

PROGRAMA PARA PROJEÇÃO DA POPULAÇÃO BRASILEIRA
UTILIZANDO REGRESSÃO LINEAR

Aluno Felipe Celestino Muros

Prof. Cleuber Nascimento

CAMPOS DOS GOYTACAZES, RJ – BRASIL.

DEZEMBRO DE 2019

LISTA DE FIGURAS

Figura 1: Tela inicial do programa	6
Figura 2: Coeficiente de correlação	7
Figura 3: Equação da reta de regressão	7
Figura 4: Estimativa calculada	8
Figura 5: Variáveis auxiliares.....	8
Figura 6: Reta de regressão linear	9

SUMÁRIO

1. INTRODUÇÃO	4
2. FORMULAÇÃO.....	5
3. RESULTADOS.....	7
4. CONCLUSÕES	10
REFERÊNCIAS BIBLIOGRÁFICAS	11
ANEXO I.....	13
DADOS CONTIDOS NO ARQUIVO POPULACAO.TXT	13
ANEXO II.....	14
CÓDIGO FONTE	14

1. INTRODUÇÃO

Este trabalho tem o objetivo de criar um programa para projeção da população brasileira utilizando o método de regressão linear simples e baseado nos dados do Instituto Brasileiro de Geografia e Estatística.

Uma breve instrução de uso do aplicativo mostra ao usuário as funções disponíveis no sistema. Como detalhamento de métodos utilizados e apresentação dos resultados obtidos será feita a análise dos resultados para apresentação da conclusão sobre o comportamento do crescimento da população brasileira.

2. FORMULAÇÃO

O programa foi escrito em linguagem C. Utilizando arquivo de entrada “Populacao.txt” o usuário pode atualizar os dados divulgados pelo IBGE ao final de cada ano, tornando o programa utilizável futuramente e não apenas para a proposta de projetar para 20 anos com os dados de 2010 a 2019.

Os dados de ano e quantidade de população são armazenados nas variáveis $x[i]$ e $y[i]$, respectivamente, através da leitura do arquivo externo e imediatamente após são feitos os cálculos das variáveis auxiliares (somatórios de $x[i]$, $y[i]$, $x[i]*y[i]$, $x[i]^2$ e $y[i]^2$).

Com os dados adquiridos e as variáveis auxiliares calculadas, o programa calcula o coeficiente de correlação (r) utilizando a fórmula:

$$r = \frac{n \sum_{i=1}^n x_i \cdot y_i - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}$$

O próximo passo é o cálculo da reta de regressão através do cálculo das constantes (a) e (b) utilizando as seguintes fórmulas:

$$a = \frac{n \sum_{i=1}^n x_i \cdot y_i - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b = \bar{Y}_n - a \cdot \bar{X}_n.$$

Com a equação da reta definida o sistema pergunta ao usuário quantos anos a partir do atual ele deseja visualizar a projeção da população ($x[i]$) e aplica este dado na equação da reta de regressão.

