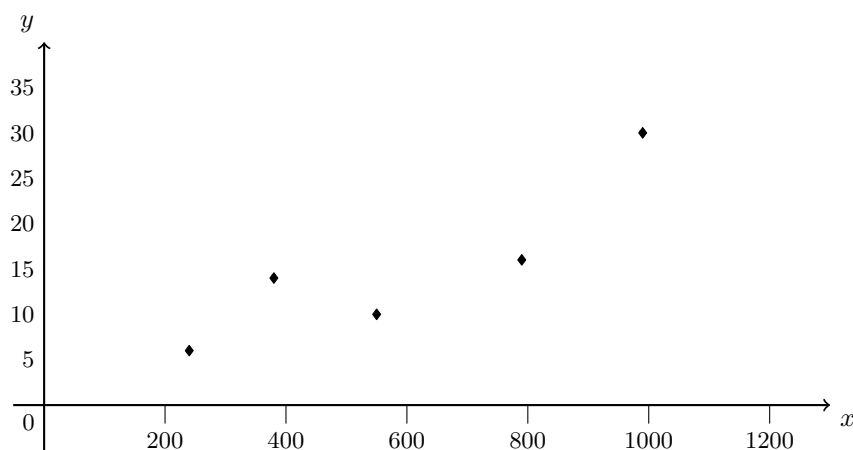


M.A.C.S. (10.º ano)  
**Análise de dados bivariados**

Exercícios de Provas Nacionais



1. O diagrama de dispersão representado na figura seguinte mostra uma forte associação linear positiva entre o preço das viagens de avião vendidas na agência de viagens Ir&Voltar, num determinado período de tempo, e as horas de voo despendidas na viagem de ida e volta.



Em cada uma das opções seguintes, são dados um valor de  $r$ , coeficiente de correlação linear, e a equação de uma reta.

Em qual das opções poderão estar representados o valor de  $r$  e uma equação da reta de regressão linear da distribuição representada na figura anterior?

- |                                       |                                       |
|---------------------------------------|---------------------------------------|
| (A) $r = -0,92$<br>$y = 0,03x - 0,86$ | (B) $r = -0,2$<br>$y = -0,03x - 0,86$ |
| (C) $r = 0,92$<br>$y = -0,03x - 0,86$ | (D) $r = 0,89$<br>$y = 0,03x - 0,86$  |

2. A venda de bilhetes para o concerto da banda *BigBand* gerou tanta procura que, na véspera do primeiro dia de venda, se formou fila para a aquisição de bilhetes à porta da bilheteira.

Ao longo do primeiro dia de venda dos bilhetes, as pessoas foram questionadas sobre o número de horas que permaneceram na fila antes da abertura da bilheteira ( $x$ ) e sobre o tempo, em horas, que decorreu desde a abertura da bilheteira até terem adquirido os bilhetes ( $y$ ).

A tabela seguinte apresenta as respostas dadas por sete das pessoas questionadas: A, B, C, D, E, F e G.

Pessoa	$x$ (horas)	$y$ (horas)
A	30	0,5
B	24	1
C	22,5	2
D	18	4
E	12	8
F	8	9
G	3	12

Admita que a relação entre as variáveis  $x$  e  $y$ , da tabela anterior, é bem aproximada por uma regressão linear, na forma  $y = ax + b$ .

Determine qual poderá ter sido o tempo que decorreu desde a abertura da bilheteira até à aquisição dos bilhetes por parte de uma pessoa que tenha estado seis horas na fila antes da abertura da bilheteira.

Apresente o resultado em horas, arredondado às unidades.

Na sua resposta, apresente a equação da reta de regressão, com os valores de  $a$  e de  $b$  arredondados com três casas decimais.

Exame – 2020, 1.ª Fase



3. Na tabela seguinte, estão registados, para cada um dos filmes, A, B, C, D, E, F e G, o custo de produção, em milhares de euros, e o número de espectadores, em milhares, que teve nas semanas de exibição em Portugal.

Filme	Custo de produção (em milhares de euros) $x$	Número de espectadores (em milhares) $y$
A	435	99
B	379	84
C	65	16
D	60	13
E	276	75
F	59	12
G	43	9

Admita que a relação entre as variáveis  $x$  e  $y$  da tabela anterior é, aproximadamente, linear, sendo modelada pela reta de regressão de equação da forma  $y = ax + b$ .

Estime o custo de produção de um filme com 52,5 milhares de espectadores.

