

## **REGRESSÃO E CORRELAÇÃO LINEARES**

**Disciplina: Probabilidade e Estatística**  
**Curso: Ciência da Computação**  
**Professor: Leandro Bordin**

### ✓ **Introdução/Conceitos iniciais**

- ✓ A regressão e a correlação são duas técnicas estreitamente relacionadas que envolvem uma forma de estimação
- ✓ Mais especificamente, a análise de regressão e correlação compreende a análise de dados amostrais para saber se e como duas ou mais variáveis estão relacionadas uma com a outra numa população

### ✓ **Introdução/Conceitos iniciais**

- ✓ A análise de correlação fornece um número (coeficiente) que resume o grau de relacionamento entre duas variáveis, enquanto que a análise de regressão tem como resultado uma equação matemática que descreve tal relacionamento
- ✓ Esta equação pode ser usada para estimar, ou prever, valores futuros de uma variável (a variável dependente) quando se conhecem valores da outra variável (a variável independente)

## **Regressão Linear**

### ✓ **Regressão Linear**

- ✓ A regressão linear constitui uma tentativa de estabelecer uma equação matemática linear (linha reta) que descreva o relacionamento entre duas variáveis, uma dependente (y) e outra, independente (x)

### ✓ **Regressão Linear**

#### ✓ **Equação Linear**

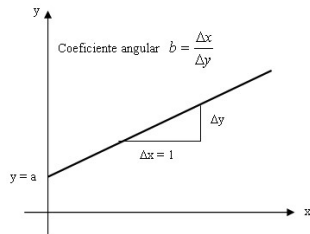
- ✓ Duas importantes características da equação linear são: (a) o coeficiente angular da reta e (2) a cota da reta em determinado ponto; uma equação linear tem a seguinte forma:

$$y = a + bx$$

## ✓ Regressão Linear

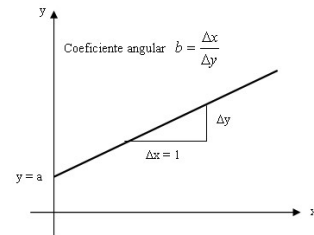
### ✓ Equação Linear

✓ O valor de  $a$  representa a cota da reta em  $x = 0$  e  $b$  é o coeficiente angular da reta, ambos determinados com base nos dados amostrais



## ✓ Regressão Linear

### ✓ Equação Linear



**Observação:** O coeficiente angular de uma reta pode ser interpretado da seguinte forma: para cada unidade de variação de  $x$ , correspondem  $b$  unidades de variação em  $y$ .

## ✓ Regressão Linear

### ✓ Diagrama de dispersão

✓ É importante ter em mente que nem todas as situações podem ser aproximadas por uma equação linear

✓ Por isso, em geral é necessário desenvolver um trabalho preliminar para determinar se um modelo linear é adequado

## ✓ Regressão Linear

### ✓ Diagrama de dispersão

✓ O processo mais simples consiste em grafar os dados e concluir se uma relação linear parece razoável

✓ O gráfico resultante deste processo denomina-se diagrama de dispersão

## ✓ Regressão Linear

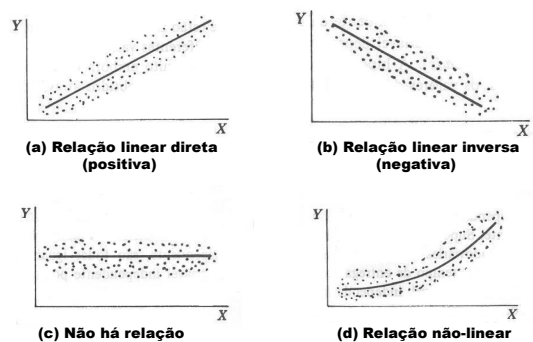
### ✓ Diagrama de dispersão

✓ Um diagrama de dispersão é um gráfico no qual cada ponto plotado representa um par observado de valores para as variáveis dependente e independente

✓ O valor da variável independente  $x$  é plotado com respeito ao eixo horizontal, e o valor da variável dependente  $y$ , com respeito ao eixo vertical

## ✓ Regressão Linear

### ✓ Diagrama de dispersão



## ✓ Regressão Linear

### ✓ Diagrama de dispersão

**Observação:** Na análise de regressão, um valor de  $y$  não pode ser estimado, se o valor de  $x$  não estiver no intervalo de valores que serviram de base para a equação de regressão

## ✓ Regressão Linear

### ✓ Determinação da equação matemática

✓ A técnica mais usada para ajustar uma linha reta a um conjunto de pontos é conhecida como método dos mínimos quadrados; a reta resultante tem duas características importantes:

- ✓ a) a soma dos desvios verticais dos pontos em relação à reta é zero
- ✓ b) a soma dos quadrados desses desvios é mínima (isto é, nenhuma outra reta daria menor soma de quadrados de tais desvios)

## ✓ Regressão Linear

### ✓ Determinação da equação matemática

✓ Considerando o método dos mínimos quadrados, os valores de  $a$  e  $b$  para a equação linear  $y = a + bx$ , são dados pelas seguintes expressões:

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{(\sum y) - b(\sum x)}{n}$$

sendo:

$x$  = valores da variável independente;

$y$  = valores da variável dependente;

$n$  = número de observações.

