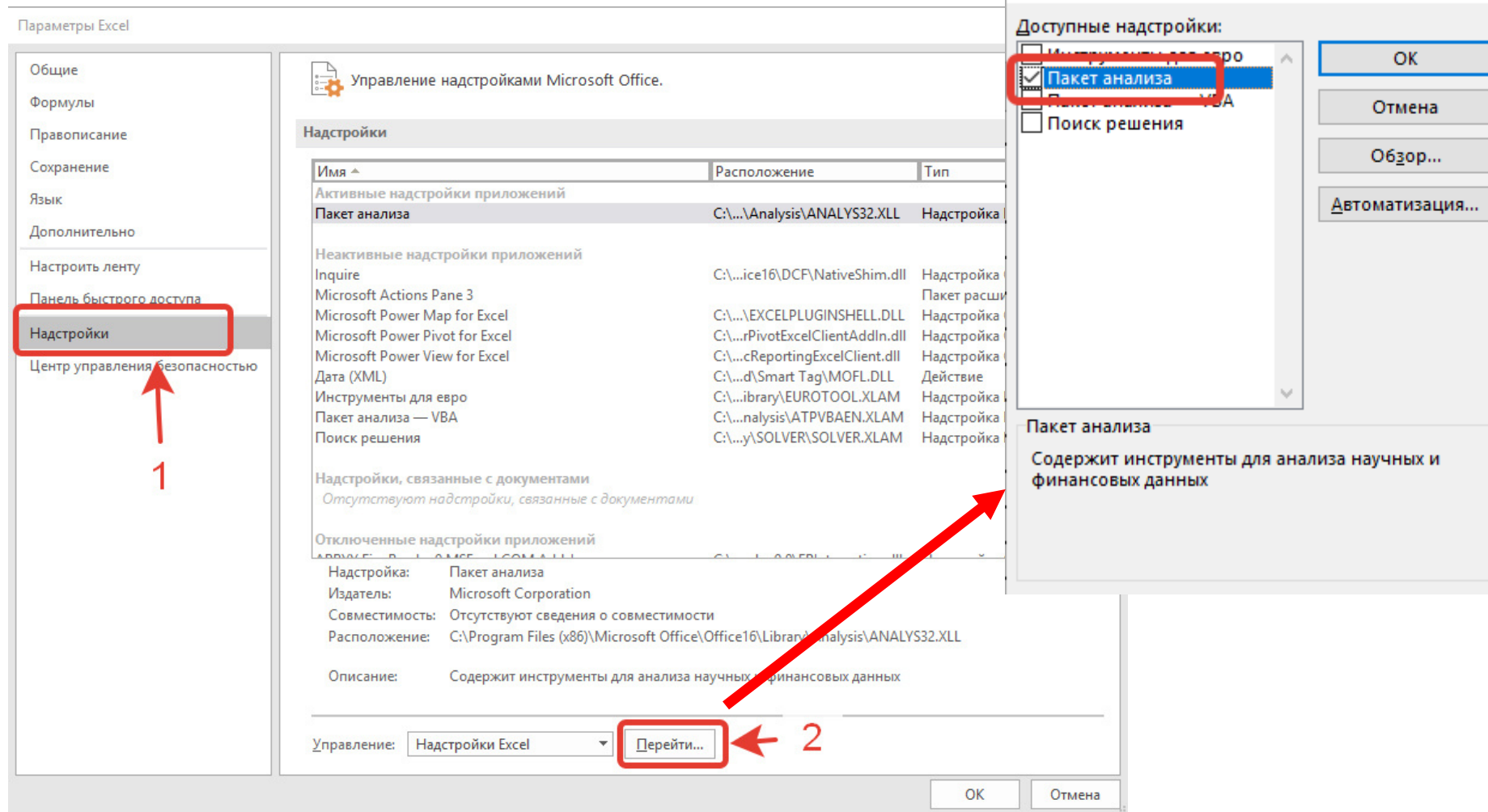


АНАЛИЗ ДАННЫХ

Настройка Excel

Выбрать вкладку **Файл - Параметры**



КОРРЕЛЯЦИОННЫЙ и РЕГРЕССИОННЫЙ АНАЛИЗ

Корреляционный и регрессионный анализ – наиболее распространенные статистические методы исследования

