

ОСНОВЫ СТАТИСТИКИ

Реализация модели множественной регрессии в MS Excel. Спецификация переменных и проблема мультиколлинеарности. Определение параметров модели



В этой лекции мы рассмотрим ключевые вопросы по реализации модели множественной регрессии в MS Excel.

Рассмотрим инструмент, позволяющий рассчитывать матрицу корреляции и ковариации, что необходимо, для того чтобы исключить интеркоррелированные факторы.

Прежде чем рассматривать инструмент, вспомним, что же такое уравнение множественной регрессии. Множественная регрессия широко используется в решении многих проблем социологии, экономики, психологии и других направлений. В настоящее время множественная регрессия – один из наиболее распространенных методов статистического анализа. Основной целью множественной регрессии является построение модели с большим числом факторов, а также определение влияния каждого фактора в отдельности и совокупного их воздействия на моделируемый показатель.

Множественный регрессионный анализ является развитием парного регрессионного анализа в случаях, когда зависимая переменная связана более чем с одной независимой переменной. Большая часть анализа является непосредственным расширением парной регрессионной модели, но здесь также появляются и некоторые новые проблемы, из которых следует выделить две.

Первая проблема касается исследования влияния конкретной независимой переменной на зависимую переменную, а также разграничения её воздействия и воздействий других независимых переменных.

Второй важной проблемой является спецификация модели, которая состоит в том, что необходимо ответить на вопрос, какие факторы следует включить в регрессию, а какие – исключить из неё. Прежде чем рассматривать модель множественной регрессии необходимо определить при помощи инструмента Корреляция те факторы, которые действительно важны в исследуемой проблеме. Если на Вашем компьютере нет или сбились настройки «Анализа данных» необходимо их установить, по тому же принципу, что рассматривали в предыдущих лекциях. Хотелось бы в дополнение рассмотреть плюсы и минусы пакета «Анализа данных». Для построения базовых моделей регрессионного анализа хорошо подходит Microsoft Excel.

Однако в пакете Excel не заложено большинство тестов или продвинутых моделей, при необходимости пользователь может сам прописать необходимые процедуры используя VBA.

Плюсы данного пакета:

1. Распространенность. В настоящее время на подавляющем количестве компьютеров стоит это программное обеспечение.
2. Кросс-платформенность. Существуют версии Excel как для ОС Windows, так и аналоги для Mac и Unix/Linux (для последних операционных систем эта программа называется Open Office).
3. Интуитивность и простота. Не нужно обладать особыми знаниями в программировании, чтобы уметь работать с этим пакетом.
4. Стабильность работы и малое количество ошибок в коде (особенно по сравнению со свободно распространяемым программным обеспечением).
5. Возможность программирования и создания пользовательских функций.

Недостатки:

1. Платное распространение. Чтобы пользоваться этим пакетом, необходимо купить лицензию.
2. Малое количество встроенных статистических функций. Пакет Excel имеет большое количество встроенных функций, но поскольку статистический анализ не является специализацией данного пакета, то количество встроенных статистических моделей не велико.

Итак, алгоритм построения регрессионной модели с помощью корреляционно-регрессионного анализа средствами табличного процессора Microsoft Excel можно разбить на следующие этапы:

1. Сбор статистических данных по исследуемым параметрам.
2. Определение причинно-следственной связи между анализируемыми параметрами.
3. Оценка тесноты связи с помощью коэффициента корреляции r_{xy} .

В табличном процессоре MS Excel для целей корреляционного анализа служит инструмент Корреляция, который позволяет получить корреляционную матрицу, содержащую коэффициенты корреляции между параметрами.

4. Расчет параметров уравнения регрессии.



Вычисление коэффициентов при независимых переменных уравнения регрессии осуществляется с помощью метода наименьших квадратов.

5. Статистическое прогнозирование возможных значений параметра y при заданных значениях параметра x по полученной модели.

Корреляция.

