

Лабораторные работы № 1-6 «МЕТОДЫ ПРОГНОЗИРОВАНИЯ ПОЖАРОВ»

Одно из важных направлений деятельности МЧС — прогнозирование числа пожаров на территории субъекта РФ. Рассмотрим прогнозирование на один месяц вперед на примере данных по Ростовской области за 2020 и 2021 года. Разберём возможные подходы для выполнения этого прогноза.

1. Метод среднего

Суть метода - прогнозируемое значение на будущий месяц находят как *среднее значение всех предшествующих величин* за прошлые месяцы года.

В январе 2020 г. в Ростовской области произошел 250 пожар.

В качестве прогнозного значения на февраль 2020 г. принимаем эту же величину - 250, т.к. для вычисления среднего мало данных.

Прогноз на март находим как среднее значение реальных показателей за два месяца (январь и февраль).

В апреле прогнозное значение определяем как среднее данных за три месяца (январь, февраль и март) и т.д.

То есть по мере появления нового месяца года прогноз будет выполнен на основе всех предшествующих значений за прошедшие месяцы года.

Выполните такой прогноз в программе Microsoft Excel:

- Скопируйте в ячейку A2 исходные данные из таблицы о пожарах в 2020 году (два столбца - месяц, число пожаров).

месяц 2020г.	число пожаров	линия прогноза
январь	250	
февраль	139	
март	151	
апрель	152	
май	210	
июнь	126	
июль	100	
август	119	
сентябрь	134	
октябрь	157	
ноябрь	162	
декабрь	168	

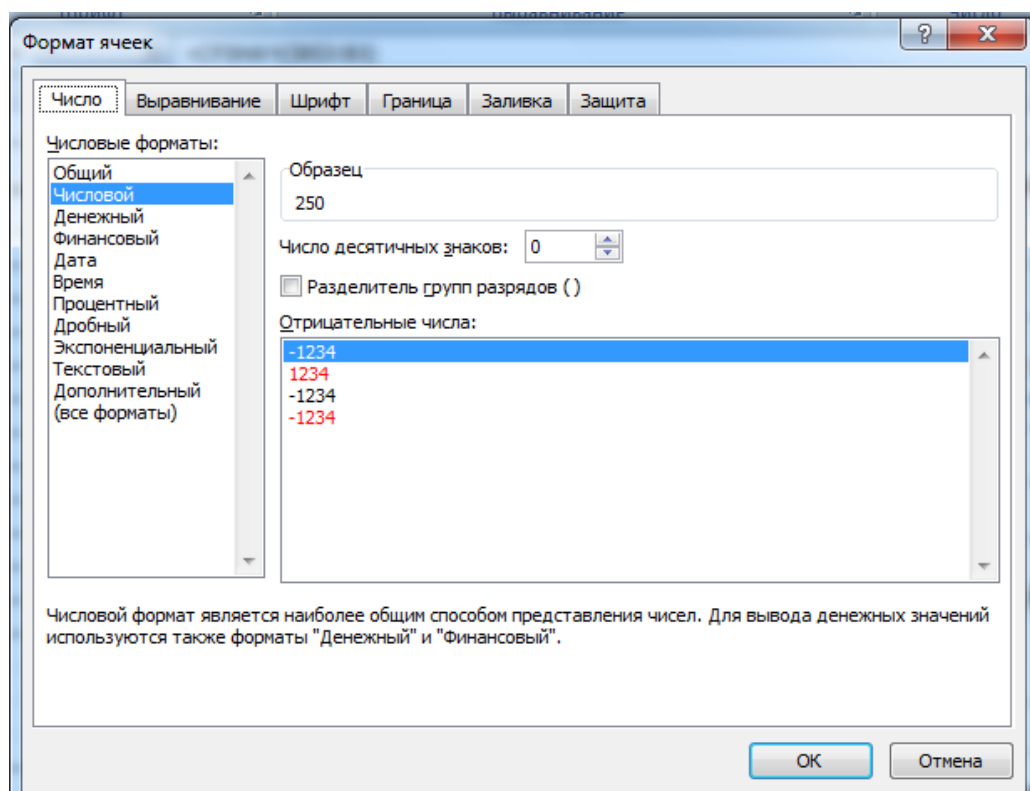
месяц 2021г.	число пожаров	линия прогноза
январь	125	
февраль	138	
март	120	
апрель	124	
май	115	
июнь	110	
июль	105	
август	90	
сентябрь	157	
октябрь	223	
ноябрь	184	
декабрь	173	

- В столбец «линия прогноза» в ячейку C5 для вычисления среднего значения примените расчетную формулу, используя функцию «СРЗНАЧ» (см. рисунок).
- Скопируйте формулу вниз на весь диапазон ячеек до C14 (потянув за правый нижний угол ячейки).

	A	B	C	D
1				
2	месяц 2020г.	число пожаров	линия прогноза	
3	январь	250		
4	февраль	139	250	
5	март	151	195	
6	апрель	152	180	
7	май	210	173	
8	июнь	126	180	
9	июль	100	171	
10	август	119	161	
11	сентябрь	134	156	
12	октябрь	157	153	
13	ноябрь	162	154	
14	декабрь	168	155	
15				

линия прогноза	
250	
195	
180	
173	
180	
171	
161	
156	
153	
154	
155	

- Для округления результатов выделите столбец с прогнозными значениями, и нажав пр.кл. мыши примените команду - *Формат ячеек*.



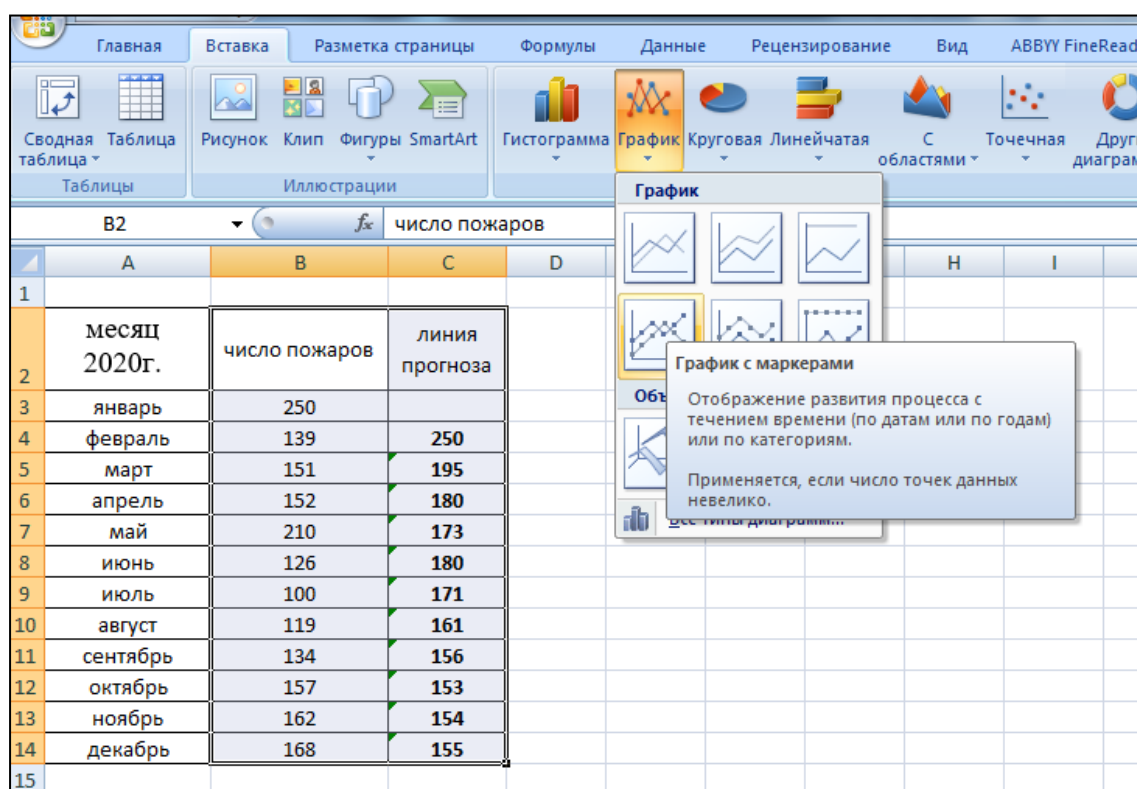
- Выполните аналогичные расчеты для данных за 2021 год.

