

ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ В СРЕДЕ MICROSOFT EXCEL

Таракановский С. Н., Картюк В.М.

Освещается возможность использования статистического аппарата (пакета «Анализ статистики, статистические функции») Microsoft Excel для обработки результатов обычных и планированных экспериментов с целью получения регрессионных моделей физических процессов.

Широкое распространение табличного процессора Excel дает большие возможности современным исследователям продолжить работы по изучению физических процессов в области строительства.

Использование статистического аппарата Excel рассматривается на примере решения задачи по уточнению режима одностадийного формирования двухслойных конструкций методом расслоения легкогобетонной смеси, изложенной в работе [1].

Был выполнен шестифакторный планированный эксперимент для выявления влияния на время расслоения легкогобетонной смеси (Y - взято в качестве нашего примера), соотношение толщины плотного и крупнопористого слоев, прочность и среднюю плотность несущего и теплоизоляционного бетонов следующих факторов, как в отдельности, так и во взаимодействии друг с другом: расхода цемента (X_1), плотного и пористого песка мелкой фракции на m^3 бетонной смеси (X_3), содержания пластифицирующей добавки С-3 (X_2), высоты формы (X_4), плотности песка (X_5) и вида крупного заполнителя (керамзитового гравия и вулканического щебня фракции 10 ... 20 мм) - X_6 .

Так как каждый из названных выше факторов может влиять на функцию «выхода» не линейно, то ее целесообразно аппроксимировать полиномом второй степени. В связи с этим, эксперимент выполнен по насыщенному плану Рехтшафтнера [2]. Выбор этого плана обусловлен также и тем, что построенные по нему квадратичные модели обладают не только хорошими статистическими характеристиками, но и малой корреляцией между оценками коэффициентов регрессии при постоянстве дисперсии этих оценок. Отметим, что пакет анализа Excel включает в себя расчет линейной регрессии, а необходимость реализации нелинейной регрессии потребовала от авторов использования специального приема, описываемого ниже.

Для обработки результатов эксперимента все исходные и результирующие данные заносятся на рабочий лист Excel в следующем

порядке:

- первые три строки выделяются для пояснений содержимого колонок, а остальные - информационные, отражают один из опытов проведенного эксперимента;

- первые восемь колонок отражают план эксперимента:

- первая колонка таблицы (колонка А) - номер опыта;

- вторая колонка (В) - индекс опыта;

- от третьей по восьмую (С - Н) - факторы (в первой строке приводится наименование «Уровни варьирования факторов», во второй строке - обозначение факторов ($X_1, X_2 \dots X_6$), в третьей строке - обозначение натуральных значений факторов (например, для расхода цемента (X_1) - Ц, кг/м³); в остальных строчках приводятся значки: (-1), (0) и (+1) - уровни варьирования факторов);

- в последующие шесть колонок вносятся кодированные значения факторов, остальные колонки используются для слагаемых типа X_1X_2 (все возможные сочетания) и $X_1^2, X_2^2, X_3^2, X_4^2, X_5^2, X_6^2$. В первой строке всех колонок пишется заглавие «Кодированные значения факторов и их слагаемых». В третьей строке в каждой ячейке - обозначение фактора или слагаемого (x_1, x_1^2, x_1x_2). Вторая строка предназначена для внесения формулы по расчету коэффициента корреляции для данной составляющей и результирующим значением (опытное значение времени расслоения смеси);

- если для шести факторов в описанные выше колонки информационных строчек вносятся их кодированные значения (+1,0,-1) по каждому опыту, то для остальных слагаемых - заносятся формулы, позволяющие рассчитать их значения, типа: для слагаемого x_1x_2 заносится формула «=I4*J4», для слагаемого x_5^2 «=M4^2»;

- за колонками матрицы эксперимента следуют: колонка AJ - отражает время расслоения, полученное в ходе эксперимента (опыта) - Y1; колонка АК - время расслоения, полученное по формуле линии регрессии, найденное в ходе исследования В.М. Карпюка [1] (Y2) и колонка AL по вновь полученному уравнению регрессии при работе в Excel(Y);

