


ANÁLISE DE DADOS ECOLÓGICOS

ISCED, Lubango, Março 2016



J. Paulo Sousa
Laboratory of Soil Ecology and Ecotoxicology
Centre for Functional Ecology
Universidade de Coimbra, Portugal
jps@zoo.uc.pt
<http://cfe.uc.pt/paulosousa>
<http://www.facebook.com/labsolos>

© J.P. Sousa Slide 1 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Tópicos do curso – Semana 1

1. Revisão dos conceitos básicos em bioestatística
 - testes de hipóteses para uma ou duas populações
 - análise de variância e desenho experimental
 - regressão linear simples e correlação
2. Regressão linear múltipla
 - exploração dos dados
 - avaliação de colinearidade entre variáveis explicativas
 - interação entre variáveis explicativas
 - interpretação dos resultados
3. Modelos Lineares Generalizados
 - GLM-Poisson e GLM-Logístico

© J.P. Sousa Slide 2 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tópicos do curso – Semana 2

4. Análise Multivariável

- princípios básicos de análise multivariável
- avaliação da estrutura subjacente aos dados: PCA, CA, NMDS, PCoA
- avaliação das diferenças entre grupos de dados/tratamentos: ANOSIM, CDA
- avaliação da relação entre variáveis de resposta e variáveis explicativas (ambientais): RDA, CCA, dbRDA
- avaliação de efeitos de tratamentos ao longo do tempo: Curvas de Resposta Principal (PRC)
- avaliação da relevância de diferentes grupos de variáveis explicativas: partição de variabilidade

© J.P. Sousa

Slide 3 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tópicos do curso – Semana 1

1. Revisão dos conceitos básicos em bioestatística

- testes de hipóteses para uma ou duas populações
- análise de variância e desenho experimental
- regressão linear simples e correlação

2. Regressão linear múltipla

- exploração dos dados
- avaliação de colinearidade entre variáveis explicativas
- interação entre variáveis explicativas
- interpretação dos resultados

3. Modelos Lineares Generalizados

- GLM-Poisson
- GLM-Logístico

© J.P. Sousa

Slide 4 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Para que (nos) serve a Bioestatística?

© J.P. Sousa Slide 5 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016



*“Se o único instrumento
que conheces é um
martelo, tenderás a ver
todos os problemas
como pregos”*

Slide 6 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

A Bioestatística...

1. Fornece os instrumentos necessários para tratar dados biológicos
2. Ao testar hipóteses, ajuda a tomar decisões e a resolver problemas
3. Ajuda a planear experiências com maior rigor


© J.P. Sousa Slide 7 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016


A Bioestatística...

1. **Recolha de dados**
➤ e.g. amostragem
2. **Apresentação de dados**
➤ e.g. gráficos e tabelas
3. **Caracterização de dados**
➤ e.g. médias

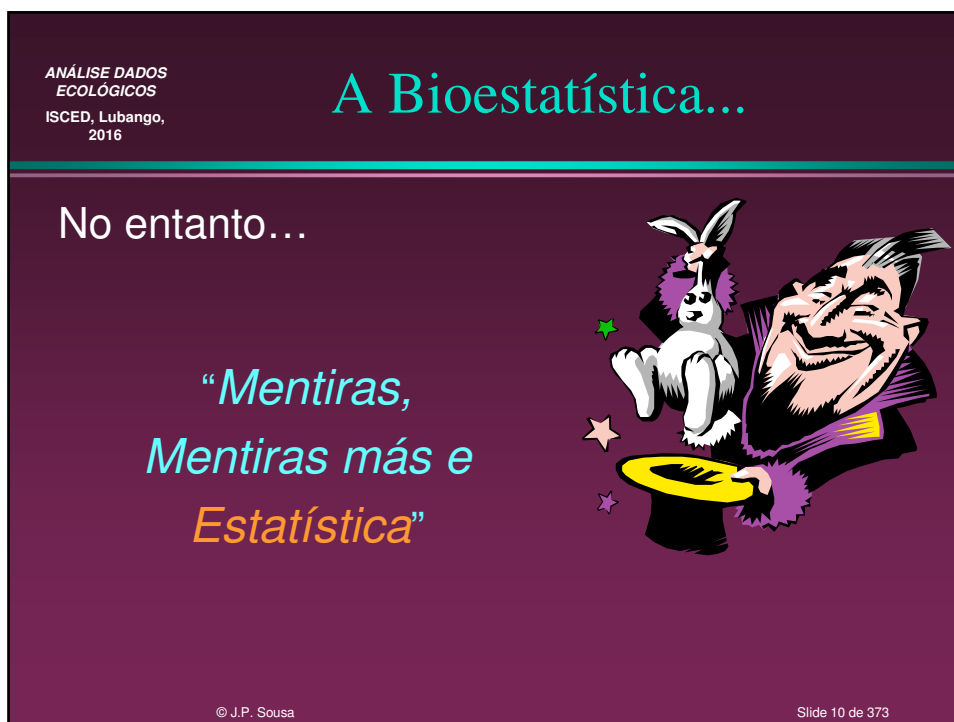
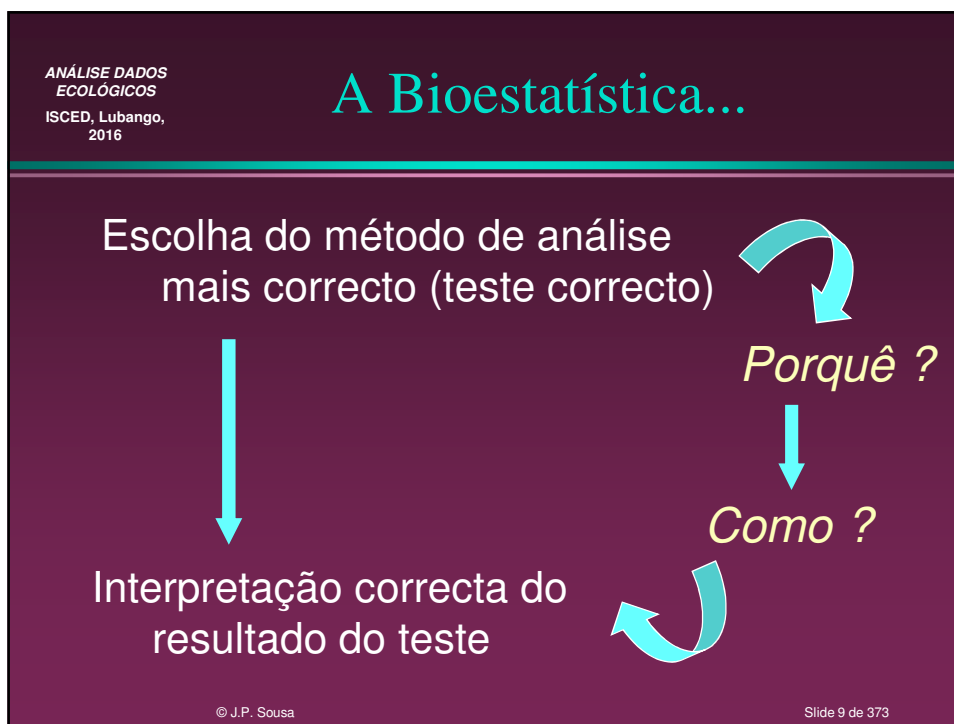
Análise de dados



Tomada de decisões



© J.P. Sousa Slide 8 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Bioestatística

Conceitos chave em bioestatística

© J.P. Sousa

Slide 11 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Métodos estatísticos

```
graph TD; A[Métodos estatísticos] --> B[Estatística Descritiva]; A --> C[Estatística inferencial];
```

Estatística Descritiva

Estatística inferencial

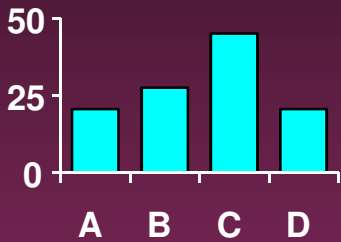
© J.P. Sousa

Slide 12 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Estatística descritiva

1. **Envolve**
 - Recolha de dados
 - Apresentação de dados
 - Caracterização de dados
2. **Objectivo**
 - Descrever os dados



Categoria	Valor
A	20
B	25
C	45
D	20

$\bar{x} = 30.5$ $S^2 = 11.3$

© J.P. Sousa Slide 13 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Estatística inferencial

1. **Envolve**
 - Estimativas
 - Testar hipóteses
2. **Objectivo**
 - Tomar decisões acerca das características da população



População ?

© J.P. Sousa Slide 14 de 373

Conceitos chave

1. População (Universo)

- Todos os elementos de interesse

2. Amostra

- Porção representativa da população

3. Elemento = Indivíduo

- Entidade em que se recolhem os dados

4. Variável

- Característica em que os indivíduos diferem (basicamente é o que se mede)

Conceitos chave (cont.)

5. Observação

- Conjunto medidas efectuadas num elemento

6. Conjunto de dados

- Medidas efectuadas num conjunto de elementos

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Conjunto de dados

variáveis

Indivíduo	Sexo	Peso	Idade
A	M	65	19
B	F	55	17
C	F	54	18
D	F	58	18
E	M	72	20
F	M	68	21

amostra

observação

© J.P. Sousa

Slide 17 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Tipos de escala (tipos de variáveis)

Escala de medida	Dados não numéricos	Dados numéricos
Fraccionária	-	<p>Intervalo constante entre unidades adjacentes (e.g., a diferença entre 14mg e 15mg é igual à diferença entre 34mg e 35mg)</p> <p>O zero encontra-se bem definido, significando a ausência da variável (e.g. 30mg é metade de 60mg)</p> <p>Exemplos: (pesos, volumes, taxas)</p>

© J.P. Sousa

Slide 18 de 373

Tipos de escala (tipos de variáveis)		
Escala de medida	Dados não numéricos	Dados numéricos
Intervalada	-	Intervalo constante entre unidades adjacentes (e.g., a diferença entre 15°C e 20°C é igual à diferença entre 35°C e 40°C) O zero não é bem definido (e.g. 20°C não é o dobro de 10°C e 0°C não significa não haver temperatura)

© J.P. Sousa

Slide 19 de 373

Tipos de escala (tipos de variáveis)		
Escala de medida	Dados não numéricos	Dados numéricos
Ordinal	A descrição permite a ordenação dos dados (e.g., o tamanho pode ser ordenado, muito embora não se conheça a diferenças reais)	A descrição permite a ordenação dos dados

© J.P. Sousa

Slide 20 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tipos de escala (tipos de variáveis)

Escala de medida	Dados não numéricos	Dados numéricos
Nominal	A descrição indica a categoria do elemento (e.g., sexo, côr)	A descrição indica a categoria do elemento (e.g., sexo, côr)

© J.P. Sousa Slide 21 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tipos de dados

A. Dados Qualitativos

- escala de medida nominal e ordinal

B. Dados Quantitativos

- escala de medida intervalada e fraccionária
 - dados contínuos (número infinito de valores dentro de um intervalo – ex: altura)
 - dados discretos ou descontínuos ou merísticos (só pode tomar determinados valores – ex: nº de indivíduos)

© J.P. Sousa Slide 22 de 373

Conceitos chave (cont.)

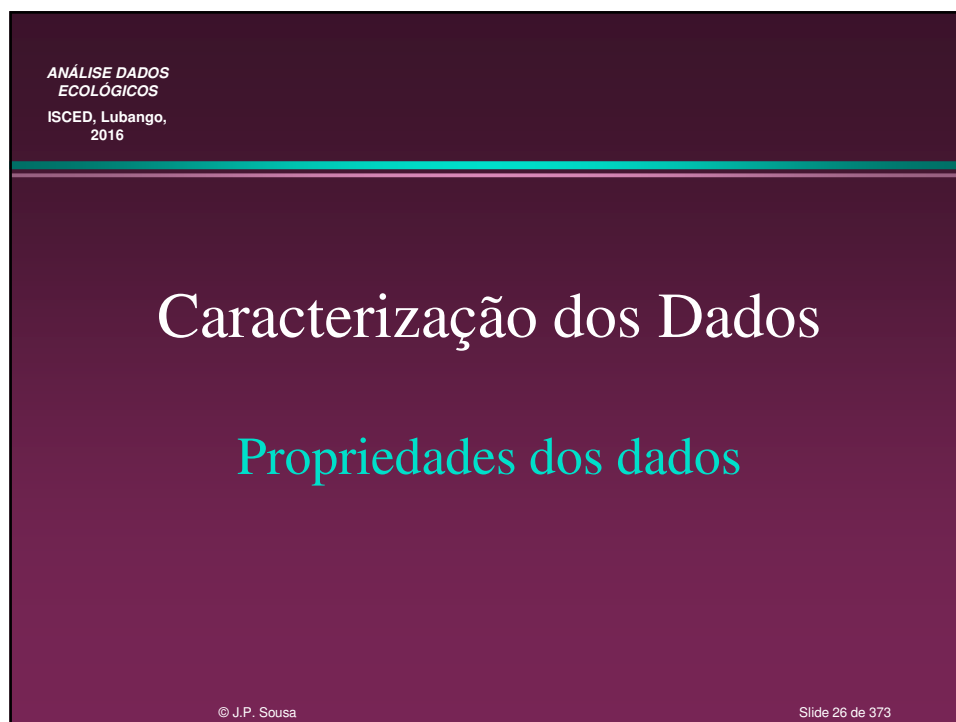
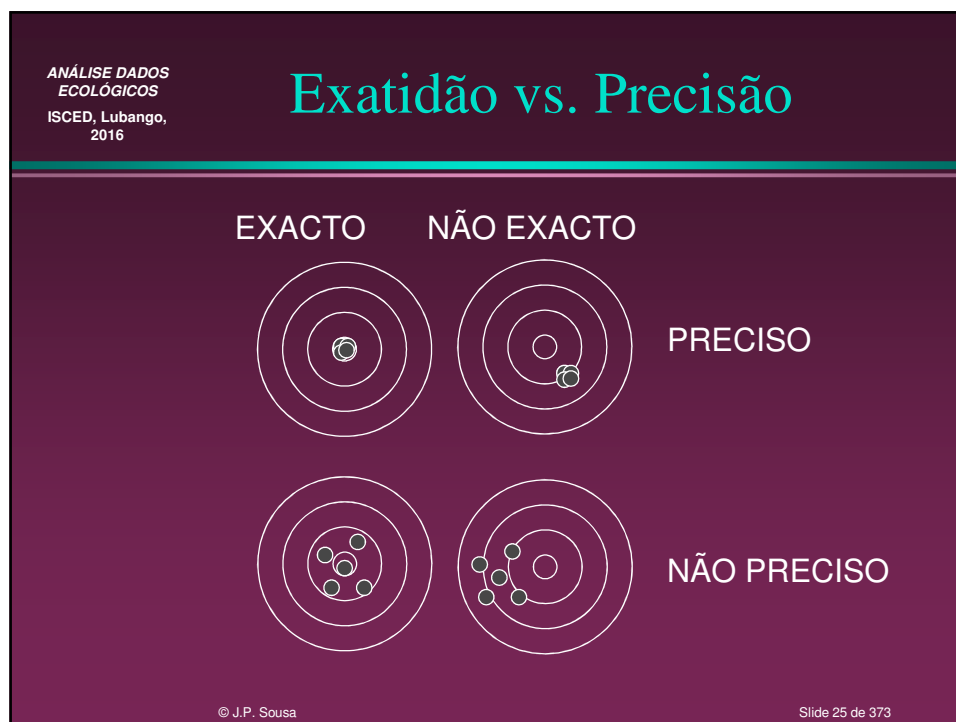
7. Exatidão (accuracy)

- Grau de aproximação da medição à realidade
- Normalmente é definida pelo observador e pelos instrumentos utilizados (se uma balança mede com uma exatidão de 0,1g, o resultado não é 5,00g mas sim 5,0g)


Conceitos chave (cont.)

8. Precisão (precision)

- Grau de aproximação entre medições repetidas (uma balança pode dar medidas muito precisas do mesmo objecto, mas muito pouco exactas, pois pode estar não calibrada)



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016



UM ESTATÍSTICO É AQUELE QUE,
TENDO A CABEÇA A ARDER E OS
PÉS ENTERRADOS NO GELO,
AINDA DIZ QUE NA MÉDIA
ESTÁ TUDO BEM!...

© J.P. Sousa

Slide 27 de 373

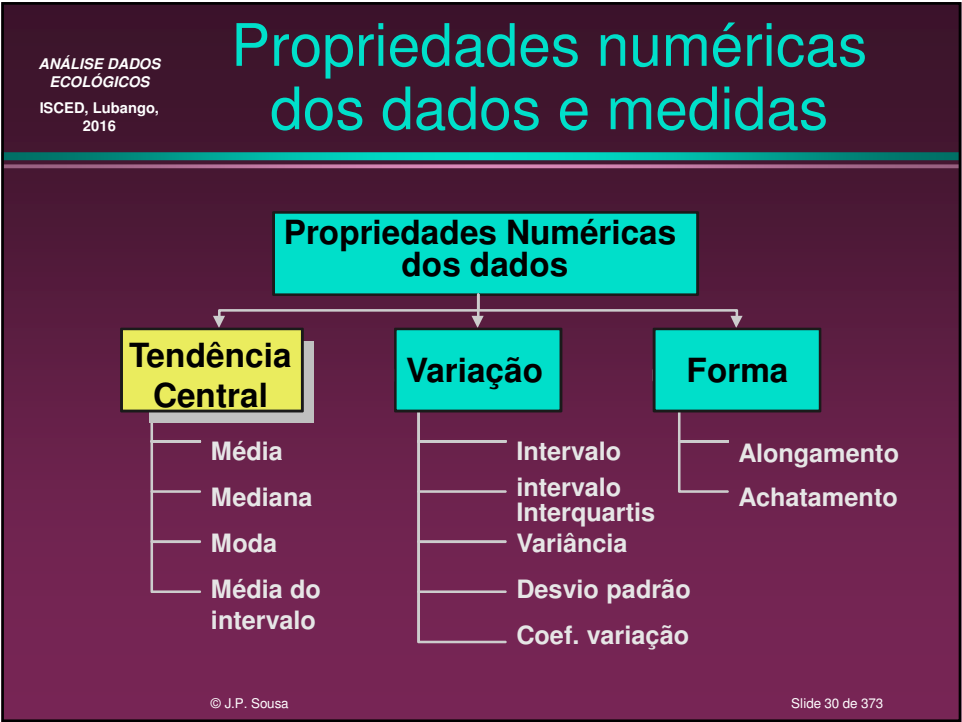
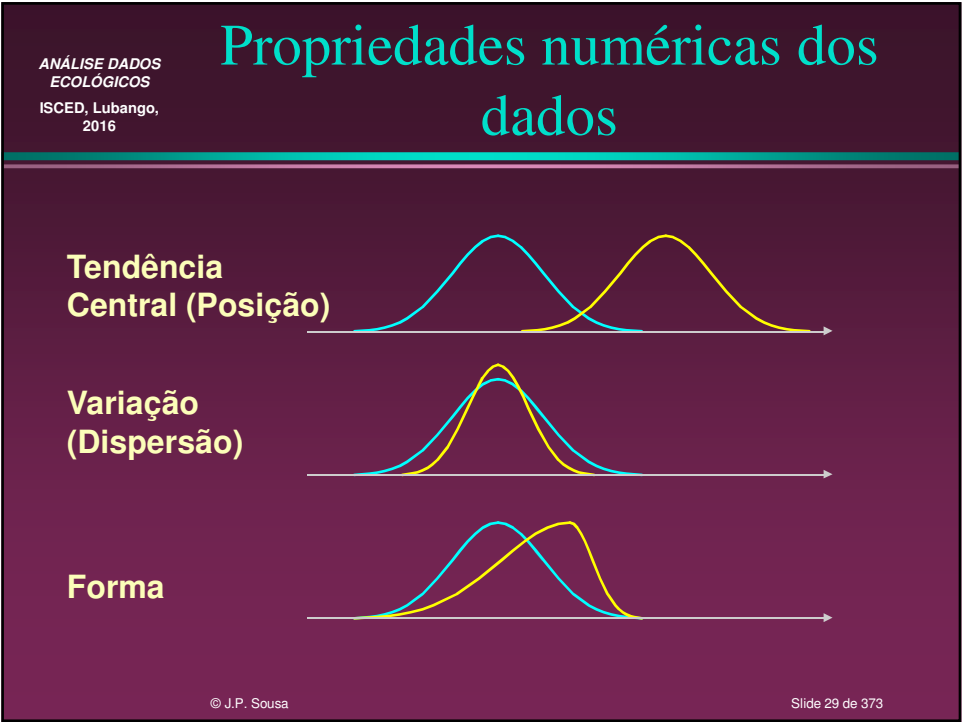
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Notação padrão

Medida	Amostra	População
Média	\bar{X}	μ_X
Desvio Padrão	S	σ_X
Variância	S^2	σ_X^2
Tamanho	n	N

© J.P. Sousa

Slide 28 de 373



Média

1. Medida de tendência central
2. Medida mais comum
3. Actua como “ponto de equilíbrio”
4. Afectada por valores extremos
5. Formula (Média aritmética da amostra)

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

© J.P. Sousa

Slide 31 de 373

Cálculo da média

Exemplo

Dados: 34 36 37 39 40 41 42 43 79

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9}{9} \\ &= \frac{34 + 36 + 37 + 39 + 40 + 41 + 42 + 43 + 79}{9} \\ &= 43,4\end{aligned}$$

© J.P. Sousa

Slide 32 de 373

Mediana

1. Medida de tendência Central
2. Valor médio de uma sequência ordenada
 - Se n for ímpar, valor médio da sequência
 - Se n for par, média dos 2 valores médios
3. Posição mediana de uma sequência de dados ordenada
4. Não é afectada por valores extremos

© J.P. Sousa

Slide 33 de 373

Cálculo da Mediana Amostra de tamanho ímpar

Dados: 24,1 22,6 21,5 23,7 22,6
 Ordenados: 21,5 22,6 **22,6** 23,7 24,1
 Posição: 1 2 **3** 4 5



$$\text{Ponto mediano} = \frac{n+1}{2} = \frac{5+1}{2} = 3,0$$

Mediana = 22,6

© J.P. Sousa

Slide 34 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Cálculo da Mediana

Amostra de tamanho par

Dados: 10,3 4,9 8,9 11,7 6,3 7,7
Ordenados: 4,9 6,3 **7,7** **8,9** 10,3 11,7
Posição: 1 2 **3** **4** 5 6

Ponto mediano $= \frac{n+1}{2} = \frac{6+1}{2} = 3,5$

Mediana $= \frac{7,7 + 8,9}{2} = 8,30$

© J.P. Sousa Slide 35 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Moda

1. Medida de tendência central
2. Valor que ocorre com mais frequência
3. Não é afectada por valores extremos
4. Pode não existir ou podem existir diversas modas

© J.P. Sousa Slide 36 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Exemplo de Moda

Nenhuma moda
Dados: 10,3 4,9 8,9 11,7 6,3 7,7

Uma moda
Dados: 6,3 4,9 8,9 6,3 4,9 4,9

Mais de uma moda
Dados: 21 28 28 41 43 43

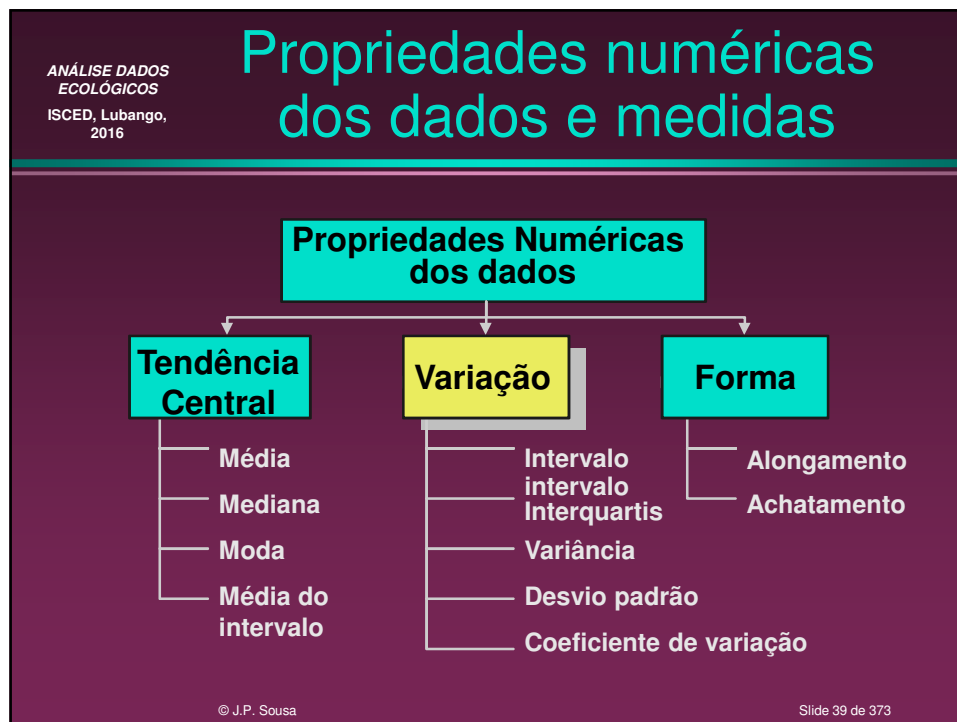
© J.P. Sousa Slide 37 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Resumo das Medidas de Tendência Central

Medida	Equação	Descrição
Média	$\Sigma X_i / n$	Ponto de equilíbrio
Mediana	Posição $\frac{(n+1)}{2}$	Valor médio quando ordenados
Moda	nenhuma	Mais frequente
Média do intervalo	$\frac{X_{menor} + X_{maior}}{2}$	Média do menor e do maior valor

© J.P. Sousa Slide 38 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Intervalo

- Medida de dispersão
- Diferença entre a menor observação e a maior observação

$$\text{Intervalo} = X_{\text{maior}} - X_{\text{menor}}$$

- Ignora a distribuição do dados

© J.P. Sousa Slide 40 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Quartis

- 1. Medida de tendência *Não-central*
- 2. Os dados ordenados são divididos em quatro partes

25	25	25	25
% Q_1 %	% Q_2 %	% Q_3 %	%

- 3. Posição do quartil de ordem i

$$\text{Posição do quartil } Q_i = \frac{i * (n+1)}{4}$$

© J.P. Sousa Slide 41 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Quartil (Q_1) Exemplo

- Dados: 10,3 4,9 8,9 11,7 6,3 7,7
- Ordenados: 4,9 **6,3** 7,7 8,9 10,3 11,7
- Posição: 1 **2** 3 4 5 6

♦

$$\text{Posição } Q_1 = \frac{1 * (n+1)}{4} = \frac{1 * (6+1)}{4} = 1,75 \cong 2$$

$$Q_1 = 6,3$$

© J.P. Sousa Slide 42 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Quartil (Q_2)

Exemplo

- Dados: 10,3 4,9 8,9 11,7 6,3 7,7
- Ordenados: 4,9 6,3 **7,7** **8,9** 10,3 11,7
- Posição: 1 2 **3** **4** 5 6

♦

$$\text{Posição } Q_2 = \frac{2 * (n+1)}{4} = \frac{2 * (6+1)}{4} = 3,5$$

$$Q_2 = \frac{7,7 + 8,9}{2} = 8,3$$

© J.P. Sousa Slide 43 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Quartil (Q_3)

Exemplo

- Dados: 10,3 4,9 8,9 11,7 6,3 7,7
- Ordenados: 4,9 6,3 7,7 8,9 **10,3** 11,7
- Posição: 1 2 3 4 **5** 6

♦

$$\text{Posição } Q_3 = \frac{3 * (n+1)}{4} = \frac{3 * (6+1)}{4} = 5,25 \approx 5$$

$$Q_3 = 10,3$$

© J.P. Sousa Slide 44 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Variância & Desvio padrão

1. Medidas de dispersão
2. Medidas mais comuns
3. Consideram a distribuição dos valores
4. Variação à volta da média (\bar{x} or μ_X)

© J.P. Sousa Slide 45 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Variância da amostra

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

$n - 1$ em denominador!
(Utilizar **N** se se tratar da variância da **População**)

$$= \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n - 1}$$

© J.P. Sousa Slide 46 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desvio padrão da amostra

$$S = \sqrt{S^2}$$

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

$$= \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1}}$$

© J.P. Sousa

Slide 47 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Variância Exemplo

Dados: 10,3 4,9 8,9 11,7 6,3 7,7

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad \text{onde} \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n} = 8,3$$

$$S^2 = \frac{(10,3 - 8,3)^2 + (4,9 - 8,3)^2 + \dots + (7,7 - 8,3)^2}{6-1}$$

$$= 6,368$$

© J.P. Sousa

Slide 48 de 373

Coeficiente de Variação

1. Medida de dispersão **Relativa**
2. Expressa sempre em %
3. Mostra a variância relativamente à média
4. Utilizada para comparar 2 ou mais grupos
5. Formula (amostra)

$$CV = \frac{S}{\bar{X}} * 100\%$$

© J.P. Sousa

Slide 49 de 373

Coeficiente de variação Exemplo

Grupo de dados 1 : 1 2 3
 Grupo de dados 2: 100 200 300

Grupo 1 $CV = \frac{S}{\bar{X}} * 100\% = \frac{1}{2} * 100\% = 50\%$

Grupo 2 $CV = \frac{S}{\bar{X}} * 100\% = \frac{100}{200} * 100\% = 50\%$

© J.P. Sousa

Slide 50 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Resumo das Medidas de Dispersão

Medida	Equação	Descrição
Intervalo	$X_{\text{maior}} - X_{\text{menor}}$	Dispersão total
Intervalo interquartis	$Q_3 - Q_1$	Dispersão dos 50% centrais
Desvio padrão (Amostra)	$\sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$	Dispersão à volta da média da amostra
Desvio Padrão (População)	$\sqrt{\frac{\sum (X_i - \mu_x)^2}{N}}$	Dispersão à volta da média da população
Variância (Amostra)	$\frac{\sum (X_i - \bar{X})^2}{n - 1}$	Quadrado da dispersão à volta da média
Coef. de variação	$(S / \bar{X}) 100\%$	Dispersão relativa

© J.P. Sousa Slide 51 de 373



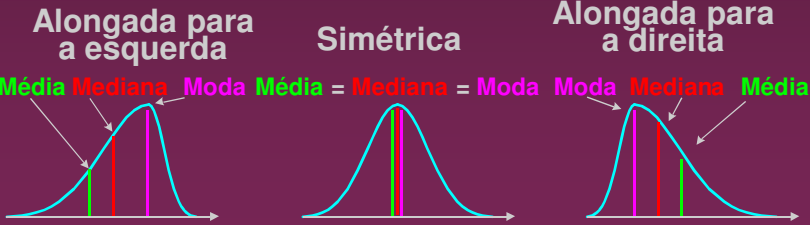
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Forma

1. Descreve a distribuição dos dados
2. Medida de forma
 - Achatamento = Grau de achatamento
 - Alongamento = Simetria

Alongada para a esquerda **Simétrica** **Alongada para a direita**

Média Mediana Moda Média = Mediana = Moda Moda Mediana Média

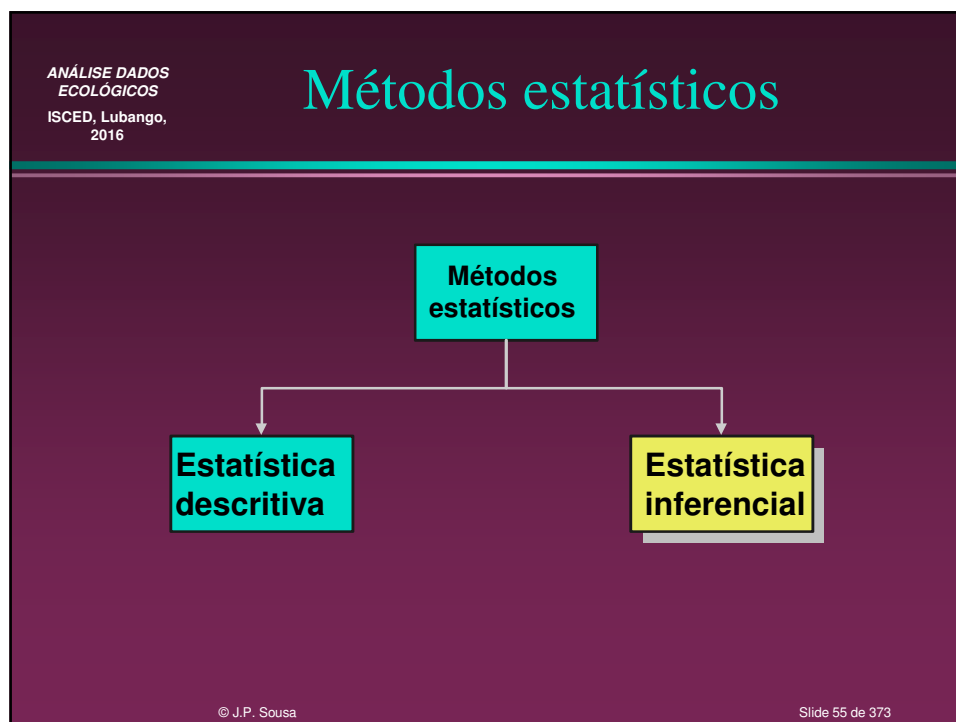


© J.P. Sousa Slide 53 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Amostragem e distribuições de amostragem

© J.P. Sousa Slide 54 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

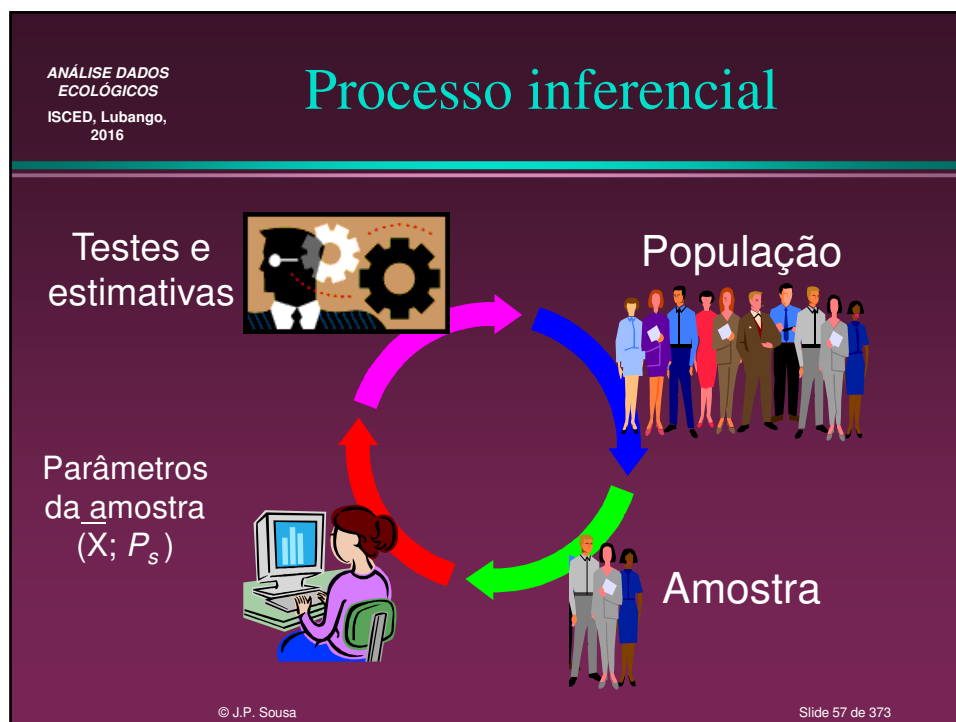
Estatística inferencial

1. **Envolve**
 - Estimativas
 - Teste de hipóteses
2. **Objectivos**
 - Tomar decisões acerca das características das populações

População?

