

Departamento *Matemática*

Unidade Curricular *Probabilidades e Estatística*

Curso *Engenharia do Ambiente*

Ano *2º*

Semestre *1º*

Ano Lectivo *2007/2008*

Folha Nº 8: Análise de Regressão e de Correlação

1. 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

$$\sum_{i=1}^n x_i = 134 \quad \sum_{i=1}^n y_i = 167.7 \quad \sum_{i=1}^n x_i^2 = 1424 \quad \sum_{i=1}^n y_i^2 = 2220.21 \quad \sum_{i=1}^n x_i y_i = 1769.6$$

- a) Construa o diagrama de dispersão e avalie se o grau de relacionamento linear entre as variáveis X e Y é forte ou fraco. Conclua se é razoável aplicar um modelo de regressão linear.
- b) Determine a equação de regressão estimada pelo método dos mínimos quadrados.
- 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) Calcule o coeficiente de correlação e interprete o seu valor. Qual a percentagem de variabilidade nos valores fornecidos pelo instrumento explicada linearmente por variações na verdadeira concentração de ácido láctico?
- e) Interprete o valor do coeficiente de regressão estimado - b_1 .
- f) Teste a hipótese nula de que a verdadeira concentração de ácido láctico não exerce uma influência significativa sobre o valor fornecido pelo instrumento, a um nível de significância de 1% (considere $SSE=20.94607$).
- g) Há evidência para afirmar que a regressão é significativa? (use o mesmo nível de significância da alínea anterior)

Unidade Curricular *Probabilidades e Estatística* **Ano** 2^o **Semestre** 1^o **Ano Lectivo** 2007/2008

2. Numa determinada experiência pretende-se estimar o nível de octanas em gasolinas que contêm diferentes concentrações de componentes A e B. Seja Y o nível de octanas, X_1 a percentagem de componente A e X_2 a percentagem de componente B. Na tabela seguinte estão registados os dados resultantes de 16 observações.

X_1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
X_2	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
Y	96.3	95.7	99.9	99.4	95.1	97.8	99.3	104.9	96.2	100.1	103.2	104.3	97.8	102.2	104.7	108.8

$$\sum_{i=1}^n y_i = 1605.7 \quad \sum_{i=1}^n x_{1i} = 56 \quad \sum_{i=1}^n x_{2i} = 56 \quad \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 = 20 \quad \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 = 20$$

a) Calcule a equação de regressão estimada pelo método dos mínimos quadrados, que permita estimar o nível de octanas em gasolinas com concentrações de componentes A e B conhecidas, tendo em conta que:

$$(X^T X)^{-1} = \frac{1}{6400} \begin{bmatrix} 8240 & -1120 & -1120 \\ -1120 & 320 & 0 \\ -1120 & 0 & 320 \end{bmatrix} \quad X^T y = \begin{bmatrix} 1605.7 \\ 5656.6 \\ 5673.6 \end{bmatrix} \quad (X^T X)^{-1} (X^T y) = \begin{bmatrix} 84.55375 \\ 1.8325 \\ 2.6825 \end{bmatrix}$$

- b) Preveja qual será o nível de octanas para uma gasolina com 3% de componente A e 5% de componente B.
- c) Interprete os valores dos coeficientes de regressão.
- d) Tendo em conta que a soma dos quadrados totais é igual a 236.25937 e que a soma dos quadrados dos resíduos em relação à equação de regressão estimada é igual a 25.18212, calcule o coeficiente de correlação e de determinação e interprete os seus valores.
- e) Teste a hipótese de que β_2 é inferior a 3, a um nível de significância de 5%.
- f) Teste a significância global do modelo (Utilize um nível de significância de 1%)

Unidade Curricular

Probabilidades e Estatística

Ano 2^o

Semestre 1^o

Ano Lectivo

2007/2008

3. Os dados apresentados no quadro seguinte representam as vendas (Y), em milhares de Euros, efectuadas por 10 empregados de uma dada empresa, o nº de anos de experiência de cada vendedor (X_1) e o respectivo score no teste de inteligência (X_2).