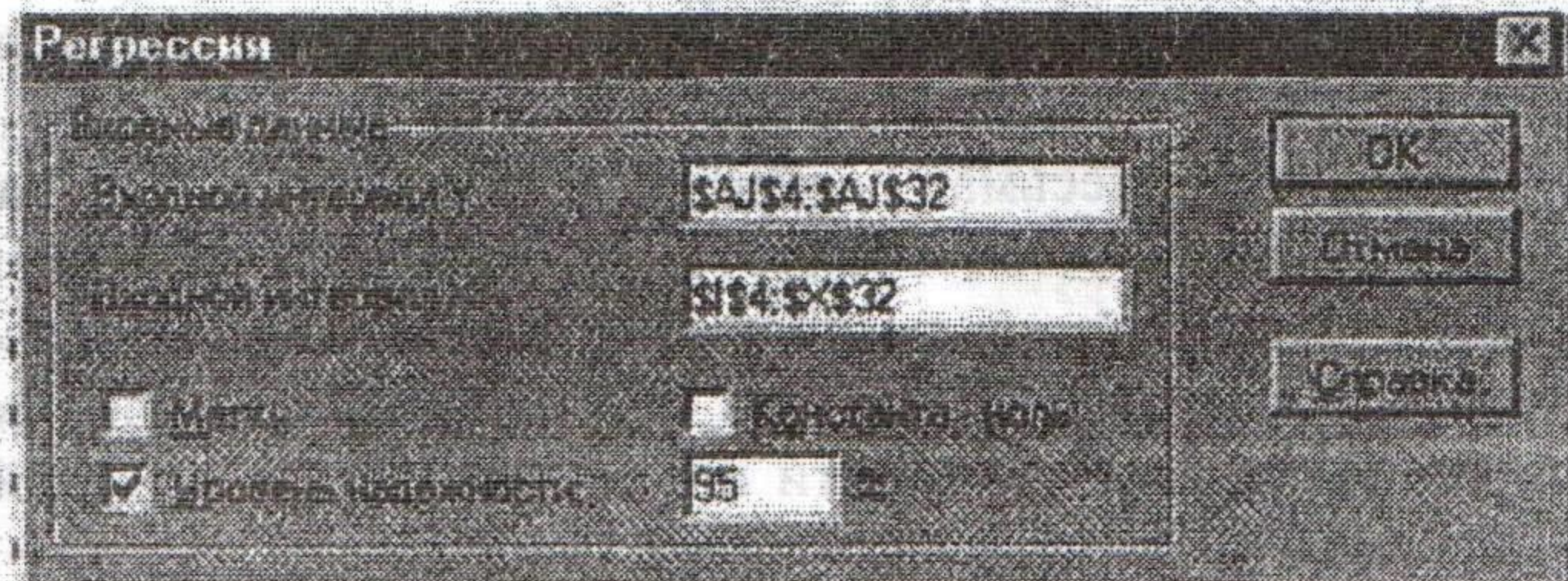
- в колонки, начиная с I до AI, второй строки вносим формулу для расчета коэффициента корреляции соответствующего столбца со значениями слагаемого из предполагаемой линии регрессии и результирующего значения Y(опыта). Например, для ячейки I2 вносится формула : «=ABS(KORREL(I4:I32;\$AJ\$4:\$AJ\$32))», по которой рассчитывается значение - 0.6308. Для внесения аналогичных формул в остальные ячейки можно выполнить операцию копирования формулы из ячейки I2.

Статистическая функция KORREL (массив1; массив2) оперирует коэффициентом корреляции между интервалами ячеек массива 1 и массива 2.

Максимальное число входных диапазонов для Excel равно 16. В связи с этим, необходимо выбрать первые наиболее значимые факторы

(слагаемые) из списка N1... N27. Колонки, начиная с I до AI, следует упорядочить по убыванию значения коэффициента корреляции.

После этого можно приступить к проведению регрессионного анализа. На рисунке приведен вид окна диалога "Регрессия" и параметров для проведения первой итерации. Ниже рисунка приведены результаты работы пакета анализа. Отметим, что коэффициент корреляции между значениями Y1 (опыты) и Y2 [1] составляет 0.98158.



