

# Estatística Descritiva III

Ciências Contábeis - FEA - Noturno

1º Semestre 2025

Profs. Leonardo T. Rolla e Nikolai Kolev

(baseado em material previamente desenvolvido  
pelo Prof. Gilberto Alvarenga Paula – CC BY-SA 4.0)

- 1 Informações sobre esta unidade
- 2 Tabelas de Contingência
- 3 Diagrama de Dispersão
- 4 Correlação Linear
- 5 Exercícios feitos em classe

Compreender e aplicar técnicas de análise estatística, incluindo tabelas de contingência, diagramas de dispersão e correlação linear, para explorar relações entre variáveis e facilitar a interpretação de dados.

Depois desta unidade, o aluno será capaz de:

- 1 Construir e interpretar tabelas de contingência para variáveis categóricas.
- 2 Construir diagramas de dispersão para visualizar relações entre variáveis numéricas.
- 3 Calcular e interpretar a correlação linear entre duas variáveis numéricas.

- 1 Interpretação de tabelas de contingência
- 2 Diagrama de dispersão: construção e análise
- 3 Correlação linear: conceito, cálculo e interpretação
- 4 Identificação de relações entre variáveis a partir do diagrama de dispersão
- 5 Interpretação da força e direção da correlação linear

O aluno pode fazer os exercícios postados no Moodle e posteriormente comparar suas respostas com as soluções apresentadas, identificando possíveis erros, e eventualmente refazer os exercícios.

Além disso, é recomendável refletir sobre as dificuldades encontradas e revisar os conceitos conforme necessário.

- 1 Informações sobre esta unidade
- 2 Tabelas de Contingência**
- 3 Diagrama de Dispersão
- 4 Correlação Linear
- 5 Exercícios feitos em classe

## Definição

Tabelas de contingência são utilizadas para descrever a distribuição (em frequências ou em porcentagens) de duas ou mais **variáveis qualitativas**.

## Exemplo: Hábito de fumar por faixa etária

Tabela de frequências absolutas:

	Sim	Não
20-30	13	25
30-40	38	10
40-50	20	5

## Exemplo: Hábito de fumar por faixa etária

Cálculo do total por variável e total geral

	Sim	Não	Total
20-30	13	25	38
30-40	38	10	48
40-50	20	5	25
Total	71	40	111

## Exemplo: Hábito de fumar por faixa etária

Cálculo do percentual por linhas:

	Sim	Não	Total
20-30	13	25	38
30-40	38	10	48
40-50	20	5	25
Total	71	40	111

	Sim	Não	Total
20-30	34	66	100
30-40	79	21	100
40-50	80	20	100
Total	64	36	100

Na faixa 20-30 há muito mais não-fumantes do que o esperado (66%, comparado com o esperado que seria 36%)

Na faixa de 30-40 há muito mais fumantes do que o esperado (79%, comparado com o esperado que seria 64%)

Na faixa de 40-50 há muito mais fumantes do que o esperado (80%, comparado com o esperado que seria 64%)

## Exemplo: Hábito de fumar por faixa etária

Cálculo do percentual por colunas:

	Sim	Não	Total
20-30	13	25	38
30-40	38	10	48
40-50	20	5	25
Total	71	40	111

	Sim	Não	Total
20-30	18	62	34
30-40	54	25	43
40-50	28	12	23
Total	100	100	100

Entre os fumantes, há muito mais indivíduos nas faixas 30-40 e 40-50 do que o esperado (54% e 28%, comparado com o esperado que seria 43% e 23%)

Entre os não-fumantes há muito mais em 20-30 do que o esperado (62%, comparado com o esperado que seria 34%)

- 1 Informações sobre esta unidade
- 2 Tabelas de Contingência
- 3 Diagrama de Dispersão**
- 4 Correlação Linear
- 5 Exercícios feitos em classe

## Definição

Os diagramas de dispersão são representações dos valores observados de duas **variáveis quantitativas** num mesmo gráfico. Um dos objetivos desse tipo de gráfico é avaliar a possível existência de alguma relação funcional entre as variáveis em estudo.

Por exemplo, verificar se há **relação linear** ou alguma **relação não linear** entre as duas variáveis quantitativas.

