

Статистика. Обработка и представление результатов исследования с помощью MS Excel

Меры центральной тенденции.

- **Мода** (Mode) – число наиболее часто встречающееся в данном множестве чисел.

Например, модой для чисел 2, 3, 3, 5, 7 и 10 будет 3.

- **Медиана** (Median) – число, которое является серединой множества чисел: половина чисел имеют значения большие, чем медиана, а половина чисел — меньшие.

Например, медианой для чисел 2, 3, 3, 5, 7 и 10 будет 4.

- **Среднее значение** (Mean, средне арифметическое) — это значение, которое является средним арифметическим, т. е. вычисляется путем сложения набора чисел с последующим делением полученной суммы на их количество.

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

Например, средним значением для чисел 2, 3, 3, 5, 7 и 10 будет 5 (результат деления суммы этих чисел, равной 30, на их количество, равное 6).

Меры изменчивости

- **Размах** (Range) – разность максимального и минимального значения.
Например, размах для чисел 2, 3, 3, 5, 7 и 10 будет 8 (Размах = 10-2 = 8).
- **Дисперсия** (Variance) – средний квадрат отклонений чисел от их средней величины

$$D = \frac{\sum(x_i - \bar{x})^2}{n-1} \text{ (для выборки)}$$

$$D = \frac{\sum(x_i - \bar{x})^2}{n} \text{ (для генеральной совокупности)}$$

Например, дисперсия для чисел 2, 3, 3, 5, 7 и 10 будет 10.4

Количество чисел 6, $n = 6$.

$$D = \frac{(2 - 6)^2 + (3 - 6)^2 + (3 - 6)^2 + (5 - 6)^2 + (7 - 6)^2 + (10 - 6)^2}{6 - 1}$$
$$D = \frac{16 + 9 + 9 + 1 + 1 + 16}{5} = 10.4$$

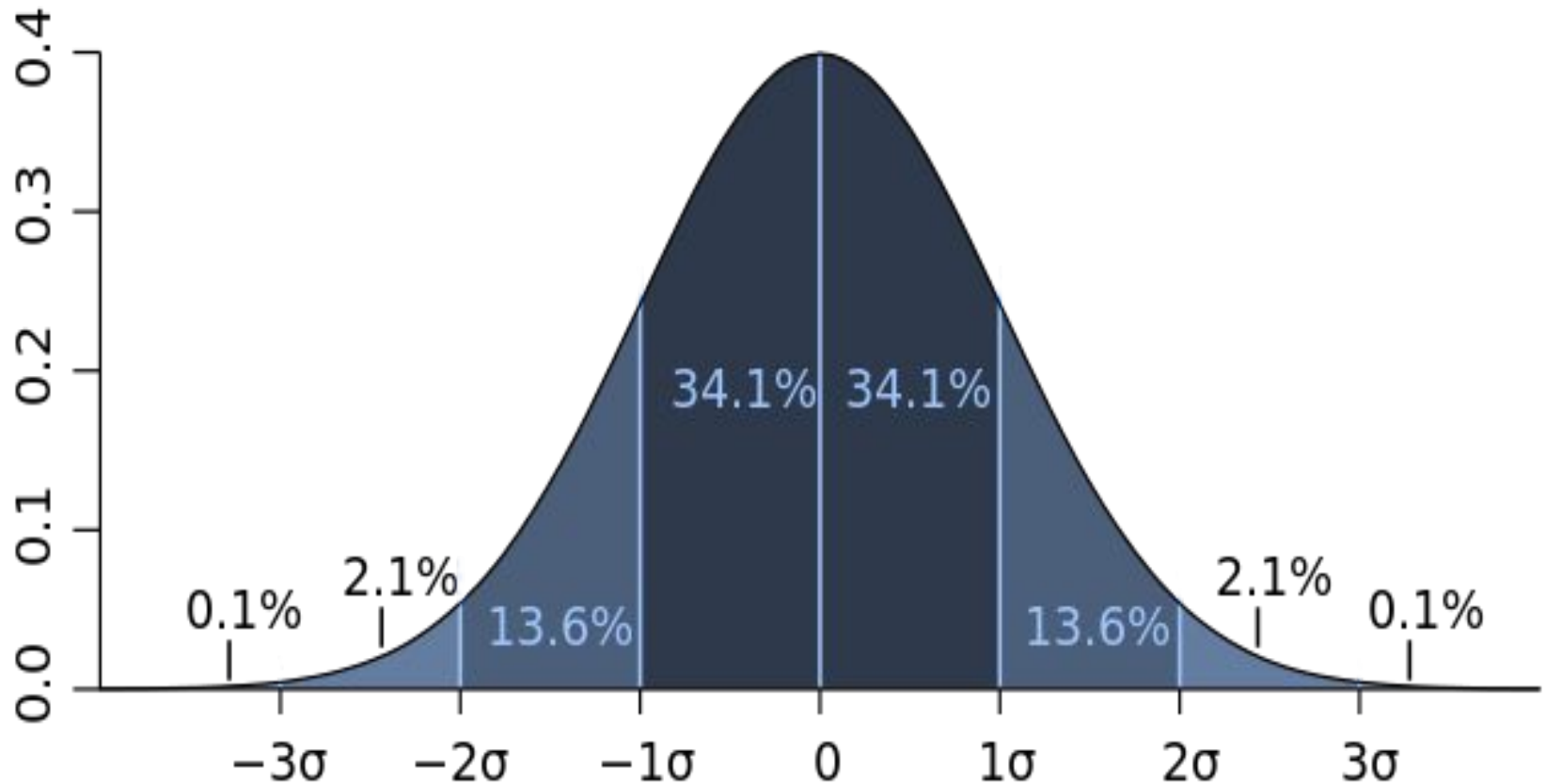
Меры изменчивости

- **Среднее квадратичное отклонение, стандартное отклонение** (для генеральной совокупности обозначается σ , а для выборки sd - standard deviation, стандартное отклонение). Вычисляется как корень из дисперсии:

$$\sigma = \sqrt{D}$$

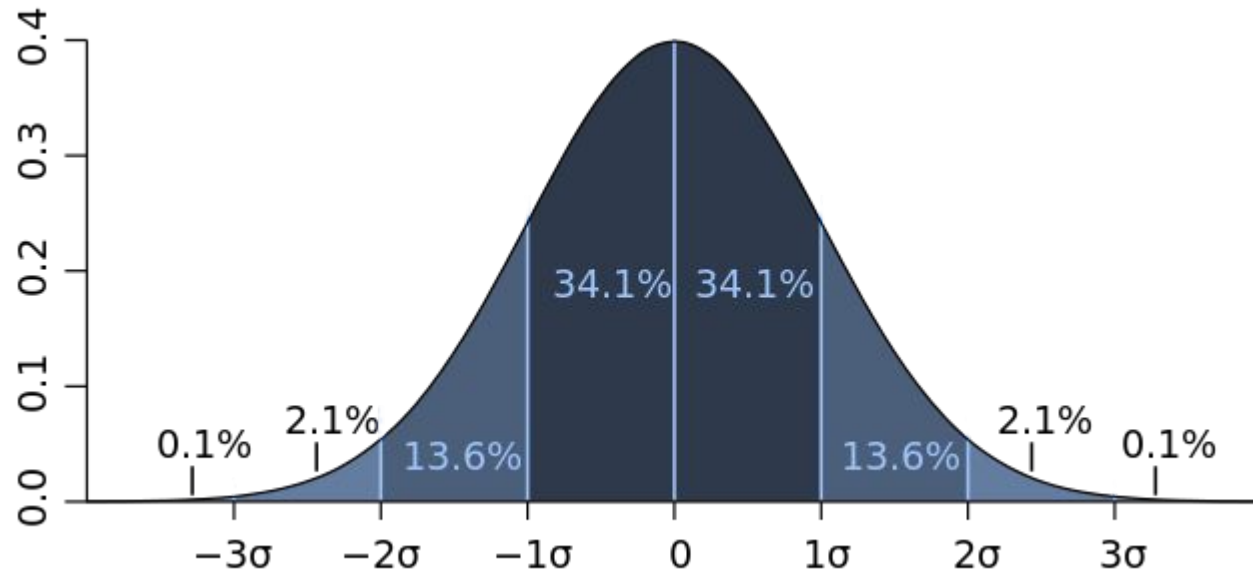
Например, среднее квадратичное отклонение для чисел 2, 3, 3, 5, 7 и 10 будет 3.22

Нормальное распределение



Этот вероятностный закон сообщает нам о том, в какой области наиболее возможно встретить измеряемые значения.

