

УДК 004:37

**КОРРЕЛЯЦИОННЫЙ АНАЛИЗ В ЭЛЕКТРОННЫХ ТАБЛИЦАХ****Курзаева Л.В.***ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова»,  
Магнитогорск, e-mail: lkurzaeva@mail.ru*

Методы аналитической статистики – одни из самых сложных в плане изучения, однако, это одни из самых востребованных методов в арсенале аналитика. Анализ взаимосвязи признаков производится в рамках решения трех основных задач: описание и понимание взаимосвязи; прогнозирование и предсказание нового наблюдения; корректировка и управление процессом. Корреляционный анализ позволяет оценить степень взаимосвязи между переменными. В статье описывается простой способ проведения корреляционного анализа в Microsoft Excel на примере реальных статистических данных. Материалы данной статьи представляют методическую и практическую ценность для преподавателей, занимающихся вопросами повышения эффективности обучения в области основ анализа данных с информационных технологий, и осуществляющие реализацию образовательного процесса в вузах и на курсах повышения квалификации.

**Ключевые слова:** анализ данных, электронные таблицы**CORRELATION ANALYSIS IN SPREADSHEETS****Kurzaeva L.V.***Nosov Magnitogorsk State Technical University, Magnitogorsk, e-mail: lkurzaeva@mail.ru*

Methods of analytical statistics are some of the hardest in terms of studying, however, it is one of the most popular tools in the Arsenal of the analyst. Analysis of the relationship of signs is made in the solution of three main objectives: description and understanding of the relationship; forecasting and prediction of new observation; adjustment and control of the process. Correlation analysis allows to estimate the degree of relationship between the variables. This paper describes a simple method of correlation analysis in Microsoft Excel in example of real statistical data. The contents of this article are of methodological and practical value to teachers working to increase the effectiveness of training in the area of foundations of data analysis with information technology, and implementing the educational process in universities and training courses.

**Keywords:** data analysis, spreadsheets

Рассмотрим возможности использования пакета Анализ данных в Microsoft Excel при проведении корреляционного анализа.

Корреляция – это взаимосвязь количественных или порядковых признаков.

Интерпретация корреляции проводится на основании:

1) коэффициента корреляции ( $r$ ) и его квадрата – коэффициент детерминации ( $R^2$ ), которые свидетельствуют о силе связи;  $R^2$  представляет собой долю вариации, общую для двух переменных (иными словами, «степень» зависимости или связанности двух переменных);

2) уровня значимости, вычисленного для каждого коэффициента корреляции, позволяющего судить о надежности корреляции;

3) визуального анализа связи.

Необходимо проанализировать статистические данные по странам Восточной Европы (рис. 1).

Для проведения корреляционного анализа нужно в меню Данные выбрать опцию Анализ данных. В появившемся окне выбрать опцию. Корреляция.

В окне «Корреляция» введите Входной интервал – те данные, которые подлежат

анализу. Группирование в нашем примере осуществляется по столбцам (столбцы содержат отдельные показатели по странам). Поле Метки в первой строке следует отметить в том случае, если входной интервал задан вместе с заголовками столбцов/строк.

В разделе окна Параметры вывода укажите, куда следует выводить корреляционную матрицу (квадратная (или прямоугольная) таблица, в которой на пересечении соответствующих строки и столбца находится коэффициент корреляции между соответствующими параметрами).

Прокомментируем полученную матрицу. Как видно из полученных результатов наибольшая корреляционная зависимость (очень сильная) наблюдается между показателями «Доля замужних женщин 15-49 лет, использующих все виды контрацептивов (%)» и «Доля замужних женщин 15-49 лет, использующих современные контрацептивы (%)», где  $r=0,921214$  – скорее всего эти факторы являются следствием одной общей причины.

Наименьшая (очень слабая) между «Доля замужних женщин 15-49 лет, использующих все виды контрацептивов (%)» и

	A	B	C	D	E	F	G	H
	Страны Восточной Европы	Коэффициент рождаемости (на 1000 жителей)	Доля городского населения, %	Коэффициент поддержки пожилых, 2010 г. *с	Доля замужних женщин 15-49 лет, использующих все виды контрацептивов (%)	Доля замужних женщин 15-49 лет, использующих современные контрацептивы (%)	ВВП на душу населения с учетом паритета покупательной способности в 2008 г., (US\$)	Число мобильных телефонов на 100 жителей
1								
2	Белоруссия	12	74	5	73	57	12150	84
3	Болгария	11	71	4	63	40	11950	138
4	Чешская республика	11	74	5	72	63	22790	134
5	Венгрия	10	67	4	77	68	17790	122
6	Молдавия	11	41	6	68	44	3210	67
7	Польша	11	61	5	49	19	17310	115
8	Румыния	10	55	5	70	38	13500	115
9	Россия	12	73	6	73	53	15630	141
10	Словакия	11	55	6	80	66	21300	102
11	Украина	11	69	5	67	48	7210	121
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								

