

Investigação Aplicada I

Aula 8

1º Semestre 2016/17

Licenciatura em Ciências Biomédicas Laboratoriais

igrodrigues@ualg.pt; ESSUAlg: gabinete 2.06

Prof. Inês Rodrigues

Inferência estatística

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

Inferência estatística

Regressão Linear Simples

Relação entre duas variáveis

Funcional:

A cada valor de X corresponde um único valor de Y

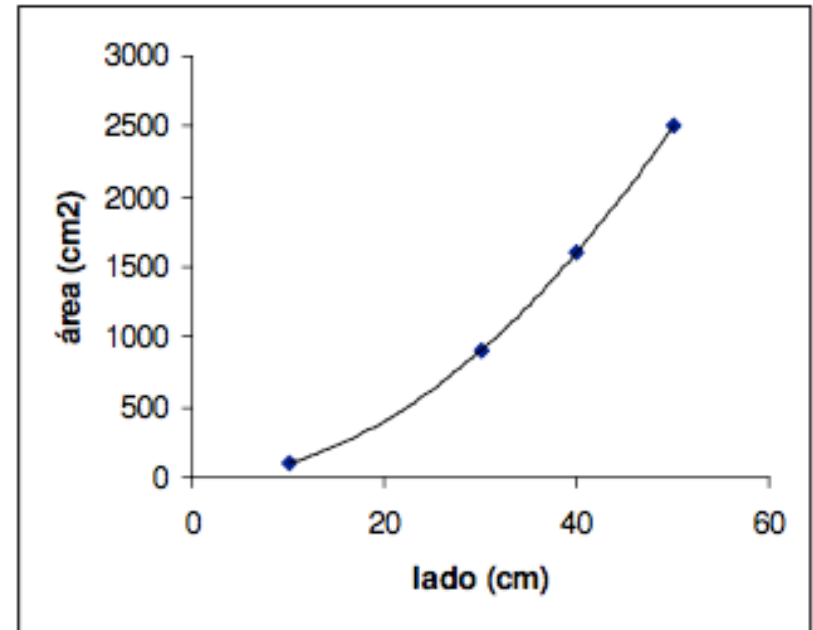
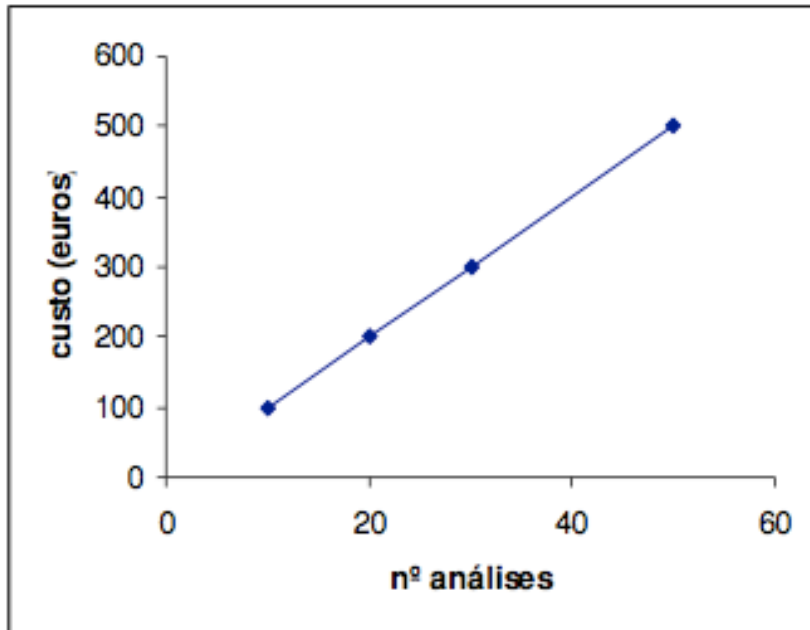
Estatística:

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

Inferência estatística

Regressão Linear Simples

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

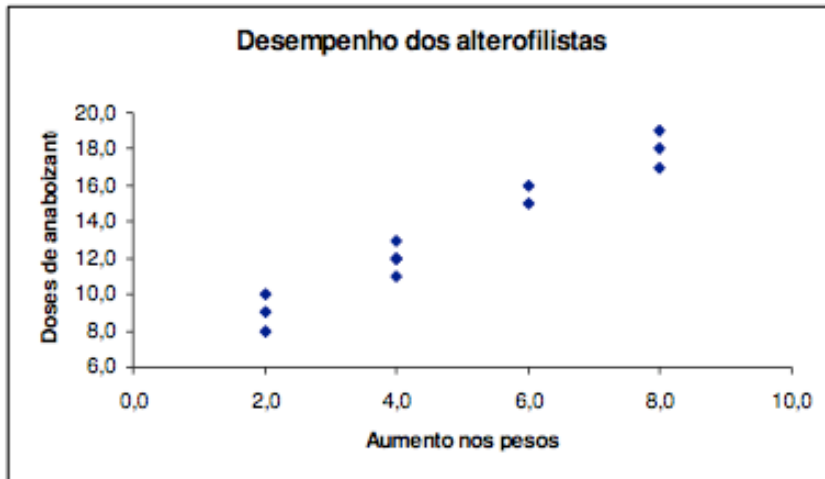


Inferência estatística

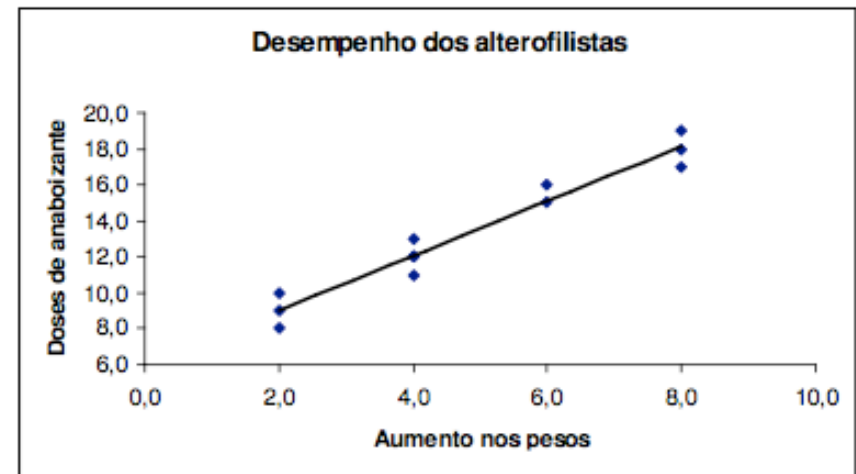
Regressão Linear Simples

Relação estatística entre duas variáveis

Diagrama de dispersão



Recta ajustada



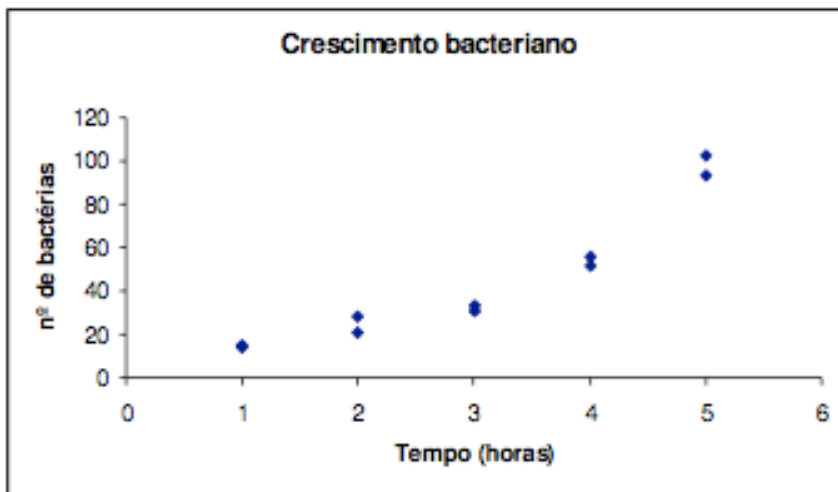
Relação Linear entre as variáveis

Inferência estatística

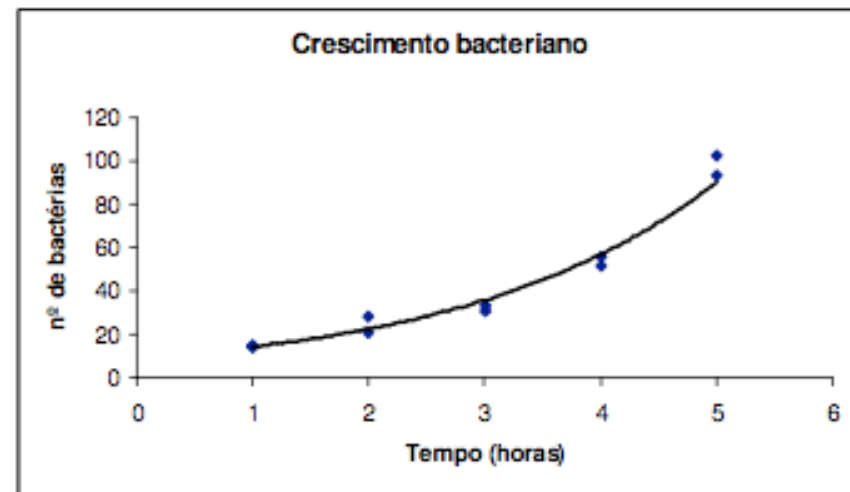
Regressão Linear Simples

Relação estatística entre duas variáveis