## Exemplos

- idade e altura das crianças
- tempo de prática de esportes e ritmo cardíaco
- tempo de estudo e nota na prova
- taxa de desemprego e taxa de criminalidade
- expectativa de vida e taxa de analfabetismo

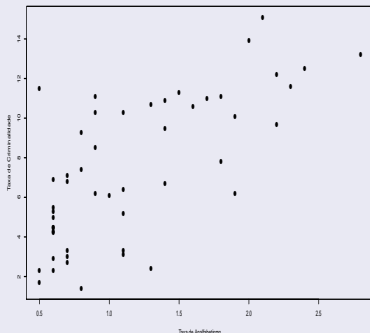
### Descrição das Variáveis

Vamos considerar como ilustração a relação entre

- Taxa de Analfabetismo (em %)
- Taxa de Criminalidade (por 100.000 habitantes)

nos 50 estados norte americanos no início dos anos 70.

## Diagrama de Dispersão



Podemos notar que, conforme aumenta a **Taxa de Analfabetismo**, a **Taxa de Criminalidade** tende a aumentar. Nota-se também uma tendência linear.

### Descrição das Variáveis

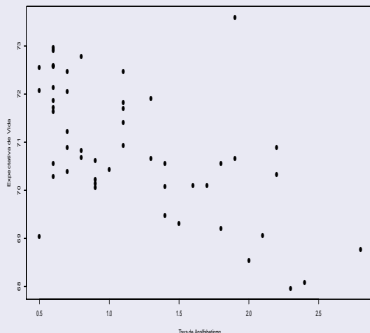
Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)
- Y: Expectativa de Vida (em anos)

nos 50 estados norte americanos no início dos anos 70.

# Exemplo: Expectativa de Vida e Analfabetismo

## Diagrama de Dispersão



- 1 Informações sobre esta unidade
- 2 Tabelas de Contingência
- 3 Diagrama de Dispersão
- 4 Correlação Linear**
- 5 Exercícios feitos em classe

## Definição

O **coeficiente de correlação linear de Pearson** é uma medida resumo que quantifica a força da relação linear entre duas **variáveis quantitativas**.

## Definição

O **coeficiente de correlação linear de Pearson** é expresso na seguinte forma:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_x \hat{\sigma}_y}$$

ou, em termos de  $s_x$  e  $s_y$ :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)s_x s_y}.$$

Lembrete:

- $\bar{x}$  e  $\bar{y}$  denotam as médias amostrais
- $s_x$  e  $s_y$  denotam os desvios padrão amostrais

## Fórmula alternativa

As seguintes fórmulas podem ser mais práticas para cálculos à mão:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}$$

ou

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{(n-1) s_x s_y}.$$

## Propriedades

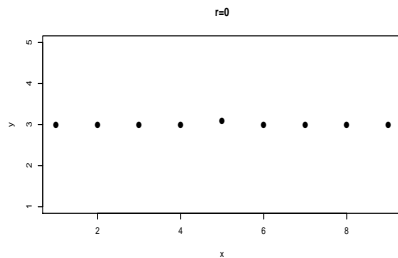
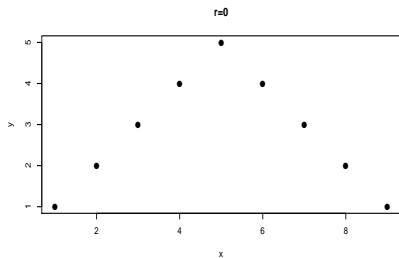
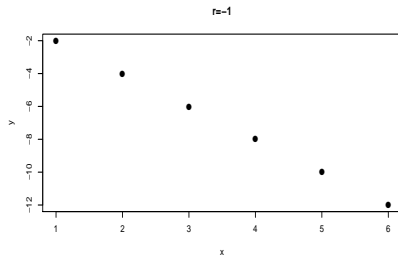
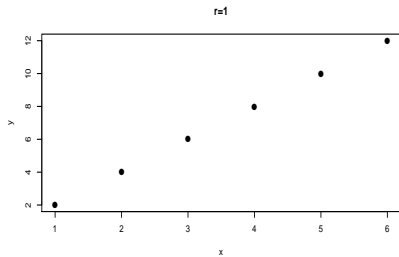
O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

