ANÁLISE DE REGRESSÃO APLICADA EM PROBLEMAS DE PREVISÃO UTILIZANDO UM PROGRAMA ITERATIVO

REGRESSION ANALYSIS APPLIED IN PREDICTION PROBLEMS USING AN ITERATIVE PROGRAM

Alfredo BONINI NETO[1]*

Talita de Souza Fabiani TAVARES[2]

Rafael Santos HERNANDES[3]

RESUMO

Neste trabalho é apresentada uma análise de Regressão pelo método dos Mínimos Quadrados que tem por finalidade prever um resultado a partir de uma sequência de dados conhecidos. É utilizado o software Matlab para criação da interface gráfica tornando o programa mais iterativo para o usuário. O programa também auxilia na resolução do sistema de equações lineares do método dos Mínimos Quadrados. Um exemplo da aplicação do método dos Mínimos Quadrados é em exercícios de previsão. Feito através da recolha dos dados que já foram medidos e através desses dados, obter uma função (1º ou 2º grau) que passe o mais próximo possível dos pontos dados. Uma observação a ser feita, é encontrar a função mais apropriada para ser utilizada, pois é esta função que passará o mais próximo possível dos pontos conhecidos. Neste trabalho é feito o diagrama de dispersão que tem por finalidade descobrir qual a função é mais apropriada para ser utilizada.

Palavras-chave: Mínimos Quadrados, Previsão, Regressão, Interface Gráfica.

ABSTRACT

This paper presents a Regression analysis by Least Squares method that aims to predict an outcome from a sequence of known data. Matlab software is used to create graphical interface making the program more iterative for the user. The program also assists in the resolution of the system of linear equations of the

[1]* Departamento de Matemática - Faculdades de Dracena/UNIFADRA. Email: alfredoboninineto@hotmail.com
[2] Departamento de Matemática - Faculdades de Dracena/UNIFADRA
[3] Departamento de Matemática - Faculdades de Dracena/UNIFADRA

method of Least Squares. An example of applying the method of Least Squares is in exercises of prevision, i.e., we can predict what will be the population of a country in a year later. This is done by collecting data that have already been measured and from this data, obtain a function (1st or 2nd degree) to pass as close as possible to the known points. Once obtained this function, just give a future value (year) and get the image of this point (the population). An observation to be made, is to find the most appropriate function to be used, since it is this function that will pass as close as possible to the known points. This work is done by the scatter diagram which aims to discover what function is more suitable for use.

Keywords: Least Squares method, Regression analysis, graphical interface

INTRODUÇÃO

Um fato que atrai pesquisadores aplicados das mais diversas áreas é a possibilidade de obter uma função real que passe nos pontos ou pelo menos passe próximo dos n pontos (x_i,y_i) dados, ou seja, (x_1, y_1) (x_2, y_2) (x_3, y_3)......(x_{n-1}, y_{n-1}) (x_n, y_n). A colocação destes pares ordenados num plano cartesiano, depende dos valores de x_i e y_i, (i=1..n) e pode fornecer um gráfico de dispersão, figura 1.