Позволяют показать зависимость какого-либо параметра от одной или нескольких независимых переменных

Парная корреляция

Парная корреляция позволяет оценить тесноту связи между двумя показателями

Теснота связи определяется коэффициентом парной корреляции:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y - \bar{y})}{n\sigma_x\sigma_y}$$

где n – число наблюдений, $\bar{x}$ и $\bar{y}$ – средние значения факторов x и y , σ_x и σ_y – средние квадратические (стандартные) отклонения факторов x и y

Парная корреляция

Парная корреляция позволяет оценить тесноту связи между двумя показателями

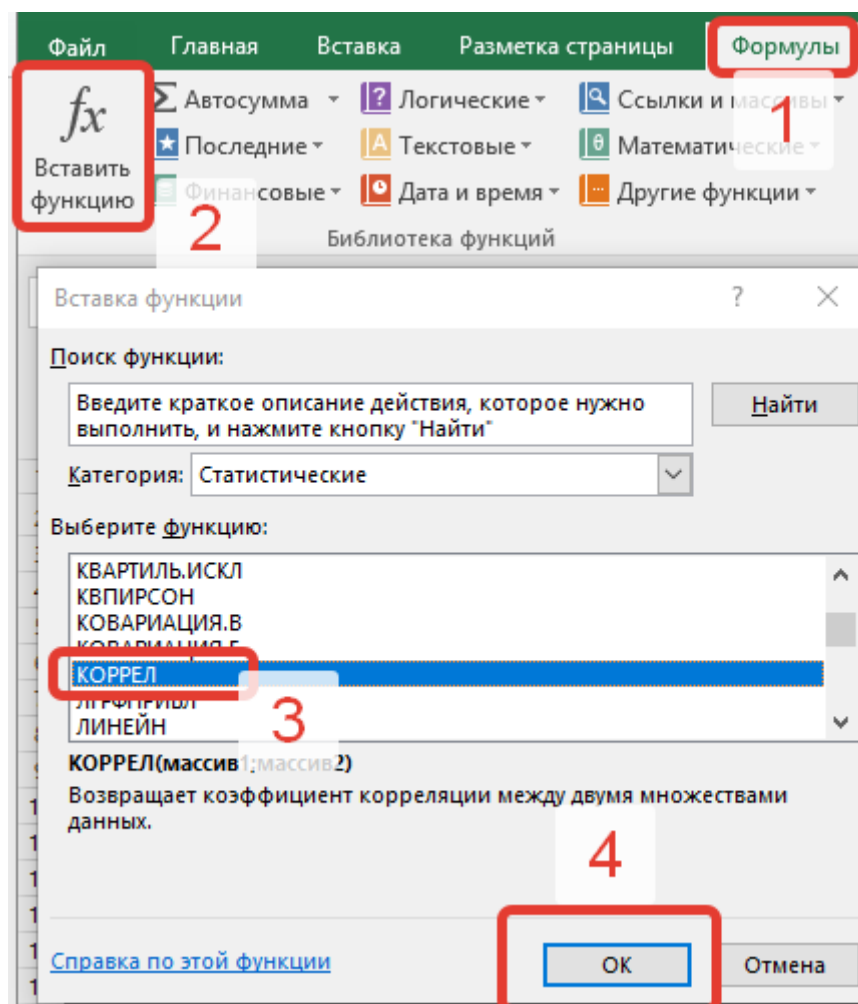
Теснота связи определяется коэффициентом парной корреляции:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y - \bar{y})}{n\sigma_x\sigma_y}$$

где n – число наблюдений, $\bar{x}$ и $\bar{y}$ – средние значения факторов x и y , σ_x и σ_y – средние квадратические (стандартные) отклонения факторов x и y

Парная корреляция

Для нахождения парной корреляции используется функция КОРРЕЛ:



Парная корреляция

O18 =КОРРЕЛ(B2:B33;E2:E33)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1																						
2	№ пробы	A ^d	SiO ₂	Al ₂ O ₃	Fe ₂ O ₃	CaO	MgO	K ₂ O	Na ₂ O	SO ₃												
3	1	8.20	44.46	4.57	37.64	4.09	1.96	0.59	0.42	4.33												
4	2	6.60	40.75	28.45	13.65	5.07	2.96	2.31	2.34	1.95												
5	3	6.80	28.13	11.23	41.13	6.34	3.42	0.82	1.70	5.84												
6	4	8.10	37.50	10.23	37.40	4.35	2.29	0.62	1.51	4.49												
7	5	8.10	46.50	14.35	86.28	3.29	2.42	1.57	2.50	2.52												
8	6	5.50	36.81	9.49	23.07	10.73	4.78	1.29	0.76	11.64												
9	7	5.80	45.11	9.86	29.14	3.85	2.88	1.54	1.87	4.80												
10	8	4.90	21.13	17.23	50.81	3.19	8.99	0.00	0.00	2.84												
11	9	8.00	24.15	17.84	46.31	2.98	2.57	0.00	0.00	3.46												
12	10	7.50	26.52	16.60	36.75	7.11	1.90	1.26	1.27	2.95												
13	11	5.10	31.16	14.40	34.64	6.70	3.61	0.93	1.45	2.93												
14	12	8.30	32.18	8.14	17.39	15.60	6.87	1.12	3.73	14.36												
15	13	7.90	30.73	10.76	31.95	11.14	3.35	1.56	1.54	5.87												
16	14	8.20	22.30	16.53	36.29	8.99	3.34	1.17	1.61	8.06												
17	15	4.50	21.65	13.53	45.38	6.83	3.56	1.25	1.71	5.06												
18	16	6.40	26.58	17.24	30.09	10.84	1.39	3.00	2.37	5.04												
19	17	7.80	34.60	27.04	24.62	4.24	2.25	0.00	0.00	3.63												
20	18	7.20	21.91	11.72	41.93	8.06	3.49	1.00	1.07	8.45												
21	19	3.90	19.70	12.24	44.77	6.15	2.10	0.96	2.22	6.86												
22	20	4.10	22.70	20.33	26.12	14.34	2.40	1.20	1.17	6.10												
23	21	7.40	29.25	3.77	50.47	5.79	2.48	0.64	0.88	5.90												
24	22	6.30	26.20	5.11	52.06	4.11	1.63	0.55	1.13	6.85												
25	23	6.40	42.90	25.16	22.52	3.66	1.28	1.76	0.92	0.95												
26	24	5.50	32.61	9.00	37.74	6.40	3.60	1.03	1.51	6.33												
27	25	8.60	20.87	5.76	49.04	9.18	1.53	0.83	0.91	11.08												
28	26	8.40	34.95	13.07	39.43	3.16	2.03	2.33	1.37	2.50												
29	27	8.10	14.72	6.15	58.70	6.44	2.02	3.00	2.95	6.64												
30	28	8.50	36.50	8.26	34.32	10.63	1.35	1.02	0.92	6.64												
31	29	5.80	26.62	6.85	52.55	5.04	1.96	0.81	1.02	4.15												
32	30	8.00	40.23	9.11	31.93	5.36	2.81	1.31	0.68	5.61												
33	31	8.60	30.21	11.07	39.72	7.35	2.18	0.93	1.35	5.61												
34																						

Аргументы функции

КОРРЕЛ 1

Массив1 B2:B33 = {"Ad":8.2;6.6;6.8;8.1;8.1;5.5;5.8;4.9;8.1;7.5;5.1;8.3;7.9;8.2;4.5;6.4;7.8;7.2;3.9;4.1;7.4;6.3;6.4;5.5;8.6;8.4;8.1;8.5;8.0;8.6}

Массив2 E2:E33 = {"Fe2O3":37.64;13.65;41.13;37.4;86.28;23.07;29.14;50.81;46.31;36.75;7.11;6.7;15.6;11.14;36.29;45.38;10.84;4.24;8.06;6.15;14.34;5.79;4.11;3.66;22.52;6.4;9.18;3.16;6.44;10.63;5.04;5.36;7.35}

Возвращает коэффициент корреляции между двумя множествами данных.