© J.P. Sousa

Slide 56 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Estimativas

1. Variáveis ao acaso utilizadas para estimar parâmetros da população
Média da amostra; Mediana da amostra; Proporção na amostra
2. A média da amostra \bar{X} é uma estimativa da média da população μ_X
 - Se $\bar{X} = 3$ então 3 é a estimativa de μ_X
3. A base teórica é a distribuição da amostragem

© J.P. Sousa Slide 58 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Distribuição de amostragem

1. Distribuição de probabilidades teórica
2. A variável ao acaso é um parâmetro estatístico da população
 - Média da amostra; Mediana da amostra; etc.
3. Resulta da obtenção de todas as amostras possíveis de um determinado tamanho
4. Contém todos os pares $[\bar{X}; P(\bar{X})]$ de valores
 - Distribuição das médias das amostras

© J.P. Sousa

Slide 59 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Distribuições de amostragem

Suponhamos uma população ...

Tamanho da população; $N = 4$

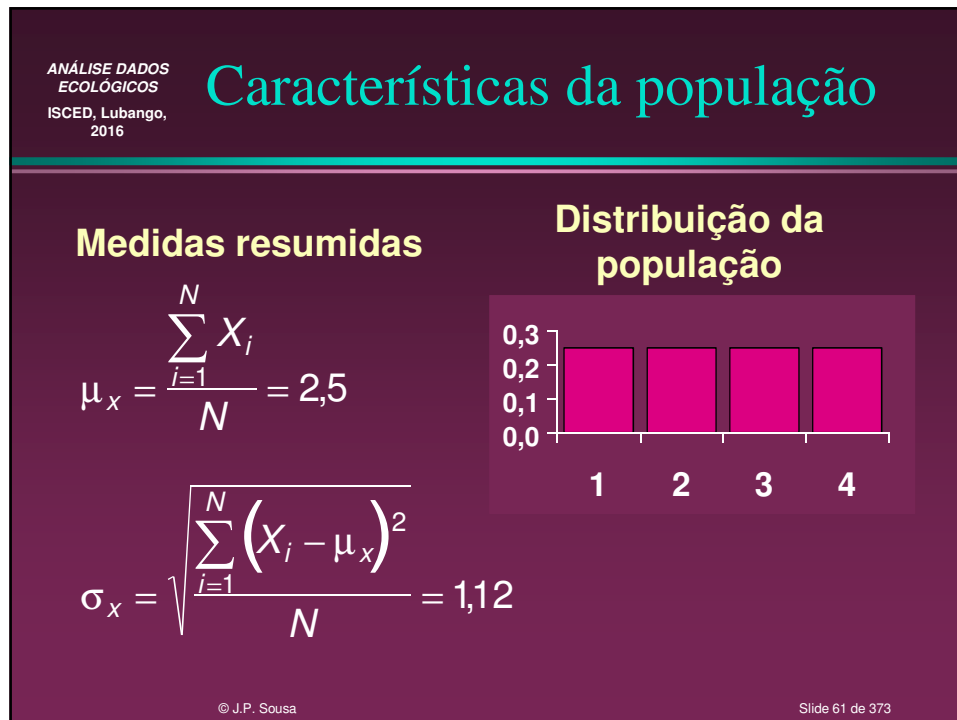
Variável ao acaso; X (e.g. peso);
Valores de peso (g): 1; 2; 3; 4

Distribuição uniforme



© J.P. Sousa

Slide 60 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Amostras possíveis com tamanho $n = 2$

16 Amostras

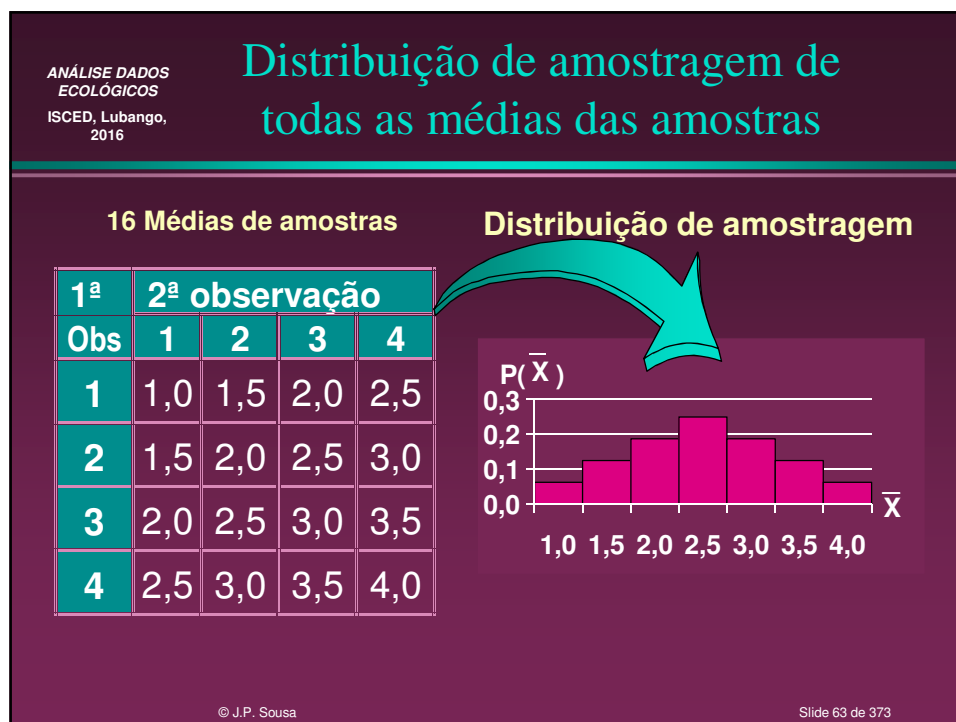
1ª	2ª observação			
Obs	1	2	3	4
1	1;1	1;2	1;3	1;4
2	2;1	2;2	2;3	2;4
3	3;1	3;2	3;3	3;4
4	4;1	4;2	4;3	4;4

16 Médias de amostras

1ª	2ª observação			
Obs	1	2	3	4
1	1,0	1,5	2,0	2,5
2	1,5	2,0	2,5	3,0
3	2,0	2,5	3,0	3,5
4	2,5	3,0	3,5	4,0

Variabilidade na amostragem !

© J.P. Sousa Slide 62 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

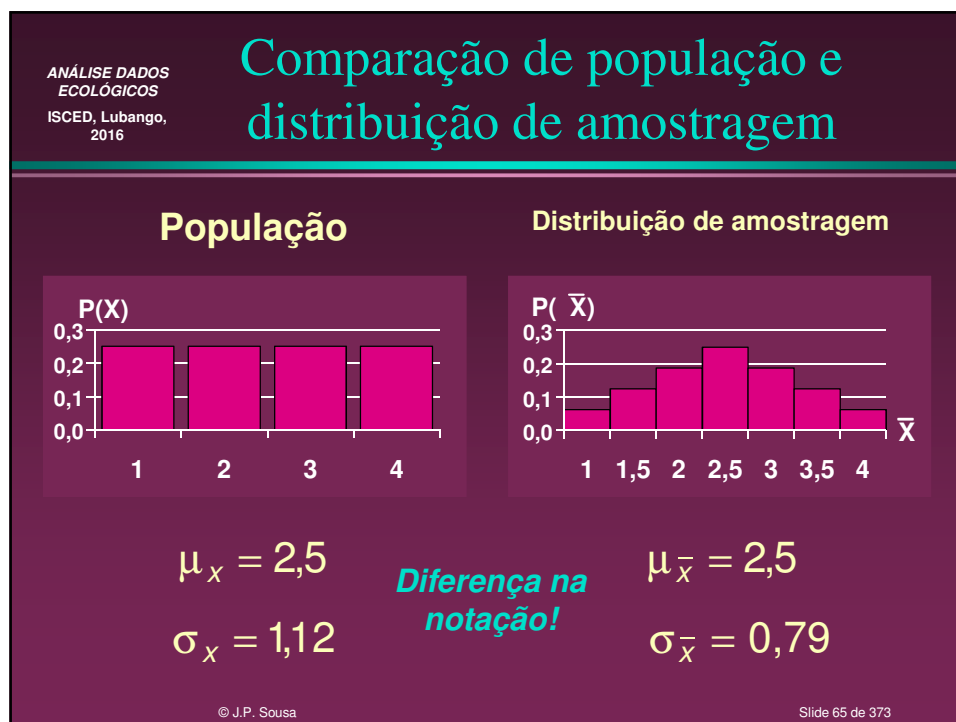
Resumo da medida de todas as médias possíveis

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^N \bar{X}_i}{N} = \frac{1,0 + 1,5 + \dots + 4,0}{16} = 2,5$$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^N (\bar{X}_i - \mu_{\bar{X}})^2}{N}}$$

$$= \sqrt{\frac{(1,0 - 2,5)^2 + (1,5 - 2,5)^2 + \dots + (4,0 - 2,5)^2}{16}} = 0,79$$

© J.P. Sousa Slide 64 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Erro padrão da média

- Desvio padrão de todas as médias possíveis; \bar{X}
 - Medida de dispersão de todas as médias das amostras;
- Inferior ao desvio padrão da população
- Fórmula (amostragem com substituição)

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^N (\bar{X}_i - \mu_{\bar{X}})^2}{N}} = \frac{\sigma_X}{\sqrt{n}}$$

© J.P. Sousa Slide 66 de 373

Propriedades da média

1. Equilibrada

- A média da distribuição de amostragem é igual à média da população

2. Eficiente

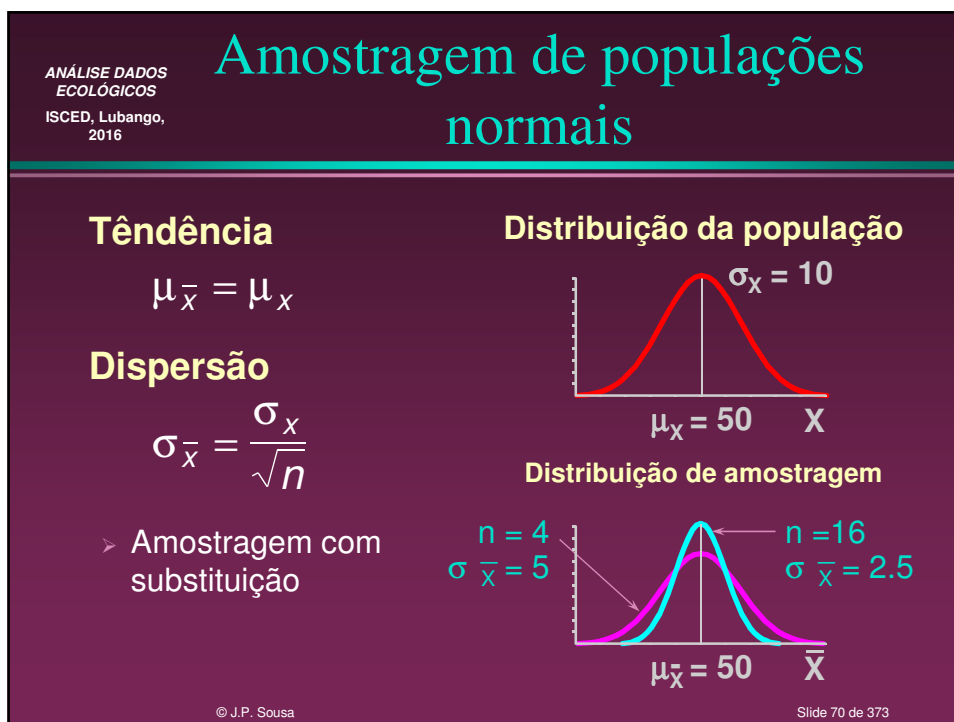
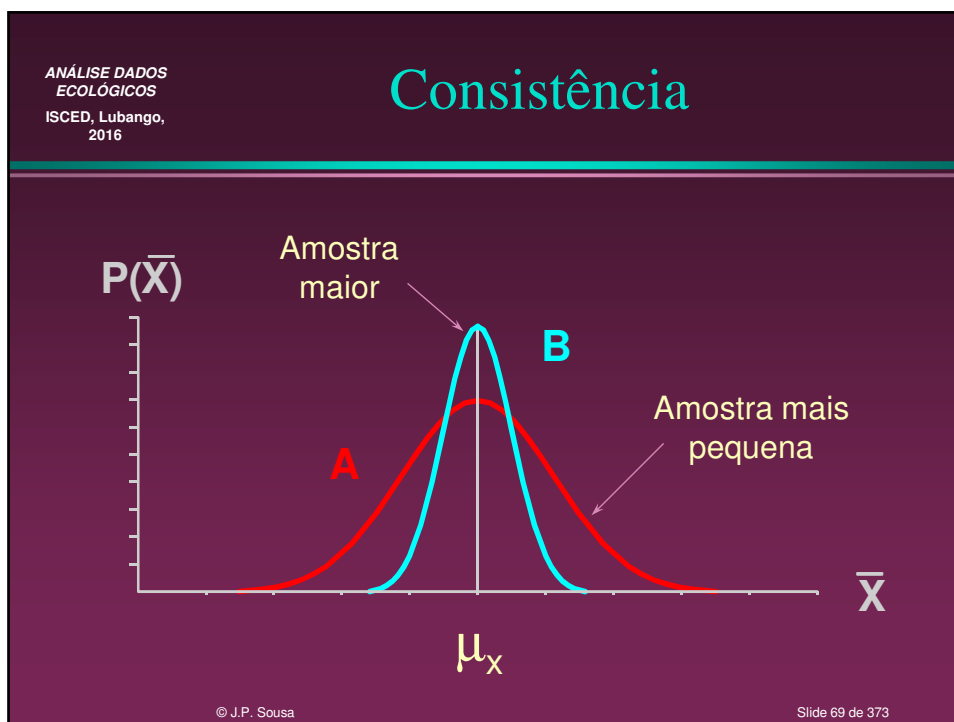
- A média da amostra aproxima-se mais da média da população que qualquer outro parâmetro

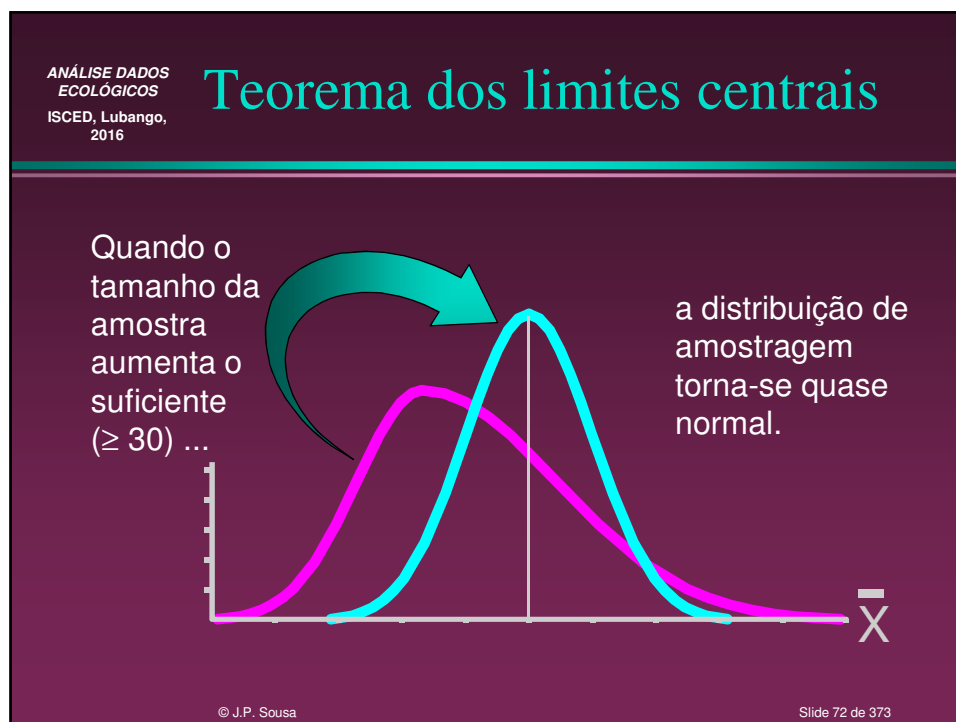
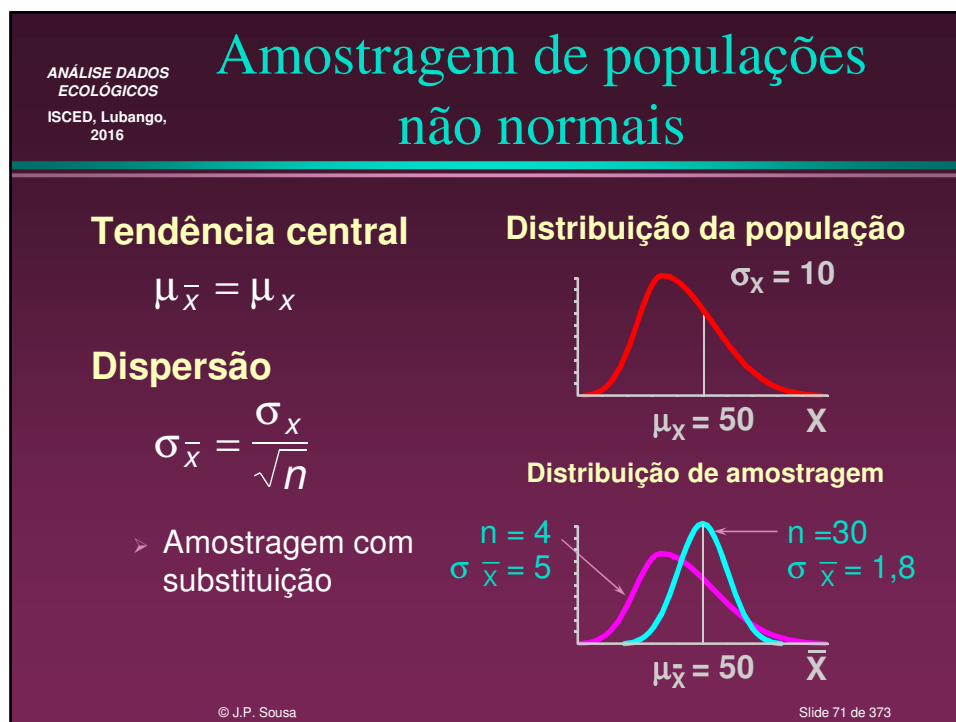
3. Consistente

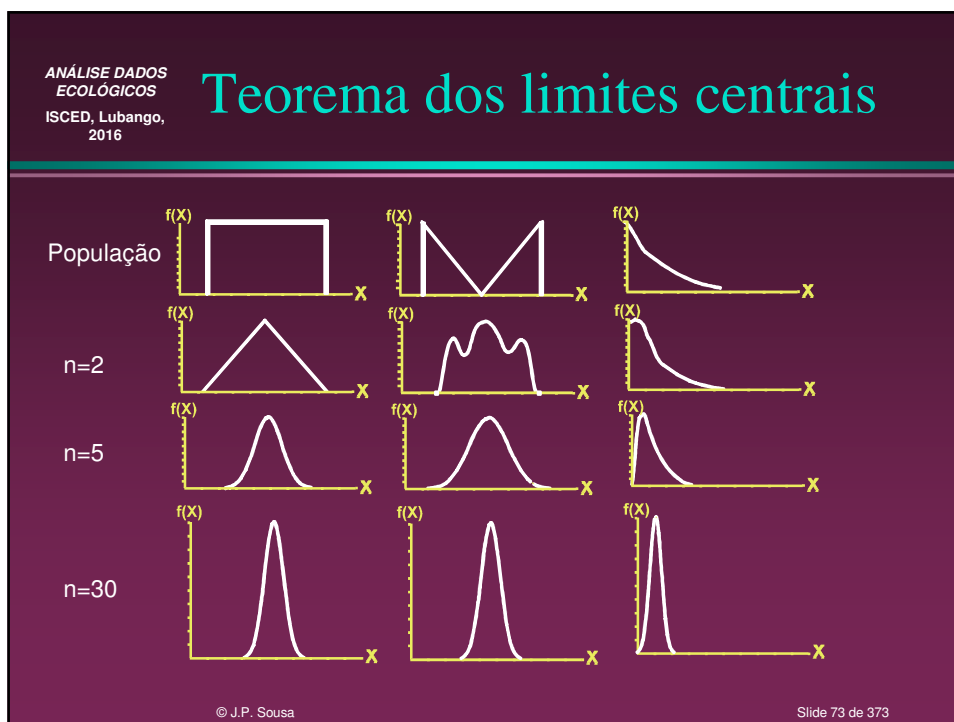
- À medida que o tamanho da amostra aumenta a variação da média da amostra relativamente à média da população reduz-se

Equilíbrio









ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teorema dos limites centrais

As médias de um elevado número de amostras retiradas ao acaso da população tendem a estar normalmente distribuídas e a “média das médias” é a média da população

© J.P. Sousa

Slide 74 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

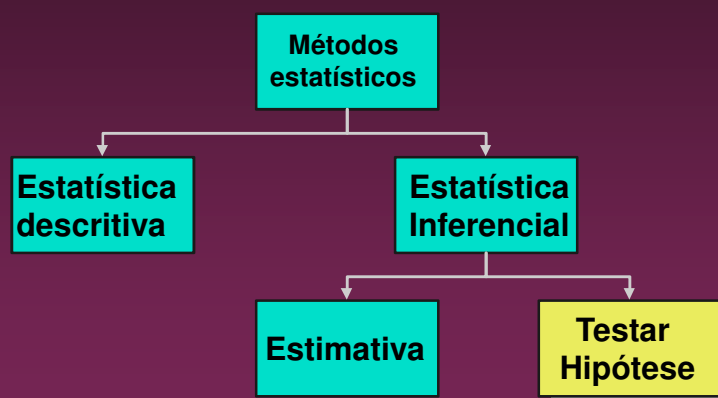
Conceitos gerais sobre os métodos de testar hipóteses

© J.P. Sousa

Slide 75 de 373

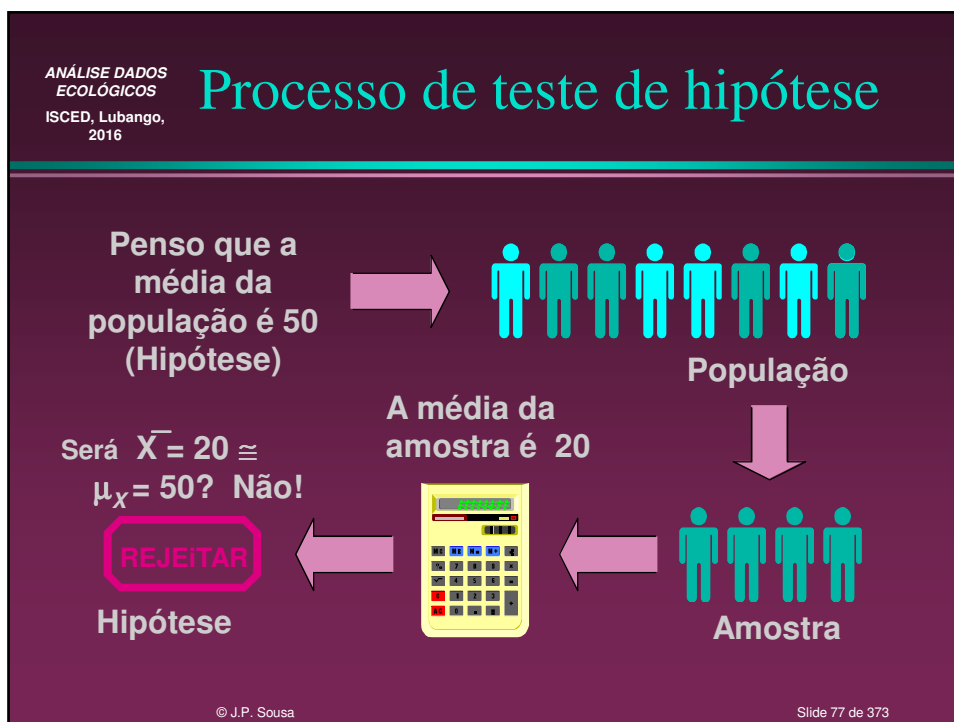
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Métodos estatísticos



© J.P. Sousa

Slide 76 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

O que é uma hipótese?

- Suposição acerca de um parâmetro da população
 - Média da **População**, Proporção, Variância
 - Deve ser definida **Antes** da análise

Penso que a idade média dos Profs do ISCED é 50 anos

© J.P. Sousa Slide 78 de 373

Hipótese nula

1. O que está em jogo
2. Pode ter consequências se a avaliação for incorrecta
3. Inclui sempre o sinal de igualdade: $=$, \leq , ou \geq
4. Designa-se por H_0
5. Enuncia-se $H_0: \mu_X \geq 3$

Hipótese alternativa

1. Oposta da hipótese nula
2. Tem sempre o sinal de desigualdade: \neq , $<$, ou $>$
3. Designa-se por H_A ou H_1
4. Enuncia-se $H_A: \mu_X < 3$

Identificação das hipóteses

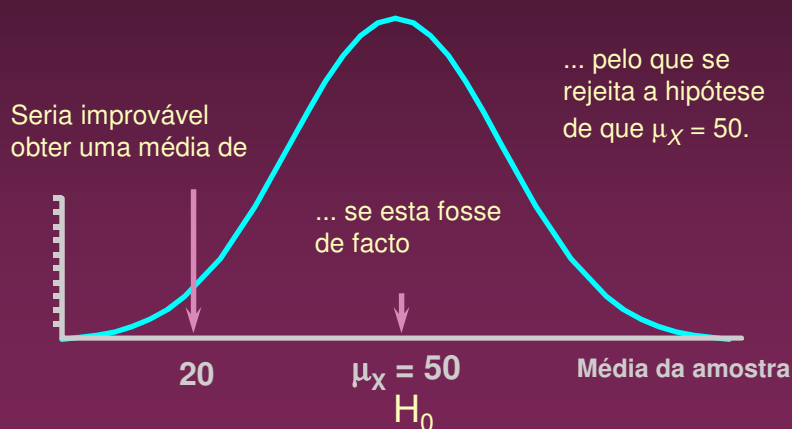
1. Problema: Testar se média da população é 50

2. Passos

- Enunciar a questão estatisticamente ($\mu_X = 50$)
- Enunciar a questão oposta estatisticamente ($\mu_X \neq 50$)
 - Devem ser mutuamente exclusivas e abrangentes
- Seleccionar a hipótese nula ($\mu_X = 50$)

Ideia básica

Distribuição de amostragem

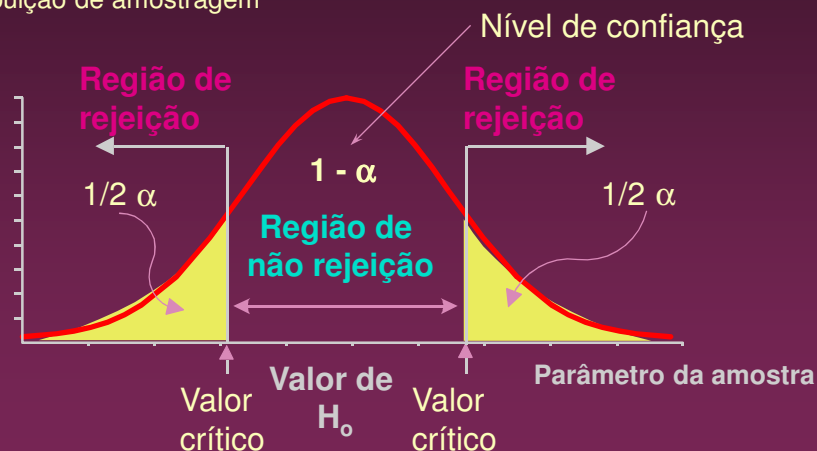


Nível de significância

1. Probabilidade
2. Define valores improváveis para os parâmetros estatísticos da amostra se a hipótese nula for verdadeira
 - Também denominada por região de rejeição da distribuição de amostragem
3. Representa-se por α
 - Valores habituais são 0,01, 0,05, 0,10
4. Nível seleccionado pelo investigador no início

Teste de duas caudas Regiões de rejeição

Distribuição de amostragem



Erros de decisão

1. Erro Tipo I

- Rejeitar uma hipótese nula verdadeira
- Tem consequências sérias
- A probabilidade de um erro tipo I é α
 - Denomina-se nível de significância

2. Erro Tipo II

- Não rejeitar uma hipótese nula falsa
- A probabilidade de erro tipo II é β

Resultados de decisão

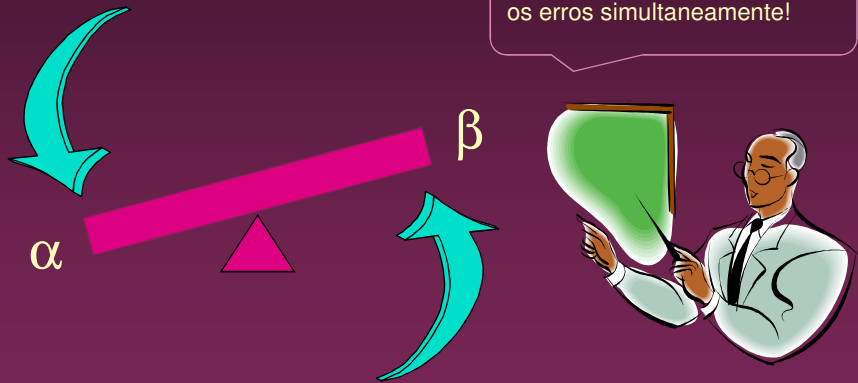
Julgamento			Testar H_0		
Situação Real			Situação correcta		
Veredito	Inocente	Culpado	Decisão	H_0 Verdadeira	H_0 Falsa
Inocente	Correcto	Erro	Não rejeitar H_0	$1 - \alpha$	Erro tipo II (β)
Culpado	Erro	Correcto	Rejeitar H_0	Erro tipo I (α)	Potência ($1 - \beta$)

 H_0 : Inocente H_A : Culpado

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

α & β têm uma relação inversa

Não é possível reduzir ambos os erros simultaneamente!



