

Министерство цифрового развития, связи и  
массовых коммуникаций Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего  
образования «Сибирский государственный университет телекоммуникаций и  
информатики» (СибГУТИ)

**Отчёт**  
по лабораторной работе №2  
по дисциплине «**Прикладные задачи теории вероятностей**»

Выполнил:

студент гр. ИС-142

«\_\_» декабря 2023 г.

\_\_\_\_\_

/Наумов А.А./

Проверил:

профессор кафедры В.С.,

«\_\_» декабря 2023 г.

\_\_\_\_\_

/Родионов А.С./

Оценка « \_\_\_\_\_ »

Новосибирск 2023

## ВЫПОЛНЕНИЕ РАБОТЫ

Целью работы является нахождение коэффициентов линейного уравнения по заданным случайным значениям  $A(x)$  между 0 и 1 и функции  $C(y(x))$ . Ошибка  $B$  будет равна сумме 12 случайных чисел - 6 (6 - среднее значение, математическое ожидание).

**Метод наименьших квадратов (МНК)** — математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от экспериментальных входных данных. Он может использоваться для «решения» переопределенных систем уравнений (когда количество уравнений превышает количество неизвестных), для поиска решения в случае обычных (не переопределенных) нелинейных систем уравнений, для аппроксимации точечных значений некоторой функции. МНК является одним из базовых методов регрессионного анализа для оценки неизвестных параметров регрессионных моделей по выборочным данным.

Для нахождения зависимости между входными данными и функцией  $Y$  с использованием метода наименьших квадратов (МНК), мы можем аппроксимировать линейное уравнение:  $Y = AX + B$ , где:

- $Y$  - функция  $Y(X)$ ,
- $X$  - входные данные,
- $B$  - коэффициент сдвига (intercept),
- $A$  - коэффициент наклона (slope).

Цель - найти значения  $A$  и  $B$ , которые минимизируют сумму квадратов ошибок между предсказанными значениями  $\hat{Y}$  и реальными значениями  $Y$ .

Сначала найдем коэффициенты  $A$  и  $B$ , используя формулы МНК:

$$A = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$
$$B = \bar{Y} - A\bar{X}$$

Где  $\bar{X}$  и  $\bar{Y}$  - средние значения  $X$  и  $Y$ .

Демонстрация таблицы значений  $A$ ,  $B$ ,  $C$  (значения  $A$  генерируются случайно при каждом вводе чего-либо или обновления страницы, поэтому фактические данные в файле example.data не будут совпадать с данной иллюстрацией):