Массив2 второй диапазон значений. Значениями могут быть числа, имена, массивы или ссылки с именами.

Значение: 0.099397382

Справка по этой функции

3

Коэффициент корреляции варьируется от «-1» (корреляция отрицательная) до «+1» (корреляция положительная), если $r_{xy} = 0$, связь отсутствует.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционное поле (поле корреляции, диаграмма рассеяния) — это графическое изображение исходных данных

Для построения поля корреляции (или диаграммы рассеивания)

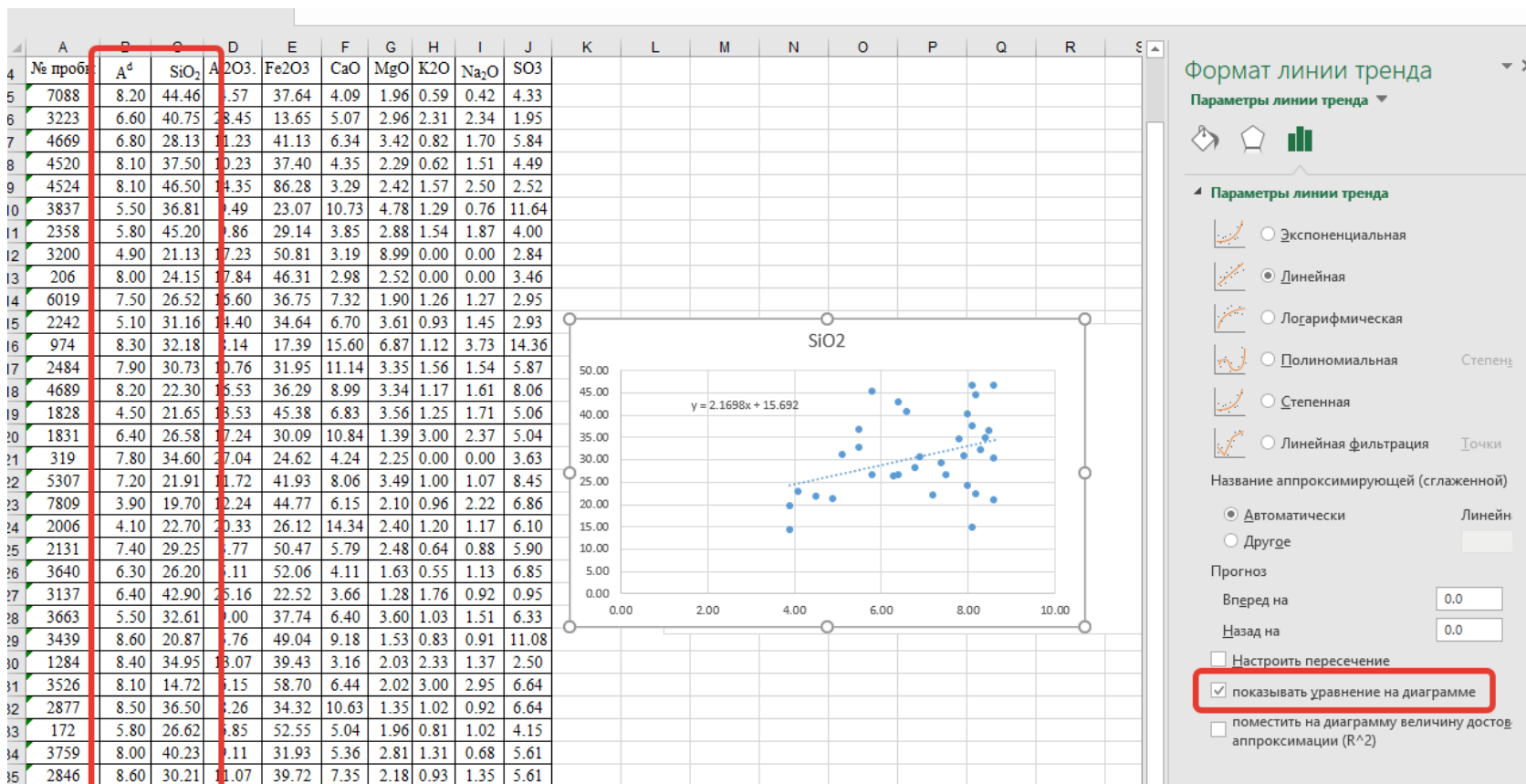
The screenshot shows the Microsoft Excel interface with the 'Вставка' (Insert) tab selected. The 'Вставка диаграммы' (Insert Chart) icon is highlighted with a red box and the number 2. The 'Точечная' (Scatter) chart type is selected from the dropdown menu, also highlighted with a red box and the number 3. The 'Пузырьковая' (Bubble) chart type is also visible. The background shows a table of chemical data.