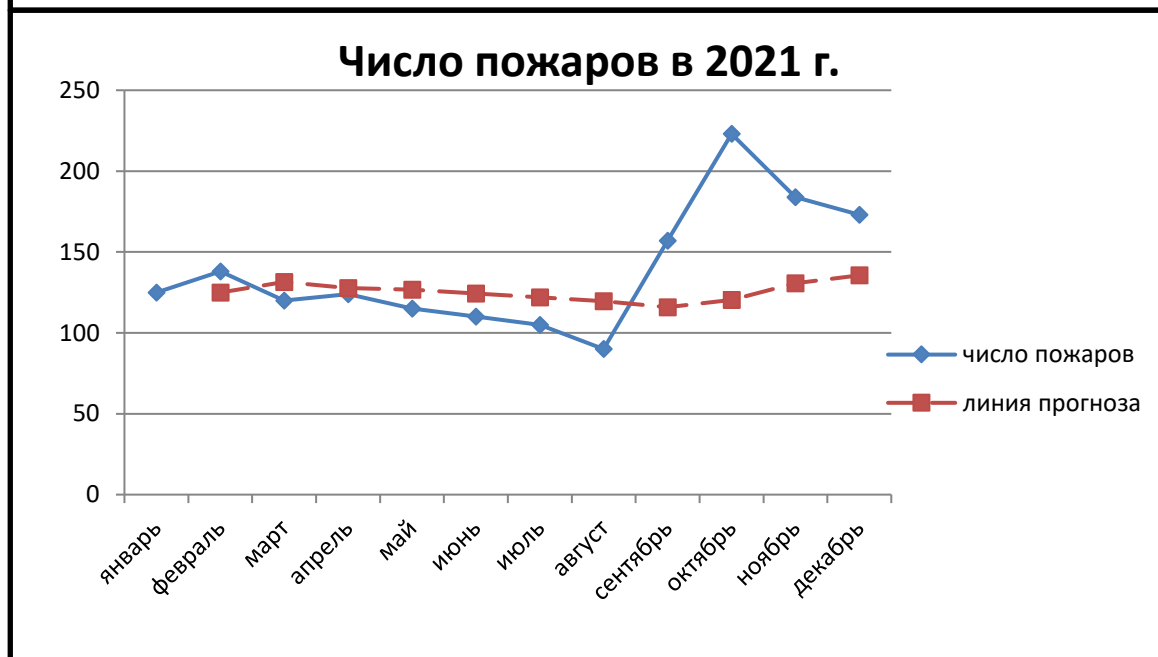
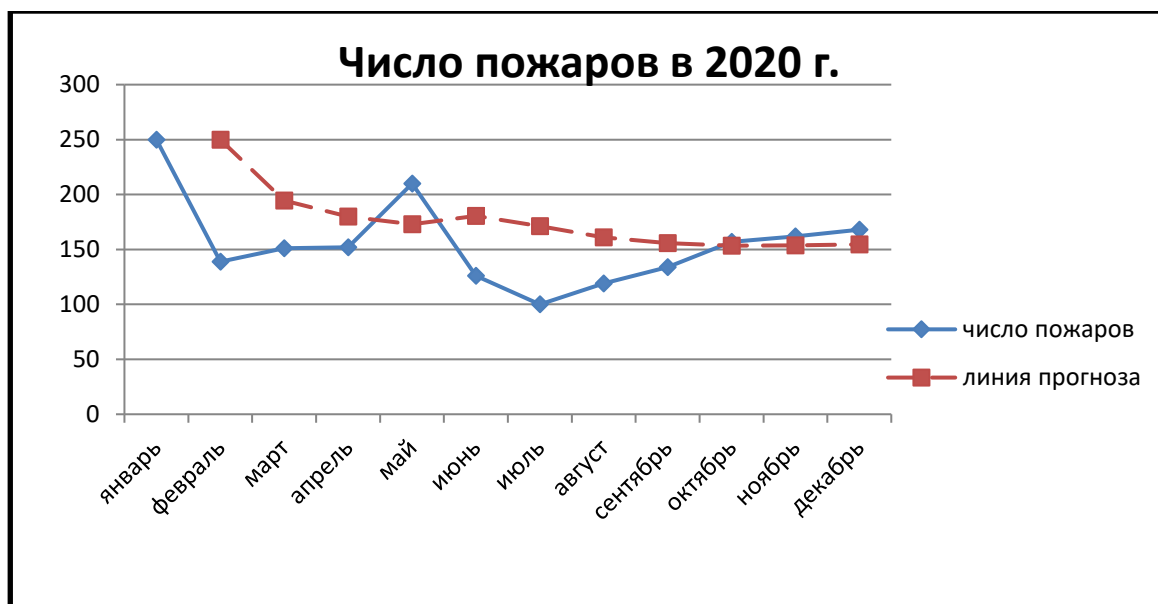
Должны получиться результаты:

месяц 2020г.	число пожаров	линия прогноза
январь	250	
февраль	139	250
март	151	195
апрель	152	180
май	210	173
июнь	126	180
июль	100	171
август	119	161
сентябрь	134	156
октябрь	157	153
ноябрь	162	154
декабрь	168	155

месяц 2021г.	число пожаров	линия прогноза
январь	125	
февраль	138	125
март	120	132
апрель	124	128
май	115	127
июнь	110	124
июль	105	122
август	90	120
сентябрь	157	116
октябрь	223	120
ноябрь	184	131
декабрь	173	136

- Постройте диаграммы для сравнения реальных данных за 2020 и 2021 года с прогнозом по методу среднего (Вставка - Диаграммы - График с маркерами).





Недостаток данного метода заключается в том, что в конце года прогнозные значения будут практически одинаковы в виду усреднения по всем предшествующим месяцам. Поэтому прогнозная линия в конце года часто будет параллельна оси абсцисс, что не совсем согласуется с практикой увеличения числа пожаров в конце года.

2. Метод скользящего среднего

Суть метода - прогнозируемое значение на будущий месяц находят как *среднее значение предшествующих величин за несколько прошлых месяцев года.*

На настоящий момент установилось представление, что для вычисления прогнозного значения достаточно использовать реальные значения **за три** предшествующих месяца:

$$\bar{y}_i = \frac{y_{i-1} + y_{i-2} + y_{i-3}}{3}.$$

Выполните такой прогноз в программе Microsoft Excel:

- Скопируйте в ячейку A4 исходные данные из таблицы о пожарах в 2020 году (два столбца - месяц, число пожаров).

- В ячейке C8 рассчитайте прогноз числа пожаров за апрель (используя функцию «СРЗНАЧ») на основе трёх предыдущих месяцев (см. рисунок).

	A	B	C	D	E
1					
2					
3					
4	месяц 2020г.	число пожаров	линия прогноза		
5	январь	250			
6	февраль	139			
7	март	151			
8	апрель	1	180		
9	май	210	147		
10	июнь	126	171		
11	июль	100	163		
12	август	119	145		
13	сентябрь	134	115		
14	октябрь	157	118		
15	ноябрь	162	137		
16	декабрь	168	151		
17					

- Выполните аналогичные расчеты для данных за 2021 год.

Должны получиться результаты:

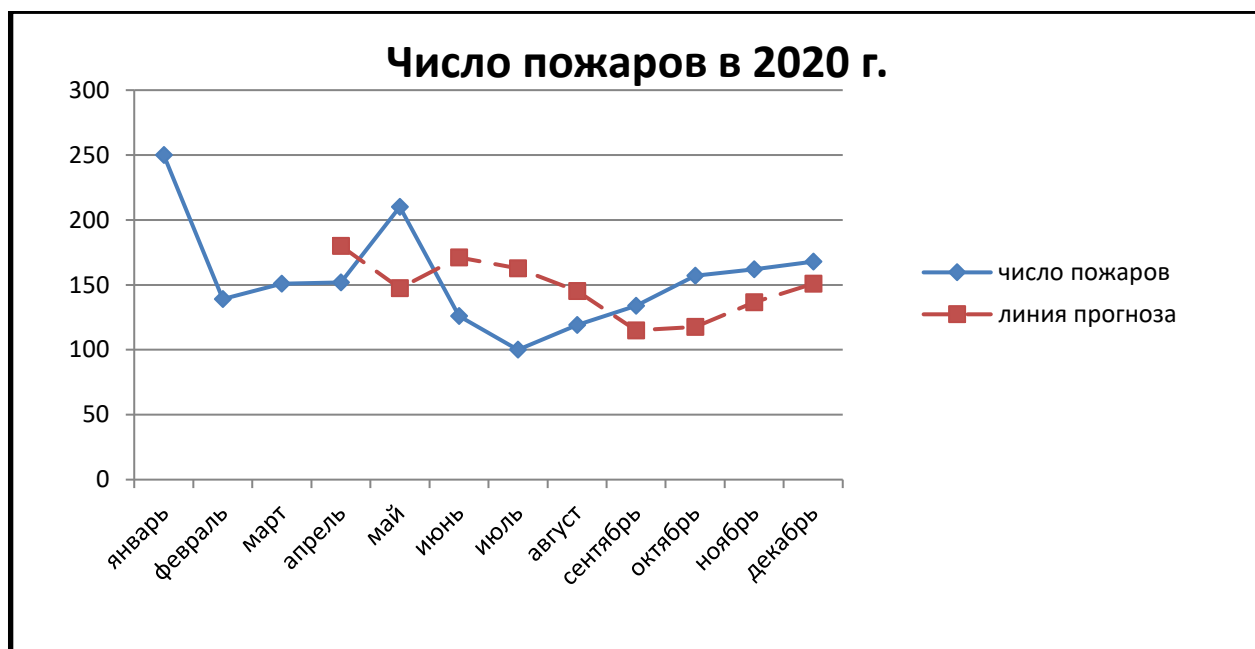
месяц 2020г.	число пожаров	Прогноз
январь	250	
февраль	139	
март	151	
апрель	152	180
май	210	147
июнь	126	171
июль	100	163
август	119	145
сентябрь	134	115

месяц 2021г.	число пожаров	Прогноз
январь	125	
февраль	138	
март	120	
апрель	124	128
май	115	127
июнь	110	120
июль	105	116
август	90	110
сентябрь	157	102

октябрь	157	118
ноябрь	162	137
декабрь	168	151

октябрь	223	117
ноябрь	184	157
декабрь	173	188

- Постройте соответствующие диаграммы, согласно рисункам ниже.



3. Метод квазислучайных чисел

Суть метода - основан на идеи *о случайности числа пожаров* за данный месяц.

В то же время на практике случайные колебания могут происходить в некотором интервале — между минимальной и максимальной значениями величины.

В Microsoft Excel функция “СЛЧИСЛ” генерирует случайные числа в диапазоне от 0 до 1.

Данный метод можно реализовать с помощью использования для прогнозного значения формулы:

$$= (\max - \min) * \text{СЛЧИСЛ} () + \min,$$

где max, min — максимальное и минимальное возможное значение соответственно.

- Скопируйте исходные данные о пожарах за 2020 год в ячейку A2.

- Добавьте ещё два столбца для расчёта максимального и минимального значений.

- В качестве *прогнозного значения на январь и февраль* 2020 г. можно взять реальное число пожаров, произошедших в январе 2020 г. (250).

- Рассчитайте максимальное значение для каждого месяца, начиная с марта (ячейка D5) и скопируйте формулу на весь диапазон вниз.

D5 fx =МАКС(B3:B4)					
	A	B	C	D	E
1					
2	месяц 2020г.	число пожаров	линия прогноза	max	min
3	январь	250	250		
4	февраль	139	250		
5	март	151		250	
6	апрель	152		151	
7	май	210		152	
8	июнь	126		210	
9	июль	100		210	
10	август	119		126	
11	сентябрь	134		119	
12	октябрь	157		134	
13	ноябрь	162		157	
14	декабрь	168		162	

- Рассчитайте минимальное значение для каждого месяца, начиная с марта (ячейка E5) и скопируйте формулу на весь диапазон вниз.

