МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ ДНР
ГОУ ВПО «Донецкий национальный технический университет»

Кафедра «Разработка месторождений полезных ископаемых»

Касьяненко Андрей Леонидович

Методические указания и контрольные задания для индивидуальной работы по дисциплине «Основы научных исследований»

ДОНЕЦК 2019

Тема индивидуальной работы:

«Установление статистической зависимости между случайными параметрами экспериментальных исследований путем прямой обработки данных»

Варианты заданий на выполнение контрольной работы представлены в Приложении А. Необходимый вариант определяется по сумме двух последних цифр шифра (номера зачетной книжки) студента

Установить вид зависимости $y_i = f(x_i)$

1. Представление исходных данных на координатной плоскости

В соответствии с заданием все представленные в табл. 1 исходные данные наносятся на координатную плоскость Х, У.

Линейный масштаб по осям координат принимается одинаковым.

Пример.

Таблица 1. Исходные данные эксперимента

№ точки		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
Вариант №1	Х	2	4	5	6	7	8	9	11	12	15	17	16	21	24	27	31
	У	4	3	5	7	9	6	8	10	14	13	18	19	21	18	22	21

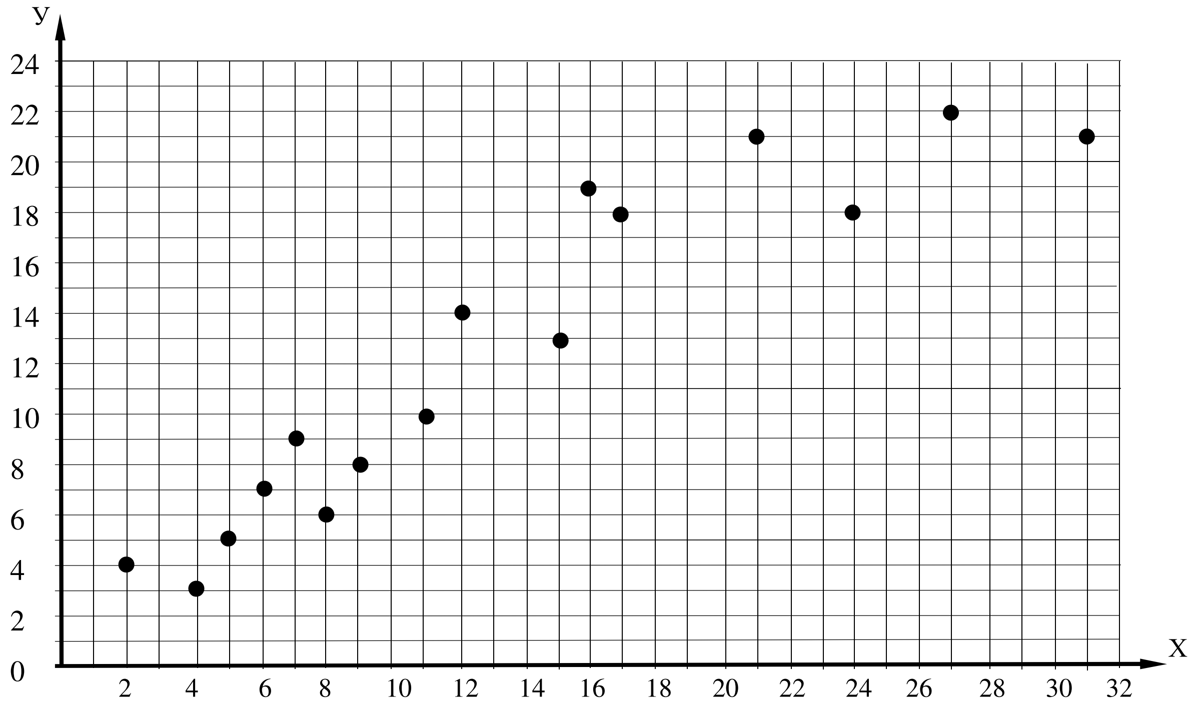

Рис. 1 Исходные данные на координатной плоскости Х, У

2. Проверка исходных данных на однородность

Обычно при решении практических горно-технических задач необходимо рассматривать целый ряд взаимовлияющих факторов [1,2]. Устанавливая зависимость между парой факторов $y_i = f(x_i)$, необходимо принимать все остальные факторы одинаковыми и не влияющими на рассматриваемую зависимость.