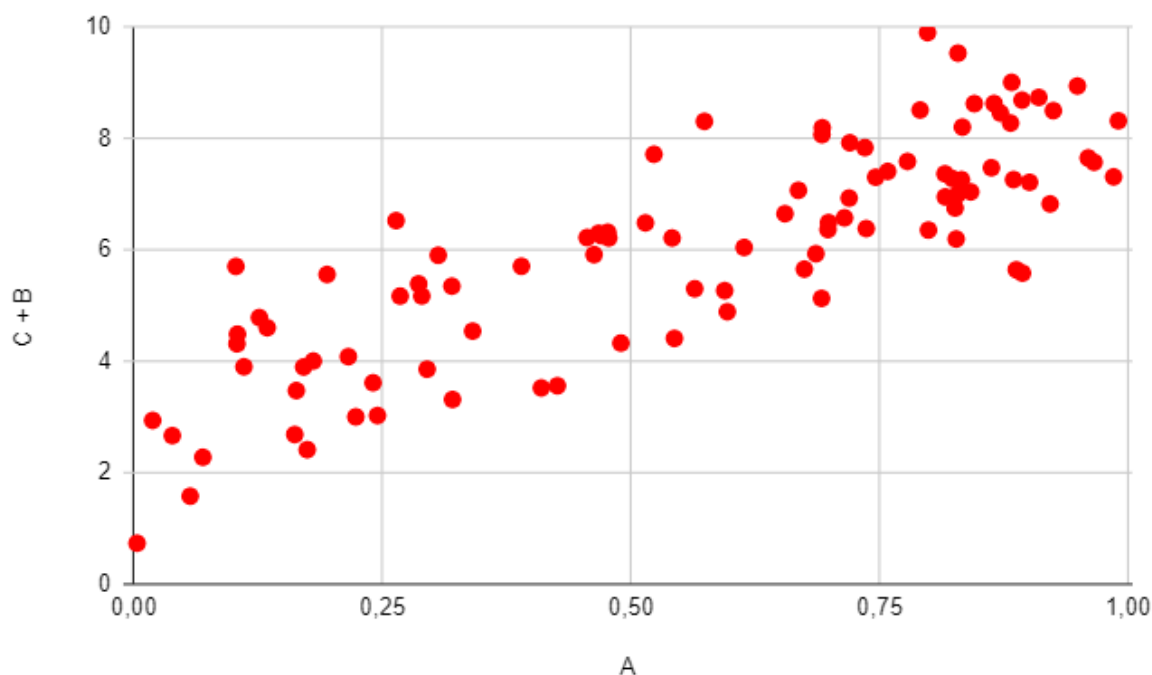
C(A) = 5A+3								
	A (random number between 0 and 1)	B (sum of random numbers - 6)	C(A) (random linear expression) + B	C(A) + 3B	C(A) + 6B	C(A) + 20B	ln(e^C + B)	ln(e^C + 6B)
1	0,3421076474	-0,3528477182	4,357690519	3,651995083	2,593451928	-2,346416126	4,707357579	4,691300744
2	0,6030591352	0,2579425862	6,273238262	6,789123434	7,562951193	11,1741474	6,015925148	6,01906658
3	0,6857441762	-0,5000612318	5,928659649	4,928537185	3,42835349	-3,572503756	6,427913199	6,423864973
4	0,4896078796	-1,123068051	4,324971347	2,078835244	-1,290368909	-17,01332163	5,443193162	5,418603305
5	0,9006680007	-0,2926773624	7,210662641	6,625307916	5,747275829	1,649792756	7,503178655	7,502371522
6	0,691825542	1,605971217	8,065098927	11,27704136	16,09495501	38,57855206	6,461639762	6,474106229
7	0,1039045926	0,9651122623	4,484635225	6,41485975	9,310196537	22,82176821	3,547702571	3,677792962
8	0,8623178202	0,1600437478	7,471632849	7,791720344	8,271851588	10,51246406	7,311695965	7,312230116
9	0,596680486	-1,095484191	4,887918239	2,696949857	-0,5895027165	-15,92628139	5,980637732	5,966698414
10	0,2154290738	0,004964717454	4,082110086	4,092039521	4,106933674	4,176439718	4,077229546	4,077650326
11	0,3202366612	-1,288042275	3,313141031	0,7370564818	-3,127070343	-21,15966219	4,588167084	4,520416463
12	0,4555734954	0,9414302079	6,219297685	8,102158101	10,92644872	24,10647164	5,282660375	5,306286228
13	0,3404841136	-0,1624825491	4,539938019	4,214972921	3,727525274	1,452769587	4,700945229	4,693535709
14	0,03837089536	-0,5269978585	2,664856618	1,610860901	0,02986732591	-7,348102693	3,169959228	3,052657091
15	0,4763973834	0,9326158496	6,314602766	8,179834466	10,97768201	24,03430391	5,386266544	5,407394188
16	0,4625849553	0,5986534725	5,911578249	7,108885194	8,904845612	17,28599423	5,315870296	5,330469138
17	0,735035	1,157576034	7,832751035	10,1479031	13,62063121	29,82669569	6,676634624	6,683900954
18	0,6914230298	-1,329500788	5,127614361	2,468612785	-1,519889579	-20,13290061	6,455026562	6,444517688
19	0,7145687889	-0,0003195498526	6,572524395	6,571885295	6,570926645	6,566452948	6,572843498	6,572841265
20	0,2861212807	0,9573377468	5,38794415	7,302619644	10,17463288	23,57736134	4,441941198	4,49676454
21	0,9655804031	-0,257647314	7,570254701	7,054960073	6,282018132	2,674955736	7,827799348	7,827285852
22	0,7905977156	1,554627644	8,507616222	11,61687151	16,28075444	38,04554146	6,954473348	6,961864312
23	0,6681512858	0,7260862407	7,066842669	8,519015151	10,69727387	20,86248124	6,342035677	6,348407486
24	0,6741342501	-0,7184744636	5,652196787	4,21524786	2,059824469	-7,998818021	6,369441177	6,363268011
25	0,1740635458	-1,458253388	2,412064341	-0,5044424345	-4,879202598	-25,29475003	3,839438717	3,66888338
26	0,2634867821	2,205608828	6,523042739	10,93426039	17,55108688	48,42961047	4,346419453	4,47994237
27	0,5940025564	-0,7031282057	5,266884576	3,860628165	1,751243548	-8,092551332	5,968215231	5,959178687
28	0,5735949306	2,435909315	8,303883968	13,1757026	20,48343054	54,58616095	5,874841228	5,908484176
29	0,8713902962	1,095595155	8,452546636	10,64373695	13,93052241	29,26885458	7,357650381	7,361137572