E5		fx =МИН(B3:B4)			
	A	B	C	D	E
1					
2	месяц 2020г.	число пожаров	линия прогноза	max	min
3	январь	250	250		
4	февраль	139	250		
5	март	151		150	139
6	апрель	152		151	139
7	май	210		152	151
8	июнь	126		210	152
9	июль	100		210	126
10	август	119		126	100
11	сентябрь	134		119	100
12	октябрь	157		134	119
13	ноябрь	162		157	134
14	декабрь	168		162	157

- Выполните прогнозные расчеты по методу квазислучайных чисел (введите соответствующую формулу в ячейку C5), см. рисунок ниже.

C5		$f_x = (D5-E5) * \text{СЛЧИС}() + E5$				
	A	B	C	D	E	F
1						
2	месяц 2020г.	число пожаров	линия прогноза	max	min	
3	январь	250	250			
4	февраль	139	250			
5	март	151	163	250	139	
6	апрель	152	141	151	139	
7	май	210	151	152	151	
8	июнь	126	177	210	152	
9	июль	100	169	210	126	
10	август	119	101	126	100	
11	сентябрь	134	109	119	100	
12	октябрь	157	128	134	119	
13	ноябрь	162	151	157	134	
14	декабрь	168	159	162	157	
15						

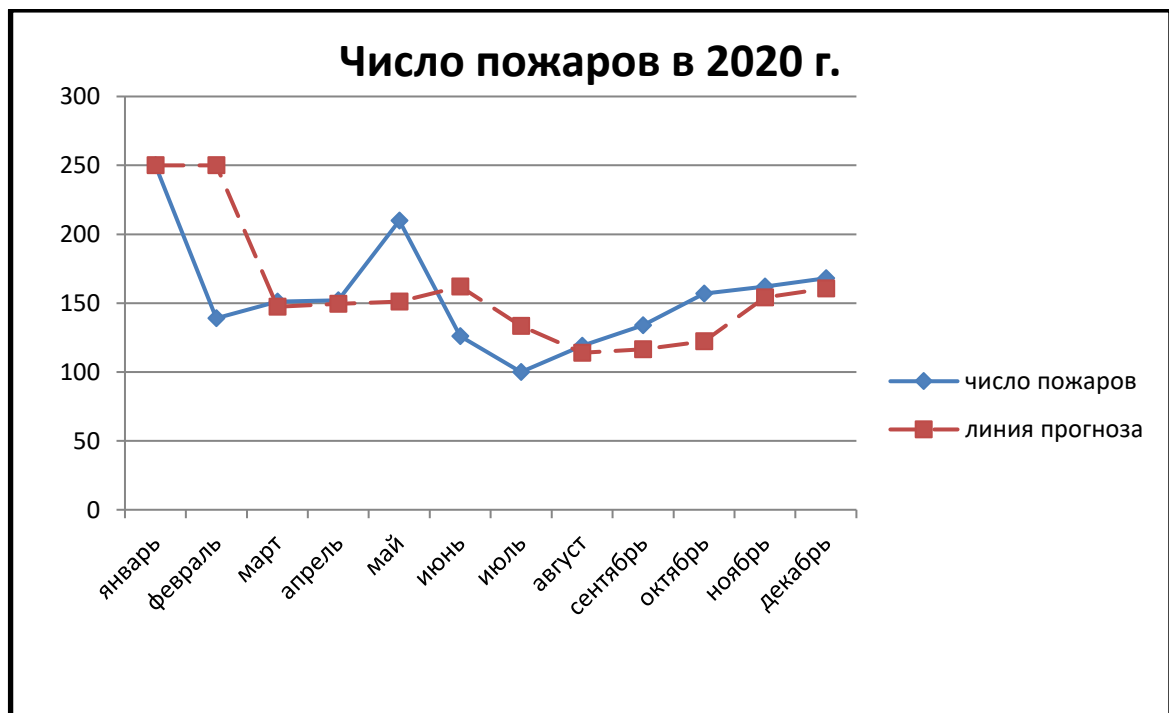
- Выполните аналогичные расчеты для данных за 2021 год.

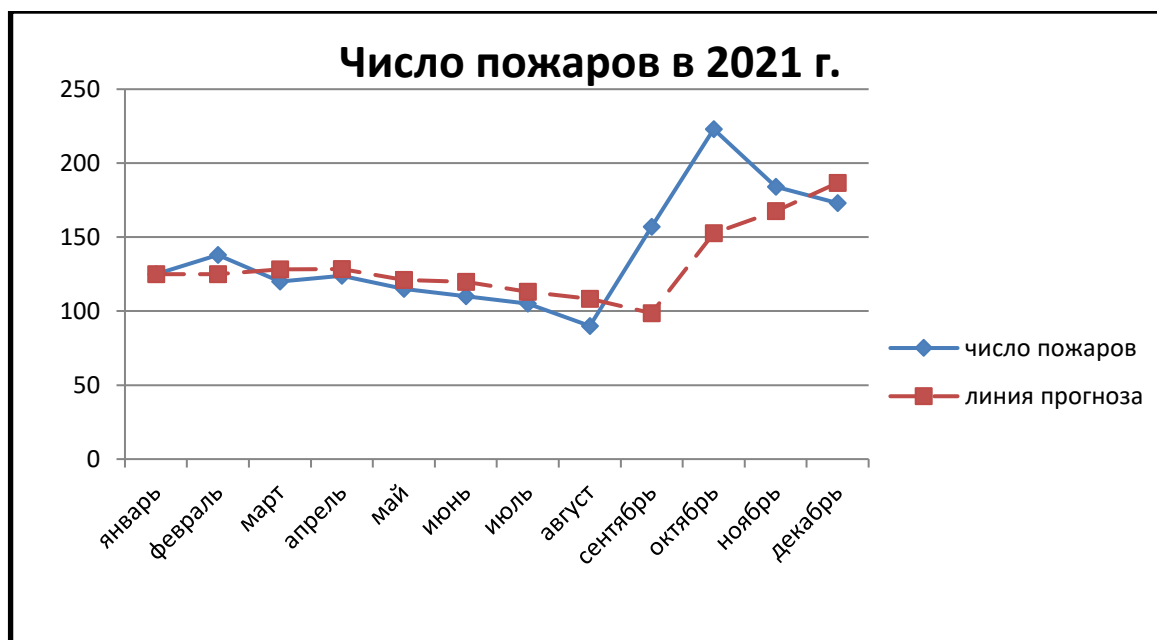
Должны получиться результаты:

месяц 2020г.	число пожаров	Прогноз
январь	250	250
февраль	139	250
март	151	193
апрель	152	146
май	210	151
июнь	126	174
июль	100	165
август	119	112
сентябрь	134	117
октябрь	157	123
ноябрь	162	156
декабрь	168	159

месяц 2021г.	число пожаров	Прогноз
январь	125	125
февраль	138	125
март	120	129
апрель	124	122
май	115	121
июнь	110	120
июль	105	114
август	90	105
сентябрь	157	93
октябрь	223	147
ноябрь	184	193
декабрь	173	218

- Постройте диаграммы.





4. Графический прогноз

В основе метода лежит построение с помощью программы Microsoft Excel графика числа пожаров и добавление к нему линий тренда.

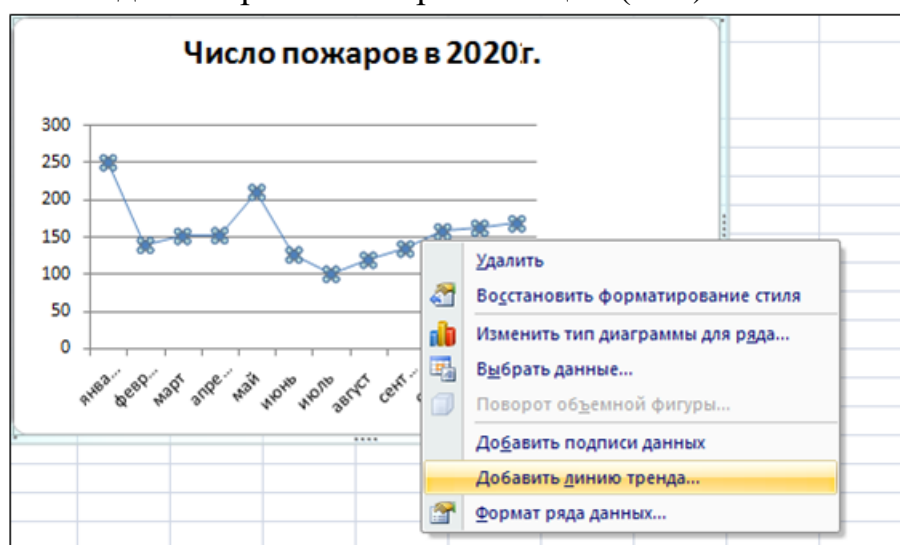
1. Скопируйте исходные данные о пожарах за 2020 год в ячейку A1 (два столбца - месяц, число пожаров).

2. Построить график по столбцу «число пожаров» и скопировать данный график (что бы всего на листе получилось 5 одинаковых графиков).

3. На каждом графике добавить одну из разновидностей линий тренда (линейную, логарифмическую, степенную, полиномиальную различных степеней от 2 до 6).

Для этого необходимо подвести указатель мыши к линии графика, нажать правую клавишу мыши и выбрать команду «добавить линию тренда».

4. В диалоговой панели «Формат линии тренда» отметить соответствующий тип линии тренда, отметить галочкой показ уравнения на диаграмме и вывод на диаграмму величины достоверности аппроксимации (R^2).