В математической статистике считается [3-5], что при выполнении условия

$$\frac{|Y_i - \overline{Y}|}{\sigma_y} \le 3 \qquad (1)$$

с вероятностью 0,997 все данные однородны и принадлежат одной совокупности. Если же какое-либо значение результата не подчиняется данному условию, то оно должно быть исключено из общего числа факторов.

В выражении (1):

$У_i$ – текущее значение фактора У;

$\overline{У}$ - среднее значение фактора У для всей совокупности данных;

$$\overline{У} = \frac{\sum_{1}^{n} У_i}{n} = \frac{4+3+5+7+9+6+8+10+14+13+18+19+21+18+22+21}{16} = 12{,}375 \approx 12{,}4$$

n – общее число исходных данных;

$\sigma_у$ – среднеквадратическое отклонение исходных данных от среднего значения фактора У,

$$\sigma_у = \sqrt{\frac{(У_1-\overline{У})^2+(У_2-\overline{У})^2+(У_3-\overline{У})^2+...+(У_n-\overline{У})^2}{n}}\,. \qquad (2)$$

Для заданных условий:

$$\sigma_у = \sqrt{\frac{(4-12{,}4)^2+(3-12{,}4)^2+(5-12{,}4)^2+(7-12{,}4)^2+...+(21-12{,}4)^2}{16}} = \sqrt{\frac{640{,}66}{16}} \approx 6{,}33.$$

Тогда, для минимального и максимального значений фактора У:

$$\frac{|3-12{,}4|}{6{,}33} = 1{,}48 \le 3;$$

$$\frac{|22-12{,}4|}{6{,}33} = 1{,}52 \le 3.$$

Таким образом, все исходные данные представленной совокупности факторов являются однородными.

3. Определение необходимого количества наблюдений для установления достоверной зависимости

Необходимое количество наблюдений определяется из выражения [3-5]:

$$n_{необх} = \frac{\sigma_у^2 \cdot t^2}{\Delta_у^2}\,, \qquad (3)$$

где $\sigma_у$ - дисперсия или среднеквадратическое отклонение, $\sigma_у = 6{,}33$;

t – стандартизованная случайная величина (параметр нормированной функции Лапласа), определяемая в зависимости от доверительной вероятности: t = 1,96 при вероятности 0,95 и t = 3 при вероятности 0,997. Обычно в инженерной практике принимается t = 1,96;

$\Delta_у$ – величина или предел ошибки, которая может быть допущена из-за недостаточного количества экспериментальных данных. Обычно в расчетах принимают:

$$\Delta_у = (0{,}05-0{,}15)\,\overline{У} = 0{,}1 \cdot 12{,}4 = 1{,}24.$$

Тогда:

$$n_{необх} = \frac{6{,}33^2 \cdot 1{,}96^2}{1{,}24^2} = \frac{40{,}07\ 3{,}84}{1{,}54} \approx 100\,.$$

Таким образом, при $n_{необх} = 100$ среднее значение результата фактора У будет отличаться от среднего при $n \to \infty$ (для генеральной совокупности) не более чем на 10% с вероятностью 0,95.

Отсюда можно сделать вывод, что представленная совокупность из 16 факторов не даст возможность получить ошибку в пределах 5 – 13%.

4. Определение величины ошибки при заданном числе наблюдений

В практике зачастую необходимо определять вид зависимости для совокупности факторов при их ограниченном количестве.

В такой ситуации величина ошибки определяется из выражения [3-5]:

$$\Delta_y = \sqrt{\frac{\sigma_y^2 \cdot t^2}{n}} \cdot \frac{100}{\overline{У}} = \sqrt{\frac{6{,}33^2 \cdot 1{,}96^2}{16}} \cdot \frac{100}{12{,}4} = 25\%\,. \qquad (4)$$

5. Установление вида зависимости

По расположению точек на координатной плоскости (поле корреляции) видно, что между рассматриваемыми факторами существует прямолинейная зависимость вида

$$У = a + b \cdot X \qquad (5)$$

Для определения параметров прямолинейно зависимости воспользуемся согласно способа наименьших квадратов системой двух нормальных уравнений вида [3-5]:

$$\left.\begin{aligned} \sum_{1n}^{n} У_i &= n \cdot a + b \cdot \sum_{1}^{n} X_i \\ \sum_{1}^{n} У_i \cdot X_i &= a \sum_{1}^{n} X_i + b \sum_{1}^{n} X_i^2 \end{aligned}\right\} \qquad (6)$$

Вычисление значений, входящих в систему уравнений (6), сведем в таблицу 2.