Apresente o resultado, em milhares de euros, arredondado às unidades.

Na sua resposta, utilize os valores de  $a$  e de  $b$  com três casas decimais.

Exame – 2017, Ép. especial



4. A maratona é a última prova em diversos encontros desportivos.

Existem diversos estudos estatísticos sobre esta prova, que envolvem dados como o tempo de conclusão da prova, a frequência cardíaca dos atletas no final da prova ou as condições climatéricas.

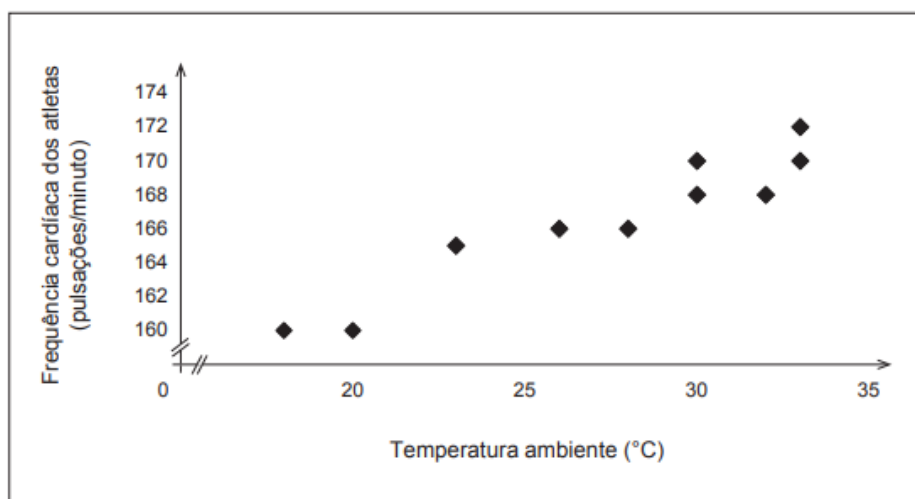
Na tabela seguinte, estão parcialmente registados os dados referentes à frequência cardíaca do atleta vencedor da maratona, no momento em que acaba a prova, nas edições do Encontro Desportivo Internacional dos últimos dez anos.

Sabe-se ainda que:

- a média dos valores constantes da tabela seguinte é 166,5;
- $P$  representa as pulsações por minuto do atleta que venceu a maratona na edição de 2012.

Ano	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Frequência cardíaca (pulsações/minuto)	165	166	166	168	170	170	$P$	160	160	168

Na figura seguinte, está representado o diagrama de dispersão da temperatura ambiente, em graus Celsius ( $^{\circ}\text{C}$ ), e da frequência cardíaca, em pulsações por minuto, dos atletas referidos na tabela anterior, no final da prova.



4.1. Considere as seguintes afirmações.

*O atleta vencedor em 2012 terminou a maratona com uma frequência cardíaca entre 169 e 171 pulsações por minuto.*

*A mediana das frequências cardíacas é 170 pulsações por minuto.*

*O coeficiente de correlação linear entre a temperatura ambiente e a frequência cardíaca dos atletas pode ser  $-0,85$ .*

Elabore uma pequena composição na qual justifique, com base nos dados apresentados, que as três afirmações são falsas.

4.2. A reta de equação  $y = 0,71x + 147,1$  é a que melhor se ajusta ao diagrama de dispersão apresentado na figura anterior, em que  $y$  representa a frequência cardíaca do atleta vencedor, em pulsações por minuto, e  $x$  representa a temperatura ambiente no final da prova, em  $^{\circ}\text{C}$ .

Conclua, atendendo à reta ajustada ao diagrama, se  $31,7^{\circ}\text{C}$  será um valor admissível da temperatura ambiente registada no final da maratona, no ano de 2006.

Exame – 2016, 2.<sup>a</sup> Fase



5. O número,  $N$ , de sócios do Grupo Desportivo de Altivo (GDA), em função do número de anos,  $t$ , que decorreram após o final de 1980, é bem aproximado pelo modelo  $N(t) = at + b$ , com arredondamento às unidades, sendo  $a$  e  $b$  duas constantes.

A tabela seguinte apresenta o número de sócios do GDA no final de alguns anos.