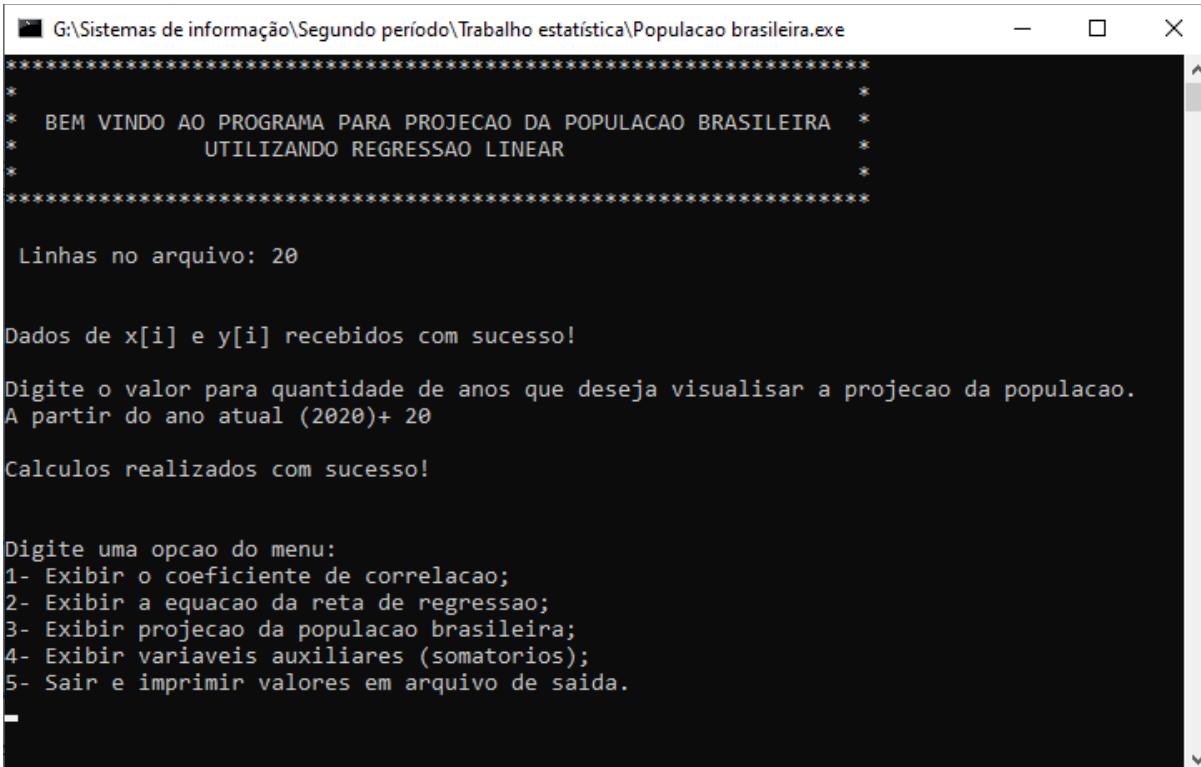
$$\hat{y}_i = ax_i + b$$

É exibido um menu ao usuário com as seguintes opções:

1. Exibir o coeficiente de correlação;
2. Exibir a equação da reta de regressão;
3. Exibir projeção da população brasileira;
4. Exibir variáveis auxiliares (somatórios);
5. Sair e imprimir valores em arquivo de saída.

Assim o usuário pode revisar os valores antes de sair do programa. Ao final o sistema gera um arquivo “Resultados.txt” onde são salvos todos os valores calculados pelo programa.

O exemplo da tela inicial do programa com o menu pode ser visto na Figura 1.



```
*****
* BEM VINDO AO PROGRAMA PARA PROJECAO DA POPULACAO BRASILEIRA *
* UTILIZANDO REGRESSAO LINEAR *
*****

Linhas no arquivo: 20

Dados de x[i] e y[i] recebidos com sucesso!

Digite o valor para quantidade de anos que deseja visualizar a projecao da populacao.
A partir do ano atual (2020)+ 20

Calculos realizados com sucesso!

Digite uma opcao do menu:
1- Exibir o coeficiente de correlacao;
2- Exibir a equacao da reta de regressao;
3- Exibir projecao da populacao brasileira;
4- Exibir variaveis auxiliares (somatorios);
5- Sair e imprimir valores em arquivo de saida.
_
```

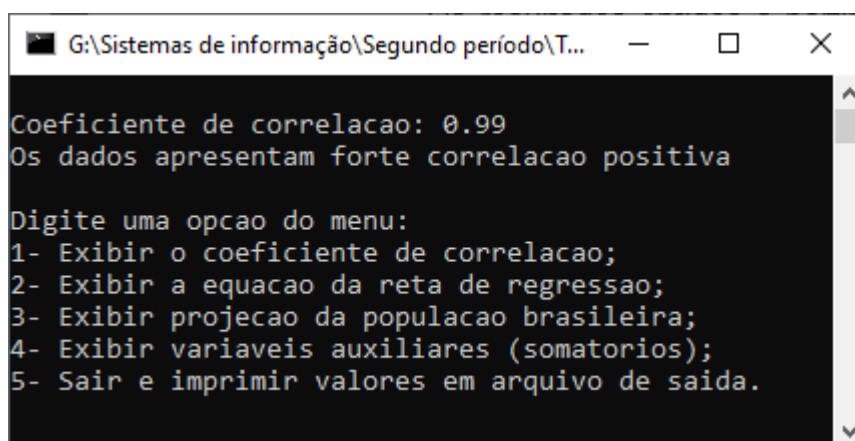
Figura 1: Tela inicial do programa

O código fonte do programa está disponível no Anexo 2.