Таблица 2. Исходные данные для расчета параметров системы (6)

№точки	У	$У^2$	Х	$Х^2$	$У \cdot Х$
1	4	16	2	4	8
2	3	9	4	16	12
3	5	25	5	25	25
4	7	49	6	36	42
5	9	81	7	49	63
6	6	36	8	64	48
7	8	64	9	81	72
8	10	100	11	121	110
9	14	196	12	144	168
10	13	169	15	225	195
11	18	324	17	289	306
12	19	361	16	256	304
13	21	441	21	441	441
14	18	324	24	576	432

15	22	484	27	729	594
16	21	441	31	961	651
Сумма	198	3120	215	4117	3471

Из табл. 2 имеем:

$$\sum_{1}^{n} У_i = 198; \ \sum_{1}^{n} X_i = 215; \ \sum_{1}^{n} У_i \cdot X_i = 3471; \ \sum_{1}^{n} X_i^2 = 4117.$$

Тогда систему уравнений (6) можно переписать в виде:

$$\left.\begin{aligned} 198 &= 16 \cdot a + 215 \cdot b \\ 3471 \cdot &= 215 \cdot a + 4117 \cdot b \end{aligned}\right\} \qquad (7)$$

Решаем систему уравнений (6) методом подстановки:

- из первого уравнения получим:

$$a = \frac{198 - 215 \cdot b}{16}; \qquad (8)$$

- подставляем выражение (8) во второе уравнение системы (7):

$$3471 = 215 \cdot \frac{198 - 215 \cdot b}{16} + 4117 \cdot b;$$

$$3471 = 215 \cdot \frac{198}{16} - \frac{215}{16} \cdot b + 4117 \cdot b;$$

$$3471 = 215 \cdot (12{,}4 - 13{,}4 \cdot b) + 4117 \cdot b;$$

$$3471 = 2666 - 2881 \cdot b + 4117 \cdot b;$$

$$1236\, b = 805$$

$$b = 0{,}65.$$

Из выражения (8) определяем значение коэффициента «а»:

$$a = \frac{198 - 215 \cdot b}{16} = \frac{198 - 215 \cdot 0{,}65}{16} = (198 - 139{,}8) \cdot \frac{1}{16} = 3{,}64.$$

Таким образом, уравнение линейной зависимости для случайных данных эксперимента (табл. 1) в соответствии с выражением (5) можно записать в виде:

$$У = 3{,}64 + 0{,}65 \cdot X \qquad (9)$$

Для построения графика прямой линии (9) на координатной плоскости достаточно получить всего лишь 2 точки:

- при У = 0 из выражения (9) получим: $0 = 3{,}64 + 0{,}65 \cdot X$; $X = -\frac{3{,}64}{0{,}65} = -5{,}6$;

- при X = 0, аналогично - $У = 3{,}64 + 0{,}65 \cdot 0$; $У = 3{,}64$.

Откладывая полученные отрезки на координатных осях Х и У и соединяя их прямой линией мы получим искомый график зависимости (9):

