

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

ФГБОУ ВО «Уральский государственный экономический университет»

В.П. Часовских

Технологии обработки больших данных

Методы регрессионного анализа в среде EXCEL

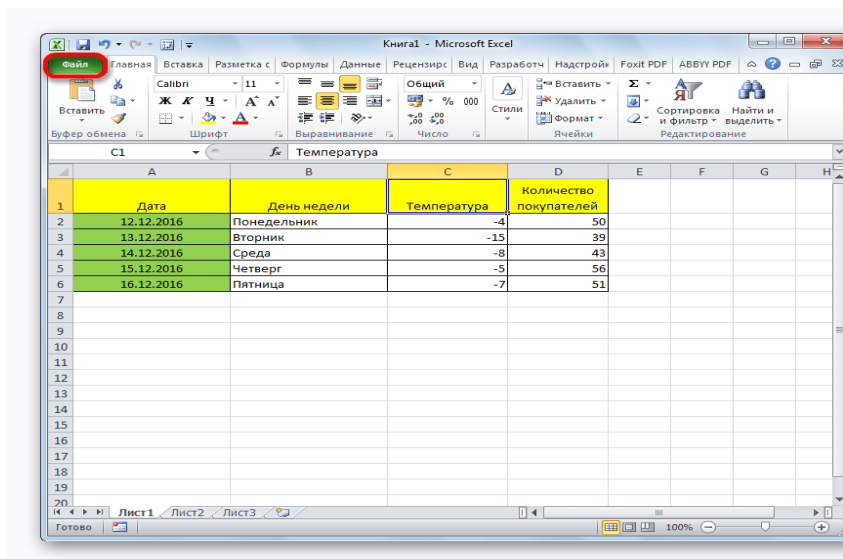
Екатеринбург 2022

Регрессионный анализ является одним из самых востребованных методов статистического исследования. С его помощью можно установить степень влияния независимых величин на зависимую переменную. В функционале Microsoft Excel имеются инструменты, предназначенные для проведения подобного вида анализа.

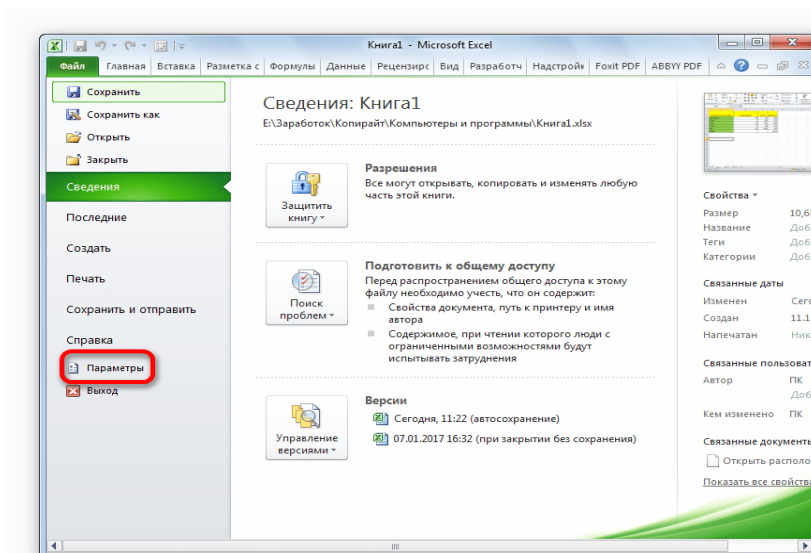
Подключение пакета анализа EXCEL

Для того, чтобы использовать функцию, позволяющую провести регрессионный анализ, прежде всего, нужно активировать Пакет анализа. Запускаем пакет EXCEL.