3. RESULTADOS

Embora o programa seja capaz de calcular a projeção conforme desejo do usuário, os resultados analisados são para uma projeção de 20 anos a partir do ano de 2019, ou seja, a projeção para 2039.

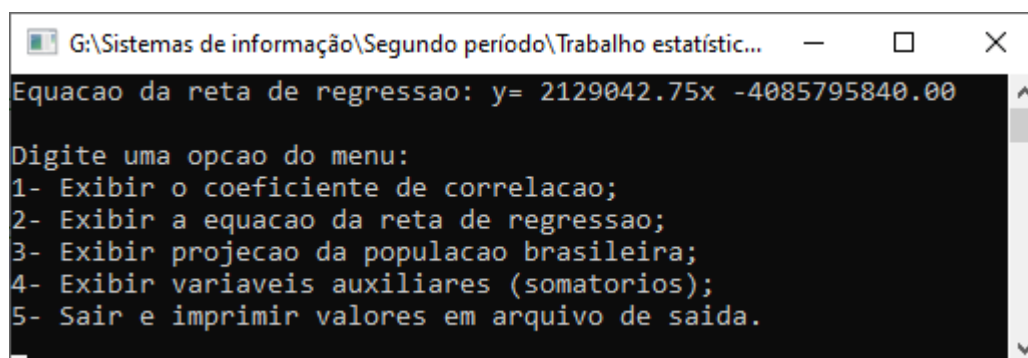
Os resultados obtidos a partir dos dados do Anexo 1 pode ser visto nas Figura 2, Figura 3, Figura 4 e Figura 6.



```
G:\Sistemas de informação\Segundo período\T...
Coeficiente de correlacao: 0.99
Os dados apresentam forte correlacao positiva

Digite uma opcao do menu:
1- Exibir o coeficiente de correlacao;
2- Exibir a equacao da reta de regressao;
3- Exibir projecao da populacao brasileira;
4- Exibir variaveis auxiliares (somatorios);
5- Sair e imprimir valores em arquivo de saida.
```

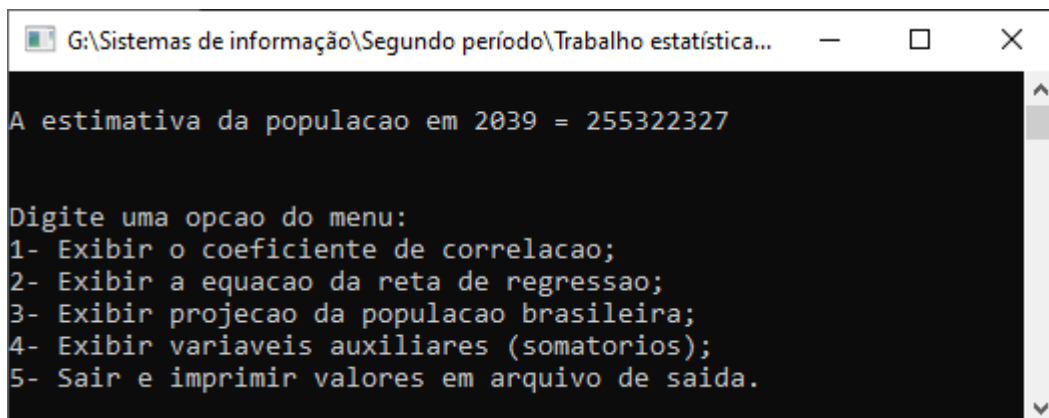
Figura 2: Coeficiente de correlação



```
G:\Sistemas de informação\Segundo período\Trabalho estatístic...
Equacao da reta de regressao: y= 2129042.75x -4085795840.00

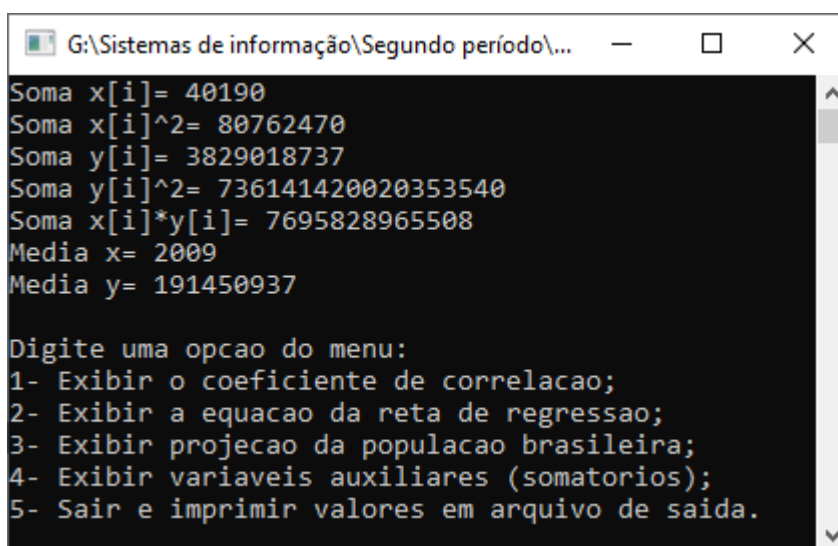
Digite uma opcao do menu:
1- Exibir o coeficiente de correlacao;
2- Exibir a equacao da reta de regressao;
3- Exibir projecao da populacao brasileira;
4- Exibir variaveis auxiliares (somatorios);
5- Sair e imprimir valores em arquivo de saida.
```