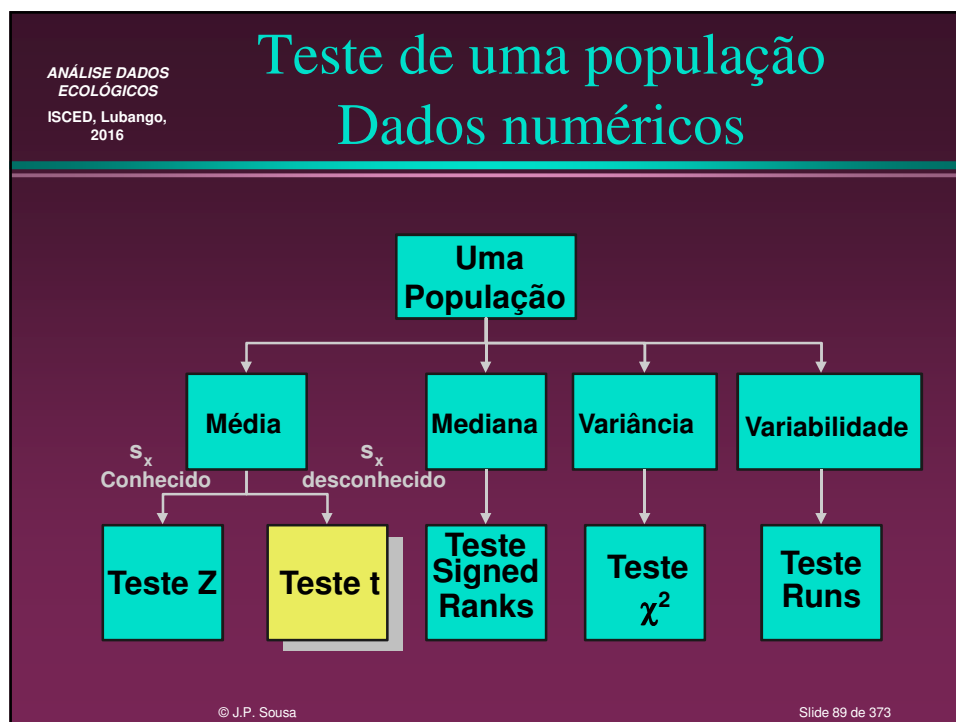
© J.P. Sousa Slide 87 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Factores que afectam β

1. Valor verdadeiro do parâmetro na população
 - Aumenta quando a diferença entre o valor teórico e o valor real na população diminui
2. Nível de significância α
 - Aumenta com o decréscimo de α
3. Desvio padrão da população σ_X
 - Aumenta quando σ_X aumenta
4. Tamanho da amostra, n
 - Aumenta quando n diminui

© J.P. Sousa Slide 88 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste t para média (σ_x desconhecido)

- Pressupostos
 - População encontra-se normalmente distribuída
 - Se não fôr normal apenas distribuições com um ligeiro alongamento ou amostras grandes ($n \geq 30$) podem ser utilizadas
- Teste paramétrico
- Teste t

$$t = \frac{\bar{X} - \mu_x}{\frac{S}{\sqrt{n}}}$$

© J.P. Sousa Slide 90 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Propriedades da distribuição t

1. Apresenta diferentes formas dependendo dos **graus de liberdade** (logo de n)
 - À medida que n aumenta vai-se aproximando da distribuição normal
2. A replicação terá que ser “verdadeira”
 - O n terá que corresponder a sujeitos verdadeiros e não várias medidas do mesmo sujeito

© J.P. Sousa

Slide 91 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t Exemplo

Será que os exemplars desta espécie pesam, em média, **368** gramas? Uma amostra ao acaso de **36** animais apresentou uma média de **372.5** e um desvio padrão de **12** gramas. Testar para um nível de significância de **0,05**.



© J.P. Sousa

Slide 92 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t Solução

H₀: $\mu_X = 368$

H_A: $\mu_X \neq 368$

$\alpha = 0,05$

df = $36 - 1 = 35$

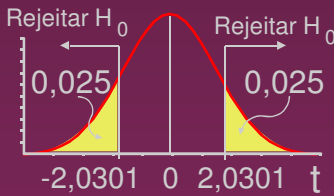
Valores críticos:

Teste estatístico:

$$t = \frac{\bar{X} - \mu_x}{\frac{S}{\sqrt{n}}} = \frac{372,5 - 368}{\frac{12}{\sqrt{36}}} = +2,25$$

Decisão:
Rejeitar para $\alpha = 0,05$

Conclusão:
A média da espécie (população) é estatisticamente diferente de 368




© J.P. Sousa
Slide 93 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t de duas caudas Problema

Mediu-se a temperatura corporal numa amostra de 25 caranguejos expostos a uma temperatura ambiente de 24,3 °C. A amostra dos caranguejos apresentava uma temperatura média de 25,03 °C e um desvio padrão de 1,34 °C. Pretende-se testar para $\alpha=0,05$ se a temperatura dos caranguejos difere da temperatura ambiente.



Slide 94 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t de duas caudas Solução*

H₀: $\mu_X = 24,3$
H_A: $\mu_X \neq 24,3$
 α = 0,05
gl = 25-1 = 24

Teste estatístico:

$$t = \frac{\bar{X} - \mu_X}{\frac{S}{\sqrt{n}}} = \frac{25,03 - 24,3}{\frac{1,34}{\sqrt{25}}} = 2,704$$

Valores críticos:



Decisão:
Rejeitar para $\alpha = 0,05$

Conclusão:
A média da população não é 24,3 °C

© J.P. Sousa Slide 95 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t de uma cauda

Pretende-se saber se a média do comprimento numa população de microlepidópetros é de **pelo menos 32 mm** para um nível de significância de **0,01**. Retira-se uma amostra de **60 indivíduos** da população e calcula-se uma média de **30,7 mm** com um desvio padrão de **3,8 mm**.



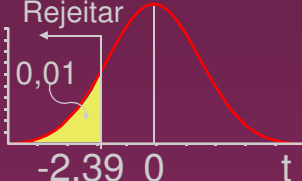
Slide 96 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste t de uma cauda Solução

H₀: $\mu_X \geq 32$
H_A: $\mu_X < 32$
 α = 0,01
gl = 60-1 = 59

Valores críticos:



Teste estatístico:

$$t = \frac{\bar{X} - \mu_X}{\frac{S}{\sqrt{n}}} = \frac{30,7 - 32}{\frac{3,8}{\sqrt{60}}} = -2,65$$

Decisão:
Rejeitar para $\alpha = 0,01$

Conclusão:
A média é estatisticamente menor de 32

© J.P. Sousa Slide 97 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Populações independentes e relacionadas

Independentes	Relacionadas
<p>1. Dados com diferentes origens</p> <ul style="list-style-type: none"> ➢ Não relacionados ➢ Independentes <p>2. Calcular diferenças entre as médias das amostras</p> <ul style="list-style-type: none"> ➢ $\bar{X}_1 - \bar{X}_2$ 	<p>1. A mesma fonte de dados</p> <ul style="list-style-type: none"> ➢ emparelhados ➢ Medições repetidas (Antes/Depois) <p>2. Utilizar diferenças entre cada par de observações</p> <ul style="list-style-type: none"> ➢ $D_n = X_{1n} - X_{2n}$

© J.P. Sousa Slide 98 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Exemplos de duas populações independentes

1. Um investigador pretende verificar se existem diferenças de tamanho entre indivíduos que habitam a alta e a baixa altitude.
2. Comparação da longevidade entre indivíduos provenientes de um local poluído e indivíduos de um local não poluído.

© J.P. Sousa

Slide 99 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Exemplo de populações relacionadas

1. Estudo da velocidade de regeneração dos fragmentos anteriores e posteriores de planárias .
2. Estudo da longevidade dos indivíduos de uma população antes e depois da exposição a um poluente.

© J.P. Sousa

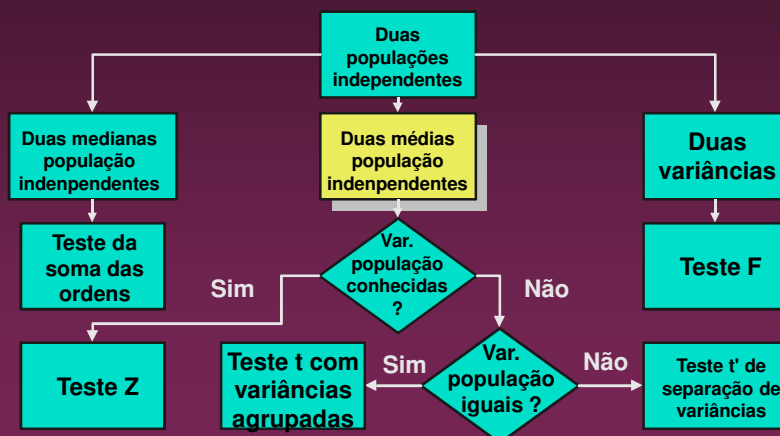
Slide 100 de 373

Problema

Estas variáveis são relacionadas ou independentes?

1. Altura de uma espécie de plantas em dois locais diferentes.
2. Altura de uma espécie de plantas de um mesmo local antes e depois de um evento de stress.

Testes para duas populações independentes



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Duas populações independentes

Hipóteses para as médias

	Questões		
Hipótese	Sem diferença	Pop 1 ≥ Pop 2	Pop 1 ≤ Pop 2
	Com diferença	Pop 1 < Pop 2	Pop 1 > Pop 2
H ₀	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 \leq 0$
H _A	$\mu_1 - \mu_2 \neq 0$	$\mu_1 - \mu_2 < 0$	$\mu_1 - \mu_2 > 0$

© J.P. Sousa

Slide 103 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Testes para duas populações independentes

```
graph TD; A[Duas populações independentes] --> B[Duas medianas população independentes]; A --> C[Duas médias população independentes]; A --> D[Duas variâncias]; B --> E[Teste da soma das ordens]; E --> F[Teste Z]; C --> G{Var. população conhecidas?}; G -- Sim --> H[Teste t com variâncias agrupadas]; G -- Não --> I{Var. população iguais?}; I -- Sim --> H; I -- Não --> J[Teste t' com separação de variâncias]; D --> K[Teste F];
```

The flowchart outlines the selection of statistical tests for two independent populations. It starts with three main branches: medianes, médias, and variâncias. The 'médias' branch includes decision points for whether population variances are known or equal, leading to different t-test variants. The 'variâncias' branch leads to an F-test.

© J.P. Sousa

Slide 104 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t para médias de duas amostras com variâncias agrupadas

1. Testa as médias de duas populações independentes com variâncias iguais
2. Teste paramétrico
3. Pressupostos
 - Ambas as populações se encontram normalmente distribuídas
 - Se não forem normais podem ser aproximadas através da normal ($n_1 \geq 30$ & $n_2 \geq 30$)
 - As variâncias das populações são **desconhecidas** mas assumidas com **iguais**

© J.P. Sousa

Slide 105 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t com variâncias agrupadas

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Diferença hipotética

$$S_P^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$gl = n_1 + n_2 - 2$$

© J.P. Sousa

Slide 106 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t com variâncias agrupadas Exemplo

Estudou-se a influência de dois tipos de alimentação (algas e leveduras) no tamanho de crustáceos à primeira reprodução:

	<u>Algas</u>	<u>Leveduras</u>
Número	21	25
Média	3,27	2,53
Des. Padrão	1,30	1,16

Assumindo variâncias **iguais** será que existe diferença no tamanho à primeira reprodução ($\alpha = 0,05$)?

© J.P. Sousa

Slide 107 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t com variâncias agrupadas Exemplo

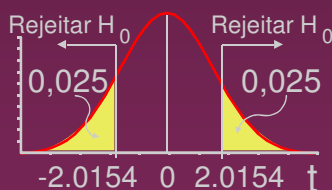
H₀: $\mu_1 - \mu_2 = 0$ ($\mu_1 = \mu_2$)

H_A: $\mu_1 - \mu_2 \neq 0$ ($\mu_1 \neq \mu_2$)

$\alpha = 0,05$

df = $21 + 25 - 2 = 44$

Valores críticos:



© J.P. Sousa

Teste estatístico:

$$t = \frac{3,27 - 2,53}{\sqrt{1,502 \cdot \left(\frac{1}{21} + \frac{1}{25}\right)}} = +2,04$$

Decisão:

Rejeitar para $\alpha = 0,05$

Conclusão:

Há diferença estatística entre as médias

Slide 108 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

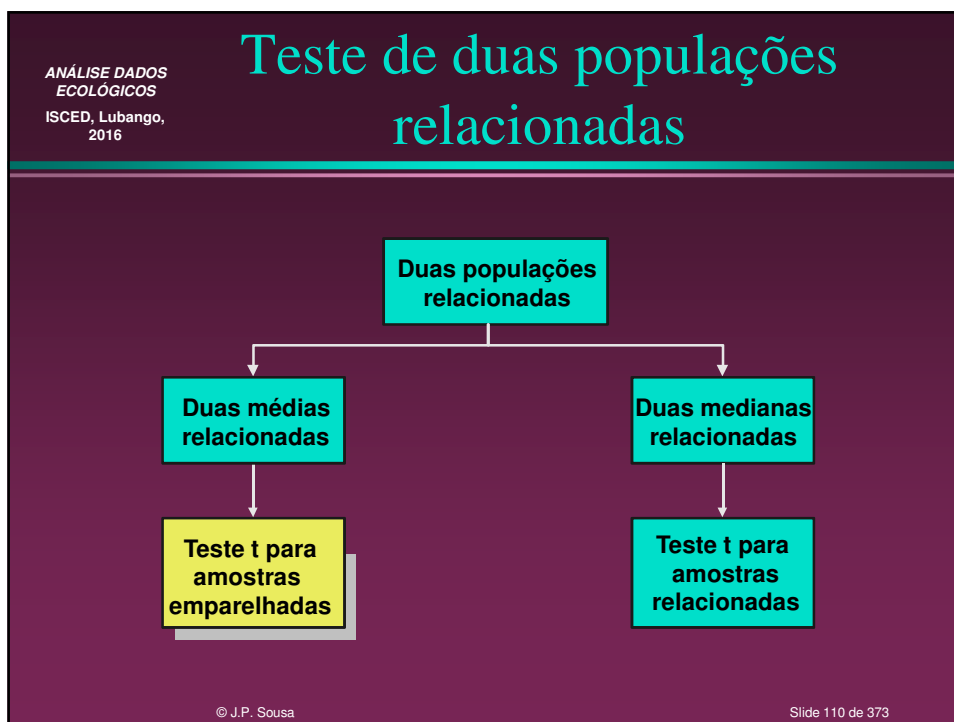
Teste estatístico Solução

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(3,27 - 2,53) - (0)}{\sqrt{1,502 \cdot \left(\frac{1}{21} + \frac{1}{25}\right)}} = +2,04$$

$$S_P^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$= \frac{(21 - 1) \cdot (1,30)^2 + (25 - 1) \cdot (1,16)^2}{(21 - 1) + (25 - 1)} = 1,502$$

© J.P. Sousa Slide 109 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t de amostras emparelhadas para diferenças de médias

1. Compara médias de duas populações relacionadas
 - Emparelhadas ou coincidentes
 - Medidas repetidas (antes/depois)
2. Elimina variações entre indivíduos
3. Pressupostos
 - Populações normalmente distribuídas
 - Se não forem normais podem ser aproximadas através da normal ($n_1 \geq 30$ & $n_2 \geq 30$)

© J.P. Sousa

Slide 111 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t para amostras emparelhadas Hipóteses

	Questões		
Hipótese	Sem diferenças	Pop 1 \geq Pop 2	Pop 1 \leq Pop 2
	Com diferenças	Pop 1 $<$ Pop 2	Pop 1 $>$ Pop 2
H_0	$\mu_D = 0$	$\mu_D \geq 0$	$\mu_D \leq 0$
H_A	$\mu_D \neq 0$	$\mu_D < 0$	$\mu_D > 0$

Nota: $D_i = X_{1i} - X_{2i}$ para a observação de ordem i

© J.P. Sousa

Slide 112 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t de amostras emparelhadas

Tabela de recolha de dados

Observação	Grupo 1	Grupo 2	Diferença
1	x_{11}	x_{21}	$D_1 = x_{11} - x_{21}$
2	x_{12}	x_{22}	$D_2 = x_{12} - x_{22}$
\vdots	\vdots	\vdots	\vdots
i	x_{1i}	x_{2i}	$D_i = x_{1i} - x_{2i}$
\vdots	\vdots	\vdots	\vdots
n	x_{1n}	x_{2n}	$D_n = x_{1n} - x_{2n}$

© J.P. Sousa

Slide 113 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t para amostras emparelhadas

$$t_{n-1} = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$$

$$gl = n - 1$$

Média da amostra

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}$$

Desvio padrão da amostra

$$S_D = \sqrt{\frac{\sum_{i=1}^n D_i^2 - n \cdot \bar{D}^2}{n - 1}}$$

© J.P. Sousa

Slide 114 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t para amostras emparelhadas Exemplo

O responsável por um herbário recorreu a um conjunto de testes para avaliar a eficiência do treino dos técnicos na medição de exemplars de herbário por dia:

<u>Nome</u>	<u>Antes (1)</u>	<u>Depois (2)</u>
João	85	94
Ana	94	87
Isabel	78	79
Manuel	87	88

Qual a eficiência do treino para um nível de 0,1?



© J.P. Sousa

Slide 115 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tabela de cálculo

Observação	Antes	Depois	Diferença
João	85	94	-9
Ana	94	87	7
Isabel	78	79	-1
Manuel	87	88	-1
Total			- 4

© J.P. Sousa

Slide 116 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Hipótese nula Solução

1. O treino foi eficaz?
2. Médias efectivas 'Antes' < 'Depois'.
3. Estatisticamente, temos que $\mu_A < \mu_B$.
4. Rearranjando os termos $\mu_A - \mu_B < 0$.
5. Definindo $\mu_D = \mu_A - \mu_B$ & substituindo em (4) obtemos $\mu_D < 0$.
6. Logo $H_0: \mu_D \geq 0$.

© J.P. Sousa

Slide 117 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste t para amostras emparelhadas Solução

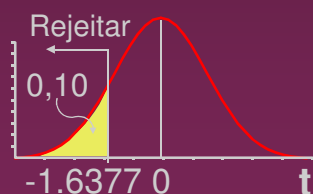
$H_0: \mu_D \geq 0$ ($\mu_D = \mu_A - \mu_B$)

$H_A: \mu_D < 0$

$\alpha = 0,10$

$df = 4 - 1 = 3$

Valores críticos:



© J.P. Sousa

Teste estatístico:

$$t = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} = \frac{-1 - 0}{\frac{6,53}{\sqrt{4}}} = -0,306$$

Decisão:

Não rejeitar para $\alpha = 0,10$

Conclusão:

Não podemos afirmar que o treino tenha sido eficaz

Slide 118 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste t para amostras emparelhadas

Problema

Estudou-se o crescimento (cm) das patas anteriores e posteriores de 8 coelhos. Pretende-se testar para um nível de **0,01** se as patas anteriores crescem menos do que as posteriores.

		(1)	(2)
	Coelho	Anter.	Post.
1	1	10	11
2	2	8	11
3	3	7	10
4	4	9	12
5	5	11	11
6	6	10	13
7	7	9	12
8	8	8	10

© J.P. Sousa Slide 119 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste t para amostras emparelhadas

Solução*

H₀: $\mu_D \geq 0$ ($\mu_D = \mu_1 - \mu_2$) **Teste estatístico:**

H_A: $\mu_D < 0$

$\alpha = 0,01$

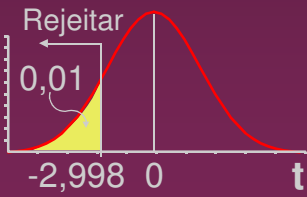
gl = 8 - 1 = 7

Valores críticos:

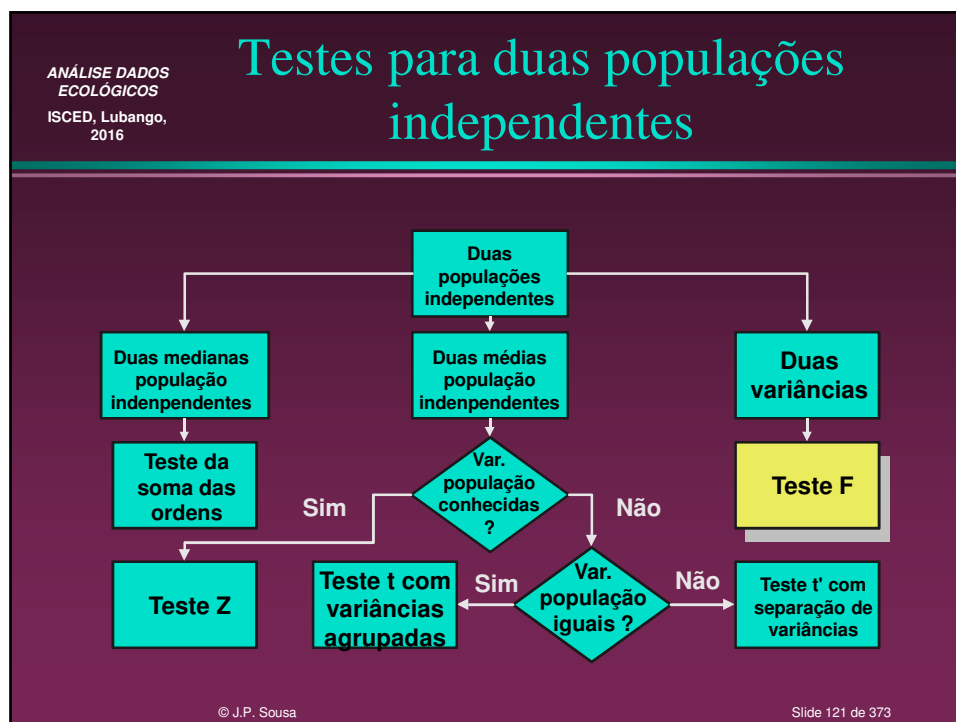
Decisão:
Rejeitar para $\alpha = 0,01$

Conclusão:
O crescimento das patas anteriores é menor do que o das patas posteriores

$$t = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} = \frac{-2,25 - 0}{\frac{1,16}{\sqrt{8}}} = -5,486$$



© J.P. Sousa Slide 120 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para as variâncias de duas populações

1. Teste para as diferenças de duas populações independentes
2. Teste paramétrico
3. Pressupostos
 - Ambas as populações têm distribuição normal
 - O teste não é muito robusto

© J.P. Sousa Slide 122 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para variância Hipóteses

1. Hipóteses

- $H_0: \sigma_1^2 = \sigma_2^2$ OU $H_0: \sigma_1^2 \geq \sigma_2^2$ (ou \leq)
- $H_A: \sigma_1^2 \neq \sigma_2^2$ $H_A: \sigma_1^2 < \sigma_2^2$ (ou $>$)

2. Teste estatístico

- $F = S_1^2 / S_2^2$ ou $F = S_2^2 / S_1^2$ (o que for maior)
- Dois conjuntos de graus de liberdade
 - $df_1 = n_1 - 1$; $df_2 = n_2 - 1$
- Segue a distribuição de F

© J.P. Sousa

Slide 123 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para variâncias Exemplo

Estudou-se a influência de dois tipos de alimentação (algas e leveduras) no tamanho de crustáceos à primeira reprodução:

	<u>Algas</u>	<u>Leveduras</u>
Número	21	25
Média	3,27	2,53
Des. Padrão	1,30	1,16

Existirá diferença de variâncias entre os dois tratamentos ($\alpha = 0,05$)?

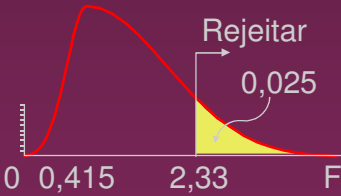
© J.P. Sousa

Slide 124 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para variâncias Solução

H₀: $\sigma_1^2 = \sigma_2^2$
H_A: $\sigma_1^2 \neq \sigma_2^2$
 $\alpha = 0,05$
gl₁ = 20 gl₂ = 24
Valores críticos:



Teste estatístico:

$$F = \frac{S_1^2}{S_2^2} = \frac{(1,30)^2}{(1,16)^2} = 1,25$$

Decisão:
 Não rejeitar para $\alpha = 0,05$

Conclusão:
 Não é evidente que existe diferença entre as variâncias

© J.P. Sousa Slide 125 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Variabilidade acerca da média

- Descrição da variabilidade de uma amostra da população
 - $\bar{x} \pm s$
- Descrição da precisão de estimativa da média da população
 - $\bar{x} \pm s_{\bar{x}}$
 - Limites de confiança

© J.P. Sousa Slide 126 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Limites de confiança para a média da população

$$P\left[-t_{0,05(2)v} \leq \frac{\bar{X} - \mu}{s_{\bar{X}}} \leq t_{0,05(2)v}\right] = 0,95$$

$$P\left[\bar{X} - t_{\alpha(2)v} s_{\bar{X}} \leq \mu \leq \bar{X} + t_{\alpha(2)v} s_{\bar{X}}\right] = 1 - \alpha$$

© J.P. Sousa

Slide 127 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016


Apresentação da variabilidade da média

- Devem indicar-se claramente
 - o tamanho da amostra
 - a medida de variabilidade apresentada
 - a unidade de medida

© J.P. Sousa

Slide 128 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016



MÃOS À OBRA 1

© J.P. Sousa Slide 129 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Análise de Variância: ANOVA

© J.P. Sousa Slide 130 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Observação & Amostragem

1. Inexistência de controle sobre factores particulares que afectam a variável de interesse
2. Exemplo
 - Estudo da densidade de carpas em 10 cursos de água (Variável de interesse). Esta pode variar consoante o teor de clorofila-a (Factor explicativo não controlada)

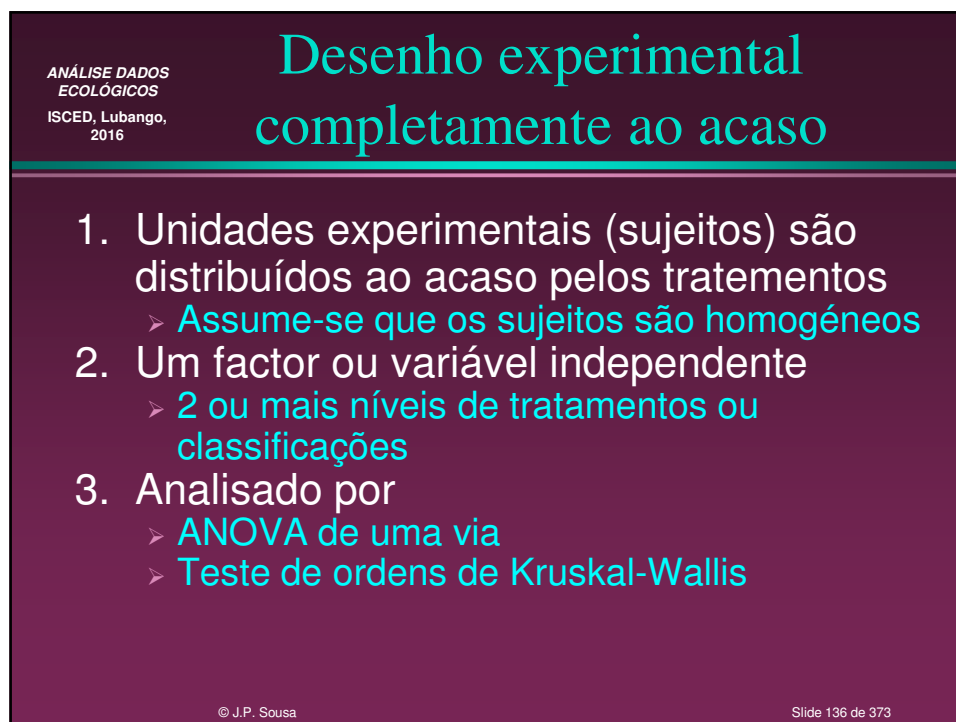
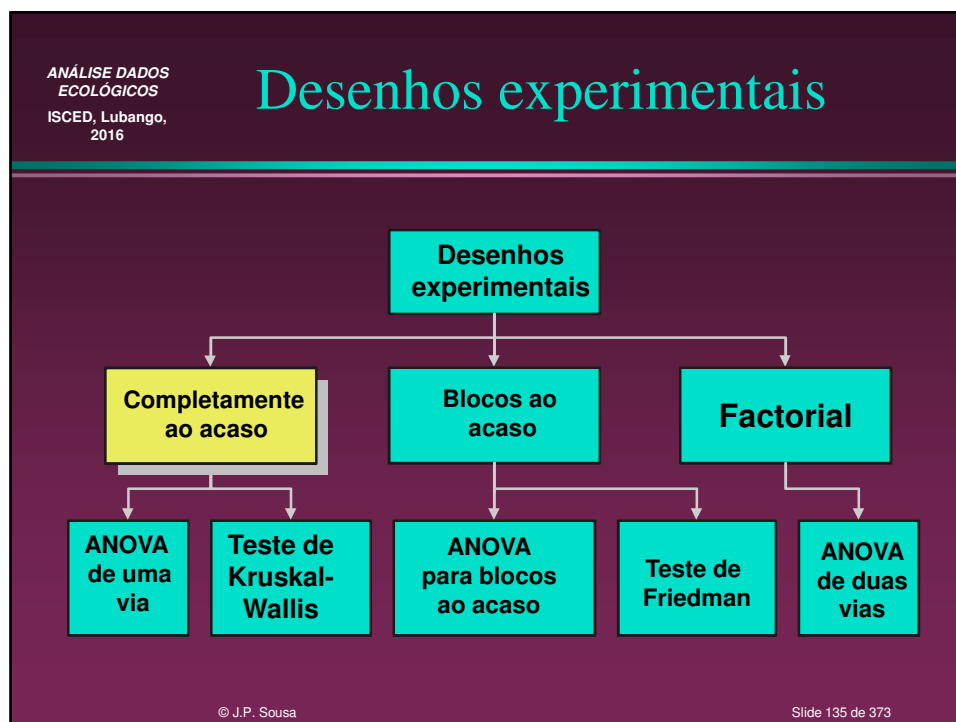
© J.P. Sousa Slide 132 de 373

Desenho experimental

1. O investigador controla uma ou mais variáveis independentes
 - Denominadas variáveis do tratamento ou **factores**
 - Contêm dois ou mais **níveis** (Subcategorias)
2. Observa efeitos sobre a variável dependente
 - Resposta para os diferentes níveis da variável independente
3. Desenho experimental - Plano utilizado para testar hipóteses




Exemplos de desenhos experimentais

1. Quarenta animais são incluídos ao acaso em grupos de 1 a 4 (**Níveis**) correspondentes a diferentes regimes alimentares (**Variável independente**) para estudar o seu efeito no peso corporal (**Variável dependente**).
2. Quinze amostras de uma planta foram distribuídas ao acaso por grupos identificados de 1 a 3 (**Níveis**) que foram atribuídos a diferentes técnicos (**Variável independente**) para estudar o conteúdo de fósforo (**Variável dependente**).