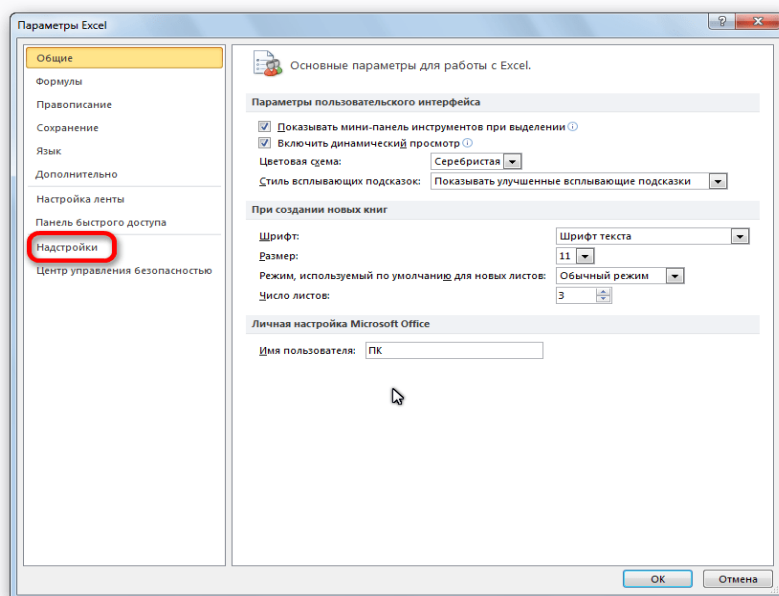
1. Перемещаемся во вкладку «Файл».



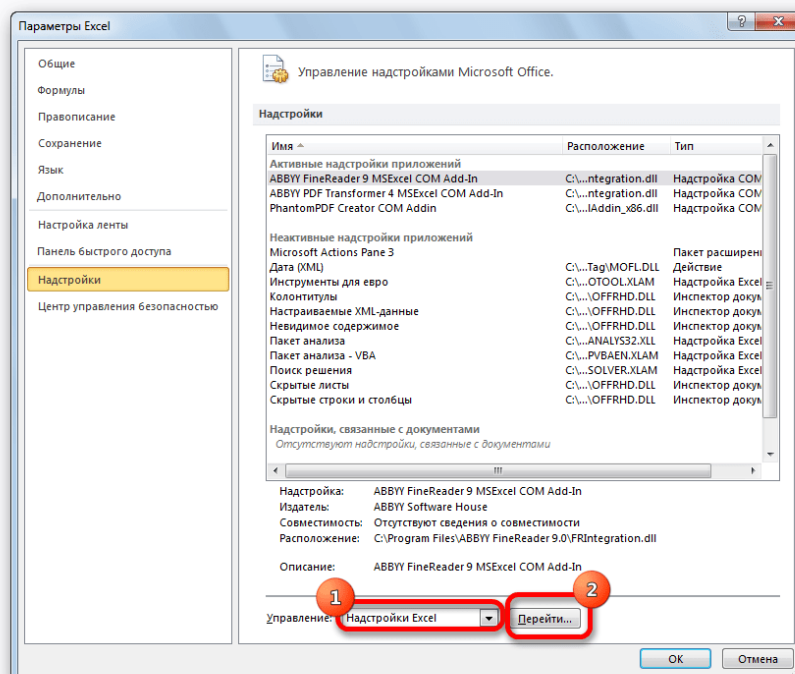
2. Переходим в раздел «Параметры».



