Подбор эмпирических формул в среде mathcad

Козлова Анастасия Сергеевна
студентка ФГОУ ВО
Курский государственный университет,
колледж коммерции, технологий и сервиса,
Россия, г. Курск

Ефимцева Ирина Борисовна,
Научный руководитель, канд. тех. наук.
преподаватель ФГОУ ВО
Курский государственный университет,
колледж коммерции, технологий и сервиса,
Россия, г. Курск

Пусть требуется найти адекватную модель зависимости величины Y от X. Другими словами, требуется смоделировать влияние фактора X на величину Y. С этой целью проводится эксперимент по схеме, изображённой на рис. 1.

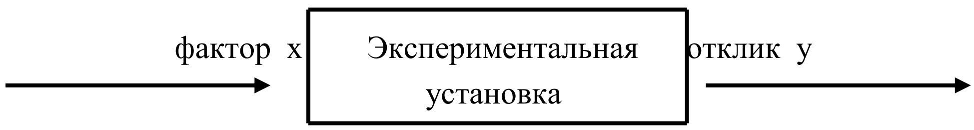

Рис. 1. Экспериментальная установка

Придавая на входе различные значения фактору X, получаем различные значения Y, которые называются откликами. Обычно сначала проводится предварительный эксперимент, целью которого является выдвижение гипотезы (предположения) о виде модели. В ходе проведения предварительного эксперимента фактору придают несколько различных значений и записывают значения отклика в таблицу наблюдений. Затем полученные данные изображают на корреляционном поле и подбирают вид модели, сравнивая с формой эталонных кривых. Вид модели может быть также определён из априорных соображений, исходя, например, из физического смысла наблюдаемых величин.

Пусть в результате предварительного эксперимента выбрана модель.

Далее проводится основной эксперимент, целью которого является получение оценок параметров модели и проверка её адекватности. Составляется план эксперимента, который заключается в оптимальном выборе множества значений фактора, которые предстоит ввести в установку при проведении основного эксперимента.

Рассматриваем наиболее простой вид линейной модели с двумя параметрами $Y = a \cdot f(X) + b$. Требуется оценить параметры a и b и проверить адекватность модели. Составляем план эксперимента в виде: берём n различных значений (уровней) фактора x_1, x_2, …, x_n и для каждого из них повторяем эксперимент m раз. При этом будет получено m значений $y_{i\,1}$, $y_{i\,2}$, …, $y_{i\,m}$ отклика Y для каждого i-го значения фактора.

Получим техническое задание:

1. Провести предварительный эксперимент. Для этого ввести в установку несколько различных значений фактора x и записать значения отклика y.

2. Запланировать эксперимент. Для этого сравнить набор полученных точек с эталонными кривыми. Выбрать из них подходящие и определить: пределы изменения для x, значения для x и число повторений опыта при одном и том же значении x.

3. Провести эксперимент. Для этого ещё раз обратиться к установке, ввести плановые значения фактора x и записать значения для отклика y. Результаты эксперимента оформить в виде таблицы.

4. Провести расчёт по экспериментальным данным. Для этого ввести таблицу наблюдений и выбранные подходящие модели поочерёдно. Записать найденные программой параметры моделей. Проверить однородность строчных дисперсий.

5. Проверить подходящие модели на адекватность и выбрать лучшую из них. Для выбранной модели найти доверительные интервалы для параметров регрессии и проверить их значимость.

Рассмотрим образец выполнения заданий в среде mathcad.

Программа содержит экспериментальную установку и все необходимые команды для проведения расчётов. Часть этих команд скрыта на странице справа

Вводим номер: $n := 22.$

Проводим предварительный эксперимент. Вводим в экспериментальную установку, изображённую в вызванном программном файле, 8 значений для фактора X и записываем значения отклика Y.