C в данном случае равна 5A+3

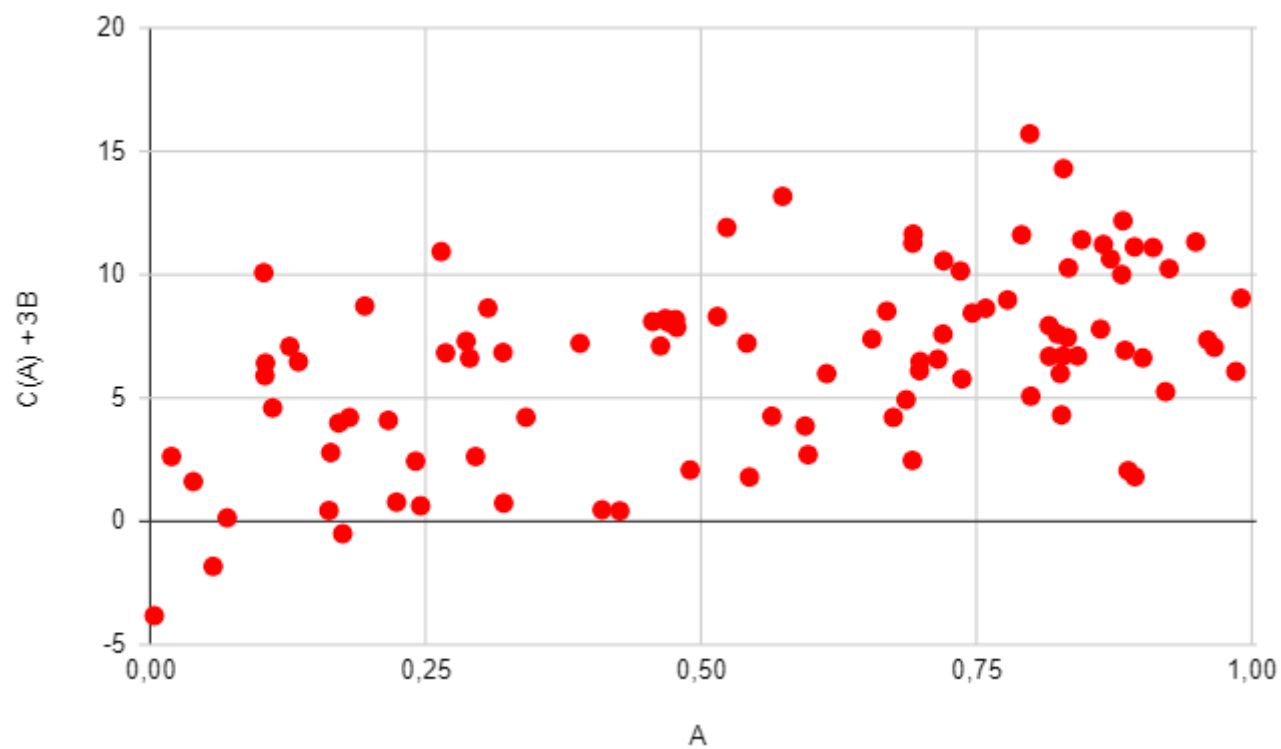
Количество точек 100

Демонстрация разброса точек для функций:

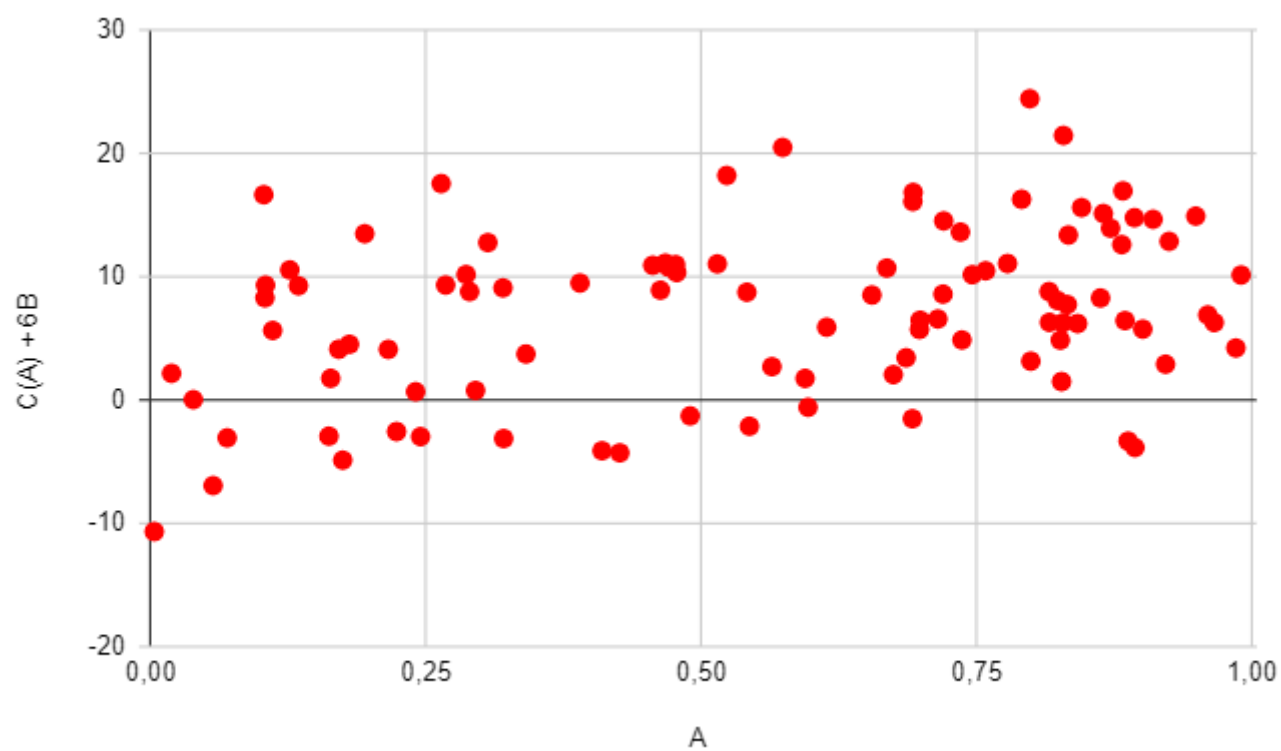
C + B



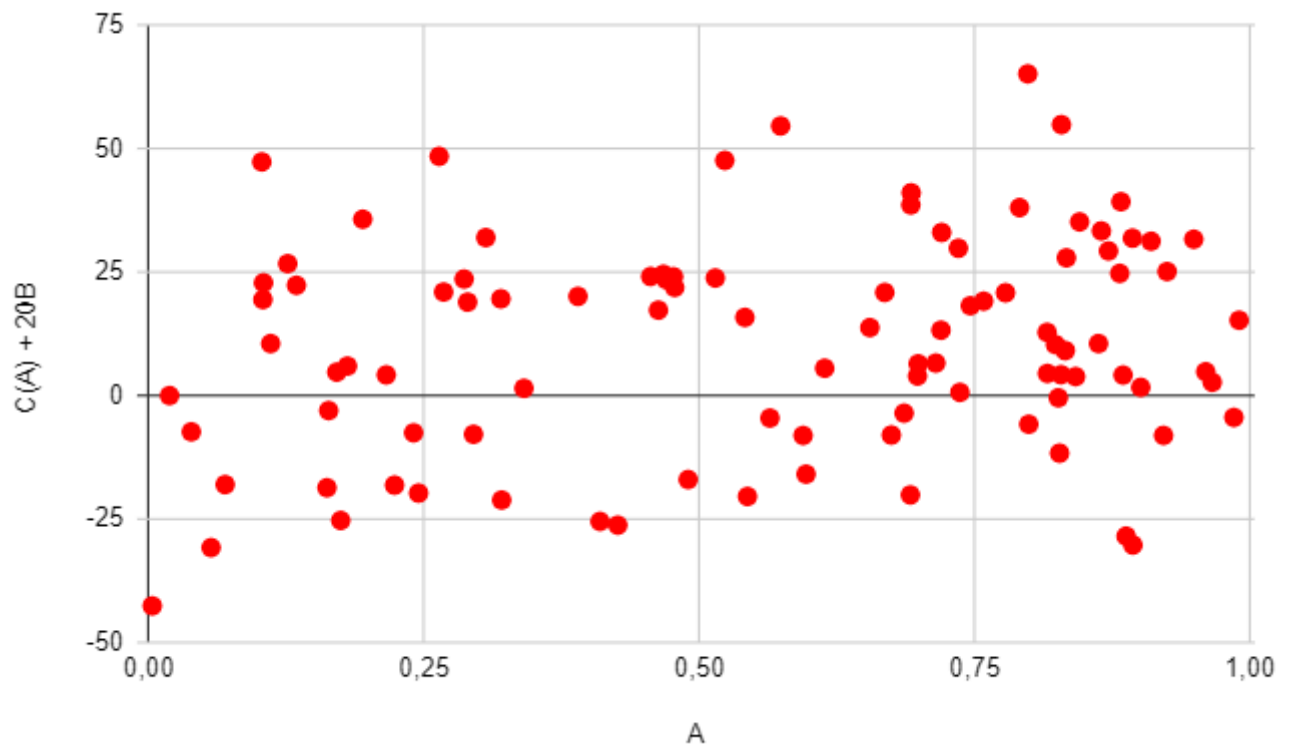
$C + 3B$



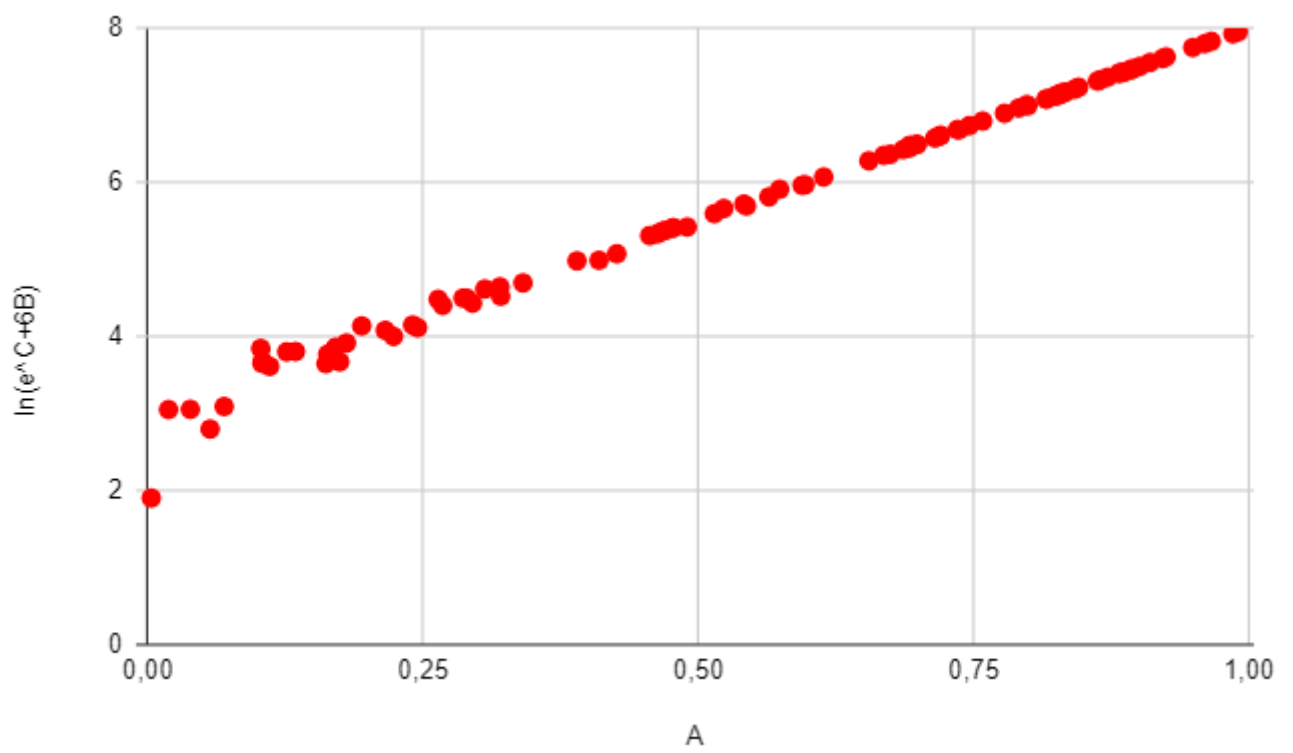
$C + 6B$



C + 20B



$\ln(e^A C + 6B)$



Программа на C++ для вычисления коэффициентов уравнения  $ax + b$ :

```
//

#include <algorithm>
#include <fstream>
#include <iostream>
#include <sstream>
#include <vector>

int main() {
    // X - random number between 0 and 1
    // Y - function of X (e.g. "2X + 1" in example.data but program doesn't know that)
    std::vector<double> X, Y;

    // DATA GETTING
    std::cout << "Enter file with X & Y data:\n> ";
    std::string path;
    std::getline(std::cin, path);
    std::ifstream file(path);

    std::string line;