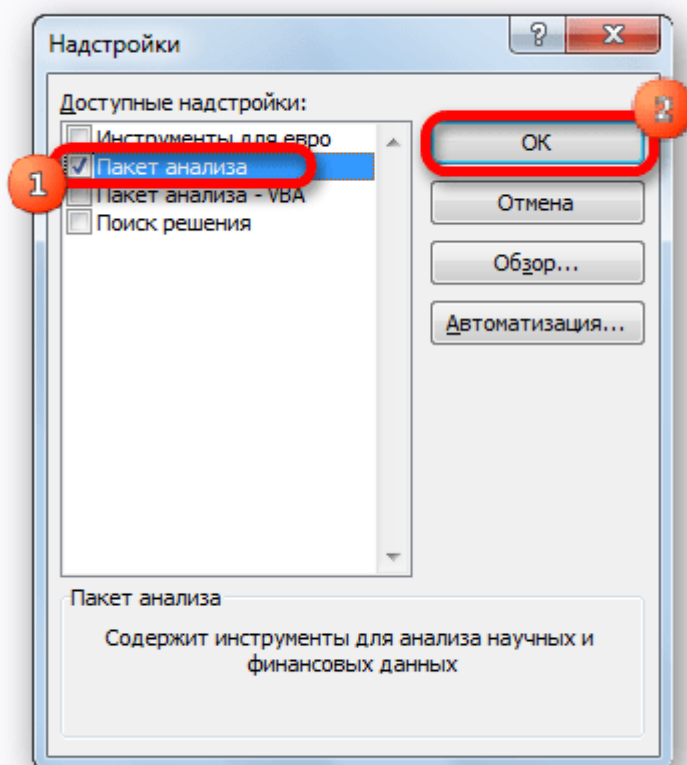
3. Открывается окно параметров Excel. Переходим в подраздел «Надстройки».



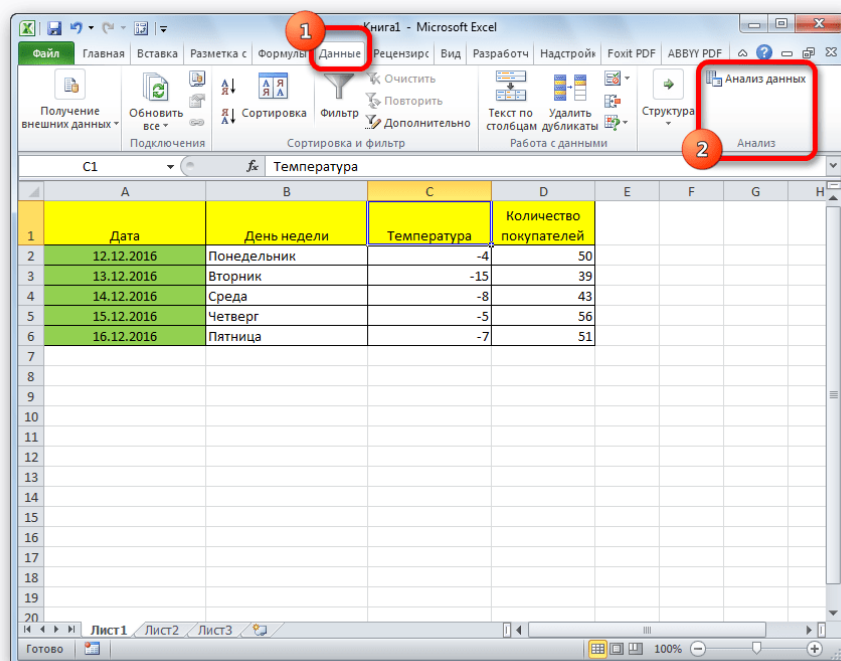
4. В самой нижней части открывшегося окна переставляем переключатель в блоке «Управление» в позицию «Надстройки Excel», если он находится в другом положении. Жмем на кнопку «Перейти».



5. Открывается окно доступных надстроек Excel. Ставим галочку около пункта «**Пакет анализа**». Жмем на кнопку «**ОК**».



Теперь, когда мы перейдем во вкладку «**Данные**», на ленте в блоке инструментов «**Анализ**» мы увидим новую кнопку – «**Анализ данных**».



Виды регрессионного анализа

Существует несколько видов регрессий:

- параболическая;
- степенная;
- логарифмическая;
- экспоненциальная;
- показательная;
- гиперболическая;
- линейная регрессия.