<i>Регрессионная статистика</i>	
Множественный R	0.980511471
R-квадрат	0.961402745
Нормированный R-квадрат	0.909939739
Стандартная ошибка	21.91201815
Наблюдения	29

	<i>df</i>	<i>55</i>	<i>MS</i>	<i>F</i>
Регрессия	16	143514.2236	8969.638975	18.68143
Остаток	12	5761.638475	480.1365396	
Итого	28	149275.8621		
	<i>Коэф.</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>Коэф. коррел</i>
Y-пересечение	116.5194754	9.064073042	12.85509007	
Переменная X1	51.36524489	4.827200339	10.6407941	0,6308
Переменная X2	-46.60376666	4.804427702	-9.700170249	0,4224
Переменная X3	23.48899595	4.816178858	4.877102085	0,2743
Переменная X4	23.29885674	4.828762456	4.82501613	0,2419
Переменная X5	-17.26994972	4.82267204	-3.580991944	0,2118
Переменная x1x2	-13.68188902	4.919356592	-2.781235465	0,1970
Переменная 2x5	2.993005257	4.92909244	0.607212239	0,1684
Переменная x5x6	7.166403947	4.929076746	1.453903909	0,1684
Переменная x3x4	15.08416569	4.929456666	3.060005739	0,1351
Переменная x4x5	2.140545466	4.92132043	0.434953484	0,1351
Переменная x1x6	-2.008490328	4.935241186	-0.406969032	0,1306
Переменная x1x3	0.263628281	4.930647354	0.053467276	0,1306
Переменная x3x5	1.938522556	4.926376596	0.393498653	0,1019

Переменная x2x3	-14.06337452	4.937126946	-2.848493602	0,0973
Переменная x2x6	-16.33549313	4.92730687	-3.315298511	0,0973
Переменная x2^2	4.705241194	10.42935843	0.451153465	0,0859

В таблицу результатов внесена колонка, отражающая значение коэффициента корреляции соответствующей переменной, полученной с помощью функции КОРРЕЛ.

Выведем из списка переменных, те у которых t-значения меньше единицы. Это переменные x2x5, x4x5, x1x6, x1x3, x3x5, x2^2.

В работе [2] отмечается весьма полезным проведение, так называемого, последовательного регрессионного анализа для удаления статистически независимых эффектов в моделях, построенных на основе планов второго порядка, что способствует их упрощению, повышению информативной способности и адекватности полиномиальных моделей, возрастанию множественного коэффициента регрессии.

Для последующих итераций регрессионного анализа выведем из списка переменную, у которой наименьшее t-значение, а введем в список одну из переменных x4x6 (0.0687), x1x5 (0,0641), x2x4 (0,0641), x6 (0.0637), имеющих схожие коэффициенты корреляции. Результаты анализа приведены ниже.

	Коэф. Y-пересечения	Стандарт. ошибка Y-пересечения	t-стат. вводимой переменной	Коэф.кор.. (множественный R)
Перемен.x4x6	120.8572085	3.800910293	0.969214925	0.975455703
Перемен.x1x5	121.431678	3.855686289	-0.657994066	0.974772658
Перемен.x2x4	121.7363937	3.212696687	-2.896858107	0.982448555
Перемен.x6	121.2579163	3.860530193	0.490086889	0.974502555

Из таблицы видно, что введение переменной x2x4 вместо x5x6 позволяет увеличить множественный коэффициент регрессии до значения 0.98245, который превышает значение коэффициента, приведенного в исследовании [1] (0.98158).

Регрессионная статистика	
Множественный R	0.982448555
R-квадрат	0.965205163
Нормированный R-квадрат	0.945874698
Стандартная ошибка	16.9869709
Наблюдения	29

	df	SS	MS	F
Регрессия	10	144081.8328	14408.18328	49.93181339
Остаток	18	5194.029246	288.5571803	