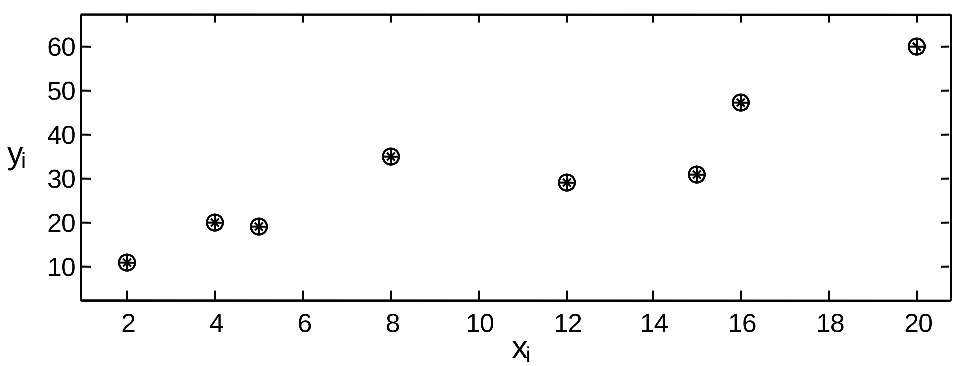

Figura 1- Diagrama de dispersão

Na Matemática existe a Teoria de Interpolação que é a área que estuda tais processos para obter funções que passam exatamente pelos pontos dados, enquanto que a Teoria de Aproximação estuda processos para obter funções que passem o mais próximo possível dos pontos dados.

É óbvio que se pudermos obter funções que passem próximas dos pontos dados e que tenham uma expressão fácil de ser manipulada, teremos obtido algo positivo e de valor científico.

Dentre os processos matemáticos que resolvem tal problema, com certeza, um dos mais utilizados é o método dos Mínimos Quadrados, que serve para gerar o que se chama em Estatística: Regressão Linear ou Ajuste Linear. As curvas mais comuns utilizadas pelos estatísticos são as funções do primeiro grau ($y = a_0 + a_1 x$) e a do segundo grau ($y = a_0 + a_1 x + a_2 x^2$).

A idéia básica para qualquer uma das funções citadas é tentar descobrir quais são os valores das incógnitas ao, a_1 para função do primeiro grau e ao, a_1 e a_2 para função do segundo grau, de tal modo que a soma dos quadrados das distâncias (tomadas na vertical) da referida curva y=f(x) a cada um dos pontos dados (y_i) seja a menor possível, daí o nome método dos Mínimos Quadrados.

Portanto, para obtenção das incógnitas das funções, basta resolver o sistema de equações lineares com duas incógnitas e duas equações para a função do primeiro grau e três incógnitas e três equações para a função do segundo grau, ambas as matrizes quadradas. Existem vários métodos para a resolução do sistema de equações lineares, como, escalonamento, Gauss-Seidel ou $\mathbf{X} = \mathbf{A}^{-1}.\mathbf{B}$, onde X representa a matriz das incógnitas $\mathbf{X} = (a_0, a_1, a_2)^t$, A^{-1} é a inversa da matriz dos coeficientes encontrados pela sequência da Regressão Linear e B é a matriz dos termos independentes, este método será utilizado nesse trabalho através do programa Matlab.

MÉTODOS DOS MÍNIMOS QUADRADOS

Seja $(x_i, y_i) = (x_1, y_1)\ (x_2, y_2)\ (x_3, y_3)......(x_{n-1}, y_{n-1})\ (x_n, y_n)$ os n pontos dados onde se deseja obter uma função que passe o mais próximo possível desse pontos. O método dos Mínimos Quadrados tem por objetivo obter essa função y=f(x) em que a soma dos quadrados das distâncias (tomadas na vertical) do referido gráfico da função y=f(x) a cada um dos pontos dados (y_i) seja a menor possível, ver Figura 2.

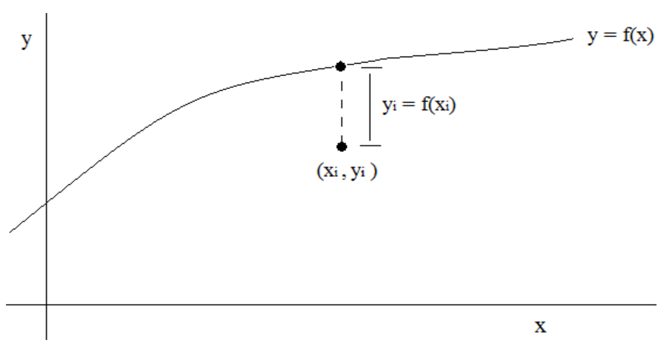

Figura 2- Distância de um ponto $(x_i; y_i)$ ao gráfico da função y = f(x)