Рассматриваемые инструменты размещены в вкладке «Данные» основного меню запускаем «Анализ данных» и выбираем необходимый инструмент, в нашем случае это корреляция. Корреляция используется для количественной оценки взаимосвязи двух наборов данных, представленных в безразмерном виде. Корреляционный анализ дает возможность установить, ассоциированы ли наборы данных по величине, то есть, большие значения из одного набора данных связаны с большими значениями другого набора (положительная корреляция), или, наоборот, малые значения одного набора связаны с большими значениями другого (отрицательная корреляция), или данные двух диапазонов никак не связаны (корреляция близка к нулю).

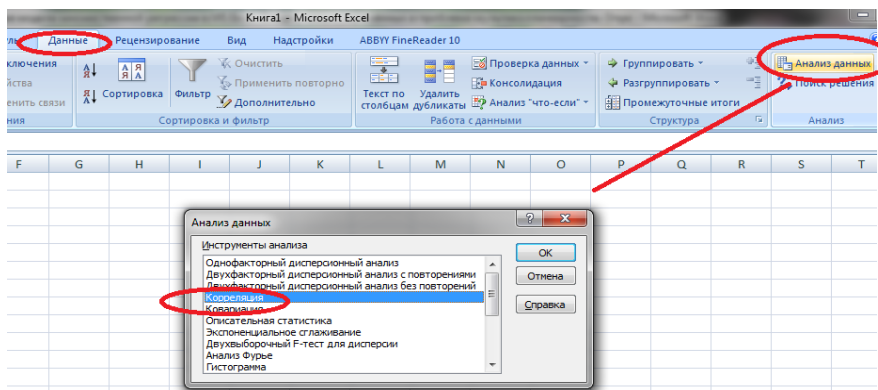


Рисунок 24.1 Диалоговое окно Пакета «Анализ данных» «Корреляция».

Указывается Входной интервал – ссылка на диапазон, содержащий анализируемые данные. Ссылка должна состоять как минимум из двух смежных диапазонов данных, организованных в виде столбцов или строк. Если вы активизируете опцию «Метки в первой строке», то можно работать с заголовками. Аналогично и для инструмента «Ковариация».

Прежде чем, начнем рассмотрение поставленной перед нами задачей вспомним ключевые моменты регрессионного анализа в MS Excel, т.е. с диалогового окна вкладки Регрессия, данное окно вы сейчас видите на экране.

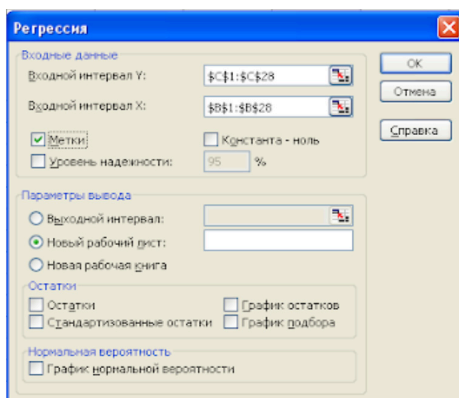


Рисунок 24.2. Диалоговое окно «Регрессии»

Диалоговое окно имеет множество дополнительных переключателей, которые приводят к выводу большого количества дополнительной информации.



Основные параметры, которые необходимо задать – это Входной интервал Y и Входной интервал X, а также Параметры вывода. Если количество данных Y и X совпадает, то выдаются итоги построения модели парной регрессии (именно этот случай будем сейчас рассматривать), а если число переменных X в несколько раз больше числа Y, то – модель множественной регрессии. В противном случае будет выдано сообщение об ошибке.

Если активизировать переключатель Метки, то во входные интервалы для X и Y можно добавить ячейки с названиями, и соответствующие метки появятся в итоговой таблице, что значительно облегчит её понимание.

Если Входной интервал Y определить как C1:C16, а Входной интервал X — B1:B16, задать некоторым образом параметры вывода, а следующая таблица.

Таблица 24.1 Таблица исходных данных (условные данные)

Номер наблюдения	Общая численность населения (млн.)	ВВП региона (ед)	Количество заключенных браков в регионе	Численность экономически активного населения
1	15,09	12,85	5,09	125,1779
2	15,21	12,26	5,03	123,8094
3	15,28	13,42	4,80	121,175
4	15,49	12,05	4,95	116,9143
5	15,54	12,70	4,88	119,8643
6	15,62	12,41	4,96	118,0681
7	15,70	13,83	5,10	123,5887
8	15,91	13,10	4,90	117,0877
9	15,92	13,07	4,72	116,1699
10	15,95	12,40	4,81	118,3436
11	16,31	12,82	4,95	116,2008
12	16,33	12,48	4,88	111,4565
13	16,60	13,20	5,02	115,1026
14	16,69	12,40	4,80	110,1056
15	16,76	12,01	4,85	110,0231

Таблица 24.2 Итоговая таблица.

