

## AULA 7

## CORRELAÇÃO E REGRESSÃO

Prof. Lupércio F. Bessegato

1. Ache os valores de  $a$  e  $b$  para a equação de regressão  $Y = a + bX$ , usando o conjunto de dados apresentados a seguir, *sem* fazer cálculos (sugestão: faça um gráfico).

Y	5	4	3	2	1
X	0	1	2	3	4

Usando a equação de regressão obtida, dê a previsão para  $X = 5$ , calcule o intervalo de confiança para a estimativa e comente seus resultados.

2. Certa empresa comprou recentemente 10 caminhões e acha que existe certa variabilidade no consumo de combustível. Como os caminhões percorrem estradas similares, suspeita-se que parte desta variabilidade é devida à forma como os motoristas dirigem, e parte é explicada pelo fato de carregarem cargas diferentes. A seguir, são apresentados dados de consumo de combustível dos caminhões para percursos similares e um registro da carga que eles carregavam.

Caminhão	Consumo (km/l)	Peso (ton)
1	7,79	1.270
2	6,76	1.678
3	6,96	1.633
4	7,44	1.451
5	6,60	1.860
6	8,01	1.179
7	7,21	1.633
8	6,99	1.542
9	7,02	1.724
10	7,60	1.225

- Existe associação entre consumo de combustível (km/l) e carga dos caminhões, e se existe, ela é forte?
  - Tendo verificado se a relação é linear, estime a equação de regressão linear, se for o caso.
  - Determine  $R^2$  e teste se a diferença entre o coeficiente angular e zero é estatisticamente significativa.
  - Interprete todos os resultados.
3. Você recebeu informações sobre a porcentagem de ações no mercado, preços como porcentagem dos preços médios do competidor, e gastos com propaganda como porcentagem dos preços médios do competidor, para uma linha específica de comida para gatos, nos últimos oito anos.

	<i>Ações do Mercado</i> (%)	<i>Preço Relativo</i> (%)	<i>Propaganda Relativa</i> (%)
1990	30	89	110
1991	31	85	115
1992	30,5	86,5	120
1993	29	92	114
1994	29,5	90	112
1995	28,5	102	115
1996	28	115	116
1997	26	125	114

- Usando um software adequado, calcule a equação de regressão múltipla para explicar as variações no mercado de ações.
  - Interprete sua resposta em função do teste de significância estatística para os coeficientes angulares.
  - Calcule o coeficiente de determinação ( $R^2$ ) de seu ajuste.
  - Compare seu resultado com o de uma regressão que ignora a variável propaganda relativa.
4. Você vende rosas vermelhas de cabo longo pelo correio, em caixas pelo preço de 9,99 dólares, incluindo o preço de remessa. Você anuncia em alguns jornais e revistas, mas está inseguro se as vendas dependem do preço, porque você vende bem. Recentemente, você experimentou variar o preço por uma semana para calcular a extensão da associação entre preços e vendas. Os resultados são dados nas duas primeiras colunas da tabela.

<i>Preço</i>	<i>Vendas</i> (un)	<i>Despesa com</i> <i>Propaganda</i>
8,99	496	1.795
9,50	465	1.756
9,99	482	2.805
10,50	459	2.218
10,99	408	1.965
11,50	382	2.285
11,99	315	1.089
12,50	363	1.888
12,99	309	1.430

- Faça um diagrama de dispersão e calcule o coeficiente de correlação para as variáveis acima. Dê especial atenção para a interpretação dos resultados e faça as pressuposições necessárias.
- Usando os dados sobre venda de rosas e os respectivos preços, ajuste uma equação de regressão e interprete os resultados.
- Inclua agora os dados de despesa com propaganda e estime, usando uma planilha de cálculo, uma regressão múltipla das vendas contra preços e despesas com propaganda. Interprete os resultados.

5. Uma academia de ginástica decidiu ilustrar uma abordagem teórica de como os exercícios aeróbicos e a ingestão de calorias podem afetar o peso. Doze dos membros estabelecidos na academia registraram cuidadosamente o número de minutos de exercícios aeróbicos que praticaram no decorrer de uma semana, juntamente com sua ingestão calórica semanal. Esses dados são apresentados na seguinte tabela.

Perda de peso (lb)	Exercício aeróbico (min)	Calorias ingeridas
0,6	112	9.560
2,8	190	7.552
1,4	171	11.981
1,4	148	8.338
2,6	193	10.202
3,8	235	7.252
3,3	237	8.097
2,5	176	8.121
2,6	185	8.300
2,0	186	11.216
3,3	228	7.212
1,1	65	7.631

- Construa dois diagramas de dispersão para ilustrar a relação entre perda de peso e exercícios aeróbicos e perda de peso e calorias ingeridas. Descreva os principais elementos dos gráficos e, assim, estime os possíveis coeficientes de correlação.
  - Calcule o coeficiente de correlação entre perda de peso e os exercícios aeróbicos. Utilizando este valor, calcule o coeficiente de determinação e defina seu significado nesse contexto.
  - Calcule o coeficiente de correlação entre perda de peso e calorias ingeridas. Comparando-o com o coeficiente de correlação encontrado em (b), determine qual dos fatores contribui mais para a perda de peso, fornecendo motivos adequados para sua escolha.
  - Calcule um modelo de regressão linear que descreva a relação entre a perda de peso e a quantidade de exercícios aeróbicos.
  - Utilize esta equação para prever a perda de peso esperada de um membro da academia que tenha praticado 150 minutos de exercícios aeróbicos na semana.
6. Os dados abaixo se referem a meses de experiência de dez digitadores e o número de erros cometidos na digitação de determinado texto.