Com isso, um sistema de equações não lineares é formado com o intuito de obter os valores das incógnitas a_0, a_1 para função do primeiro grau e a_0, a_1 e a_2 para função do segundo grau Aguiar e Moreira, (2009). O sistema a seguir é formado para obtenção da função do primeiro grau $y = a_0 + a_1 x$:

$$\begin{cases} n a_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{1}$$

onde n é o número de ponto dados, e os coeficientes são:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + ... + x_n$$

(2)

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + y_3 + ... + y_n$$

(3)

$$\sum_{i=1}^{n} x_i^{\ 2} = x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} + ... + x_n^{\ 2}$$

(4)

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + ... + x_n y_n$$

(5)

Existem vários métodos na literatura que resolvem sistemas lineares. Neste trabalho será aplicado o método $\mathbf{A}.\mathbf{X} = \mathbf{B}$, pois a matriz dos coeficientes (A) sempre possui o mesmo numero de linhas e colunas (quadrada), permitindo assim um modo rápido de resolução $\mathbf{X} = \mathbf{A}^{-1}.\mathbf{B}$. Portanto, para resolver o sistema de equações (1) será utilizado o programa Matlab, através da forma matricial:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\ 2} \end{bmatrix} . \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

ou ainda,

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\ 2} \end{bmatrix}^{-1} . \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

(6)

O sistema a seguir é formado para obtenção da função do segundo grau $y = a_0 + a_1 x + a_2 x^2$:

$$\begin{cases} n a_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\ 2}\right) a_2 = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^{\ 2}\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\ 3}\right) a_2 = \sum_{i=1}^{n} x_i y_i \\ \left(\sum_{i=1}^{n} x_i^{\ 2}\right) a_0 + \left(\sum_{i=1}^{n} x_i^{\ 3}\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\ 4}\right) a_2 = \sum_{i=1}^{n} x_i^{\ 2} y_i \end{cases}$$

(7)

onde os demais coeficientes são:

$$\sum_{i=1}^{n} x_i^{\ 3} = x_1^{\ 3} + x_2^{\ 3} + x_3^{\ 3} + ... + x_n^{\ 3}$$

(8)

$$\sum_{i=1}^{n} x_i^{\ 4} = x_1^{\ 4} + x_2^{\ 4} + x_3^{\ 4} + ... + x_n^{\ 4}$$

(9)

$$\sum_{i=1}^{n} x_i^{\ 2} y_i = x_1^{\ 2} y_1 + x_2^{\ 2} y_2 + x_3^{\ 2} y_3 + ... + x_n^{\ 2} y_n$$

(10)

A solução de (7) será:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\ 2} \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\ 2} & \sum_{i=1}^{n} x_i^{\ 3} \\ \sum_{i=1}^{n} x_i^{\ 2} & \sum_{i=1}^{n} x_i^{\ 3} & \sum_{i=1}^{n} x_i^{\ 4} \end{bmatrix} . \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^{\ 2} y_i \end{bmatrix}$$

ou

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\ 2} \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\ 2} & \sum_{i=1}^{n} x_i^{\ 3} \\ \sum_{i=1}^{n} x_i^{\ 2} & \sum_{i=1}^{n} x_i^{\ 3} & \sum_{i=1}^{n} x_i^{\ 4} \end{bmatrix}^{-1} . \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^{\ 2} y_i \end{bmatrix}$$

(11)

O SOFTWARE MATLAB

O Matlab é uma linguagem de programação apropriada ao desenvolvimento de aplicativos de natureza técnica. Para isso, possui facilidades de computação, programação e baixo custo, dentro de um ambiente amigável e de fácil aprendizado (Huang, Zhang, 2000). Com o Matlab é possível resolver problemas computacionais mais rápido do que com linguagens de programação tradicionais, como C, C + + e Fortran (Mathworks, 2009). O Matlab foi desenvolvido no início da década de 80 por Cleve Moler, no Departamento de Ciência da Computação da Universidade do Novo México, EUA. As versões posteriores ao Matlab 4.0, foram desenvolvidas na firma comercial MathWorks Inc., que detêm os direitos de autores destas implementações. O Matlab foi originalmente desenvolvido para prover um acesso amigável ao tratamento de vetores e matrizes. Atualmente o Matlab dispõe de uma biblioteca bastante abrangente de funções matemáticas, geração de gráficos e manipulação de dados que auxiliam muito o trabalho do programador. E ainda possui uma vasta coleção de bibliotecas denominadas toolboxes para áreas específicas como: equações diferenciais ordinárias, estatística, processamento de imagens, processamento de sinais, finanças, entre outras.

