ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ СТАТИСТИЧЕСКИМИ СРЕДСТВАМИ EXCEL

THE EXPERIMENTAL DATA BY STATISTICAL MEANS EXCEL

Шевченко И.М.
Костанайский государственный педагогический институт, г.Костанай, Казахстан

Нестационарные процессы диффузии, теплопроводности и многие другие описываются уравнением параболического типа, а стационарные – уравнением Лапласа (эллиптического типа). Колебательные и волновые процессы самой различной природы (электромагнитные, акустические, гидродинамические и др.) описываются одним и тем же волновым уравнением гиперболического типа [1, с. 17].

Приведенные и подобные им примеры, которых на практике очень много, делают соблазнительным изучение исследуемых процессов на основе самых общих их физических особенностей, описываемых соответствующим математическим аппаратом. Частично эта тенденция получила практическое применение даже при изучении физики. Например, в Берклеевском курсе физики все волновые явления независимо от их физической природы излагаются в одном томе «Волны» [2].

Итак, многие явления различной физической природы имеют аналогичные количественные закономерности и описываются с помощью одного и того же математического аппарата. Это обстоятельство делает возможным количественное описание некоторого явления путем исследования другого процесса совершенно иной физической природы.

При анализе данных непланируемых экспериментов, например, при наблюдении неконтролируемых явлений используют регрессионный анализ.

В регрессионном анализе предполагается, что можно прямо или косвенно контролировать одну или несколько независимых переменных $x_1, x_2, ..., x_k$, и их значения вместе с множеством параметров $\theta_1, \theta_2, ..., \theta_k$ определяют математическое ожидание зависимой переменной y. Задача состоит в вычислении оценок параметров с помощью выборочных данных [3, с. 447].

Регрессионная модель должна аппроксимировать совокупность экспериментальных данных. В некоторых случаях исследователю известен точный вид истинной функциональной зависимости между y и $x_1, x_2, ..., x_k$, скажем, $y=\varphi(x_1, x_2, ..., x_k)$. Однако чаще всего истинная функциональная связь неизвестна, и экспериментатору приходится выбирать подходящую функцию для аппроксимации φ. Для аппроксимации широко используются полиномиальные модели [4, с. 282].

Аппроксимируем зависимость давления насыщенного пара от температуры (в мм.рт. ст.), которая была взята из справочника по физике и представлена в следующей таблице.

Таблица 1. Зависимость давления насыщенного пара от температуры

T, ^{0}C	20	40	60	80	90	100	120	140	200	300
p, мм.рт.ст.	17,5	55	149	355	526	760	1489	2711	11660	64450

Для обработки данных будем использовать MS Excel версии 2007 года. Занесем данные таблицы в ячейки A9:B18, разместив таблицу вертикально (это в дальнейшем будет использоваться для расчетов). Сначала построим точечный график функции от таблицы данных p(T) (см. рисунок 1).