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenho experimental
completamente ao acaso

	Factor (Local-Floresta)		
Níveis (Tratamentos)	Nível 1 Floresta 1	Nível 2 Floresta 2	Nível 3 Floresta 3
Unidades Experimentais			
Variável Dependente (Resposta)	21	17	31
	27	25	28
	29	20	22

© J.P. Sousa

Slide 137 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais

Desenhos
experimentais

Completamente
ao acaso

ANOVA
de uma
via

Teste de
Kruskal-
Wallis

Blocos ao
acaso

Teste F
para blocos
ao acaso

Factorial

Teste de
Friedman

ANOVA
de duas
vias

© J.P. Sousa

Slide 138 de 373

Teste F de uma via

1. Testa a igualdade de duas ou mais médias da população
2. Variáveis
 - Uma variável nominal independente
 - 2 ou mais níveis de tratamento ou classificações
 - Uma variável dependente intervalada ou fracionária
3. Utilizado para analisar desenhos experimentais completamente ao acaso

Pressupostos do teste F para a ANOVA de uma via

1. Erro ao acaso e independente
 - As amostras devem ser independentes e obtidas ao acaso
2. Normalidade
 - Populações normalmente distribuídas
3. Homogeneidade de variância
 - As populações têm variâncias iguais

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

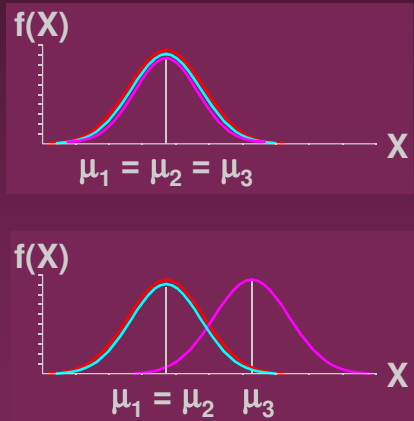
Hipóteses do Teste F para ANOVA de uma via

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_c$

- Todas as médias da população são iguais
- Não há efeito dos tratamentos

H_1 : Nem todas μ_j são iguais

- Pelo menos uma média é diferente
- Há efeito dos tratamentos
- $\mu_1 \neq \mu_2 \neq \dots \neq \mu_c$ é **errado**



© J.P. Sousa

Slide 141 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

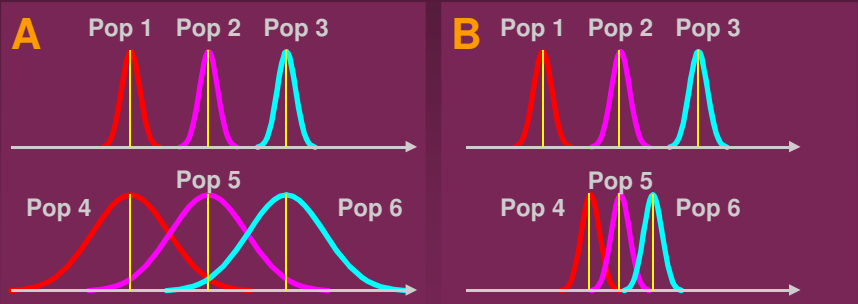
O valor das variâncias afecta as decisões acerca da igualdade das médias

Mesma variação dos tratamentos
Variação ao acaso é diferente

Variação diferente dos tratamentos
Mesma variação ao acaso

A Pop 1 Pop 2 Pop 3
Pop 4 Pop 5 Pop 6
Variâncias **dentro** diferem

B Pop 1 Pop 2 Pop 3
Pop 4 Pop 5 Pop 6
Variâncias **entre** diferem



© J.P. Sousa

Slide 142 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de uma via Ideia Básica

1. Compara 2 tipos de variação para testar a igualdade das médias
2. A comparação baseia-se no quociente de variâncias
3. Se a variação dos tratamentos for significativamente maior que a variância ao acaso então as médias **não são** iguais
4. As medidas de variação são obtidas por 'partição' da variância total

© J.P. Sousa Slide 143 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de uma via Partição da variação total

```
graph TD; A[Variação total] --> B[Variação devida ao tratamento]; A --> C[Variação devida a amostragem ao acaso];
```

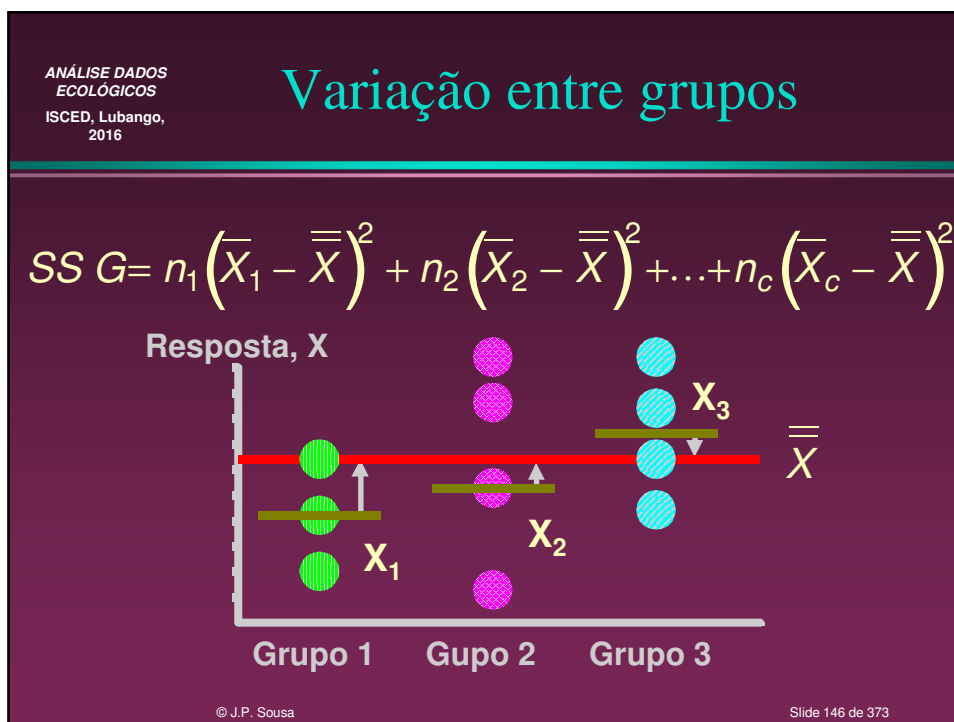
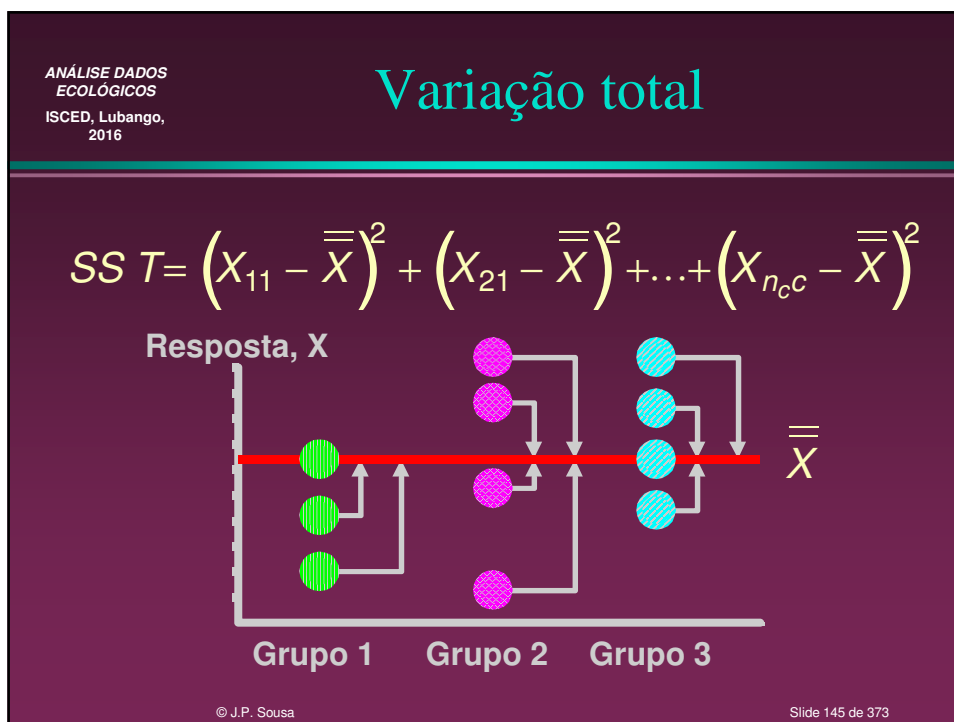
Variação devida ao tratamento

- Soma de quadrados entre
- Soma de quadrados do modelo
- Variação entre grupos

Variação devida a amostragem ao acaso

- Soma de quadrados dentro
- Soma de quadrados do erro
- Variação dentro dos grupos

© J.P. Sousa Slide 144 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Variação dentro dos grupos

$$SS E = (X_{11} - \bar{X}_1)^2 + (X_{21} - \bar{X}_1)^2 + \dots + (X_{nc} - \bar{X}_c)^2$$

Resposta, X

Grupo 1 Grupo 2 Grupo 3

© J.P. Sousa Slide 147 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para ANOVA de uma via

Teste estatístico

1. Teste Estatístico
 - $F = MS \text{ Grupo} / MS \text{ Erro}$
 - $MS \text{ Grupo}$ é a média de quadrados **entre**
 - $MS \text{ Erro}$ é a média de quadrados **dentro**
2. Graus de liberdade
 - $gl_1 = c - 1$
 - $gl_2 = n - c$
 - $c = \# \text{ colunas (Populações, Grupos, ou níveis)}$
 - $n = \text{Tamanho total da amostra}$

© J.P. Sousa Slide 148 de 373

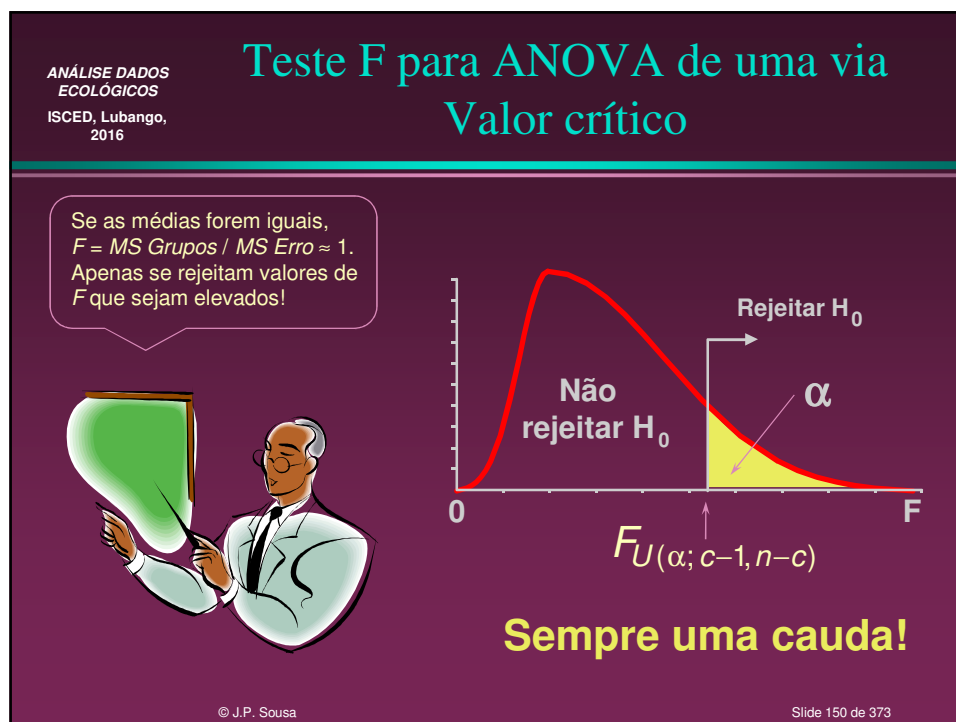
ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

ANOVA de uma via

Tabela de resumo

Fontes de variação	Graus de liberdade	Soma de Quadrados	Quadrados médios (variância)	F
Entre (Factor)	$c - 1$	SS grupo (SS G)	$MS \text{ grupo} = SS G / (c - 1)$	$\frac{MS \text{ Grupo}}{MS \text{ Erro}}$
Dentro (Erro)	$N - c$	SS erro (SS E)	$MS \text{ Erro} = SS E / (n - c)$	
Total	$N - 1$	$SS T = SS G + SS E$		

© J.P. Sousa Slide 149 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA de uma via Exemplo

Estudou-se o efeito da ração no peso de javalis. Dezanove animais foram distribuídos ao acaso por quatro grupos correspondentes a dietas diferentes. Pretende-se testar se os pesos dos animais (kg) é o mesmo para todas as dietas?

<u>Dieta 1</u>	<u>Dieta 2</u>	<u>Dieta 3</u>	<u>Dieta 4</u>
60,8	68,7	102,6	87,9
57,0	67,7	102,1	84,2
65,0	74,0	100,2	83,1
58,6	66,3	96,5	85,7
61,7	69,8		90,3



© J.P. Sousa

Slide 151 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tabela de resumo Solução

Fontes de variação	Graus de Liberdade	Soma de Quadrados	Média de Quadrados (Variância)	F
Entre (Rações)	$4 - 1 = 3$	4226,348	1408,783	165
Dentro (Erro)	$19 - 4 = 15$	128,350	8,557	
Total	$19 - 1 = 18$	4354,698		

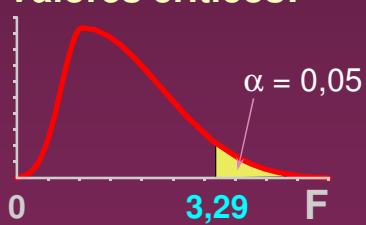
© J.P. Sousa

Slide 152 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para ANOVA de uma via Solução

H₀: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_A: Nem todos são iguais
 α = 0,05
gh = 3 **gl** = 15
Valores críticos:



Teste estatístico:

$$F = \frac{MS \text{ Grupos}}{MS \text{ erro}} = \frac{1408,783}{8,557} = 165$$

Decisão:
Rejeitar para $\alpha = 0,05$

Conclusão:
Há evidências de que existem diferenças entre as médias das populações

© J.P. Sousa Slide 153 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Comparações múltiplas

© J.P. Sousa Slide 154 de 373

Comparações múltiplas

1. Utilizam-se geralmente após ANOVA modelo I
2. Pressupostos
 - Populações normais
 - Homogeneidade de variâncias
3. Paramétricas
4. Não paramétricas
 - Violação dos pressupostos

Testes para comparações múltiplas

1. Testes mais robustos e mais utilizados
 - Teste de Tukey
 - Teste de Newman-Keuls
2. Outros métodos
 - LSD (Diferença mínima de quadrados)
 - LSD com ajustamento de Bonferroni (Teste de Bonferroni)
 - Teste de Duncan

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Testes para comparações múltiplas com um controle

1. Teste de Dunnet

- Não se fazem todas as comparações possíveis
- Comparam-se apenas os grupos com um controle

© J.P. Sousa

Slide 157 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

1. Teste das Diferenças Completamente Significativas

2. Hipóteses

- $H_0: \mu_B = \mu_A$
- $H_A: \mu_B \neq \mu_A$

3. Comparações de k grupos

- $k(k-1)/2$

© J.P. Sousa

Slide 158 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

Aplicação

- Ordenação das médias por ordem crescente
- Comparações efectuadas entre os maiores e os menores valores
- Não existindo diferenças entre duas médias, então as médias situadas entre elas não são testadas

© J.P. Sousa

Slide 159 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

Teste Estatístico

$$q = \frac{\bar{X}_A - \bar{X}_B}{SE}$$

(para grupos de igual tamanho, n)

$$SE = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{MSE}{n}}$$

(para grupos de tamanho desigual, n_A, n_B)

$$SE = \sqrt{\frac{MSE}{2} \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}$$

© J.P. Sousa

Slide 160 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

Exemplo

Determinou-se a concentração de do metal estrôncio (Sr) em cinco cursos de água (6 amostras cada). Pretende-se verificar se existem diferenças de concentração entre os locais e em caso afirmativo quais os locais que diferem entre si.



© J.P. Sousa

Slide 161 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

Exemplo

	Grayson	Beaver	Angler	Appletree	Rock
	28,2	39,6	46,3	41,0	56,3
	33,2	40,8	42,1	44,1	54,1
	36,4	37,9	43,5	46,4	59,4
	34,6	37,1	48,8	40,2	62,7
	29,1	43,6	43,7	38,6	60,0
	31,0	42,4	40,1	36,3	57,3
ΣX_i	192,5	241,4	264,5	246,6	349,8
ΣX_i^2	6227,4	9744,3	11707,5	10202,5	20439,4
n_i	6	6	6	6	6
\bar{X}_i	32,1	40,2	44,1	41,1	58,3

© J.P. Sousa

Slide 162 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Análise de variância

	Grayson	Beaver	Angler	Appletree	Rock
SS_i	51,4	32,0	47,4	67,2	46,1
$\bar{x}_i^2 \cdot n_i$	6176,0	9712,3	11660,0	10135,3	20393,4

$$\bar{x}^2 \cdot N = 30 (43,2)^2 = 55883,6$$

$$SS\ E = \sum SS_i = 244,1$$

$$SS\ G = \sum \bar{x}_i^2 \cdot n_i - \bar{x}^2 \cdot N = 58077,0 - 55883,6 = 2193,4$$

$$SS\ T = SS\ E + SS\ G = 244,1 + 2193,4 = 2437,5$$

© J.P. Sousa

Slide 163 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Tabela de Análise de Variância

Fontes de variação	Graus de liberdade	Soma de Quadrados	Quadrados médios (variância)	F
Entre (Grupo)	5 - 1	2193,4	548,4	$\frac{548,4}{9,8}$
Dentro (Erro)	30 - 5	244,1	9,8	= 56,2
Total	30 - 1	2437,6		

© J.P. Sousa

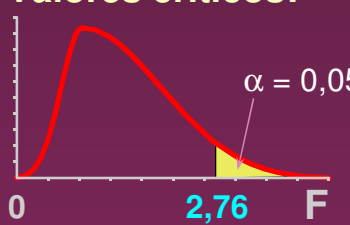
Slide 164 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para ANOVA de uma via Solução

H₀: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
H_A: Nem todas são iguais
 $\alpha = 0,05$
gh = 4 gl = 25

Valores críticos:



$\alpha = 0,05$
2,76 F

Teste estatístico:

$$F = \frac{MS \text{ Grupos}}{MS \text{ erro}} = \frac{548,4}{9,8} = 56,2$$

Decisão:
Rejeitar para $\alpha = 0,05$

Conclusão:
Há evidências de que existem diferenças entre as médias das populações

© J.P. Sousa Slide 165 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste de Tukey Exemplo

Como a hipótese nula foi rejeitada somos forçados a concluir que nem todas as médias são iguais pelo que é necessário aplicar o teste de Tukey para comparações múltiplas.

Grupos de tamanho igual (n=6)

$$SE = \sqrt{\frac{9,8}{6}} = \sqrt{1,63} = 1,28$$

$$q_{0,05, 25, 5} \cong q_{0,05, 24, 5} = 4,166$$

© J.P. Sousa Slide 166 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

	Grayson	Beaver	Angler	Appletree	Rock
\bar{X}_i	32,1	40,2	44,1	41,1	58,3

Médias ordenadas

Ordem (i)	1	2	4	3	5
\bar{X}_i	<u>32,1</u>	<u>40,2</u>	<u>41,1</u>	<u>44,1</u>	<u>58,3</u>

© J.P. Sousa Slide 167 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Tukey

B vs A	$\bar{X}_A - \bar{X}_B$	SE	q	$q_{0,05, 24, 5}$	Conclusão
5 vs 1	58,3-32,1=26,2	1,28	20,47	4,166	Rejeitar $H_0: \mu_5=\mu_1$
5 vs 2	58,3-40,2=18,1	1,28	14,14	4,166	Rejeitar $H_0: \mu_5=\mu_2$
5 vs 4	58,3-41,1=17,2	1,28	13,44	4,166	Rejeitar $H_0: \mu_5=\mu_4$
5 vs 3	58,3-44,1=14,2	1,28	11,09	4,166	Rejeitar $H_0: \mu_5=\mu_3$
3 vs 1	44,1-32,1=12,0	1,28	9,38	4,166	Rejeitar $H_0: \mu_3=\mu_1$
3 vs 2	44,1-40,2= 3,9	1,28	3,05	4,166	Aceitar $H_0: \mu_3=\mu_2$
3 vs 4	não testar				
4 vs 1	41,1-32,1= 9,0	1,28	7,03	4,166	Rejeitar $H_0: \mu_4=\mu_1$
4 vs 2	não testar				
2 vs 1	40,2-32,1= 8,1	1,28	6,33	4,166	Rejeitar $H_0: \mu_2=\mu_1$

© J.P. Sousa Slide 168 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Newman-Keuls

B vs A	\bar{X}_A	\bar{X}_B	SE	q	p	$q_{0,05, 24, p}$	Número de grupos a comparar	Conclusão
5 vs 1	58,3	32,1	26,2	1,28	20,47	5	4,166	Rejeitar $H_0: \mu_5 = \mu_1$
5 vs 2	58,3	40,2	18,1	1,28	14,14	4	3,901	Rejeitar $H_0: \mu_5 = \mu_2$
5 vs 4	58,3	41,1	17,2	1,28	13,44	3	3,532	Rejeitar $H_0: \mu_5 = \mu_4$
5 vs 3	58,3	44,1	14,2	1,28	11,09	2	2,919	Rejeitar $H_0: \mu_5 = \mu_3$
3 vs 1	44,1	32,1	12,0	1,28	9,38	4	3,901	Rejeitar $H_0: \mu_3 = \mu_1$
3 vs 2	44,1	40,2	3,9	1,28	3,05	3	3,532	Aceitar $H_0: \mu_3 = \mu_2$
3 vs 4	não testar							
4 vs 1	41,1	32,1	9,0	1,28	7,03	3	3,532	Rejeitar $H_0: \mu_4 = \mu_1$
4 vs 2	não testar							
2 vs 1	40,2	32,1	8,1	1,28	6,33	2	2,919	Rejeitar $H_0: \mu_2 = \mu_1$

© J.P. Sousa

Slide 169 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Comparações múltiplas Intervalos de confiança

- Fórmula geral (k=2)

$$\bar{X}_i \pm t_{\alpha(2), v} \sqrt{\frac{s^2}{n_i}}$$

- Para k>2

s^2 - MSE
 v - graus de liberdade da análise de variância

© J.P. Sousa

Slide 170 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Comparações múltiplas Intervalos de confiança

Se duas ou mais médias não diferem entre si

$$\bar{x}_p = \frac{\sum n_i \bar{x}_i}{\sum n_i}$$

$$\bar{x}_p \pm t_{\alpha(2),v} \sqrt{\frac{s^2}{n_p}}$$

© J.P. Sousa Slide 171 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Intervalos de confiança

	Grayson	Beaver	Angler	Appletree	Rock
\bar{x}_i	32,1	40,2	44,1	41,1	58,3

Médias ordenadas

Ordem (i)	1	2	4	3	5
\bar{x}_i	<u>32,1</u>	<u>40,2</u>	41,1	<u>44,1</u>	<u>58,3</u>
n_i	6	6	6	6	6

© J.P. Sousa Slide 172 de 373

Intervalos de confiança

Médias ordenadas

Ordem (i)	1	2	4	3	5
\bar{x}_i	<u>32,1</u>	<u>40,2</u>	<u>41,1</u>	<u>44,1</u>	<u>58,3</u>
n_i	6	6	6	6	6

$$95\% \text{ IC para } \mu_1 = \bar{x}_1 \pm t_{0,05(2),25} \sqrt{\frac{s^2}{n_1}} = 32,1 \pm 2,060 \sqrt{\frac{9,7652}{6}} = 32,1 \text{ mg / ml} \pm 2,6 \text{ mg / ml}$$

$$95\% \text{ IC para } \mu_5 = \bar{x}_5 \pm t_{0,05(2),25} \sqrt{\frac{s^2}{n_5}} = 58,3 \pm 2,060 \sqrt{\frac{9,7652}{6}} = 58,3 \text{ mg / ml} \pm 2,6 \text{ mg / ml}$$

Intervalos de confiança

Médias ordenadas

Ordem (i)	1	2	4	3	5
\bar{x}_i	<u>32,1</u>	<u>40,2</u>	<u>41,1</u>	<u>44,1</u>	<u>58,3</u>
n_i	6	6	6	6	6

$$\bar{x}_p = \bar{x}_{2,4,3} = \frac{n_2 \bar{x}_2 + n_4 \bar{x}_4 + n_3 \bar{x}_3}{n_2 + n_4 + n_3} = \frac{(6)(40,2) + (6)(41,1) + (6)(44,1)}{6 + 6 + 6} = 41,8 \text{ mg / ml}$$

$$95\% \text{ IC para } \mu_{2,4,3} = \bar{x}_{2,4,3} \pm t_{0,05(2),25} \sqrt{\frac{s^2}{n_5}} = 41,8 \pm 2,060 \sqrt{\frac{9,7652}{6 + 6 + 6}} = 41,8 \text{ mg / ml} \pm 1,5 \text{ mg / ml}$$

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Comparação múltiplas com um controle Teste de Dunnet

Comparação de cada tratamento com um considerado como controle

- (k-1 combinações)

Teste de Dunnet

- Fórmula
- Valor crítico ($q'_{\alpha, v, k}$)

© J.P. Sousa

Slide 175 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Dunnet Aplicação

- Ordenação das médias por ordem crescente
- Comparações efectuadas entre os maiores e os menores valores
- Não existindo diferenças entre duas médias, então as médias situadas entre elas não são testadas

© J.P. Sousa

Slide 176 de 373

Teste de Dunnet

$$q = \frac{\bar{X}_{\text{controle}} - \bar{X}_A}{SE}$$

(para grupos de
igual tamanho, n)

(para grupos de tamanho
desigual, n_A , n_B)

$$SE = \sqrt{\frac{2 s^2}{n}} = \sqrt{\frac{2 \text{MSE}}{n}}$$

$$SE = \sqrt{\text{MSE} \left(\frac{1}{n_{\text{controle}}} + \frac{1}{n_A} \right)}$$

Teste de Dunnet Exemplo

A produção de batata em diversos lotes de terreno foi determinada após aplicação de um fertilizante corrente. Noutros lotes efectuou-se a aplicação de quatro outros fertilizantes. O fabricante pretende provar que os novos fertilizantes aumentam a produção.

Assim 24 lotes (grupo 2) foram adubados com o fertilizante corrente enquanto cada um dos novos fertilizantes foi aplicado em 14 lotes cada um (grupos 1, 3, 4, 5).

A análise de variância dos dados provou que existiam diferenças entre as médias e que a variância do erro é de 10,42.



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Dunnet Exemplo

Erro padrão

$$SE = \sqrt{10,42 \left(\frac{1}{24} + \frac{1}{14} \right)} = 1,1 \text{ ton / ha}$$

Ordem das médias	1	2	3	4	5
Média dos grupos	17,3	21,7	22,1	23,6	27,8

Hipóteses

- $H_0: \mu_2 \geq \mu_A$
- $H_A: \mu_2 < \mu_A$

© J.P. Sousa Slide 179 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Dunnet Exemplo

2 vs A	$\bar{X}_2 - \bar{X}_A$	SE	q	$q'_{0,05, 75, 5}$	Conclusão
2 vs 1	21,7-17,3=4,4	Já que	$\mu_2 > \mu_1$, então...		Aceitar $H_0: \mu_2 \geq \mu_1$
2 vs 5	21,7-27,8=-6,1	1,1	-5,55	2,21	Rejeitar $H_0: \mu_2 \geq \mu_5$
2 vs 4	21,7-23,6=-1,9	1,1	-1,73	2,21	Aceitar $H_0: \mu_2 \geq \mu_4$
2 vs 3	não testar				

© J.P. Sousa Slide 180 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Avaliação da Homogeneidade de Variâncias

© J.P. Sousa

Slide 181 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Homogeneidade de Variâncias

- Três ou mais amostras
- Todas as amostras provêm de populações com variâncias idênticas
 - $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$
- Homoscedasticidade
- Heterocedasticidade
- Teste estatístico
 - Teste de Bartlett

© J.P. Sousa

Slide 182 de 373

Teste de Bartlett

1. Pressupostos

- Amostras retiradas de populações normais
- $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$

2. Equivalente ao quociente de variâncias

- $k=2$ e $n_1=n_2$

Teste de Bartlett

$$B = (\ln s_p^2) \left(\sum_{i=1}^k v_i \right) - \sum_{i=1}^k v_i \ln s_i^2$$

$$v = n_i - 1$$

Valor crítico de B segue a distribuição do χ^2 com $n = k-1$

$$s_p^2 = \frac{\sum_{i=1}^k SS_i}{\sum_{i=1}^k v_i}$$

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Bartlett para a homogeneidade de variâncias

Estudou-se o efeito da ração no peso de porcos selvagens. Dezanove animais foram distribuídos ao acaso por quatro grupos correspondentes a dietas diferentes. Pretende-se testar se a variância dos pesos dos animais (kg) é o mesmo para todas as dietas?