ВЫВОД ИТОГОВ	
Регрессионная статистика	
Множественный R	0,887036
R-квадрат	0,786833
Нормированный R-квадрат	0,770435
Стандартная ошибка	2,264609
Наблюдения	15

Дисперсионный анализ					
	df	SS	MS	F	Значимость
Регрессия	1	246,0889	246,0889	47,985	F
Остаток	13	66,66991	5,128455		1,04E-05



Итого	14	312,7588			
-------	----	----------	--	--	--

	Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95 %	Верхние 95 %
Y-пересечение	240,142	17,70861	13,56075	4,76E-09	201,8849	278,3991
Экономически активное население	-7,71453	1,113671	-6,92712	1,04E-05	-10,1205	-5,30859

Эта таблица содержит большое количество информации, поэтому будем изучать её содержимое постепенно, в нескольких последующих работах.

Представленные в этой таблице данные можно условно разделить на три раздела:

- регрессионная статистика
- дисперсионный анализ
- коэффициенты.

Весь раздел регрессионная статистика посвящен описанию коэффициента детерминации и его различным характеристикам. В пунктах множественный R и R-квадрат выводится значение коэффициента детерминации и его квадрата. Кроме этого выдается общее количество наблюдений.

Рассмотрим раздел дисперсионный анализ.

В столбце SS выдаются все виды сумм квадратов отклонений. В данном случае в первой строке, которая соответствует надписи Регрессия, выдается объясненная сумма квадратов отклонений RSS.

В строке – Остаток – выдается необъясненная (остаточная) сумма квадратов отклонений ESS.

В строке – Итого – выдается общая сумма квадратов отклонений TSS.

В последнем разделе, который не имеет названия, будет интерпретироваться как раздел – коэффициенты, содержится полная информация по коэффициентам.

В диалоговом окне Регрессия имеется целый раздел переключателей для получения дополнительной информации по остаткам.

В таблице получены результаты предсказанных значений и значения остатков отдельно для каждого наблюдения. Указав опции График подбора, График остатков и График нормального распределения можно получить множество дополнительной информации и некоторые диаграммы.

Рассмотрим ключевой вопрос.

Это спецификация переменных и проблемы мультиколлинеарности. Эффект мультиколлинеарности возникает тогда, когда коэффициент корреляции между регрессорами близок к единице, в то время как коэффициент корреляции между регрессором и зависимой переменной мал. Как правило, выделяют значение в 0,8.

Для анализа влияния мультиколлинеарности можно проанализировать значение корреляционной матрицы, которую можно получить, используя надстройку Корреляция.

Соответственно матрица корреляции может быть представлена следующим образом.

Таблица 24.3 Матрица корреляции

	Общая численность населения региона x^1	Общая численность населения региона x^1	ВВП региона x^2	Количество заключенных браков в регионе x^3	Численность экономичес и активного населения.



Общая численность населения региона x^1	1	1			
ВВП региона x^2	-0,17012	-0,17012			
Средний доход населения x^3	-0,33828	-0,33828	1		
Численность, экономически активного населения y	-0,88704	-0,88704	0,195127		
			0,423668	1	
				0,555581	
					1

Начнем рассмотрение результатов данной таблицы с последней строки, в которой находятся частные коэффициенты корреляции зависимой переменной y (экономически активное население рассматриваемого региона) и регрессоров X_1 -общая численность населения региона, X_2 – ВВП региона, X_3 – количество заключенных браков в регионе). Имеется некоторая взаимосвязь между каждым регрессором и численностью населения, при этом максимальное значение коэффициента корреляции равно $-0,887$ и говорит о существенной связи экономически активного населения от общей численности населения. Именно эта парная регрессионная модель строилась в предыдущей лекции. Все остальные коэффициенты корреляции значительно меньше, поэтому нет оснований утверждать, что присутствует эффект мультиколлинеарности, однако в некоторой незначительной степени этот эффект имеет место.

