cálculo da Soma de Quadrados**

- SQ Clones: $\frac{5.28^2 + 7.00^2 + 8.85^2}{12} - TC = \frac{155.2009}{12} - TC$

  : 12.933 - 12.4 = 0.533

  Notar: O denominador 12 foi obtido realizando a seguinte pergunta: De quantos dados foi obtida cada soma para cada clone².

- SQ Plantas: $\frac{6.98^2 + 7.11^2 + 7.04^2}{12} - Tc$

  : 12.403 - 12.4 = 0.003

- SQ Plantas x Clones: $\frac{1.79^2 + \ldots + 3.02^2}{4}$ - (SQ*Clones + SQ* Plantas)+TC

  Notar: Denominador = 4: De quantos dados foi obtida cada soma de interação ?

  * Trata-se de SQ sem corrigir (antes de subtrair o TC)

  SQ plantas x clones: 12.945 - (12.933 + 12.403) + TC

  ≈ 0.009

- SQ Total: $0.42^2 + 0.46^2 + \ldots + 0.74^2 - TC$

  ≈12.965 - 12.4 = 0.565

- SQ Erro : Calculado por diferença

  SQ Erro = SQ Total - (SQ Clones + SQ Plantas + SQ Clones x Plantas)

  = 0.565 - (0.533 + 0.003 + 0.009)

  = 0.02

4. Cálculo dos quadrados médios.

O quadrado médio e na realidade a variância, é calculado dividindo a soma de quadrados de cada fator pelos graus de liberdade de cada fator.

Q.M. Clones:	0.533/2 = 0.2665
Q.M. Plantas:	0.003/2 = 0.00015
Q.M. Clones x Plantas:	0.009/4 = 0.00225
Erro (Resíduo):	0.02/27 = 0.00074

5. Cálculo de F

O valor de F é calculado dividindo cada valor de quadrado médio pelo quadro médio do erro (resíduo).

F clones	0.2665/0.00074 = 360.13
F plantas	0.00015/0.00074 = 0.203
F clones x plantas	0.00225/0.00074 = 3.040

6. F das Tabelas

Para obter os valores de F das tabelas se considera os G.L. de cada fator (numerador) e os G.L. do erro (denominador)

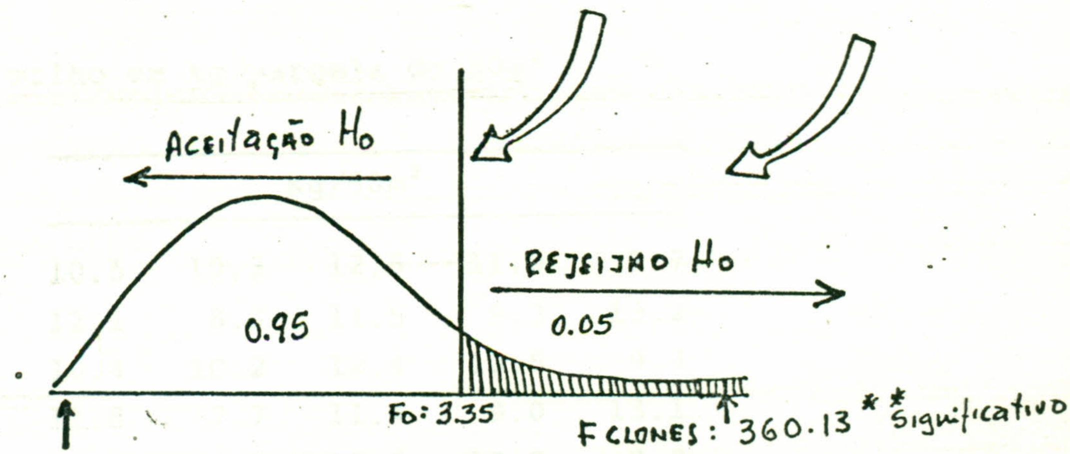
F_t Clones: GL numerador: 2; denominador 27 = 3.35

F_t Plantas: GL numerador: 2; denominador 27 = 3.35

F_t Clones x Plantas : GL numerador: 4; denominador 27 = 2.73

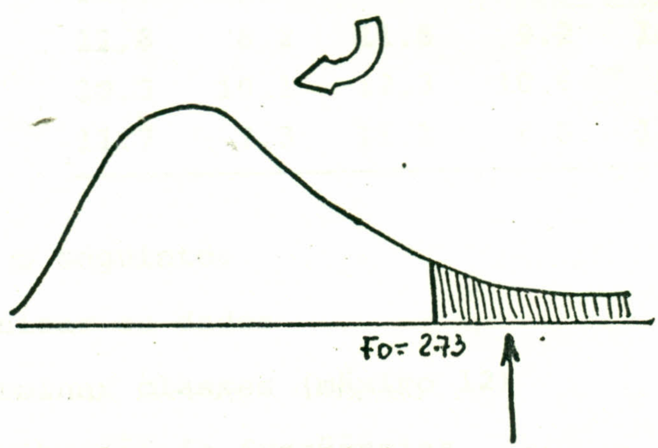
7. Para saber si um valor de F é significativo, se compara com seu valor da tabela que corresponde a um valor da distribuição de F para os respectivos graus de liberdade.