<u>Dieta 1</u>	<u>Dieta 2</u>	<u>Dieta 3</u>	<u>Dieta 4</u>
60,8	68,7	102,6	87,9
57,0	67,7	102,1	84,2
65,0	74,0	100,2	83,1
58,6	66,3	96,5	85,7
61,7	69,8		90,3

© J.P. Sousa

Slide 185 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Bartlett Solução

	<u>Dieta 1</u>	<u>Dieta 2</u>	<u>Dieta 3</u>	<u>Dieta 4</u>	
	60,8	68,7	102,6	87,9	
	57,0	67,7	102,1	84,2	
	65,0	74,0	100,2	83,1	
	58,6	66,3	96,5	85,7	
	61,7	69,8		90,3	
SS _i	37,57	34,26	22,97	33,55	$\Sigma SS_i = 128,35$
v_i	4	4	3	4	$\Sigma v_i = 15$
s_i^2	9,39	8,56	7,66	8,39	
$\ln s_i^2$	2,239	2,147	2,036	2,127	
$v_i \ln s_i^2$	8,956	8,588	6,108	8,508	$\Sigma v_i \ln s_i^2 = 32,16$

© J.P. Sousa

Slide 186 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Bartlett Solução

$$s_p^2 = \frac{\sum SS_i}{\sum \nu_i} = \frac{128,35}{15} = 8,56$$

$$B = (\ln s_p^2) \left(\sum_{i=1}^k \nu_i \right) - \sum_{i=1}^k \nu_i \ln s_i^2$$

$$= (\ln 8,56)(15) - 32,163$$

$$= 32,2065 - 32,163$$

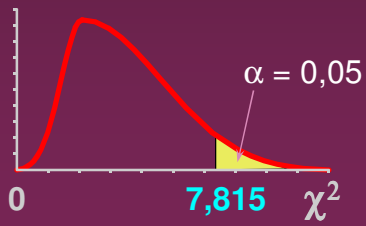
$$= 0,0435$$

© J.P. Sousa Slide 187 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste de Bartlett Solução

H₀: $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$
H_A: As variâncias são heterogêneas
 α = 0,05
gl = 4-1 = 3
Valores críticos:



Teste estatístico:
 $B = 0,0435$

Decisão:
 Não rejeitar para $\alpha = 0,05$

Conclusão:
 Não há evidências de que as variâncias sejam heterogêneas

© J.P. Sousa Slide 188 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais

Desenhos factoriais

ANOVA de 2 vias

© J.P. Sousa
Slide 189 de 373

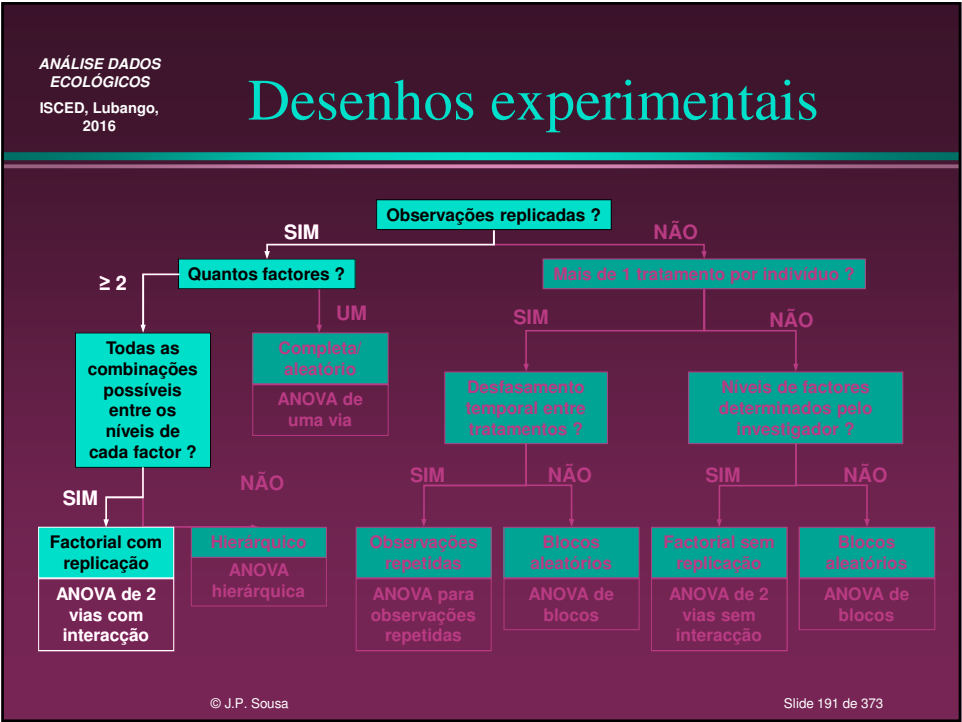
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais

```

graph TD
    A[Observações replicadas?] -- SIM --> B[Quantos factores?]
    A -- NÃO --> C[Mais de 1 tratamento por indivíduo?]
    B -- > 2 --> D[Todas as combinações possíveis entre os níveis de cada factor?]
    B -- UM --> E[Completa/aleatório]
    E --> F[ANOVA de uma via]
    D -- SIM --> G[Factorial com replicação]
    D -- NÃO --> H[Hierárquico]
    G --> I[ANOVA de 2 vias com interacção]
    H --> J[ANOVA hierárquica]
    C -- SIM --> K[Desfasamento temporal entre tratamentos?]
    C -- NÃO --> L[Níveis de factores determinados pelo investigador?]
    K -- SIM --> M[Observações repetidas]
    K -- NÃO --> N[Blocos aleatórios]
    M --> O[ANOVA para observações repetidas]
    N --> P[ANOVA de blocos]
    L -- SIM --> Q[Factorial sem replicação]
    L -- NÃO --> R[Blocos aleatórios]
    Q --> S[ANOVA de 2 vias sem interacção]
    R --> T[ANOVA de blocos]
    
```

© J.P. Sousa
Slide 190 de 373




© J.P. Sousa

Slide 191 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenho factorial
Exemplo




Factor 1 (Hormona)		
Nível 1 (sem)	Nível 2 (C1)	Nível 3 (C2)
14,5 µg/ml	25.4 µg/ml	36.7 µg/ml
11.0 µg/ml	30.1 µg/ml	39.2 µg/ml
16.0 µg/ml	35.2 µg/ml	42.4 µg/ml
18.3 µg/ml	38.4 µg/ml	46.3 µg/ml

© J.P. Sousa

Slide 192 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Desenho factorial Exemplo



		Factor 1 (Hormona)		
		Nível 1 (sem)	Nível 2 (C1)	Nível 3 (C2)
Factor 2 (Sexo)	Nível 1 (Machos)	14,5 µg/ml 11.0 µg/ml	25.4 µg/ml 30.1 µg/ml	36.7 µg/ml 39.2 µg/ml
	Nível 2 (Fêmeas)	16.0 µg/ml 18.3 µg/ml	35.2 µg/ml 38.4 µg/ml	42.4 µg/ml 46.3 µg/ml

© J.P. Sousa Slide 193 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Desenhos Factoriais

- 1.As unidades experimentais (**Sujeitos**) são distribuídos ao acaso pelos tratamentos
 - Os sujeitos assumem-se como homogêneos
- 2.Dois ou mais **Factores** ou variáveis independentes
 - Cada um tem 2 ou mais Tratamentos (**Níveis**)
- 3.Analisado por **ANOVA de duas vias**

© J.P. Sousa Slide 194 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Vantagens dos desenhos factoriais

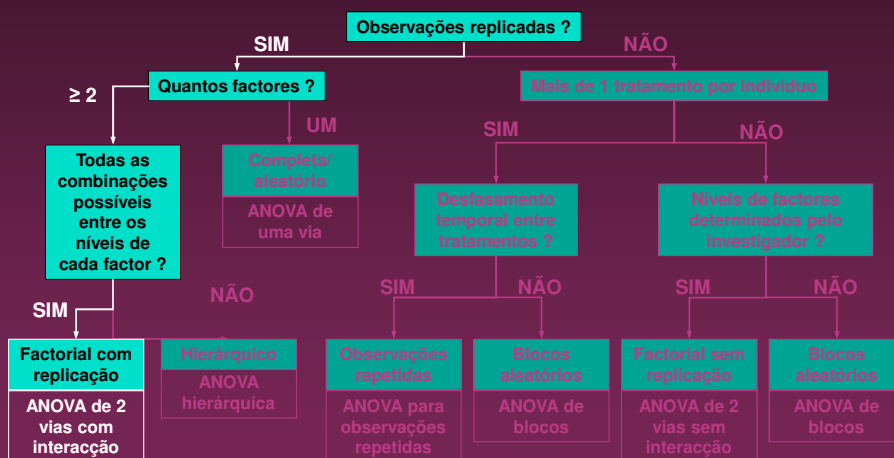
- 1. Pouparam tempo e esforço
➤ e.g., Poder-se-iam utilizar desenhos experimentais completamente ao acaso para cada variável
- 2. Torna os resultados mais claros pela introdução de outras variáveis no modelo
- 3. Pode explorar a interacção entre variáveis

© J.P. Sousa

Slide 195 de 373

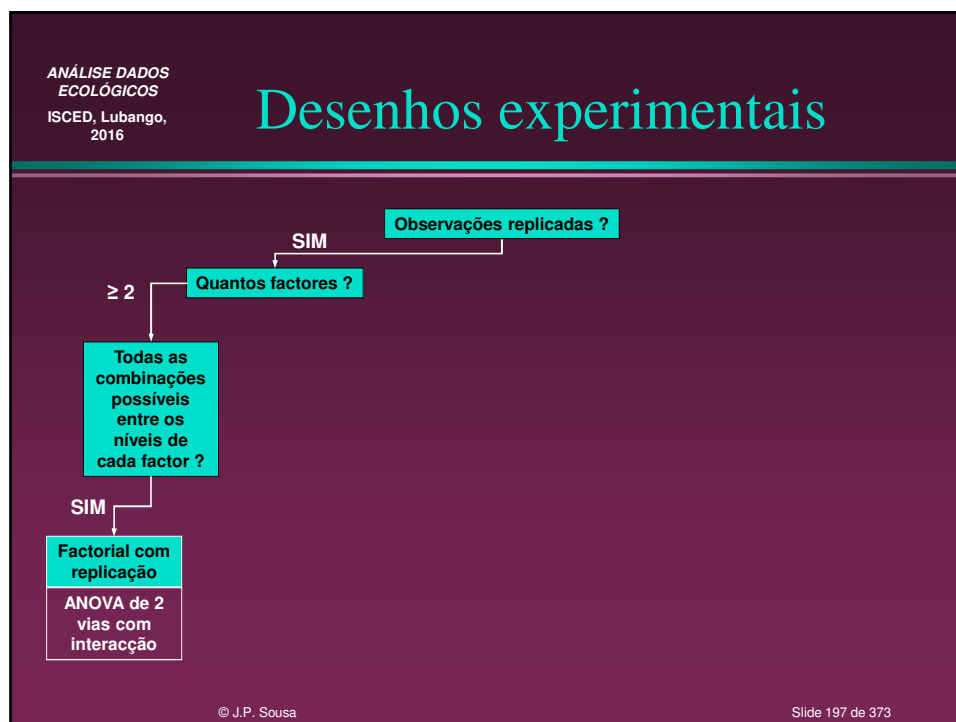
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais



© J.P. Sousa

Slide 196 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

ANOVA de duas vias

- 1. Testa a igualdade de 2 ou mais (μ) Médias de População quando se utilizam diversas variáveis independentes
- 2. Mesmos resultados que a ANOVA de uma via para cada Variável
 - Nenhuma interacção pode ser testada
- 3. Utilizada para analisar desenhos factoriais

© J.P. Sousa Slide 198 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias Pressupostos

- 1. Normalidade
 - As populações encontram-se normalmente distribuídas
- 2. Homogeneidade de variância
 - Populações têm variâncias iguais
- 3. Independência dos erros
 - Utilizam-se amostras ao acaso

© J.P. Sousa

Slide 199 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias Tabela de Dados

Factor	Factor B			
A	1	2	...	b
1	X_{111}	X_{121}	...	X_{1b1}
	X_{112}	X_{122}	...	X_{1b2}
2	X_{211}	X_{221}	...	X_{2b1}
	X_{212}	X_{222}	...	X_{2b2}
:	:	:	:	:
a	X_{a11}	X_{a21}	...	X_{ab1}
	X_{a12}	X_{a22}	...	X_{ab2}



© J.P. Sousa

Slide 200 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

ANOVA de duas vias

Hipóteses nulas

- 1. Não há diferenças de médias devido ao Factor A
 $\text{➤ } H_0: \mu_{1..} = \mu_{2..} = \dots = \mu_{a..}$
- 2. Não há diferenças de médias devido ao Factor B
 $\text{➤ } H_0: \mu_{.1.} = \mu_{.2.} = \dots = \mu_{.b.}$
- 3. Não há interacção entre os factores A & B
 $\text{➤ } H_0: AB_{ij} = 0$

© J.P. Sousa
Slide 201 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias
Desenho experimental

Factor A

Factor B

Nível 1Nível 2Nível 3Nível 4

Nível 1

X

X X

X

X X

X

X X

X

X X

Nível 2

X

X X

X

X X

X

X X

X

X X

Nível 3

X

X X

X

X X

X

X X

X

X X

© J.P. Sousa

Slide 203 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias
Somadas de quadrados total (SST)

Factor A

Factor B

Nível 1Nível 2Nível 3Nível 4

Nível 1

X

X X

X

X X

X

X X

X

X X

Nível 2

X

X X

X

X X

X

X X

X

X X

Nível 3

X

X X

X

X X

X

X X

X

X X

A soma de quadrados total é calculada a partir de todas as observações individuais.

© J.P. Sousa

Slide 204 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias

Somadas de quadrados total (SST)

$$SST = (X_{111} - \bar{\bar{X}})^2 + (X_{112} - \bar{\bar{X}})^2 + \dots + (X_{ijk} - \bar{\bar{X}})^2$$

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (X_{ijk} - \bar{\bar{X}})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk} \right)^2}{N}$$

GL Total = N-1

© J.P. Sousa

Slide 205 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias

Somadas de quadrados do factor A (SSFA)

•Factor B

•Nível 1 Nível 2 Nível 3 Nível 4

Factor A	Nível 1	X	X	X	X
		X	X	X	X
	Nível 2	X	X	X	X
		X	X	X	X
	Nível 3	X	X	X	X
		X	X	X	X

A soma de quadrados do factor A é calculada ignorando a variabilidade interna do factor A.

© J.P. Sousa

Slide 206 de 373

ANOVA de duas vias
Somos de quadrados do factor A (SSFA)

$$SSFA = bn \left[(\bar{X}_1 - \bar{\bar{X}})^2 + (\bar{X}_2 - \bar{\bar{X}})^2 + \dots + (\bar{X}_i - \bar{\bar{X}})^2 \right]$$

$$SSFA = bn \sum_{i=1}^a (\bar{X}_i - \bar{\bar{X}})^2 = \frac{\sum_{i=1}^a \left(\sum_{j=1}^b \sum_{k=1}^n X_{ijk} \right)^2}{bn} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk} \right)^2}{N}$$

$$GL \text{ Factor A} = a-1$$

ANOVA de duas vias
Somos de quadrados do factor B (SSFB)

•Factor B

•Nível 1 Nível 2 Nível 3 Nível 4

Factor A	Nível 1	<div>X X X</div>	<div>X X X</div>	<div>X X X</div>	<div>X X X</div>
	Nível 2	<div>X X X</div>	<div>X X X</div>	<div>X X X</div>	<div>X X X</div>
	Nível 3	<div>X X X</div>	<div>X X X</div>	<div>X X X</div>	<div>X X X</div>

A soma de quadrados do factor B é calculada ignorando a variabilidade interna do factor B.

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias
Somas de quadrados do factor B (SSFB)

$$SSFB = an \left[\left(\bar{X}_{.1} - \bar{\bar{X}} \right)^2 + \left(\bar{X}_{.2} - \bar{\bar{X}} \right)^2 + \dots + \left(\bar{X}_{.j} - \bar{\bar{X}} \right)^2 \right]$$
$$SSFB = an \sum_{j=1}^b \left(\bar{X}_{.j} - \bar{\bar{X}} \right)^2 = \frac{\sum_{j=1}^b \left(\sum_{i=1}^a \sum_{k=1}^n X_{ijk} \right)^2}{an} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk} \right)^2}{N}$$

GL Factor B = $b-1$

© J.P. Sousa

Slide 209 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias
Somas de quadrados do erro (SSE)

Factor B

Nível 1

Nível 2

Nível 3

Nível

Nível 1

Nível 2

Nível 3

Factor A

4

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

X

X X

A soma de quadrados do erro é igual à variabilidade interna das células.

© J.P. Sousa

Slide 210 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA de duas vias

Somas de quadrados do erro (SSE)

$$SSE = (X_{111} - \bar{X}_{11})^2 + (X_{112} - \bar{X}_{11})^2 + \dots + (X_{ijk} - \bar{X}_{ij})^2$$

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \left[\sum_{k=1}^n (X_{ijk} - \bar{X}_{ij})^2 \right] = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - \frac{\sum_{i=1}^a \sum_{j=1}^b \left(\sum_{k=1}^n X_{ijk} \right)^2}{n}$$

$$GL \text{ Erro} = GL \text{ Total} - GL \text{ Células} = ab \cdot (n-1)$$

© J.P. Sousa

Slide 211 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

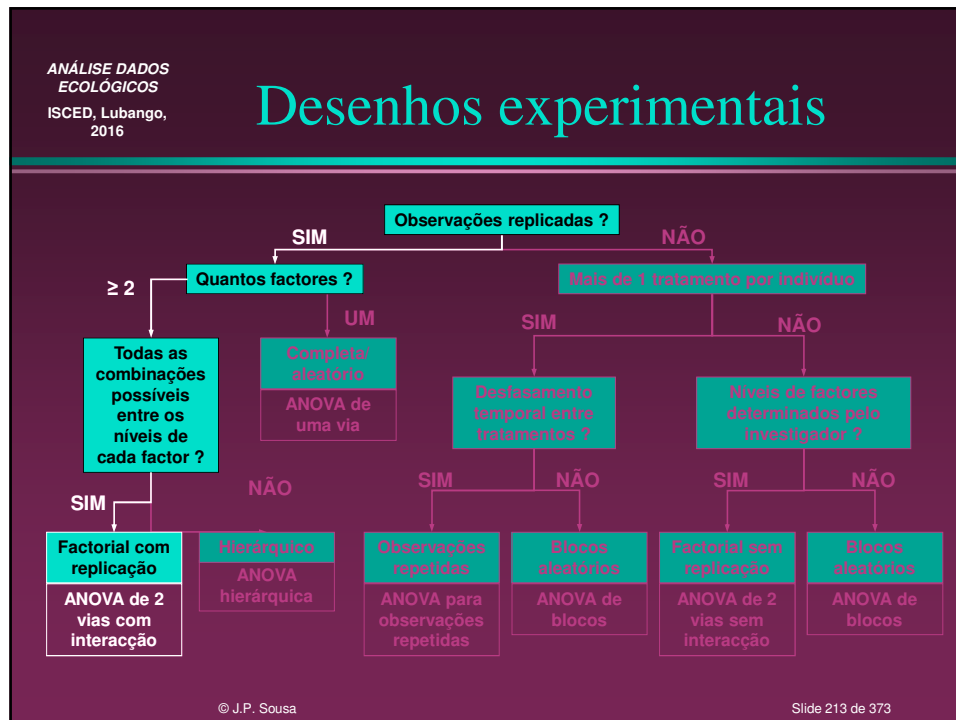
ANOVA de duas vias

Tabela de resumo

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Média de Quadrados	F
Factor A (Linha)	a - 1	SSFA	MSFA	$\frac{MSFA}{MSE}$
Factor B (Coluna)	b - 1	SSFB	MSFB	$\frac{MSFB}{MSE}$
AB (Interacção)	(a-1)(b-1)	SSAB	MSAB	$\frac{MSAB}{MSE}$
Erro	a·b·(n'-1)	SSE	MSE	
Total	a·b·n' - 1	SST	A mesma dos outros desenhos	

© J.P. Sousa

Slide 212 de 373




ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

ANOVA de duas vias

Problema

- Os dados correspondem à concentração de plasma em coelhos de ambos os sexos em que metade recebeu tratamento hormonal e a outra metade não.



© J.P. Sousa

Slide 214 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

ANOVA de duas vias Dados

	Fêmea	Macho		
	16.5	14.5		
	18.4	11		
S/ Hormona	12.7	10.8	$\Sigma x =$	74.4 60.6
	14	14.3	$\bar{x} =$	14.88 12.12
	12.8	10		
	39.1	32		
	26.2	23.8		
C/ Hormona	21.3	28.8	$\Sigma x =$	162.6 138.9
	35.8	25	$\bar{x} =$	32.52 27.78
	40.2	29.3		

S/ Hormona

C/ Hormona

© J.P. Sousa Slide 215 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

ANOVA de duas vias Tabela Resumo

Fonte de variação	SS	gl	MS	F	
Total	1827,70	19			
Factor A	1386,11	1	1386,11	60,53	*
Factor B	70,31	1	70,31	3,07	NS
Interação	4,90	1	4,90	0,21	NS
Erro	366,37	16	22,90		

$F_{0.05(1),1,16} = 4.49$

© J.P. Sousa Slide 216 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais

ANOVA de Blocos ao acaso

© J.P. Sousa

Slide 217 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Agora vamos falar de...



1. Animais bastante selectivos na sua dieta
2. Capazes de escolher o alimento conforme o seu valor nutritivo e o conteúdo de defesas químicas e estruturais
3. Efeito do factor aprendizagem na escolha da dieta

© J.P. Sousa

Slide 218 de 373

Blocos ao acaso

Problema

Pretendemos saber qual a preferência alimentar do corço (*Capreolus capreolus*) testando quatro dietas com diferentes valores nutritivos e de defesas (químicas e estruturais). Para tal utilizaram-se 20 corços de pesos idênticos. A variável de resposta é a quantidade de alimento consumido (Kg/20 dias).

Dieta 1 - Alto valor nutritivo

Dieta 2 - Alto valor nutritivo + ácido tânico (60mg/Kg folha)

Dieta 3 - Alto valor nutritivo + ácido tânico (40mg/Kg folha)

Dieta 4 - Alto valor nutritivo + ácido tânico (20mg/Kg folha)

Blocos ao acaso

TRATAMENTOS

BLOCOS	Dieta 1	Dieta 2	Dieta 3	Dieta 4
1	7.0	5.3	4.9	8.8
2	9.9	5.7	7.6	8.9
3	8.5	4.7	5.5	8.1
4	5.1	3.5	2.8	3.3
5	10.3	7.7	8.4	9.1

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenho em blocos ao acaso

1. As unidades experimentais (Sujeitos) são distribuídas ao acaso pelos tratamentos
2. Utiliza uma **Variável de Bloco** além da variável independente (Tratamento)
 - Permite uma melhor atribuição do tratamento
3. Analisado pelo
 - Teste F para blocos ao acaso

© J.P. Sousa

Slide 221 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso

1. Testa a igualdade de duas ou mais médias da população quando se utiliza uma variável de bloco
2. Análise mais eficiente que a ANOVA de uma via
 - A variação do erro é reduzida

© J.P. Sousa

Slide 222 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso

Pressupostos

1. Normalidade
 - As populações encontram-se normalmente distribuídas
2. Homogeneidade de variâncias
 - As populações têm variâncias iguais
3. Independência dos erros
 - Utilizam-se amostras ao acaso independentes
4. Não existem interações entre os blocos e os tratamentos

© J.P. Sousa

Slide 223 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso

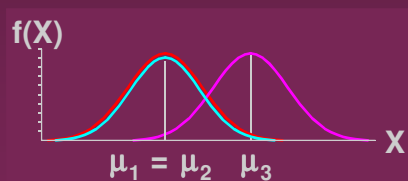
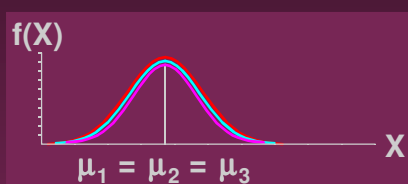
Hipóteses

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a$$

- As médias são todas iguais
- Não há efeito do tratamento

$$H_A: \text{Nem todas as } \mu_i \text{ são iguais}$$

- Pelo menos uma média da População é diferente
- Há efeito do tratamento
- $\mu_1 \neq \mu_2 \neq \dots \neq \mu_a$ é **Incorrecto**



© J.P. Sousa

Slide 224 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso

Ideias básicas

1. SST & SSG são as mesmas do desenho experimental completamente ao acaso
2. A variação do erro (SSE) é diferente
 - O efeito dos blocos (SSB) é retirado da variação dentro dos grupos reduzindo o SSE
 - No desenho experimental de 1 via a variação do erro inclui o efeito de bloco
3. Ao reduzirmos o erro, F pode aumentar

© J.P. Sousa

Slide 225 de 373

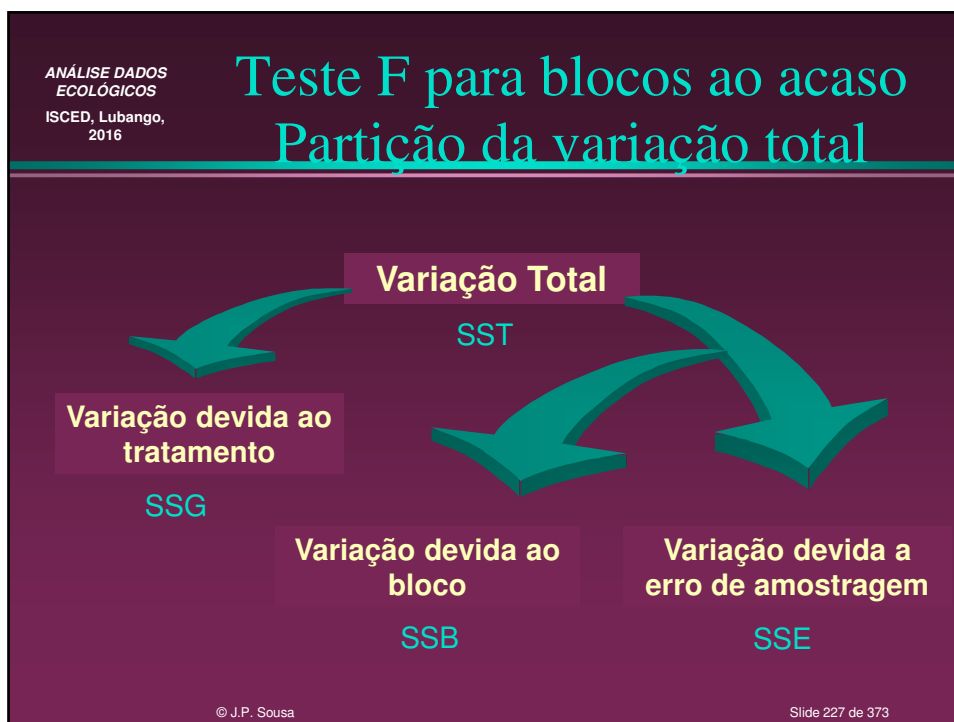
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenho de blocos ao acaso

Blocos	Dieta 1	Dieta 2	Dieta 3	Dieta 4	Totais
1	X_{11}	X_{21}	X_{31}	X_{41}	$B1$
2	X_{12}	X_{22}	X_{32}	X_{42}	$B2$
3	X_{13}	X_{23}	X_{33}	X_{43}	$B3$
4	X_{14}	X_{24}	X_{34}	X_{44}	$B4$
5	X_{15}	X_{25}	X_{35}	X_{45}	$B5$
Totais	$G1$	$G2$	$G3$	$G4$	

© J.P. Sousa

Slide 226 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para blocos ao acaso

Fórmulas

Os mesmos do desenho completamente ao acaso

Total	$SST = \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N}$	GL Total = N-1
Grupos (tratamento)	$G_i = \sum_{j=1}^b X_{ij}$ $SSG = \frac{\sum_{i=1}^a G_i^2}{b} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N}$	GL Grupos = a-1
Blocos	$B_j = \sum_{i=1}^a X_{ij}$ $SSB = \frac{\sum_{j=1}^b B_j^2}{a} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N}$	GL Blocos = b-1
Erro	$SSE = SST - SSG - SSB$	GL Erro = (a-1)*(b-1)

© J.P. Sousa Slide 228 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para blocos ao acaso

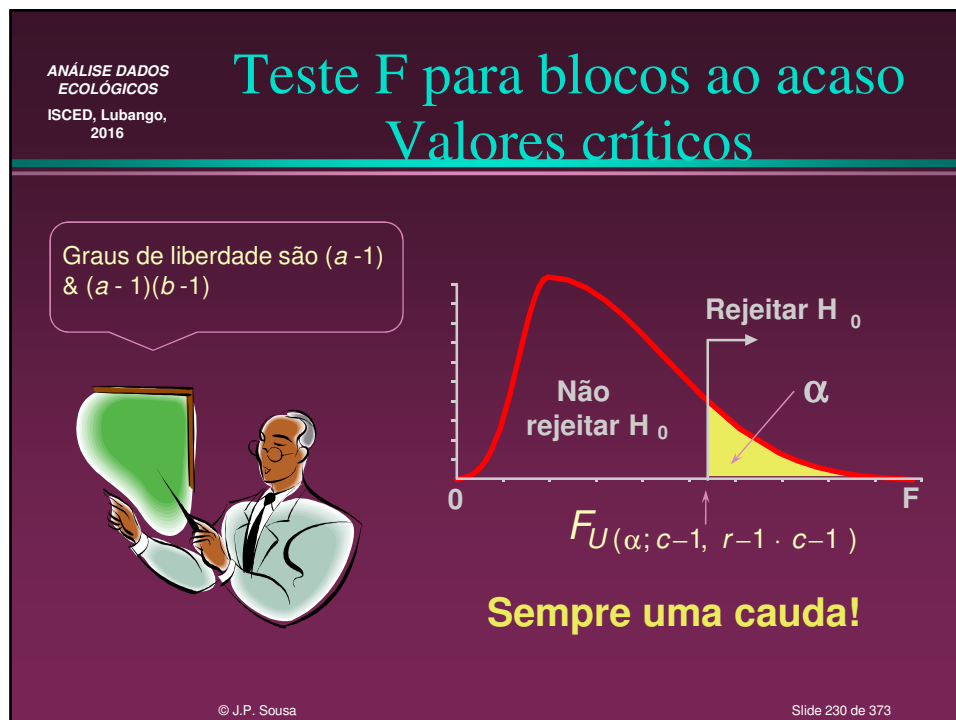
Tabela de resumo

Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrados Médios (Variância)	F
Tratamentos	$a - 1$	SSG	MSG	$\frac{MSG}{MSE}$
Blocos	$b - 1$	SSB	MSB	$\frac{MSB}{MSE}$
Erro (restante)	$(b-1)(a-1)$	SSE	MSE	
Total	$b \cdot a - 1$	SST		

Os mesmos do desenho completamente ao acaso

$b - n^{\circ}$ de blocos $a - n^{\circ}$ de tratamentos

© J.P. Sousa Slide 229 de 373



Blocos ao acaso

Problema

Pretendemos saber qual a preferência alimentar do corço (*Capreolus capreolus*) testando quatro dietas com diferentes valores nutritivos e de defesas (químicas e estruturais). Para tal utilizaram-se 20 corços de pesos idênticos. A variável de resposta é a quantidade de alimento consumido (Kg/20 dias).