Вводим данные предварительного эксперимента:

Число опытов: $m := 8.$

Значения фактора X: $X := (-2\ \ -1\ \ 0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5).$

$xmin := \min(X);\ xmax := \max(X).$

Значения отклика Y:

$Y := (0.096\ \ 0.292\ \ 0.493\ \ 0.71\ \ 0.892\ \ 1.09\ \ 1.301\ \ 1.502).$

Строим графическое изображение результатов предварительного эксперимента:

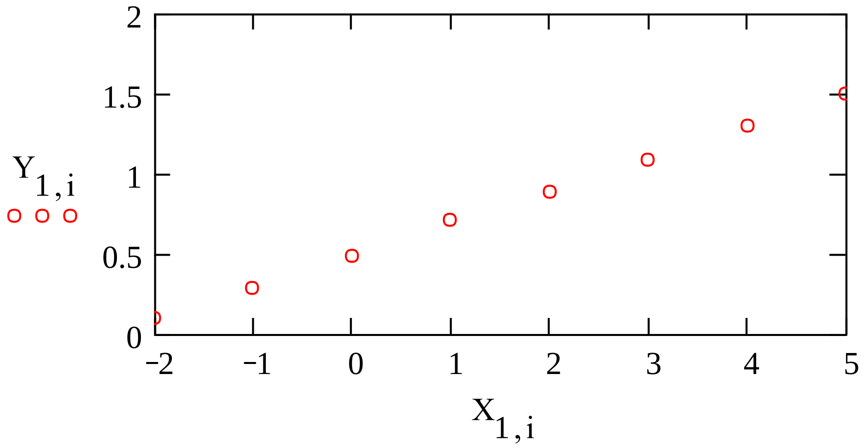

Рис. 2. Расположение точек в предварительном эксперименте.

Сравниваем расположение с эталонными кривыми: $a := 1;\ b := 1.$

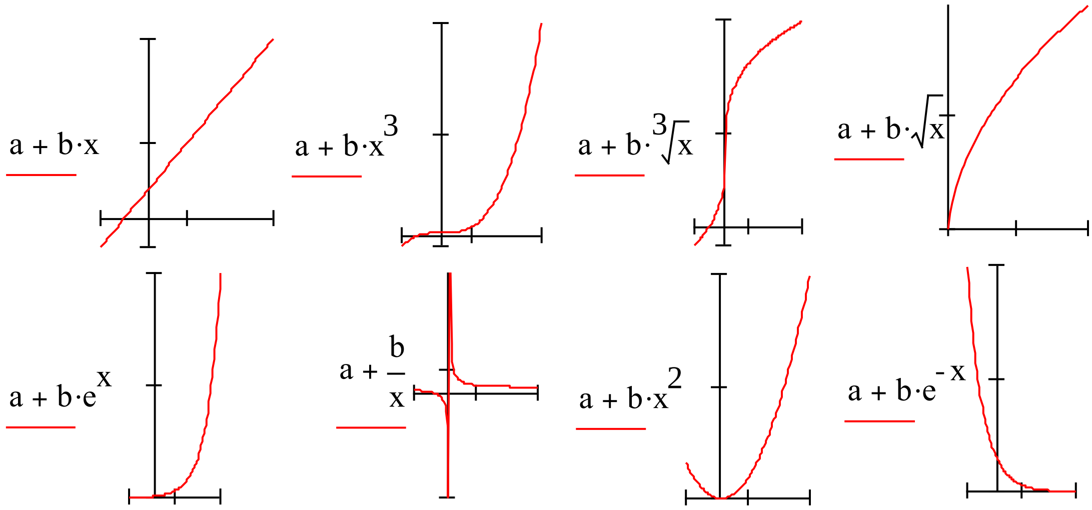

Рис. 3. Таблица эталонных моделей

Выбираем, как наиболее подходящую, первую модель. Далее, планируем основной эксперимент и возвращаемся к экспериментальной установке. Проводим эксперимент и продолжаем дальше обрабатывать результаты эксперимента. Вводим в поле MATHCAD-а данные эксперимента:

Количество различных значений (уровней фактора) для X:

$$m := 8.$$

Число повторений опытов в каждой точке (одинаково для всех точек):

$$k := 5.$$

Строку значений фактора X:

$$X := (-4 \;\; -3 \;\; -2 \;\; -1 \;\; 0 \;\; 1 \;\; 2 \;\; 3).$$

Матрицу значений отклика Y (k строк и m столбцов):

$$Y := \begin{bmatrix} -0.3 & -0.108 & 0.098 & 0.31 & 0.492 & 0.699 & 0.901 & 1.102 \\ -0.31 & -0.107 & 0.092 & 0.308 & 0.493 & 0.691 & 0.903 & 1.101 \\ -0.304 & -0.1 & 0.092 & 0.309 & 0.49 & 0.691 & 0.909 & 1.11 \\ -0.309 & -0.107 & 0.09 & 0.302 & 0.492 & 0.69 & 0.902 & 1.1 \\ -0.296 & -0.108 & 0.093 & 0.309 & 0.498 & 0.692 & 0.902 & 1.101 \end{bmatrix}.$$