Diagrama de dispersão



Curva ajustada



Relação NÃO Linear entre as variáveis

Inferência estatística

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

Inferência estatística

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 |

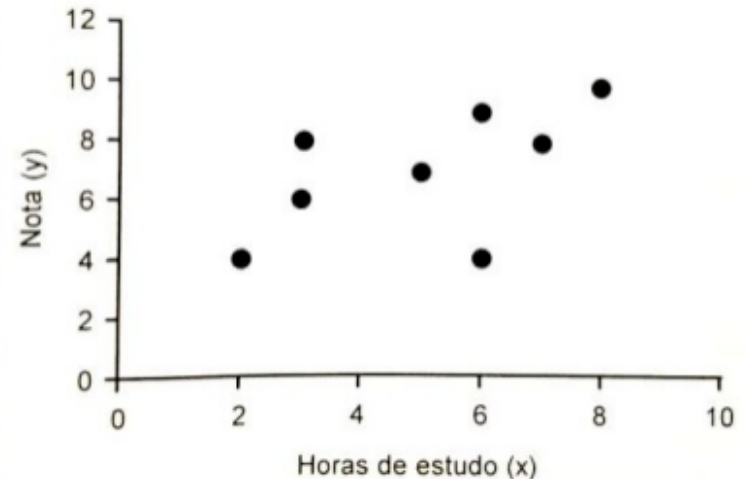


Diagrama de dispersão

Inferência estatística

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

Inferência estatística

Coeficiente de correlação linear

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

Inferência estatística

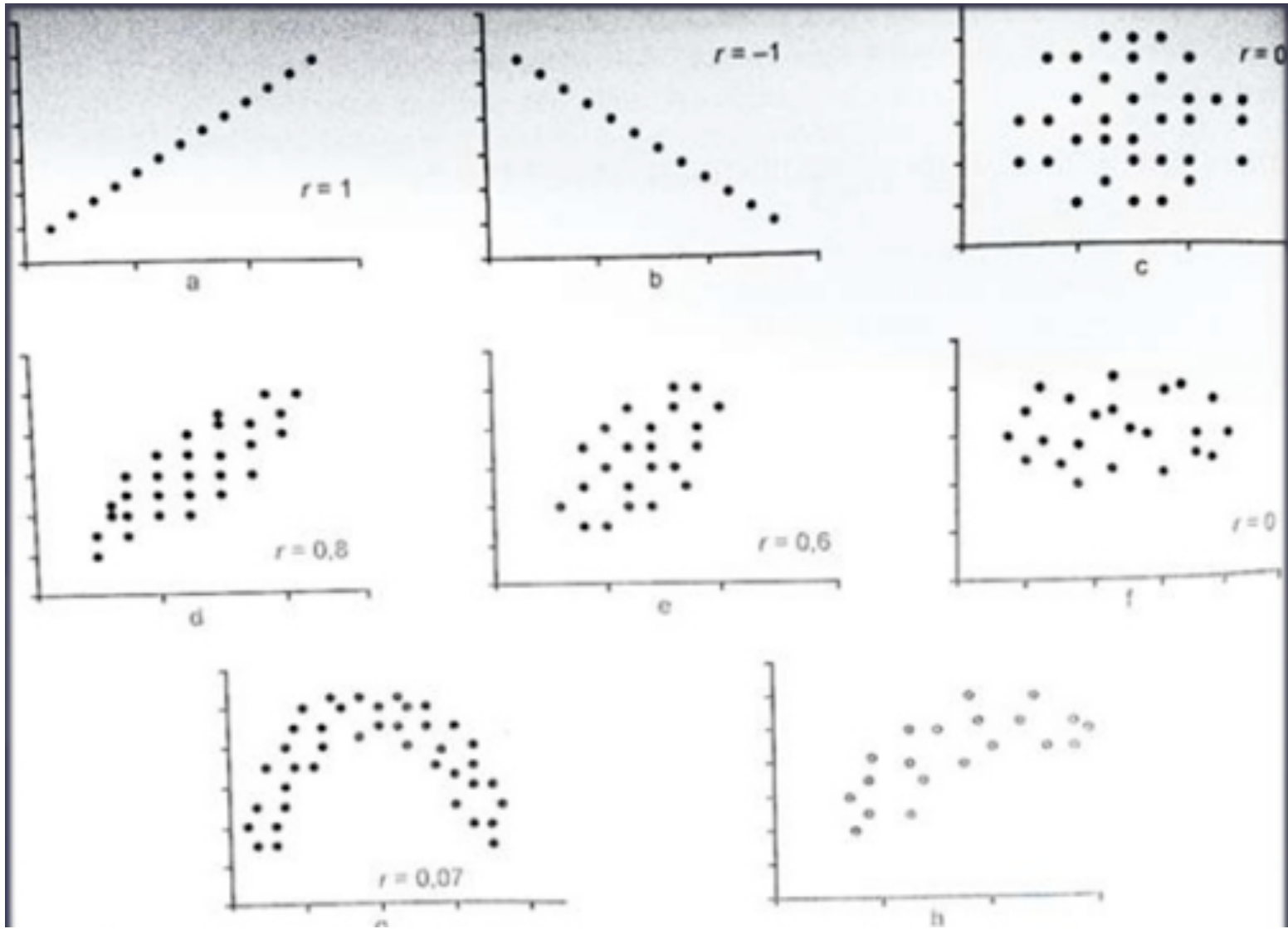
Coeficiente de correlação linear

