

# Investigação Aplicada I

## Aula 8

1º Semestre 2020/21

Licenciatura em Ciências Biomédicas Laboratoriais

## **Sumário**

Testes à associação entre variáveis qualitativas

Avaliação de correlações

# Testes à associação entre variáveis qualitativas

## Testes não paramétricos

- Os dados não seguem uma distribuição normal
- Amostras pequenas
- Não são tão potentes quanto os testes paramétricos, ou seja, não se encontram tantas diferenças entre os dados quando estas diferenças realmente existem

# Testes à associação entre variáveis qualitativas

## Testes não paramétricos

### Tabelas de contingência e Teste Qui-quadrado

- Análise de variáveis qualitativas
  - Utiliza-se quando queremos saber se existem diferenças (ou relação) estatisticamente significativa entre duas variáveis.
1. Estabelecer as hipóteses estatísticas
  2. Estabelecer nível de significância
  3. Calcular o valor do teste → valor estatístico do teste:  $p$

# Testes à associação entre variáveis qualitativas

## Testes não paramétricos

### Tabelas de contingência (2x2) e Teste Qui-quadrado ( $\chi^2$ )

Exemplo:

Exposição a um agente tóxico em que há apenas 2 resultados em relação ao estado (0- são; 1- doente)

Questão: há associação entre a exposição ao agente tóxico e a doença?

$H_0$ : Não há relação entre as 2 variáveis (são independentes)

$H_1$ : Há relação entre as 2 variáveis

# Testes à associação entre variáveis qualitativas

## Testes não paramétricos

### Teste Qui-quadrado

#### Valores observados

		Com patologia	Sem patologia	Total
Exposição ao tóxico	Sim	42	34	76
	Não	38	166	204
Total		80	200	280

Compara as frequências que observamos em determinadas variáveis (*observed*) COM as frequências que obteríamos se não houvesse efeito (*expected*)

# Testes à associação entre variáveis qualitativas

## Testes não paramétricos

### Teste Qui-quadrado

#### Valores esperados (caso não exista relação entre variáveis)

		Com patologia	Sem patologia	Total
Exposição ao tóxico	Sim	$80 \times (76/280) =$ <b>21,7</b>	$200 \times (76/280) =$ <b>54,3</b>	76
	Não	$80 \times (204/280) =$ <b>58,3</b>	$200 \times (204/280) =$ <b>145,7</b>	204
Total		80	200	280

# Testes à associação entre variáveis qualitativas

## Testes não paramétricos

### Teste Qui-quadrado

Comparar as frequências que observamos (*observed*) COM as frequências que obteríamos se não houvesse efeito (*expected*)

$$\chi^2 = \sum \frac{(\text{observed}_{ij} - \text{expected}_{ij})^2}{\text{expected}_{ij}}$$

- Valor crítico para o teste do  $\chi^2$
- Consultar a tabela de distribuição de probabilidades do Qui-Quadrado para verificar o valor estatístico ( $p$ )

# Inferência estatística

## Testes não paramétricos

### Teste Qui-quadrado - Pressupostos

Nenhuma frequência esperada é inferior a 1

Não mais de 20% das células têm frequência esperada  $< 5$

Em tabelas 2x2 considera-se, em geral, ser necessário que as frequências esperadas sejam maiores ou iguais a 5

# Testes à associação entre variáveis qualitativas

## Teste Qui-quadrado

Exemplo:

Registou-se relativamente a um grupo de 77 controlos e a um grupo de 75 pacientes com cancro, quais destes poderiam ser considerados obesos.

**Questão 1:** Há associação entre a obesidade e o cancro?

# Testes à associação entre variáveis qualitativas

**Questão** : Há associação entre a obesidade e o cancro?

Perante duas variáveis qualitativas pretende-se averiguar se existe associação entre elas ou se, ao invés, estas são independentes

Tabelas de contingência e testes de associação:

*Analyze > Descriptive Statistics > Crosstabs*

Colocam-se as variáveis *Obesidade* e *Grupo* nos campos *Rows* e *Column*, respetivamente

Na opção *Statistics*, seleciona-se *Chi-square* e *Risk*

Na opção *Cells*, seleciona-se *Observed* e *Expected*

**Chi-square:** mede a associação entre variáveis

**Risk:** Mede o risco de uma pessoa com obesidade desenvolver cancro

# Testes à associação entre variáveis qualitativas

Questão 1: Há associação entre a obesidade e o cancro?

