Шашков В.Б.
Оренбургский государственный университет

КОМПЬЮТЕРНАЯ МОДЕЛЬ ЭКСПЕРИМЕНТА

Разработана методика моделирования эксперимента с заданной дисперсией воспроизводимости $s^2_в$ отклика объекта исследования. Проведен компьютерный эксперимент в широком диапазоне дисперсий $s^2_в$ (от 200 до 1600) с десятью параллельными векторами откликов для каждой задачи, позволивший уточнить некоторые закономерности, определяющие точность эксперимента и полиномов регрессии.

Состояние сложных объектов исследования определяется множеством факторов. При рекомендованной форме полинома регрессии, включающей, в частности, все сочетания аргументов-факторов в первой степени [1], при увеличении количества факторов в полиноме количество коэффициентов регрессии в нем быстро нарастает. Так, для пятифакторной задачи полная кубическая форма полинома содержит сорок два коэффициента, а при десяти факторах – уже четыре тысячи сто двадцать. Это означает, что таблица экспериментальных данных должна содержать не менее четырех тысяч сто двадцать одной строки. Поэтому большое количество факторов, включенных в модель эксперимента при его планировании, не только затрудняет проведение эксперимента, но в ряде случаев делает его неосуществимым. В силу этого модель объекта исследования является обычно заведомо неполной, и значение отклика объекта исследования Yg на данной g-ой строке эксперимента определяется соотношением

$$Yg=\varphi(\overline{Xg})+\delta(\overline{Wg}), \qquad (1)$$

где $\varphi(\overline{Xg})$ – функция факторов, включенных в эксперимент,

$\delta(\overline{Wg})$ – функция шума (или функция всех прочих факторов).

Наличие последней составляющей делает последующее значение отклика Yg при наблюдении непредсказуемым и более того – при повторных наблюдениях по режиму данной g-ой строки мы получим ряд разных результатов, т. к. среди факторов *Wg* всегда будут присутствовать факторы со временным дрейфом. Таким образом, за каждым единичным значением отклика Yg на данной строке таблицы экспериментальных данных будет «прятаться» массив значений, являющихся случайной величиной. Согласно условиям процедуры регрессионного анализа, этот массив имеет нормальное распределение с математическим ожиданием M{Yg} и дисперсией воспроизводимости $\sigma^2_в$. Отличаться массивы разных строк таблицы экспериментальных данных будут только математическим ожиданием. Дисперсия $\sigma^2_в$ зависит только от природы объекта исследования и для всех строк таблицы будет одинаковой. С учетом этих условий и будем строить модель эксперимента.

Для начала следует построить табличную **функциональную** зависимость между аргументами **x** и функцией **y**. Примем пятифакторную модель с объемом каждого фактора n=90, связанную полиномом седьмой степени с сорока шестью коэффициентами регрессии. Ранее мы описывали сокращенную запись степенных полиномов [2] в индексах при коэффициентах регрессии. Например, члены полинома

$$\ldots+b_{45}\times x4\times x5+b_{123}\times x1\times x2\times x3+\ldots$$

записываются как *…45 123…*. Тогда принятая нами форма полинома может быть записана следующим образом:

0 1 2 3 4 5 12 13 14 15 23 24 25 34 35
45 11 22 33 44 55 111 222 333 444 555
1111 2222 3333 4444 5555 155 255 355
455 500 166 266 366
466 566 177 277 377 477 577. (2)

Для начала тем или иным способом компануем пять векторов **x**, следя чтобы парные коэффициенты корреляции векторов были менее единицы (по модулю). Затем, используя программу регрессионного анализа, «подгоняем» их под уравнение (2), формируя функцию **y**. В итоге получаем табличную функцию, представленную в таблице 1 (для сокращения в таблице представлена каждая шестая строка).

В данной работе для оценки точности полиномов использовали следующие характеристики.