Vendedor	Vendas (Y) (em milhares de Euros)	Anos de experiência como vendedor (X_1)	Score no teste de inteligência (X_2)
1	9	6	3
2	6	5	2
3	4	3	2
4	3	1	1
5	3	4	1
6	5	3	3
7	8	6	3
8	2	2	1
9	7	4	2
10	4	2	2

$$(X^T X)^{-1} = \frac{1}{944} \begin{bmatrix} 776 & -56 & -240 \\ -56 & 60 & -80 \\ -240 & -80 & 264 \end{bmatrix} \quad (X^T Y) = \begin{bmatrix} 51 \\ 214 \\ 116 \end{bmatrix} \quad (X^T X)^{-1} (X^T y) = \begin{bmatrix} -0.262712 \\ 0.745763 \\ 1.338983 \end{bmatrix}$$

Pretende-se determinar se o sucesso das vendas pode ser medido em função das duas variáveis explicativas X_1 e X_2 .

- a) Calcule a equação de regressão estimada.
- b) Diga qual o volume de vendas previsto para um vendedor com 4 anos de experiência e com score 3 no teste de inteligência.
- c) Interprete os valores dos coeficientes de regressão.
- d) Teste a hipótese nula de que o coeficiente de regressão associado a X_1 seja zero, ao nível de significância de 5%. Comente o resultado obtido.
- e) Posso rejeitar a hipótese nula de que o coeficiente β_0 é nulo, ao nível de significância de 5%? Diga qual o menor nível de significância que me permite rejeitar a hipótese anterior.
- f) Sabendo que a soma dos quadrados da regressão é 41.41695, calcule o coeficiente de correlação e de determinação e interprete os seus valores.
- g) Teste a hipótese nula da regressão não ser significativa. Use um nível de significância de 5%. Que pode concluir?

Unidade Curricular

Probabilidades e Estatística

Ano 2^o

Semestre 1^o

Ano Lectivo

2007/2008

4. A tabela seguinte indica as idades (X) e os gastos diários em centenas de escudos (Y) de 10 pessoas.

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

- Construa o diagrama de dispersão e avalie se o grau de relacionamento linear entre as variáveis X e Y é forte ou fraco. Conclua se é razoável aplicar um modelo de regressão linear.
- Determine a equação de regressão estimada pelo método dos mínimos quadrados.
- Avalie o gasto diário de uma pessoa com 24 anos. Interprete o valor encontrado.
- Calcule o coeficiente de correlação e de determinação e interprete os seus valores.
- Interprete o valor do coeficiente de regressão estimado - b_1 .
- Teste a hipótese de que o declive populacional excede o valor 0.3, a um nível de significância de 5%.
- Teste a hipótese nula do coeficiente de regressão associado à variável X ser nulo, a um nível de significância de 0.1%. Que pode dizer acerca da significância da regressão?

Unidade Curricular

Probabilidades e Estatística

Ano 2^o

Semestre 1^o

Ano Lectivo

2007/2008

5. Pretende-se encontrar um modelo linear a partir do qual seja possível prever o gasto de gasolina de um automóvel baseado no seu peso e na temperatura ambiente. Seja Y as milhas percorridas com um galão de gasolina (1milha=1609 metros; 1 galão=4.545 litros), X_1 o peso do automóvel (em toneladas) e X_2 a temperatura ambiente (em °F). Os dados obtidos estão registados na tabela que se segue.

Nº do automóvel	Milhas percorridas por cada galão de gasolina (Y)	Peso do automóvel em toneladas (X_1)	Temperatura ambiente em °F (X_2)
1	17.9	1.35	90
2	16.5	1.9	30
3	16.4	1.7	80
4	16.8	1.8	40
5	18.8	1.3	35
6	15.5	2.05	45
7	17.5	1.6	50
8	16.4	1.8	60
9	15.9	1.85	65
10	18.3	1.4	30

$$\sum_{i=1}^n y_i = 170 \quad \sum_{i=1}^n x_{1i} = 16.75 \quad \sum_{i=1}^n x_{2i} = 525 \quad \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 = 0.58 \quad \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 = 3912.5$$