Нормальное распределение



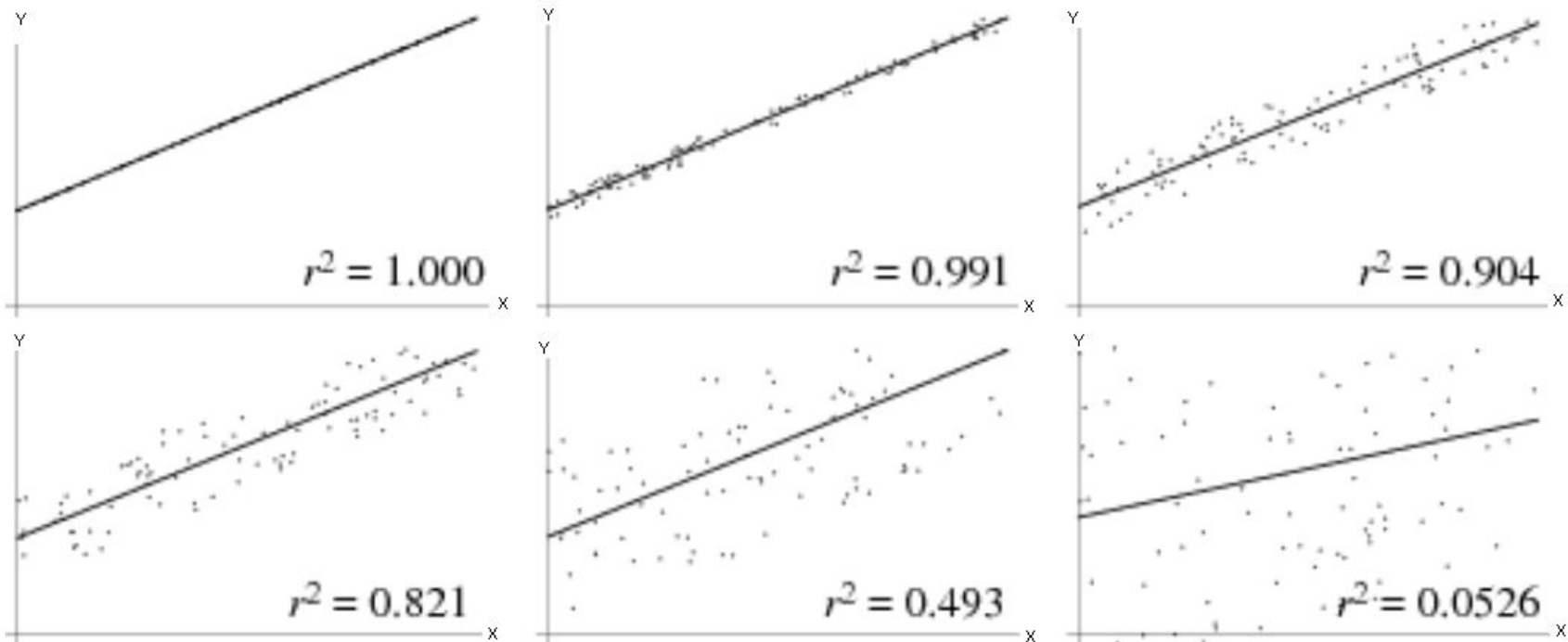
Из рисунка видно, что значения, укладывающиеся от 0 до $\pm 1\sigma$ будут встречаться с вероятностью в 68%, от $\pm 1\sigma$ до $\pm 2\sigma$ с вероятностью в 27%, от $\pm 2\sigma$ до $\pm 3\sigma$ – 4%, а более $\pm 3\sigma$ – имеет незначительную вероятность.

Из вышесказанного следует, что вероятность найти значения укладывающиеся в интервал от 0 до $\pm 2\sigma$ - 95%, а от 0 до $\pm 3\sigma$ практически 100%.

Коэффициент корреляции Пирсона

- Соотношение x и y линейное, если прямая линия, проведенная через центральную часть скопления точек, дает наиболее подходящую аппроксимацию наблюдаемого соотношения.
- Можно измерить, как близко находятся наблюдения к прямой линии, которая лучше всего описывает их линейное соотношение путем вычисления коэффициента корреляции Пирсона, обычно называемого просто коэффициентом корреляции

Графические примеры зависимости разброса значений и тренда коэффициента корреляции



Коэффициент корреляции Пирсона

Из свойств коэффициента корреляции r следует выделить:

- r изменяется в интервале от -1 до $+1$.
- знак r означает, увеличивается ли одна переменная по мере того, как увеличивается другая (положительный r), или уменьшается ли одна переменная по мере того, как увеличивается другая (отрицательный r).
- величина r указывает, как близко расположены точки к прямой линии.

Обработка и представление результатов исследования с помощью MS Excel

Температура кипения спиртов

Спирты	Вр (°C) а	Predicted b Вр (°C)
1	82.3	85.4
2	97.2	97.5
3	117.7	117.1
4	107.8	107.7
5	102.4	103
6	119.3	124.6
7	131.1	127.3
8	128	127.3
9	137.9	136.7
10	120.4	116
11	121.1	122.6
12	122.4	122.6
13	136.5	128.1
14	131.6	134.8
15	126.5	134.8
16	144.5	137.5
17	149	146.9
18	157.6	156.3
19	138.7	145
20	159	157.1
21	142	142.2
22	176.4	175.9
23	161.0	161.8
24	163.0	161.8
25	188.0	186.1
26	195.1	195.5
27	178.0	183.3
28	182.0	181.4

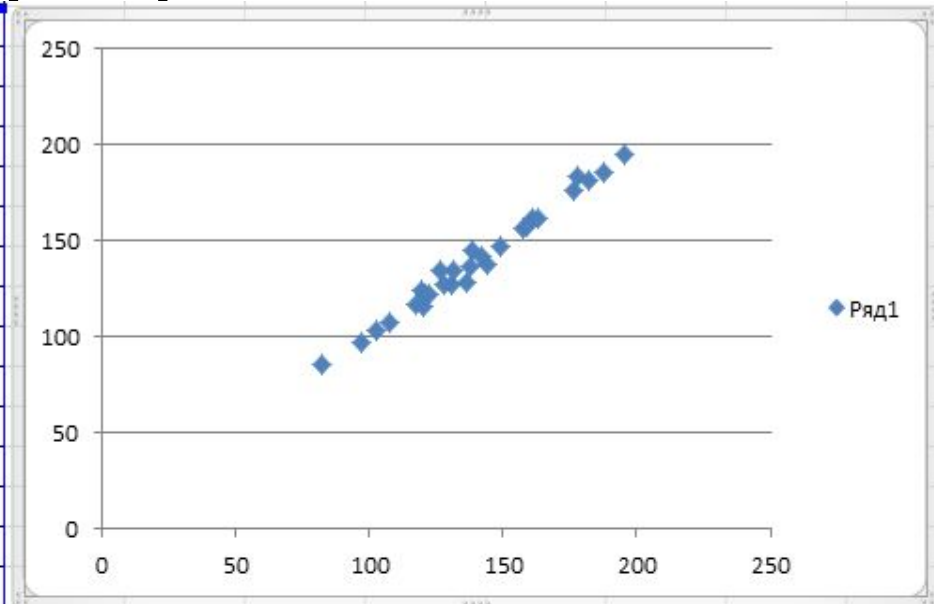
Построение корреляционного графика

3	1	82.3	85.4
4	2	97.2	97.5
5	3	117.7	117.1
6	4	107.8	107.7
7	5	102.4	103
8	6	119.3	124.6
9	7	131.1	127.3
10	8	128	127.3
11	9	137.9	136.7
12	10	120.4	116
13	11	121.1	122.6
14	12	122.4	122.6
15	13	136.5	128.1
16	14	131.6	134.8
17	15	126.5	134.8
18	16	144.5	137.5
19	17	149	146.9
20	18	157.6	156.3
21	19	138.7	145
22	20	159	157.1
23	21	142	142.2
24	22	176.4	175.9
25	23	161	161.8
26	24	163	161.8
27	25	188	186.1
28	26	195.1	195.5
29	27	178	183.3
30	28	182	181.4
31			

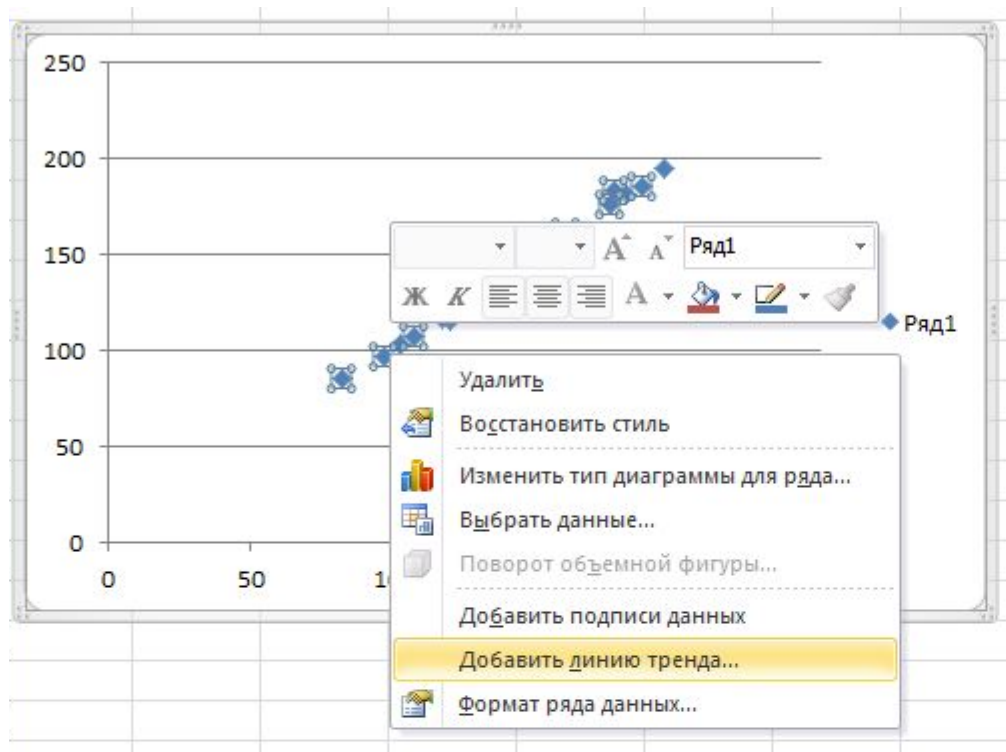
Для построения графика зависимости экспериментальных и рассчитанных значений выделяем области с численными значениями свойства. В меню «Вставка» в подменю «Диаграмма» выбираем «Точечная» → «Точечная с