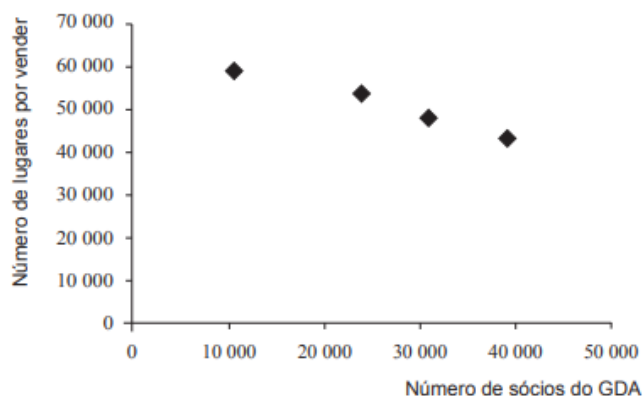
Ano	N.º de sócios
1980	10 585
1990	23 150
2000	30 981
2010	39 125

- 5.1. Estime o número de sócios que o GDA atingiu no final de 2005 com base no modelo  $N(t)$ , admitindo que este se mantém válido entre o final de 1980 e o final de 2010.

Apresente o resultado com arredondamento às unidades.

Caso proceda a arredondamentos nos cálculos intermédios, conserve, no mínimo, uma casa decimal.

- 5.2. O diagrama de dispersão representado na figura seguinte mostra uma forte associação linear negativa entre o número de sócios do do Grupo Desportivo de Altivo (GDA), e o número de lugares por vender nos jogos de futebol.



Relativamente à figura anterior, apenas uma das opções seguintes está correta.

Opção I	Opção II	Opção III
$r = -0,987$	$r = -0,987$	$r = -0,087$
$a = 1,744$	$a = -0,558$	$a = -1,744$
$b = 10\,354,123$	$b = 65\,346,152$	$b = 65\,346,152$

Em cada uma das opções,  $r$  representa o coeficiente de correlação linear, e  $a$  e  $b$  representam os parâmetros da reta de regressão linear  $y = ax + b$

Elabore uma composição na qual:

- identifique a opção correta;
- apresente uma razão para rejeitar cada uma das restantes opções.

6. Nas tabelas seguintes encontram-se o número total de pontos de acesso à rede postal e a densidade postal (número de habitantes / número de pontos de acesso), de 2001 a 2009, em Portugal.

Ano	Número total de pontos de acesso à rede postal	Ano	Densidade postal (habitantes / pontos de acesso)
2001	19 775	2001	471,3
2002	21 758	2002	481,4
2003	21 008	2003	501,2
2004	20 630	2004	512,5
2005	20 457	2005	517,9
2006	20 215	2006	525,5
2007	19 897	2007	534,1
2008	19 155	2008	554,8
2009	18 394	2009	563,2

Ao serem representados os dados que constam das tabelas anteriores num diagrama de dispersão, verificou-se que o ponto (19 775; 471,3), correspondente ao ano 2001, se encontrava fora do contexto dos restantes, se se pretendesse ajustar um modelo de regressão linear, sendo considerado um *outlier*.

Explique o efeito da exclusão do *outlier* no valor do coeficiente de correlação linear e na reta de regressão quando se pretende fazer previsões.

Na sua resposta, deve:

- apresentar os dados das tabelas num diagrama de dispersão, incluindo o ano 2001;
- determinar o valor do coeficiente de correlação linear entre as variáveis número total de pontos de acesso à rede postal ( $x$ ) e densidade postal ( $y$ ), incluindo o ano 2001;
- apresentar a simulação do efeito da exclusão do ano 2001 no valor do coeficiente de correlação linear e na reta de regressão  $y = ax + b$ ;
- referir o efeito da exclusão do *outlier* quando se pretende fazer previsões.

Apresente os resultados com arredondamento às milésimas.

Exame – 2013, 2.ª Fase



7. O professor da disciplina de Matemática Aplicada às Ciências Sociais da escola de Xisto estudou a existência de uma correlação linear entre as classificações dos alunos na disciplina de Matemática Aplicada às Ciências Sociais no final do 3.º período de 2010 (CI) e as classificações desses mesmos alunos no exame nacional da disciplina de Matemática Aplicada às Ciências Sociais (CE).  
Os dados recolhidos encontram-se organizados na tabela seguinte.

N.º do aluno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
(CI) $x$	9	9	10	13	14	9	11	11	14	8	14	16	16	19	14	11	16	10
(CE) $y$	7	8,2	9,5	12,1	13,7	7,5	14,2	10,6	14,5	6,3	12,1	17,2	14,8	4,2	11,7	9,5	15,2	8,5

- 7.1. Uma aluna, após calcular o coeficiente de correlação dos dados organizados na tabela anterior, afirma:  
«A correlação linear entre as classificações dos alunos da escola de Xisto na disciplina de Matemática Aplicada às Ciências Sociais no final do 3.º período de 2010 (CI) e as classificações obtidas no exame (CE) é fraca. No entanto, se excluirmos as classificações do aluno número 14, essa correlação linear aumenta substancialmente.»