Teste Qui-quadrado (Chi-square):

	Chi-Square Tests				
	Value	df	Asymptotic Significance (2-tailed)	Exact Sig. (2-tailed)	Exact Sig. (1-sided)
Pearson Chi-Square	7.016	1	.008		
Continuity Correction	6.132	1	.013		
Likelihood Ratio	7.113	1	.008		
Fisher's Exact test				.010	.006
Linear-by-linear association	6.970	1	.008		
N of valid cases	152				

**Há associação estatisticamente significativa entre as variáveis,  
faz sentido avaliar o risco!**

# Testes à associação entre variáveis qualitativas

**Questão:** Há associação entre a obesidade e o cancro?

## Chi-square & Risk:

	Risk Estimate		
	Value	95% confidence Interval	
		Lower	Upper
Odds Ratio for Obesidade (Não/Sim)	2.559	1.265	5.175
For cohort Grupo = Controlo	1.530	1.135	2.063
For cohort Grupo = Cancro	.598	.392	.911
N of valid cases	152		

Sendo o valor de “ausência de cancro”= 0 e o valor da “presença de cancro”=1, o valor de *Odds Ratio*, não deve ser superior a 1.

Observando a primeira linha da tabela, a estimativa de de *Odds Ratio* entre pessoas com e sem obesidade é de 2.559, o que nos indica que a hipótese de ter cancro é diferente para pessoas com e sem obesidade, sendo 2.559 vezes superior para pessoas com obesidade

# Testes à associação entre variáveis qualitativas

**Questão 1:** Há associação entre a obesidade e o cancro?

Resultado

*Avaliou-se a eventual existência de associação entre a obesidade e o cancro recorrendo a um teste de independência Qui-quadrado.*

*Verificou-se existir associação estatisticamente significativa entre a obesidade e o cancro ( $p=0.008$ ). O risco de ter cancro está aumentado cerca de 2.6 vezes nas pessoas com obesidade ( $OR=2.559$ , com IC 95%)*

# **Avaliação de correlações**

# Avaliação de correlações

## Correlação e Regressão

- Análise bi-variada de correlação e regressão:  
associação de duas variáveis (regressão linear simples)
- Análise multi-variada de correlação e regressão:  
associação de múltiplas variáveis

# Avaliação de correlações

## Correlação e Regressão

### Regressão Linear Simples

Relação entre duas variáveis

#### **Funcional:**

A cada valor de X corresponde um único valor de Y

#### **Estatística:**

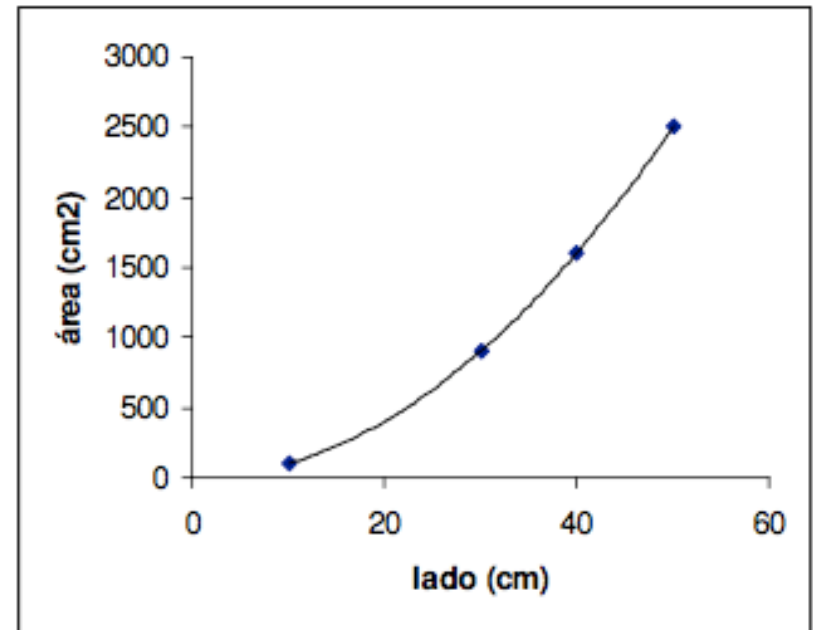
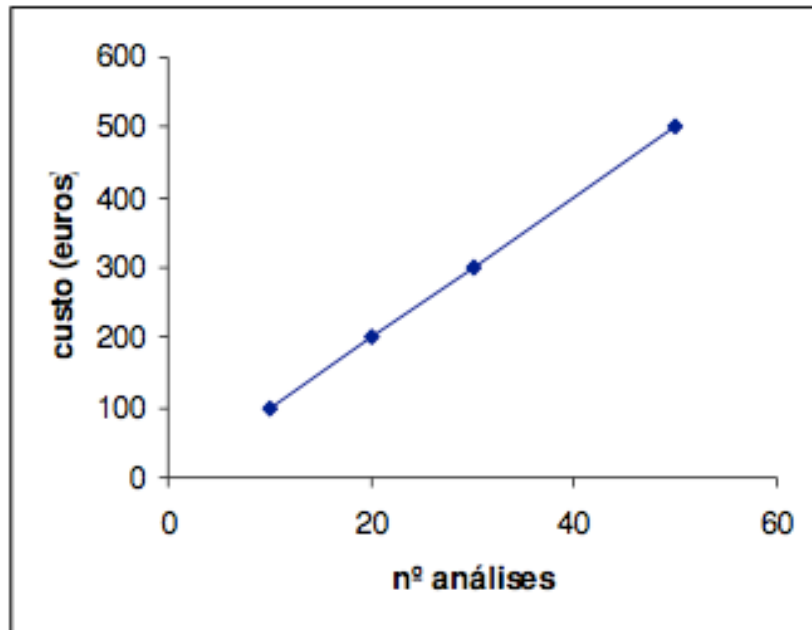
A cada valor de X não corresponde um único valor de Y