Формат линии тренда

Параметры линии тренда
 Цвет линии
 Тип линии
 Тень

Параметры линии тренда
 Построение линии тренда (аппроксимация и сглаживание)

☐ Экспоненциальная
 ☒ Линейная
 ☐ Логарифмическая
 ☐ Полиномиальная Степень: 2
 ☐ Степенная
 ☐ Линейная фильтрация Точки: 2

 Название аппроксимирующей (сглаженной) кривой

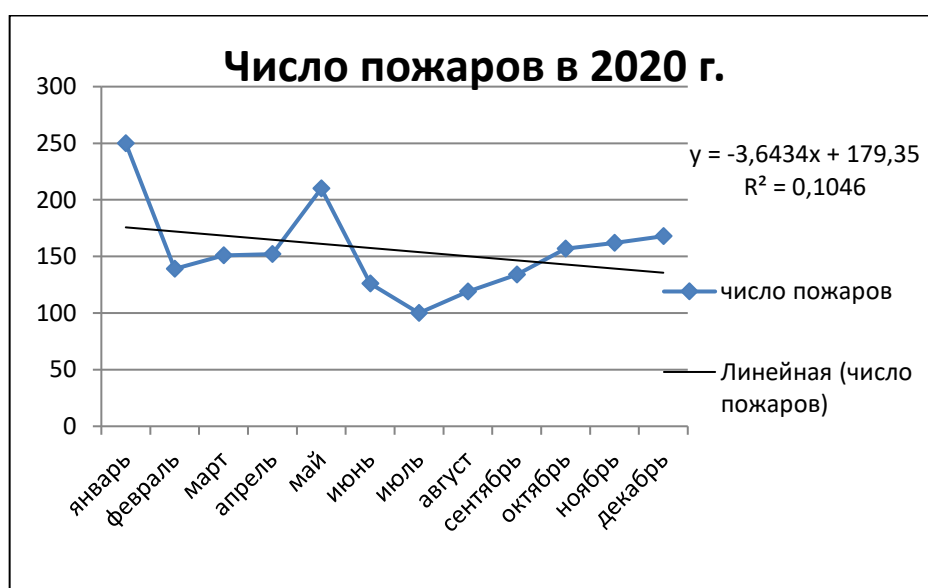
☒ автоматическое: Линейная (число пожаров)
 ☐ другое:

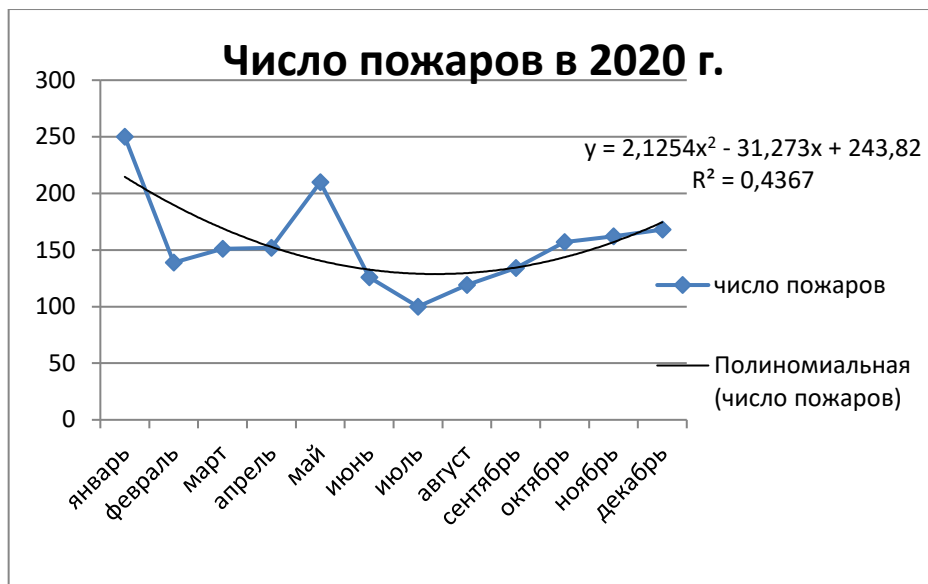
 Прогноз

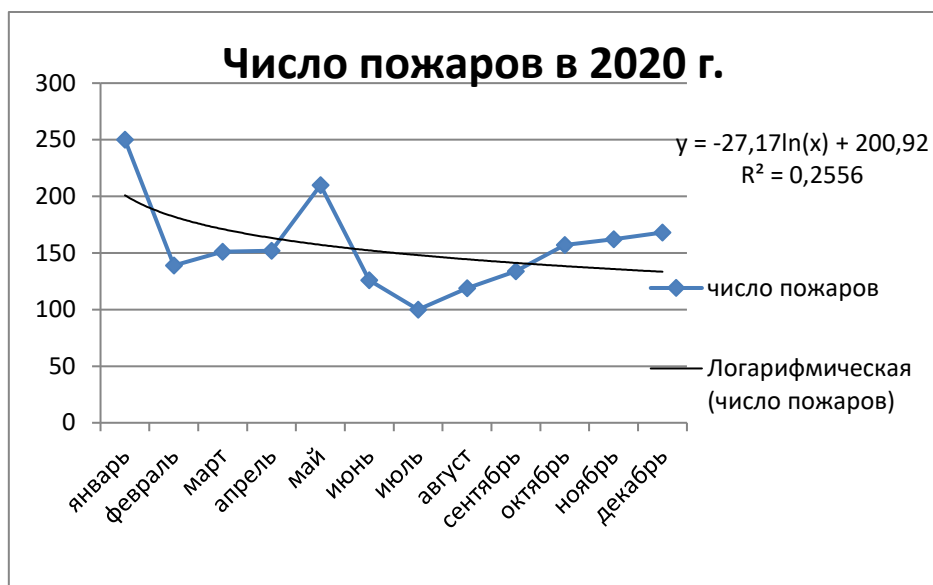
вперед на: 0,0 периодов
 назад на: 0,0 периодов

☐ пересечение кривой с осью Y в точке: 0,0
 ☒ показывать уравнение на диаграмме
 ☒ поместить на диаграмму величину достоверности аппроксимации (R^2)

Заккрыть





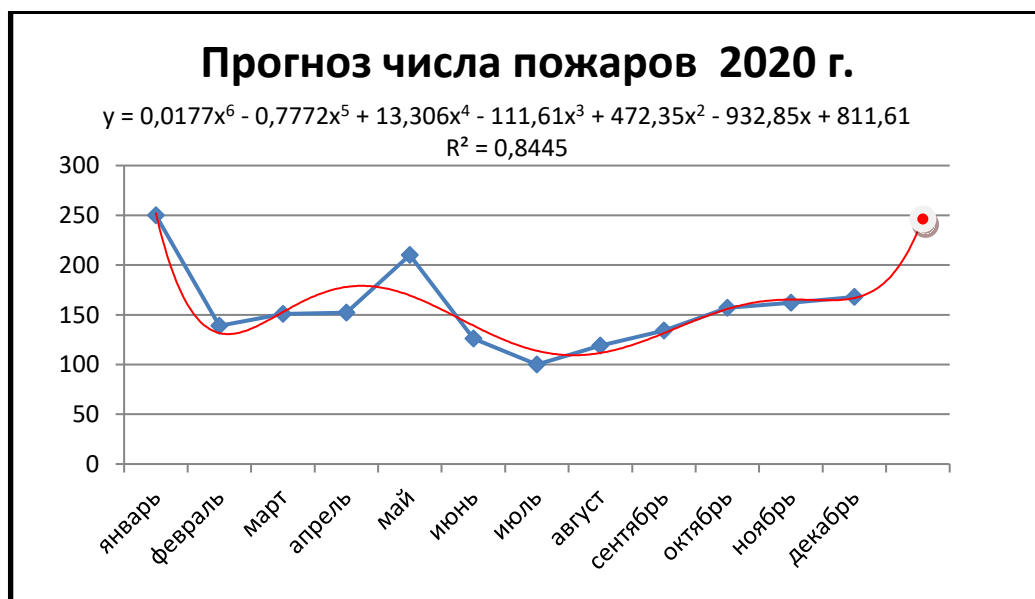


Величина коэффициента R^2 показывает, насколько линия тренда больше совпадает с данным графиком.

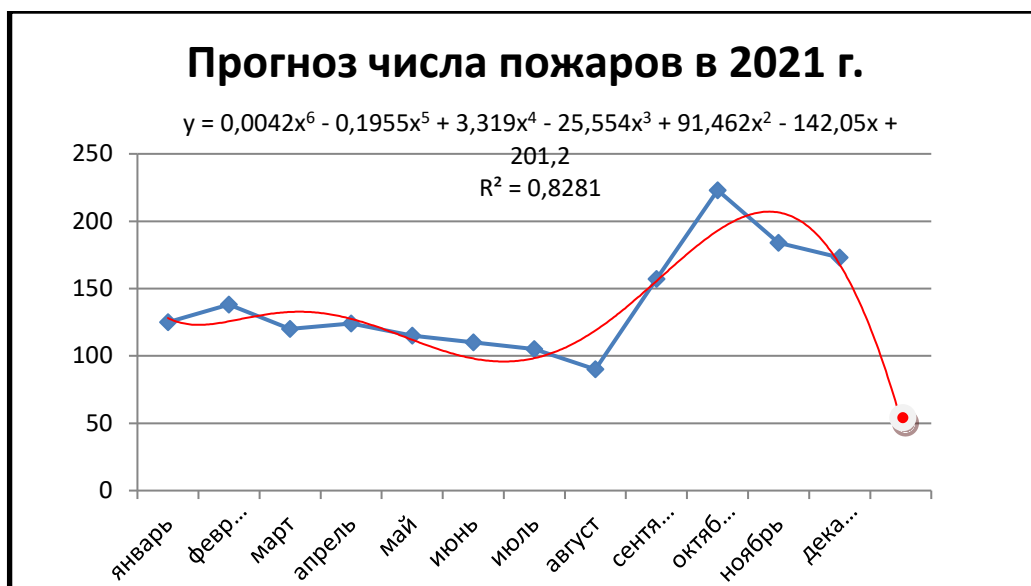
Чем он **более близок по значению к 1**, тем лучше линия тренда описывает рассматриваемый процесс.

Как видим из значений коэффициента детерминации R^2 он принимает наибольшее значение близкое к 1 при выборе в качестве линии тренда **полиномиальной функции степени 6**.

Для графика с этой линией тренда выполните прогноз количества пожаров для на 1 месяц вперёд, для этого в диалоговой панели «Формат линии тренда» указывается период прогноза (в нашем случае нужно указать на один период вперед).