Distribuição de F a nível de P = 0,05 e GL = 2/27.



F plantas: 0.203 → NS - Não significativo

Para GL 4/27



F Clones x plantas → 3.040* Significativo

8. Conclusão

A. Existem diferenças significativas entre clones quanto ao índice alométrico L/C.
(Aceitou-se $\mu_a = (\mu_i \neq \mu_2 \ldots \neq \mu_6)$

B. Não existem diferenças entre plantas dentro de clones.

C. Os índices alométricos das plantas variam de acordo com o clone.

ESTATÍSTICA EXPERIMENTAL

Distribuição de Freqüências

PRÁTICA Nº 1

Produção de milho em kg/parcela de $50m^2$

$Kg/50m^2$				
10.5	10.3	12.5	11.0	9.7
12.1	8.3	11.5	9.3	13.2
10.4	10.2	12.4	10.8	9.4
11.8	7.7	11.4	9.0	13.1
10.3	9.8	12.0	10.5	9.3
11.7	7.0	11.1	13.0	8.4
10.4	10.2	12.4	11.0	9.6
11.8	8.2	11.5	9.2	13.2
10.3	10.1	12.3	10.6	9.4
11.7	7.3	11.1	8.5	13.1

Realize o seguinte:

1. Organizar os dados
2. Determinar classes (máximo 12)
3. Distribuição de freqüências
4. Elaborar tabela de freqüências ascendentes e descendentes e de freqüências relativas ascendentes e descendentes.
5. Descreva 3 aspectos que considere importantes desta amostra de dados.

ESTATÍSTICA EXPERIMENTAL

Distribuição de freqüências

Prática nº 2

A seguinte tabela mostra uma distribuição de freqüência de horas trabalhadas na colheita de frutos de laranja num certo período.

Determinar:

a. Limite superior da quinta classe
b. Limite inferior da oitava classe
c. Ponto médio da sétima classe
d. Freqüência da quarta classe
e. Freqüência relativa da sexta classe
f. Porcentagem de trabalhadores que não trabalharam mais de 600 hs
g. Porcentagem de trabalhadores que trabalham 950 horas ou mais
h. Porcentagem de trabalhadores que trabalham pelo menos 500 horas porém menos que 1000 horas.

Trabalho colheita (Horas)	Número de trabalhadores
300 - 399	14
400 - 499	46
500 - 599	58
600 - 699	76
700 - 799	68
800 - 899	62
900 - 999	48
1000 -1099	22
1100 -1199	6

Nota: Elabore uma figura para responder o item g.

ESTATÍSTICA EXPERIMENTAL

MEDIDAS DE TENDÊNCIA CENTRAL

PRÁTICA Nº 3

1 . Com os dados da prática 2, determinar:

a) Média

b) Mediana

c) Moda

2. Seja a seguinte amostra {4, 4, 6, 7, 8, 9, 10}

a) demonstre que os desvios com relação a média e = 0

b) Demonstre que os desvios quadráticos com relação a média e um mínimo.

Nota: As propriedades da média do item 2, são fundamentais, para compreender as propriedades das medidas da dispersão.

ESTATÍSTICA EXPERIMENTAL

MEDIDAS DE DISPERSÃO

PRÁTICA Nº 4

• Sejam as seguintes amostras de uma variável qualquer.

A = {2,0, 5, 3, 3, 2, 4, 3, 5, 2, 2, 4, 1}

B = {23, 21, 24, 22, 23, 25, 22, 20, 22, 23, 24, 22, 25}

Calcular o seguinte de cada conjunto:

a) Média aritmética

b) Desvio padrão

c) Coeficiente de variação

• Responda o seguinte:

1. Qual das médias é mais consistente com base ao desvio padrão
2. Com base ao coeficiente de variação
3. Discuta as conclusões do item 1 e 2.

