УДК 681.3.06

Н.К. Кадыркулова – ст.преп.ОшТУ
N.K. Kadyrkulova- senior teacher OshTU

ВОЗМОЖНОСТИ ПАКЕТА MS EXCEL ПРИ ОБРАБОТКЕ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Изучение возможностей пакета MS Excel при решении задач обработки экспериментальных данных. Приобретение навыков обработки результатов эксперимента.

Ключевые слова: пакет MS EXCEL, математическая задача, уравнения, расчет, линейная функция.

MS EXCEL POSSIBLE PACKAGE FOR PROCESSING EXPERIMENTAL DATA

The study of MS Excel package options for solving problems of experimental data processing. The acquisition of skills in the processing of the experimental results.

Keywords: MS EXCEL package, mathematical problem, equation, calculation, linear function.

Одной из распространенных задач в науке, технике, экономике является аппроксимация экспериментальных данных, алгебраических данных аналитическими выражениями. Возможность подобрать параметры уравнения таким образом, чтобы его решение совпало с данными эксперимента, зачастую является доказательством (или опровержением) теории.

Рассмотрим следующую математическую задачу. Известные значения некоторой функции ***f*** образуют таблицу:

Таблица 1.1

x	x_1	x_2	. . .	x_n
f(x)	y_1	y_2	. . .	y_n

Необходимо построить аналитическую зависимость **y=*f*(x)**, наиболее близко описывающую результаты эксперимента. Построим функцию $y=f(x, a_0, a_1, ..., a_k)$ таким образом, чтобы сумма квадратов отклонений измеренных значений y_i от расчетных $f(x_i, a_0, a_1, ..., a_k)$ была наименьшей (см. рис.1.).

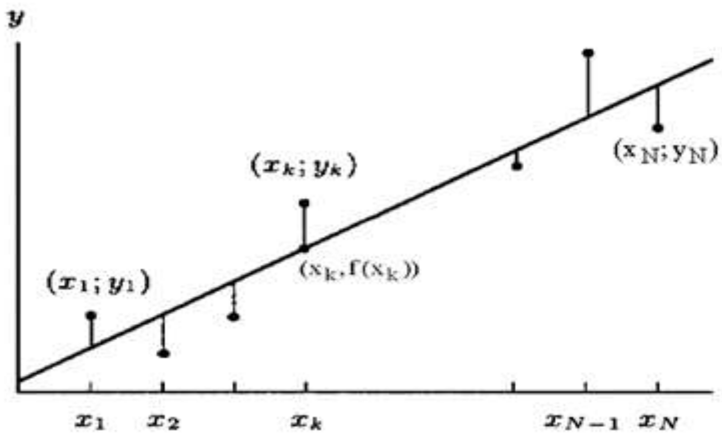


Рис.1.

Математически эта задача равносильна следующей: найти значение параметров **a_0, a_1, a_2, ...,a_k**, при которых функция принимала бы минимальное значение.

$$S(a_0,\ a_1,\ \ldots,\ a_k) = \sum_{i=1}^{n}[y_i - f(x_i,\ a_0,\ a_1,\ \ldots,\ a_k)]^2 \qquad 1.1)$$

Эта задача сводится к решению системы уравнений:

$$\partial s/\partial a_0 = 0; \quad \partial s/\partial a_1 = 0; \quad \partial s/\partial a_k = 0; \qquad 1.2)$$

Если параметры a_i входят в зависимость **$y = f(x, a_0, a_1, \ldots, a_k)$** линейно, то мы получим систему линейных уравнений:

$$\frac{\partial s}{\partial a_j} = 0; \quad \sum_{i=1}^{n}(-f(x_i,\ a_0,\ a_1,\ \ldots,\ a_k))\ \frac{\partial f}{\partial a_j} = 0; \quad j = 0,1,\ldots,k \qquad 1.3)$$

Линейная функция (линия регрессии)

