

Correlação e Regressão

Aplicações

Aplicações

1. Fez-se um estudo acerca do volume de oxigénio (VO_2) máximo em litros/minuto consoante a idade em anos de 11 indivíduos. Os resultados obtidos estão representados na tabela:

Idade	20	18	32	45	38	39	21	42	24	35	27
VO_2 Máx	4,2	4,0	3,6	2,8	3,4	3,0	4,1	3,1	3,9	3,5	3,9

- a) Parece-lhe admissível a existência de uma relação linear entre as variáveis consideradas?
- b) Determine a equação da reta de regressão do volume de oxigénio máximo sobre a idade;
- c) Estime o volume de oxigénio de um indivíduo de 23 anos.

Aplicações

2. Na tabela seguinte estão registados os resultados de uma experiência realizada com amostras de sangue do mesmo indivíduo, para conhecer como varia o tempo que demora a coagular o sangue humano a diferentes temperaturas:

Temperatura (°C)	32	45	38	39	21	42	24	35	27
Tempo de coagulação (s)	3,6	2,8	3,4	3,0	4,1	3,1	3,9	3,5	3,9

- a) Interprete os valores do coeficiente de correlação e de determinação;
- b) Determine a equação da reta de regressão linear;
- c) Qual o tempo de coagulação previsível para uma temperatura de 22 °C?

Aplicações

3. Foi registada a taxa de consumo de oxigénio de aves de certa espécie, a diferentes temperaturas ($^{\circ}\text{C}$), e registaram-se os valores na seguinte tabela:

Taxa O_2	5,2	4,7	4,5	3,6	3,4	3,1	2,7	1,8
Temp	-18	-15	-10	-5	0	5	10	19

- a) Parece-lhe admissível a existência de uma relação linear entre as variáveis consideradas?
- b) Determine a equação da reta de regressão, da taxa de oxigénio sobre a temperatura;
- c) Estime a taxa de consumo de oxigénio, para uma ave da população, à temperatura de -12°C .

Aplicações

4. Um grupo de 10 pacientes idosos, sujeitos a cirurgia da anca, foi avaliado na função locomotora após aplicação de uma metodologia experimental de fisioterapia ao longo de períodos diferenciados para cada doente, através de uma taxa de mobilidade. Os dados recolhidos foram os seguintes:

Taxa de mobilidade (%)	54	50	65	53	74	80	71	61	70	77
Duração do tratamento (dias)	10	12	15	11	20	30	30	14	18	25

Aplicações

4.

- a) Parece-lhe admissível a existência de uma relação linear entre as variáveis consideradas?
- b) Interprete o valor do coeficiente de determinação;
- c) Ajuste uma reta de regressão linear aos dados;
- d) Aplicou-se a referida metodologia a um paciente durante 28 dias. Efectue uma previsão relativa à taxa de mobilidade desse paciente.

Aplicações

5. A tabela seguinte indica as idades, em anos, e os gastos diários, em euros, de 10 pessoas:

Idade	16	22	22	28	30	31	36	44	55	60
Gastos diários	5	8	10	11	12	12	15	20	22	21

- a) Calcule o coeficiente de correlação e de determinação e interprete os seus valores;
- b) Determine a equação da reta de regressão;
- c) Interprete o valor do coeficiente de regressão estimado:
a;
- d) Avalie o gasto diário de uma pessoa com 24 anos.

Regressão

CUIDADO!!!!

Um conjunto de pontos dá evidência de linearidade apenas para os valores de X cobertos pelo conjunto de dados. Para valores de X que saem fora dos que foram cobertos não há qualquer evidência de linearidade. Por isso é arriscado usar uma recta de regressão estimada para prever valores de Y correspondentes a valores de X que saem fora do âmbito dos dados.

O perigo de extrapolar para fora do âmbito dos dados amostrais é que a mesma relação possa não mais se verificar.