# Avaliação de correlações

## Correlação e Regressão

### Regressão Linear Simples

Relação funcional entre duas variáveis  $Y=f(X)$



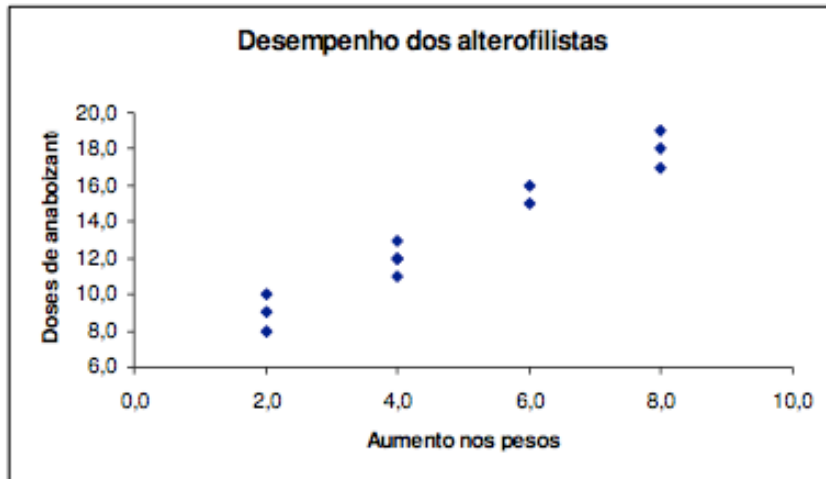
# Avaliação de correlações

## Correlação e Regressão

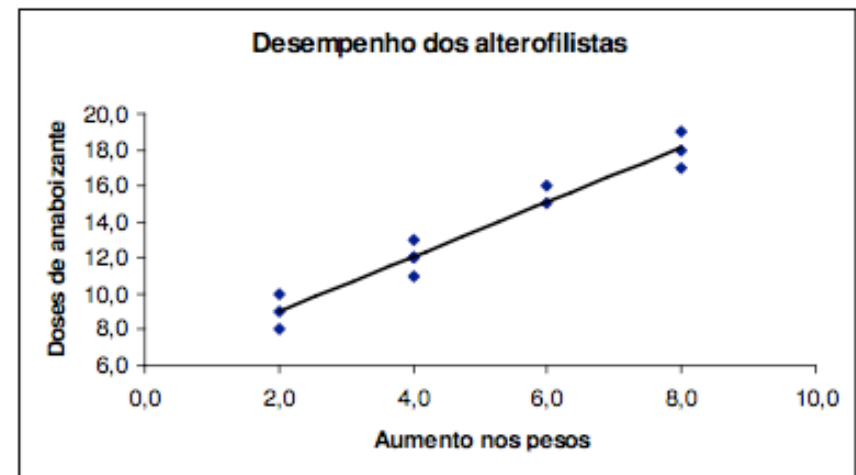
### Regressão Linear Simples

Relação estatística entre duas variáveis

Diagrama de dispersão



Recta ajustada



Relação Linear entre as variáveis

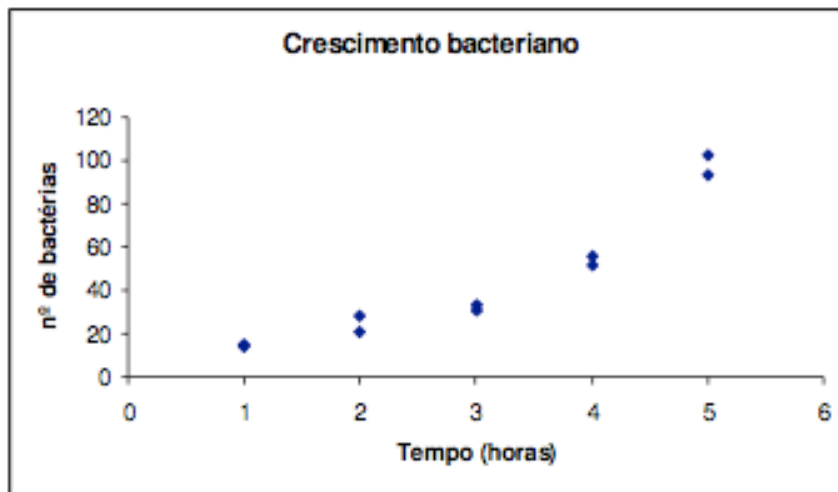
# Avaliação de correlações

## Correlação e Regressão

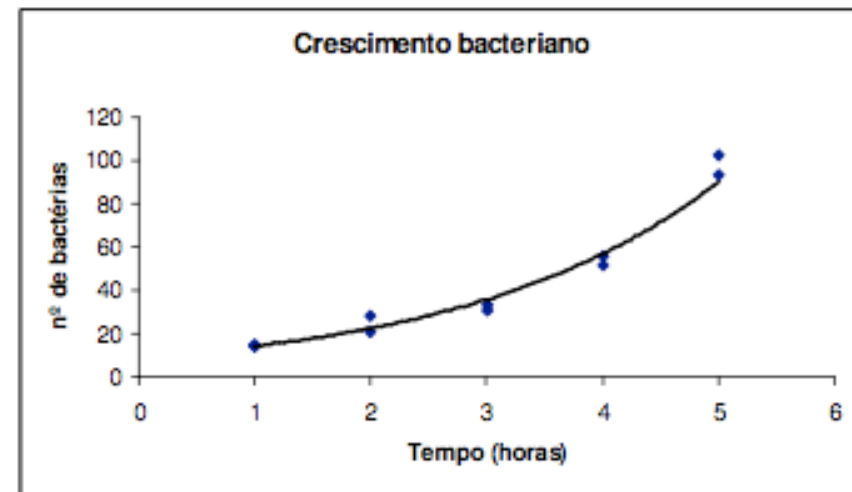
### Regressão Linear Simples

Relação estatística entre duas variáveis

Diagrama de dispersão



Curva ajustada



Relação NÃO Linear entre as variáveis

# Avaliação de correlações

## Correlação e Regressão

### Regressão Linear Simples

Modelo de regressão simples:

- A cada valor da variável independente  $X$  está associada uma distribuição de probabilidades  $Y$
- As médias destas distribuições de probabilidade varia de forma sistemática com  $X$

# Avaliação de correlações

## Coeficiente de correlação linear

Avalia se existe associação entre duas características quantitativas

Quando se constata que duas variáveis quantitativas variam juntas, diz-se que estão correlacionadas