Построение корреляционного графика

3	1	82.3	85.4
4	2	97.2	97.5
5	3	117.7	117.1
6	4	107.8	107.7
7	5	102.4	103
8	6	119.3	124.6
9	7	131.1	127.3
10	8	128	127.3
11	9	137.9	136.7
12	10	120.4	116
13	11	121.1	122.6
14	12	122.4	122.6
15	13	136.5	128.1
16	14	131.6	134.8
17	15	126.5	134.8
18	16	144.5	137.5
19	17	149	146.9
20	18	157.6	156.3
21	19	138.7	145
22	20	159	157.1
23	21	142	142.2
24	22	176.4	175.9
25	23	161	161.8
26	24	163	161.8
27	25	188	186.1
28	26	195.1	195.5
29	27	178	183.3
30	28	182	181.4

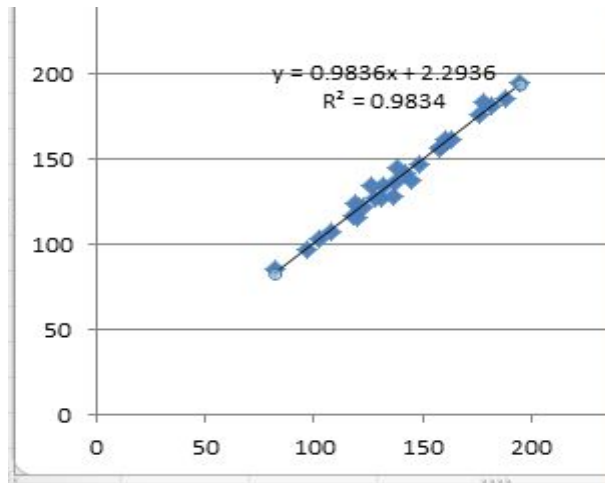


Построение корреляционного графика



Выбрав на появившемся графике любой маркер, нажимаем правую кнопку мыши, и в открывшемся меню выбираем «Добавить линию тренда». В открывшемся окне «Формат линии тренда» нужно установить галочки напротив «Показать уравнение на диаграмме» и «Поместить на диаграмму величину достоверности аппроксимации (R^2)»

Построение корреляционного графика



Формат линии тренда

Параметры линии тренда

Цвет линии

Тип линии

Тень

Свечение и сглаживание

Параметры линии тренда

Построение линии тренда (аппроксимация и сглаживание)

- Экспоненциальная
- Линейная
- Логарифмическая
- Полиномиальная Степень: 2
- Степенная
- Линейная фильтрация Точки: 2

Название аппроксимирующей (сглаженной) кривой

- автоматическое: Линейная (Ряд 1)
- другое:

Прогноз

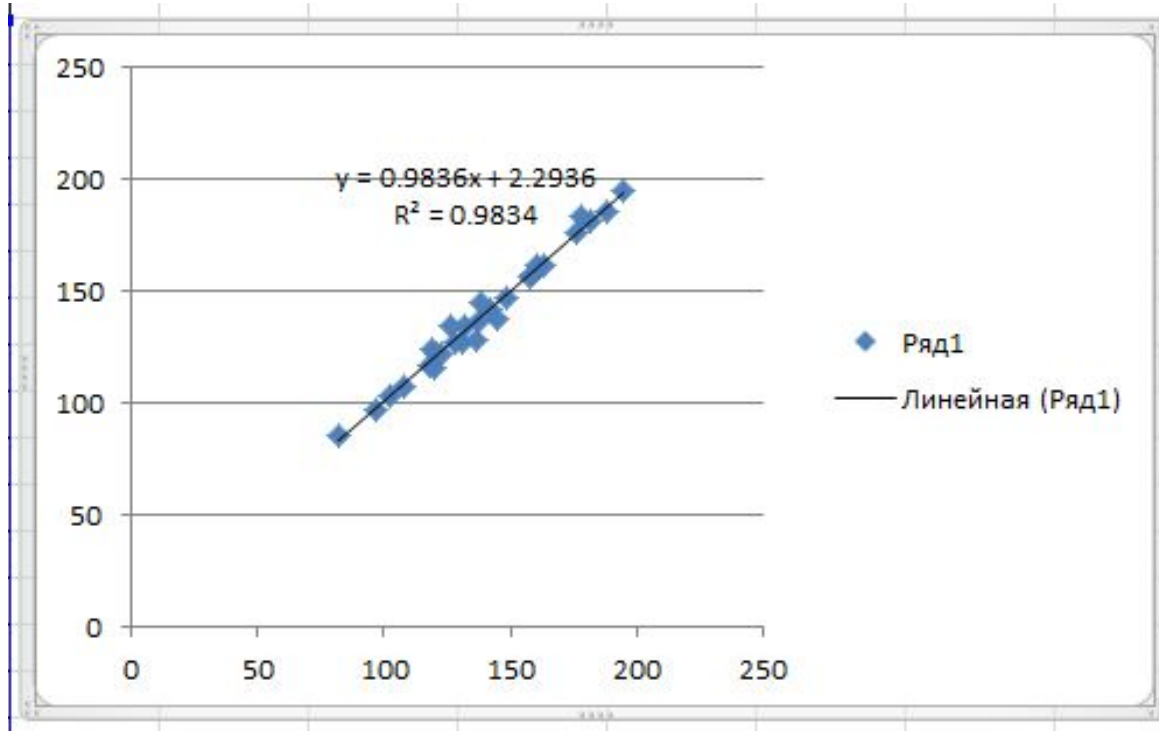
вперед на: 0.0 периодов

назад на: 0.0 периодов

- пересечение кривой с осью Y в точке: 0.0
- показывать уравнение на диаграмме
- поместить на диаграмму величину достоверности аппроксимации (R^2)