    while (std::getline(file, line)) {
        std::istringstream iss(line);
        std::string value;
        double x, y;
        while (std::getline(iss, value, '\t')) { // reading X, tab is separator
            std::replace(value.begin(), value.end(), ',', '.'); // replace , to .
            if (!(std::istringstream(value) >> x)) {
                std::cerr << "Error while reading X from file\n";
                return 1;
            }
            if (std::getline(iss, value, '\t')) { // reading Y
                std::replace(value.begin(), value.end(), ',', '.');
                if (!(std::istringstream(value) >> y)) {
                    std::cerr << "Error while reading Y from file\n";
                    return 1;
                }
            }
            X.push_back(x);
            Y.push_back(y);
        }
    }

    // DATA PROCESSING
    // mean/error/epsilon/dispersion/whatever of X and Y
    double meanX = 0.0, meanY = 0.0;
    for (size_t i = 0; i < X.size(); i++) {
        meanX += X[i];
        meanY += Y[i];
    }
    meanX /= X.size();
```

```

meanY /= Y.size();

// calculate coefficients
double numerator = 0.0, denominator = 0.0;
for (size_t i = 0; i < X.size(); i++) {
    numerator += (X[i] - meanX) * (Y[i] - meanY);
    denominator += (X[i] - meanX) * (X[i] - meanX);
}
double A = numerator / denominator;
double B = meanY - A * meanX;

std::cout << "Y = " << A << "X + " << B << "\n";
return 0;
}

//