$$-1 \leq r \leq 1.$$

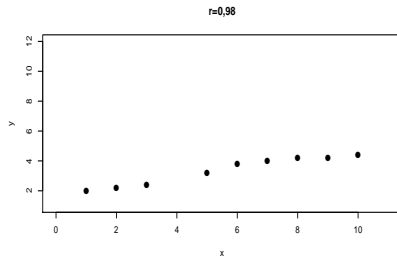
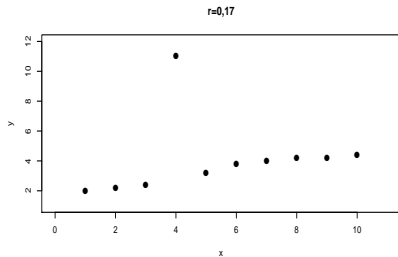
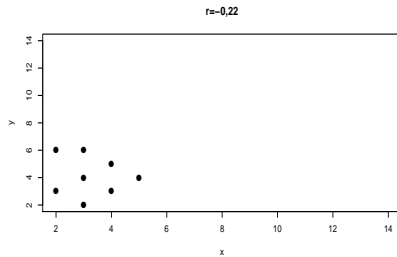
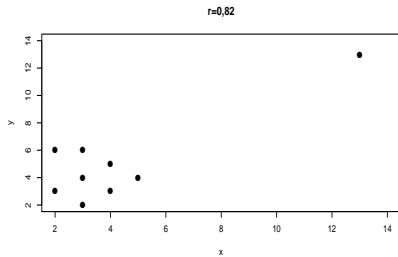
## Casos particulares

- $r = 1$ : correlação linear positiva e perfeita
- $r = -1$ : correlação linear negativa e perfeita
- $r = 0$ : ausência de correlação linear

# Interpretação



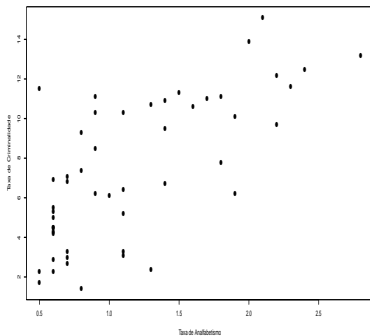
# Interpretação



## Observações

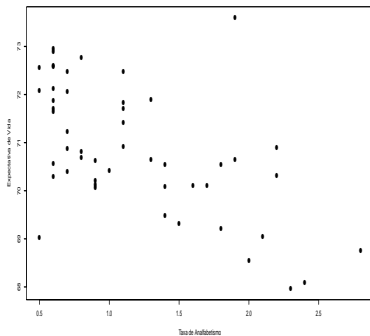
- Correlação linear próxima de zero significa ausência de **relação linear** entre as variáveis quantitativas, contudo pode haver **outro tipo de relação** entre as variáveis
- Pontos extremos no diagrama de dispersão podem influenciar o valor do coeficiente de correlação linear. Assim, deve-se olhar conjuntamente o diagrama de dispersão e o coeficiente de correlação linear

# Exemplo: Criminalidade e Analfabetismo



$$r = 0,702$$

# Exemplo: Expectativa de Vida e Analfabetismo



$$r = -0,59$$

- 1 Informações sobre esta unidade
- 2 Tabelas de Contingência
- 3 Diagrama de Dispersão
- 4 Correlação Linear
- 5 Exercícios feitos em classe**

## Execício: preferência de câmbio vs gênero

Com base na tabela de frequência entre Gênero e Preferência, calcule os percentuais por linha e por coluna. O que eles revelam sobre a distribuição das preferências entre os grupos?

**Tabela:** Tabela de frequências entre Gênero e Preferência

Gênero	Preferência		Total
	automática	manual	
Feminino	158	56	214
Masculino	160	147	307
Total	318	203	521

## Execício: preferência de câmbio vs gênero

Tabela: Tabela de frequências entre Gênero e Preferência

Gênero	Preferência		Total
	automática	manual	
Feminino	158 (74%)	56 (26%)	214 (100%)
Masculino	160 (52%)	147 (48%)	307 (100%)
Total	318 (61%)	203 (39%)	521 (100%)

## Execício: preferência de câmbio vs gênero

Tabela: Tabela de frequências entre Gênero e Preferência

Gênero	Preferência		Total
	automática	manual	
Feminino	158 (50%)	56 (28%)	214 (41%)
Masculino	160 (50%)	147 (72%)	307 (59%)
Total	318 (100%)	203 (100%)	521 (100%)

### Problema

Os dados abaixo correspondem aos valores do comprimento (em cm) da asa e a idade (em dias) de 13 pardais. O objetivo é investigar a possível associação entre essas medidas após a eclosão.