- Indique a matriz significativa do modelo (X) e o vector das observações da variável dependente (y).
- Escreva na forma matricial o sistema de equações resultante da aplicação do modelo às 10 observações.
- Sabendo que:

$$(X^T X)^{-1} = \begin{bmatrix} 6.070769 & -3.02588 & -0.0171888 \\ -3.02588 & 1.738599 & 0.002166306 \\ -0.0171888 & 0.002166306 & 0.0002582903 \end{bmatrix} \quad X^T y = \begin{bmatrix} 170 \\ 282.405 \\ 8887 \end{bmatrix} \quad (X^T X)^{-1} (X^T y) = \begin{bmatrix} 24.748874 \\ -4.159335 \\ -0.014895 \end{bmatrix}$$

Calcule a equação de regressão estimada pelo método dos mínimos quadrados.

- Estime o gasto de um carro que pese 1.5 toneladas num dia em que a temperatura é de 70°F.
- Interprete os valores dos coeficientes de regressão b_1 e b_2 .
- Sabendo que a soma dos quadrados da regressão em relação à equação de regressão estimada é igual a 10.31965 e que a soma dos quadrados totais é 10.46, calcule o coeficiente de correlação e de determinação e interprete os seus valores.
- Averigüe se a variável temperatura ambiente exerce uma influência significativa sobre a quantidade de milhas percorridas por cada galão de gasolina. (Use um nível de significância de 0.1%)
- Teste a hipótese nula de que nenhuma das variáveis explicativas contribui para a explicação da variação em torno da média da variável dependente. (Use um nível de significância de 1%)

Unidade Curricular	<i>Probabilidades e Estatística</i>	Ano	2 ^o	Semestre	1 ^o	Ano Lectivo	2007/2008
---------------------------	-------------------------------------	------------	----------------	-----------------	----------------	--------------------	-----------

6. Uma determinada empresa possui para a produção de um certo bem dois tipos de máquinas alternativas (A e B). O director de produção da empresa pretende conhecer se existem vantagens, em termos de custo marginal, na utilização de uma ou de outra máquina. Com esse objectivo decidiu estudar o seguinte modelo de regressão: $Y_i = \beta_0 + \beta_1 x_i + E_i$, onde Y_i e X_i representam respectivamente o custo de produção (em unidades monetárias) e a quantidade produzida do bem no dia i (em centenas).

Após 12 ensaios, obtiveram-se os seguintes resultados para a máquina A:

$$(X^T X)^{-1} = \begin{bmatrix} 0.15 & -0.2 \\ -0.2 & 0.6 \end{bmatrix} \quad X^T Y = \begin{bmatrix} 25 \\ 13 \end{bmatrix}.$$

Para a máquina B o modelo de regressão linear que melhor se ajustou foi: $\hat{y}_i = 1.6315 + 2.1255x_i$.

- Com base nos resultados apresentados, indique o modelo de regressão estimado para a máquina A.
- Diga qual o significado do valor 2.1255 no modelo de regressão estimado para a máquina B.
- Para o modelo estimado para a máquina A, analise da forma que julgar mais conveniente se a variável “quantidade de bem produzido num dia” é estatisticamente significativa (utilize um nível de significância de 5% e suponha que a soma dos quadrados dos resíduos é 0.169).
- Admita que o director de produção prevê produzir este mês 200 unidades por dia para fazer face às encomendas. Sabendo que a qualidade de produção é idêntica nas duas máquinas, qual das máquinas recomendaria? Justifique convenientemente a sua resposta e interprete os valores em que se basear para dar a resposta.
- Suponha agora que para utilizar a máquina B é necessário dispor de mais um trabalhador na linha de fabrico e que o custo/dia de cada trabalhador é de 10 unidades monetárias. Em que medida é que este novo dado poderia influenciar a resposta que deu na alínea anterior?
- Com auxílio da representação gráfica das rectas de regressão, indique para que níveis de produção seria preferível utilizar a máquina A.