Итого	28	149275.8621		
	Коэф.	Стандарт. ошибка	t- статис-а тик	P-Значение
Y-пересеч.	121.7363937	3.212696687	37.8922773	1.27512E-18
Перемен.х1	53.3764789	3.662428863	14.57406571	2.0853E-11
Перемен.х2	-47.0119693	3.6680212	-12.81671144	1.73604E-10
Перемен.х3	24.9806738	3.664852052	6.81628438	2.21329E-06
Перемен.х4	22.60443617	3.664852052	6.16789869	8.00537E-06
Перемен.х5	-18.32321026	3.652141675	-5.017113764	8.94669E-05
Перемен.х1х2	-13.57663187	3.755161294	-3.615459045	0.001977988
Перемен.х3х4	15.75148025	3.769532534	4.178629607	0.000564391
Перемен.х2х3	-13.40608894	3.756254556	-3.569004373	0.002193295
Перемен.х2х6	-15.25192765	3.744539105	-4.073112129	0.000713669
Перемен.х2х4	-10.88133646	3.756254556	-2.896858107	0.00961038

Используя данные регрессионной статистики получаем математическую модель времени расслоения бетона (Y): $Y = 122 + 53 X_1 - 47X_2 + 25 X_3 + 23 X_4 - 18 X_5 - 14 X_1X_2 + 16 X_3X_4 - 13 X_2X_3 - 15 X_2X_6 - 11 X_2X_4$

Среднее значение выходного параметра равно 118. Следовательно, целесообразно пересчитать коэффициенты при переменных регрессии, когда свободный член равен 118. С этой целью воспользуемся средством поиска решений из Excel 5.0. Чтобы построить модель для пакета «Поиск решений», вставим дополнительно две колонки (I,J), где укажем:

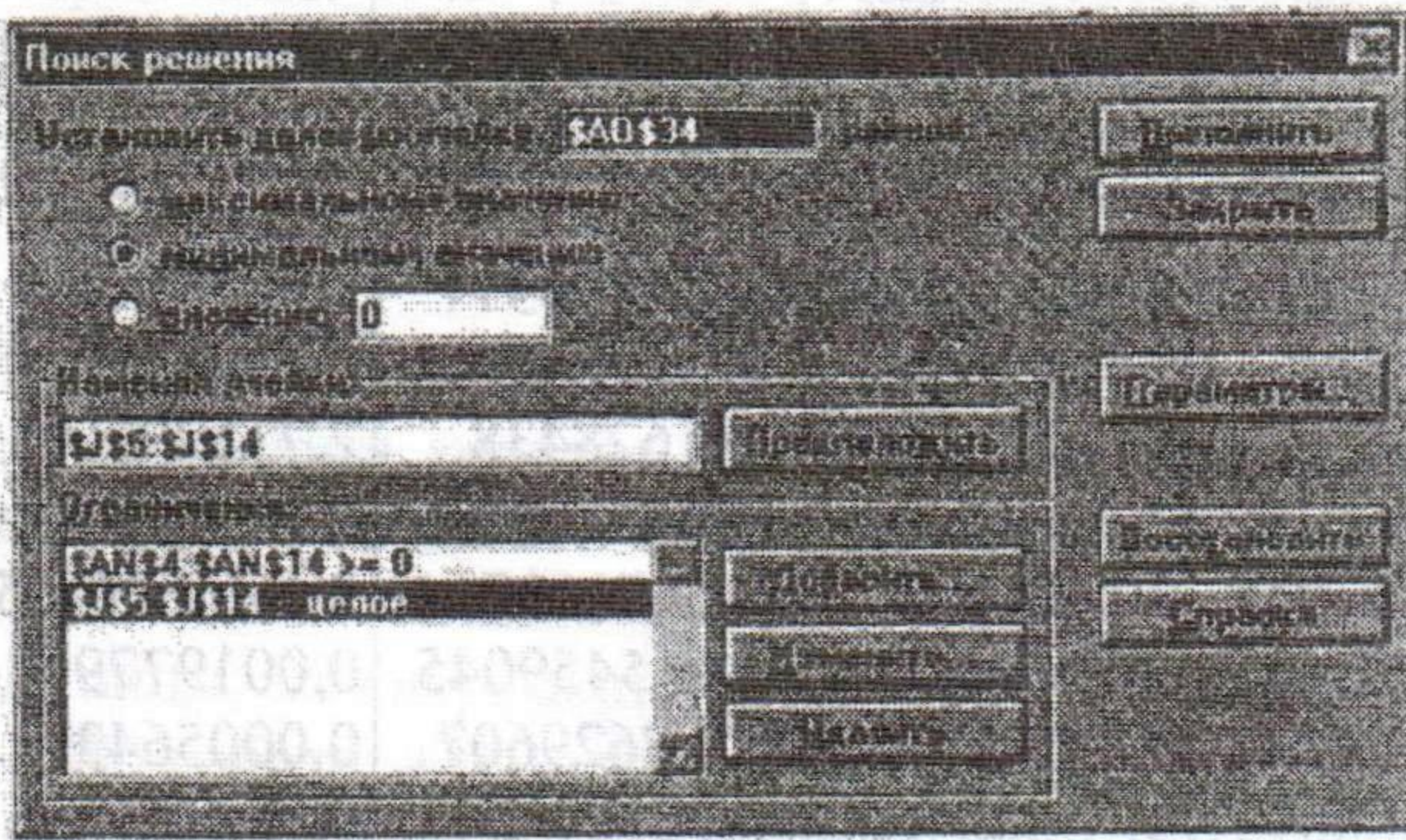
- в колонке I наименования Y-пересечения и переменных линии регрессии (построчно, начиная с 4 по 14 строку);
- в колонке J значения коэффициентов регрессии (в J4 заносим 118, а в J5 - J14 вносятся коэффициенты приведенного выше уравнения в качестве исходных);