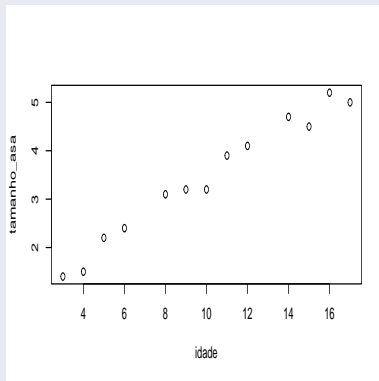
## Exercício: Asa dos Pardais

Idade (dias)	Tamanho da asa (cm)
3	1.4
4	1.5
5	2.2
6	2.4
8	3.1
9	3.2
10	3.2
11	3.9
12	4.1
14	4.7
15	4.5
16	5.2
17	5.0

## Exercício: Asa dos Pardais

- (a) Construa manualmente o diagrama de dispersão entre idade e tamanho da asa. O que o gráfico sugere?
- (b) Calcule o coeficiente de correlação linear de Pearson entre as variáveis.

### Visualização



### Interpretação

Observa-se que, conforme a **idade** aumenta, o **tamanho da asa** também tende a crescer, sugerindo uma relação linear positiva.

### Fórmula

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}$$

$$s_x^2 = \frac{\sum_{i=1}^n x_i^2 - n \bar{x}^2}{(n-1)}$$

## Exercício: Asa dos Pardais

$x$	$y$	$x^2$	$y^2$	$xy$
3	1.4	9	1.96	4.2
4	1.5	16	2.25	6
5	2.2	25	4.84	11
6	2.4	36	5.76	14.4
8	3.1	64	9.61	24.8
9	3.2	81	10.24	28.8
10	3.2	100	10.24	32
11	3.9	121	15.21	42.9
12	4.1	144	16.81	49.2
14	4.7	196	22.09	65.8
15	4.5	225	20.25	67.5
16	5.2	256	27.04	83.2
17	5.0	289	25	85
–	44.4	1562	171.3	514.8

### Médias e Desvios-Padrão

$$\bar{x} = \frac{130}{13} = 10, \quad \bar{y} = \frac{44.4}{13} = 3.42$$

$$s_x^2 = \frac{1562 - 13 \times 10^2}{13 - 1} = 21.83 \Rightarrow s_x = \sqrt{21.83} = 4.67$$

$$s_y^2 = \frac{171.3 - 13 \times 3.42^2}{13 - 1} = 1.60 \Rightarrow s_y = \sqrt{1.60} = 1.27$$

### Coefficiente de Correlação

$$\begin{aligned} r &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n - 1) s_x s_y} \\ &= \frac{514.8 - 13 \times 10 \times 3.42}{12 \times 4.67 \times 1.27} \\ &= 0.988 \end{aligned}$$

### Problema

A tabela abaixo apresenta os dados sobre a preferência por computadores entre homens e mulheres. Calcule os percentuais por linha e por coluna. O que eles revelam sobre a distribuição das preferências entre os grupos?

	PC	Mac
Homem	32	10
Mulher	42	40

## Exercício: Preferência de Computador por Gênero

	PC	Mac	Total
Homem	32	10	42
Mulher	42	40	82
Total	74	50	124

## Exercício: Preferência de Computador por Gênero

	PC	Mac	Total
Homem	32	10	42
Mulher	42	40	82
Total	74	50	124

	PC	Mac	Total
Homem	76	24	100
Mulher	51	49	100
Total	60	40	100

## Exercício: Preferência de Computador por Gênero

	PC	Mac	Total
Homem	32	10	42
Mulher	42	40	82
Total	74	50	124

	PC	Mac	Total
Homem	43	20	34
Mulher	57	80	66
Total	100	100	100