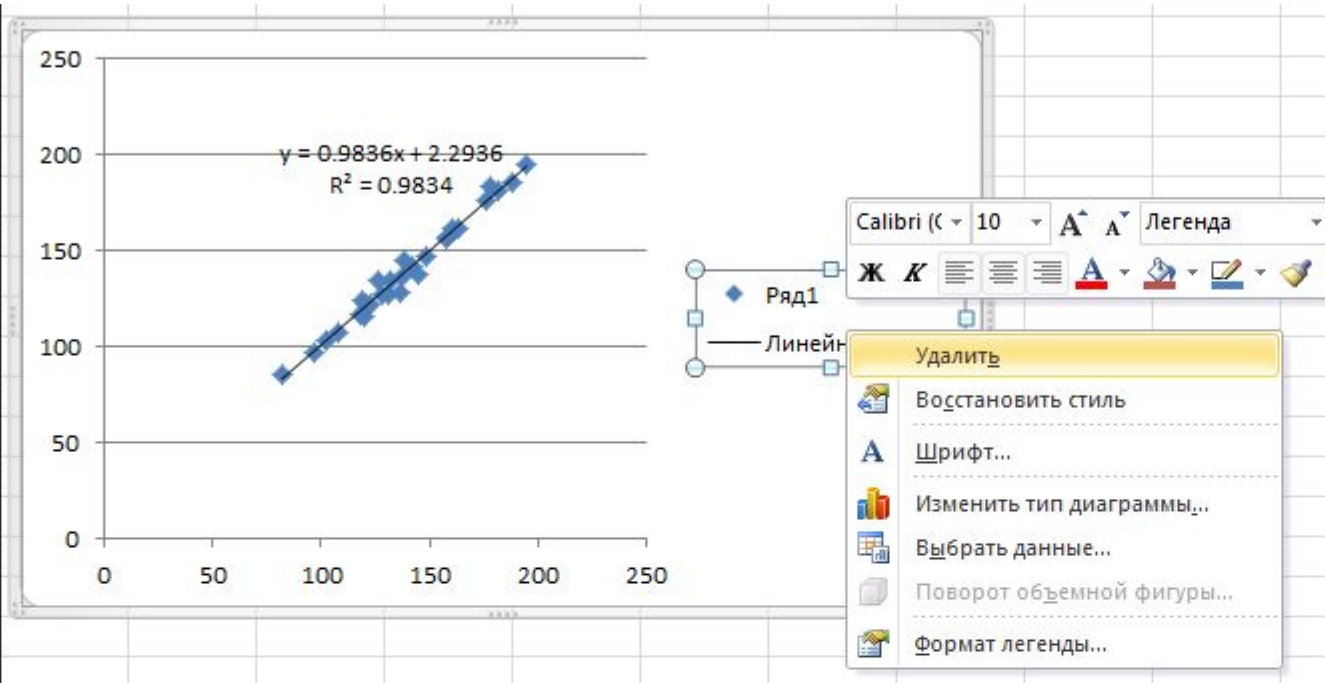
Заккрыть

Редактирование корреляционного графика



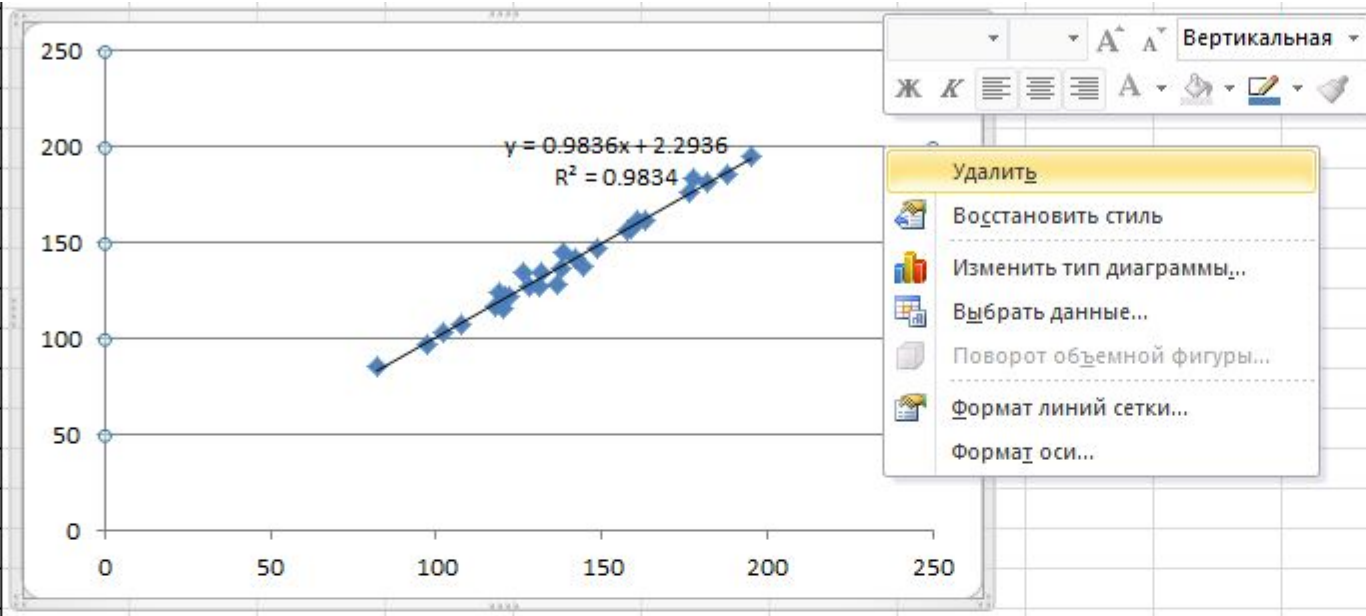
Редактируем представление графика: начало и конец отсчёта графика, отображение промежуточных делений и их направление, фон заливки и прочее.

Редактирование корреляционного графика



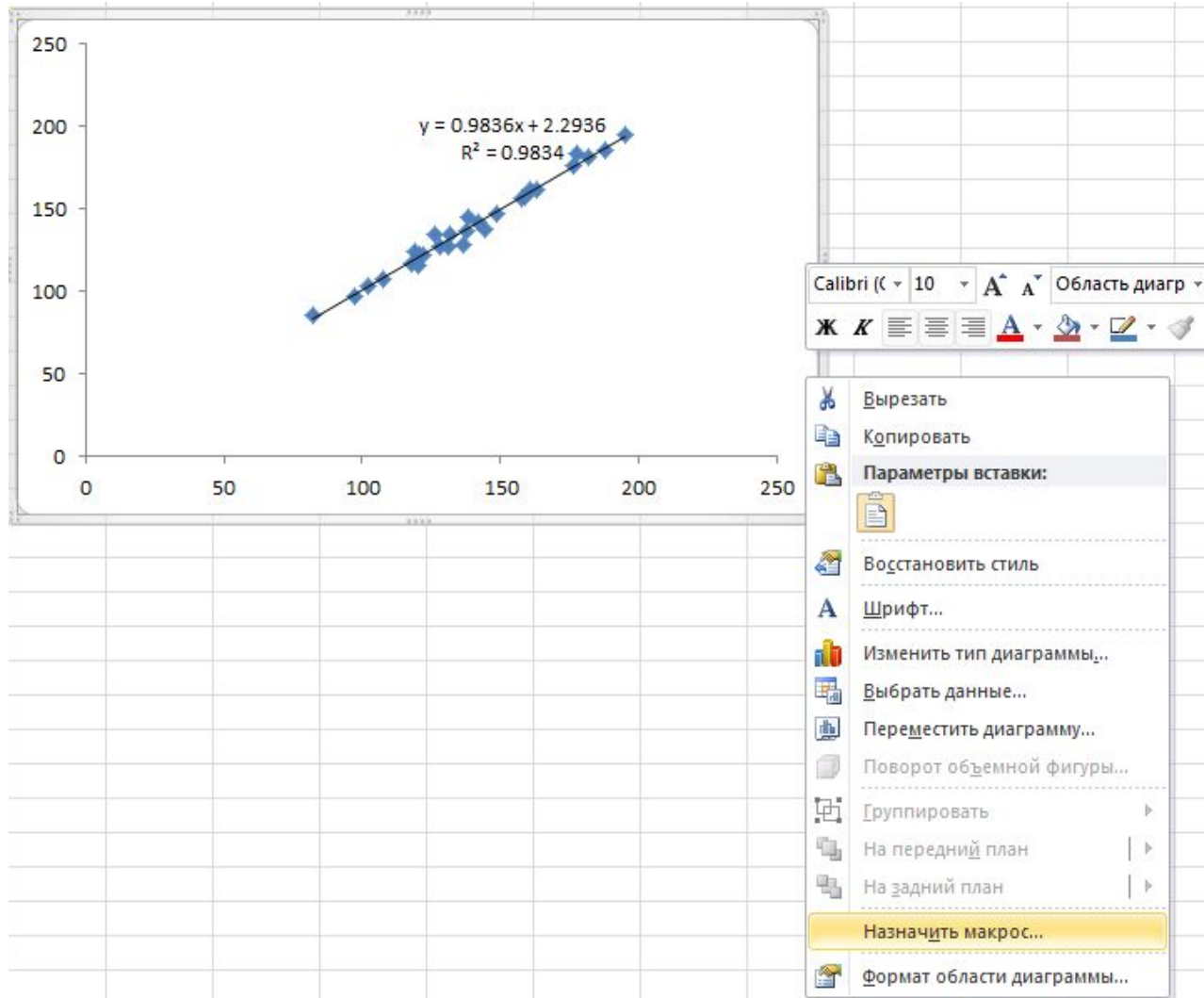
Удаляем
«легенду»

Редактирование корреляционного графика



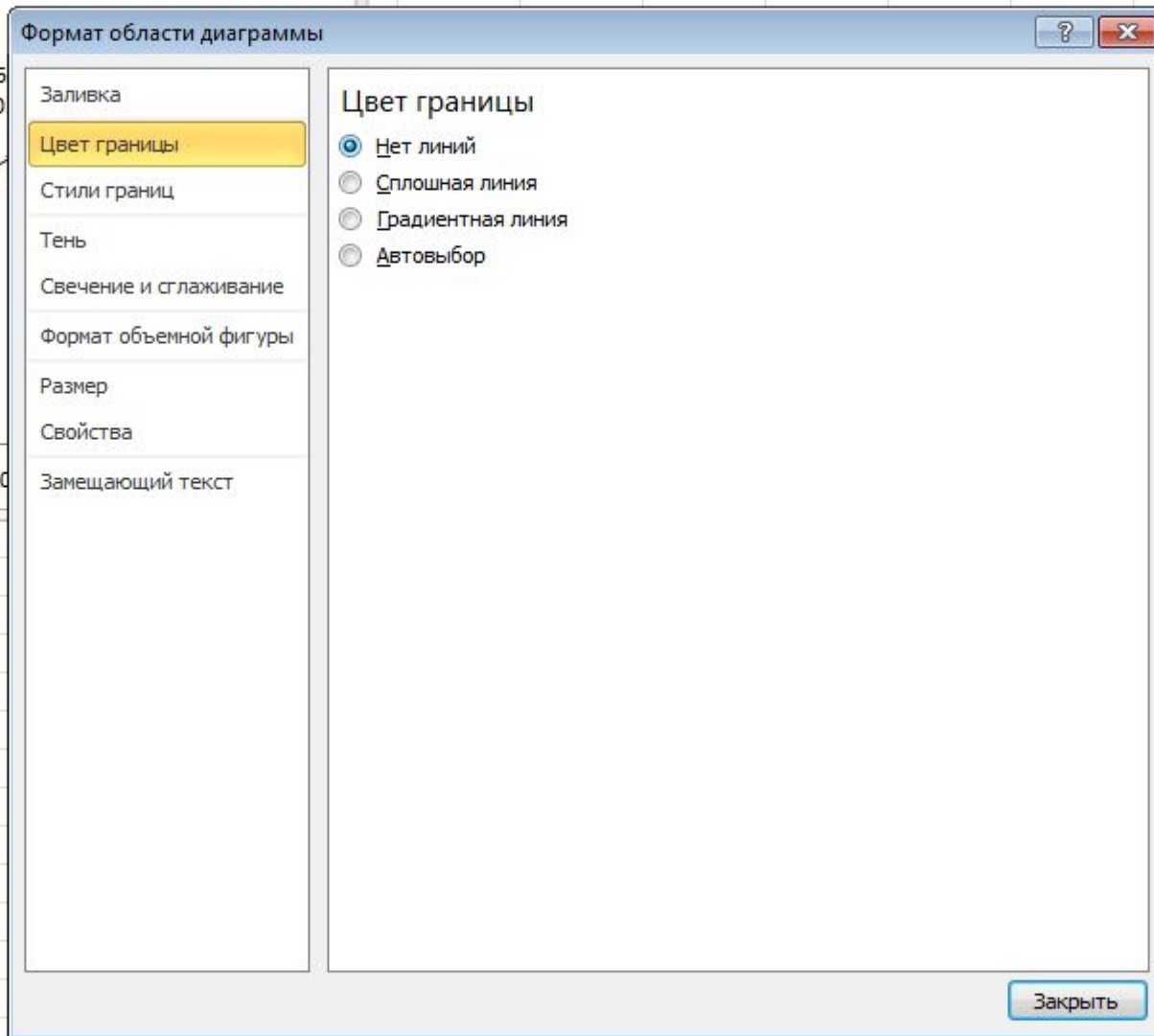
Удаляем
линии
сетки,
нажав
правой
кнопкой
на любой
из линий
сетки