ALUNO	X (horas)	Y (nota)
A	8	10
B	7	8
C	6	4
D	3	8
E	3	6
F	6	9
G	5	7
H	2	4

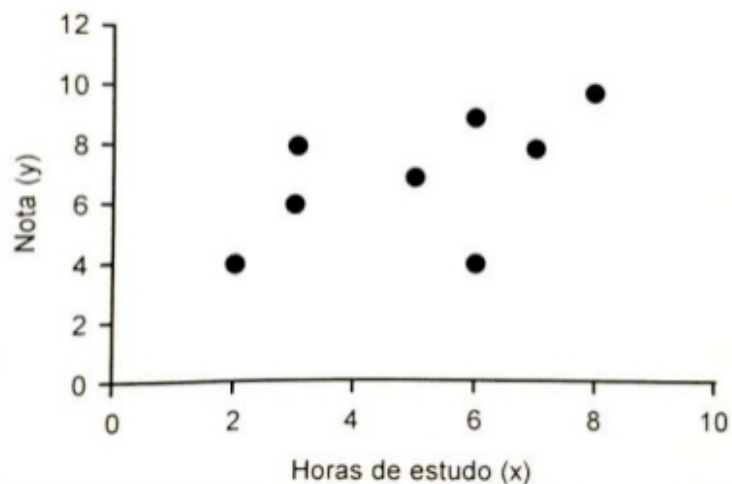


Diagrama de dispersão

# Avaliação de correlações

## Coeficiente de correlação linear

- Para avaliar a correlação utiliza-se um coeficiente
- Mede a intensidade da associação existente entre duas variáveis quantitativas
- Foi proposto por Karl Pearson em 1896 - coeficiente de correlação de Pearson ( $r$ )
- $r$  pode variar entre -1 e +1
  - ✓ Valores de  $r$  negativos = correlação inversa
  - ✓ Valores de  $r$  positivos = correlação direta

# Avaliação de correlações

## Coeficiente de correlação linear

### Intensidade da correlação de Pearson

$ r $	Intensidade
0	nula
0 – 0,3	Fraca