Аналогичным образом постройте линии тренда и для показаний за 2021 год.



Недостаток данного метода: имеет место сдвиг прогнозных значений на один период в будущее относительно исходных данных.

В целом метод графического прогнозирования может быть полезен для прогнозирования показателей деятельности подразделений.

5. Метод переменного множителя

Данный метод основан на вычислении множителя по формуле:

$$K_i = y_i / y_{i-1},$$

где y_i — реальное число пожаров на данный месяц; y_{i-1} — данные за прошлый месяц.

Прогнозное значение на следующий месяц получают путем умножения реальных данных за i -й месяц на множитель K с помощью уравнения:

$$y_{i+1} = y_i K_i,$$

где K — значение переменного множителя за i -й месяц.

- Скопируйте исходные данные о пожарах за 2020 год в ячейку A2.
- Добавьте ещё два столбца для расчёта переменного множителя - **K** и линии прогноза.
- Введите в ячейку C4 формулу для расчёта переменного множителя для февраля (как деление значений количества пожаров за данный месяц и предыдущий).
- Скопируйте формулу на весь диапазон вниз.
- Вычислите прогноз, начиная с марта (как произведение значений количества пожаров и переменного множителя **K** за предыдущий месяц), м. рисунок.

C4 fx =B4/B3					
	A	B	C	D	E
1					
2	месяц 2020г.	число пожаров	к	линия прогноза	
3	январь	250			
4	февраль	139	0,56		
5	март	151	1,09		
6	апрель	152	1,01		
7	май	210	1,38		
8	июнь	126	0,60		
9	июль	100	0,79		
10	август	119	1,19		
11	сентябрь	134	1,13		
12	октябрь	157	1,17		
13	ноябрь	162	1,03		
14	декабрь	168	1,04		
15					

D5 fx =B4*C4					
	A	B	C	D	E
1					
2	месяц 2020г.	число пожаров	к	линия прогноза	
3	январь	250		250	
4	февраль	139	0,56	250	
5	март	151	1,09	77	
6	апрель	152	1,01	164	
7	май	210	1,38	153	
8	июнь	126	0,60	290	
9	июль	100	0,79	76	
10	август	119	1,19	79	
11	сентябрь	134	1,13	142	
12	октябрь	157	1,17	151	
13	ноябрь	162	1,03	184	
14	декабрь	168	1,04	167	
15					

- Выполните аналогичные расчеты для 2021 года.

Должны получиться результаты:

месяц 2020г.	число пожаров	к	линия прогноза
январь	250		250
февраль	139	0,56	250
март	151	1,09	77
апрель	152	1,01	164
май	210	1,38	153
июнь	126	0,60	290
июль	100	0,79	76
август	119	1,19	79
сентябрь	134	1,13	142
октябрь	157	1,17	151
ноябрь	162	1,03	184
декабрь	168	1,04	167

месяц 2021г.	число пожаров	к	линия прогноза
январь	125		125
февраль	138	1,10	125
март	120	0,87	152
апрель	124	1,03	104
май	115	0,93	128
июнь	110	0,96	107
июль	105	0,95	105
август	90	0,86	100
сентябрь	157	1,74	77
октябрь	223	1,42	274
ноябрь	184	0,83	317
декабрь	173	0,94	152

Постройте диаграммы





6. Метод Холта

Для реализации данного метода сначала для каждого месяца выделяется трендовая составляющая по уравнению:

$$T_i = (1 - \beta)(y_i - y_{i-1}) + \beta T_{i-1}, \quad i > 1,$$

где y_i — число пожаров в данном месяце;

i — порядковый номер месяца;

β — коэффициент сглаживания ($0 < \beta < 1$).

Сглаженное число пожаров вычислим по формуле:

$$\bar{y}_i = (1 - \alpha)y_i + \alpha(y_{i-1} + T_{i-1}),$$

где α — коэффициент сглаживания ($0 < \alpha < 1$).

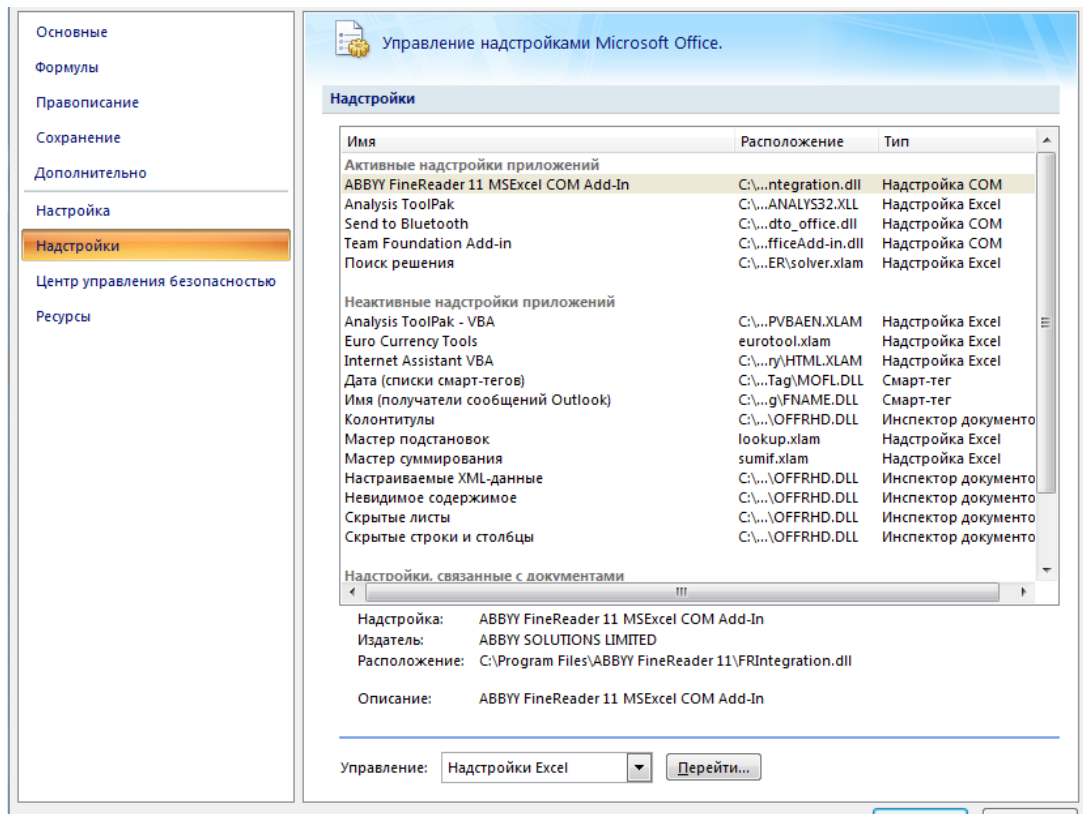
Прогноз на один месяц вперед выполняется по формуле:

$$\bar{y}_{n+1} = y_n + T_n.$$

Excel непосредственно поддерживает данный метод прогнозирования с помощью средства «**Экспоненциальное сглаживание**» в надстройке «Пакет анализа».

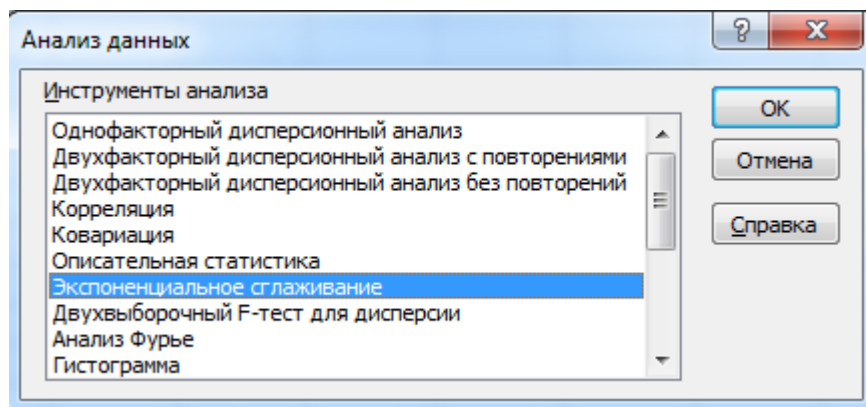
Если команды «Анализ данных» из этого пакета нет на панели «Данные», то её нужно включить:

1. Файл - Параметры.
2. Надстройки.
3. Выбрать - Пакет анализа
4. Перейти
4. Отметить галочками надстройки, связанные с Анализом данных.



- В программе Microsoft Excel сначала в ячейку A1 скопируйте исходную таблицу с данными о числе пожаров за 2020 год.

- Выполните расчеты прогноза, выбрав команду «Данные» – «Анализ данных», затем активизируйте средство «Экспоненциальное сглаживание»,



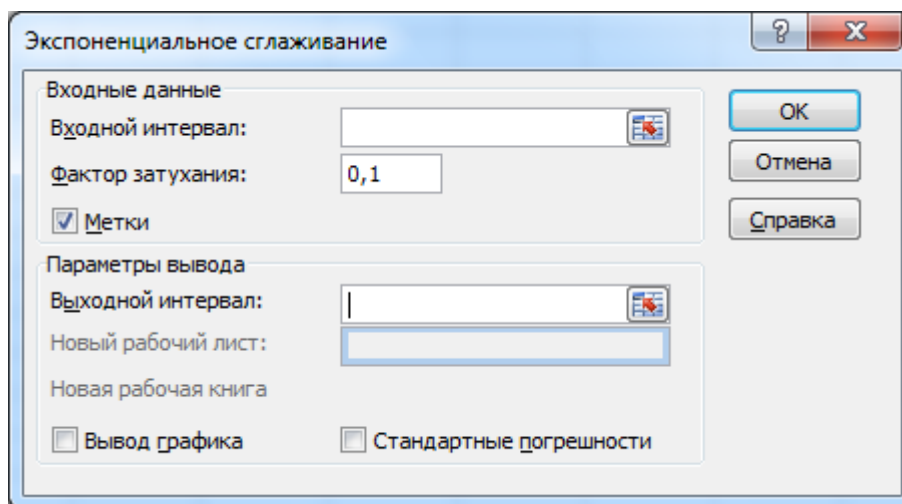
- В качестве параметра *Входной интервал* используйте весь диапазон ячеек с данными о количестве пожаров за все 12 месяцев 2020 года (B2:B13),

- Установите флажок *Метки*, а значение 0,1 — в качестве параметра *Фактор затухания*.