C	D	E	F	G	H	I	J	K
SiO ₂	Al ₂ O ₃	Fe ₂ O ₃	CaO	MgO	K ₂ O	Na ₂ O	SO ₃	
44.46	4.57	37.64	4.09	1.96	0.59	0.42	4.33	
40.75	28.45	13.65	5.07	2.96	2.31	2.34	1.95	
28.13	11.23	41.13	6.34	3.42	0.82	1.70	5.84	
37.50	10.23	37.40	4.35	2.29	0.62	1.51	4.49	
46.50	14.35	86.28	3.29	2.42	1.57	2.50	2.52	

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционное поле

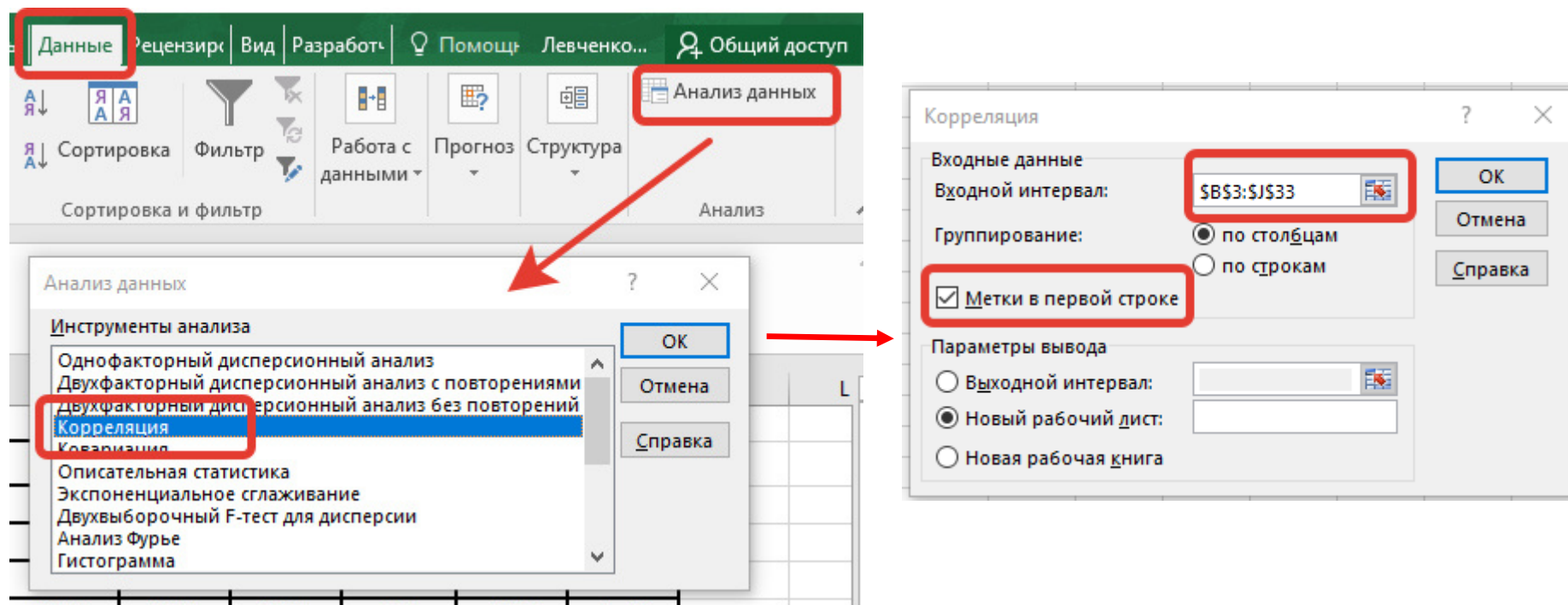
Предварительно необходимо выделить колонки значений, для которых необходимо построить поле корреляций



КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционная матрица — это квадратная (или прямоугольная) таблица, в которой на пересечении соответствующих строки и столбца находится **коэффициент корреляции** между соответствующими параметрами.

Вызов функции корреляционного анализа



КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционная матрица

A	B	C	D	E	F	G	H	I	J
	<i>Ad</i>	<i>SiO2</i>	<i>Al2O3</i>	<i>Fe2O3</i>	<i>CaO</i>	<i>MgO</i>	<i>K2O</i>	<i>Na2O</i>	<i>SO3</i>
Ad	1								
SiO2	0.232689	1							
Al2O3	-0.20141	0.113102	1						
Fe2O3	0.099397	-0.24137	-0.39047	1					
CaO	-0.02637	-0.2741	-0.11387	-0.40438	1				
MgO	-0.25476	-0.12857	0.026438	-0.11126	0.171825	1			
K2O	0.051729	0.073681	0.105543	-0.11317	0.195189	-0.27046	1		
Na2O	0.009942	-0.03466	-0.10095	0.008759	0.370347	0.043667	0.634631	1	
SO3	0.120413	-0.31317	-0.50463	-0.14241	0.706934	0.270502	-0.07232	0.280965	1

РЕГРЕССИОННЫЙ АНАЛИЗ

Показывает влияние одних значений (самостоятельных, независимых) на зависимую переменную

Виды регрессии:

- линейная - $y = a + bx$;
- параболическая - $y = a + bx + cx^2$;
- экспоненциальной - $y = a * \exp(bx)$;
- степенная - $y = a * x^b$;
- гиперболическая - $y = b/x + a$;
- логарифмическая - $y = b * \ln(x) + a$;
- показательная - $y = a * b^x$.

РЕГРЕССИОННЫЙ АНАЛИЗ

Модель линейной регрессии (множественной) имеет, в общем виде определяется выражением:

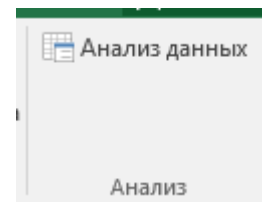
$$y = a_0 + a_1x_1 + \dots + a_kx_k,$$

где a – коэффициенты регрессии, x – влияющие переменные, k – число факторов

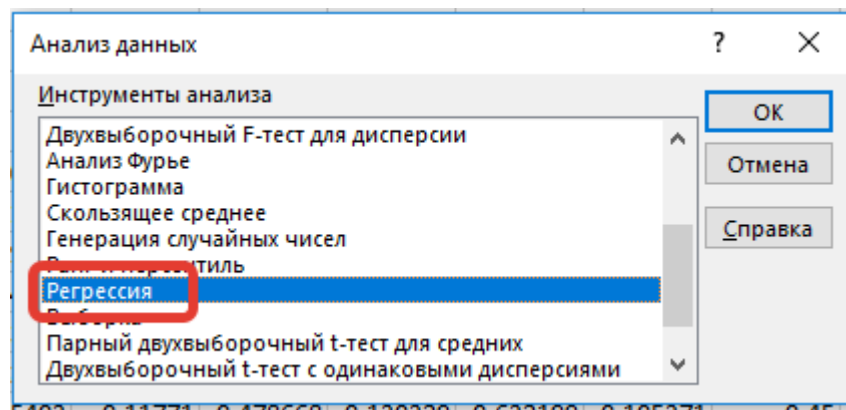
РЕГРЕССИОННЫЙ АНАЛИЗ

Вызов функции «Регрессия»