Justifique a veracidade da afirmação da aluna.

Na sua resposta, deve:

- determinar o coeficiente de correlação linear incluindo as classificações do aluno número 14;
- determinar o coeficiente de correlação linear excluindo as classificações do aluno número 14;
- relacionar os valores dos dois coeficientes, justificando a veracidade da afirmação da aluna.

Apresente os coeficientes de correlação com arredondamento às milésimas.

- 7.2. Admita que a reta de regressão  $y = 1,0927x - 1,8476$ , no intervalo  $[8,18]$ , se ajusta, aproximadamente, aos pontos do diagrama de dispersão das variáveis:

$x$ : «classificação dos alunos da escola de Xisto na disciplina de Matemática Aplicada às Ciências Sociais no final do 3.º período de 2010 (CI)»

$y$ : «classificação dos alunos da escola de Xisto na disciplina de Matemática Aplicada às Ciências Sociais no exame nacional de 2010 (CE)»

Calcule a estimativa do valor da classificação no exame da disciplina de Matemática Aplicada às Ciências Sociais de um aluno da escola de Xisto que, no final do 3.º período de 2010, tenha obtido a classificação de 12.

Apresente o resultado com arredondamento às décimas.

Caso proceda a arredondamentos nos cálculos intermédios, conserve, no mínimo, quatro casas decimais.



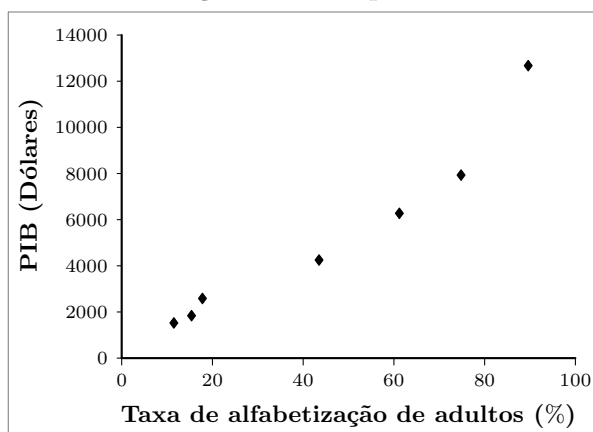
8. Um grupo de alunos está interessado em estudar o grau de desenvolvimento de sete países. As variáveis estudadas foram analisadas individualmente e através de associações entre elas.

A tabela e o diagrama de dispersão seguintes apresentam, para os sete países, num determinado ano, duas das variáveis: a Taxa de Alfabetização de Adultos (TAA), em percentagem; e o Produto Interno Bruto per capita (PIB), em dólares.

Tabela

País	TAA ( $x$ )	PIB ( $y$ )
A	15,4	1839
B	74,8	7932
C	43,5	4251
D	17,8	2586
E	11,5	1524
F	89,6	12 674
G	61,2	6275

Diagrama de dispersão



Nos dois itens seguintes, pode recorrer à calculadora. Sempre que recorrer a estatísticas obtidas na calculadora (média, desvio padrão, coeficiente de correlação, declive e ordenada na origem de uma reta de regressão, etc.), apresente a(s) lista(s) que introduziu na calculadora para a(s) obter.

- 8.1. Admita um modelo em que a associação entre as variáveis TAA ( $x$ ) e PIB ( $y$ ) é, aproximadamente, linear.

Considere a equação  $y = ax + b$ , da reta de regressão linear, e também os dados da tabela anterior.

Indique os valores de  $a$ , de  $b$  e do coeficiente de correlação linear,  $r$ , recorrendo à calculadora.

Apresente os resultados com quatro casas decimais.

- 8.2. Ao analisar os dados que constam da tabela anterior, um economista reparou que o PIB do país F é de 12 674 dólares e que a TAA é de 89,6%. Segundo alguns analistas, este ponto encontra-se fora do contexto dos restantes quando se pretende ajustar um modelo de regressão linear, sendo considerado um *outlier*.