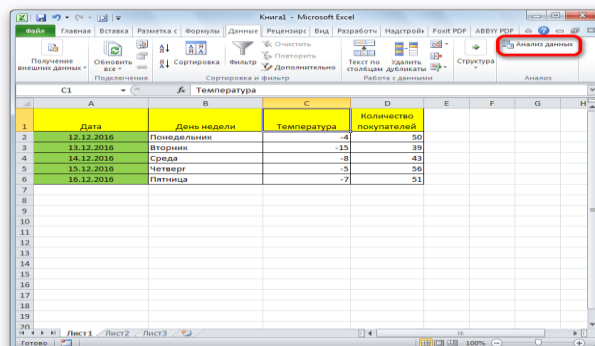
О выполнении последнего вида регрессионного анализа в Экселе мы подробнее поговорим далее.

Линейная регрессия в программе Excel

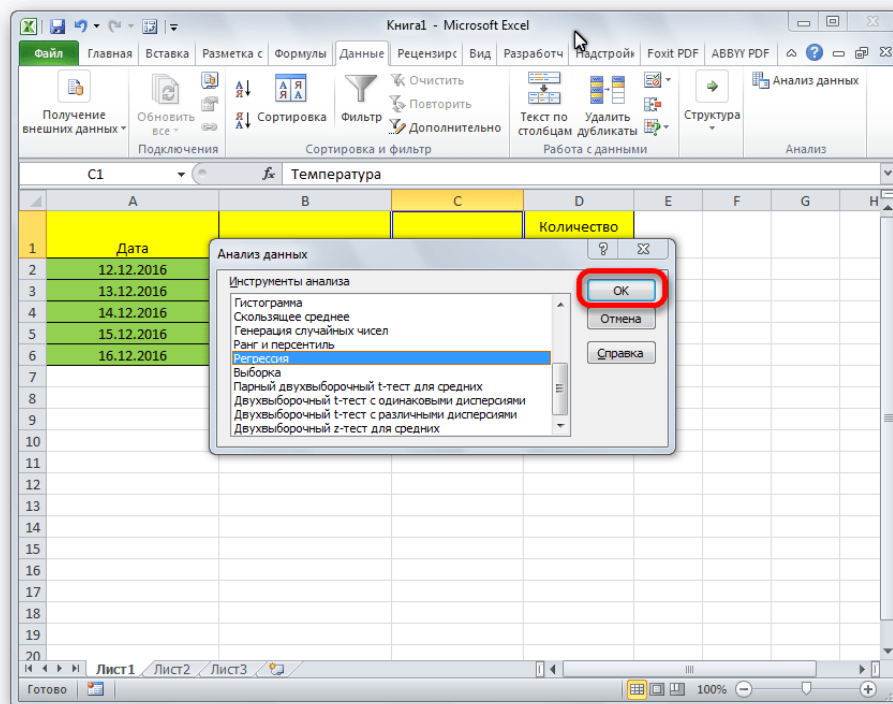
В качестве примера, представлена таблица, в которой указана среднесуточная температура воздуха на улице, и количество покупателей магазина за соответствующий рабочий день. Давайте выясним при помощи регрессионного анализа, как именно погодные условия в виде температуры воздуха могут повлиять на посещаемость торгового заведения.

Общее уравнение регрессии линейного вида выглядит следующим образом: $y = a_0 + a_1 * x_1 + \dots + x_k$ В этой формуле Y означает переменную, влияние факторов на которую мы пытаемся изучить. В нашем случае, это количество покупателей. Значение x – это различные факторы, влияющие на переменную. Параметры a являются коэффициентами регрессии. То есть, именно они определяют значимость того или иного фактора. Индекс k обозначает общее количество этих самых факторов.

1. Кликаем по кнопке **«Анализ данных»**. Она размещена во вкладке **«Главная»** в блоке инструментов **«Анализ»**.



