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.

Перемещаемся во вкладку «Файл».

🗶 🛃 🤊 • (** - 🗊 =	ł	Книга1 - Microsoft Exc	el			×
Файл Главная Вставка Ра	зметка с Формулы Данные	Рецензирс Вид Раз	работч Надстройк	Foxit PDF ABBYY PI	DF 🗠 🕜 🗆 🗗	E3 6
Вставить вставить Буфер обмена г≤ Шрифт	 11 ● ● ● ● ▲、、 ● ● ● ● ● ●	Общий • • • % 000 • * % 000 • * * % • Число Б	▲ Вставить ▼	∑ • ЯТ 	Найти и выделить т ание	
c1 + (e)	<i>J</i> _* Температура					¥
A	B	C	D	E F	G	H
1	0	T	Количество			
2 12 12 2016	Понедельник	-4	бокупателей			
3 13.12.2016	Вторник	-15	39			
4 14.12.2016	Среда	-8	43			
5 15.12.2016	Четверг	-5	56			
6 16.12.2016	Пятница	-7	51			_
7						_
8						
10						
11						
12						
13						
14						_
15						_
16						_
17						_
19						
20						-
Готово	Лист3 / 😏 /					
						· .::

2. Переходим в раздел «Параметры».

 Сохранить Сохранить как Открыть 	Сведения: Е:\Заработок\Коп	Книга1 играй\Компьютеры и программы\Книга1.xtxx		1
📑 Закрыть Сведения	20000000	Разрешения Все могут открывать, копировать и изменять любую часть этой книги.	Свойства т	-
Последние	книгу т		Размер Название	
Создать		Подготовить к общему доступу	Категории	
Сохранить и отправить	Поиск проблем т	 перед распространением общего доступа к этому файлу необходимо учесть, что он содержит: Свойства документа, путь к принтеру и имя автора 	Связанные дат Изменен Создан	ы
Справка		 Содержимое, при чтении которого люди с ограниченными возможностями будут испытывать затруднения 	Напечатан	162
Параметры			Автор	
выход		Версии 🔊 Сегодня, 11:22 (автосохранение)	Кем изменено	
	Управление версиями *	🔊 07.01.2017 16:32 (при закрытии без сохранения)	Связанные док	ум
			Открыть ра	300
			Показать все се	301

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

Іараметры Ехсеі		? <mark>×</mark>
Общие Формулы	🔯 Основные параметры для работы с Excel.	
Правописание	Параметры пользовательского интерфейса	
Сохранение Язык Дополнительно	 ✓ Показывать мини-панель инструментов при выделении () ✓ Включить динамический просмотр () Цветовая судма: Серебристая ▼ 	
Настройка ленты	Стиле всплывающих подсказок. Показывать улучшенные всплывающие подсказки При создании новых книг	
Надстройки Центр управления безопасностью	Щрифт текста <u>Размер:</u> 11 — Режим, используемый по умолчани <u>к</u> для новых листов: Эколо листов: <u>3</u> —	•
	Личная настройка Microsoft Office	
	ОК	Отмена

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»**.

Общие	Управление налотройками Microsoft Off	ice	
Формулы	эправление надстроиками містозон он	ice.	
Правописание	Надстройки		
Сохранение	Имя 🔺	Расположение	Тип
_	Активные налстройки приложений		[
Язык	ABBYY EineReader 9 MSExcel COM Add-In	C:\ntegration.dll	Налстройка СОМ
0	ABBYY PDF Transformer 4 MSExcel COM Add-In	C:\ ntegration dll	Налстройка СОМ
дополнительно	PhantomPDE Creator COM Addin	C) Addin x86 dll	Наястройка СОМ
Hactroukea sente	Thankom briereator com Addin	crimina a a million a a million a mi	nagerporkd Coli
пастроика ленты	Неактивные налотройки приложений		
Панель быстрого доступа	Microsoft Actions Pane 3		Пакет расширени
	/Jata (XMI)	C:\Tag\MOELDU	Лействие
Надстройки	Инструменты для евро	CI\ OTOOL XIAM	Наястройка Evcel
	Колонтитулы	CIV VOEEBHD DU	Инспектор докум
Центр управления безопасностью	Настраиваемые УМ ванные	CI. JOFFRHD DU	Инспектор докул
	Настраиваетые холоржинов		Инспектор докул
	Певидимое содержимое		Инспектор докук
		CI\AINALTSSZIALL	Надстройка Ехсег
	Пакет анализа - УБА	C:\PVBAEN.XLAM	надстройка Ехсеі
	Поиск решения	CI\SOLVER.XLAM	падстроика Ехсег
	Скрытые листы	C:\\OFFRHD.DLL	инспектор докум
	Скрытые строки и стольцы	C:\\OFFRHD.DLL	инспектор докум
	падстроики, связанные с документами		
	Отсутствуют наостроаха, связанные с оокумента	imu	
	< III		•
	Надстроика: ABBYY FineReader 9 MSExcel CO	M Add-In	
	Издатель: ABBYY Software House		
	Совместимость: Отсутствуют сведения о совмес	тимости	
	Расположение: C:\Program Files\ABBYY FineRead	der 9.0\FRIntegration.dll	
	Описание: ABBYY FineReader 9 MSExcel CO	M Add-In	
		2	
	Управление: Надстройки Ехсе!	рейти	
	The second		