Dieta 1 - Alto valor nutritivo

Dieta 2 - Alto valor nutritivo + ácido tânico (60mg/Kg folha)

Dieta 3 - Alto valor nutritivo + ácido tânico (40mg/Kg folha)

Dieta 4 - Alto valor nutritivo + ácido tânico (20mg/Kg folha)

Blocos ao acaso Solução

TRATAMENTOS

BLOCOS

	Dieta 1	Dieta 2	Dieta 3	Dieta 4
1	7.0	5.3	4.9	8.8
2	9.9	5.7	7.6	8.9
3	8.5	4.7	5.5	8.1
4	5.1	3.5	2.8	3.3
5	10.3	7.7	8.4	9.1

Blocos ao acaso
Solução

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

TRATAMENTOS

BLOCOS	Dieta 1	Dieta 2	Dieta 3	Dieta 4	B_j
1	7.0	5.3	4.9	8.8	26.0
2	9.9	5.7	7.6	8.9	31.2
3	8.5	4.7	5.5	8.1	26.8
4	5.1	3.5	2.8	3.3	14.7
5	10.3	7.7	8.4	9.1	35.5
G_i	40.8	26.9	29.2	38.2	135.1

© J.P. Sousa Slide 233 de 373

Blocos ao acaso
Solução

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Total	$SST = \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N} = 1011.95 - \frac{(135.1)^2}{20} = 1011.95 - 912.60 = 99.35$
Grupos (tratamentos)	$SSG = \frac{\sum_{i=1}^a G_i^2}{b} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N} = \frac{(40.8)^2 + \dots + (38.2)^2}{5} - 912.60 = 27.43$
Blocos	$SSB = \frac{\sum_{j=1}^b B_j^2}{a} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N} = \frac{(26.0)^2 + \dots + (35.5)^2}{4} - 912.60 = 62.65$
Erro	$SSE = SST - SSG - SSB = 99.35 - 27.43 - 62.65 = 9.28$

© J.P. Sousa Slide 234 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para blocos ao acaso

Solução

Fonte de variação	Graus de Liberdade	Soma de quadrados	Quadrados médios (Variância)	F
Grupos	3	27,43	9,14	11,8
Blocos	4	62,65	15,66	
Erro (restante)	12	9,28	0,77	
Total	19	99,35		

© J.P. Sousa Slide 235 de 373

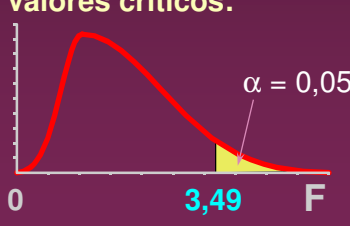
ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para blocos ao acaso

Solução

H₀: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_A: Nem todas são iguais
 $\alpha = 0,05$
 $df_1 = 3$ $df_2 = 12$

Valores críticos:



Teste Estatístico:

$$F = \frac{MSG}{MSE} = \frac{9,1418}{0,7731} = 11,8$$

Decisão:
Rejeitar para $\alpha = 0,05$

Conclusão:
Há evidências de que o consumo foi diferente

© J.P. Sousa Slide 236 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso Solução

Comparações múltiplas - Teste de Tukey

Erro padrão

$$SE = \sqrt{\frac{0.77}{5}} = 0.392$$

Tukey

$$q = \frac{\bar{X}_B - \bar{X}_A}{SE}$$

Ordem das médias	Dieta 2	Dieta 3	Dieta 4	Dieta 1
Média dos grupos	26.9	29.2	38.2	40.8

Hipóteses

➤ $H_0: \mu_B = \mu_A$

➤ $H_A: \mu_B \neq \mu_A$

© J.P. Sousa

Slide 237 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso Solução

Comparações múltiplas - Teste de Tukey

Comp	Difs	SE	q	$q_{0.05, 12, 4}$	
1 vs 2	13.9	0.392	35.45	4.199	Rejeitar $H_0: \mu_1 = \mu_2$
1 vs 3	11.6	0.392	29.59	4.199	Rejeitar $H_0: \mu_1 = \mu_3$
1 vs 4	2.6	0.392	6.63	4.199	Rejeitar $H_0: \mu_1 = \mu_4$
4 vs 2	11.3	0.392	28.82	4.199	Rejeitar $H_0: \mu_4 = \mu_2$
3 vs 3	9.0	0.392	22.96	4.199	Rejeitar $H_0: \mu_4 = \mu_3$
3 vs 2	2.3	0.392	5.86	4.199	Rejeitar $H_0: \mu_3 = \mu_2$

Ordem das médias	Dieta 2	Dieta 3	Dieta 4	Dieta 1
Média dos grupos	26.9	29.2	38.2	40.8

© J.P. Sousa

Slide 238 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA Blocos vs. ANOVA 1 via

ANOVA Blocos

Fonte	SS	GL	MS	F	F _{0.05(1),3,12}
Total	99.35	19			
Grupos	27.43	3	9.14	11.82	3.49
Blocos	62.65	4			
Erro (rest.)	9.28	12	0.77		

F > F_{0.05(1),3,12}

Rejeitar H₀

F ≤ F_{0.05(1),3,12}

Aceitar H₀

ANOVA 1 via

Fonte	SS	GL	MS	F	F _{0.05(1),3,16}
Total	99.35	19			
Grupos	27.43	3	9.14	2.034	3.24
Erro (rest.)	71.93	16	4.50		

© J.P. Sousa

Slide 239 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais

ANOVA

Observações repetidas

© J.P. Sousa

Slide 240 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Observações repetidas

TRATAMENTOS

SUJEITOS	Dieta 1	Dieta 2	Dieta 3	Dieta 4
Sujeito 1	7.0	5.3	4.9	8.8
Sujeito 2	9.9	5.7	7.6	8.9
Sujeito 3	8.5	4.7	5.5	8.1
Sujeito 4	5.1	3.5	2.8	3.3
Sujeito 5	10.3	7.7	8.4	9.1

© J.P. Sousa

Slide 241 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para observações repetidas

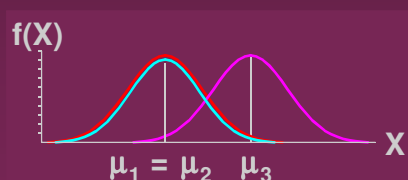
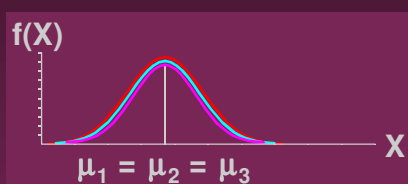
Hipóteses

$$H_0: \mu_1 = \mu_2 = \dots = \mu_c$$

- As médias são todas iguais
- Não há efeito do tratamento

$$H_1: \text{Nem todas as } \mu_j \text{ são iguais}$$

- Pelo menos uma média da População é diferente
- Há efeito do tratamento
- $\mu_1 \neq \mu_2 \neq \dots \neq \mu_c$ é **Incorrecto**



© J.P. Sousa

Slide 242 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para observações repetidas

Ideias básicas

1. SST & SSG são as mesmas do desenho experimental completamente ao acaso
2. A variação do erro (SSE) é diferente
 - O efeito dos grupos (SSG) é retirado da variação dentro dos sujeitos reduzindo o SSE
 - Neste desenho experimental a variação do erro (dentro dos grupos) inclui o efeito tratamento
4. Ao reduzirmos o erro, F pode aumentar

© J.P. Sousa

Slide 243 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Observações repetidas

Sujeitos	Dieta 1	Dieta 2	Dieta 3	Dieta 4	Totais
1	X_{11}	X_{21}	X_{31}	X_{41}	$S1$
2	X_{12}	X_{22}	X_{32}	X_{42}	$S2$
3	X_{13}	X_{23}	X_{33}	X_{43}	$S3$
4	X_{14}	X_{24}	X_{34}	X_{44}	$S4$
5	X_{15}	X_{25}	X_{35}	X_{45}	$S5$
Totais	$G1$	$G2$	$G3$	$G4$	

© J.P. Sousa

Slide 244 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para observações repetidas Fórmulas

Total	$SST = \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N}$	GL Total = N-1
Sujeitos	$S_j = \sum_{i=1}^a X_{ij}$ $SSS = \frac{\sum_{j=1}^b S_j^2}{a} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N}$	GL Sujeitos = b-1
Sujeitos (dentro)	$SSD = SST - SSS$	GL Dentro = a*(b-1)
Grupos (tratamento)	$G_i = \sum_{j=1}^b X_{ij}$ $SSG = \frac{\sum_{i=1}^a G_i^2}{b} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N}$	GL Grupos = a-1
Erro	$SSE = SST - SSG - SSS$	GL Erro = (a-1)*(b-1)

© J.P. Sousa Slide 246 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para observações repetidas

Tabela de resumo

Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrados Médios (Variância)	F
Sujeitos	$b - 1$	SSS	MSS	
Dentro dos Sujeitos	$b(a - 1)$	SSD	MSD	
Grupos	$a - 1$	SSG	MSG	$\frac{MSG}{MSE}$
Erro (restante)	$(b-1)(a-1)$	SSE	MSE	
Total	$b \cdot a - 1$	SST		

Os mesmos do desenho completamente ao acaso

$b - n^{\circ}$ de sujeitos $a - n^{\circ}$ de tratamentos

© J.P. Sousa Slide 247 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Observações repetidas

Solução

		TRATAMENTOS			
SUJEITOS		Dieta 1	Dieta 2	Dieta 3	Dieta 4
1		7.0	5.3	4.9	8.8
2		9.9	5.7	7.6	8.9
3		8.5	4.7	5.5	8.1
4		5.1	3.5	2.8	3.3
5		10.3	7.7	8.4	9.1
G_i		40.8	26.9	29.2	38.2
	S_j	26.0	31.2	26.8	14.7
		35.5	135.1		

© J.P. Sousa Slide 248 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Observações repetidas Solução

Total	$SST = \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N} = 1011.95 - \frac{(135.1)^2}{20} = 1011.95 - 912.60 = 99.35$
Sujeitos	$SSS = \frac{\sum_{i=1}^b S_i^2}{a} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N} = \frac{(26.0)^2 + \dots + (35.5)^2}{4} - 912.60 = 62.65$
Sujeitos (dentro)	$SSD = SST - SSS = 99.35 - 62.65 = 36.7$
Grupos (tratamento)	$SSG = \frac{\sum_{i=1}^a G_i^2}{b} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b X_{ij} \right)^2}{N} = \frac{(40.8)^2 + \dots + (38.2)^2}{5} - 912.60 = 27.43$
Erro	$SSE = SSD - SSG = 36.7 - 27.43 = 9.27$

© J.P. Sousa Slide 249 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste F para observações repetidas Tabela de resumo

Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrados Médios (Variância)	F
Sujeitos	4	62,65		
Dentro dos Sujeitos	15	36,7		
Dietas	3	27,43	9,14	$\frac{9,14}{0,77} = 11,8$
Erro (restante)	12	9,27	0,77	
Total	19	99,35		

© J.P. Sousa Slide 250 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para blocos ao acaso Solução

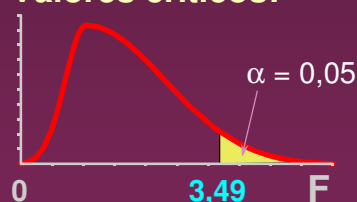
H₀: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_A: Nem todas são iguais

α = 0,05

df₁ = 3 **df₂** = 12

Valores críticos:



Teste Estatístico:

$$F = \frac{MSG}{MSE} = \frac{9,1418}{0,7731} = 11,8$$

Decisão:

Rejeitar para $\alpha = 0,05$

Conclusão:

Há evidências de que o consumo foi diferente

© J.P. Sousa

Slide 251 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenho para observações repetidas Observações

1. Os cálculos são semelhantes aos dos “Blocos ao acaso”. No entanto o modelo de partição da variabilidade é diferente
2. Tal como o desenho em “Blocos ao acaso” possui vantagens em relação à ANOVA de 1 via quando existe uma forte relação de consequência entre os valores de cada “linha”.
3. Possui vantagens em relação ao desenho de “Blocos ao acaso” quando existir uma forte relação dentro dos sujeitos que dentro dos blocos
4. Possui desvantagens se houver efeitos relativos à sequência dos tratamentos e/ou se o tempo entre a administração dos tratamentos for insuficiente.

© J.P. Sousa

Slide 252 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Desenhos experimentais

ANOVA Hierárquica
“Nested” ANOVA

© J.P. Sousa

Slide 253 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Exemplo de desenho experimental

Nível 1

Local A Local G

X

X X

X

X X

Nível 2

Local F Local B

X

X X

X

X X

Nível 3

Local E Local H

X

X X

X

X X

Factor A = Nível de pesticida
Factor B = Local

X

i

j

k

Observação k

Nível i
Factor A

Nível j
Factor B

© J.P. Sousa

Slide 254 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para desenhos hierárquicos (“nested”)

1. Testa a igualdade de duas ou mais médias da população quando se utiliza desenho experimental hierárquico (não cruzado)
 - não existem todas as combinações de níveis possíveis
2. Utilizados por vezes para testar hipóteses acerca das amostras
3. A introdução de factores hierárquicos aleatórios para evidenciar a variabilidade dentro dos grupos
 - factor principal de interesse geralmente fixo

© J.P. Sousa

Slide 255 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

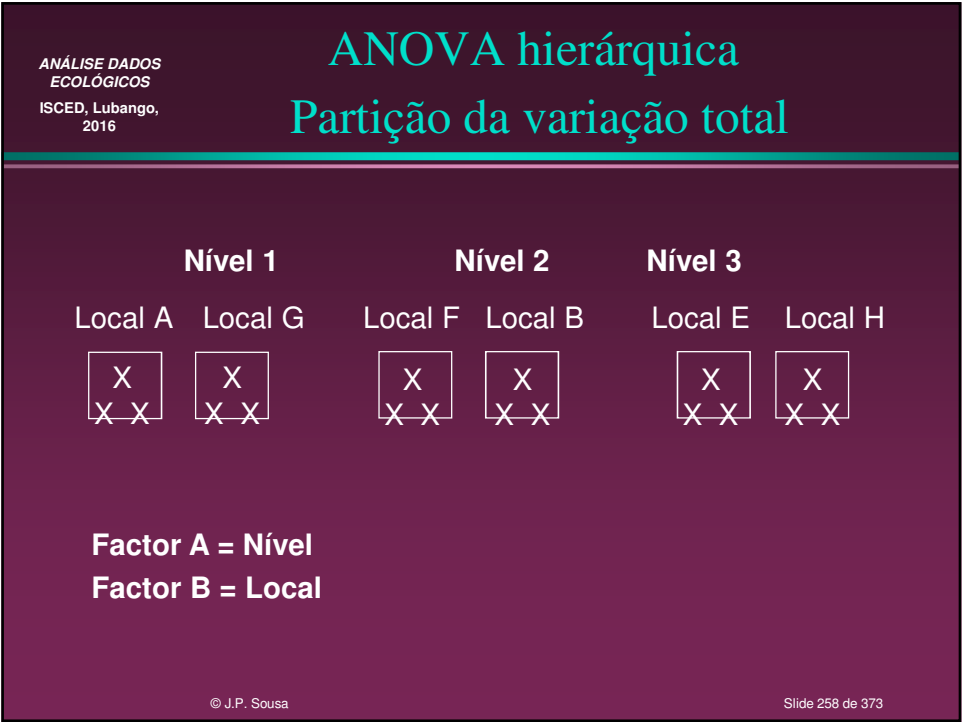
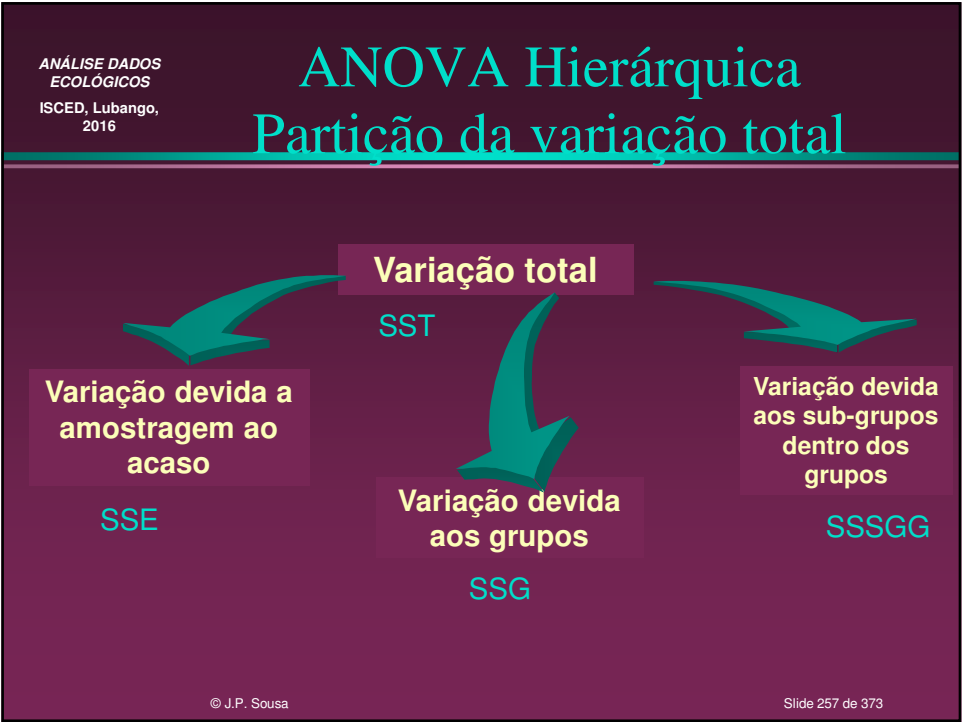
Teste F para ANOVA hierárquica

Pressupostos

1. Normalidade
 - As populações encontram-se normalmente distribuídas
2. Homogeneidade de variâncias
 - As populações têm variâncias iguais
3. Independência dos erros
 - Utilizam-se amostras ao acaso independentes

© J.P. Sousa

Slide 256 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA hierárquica

Ideias básicas

1. SST & SSG são as mesmas do desenho experimental completamente ao acaso
2. A variação entre todos os sub-grupos ($SSSG$) tem dois componentes
 - variação devida aos factores (SSG)
 - variação entre sub-grupos dentro de cada factor

© J.P. Sousa

Slide 259 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica

Somas de quadrados total (SST)

Nível 1		Nível 2		Nível 3	
Local A	Local G	Local F	Local B	Local E	Local H
X	X	X	X	X	X
X X	X X	X X	X X	X X	X X

Factor A = Nível
Factor B = Local

$$[X_{ijk} - \bar{X}]^2$$

© J.P. Sousa

Slide 260 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Somas de quadrados total (SST)

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk}^2 - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk} \right)^2}{N}$$

GL Total = N-1

© J.P. Sousa

Slide 261 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Somas de quadrados entre grupos (SSG)

Nível 1

Local A Local G

X X

X X X X

Nível 2

Local F Local B

X X

X X X X

Nível 3

Local E Local H

X X

X X X X

Factor A = Nível
Factor B = Local

$$[\bar{X}_i - \bar{X}]^2$$

© J.P. Sousa

Slide 262 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Somos de quadrados entre grupos (SSG)

$$SSG = \frac{\sum_{i=1}^a \left(\sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk} \right)^2}{n_i} - \frac{\left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk} \right)^2}{N}$$

GL Grupos = $a-1$

© J.P. Sousa

Slide 263 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Somos de quadrados entre sub-grupos dentro
de cada grupo (SSSG)

Nível 1

Local A Local G

X

X X

X

X X

Nível 2

Local F Local B

X

X X

X

X X

Nível 3

Local E Local H

X

X X

X

X X

Factor A = Nível

Factor B = Local

$$[\bar{X}_{ij} - \bar{X}_i]^2$$

© J.P. Sousa

Slide 264 de 373

Notas para os alunos

© Paulo Sousa

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Somos de quadrados entre sub-grupos dentro
de cada grupo (SSSGG)

$$SSSGG = \sum_{i=1}^a \sum_{j=1}^b \frac{\left(\sum_{k=1}^{n_{ij}} X_{ijk} \right)^2}{n_{ij}} - \frac{\sum_{i=1}^a \left(\sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk} \right)^2}{n_i}$$

GL SGG = a(b-1)

© J.P. Sousa

Slide 265 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica
Somos de quadrados do erro (SSE)

Nível 1

Local A Local G

X

X X

X

X X

Nível 2

Local F Local B

X

X X

X

X X

Nível 3

Local E Local H

X

X X

X

X X

Factor A = Nível
Factor B = Local

$$[X_{ijk} - \bar{X}_{ij}]^2$$

© J.P. Sousa

Slide 266 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

ANOVA hierárquica Somadas de quadrados do erro (SSE)

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk}^2 - \sum_{i=1}^a \sum_{j=1}^b \frac{\left(\sum_{k=1}^{n_{ij}} X_{ijk} \right)^2}{n_{ij}}$$

$$SSE = SST - SSG - SSSGG$$

$$GL \text{ Erro} = N - ab$$

© J.P. Sousa

Slide 267 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA hierárquica Tabela de resumo

Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrados Médios (Variância)	F
Erro	N - a.b	SSE	MSE	
Grupos	a - 1	SSG	MSG	$\frac{MSG}{MSSGG}$
Sub-grupos	a.(b-1)	SSSGG	MSSGG	$\frac{MSSGG}{MSE}$
Total	N - 1	SST		

a - nº de grupos

b - nº de subgrupos em cada grupo

© J.P. Sousa

Slide 268 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA hierárquica Problema

A concentração de colesterol no sangue foi medida em mulheres (mg / 100 ml de plasma). Esta variável foi determinada após administração de três medicamentos (drogas) diferentes provenientes cada uma delas de dois locais diferentes.



Droga 1		Droga 2		Droga 3	
Local A	Local Q	Local D	Local B	Local L	Local S
102	103	108	109	104	105
104	104	110	108	106	107

© J.P. Sousa

Slide 269 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA hierárquica Resolução

Hipóteses

- H_0 : Não existem diferenças no efeito de drogas provenientes de locais diferentes
- H_A : Existe diferença de efeito em drogas provenientes de diferentes origens

- H_0 : Não existem diferenças nas concentrações de colesterol devidas às três drogas
- H_A : Existem diferenças nas concentrações de colesterol devidas às três drogas

© J.P. Sousa

Slide 270 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA hierárquica Tabela de Variância

Fonte de variação	SS	gl	MS
Total	71,67	11	
Grupos	61,17	2	30,58
Sub-grupos	1,50	3	0,50
Erro	9,00	6	1,50

© J.P. Sousa

Slide 271 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste F para ANOVA hierárquica Resolução

Hipóteses

- H_0 : Não existem diferenças no efeito de drogas provenientes de locais diferentes
- H_A : Existem diferenças de efeito em drogas provenientes de diferentes origens

$$F = \frac{0,50}{1,50} = 0,33 \quad F_{0,05(1),3,6} = 4,76 \quad (\text{Não rejeitar } H_0)$$


- H_0 : Não existem diferenças nas concentrações de colesterol devidas às três drogas
- H_A : Existem diferenças nas concentrações de colesterol devidas às três drogas

$$F = \frac{30,58}{0,50} = 61,16 \quad F_{0,05(1),2,3} = 9,55 \quad (\text{Rejeitar } H_0)$$

© J.P. Sousa

Slide 272 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016



MÃOS À OBRA 2

© J.P. Sousa Slide 273 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Regressão Linear Simples

© J.P. Sousa Slide 274 de 373

Modelos

1. Representação de alguns fenómenos.

Por exemplo:

- Conhecer a **relação entre duas** ou mais **variáveis** (ex. relação entre o consumo de O_2 e o peso dos indivíduos)
- Conhecer a variação **ao longo do tempo** da taxa de decomposição da matéria orgânica;
- Saber que características de um determinado alimento condicionam o seu consumo

Modelos

2. Um modelo matemático é a expressão matemática de um fenómeno

3. Tipos

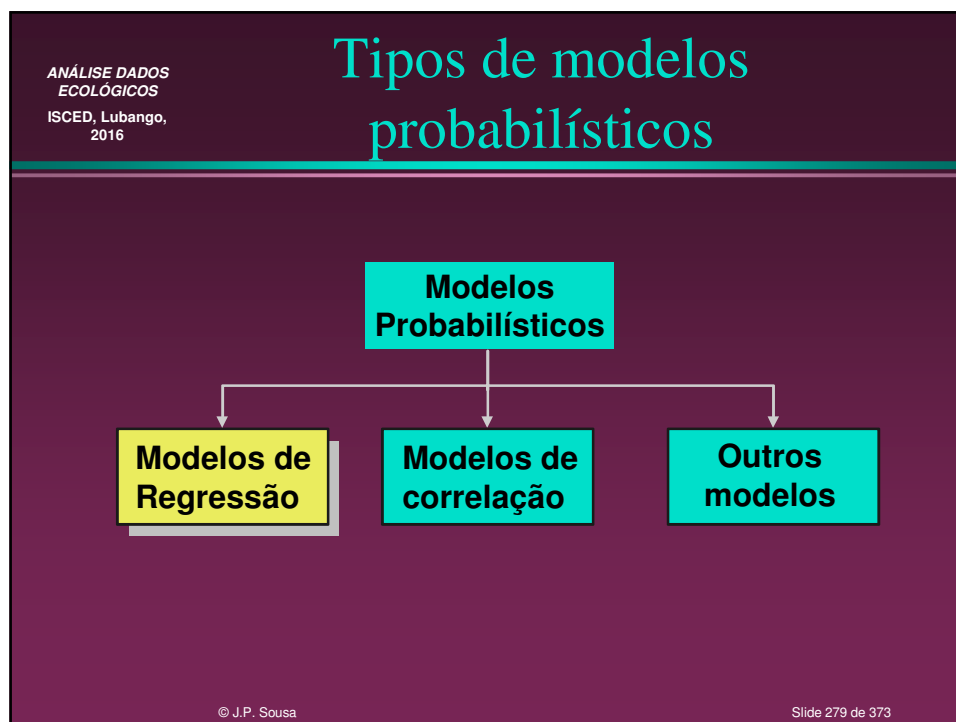
- Modelos Determinísticos
- Modelos Probabilísticos

Modelos Determinísticos

1. Prevê relações hipotéticas exactas
2. Adequado quando o erro das previsões é negligenciável
3. Ex: Força é exactamente igual à massa vezes a aceleração
 - $F = m \cdot a$

Modelos Probabilísticos

1. Prevê dois componentes
 - Determinístico
 - Erro aleatório
2. A altura (Y) é o dobro do comprimento dos membros posteriores (X) mais o erro aleatório
 - $Y = 2 X + \epsilon$
 - O erro aleatório pode ser devido a outros factores que não o comprimento dos membros posteriores



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Regressão vs. Correlação

1. A **Correlação** mostra apenas a relação entre duas variáveis (coeficiente de correlação 'r')
2. A **Regressão** permite ainda prever o valor de uma variável em função da(s) outra(s)

Dependência funcional entre as variáveis

© J.P. Sousa Slide 280 de 373

Modelos de Regressão

Equação utilizada

- 1 variável numérica dependente (**Resposta**)
 - O que se pretende prever (Y)
 - Ex: consumo de O₂
- 1 ou mais variáveis independentes (**Explicativas**)
 - O que prevê (X)
 - Ex: Peso do animal (Kg)

Modelos de regressão Passos de cálculo

1. Definir o problema ou questão
2. Especificar o modelo
3. Recolher os dados
4. Fazer uma análise descritiva dos dados
5. Estimar parâmetros desconhecidos
6. Avaliar o modelo
7. Utilizar o modelo para fazer previsões

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Modelos de regressão Passos de cálculo

1. Definir o problema ou questão
2. Especificar o modelo
3. Recolher os dados
4. Fazer uma análise descritiva dos dados
5. Estimar parâmetros desconhecidos
6. Avaliar o modelo
7. Utilizar o modelo para fazer previsões

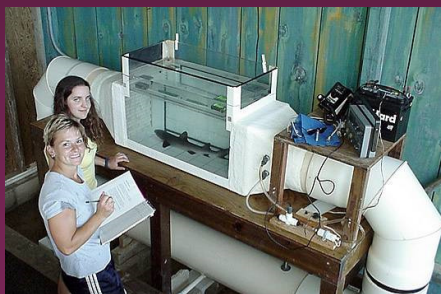
© J.P. Sousa

Slide 283 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Definição do problema

Conhecer e prever o consumo
de O_2 em várias espécies de
peixes



© J.P. Sousa

Slide 284 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Modelos de regressão Passos de cálculo

1. Definir o problema ou questão
- 2. Especificar o modelo**
3. Recolher os dados
4. Fazer uma análise descritiva dos dados
5. Estimar parâmetros desconhecidos
6. Avaliar o modelo
7. Utilizar o modelo para fazer previsões

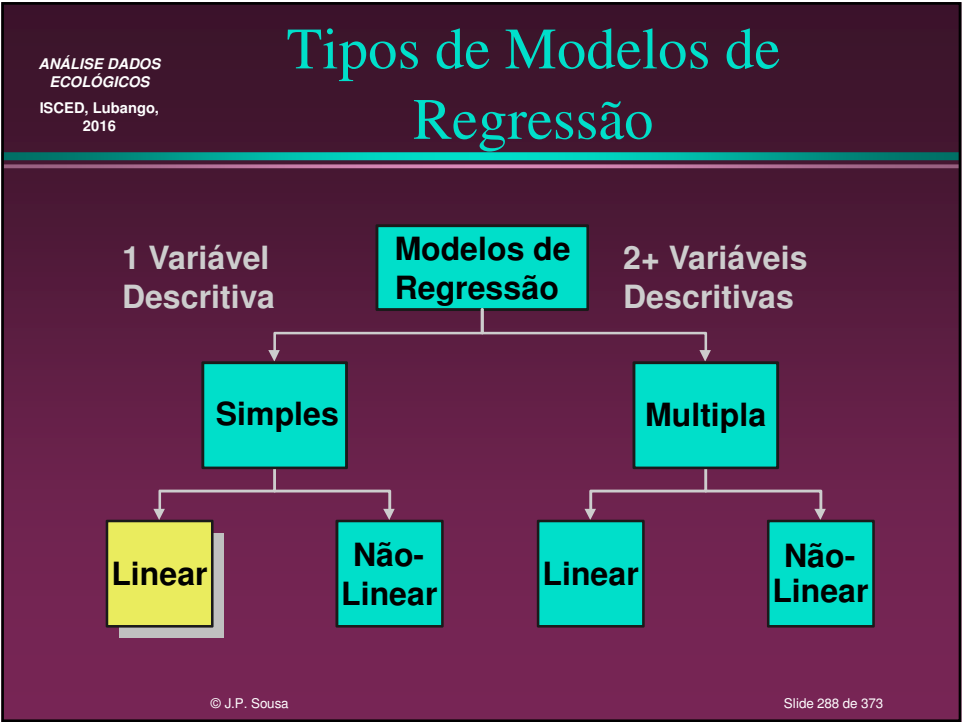
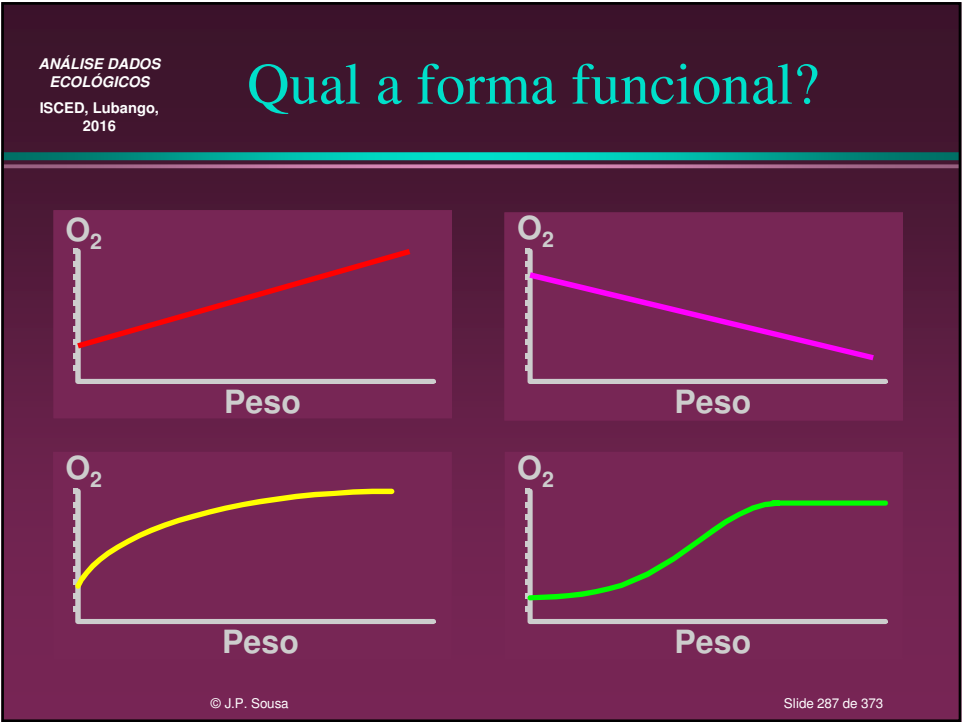
© J.P. Sousa Slide 285 de 373

