Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Cadernos de Matemática e Estatística
Série B: Trabalho de Apoio Didático

Elementos Básicos de Estatística

João Riboldi

Série B, nº 14,
Porto Alegre, janeiro de 1993

ELEMENTOS BÁSICOS DE ESTATÍSTICA

JOÃO RIBOLDI

P R E F Á C I O

Estas notas não têm por objetivo substituir nenhum texto de Estatística Elementar e sim organizar, de forma simplificada, alguns conceitos básicos de Estatística.

Elas se destinam à revisão de Conceitos de Estatística e, como são apresentadas de forma muito particular, especificamente são úteis para a disciplina de Análise Estatística que ministramos nos cursos de Pós-Graduação em Agronomia, muito embora possam ter alguma utilidade noutras situações.

Agradecemos à Stela e ao Flávio pelo trabalho de digitação, sem o qual não seria possível a publicação das notas.

Porto Alegre, fevereiro de 1993.

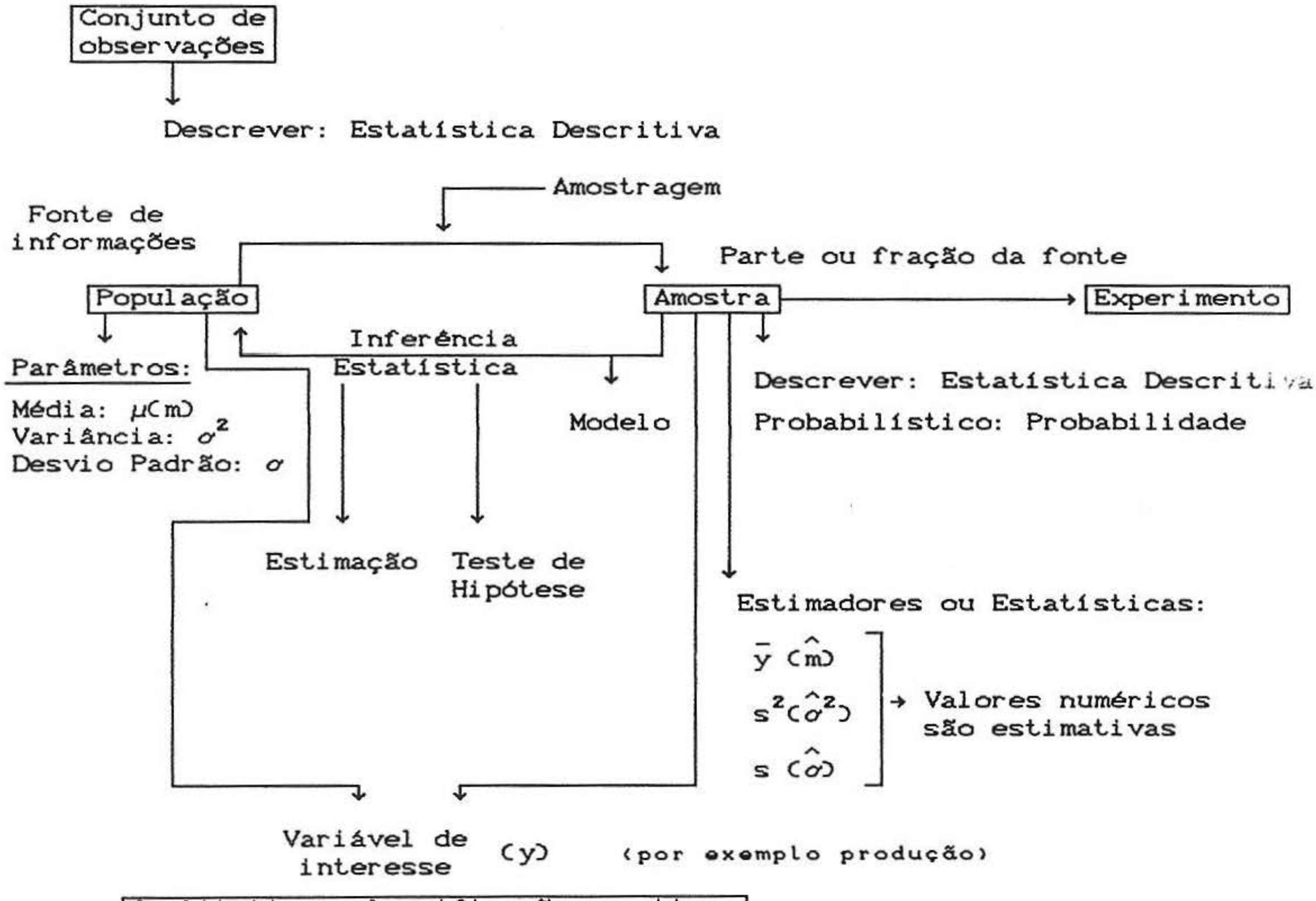
Prof. João Riboldi

I N D I C E

0. E S T A T Í S T I C A

- Matemática aplicada aos dados de observação;
- Quando orientada para a área de investigação, dentro do chamado método científico, é definida como a ciência que se ocupa da experimentação no que diz respeito a sua

* Planificação (Planejamento de Experimentos)
* Execução (Instalação, condução e coleta de informações de experimentos)
* Análise dos seus resultados



Qualitativa: classificação por tipos
Quantitativa: discreta → contagem; contínua → medição

1. ESTATÍSTICA DESCRITIVA

..1 - Medidas de Tendência Central ou de Posição

* * * * * * * * * * * * * *

↑

VALOR CENTRAL

↓

usado para descrever ou representar o conjunto de observações

.1.1 - Média : é a mais importante das medidas de tendência central.

→ Média Aritmética
Média Ponderada
Média Harmônica
Média Geométrica

(a) Média Aritmética (média): é a que mais interessa pois de :ma maneira geral é a mais representativa

na amostra: $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n} = \dfrac{y_1 + y_2 + \ldots + y_n}{n}$

na população: $\mu = \dfrac{\sum_{i=1}^{N} y_i}{N} = \dfrac{y_1 + y_2 + \ldots + y_N}{N}$

(b) Média Ponderada (média aritmética ponderada):

Em algumas situações a média ponderada é mais recomendável e la é dada por

$$\bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} = \frac{w_1 y_1 + w_2 y_2 + \ldots + w_n y_n}{w_1 + w_2 + \ldots + w_n}$$

onde w_i é o peso associado à observação y_i wu

$$\bar{y} = \frac{\sum_{i=1}^{n} f_i y_i}{\sum_{i=1}^{n} f_i} = \frac{f_1 y_1 + f_2 y_2 + \ldots + f_n y_n}{f_1 + f_2 + \ldots + f_n}$$

onde f_i é a freqüência associada à observação y_i.

Exemplo 1: Dados de produtividade em t/ha, referentes a uma certa variedade de cana-de-açúcar.

80,7	90,6
81,6	91,8
83,5	92,4
83,8	95,6
87,5	100,4

$n=10 \quad \bar{y} = \frac{80{,}7+81{,}6+\ldots+100{,}4}{10} = \frac{887{,}9}{10} =$

$= 88{,}79$ t/ha

↓

estimativa da produtividade média verdadeira da variedade

Exemplo 2 : Intensidade média de infestação de "broca-podridões" da cana-de-açúcar numa determinada usina e o nº de talhões infestados de cada variedade:

Variedade	Nº de talhões	% de infestação
1	12	9,10
2	40	14,57
3	4	3,20
4	2	2,89
5	6	8,74
6	18	11,70
7	21	10,10
8	10	7,15

Média Aritmética:

$$\bar{y} = \frac{9,10+14,57+\ldots+7,15}{8} = 8,43\%$$

Irreal dada a grande variação no nº de talhões infestados por variedade.

Média Ponderada:

$$\bar{y} = \frac{12(9,10)+40(14,57)+\ldots+10(7,15)}{12+40+\ldots+10} = 11,12\%$$

(c) Média Geométrica:

$$g \text{ ou } \bar{y}_g = \sqrt{y_1 \times y_2 \times \ldots \times y_n}$$

(d) Média Harmônica:

$$h \text{ ou } \bar{y}_h = \frac{n}{\frac{1}{y_1} + \frac{1}{y_2} + \ldots + \frac{1}{y_n}}$$

$$\boxed{h \leq g \leq \bar{y}}$$

1.1.2 - Mediana: é o valor central de uma seqüência de observações ordenadas de forma crescente.

Se n é par e $n = 2k$

$$\text{Mediana} = \frac{y_k + y_{k+1}}{2}$$

Se n é ímpar e $n = 2k+1$

$$\text{Mediana} = y_{k+1}$$

Exemplo 1: $n = 10 \leftarrow$ par $\quad n = 2.k \longrightarrow k = 5$

$$m_d = \frac{87,5 + 90,6}{2} = 89,05 \text{ t/ha}$$

$n = 11 \qquad y_{11} = 101,2$

$$m_d = y_6 = 90,6$$

1.1.3 - Moda: é a observação que ocorre com maior freqüência.

Exemplo 1: não tem moda.

Exemplo 2: 10,10%

1.1.4 - Representatividade das medidas de tendência central:

* * * * * * * *   * *

$\uparrow$ Mediana $\uparrow \bar{y}$

1.2 - Medidas de Variação ou Dispersão

* * * * * * * * * * *

Medida de tendência central

Distância: medida de dispersão

Amostra 1	Amostra 2	Amostra 3	Amostra 4
10	9	0	5
10	10	10	5
10	10	10	5
10	11	20	25
$\bar{y}_1 = 10$	$\bar{y}_2 = 10$	$\bar{y}_3 = 10$	$\bar{y}_4 = 10$
Variação nula	Variação pequena	Variação alta	
$A_1 = 0$	$A_2 = 2$	$A_3 = 20$	$A_4 = 20$
$s_1^2 = 0$	$s_2^2 = 0,67$	$s_3^2 = 66,67$	$s_4^2 = 100$
$s_1 = 0$	$s_2 = 0,82$	$s_3 = 8,16$	$s_4 = 10$
$cv_1 = 0\%$	$cv_2 = 8,2\%$	$cv_3 = 81,6\%$	$cv_4 = 100\%$

.2.1 - Amplitude Total: diferença entre as observações extremas.

A = maior valor - menor valor

Medida incompleta de variação pois não considera a variação nterna.

Exemplo 1: A = 100,4 - 80,7 = 19,7 t/ha

.2.2 - Variância: é a melhor medida de dispersão pois leva em consideração todas as observações.

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

$\sum_{i=1}^{n} (y_i - \bar{y})^2 \rightarrow$ Soma dos Quadrados dos desvios das observações em relação a média (SQ)

$n-1 \rightarrow$ Graus de liberdade (GL)

Quadrado médio (QM): $QM = \dfrac{SQ}{GL}$

Na população: $\sigma^2 = \dfrac{SQ}{N} = \dfrac{\sum_{i=1}^{N} (y_i - \mu)^2}{N}$

$$\sum (y_i - \bar{y}) = 0$$

$$\sum (y_i - \bar{y}) = (y_1 - \bar{y}) + (y_2 - \bar{y}) + \ldots + (y_n - \bar{y}) = y_1 + y_2 + \ldots + y_n - n\bar{y} =$$

$$= \sum y_i - n \frac{\sum y_i}{n} = \sum y_i - \sum y_i = 0$$

O princípio de GL, constantemente usado na metodologia estatística, pode ser conceituado de ≠s maneiras, sendo uma delas a seguinte: "Para n observações, se impusermos a restrição da média $\bar{y}$, pode-se escolher n-1 observações livremente, ficando a última condicionada a satisfazer a condição de $\bar{y}$, ou seja que $\sum (y_i - \bar{y}) = 0$."

$$s^2 = \frac{SQ}{GL} \quad ; \quad SQ = \sum (y_i - \bar{y})^2 = \underbrace{\sum y_i^2}_{\text{Soma dos Quadrados das observações}} - \underbrace{\frac{\overbrace{(\sum y_i)^2}^{\text{Quadrado da soma das observações.}}}{n}}_{\text{Correção ou Fator de Correção}}$$

Propriedades da média e da variância

	MÉDIA	VARIANCIA
Soma ou subtração por k	+ ou − por k	inalterada
Multiplicação ou divisão por k	× ou ÷ por k	× ou ÷ por k^2

Exemplo 1: $\sum y = 80,7 + \ldots + 100,4 = 887,9$

$$\sum y^2 = (80,7)^2 + (81,6)^2 + \ldots + (100,4)^2 = 79214,87$$

$$s^2 = \frac{79214,87 - (887,9)^2/10}{9} = 42,0254$$

.2.3 - Desvio Padrão:

$$\boxed{s = \sqrt{s^2} = \sqrt{\frac{SQ}{GL}}}$$

Tem a mesma unidade de medida dos dados originais, ou seja, a média.