Необходимо определить параметры функции **y = ax+b**. Составим функцию **S**:

$$S = \sum_{i=1}^{n}[y_i - ax_i - b]^2 \qquad (1.4)$$

Продифференцируем выражение (8.4) по **a** и **b**, сформируем систему линейных уравнений, решив которую мы получим следующие значения параметров:

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}; \qquad b = \frac{\sum x_i^2 y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \qquad 1.5)$$

Подобранная прямая называется **линией регрессии** y на x, a и b называются **коэффициентами регрессии**.

$$Z = \sum_{i=1}^{n}(y_i - ax_i - b)^2, \qquad (1.6)$$

Чем меньше величина тем более обосновано предположение, что табличная зависимость описывается линейной функцией. Существует показатель, характеризующий тесноту линейной связи между **x** и **y**. Это **коэффициент корреляции**. Он рассчитывается по формуле:

$$r = \frac{\sum(x_i - M_x)(y_i - M_y)}{\sqrt{\sum(x_i - M_x)^2 \sum(y_i - M_y)^2}} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum(x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}} \qquad (1.7)$$

Коэффициент корреляции **r** и коэффициент регрессии **a** связаны соотношением:

$$a = \frac{Dy}{Dx} r, \qquad (1.8)$$

где **Dy, Dx** - среднеквадратичное отклонение значений **x** и **y**.

$$Dx = \frac{\sum(x_i - Mx)^2}{n}, \quad \text{где} \quad Mx = \frac{\sum x_i}{n}; \quad Dx = \frac{\sum(y_i - My)^2}{n}; \quad My = \frac{\sum y_i}{n} \qquad (1.9)$$

Значение коэффициента корреляции удовлетворяет соотношению **-1 ≤ r ≤ 1**. Чем меньше отличается абсолютная величина **r** от единицы, тем ближе к линии регрессии располагаются экспериментальные точки. Если коэффициент корреляции равен нулю, то переменные **x, y** называются **некоррелированными**. Если **r** = 0, то это только означает, что между **x, y** не существует линейной связи, но между ними может существовать зависимость, отличная от линейной.

Для того чтобы проверить, значимо ли отличается от нуля коэффициент корреляции, можно использовать **критерий Стьюдента**. Вычисленное значение критерия определяется по формуле:

$$t = \frac{\sqrt{n - 2}}{\sqrt{1 - r^2}} \qquad (1.10)$$

Значение **t** сравнивается со значением, взятым из таблицы распределения Стьюдента в соответствии с уровнем значимости **a** и числом степеней свободы **n-2**. Если **t** больше табличного, то коэффициент корреляции значимо отличен от нуля.

Решение поставленной задачи средствами MS Excel

Вычисление коэффициентов регрессии осуществляется с помощью функции ЛИНЕЙН():

ЛИНЕЙН(Значения_y; Значения_x; Конст; статистика)

Значения_y - массив значений y.

Значения_x- необязательный массив значений **x**, если массив **x** опущен, то предполагается, что это массив **{1;2;3;...}** такого же размера, как и **Значения_y**.

Конст - логическое значение, которое указывает, требуется ли, чтобы константа **b** была равна 0. Если **Конст** имеет значение **ИСТИНА** или опущено, то **b** вычисляется обычным образом. Если аргумент **Конст** имеет значение **ЛОЖЬ**, то **b** полагается равным 0 и значения **a** подбираются так, чтобы выполнялось соотношение **y = ax**.

Статистика - логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии. Если аргумент статистика имеет значение **ИСТИНА**, то функция **ЛИНЕЙН** возвращает дополнительную регрессионную статистику. Если аргумент статистика имеет значение **ЛОЖЬ** или опущен, то функция **ЛИНЕЙН** возвращает только коэффициент **a** и постоянную **b**.

Для вычисления множества точек на линии регрессии используется функция **ТЕНДЕНЦИЯ**.

ТЕНДЕНЦИЯ(Значения_y; Значения_x; Новые_значения_x; Конст)

Значения_y- массив значений **y**, которые уже известны для соотношения **y = ax + b**.

Значения_x- массив значений **x**.