0,3  – 0,6	Regular
0,6  – 0,9	Forte
0,9  – 1	Muito forte
1	Plena ou perfeita

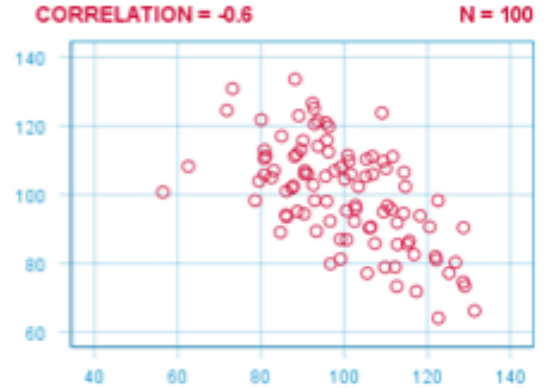
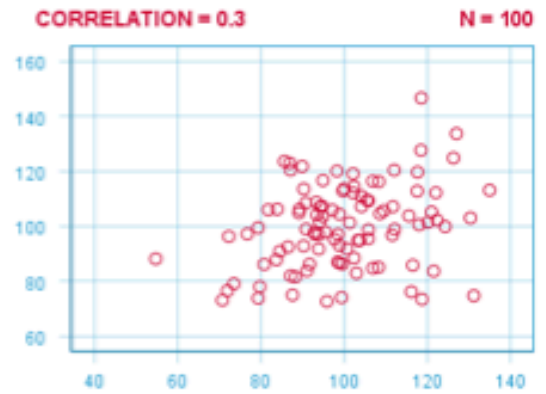
# Avaliação de correlações

## Intensidade da correlação

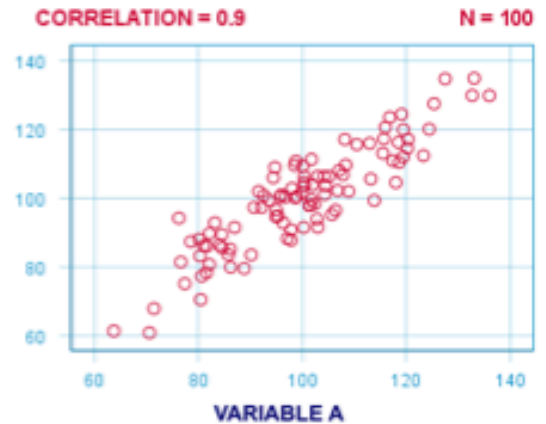
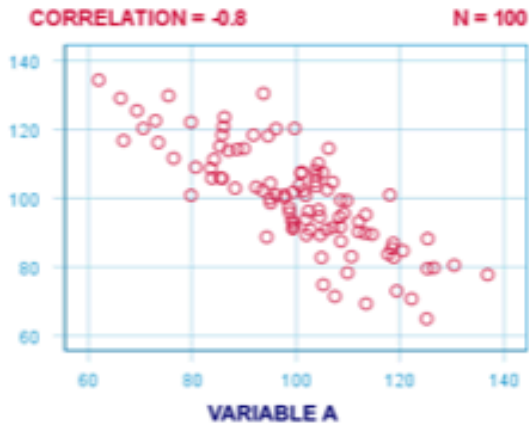
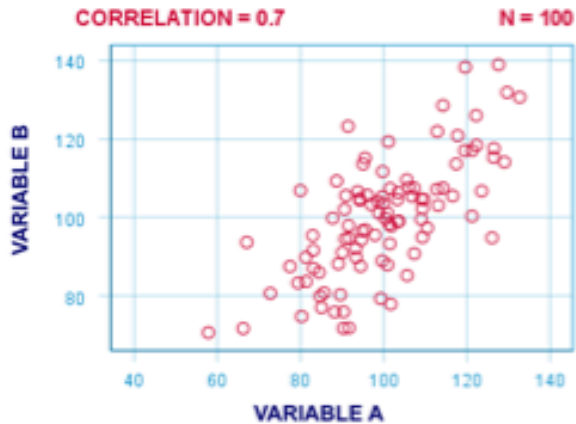
- Valores máximos de correlação: reta inclinada
- Correlação nula: nuvens circulares ou nuvem elíptica paralela a um dos eixos do gráfico
- Valores intermédios de correlação: nuvens elípticas inclinadas (quanto mais estreitas, maior a correlação)
- Situações especiais: pontos formam uma nuvem cujo o eixo principal é uma curva. Solução: análise não paramétrica