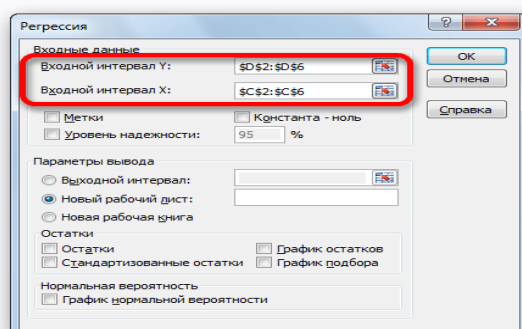
2. Открывается небольшое окошко. В нём выбираем пункт **«Регрессия»**. Жмем на кнопку **«ОК»**.



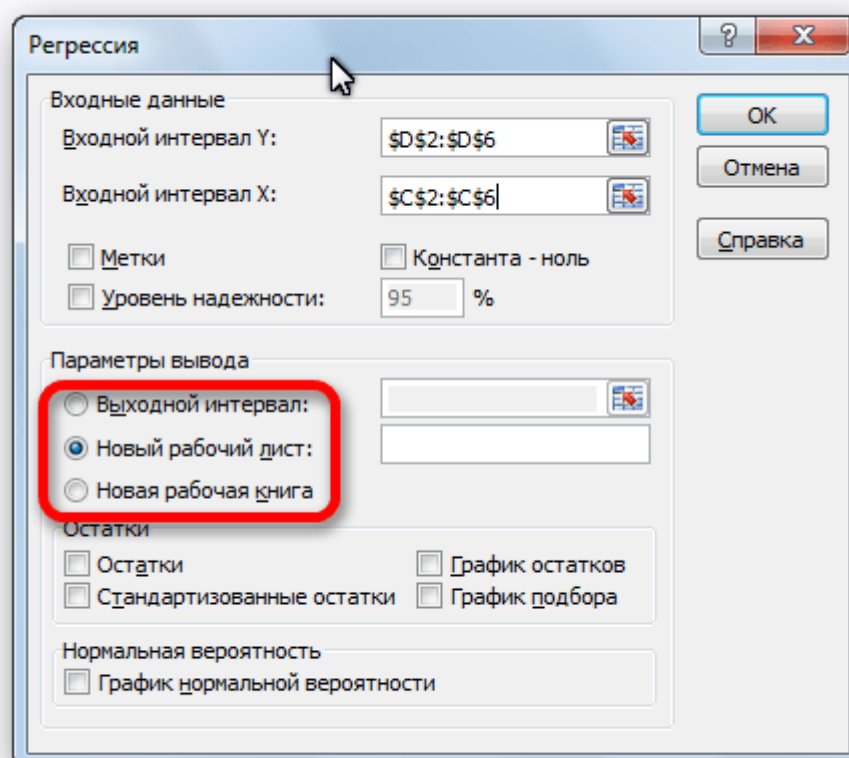
3. Открывается окно настроек регрессии. В нём обязательными для заполнения полями являются **«Входной интервал Y»** и **«Входной интервал X»**. Все остальные настройки можно оставить по умолчанию.

В поле **«Входной интервал Y»** указываем адрес диапазона ячеек, где расположены переменные данные, влияние факторов на которые мы пытаемся установить. В нашем случае это будут ячейки столбца «Количество покупателей». Адрес можно вписать вручную с клавиатуры, а можно, просто выделить требуемый столбец. Последний вариант намного проще и удобнее.