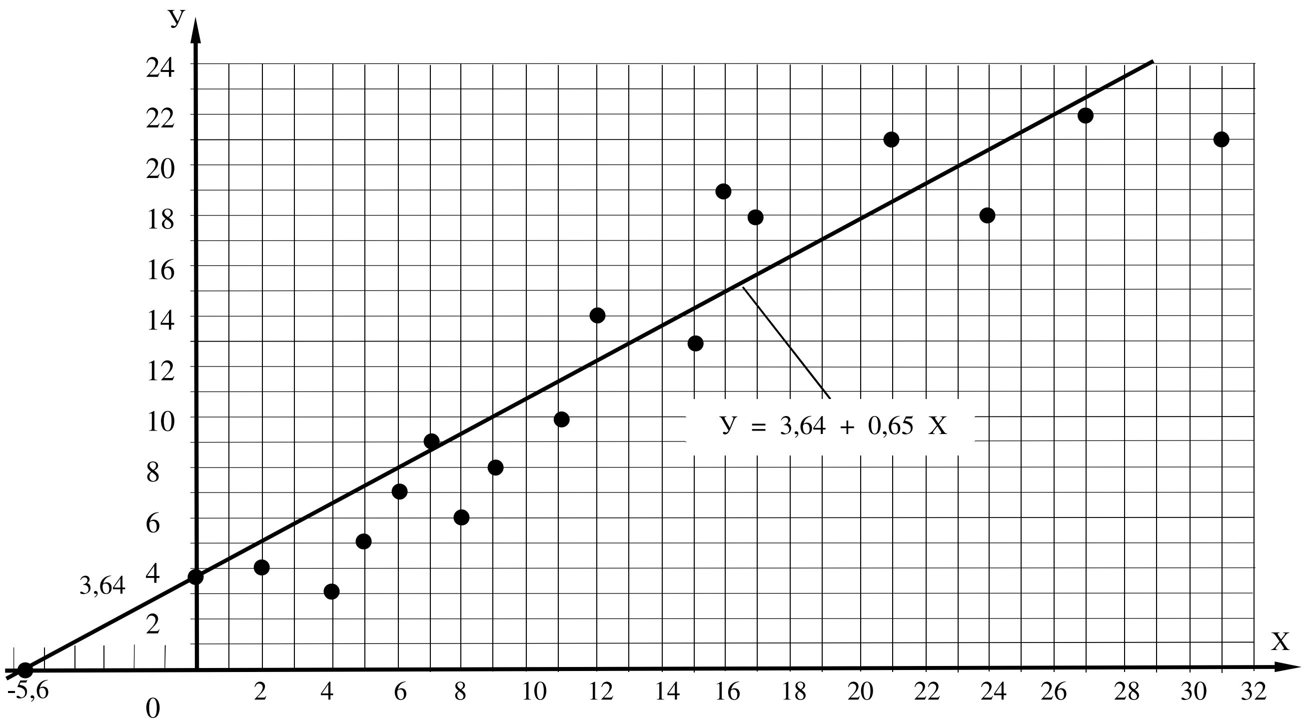

Рис. 2 График зависимости У = 3,64 + 0,65 · Х

6. Установление тесноты связи между рассматриваемыми случайными параметрами эксперимента

В качестве показателя тесноты связи между исследуемыми параметрами при прямолинейной зависимости принят коэффициент корреляции, который определяется из выражения [3-5]:

$$r = \frac{n \cdot \sum_{1}^{n}(У_i \cdot X_i) - \sum_{1}^{n} У_i \cdot X_i}{\sqrt{n \cdot \sum_{1}^{n} У_i^2 - \left(\sum_{1}^{n} У_i\right)^2} \cdot \sqrt{n \cdot \sum_{1}^{n} X_i^2 - \left(\sum_{1}^{n} X_i\right)^2}} \qquad (10)$$

Подставляя исходные данные из таблицы (2) получим:

$$r = \frac{16 \cdot 3471 - 198 \cdot 215}{\sqrt{16 \cdot 3120 - (198)^2} \cdot \sqrt{16 \cdot 4117 - (215)^2}} = \frac{55536 - 42570}{\sqrt{49920 - 39204} \cdot \sqrt{65872 - 46225}} =$$

$$= \frac{12966}{103{,}5 \cdot 140{,}2} = \frac{12966}{14510{,}7} = 0{,}89\,.$$

7. Определение надежности тесноты связи между рассматриваемыми величинами

Надежность связи оценивается погрешностью коэффициента корреляции и определяется из выражения [3-5]:

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}} \tag{11}$$

$$\sigma_r = \frac{1 - 0{,}89^2}{\sqrt{16}} = \frac{1 - 0{,}7921}{4} = \frac{0{,}2079}{4} = 0{,}052 .$$

Связь между параметрами зависимости считается надежной, если выполняется условие:

$$\mu_r = \frac{|\, r \,|}{\sigma_r} \geq 3 .$$

В нашем случае:

$$\mu_r = \frac{|\, 0{,}89 \,|}{0{,}052} = 17{,}1 \geq 3 ,$$

значит в данном случае связь между случайными данными эксперимента (табл.. 1) является надежной.

8. Определение доверительных интервалов коэффициента корреляции и параметров зависимости

Для установления параметров зависимости использовалось ограниченное количество исходных данных, поэтому для оценки величины коэффициента корреляции ρ (определенного по генеральной совокупности) используются доверительные интервалы, которые определяются из выражений [3-5]:

- для коэффициента корреляции:

$$r - t \cdot \sigma_r \prec \rho \prec r + t \cdot \sigma_r \tag{12}$$

- для коэффициентов регрессии (зависимости):

$$\left.\begin{aligned} a - t \cdot \sigma_a \prec \alpha \prec r + t \cdot \sigma_a \\ b - t \cdot \sigma_b \prec \beta \prec r + t \cdot \sigma_b \end{aligned}\right\} \tag{13}$$