Os recursos instalados também podem ser estendidos pelo usuário através da implementação de funções Matlab (M-files) ou de rotinas escritas em linguagem C ou Fortran.

A necessidade de um programa em língua portuguesa para atender as necessidades das aulas e permitir ao aluno de graduação um primeiro contato com os computadores e programas de computadores relacionados à pesquisa científica. Além de permitir o uso do programa por pesquisadores (alunos e professores) dos cursos de pós-graduação.

CARACTERÍSTICAS DOS RECURSOS GRÁFICOS DO MATLAB

Existem muitos comandos para criação da interface gráfica no Matlab, citaremos alguns dos comandos. Podemos criar uma janela através da função Figure e formatar essa janela através de seus parâmetros (alguns destes parâmetros serão mostrados a seguir). A figura 3 mostra uma janela feita com a função Figure e seus parâmetros devidamente configurado.

```
dx=0.2850;
dy=0.2200;
pos = [(1-dx)*0.5,  (1-dy)*0.5,  dx, dy];
h0 = figure('Color',[  0.800  0.800  0.800],  ...
                  'Units','normalized',  ...
                  'MenuBar','none', ...
                  'NumberTitle','off',  ...
                  'Position',pos, ...
                  'Resize','off',  ...
                  'name','');
```

Figura 3 - Exemplo da função F*igure.*

Parâmetros da Função Figure:

Color: Representa a cor de fundo da janela. É um vetor com os componentes RGB. Exemplo: a seqüência Color [0 0 0] equivale a cor preta e a seqüência Color[1 1 1] equivale a cor branca.

Units: É uma unidade usada para posicionar o controle. A posição e tamanho de um controle dentro da janela, que são feitos através de coordenadas como: Normalized (máximo e mínimo da janela correspondendo a 0 e 1) e Pixels(pontos gráficos).

MenuBar: Se o valor dessa propriedade for 'none' nenhum menu é mostrado na janela. Se for 'Figure' a janela terá o menu padrão de figuras.

NumberTitle: Se o valor dessa propriedade for 'on' aparecerá o nome e o número da janela. Se for 'off' a barra de título aparece em branco.

Position: Especifica a posição e tamanho da janela através das propriedades: [left, bottom, width, height].

Resize: Se estiver em 'on' a janela pode ter seu tamanho alterado. Se tiver em 'off' o tamanho da janela não pode ser alterado.

Name: Dá um nome para a janela. O valor desta propriedade deve ser uma string.

Controlando os Controles

No *Matlab* existe uma maneira muito prática de se programar a resposta de um controle ao usuário. Por exemplo, ao apertar-se um botão queremos que seja plotado um gráfico, ou fechar a janela que se está operando. Os controles também servem para retornar algum valor para o usuário de maneira mais amigável.

O Comando Uicontrol

O *Uicontrol* é um comando de controle para a janela que está ativa. Para criar os controles, deve-se configurar apropriadamente seus parâmetros. A figura 4 apresenta o exemplo de *BackgroundColor*.

```
h0 = figure('Color',[ 0.800 0.800
0.800], ...
'Units','normalized', ...
'MenuBar','none', ...
'NumberTitle','off', ...
'Position',pos, ...
'Resize','off', ...
'name','');
h1 = uicontrol('Parent',h0, ...
`Units','normalized', ...
'BackgroundColor',[ 1 1 1], ...
```

```
'ForegroundColor',[0.000 0.000
0.502],...

'HorizontalAlignment','center',
...
'Position',[0.0787 0.7678 0.8287
0.1547], ...
'String','Sistemas:', ...
'FontSize',17,...
'Fontname','Arial',...
'Style','text', ...
'Tag','StaticText1');
```