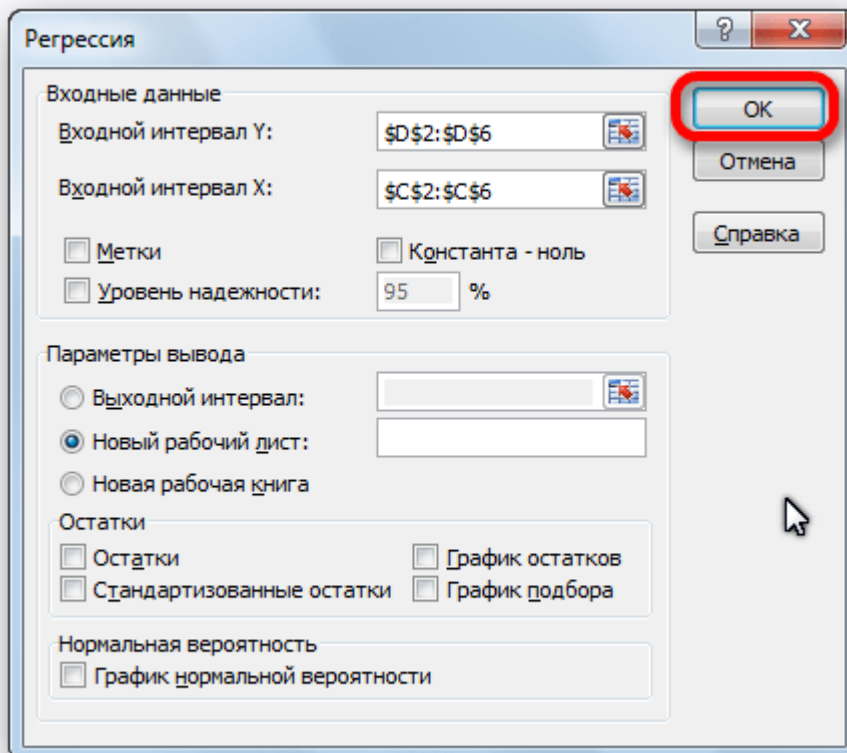
В поле **«Входной интервал X»** вводим адрес диапазона ячеек, где находятся данные того фактора, влияние которого на переменную мы хотим установить. Как говорилось выше, нам нужно установить влияние температуры на количество покупателей магазина, а поэтому вводим адрес ячеек в столбце «Температура». Это можно сделать теми же способами, что и в поле «Количество покупателей».



С помощью других настроек можно установить метки, уровень надёжности, константу-ноль, отобразить график нормальной вероятности, и выполнить другие действия. Но, в большинстве случаев, эти настройки изменять не нужно. Единственное на что следует обратить внимание, так это на параметры вывода. По умолчанию вывод результатов анализа осуществляется на другом листе, но переставив переключатель, вы можете установить вывод в указанном диапазоне на том же листе, где расположена таблица с исходными данными, или в отдельной книге, то есть в новом файле.



После того, как все настройки установлены, жмем на кнопку «**ОК**».



Результаты анализа

Результаты регрессионного анализа выводятся в виде таблицы в том месте, которое указано в настройках.

Регрессионная статистика						
Множественный R	0,839793663					
R-квадрат	0,705253396					
Нормированный R-кв	0,607004528					
Стандартная ошибка	4,237911402					
Наблюдения	5					

Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	1	128,9203209	128,9203209	7,178234331	0,075098537	
Остаток	3	53,87967914	17,95989305			
Итого	4	182,8				

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	58,04010695	4,266145634	13,60481145	0,000859033	44,46332754	71,61688635
Переменная X 1	1,312834225	0,490005635	2,67922271	0,075098537	-0,246582397	2,872297049

Одним из основных показателей является **R-квадрат**. В нем указывается качество модели. В нашем случае данный коэффициент равен 0,705 или около 70,5%. Это приемлемый уровень качества. Зависимость менее 0,5 является плохой.

Ещё один важный показатель расположен в ячейке на пересечении строки «**Y-пересечение**» и столбца «**Коэффициенты**». Тут указывается какое значение будет у Y, а в нашем случае, это количество покупателей, при всех остальных факторах равных нулю. В этой таблице данное значение равно 58,04.

Значение на пересечении граф «**Переменная X1**» и «**Коэффициенты**» показывает уровень зависимости Y от X. В нашем случае — это уровень зависимости количества клиентов магазина от температуры. Коэффициент 1,31 считается довольно высоким показателем влияния.

Как видим, с помощью программы Microsoft Excel довольно просто составить таблицу регрессионного анализа.