# Avaliação de correlações

# Intensidade da correlação

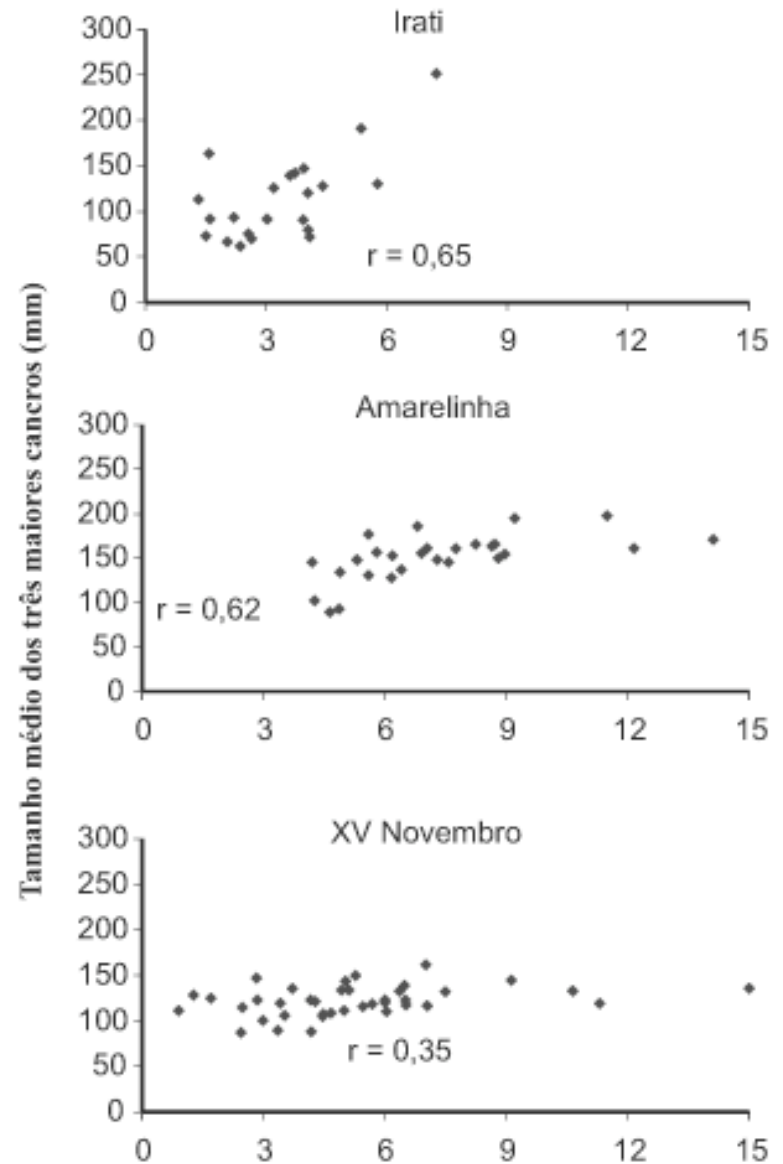
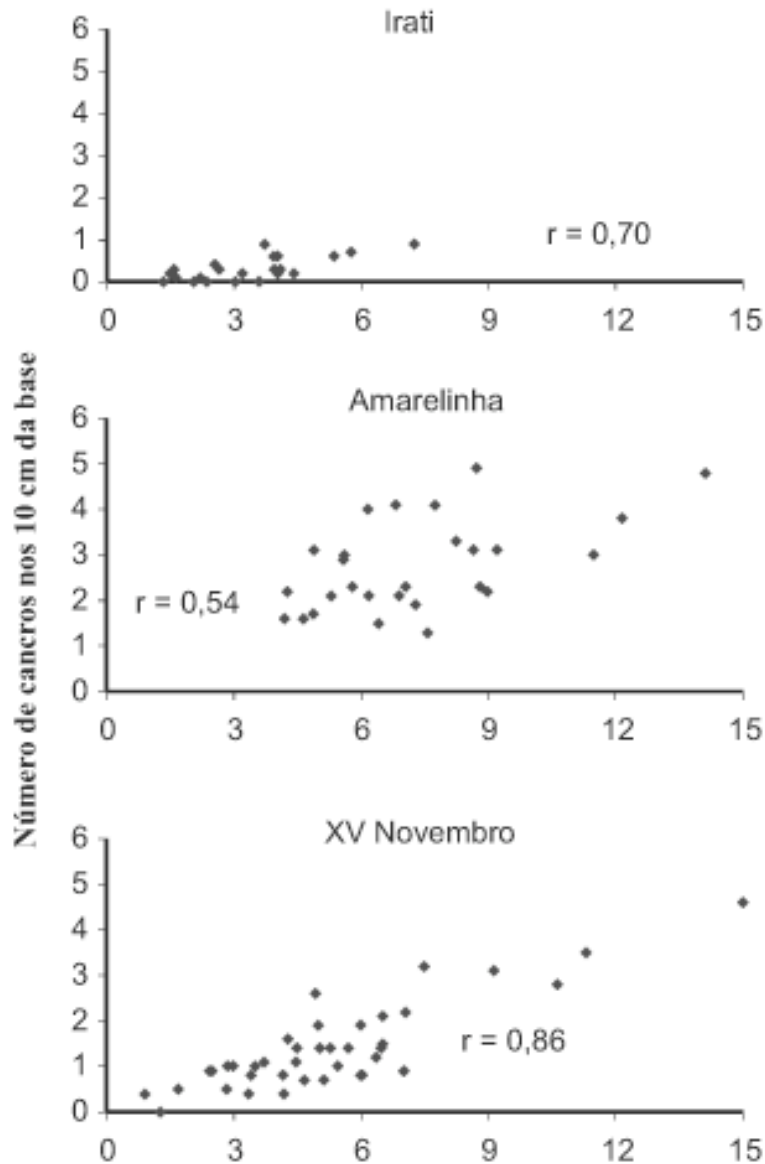


