

CAPÍTULO 14 – REGRESSÃO

Problemas

PROBLEMA 14.1

O montante global dos seguros de vida efectuados pelas famílias de um determinado país depende do rendimento anual do agregado familiar. Na tabela seguinte apresentam-se os valores destas variáveis, expressas em unidades monetárias do país em causa, para um conjunto de 12 famílias considerado representativo da população.

Rendimento anual [1000 u.m.]	Capital seguro [1000 u.m.]
14	31
19	40
23	49
12	20
9	21
15	34
22	54
25	52
15	28
10	21
12	24
16	34

- (i) Estime a relação entre as duas variáveis.
- (ii) Calcule a probabilidade de uma família qualquer com um rendimento anual de 20.000 u.m. ter seguros de vida num montante que excede as 45.000 u.m.

PROBLEMA 14.2

Na tabela seguinte apresenta-se, para os últimos 9 anos, o volume de produção de trigo numa determinada região (em milhares de toneladas).

Ano (X)	1997	1998	1999	2000	2001	2002	2003	2004	2005
Vol. de Produção (Y)	285	270	294	279	260	262	258	272	255

Com base nos valores inscritos na tabela, calcularam-se as seguintes estatísticas:

$$\bar{x} = 2001, \quad \bar{y} = 270.(5);$$

$$s_{xx} = 60 \quad s_{yy} = 1416.(2) \quad s_{xy} = -203.$$

- (i) Estime e teste, ao nível de significância de 5%, a relação que define a evolução de produção de trigo em função do tempo.

- (ii) Estime a probabilidade de o volume de produção em 2006 ser superior ao de 2005.
- (iii) Estime a probabilidade de o volume de produção em 2007 ser superior ao de 2006.

PROBLEMA 14.3

No âmbito de uma auditoria internacional a uma Agência Governamental, tentou verificar-se a fiabilidade do Índice de Preços no Consumidor publicado pela Agência (o IPC oficial). Para o efeito, recalculou-se aquele índice seguindo uma metodologia considerada correcta. Na tabela seguinte apresentam-se os valores dos dois índices, o “correcto” e o oficial, para um período de seis anos.

<i>IPC correcto (X)</i>	<i>IPC oficial (Y)</i>
112.2	108.8
123.5	119.7
131.0	128.0
138.6	135.8
145.7	142.5
149.4	147.0

Com base nestes valores, calcularam-se as seguintes estatísticas:

$$\bar{x} = 133.4 \quad \bar{y} = 130.3$$

$$s_{XX} = 987.54 \quad s_{XY} = 1012.12 \quad s_{YY} = 1037.88$$

Face a estes dados, estime, teste e interprete a relação entre os dois índices. Nos testes hipóteses que efectuar, adopte o nível de significância de 5%.

PROBLEMA 14.4

No âmbito de um estudo de tráfego efectuado para apoiar o projecto de um túnel rodoviário, registaram-se, em 20 ocasiões diferentes, os valores observados das duas variáveis seguintes:

X: densidade do tráfego, expressa em número de veículos por quilómetro

Y: velocidade média dos veículos, expressa em km/hora.

A partir das mediações efectuadas, foram calculadas as seguintes estatísticas:

$$\bar{x} = 60 \quad \bar{y} = 25$$

$$s_{XX} = 5000 \quad s_{XY} = -1500 \quad s_{YY} = 630.$$

- (i) Determine a recta de regressão que permite prever a velocidade a partir da densidade. Verifique (com $\alpha = 5\%$) se o coeficiente angular da recta é significativo.
- (ii) Admitindo que a densidade de tráfego é de 70 veículos por quilómetro, defina o intervalo de previsão a 95% para a velocidade média.

- (iii) Estime o nível ao qual deve ser controlada a densidade de tráfego, de forma a maximizar o volume (V) de tráfego do túnel (com $V = X \cdot Y$ veículos por hora). Qual o valor do desvio padrão do estimador da densidade óptima?