```

Подав в программу нужные значения X и Y (A и C) получаем результат для функций:

C + B

```

Enter file with X & Y data:
> c+b.dat
Y = 4.39187X + 3.29362

```

C + 3B

```

Enter file with X & Y data:
> c+3b.dat
Y = 3.1756X + 3.88086

```

C + 6B

```

Enter file with X & Y data:
> c+6b.dat
Y = 1.35119X + 4.76172

```

C+20B

```

Enter file with X & Y data:
> c+20b.dat
Y = -7.16269X + 8.87239

```

$\ln(e^C + B)$

```

Enter file with X & Y data:
> ln(e^C + B).dat
Y = 5.002X + 2.99905

```

$\ln(e^C + 6B)$

```

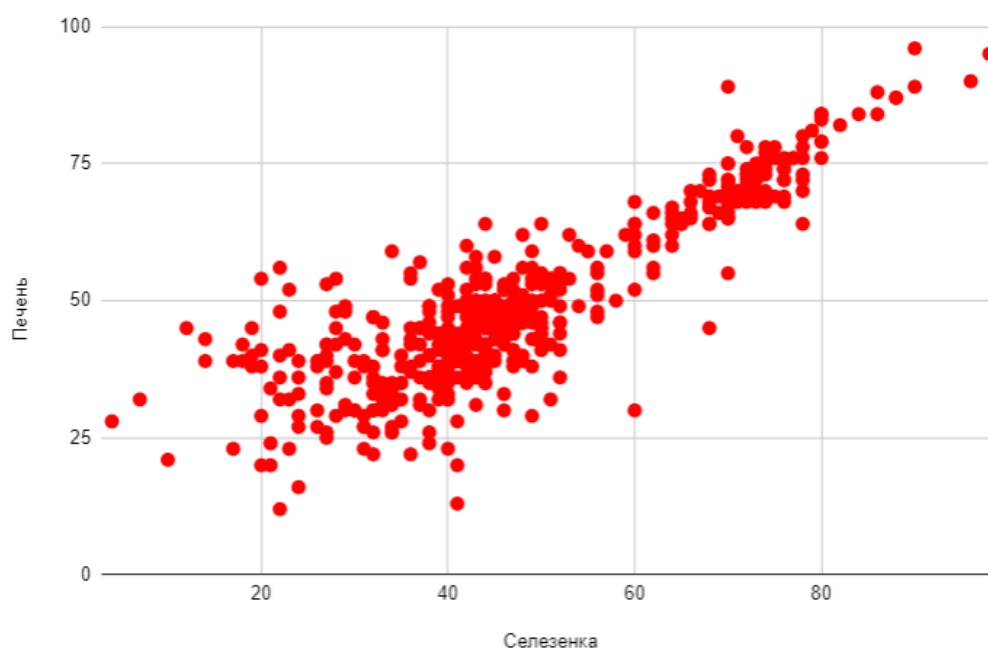
Enter file with X & Y data:
> ln(e^C + 6B).dat
Y = 5.04008X + 2.97588