Figura 3: Equação da reta de regressão



```
G:\Sistemas de informação\Segundo período\Trabalho estatística...  
A estimativa da populacao em 2039 = 255322327  
  
Digite uma opcao do menu:  
1- Exibir o coeficiente de correlacao;  
2- Exibir a equacao da reta de regressao;  
3- Exibir projecao da populacao brasileira;  
4- Exibir variaveis auxiliares (somatorios);  
5- Sair e imprimir valores em arquivo de saida.
```

Figura 4: Estimativa calculada



```
G:\Sistemas de informação\Segundo período\...  
Soma x[i]= 40190  
Soma x[i]^2= 80762470  
Soma y[i]= 3829018737  
Soma y[i]^2= 736141420020353540  
Soma x[i]*y[i]= 7695828965508  
Media x= 2009  
Media y= 191450937  
  
Digite uma opcao do menu:  
1- Exibir o coeficiente de correlacao;  
2- Exibir a equacao da reta de regressao;  
3- Exibir projecao da populacao brasileira;  
4- Exibir variaveis auxiliares (somatorios);  
5- Sair e imprimir valores em arquivo de saida.
```

Figura 5: Variáveis auxiliares

Para ajuda na análise gráfica foi utilizado o software Excel para elaboração do gráfico, utilizando os mesmos dados do programa objeto deste estudo. O resultado pode ser visto na Figura 6 cujo eixo x representa os anos e o eixo y a população brasileira.