Meses ( $x$ )	1	2	3	4	5	6	7	8	9	10
Erros ( $y$ )	30	28	24	20	18	14	13	10	7	6

Dados:  $\sum x_i = 60$ ,  $\sum x_i^2 = 460$ ,  $\sum y_i = 170$ ,  $\sum x_i y_i = 768$ .

- Represente graficamente esse conjunto de dados.
- Assumindo que um modelo de regressão linear é adequado, determine os coeficientes da equação pelo método dos mínimos quadrados.
- Represente a reta de regressão no gráfico feito anteriormente.
- Qual a posição do ponto  $(\bar{x}, \bar{y})$  em relação à reta de regressão?
- Qual o número esperado de erros para um digitador com 5 meses de experiência?

7. Uma empresa de transportes forneceu os seguintes dados com relação a uma amostra de viagens feitas, dando a distância viajada e o tempo gasto. A empresa está interessada em desenvolver um modelo para prever o tempo gasto com uma viagem, se a distância a ser viajada for conhecida.

Distância (km)	Tempo (horas)
200	3,2
120	2,0
175	3,0
150	2,0
300	4,7
320	5,5
240	3,8
180	2,8
210	3,4
260	4,5

- Apresente os dados utilizando um diagrama de dispersão apropriado.
  - Encontre o coeficiente de correlação e a equação da linha de regressão.
  - Dois caminhões estão prestes a deixar a garagem. Um fará uma viagem de 90 km, enquanto o outro viajará 220 km. Utilizando sua equação de regressão, estime o tempo de viagem para cada caminhão. Quanto de confiança você teria em cada uma destas respostas?
8. Os dados abaixo correspondem às variáveis: renda familiar e gasto com alimentação numa amostra de dez famílias, representadas em reais.

Renda familiar (x)	Gasto com alimentação (y)
3	1,5
5	2,0
10	6,0
20	10,0
30	15,0
50	20,0
70	25,0
100	40,0
150	60,0
200	80,0

Obtenha a equação de regressão  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ .

- Qual a previsão do gasto com alimentação para uma família com renda de 170 reais?
- Qual a previsão do gasto para famílias com excepcional renda, por exemplo, 1.000 reais? Você acha este valor razoável? Por quê?
- Se você respondeu que o valor obtido em (b) não é razoável, encontre uma explicação para o ocorrido. (Sugestão: Interprete a natureza das variáveis X e Y e o comportamento de Y para grandes valores de X).

9. Um jornal quer verificar a eficácia de seus anúncios na venda de carros usados. A tabela abaixo mostra o número de anúncios publicados e o correspondente número de carros vendidos por seis companhias que usaram apenas esse jornal como veículo de propaganda.

Companhia	A	B	C	D	E	G
Anúncios	74	45	48	36	27	16
Carros vendidos	139	108	98	76	62	57

Ajustando-se a reta de regressão obteve-se,  $y = 1,516x + 27,844$  e  $F = 70,17$ . Como você argumentaria com a companhia G para que ela aumentasse o número de anúncios, aumentando a venda de carros?

10. A indústria farmacêutica MIMI vende um remédio para combater resfriado. Após dois anos de operação, ela coletou as seguintes informações trimestrais:

Trimestre	Vendas (10.000) (Y)	Despesas c/ Propaganda (X)	Temperatura Média do Trimestre (Z)
1	25	11	2
2	13	5	13
3	8	3	16
4	20	9	7
5	25	12	4
6	12	6	10
7	10	5	13
8	15	9	4

$$\Sigma y = 128$$

$$\Sigma x = 60$$

$$\Sigma z = 69$$

$$\Sigma y^2 = 128$$

$$\Sigma x^2 = 60$$

$$\Sigma z^2 = 69$$

$$\Sigma yx = 1.101$$

$$\Sigma yz = 897$$

$$\Sigma xz = 397$$

- Faça os gráficos  $(x, y)$  e  $(z, y)$ .
- Encontre as retas  $\hat{y} = a + bx$  e  $\hat{y} = c + dz$ .
- Qual das duas retas você acha estatisticamente mais adequada para prever as vendas? Por quê?
- De acordo com a decisão acima, qual a previsão de vendas para um trimestre em que a despesa de propaganda será 8 e a temperatura prevista 10?