Unidade Curricular

Probabilidades e Estatística

Ano 2^o

Semestre 1^o

Ano Lectivo

2007/2008

7. Num estudo para determinar de que modo a habilidade para executar uma determinada tarefa complexa é influenciada pela quantidade de 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:

$$\bar{x} = 7.2 \quad \bar{y} = 45.6 \quad \sum_{i=1}^{15} x_i^2 = 811.2 \quad \sum_{i=1}^{15} y_i^2 = 31350.6 \quad \sum_{i=1}^{15} x_i y_i = 4867.6$$
$$SSE=62.824 \quad SST=160.2$$

- Determine a equação de regressão estimada pelo método dos mínimos quadrados.
- Os dados suportam a hipótese de pesquisa de que o aumento da duração do treino reduz o tempo de execução da tarefa? (use $\alpha=0.01$)
- Calcule o coeficiente de correlação e interprete o valor encontrado.
- O que pode dizer relativamente ao tempo de execução da tarefa de um indivíduo que tenha sido sujeito a 35 horas de treino?

Unidade Curricular

Probabilidades e Estatística

Ano 2^o

Semestre 1^o

Ano Lectivo

2007/2008

8. Para 20 casas de habitação que se encontravam à venda numa determinada cidade do Norte da América, foram registados dados relativos a:

X_1 : área total (em centenas de pés quadrados – ft²)

X_2 : valor da casa atribuído por avaliação de um perito (em milhares de dólares)

Y : preço de venda (em milhares de dólares)

Para relacionar o preço de venda com a área total da casa e com o valor atribuído a esta por avaliação de um perito resolveu-se considerar o seguinte modelo de regressão:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + E_i$$

Os dados recolhidos permitiram determinar:

$$(X^T X)^{-1} = \begin{bmatrix} 1.9961 & -0.08955 & -0.011459 \\ -0.08955 & 0.05118 & -0.017199 \\ -0.011459 & -0.017199 & 0.0067445 \end{bmatrix} \quad (X^T X)^{-1} (X^T y) = \begin{bmatrix} 11.87 \\ 2.6344 \\ 0.04518 \end{bmatrix}$$

$$SSE = 204.995$$

$$SST = 1237.87$$

- Diga, recorrendo a um indicador adequado, se é razoável aplicar o modelo de regressão linear para relacionar Y com X_1 e X_2 .
- Qual a percentagem de variação no preço de venda que não se consegue explicar linearmente através da área total e do valor atribuído à casa?
- Um especialista em compra e venda de imóveis afirma que o valor da casa atribuído por avaliação de um perito não influencia significativamente o preço de venda da casa. Comente esta afirmação. (Use um nível de significância de 5%)
- O mesmo especialista quer prever o preço de uma casa com área total 15.31 ft² e avaliada em 37.3 milhares de dólares. Que valor lhe indicaria e qual o seu significado?
- Uma casa com as mesmas características da alínea anterior (área total=15.31 ft² e avaliada em 37.3 milhares de dólares) foi vendida por 55 milhares de dólares. É este preço diferente do que sugeriu na alínea anterior? Se sim diga a que se deve essa diferença (na sua explicação refira-se a E_i do modelo considerado).

Unidade Curricular

Probabilidades e Estatística

Ano 2^o

Semestre 1^o

Ano Lectivo

2007/2008

9. Numa empresa pretende-se relacionar o volume de vendas (em quantidade) com o preço de um determinado produto e com o preço do mesmo produto praticado pelo concorrente mais importante da empresa. O modelo apresenta-se na forma $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + E$ em que:

Y: vendas da empresa

X₁: preços praticados pela empresa (em euros)

X₂: preços praticados pelo concorrente (em euros)

Apresentam-se a seguir os dados obtidos de 20 observações.

i	1	2	3	4	5	6	7	8	9	10
Y	168	183	192	204	228	240	216	192	180	156
X ₁	22,5	22,2	22,5	22,7	20,6	20,9	21,3	21,5	22,1	22,8
X ₂	20,5	20,8	23,5	24,9	23,9	22,8	21,7	20,4	19,5	18,3

i	11	12	13	14	15	16	17	18	19	20
Y	215	193	174	212	235	217	177	210	190	177
X ₁	20,5	21,6	22,8	21,5	20,4	21,3	22,8	20,9	21,9	22,7
X ₂	19,8	20,5	21,5	22,6	22,4	22,7	21,4	20,8	20,7	21,4

$$(X^T X)^{-1} (X^T y) = \begin{bmatrix} 532.019 \\ -22.014 \\ 6.756 \end{bmatrix} \quad SSR = 9705.832 \quad SSE = 509.118 \quad s_{\hat{\beta}_1} = s\sqrt{c_{11}} = 1.517$$