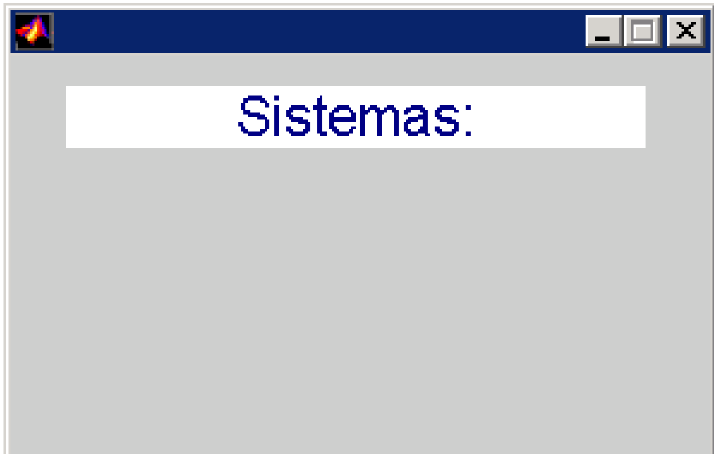

Figura 4 - Exemplo do *BackgroundColor*.

RESULTADOS E DISCUSSÃO

Serão apresentados exercícios de aplicações, onde são apresentados exercícios de previsão para um valor posterior e outro em que encontraremos valores entre dois pontos conhecidos.

Um exemplo de aplicação para obtenção de um ponto posterior seria para ajustar uma curva aos dados da população brasileira entre os anos de 1872 e 2000; com isso podemos prever qual será a população em um ano posterior.

Considere a tabela 1 a seguir apresenta a população brasileira (em milhões) obtidas pelo censo (IBGE, 2009):

Tabela 1 - Populações brasileiras entre 1872 e 2010 (em milhões)

Censo de 1872	10,112 habitantes.
Censo de 1890	14,333 habitantes.
Censo de 1900	17,318 habitantes.
Censo de 1920	30,635 habitantes.
Censo de 1940	41,165 habitantes.
Censo de 1950	51,941 habitantes.
Censo de 1960	70,070 habitantes.
Censo de 1970	93,139 habitantes.
Censo de 1980	119,002 habitantes.
Censo de 1991	146,825 habitantes.
Censo de 2000	169,799 habitantes.
*** ano de 2015**	***210,405 habitantes.**

*obtido pelo método dos mínimos quadrados através do programa proposto

Para resolver o exercício anterior foi proposto um programa por meio de uma interface gráfica para obter a função y. A figura 5 apresenta a tela inicial do programa. Ao clicar no botão “MÉTODO DOS MÍNIMOS QUADRADOS” outra tela é aberta (figura 6), esta tela representa a figura de entrada de dados do exercício.

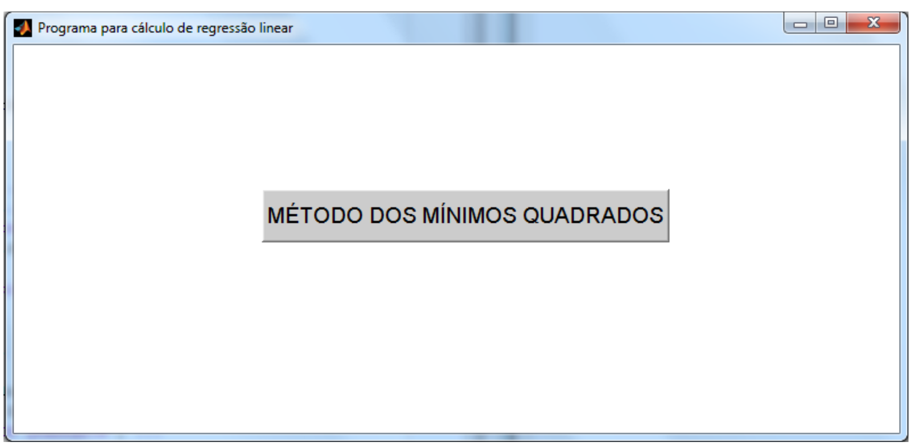

Figura5- Tela inicial do programa.