Среднеквадратичное отклонение коэффициентов регрессии (параметров зависимости) определяется из выражения:

$$\sigma_a = \sigma_b = \frac{\sigma_y}{\sigma_x} \cdot \frac{1 - r^2}{\sqrt{n}} \tag{14}$$

По аналогии с выражением (2) величина σ_x определяется как:

$$\sigma_x = \sqrt{\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + (X_3 - \overline{X})^2 + ... + (X_n - \overline{X})^2}{n}} . \qquad (2)$$

Для заданных условий:

$$\sigma_x = \sqrt{\frac{(2-13,4)^2 + (4-13,4)^2 + (5-13,4)^2 + (6-13,4)^2 + ... + (31-13,4)^2}{16}} = \sqrt{\frac{943,06}{16}} \approx 7,68.$$

Тогда:

$$\sigma_a = \sigma_b = \frac{6,33}{7,68} \cdot \frac{1-0,89^2}{\sqrt{16}} = 0,824 \cdot 0,052 = 0,043 .$$

Тогда из выражения (12) получим:

$$0,89 - 1,96 \cdot 0,052 = 0,79 \prec \rho \prec 0,89 + 1,96 \cdot 0,052 = 0,99$$

$$0,79 < \rho < 0,99.$$

Из выражения (13):

$$3,64 - 1,96 \cdot 0,043 = 3,56 \prec \alpha \prec 3,64 + 1,96 \cdot 0,043 = 3,72;$$

$$0,65 - 1,96 \cdot 0,043 = 0,57 \prec \beta \prec 0,65 + 1,96 \cdot 0,043 = 0,73.$$

или:

$$3,56 \prec \alpha \prec 3,72;$$

$$0,57 \prec \beta \prec 0,73.$$

Таким образом, приняв t = 3 при доверительной вероятности 0,95 найдем генеральный коєффициент корреляции, который изменяется в пределах:

$$0,79 < \rho < 0,99.$$

ЛИТЕРАТУРА

1. **Кожухар, В. М.** Основы научных исследований [Электронный ресурс] : учебное пособие для вузов / В.М. Кожухар ; гл. ред. А.Е. Илларионова. - 1 Мб. - Москва : Изд.-торг. корпорация «Дашков и К», 2010. - 1 файл. - Режим доступа: http://ed.donntu.org/books/19/cd9317.pdf - Загл. с экрана.
2. **Пономарев, А. Б.** Методология научных исследований [Электронный ресурс] : учебное пособие для вузов / А.Б. Пономарев, Э.А. Пикулева ; ФГБОУ ВПО «Перм. нац. исслед. политехн. ун-т». - 1 Мб. - Пермь : Изд-во Перм. нац. исслед. политехн. ун-та, 2014. - 1 файл. - Режим доступа: http://ed.donntu.org/books/cd5139.pdf - Загл. с экрана
3. **Бурда, А. Г.** Основы научно-исследовательской деятельности [Электронный ресурс] : учебное пособие (курс лекций) / А.Г. Бурда ; ФГБОУ ВПО "Кубан. гос. аграрный ун-т". - 1 Мб. - Краснодар : [б.и.], 2015. - 1 файл. - Режим доступа: http://ed.donntu.org/books/19/cd9326.pdf - Загл. с экрана.
4. Основы научных исследований и патентоведение [Электронный ресурс] : учебно-методическое пособие для вузов / Новосиб. гос. аграрн. ун-т, Инж. ин-т ; сост.: С.Г. Щукин и др.. - 1 Мб. - Новосибирск : НГАУ, 2013. - 1 файл. - Режим доступа: http://ed.donntu.org/books/20/cd9707.pdf - Загл. с экрана.