Aplicações

6. Pretende-se testar um instrumento que mede a concentração de ácido láctico no sangue. Para isso foram utilizadas 20 amostras sanguíneas para as quais se conhece essa concentração e registou-se o valor da concentração fornecido pelo instrumento. Seja X a concentração conhecida de ácido láctico e Y a concentração de ácido láctico fornecida pelo instrumento. Os dados obtidos estão registados na tabela seguinte:

X	1	1	1	1	3	3	3	3	3	5	5	5	10	10	10	10	15	15	15	15
Y	1.1	0.7	1.8	0.4	3	1.4	4.9	4.4	4.5	7.3	8.2	6.2	12	13.1	12.6	13.2	18.7	19.7	17.4	17.1

Aplicações

6. $\sum_{i=1}^{20} x_i = 134$ $\sum_{i=1}^{20} y_i = 167.7$ $\sum_{i=1}^{20} x_i^2 = 1424$ $\sum_{i=1}^{20} y_i^2 = 2220,21$ $\sum_{i=1}^{20} x_i y_i = 1769.6$

- a) Calcule o coeficiente de correlação e interprete o seu valor. Qual a percentagem de variabilidade nos valores fornecidos pelo instrumento pode ser explicada linearmente por variações na verdadeira concentração de ácido láctico?
- b) Determine a equação de regressão;
- c) Preveja qual será o valor da concentração de ácido láctico fornecido pelo instrumento para uma amostra sanguínea cuja concentração conhecida é 1. Interprete o valor encontrado;
- d) Interprete o valor do coeficiente de regressão estimado - a.

Aplicações

7. Num estudo para determinar de que modo a habilidade para executar uma determinada tarefa complexa é influenciada pela duração do treino, foram usados 15 indivíduos aos quais foi dado um treino que variava de 3 a 12 horas. Depois do treino foram registados os tempos que cada um deles gastou a executar a tarefa. Representando por X a duração do treino (em horas) e por Y o tempo gasto na execução da tarefa (em minutos), os resultados obtidos resumem-se seguidamente:

$$\bar{x} = 7.2 \quad \bar{y} = 45.6 \quad \sum_{i=1}^{15} x_i^2 = 811.2 \quad \sum_{i=1}^{15} y_i^2 = 31350.6 \quad \sum_{i=1}^{15} x_i y_i = 4867.6$$

Aplicações

7.

- a) Calcule o coeficiente de correlação e de determinação e interprete os seus valores;
- b) Determine a equação de regressão;
- c) O que pode dizer relativamente ao tempo de execução da tarefa por um indivíduo que tenha sido sujeito a 35 horas de treino?

Aplicações

8. Uma amostra de alunos selecionada ao acaso de um curso com as disciplinas de Matemática e Estatística produziu as seguintes classificações em testes realizados no final do ano lectivo (escala 0-100):

Matemática X	56	50	72	67	31	50	65	40	80	61
Estatística Y	60	50	67	75	44	56	72	48	76	62

Aplicações

8. a) Parece-lhe admissível a existência de uma relação linear entre as duas disciplinas?
- b) Ajuste uma reta de regressão linear aos dados;
- c) A Joana, o António e a Maria obtiveram 60, 95 e 20 em Matemática, respectivamente, mas faltaram ao teste de Estatística. Poderá sugerir valores para as notas esperadas no teste de Estatística dos alunos que faltaram? Justifique a sua resposta;
- d) Suponha que o João obteve 70 em Estatística e faltou a Matemática. Obtenha uma nova recta de regressão que permita estimar uma nota para o teste de Matemática deste aluno e indique esse valor predito.
-

Aplicações

9. Os dados referem-se ao preço de venda de 14 carros usados da mesma marca e modelo no último trimestre:

$$\sum x_i^2 = 21825$$

$$\sum y_i^2 = 39960000$$

$$\sum x_i y_i = 640000$$

Quilometragem X (1000 km)	Preço de venda Y (dezenas de euros)
40	1000
30	1500
30	1200
25	1800
50	800
60	1000
65	500
10	3000
15	2500
20	2000
55	800
40	1500
35	2000
30	2000

Aplicações

9.