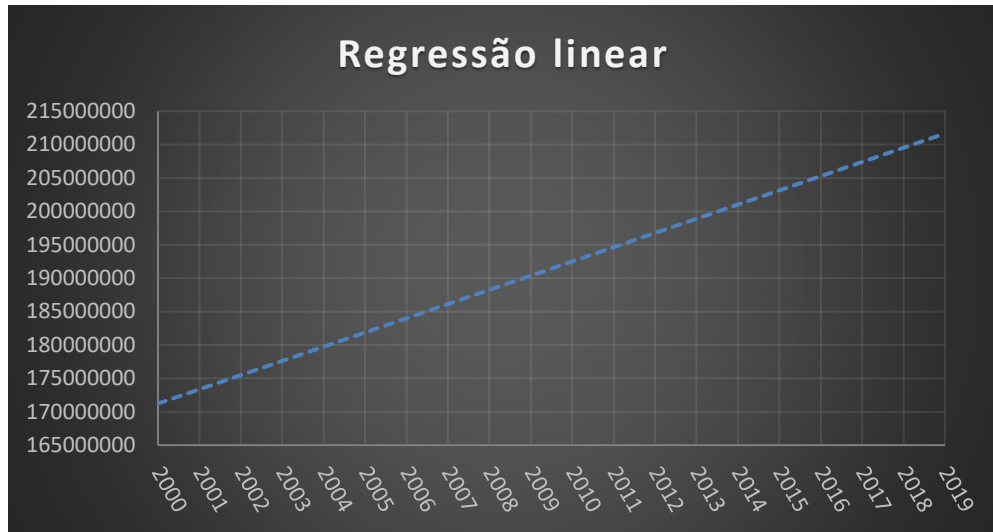


Figura 6: Reta de regressão linear

4. CONCLUSÕES

O coeficiente de correlação apresentou valor de aproximadamente $r=0,99$ o que indica uma forte correlação positiva, ou seja, ao aumentar o valor de $x[i]$ tem-se proporcionalmente (considerando a fórmula da reta de regressão) o aumento de $y[i]$.

Dado esta correlação, temos uma população estimada para 2039 igual a 255.321.856 mantendo o ritmo estatístico de crescimento.

A projeção gráfica da reta de regressão (Figura 6) comparada com os dados registrados da população brasileira entre 2000 e 2019, também comprova esta tendência de crescimento linear, com valores bem próximos à reta. O que pode ser confirmado ao analisar o coeficiente (a) da reta de regressão que aponta um erro de cerca de 2.000 pessoas, isso que representa uma média de 1% de erro num universo de 200.000.000 de habitantes.

REFERÊNCIAS BIBLIOGRÁFICAS

ROCHA, Nei. **Probabilidade e estatística**. Instituto de Matemática – UFRJ. Rio de Janeiro, 2018.

IBGE (2000 a 2003): **Estimativas das populações residentes municipais calculadas com base na Projeção Populacional para o Brasil - Revisão 2000**, Disponível em: ftp://ftp.ibge.gov.br/Estimativas_Projecoes_Populacao/Revisao_2000_Projecoes_1980_2050/. Acesso em dezembro de 2019.

IBGE (2004 a 2006): **Estimativas das populações residentes municipais calculadas com base na Projeção Populacional para o Brasil - Revisão 2004**, Disponível em: ftp://ftp.ibge.gov.br/Estimativas_Projecoes_Populacao/Revisao_2004_Projecoes_1980_2050/. Acesso em dezembro de 2019.

IBGE (2007): **População residente obtida da Contagem Populacional 2007 para 5543 municípios e populações estimadas para 129 municípios**, Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/contagem2007/default.shtm>. Acesso em dezembro de 2019.

IBGE (2008 e 2009): **Estimativas das populações residentes municipais calculadas com base na Projeção Populacional para o Brasil - Revisão 2008**. Disponível em: http://www.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2008/projecao.pdf. Acesso em dezembro de 2019.

IBGE (2010): **População residente obtida do Censo Demográfico 2010**. Disponível em: http://www.ibge.gov.br/home/estatistica/populacao/censo2010/primeiros_resultados/default_primeiros_resultados.shtm. Acesso em dezembro de 2019.

IBGE (2011 e 2012): **Estimativas das populações residentes municipais calculadas com base na Projeção Populacional para o Brasil**. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/estimativa2012/default.shtm>.

Acesso em dezembro de 2019.

IBGE (2013): **Projeção da população para o Brasil e Unidades da Federação, por sexo e idade**, 2013. Disponível em: http://www.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2013/default.shtm. Acesso em dezembro de 2019.

IBGE (2013 a 2019): **Projeção da população do Brasil e das Unidades da Federação**, 2019. Disponível em: <https://www.ibge.gov.br/apps/populacao/projecao/index.html>. Acesso em dezembro de 2019.

