

Universidade Federal de Goiás

Bioestatística

Prof. Adriano Sanches Melo - Dep. Ecologia – ICB
asm.adrimelo@gmail.com

Página do curso: <http://www.ecologia.ufrgs.br/~adrimelo/bioestat>

Mais um pouco de probabilidade...

1. Uma doença atinge 1 em 1000, ou seja, 0,1%
2. A probabilidade de um falso-positivo no exame é de 5%
3. Uma pessoa recebeu um resultado positivo
4. Qual a probabilidade da pessoa ter a doença

RESPOSTA →

Medidas de dispersão
Correlação

Medidas de dispersão

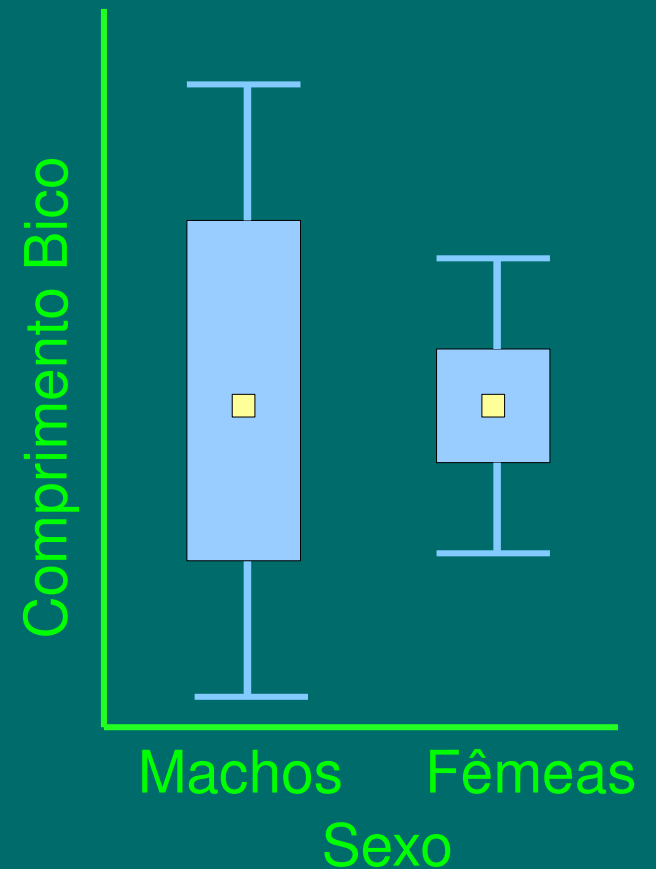
4.1.1 – Mínimo, Máximo e Amplitude

4.1.2 – Distância Interquartil

Percentil

Quartil

Diagrama de caixa (*Box plot*)



Variância e Desvio Padrão

Desvio = observação – média (vimos na aula passada)

Soma dos desvios é igual a zero (negativos anulam positivos)

Para obter uma medida de variabilidade, poderíamos obter a Soma dos Quadrados dos Desvios (SQ).

Problema: maior número de observações, maior o valor de SQ
Solução: dividir por graus de liberdade:

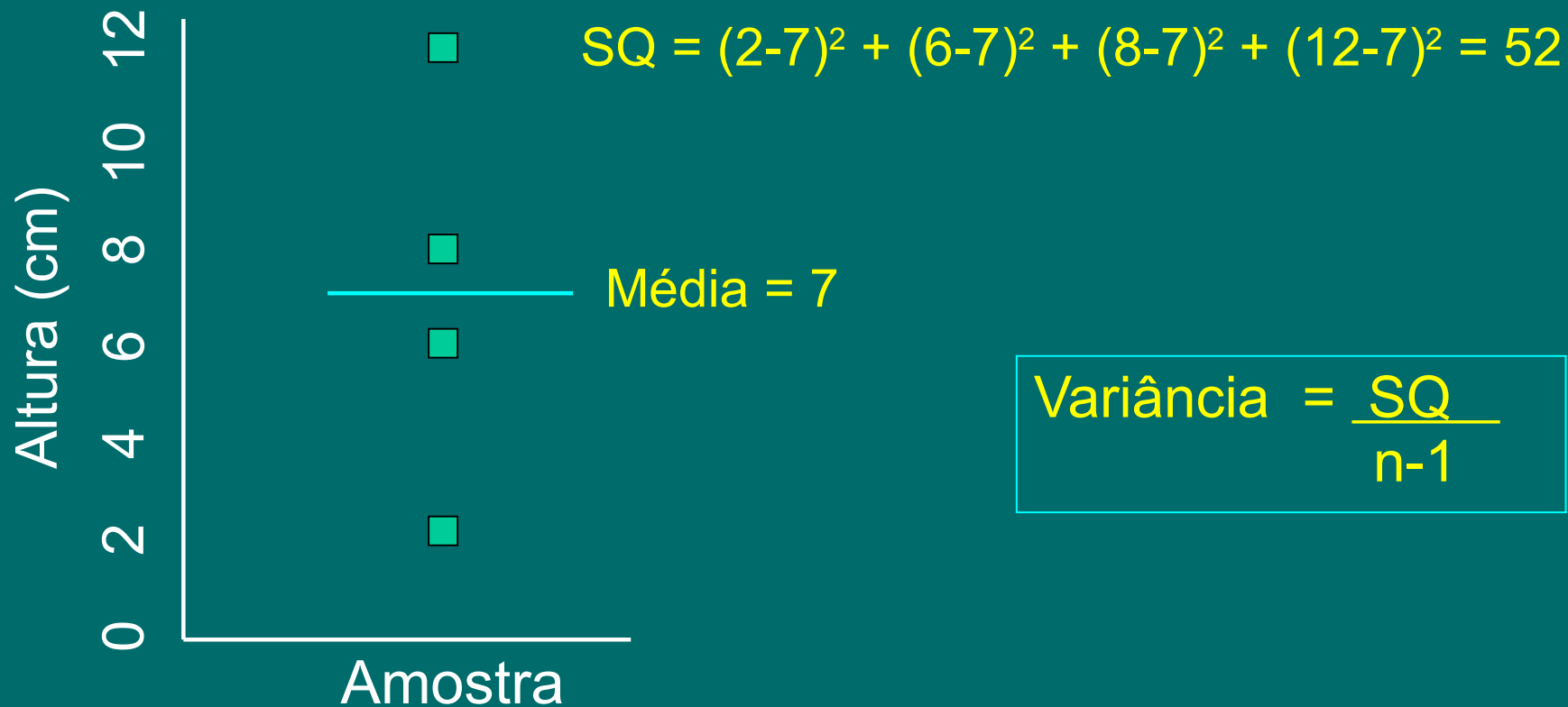
$$S^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Variância e Desvio Padrão

Problema: Unidade está em 'quadrados'

Solução: Tirar raiz quadrada:

$$s = \sqrt{\text{variância}} = \sqrt{\left(\frac{\sum (x - \bar{x})^2}{n-1}\right)}$$



$$\text{Variância} = \frac{SQ}{n-1}$$

Coeficiente de variação

Uma medida de variação relativa à média:

$$CV = \frac{s}{\bar{x}} \times 100$$

Exemplo: idade de dois grupos de pessoas

Grupo1: 3, 1 e 5 → $\bar{x} = 3; s = 2; CV = \frac{2}{3} \times 100 = 66,67 \%$

Grupo2: 55, 57 e 53 → $\bar{x} = 55; s = 2; CV = \frac{2}{55} \times 100 = 3,64 \%$

A idéia de graus de liberdade

$$\text{Variância} = s^2 = \frac{SQ}{n-1}$$

-- Por que dividir por (n-1) ?