Новые_значения_x- новый массив значений, для которых **ТЕНДЕНЦИЯ** возвращает соответствующие значения **y**. Если **Новые_значения_x** опущены, то предполагается, что они совпадают с массивом значений **x**.

Конст - логическое значение, которое указывает, требуется ли, чтобы константа **b** была равна 0. Если **Конст** имеет значение **ИСТИНА** или опущено, то **b** вычисляется обычным образом. Если **Конст** имеет значение **ЛОЖЬ**, то **b** полагается равным 0, и значения **a** подбираются таким образом, чтобы выполнялось соотношение **y = ax**. Необходимо помнить, что результатом функций **ЛИНЕЙН, ТЕНДЕНЦИЯ** является множество значений - массив.

Для расчета коэффициента корреляции используется функция **КОРРЕЛ**, возвращающая значения коэффициента корреляции:

КОРРЕЛ(Массив1; Массив2)

Массив1 - массив значений **y**.

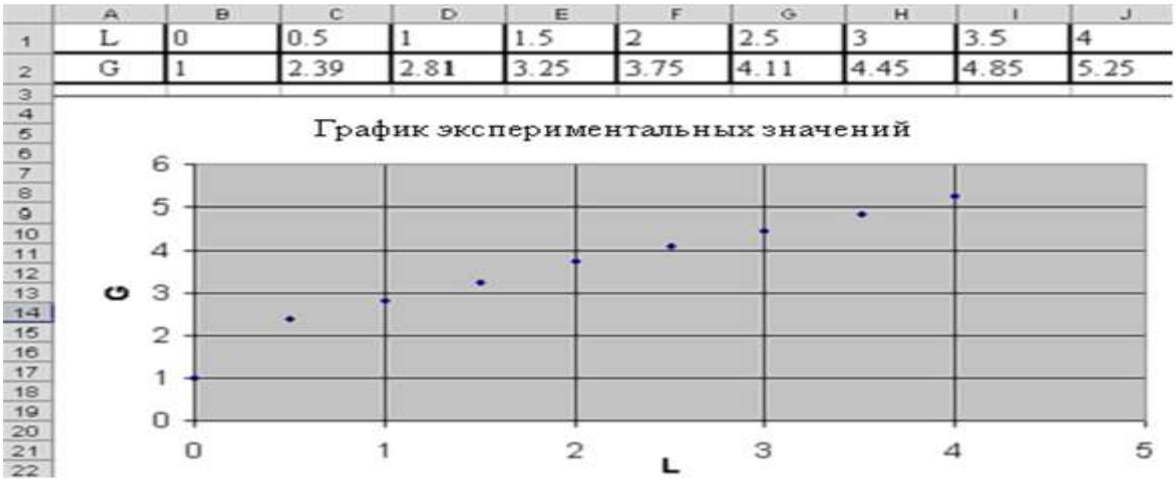
Массив2 - массив значений **y**.

Массив1 и **Массив2** должны иметь одинаковое количество точек данных.

ПРИМЕР 1. Известна табличная зависимость **G(L)**. Построить линию регрессии и вычислить ожидаемое значение в точках 0, 0.75, 1.75, 2.8, 4.5.

L	0	0,5	1	1,5	2	2,5	3	3,5	4
G	1	2,39	2,81	3,25	3,75	4,11	4,45	4,85	5,25

Введем таблицу значений в лист MS Excel и построим точечный график. Рабочий лист примет вид изображенный на рис.2.