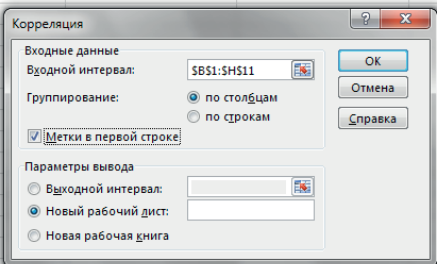


Рис. 1. Окно «Корреляция»

	A	B	C	D	E	F	G	H
		Коэффициент рождаемости (на 1000 жителей)	Доля городского населения, %	Коэффициент поддержки пожилых, 2010 г. *с	Доля замужних женщин 15-49 лет, использующих все виды контрацептивов (%)	Доля замужних женщин 15-49 лет, использующих современные контрацептивы (%)	ВВП на душу населения с учетом паритета покупательной способности в 2008 г., (US\$)	Число мобильных телефонов на 100 жителей
1								
2	Коэффициент рождаемости (на 1000 жителей)	1						
3	Доля городского населения, %	0,383212087	1					
4	Коэффициент поддержки пожилых, 2010 г. *с	0,451753951	-0,4570304	1				
5	Доля замужних женщин 15-49 лет, использующих все виды контрацептивов (%)	-0,01935138	0,06288497	0,206312624	1			
6	Доля замужних женщин 15-49 лет, использующих современные контрацептивы (%)	0,044168302	0,2681051	0,053873653	0,921214401	1		
7	ВВП на душу населения с учетом паритета покупательной способности в 2008 г., (US\$)	-0,09688479	0,37940628	-0,096863385	0,225435419	0,359134001	1	
8	Число мобильных телефонов на 100 жителей	-0,08424092	0,66587291	-0,405298193	-0,081400205	0,020960385	0,489485881	1

Рис. 2. Корреляционная матрица

«Коэффициент рождаемости (на 1000 жителей)», где  $r = -0,0193514$ . В то время как взаимосвязь признаков «Доля городского населения, %» и «Число мобильных телефонов на 100 жителей» можно оценить как сильную, т.к.  $r = 0,66587291$ , при этом, скорее всего второй признак является следствием

первого, но, скорее всего, на «Число мобильных телефонов на 100 жителей» оказывают влияние еще какие-то факторы.

Для демонстрации возможностей графического метода построим точечную диаграмму (диаграмму рассеяния) по этим признакам.