Редактирование корреляционного графика



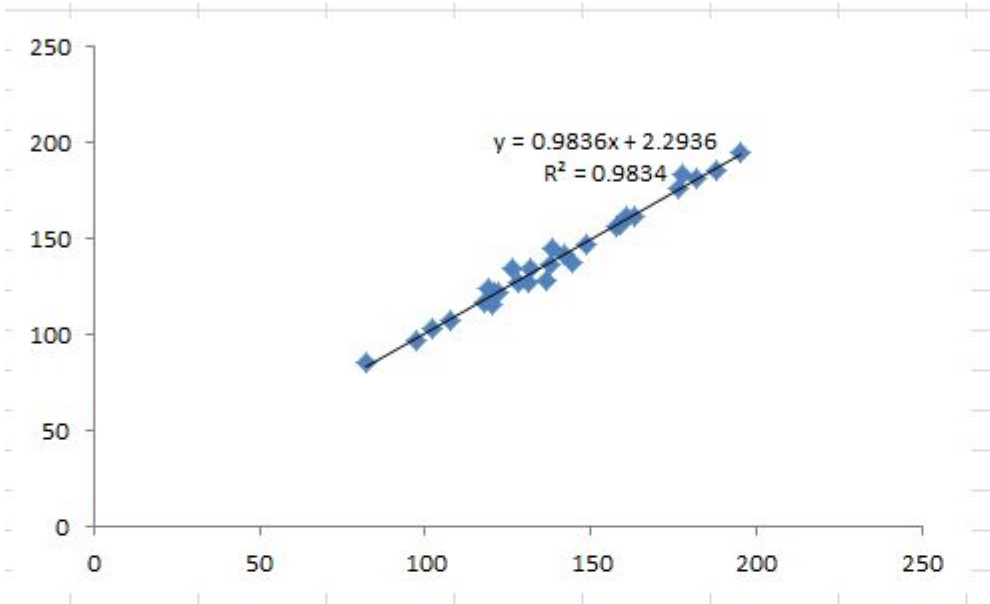
Правой кнопкой нажимаем на пустом месте графика, в появившейся всплывающемся меню выбираем «формат области диаграммы»

Редактирование корреляционного графика



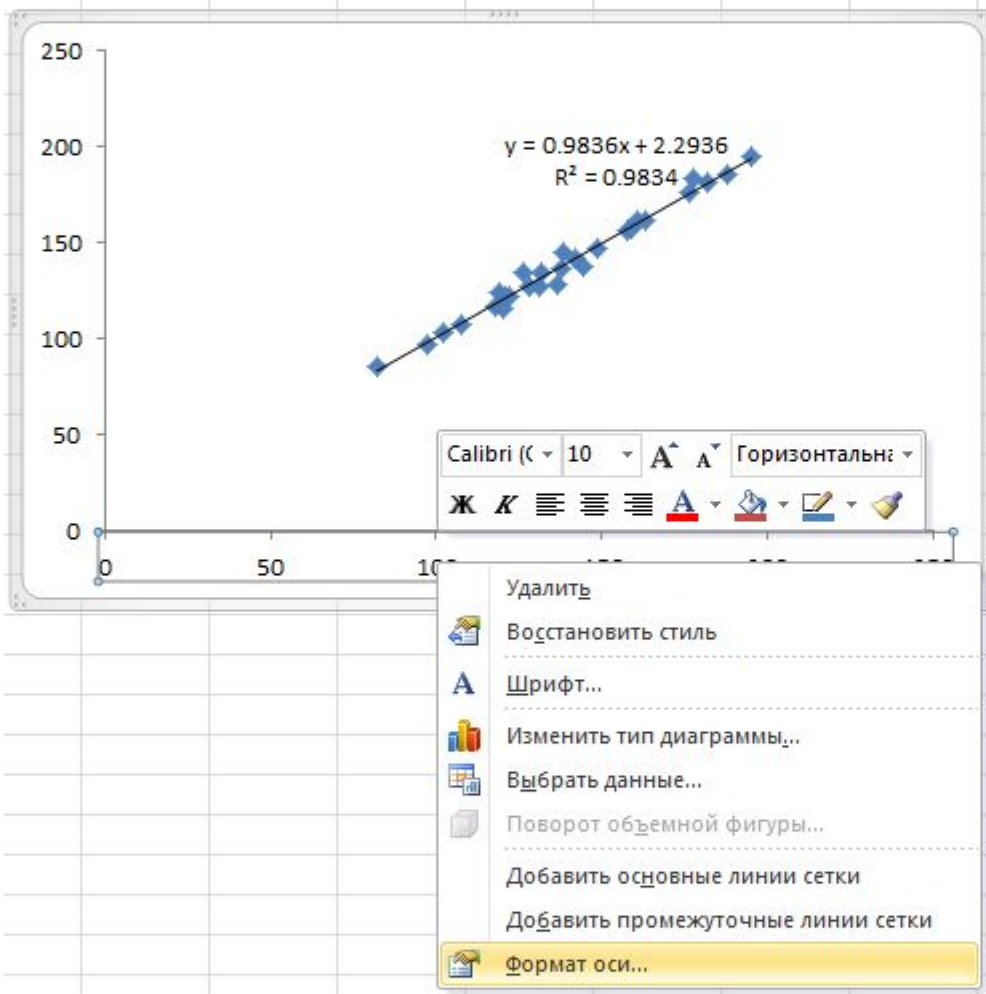
Во вкладке «Цвет границы» устанавливаем значение «Нет линий».

Редактирование корреляционного графика

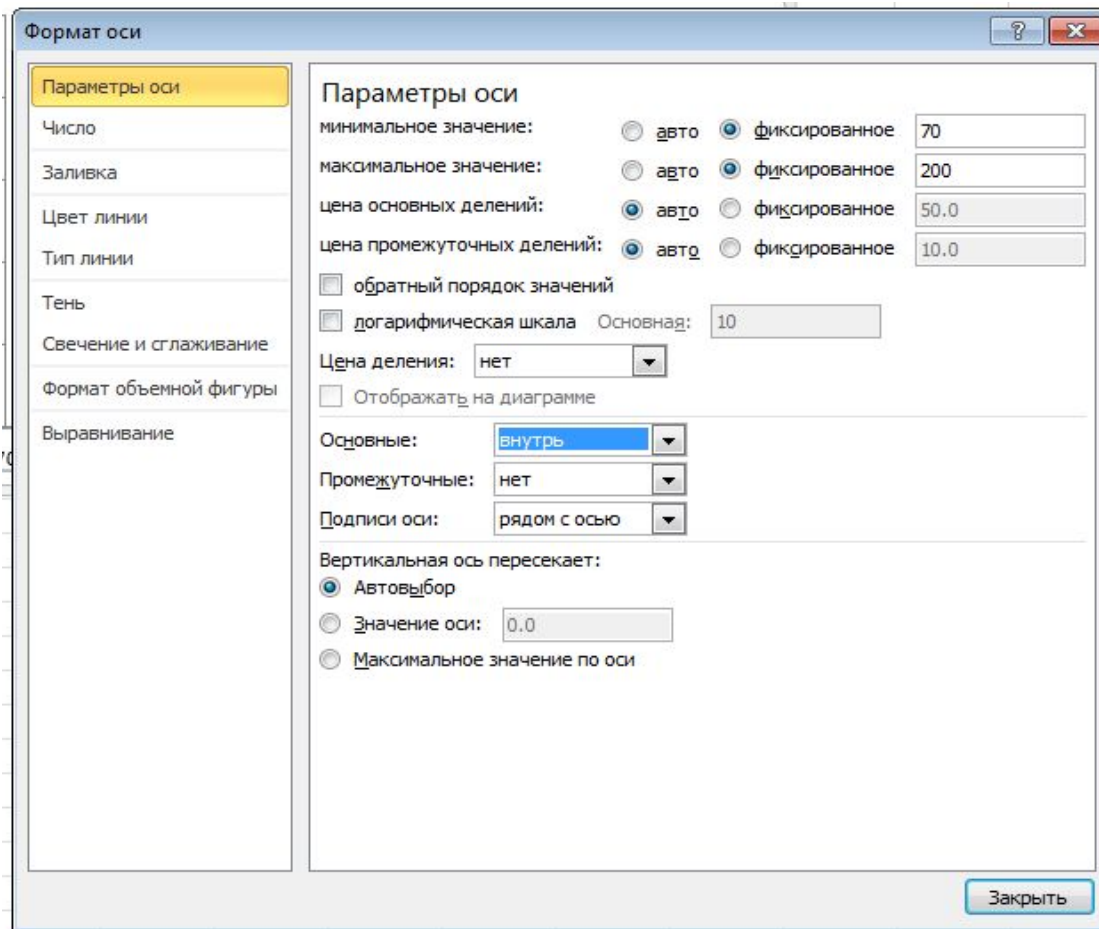


Редактируем оси.
Нажимаем на
числовое
значение на
шкале оси
правой кнопкой.
Во всплывающей
вкладке
выбираем
«Формат оси»