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 |

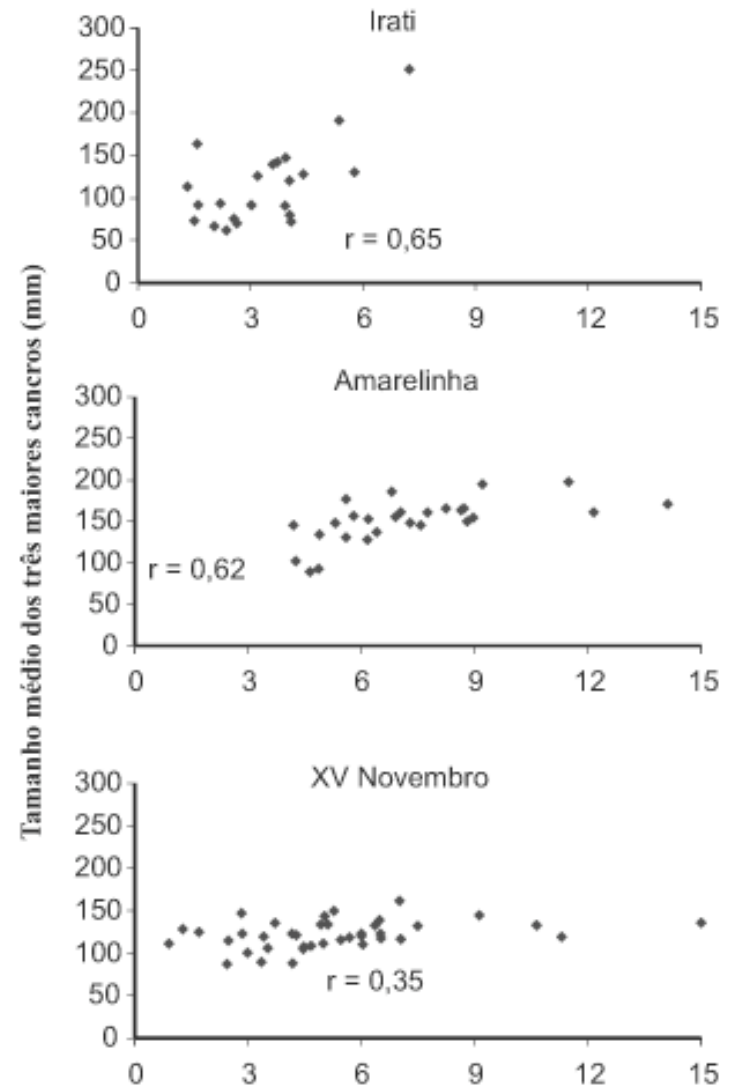
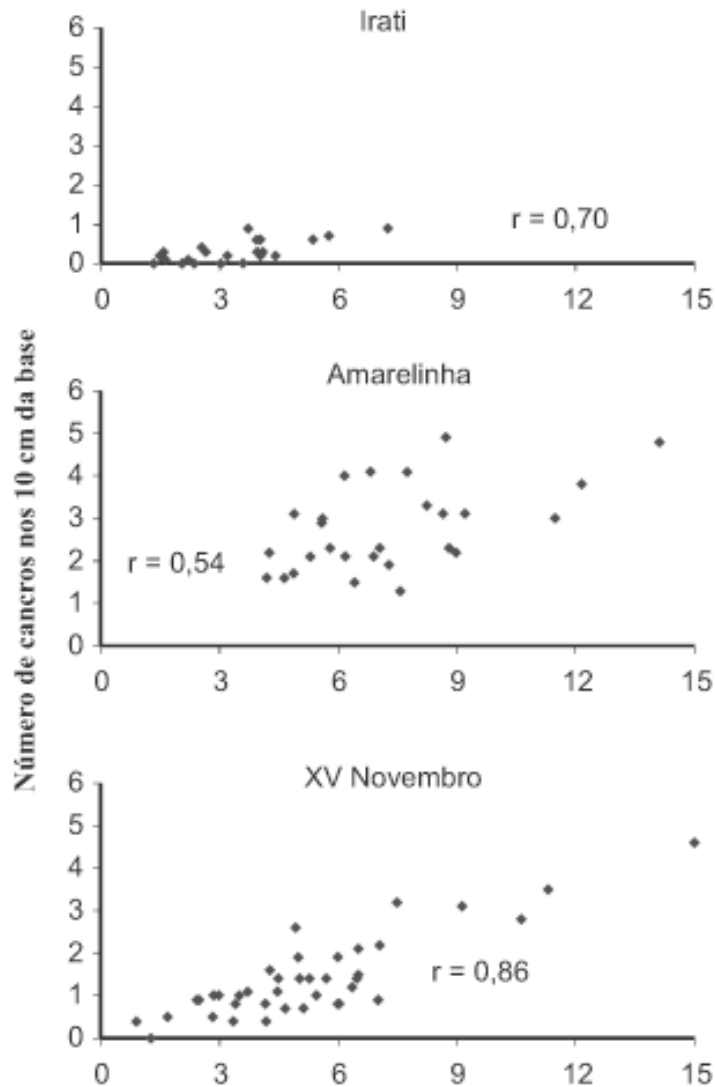
Inferência estatística

Coeficiente de correlação linear



Inferência estatística

Coeficiente de correlação linear



Inferência estatística

Coeficiente de correlação linear

$$\text{Covariação} = (X_i - E\{\bar{X}\})(Y_i - E\{\bar{Y}\})$$

$$r = \frac{\text{COV}_{xy}}{S_x S_y}, \text{ onde } \text{cov}_{xy} = \sum \frac{(x - \bar{x})(y - \bar{y})}{n - 1}$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \frac{SP_{xy}}{\sqrt{SQ_x SQ_y}}$$

$$r = 0,58$$

Inferência estatística

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 á 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

Inferência estatística

Coeficiente de correlação linear

Teste de Hipóteses sobre a correlação

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

1. Elaboração de Hipóteses: H_0 : correlação = 0 ; H_a : correlação diferente de 0
2. Escolha do nível de significância : $\alpha=0,05$
3. Determinação do valor crítico do teste
4. Consultar a tabela t para saber o valor de p
5. $p= 2,45$ (superior a 0,05) \rightarrow Não e rejeita H_0