[www.spss-tutorials.com](http://www.spss-tutorials.com)



# Avaliação de correlações

# Intensidade da correlação



# Avaliação de correlações

## Coeficiente de correlação linear $r$

- Não há necessidade de satisfazer nenhuma preposição para calcular o  $r$  entre duas variáveis quantitativas
- Os pressupostos aplicam-se unicamente à realização do teste estatístico

### Pressupostos

1. As variáveis têm uma distribuição normal
2. A variância de  $X$  é a mesma para vários valores de  $Y$
3. A variância de  $Y$  é a mesma para vários níveis de  $X$

# Avaliação de correlações

## Teste de Hipóteses sobre a correlação

$r$  da amostra é uma estimativa da verdadeira correlação entre X e Y existente na população

➤ Elaboração de Hipóteses:

$H_0$ : correlação = 0 ;  $H_a$ : correlação diferente de 0

➤ Escolha do nível de significância :  $\alpha=0,05$

➤ Verificar valor de Sig., se  $p > 0,05 \rightarrow$  Não e rejeita  $H_0$

# Avaliação de correlações

## Coeficiente de correlação linear

## Coeficiente de determinação

- $r^2$  é o quadrado do coeficiente de correlação
- Informa que fração da variabilidade de uma característica é explicada estatisticamente pela outra variável

# Avaliação de correlações

## Coeficiente de correlação linear

### Em Resumo:

- $r$  mede uma associação e não uma relação causa efeito
- Pode haver outros factores que determinam o nível de correlação das duas variáveis independentemente
- Pode haver correlação fraca e estatisticamente significativa (i.e.:  $n=900$ ;  $r=0,15$  ;  $p < 0,001$ )