В ячейку AN4 колонки AN заносим формулу уравнения регрессии, где в качестве коэффициентов регрессии используются значения соответствующих ячеек колонки J, а в качестве переменных - кодовые значения самих переменных. Формула копируется на все строки с 5 по 32. При этом, в колонке AN будет получено прогнозируемое значение времени расслоения, полученное по установленной регрессии.

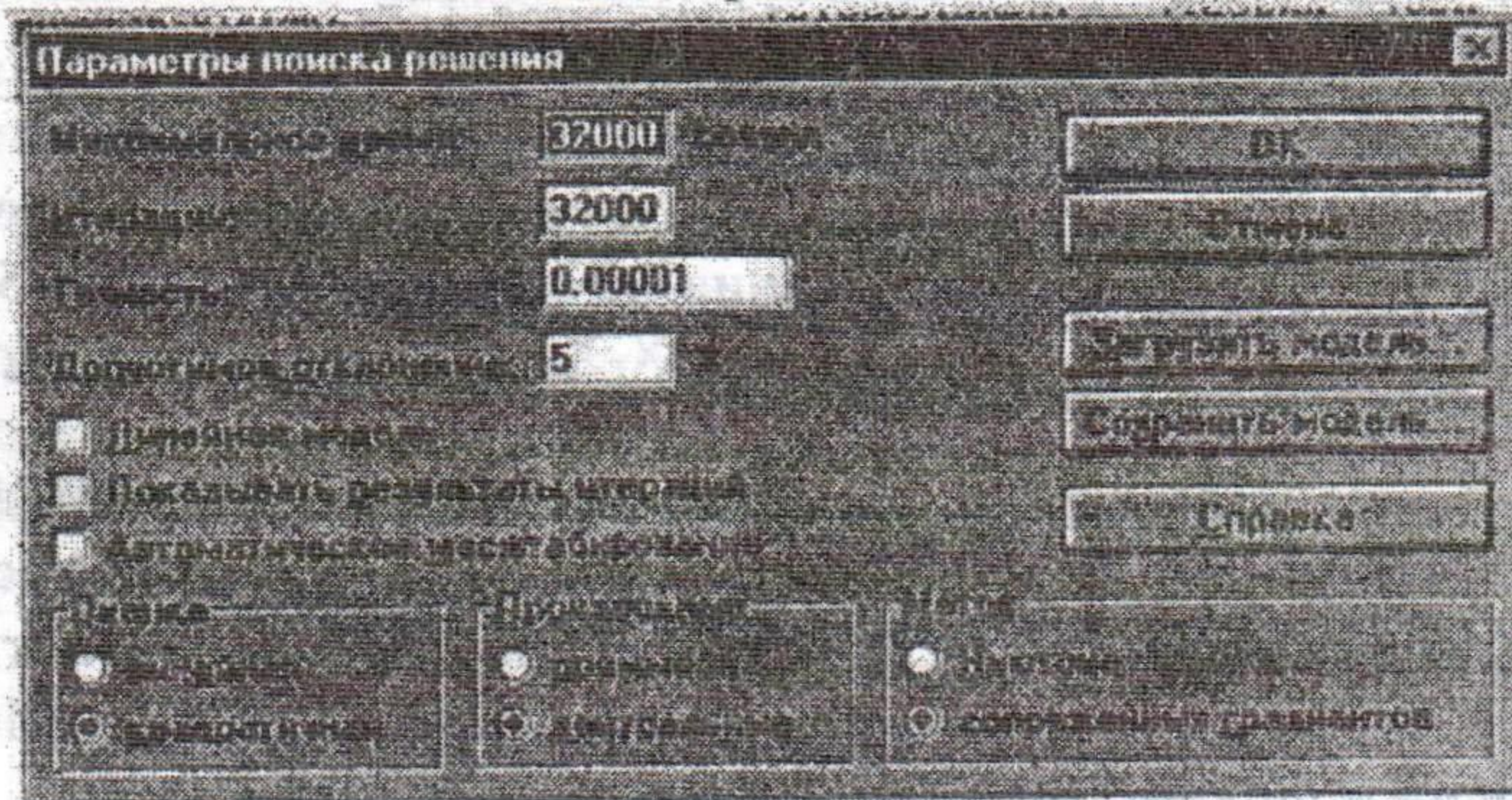
В ячейку A04 вносим формулу расчета квадрата отклонения наблюдаемого значения (AL4) и расчетного значения (AN4) для первого опыта эксперимента: $=(AL4-AN4)^2$. Формула копируется на ячейки A05 - A032. В ячейку A034 заносим формулу для расчета суммы квадратов отклонений:

$=СУММ(A04:A033)$.

Приведенные ниже панели диалога позволяют сформировать модель задачи поиска решения и указать параметры расчета.



Окно диалога "Поиск решения"



Окно диалога "Параметры поиска решения"

Можно изменять условия и варианты поиска решения для линейных и нелинейных задач, а также загружать и сохранять оптимизируемые модели. Значения и состояния элементов управления, используемые по умолчанию, подходят для решения большинства задач. Окно ввода «Точность» служит для задания точности, с которой определяется соответствие ячейки целевому значению или приближение к указанным границам. Низкая точность соответствует введенному числу, содержащему меньше количество десятичных знаков, чем число, используемое по умолчанию, например, 0,0001. Высокая точность увеличит время, которое требуется для того, чтобы сошелся процесс оптимизации. Можно существенно ускорить процесс поиска решения, если задать исходные значения влияющих ячеек модели близкими к искомым результатам.

В данном примере воспользовались флажком «квадратичная» с целью использования квадратичной экстраполяции, которая дает лучшие результаты при решении нелинейных задач. Остальные флажки пользователь может расставить в соответствии с требованиями задачи (случаи их использования объясняются во встроенной системе помощи).

В процессе оптимизации было получены коэффициенты: $Y = 118 + 53$

$$X_1 - 47 X_2 + 25 X_3 + 22 X_4 - 19 X_5 - 13 X_2 X_3 - 13 X_1 X_2 - 11 X_2 X_4 - 15 X_2 X_6 + 16 X_3 X_4$$

В работе [1] В.М. Карпюк получил следующую математическую модель времени расслоения бетонной смеси:

$$Y = 118 + 53 X_1 - 47 X_2 + 25 X_3 + 23 X_4 - 17 X_5 - 13 X_2 X_3 - 13 X_1 X_2 - 13 X_2 X_4 - 16 X_2 X_6 + 15 X_3 X_4$$

Приведенные модели вполне сопоставимы. Таким образом, Microsoft Excel, наряду с существующими является ценным инструментом для исследователя, работающего в области изучения физических процессов, носящих случайный характер.

Литература.

1. Карпюк В.М., Майборода В.Ф. Параметры одностадийного формования комплексных конструкций методом расслоения легкогобетонной смеси. Одесский ЦНТИ, Информлисток №140-86, серия44, 1986. - 4с.
2. Вознесенский В.А. Статистические методы планирования эксперимента в технико-экономических исследованиях. -М: Наука, 1982, - 250с.