Inferência estatística

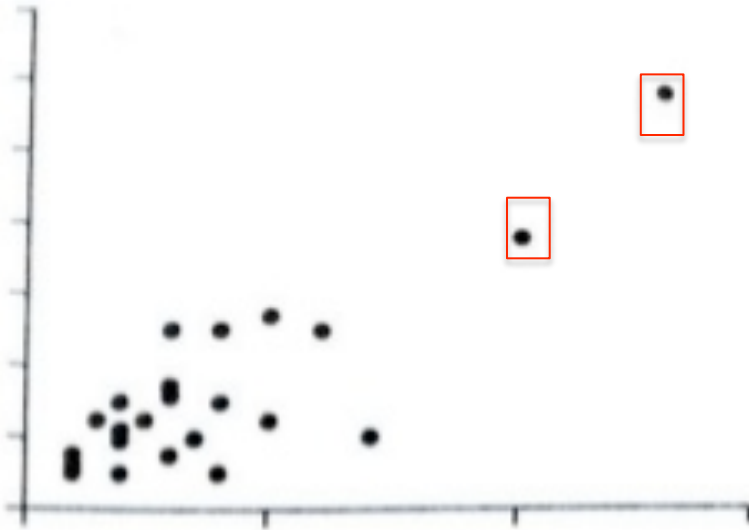
Coeficiente de correlação linear

Coeficiente de determinação

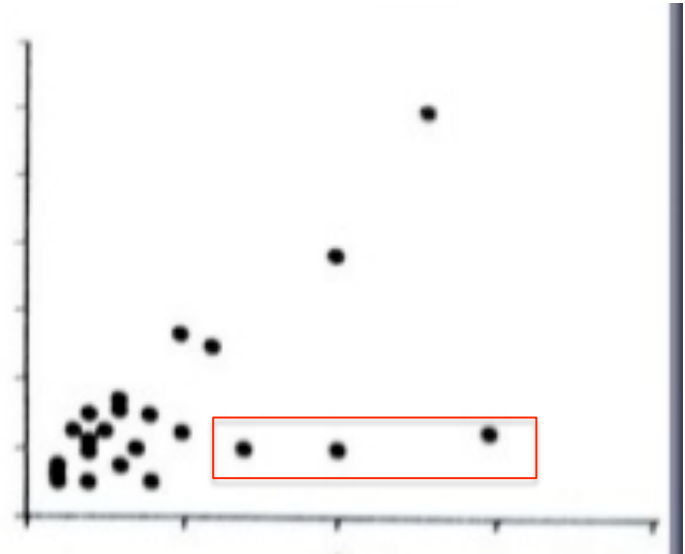
- r^2 é o quadrado do coeficiente de correlação
- Informa que fração da variabilidade de uma característica é explicada estatisticamente pela outra variável

Inferência estatística

Coeficiente de correlação linear



$r = 0,84$ (todos os pontos)
 $r = 0,46$ (excluindo *outliers*)



A variância de X NÃO é a mesma
para vários valores de Y

Inferência estatística

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.: $n=900$; $r=0,15$; $p < 0,001$)

Inferência estatística

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

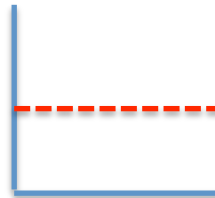
Designa-se por r_s

Inferência estatística

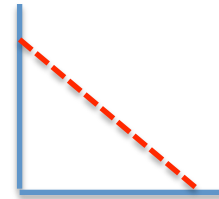
Coeficiente de correlação linear

Coeficiente de correlação de Spearman

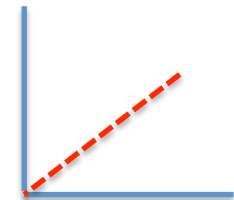
$r_s = 0 \rightarrow$ Ausência de correlação



$r_s = -1 \rightarrow$ 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'

Exercício Exemplo

Um estudo clínico foi desenvolvido para avaliar as diferenças nos níveis de albumina em adultos após terem sido submetidos a dietas com diferentes quantidades de proteína. Dietas de baixa proteína são frequentemente prescritas a doentes com insuficiência renal. A albumina é a proteína mais abundante no sangue e a sua concentração no soro é medida em gramas por decilitro (g / dl).