- Utilizando o método dos mínimos quadrados determine a equação de regressão estimada e interprete-a.
- Determine a percentagem da variação total das vendas que pode ser atribuída à variação dos preços praticados pela empresa e pelo concorrente.
- Teste se a relação entre as vendas e o preço da empresa é estatisticamente significativa. (Use um nível de significância de 5%)
- Teste a significância global do modelo. (Use um nível de significância de 5%)

Unidade Curricular	<i>Probabilidades e Estatística</i>	Ano	2 ^o	Semestre	1 ^o	Ano Lectivo	2007/2008
---------------------------	-------------------------------------	------------	----------------	-----------------	----------------	--------------------	-----------

Soluções da ficha n.º 8

1.b) $\hat{y} = 0.159483 + 1.227689x$;

c) $\hat{y} = \hat{\mu}_{Y/X=1} = 1.387$: estimativa para o valor médio de ácido láctico medido pelo instrumento quando a concentração conhecida é 1;

d) $r^2 = 0.97427$: cerca de 97.4% da variação observada nos valores da concentração de ácido láctico apresentados pelo instrumento, são explicados linearmente pela variação da verdadeira concentração de ácido láctico no sangue;

$r = 0.98705$: indica a existência de uma relação linear forte e positiva entre as duas variáveis em estudo;

e) Em média, um acréscimo de uma unidade na verdadeira concentração de ácido láctico no sangue, aumenta o valor da concentração fornecido pelo instrumento numa quantidade que pode ser estimada em 1.227689 unidades;

f) O valor observado da estatística do teste (T_{Obs}) é 26.1064, R.C. = $]-\infty, -2.878] \cup [2.878, +\infty[$.
Rejeita-se H_0 ;

g) Sim, pelo resultado obtido na alínea f).

2.a) $\hat{y} = 84.55375 + 1.8325x_1 + 2.6825x_2$;

b) $\hat{y} = \hat{\mu}_{Y/X_1=3, X_2=5} = 103.46125$;

d) $r = 0.94521$ e $r^2 = 0.89341$;

e) $T_{Obs} = -1.0202$, R.C. = $]-\infty, -1.771]$. Não se rejeita H_0 ;

f) $F_{Obs} = 54.48318$, R.C. = $[6.70, +\infty[$. Rejeita-se H_0 .

3.a) $\hat{y} = -0.262712 + 0.745763x_1 + 1.338983x_2$;

b) $\hat{y} = \hat{\mu}_{Y/X_1=4, X_2=3} = 6.737289$

d) $T_{Obs} = 2.861$, R.C. = $]-\infty, -2.365] \cup [2.365, +\infty[$. Rejeita-se H_0

Unidade Curricular	Probabilidades e Estatística	Ano	2º	Semestre	1º	Ano Lectivo	2007/2008
---------------------------	------------------------------	------------	----	-----------------	----	--------------------	-----------

e) $T_{Obs} = -0.28$, R.C. = $]-\infty, -2.365] \cup [2.365, +\infty[$. Não se rejeita H_0 .

$\alpha = 2P(t_7 \geq 0.28)$ e como as tabelas da t-Student disponíveis não são suficientes para calcular este α , apenas podemos concluir que $0.25 \leq \frac{\alpha}{2} \leq 0.4$;

f) $r = 0.92031$ e $r^2 = 0.84697$;

g) $F_{Obs} = 19.37$, R.C. = $[4.74, +\infty[$. Rejeita-se H_0 .