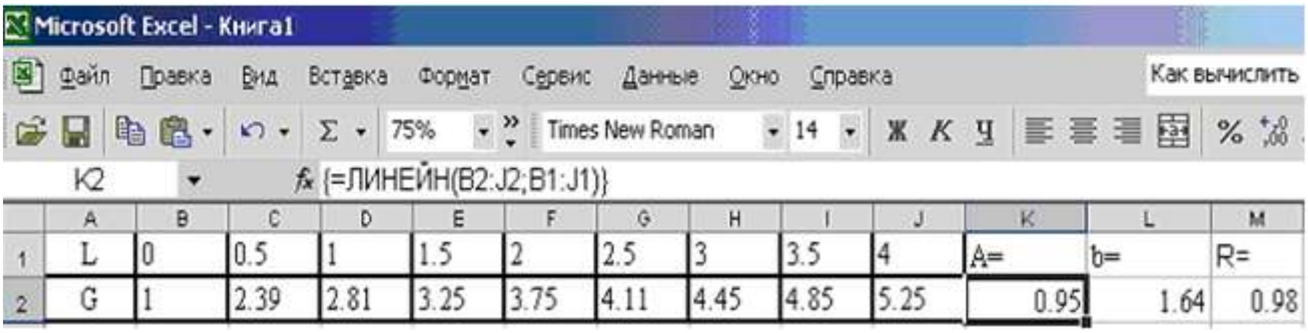
	A	B	C	D	E	F	G	H	I	J
1	L	0	0.5	1	1.5	2	2.5	3	3.5	4
2	G	1	2.39	2.81	3.25	3.75	4.11	4.45	4.85	5.25

Рис.2. График экспериментальных данных

Для того, чтобы рассчитать значения коэффициентов регрессии **a** и **b** выделим ячейки **K2:L2**, обратимся к мастеру функций и в категории **Статистические** выберем функцию **ЛИНЕЙН**. Заполним появившееся диалоговое окно так, как показано на рис.3. и нажмем Ок.

В результате вычисленное значение появится только в ячейке **K2** (см. рис.3). Для того чтобы вычисленное значение появилось и в ячейке **L2** необходимо войти в режим редактирования, нажав клавишу **F2**, а затем нажать комбинацию клавиш **CTRL+SHIFT+ENTER**.

Для расчета значения коэффициента корреляции в ячейку **M2** была введена следующая формула: **M2 = КОРРЕЛ(B1:J1;B2:J2)** (см. рис.3).



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	L	0	0.5	1	1.5	2	2.5	3	3.5	4	A=	b=	R=
2	G	1	2.39	2.81	3.25	3.75	4.11	4.45	4.85	5.25	0.95	1.64	0.98

Рис.3. Диалоговое окно MS Excel

Для вычисления ожидаемого значения в точках 0, 0.75, 1.75, 2.8, 4.5 занесем их в ячейки **L9:L13**. Затем выделим диапазон ячеек **M10:M13** и введем формулу:

= ТЕНДЕНЦИЯ(B2:J2;B1:J1;L9:L13).

Для того чтобы вычисленные значения появились и в ячейках **M10:M13** необходимо нажать комбинацию клавиш **CTRL+SHIFT+ENTER**.

