

4. Na Tabela 5 e na Tabela 6, encontram-se, respetivamente, 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.

Tabela 5

Ano	Número total de pontos de acesso à rede postal
2001	19 775
2002	21 758
2003	21 008
2004	20 630
2005	20 457
2006	20 215
2007	19 897
2008	19 155
2009	18 394

Tabela 6

Ano	Densidade postal (habitantes / pontos de acesso)
2001	471,3
2002	481,4
2003	501,2
2004	512,5
2005	517,9
2006	525,5
2007	534,1
2008	554,8
2009	563,2

- 4.1. Ao serem representados os dados que constam da Tabela 5 e da Tabela 6 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 5 e 6 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.

- 4.2. Em 2004, num determinado concelho com doze pontos de acesso à rede postal, a média do número de habitantes por cada ponto de acesso foi 512,5. No seguinte conjunto de números relativos aos habitantes servidos por cada um dos doze pontos de acesso nesse concelho, no ano de 2004, falta o número de habitantes, a , servidos por um ponto de acesso.

531 518 481 535 493 500 490 a 525 502 493 550

Determine o valor do desvio padrão do número de habitantes servidos por cada um dos pontos de acesso desse concelho, em 2004.

Comece por calcular o valor de a .

Apresente o valor do desvio padrão arredondado às unidades.

- 4.3. O Portal do Consumidor tem recebido queixas pelo aumento, verificado entre 2001 e 2009, da densidade postal (número de habitantes / número de pontos de acesso). Em 2012, uma recolha aleatória em 200 desses pontos indicou uma média amostral de \bar{x} habitantes por cada ponto de acesso e um desvio padrão amostral de s habitantes.

Tendo em conta os valores de \bar{x} e de s , obteve-se o intervalo $546, 554[$ para estimar o número médio de habitantes servidos por cada ponto de acesso à rede postal, em 2012, com uma confiança de 90%

Determine os valores de \bar{x} e de s , com os quais se obteve o intervalo $546, 554[$

Apresente os valores de \bar{x} e de s arredondados às unidades.

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

4.1 Inserindo na calculadora gráfica as listas dos dados das duas tabelas, incluindo os valores referentes ao ano 2001, temos:

Número total de pontos de acesso à rede postal	Densidade postal (habitantes / posto de acesso)
19775	471,3
21758	481,4
21008	501,2
20630	512,5
20457	517,9
20215	525,5
19897	534,1
19155	554,8
18394	563,2

De onde se obtém o valor do coeficiente de correlação de -0,728:

LinReg
 $y = ax + b$
 $a = -0,022450583$
 $b = 970,2159704$
 $r^2 = 0,5295981746$
 $r = -0,7277349618$

Retirando os dados relativos ao ano de 2001, e refazendo o cálculo obtemos o valor do coeficiente de correlação de -0,992.

LinReg
 $y = ax + b$
 $a = -0,0253782513$
 $b = 1036,192861$
 $r^2 = 0,9832699368$
 $r = -0,9915996858$

Pela análise dos dados obtidos, verifica-se que a exclusão do *outlier* indica uma correlação mais forte, ou seja um coeficiente de correlação mais distante de 0. Assim, conclui-se que, quando se exclui o *outlier*, o ajuste da reta de regressão à nuvem de pontos é maior, e as previsões serão mais fiáveis.

4.2 Podemos calcular o valor de a , recorrendo ao valor da média:

$$\frac{531 + 518 + 481 + 535 + 493 + 50 + 490 + a + 525 + 502 + 493 + 550}{12} = 512,5 \Leftrightarrow$$

$$\Leftrightarrow \frac{5618 + a}{12} = 512,5 \Leftrightarrow$$

$$\Leftrightarrow 5618 + a = 512,5 \times 12 \Leftrightarrow$$

$$\Leftrightarrow a = 6150 - 5618 \Leftrightarrow$$

$$\Leftrightarrow a = 532$$

Inserindo os valores dados (e o valor determinado para a) numa lista da calculadora gráfica, e fazendo os cálculos relativos a uma variável, temos:

L1	L2	L3	1
490			
532			
535			
502			
493			
550			
L1(13) =			

1-Var Stats
List: L1
FreqList:
Calculate

1-Var Stats
 $\bar{x} = 512,5$
 $\sum x = 6150$
 $\sum x^2 = 3157222$
 $Sx = 22,04746945$
 $\sigma x = 21,10884491$
 $\downarrow n = 12$

Pelo que se conclui que, o valor do desvio padrão é de 21.

4.3. Sabemos que $I =]546,554[= \left[\bar{x} - z \frac{s}{\sqrt{n}}, \bar{x} + z \frac{s}{\sqrt{n}} \right]$, em que $n = 200$ e $z = 1,645$

Sabemos ainda que $\bar{x} = \frac{546+554}{2} = 550$ por ser o valor médio do intervalo.

Recorrendo a um dos extremos do intervalo, por exemplo o extremo superior, temos:

$\bar{x} + z \frac{s}{\sqrt{n}} = 554$ e substituindo os valores indicados, obtemos:

$$550 + 1,645 \frac{s}{\sqrt{200}} = 554 \Leftrightarrow 1,645 \frac{s}{\sqrt{200}} = 4 \Leftrightarrow$$

$$\Leftrightarrow \frac{s}{\sqrt{200}} \approx 2,432 \Leftrightarrow$$

$$\Leftrightarrow s \approx 34,388$$

Logo, os valores são $\bar{x} = 550$ e $s \approx 34$