Если было подтверждено наличие эффекта мультиколлинеарности, то один из возможных способов её устранения либо в укрупнении регрессоров, либо в их исключении.

На основании полученных коэффициентов частной корреляции нетрудно самостоятельно подсчитать значение коэффициента детерминации R^2 , или, как его ещё иногда называют, множественного коэффициента корреляции.

Рассмотрим теперь более подробно тот факт, как и почему изменяется значение коэффициентов регрессии в зависимости от того, какая модель рассматривается. Ещё раз напомним, что в парном случае модель имела вид:

$$y = 239,96 - 7,703x^1,$$

А во множественном случае лучшая модель будет иметь вид:

$$y = 142,21 - 6,61 x^1 + 2,24 x^2 + 10,56 x^3.$$

Это модель с отсутствующей переменной, когда вместо двух, реально присутствующих в модели регрессоров, рассматривается парный случай, и модель с лишней переменной, когда исходная модель является парной, а она рассматривается как множественная. В случае лишней переменной происходит только потеря эффективности, в случае отсутствующей происходит нарушение наиболее важного свойства, а именно, нарушается несмещённость оценки. При этом показано, что математическое ожидание коэффициента в парном случае будет иметь вид:

Начнем рассмотрение результатов данной таблицы с последней строки, в которой находятся частные коэффициенты корреляции зависимой переменной y (экономически активное население рассматриваемого региона) и регрессоров X^1 -общая численность населения региона, X^2 – ВВП региона, X_3 -количество заключенных браков в регионе). Имеется некоторая взаимосвязь между каждым регрессором и численностью населения, при этом максимальное значение коэффициента корреляции равно $-0,887$ и говорит о существенной связи экономически активного населения от общей численности населения. Именно эта парная регрессионная модель строилась в предыдущей лекции. Все остальные коэффициенты корреляции значительно меньше, поэтому нет оснований утверждать, что присутствует эффект



мультиколлинеарности, однако в некоторой незначительной степени этот эффект имеет место.

Если было подтверждено наличие эффекта мультиколлинеарности, то один из возможных способов её устранения либо в укрупнении регрессоров, либо в их исключении.

На основании полученных коэффициентов частной корреляции нетрудно самостоятельно подсчитать значение коэффициента детерминации R^2 , или, как его ещё иногда называют, множественного коэффициента корреляции.

Рассмотрим теперь более подробно тот факт, как и почему изменяется значение коэффициентов регрессии в зависимости от того, какая модель рассматривается. Ещё раз напомним, что в парном случае модель имела вид:

$$y = 239,96 - 7,703x^1,$$

А во множественном случае лучшая модель будет иметь вид:

$$y = 142,21 - 6,61 x^1 + 2,24 x^2 + 10,56 x^3.$$

Это модель с отсутствующей переменной, когда вместо двух, реально присутствующих в модели регрессоров, рассматривается парный случай, и модель с лишней переменной, когда исходная модель является парной, а она рассматривается как множественная. В случае лишней переменной происходит только потеря эффективности, в случае отсутствующей происходит нарушение наиболее важного свойства, а именно, нарушается несмещённость оценки. При этом показано, что математическое ожидание коэффициента в парном случае будет иметь вид:

Формула 24.1

$$M[b_1] = \beta_1 + \beta_3 \cdot \frac{Cov_n(x^1, x^3)}{Var_n(x^1)} + \beta_4 \cdot \frac{Cov_n(x^1, x^4)}{Var_n(x^1)}$$

Для вычисления найдем ковариационную матрицу, которая представлена в таблице:

Таблица 24.4 Матрица корреляции.