Exemplo 1:

$$s = \sqrt{42,0254} = 6,48 \text{ t/ha}$$

.2.4 - Coeficiente de variação:

É um desvio padrão relativo ou proporcional, pois expressa o desvio padrão como proporção da média.

Assim

$$\boxed{cv = \frac{s}{\bar{y}} \cdot 100}$$

O cv é utilizado, dentre outras utilizações, para expressar precisão de experimentos.

Considera-se

cv	
Baixos:	< 10%
Médios:	10 a 20%
Altos:	20 a 30%
Muito altos:	> 30%

Exemplo 1:

$$cv = \frac{s}{\bar{y}} \cdot 100 = \frac{6,48}{88,79} \cdot 100 = 7,30\%$$

.3 - Distribuições de Freqüências

Organização, síntese e descrição da informação.

TABELAS

Tabela de Freqüências (Distribuição de freqüências): Distribuição dos dados de observação em classes associando a cada classe uma freqüência.

GRÁFICOS

Representação no sistema de eixos cartesianos.

* Variáveis qualitativas e quantitativas discretas:
  - Diagrama de Colunas
  - Diagrama de Barras.
* Variável quantitativa contínua:
  - Histograma
  - Polígono de Freqüência
  - Ogiva.

DIAGRAMA DE COLUNAS

Variável Qualitativa

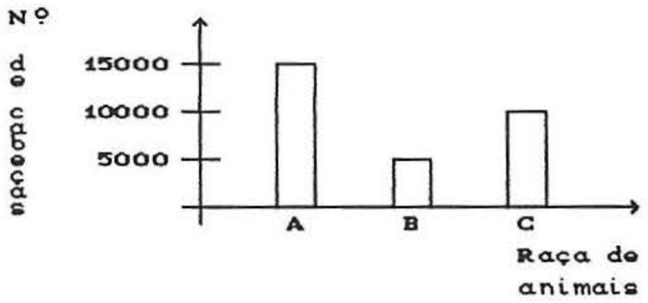
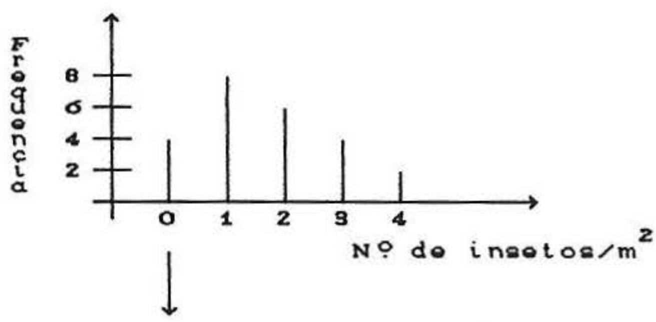


DIAGRAMA DE BARRAS

Variável Quantitativa Discreta



Classes: intervalos de valores.

Variável quantitativa contínua

Classes representam intervalos de valores.

- Nº de classes depende →
  - nº de observações
  - amplitude de variação
  - grau de condensação

- Determinação do nº classes: (c)

  * Fórmulas empíricas: $c = \sqrt{n}$

  $c \cong 1 + 3{,}22 \log n$

  * Usando a relação A/s que depende do nº de observações e lependendo do grau de condensação desejado usa-se como intervalo le classe 1/2 ; 1/3 ou 1/4 do s.

Exemplo 3: Os dados seguintes são referentes ao teor de N(%) em cana planta.

,99	2,04	2,07	2,09	2,13
,00	2,04	2,07	2,09	2,14
,00	2,04	2,08	2,09	2,14
,01	2,04	2,08	2,10	2,15
,01	2,06	2,08	2,10	2,16
,02	2,06	2,08	2,11	2,16
,02	2,06	2,09	2,12	2,16
,03	2,07	2,09	2,13	2,17

$n = 40$

º de classes: $C = \sqrt{40} = 6{,}32 \cong 7$

mplitude Total : $A = 2{,}17 - 1{,}99 = 0{,}18$

ntervalo ou amplitude de classe (a):

$$a = \frac{A}{C} = \frac{0{,}18}{7} = 0{,}0257 \cong 0{,}03$$

$a \cdot C \geq A \quad \Rightarrow \quad 0.03 \cdot 7 = 0{,}21 > A = 0{,}18$

$$0{,}21 - 0{,}18 = 0.03 \begin{cases} 0{,}015 \text{ abaixo} \\ 0{,}015 \text{ acima} \end{cases}$$

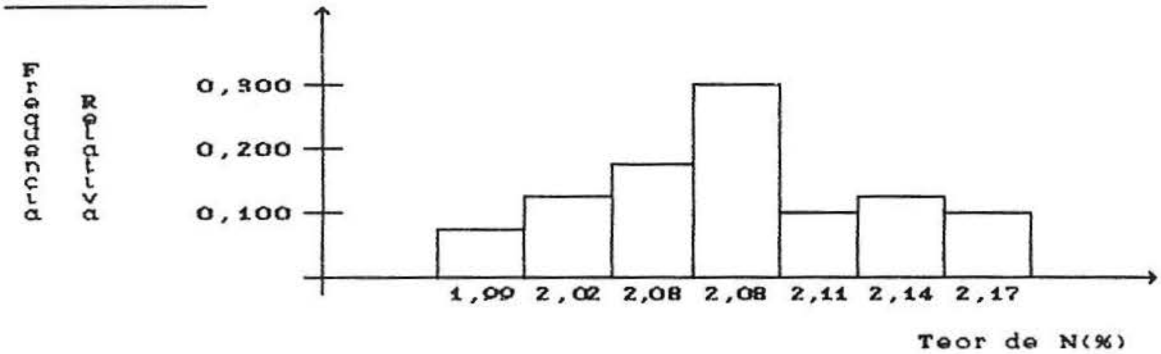
Início: $1{,}99 - 0{,}015 = 1{,}975$

Fim: $2{,}17 + 0{,}015 = 2{,}185$

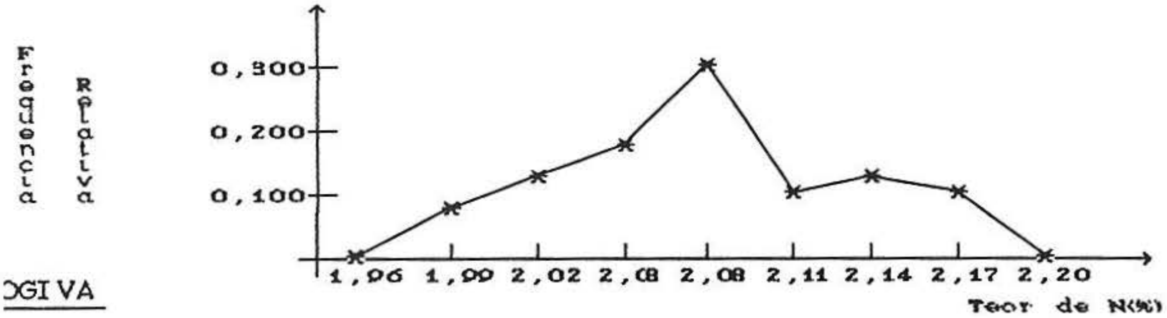
istribuição de Freqüências — Ponto Médio

Classes	Centro de classe	Freqüências	Freqüência Relativa	Freq. Relativa Acumulada
1,975 - 2,005	1,99	3	0,075	0,075
2,005 - 2,035	2,02	5	0,125	0,200
2,035 - 2,065	2,05	7	0,175	0,375
2,065 - 2,085	2,08	12	0,300	0,675
2,085 - 2,125	2,11	4	0,100	0,775
2,125 - 2,155	2,14	5	0,125	0,900
2,155 - 2,185	2,17	4	0,100	1,000
Total	--	40	1,000	--

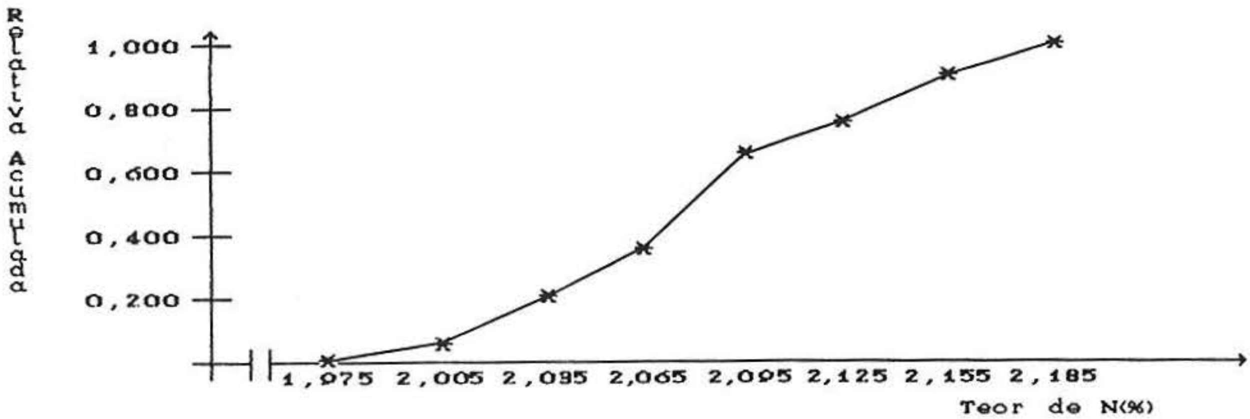
HISTOGRAMA



POLÍGONO DE FREQÜÊNCIAS



OGIVA



2. DISTRIBUIÇÃO DE PROBABILIDADE

2.1 - Algumas noções de probabilidade:

Experimentos aleatórios → Evento: um acontecimento em particular

Experimentos aleatórios → Espaço Amostral (s): conjunto dos possíveis resultados

↓

Não se sabe "a priori" qual resultado acontecerá.

Exemplo 4

Lançamento de uma moeda honesta e observar o resultado da face para cima.

s = {cara, coroa}

Variável y = nº de caras

↓

Variável Aleatória

Evento	y	Probabilidade
Cara	1	1/2
Coroa	0	1/2

Nos casos onde cada elemento do espaço amostral tem a mesma probabilidade de ocorrer (espaço equiprovável), a probabilidade de ocorrer o evento A é dada por

$$\boxed{P(A) = \frac{\text{nº de casos favoráveis}}{\text{nº de casos totais}}}$$

Experimento

Dose de N kg/ha → (UE) $\boxed{\text{Parcela de Campo}}$ → Efeito → y (rendimento de uma cultura)

(0;50;100;150 kg/ha)

↓

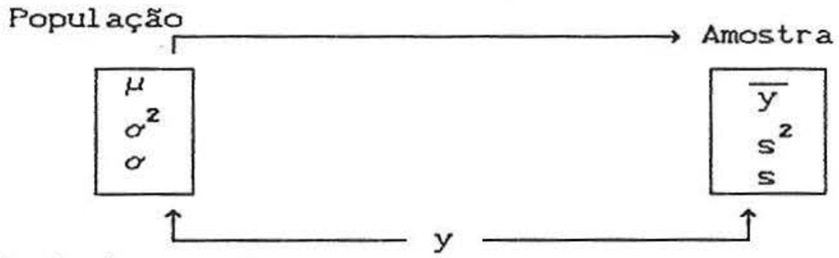
Variável Aleatória [0;□]

Variáveis Aleatórias:
(v.a.)

Discretas: Nº enumerável de valores
Qualitativas; Quantitativas de contagem
Sexo, raça, nº de insetos/m^2, ...

Contínuas: nº não enumerável de valores
Quantitativas de medição.

.2 - Distribuição de Probabilidade



Parâmetros que caracterizam a população e a distribuição de Probabilidade da v.a de interesse.