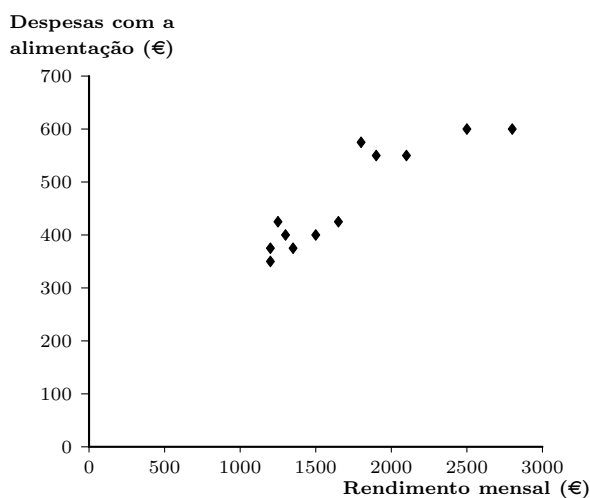
A partir dos dados apresentados na tabela anterior e no Diagrama de dispersão, ao excluir os valores relativos ao país F, o valor do coeficiente de correlação linear entre as variáveis  $x$  e  $y$  é, aproximadamente,  $r = 0,9937$ , e a equação da reta de regressão linear, considerando os coeficientes com quatro casas decimais, é  $y = 96,6098x + 457,8482$ .

Num pequeno texto, explique o efeito da exclusão do *outlier* no valor do coeficiente de correlação linear e na reta de regressão, quando se pretende fazer previsões.





9. As despesas de um agregado familiar com a alimentação dependem de muitos fatores. Do ponto de vista sociológico, pode ser estudada a relação entre as despesas mensais com a alimentação e o rendimento mensal. Para conhecer esta relação, recolheram-se, aleatoriamente, os dados relativos a doze agregados familiares. Obtiveram-se os dados representados no diagrama de dispersão e constantes da tabela.



Agregado familiar	Rendimento mensal (€) $x$	Despesas com a alimentação (€) $y$
A	1250	425
B	2800	600
C	1900	550
D	1650	425
E	1300	400
F	1800	575
G	1200	375
H	2500	600
I	1350	375
J	2100	550
K	1200	350
L	1500	400

- 9.1. Admita que a associação entre as variáveis  $x$  e  $y$  é linear.

Classifique o tipo e o grau de associação entre as variáveis  $x$  e  $y$ , a partir da interpretação do valor do coeficiente de correlação.

Na sua resposta deve:

- apresentar o valor do coeficiente de correlação, com arredondamento às décimas;
- classificar o tipo e o grau de associação linear entre as variáveis;
- justificar a forma como classificou o tipo e o grau de associação linear entre as variáveis  $x$  e  $y$ .

- 9.2. Para responder aos itens seguintes, considere a equação  $y = ax + b$ , da reta de regressão linear das variáveis em estudo, e também os dados da tabela.

- i. Determine os valores de  $a$  e  $b$ , **recorrendo à calculadora**.

Apresente os resultados com quatro casas decimais.

- ii. Faça a estimativa do valor das despesas mensais com a alimentação de um agregado familiar cujo rendimento mensal é de €1750.

Apresente o resultado, em euros, com arredondamento às unidades.

Caso não tenha respondido ao item anterior, e somente nesse caso, considere  $a = 0,1705$  e  $b = 177,0151$ .

Exame – 2009, 1.ª Fase



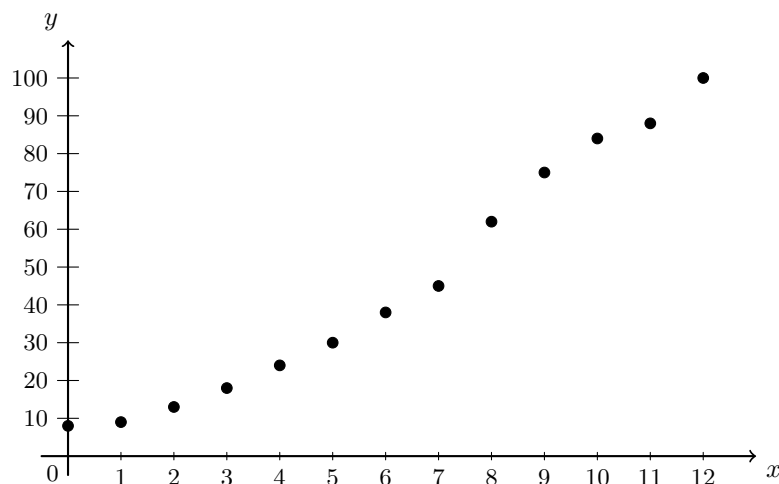
10. Na atualidade, há uma crescente preocupação com a preservação da natureza, nomeadamente, quanto à necessidade de proteger espécies que se encontram em vias de extinção.

