Математическое моделирование

Задачи оптимизации

ЗНАЧИМОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

После определения оценок коэффициентов регрессии необходимо проверить гипотезу о значимости коэффициентов b_i. Лучше всего это сделать в виде *нуль-гипотезы*, т.е. гипотезы о равенстве $b_i = 0$.

Если она подтвердится, то коэффициент b_i следует признать статистически незначимым и отбросить из искомой модели; если гипотеза не подтвердится, то соответствующий коэффициент b_i следует признать значимым и включить в модель.

ЗНАЧИМОСТЬ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Проверка гипотезы проводится с помощью *t* - критерия Стьюдента, который при проверке нуль-гипотезы формируется в виде

$$t = \frac{|b_i|}{\sqrt{S^2_{b_i}}} = \frac{|b_i|}{\sqrt{S^2_y}} \cdot \sqrt{N}$$

где – дисперсия ошибки определения коэффициента b_i.
При полном и дробном факторном планировании для всех *i*

$$S^2_{b_i} = \frac{S^2_y}{N}$$

доверительные интервалы $\Delta b_i = \pm t S_{bj}$
Некоторые значения *t*-критерия представлены в табл.

РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

f	Доверительная вероятность						
	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	**1,963**	**3,078**	**6,314**	**12,706**	**31,821**	**63,657**	**636,619**
2	**1,336**	**1,886**	**2,920**	**4,303**	**6,965**	**9,925**	**31,598**
3	**1,250**	**1,638**	**2,353**	**3,182**	**4,541**	**5,841**	**12,941**
4	**1,190**	**1,533**	**2,132**	**2,776**	**3,747**	**4,604**	**8,610**
5	**1,156**	**1,476**	**1,943**	**2,571**	**3,365**	**4,032**	**6,859**
6	**1,134**	**1,440**	**1,895**	**2,447**	**3,143**	**3,707**	**5,959**
7	**1,119**	**1,415**	**1,860**	**2,365**	**2,998**	**3,499**	**5,405**
8	**1,108**	**1,397**	**1,833**	**2,306**	**2,896**	**3,355**	**4,781**

Если величина коэффициента регрессии превышает Δb_j, найденное для q % -го уровня значимости и числа степеней свободы, $f = N(k-1)$, где N – число серий параллельных опытов, k – число параллельных опытов, нуль-гипотеза отвергается, коэффициент считается значимым и его следует включить в искомую модель

РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Статистическая незначимость коэффициента b_i может быть обусловлена следующими причинами:

- уровень базового режима близок к точке частного экстремума по переменной X_i или по произведению переменных;
- шаг варьирования Δx_i выбран малым;
- данная переменная (или произведение переменных) не имеет функциональной связи с выходным параметром y;
- велика ошибка эксперимента вследствие наличия неуправляемых и неконтролируемых переменных.

РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Поскольку ортогональное планирование позволяет определять доверительные границы для каждого из коэффициентов регрессии в отдельности, то если какой-либо из коэффициентов окажется незначимым, он может быть отброшен без пересчета всех остальных. После этого математическая модель объекта составляется в виде уравнения связи выходного параметра y и переменных X_i, включающего только значимые коэффициенты.

ПРОВЕРКА АДЕКВАТНОСТИ

Первый вопрос, который нас интересует после вычисления коэффициентов регрессии, – это проверка ее пригодности или проверка адекватности модели.

Для этого определяют дисперсию адекватности (остаточную дисперсию)

$$S_{ад}^2 = \frac{1}{f}\sum_i \Delta y_i^2$$

где f – число степеней свободы (разность между числом опытов и числом коэффициентов (констант), которые уже вычислены по результатам этих опытов независимо друг от друга)

$$f = N-(k+1),$$

где N – число опытов;

k – число коэффициентов регрессии b_i.

ПРОВЕРКА АДЕКВАТНОСТИ

Так для ПФЭ 2^3 $f = 8–(3+1) = 4$;

для ПФЭ 2^2 $f = 4–(2+1) = 1$.

Вычислим разность $\Delta y_i^2 = (y_i^{эксп} - y_i^{расч})^2$

После этого для проверки адекватности используем критерий Фишера:

$$F = \frac{S_{ад}^2}{S^2} \geq F_{табл}$$

Для определения $F_{\text{табл}}$ необходимо знать число степеней свободы для дисперсий воспроизводимости и адекватности. В табл. представлены некоторые значения критерия Фишера с доверительным уровнем вероятности 95 %.

ФИШЕРОВСКОЕ РАСПРЕДЕЛЕНИЕ ВЕЛИЧИН ДЛЯ *P* = 0,95

Число степеней свободы дисперсии воспроизводимости	Число степеней свободы дисперсии адекватности										
		1	2	3	4	5	6	7	8	9	10
	1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
	2	18,51	19,00	19,16	19,75	19,30	19,33	19,35	19,37	19,38	19,40
	3	10,13	9,55	9,28	9,12	9,01	8,44	8,89	8,85	8,81	8,79
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98

ФИШЕРОВСКОЕ РАСПРЕДЕЛЕНИЕ

Если расчетное значение критерия Фишера не превышает табличного, то с соответствующей доверительной вероятностью модель можно считать адекватной.