- В качестве параметра *Выходной интервал* примените ячейку, соответствующую прогнозу на февраль (C3).

- Нажмите кнопку - ОК

- Замените появившееся значение для прогноза на февраль - **Н/Д** (нехватка данных) на 250 - исходные данные за январь.



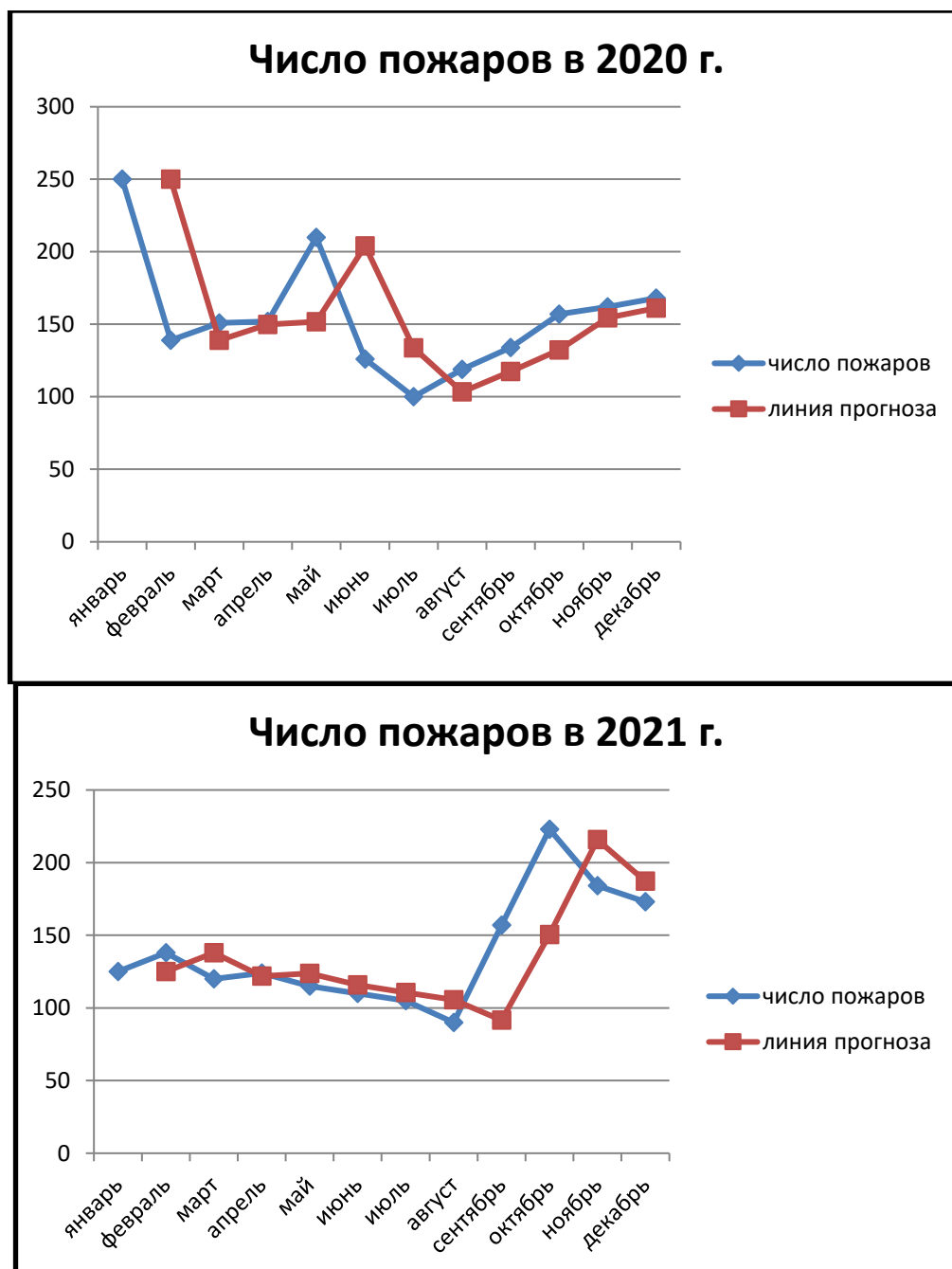
- Аналогичные действия произведите для данных за 2021 год.

Должны получиться результаты:

месяц 2020г.	число пожаров	линия прогноза
январь	250	
февраль	139	250,00
март	151	139,00
апрель	152	149,80
май	210	151,78
июнь	126	204,18
июль	100	133,82
август	119	103,38
сентябрь	134	117,44
октябрь	157	132,34
ноябрь	162	154,53
декабрь	168	161,25

месяц 2021г.	число пожаров	линия прогноза
январь	125	
февраль	138	125,00
март	120	138,00
апрель	124	121,80
май	115	123,78
июнь	110	115,88
июль	105	110,59
август	90	105,56
сентябрь	157	91,56
октябрь	223	150,46
ноябрь	184	215,75
декабрь	173	187,17

- Постройте графики.



Прогнозная кривая совпадает по форме с реальными данными.
Недостаток метода: сдвиг прогнозной кривой на один период в будущее.

Выводы

В результате проведенного исследования на основе реальных данных по числу пожаров за 2020 и 2021 гг. был выполнен прогноз обстановки с пожарами по месяцам с помощью различных методов прогнозирования. В итоге сравнения минимальное среднее абсолютное отклонение получено в методах Холта, скользящего среднего и квазислучайного числа.

Методы Холта, скользящего среднего и квазислучайного числа перспективными для прогнозирования показателей деятельности подразделений МЧС.

Лабораторные работы № 7-9
«МЕТОДЫ АНАЛИЗА ДАННЫХ В СРЕДЕ MS EXCEL»

Рассмотрим различные методы и способы анализа массива данных в среде MS Excel. В качестве исследуемой информации возьмем показатели городских пожаров в районах одного из субъектов РФ за 2021 год.

Район	Кол-во пожаров, шт.	Ущерб, тыс. руб.	Уничтожено зданий, шт.	Травмировано чел.	Погибло чел.
1. Ленинский	41	2921	6	4	7
2. Октябрьский	105	275	3	7	8
3. Железнодорожный	85	1340	13	6	7
4. Первомайский	68	3687	7	8	6
5. Башмаковский	22	182	7	8	8
6. Бековский	18	545	9	0	2
7. Белинский	28	778	16	1	8
8. Бессоновский	62	651	20	4	11
9. Вадинский	15	515	4	2	2
10. Городищенский	62	550	26	3	6
11. Земетчинский	38	2076	43	1	4
12. Иссинский	16	782	4	1	1
13. Каменский	62	1910	11	4	10
14. Камешкирский	20	888	6	0	2
15. Кольшлейский	22	1676	5	2	3
16. Кузнецкий	120	391	20	10	9
17. Лопатинский	18	272	6	3	4
18. Лунинский	35	4468	40	0	3
19. Мокшанский	36	884	3	1	6
20. Малосердобинский	22	596	5	1	4
21. Наровчатский	16	383	5	0	3
22. Неверкинский	15	369	6	0	1
23. Нижнеломовский	39	2169	10	5	9
24. Никольский	39	1064	25	4	5
25. Пачелмский	28	1390	4	0	6
26. Пензенский	67	1456	27	4	7
27. Сердобский	39	732	4	8	6
28. Сосновоборский	15	644	7	2	3
29. Спасский	16	549	6	1	3
30. Тамалинский	18	330	10	1	2
31. Шемышейский	15	696	8	1	2

7. Описательная статистика

Это техника сбора и суммирования количественных данных, которая используется для превращения массы цифровых данных в форму, удобную для восприятия и обсуждения.

Цель описательной статистики - обобщить первичные результаты, полученные в результате наблюдений и экспериментов.

- Скопируйте на первый лист книги MS Excel в ячейку B3 таблицу значений количества пожаров из исходной таблицы.

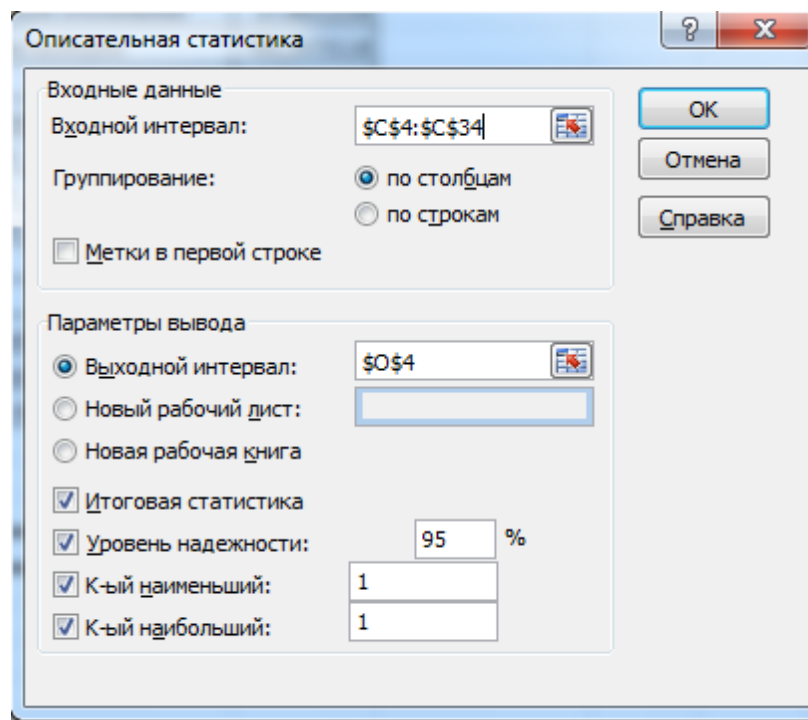
- Выберите в меню *Данные – Анализ данных* и "*Описательная статистика*"

- Выполните Описательную статистику:

Входной интервал - все исходные значения показателей количества пожаров (диапазон ячеек C4:C34)

Выходной интервал - любая пустая ячейка на листе.