	Общая численность населения региона x^1	ВВП региона x^2	Количество заключенных браков в регионе x^3	Численность экономически активного населения.
Общая численность населения региона x^1	0,275665			
ВВП региона x^2	-0,04468	0,250289		
Средний доход населения x^3	-0,01923	0,010569	0,011722	
Численность, экономически активного населения y	-2,12663	0,967845	0,274663	20,85059

Используя полученные данные, нетрудно вычислить, что полностью совпадает со значением, полученным в парном случае.

$$M[b_1] = -6,61 + 2,24 \cdot \frac{-0,04468}{0,2756} + 10,56 \cdot \frac{0,01923}{0,2756} = -7,71$$

Если сравнивать случаи с тремя регрессорами, которая была признана нами наилучшей, и общим случаем, то очевидно, что в общем случае стандартные ошибки коэффициентов больше, следовательно, оценки менее эффективные.

Автоматизация оценки фиктивных переменных. Для описания категорий необходимо ввести совокупность фиктивных переменных. До сих пор нами рассматривался только случай количественных регрессоров,



поскольку значение экономически активного населения и общая численность населения являются числами. Однако может возникнуть ситуация, когда необходимо учесть некоторую специфическую информацию. Рассматривая модель оценки экономически активного населения региона, можно предположить, среди общей численности имеются мужчины и женщины. Конечно, можно построить две различные модели, и посмотреть разницу между ними, однако нас будет интересовать общая модель. В этом случае в модель необходимо вносить качественный регрессор, для чего нужно использовать фиктивную переменную. Данная переменная может принимать только два значения 0 или 1, в зависимости от отсутствия или наличия нового качества. В этом случае можно строить модель с фиктивной переменной наклона и сдвига. Работа с фиктивными переменными ничем не отличается от построения регрессионной модели.

Поэтому рассмотрим задачу. Значение x_1 и y на две различные категории, которые мы условно назовем «мужской пол» и «женский», представлены следующей таблицей:

Таблица 24.5 Таблица данных

Номер наблюдения	Вид	Общая численность населения региона x_1	Численность, экономически активного населения y
1	мужской	15,09	125,1779
2	мужской	15,21	123,8094
3	женский	15,28	121,175
4	женский	15,49	116,9143
5	женский	15,54	119,8643
6	женский	15,62	118,0681
7	мужской	15,70	123,5887
8	мужской	15,91	117,0877
9	женский	15,92	116,1699
10	мужской	15,95	118,3436
11	мужской	16,31	116,2008
12	женский	16,33	111,4565
13	мужской	16,60	115,1026
14	женский	16,69	110,1056
15	женский	16,76	110,0231

В электронной таблице Excel имеются возможности для быстрого задания значений фиктивной переменной. Для этого необходимо вставить столбец между колонками с названиями «Экономически активное население» и «Общая численность населения региона». Фиктивная переменная, и для определения значений будем использовать логическую функцию ЕСЛИ. Данная функция имеет три аргумента. Первый – это логическое выражение, которое может принимать истинное или ложное значение. Вторым аргументом идет то значение, которое появляется в ячейке при истинности условия, а соответственно в третьем аргументе – значение, которое появляется в противном случае. выполнив данные действия, получим расчетную таблицу.

Таблица 24.6 Расчетная таблица.

Номер наблюдения	Вид	Фиктивная переменная	Численность населения	Экономически активное население
1	мужской	=ЕСЛИ(B2=»мужской»;1;0)	15,09 т.	125,1779

В столбце фиктивной переменной появится значение 1, если в предыдущем столбце находилось слово «мужчина», и 0 в противоположном случае. Теперь наша задача заключается в определении степени влияния фиктивной переменной. А именно, влияет ли это значение на свободный член (в этом случае при



изменении качества можно говорить о том, что экономическое активное население изменится на какое-то количество) или на наклон линии регрессии, или на оба эти значения сразу.

Вначале оценим регрессию, при условии, что фиктивная переменная влияет только на значение свободного члена. В этом случае итоговая таблица после выполнения надстройки Регрессии, будет иметь вид:

Таблица 24.7 Результаты анализа.