- a) Avalie se é razoável aplicar um modelo de regressão linear para descrever a relação entre a quilometragem e o preço de venda de um carro;
- b) Determine a equação de regressão;
- c) Qual a percentagem de variação no preço de venda dos carros que é explicada por variações na quilometragem?
- d) Qual o efeito no preço de venda de um acréscimo na quilometragem de 1000 Km? E de 2000 Km?

Aplicações

10. É esperado que a massa muscular de uma pessoa diminua com a idade. Para estudar essa relação, uma nutricionista selecionou 18 mulheres, com idades entre 40 e 79 anos, e observou em cada uma delas a idade e a massa muscular:

Idade	71	64	43	67	56	73	68	56	76	65	45	58	45	53	49	78	73	68
Massa muscular	82	91	100	68	87	73	78	80	65	84	116	76	97	100	105	77	73	78

Aplicações

10.

- a) Calcule o coeficiente de correlação e interprete o seu valor;
- b) Qual a percentagem de variabilidade nos valores da massa muscular pode ser explicada linearmente por variações na idade?
- c) Ajuste uma reta de regressão para a relação entre a massa muscular e a idade;
- d) Preveja um valor médio para a massa muscular de mulheres com 50 anos;
- d) Qual a consequência no valor da massa muscular por cada ano que passa.

Aplicações

11. A tabela seguinte apresenta o tempo de serviço (em anos) de 10 funcionários de uma companhia de seguros e o número de clientes que cda um possui:

Tempo de serviço	2	3	4	5	4	6	7	8	8	10
Nº de clientes	48	50	56	52	43	60	62	58	64	72

Aplicações

11.

- a) Parece-lhe admissível a existência de uma relação linear entre o tempo de serviço e o número de clientes?
- b) Calcule o coeficiente de determinação e interprete o seu valor;
- c) Ajuste uma reta de regressão linear aos dados;
- d) Interprete o valor do coeficiente da variável independente;
- e) Suponha que o Francisco trabalha há 9 anos para a companhia de seguros. Qual seria o número previsto de clientes que o Francisco devia ter.

Aplicações

12. Da análise do consumo médio de energia por agregado familiar durante 10 dias de um mês de Inverno numa cidade obtiveram-se os seguintes resultados:

Dia	1	2	3	4	5	6	7	8	9	10
Temperatura (°C)	15	14	12	14	12	11	11	10	12	13
Consumo de energia (kW)	4,3	4,4	5,3	4,6	5,5	5,9	5,7	6,2	5,2	5,0

Aplicações

12.

- a) Parece-lhe admissível a existência de uma relação linear entre a temperatura e o consumo de energia?
- b) Calcule o coeficiente de determinação e interprete o seu valor;
- c) Ajuste uma reta de regressão linear aos dados;
- d) Interprete o valor do coeficiente da variável independente;
- e) Qual o valor previsto para o consumo médio num dia de temperatura média igual a 10°C ? Que responderia se lhe fosse pedida uma previsão do consumo médio para um dia com temperatura média de 20°C ?

Aplicações

- 13.** Num estudo sobre o consumo de gasolina de vários modelos de automóveis ligeiros de passageiros e a cilindrada do respetivo motor, foi estabelecida a seguinte equação da recta de regressão dos mínimos quadrados

$$Y = 3,5 + 2 X$$

em que x é a cilindrada (em 10^3 cm^3) e y é o consumo (em litros por 100 km percorridos). Sabe-se que a precisão desta recta é de 0,803.