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

<u>Д</u> оступные надстройки:		
Инструменты для евро	~	ок
Пакет анализа Пакет анализа - VBA		Отмена
		Об <u>з</u> ор
		Автоматизация
	-	
Пакет анализа		
Содержит инструменты, финансовь	для ан іх дані	ализа научных и ных

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

X	🚽 🍠 🕶 (🗎 👻 🖬 😾	1-	Книга1 - Microsoft E	xcel			- 0	×
Фа	йл Главная Вставка В	Разметка с Формулы Данны	Рецензирс Вид Р	азработч Надстрой	Foxit PDF	ABBYY PDF	⇔ 🕜 ⊏	6 B
Г внец	Солучение риих данных т Подключен	АЦ АЛ АЦ Сортировка Фильтр ия Сортировка	К Очистить Б Повторить У Дополнительно и фильтр	Текст по Удалить столбцам дубликат Работа с данны	Щарана Парадана Парадана Парадана Парадана Стр МИ	▶ □	Анализ данн Анализ	ых
	C1 • 🤇	<i>f</i> _x Температура						
1	А	В	С	D	E	F	G	H
1	Дата	День недели	Температура	Количество покупателей				
2	12.12.2016	Понедельник		-4 50				
3	13.12.2016	Вторник	-1	15 39				
4	14.12.2016	Среда		-8 43				
5	15.12.2016	Четверг		-5 56				
6	16.12.2016	Пятница		-7 51				
7								
8								
9								
10								
12								
13								
14								
15								
16								
17								
18								
19								
20	b. Ы. Пист1 Пист2 (Bucz2 / \$1						•
		/mero / w						- 1

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 · · · · ·	- DI 📼				Книга1 - Мі	crosoft l	Excel					100 X
Фаí	йл Главная	Вставка Р	азметка с	Формулы	Данные	Рецензирс	Вид	Разработ	ч Надстрой	Foxit	PDF ABBY		- 57
п	олучение них данных *	Обновить все т	월 종1 1 전 1 전	<u>я</u>	Фильтр	🛠 Очистить 🚡 Повтори 妤 Дополни	, ть тельно	Текст г	і Удалить ам дубликаті	- 100 - 100 - 100		Ц Анализ	данных
		Подключен	1/1	Copt	ировка и	фильтр		Pa	бота с данны	ми		Анал	43
	C1		fx	Темпе	ратура								
		A		в		c			D	E	F	G	H
								Ko	личество				
1	д	ата	Д	ень неде	ли	Темпер	оатура	пок	упателей				
2	12.12	.2016	Понеде	льник				-4	50				
3	13.12	2016	Вторния	к				15	39				
4	14.12	2.2016	Среда					-8	43				
5	15.12	2.2016	Четверг					-5	56				
•	10.12	.2016	плтница	3				-7	51				
2													
9													
10													
11													
12													
13													
14													
15													
16													
17													
10													
20													
4 4	н н Лист:	1 Лист2 /	Лист3	2/				14			18		▶ []