# Correlação Linear

## ✓ Correlação Linear

✓ O objetivo do estudo correlacional é a determinação da força do relacionamento entre duas observações (variáveis)

✓ A correlação indica até que ponto os valores de uma variável estão relacionados com os de outra

## ✓ Correlação Linear

### ✓ Coeficiente de correlação linear

✓ O grau de relacionamento entre duas variáveis é sintetizado por um coeficiente de correlação, dado pela seguinte expressão

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{(\sqrt{n(\sum x^2) - (\sum x)^2})(\sqrt{n(\sum y^2) - (\sum y)^2})}$$

sendo:

$x$  = valores da variável independente;

$y$  = valores da variável dependente;

$n$  = número de observações.

## ✓ Correlação Linear

### ✓ Coeficiente de correlação linear

- ✓ O coeficiente de correlação tem duas propriedades que caracterizam a natureza de uma relação entre duas variáveis
- ✓ Uma é o sinal (+ ou-) e a outra é sua magnitude ( $r$ )
- ✓ O sinal é o mesmo que o do coeficiente angular de uma reta imaginária que se “ajustasse” aos dados se fosse traçada num diagrama de dispersão, e a magnitude de  $r$  indica quão próximos da “reta” estão os pontos individuais

## ✓ Correlação Linear

### ✓ Coeficiente de correlação linear

- a) o valor de  $r$  varia de  $-1,00$  a  $+1,00$
- b) um relacionamento positivo ( $r = +$ ) entre duas variáveis indica que a valores altos (baixos) de uma das variáveis, correspondem valores altos (baixos) da outra
- c) um relacionamento negativo ( $r = -$ ) entre duas variáveis indica que a valores altos (baixos) de uma das variáveis, correspondem valores baixos (altos) da outra

## ✓ Correlação Linear

### ✓ Coeficiente de correlação linear

- d) um relacionamento zero ( $r \approx 0$ ) indica que alguns valores altos estão em correspondência com valores baixos e outros estão em correspondência com valores altos
- e) o sinal de  $r$  é sempre o mesmo do sinal de  $b$ , o coeficiente angular de uma reta imaginária ajustada aos dados. Note-se que não é necessário calcular esta reta

## ✓ Correlação Linear

### ✓ Coeficiente de correlação linear

- ✓ A classificação do relacionamento entre variáveis é feita da seguinte forma:

Valor de $r$	Descrição do relacionamento
$+1,00$	Relacionamento positivo perfeito
Cerca de $+0,70$	Relacionamento positivo moderado
$0,00$	Ausência de relacionamento
Cerca de $-0,70$	Relacionamento negativo moderado
$-1,00$	Relacionamento negativo perfeito

## ✓ Correlação Linear

### ✓ Coeficiente de Determinação ( $r^2$ )

- ✓ O coeficiente de determinação (ou poder de explicação) tem como objetivo dizer o quanto a variável  $x$  (variável independente) explica ou é responsável pelo valor de  $y$  (variável dependente); o coeficiente de determinação é expresso da seguinte forma:

$$r^2 = (r)^2$$

sendo:  
 $r$  = coeficiente de correlação linear

## ✓ Regressão e Correlação Lineares

### Exemplo

1. Suponha que um analista toma uma amostra aleatória de 10 carregamentos feitos por via rodoviária e anota a distância (em Km) e o tempo de entrega (em dias), conforme tabela abaixo. Com base nestes dados:
  - a) determinar a equação de regressão;
  - b) estimar o tempo de entrega para um carregamento de 1.000 Km;
  - c) concluir se esta equação de regressão pode ser usada para estimar o tempo de entrega de um carregamento de 2.500 Km;
  - d) determinar o coeficiente de correlação;
  - e) determinar o coeficiente de determinação.