ВЫВОД ИТОГОВ	
Регрессионная статистика	
Множественный R	0,963696
R-квадрат	0,928711
Нормированный R-квадрат	0,916830
Стандартная Ошибка	1,363084
Наблюдения	15

Дисперсионный анализ

	df	SS	MS	F	Значимость F	
Регрессия	2	290,4628387	145,231419	78,16547142	1,31E-07	
Остаток	12	22,29599593	1,85799966			
Итого	14	312,7588347				
	Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95 %	Верхние 95 %
Y-пересечение	232,0028	10,78827	21,5051052	5,9691E-11	208,49	255,508
Фиктивная переменная	3,474500	0,7109700	4,8869856	0,00037407	1,9254	5,02357
Цена x(t.)	-7,30442	0,675558	-10,8124125	1,5303E-07	-8,77634	-5,83251

Регрессионная модель имеет вид: $y = 232 + 3,47D - 7,304x$

Поскольку значение фиктивной переменной D равно 1 для «мужчин» вида и 0 для «женщин», то данную модель можно отдельно расписать для каждого случая.

$y = 232 - 7,304x$ – женщины,

$y = 235,47 - 7,304x$ – мужчины.

До сих пор нами рассматривался случай, когда имеются всего два значения качества. Однако нередки случаи, когда необходимо проанализировать спрос для различных продуктов. Тогда необходимо вводить набор категорий, как конечный набор взаимоисключающих событий, полностью описывающий все возможности. Предположим, что исследуется влияние общей численности населения на численность экономически активного населения при следующих переменных оценивающих население «мужчины», «женщины», «безработные», «занятые».

В этом случае для описания этих категорий необходимо вводить набор фиктивных переменных по следующему правилу.

Число фиктивных переменных должно быть на единицу меньше, чем число категорий. В данном случае имеется четыре категории, следовательно, необходимо ввести три фиктивные переменные, которые мы обозначим D1, D2, D3.

Выбрать произвольную категорию в качестве эталонной. Именно с этой категорией, в последствии, будут сравниваться все остальные. Для эталонной категории необходимо, чтобы значения всех фиктивных переменных равнялись нулю.



Для всех остальных категорий необходимо, чтобы одна из фиктивных переменных равнялась 1, в то время как значение всех остальных равно 0.

Достаточно легко можно расставить значения фиктивных переменных, используя ту же условную функцию ЕСЛИ. При наличии четырёх различных видов продукции необходимо вставить три дополнительных столбца, в которых будут находиться фиктивные переменных. Задать логические функции можно так, как показано в таблице.

Таблица 24.8 Логические функции.

Номер наблюдения	Вид	Фиктивная переменная D1	Фиктивная переменная D2	Фиктивная переменная D3	Численность населения	Экономически активное населени
1		=ЕСЛИ(B2=«обычный»;1;0)	=ЕСЛИ(B2=«новой»;1;0)	=ЕСЛИ(B2=«самой новый»;1;0)	15,09	125

После копирования данных функций вниз для значения старой все фиктивные переменные будут равны нулю, для обычной – только значение первой фиктивной переменной будет равно 1 и т. д.

После этого можно вызвать надстройку Регрессия, у которой в качестве входного интервала X, необходимо указать значения всех фиктивных переменных D и нефиктивной переменной X, то есть задать Входной интервал X.

Полученные результаты поддаются достаточно простой интерпретации. Значение, находящееся напротив фиктивной переменной D1, показывает, насколько изменилась численность экономически активного населения рассматриваемого региона при переходе от эталонной к первой категории, то есть насколько различен результат численности экономически активного населения между мужчинами и женщинами рассматриваемого региона. Аналогично интерпретируются значения, стоящие напротив других фиктивных переменных.

Кроме инструмента Регрессия для получения оценочных показателей уравнения множественной регрессии можно использовать функцию ЛИНЕЙН.

Функция ЛИНЕЙН рассчитывает статистику для ряда с применением метода наименьших квадратов, вычисляя прямую линию, которая наилучшим образом аппроксимирует имеющиеся данные. Задача отыскания функциональной зависимости очень важна, поэтому для ее решения в MS Excel введен набор функций, основанных на методе наименьших квадратов. В качестве результата выдаются не только коэффициенты функции, приближающей данные, но и статистические характеристики полученных результатов.

Функция рабочего листа ЛИНЕЙН помогает определить характер линейной связи между результатами наблюдений и временем их фиксации и дать ей математическое описание, наилучшим образом аппроксимирующее исходные данные. Для построения трендовой модели она использует уравнение вида $y = mx + b$, где y — исследуемый показатель; $x = t$ — временной тренд; b, m — параметры уравнения, характеризующие соответственно y -пересечение и наклон линии тренда.