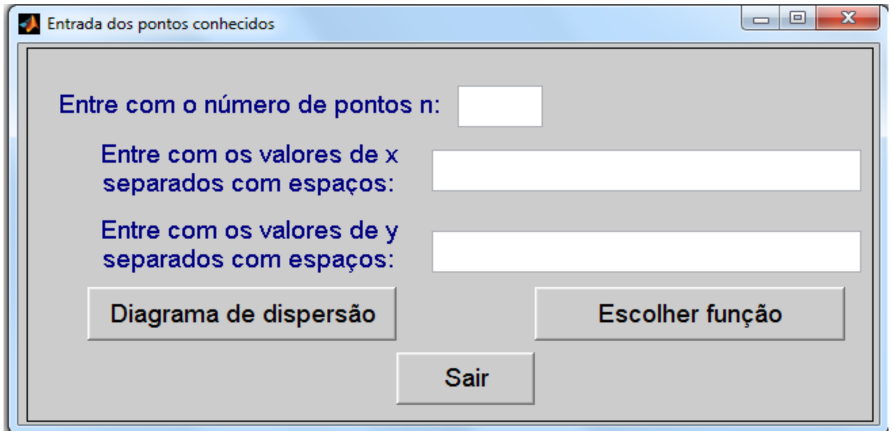

Figura 6- tela de entrada de dados.

Observam-se na figura 6 as opções que o usuário tem de escolher. Nesta tela o usuário entra com o valor de n que é o número de pontos conhecidos e em seguida com os valores de x e de y dos n pontos. Caso o usuário queira montar o diagrama de dispersão para analisar qual a função mais adequada para os pontos basta clicar no botão “Diagrama de dispersão”. A figura 7 representa o diagrama de dispersão para os pontos do exercício anterior.

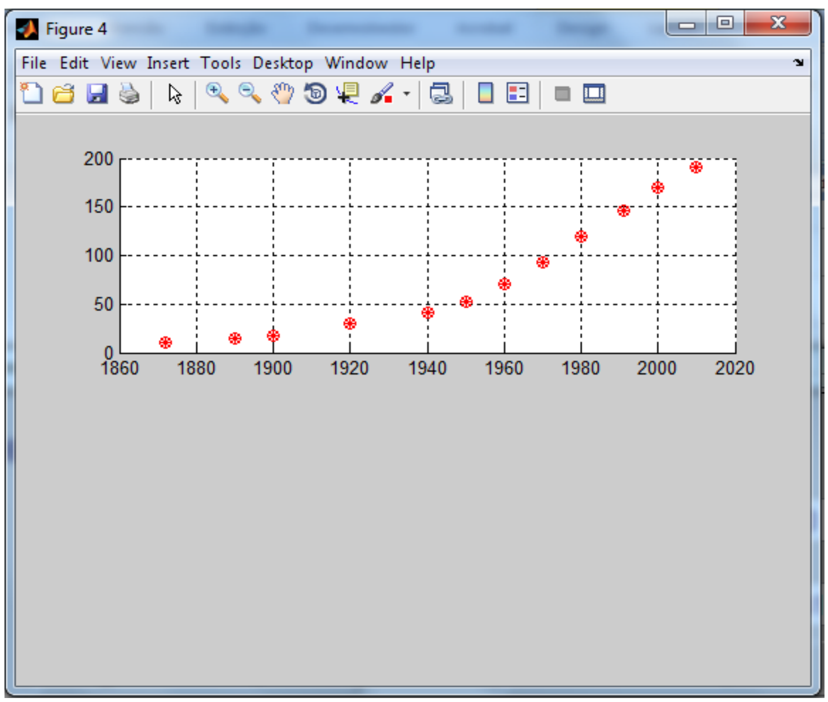

Figura 7- Diagrama de dispersão do exercício de população

Observando a figura 7, constata-se que os pontos representam melhor uma parábola, assim o usuário pode clicar no botão “Escolher função” (figura 6), a qual abrirá uma tela para escolher a função, reta para a função do primeiro grau e parábola para a função do segundo grau. Clicando no botão “Parábola”, a função e o gráfico serão mostrados na figura 7, ver figura 9.