PROBLEMA 14.5

Uma companhia de aviação efectua a manutenção preventiva de um determinado instrumento dos aviões do tipo AA após cada 25 horas de voo dos aparelhos. No entanto, foi sugerido que a desregulação dos instrumentos em causa dependeria mais do tempo que decorre entre manutenções sucessivas do que do número de horas de voo.

Para examinar a validade desta sugestão, efectuaram-se, para os aviões do tipo AA, 20 observações das variáveis seguintes:

X_1 : número de horas de voo decorridas desde a última operação de manutenção até à verificação do instrumento

X_2 : número de dias decorridos desde a última operação de manutenção até à verificação do instrumento

Y : medida de desregulação do instrumento.

A partir dos resultados obtidos foram calculadas as seguintes estatísticas:

$$\bar{x}_1 = 12 \quad \bar{x}_2 = 10 \quad \bar{y} = 2$$

$$s_{x_1x_1} = 1500 \quad s_{x_1x_2} = 500 \quad s_{x_2x_2} = 300 \quad s_{x_1Y} = 120 \quad s_{x_2Y} = 60 \quad s_{YY} = 30.$$

- Analise os resultados e verifique se as manutenções devem ser programadas de acordo com X_1 , de acordo com X_2 , ou de acordo com as duas variáveis.
- Admita que, por razões de segurança, a mediada de desregulação do instrumento em causa não deve ultrapassar o valor de 5.0. Viajaria tranquilo nesta companhia de aviação se soubesse que a nova política de manutenção consistia em regular o aparelho logo que decorressem 15 dias desde a intervenção anterior?

PROBLEMA 14.6

Para estudar os efeitos exercidos pela temperatura de reacção (T) e pela presença de um determinado catalisador no rendimento (R) de um processo químico, efectuou-se uma experiência cujos resultados se incluem na tabela seguinte.

Presença do catalisador	Temperatura [$^{\circ}\text{C}$]	Rendimento [%]
não	20	70
não	22	71
não	26	80
não	29	82
sim	21	72
sim	23	78
sim	27	88
sim	28	89

Modele a relação entre o rendimento do processo e os dois factores analisados (temperatura de reacção e ausência/presença do catalisador).

Na realização do teste de significância, admita que, pela natureza da reacção envolvida,

- O rendimento não pode baixar com o aumento da temperatura e
- A presença do catalisador não pode contribuir para a diminuição do efeito da temperatura sobre o rendimento.

PROBLEMA 14.7

Na tabela seguinte apresentam-se as vendas trimestrais (VV) de um produto que foi lançado no primeiro trimestre (trimestre $t = 1$).

<i>Trimestre (t)</i>	<i>Volume de vendas (VV)</i>
1	1206
2	11356
3	36888
4	39636
5	53887
6	77806
7	91627
8	80147
9	89646
10	101677

Efectue previsões das vendas para o trimestre 11, calculando o intervalo de previsão a 95%.

PROBLEMA 14.8

O preço de um veículo usado é função de diversos factores, entre os quais figuram os seguintes: o modelo do veículo em causa, o preço do veículo novo, a idade do veículo usado, o seu estado de conservação e a relação procura/oferta no mercado de veículos em segunda mão.

Considerando apenas o caso de veículos devidamente conservados e supondo que o factor procura/oferta do mercado em segunda mão não varia significativamente, o preço de um veículo usado de determinado modelo pode ser explicado através de uma relação do tipo

$$p = \frac{PVU}{PVN} = f(I)$$

onde

PVU : preço do veículo usado

PVN : preço do veículo novo

I : idade do veículo usado.

Na tabela seguinte apresentam-se os valores de p e I para veículos de um determinado modelo, que se supõem obtidos nas condições acima enunciadas.

$I[\text{anos}]$	1	2	3	4	5	6	7	8	9
$p = PVU/PVN$	0.843	0.753	0.580	0.520	0.452	0.414	0.346	0.264	0.241

- (i) Começando por representar graficamente os dados, defina a função que melhor traduz a relação entre p e I .
- (ii) Calcule a probabilidade de p ultrapassar o valor 0.65, para um veículo que, tendo 3 anos de idade, esteja nas condições definidas anteriormente.