```

Проведем то же исследование с использованием медицинских данных:

Селезенка	Печень
X	Y
24	29
32	47
42	36
45	44
27	42
26	27
48	46
43	31
29	43
40	36
50	42
32	36
38	26
42	60
43	39
47	54
32	22
34	27
14	39
39	52
44	54
24	16
40	43
22	12

График распределения точек на графике:



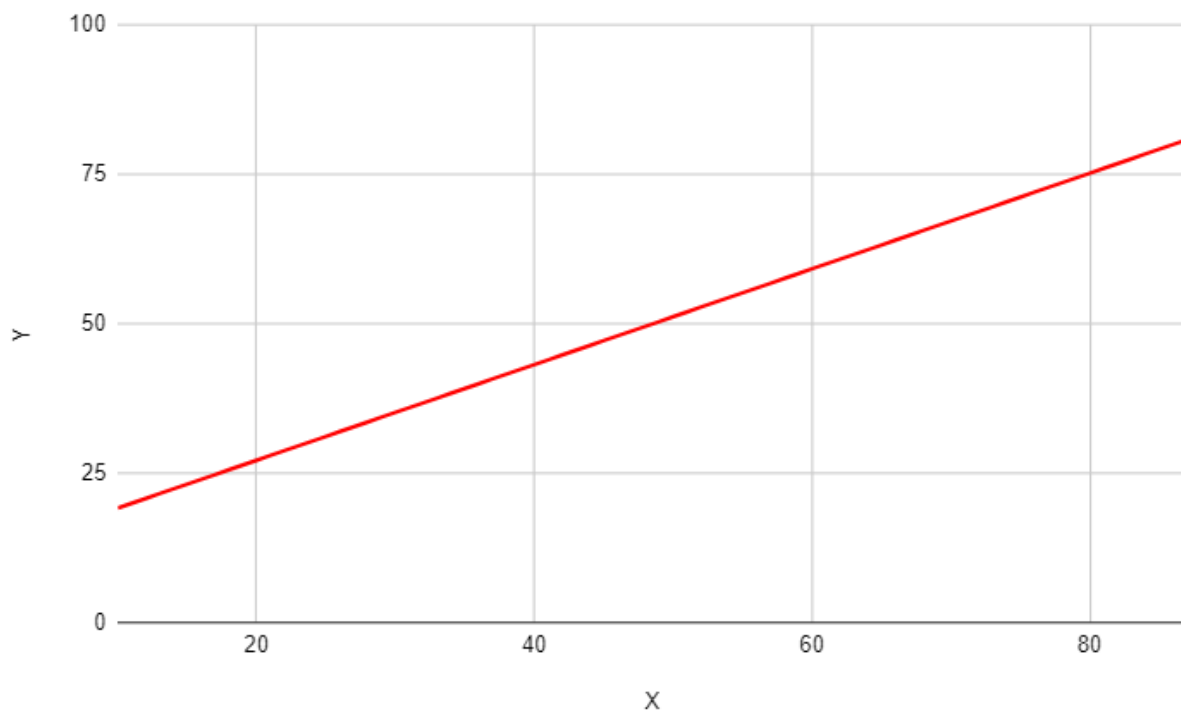


Предполагаемая функция:

```
Enter file with X & Y data:
> medical.dat
Y = 0.800538X + 11.1396
```

предполагаемая функция 0.800538X+11.1396

X	Y
10	19,14498
11	19,945518
12	20,746056
13	21,546594
14	22,347132
15	23,14767
16	23,948208
17	24,748746
18	25,549284
19	26,349822
20	27,15036
21	27,950898
22	28,751436
23	29,551974
24	30,352512
25	31,15305
26	31,953588
27	32,754126



Вывод: меньшая ошибка = большая точность, приближение к оригинальной линейной функции С, большая ошибка = меньшая точность, отдаление от оригинальной функции С.