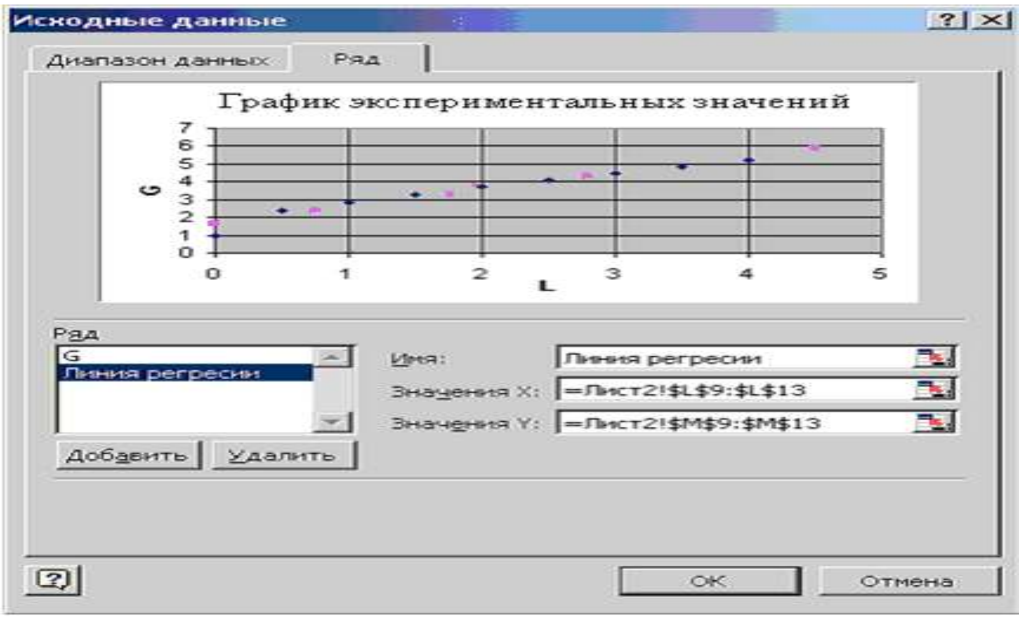
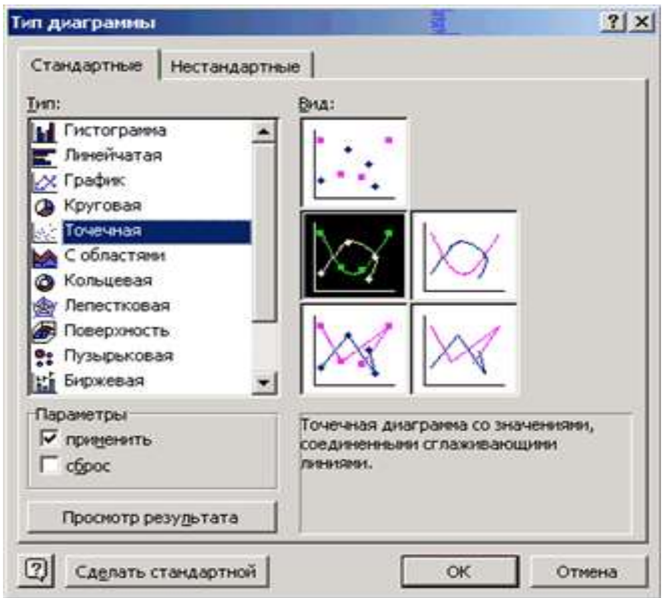


Рис.4. Линию регрессии на диаграмме

Изобразим линию регрессии на диаграмме. Для этого выделим экспериментальные точки на графике, щелкнем правой кнопкой мыши и выберем команду **Исходные данные**. В появившемся диалоговом окне (см. рис.5), для добавления линии регрессии щелкнем по кнопке **Добавить**.

В качестве имени введем **Линия регрессии**, в качестве **Значения X: L9:L13**, в качестве **Значения Y: M9:M13**. Далее выделяем линию регрессии, для изменения ее типа щелкаем правой кнопкой мыши и выбираем команду **Тип диаграммы** (см. рис.5). Для форматирования линии регрессии (можно изменить толщину линии, цвет, тип маркера и т.п) дважды щелкаем по ней (см. рис.6).