Вызвать функцию ЛИНЕЙН можно в диалоговом окне Мастера функций (категория «Статистические»), расположенном на панели инструментов Стандартная.

Используя метод наименьших квадратов, функция ЛИНЕЙН создает массив значений, который описывает искомую модель тренда. Учитывая, что создается массив значений, функция должна задаваться пользователем в виде формулы массива. Поэтому перед началом работы с ЛИНЕЙН необходимо на рабочем листе выделить диапазон ячеек, достаточный для размещения создаваемого ею массива значений. После того, как выделен выходной диапазон и пользователь определился с аргументами функции посредством диалогового окна ЛИНЕЙН, следует нажать на клавиатуре кнопки Ctrl+Shift+Enter.

Функция ЛИНЕЙН имеет четыре аргумента:

ЛИНЕЙН(известные_значения_y; [известные_значения_x]; [конст]; [статистика]).

1) Известные значения y – это множество уже известных значений исследуемого показателя, на основе которых будет производиться оценка параметров уравнения тренда.

2) Известные значения x – при построении трендовой модели представляют собой временной ряд, соответствующий по размерам первому аргументу.



3) Конст – логическое значение, которое указывает на необходимость расчета параметра b (свободного члена) при построении модели тренда. Если Конст имеет значение ИСТИНА, то параметр b вычисляется. Если Конст имеет значение ЛОЖЬ, то параметр b принимается равным нулю.

4) Статистика – логическое значение, которое указывает на необходимость отражения на рабочем листе дополнительной статистической информации, позволяющей судить о качестве построенной модели. Если этот аргумент имеет значение ЛОЖЬ или ссылка на него отсутствует, то функция ЛИНЕЙН не рассчитывает статистические характеристики. Если Статистика задана значением ИСТИНА, то массив, создаваемый функцией, содержит значения следующих статистических величин.

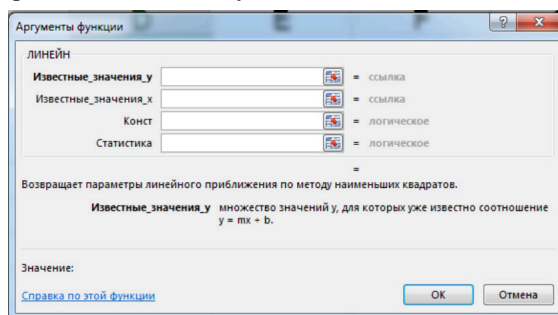


Рисунок 24.3 Диалоговое окно функции ЛИНЕЙН.

На выходе мы получим следующую таблицу:

Таблица 24.7 Значения статистических величин функции ЛИНЕЙН

Стандартная ошибка для параметра b	Стандартная ошибка для свободного члена a
Квадрат коэффициента корреляции (r^2)	Стандартная ошибки для y
F-критерий (F)	Степень свободы (df)
Сумма квадратов регрессии (SSp)	Остаточная сумма квадратов ($SS0$)

Т.е. мы получили, во-первых, параметры модели, затем стандартные ошибки. Значение F-критерия Фишера, сумму квадратов регрессии и остаточную сумму квадратов, что позволит нам дать оценку уравнения регрессии. Данную функцию можно использовать и для модели множественной регрессии, это вам будет необходимо рассмотреть самостоятельно.

Используя модель, которую мы рассмотрели в начале нашей лекции, мы видим, что результаты, полученные при помощи инструмента Регрессия, совпадают с результатами, полученными при помощи инструмента функции ЛИНЕЙН.

Подводя итоги можно сказать, что можно сделать полноценный корреляционно-регрессионный анализ модели множественной регрессии при помощи встроенных возможностей MS Excel, в частности используя надстройку «Анализ данных» – корреляция и регрессия и статистическую функцию ЛИНЕЙН, в последнем случае необходимо использование дополнительных функций, которые позволят рассчитать доверительные интервалы, значение критерия Стьюдента, остатки. Как и все имеет свои достоинства и недостатки, о которых мы уже говорили. Для закрепления материала, необходимо выполнить домашнее задание.