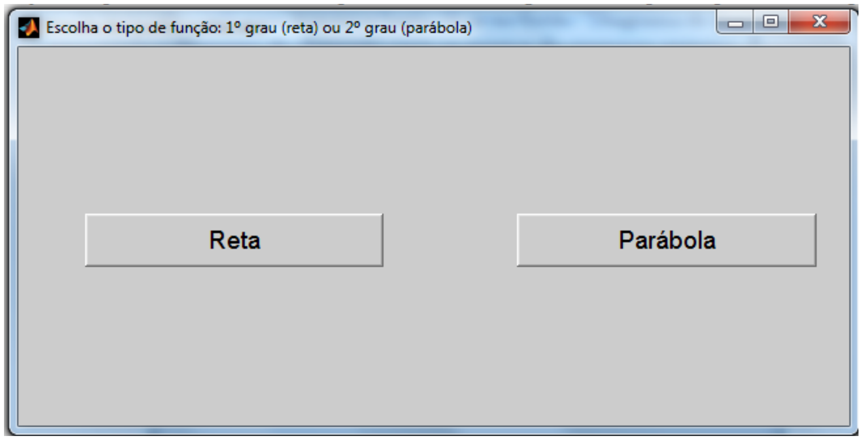

Figura 8 - Tela para escolha da função, 1º ou 2º grau.

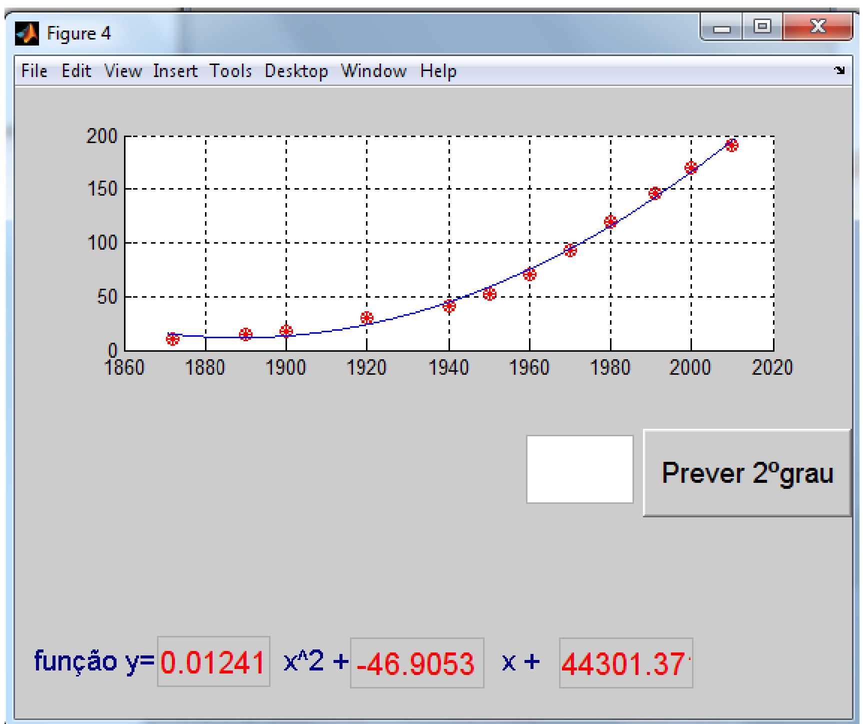

Figura 9 - Função e gráfico obtidos pelo método dos Mínimos Quadrados do exercício de população.