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

Distância (Km)	Tempo de entrega (dias)
825	3,5
215	1,0
1070	4,0
550	2,0
480	1,0
920	3,0
1350	4,5
325	1,5
670	3,0
1215	5,0

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

a) determinar a equação de regressão

$$y = a + bx$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{(\sum y) - b(\sum x)}{n}$$

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

a) determinar a equação de regressão

Distância (x)	Tempo (y)	x.y	x <sup>2</sup>
825	3,5	2887,5	680625
215	1,0	215	46225
1070	4,0	4280	1144900
550	2,0	1100	302500
480	1,0	480	230400
920	3,0	2760	846400
1350	4,5	6075	1822500
325	1,5	487,5	105625
670	3,0	2010	448900
1215	5,0	6075	1476225
Σ 7620	Σ 28,5	Σ 26370	Σ 7104300

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

a) determinar a equação de regressão

$$y = a + bx$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{10(26370) - (7620)(28,5)}{10(7104300) - (7620)^2} = 0,0036$$

$$a = \frac{(\sum y) - b(\sum x)}{n} = \frac{28,5 - (0,0036)(7620)}{10} = 0,11$$

$$Y = 0,11 + 0,0036x$$

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

b) estimar o tempo de entrega para um carregamento de 1.000 Km

$$Y = 0,11 + 0,0036x$$

$$Y = 0,11 + 0,0036(1000)$$

$$Y = 3,7 \text{ dias}$$

c) concluir se esta equação de regressão pode ser usada para estimar o tempo de entrega de um carregamento de 2.500 Km

Não é apropriado usar a equação para x = 2500 km porque os dados da amostra incluem distâncias no intervalo de 215 a 1350 km

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

d) o coeficiente de correlação

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{(\sqrt{n(\sum x^2) - (\sum x)^2})(\sqrt{n(\sum y^2) - (\sum y)^2})}$$

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

d) o coeficiente de correlação

Distância (x)	Tempo (y)	x.y	x <sup>2</sup>	y <sup>2</sup>
825	3,5	2887,5	680625	12,25
215	1,0	215	46225	1
1070	4,0	4280	1144900	16
550	2,0	1100	302500	4
480	1,0	480	230400	1
920	3,0	2760	846400	9
1350	4,5	6075	1822500	20,25
325	1,5	487,5	105625	2,25
670	3,0	2010	448900	9
1215	5,0	6075	1476225	25
Σ 7620	Σ 28,5	Σ 26370	Σ 7104300	Σ 99,75

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

d) o coeficiente de correlação

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

$$r = \frac{10(26370) - (7620)(28,5)}{\sqrt{10(7104300) - (7620)^2} \sqrt{10(99,75) - (28,5)^2}}$$

r = 0,95 (relação positiva quase perfeita)

## ✓ Regressão e Correlação Lineares

### Exemplo

1.

e) o coeficiente de determinação

$$r^2 = (r)^2$$

$$r^2 = (0,95)^2$$

$$r^2 = 0,9025$$

$$r^2 = 90,25\%$$

90,25% da variação no tempo de entrega é explicada pela variação na distância percorrida; os outros 9,75% são explicados por fatores não considerados na pesquisa

## ✓ Regressão e Correlação Lineares

### Exemplo

2. Certa empresa estudando a variação da demanda (em unidades) de seu produto em relação à variação de preço de venda (R\$), obteve os resultados da tabela abaixo. Com base nestes dados:
- determinar a equação de regressão;
  - estimar a demanda esperada se o preço de venda é de R\$ 60,00 e R\$ 120,00;
  - determinar o coeficiente de correlação;
  - determinar o coeficiente de determinação.

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

Preço de venda (R\$)	Demanda (unidades)
38	350
42	325
50	297
56	270
59	256
63	246
70	238
80	223
95	215
110	208

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

a) determinar a equação de regressão

$$y = a + bx$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{(\sum y) - b(\sum x)}{n}$$

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

a) determinar a equação de regressão

Preço (x)	Demanda (y)	$\sum xy$	$\sum x^2$
38	350	13300	1444
42	325	13650	1764
50	297	14850	2500
56	270	15120	3136
59	256	15104	3481
63	246	15498	3969
70	238	16660	4900
80	223	17840	6400
95	215	21425	9025
110	208	22880	12100
<b><math>\sum 663</math></b>	<b><math>\sum 2628</math></b>	<b><math>\sum 165327</math></b>	<b><math>\sum 48719</math></b>

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

a) determinar a equação de regressão

$$y = a + bx$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{10(165327) - (663)(2628)}{10(48719) - (663)^2} = -1,87$$

$$a = \frac{(\sum y) - b(\sum x)}{n} = \frac{2628 - (-1,87)(663)}{10} = 386,78$$

$$Y = 386,78 - 1,87x$$

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

b) estimar a demanda esperada se o preço de venda é de R\$ 60,00

$$Y = 386,78 - 1,87x$$

$$Y = Y = 386,78 - 1,87(60)$$

$$Y = 275 \text{ unidades}$$

b) estimar a demanda esperada se o preço de venda é de R\$ 120,00

Y não pode ser estimado porque x está fora do intervalo 38/110

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

c) o coeficiente de correlação

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{(\sqrt{n(\sum x^2) - (\sum x)^2}) \sqrt{n(\sum y^2) - (\sum y)^2}}$$