1. Перейти на вкладку «Данный»
2. Открыт группу команд «Анализ данных»



3. В открывшемся списке команд выбрать «Регрессия»



РЕГРЕССИОННЫЙ АНАЛИЗ

Рассмотрим зависимость концентрации As от концентрации Fe

Mn, %	Fe, %	Li	Be	Sc	V	Ti	Cr	Co	Ni	Cu	Zn	Ga	Ge	As	Rb
0.12	1.10	12.00	1.90	3.70	35.00	570.00	35.00	6.10	14.00	59.00	46.00	3.00	0.13	6.00	7.10
0.03	0.92	1.70	2.10	0.90	2.00	42.00	460.00	3.50	29.00	24.00	15.00	1.30	0.13	35.00	0.81
0.01	1.30	5.00	1.20	1.70	11.00	20.00	710.00	5.30	18.00	150.00	71.00	1.40	0.06	18.00	3.10
0.01	1.20	2.10	4.20	3.40	12.00	9.30	16.00	9.10	44.00	44.00	200.00	1.50	1.10	18.00	1.90
0.05	4.20	1.90	5.00	4.70	9.80	130.00	64.00	5.80	10.00	32.00	110.00	4.60	1.60	66.00	1.40
0.05	1.00	0.40	5.50	0.90	6.20	0.90	23.00	2.80	15.00	23.00	11.00	1.90	0.06	14.00	0.19
0.02	1.30	1.60	3.10	1.80	12.00	84.00	9.20	2.90	5.00	35.00	17.00	2.00	0.45	4.40	1.00
0.02	1.80	1.30	2.60	0.90	24.00	32.00	99.00	3.50	18.00	20.00	21.00	0.82	0.06	15.00	0.13
0.01	1.60	1.30	6.10	3.10	16.00	18.00	31.00	4.70	11.00	25.00	43.00	2.40	0.93	27.00	1.10
0.01	0.77	2.70	4.00	2.40	36.00	190.00	63.00	2.30	22.00	32.00	47.00	2.70	0.19	4.10	3.10
0.03	1.70	0.46	6.20	0.90	2.00	0.90	9.10	3.50	5.00	33.00	25.00	2.10	0.11	23.00	0.43
0.11	4.20	3.20	5.20	4.60	5.20	35.00	170.00	11.00	5.00	28.00	140.00	3.80	1.10	52.00	1.10
0.03	1.90	4.10	3.50	2.00	8.60	95.00	13.00	3.30	5.00	44.00	32.00	1.50	0.06	6.80	2.80
0.04	4.20	17.00	2.50	1.80	20.00	390.00	72.00	3.40	5.00	42.00	32.00	1.20	0.06	4.90	5.30
0.01	3.90	12.00	4.20	1.80	5.80	74.00	260.00	6.70	13.00	40.00	39.00	1.60	0.16	34.00	2.50
0.00	0.39	9.90	2.30	6.80	9.30	0.90	39.00	1.30	5.00	30.00	18.00	0.87	0.06	1.40	11.00
0.03	2.70	1.20	5.30	1.80	2.00	17.00	170.00	4.40	22.00	16.00	25.00	2.50	0.43	20.00	0.24
0.00	1.10	2.20	4.40	1.70	18.00	67.00	60.00	2.70	5.00	22.00	17.00	1.30	0.16	15.00	1.40
0.00	1.40	1.90	5.90	1.70	25.00	18.00	34.00	6.30	23.00	27.00	35.00	1.40	0.43	21.00	0.86
0.01	1.20	7.50	4.20	2.20	13.00	91.00	82.00	6.70	19.00	54.00	47.00	2.10	0.71	13.00	1.60
0.02	1.30	5.90	5.30	1.10	150.00	110.00	210.00	3.90	88.00	110.00	110.00	1.20	0.15	10.00	1.80

РЕГРЕССИОННЫЙ АНАЛИЗ

Рассмотрим зависимость концентрации As от концентрации Fe

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

☐ Стандартизованные остатки

☐ График остатков

☐ График подбора

Нормальная вероятность

☐ График нормальной вероятности

OK

Отмена

Справка

В качестве входного интервала выделяем столбец с данными о концентрации As. Выходной интервал – соответствующие ячейки концентраций Fe.