-- A idéia de *graus de liberdade* (Crawley p.36)

Suponha termos 5 números e que sua média seja 4. A soma dos números, portanto, deve ser 20. Vamos ver quais números poderiam ser:

1) O primeiro número pode ser qualquer um; por exemplo o 2

2				
---	--	--	--	--

2) O segundo número pode ser qualquer um; por exemplo o 7

2	7			
---	---	--	--	--

3) O terceiro número pode ser qualquer um; por exemplo o 4

2	7	4		
---	---	---	--	--

4) O quarto número pode ser qualquer um; por exemplo o 0

2	7	4	0	
---	---	---	---	--

5) Não temos escolha para o quinto número; ele DEVE ser 7

2	7	4	0	7
---	---	---	---	---

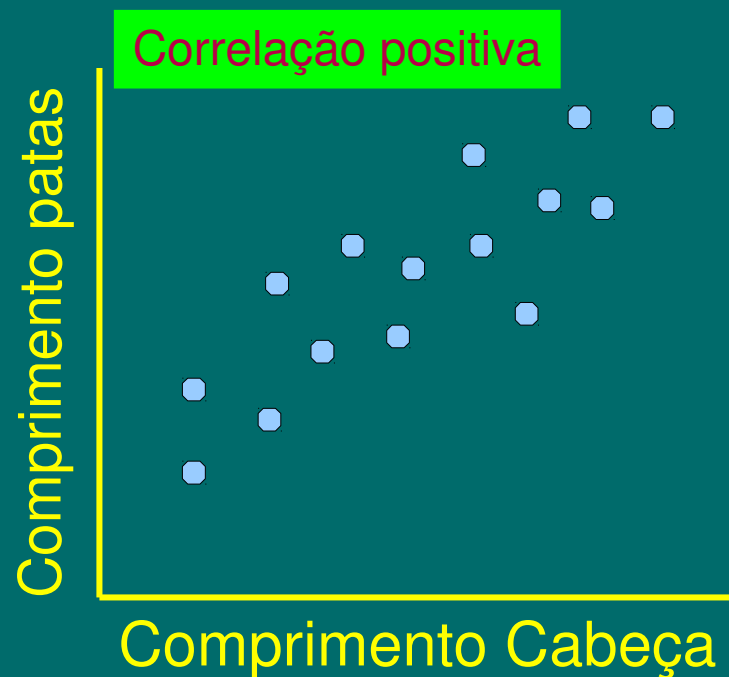
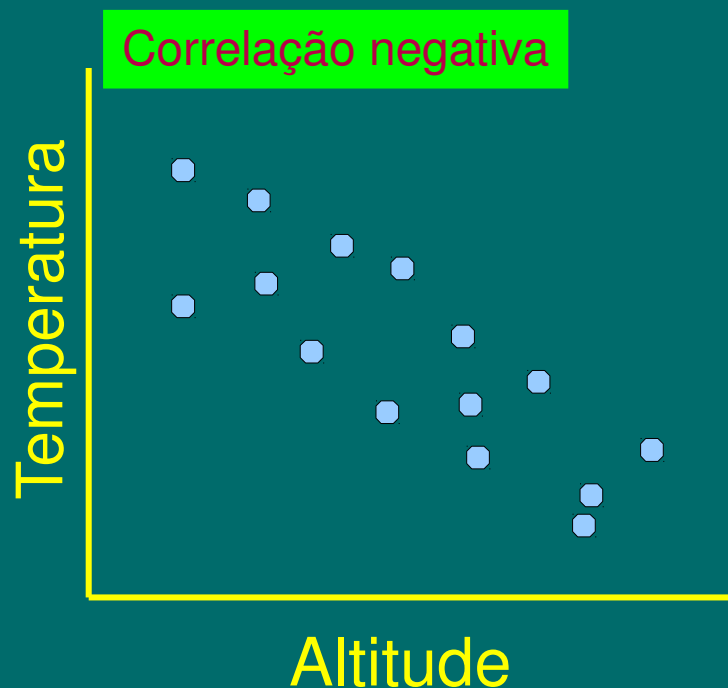
Correlação

Uma medida de associação entre duas variáveis:

- Pressão arterial aumenta com idade
- Temperatura diminui com altitude
- Indivíduos com maiores patas possuem cabeças maiores