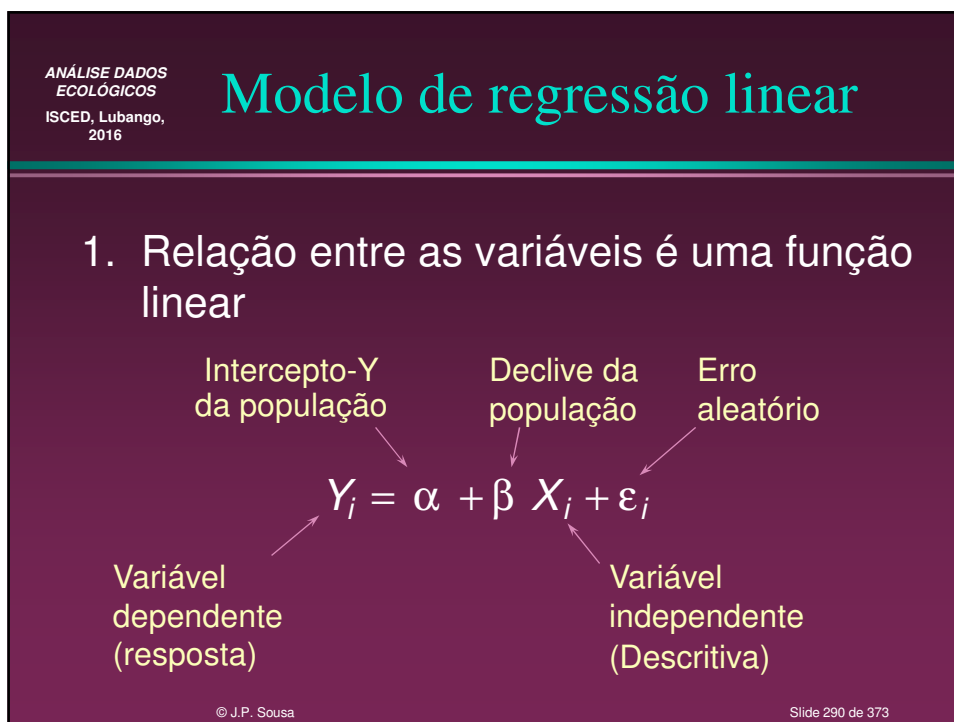
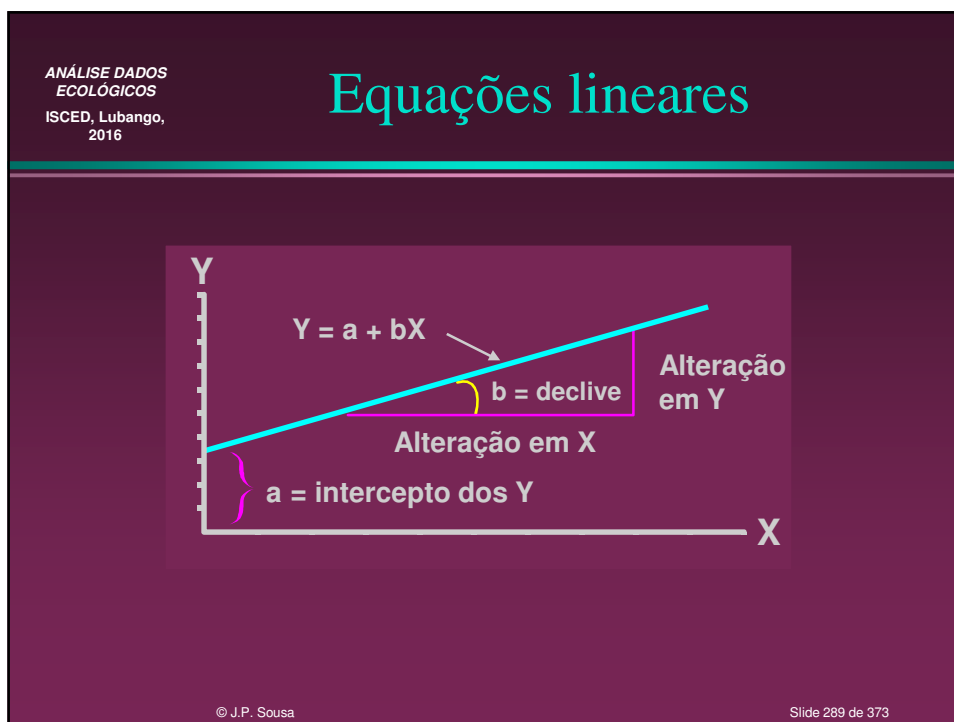
ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Especificar o modelo

1. Definir as variáveis
 - Consumo de O_2 (ml/hora) - Y
 - Peso do animal (Kg) - X
2. Teorizar a natureza da relação
 - Efeitos esperados (i.e., sinais dos coeficientes)
 - Forma funcional (Linear ou Não-Linear)
 - Interações

© J.P. Sousa Slide 286 de 373





ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Modelos de regressão Passos de cálculo

1. Definir o problema ou questão
2. Especificar o modelo
- 3. Recolher os dados**
4. Fazer uma análise descritiva dos dados
5. Estimar parâmetros desconhecidos
6. Avaliar o modelo
7. Utilizar o modelo para fazer previsões

© J.P. Sousa Slide 291 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

População & Amostra Modelos de Regressão

População

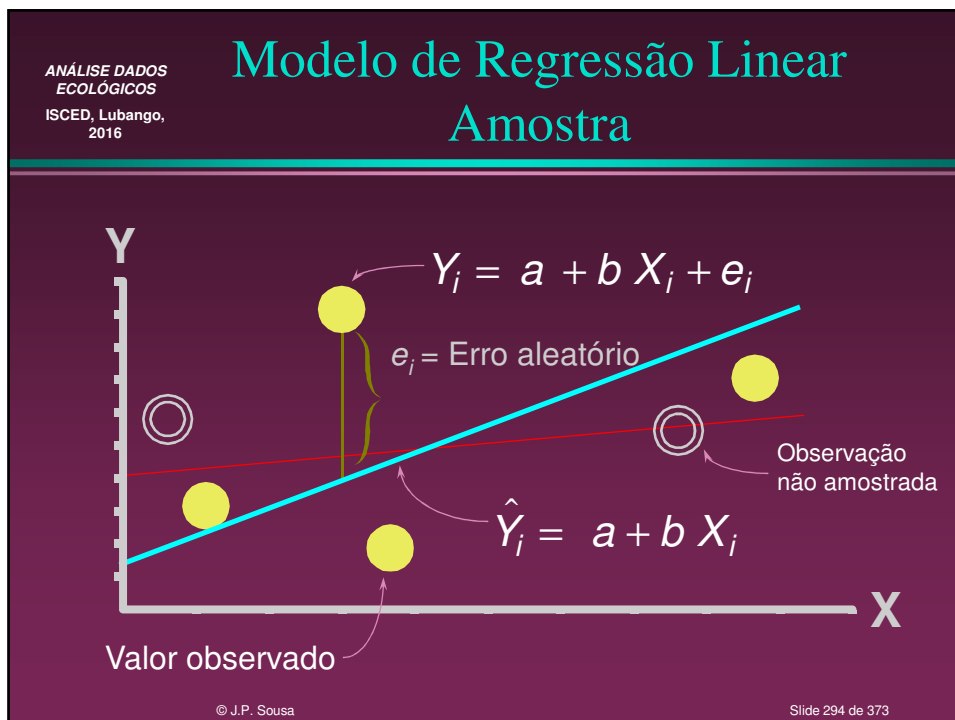
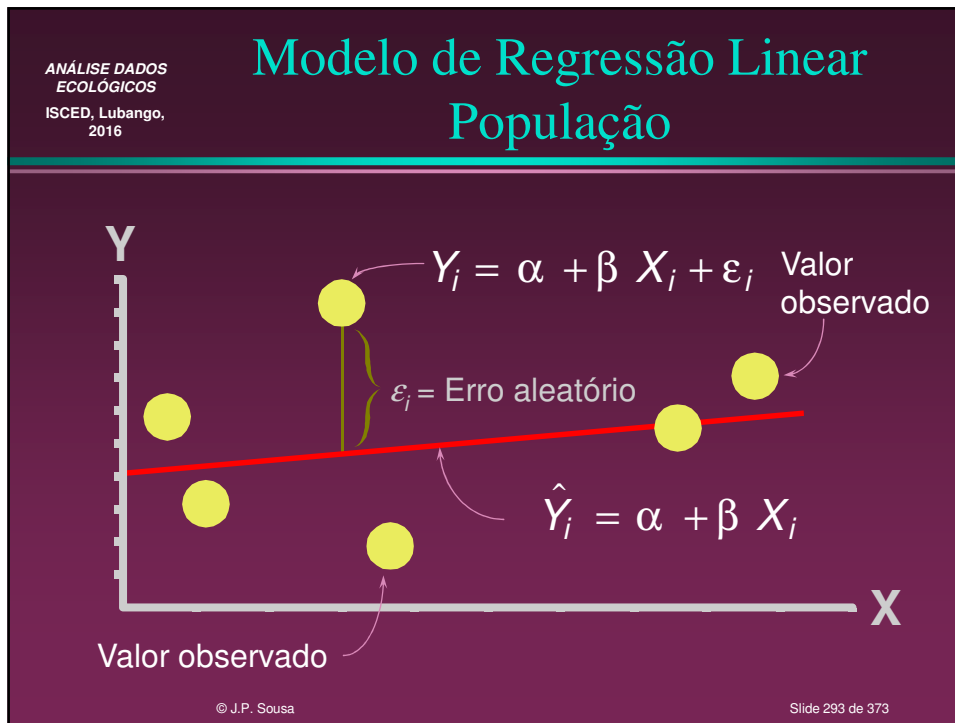
Relação Desconhecida

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Variável aleatória

$$Y_i = a + b X_i + e_i$$

© J.P. Sousa Slide 292 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Modelos de regressão

Passos de cálculo

1. Definir o problema ou questão
2. Especificar o modelo
3. Recolher os dados
- 4. Fazer uma análise descritiva dos dados**
5. Estimar parâmetros desconhecidos
6. Avaliar o modelo
7. Utilizar o modelo para fazer previsões

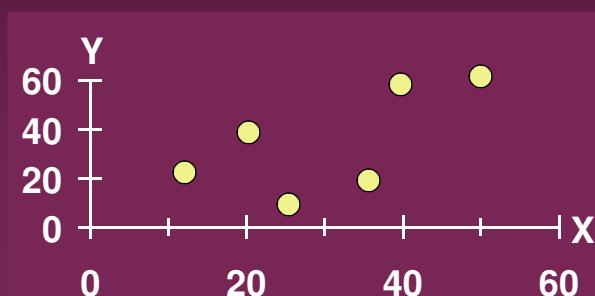
© J.P. Sousa

Slide 295 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Diagrama de dispersão

1. Gráfico de todos os pares (X_i, Y_i)
2. Sugere até que ponto o modelo se ajustará



© J.P. Sousa

Slide 296 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Modelos de regressão Passos de cálculo

1. Definir o problema ou questão
2. Especificar o modelo
3. Recolher os dados
4. Fazer uma análise descritiva dos dados
- 5. Estimar parâmetros desconhecidos**
6. Avaliar o modelo
7. Utilizar o modelo para fazer previsões

© J.P. Sousa Slide 297 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Problema

Como desenharíamos uma linha através dos pontos?
Como determinaríamos qual a recta que corresponde ao
'melhor ajustamento'?

X	Y
10	25
20	40
25	10
35	20
40	60
45	65
50	60

© J.P. Sousa Slide 298 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Método dos quadrados mínimos (LS) “ Ordinary Least Squares”

1. ‘Melhor ajustamento’ significa ajustamento entre os valores reais (Y_i) e os valores previstos (\hat{Y}_i).

É um mínimo

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$$

2. O método minimiza a soma dos quadrados das diferenças (ou erros)

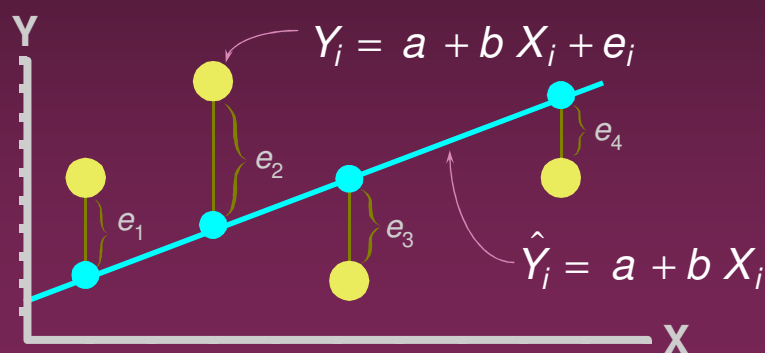
© J.P. Sousa

Slide 299 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Método dos quadrados mínimos (LS) Graficamente

LS Minimiza $\sum_{i=1}^n e_i^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2$



© J.P. Sousa

Slide 300 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Coeficientes das equações

Equação de regressão da amostra

$$\hat{Y}_i = a + b X_i$$

declive da amostra

$$b = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n (\bar{X})^2}$$

Intercepto-Y da amostra

$$a = \bar{Y} - b \bar{X}$$

pares # (X_i, Y_i)

média do X_i ao quadrado

© J.P. Sousa Slide 301 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Tabela de cálculo

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
X_1	Y_1	X_1^2	Y_1^2	$X_1 Y_1$
X_2	Y_2	X_2^2	Y_2^2	$X_2 Y_2$
\vdots	\vdots	\vdots	\vdots	\vdots
X_n	Y_n	X_n^2	Y_n^2	$X_n Y_n$
ΣX_i	ΣY_i	ΣX_i^2	ΣY_i^2	$\Sigma X_i Y_i$

© J.P. Sousa Slide 302 de 373

Interpretação dos coeficientes

1. Declive (b)

- Alteração estimada para Y a partir de ' a ' para cada incremento de 1 unidade em X
 - Se $b = 2$, então (Y) aumentará duas unidades por cada unidade de aumento em (X)

2. Intercepto- Y (a)

- Valor médio de Y quando $X = 0$
 - Se $a = 4$, então espera-se que (Y) seja 4 quando (X) for 0

Pressupostos

1. Normalidade

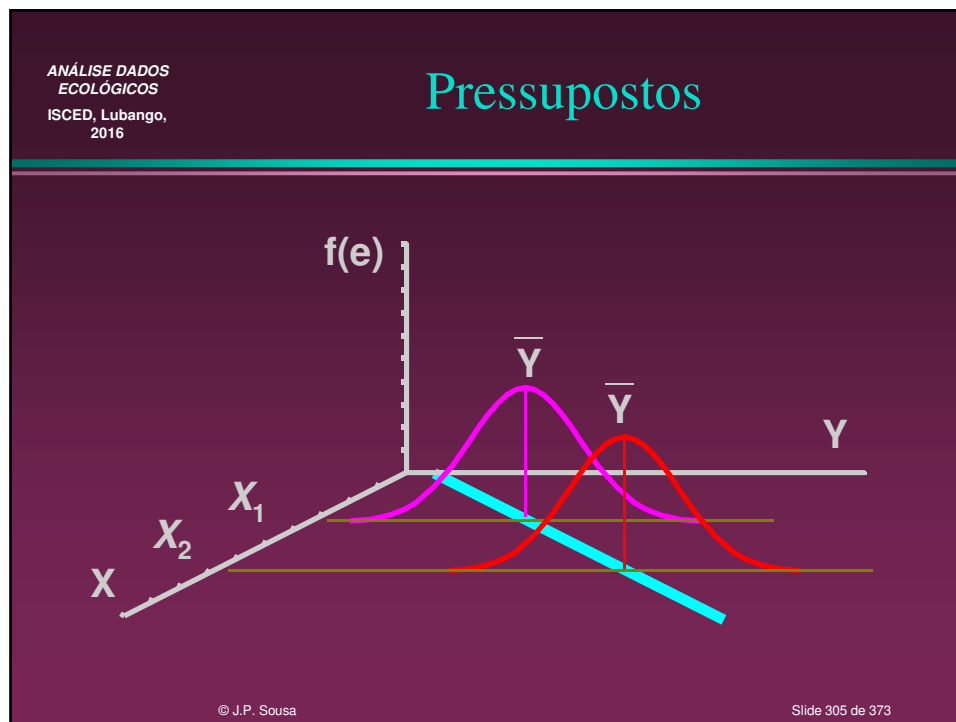
- Para cada valor de X_i existe, na população, uma série de valores de Y_i com distribuição normal

2. Homoscedastia

- As variâncias destes valores de Y_i devem ser iguais

3. Relação linear na população

- As médias de Y_i , para cada X_i , ficam localizadas na recta



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Pressupostos

5. Aleatoriedade
 - Os valores de Y_i provêm de uma amostragem aleatória e são independentes
6. Valores de X obtidos sem erro
 - Ou pelo menos que o seu erro seja negligenciável, quando comparado com o erro na obtenção dos valores de Y_i .

© J.P. Sousa


Slide 306 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Cálculo de parâmetros Exemplo

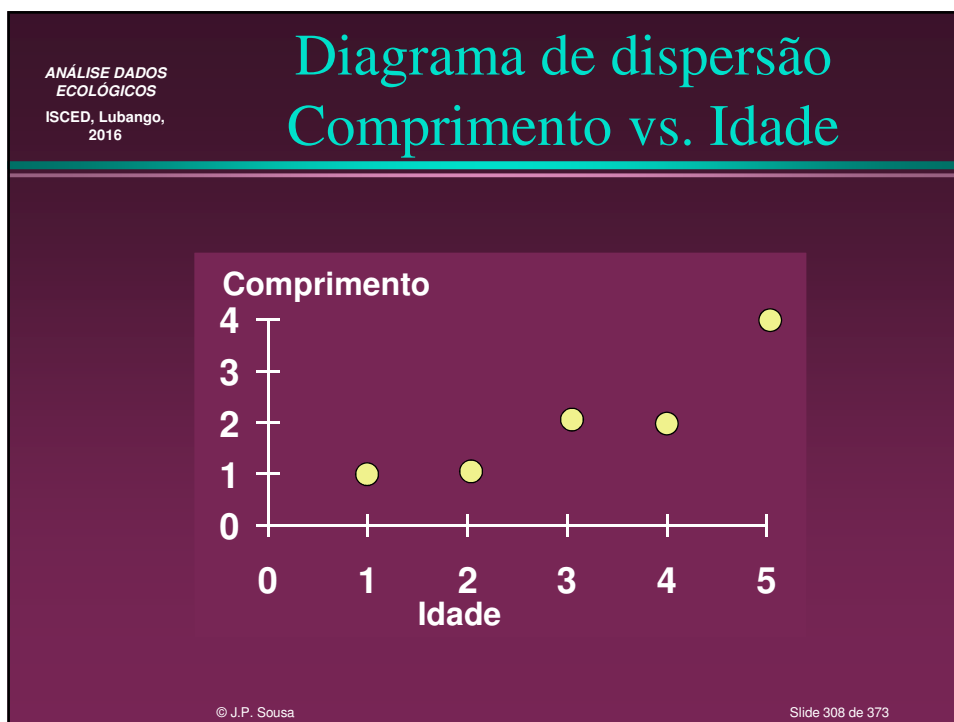
Estudou-se o crescimento de uma espécie de planária ao longo de 5 dias:

<u>Idade (d)</u>	<u>Comp. (mm)</u>
1	1
2	1
3	2
4	2
5	4



Qual a **relação** entre crescimento e a idade?

© J.P. Sousa Slide 307 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Cálculo de parâmetros Tabela de soluções

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

© J.P. Sousa Slide 309 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Cálculo de parâmetros Solução

$$b = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n (\bar{X})^2} = \frac{37 - (5)(3)(2)}{55 - (5 * 9)} = 0,70$$

$$a = \bar{Y} - b \bar{X} = 2 - (0,70)(3) = -0,10$$

$$\hat{Y}_i = -0,10 + 0,70 X_i$$

© J.P. Sousa Slide 310 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Cálculo de parâmetros Resultado do computador

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Param=0	Prob> T
INTERCEP	1	-0.1000	0.6350	-0.157	0.8849
SLOPE	1	0.7000	0.1914	3.656	0.0354

a points to the SLOPE row, and *b* points to the INTERCEP row.


© J.P. Sousa Slide 311 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Cálculo de Parâmetros Problema

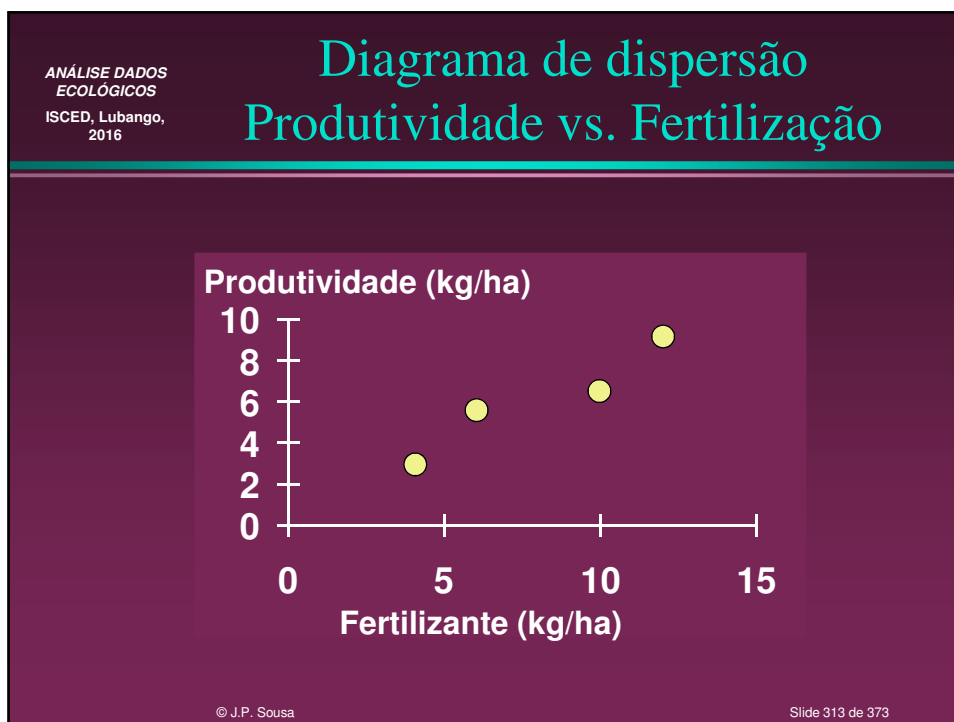
Uma cooperativa pretendeu estudar a eficácia do processo de adubação na produtividade das culturas tendo reunido os seguintes dados (Kg/ha):

<u>Fertilizante</u>	<u>Produção</u>
4	3,0
6	5,5
10	6,5
12	9,0



Qual a **relação** entre a fertilização e a colheita?

© J.P. Sousa Slide 312 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Cálculo dos parâmetros Tabela de solução*

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
4	3,0	16	9,00	12
6	5,5	36	30,25	33
10	6,5	100	42,25	65
12	9,0	144	81,00	108
32	24,0	296	162,50	218

© J.P. Sousa

Slide 314 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Cálculo dos parâmetros Solução*

$$b = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n (\bar{X})^2} = \frac{218 - (4)(8)(6)}{296 - (4)(64)} = 0,65$$

$$a = \bar{Y} - b \bar{X} = 6 - (0,65)(8) = 0.80$$

$$\hat{Y}_i = 0,80 + 0,65 X_i$$

© J.P. Sousa

Slide 315 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Modelos de regressão Passos de cálculo

1. Definir o problema ou questão
2. Especificar o modelo
3. Recolher os dados
4. Fazer uma análise descritiva dos dados
5. Estimar parâmetros desconhecidos
- 6. Avaliar o modelo**
7. Utilizar o modelo para fazer previsões

© J.P. Sousa

Slide 316 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Avaliação do modelo

1. Até que ponto o modelo descreve correctamente a relação entre as variáveis?
2. Ajustamento dos pontos à recta
 - Quanto mais próximos da recta melhor
3. Pressupostos cumpridos
4. Significância da estimativa dos parâmetros
5. “Outliers” (Observações irregulares)

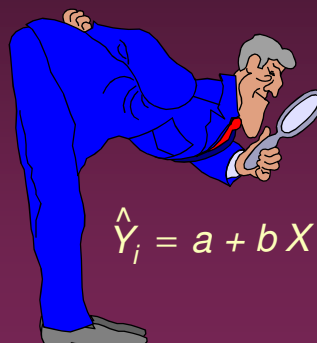
© J.P. Sousa

Slide 317 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Passos de avaliação do modelo

1. Exame das medidas de variação
2. Fazer análise de resíduos
3. Testar significância dos coeficientes
4. Realizar análise de influências



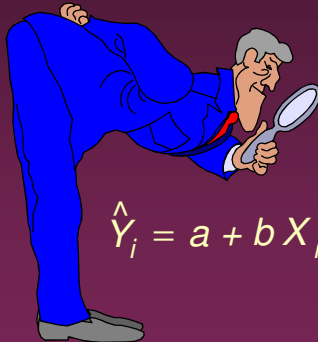
© J.P. Sousa

Slide 318 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Passos de avaliação do modelo

1. Exame das medidas de variação
2. Fazer análise de resíduos
3. Testar significância dos coeficientes
4. Realizar análise de influências


$$\hat{Y}_i = a + b X_i$$

© J.P. Sousa Slide 319 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Variação aleatória do erro

1. Variação dos valores de Y observados relativamente aos previstos (na recta)
2. Medição do “erro padrão da estimativa”
 - Desvio padrão do ‘erro’ ou dos ‘resíduos’
 - Representado por S_{YX}
3. Afecta diversos factores
 - Significância dos parâmetros
 - Precisão das previsões

© J.P. Sousa Slide 320 de 373

Variação aleatória do erro

1. Como se fará a análise dessa variação ?



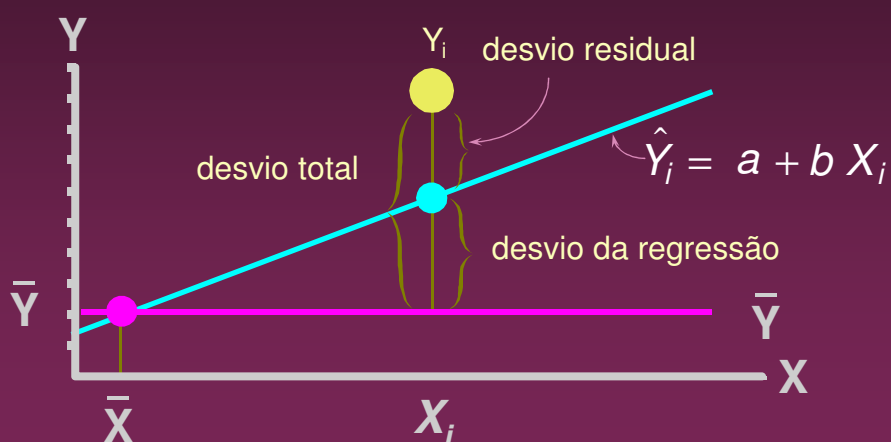
ANOVA

1. Qual será a hipótese nula ?

$$b = 0$$

$$\beta = 0 \quad \text{SIM !}$$

Medidas de variação



Algumas notações !

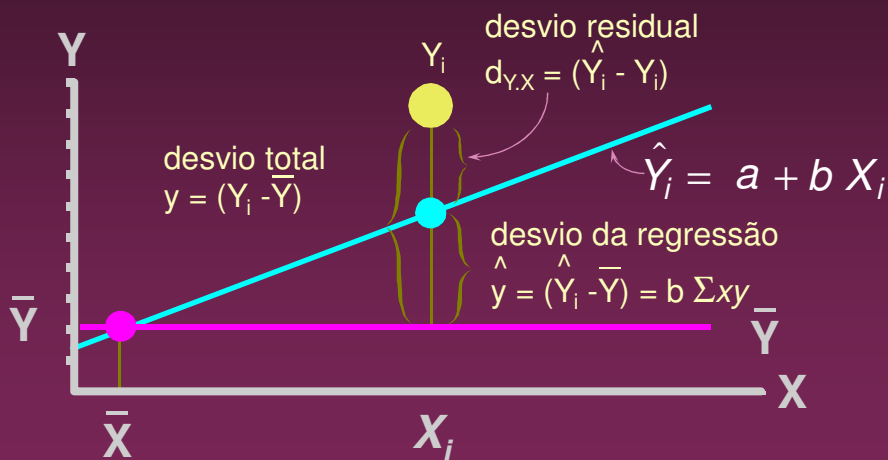
- $\sum x = \sum (X_i - \bar{X})$
- $\sum x^2 = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n(\bar{X})^2$
- $\sum xy = \sum (X_i - \bar{X}) (Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y}$

$$b = \frac{\sum xy}{\sum x^2} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n(\bar{X})^2}$$

© J.P. Sousa

Slide 323 de 373

Medidas de variação



© J.P. Sousa

Slide 324 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Medidas de variação em regressão

1. Soma de Quadrados Totais (SST)
 - Mede a variação dos valores Y_i à volta de \bar{Y}
2. Variação explicada pela regressão (SSR)
 - Variação devida à relação entre X & Y
3. Variação residual (SSE)
 - Variação devida a outro factores

© J.P. Sousa

Slide 325 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Medidas de variação em regressão

Fonte de variação	Soma de quadrados (SS)	gl	Média de Quadrados (MS)
Total	Σy^2	$n - 1$	s_Y^2
Regressão linear	$\Sigma \hat{y}^2 = \frac{(\Sigma xy)^2}{\Sigma x^2}$	1	$s_{\hat{Y}}^2$
Residual	$\Sigma d_{Y.X}^2 = \Sigma y^2 - \Sigma \hat{y}^2$	$n - 2$	$s_{Y.X}^2$

© J.P. Sousa

Slide 326 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Regressão linear Significância

1. Hipóteses
 - $H_0: \beta = 0$
 - $H_A: \beta \neq 0$
2. Teste estatístico
 - $F = \frac{MSR}{MSE}$
3. Valor crítico
 - $F_{\alpha(1),1,(n-2)}$

© J.P. Sousa Slide 327 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016


Coeficiente de determinação

1. **Proporção** da variação explicada pela relação entre X & Y

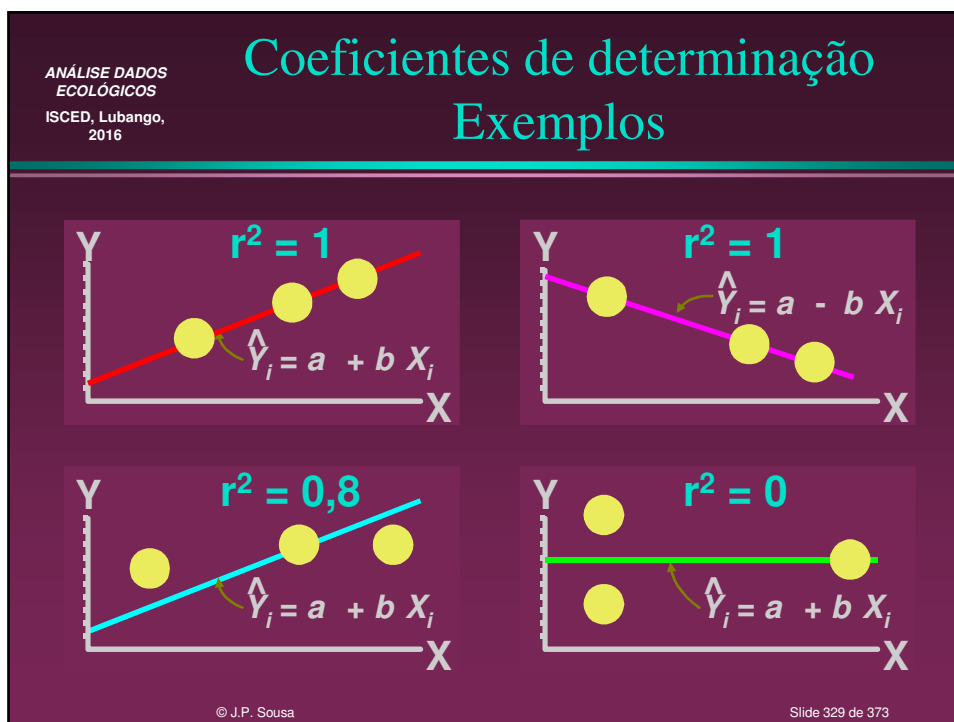
$0 \leq r^2 \leq 1$

$$r^2 = \frac{\text{Variação explicada}}{\text{Variação Total}} = \frac{SSR}{SST}$$

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \cdot \sum y^2}$$



© J.P. Sousa Slide 328 de 373



ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Coeficiente de determinação ajustado

1. Proporção da variação “explicada” pela relação entre X e Y
2. Reflecte
 - Tamanho da amostra
 - Número de variáveis independentes
3. Equação

$$r_{adj}^2 = 1 - \left[(1 - r^2) \cdot \frac{n - 1}{n - 2} \right]$$

© J.P. Sousa


Slide 330 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Coeficiente de determinação Exemplo

Estudou-se o crescimento de uma planária ao longo de 5 dias. A relação entre as duas variáveis correspondia a um valor de $a = -0,1$ e de $b = 0,7$.