Приложение А

ВАРИАНТЫ ЗАДАНИЙ

№ вар.	Параметр	Значение точек																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	X	1	2	2,5	0,5	3	3,5	4	5	6	7	6,5	6,1	5,2	4,7	3,5	1,5	1,7
	У	0,3	0,7	1	0,2	1,1	1,2	1,5	1,6	2,3	2,6	2	2,3	1,9	1,3	1,4	0,6	0,5
2	X	1	2	3	4	5	6	7	0,5	1,1	2,1	3,1	4,1	5,1	6,1	7,1	2,5	3,5
	У	0,9	1,8	2,8	3,7	4,7	5,7	6,7	0,5	0,9	2	2,6	3,8	4,5	5,9	6,5	2,2	3,2
3	X	1	2	3	4	5	6	7	8	1,5	2,5	3,5	4,5	5,5	6,5	7,5	0,5	0,7
	У	0,5	1,2	1,4	2,1	2,3	3,1	3,3	4	0,9	1,2	1,8	2,4	2,4	3,2	3,6	0,3	0,5
4	X	0,9	1,9	2,6	0,5	3,1	3,4	3,8	6	5,1	7	6,5	5,1	6,1	4,7	3,5	1,5	1,8
	У	0,3	0,7	1	0,2	1,1	1,2	1,6	2,2	1,7	2,5	2,1	1,9	2,2	1,4	1,6	0,7	0,6
5	X	0,8	1,8	2,8	3,8	4,8	5,8	6,8	0,8	1	2	3	4	5	6	7	2,5	3,5
	У	0,8	1,8	2,6	3,5	4,5	5,4	6,4	0,5	0,8	2	2,6	3,8	4,5	5,9	6,5	2,1	3,1
6	X	10	25	20	5	3	35	40	50	60	70	65	61	52	47	35	15	17
	У	3	10	7	2	11	12	15	16	23	26	20	23	19	13	14	6	5
7	X	5	12	10	2,5	1,5	17	20	25	30	35	32	30	26	23	17	8	8
	У	1,5	5	3,5	1	5,5	6	8	8	11	13	10	12	9	7	7	3	2,5
8	X	10	20	30	40	50	60	70	5	11	21	31	41	51	61	71	25	35
	У	9	18	28	37	47	57	67	5	9	20	26	38	45	59	65	22	32
9	X	0,1	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	0,7	2,3
	У	0,4	1,2	2,1	2,8	3,7	3,8	4,2	4,5	4,6	5,3	5,4	5,7	5,8	5,9	6,3	1,5	3,1
10	X	0.05	0.25	0,5	0,7	1	1,25	1,5	1,75	2	2,2	2,5	2,8	3	3,2	3,5	0,35	1,1
	У	0,2	0,6	1,05	1,4	1,8	1,9	2,1	2,2	2,3	2,6	2,7	2,8	2,9	2,9	3,1	0,75	1,5
11	X	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	1,2	2,2
	У	0,2	0,3	0,5	0,5	0,6	0,7	1	1,1	1,3	1,8	2	2,4	2,5	3,1	3,2	0,4	0,5
12	X	5	10	15	10	25	30	35	40	45	50	55	60	65	70	75	12	22
	У	2	3	5	5	6	7	10	11	13	18	20	24	25	31	32	4	5
13	X	1	5	10	15	20	25	30	35	40	45	50	55	60	65	70	7	23
	У	4	12	21	28	37	38	42	45	46	53	54	57	58	59	63	15	31
14	X	0,5	25	5	7	10	12	15	17	20	22	25	28	30	32	35	3,5	11
	У	2	6	12	14	18	19	21	22	23	36	27	28	29	29	32	7	16
15	X	2,5	5	7	5	12	15	17	20	22	25	27	30	32	35	37	6	11
	У	1	1,5	2,5	2,5	3	3,5	5	5,5	6,5	9	10	12	12	16	16	2	2,5
16	X	0,5	0,7	0,8	0,4	1,1	1,5	2	2,5	3	3,2	3,6	4	5,5	6	6,5	7	0,1
	У	0,6	0,7	0,7	0,2	1	1,7	2	2,7	2,8	3	3,8	4	5,2	6,3	6,4	7,1	0,1
17	X	10	20	30	40	50	60	70	80	90	5	15	25	35	45	55	65	75
	У	4,3	5,7	6,5	7	7,3	7,8	8	8	8,1	3	5,3	6,2	6,8	7,3	7,7	7,9	8
18	X	5	10	15	20	25	30	35	40	45	2,5	1,5	12	17	22	27	32	38
	У	2,1	2,8	3,2	3,5	3,6	3,9	4	4,1	4	1,5	2,7	3,1	3,4	3,6	3,8	3,8	4
19	X	2,5	5	7	10	12	15	17	20	22	1,2	3,8	6	8	11	14	16	19
	У	1,05	1,4	1,6	1,7	1,8	1,9	2	2,1	2	0,7	1,4	1,6	1,7	1,8	1,9	1,9	2
20	X	5	10	15	20	25	30	35	40	45	2,5	7,5	12	17	22	28	30	38
	У	10,5	14	16	17	18	19	20	21	20	7	14	16	17	18	19	19	20