Aplicações

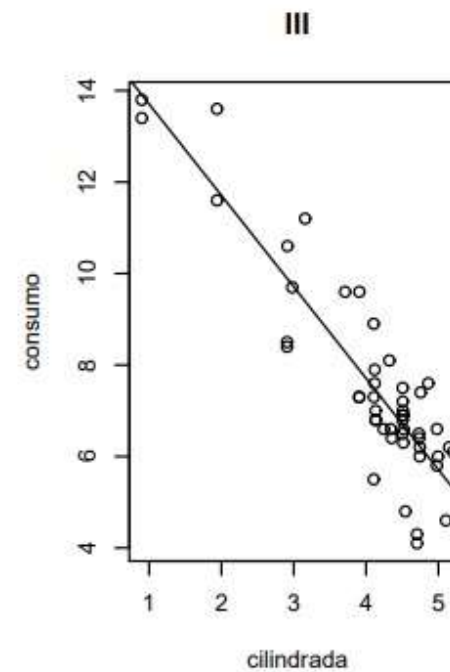
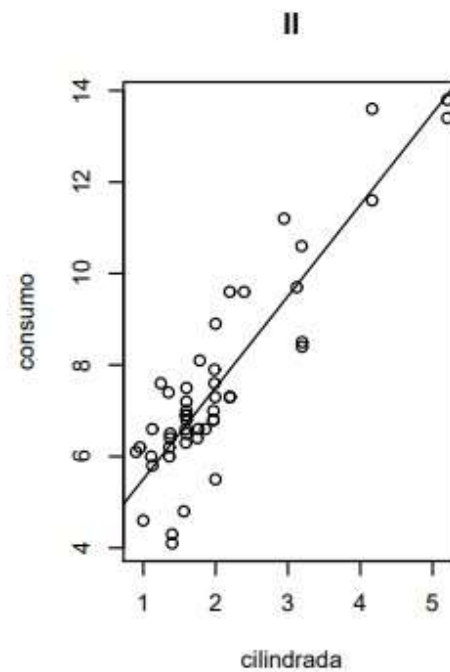
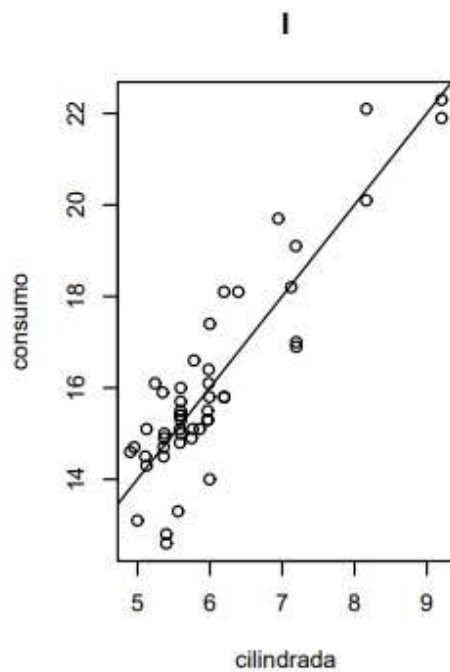
13.

- a) Qual é a variação esperada para o consumo de gasolina quando se aumenta a cilindrada de 1000 cm^3 ?
- b) Parece-lhe adequada a utilização do modelo linear para descrever a relação entre o consumo de gasolina e a cilindrada do motor nos modelos de automóveis analisados? Justifique;

Aplicações

13.

c) Qual dos seguintes gráficos corresponde à nuvem de pontos e à respectiva reta de regressão do estudo descrito?



Aplicações

14. Uma pesquisa foi realizada com o objectivo de verificar se existe uma associação entre a falta de sono e a capacidade de as pessoas resolverem problemas simples. Foram testadas 10 pessoas, mantendo-se sem dormir por um determinado número de horas. Após cada um destes períodos, cada pessoa teve de resolver um teste com adições simples, anotando-se então os erros cometidos. Os dados resultantes são os seguintes:

Nº de erros	6	8	6	10	8	14	12	14	12	16
Nº de horas sem dormir	8	8	12	12	16	16	20	20	24	24

Aplicações

14.

- a) Parece-lhe admissível a existência de uma relação linear entre o número de erros e o número de horas sem dormir?
- b) Calcule o coeficiente de determinação e interprete o seu valor;
- c) Ajuste uma reta de regressão linear aos dados;
- d) Interprete o valor do coeficiente da variável independente;
- e) Em média, qual o número previsto de erros cometidos por uma pessoa que esteve 10 horas sem dormir?