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

K 🖵	17 · (* ·					Книга1 - Міс	rosoft Ex	cel	N			_		23
Файл	Главная	Вставка	Разметка	с Формулы	Данные	Рецензирс К Очистить	Вид Ра	азработч	Надстрой в	Foxit P	DF ABBY	(PDF 🗠 🌍) — 🗟 данных	pΣ
Пол нешни	пучение их данных ≚	Обновить все т Подключе	ния	Сортировка Сорт	Фильтр тировка и	У Дополни фильтр	гельно	Текст по столбцая Рабо	Удалить и дубликать ота с данны	ии ии	Структура	Аналі	13	
	C1	- (0	<i>f</i> _x Темпе	ература									
1	A	4		В		C			D	E	F	G		Н
2	12.12 13.12 14.12 15.12 16.12	.2016 .2016 .2016 .2016 .2016 .2016	Анали <u>И</u> нст Сосо Ген Ран Ран Выб Пар Дву Дву Дву	з данных грументы анал тограмма льзящее сред ерация случай ерация случай горка ный двухвыбо гхвыборочный хвыборочный	иза нее іных чисел рочный t t-тест с ол t-тест с ра z-тест для	гест для сред динаковыми д азличными дии а средних	них исперсиям		В ОК Отмен	a a				
1 2 .3 .4 .5 .6 .7 .8														
	ы Лист1	. / Лист2	/Лист3	<u></u>				14			III 100% (•	. I

 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 Х»** 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».

Входные данные		ОК
Входной интервал Y:	\$D\$2:\$D\$6	
В <u>х</u> одной интервал X:	\$C\$2:\$C\$6	Отмена
<u>М</u> етки	Константа - ноль	<u>С</u> правка
Уровень надежности:	95 %	
Параметры вывода		
💿 Выходной интервал:	1	
Новый рабочий дист:		
💿 Новая рабочая книга		
Остатки		
Остатки	рафик остатков	
Стандартизованные ост	атки 🔄 График <u>п</u> одбора	
Нормальная вероятность		
П График нормальной веро	ятности	

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.

Входные данные		ОК
<u>В</u> ходной интервал Y:	\$D\$2:\$D\$6	
В <u>х</u> одной интервал X:	\$C\$2:\$C\$6	Отмена
Метки	Константа - ноль	<u>С</u> правка
Уровень надежности:	95 %	
 Выходной интервал: Новый рабочий <u>л</u>ист: Новая рабочая <u>к</u>нига Остатки 		
Остатки Стандартизованные ост	[] График остатков татки]] График подбора	
Нормальная вероятность	оятности	

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

Входные данные		ОК
<u>В</u> ходной интервал Y:	\$D\$2:\$D\$6	070002
В <u>х</u> одной интервал X:	\$C\$2:\$C\$6	Отмена
<u>М</u> етки	Константа - ноль	<u>С</u> правка
Уровень надежности:	95 %	
Параметры вывода		
Выходной интервал:	.	
Новый рабочий <u>л</u> ист:		
🔘 Новая рабочая <u>к</u> нига		
Остатки		L3
Остатки	🔲 График остатков	
С <u>т</u> андартизованные ост	атки 🔲 График <u>п</u> одбора	
Нормальная вероятность		
График нормальной верс	ятности	

Результаты анализа

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

X 🛃 🌖 • (* - 🞲 =		Книга1 - Microsof	t Excel			• X
Файл Главная Вставка	Разметка с Формул	ы Данные Рецензирс Вид	Разработч Надстро	йк Foxit PDF A	BBYY PDF 🗠 🕜	- # 2
Получение внешних данных *		ка Фильтр У Очистить Фильтр У Дополнительн	 Текст по Удалит толбцам дублика Работа с данн 	ыми	ура	анных
E10 -	(fx					
A	B	С	D	E	F	G
1 ВЫВОД ИТОГОВ		_	-			
2						
3 Регрессионная	статистика					
Множественный R	0,839793663					
R-квадрат	0,705253396					
6 Нормированный R-к	в 0,607004528					
7 Стандартная ошибка	4,237911402					
8 Наблюдения	5					
9						
10 Дисперсионный ана	лиз				<u></u>	
11	df	SS	MS	F	Значимость F	
12 Регрессия	1	128,9203209	128,9203209	7,178234331	0,075098537	
13 Остаток	3	53,87967914	17,95989305			
14 11000	4	182,8				
15	Kodobuuuaumu		t.cmamucmuwa	D_2µqµqµµq	Hunkung 95%	Repyuu
17 У-пересечение	58.04010695	4.266145634	13.60481145	0.000859033	44.46332754	71.616
18 Переменная Х 1	1.312834225	0,490005635	2,67922271	0.075098537	-0.246582397	2.8722
19	-,	-,	_,••••	-,	-,	
20						
21		18-2				
Готоро	. / Лист2 / Лист3 /					
					• • • •	Ð

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 **«Ү-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

Х1» 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.