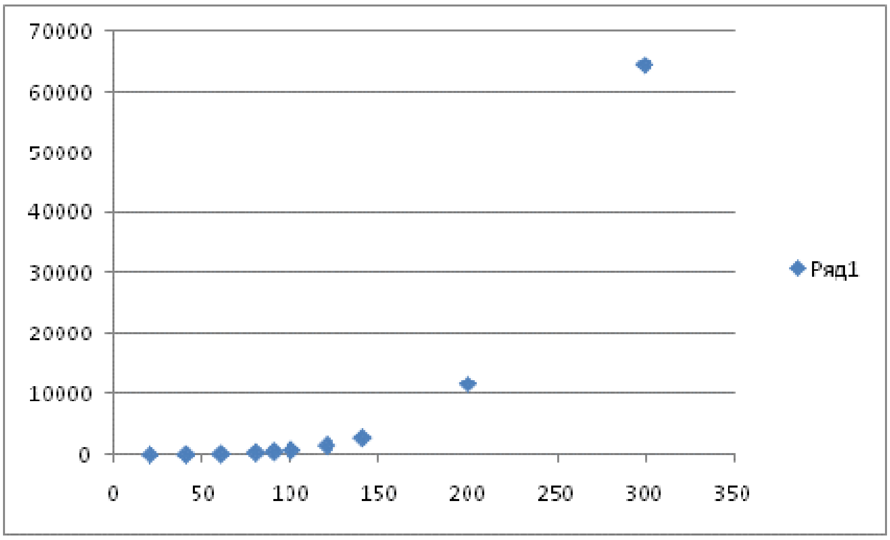

Рисунок 1. График экспериментальных данных

Определимся с типом регрессии. Значения давления увеличиваются с температурой в геометрической прогрессии, а значит нужно использовать степенную регрессию, но учитывая отдаленность последнего значения давления при температуре 300 ^{0}С, воспользуемся полиномиальной регрессией.

Полиномиальная регрессия означает приближение данных (x_i, y_i) полиномом k-й степени.

$$A(x) = a + bx + cx^2 + dx^3 +... + hx^k.$$

При k = 1 полином является прямой линией, при k = 2 – параболой, при k = 3 – кубической параболой и т. д. Как правило, на практике применяются к<5. Для построения регрессии полиномом k-й степени необходимо наличие, по крайней мере, (k +1) точек данных.

Приступим к обработке данных. Для начала найдем коэффициенты полинома. Наиболее подходящее значение k=3, т.е. график функции наших данных приближен к кубической параболе. Проверим данное утверждение.

При k=3, полином будет иметь вид:

$$A(x) = a + bx + cx^2 + dx^3.$$

Где a, b, c, d - коэффициенты полинома. Вычислим их.

В ячейках E11:E14запишем обозначения T0, T1, T2, T3. Для того, чтобы рассчитать значения коэффициентов полинома a, b, c, d выделим ячейку F11, обратимся к мастеру функций и выберем функцию ИНДЕКС. Заполним появившееся диалоговое окно.

Вычисление коэффициентов регрессии осуществляется с помощью функции ЛИНЕЙН():

ЛИНЕЙН(Значения_y; Значения_x; Конст; статистика).

Где значения x и y - это экспериментальные значения T и p.

В результате вычисленное значение появится в ячейке F11.

В итоге, в строке формул должно оказаться следующее:

=ИНДЕКС(ЛИНЕЙН($B$9:$B$18;$A$9:$A$18^{1;2;3});1;4).

Теперь повторим вычисления и найдем значения коэффициентов 1-ой степени:

=ИНДЕКС(ЛИНЕЙН($B$9:$B$18;$A$9:$A$18^{1;2;3});1;3)

2-ой степени:

=ИНДЕКС(ЛИНЕЙН($B$9:$B$18;$A$9:$A$18^{1;2;3});1;2)

и 3-ей степени:

=ИНДЕКС(ЛИНЕЙН($B$9:$B$18;$A$9:$A$18^{1;2;3});1;1).

Теперь мы можем записать общий вид нашего полинома:

$$A(x) = 0,005x^3 - 1,143x^2 + 83,64x - 1573$$

Для проверки нашей функции, вычислим значения полинома в точках, которые трудно представить на графике. Тем самым, мы выполним предсказывания значения давления при промежуточных температурах.

В ячейках A21:D21 сделаем надписи «T», «T^2», «T^3», «полином». Под надписью «T» укажем значения температуры, в которых предсказываем давление: 160, 180, 220, 240, 250, 260, 270, 280, 290.

Под надписью «T^2» вычисляем квадрат температуры:

=A22*A22

и протягиваем маркер для всех значений T^2.

Под надписью «T^3» вычисляем куб температуры:

=B22*A22

и протягиваем маркер для всех значений T^3.

Используя данные приготовления, мы сможем легко вычислить значения полинома при новых температурах. Для этого используем функцию:

=C22*$F$14+B22*$F$13+A22*$F$12+$F$11

где $F$14, $F$13, $F$12, $F$11 - абсолютные адресации на коэффициенты полинома. Протянем маркер для всех имеющихся значений (см. рисунок 2).