Находим строчные средние и дисперсии

$$M(i) := \mathrm{mean}\left(Y^{\langle i\rangle}\right);\qquad D_{1,i} := \frac{\sum_{j=1}^{k}\left[M(i) - \left(Y^{\langle i\rangle}\right)_{j,1}\right]^2}{k-1}.$$

Находим максимальную строчную дисперсию

$$D\max := \max(D).$$

Проверяем однородность строчных дисперсий по критерию Кочрена.

Проверим нулевую гипотезу H_o: D[X1] = D[X2] = ... = D[Xm].

Выбираем доверительную вероятность:

$$P := 0.95.$$

Находим наблюдаемое значение критерия:

$$S := \mathrm{mean}(D)\cdot m;\quad G\text{набл} := \frac{D\max}{S};\ G\text{набл} = 0.305.$$

Чтобы найти табличное значение критерия Кочрена, можно воспользоваться таблицами, взятыми изучебных пособий, так как в версии MATHCAD-а может и не быть специальной функции для его вычисления. Итак, находим табличное значение:

$$G\text{крит} = 0.56.$$

Сравниваем Gнабл с Gкрит. Так как Gнабл < Gкрит, то гипотезу принимаем. (Если бы дисперсии оказались неоднородны, то увеличили бы число опытов в каждой точке и повторили эксперимент).

Вводим в поле MATHCAD-а модель:

$$f(a,b,x) := a + b\cdot x.$$

Теперь находим параметры a и b методом наименьших квадратов.

Находим средние значения отклика Y при каждом значении фактора:

$$i := 1..m;\qquad MY(i) := \mathrm{mean}\left(Y^{\langle i\rangle}\right);$$

$Ycp_{1,i} := MY(i).$

Введём начальные значения искомых параметров (оценивая по графику на глаз):

$$a := 0.9\ ;\ b := 0.2.$$

Вводим минимизируемую функцию (сумму квадратов отклонений) и обозначаем через R вектор искомых параметров.

$$g(a,b) := \sum_{i=1}^{k} \left(f\left(a,b,X_{1,i}\right) - Ycp_{1,i}\right)^2 \ ; \quad R := \begin{bmatrix} a \\ b \end{bmatrix}.$$

Находим параметры a и b при помощи следующей MATHCAD-программы:

```
R := | R ← R
     | for h ∈ 0.01, 0.001
     |   for i ∈ 0..200
     |     for da ∈ -h, 0, h
     |       for db ∈ -h, 0, h
     |         | a ← R₁
     |         | b ← R₂
     |         | R ← | R + [da; db]  if g(a + da, b + db) < g(a, b)
     |         |     | R  otherwise
;
```

$$a := R_1;\ b := R_2.$$

Получаем значения a = 0.499 ; b = 0.201.

Сначала строим график линии с найденными параметрами и оцениваем визуально адекватность модели судя по отклонениям точек от линии.

$$y(x) := f(a,b,x);\ h := \frac{(xmax - xmin)}{100};\ x := xmin, xmin + h .. xmax$$