Редактирование корреляционного графика

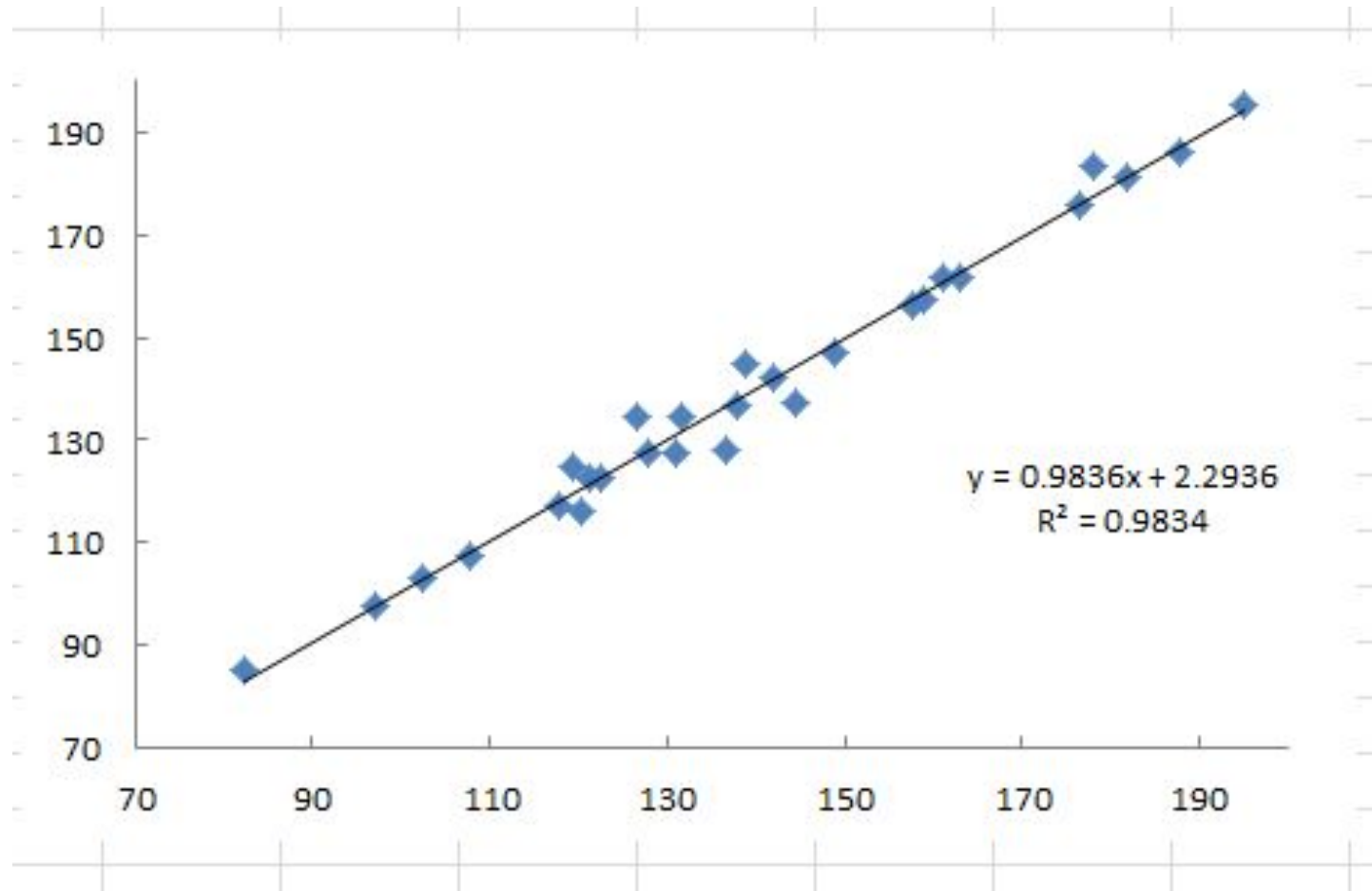


Редактирование корреляционного графика



В меню «Параметры оси» устанавливаем интервалы оси, так чтобы все точки остались в отображаемой зоне. Меняем отображение делений оси на значение «внутри». Нажимаем «Заккрыть». Так же настраиваем ось «у»

Редактирование корреляционного графика



Редактирование корреляционного графика

Подпишем оси. При активном режиме графика, на главной панели отображается меню «Работа с диаграммами», которое имеет 3 подменю. Выбираем подменю «Макет», вкладку «Название осей», во всплывающем меню выбираем нужные значения

Книга1 - Microsoft Excel

Работа с диаграммами

Конструктор Макет Формат

Область построения

Формат выделенного Восстановить стиль Текущий фрагмент

Рисунок Фигуры Надпись

Название диаграммы Названия осей Легенда Подписи данных Таблица данных

Оси Сетка

Область построения Стенка Основание Поворот Линия тренда

Название основной горизонтальной оси

Название основной вертикальной оси

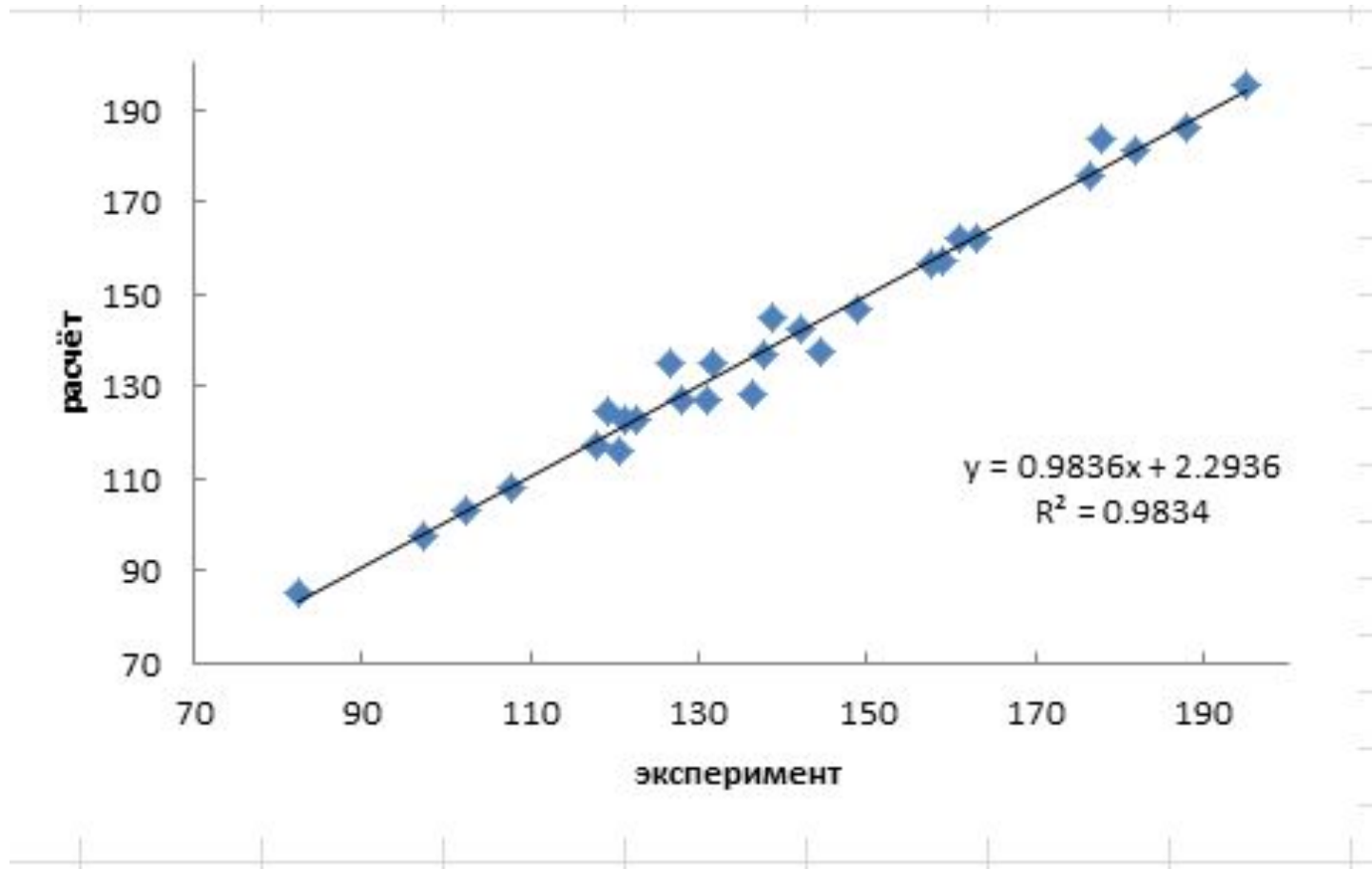
Нет
Не показывать название оси

Название под осью
Размещение названия под горизонтальной осью с соответствующим изменением размеров диаграммы

Дополнительные параметры названия основной горизонтальной оси...

	A	B	C	D	E	F	G	H
			Predicted b					
	Спирты	Вр (°C) a	Вр (°C)					
1								
2	1	82.3	85.4					
3								

Редактирование корреляционного графика



Построение гистограммы распределения частот ошибок

2	Смещение	Вр (°C)	Вр (°C)	Разность
3	1	82.3	85.4	=B3-C3
4	2	97.2	97.5	
5	3	117.7	117.1	
6	4	107.8	107.7	
7	5	102.4	103	
8	6	119.3	124.6	
9	7	131.1	127.3	
10	8	128	127.3	
11	9	137.9	136.7	
12	10	120.4	116	
13	11	121.1	122.6	
14	12	122.4	122.6	
15	13	136.5	128.1	
16	14	131.6	134.8	
17	15	126.5	134.8	
18	16	144.5	137.5	
19	17	149	146.9	
20	18	157.6	156.3	
21	19	138.7	145	
22	20	159	157.1	
23	21	142	142.2	
24	22	176.4	175.9	
25	23	161	161.8	
26	24	163	161.8	
27	25	188	186.1	
28	26	195.1	195.5	
29	27	178	183.3	
30	28	182	181.4	