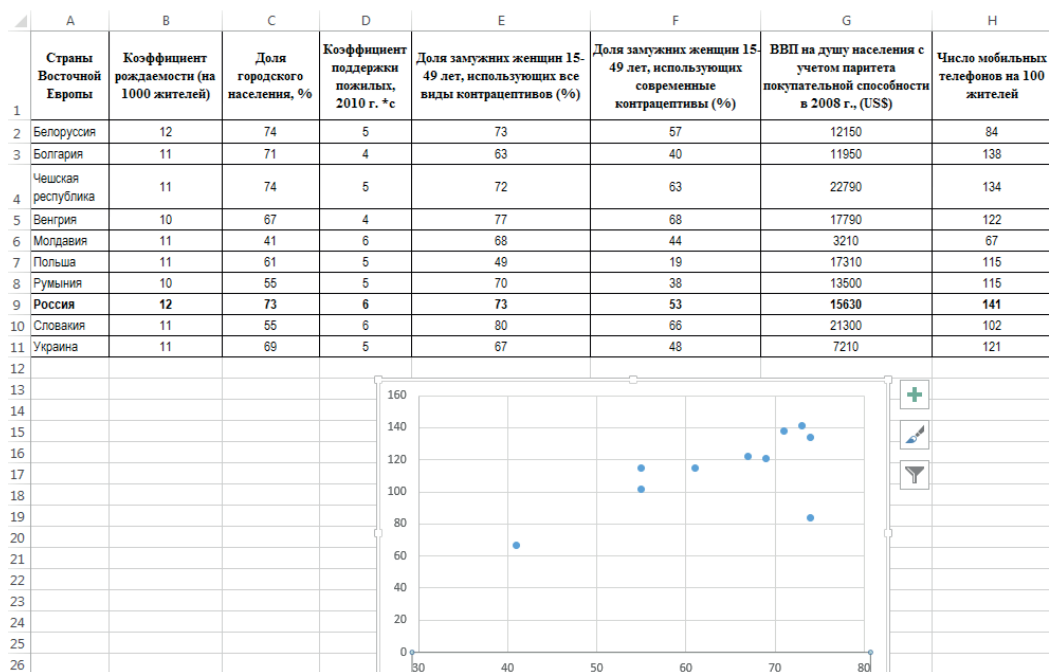


Рис. 3. Построение диаграммы рассеяния (точечной диаграммы)

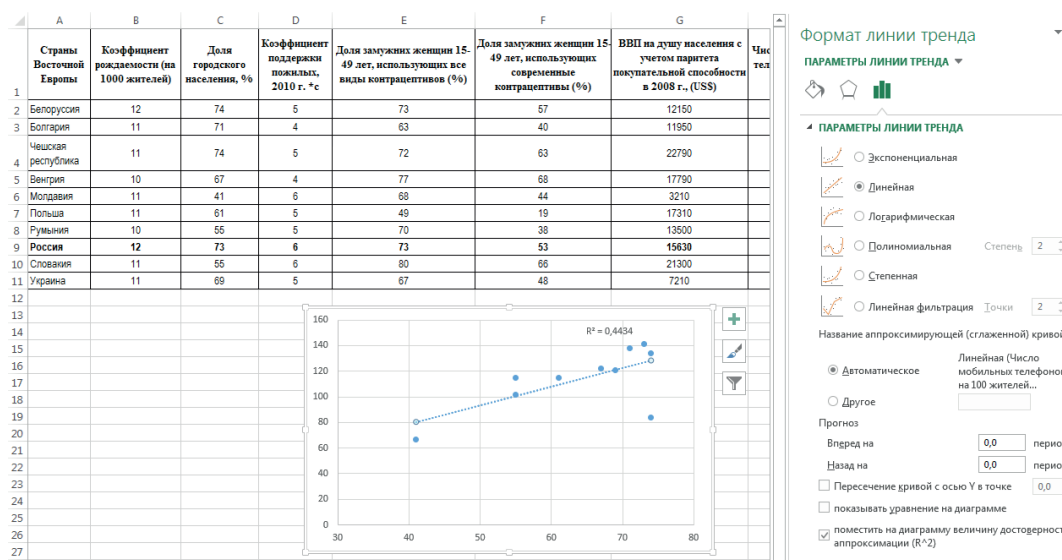


Рис. 4. Добавление линии тренда

Полученную диаграмму дополним линией регрессии и коэффициентом достоверности аппроксимации, щелкнув правой кнопкой мыши по точкам диаграммы и выбрав в контекстном меню пункт Линия тренда (рис. 4).

Полученная точечная диаграмма позволяет судить не только о разбросе точек вокруг предполагаемой линии тренда, но и увидеть аномальные совместные проявления признаков (рис. 5).

Так, выделенная точка построена по значениям показателей «Доля городского населения, %» и «Число мобильных телефонов на 100 жителей» Белоруссии. Анализ и интерпретация появления аномалий должны производиться исходя из знаний о социально-экономической природе рассматриваемого объекта.

#### Список литературы

1. Овчинникова И.Г., Варфоломеева Т.Н., Гусева Е.Н. Учебно-методическое пособие для подготовки к вступительным экзаменам по информатике. -Магнитогорск, 2002. -С. 119
2. Овчинникова И.Г., Варфоломеева Т.Н., Корнещук Н.Г. Учебное пособие для подготовки к централизованному тестированию по информатике. -Магнитогорск, 2002. -С.205
3. Курзаева Л.В. Дистанционный курс «Основы математической обработки информации»: электронный учебно-методический комплекс // Хроники объединенного фонда электронных ресурсов Наука и образование. - 2014. -Т. 1. - № 12 (67). - С. 117
4. Курзаева Л.В. Введение в теорию систем и системный анализ: учеб. пособие/Л.В. Курзаева. -Магнитогорск: МаГУ, 2015. -211 с.
5. Курзаева Л.В. Введение в методы и средства получения и обработки информации для задач управления социальными и экономическими системами: учеб. пособие/Л.В. Курзаева, И.Г. Овчинникова, Г.Н. Чусавитина. -Магнитогорск: Магнитогорск. гос. техн. ун-та им. Г.И. Носова, 2016. -118 с.