Exemplo 5

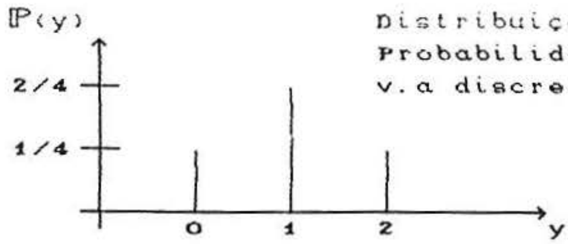
Lançamento de uma moeda honesta 2 vezes e observando o esultado da face voltada para cima.

v.a   y = nº de caras

Evento	y
cc	2
cç	1
çc	1
çç	0

y	$\mathbb{P}(y)$
0	1/4
1	2/4
2	1/4

Distribuição de Probabilidade



Distribuição de Probabilidade da v.a discreta

Esperança matemática

↓

Média: $\mu = \sum y\, \mathbb{P}(y) = 0 \,.\, 1/4 + 1 \,.\, 2/4 + 2 \,.\, 1/4 = 1$ cara

y:	y_1	y_2	...	y_N
$\mathbb{P}(y)$:	1/N	1/N	...	1/N

$$\mu = \frac{1}{N} \sum y_i = \frac{\sum y_i}{N}$$

.2.1 - Distribuição Binomial:

Distribuição Clássica de v.a. discreta, caracterizando um rocesso de contagem com 2 categorias, como por exemplo:

Contagem de plantas → infectados / não-infectados

Contagem de sementes → germinadas / não-germinadas

Na Binomial

Resultados	Probabilidade	
Sucesso	p	$p + q = 1$ → constantes
Insucesso	q	

a v.a Y = nº de sucessos.

A probabilidade de ocorrer y sucessos em n realizações é dada or

$$\boxed{\mathbb{P}(Y=y) = C_n^y\, p^y\, q^{n-y}}$$

(Função de Probabilidade da Distribuição Binomial)

$\mu = np$

$\sigma^2 = npq$

$\sigma = \sqrt{npq}$

Exemplo 6: Germinação de sementes

$n = 5 \quad G = 80 \quad (p = 0{,}80\ ;\ q = 0{,}20)$

y = nº de sementes germinadas

Qual a Probabilidade de 2 sementes germinarem?

$$\mathbb{P}(y=2) = C_5^2 \cdot (0{,}80)^2 \cdot (0{,}20)^3 = \frac{5 \cdot 4}{2} \cdot (0{,}64)\cdot(0{,}008) = 0{,}0512$$

Qual a Probabilidade de no máximo 1 semente germinar?

$$\mathbb{P}(y \leq 1) = \mathbb{P}(y=0) + \mathbb{P}(y=1) = C_5^0 \cdot (0{,}80)^0 \cdot (0{,}20)^5 + C_5^1 \cdot (0{,}80)^1 \cdot (0{,}20)^4$$

$$= 0{,}00032 + 0{,}0064 = 0{,}00672$$

Qual a Probabilidade de pelo menos 1 semente germinar?

$$\mathbb{P}(y \geq 1) = \mathbb{P}(y=1) + \mathbb{P}(y=2) + \mathbb{P}(y=3) + \mathbb{P}(y=4) + \mathbb{P}(y=5)$$

$$= 1 - \mathbb{P}(y=0) = 1 - 0{,}00032 = 0{,}99968$$

.2.2 - Distribuição de Poisson:

p pequeno; freqüência de sucesso é baixa.

Enquadram-se nesse caso contagem de certa característica por nidade de área como por exemplo nº de insetos/m^2; nº de lantas/m^2, ...

Função de Probabilidade: $\boxed{\mathbb{P}(Y=y) = \dfrac{k^y\ e^{-k}}{y!}}$

k = média ou esperança da distribuição

$$\left.\begin{aligned} \mu &= k \\ \sigma^2 &= k \\ \sigma &= \sqrt{k} \end{aligned}\right] \rightarrow \quad \text{Relação entre média e variância}$$

.2.3 - Distribuição Normal:

(a) Caracterização:

Para dados de medição é razoável admitir que segue uma istribuição normal ou aproximadamente normal.

$$Y \cap N(\mu,\sigma^2)$$

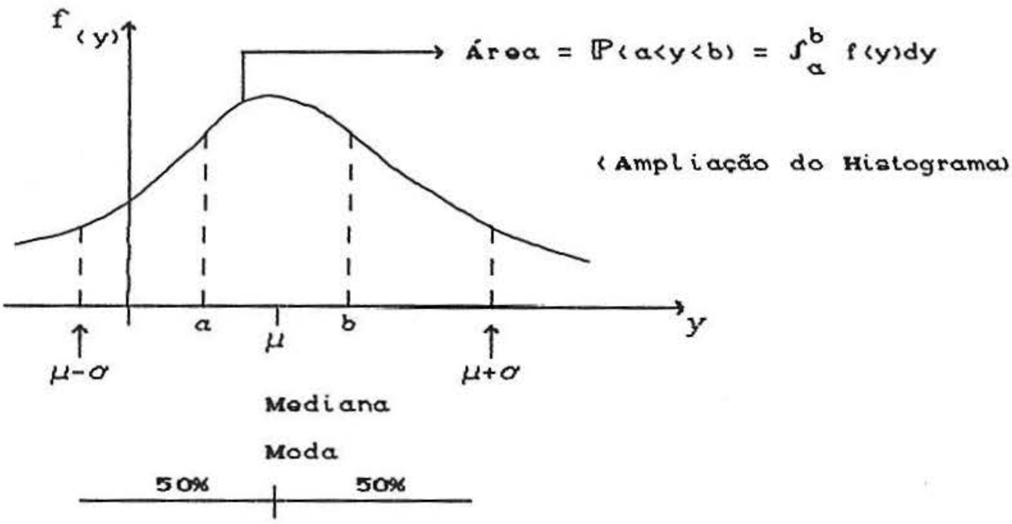
Função de Densidade

$$\boxed{f(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(y-\mu)^2}{2\sigma^2}}}$$

μ = média

σ^2 = variância

σ = desvio padrão

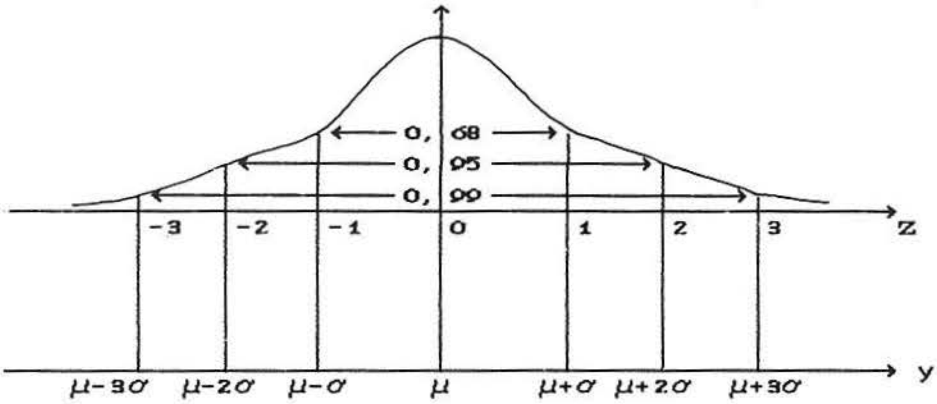


Curva Simétrica
Mesocúrtica

(b) Distribuição Normal Padrão ou Reduzida:

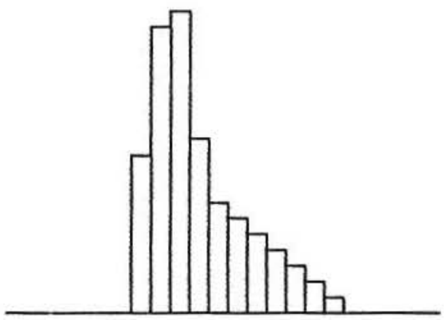
De $Y \cap N(\mu,\sigma^2)$ fazendo a transformação $z = \frac{y-\mu}{\sigma}$ temos $\cap N(0,1)$ (Normal Padrão ou Reduzida)

$$\boxed{f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}}$$



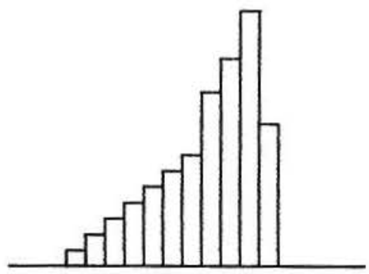
Assimetria e Curtose:

istribuição alongada à direita (Assimetria Positiva)



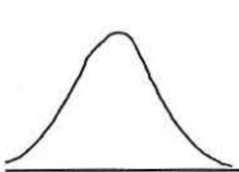
$g_1>0$

Distribuição alongada à esquerda (Assimetria Negativa)

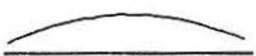


$g_1<0$

Curtose (achatamento)



.eptocúrtica ($g_2>3$)



Platicúrtica ($g_2<3$)

Mesocúrtica (normal) ($g_2=3$)

Coeficiente de Assimetria:

$$g_1 = \frac{m_3}{s^3} = \frac{m_3}{m_2\sqrt{m_2}}$$

Normal: $g_1 = 0$

Coeficiente de Curtose:

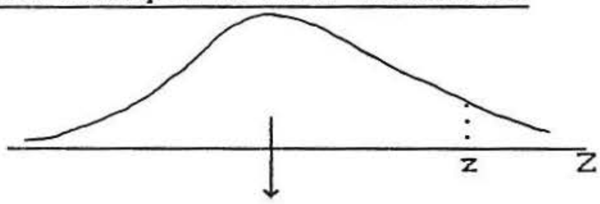
$$g_2 = \frac{m_4}{s^4} = \frac{m_4}{m_2^{\,2}}$$

Normal: $g_2 = 3$

$$m_t = \frac{\Sigma\,(y_i - \bar{y})^t}{n}$$

[momento centrado (em relação a $\bar{y}$) de ordem t]

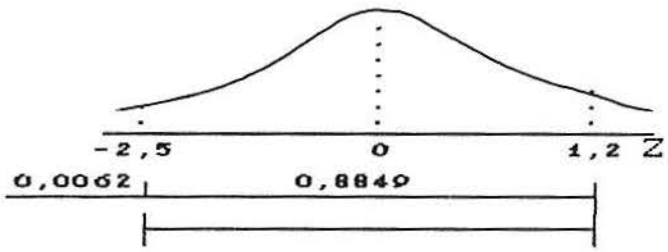
(c) Tabela da Distribuição Normal Reduzida:



$\mathbb{P}(Z \leq z)$ → Valor que a tabela fornece (Tabela 1A)

Exemplo 6:

Z	Área
-2,5	0,0062
1,2	0,8849



$\mathbb{P}(Z \leq -2{,}5) = 0{,}0062$; $\mathbb{P}(Z \geq -2{,}5) = 1 - \mathbb{P}(Z \leq -2{,}5) = 1 - 0{,}0062 = 0{,}9938$

$\mathbb{P}(Z \leq 1{,}2) = 0{,}8849$; $\mathbb{P}(Z \geq 1{,}2) = 1 - \mathbb{P}(Z \leq 1{,}2) = 1 - 0{,}8849 = 0{,}1151$

$\mathbb{P}(-2{,}5 \leq Z \leq 1{,}2) = \mathbb{P}(Z \leq 1{,}2) - \mathbb{P}(Z \leq -2{,}5) = 0{,}8849 - 0{,}0062 = 0{,}8787$

(d) Cálculo de probabilidade usando a distribuição normal adrão:

Exemplo 7: Vamos admitir que os teores de N(%) em cana planta enham distribuição normal com média

$$\mu = \frac{83,17}{40} = 2,08\% \quad ; \quad \sigma^2 = 0,0024 \quad ; \quad \sigma = 0,05\%$$

i) Qual a probabilidade de termos plantas com % N inferior a 2,03.

ii) Superior a 2,10.

iii) Entre 2,03 e 2,10.

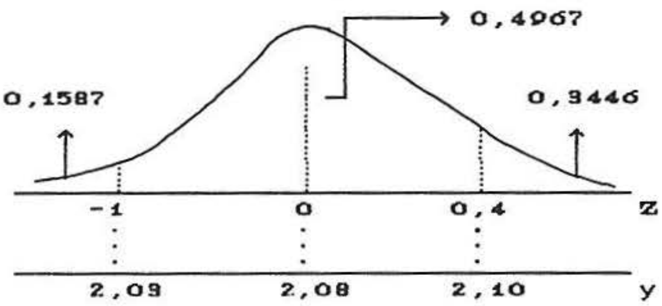
iv) Qual é o valor de % de N tal que 30% das plantas apresentam % N inferior a ele.

v) Qual é o valor de % de N superado por 10% das plantas.