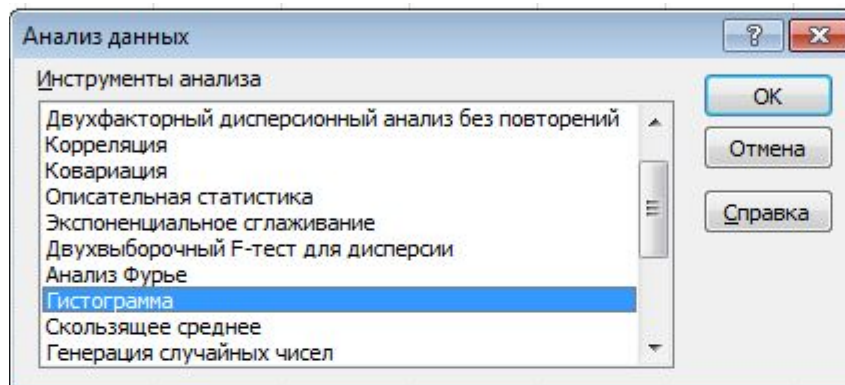
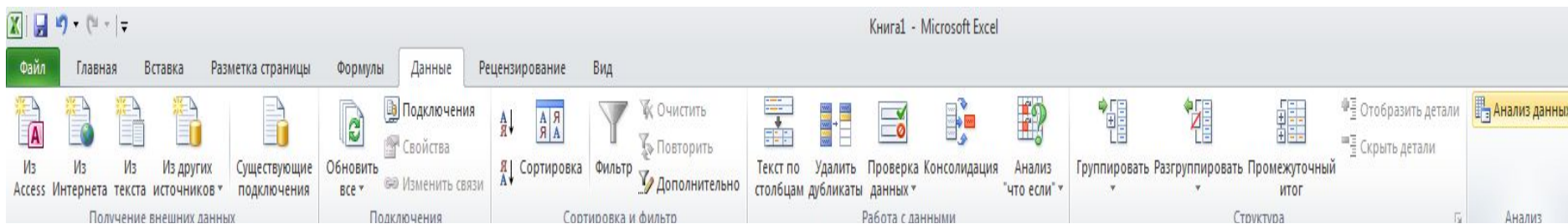
Находим разность между экспериментальными и расчетными данными, отнимая от экспериментальных значений рассчитанные.

Построение гистограммы распределения частот ошибок

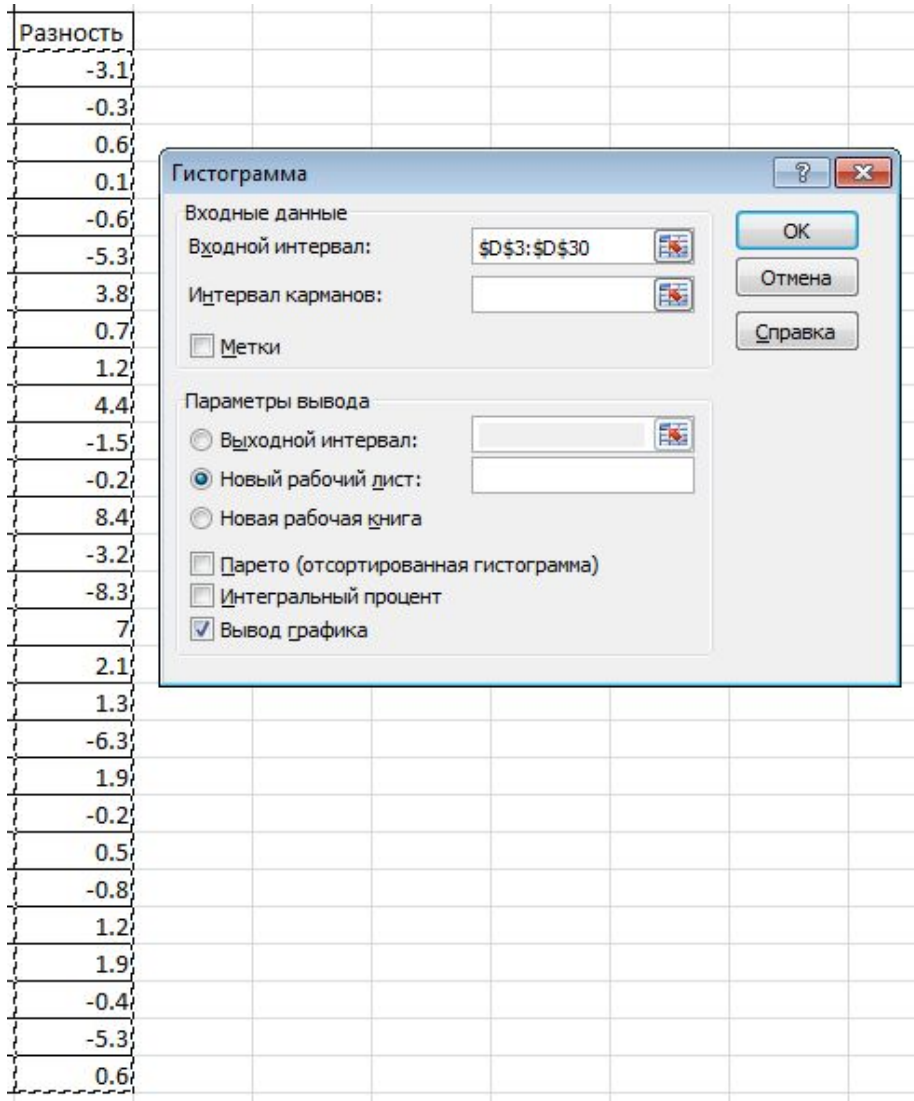
2	Спирты	Вр (°C) а	Вр (°C)	Разность
3	1	82.3	85.4	-3.1
4	2	97.2	97.5	-0.3
5	3	117.7	117.1	0.6
6	4	107.8	107.7	0.1
7	5	102.4	103	-0.6
8	6	119.3	124.6	-5.3
9	7	131.1	127.3	3.8
10	8	128	127.3	0.7
11	9	137.9	136.7	1.2
12	10	120.4	116	4.4
13	11	121.1	122.6	-1.5
14	12	122.4	122.6	-0.2
15	13	136.5	128.1	8.4
16	14	131.6	134.8	-3.2
17	15	126.5	134.8	-8.3
18	16	144.5	137.5	7
19	17	149	146.9	2.1
20	18	157.6	156.3	1.3
21	19	138.7	145	-6.3
22	20	159	157.1	1.9
23	21	142	142.2	-0.2
24	22	176.4	175.9	0.5
25	23	161	161.8	-0.8
26	24	163	161.8	1.2
27	25	188	186.1	1.9
28	26	195.1	195.5	-0.4
29	27	178	183.3	-5.3
30	28	182	181.4	0.6

Построение гистограммы распределения частот ошибок

Строим по полученным значениям гистограмму частоты распределения ошибок. Вкладка «Данные → Анализ данных → Гистограмма».

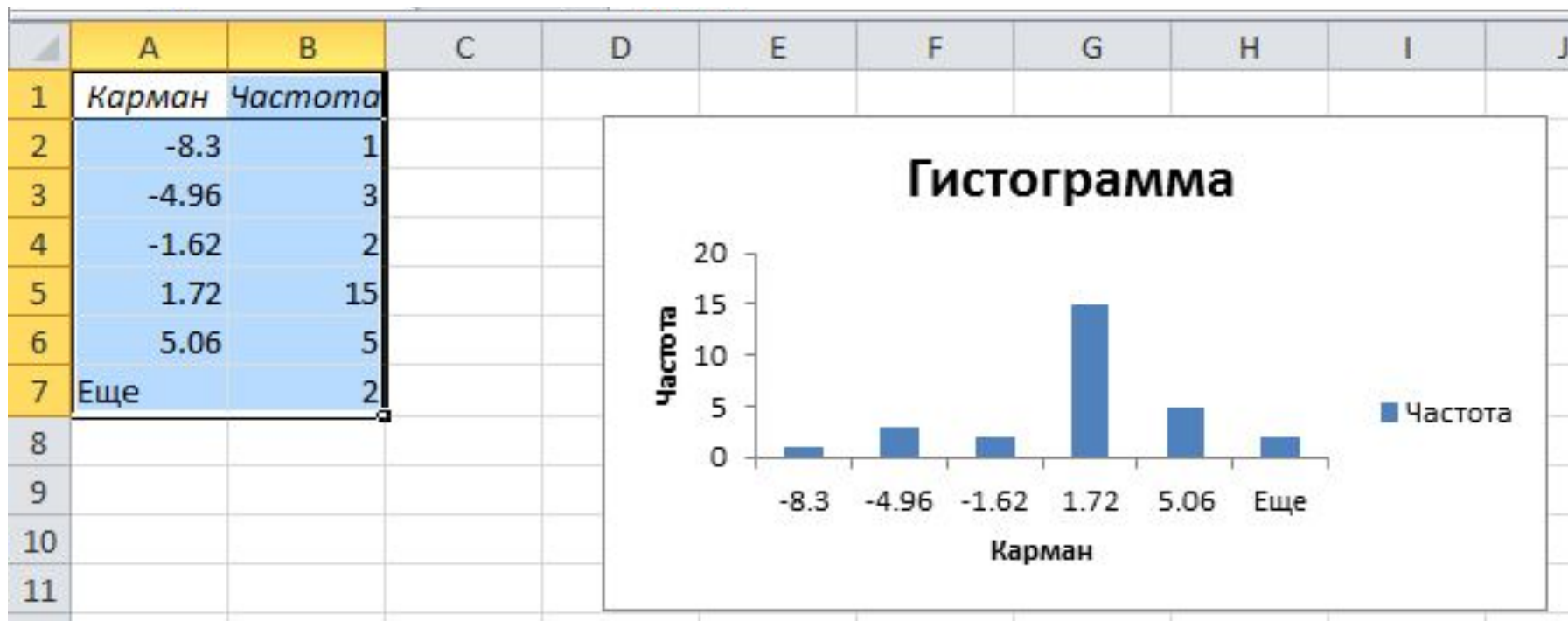


Построение гистограммы распределения частот ошибок



В появившемся окне, в поле «Входной интервал» вводим весь интервал разности и устанавливаем галочку напротив «Вывод графика».