•• Seja o seguinte conjunto {5, 4, 3, 2, 2, 6, 7}

a) Calcular a variância usando as seguintes fórmulas:

$$S^2 = \frac{\Sigma\,(x-\bar{x})^2}{N} \quad ; \quad S^2 = \frac{\Sigma x^2 - (\Sigma x)^2/h}{N} \quad ; \quad S^2 = \frac{\Sigma x^2}{N} - \bar{x}^2$$

Verifique se os resultados são iguais.

ESTATÍSTICA EXPERIMENTAL

CORRELAÇÃO E REGRESSÃO

PRÁTICA Nº 5

As seguintes cinco variáveis, correspondo a uma avaliação de cinco características em 5 plantas.

Variável (Característica)				
X_1	X_2	X_3	X_4	X_5
12	46	10	15	85
6	19	18	25	46
4	12	22	33	23
10	35	9	12	80
8	40	25	38	75

Calcular os coeficientes de correlação seguintes:

a) X_1 X_2 b) X_3 X_5 c) X_3 X_4

Se algum valor r for significativo, calcular a equação linear e representar graficamente.

ESTATÍSTICA EXPERIMENTAL

AMOSTRAGEM

PRÁTICA Nº 6

Comprimento em metros do Raquis da folha 17 de 50 plantas diferentes de Dendê de 3 1/2 anos de idade. Material. *Elaeis oleifera*.

(1)	3.41	(11)	3.84	(21)	3.51	(31)	3.33	(41)	3.84
(2)	4.17	(12)	3.97	(22)	4.11	(32)	3.28	(42)	4.00
(3)	3.70	(13)	3.84	(23)	4.31	(33)	3.79	(43)	3.94
(4)	3.64	(14)	4.01	(24)	4.08	(34)	3.60	(44)	3.81
(5)	3.65	(15)	4.30	(25)	4.23	(35)	3.61	(45)	4.08
(6)	3.37	(16)	3.39	(26)	3.96	(36)	3.59	(46)	3.74
(7)	3.21	(17)	3.85	(27)	4.08	(37)	3.60	(47)	4.06
(8)	3.69	(18)	3.77	(28)	3.65	(38)	3.76	(48)	3.88
(9)	3.60	(19)	3.78	(29)	3.95	(39)	3.59	(49)	3.76
(10)	3.41	(20)	4.05	(30)	4.24	(40)	3.62	(50)	4.09

- Do quadro anterior, extrair uma amostra ao acaso de 5 plantas e determine o tamanho da amostra sendo a média da população de 3.80m e o desvio padrão de: 0.267.

- Determine também a confiabilidade da estimativa a partir da amostra e erro de amostragem.

ESTATÍSTICA EXPERIMENTAL

TESTE DE HIPÓTESE

PRÁTICA Nº 7

Durante uma visita de campo a um plantio de Eucalipto, alguém observou que certa quadra tinha em aparência maior altura que outra quadra visitada anteriormente. Porém, outras pessoas afirmaram o contrário. O gerente incomodado por não dispor das informações necessárias para esclarecer a dúvida, mandou a medir duas amostras ao acaso de 50 árvores de cada quadra. Em ambas as duas amostras, mediu-se a altura do tronco, e os resultados foram os seguintes:

	Quadra A	Quadra B
$\bar{X}$	= 5,70m	5,03m
S^2	= 0.588m	0.481m

É realmente a altura da quadra A <u>superior</u> a quadra B?

ESTATÍSTICA EXPERIMENTAL

ANÁLISES DE CONTAGEM

PRÁTICA Nº 8

Em um lote de sementes germinadas encontrou-se uma infecção de fungos *Schizophyllum comune*.

Para combater a doença se aplica na metade do lote atacado uma mistura de 2:1 dos fungicidas Dithane + Benlate e se obteram os seguintes resultados:

Contagem 15 dias após o tratamento.

	Recuperadas	Recuperando-se	Mortas	Total
Tratadas	50	25	15	90
Não tratadas	55	17	13	85
Total	105	42	28	175

- Está relacionada a recuperação das sementes com o tratamento de fungicidas aplicados?

ESTATÍSTICA EXPERIMENTAL

CONCEITOS GERAIS PARA ANÁLISE DE VARIÂNCIA

PRÁTICA Nº 9

A seguir temos uma tabela de análise de variância incompleta. Favor completá-la e comentar os resultados.

ANOVA — Número de folhas de guaraná por m de ramo

F.V.	G.L.	SQ	QM	F_c	F(P=0.05)
Tipos de Ramos		0,017			
Plantas	4				
Ramos x Plantas		23,395			
Erro		562,159			
	99	625,499			

NOTA: Testou-se: Ramos sombreados vs não sombreados

Amostra : 5 plantas e 10 ramos sombreados e 10 não sombreados por planta.