Поставьте галочки на выводимых характеристиках описательной статистики.



В состав описательной статистики входят такие характеристики: среднее значение; стандартная ошибка; медиана; мода; стандартное отклонение; дисперсия выборки; эксцесс; асимметричность; интервал; минимум; максимум; сумма; счет.

<i>Кол-во пожаров</i>	
Среднее	38,77419
Стандартная ошибка	4,96844
Медиана	28
Мода	15
Стандартное отклонение	27,6631
Дисперсия выборки	765,2473
Эксцесс	1,723464
Асимметричность	1,465648
Интервал	105
Минимум	15
Максимум	120
Сумма	1202
Счет	31
Наибольший(1)	120
Наименьший(1)	15
Уровень надежности(95,0%)	10,14691

Рассмотрим, что же представляют собой характеристики описательной статистики:

Измерение **центральной тенденции** заключается в выборе числа, которое наилучшим способом описывает все значения признака набора данных. Такое число имеет как свои достоинства, так и недостатки. Мы рассмотрим две характеристики этого измерения, а именно: *среднее значение и медиану*.

Главная цель *среднего* – представление набора данных для последующего анализа, сопоставления и сравнения. Среднее значение рассчитывается как среднее арифметическое набора данных: сумма всех значений выборки, деленная на количество выборки.

Среднее значение очень информативно и позволяет делать вывод относительно всего исследуемого набора данных. При помощи среднего мы получаем возможность сравнивать несколько наборов данных или их частей.

С увеличением размера выборки точность оценки среднего возрастает. С увеличением разброса значений выборки надежность среднего падает. Если размер выборки достаточно большой, качество среднего увеличивается независимо от выполнения предположения нормальности выборки.

Медиана - точная середина выборки, которая делит ее на две равные части по числу наблюдений.

Обязательным условием нахождения медианы является упорядоченность выборки. Таким образом, для нечетного количества наблюдений медианой выступает наблюдение с номером $(n+1)/2$, где n - количество наблюдений в выборке.

Для четного числа наблюдений медианой является среднее значение наблюдений $n/2$ и $(n+2)/2$.

Наиболее простыми характеристиками выборки являются максимум и минимум.

Минимум – наименьшее значение выборки.

Максимум – наибольшее значение выборки.

Размах – разница между наибольшим и наименьшим значениями выборки.

Дисперсия - среднее арифметическое квадратов отклонений значений от их среднего.

Стандартное отклонение - квадратный корень из дисперсии выборки - мера того, насколько широко разбросаны точки данных относительно их среднего.

Эксцесс показывает "остроту пика" распределения, характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. Положительный эксцесс обозначает относительно остроконечное распределение (пик заострен). Отрицательный эксцесс обозначает относительно сглаженное распределение (пик закруглен).

Если эксцесс существенно отличается от нуля, то распределение имеет или более закругленный пик, чем нормальное, или, напротив, имеет более острый пик (возможно, имеется несколько пиков). Эксцесс нормального распределения равен нулю.

Асимметричность показывает отклонение распределения от симметричного. Если асимметрия существенно отличается от нуля, то распределение несимметрично, нормальное распределение абсолютно симметрично. Если распределение имеет длинный правый хвост, асимметрия положительна; если длинный левый хвост - отрицательна.

Выбросы – данные, резко отличающиеся от основного числа данных. При обнаружении выбросов перед исследователем стоит дилемма: оставить наблюдения-выбросы либо от них отказаться. Второй вариант требует серьезной аргументации и описания. Полезным будет провести анализ данных с выбросами и без и сравнить результаты.

Следует помнить, что при применении классических методов статистического анализа, которые, как правило, не являются устойчивыми, наличие выбросов в наборе данных приводит к некорректным результатам. Если набор данных относительно мал, исключение данных, которые считаются выбросами, может заметно повлиять на результаты анализа.

Наличие выбросов в наборе данных может быть связано с появлением так называемых "сдвинутых" значений, связанных с систематической ошибкой, ошибок ввода, ошибок сбора данных и т.д. Иногда к выбросам могут относиться наименьшие и наибольшие значения набора данных.

8. Частотный анализ

При анализе данных часто возникает вопрос, как часто встречаются некоторые показатели в заданных интервалах значений.

Функция **ЧАСТОТА** относится к категории статистических функций и показывает распределение частот в виде вертикального массива.

Синтаксис задания функции: **ЧАСТОТА** (массив_данных; массив_карманов).

Для данного множества значений и заданного множества интервалов (карманов) частотное распределение подсчитывает, сколько значений попадает в каждый интервал.

- Скопируйте на новый лист книги MS Excel в ячейку B3 таблицу значений количества пожаров из исходной таблицы.

- Для подсчета показателей частоты подготовьте рядом с таблицей, содержащей исходные данные, таблицу для вывода результатов работы, в которой в столбик запишите границы рассматриваемых интервалов

от 0 до 20, от 20 до 100, от 100 до 500;

Например, как показано ниже на рисунке

	A	B	C	I	J	K	L
1							
2		Район	Кол-во пожаров,			Кол-во пожаров,	
3							
4	1	Ленинский	41		0		
5	2	Октябрьский	105		20		
6	3	Железнодорожный	85		50		
7	4	Первомайский	68		100		
8	5	Башмаковский	22		100		
9	6	Бековский	18		500		
10	7	Белинский	28				
11	8	Бессоновский	62				
12	9	Вадинский	15				
13	10	Городищенский	62				
14	11	Земетчинский	38				
15	12	Иссинский	16				

- Выделите диапазон ячеек K4:K5 и введите в строке формул =**ЧАСТОТА**(

- Далее выделите указателем мыши диапазон ячеек, содержащих данные о количестве пожаров **C4:C34**.

- Поставьте точку с запятой и выделите диапазон с границами рассматриваемого интервала **J4:J5**. Закройте скобку.

ЧАСТОТА X ✓ fx =ЧАСТОТА(C4:C34;J4:J5)							
	A	B	C	I	J	K	L
1							
2		Район	Кол-во пожаров,			Кол-во пожаров,	
3							
4	1	Ленинский	41		0	=ЧАСТОТА(C4:C34;J4:J5)	
5	2	Октябрьский	105		20		
6	3	Железнодорожный	85		50		
7	4	Первомайский	68		100		
8	5	Башмаковский	22		100		
9	6	Бековский	18		500		
10	7	Белинский	28				
11	8	Бессоновский	62				
12	9	Вадинский	15				

- Не выходя из строки формул, нажмите одновременно комбинацию клавиш <Ctrl/ Shift/ Enter> для расчета элементов массива.

- Полученный результат:

K4		{=ЧАСТОТА(C4:C34;J4:J5)}				
	A	B	C	I	J	K
1						
2		Район	Кол-во пожаров,			Кол-во пожаров,
3						
4	1	Ленинский	41		0	0
5	2	Октябрьский	105		20	11
6	3	Железнодорожный	85		50	
7	4	Первомайский	68		100	
8	5	Башмаковский	22		100	
9	6	Бековский	18		500	
10	7	Белинский	28			
11	8	Бессоновский	62			
12	9	Вадинский	15			
13	10	Городищенский	62			
14	11	Земетчинский	38			
15	12	Иссинский	16			

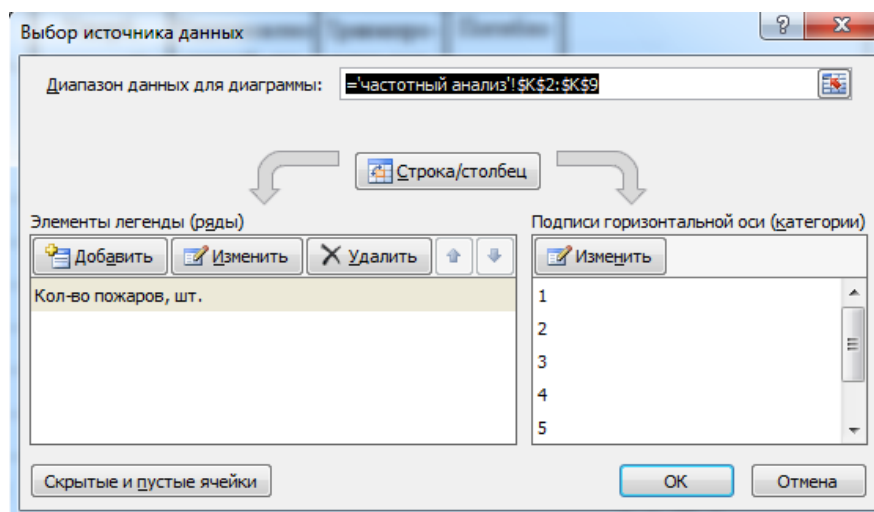
- Для остальных интервалов проделайте аналогичный алгоритм действий. В результате должны получиться следующие показатели:

	Кол-во пожаров,	
0	0	количество нулевых показателей
20	11	кол-во показателей от 0 до 20
50	23	кол-во показателей от 0 до 50
100	6	кол-во показателей от 50 до 100
100	29	кол-во показателей от 0 до 100
500	2	кол-во показателей свыше 100 (от 100 до 500)

- Постройте диаграмму по полученным результатам.

Для этого выделите диапазон с заголовком и результатами вычислений K2:K9 и выберите команды *Вставка-Гистограмма*.

- Для редактирования и форматирования диаграммы нажмите пр. кл. мыши на пустой области диаграммы и в контекстном меню выделите команду *Выбрать данные*.