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

c) o coeficiente de correlação

Preço (x)	Demanda (y)	$\sum xy$	$\sum x^2$	$\sum y^2$
38	350	13300	1444	122500
42	325	13650	1764	105625
50	297	14850	2500	88209
56	270	15120	3136	72900
59	256	15104	3481	65536
63	246	15498	3969	60516
70	238	16660	4900	56644
80	223	17840	6400	49729
95	215	21425	9025	46225
110	208	22880	12100	43264
<b><math>\sum 663</math></b>	<b><math>\sum 2628</math></b>	<b><math>\sum 165327</math></b>	<b><math>\sum 48719</math></b>	<b><math>\sum 711148</math></b>

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

c) o coeficiente de correlação

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{(\sqrt{n(\sum x^2) - (\sum x)^2}) \sqrt{n(\sum y^2) - (\sum y)^2}}$$

$$r = \frac{10(165327) - (663)(2628)}{\sqrt{10(48719) - (663)^2} \sqrt{10(711148) - (2628)^2}}$$

$$r = -0,90 \text{ (relação negativa quase perfeita)}$$

## ✓ Regressão e Correlação Lineares

### Exemplo

2.

d) o coeficiente de determinação

$$r^2 = (r)^2$$

$$r^2 = (-0,9)^2$$

$$r^2 = 0,81$$

$$r^2 = 81\%$$

81% da variação na demanda é explicada pelo preço de venda; os outros 19% são explicados por fatores não considerados na pesquisa

## ✓ Regressão e Correlação Lineares

### Exercícios

1. A tabela abaixo apresenta dados referentes ao número de horas de estudo extra-classe para uma amostra aleatória de 8 alunos de um curso de Estatística, bem como os graus obtidos em um exame aplicado no final do curso. Com base nestes dados:

- determinar a equação de regressão;  $Y = 40 + 1,5x$
- estimar o grau no exame obtido por um estudante que dedicou 30 horas de estudo extra-classe;  $Y = 85$  (nota)
- determinar o coeficiente de correlação;  $r = 0,86$
- determinar o coeficiente de determinação.  $r^2 = 73,96\%$

Horas de estudo (x)	20	16	34	23	27	32	18	22
Grau no exame (y)	64	61	84	70	88	92	72	77

## ✓ Regressão e Correlação Lineares

### Exercícios

2. A tabela abaixo apresenta dados relacionados com o número de semanas de experiência na montagem de pequenos componentes eletrônicos, bem como o número de tais componentes que foram rejeitados durante a última semana, para uma amostra de 12 trabalhadores aleatoriamente selecionados. Com base nestes dados, determinar:

- a equação de regressão;  $Y = 35,49 - 1,39x$
- estimar o número de componentes eletrônicos que serão rejeitados se o tempo de experiência for de 13 semanas;  $Y = 17$  componentes
- o coeficiente de correlação;  $r = -0,91$
- o coeficiente de determinação.  $r^2 = 82,81\%$

Semanas de experiência (x)	7	9	6	14	8	12	10	4	2	11	1	8
Componentes rejeitados (y)	26	20	28	16	23	18	24	26	38	22	32	25

## ✓ Regressão e Correlação Lineares

### Exercícios

3. Os dados da tabela abaixo descrevem a relação entre a frequência de acidentes e o nível de esforço preventivo (educacional) para uma amostra aleatória de 8 observações.

- determinar a equação de regressão;  $Y = 8,45 - 0,0059x$
- estimar o número de acidentes se o esforço preventivo for de 250 horas;  $Y = 7$  acidentes
- determinar o coeficiente de correlação;  $r = -0,95$
- determinar o coeficiente de determinação.  $r^2 = 90,25\%$

Horas em educação (x)	200	500	450	800	900	150	300	600
Número de acidentes (y)	7,0	6,4	5,2	4,0	3,1	8,0	6,5	4,4

## ✓ Regressão e Correlação Lineares

### Exercícios

4. O conjunto de dados apresentados na tabela abaixo descreve a relação entre as vendas e o lucro, ambos dados em R\$, de uma determinada empresa. Com base nestes dados:

- determinar a equação de regressão;  $Y = 13,31 + 0,02x$
- estimar o lucro da empresa se o total de vendas for de R\$ 250.000,00;  $Y = 18310$  (R\$)
- determinar o coeficiente de correlação;  $r = 0,93$
- determinar o coeficiente de determinação.  $r^2 = 86,49\%$

Vendas -R\$1.000,00- (x)	201	225	305	380	560	600	685	735	510	725	450	370	150
Lucro -R\$1.000,00- (y)	17	20	21	23	25	24	27	27	22	30	21	19	15