

Estatística Descritiva III

Ciências Contábeis - FEA - Noturno

1º Semestre 2023

Profs. Leonardo T. Rolla e Nikolai Kolev

(baseado em material previamente
desenvolvido pelo Prof. Gilberto Alvarenga Paula)

- 1 Tabelas de Contingência
- 2 Diagrama de Dispersão
- 3 Correlação Linear

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	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	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%)

Exercício: preferência de computador por gênero

	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	76	24	100
Mulher	51	49	100
Total	60	40	100

Exercício: preferência de computador por gênero

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

1 Tabelas de Contingência

2 Diagrama de Dispersão

3 Correlação Linear

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

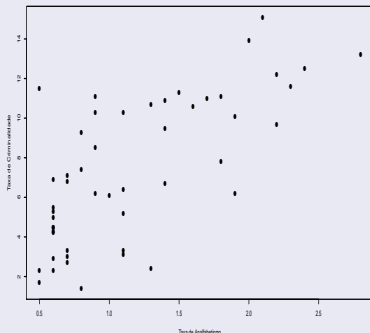
Vamos considerar como ilustração a relação entre

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

nos 50 estados norte americanos no início dos anos 70.

Exemplo: Criminalidade e Analfabetismo

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

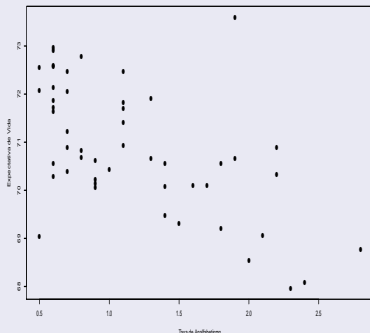
Vamos considerar como ilustração a relação entre

- 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



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

- 1 Tabelas de Contingência
- 2 Diagrama de Dispersão
- 3 Correlação Linear**

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

A seguinte fórmula pode ser mais prática para cálculos à mão:

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

Propriedades

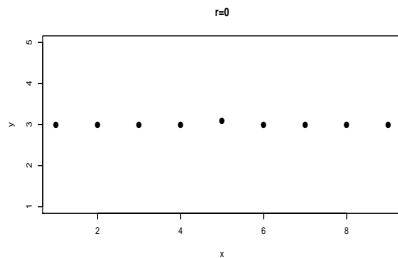
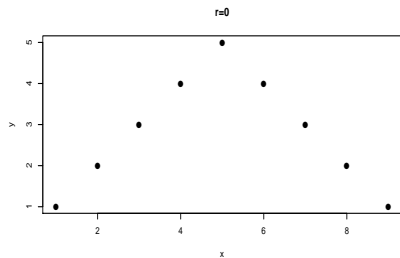
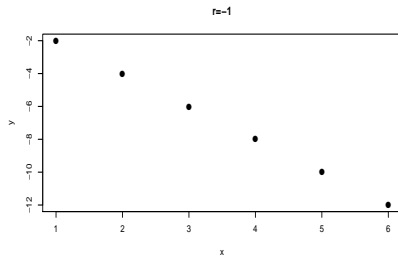
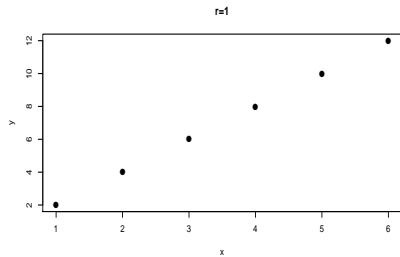
O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