Compararam-se três dietas em função da percentagem de proteína. Os níveis de albumina dos participantes, registados após cada dieta foram:


| | | | | | | | |
|--------------|-----|-----|-----|-----|------|-----|-----|
| Proteína 5%: | 3,1 | 2,6 | 2,9 | 2,7 | 8,4 | 4,5 | 1,2 |
| | 8,7 | 4,5 | 6,9 | 4,7 | 8,8 | 2,5 | 5,6 |
| | 4,7 | 8,1 | 1,1 | 7,5 | 14,0 | | |

| | | | | | | | |
|---------------|-----|-----|-----|-----|-----|-----|-----|
| Proteína 10%: | 3,8 | 4,1 | 2,9 | 3,4 | 4,2 | 1,7 | 7,8 |
| | 8,8 | 6,9 | 6,6 | 3,1 | 1,2 | 0,8 | 7,8 |

| | | | | | | | |
|---------------|-----|------|-----|-----|-----|-----|-----|
| Proteína 15%: | 4,0 | 5,5 | 5,0 | 4,8 | 4,4 | 8,7 | 9,5 |
| | 5,5 | 2,1 | 1,3 | 3,1 | 5,4 | 4,4 | 6,9 |
| | 1,2 | 12,2 | 1,1 | 7,1 | 4,4 | 5,8 | 2,2 |

Existirão diferenças nos níveis médios de albumina sérica nas três dietas? Justifique.


Exercício Exemplo

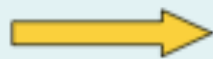
Correlação  permite-nos saber se existe ou não uma relação linear entre duas variáveis.




medir o grau da relação

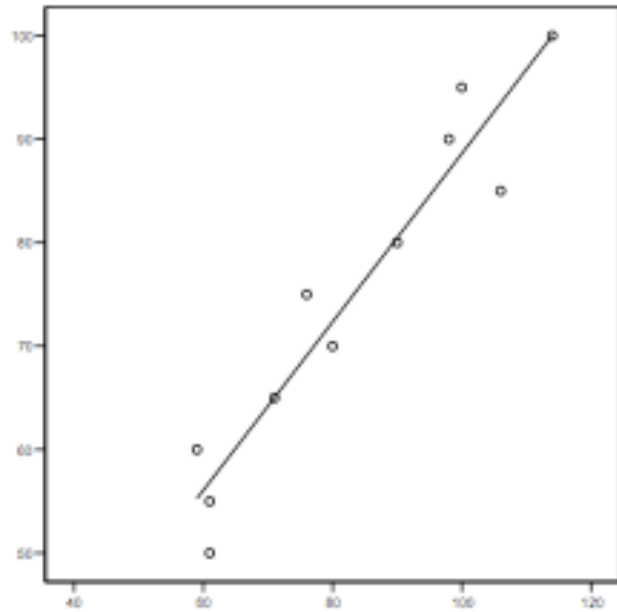
Coeficiente de correlação - ρ ($-1 \leq \rho \leq 1$)

Se $\rho > 0$  relação linear directa entre as variáveis

Se $\rho < 0$  relação linear inversa entre as variáveis

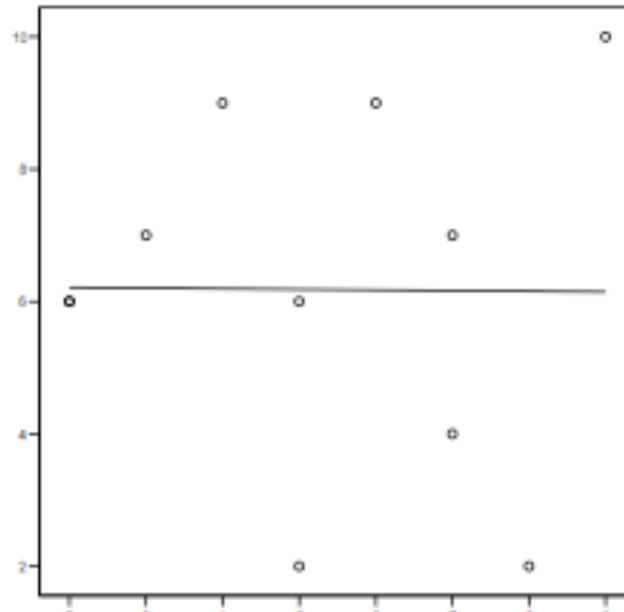
Se $\rho = 0$  independência linear entre as variáveis