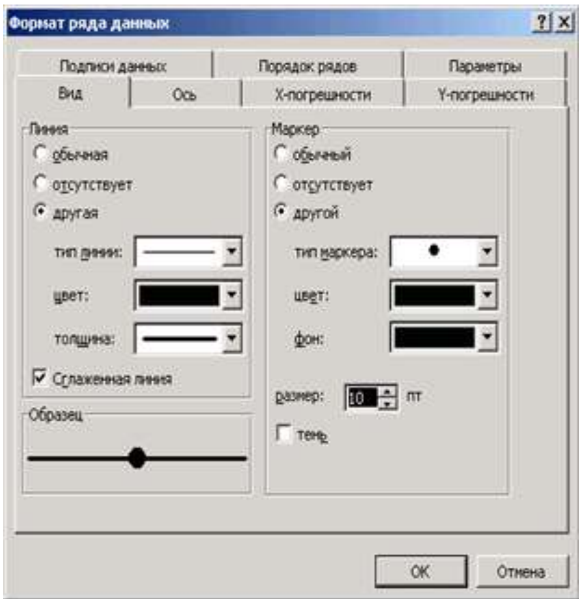
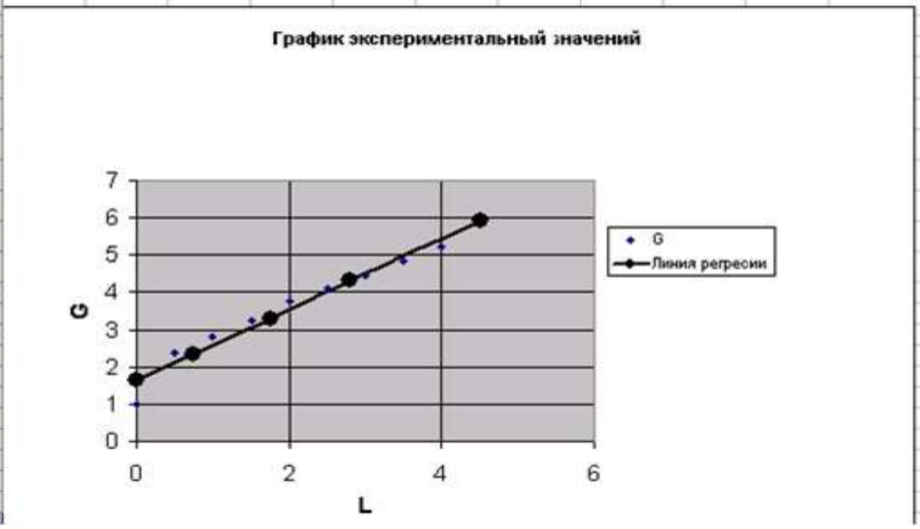


Рис.5. Тип диаграммы

Рис.6. Форматирование линии регрессии

После форматирования графика рабочий лист примет вид, изображенный на рис.7.

L	0	0.5	1	1.5	2	2.5	3	3.5	4	A=	b=	R=
G	1	2.39	2.81	3.25	3.75	4.11	4.45	4.85	5.25	0.95	1.64	0.98



X	Y
0	1.64
0.75	2.35
1.75	3.30
2.8	4.30
4.5	5.92

Построение различных аппроксимирующих зависимостей в MS Excel реализовано в виде свойства диаграммы - **линия тренда**.

Вывод

Рассмотренный пример показывает, возможности пакета MS Excel при обработке экспериментальных данных. Приобретение навыков обработки результатов эксперимента требует произвести немало вычислений с использованием значительного количества числовых значений, порождаемых главным образом в процессе самого счета. Исходными данными здесь являются только числа из таблицы. Ручное вычисление требует много времени и зависит от аккуратности человека. Поэтому, наиболее подходящим вычислительным средством в этом случае является программа, составленная на компьютерных языках. Это значительно уменьшает время нахождения функции, и в результате получим более точный результат.

Литература:

1. Зайдель А.Н. Ошибки измерения физических величин. – Л.:Наука, 1974. –108 с.
2. Кассандрова О.Н., Лебедев В.В. Обработка результатов наблюдений. – М.: Наука, 1970. – 104с.
3. Колесников А.Ф. Основы математической обработки результатов измерений. – Томск: ТГУ, 1963. – 49 с.
4. Плескунин В.И., Воронина Е.Д. Теоретические основы организации и анализа выборочных данных в эксперименте. Учебное пособие. – Л.:ЛЭУ, 1979. – 232 с.
5. Румшинский Л.З. Математическая обработка результатов эксперимента. Справочное руководство. – М.: Наука, 1971. – 192 с.
6. Рыжов Э.В., Горленко О.А. Математические методы в технологических исследованиях. – Киев:Наук. думка, 1990. – 184 с.
7. Сухов А.Н. Математическая обработка результатов измерений. Учебное пособие. – М.: МИСИ, 1982. – 89 с.
8. Чкалова О.Н. Основы научных исследований. – Киев:Вища школа, 1978. –120 с.
9. Вучков М., Бауживева Л., Солаков С. «Прикладной регрессионный анализ», М.: "Радио и связь", 1987 г.
10. Б. П. Демидович «Задачи и упражнения по математическому анализу», М.: "Наука", 1970 г.