Solução:

i) $\mu = 2,08 \quad ; \quad \sigma = 0,05 \qquad Z = \dfrac{y - \mu}{\sigma}$

$$y = 2,02 \Rightarrow Z = \frac{2,03 - 2,08}{0,05} = \frac{-0,05}{0,05} = -1$$



$\mathbb{P}(y<2,03) = \mathbb{P}(Z<-1) = 0,1587$

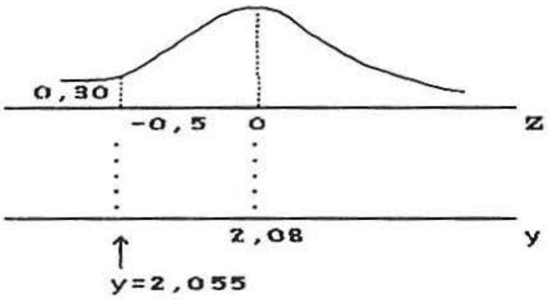
ii) $y = 2{,}10 \Rightarrow Z = \dfrac{2{,}10 - 2{,}08}{0{,}05} = \dfrac{0{,}02}{0{,}05} = 0{,}4$

$\mathbb{P}(y>2{,}10) = \mathbb{P}(Z>0.4) = 1 - \mathbb{P}(Z<0{,}4) = 1 - 0{,}6554 = 0{,}3446$

iii) $\mathbb{P}(2{,}03<y<2{,}10) = \mathbb{P}(-1<Z<0{,}4) = \mathbb{P}(Z<0{,}4) - \mathbb{P}(Z<-1)$

$= 0{,}6554 - 0{,}1587 = 0{,}4967$

iv)

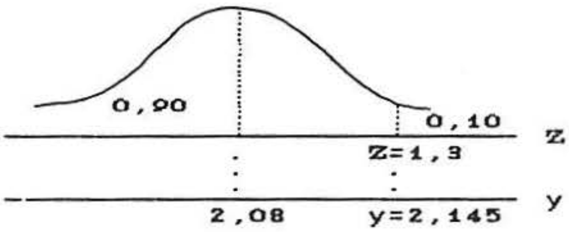


z	Área
-0,5	0,3055

$Z = \dfrac{y - \mu}{\sigma} \Rightarrow y = \sigma_z + \mu$

$y = (0{,}05)(-0{,}5) + 2{,}08 = -0{,}025 + 2{,}08$
$= 2{,}05511$

(v)



$y = \sigma_z + \mu = (0{,}05)(1{,}3) + 2{,}08 = 0{,}065 + 2{,}08 = 2{,}145$

3. AMOSTRAGEM, DISTRIBUIÇÕES AMOSTRAIS, DISTRIBUIÇÕES DE t

.1 - Amostragem

É a parte da estatística que estabelece critérios de seleção e frações (amostras) representativas de populações.

População N	→	Amostra n
	← Inferência Estatística	

População N: Universo fonte de observações

Amostra n: Parte do universo selecionada conforme algum critério que varia conforme a complexidade da população, buscando representatividade.

↓ Parâmetros: constantes que caracterizam

μ

σ^2

σ

↓ Estimadores ou Estatísticas: Variáveis

$\bar{y}$

s^2

s

→ Valores numéricos são estimativas.

Exemplos

Propriedades agrícolas: estimativa de rendimento.

Germinação de sementes: poder germinativo.

Amostragem de solo: teor de nutrientes.

Experimentos.

Tamanho da amostra: variabilidade da população

População finita
- → amostragem com reposição ⇒ admite-se população infinita
- → amostragem sem reposição ⇒ população finita

Tipos de amostragem

- Aleatória ou Probabilística: amostragem com escolha casual (por sorteio) dos elementos.
- Não-aleatória: escolha não casual dos elementos.

Amostragem aleatória simples: Todos os elementos da opulação têm igual e independente probabilidade de pertencer à mostra.

Identifica-se os ele..entos da população e sorteia-se os lementos 1 a 1 até se ter os n elementos da amostra, tilizando-se, geralmente, dispositivos construídos para tal inalidade como a tabela de números aleatórios.

Amostragem sistemática: De acordo com algum sistema.

Escolha de uma amostra sistemática:

i) Determina-se o intervalo de escolha: h = N/n;

ii) Determina-se o início aleatório, sorteando-se um elemento os primeiros h's elementos;

iii) A cada intervalo h escolhe-se um novo elemento para ertencer à amostra.

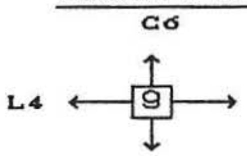
Amostragem estratificada: Utilizada quando a população é eterogênea. Primeiro subdivide-se a população em partes omogêneas (estratos) e depois retira-se uma amostra aleatória imples de cada estrato, que poderá ser uniforme, proporcional ao amanho do estrato ou de partilha ótima (variância e/ou custo).

Exemplo 8:

População de 100 leitões

$\boxed{N = 100}$ identificados 00 a 99

Amostra aleatória simples: n = 8



93 76 73 55 10 13 35 18 30 ...

Tabela de números aleatórios

Nº do leitão	Ganho de peso, kg
93	39
76	49
73	46
55	61
10	48
13	64
35	66
18	50

Exemplo 9:

População

2
4
6

$N = 3$

Média: $\mu = \frac{12}{3} = 4$

Variância: $\sigma^2 = \frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3} = \frac{8}{3}$

$$\sigma^2 = \frac{\Sigma (y-\mu)^2}{N}$$

Desvio Padrão: $\sigma = \sqrt{8/3}$

Vamos retirar todas as amostras possíveis de tamanho n = 2; onsiderando amostragem com reposição.

Nº	Amostra	$\bar{y}$	s^2
1	2,2	2	0
2	2,4	3	2
3	2,6	4	8
4	4,2	3	2
5	4,4	4	0
6	4,6	5	2
7	6,2	4	8
8	6,4	5	2
9	6,6	6	0

$$\bar{y} = \frac{\Sigma y}{n}$$

$$s^2 = \frac{\Sigma (y-\bar{y})^2}{n-1}$$

$\bar{y}$	f	f.$\bar{y}$	fr
2	1	2	1/9
3	2	6	2/9
4	3	12	3/9
5	2	10	2/9
6	1	6	1/9
Total	9	36	1

População de Médias

$$\boxed{\mu_{s^2} = \frac{24}{9} = \frac{8}{3} = \sigma^2}$$

Média das médias amostrais:

$$\mu_{\bar{y}} = 36/9 = 4 \qquad \left[\mu_{\bar{y}} = \frac{\Sigma f\bar{y}}{\Sigma f}\right]$$

Variância das médias amostrais (variância da média):

$$\sigma^2_{\bar{y}} = \frac{\Sigma f(\bar{y}-\mu_{\bar{y}})^2}{\Sigma f} = \frac{2(2-4)^2 + 2(3-4)^2 + \ldots + 1(6-4)^2}{9} = \frac{12}{9} = \frac{4}{3} = \frac{8/3}{2}$$

Desvio ou erro padrão das médias amostrais (erro padrão da média):

$$\sigma_{\bar{y}} = \sqrt{\sigma^2_{\bar{y}}} = \sqrt{\frac{8/3}{2}} = \frac{\sqrt{8/3}}{\sqrt{2}}$$

-Propriedades da distribuição amostral da média

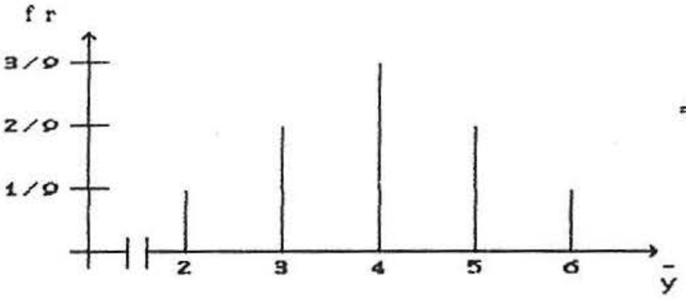
i) $\mu_{\bar{y}} = \mu$ ($\bar{y}$ é estimador imparcial; não-tendencioso de μ)

ii) $\sigma^2_{\bar{y}} = \dfrac{\sigma^2}{n}$

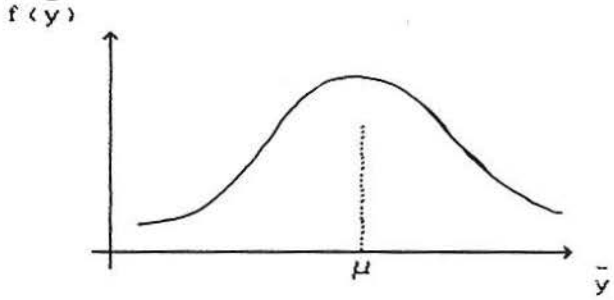
iii) $\sigma_{\bar{y}} = \dfrac{\sigma}{\sqrt{n}}$

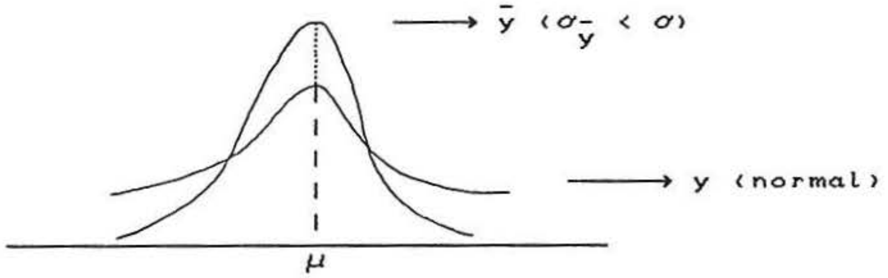
→ (Precisão de $\bar{y}$)

iv)



⇒

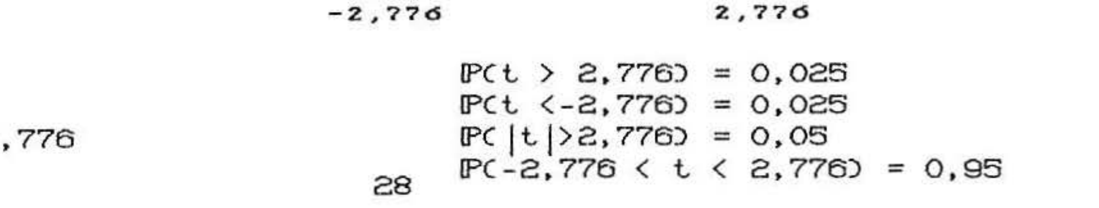




$\bar{y} \cap N\left[\mu, \dfrac{\sigma^2}{n}\right] \Rightarrow Z = \dfrac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}} = \dfrac{\bar{y} - \mu}{\sigma/\sqrt{n}} \cap N(0,1)$

$t = \dfrac{\bar{y} - \mu}{s_{\bar{y}}} = \dfrac{\bar{y} - \mu}{s/\sqrt{n}} \cap \underbrace{t(n-1)}_{\text{Distribuição de t com n-1 GL}}$

p/ n=5 ⇒ GL=4
(tabela 1)



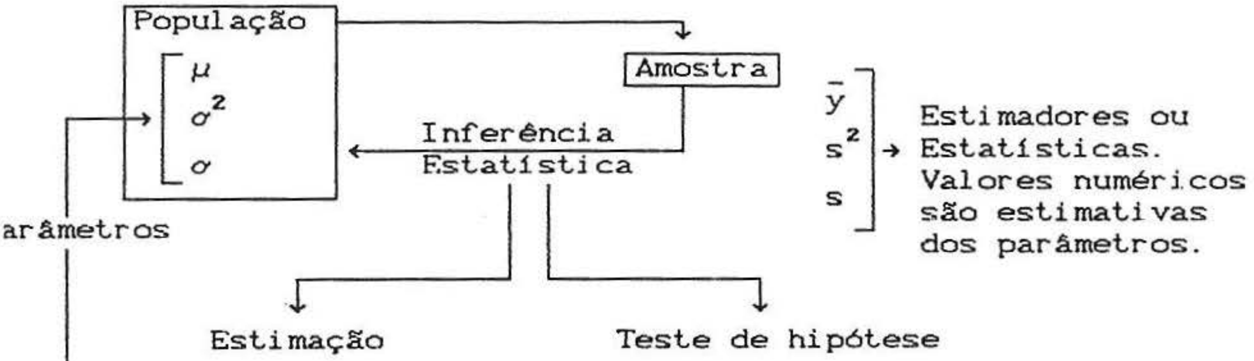
$t_{\alpha}(n-1)$

$t_{.05}(4) = 2,776$

$P(t > 2,776) = 0,025$
$P(t < -2,776) = 0,025$
$P(|t| > 2,776) = 0,05$
$P(-2,776 < t < 2,776) = 0,95$

4. ESTIMAÇÃO

.1 - Estimadores e Estimativas:



→ Podem ter diferentes estimadores. Busca-se o melhor estimador. Melhor estimador de um parâmetro é aquele
⟶ não tendencioso
⟶ consistente
⟶ eficiente ou de variância mínima.