<u>Idade (d)</u>	<u>Comp. (mm)</u>
1	1
2	1
3	2
4	2
5	4



Qual **coeficiente de determinação**?

© J.P. Sousa Slide 331 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Coeficiente de determinação Tabela de cálculo

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

© J.P. Sousa Slide 332 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Coeficiente de determinação Solução

$$\sum x^2 = 10 \quad \sum y^2 = 6 \quad \sum xy = 7$$

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2}$$

$$= \frac{(7)^2}{(10)(6)} = 0,8167$$

81,67% da
variação do
comprimento é
devida à idade

© J.P. Sousa

Slide 333 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Coeficiente de determinação Resultado de computador

Root MSE	0.60553	R-square	0.8167
Dep Mean	2.00000	Adj R-sq	0.7556
C.V.	30.27650		

S_{YX}

r^2 ajustado para o
número de variáveis e o
tamanho da amostra

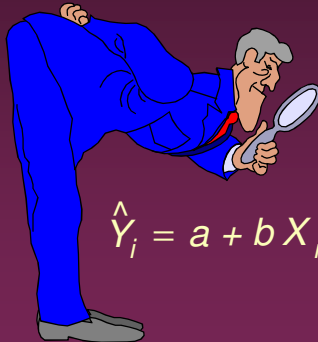
© J.P. Sousa

Slide 334 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Passos de avaliação do modelo

1. Exame das medidas de variação
- 2. Fazer análise de resíduos**
3. Testar significância dos coeficientes
4. Realizar análise de influências


$$\hat{Y}_i = a + b X_i$$

© J.P. Sousa Slide 335 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Análise de resíduos

1. Análise gráfica de resíduos
 - Gráficar Resíduos vs. os valores X_i
 - Médias dos Erros Residuais
 - Diferença entre os valores Y_i reais & os valores Y_i previstos
2. Objectivos
 - Examinar a forma funcional (Modelos Lineares vs. Modelos Não-Lineares)
 - Avaliar violação dos pressupostos

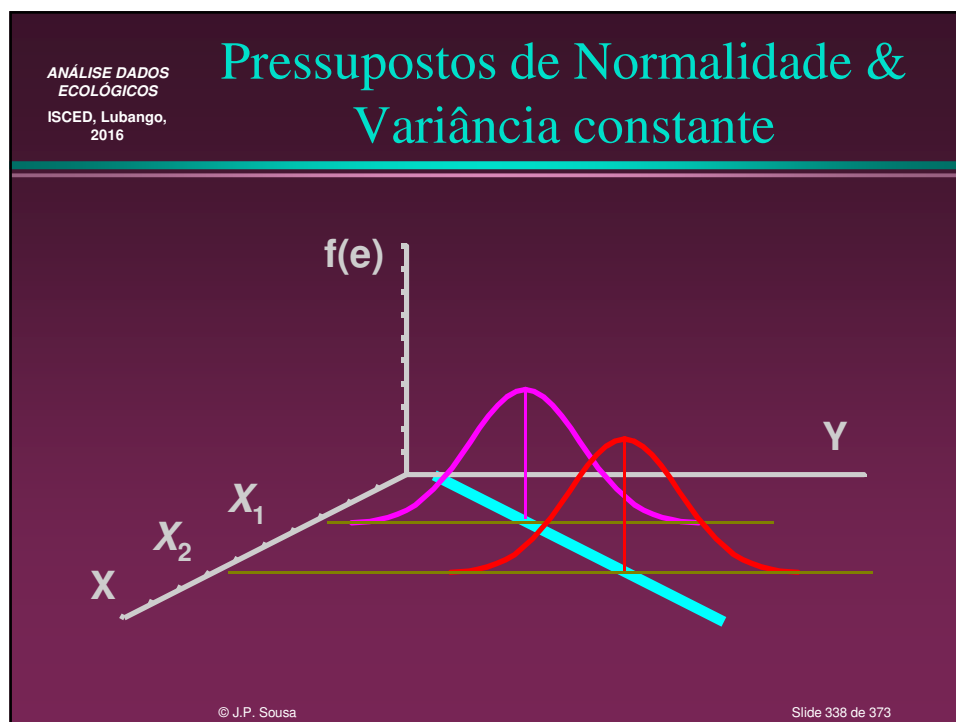
© J.P. Sousa Slide 336 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Regressão Linear Pressupostos

1. Normalidade
 - Valores de Y normalmente distribuídos para todos os valores de X
 - Distribuição normal das probabilidade do erro
2. Homoscedasticidade (Variância constante)
3. Independência dos Erros
4. Linearidade

© J.P. Sousa Slide 337 de 373



ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Gráfico de resíduos para a forma funcional

➤ Se os pressupostos forem cumpridos, especialmente a homogeneidade de variâncias, os resíduos aparecem em bandas paralelas em torno da sua média (0)




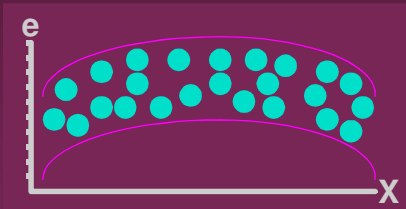
Especificação correcta

© J.P. Sousa Slide 339 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

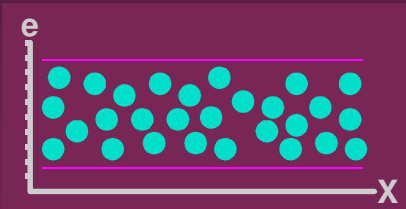
Gráfico de resíduos para a forma funcional

 **Utilizar o termo X^2**

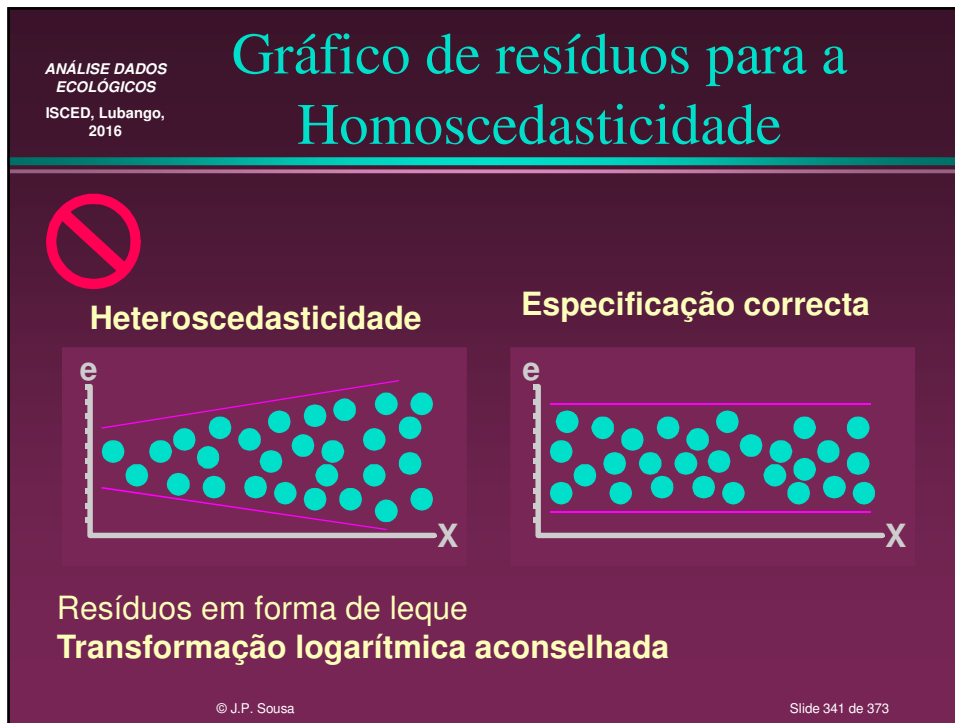


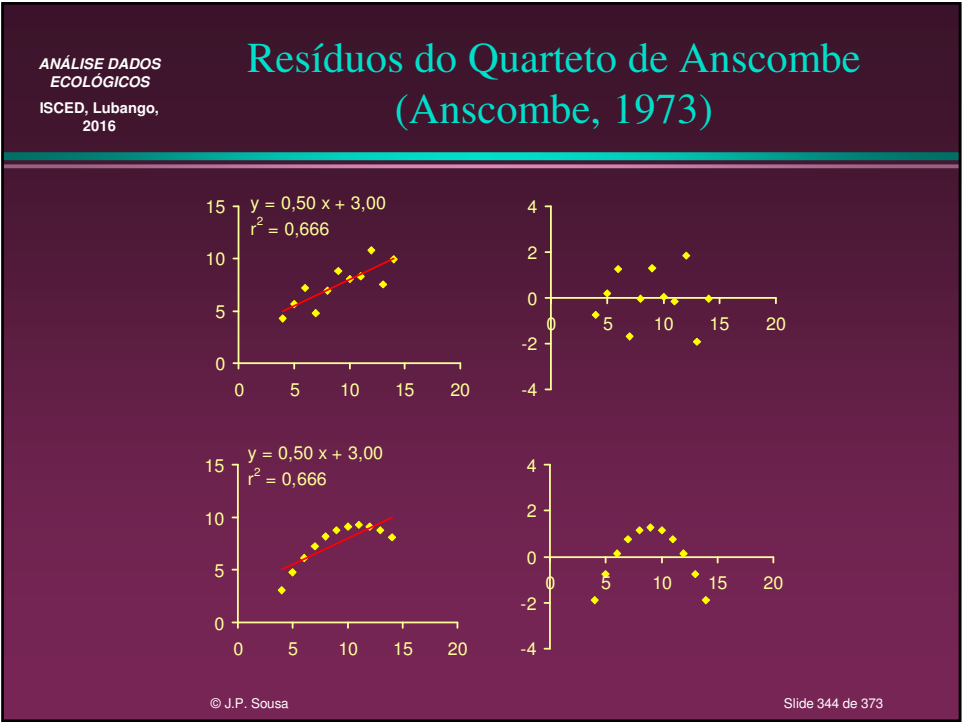
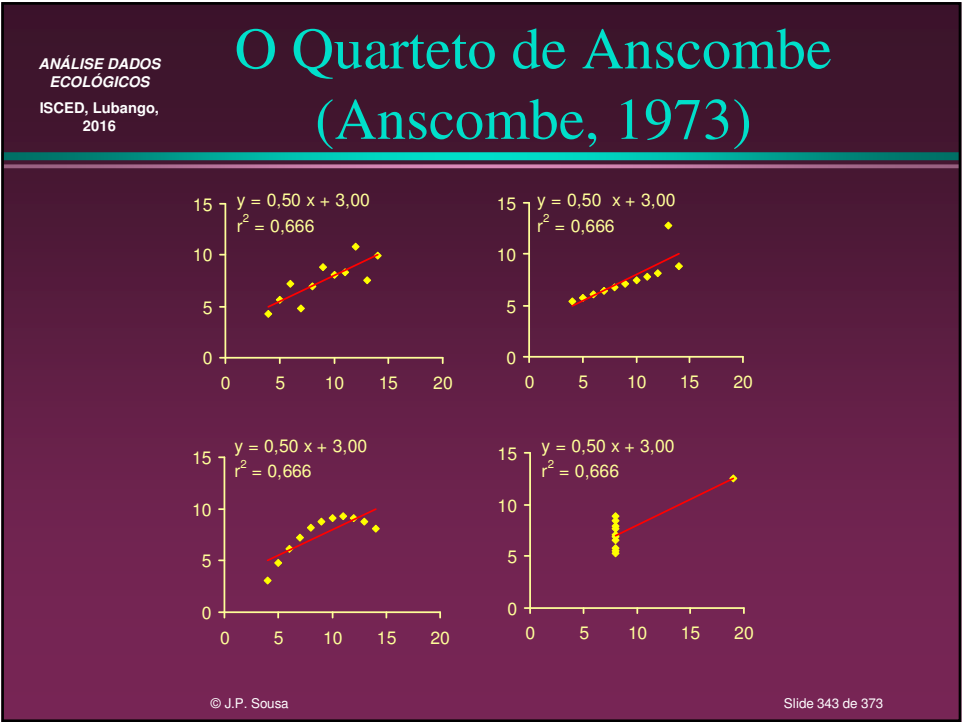
**Regressão quadrática
aconselhada**

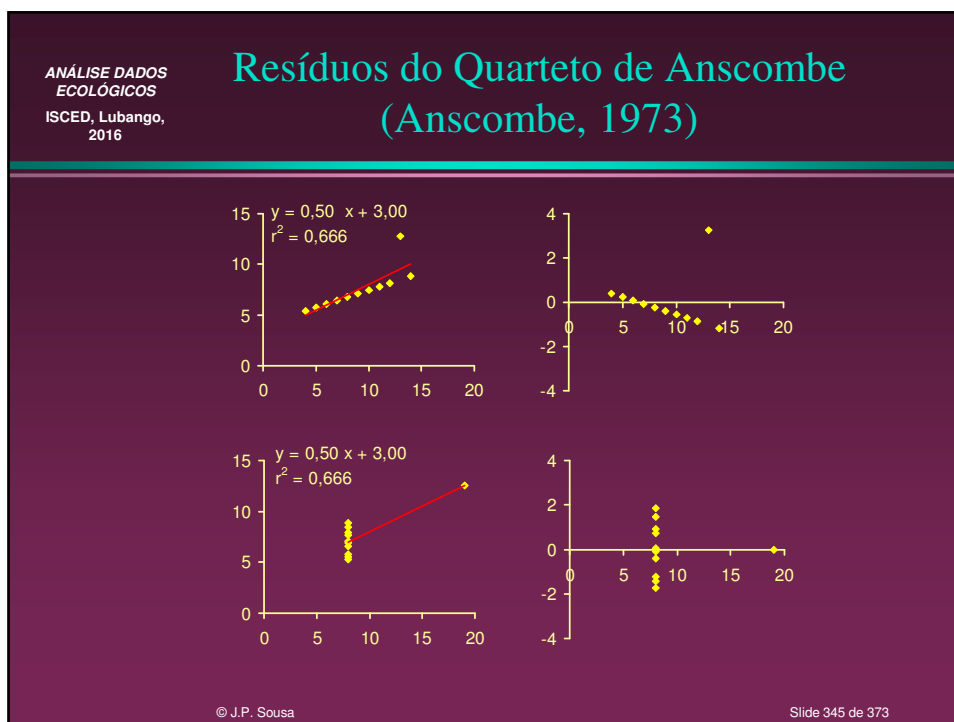
Especificação correcta



© J.P. Sousa Slide 340 de 373







ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Passos de avaliação do modelo

1. Exame das medidas de variação
2. Fazer análise de resíduos
- 3. Testar significância dos coeficientes**
4. Realizar análise de influências

$\hat{Y}_i = a + bX_i$

© J.P. Sousa

Slide 346 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Testes ao coeficiente de regressão

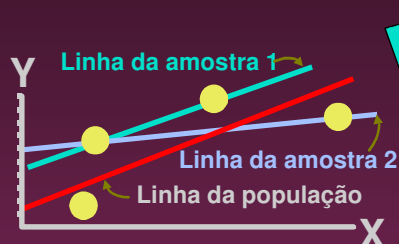
1. Testa a existência de uma relação linear entre X & Y
2. Envolve o declive da população β
3. Hipóteses
 - $H_0: \beta = 0$ (não existe relação linear)
 - $H_A: \beta \neq 0$ (existe relação linear)
4. A base teórica é a distribuição da amostragem dos declives

© J.P. Sousa

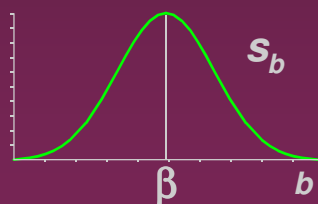
Slide 347 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Distribuição de amostragem dos declives



Distribuição de amostragem



© J.P. Sousa

Todos os declives possíveis para as amostras

Amostra 1: 2,5

Amostra 2: 1,6

Amostra 3: 1,8

Amostra 4: 2,1

Um número muito grande de declives da amostra

Slide 348 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste do coeficiente de regressão Teste Estatístico

$$t = \frac{b - \beta_0}{S_b}$$

onde

$$S_b = \frac{S_{YX}}{\sqrt{\sum_{i=1}^n X_i^2 - n(\bar{X})^2}} = \frac{S_{YX}}{\sqrt{\sum X^2}}$$

© J.P. Sousa

Slide 349 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste do coeficiente de regressão Exemplo

Estamos a estudar a relação entre o tamanho de planárias e a idade. Encontramos $a = -0,1$; $b = 0,7$ & $S_{YX} = 0,60553$.

<u>Idade</u>	<u>Comp. (mm)</u>
1	1
2	1
3	2
4	2
5	4



Existe uma relação **significativa** para um nível de **0,05**?

© J.P. Sousa

Slide 350 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste do coeficiente de regressão Tabela de cálculo

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

© J.P. Sousa

Slide 351 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Teste Estatístico Solução

$$S_b = \frac{S_{YX}}{\sqrt{\sum_{i=1}^n X_i^2 - n(\bar{X})^2}} = \frac{0,60553}{\sqrt{55 - (5)(3)^2}} = 0,1915$$

$$t = \frac{b - \beta_0}{S_b} = \frac{0,70 - 0}{0,1915} = 3,656$$

© J.P. Sousa


Slide 352 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste do coeficiente de regressão

Solução

H₀: $\beta = 0$
H_A: $\beta \neq 0$
 α = 0,05
gl = 5 - 1 - 1 = 3
Valor crítico:



Teste estatístico:

$$t = \frac{b - \beta_0}{S_b} = \frac{0,70 - 0}{0,1915} = +3,655$$

Decisão:
Rejeitar para $\alpha = 0,05$

Conclusão:
Há evidência de existir uma relação entre X & Y

© J.P. Sousa Slide 353 de 373

ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016

Teste do declive

Resultado do computador

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0:	Param=0	Prob> T
INTERCEP	1	-0.1000	0.6350	-0.157	0.8849	
Slope	1	0.7000	0.1914	3.656	0.0354	

Diagram illustrating the relationship between the parameters and the test results:

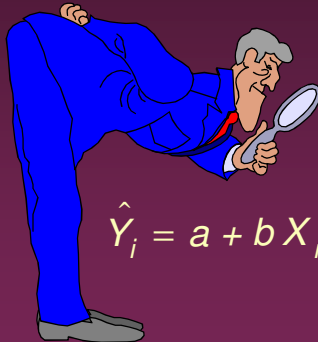
- b (intercept) points to the Estimate for INTERCEP.
- S_b (standard error of the slope) points to the Standard Error for Slope.
- $t = b / S_b$ points to the T for H0 for Slope.
- Valor de p points to the Prob>|T| for Slope.

© J.P. Sousa Slide 354 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Passos de avaliação do modelo

1. Exame das medidas de variação
2. Fazer análise de resíduos
3. Testar significância dos coeficientes
4. **Realizar análise de influências**


$$\hat{Y}_i = a + b X_i$$

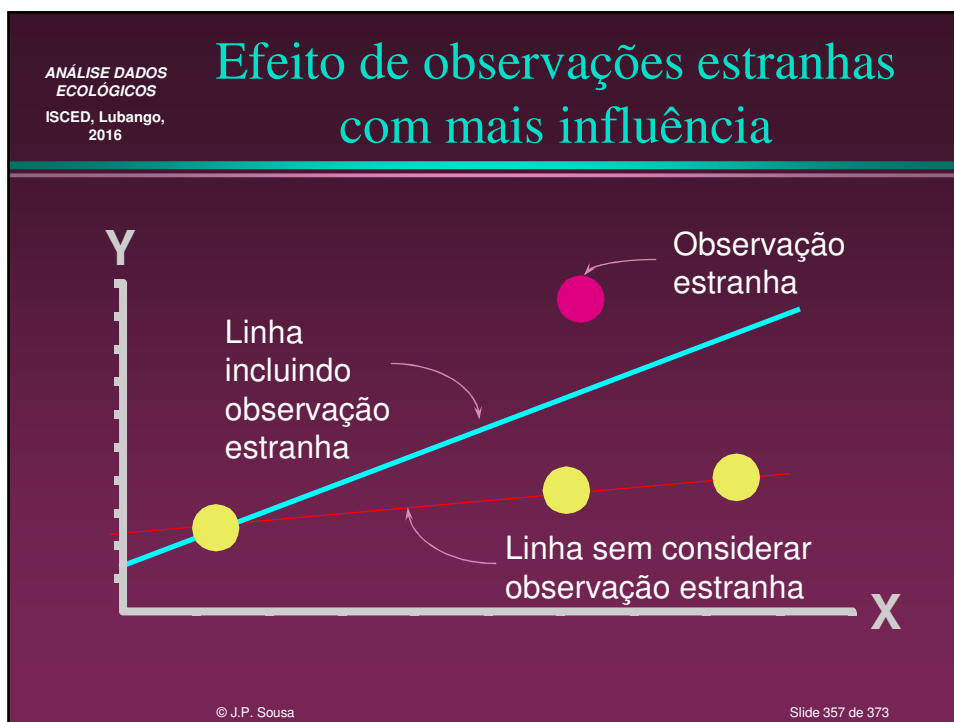
© J.P. Sousa Slide 355 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Análise de influências

1. Exame das observações que mais afectam os valores dos coeficientes
2. Exemplo - Durante o estudo ocorreu uma infecção das culturas
3. Devemos tentar perceber as causas da ocorrência destas observações
4. **Cuidadosamente** ponderar a eliminação da observação

© J.P. Sousa Slide 356 de 373



- ANÁLISE DADOS ECOLÓGICOS
ISCED, Lubango, 2016
- ### Modelos de regressão
- #### Passos de cálculo
1. Definir o problema ou questão
 2. Especificar o modelo
 3. Recolher os dados
 4. Fazer uma análise descritiva dos dados
 5. Estimar parâmetros desconhecidos
 6. Avaliar o modelo
 - 7. Utilizar o modelo para fazer previsões**
- © J.P. Sousa
- Slide 358 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Previsões utilizando modelos de regressão

1. Tipos de previsões

- Estimativas de pontos
- Estimativas de intervalos

2. O que se prevê ?

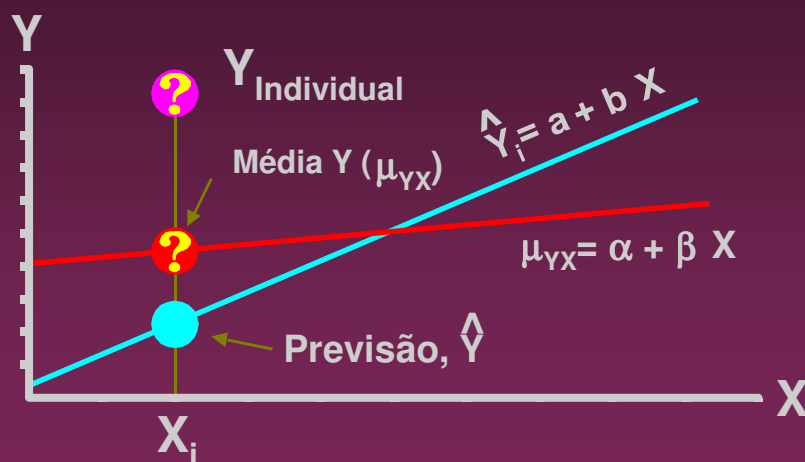
- Respostas médias da população (μ_{YX}) para um determinado valor de X
 - Ponto na linha de regressão da população
- Resposta individual (Y_i) para um determinado valor de X

© J.P. Sousa

Slide 359 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

O que se prevê ?



© J.P. Sousa

Slide 360 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Intervalo de *Confiança* Estimativa da média Y (μ_{YX})

$$\hat{Y} - t_{\alpha(2),n-2} \cdot S_{\hat{Y}} \leq \mu_{YX} \leq \hat{Y} + t_{\alpha(2),n-2} \cdot S_{\hat{Y}}$$

onde

$$S_{\hat{Y}} = S_{YX} \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n X_i^2 - n(\bar{X})^2}}$$

© J.P. Sousa Slide 361 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Factores que afecta a largura do intervalo

1. Nível de confiança ($1 - \alpha$)
 - A largura aumenta com o aumento da confiança
2. Dispersão de dados (S_{YX})
 - A largura aumenta com o aumento da variabilidade
3. Tamanho da amostra
 - A largura diminui com o aumento do tamanho da amostra
4. Distância de X_i relativamente à média \bar{X}
 - A largura aumenta com o aumento da distância

© J.P. Sousa Slide 362 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Intervalo de confiança da estimativa Exemplo

Estamos a estudar a relação entre o tamanho de planárias e a idade. Encontramos $a = -0,1$; $b = 0,7$ & $S_{YX} = 0,60553$.

Idade	Comp. (mm)
1	1
2	1
3	2
4	2
5	4



Qual a estimativa **média** do comprimento com **4 dias** de idade para um nível de **0,05**?

© J.P. Sousa

Slide 363 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Intervalo de confiança da estimativa Tabela de cálculo

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

© J.P. Sousa

Slide 364 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Intervalo de *Confiança* da estimativa
Solução

$$\hat{Y} - t_{\alpha(2),n-2} \cdot S_{\hat{Y}} \leq \mu_{YX} \leq \hat{Y} + t_{\alpha(2),n-2} \cdot S_{\hat{Y}}$$
$$\hat{Y} = -0,1 + (0,7)(4) = 2,7$$
$$S_{\hat{Y}} = 0,60553 \cdot \sqrt{\frac{1}{5} + \frac{(4-3)^2}{55 - (5)(3)}} = 0,3316$$
$$2,7 - (3,1824)(0,3316) \leq \mu_{YX} \leq 2,7 + (3,1824)(0,3316)$$
$$1,6445 \leq \mu_{YX} \leq 3,7553$$

© J.P. Sousa

Slide 365 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Intervalos de confiança
Resultado de computador

Dep	Var	Pred	Std Err	Low95%	Upp95%	Low95%	Upp95%
Obs	SALES	Value	Predict	Mean	Mean	Predict	Predict
1	1.000	0.600	0.469	-0.892	2.092	-1.837	3.037
2	1.000	1.300	0.332	0.244	2.355	-0.897	3.497
3	2.000	2.000	0.271	1.138	2.861	-0.111	4.111
4	2.000	2.700	0.332	1.644	3.755	0.502	4.897
5	4.000	3.400	0.469	1.907	4.892	0.962	5.837

Y previsto
quando X = 4

S_Ŷ

Intervalo de
confiança

Intervalo de
previsão

© J.P. Sousa

Slide 366 de 373

ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016


Previsão do intervalo de resposta individual

$$\hat{Y} - t_{\alpha(2), n-2} \cdot S_{(Y-\hat{Y})} \leq Y_P \leq \hat{Y} + t_{\alpha(2), n-2} \cdot S_{(Y-\hat{Y})}$$

onde

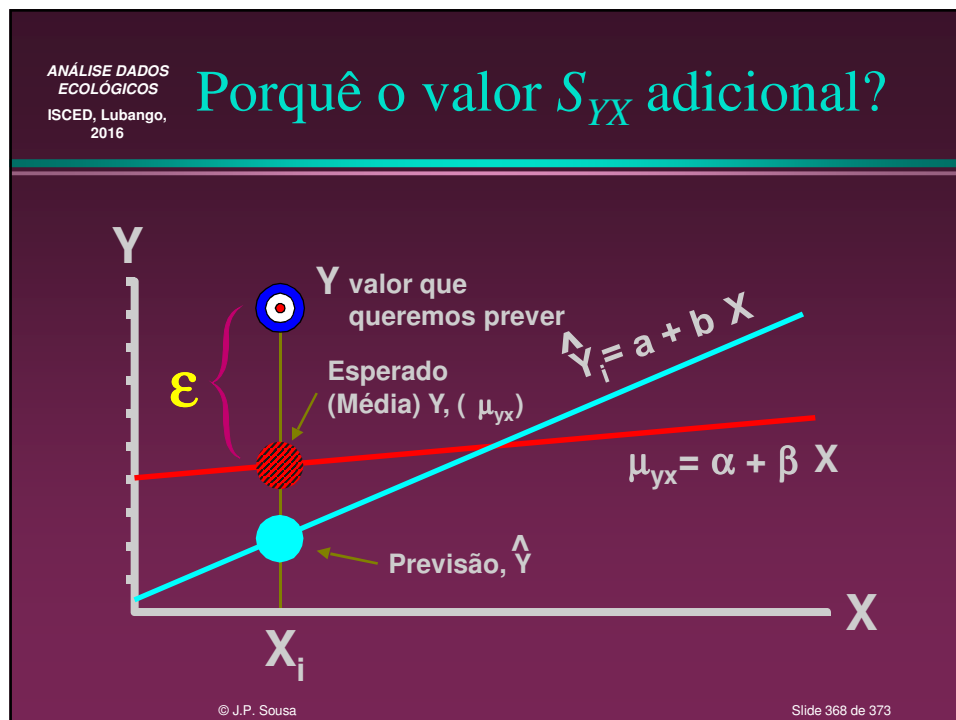
$$S_{(Y-\hat{Y})} = S_{YX} \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n X_i^2 - n(\bar{X})^2}}$$

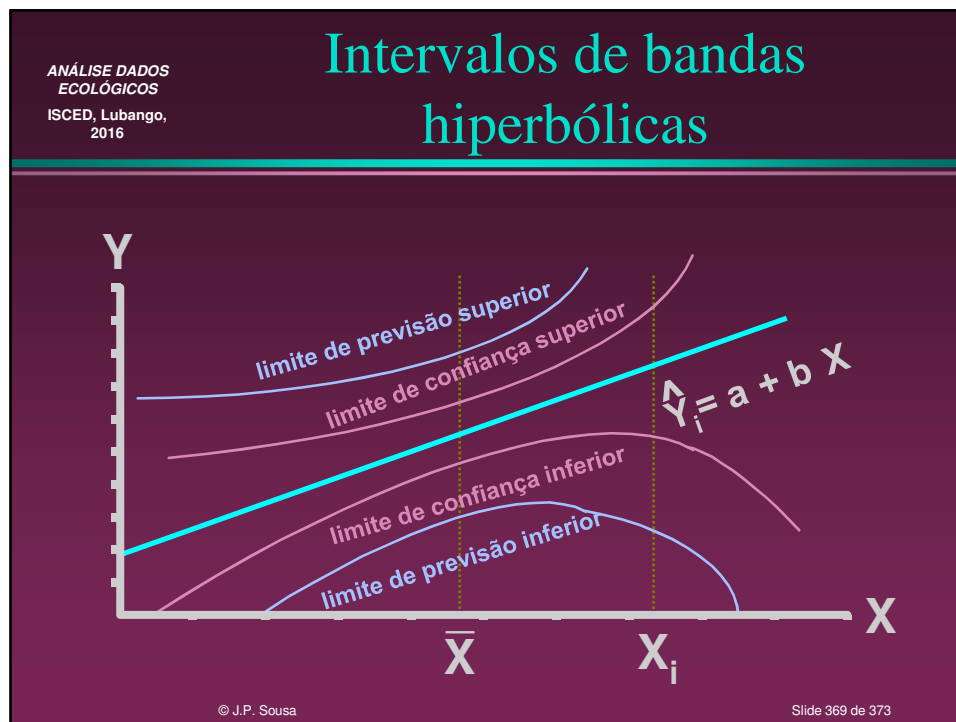
#1



Atenção!

© J.P. Sousa
Slide 367 de 373





ANÁLISE DADOS
ECOLÓGICOS
ISCED, Lubango,
2016

Precauções em regressão

1. Violação dos pressupostos
2. Relevância dos dados
3. Nível de significância
4. Extrapolação
5. Causa e efeito

ATENÇÃO!

© J.P. Sousa

Slide 370 de 373