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

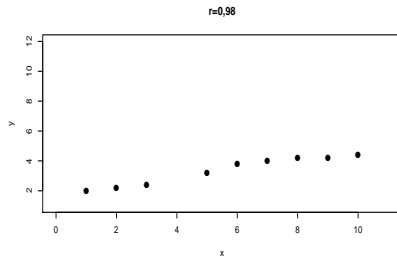
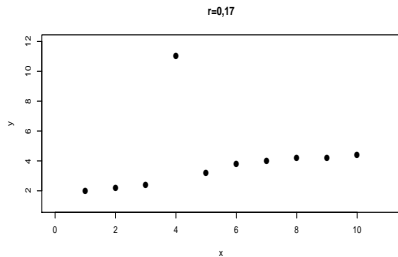
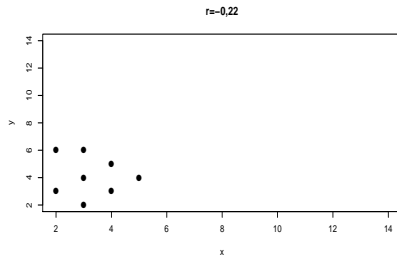
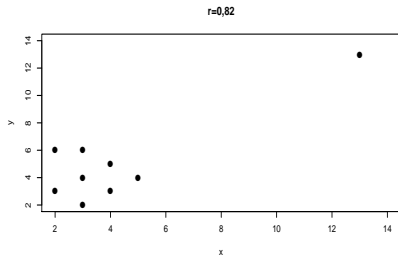
Casos particulares

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

Interpretação



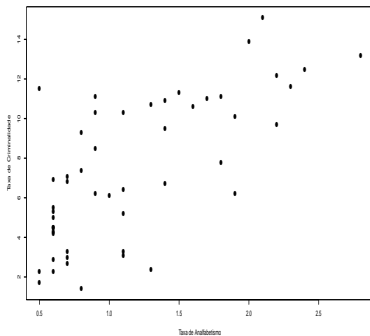
Interpretação



Observações

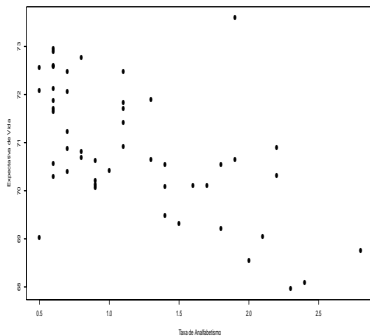
- Correlação linear próxima de zero significa ausência de **relação linear** entre as variáveis quantitativas, contudo pode haver **outro tipo de relação** entre as variáveis
- Pontos extremos no diagrama de dispersão podem influenciar o valor do coeficiente de correlação linear. Assim, deve-se olhar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

Exemplo: Criminalidade e Analfabetismo



$$r = 0,702$$

Exemplo: Expectativa de Vida e Analfabetismo



$$r = -0,59$$

Exercício: Tempo de Estudo e Nota na Prova

Descrição das Variáveis

Vamos considerar dados hipotéticos para as variáveis quantitativas

- X: Tempo de Estudo (em horas)
- Y: Nota na Prova

para uma amostra de $n = 5$ estudantes.

Descrição dos Dados

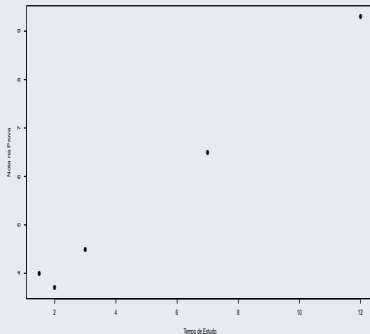
x	3,0	7,0	2,0	1,5	12,0
y	4,5	6,5	3,7	4,0	9,3

Exercício

Desenhar o diagrama de dispersão.

Calcular o coeficiente de correlação.

Diagrama de Dispersão



Podemos notar que, conforme aumenta o **Tempo de Estudo**, a **Nota na Prova** tende a aumentar. Nota-se também uma tendência linear.

Algumas Quantidades

$$\bar{x} = 5,1; \bar{y} = 5,6; \sum x_i^2 = 208,25; \sum y_i^2 = 178,68 \text{ e } \sum x_i y_i = 184.$$

Desvios Padrão Amostrais

$$s_x^2 = \frac{208,25 - 5 \times (5,1)^2}{4} = 19,55 \rightarrow s_x = \sqrt{19,55} = 4,42.$$

$$s_y^2 = \frac{178,68 - 5 \times (5,6)^2}{4} = 5,47 \rightarrow s_y = \sqrt{5,47} = 2,34.$$

Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

$$\begin{aligned} r &= \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y} \\ &= \frac{184 - 5 \times 5,1 \times 5,6}{4 \times 4,42 \times 2,34} \\ &= \frac{184 - 142,8}{41,37} \\ &= \frac{41,20}{41,37} = 0,996. \end{aligned}$$