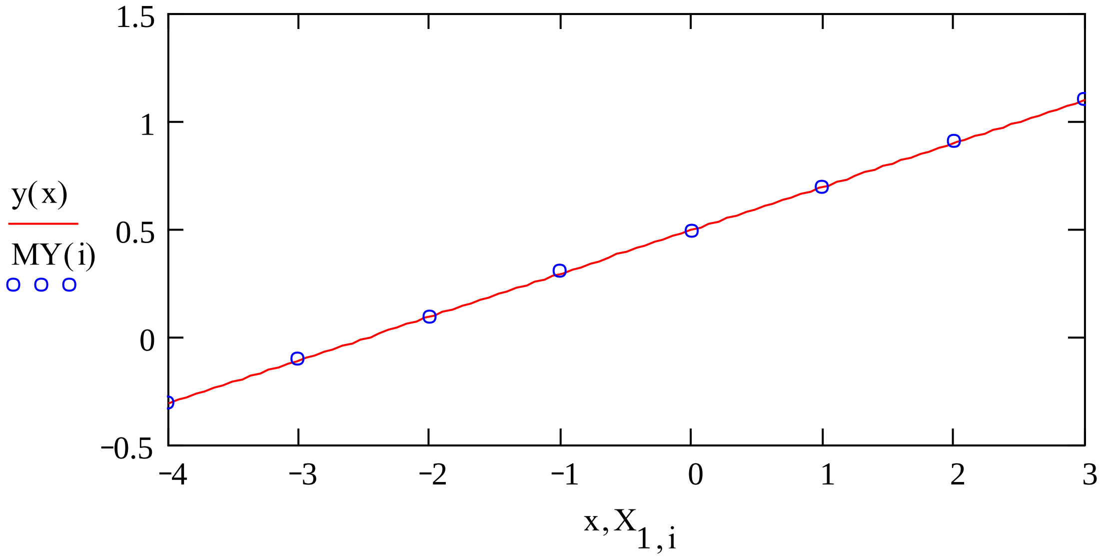

Рис. 4. Взаиморасположение экспериментальных точек и линии

По рис. 4. видно, что расхождения построенной модели (линии) и результатов эксперимента (точек) незначительны. Нужно провести более строгий анализ адекватности модели по критерию Фишера. Находим наблюдаемое и табличное значения критерия:

$$DO := \frac{k}{m-3} \cdot \sum_{i=1}^{m} \left(Ycp_{1,i} - f\left(a,b,X_{1,i}\right)\right)^2;\quad DS := \frac{S}{m}$$

;

$$Fнабл := \frac{DO}{DS \cdot m};$$

$$Fкрит := qF(P, m-3, m\cdot(k-1));$$

$$Fнабл = 1.814;\ Fкрит = 2.512\ .$$

Так как Fнабл < Fкрит - модель адекватна. (Если бы это неравенство оказалось невыполненным, то попробовали бы взять другую модель).

Проверяем значимость параметров a и b, если они близки к нулю. Для значимых параметров строим доверительные интервалы. Находим концы доверительных интервалов и длину этих интервалов. Предварительно находим табличное значение критерия Стьюдента.

$$Ткрит := qt\left[\frac{(P+1)}{2}, m\cdot(k-1)\right];\ \Delta := \sqrt{\frac{DS}{m}}\cdot Ткрит;$$

$$ar := a + \Delta\ ;\ al := a - \Delta\ ;\ br := b + \Delta\ ;\ bl := b - \Delta.$$

Итак, найдены оптимальные значения параметров:

$$a = 0.499\ ;\ b = 0.201.$$

В данном случае проверку значимости параметров проводить не нужно. Однако если бы значения этих параметров были близки к нулю, то проверка гипотезы о значимости a и b (если они близки к нулю) проводится следующим образом:

Для a: $Тнабл := \frac{|a|}{\sqrt{\frac{DS}{m}}}$; $Тнабл = 371.641$; $Ткрит = 2.037$.

Для b: $Тнабл := \frac{|b|}{\sqrt{\frac{DS}{m}}}$; $Тнабл = 149.693$; $Ткрит = 2.037$.

Если для какого-либо из этих коэффициентов выполняется неравенство Тнабл < Ткрит, то этот коэффициент незначим и его следует считать равным нулю.

Построим доверительные интервалы для параметров a и b. Значения на концах интервалов вычислены и равны:

$$al = 0.496;\ ar = 0.502;\ bl = 0.198;\ br = 0.204.$$

Доверительные интервалы имеют вид

$$0.496 < a < 0.502;$$

$$0.198 < b < 0.204.$$

Список литературы

1. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высш. шк., 1977.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высш. шк., 1979.
3. Дьяконов В.П., Абраменкова И.В. Mathcad 7.0 в математике, физике и в Internet. М.: Нолидж, 1998.
4. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. М.: Наука, 1976.
5. Бондарь А.Г., Статюха Г.А. Планирование эксперимента в химической технологии. Киев: Высш. шк., 1976.