# Avaliação de correlações

## Coeficiente de correlação linear

## Coeficiente de correlação de Spearman (1904)

Utilizado para avaliar o grau de correlação entre variáveis quantitativas quando as exigências para o teste de Pearson não são satisfeitas

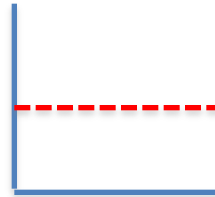
Designa-se por  $r_s$

# Avaliação de correlações

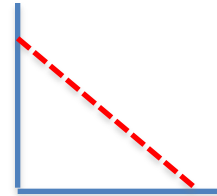
Coeficiente de correlação linear

Coeficiente de correlação de Spearman

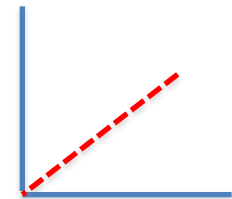
$r_s = 0 \rightarrow$  Ausência de correlação



$r_s = -1 \rightarrow$  Correlação negativa perfeita



$r_s = +1 \rightarrow$  Correlação positiva perfeita



- Cálculo de  $r_s$  baseia-se nas diferenças entre os valores de X e Y.
- Consiste em averiguar quantas vezes é que os valores estão 'empatados'

# Avaliação de correlações

- ✓ Avalia a correlação entre duas variáveis quantitativas ou ordinais (com um número significativo de categorias: aconselhado  $\leq$  a 5), determinando o coeficiente de **correlação de Pearson** ou de **Spearman** através de testes de significância.

Exemplo:

Registaram-se relativamente a 30 pacientes, as suas idades, níveis de colesterol e frequências cardíacas máximas.

**Questão 1:** Os níveis de colesterol estão relacionados com a frequência cardíaca máxima?

**Questão 2:** Existe alguma relação entre a idade e os níveis de colesterol?

# Avaliação de correlações

## Testes

- Coeficiente de correlação de Pearson e teste à sua significância:  
Distribuição Normal
- Coeficiente de correlação de Spearman e teste à sua significância:  
Distribuição não normal

**Questão 1:** Os níveis de colesterol estão relacionados com a frequência cardíaca máxima?

- Estudo da Normalidade:

Colesterol:  $p=0,965$

Grupo pacientes:  $p=0.501$

As variáveis seguem uma distribuição normal → **Coeficiente de correlação de Pearson**

# Avaliação de correlações

## ➤ Coefficiente de correlação de Pearson

Analyze > Correlate > Bivariate

Colocam-se as variáveis Colestrol e Freq.Cardíaca no campo Variables

		Colesterol	Frequência cardíaca máxima
Colesterol	Pearson Correlation	1	<b>.687**</b>
	Sig (2-tailed)		<b>.000</b>
	N	30	30
Frequência cardíaca máxima	Perason Correlation	<b>.687**</b>	1
	Sig (2-tailed)	<b>.000</b>	
	N	30	30

\*\* Correlation is significant at the 0.01 level (2-tailed)

Verificou-se que existe uma correlação estatisticamente significativa entre o colesterol e a frequências cardíaca máxima, sendo esta positiva e forte ( $r(30)=0.687$ ,  $p<0.001$ )

# Avaliação de correlações

**Questão 2:** Existe alguma relação entre a idade e os níveis de colesterol?

➤ Estudo da Normalidade:

Idade:  $p=0,010$

Colesterol:  $p=0.965$

A variável Idade não seguem uma distribuição normal → **Coefficiente de correlação de Spearman**

*Analyze > Correlate > Bivariate*

Colocam-se as variáveis *Idade* e *Colestreol* no campo *Variables*

Selecionar a opção *Spearman* em vez de *Pearson*

# Avaliação de correlações

**Questão 2:** Existe alguma relação entre a idade e os níveis de colesterol?

## Coeficiente de correlação de Spearman

			Correlations	
			Idade	Colesterol
Spearman's rho	Idade	Correlation Coefficient	1.000	.068
		Sig (2-tailed)	-	<b>.721</b>
		N	30	30
	Colesterol	Correlation Coefficient	.068	1.000
		Sig (2-tailed)	.721	-
		N	30	30

Verificou-se que não existe uma correlação estatisticamente significativa entre a idade e o colesterol ( $\rho(30)=0.068$  e  $p=0.721$ )