.2 - Estimação:

Por ponto: $\bar{y}$ ⟵ um ponto para estimar o parâmetro

$\bar{y} \pm \boxed{s_{\bar{y}}}$ → precisão

Por intervalo: Intervalo de Confiança (IC)

Calcula-se um intervalo dentro do qual tem-se eterminada confiança de que o parâmetro esteja incluso. Usa-se a istribuição amostral do estimador

IC para μ

σ desconhecido: $\underbrace{\text{IC } 100(1-\alpha)\%}_{\text{coeficiente de confiança}} = \bar{y} \pm \underbrace{t\alpha(n-1)}_{\text{valor tabelado de t com probabilidade } \alpha \text{ e } (n-1) \text{ GL}} s_{\bar{y}}$ ⟶ precisão do estimador

(→ estimador: $\bar{y}$)

$$\frac{\bar{y} - \mu}{s_{\bar{y}}} \cap t(n-1) \Rightarrow \mathbb{P}[\bar{y} - t\alpha(n-1)s_{\bar{y}} \leq \mu \leq \bar{y} + t\alpha(n-1)s_{\bar{y}}] = 1-\alpha$$

· σ conhecido ou Grandes Amostras (n > 30)

$$IC\ 100(1-\alpha)\% = \bar{y} \pm Z_{\alpha/2}\ \sigma_{\bar{y}} \longleftarrow s_{\bar{y}} \longrightarrow \frac{s}{\sqrt{n}}$$

$$\sigma_{\bar{y}} \longrightarrow \frac{\sigma}{\sqrt{n}}$$

<u>I.C:</u>

$\alpha = 0{,}05 \Rightarrow$ IC 95% $\Rightarrow t_{.05(n-1)}$; $Z_{.025} = 1{,}96$

$\alpha = 0{,}01 \Rightarrow$ IC 99% $\Rightarrow t_{.01(n-1)}$; $Z_{.005} = 2{,}58$

<u>Exemplo 10:</u> Os dados seguintes referem-se a Brix no Caldo da variedade CB 41-14:

19,68	19,58	19,88	20,01	n = 12
19,91	19,31	20,11	19,31	
19,61	19,61	20,21	19,41	GL = 12-1 = 11

$\bar{y} = 19{,}72 \longleftarrow$ estimativa da verdadeira média de Brix no Caldo da variedade CB 41-14

$s = 0{,}30$

$s_{\bar{y}} = s/\sqrt{n} = 0{,}30/\sqrt{12} = 0{,}30/3{,}46 = 0{,}09$; $\bar{y} \pm s_{\bar{y}} = 19{,}72 \pm 0{,}09$

Estimação por intervalo

$$IC\ 95\%\ p/\ \mu = \bar{y} \pm t_{.05(11)}\ s_{\bar{y}} = 19{,}72 \pm (2{,}201)(0.09)$$

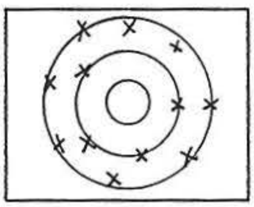
$$= 19{,}72 \pm 0{,}20 \begin{cases} \rightarrow 19{,}92 \\ \rightarrow 19{,}52 \end{cases}$$

$$\therefore \quad 19{,}52 \leq \mu \leq 19{,}92$$

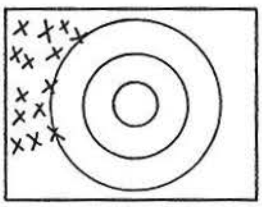
Temos uma confiança de 95% que a verdadeira média de Brix no aldo da variedade CB 41-14 esteja no intervalo [19,52;19,92].

<u>Precisão:</u> Variabilidade

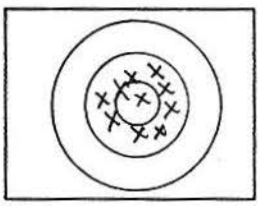
<u>Exatidão:</u> Imparcialidade; Não-tendenciosidade; Sem vício



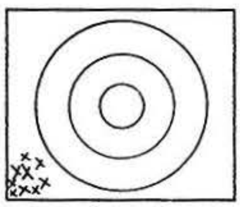
Não-tendenciosa
Baixa precisão



Tendenciosa
Baixa precisão

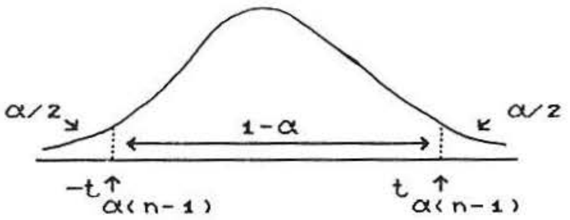


Não-tendenciosa
Alta precisão



Tendenciosa
Alta-precisão

Estruturação do intervalo de confiança para μ:



$$t = \frac{\bar{y} - \mu}{s_{\bar{y}}} = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

$$\mathbb{P}\left[-t_{\alpha(n-1)} < t < t_{\alpha(n-1)}\right] = 1 - \alpha$$

$$\mathbb{P}\left[-t < \frac{\bar{y} - \mu}{s_{\bar{y}}} < t\right] = 1 - \alpha$$

$$\mathbb{P}\left[-ts_{\bar{y}} < \bar{y} - \mu < ts_{\bar{y}}\right] = 1 - \alpha$$

$$\mathbb{P}\left[-ts_{\bar{y}} - \bar{y} < -\mu < ts_{\bar{y}} - \bar{y}\right] = 1 - \alpha$$

$$\mathbb{P}\left[\bar{y} + ts_{\bar{y}} > \mu > \bar{y} - ts_{\bar{y}}\right] = 1 - \alpha$$

$$\mathbb{P}\left[\bar{y} - t_{\alpha(n-1)}\, s_{\bar{y}} > \mu > \bar{y} + t_{\alpha(n-1)}\, s_{\bar{y}}\right] = 1 - \alpha$$

$$\therefore \boxed{\text{IC } 100(1-\alpha)\% \text{ p/ } \mu = \bar{y} \pm t_{\alpha(n-1)}\, s_{\bar{y}}}$$

<u>Exemplo 11:</u>

População de 100 leitões $\begin{cases} \mu = 51 \text{ kg} \\ \sigma = 10 \text{ kg} \end{cases}$ (Exemplo didático pois se se conhece o parâmetro não há necessidade de se estimar)

Amostra aleatória simples de tamanho $n = 8$

$\bar{y} = 52{,}9$ kg

$s = 9{,}6$ kg

$$s_{\bar{y}} = \frac{s}{\sqrt{n}} = \frac{9{,}6}{\sqrt{8}} = 3{,}4 \qquad \text{GL} = 7$$

$$\text{IC } 95\% \text{ p/ } \mu = \bar{y} \pm t_{.05(7)}\, s_{\bar{y}}$$

$$= 52{,}9 \pm (2{,}365)(3{,}4)$$

$$= 52{,}9 \pm 8 \begin{cases} \to 60{,}9 \text{ kg} \\ \to 44{,}9 \text{ kg} \end{cases}$$

$$\text{IC } 95\% \text{ p/ } \mu = [44{,}9 \text{ kg} \;;\; \underset{\substack{\uparrow \\ 50 \text{ kg}}}{} 60{,}9 \text{ kg}]$$

Tem-se uma confiança de 95% que o ganho médio de peso da opulação de leitões esteja no intervalo [44,9 kg ; 60,9 kg].

.3 - Tamanho da amostra para estimar a média μ:

i) Quando σ é conhecido

$$IC = \bar{y} \pm \underbrace{Z \frac{\sigma}{\sqrt{n}}}_{d}$$

d → semiamplitude do IC = precisão

$$d = \frac{Z\,\sigma}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} = \frac{Z\,\sigma}{d} \quad \Rightarrow \quad n = \left[\frac{Z\,\sigma}{d}\right]^2$$

Exemplo 12: Qual o tamanho da amostra necessária para se estimar média μ de uma população, com 95% de confiança, sabendo-se que $= 4$ e $d = 0,5$.

$$IC\ 95\% \Rightarrow \alpha = 0,05 \qquad Z_{\alpha/2} = Z_{.025} = 1,96$$

$$n = \left[\frac{(1,96)(4)}{0,5}\right]^2 = 246$$

i) Quando σ é desconhecido

$$IC = \bar{y} \pm \underbrace{t \frac{s}{\sqrt{n}}}_{d} \qquad d = t \frac{s}{\sqrt{n}} \quad \Rightarrow \quad n = \left[\frac{t\ s}{d}\right]^2$$

$t \rightarrow \alpha$, $GL = n-1$

Não retiramos a amostra e precisamos de s e do GL. Então etira-se uma amostra de tamanho n' obtem-se uma estimativa s' e sando GL = n'-1 determina-se n por

$$\boxed{n = \left[\frac{t_{\alpha(n'-1)}\ s'}{d}\right]^2}$$

Se $n \leq n'$ a amostra já é suficiente, caso contrário devemos mplementar a amostra.

Exemplo 13: Leitões ; $n' = 8$ $s' = 9{,}6$ kg . A amostra é uficiente p/ estimar μ com 95% de confiança e precisão de 3 kg.

$d = 3$ kg ;

$$n = \left[\frac{(2{,}365)(9{,}6)}{3} \right]^2 = 57$$

$t_{.05(n'-1)} = t_{.05(7)} = 2{,}365$ Necessita-se mais 49 animais.

5-TESTE DE HIPÓTESE

5.1-Hipóteses, Erros de Conclusão e suas probabilidades

H_o: Hipótese de nulidade (Hipótese Estatística definida operacionalmente)

Não há diferença

Ex.: $H_o\ \mu: = \mu_o$ —— padrão

└─ constante qualquer

H_a: Hipótese Alternativa (Definição Operacional da Hipótese de Pesquisa)

Há diferença

Ex : H_a: $\mu \neq \mu_o$ (+ geral)

H_a: $\mu > \mu_o$

H_a: $\mu < \mu_o$

H_o: $\mu = \mu_o$ (verdadeira)
- Aceitação de H_o: conclusão correta
- Rejeição de H_o: Erro de conclusão do tipo I
  - Probabilidade = $\alpha \Rightarrow$ nível de probabilidade ou nível de significância

H_o: $\mu = \mu_o$ (falsa)
- Aceitação de H_o: Erro de conclusão do tipo II
  - probabilidade = β
- Rejeição de H_o : conclusão correta
  - poder do teste = $1 - \beta$

Níveis de significância usuais:

$\left\{\begin{array}{l}\alpha = 0{,}05 \text{ ou } 5\% \Rightarrow \text{condições normais de experimentação} \\ \alpha = 0{,}01 \text{ ou } 1\% \Rightarrow \text{trabalhos de grande responsabilidade}\end{array}\right.$

fixados "a priori"

Relação entre α e β

α : Minimizar

β : Manter em níveis razoáveis

Teste de hipótese

Teste de H_0 (formulada para a população) com base na evidência da amostra decisão → rejeita-se $H_0 \Rightarrow$ Falsidade de H_0
→ aceita-se $H_0 \Rightarrow$ Falta de evidência para se rejeitar H_0

usa-se distribuição amostral do estimador

Hipótese Estatística : É a especificação do valor de um parâmetro ou do relacionamento entre diferentes parâmetros.