РЕГРЕССИОННЫЙ АНАЛИЗ

Результаты расчета

Регрессионная статистика								
Множественный R	0.61778763							
R-квадрат	0.381661556							
Нормированный R-квадрат	0.349117428							
Стандартная ошибка	13.12983165							
Наблюдения	21							
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	1	2021.734324	2021.734324	11.72750884	0.002842143			
Остаток	19	3275.457105	172.3924792					
Итого	20	5297.191429						
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95.0%	Верхние 95.0%
Y-пересечение	4.02301897	5.340545219	0.753297427	0.460504448	-7.154870638	15.20090858	-7.154870638	15.20090858
Переменная X 1	8.272501318	2.415649724	3.424545056	0.002842143	3.216488339	13.3285143	3.216488339	13.3285143

R-квадрат – коэффициент детерминации.

В нашем примере – 0,38, или 38%. Это означает, что расчетные параметры модели на 38% объясняют зависимость между изучаемыми параметрами.

Чем выше коэффициент детерминации, тем качественнее модель.

Хорошо – выше 0,8. Плохо – меньше 0,5

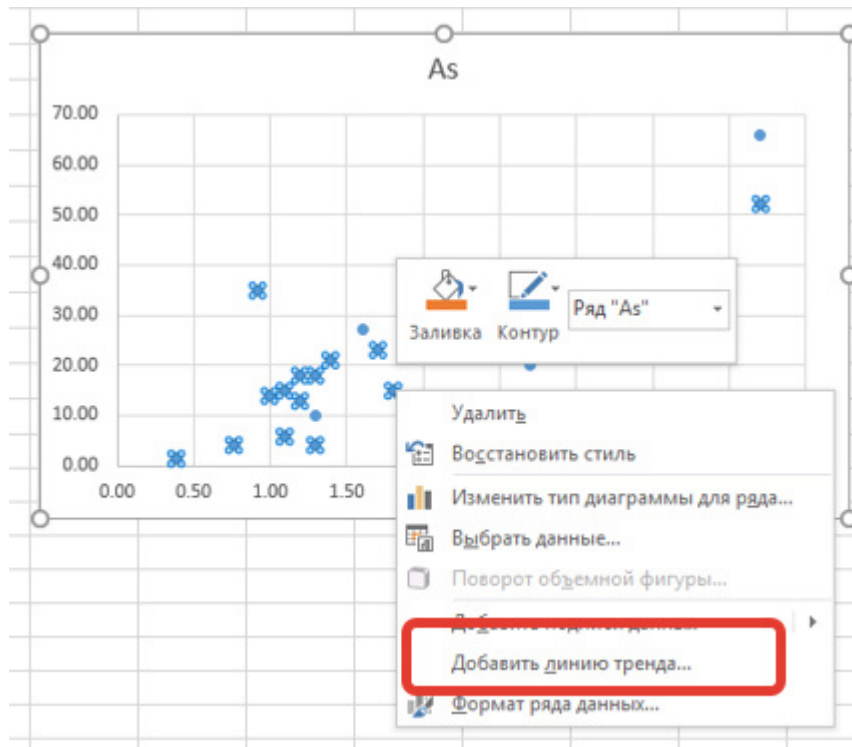
РЕГРЕССИОННЫЙ АНАЛИЗ

КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Скопируем столбцы со значениями концентраций Fe и As на отдельный лист.

Строим корреляционное поле: «Вставка» - «Диаграмма»
Точечная диаграмма» (дает сравнивать пары).

Диапазон значений – Столбцы Fe и As.



Левой кнопкой выделим точку на диаграмме.

Правой – откроем контекстное меню и выбираем – Добавить линию тренда
Выбираем параметры линии (в данном случае – линейная)

Включаем – Показать уравнение регрессии

РЕГРЕССИОННЫЙ АНАЛИЗ

Линия тренда