Importante: Uma medida de associação; NÃO implica causa-e-efeito

Representação gráfica: Diagrama de dispersão

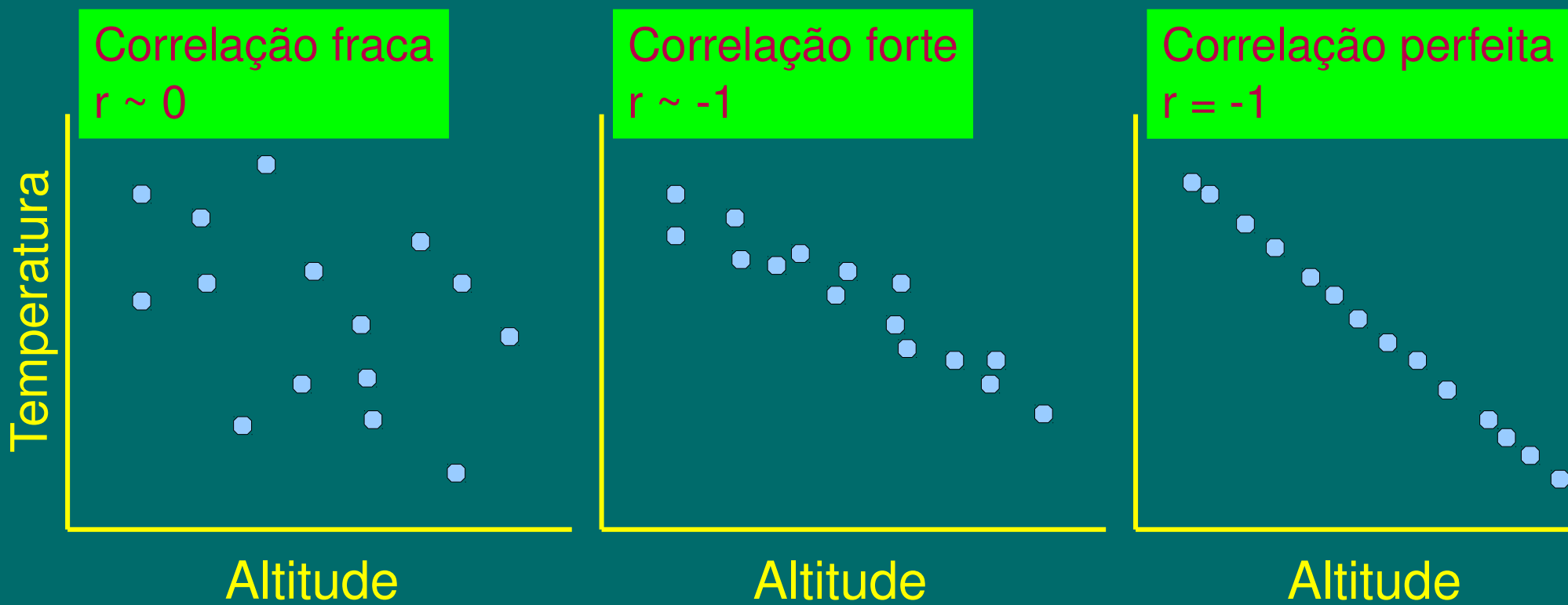


Correlação

Como medir?

$$r = \frac{\sum (XY) - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{[\sum X^2 - \frac{(\sum X)^2}{n}][\sum Y^2 - \frac{(\sum Y)^2}{n}]}}$$

Coeficiente de correlação de Pearson
Varia de -1 a 1



Exercícios

No Caderno

Com planilha eletrônica (demonstração)

Uso de fórmulas

Cálculo de correlação

O probleminha do início da aula...

Mais um pouco de probabilidade...

1. Uma doença atinge 1 em 1000, ou seja, 0,1%
2. A probabilidade de um falso-positivo no exame é de 5%
3. Uma pessoa recebeu um resultado positivo
4. Qual a probabilidade da pessoa ter a doença

RESPOSTA → 2%

Consegue explicar?

Sugestão de estudo:

Vieira, S. 2008. Introdução à Bioestatística. 4a ed. Elsevier
Caps: 5 e 6