Erro de conclusão mais grave : Aquele de consequência mais grave.

Exemplo 14 :

(I) H_0 : O réu é inocente

H_0	decisão	
	aceitação	rejeição
Verdadeira	conclusão correta	erro tipo I
Falsa	erro tipo II	conclusão correta

Rejeitar H_0 verdadeira (erro tipo I) $\Rightarrow$ inocente condenado $\Rightarrow$ mais grave , pois perdas não poderão ser reparadas. Aceitação de H_0 falsa (erro tipo II) $\Rightarrow$ culpado absolvido $\Rightarrow$ novo julgamento $\Rightarrow$ erro poderá ser reparado.

(II) Ho : O aluno deve ser aprovado

Erro tipo I ⇒ mais grave sem reparo (reprovam quem deve ser aprovado.

Ho	decisão	
	aceitação	rejeição
Verdadeira	conclusão correta	erro tipo I
Falsa	erro tipo II	conclusão correta

Erro tipo II ⇒ menos grave ⇒ com reparo (aprovar quem deve ser reprovado).

(III) Ho: Duas cultivares de trigo (nova e padrão) não diferem em rendimento ($\mu_N = \mu_P$)

Ho	decisão	
	aceitação	rejeição
Verdadeira	conclusão correta	erro tipo I
Falsa	erro tipo II	conclusão correta

Num experimento (amostra) Nova variedade apresentou rendimento médio 10% superior a padrão.

Rejeitar Ho verdadeira ⇒ substituir padrão pela nova quando as duas não são diferentes ⇒ com o tempo produtor vai verificar que substituição não mostrou melhoria assegurada pela pesquisa ⇒ perda de confiabilidade nos resultados pela pesquisa ⇒ consequência grave.

Aceitar Ho falsa ⇒ considerar duas cultivares não diferentes quando são. Em novos experimentos com maior precisão pode-se identificar diferenças e daí concluir-se corretamente ⇒ consequência - grave.

5.2-Etapas da execução de um teste de hipótese

(a) Formulação das Hipóteses : Ho

Ha

(b) Especificação do nível de significância

$\alpha = 0{,}05$ ou 5%

$\alpha = 0{,}01$ ou 1%

(c) Escolha da estatística para testar Ho com base na distribuição amostral do estimador.

$$z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \cap N(0,1) \qquad t = \frac{\bar{y} - \mu}{s / \sqrt{n}} \cap t(n-1)$$

(d) Estabelecimento da regra de decisão ou seja determinação da região de rejeição de Ho.

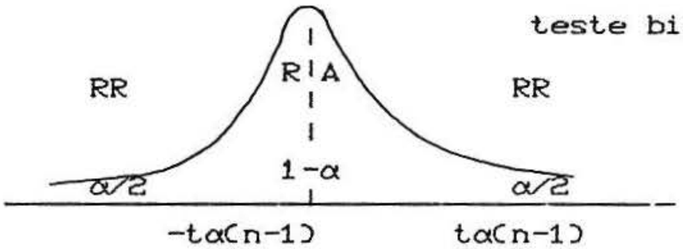
Com base na distribuição teórica da estatística escolhida para o teste.

estatística t

α

Ha : $\neq$

teste bilateral



RA : região de aceitação de Ho

RR : região de rejeição de Ho

Rejeita-se Ho se $|t \text{ calculado}| > t\alpha(n-1)$

(e) Obtenção dos dados de observação e cálculo da estatística escolhida para o teste.

(f) Conclusão estatística (decisão) : Rejeição de Ho , quando a estatística calculada se situar em RR.

(g) Conclusão prática ou em termos do material estudado.

5.3-<u>Comparação da média com um valor padrão</u>

população		amostra	
μ	$\longrightarrow$	$\bar{y}$, s, n	Ho : $\mu = \mu_o$ Ha : $\mu \neq \mu_o$

$$\text{como} \quad \frac{\bar{y} - \mu}{s_{\bar{y}}} \underset{\text{Ho}}{\overset{\text{sob}}{=}} \frac{\bar{y} - \mu_o}{s_{\bar{y}}} \quad \cap\ t\ (n-1)$$

$$\text{onde } s_{\bar{y}} = s \,/\, \sqrt{n}$$

Rejeita-se Ho se |t calculado| > t_α (n-1)

Se σ é conhecido (ou para grandes amostras , n > 30) usa-se z , onde

$$z = \frac{\bar{y} - \mu_o}{\sigma_{\bar{y}}} \quad \text{onde } \sigma_{\bar{y}} = \sigma \,/\, \sqrt{n}$$

Rejeita-se Ho μ |z calculado| > $z_{\alpha/2}$

∘ No teste de hipótese faz-se comparação de valores e para o caso comparam-se $\bar{y}$ e μ_o e a diferença observada poderá ser considerada :

(1) Real : → Diferença significativa ; $\alpha = 0{,}05$ ($P < 0{,}05$ de se obter diferença de tal magnitude por mero acaso)
→ Diferença muito significativa , $\alpha = 0{,}01$ ($P < 0{,}01$ de se obter diferença de tal magnitude por mero acaso)

] rejeita-se Ho

(2) Devida a variação de amostragem : Diferença não significativa ($P > 0{,}05$ de se obter diferença de tal magnitude por mero acaso)

] aceita-se Ho

Exemplo 15: Consideremos que o padrão médio de Brix no caldo para variedades de cana-de-açucar seja 20. Verifiquemos com base nos dados de uma amostra de tamanho 12 (exemplo de estimação) se a variedade CB 41-14 difere ou não do padrão.

(I) $H_o : \mu = 20$ [A média de Brix da variedade CB 41-14 não difere da média padrão.]

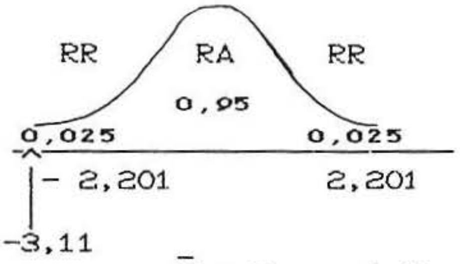
$H_a : \mu \neq 20$ [A média de Brix da variedade CB 41-14 difere da média padrão.]

(II) $\alpha = 0{,}05$ ou 5%

(III) $t = \dfrac{\bar{y} - \mu}{s_{\bar{y}}} \cap t\,(n-1)$

(IV) $n = 12$; $\alpha = 0{,}05$ e t ; $H_a \neq$ (teste bilateral)

$t_{.05(11)} = 2{,}201$



rejeita-se Ho se
|t calculado| > t.05(11) = 2,201

$$(v)\quad t = \frac{\bar{y} - \mu}{s_{\bar{y}}} \xrightarrow[\mu=\mu_o=20]{\text{sob Ho}} t = \frac{\bar{y} - \mu_a}{s_{\bar{y}}} = \frac{\bar{y} - 20}{s_{\bar{y}}}$$

Amostra: $n = 12$; $\bar{y} = 19,72$; $s = 0,30$; $s_{\bar{y}} = s/\sqrt{n} = 0,30/\sqrt{12} = 0,09$

$$t = \frac{19,72 - 20}{0,09} = \frac{-0,28}{0,09} = -3,11$$

(VI) $|t| \geq 3,11 > t_{.05}(11) = 2,201$

A média de Brix da variedade CB 41-14 difere da média padrão. Rejeita-se Ho

(VII) A variedade CB 41-14 se distingue do padrão médio de Brix das variedades de cana-de-açucar , apresentando , Brix médio inferior ao padrão .

IC 95% para μ : $\left[19,52 \; ; \; 19,92\right]^{\overset{20}{\uparrow}}$

5.4-Comparação de dois tratamentos

Experimento

tratamento → unidade (UE) experimental → efeito → y

y — variável aleatória

y — variável resposta

- tratamento: -doses de N; -rações
- unidade experimental: -parcela de campo; -animal
- y: -rendimento de trigo; -ganho de peso

usa-se →

comparação emparelhadas

- **grupos independentes**
  - **obrigatórios**
    - (1) <u>Repetições</u>: Mais de uma UE por tratamento . Para poder verificar diferenças entre tratamentos.
    - (2) <u>Casualização</u>: Atribuição por sorteio dos tratamentos às UE. Comparação válida dos tratamentos.
- **opcional**
  - (3) <u>Bloqueamento</u> <u>ou</u> <u>pareamento</u>: Formação de blocos (pares) homogêneos de UE , antes da aplicação dos tratamentos . Para controlar variabilidade intrínseca das UE ⇒ maior precisão.

5.4.1-Comparação de dois tratamentos em grupos independentes ou amostras independentes.

conjunto de UE			tratamentos
1 2 3 4 5 6	repetição	casualização	A B

tratamentos	
A	B
2	1
4	3
5	6

Escolha casual das UE que serão submetidas aos tratamentos A e B.

população 1

μ_1, σ_1^2

amostra 1(grupo 1)

tratamento 1

n_1 repetições

$y_{1j} \rightarrow$ observação j do tratamento 1

$\bar{y}_1$

SQ_1

s_1^2

população 2

μ_2, σ_2^2

amostra 2(grupo 2)

tratamento 2

n_2 repetições

$y_{2j} \rightarrow$ observação j do tratamento 2

$\bar{y}_2$

SQ_2

s_2^2

suposição : $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$H_o : \mu_1 = \mu_2 \Leftrightarrow H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 \neq \mu_2$

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_d} \cap t(n_1 + n_2 - 2)$$

$s_d \longrightarrow$ erro padrão da diferença entre duas médias

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_d} \xrightarrow[H_o]{\text{sob}} t = \frac{\bar{y}_1 - \bar{y}_2}{s_d}$$

$$\text{onde } s_d = \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_1}\right)}$$

$s^2 \longrightarrow$ variância ponderada para as 2 amostras.

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{SQ_1 + SQ_2}{n_1 + n_2 - 2}$$

$$\boxed{s^2 = \frac{SQ}{GL} \quad SQ = (GL)s^2}$$

se $|t \text{ calculado}| > t\alpha(n_1 + n_2 - 2) \Rightarrow$ rejeita-se H_o

se $n_1 = n_2 = n \Rightarrow s_d = \sqrt{2s^2/n}$

$$s^2 = \frac{(n - 1)s_1^2 + (n - 1)s_2^2}{2(n - 1)} = \frac{s_1^2 + s_2^2}{2} = \frac{SQ_1 + SQ_2}{2(n - 1)}$$

então rejeita-se H_o se $|t \text{ calculado}| > t\alpha\left[2(n - 1)\right]$

Exemplo 16:

Experimento com duas variedades de cana-de-açúcar (1 e 2)
y = Kg / parcela

Amostras independentes ou grupos independentes.

<u>vairedade 1</u>	<u>variedade 2</u>
518	458
524	550
420	384
486	494
515	418
$n_1=5$	$n_2=5$
$\bar{y}_1=492,6$	$\bar{y}_2=460,8$
$SQ_1=7447,2$	$SQ_2=16796,8$
$s_1^2=1861,8$	$s_2^2=4199,2$
$s_1=43,15$	$s_2=64,80$

[suposição: $\sigma_1^2 = \sigma_2^2$]

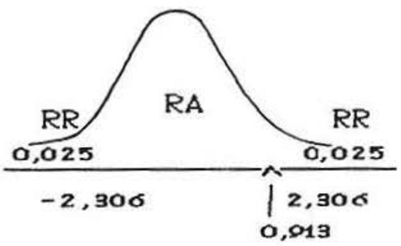
(I) $H_o : \mu_1 = \mu_2$ $\quad (H_o : \mu_1 - \mu_2 = 0)$
$\quad H_a : \mu_1 \neq \mu_2$

(II) $\alpha = 0,05$

(III) $$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_d} \cap t(n_1 + n_2 - 2)$$

$s_d \longrightarrow$ erro padrão da diferença entre duas médias.