Caso o usuário queira saber um valor futuro para a população, por exemplo, 2015, basta inserir 2015 (ano) no quadro da figura 9 e basta clicar no botão “Prever 2º grau” obtém-se a resposta logo abaixo do botão, o valor obtido é aproximadamente 210,405 milhões de habitantes, ver figura 10.

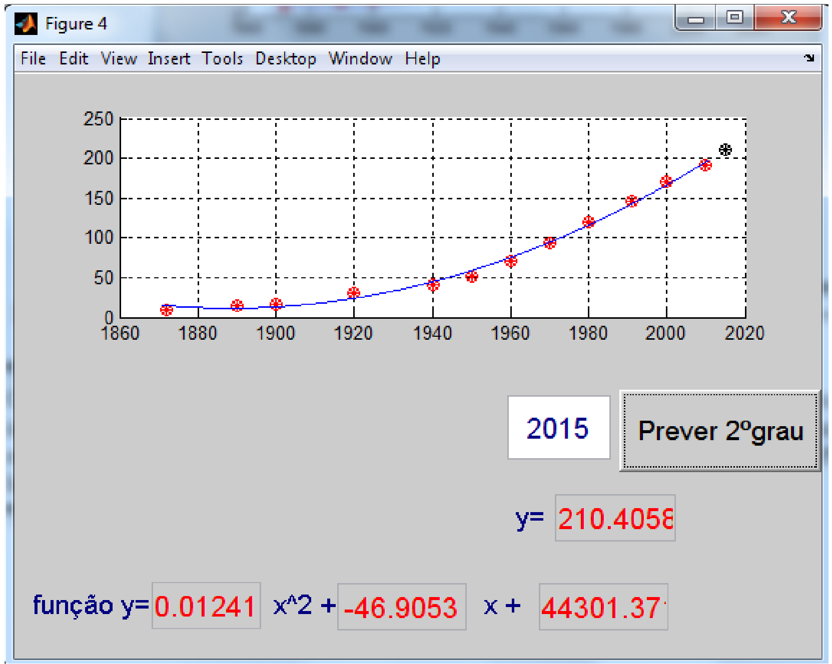

Figura 10 - Valor previsto para a população no ano de 2015.

CONCLUSÕES

Neste trabalho foi apresentado o método dos Mínimos Quadrados desenvolvido a partir de uma interface gráfica que tem por finalidade prever um resultado a partir de uma sequência de dados conhecidos. Foi utilizado o software Matlab para criação da interface gráfica tornando o programa mais iterativo para o usuário. O programa também auxiliou na resolução do sistema de equações lineares do método dos Mínimos Quadrados. Os exemplos da aplicação do método dos Mínimos Quadrados foram baseados em exercícios de previsão, ou seja, previmos qual foi a população de um determinado país em um ano posterior. Isso foi feito colhendo os dados que já foram medidos e através desses dados, foi obtida uma função (1º ou 2º grau) que passe o mais próximo possível dos pontos dados. A ideia deste trabalho é prever qualquer tipo de problema que envolva pontos de duas variáveis (x_i, y_i). Neste trabalho também foi feito o diagrama de dispersão que tem por finalidade descobrir qual a função mais apropriada para ser utilizada.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR, F. L. E MOREIRA, W. I. Ajuste de curvas por quadrados mínimos lineares. Disponível em: < http://www.mat.ufmg.br/gaal/aplicacoes/quadrados_minimos.pdf>. Acesso em: 06 out. 2009.

HUANG, G.M.; ZHANG, H. A. New Education Matlab Software for Teaching Power Analysis that Involves the Slack Bus Concept and

Allocation Issues. Power Engineering Society Winter Meeting. IEEE, v. 2, 23-27, p. 1150- 1158, Jan 2000.

MATHWORKS. Disponível em: < http://www.mathworks.com>. Acesso em: 15 mar. 2009.

IBGE. Instituto Brasileiro de Geografia e Estatística. Disponível em: <http://www.ibge.gov.br/home/>. Acesso em: 06 out. 2009.