PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS

ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO

CURSO DE ENGENHARIA DA COMPUTAÇÃO

AED V – AJUSTE DE CURVAS

Goiânia

2018

LUCAS MACEDO DA SILVA

AED VI – AJUSTE DE CURVAS

Trabalho apresentado por Lucas Macedo da Silva e, para avaliação e aplicação dos conceitos aprendidos em atividade extra disciplinar.

Goiânia

2018

1 ENUNCIADO DO PROBLEMA

A aplicação de ajustes de curvas se deu na resolução do problema descrito na figura a seguir.

Figura 1 – Enunciado do Problema.

(a) Construa um modelo de regressão linear múltiplca para os dados da Tabela 1

(b) Calcule um intervalo de confiança de 95% para uma resposta média de sobrevivência quando $x_1 = 3\%$,$x_2 = 8\%$, e $x_3 = 9\%$ Sabendo que A $100(1-\alpha)\%$ μ_Y é

$$\hat{y}_0 - t_{\alpha/2} s \sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} < \mu_{Y|3,8,9} < \hat{y}_0 + t_{\alpha/2} s \sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

- E
$$s^2 = \frac{SSE}{n-k-1}, \text{ onde } SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$
- n é o número de linhas e k é o número de coluna da Tabela 1.
- O valor de $t_{\alpha/2}$ deve ser obtido na tabela de distribuição de $t-$Student como mostrado na construção da regressão linear simples.

Fonte: Atividade extra disciplinar 5 (2018)

A tabela 1 se encontra na figura 2 a seguir.

Figura 2 – Tabela 1

Tabela 1. Sêmen animal.

y % Survival	x_1 peso %	x_2 peso %	x_3 peso %
25.5	1.74	5.30	10.80
31.2	6.32	5.42	9.40
25.9	6.22	8.41	7.20
38.4	10.52	4.63	8.50
18.4	1.19	11.60	9.40
26.7	1.22	5.85	9.90
26.4	4.10	6.62	8.00
25.9	6.32	8.72	9.10
32.0	4.08	4.42	8.70
25.2	4.15	7.60	9.20
39.7	10.15	4.83	9.40
35.7	1.72	3.12	7.60
26.5	1.70	5.30	8.20

Fonte: Atividade extra disciplinar 5 (2018)

2 CÓDIGOS DESENVOLVIDOS

Para a resolução do problema proposto foram desenvolvidos dois códigos. O código “regressao_multipla” calcula os coeficientes da regressão, já o código “intervalo_confianca” calcula o intervalo de confiança neste caso de 95% para o ponto (3,8,9). O valor de $t\ \alpha/2$, foi obtido a partir da tabela de T-student, como o intervalo é de 95% o valor é de 1.96.

2.1 Código para cálculo dos coeficientes da regressão

```
function [a0, a1, a2, a3] = regressao_multipla (xi, y)
  n  = length(xi); %%n, quantidade de elementos
  %%Determina a quantidade de linhas da matriz
  k = size(xi);
  a = k(1) + 1; %%Números linhas da matriz M
  c = k(2);
  %%Matriz que conterá os coeficientes
  M = [];
  %%Vetor dos coeficentes constantes
  b = [];
  %%Preenchendo x1,i com 1's
  x = [];
  for i = 1: a
    for j = 1: c
      if i == 1
        x(i, j) = 1;
      else
        x(i, j) = xi(i-1,j);
      endif
    endfor
  endfor

  for i = 1: a
    for j = 1: i
      S = 0;
```

```
    for l = 1: n
      S = S + x(i,l)*x(j,l); %%Calculo dos termos da matriz M
    endfor
    M(i, j) = S;
    M(j,i) = S;
  endfor
  S = 0;
  for k = 1: n
    S = S + y(k) * x(i,k); %%Calculo do vetor de constantes
  endfor
  b(i) = S;
 endfor

 %%Calculo dos coeficentes da regresao
 a = [];
 a = b/M;
 a0 = a(1);
 a1 = a(2);
 a2 = a(3);
 a3 = a(4);
endfunction
```

2.2 Código para cálculo do intervalo de confiança

```
function [intervalo1, intervalo2] = intervalo_confianca (xi, y)
 tam_y = length(y)
 [a0, a1, a2, a3] = regressao_multipla(xi, y)
 %%Calculo do SSE Soma do quadrado dos erros
 SSE = 0;
 for i = 1: tam_y
   SSE = SSE + (y(i) - (a0 + a1*xi(1,i) + a2*xi(2,i) + a3*xi(3,i))).^2;
 endfor
 %%Calculo do s, desvio padrão
```

```
  k = size(xi');
  a = k(1); %%Números linhas da matriz
  %%Ocorreu a soma + 1, devido a tabela 1, englobar também
  %%os valores de y
  c = k(2) + 1; %%Número de colunas
  s = sqrt (SSE / (a - c - 1));
  %% Vetor com um ponto da amostra
  x0 = [3, 8, 9];
  %%A partir da tabela de T-Student, com 95% de confiança obtemos
  %%ta/2 = 1.96
  y0 = (a0 + a1*x0(1) + a2*x0(2) + a3*x0(3));

  intervalo1 = y0 - 1.96 * s * sqrt (x0 * ((xi*xi')^-1) * x0');
  intervalo2 = y0 + 1.96 * s * sqrt (x0 * ((xi*xi')^-1) * x0');
endfunction
```

3 RESULTADOS OBTIDOS

A execução dos códigos gerou os seguintes resultados.

Sendo a regressão múltipla dada por:

$y = a0 + a1x1 + a2x2 + a3x3$

O resultado dos coeficientes são:

a0 = 39.1573499549779

a1 = 1.01610044077931

a2 = -1.86164920272855

a3 = -0.343260492573456

Com isso o modelo da regressão linear múltipla é dado por:

$$y = 39.1573499549779 + 1.01610044077931x1 + -1.86164920272855x2 - 0.343260492573456x3$$

O A soma dos quadrados dos erros (SSE - *sum of squares of errors*) resultou:

SSE = 38.6764020563198

O desvio padrão s, resultou em:

s = 2.19876107320463

O intervalo de confiança é:

(22.6351188337611; 25.8111076108915)

Denotados no código “intervalo_confianca”, por intervalo1 e intervalo2 respectivamente. Tais valores também podem ser vistos no Anexo I – *Print Screen* da execução do código.

4 ANEXO I *Print Screen* da execução do código

Figura 3 - *Print Screen* dos resultados da execução do código

```
>> [intervalo1, intervalo2] = intervalo_confianca (xi, y)
a0 =  39.1573499549779
a1 =  1.01610044077931
a2 = -1.86164920272855
a3 = -0.343260492573456
SSE =  38.6764020563198
s =  2.19876107320463
intervalo1 =  22.6351188337611
intervalo2 =  25.8111076108915
```

Fonte: O autor (2018)

Figura 4 - *Print Screen* do vetor de constantes e da matriz da execução do código

```
b =

   377.500000000000   1877.567000000000   2246.661000000000   3337.780000000000

M =

     13.000000000000     59.430000000000     81.820000000000    115.400000000000
     59.430000000000    394.725500000000    360.662100000000    522.078000000000
     81.820000000000    360.662100000000    576.726400000001    728.309999999999
    115.400000000000    522.078000000000    728.309999999999   1035.959999999998
```

Fonte: O autor (2018)