(IV) $H_a \neq$; $\alpha = 0,05$; $n_1 + n_2 - 2 = 10 - 2 = 8$
$t_{.05(8)} = 2,306$



Rejeita-se H_o se
$|t\ \text{calculado}| > t_{.05(8)} = 2,306$

(v) $t = \dfrac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{s_d} \xrightarrow[Ho]{sob} t = \dfrac{\bar{y}_1 - \bar{y}_2}{s_d}$

$$s_d = \sqrt{\underbrace{s^2}_{\text{variância ponderada}}\left(\frac{1}{n1} + \frac{1}{n2}\right)} \quad ; \quad s^2 = \frac{SQ1 + SQ2}{n1 + n2 - 2}$$

$$s_d = \sqrt{s^2\left(\frac{1}{n1} + \frac{1}{n2}\right)} = \sqrt{\frac{2.3030,5}{5}} = 34,82$$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_d} = \frac{492,6 - 460,8}{34,82} = \frac{31,8}{34,82} = 0,913$$

(VI) $|t| = 0,913 < t.05(8) = 2,306$

$\bar{y}_1$ e $\bar{y}_2$ não diferem significativamente

não se rejeita Ho.

(VII) As evidências amostrais não comprovam que as variedades 1 e 2 de cana-de-açúcar se diferenciam em termos de rendimento.

IC $\underbrace{100(1 - \alpha)\%}_{95\%}$ para $\mu_1 - \mu_2 = \bar{y}_1 - \bar{y}_2 \pm t.05(8)\ sd$

$=(492,6-460,8)\pm(2,306)34,82$

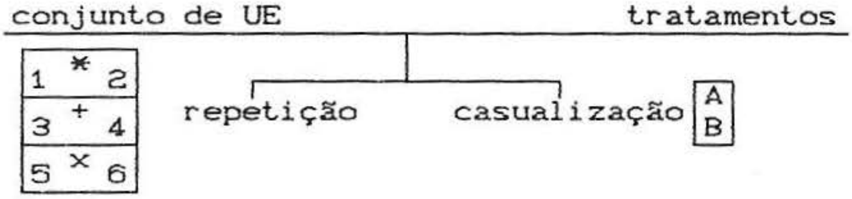
$=31,8\pm80,3$ — 112,8 / -48,5

$-48,5 \leq \mu_1 - \mu_2 \leq 112,1$

→ inclui o zero $\Rightarrow$ Aceitação de Ho : $\mu_1 = \mu_2 \Leftrightarrow$ Ho: $\mu_1 - \mu_2 = 0$

5.4.2- Comparação emparelhada de dois tratamentos

Tratamentos são aplicados a pares homogêneos de UE , sendo a atribuição por sorteio dentro de cada par.



par	tratamentos A	tratamentos B
*	2	1
+	3	4
×	6	5

par(j)	tratamento 1	tratamento 2	diferença $D_j = y_{1j} - y_{2j}$
1	y_{11}	y_{21}	$D_1 = y_{11} - y_{21}$
2	y_{12}	y_{22}	$D_2 = y_{12} - y_{22}$
:	:	:	:
:	:	:	:
n	y_{1n}	y_{2n}	$D_n = y_{1n} - y_{2n}$

Ho: $\mu_D = 0$

Ha: $\mu_D \neq 0$

$$\frac{\bar{D} - \mu_D}{s_{\bar{D}}} \sim t(n - 1)$$

$$t = \frac{\bar{D} - \mu_D}{s_{\bar{D}}} \xrightarrow[H_0]{sob} t = \frac{\bar{D}}{s_{\bar{D}}}$$

$\bar{D}$ → diferença média ou média das diferenças

$s_{\bar{D}}$ → erro padrão da diferença média

$$\bar{D} = \frac{\Sigma D_j}{n} = \bar{y}_1 - \bar{y}_2$$

$$s_{\bar{D}} = \sqrt{\frac{s_D^2}{n}} = s_D / \sqrt{n}$$

s_D → desvio padrão das diferenças

$$s_D^2 = \frac{\Sigma (D_j - \bar{D})^2}{n - 1} = \frac{\Sigma D_j^2 - (\Sigma D_j)^2 / n}{n - 1}$$

s_D^2 → variância das diferenças

Rejeita-se Ho se $|t \text{ calculado}| > t\alpha(n - 1)$

Exemplo 17:

Num estudo de nutrição de leitões , empregou-se o método do emparelhamento para estudar o efeito da vitamina B_{12} sobre o aumento de peso dos animais. Os pares de leitões eram irmãos da mesma ninhada. Os aumentos de peso observados , foram os seguintes , em Kg :

PARES	1	2	3	4	5	6	7	8
com vit B_{12}	80	84	82	87	87	89	78	88
sem vit B_{12}	72	75	75	76	80	78	89	78
diferença	8	8	7	11	7	11	-11	10

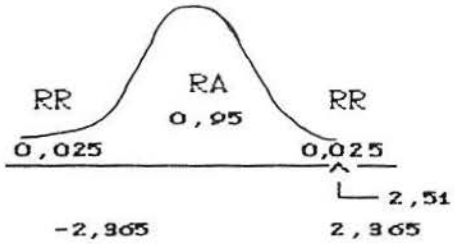
(I) $H_0 : \mu_D = 0$

$H_a : \mu_D \neq 0$

(II) $\alpha = 0,05$

(III) $\frac{\bar{D} - \mu_D}{s_{\bar{D}}} \cap t(n - 1)$

(IV) $\alpha = 0{,}05$, t , GL = n − 1 = 8 − 1 = 7



$t_{.05(7)} = 2{,}365$

Rejeita-se Ho se

$|t \text{ calculado}| > t_{.05(7)} = 2{,}365$

(V) $\bar{D} = 6{,}4$ Kg $\quad s_D = 7{,}2$

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{7{,}2}{\sqrt{8}} = 2{,}55 \qquad t = \frac{\bar{D}}{s_{\bar{D}}} = \frac{6{,}4}{2{,}55} = 2{,}51$$

(VI) $|t| = 2{,}51 > t_{.05(7)} = 2{,}365$. $\bar{D} \neq$ significativamente de zero

Rejeita-se o Ho.

(VII) A vitamina B12 tem efeito sobre o ganho de peso de leitões , ou seja a vitamina B12 é eficiente para incrementar ganho de peso de leitões .

IC 95% para $\mu_D = \bar{D} \pm t_{.05(n-1)} s_{\bar{D}}$
$= 6{,}4 \pm t_{.05(7)}(2{,}55)$
$= 6{,}4 \pm (2{,}365)(2{,}55)$
$= 6{,}4 \pm 6 \begin{cases} 12{,}4 \\ 0{,}4 \end{cases}$

$0{,}4 \text{ Kg} \leq \mu_D \leq 12{,}4 \text{ Kg}$

0 ⇒ vit B12 eficiente

<u>Exemplo 18:</u>

	com herbicida	sem herbicida
par1		
:		
:		
par15		

$n = 15$

$\bar{D} = 400$ Kg/ha

$s_{\bar{D}} = 80$ Kg/ha

Se o custo do herbicida é equivalente a 200 Kg/ha é conveniente usar herbicida ?

(I) Ho : $\mu_D = 200$

Ha : $\mu_D \neq 200$

(II) $\alpha = 0{,}05$

(III) $\dfrac{\bar{D} - \mu_D}{s_{\bar{D}}} \cap t(n-1)$ se $\mu_D = 0 \Rightarrow \dfrac{\bar{D}}{s_{\bar{D}}}$

se $\mu_D \neq 0 \Rightarrow \dfrac{\bar{D} - \mu_D}{s_{\bar{D}}}$

(IV) $t_{.05}(14) = 2{,}145 \Rightarrow$ Rejeita-se Ho se

$|t \text{ calculado}| > t_{.05}(14) = 2{,}145$

(V) $t = \dfrac{\bar{D} - \mu_D}{s_{\bar{D}}} = \dfrac{400 - 200}{80} = \dfrac{200}{80} = 2{,}5$

(VI) Decisão :

$|t| = 2{,}50 > t_{.05}(14) = 2{,}145$

A diferença média , difere significativamente de 200

($P < 0{,}05$)

Rejeita-se Ho

(VII) Conclusão :

É conveniente o uso do herbicida , dado que o acréscimo na produção decorrente de seu uso supera o seu custo.

$$\begin{aligned} \text{IC 95\% para } \mu_D &= \bar{D} \pm t_{.05}(14)\, s_{\bar{D}} \\ &= 400 \pm (2{,}145)(80) \\ &= 400 \pm 172 \begin{cases} 572 \\ 228 \end{cases} \end{aligned}$$

$228 \text{ Kg/ha} \leq \mu_D \leq 572 \text{ Kg/ha}$

200 Kg/ha $\Rightarrow$ Conveniente uso de herbicida

5.5- Testes Unilaterais

- Critério no seu uso : Bilaterais são gerais ; alguns casos é válido usar unilaterais.

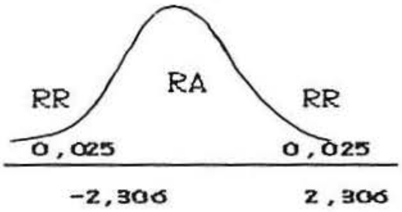
- Mais eficientes para rejeição de Ho.

Exemplo 19:

$H_o : \mu_1 = \mu_2$; GL = 8 ; $\alpha = 0{,}05$

→ $H_a : \mu_1 \neq \mu_2$

$t_{.05(8)} = 2{,}306$ } bilateral ou desconsiderando o sinal



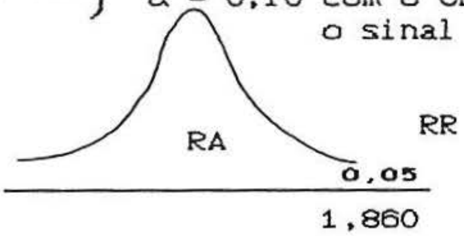
Rejeita-se H_o se

$|t \text{ calculado}| > 2{,}306$

→ $H_a : \mu_1 > \mu_2$

$t = 1{,}860$ } $\alpha = 0{,}05$ com 8 GL unilateral ou considerando o o sinal

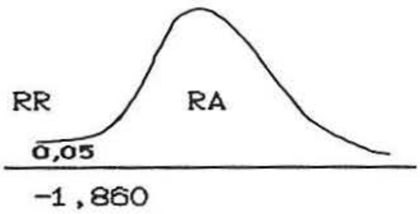
$\alpha = 0{,}10$ com 8 GL bilateral ou desconsiderando o sinal



Rejeita-se H_o se

$t > 1{,}860$

→ $H_a : \mu_1 < \mu_2$



Rejeita-se H_o se

$t < -1{,}860$

Exemplo 20:

Os aumentos de peso de 15 frangos alimentados com uma ração especial em um período de um mês foram os seguintes em Kg :1,3 ; 1,0 ; 1,1 ; 0,8 ; 1,1 ; 1,1 ; 0,9 ; 0,8 ; 0,8 ; 1,1 ; 0,9 ; 1,0 ; 0,9; 0,9 ; 1,1 . A experiência tem mostrado que com uma dieta-padrão , o ganho médio de peso em frangos em igual período é de 0,9 Kg. Pode-se concluir que a ração especial incrementa o ganho de peso.

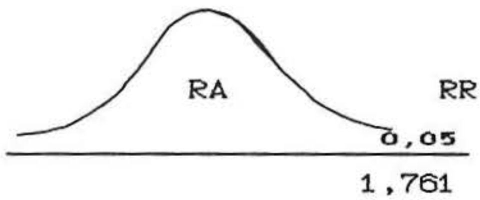
(I) $H_o : \mu = 0,9$ Kg

$H_a : \mu > 0,9$ Kg

(II) $\alpha = 0,05$

(III) Estatística t

(IV) $\alpha = 0,05$ GL = 15 - 1 = 14 t unilateral = 1,761



Rejeita-se H_o se t calculado > 1,761

(V) $\bar{y} = 0,99$ Kg ; $s = 0,15$ Kg ; $n = 15$

$$s_{\bar{y}} = \frac{s}{\sqrt{n}} = \frac{0,15}{\sqrt{15}} = 0,04$$

$$t = \frac{\bar{y} - \mu}{s_{\bar{y}}} = \frac{0,99 - 0,9}{0,04} = \frac{0,09}{0,04} = 2,25$$

(VI) Decisão :

$t = 2,25 > 1,761 \Rightarrow$ A $\neq$ entre $\bar{y}$ e μ é significativa ($P<0,05$). Rejeita-se H_o.

(VII) Conclusão :

A ração especial é eficiente na alimentação de frangos , incrementando o ganho médio de peso.