Considere que uma certa espécie animal se encontrava em vias de extinção. Para a proteger, tomaram-se medidas protecionistas, designadamente, a criação de uma área protegida, no seu habitat natural.

Admita que, no início, apenas existiam 8 animais da espécie nessa área. A tabela seguinte traduz a contagem anual do número de animais nela existentes.

Anos decorridos desde a criação da área protegida ( $x$ )	Número de animais existentes na área protegida ( $y$ )
0	8
1	9
2	13
3	18
4	24
5	30
6	38
7	45
8	62
9	75
10	84
11	88
12	100

O gráfico seguinte representa os dados da tabela, através de uma nuvem de pontos.



Com recurso à calculadora, determine o modelo de regressão linear, de equação  $y = ax + b$ , que se ajusta à nuvem de pontos apresentada.

Indique os valores de  $a$  e de  $b$ , com uma aproximação às décimas.



11. Num dos muitos *sites* em que se joga xadrez *online*, na Internet, a entrada de um jogador é condicionada pelo gestor do *site*, com probabilidade fixa igual a 0,8, em cada tentativa de entrada na sala de jogo.

O gestor do *site* decidiu estudar a evolução do número de jogadores de xadrez, desde o lançamento do *site* até à sexagésima semana, para o que foi registando o número de jogadores, de cinco em cinco semanas, tendo obtido a tabela seguinte:

Tempo (em semanas) ( $x$ )	Número de jogadores (em milhares) ( $y$ )
5	20
10	46
15	58
20	82
25	110
30	128
35	136
40	163
45	170
50	194
55	210
60	245

Represente **na sua calculadora** o diagrama de dispersão dos dados e determine a equação da reta de regressão linear,  $y = ax + b$ , indicando os valores de  $a$  e  $b$  com uma **aproximação às centésimas**.

Transcreva para a sua folha de prova um esboço, no mesmo referencial, dos gráficos obtidos (diagrama de dispersão e reta de regressão linear).

Exame – 2007, 1.<sup>a</sup> Fase

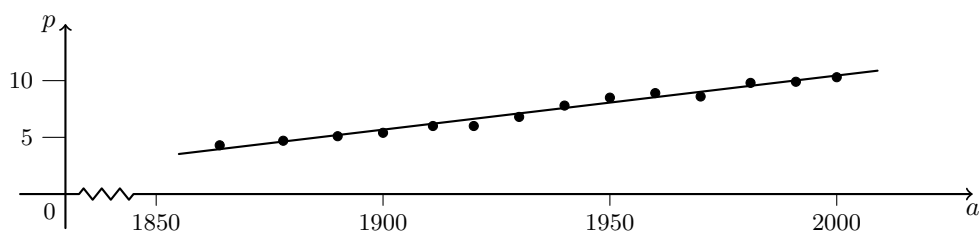


12. Na tabela seguinte, estão alguns dados sobre a população residente em Portugal, desde 1864 até ao final do século XX.

Ano ( $a$ )	População, em milhões ( $p$ )
1864	4,3
1878	4,7
1890	5,1
1900	5,4
1911	6,0
1920	6,0
1930	6,8
1940	7,8
1950	8,5
1960	8,9
1970	8,6
1981	9,8
1991	9,9
2000	10,3

Na figura abaixo está representado o diagrama de dispersão relativo aos dados apresentados na tabela, assim como a respetiva reta de regressão, cuja equação é

$$p = 0,0477a - 84,95$$



- 12.1. Com recurso à calculadora, determine o coeficiente de correlação linear das variáveis  $a$  e  $p$ , tendo em conta a tabela apresentada. Apresente o valor pedido na forma de dízima, arredondado às milésimas. Explique como procedeu e interprete esse valor, tendo em conta o diagrama de dispersão apresentado.
- 12.2. Explique porque razão o modelo linear acima apresentado (reta de regressão) não pode ser adequado para:
- estimar o número de habitantes, em Portugal, há alguns séculos (três ou mais);
  - prever a evolução da população portuguesa, a muito longo prazo (relacione uma tal previsão com os recursos, alimentares e outros, necessariamente limitados).

Exame – 2006, 2.<sup>a</sup> Fase