Exercício Exemplo



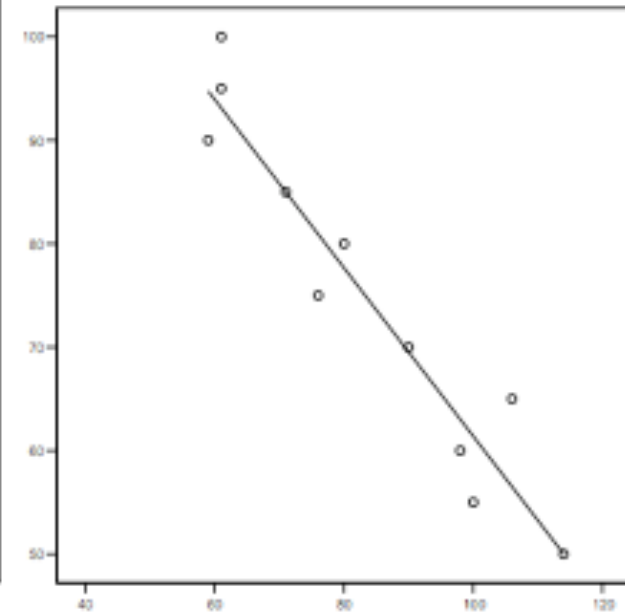
$$\rho > 0$$

**Relação Linear
Direta**



$$\rho = 0$$

**Independência
Linear**



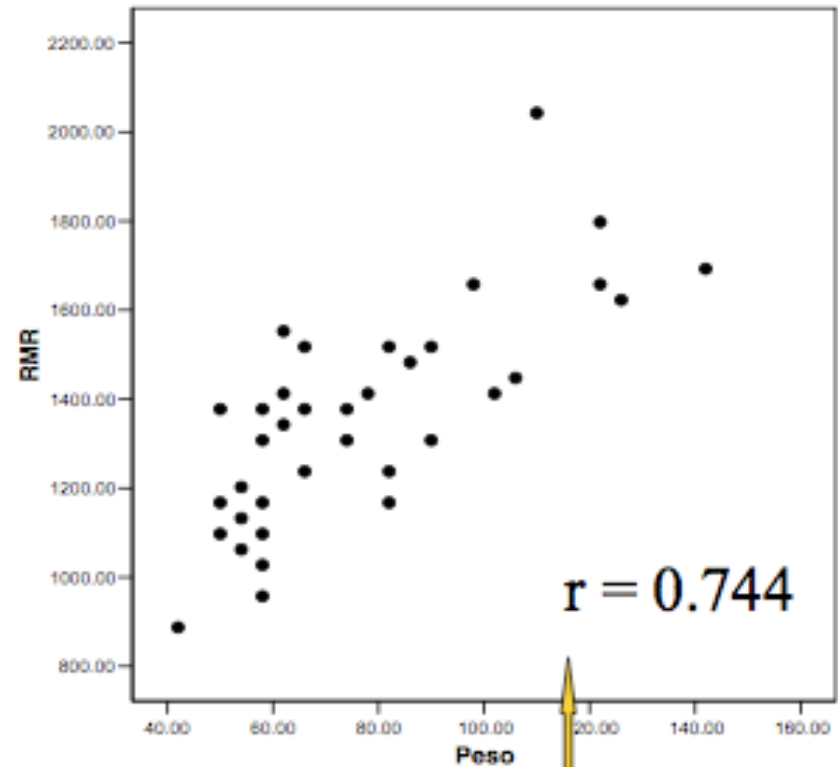
$$\rho < 0$$

**Independência
Inversa**

Exercício Exemplo

Exemplo:

Estudo da relação entre o peso (kg) e o índice metabólico em repouso (RMR) (kcal/24 hr)



Correlations

| | | Body weight | Resting metabolic rate |
|------------------------|---------------------|-------------|------------------------|
| Body weight | Pearson Correlation | 1 | .744** |
| | Sig. (2-tailed) | . | .000 |
| | N | 44 | 44 |
| Resting metabolic rate | Pearson Correlation | .744** | 1 |
| | Sig. (2-tailed) | .000 | . |
| | N | 44 | 44 |

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

Estimativa pontual
de ρ

Testes de Correlação

- ❖ Coeficiente de correlação de Pearson (Normalidade)
 - ❖ Coeficiente de correlação de Spearman
 - ❖ Coeficiente de correlação de Kendall
- } não paramétricos

Exercício 3