ANEXO I**DADOS CONTIDOS NO ARQUIVO POPULACAO.TXT**

2000 169799170
2001 172385826
2002 174632960
2003 176871437
2004 181581024
2005 184184264
2006 186770562
2007 183989711
2008 189612814
2009 191480630
2010 190747855
2011 192379287
2012 193946886
2013 201032714
2014 202768562
2015 204450649
2016 206081432
2017 207660929
2018 208494900
2019 210147125

ANEXO II

CÓDIGO FONTE

```

#include <stdio.h>
#include <math.h>
#include <time.h>

float coefcorrel (int x[], double y[], int n, FILE *arqresults); //função para
calcula do coeficiente de correlação
float calc_a (int x[], double y[], int n); //função para calculo de a
float calc_b (int x[], double y[], int n, float a); //função para calculo de a

int main ()
{
    int n=20,i=0, x[20], ano;
    float r=0,a=0,b=0;
    double y[20];
    struct tm *data_hora_atual;
    FILE *seriepopul, *arqresults;
    seriepopul = fopen ("Populacao.txt","r"); // leitura do arquivo de entrada
    arqresults = fopen ("Resultados.txt","w"); //criação do arquivo de saída
    if (seriepopul==NULL || arqresults==NULL)
    {
        printf ("\nErro na abertura do arquivo.\n Certifique-se que o
arquivo Populacao.txt esta na mesma pasta que este programa\n\n");
        return 0;
    }
    do //Ler os dados do arquivo em xi e yi
    {
        fscanf (seriepopul,"%d %lf",&x[i],&y[i]);
        i++;
    }
    while (!feof(seriepopul));

```

```

r=coefcorrel(x,y,n,arqresults);
printf ("\nCoeficiente de correlacao: %.2f \n\n",r);
fprintf (arqresults,"Coeficiente de correlacao: %.2f \n",r);
a=calc_a(x,y,n);
b=calc_b(x,y,n,a);
if (b>0)
{
    printf ("Equacao da reta de regressao: y=%.2fx+%.2f\n",a,b);
    fprintf (arqresults,"Equacao da reta de regressao:
y=%.2fx+%.2f\n",a,b);
}
else
{
    printf ("Equacao da reta de regressao: y=%.2fx%.2f\n",a,b);
    fprintf (arqresults,"Equacao da reta de regressao:
y=%.2fx%.2f\n",a,b);
}
time_t segundos;
time(&segundos);
data_hora_atual = localtime(&segundos);
printf ("\nDigite o valor para quantidade de anos que deseja visualizar a
projecao da populacao: ");
scanf ("%d",&ano);
ano=data_hora_atual->tm_year+1900+ano; //define para qual ano será
a projeção
printf ("\nA estimativa da populacao em %d = %d \n\n",ano,(a*ano)+b);
fprintf (arqresults,"\nA estimativa da populacao em %d =
%d\n\n",ano,(a*ano)+b);
fclose (seriepopul);
fclose (arqresults);
return 0;
}

float coefcorrel(int x[], double y[], int n, FILE *arqresults) //função para calculo
do coeficiente de correlação

```

```

{
    long int somax2=0,somax=0;
    double somaxy=0, somay=0, somay2=0;
    int i=0;
    float r=0;
    for (i=0;i<n;i++)
    {
        somax+=x[i];
        somay+=y[i];
        somaxy=somaxy+(x[i]*y[i]);
        somax2+=x[i]*x[i];
        somay2+=y[i]*y[i];
    }
    r=(n*somaxy-(somax*somay))/((sqrt(n*somax2-
(somax*somax)))*(sqrt(n*somay2-(somay*somay))));
    fprintf (arqresults,"Somatorio xi= %d\nSomatorio yi= %lf\nSomatorio
xi*yi= %lf\nSomatorio x^2= %d\nSomatorio y^2=
%lf\n",somax,somay,somaxy,somax2,somay2);
    return r;
}

float calc_a (int x[], double y[], int n)
{
    long int somax2=0,somax=0;
    double somaxy=0, somay=0;
    int i=0;
    float a;
    for (i=0;i<n;i++)
    {
        somax+=x[i];
        somay+=y[i];
        somaxy+=x[i]*y[i];
        somax2+=x[i]*x[i];
    }
    a=(n*somaxy-(somax*somay))/(n*somax2-(somax*somax));

```



```
        return a;
    }
float calc_b (int x[], double y[], int n, float a)
{
    double somay=0;
    int i=0,somax=0;
    float b;
    for (i=0;i<n;i++)
    {
        somax+=x[i];
        somay+=y[i];
    }
    b=(somay/n)-(a*(somax/n));
    return b;
}
```