21	T	T^2	T^3	полином
22	160	25600	4096000	4346,996168
23	180	32400	5832000	7489,082221
24	220	48400	10648000	18186,97495
25	240	57600	13824000	26254,11644
26	250	62500	15625000	31079,09047
27	260	67600	17576000	36474,27791
28	270	72900	19683000	42471,6372
29	280	78400	21952000	49103,12677
30	290	84100	24389000	56400,70503
31				

Рисунок 2. Предсказывание новых значений давления

Достроим на имеющимся уже графике вычисленные нами точки, выставив для их отображения маркер «+». Как мы видим, только что построенные точки являются продолжением наших экспериментальных данных.

В инструментах Excel имеется возможность предсказывать вид аппроксимирующей функции при помощи линий тренда.

Если правой кнопкой мышки щелкнуть по одной из точек графика, то в контекстном меню можно увидеть команду «Добавить линию тренда». Это и есть необходимая нам линия. Добавляется она в два шага. На первом выбирается нужный нам тип (полиномиальная) и выставляется степень 3, на втором – параметры. Нам важно поставить галочки против слов: «показывать уравнение» и «поместить величину достоверности».

Для оценки достоверности используется квадрат коэффициента корреляции Пирсона (R^2). Если он равен 1, то имеет место полная корреляция с моделью, т.е. точки лежат строго на прямой. В нашей случае, R^2=0,999. Т.е. почти стопроцентное совпадение с экспериментальными данными.

Вид графика после добавления линии тренда представлен на рисунке 3. Выведенное уравнение прямой совпадает с рассчитанным ранее.

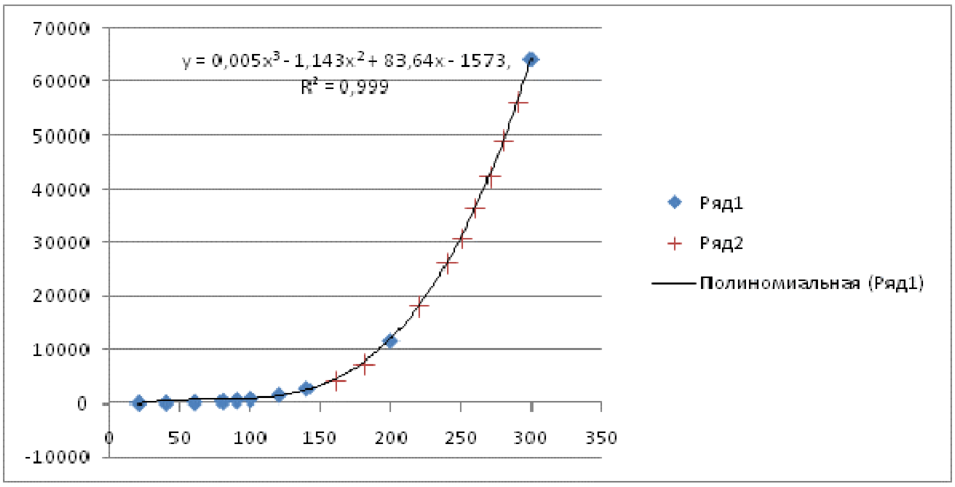

Рисунок 3. Полиномиальная регрессия

Таким образом, можно сделать вывод о том, что зависимость давления насыщенного пара от температуры может быть представлена в виде формулы, которая является вычисленным полиномом:

$$p(T) = 0,005T^3 - 1,143T^2 + 83,64T - 1573.$$

СПИСОК ЛИТЕРАТУРЫ

1. Блохин В.Г., Глудкин О.П., Гуров А.И. и др. Современный эксперимент: подготовка, проведение, анализ результатов. – М.: Радио и связь, 1997. – 232 с.
2. Крауфорд Ф., Берклеевский курс физики: волны (Том 3). – М.: Наука, 1974. – 529 с.
3. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. Методы обработки данных. – М.: Мир, 2002. – 611 с.
4. Монтгомери Д. К. Планирование эксперимента и анализ данных. Л.: Судостроение, 1980. – 384 с.