Если гипотеза адекватности отвергается, то модель признается неадекватной экспериментальным данным. *Неадекватность модели не означает ее неправильности!*

Неадекватность модели может означать, что не весь перечень влияющих факторов был принят во внимание, или что необходимо перейти к более сложной форме уравнения связи, или выбрать другой шаг варьирования по одному или нескольким факторам и т. п. Однако все достижения неадекватной модели (отсев незначимых факторов, оценка дисперсии эксперимента и другие) остаются в силе

ПРИМЕР

Рассмотрим химический процесс, в котором выход продукта реакции y (%) зависит от температуры реакционной смеси x_1 (°С) и концентрации реагента x_2 (%). Требуется с помощью полного факторного эксперимента найти математическое описание этого процесса в окрестности точки факторного пространства с координатами $x_{01} = 50$ °С и $x_{02} = 25$ %.

При проведении полного факторного эксперимента зададимся условиями, приведенными в табл.

Характеристика эксперимента	x_1, °С	x_2, %
Основной уровень	50	25
Интервал варьирования	5	1
Верхний уровень	55	26
Нижний уровень	45	24

ПРИМЕР

Решение. Математическое описание рассматриваемого процесса будем искать в виде уравнения регрессии $y = b_0 + b_1X_1 + b_2X_2$, где кодированные переменные связаны с температурой и концентрацией следующими соотношениями:

$$X_1 = \frac{x_1 - x_{01}}{\Delta x_1} \qquad X_2 = \frac{x_2 - x_{02}}{\Delta x_2}$$

Матрица планирования и результаты полного факторного эксперимента представлены в таблице

Номер опыта	X_1	X_2	x_1, ^0C	x_2, %	y, %
1	–1	–1	45	24	35,5
2	+1	–1	55	24	38,7
3	–1	+1	45	26	32,6
4	+1	+1	55	26	36,2

ПРИМЕР

На основании результатов полного факторного эксперимента рассчитаем коэффициенты регрессии:

b_0 **= ¼(35,5+38,7+32,6+36,2) = 35,6,**

b_1 **= ¼(–35,5+38,7–32,6+36,2) = 1,95,**

b_2 **= ¼(–35,3–38,7+32,6+36,2)= –1,35.**

Будем считать, что оценка дисперсии среднего значения равна 0,42 (предыдущий пример). Примем также, что с этой величиной связаны три степени свободы

$$f = N(k-1) = 3(2-1) = 3.$$

ПРИМЕР

Ошибку в определении коэффициентов регрессии вычислим как

$$S_b = \sqrt{\frac{S_y^2}{N}} = \sqrt{\frac{0,42}{4}} = 0,32$$

Пользуясь табл. Стьюдента, найдем, что для доверительной вероятности $P = 0,95$ и трех степеней свободы значение критерия Стьюдента $t = 3,18$. Тогда $S_b\, t = 0,32 \cdot 3,18 = 1,03$.

Для оценки значимости коэффициентов регрессии рассмотрим следующие соотношения:

$|b_0| = 35,6 > S_б t$; $|b_1| = 1,95 > S_б t$; $|b_2| = 1,35 > S_б t$.

Видно, что все коэффициенты регрессии значимы. Следовательно, искомое уравнение имеет вид $y = 35,6 + 1,95X_1 - 1,35X_2$.

ПРИМЕР

Для проверки адекватности уравнения регрессии найдем расчетные значения функции отклика:

$\boldsymbol{y}_{\mathbf{1}}^{\text{расч}}$ = 35,6+1,95(–1)–1,35(–1) = 35,0;

$\boldsymbol{y}_{\mathbf{2}}^{\text{расч}}$ = 35,6+1,95(+1)–1,35(–1) = 38,9,

$\boldsymbol{y}_{\mathbf{3}}^{\text{расч}}$ = 35,6+1,95(–1)–1,35(+1) = 32,3,

$\boldsymbol{y}_{\mathbf{4}}^{\text{расч}}$ = 35,6+1,95(+1)–1,35(+1) = 36,2.

ПРИМЕР

Вычислим оценку дисперсии адекватности:

$$S_{ад}^2 = \frac{1}{N-(k+1)}\sum_{j=1}^{N}\left(y_j^{эксп} - y_j^{расч}\right)^2 =$$

$$= \frac{1}{4-3}\left[(35{,}5-35{,}0)^2 + (38{,}7-38{,}9)^2 + (32{,}6-32{,}3)^2 + (36{,}2-36{,}2)^2\right] = 0{,}38.$$

С ней связано число степеней свободы $f = N-(k+1) = 4-3 = 1$. Расчетное значение критерия Фишера

$$F = \frac{S_{ад}^2}{S^2} = \frac{0{,}38}{0{,}42} = 0{,}905$$

Оно не превосходит табличного значения ($F_{таб} = 10{,}13$), следовательно, нельзя сказать, что уравнение регрессии неадекватно.

ИСПОЛЬЗУЕМАЯ ЛИТЕРАТУРА

1. Основы научных исследований. Курс лекций (для студентов инженерных специальностей) / Сост. Н. Г. Бойко, О. В. Федоров – Донецк: ДонНТУ, 2007. – 76 с.