4.b) $\hat{y} = 0.373716 + 0.384485x$; **c)** $\hat{y} = \hat{\mu}_{Y/x=24} = 9.6$; **d)** $r = 0.96825$ e $r^2 = 0.9375$;

e) $T_{Obs} = 2.407$, R.C. = $[1.86, +\infty[$. Rejeita-se H_0 .

g) $T_{Obs} = 10.955$, R.C. = $]-\infty, -5.041] \cup [5.041, +\infty[$. Rejeita-se H_0 . A regressão é significativa.

5.a) $X = \begin{bmatrix} 1 & 1.35 & 90 \\ 1 & 1.9 & 30 \\ 1 & 1.7 & 80 \\ 1 & 1.8 & 40 \\ 1 & 1.3 & 35 \\ 1 & 2.05 & 45 \\ 1 & 1.6 & 50 \\ 1 & 1.8 & 60 \\ 1 & 1.85 & 65 \\ 1 & 1.4 & 30 \end{bmatrix}$ $y = \begin{bmatrix} 17.9 \\ 16.5 \\ 16.4 \\ 16.8 \\ 18.8 \\ 15.5 \\ 17.5 \\ 16.4 \\ 15.9 \\ 18.3 \end{bmatrix}$

b) $y = X\beta + \varepsilon$, onde X e y são dados na alínea a), $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ e $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{10} \end{bmatrix}$.

c) $\hat{y} = 24.748874 - 4.159335x_1 - 0.014895x_2$;

d) $\hat{y} = \hat{\mu}_{Y/x_1=1.5, x_2=70} = 17.4672$;

f) $r = 0.99327$ e $r^2 = 0.98658$;

g) $T_{Obs} = -6.55$, R.C. = $]-\infty, -5.408] \cup [5.408, +\infty[$. Rejeita-se H_0 ;

h) $F_{Obs} = 257.348$, R.C. = $[9.55, +\infty[$. Rejeita-se H_0 .

Unidade Curricular	<i>Probabilidades e Estatística</i>	Ano	2 ^o	Semestre	1 ^o	Ano Lectivo	2007/2008
---------------------------	-------------------------------------	------------	----------------	-----------------	----------------	--------------------	-----------

6.a) $\hat{y}=1.15+2.8x$;

b) Em média, um acréscimo de uma centena na quantidade do bem produzido pela máquina A, irá levar a um aumento no custo de produção, que pode ser estimado em 2.1255 unidades monetárias;

c) $T_{Obs}=28$, R.C.= $]-\infty, -2.228] \cup [2.228, +\infty[$. Rejeita-se $H_0: \beta_1=0$, logo há evidência de que a quantidade de bem produzido num dia exerce uma influência significativa sobre o custo de produção.

d) É preferível a máquina B ($5.8825 < 6.75$);

e) É preferível a máquina A ($6.75 < 15.8825$);

f) Para níveis de produção inferiores a 71 unidades.

7.a) $\hat{y}=57.857-1.70238x$; **b)** $T_{Obs}=-4.489$, R.C.= $]-\infty, -2.65]$. Rejeita-se H_0 ;

c) $r=-0.779$; **d)** Não é sensato utilizar a equação de regressão obtida para prever o tempo de execução da tarefa para as 35 horas de treino, uma vez que 35 está fora do âmbito dos dados.

8.a) $r=0.91345$: indica uma associação linear forte entre Y e as variáveis X_1 e X_2 , logo é razoável aplicar o modelo de regressão linear para relacionar a variável Y com as variáveis X_1 e X_2 ;

b) Cerca de 16.56%;

c) (**usando $\alpha=0.05$**) $T_{Obs}=0.1584$, R.C.= $]-\infty, -2.11] \cup [2.11, +\infty[$. Não se rejeita $H_0: \beta_2=0$, e, como disse o perito, X_2 não influencia significativamente Y;

d) 53.8785 :estimativa para o preço médio de uma casa com área total 15.31 ft² e avaliada em 37.3 milhares de dólares.

9.a) $\hat{y}=532.019-22.014 x_1+6.756 x_2$

b) 95%

c) $ET_{Obs}=-14.51$, R.C.= $]-\infty, -2.11] \cup [2.11, +\infty[$. Rejeita-se $H_0: \beta_1=0$

d) $F_{obs}=162.044$, R.C.= $[3.59, +\infty[$. Rejeita-se H_0 .