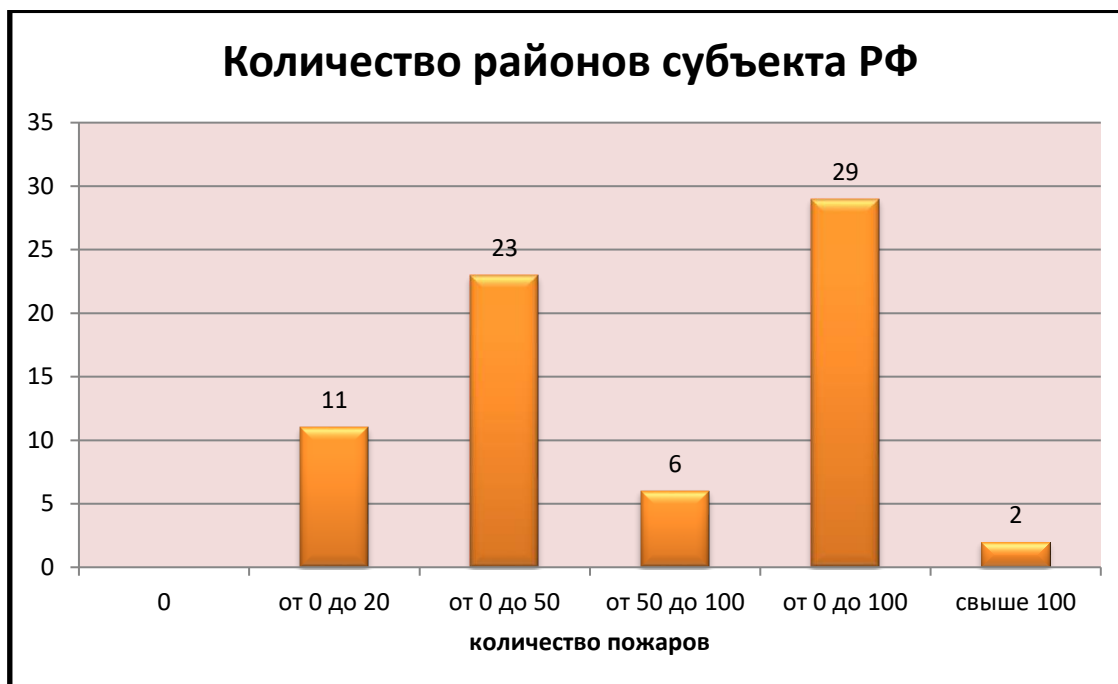
Если в левом разделе *Элементы легенды* находятся названия рядов, то в этом разделе необходимо нажать кнопку *Изменить* и выделить в исходной таблице диапазон ячеек с названиями рассматриваемых данных (в нашем случае это ячейки K2:K3).

В разделе *Подписи горизонтальной оси* нажмите кнопку *Изменить*.

В поле *Диапазон подписей оси* впишите строку:

={"0","от 0 до 20","от 0 до 50","от 50 до 100","от 0 до 100","свыше 100"}

- Далее выполните форматирование различных областей диаграммы по своему вкусу, нажимая на них пр. клавишей мыши и выбирая соответствующие команды.

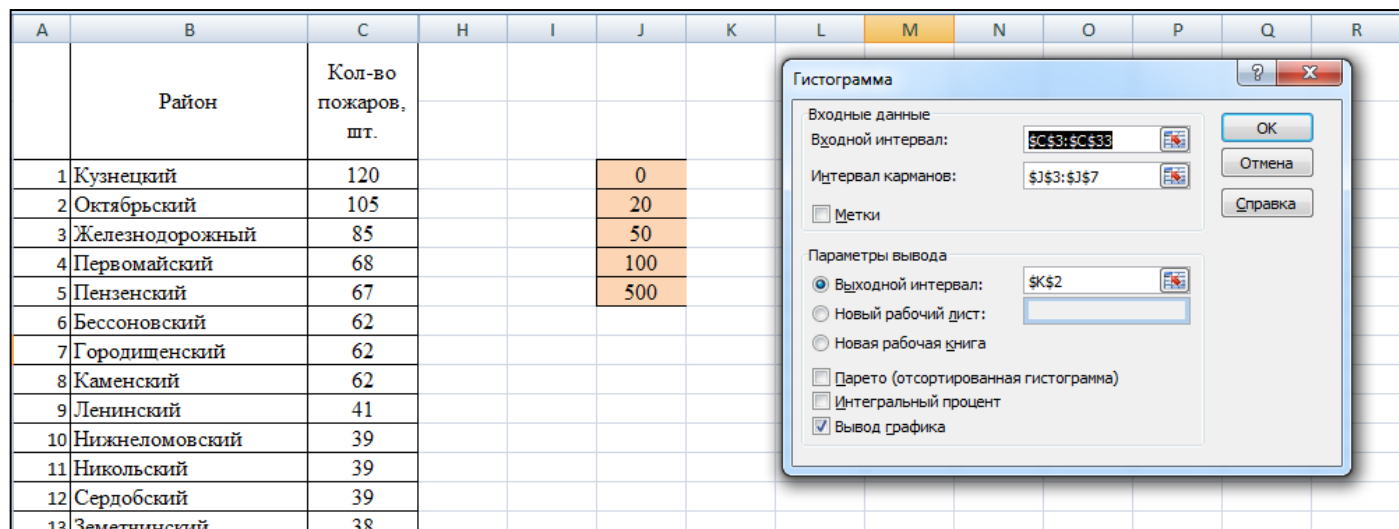


- Частотный анализ можно так же произвести с помощью инструмента *Анализа данных* - «Гистограмма».

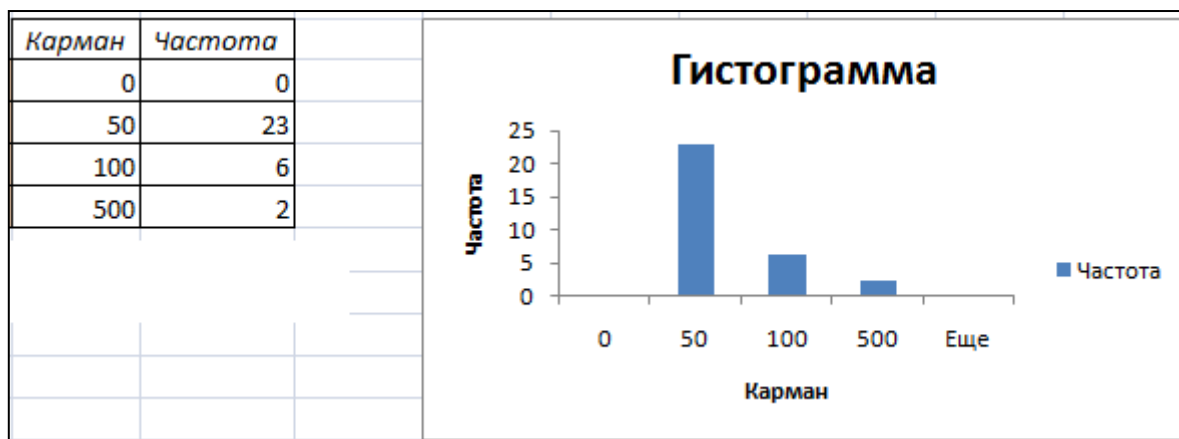
Выполните его на отдельном листе (подготовьте исходную таблицу и таблицу с промежутками исследования), см. рисунок.

- Выберите команду *Данные – Анализ данных – Гистограмма*.

- Задайте *Входной интервал* данных, *Интервал карманов* (промежутков исследования), *Выходной интервал* (любая пустая ячейка листа) и поставьте галочку для вывода графика, см. рисунок.



Выходные данные будут немного отличаться от результатов предыдущего способа анализа.



Оценим результаты анализа:

23 – количество районов с показателями пожаров от 0 до 50,

6 – количество районов с показателями пожаров от 50 до 100,

2 – количество районов с показателями пожаров от 100 до 500.

Недостатки выведенной диаграммы – невозможность ее редактирования.

9. Ранжирование показателей

Для сравнения различных показателей некоторой величины друг с другом в программе MS Excel используется функция РАНГ, которая находит ранг числа в списке данных, то есть его номер по величине относительно других значений в списке. Если несколько значений имеют одинаковый ранг, отображается среднее.

- Скопируйте на новый лист в ячейку В3 исходные данные о количестве пожаров
- Добавьте к исходной таблице столбец «Ранг по количеству пожаров».
- В первую ячейку этого столбца для первого района впишите формулу =РАНГ(С4;\$С\$4:\$С\$34).

Где С4 – ячейка с показателями по количеству пожаров в первом районе, а \$С\$4:\$С\$34 – абсолютная ссылка на весь диапазон данных по столбцу «Кол-во пожаров».

- Скопируйте данную формулу маркером автозаполнения вниз до конца диапазона. В результате напротив каждого района отобразится его ранг (номер) в зависимости от величины количества пожаров (в порядке убывания, 1 – район с наибольшим количеством пожаров и т.д.)

- Выполните сортировку данных в столбце о количестве пожаров по возрастанию (Данные - Сортировка).

- Постройте гистограмму ранжирования рассматриваемых показателей.



Вывести таблицу, содержащую порядковый и процентный ранги для каждого значения в наборе данных можно так же с помощью инструмента анализа данных РАНГ И ПЕРСЕНТИЛЬ.

Данная процедура может быть применена для анализа относительного взаиморасположения данных в наборе.

- Для исходной таблицы о количестве пожаров выберите *Данные – Анализ данных – Ранг и персентиль*.

- *Входной интервал* - исходные данные о количестве пожаров.

- *Выходной интервал* - любая ячейка листа

Точка	Столбец1	Ранг	Процент
16	120	1	100,00%
2	105	2	96,60%
3	85	3	93,30%
4	68	4	90,00%
26	67	5	86,60%
8	62	6	76,60%
10	62	6	76,60%
13	62	6	76,60%
1	41	9	73,30%
23	39	10	63,30%
24	39	10	63,30%
27	39	10	63,30%
11	38	13	60,00%
19	36	14	56,60%
18	35	15	53,30%
7	28	16	46,60%
25	28	16	46,60%
5	22	18	36,60%
15	22	18	36,60%
20	22	18	36,60%
14	20	21	33,30%
6	18	22	23,30%
17	18	22	23,30%
30	18	22	23,30%
12	16	25	13,30%
21	16	25	13,30%
29	16	25	13,30%
9	15	28	0,00%
22	15	28	0,00%
28	15	28	0,00%
31	15	28	0,00%