5.6- Grupos independentes com variâncias desiguais

Quando não é aceita a pressuposição de que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, isto é quando é aceito que $\sigma_1^2 \neq \sigma_2^2$, a mecânica apresentada para o teste t para comparar a diferença entre as médias sofre modificação.

Ho : $\mu_1 = \mu_2$

Ha : $\mu_1 \neq \mu_2$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_d} \;, \quad s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Rejeita-se Ho quando |t calculado | > $t_\alpha(n')$, onde

$$n' = \frac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]}{\dfrac{\left[\dfrac{s_1^2}{n_1}\right]^2}{n_1 - 1} + \dfrac{\left[\dfrac{s_2^2}{n_2}\right]^2}{n_2 - 1}}$$

se $\dfrac{\text{maior variância}}{\text{menor variância}} \leq 4 \Rightarrow$ Aceita-se que $\sigma_1^2 = \sigma_2^2$

Exemplo 21:

Deseja-se saber se duas máquinas de empacotar café estão fornecendo o mesmo peso por pacote. Entretanto , como uma das máquinas é nova e a outra é velha é razoável supor-se que trabalham com deferentes variabilidades dos pesos colocados nos pacotes . As amostras disponiveis constam de sete pacotes produzidos pela máquina nova e nove produzidos pela máquina velha. Os pesos , em Kg ,desses pacotes são :

máquina nova: 0,82 ; 0,83 ; 0,79 ; 0,81 ; 0,81 ; 0,80

máquina velha: 0,79 ; 0,82 ;0,73 ; 0,74 ; 0,80 ; 0,77 ; 0,75 ; 0,84 ; 0,78

(I) $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$

(II) $\alpha = 0,05$

(III) Estatística t

(IV) N : $\bar{y}_1 = 0,81$ $\quad s_1^2 = 0,00020$ $\quad n_1 = 6$

V : $\bar{y}_2 = 0,78$ $\quad s_2^2 = 0,00135$ $\quad n_2 = 9$

$$\frac{\text{maior variância}}{\text{menor variância}} = \frac{0,00135}{0,00020} = 6,75 \Rightarrow \sigma_1^2 \neq \sigma_2^2$$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_d} = \frac{0,81 - 0,78}{0,01354} = 2,216$$

$$s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0,00020}{6} + \frac{0,00135}{9}} = 0,01354$$

(V)
$$n' = \frac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]^2}{\dfrac{\left[\dfrac{s_1^2}{n_1}\right]^2}{n_1 - 1} + \dfrac{\left[\dfrac{s_2^2}{n_2}\right]^2}{n_2 - 1}} = \frac{\left[\dfrac{0,00020}{6} + \dfrac{0,00135}{9}\right]^2}{\dfrac{\left[\dfrac{0,00020}{6}\right]^2}{5} + \dfrac{\left[\dfrac{0,00135}{9}\right]^2}{8}} = 11$$

(VI) Decisão :

$t_{.05}(n') = t_{.05}(11) = 2,201$

$|t| = 2,216 > t_{.05}(11) = 2,201$

A diferença entre $\bar{y}_1$ e $\bar{y}_2$ é significativa $(P < 0,05)$

(VII) Conclusão :

As duas maquinas de empacotar café se diferenciam , fornecendo a máquina nova um peso médio superior.

5.7- Eficiência relativa de grupos independentes e comparações emparelhadas

C.E.	G.I.
n pares	
(n -1) GL	2(n - 1)GL

CE geralmente mais eficiente que GI , e tanto mais eficiente quanto mais semelhantes os pares , isto é quando a variação entre pares é maior do que a variação dentro do par.

Par $\boxed{y_1}$ $\boxed{y_2}$ $\qquad D = y_1 - y_2$

$$\sigma^2_D = \underbrace{v(y_1)}_{\sigma^2_1} + \underbrace{v(y_2)}_{\sigma^2_2} - 2\underbrace{cov(y_1, y_2)}_{\text{medida da variação simultânea de } y_1 \text{ e } y_2}$$

$v(y_1)$= variância y_1
$v(y_2)$= variância y_2
cov (y_1, y_2) = covariância (y_1, y_2)

$$\rho_{12} = \frac{cov\ (y_1, y_2)}{\sqrt{v(y_1)v(y_2)}} = \frac{cov\ (y_1, y_2)}{\sqrt{(\sigma^2_1)(\sigma^2_2)}} = \frac{cov\ (y_1, y_2)}{\sigma_1 \sigma_2}$$

$$\therefore \rho_{12}\sigma_1\sigma_2 = cov\ (y_1, y_2)$$

coeficiente de correlação
entre y_1 e y_2 ; $-1 \leq \rho \leq 1$

Supondo $\sigma^2_1 = \sigma^2_2 = \sigma^2$, temos

$$\sigma^2_D = \underbrace{\sigma^2_1}_{\sigma^2} + \underbrace{\sigma^2_2}_{\sigma^2} - 2\ \rho_{12}\ \underbrace{\sigma_1}_{\sigma}\ \underbrace{\sigma_2}_{\sigma}$$

$$\therefore \sigma^2_D = 2\ \sigma^2 - 2\ \rho_{12}\ \sigma^2$$

Se $\rho_{12} \not\to 0 \Rightarrow$ CE eficiência = GI

Se $\rho_{12} \to 1 \Rightarrow$ eficiência CE maior do que eficiência GI e é tanto mais eficiente quanto mais próximo s_{12} de 1 , ou seja quanto mais semelhante os pares.

A eficiência do planejamento 1 em relação ao planejamento 2 é dada por

$$ER = \frac{1}{s_1^2 / s_2^2} \times 100 = \frac{s_2^2}{s_1^2} \times 100$$

Logo ER de CE em relação a GI : $ER = \dfrac{s'^2}{s_D^2} \times 100$

Onde s'^2 é a variância para grupos independentes estimada a partir de comparações emparelhadas e é dada por:

$$s'^2 = 2\, s^2 - (2\, s^2 - s_D^2) / \, 2\,(n - 1)$$

Fator de correção para o fato dos dois grupos não constituirem amostras independentes

→ variância combinada das unidades experimentais dos dois grupos

$$s^2 = \frac{SQ_1 + SQ_2}{2(n - 1)} = \frac{s_1^2 + s_2^2}{2}$$

<u>Exemplo 22:</u> exemplo de vit B_{12} ; n = 8

$$s^2 = \frac{5{,}0427 + 4{,}0333}{2} = 4{,}3380$$

$$s_D^2 = 7{,}2099$$

$$s'^2 = 2(4{,}3380) - \left[2(4{,}3380) - 7{,}2099\right] / \, 2(8 - 1)$$

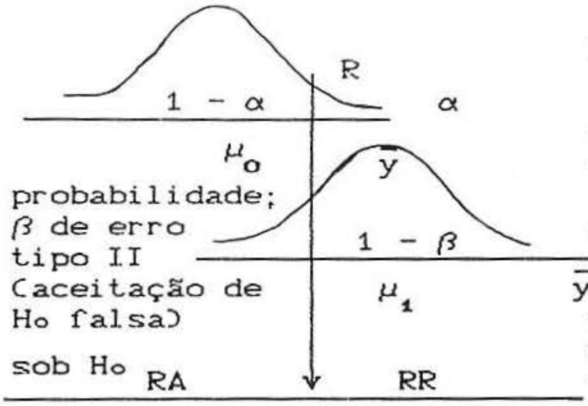
$$= 8{,}9427$$

ER CE em relação a GI

$$ER = \frac{s'^2}{s_D^2} \times 100 = \frac{8{,}9427}{7{,}2099} \times 100 = 124\ \%$$

Ganho de 24% em eficiência pelo uso de CE relativamente a GI. GI exigiria 24% a mais de repetições por tratamento para garantir uma eficiência compatível a CE. No caso 24(8) / 100 = 1,92 ≅ 2 repetições a mais deveriam ser usadas , isto é 10 repetições no total .

5.8- Relação entre α e β



distribuição de $\bar{y}$ sendo $H_o: \mu=\mu_o$ verdadeira

probabilidade de erro tipo I (rejeição de H_o quando verdadeira)

distribuição de $\bar{y}$ sendo $H_a: \mu=\mu_1$ verdadeira

poder do teste ; probabilidade de rejeitar H_o quando falsa ou aceitar H_a verdadeira

(I) Movendo a reta à direita α diminui , mas β aumenta , diminuindo consequentemente o poder do teste.

(II) Mantendo fixa a reta R e consequentemente α e movendo μ_1 para a direita diminui β e β será tanto menor quanto maior for a diferença $\mu_1 - \mu_o$.

(III) Mantendo μ_o e μ_1 fixos α e β poderão ser reduzidos aumentando o tamanho n da amostra , uma vez que quando n aumenta o erro padão da distribuição amostral de $\bar{y}$ diminui , isto é

$\sigma_{\bar{y}} = \sigma/\sqrt{n}$ diminui , fazendo com que a distribuição de $\bar{y}$ tome uma forma mais aguda.

5.9- Número de repetições a usar num experimento

Número adequado de repetições é importante no planejamento de um experimento :

- Poucas repetições $\Rightarrow$ pode-se n descobrir diferenças importantes
- Muitas repetições $\Rightarrow$ desperdicio de tempo e material .
- Deve-se ter número suficiente de repetições para detectar como significativa a diferença no efeito de dois tratamentos , se ela existir.

Para se determinar o número de repetições necessita-se:

- Estimativa de variabilidade : s^2 ou CV.
- Tamanho da diferença entre médias a ser detectada como significativa : δ , expressa com % da média geral .

- Nível de significância : α

- Segurança com que se deseja detectar a diferença : poder do teste , $P = 1 - \beta$.

- Teste unilateral ou bilateral.

Dois grupos independentes :

população 1

μ_1, σ_1^2

$\downarrow$

amostra 1

n_1

$\bar{y}_1, s_1^2, SQ_1$

população 2

μ_2, σ_2^2

$\downarrow$

amostra 2

n_2

$\bar{y}_2, s_2^2, SQ_2$

suposição

$\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$Ho : \mu_1 = \mu_2 \Leftrightarrow Ho : \mu_1 - \mu_2 = 0$$

$$Ha : \mu_1 \neq \mu_2 \Leftrightarrow Ha : \mu_1 - \mu_2 \neq 0 \quad \left[\text{bilateral}\right]$$

$$Ha : \mu_1 > \mu_2 \Leftrightarrow Ha : \mu_1 - \mu_2 > 0 \quad \left[\text{unilateral}\right]$$

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{sd} \cap t(n_1 + n_2 - 2)$$

$\hookrightarrow$ erro padrão da diferença entre duas médias

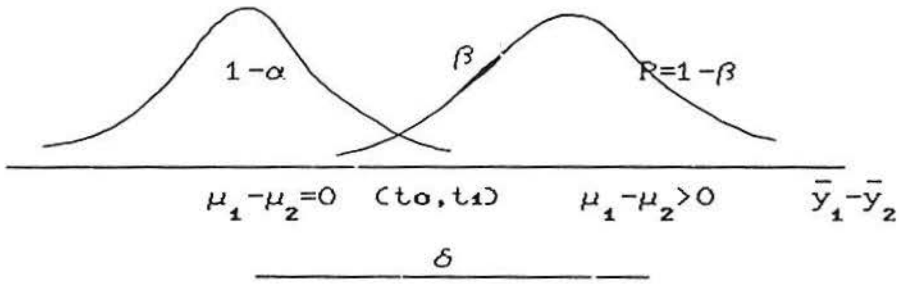
$$sd = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$\downarrow$

variância ponderada

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{SQ_1 + SQ_2}{n_1 + n_2 - 2}$$

se $n_1 = n_2 = n \Rightarrow s_d = \sqrt{2s^2/n}$ e $s^2 = \dfrac{s_1^2 + s_2^2}{2}$



DECISÕES:

(I) Ho será rejeitada quando t calculado > to

↓

[valor tabelado de t ignorando o sinal ⇒ t bilateral]

sob Ho $t = \dfrac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2 = 0)}{s_d} = \dfrac{\bar{y}_1 - \bar{y}_2}{s_d} = t_0$

(II) Ha será aceita quando t calculado > t_1

↓

[o valor tabelado de t considerando o sinal ⇒ t unilateral (do lado esquerdo da curva)]

sob Ha $t = \dfrac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{s_d} = t_1$

$t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{s_d} \Rightarrow t_0\, s_d = \bar{y}_1 - \bar{y}_2$ (1)