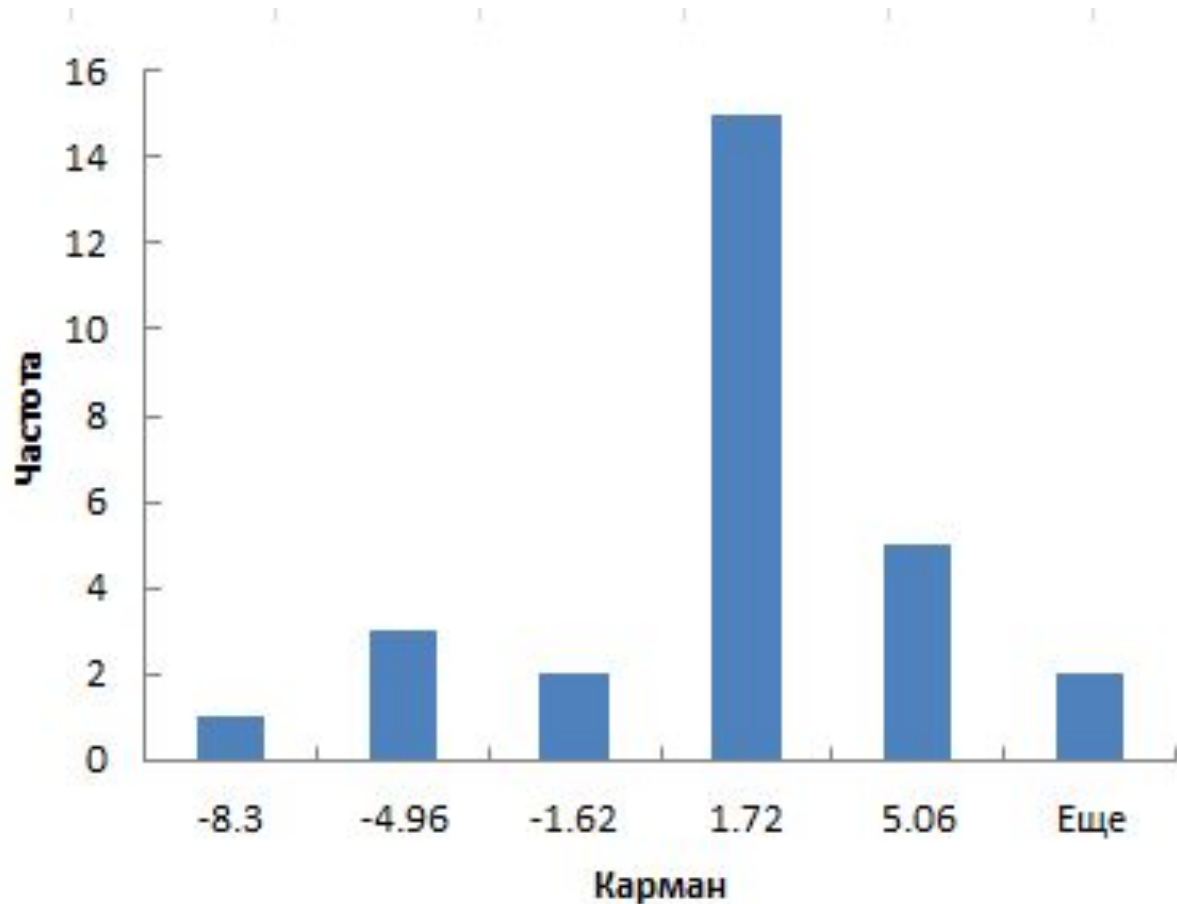
Редактирование гистограммы распределения частот ошибок



Редактирование гистограммы распределения частот ошибок

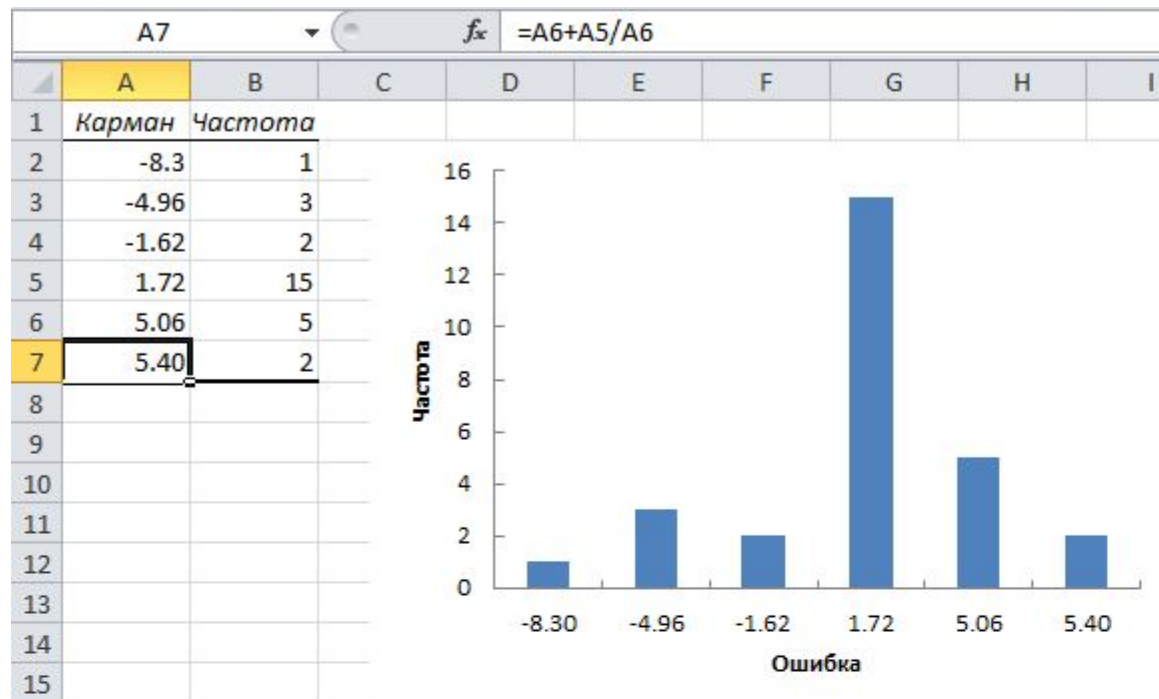
- Редактируем гистограмму так же, как и корреляционный график:
- Удаляем «легенду»
- Удаляем название диаграммы «Гистограмма»
- Удаляем цвет границы
- Выставляем деления осей «Внутрь»

Редактирование гистограммы распределения частот ошибок



Меняем
название
оси «х» со
слова
«Карман»
на
«Ошибка»

Редактирование гистограммы распределения частот ошибок



В таблице
меняем
выражение
«ещё» на
уравнение вида:
(последнее
значение +
предпоследнее
значение) / на
последнее
значение
столбца
«Ошибка»

Нахождение статистических параметров

- статистические параметры результатов исследования находятся, используя команду «регрессия» во вкладке «данные → анализ данных». В входном интервале Y указывается интервал значений экспериментальных данных, а в входном интервале X берётся интервал рассчитанных значений.

Нахождение статистических параметров

1	82.3	85.4	-3.1
2	97.2	97.5	-0.3
3	117.7	117.1	0.6
4	107.8	107.7	0.1
5	102.4	103	-0.6
6	119.3	124.6	-5.3
7	131.1	127.3	3.8
8	128	127.3	0.7
9	137.9	136.7	1.2
10	120.4	116	4.4
11	121.1	122.6	-1.5
12	122.4	122.6	-0.2
13	136.5	128.1	8.4
14	131.6	134.8	-3.2
15	126.5	134.8	-8.3
16	144.5	137.5	7
17	149	146.9	2.1
18	157.6	156.3	1.3
19	138.7	145	-6.3
20	159	157.1	1.9
21	142	142.2	-0.2
22	176.4	175.9	0.5
23	161	161.8	-0.8
24	163	161.8	1.2
25	188	186.1	1.9
26	195.1	195.5	-0.4
27	178	183.3	-5.3
28	182	181.4	0.6

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK Отмена Справка

Статистические параметры исследования

Результатом является «таблица» в которой представлены следующие параметры «Множественный R» - коэффициент корреляции, «R-квадрат» - квадрат коэффициента корреляции, «Нормированный R-квадрат» - скорректированный индекс множественной детерминации, содержит поправку на число степеней свободы. «Стандартная ошибка» - погрешность измерения, оценка отклонения величины измеренного значения величины от её истинного значения. «Наблюдения» - число измерений

Статистические параметры исследования

1	ВЫВОД ИТОГОВ									
2										
3	<i>Регрессионная статистика</i>									
4	Множественный R	0.991689137								
5	R-квадрат	0.983447344								
6	Нормированный R-квадрат	0.982810703								
7	Стандартная ошибка	3.71700009								
8	Наблюдения	28								
9										
10	<i>Дисперсионный анализ</i>									
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>				
12	Регрессия	1	21342.33	21342.33	1544.745	1.09E-24				
13	Остаток	26	359.2183	13.81609						
14	Итого	27	21701.55							
15										
16		<i>Коэффициенты</i>	<i>стандартная ошибка</i>	<i>t-Значения</i>	<i>значимость</i>	<i>нижние 95%</i>	<i>верхние 95%</i>	<i>нижние 95%</i>	<i>верхние 95%</i>	
17	Y-пересечение	0.021974119	3.62698	0.006059	0.995212	-7.43339	7.477338	-7.43339	7.477338	
18	Переменная X 1	0.999893962	0.02544	39.30324	1.09E-24	0.9476	1.052188	0.9476	1.052188	
19										