Таблица 1. Исходная табличная функция

g	X1	X2	X3	X4	X5	Yg
1	99.67	62.61	89.18	2.59	102.30	230.84
6	94.38	32.27	85.05	7.52	96.87	162.73
12	88.75	8.86	80.49	13.36	90.44	139.32
18	82.89	1.82	77.35	19.15	83.90	125.06
24	76.59	12.95	75.16	25.33	77.41	126.74
30	70.67	38.64	68.22	36.57	70.90	140.66
36	64.06	7.51	67.68	38.66	55.11	122.02
42	58.12	8.21	63.07	29.66	60.66	124.33
48	52.63	9.40	81.17	20.66	65.32	116.55
54	46.56	10.60	100.34	11.66	72.78	136.59
60	40.86	11.80	121.21	17.36	78.86	172.64
66	34.44	22.22	62.50	26.36	85.39	109.70
72	28.03	34.22	46.39	44.72	92.01	136.62
78	22.93	48.03	34.03	65.28	98.59	250.03
84	43.53	63.63	22.00	30.32	105.12	253.94
90	129.97	81.03	10.00	12.20	111.63	264.01

Остаточная дисперсия полинома

$$S^2_{ost} = \frac{\sum_{g=1}^{n}(Yg - YRg)^2}{n-(k+1)}, \qquad (3)$$

где *Yg* и *YRg* – табличное и расчетное значение отклика;

(k+1) – количество коэффициентов в полиноме.

Корреляционное отношение

$$\theta = \sqrt{\frac{\sum_{g=1}^{n}(Yg - YRg)^2}{\sum_{g=1}^{n}(Yg - \overline{Yg})^2}}, \qquad (4)$$

где $\overline{Yg}$ – среднее арифметическое по табличному вектору откликов.

Интервальная оценка M{YRg} в виде разности «правая граница минус левая граница» интервальной оценки.

Решение задачи из таблицы 1 с использованием уравнения (2) показало, что остаточная дисперсия и интервальная оценка равны нулю, а корреляционное отношение – единице. Таким образом, каждое значение *Yg* в таблице 1 можно считать математическим ожиданием будущего массива случайной величины на g-той строке таблицы. Задача моделирования эксперимента, таким образом, сводится к генерации случайной величины на g-той строке таблицы с нормальным распределением, математическим ожиданием M{Yg} и дисперсией воспроизводимости $\sigma^2_в$.

Центральная предельная теорема А.М. Ляпунова гласит, что если случайная величина представляет собой сумму большого числа независимых случайных величин, влияние каждой из которых на всю сумму ничтожно, то сумма имеет распределение, близкое к нормальному. Тогда каждое значение генерируемой случайной величины будет суммой ряда значений независимых равномерно распределенных в интервале (0,1) случайных величин с математическим ожиданием n/2 и дисперсией n/12 (где n – количество слагаемых).

Для каждой g-той строки таблицы 1 (всего девяносто строк) составляем массив из w значений, каждое из которых является суммой n слагаемых. Каждое значение из массива w нормируем по соотношению

$$u[i] = \frac{w[i] - n/2}{\sqrt{n/12}}, \qquad (5)$$

и тогда в случае нормального распределения при достаточно большом объеме массива U математическое ожидание M{U} будет близко нулю, а дисперсия – единице.

Теперь создаем массив случайной величины на каждой g-той строке таблицы 1

$$Yg[i] = \sigma^2_в \times U[i] + M\{Yg\}, \qquad (6)$$

задаваясь дисперсией воспроизводимости $\sigma^2_в$ и используя значение *Yg* на этой строке в качестве *M{Yg}*. Каждый массив значений проверяем на соответствие начальным условиям: математическое ожидание и дисперсия нормированной величины должны быть равны нулю и единице (в пределах ±0,05), а практически полученное значение *M{Yg}* и $\sigma^2_в$ должны быть близки заданным. Для $\sigma^2_в$, например, допускалось отклонение от заданного значения не более чем на ±5. Соответствие распределения случайной величины нормальному проверялось также по коэффициенту асимметрии $\alpha\{Yg\}$ и эксцессу $\varepsilon\{Yg\}$, которые должны были отличаться от нуля менее чем на 0,1. Если эти условия не выполнялись, массив браковался.

Из каждого массива случайным образом производилась выборка объемом десять значений; при повторе этой процедуры девяносто раз, в соответствии с таблицей 1, создавалось десять параллельных векторов откликов Yg для одной и той же задачи регрессии. В таблице 2 частично представлены описанные данные для массива с $\sigma^2_в$, равной 400.

Среднее арифметическое по столбцу 7 для $\sigma^2_в$ составило 400,67. Решив задачу с вектором $M\{Yg\}$ в качестве вектора откликов (столбец 6), получили данные, представленные в таблице 3.

Они свидетельствуют, что полученные практически математические ожидания массивов $M\{Yg\}$ близки функции Yg из таблицы 1.

Было создано десять комплексов, аналогичных приведенным в таблице 2, с $\sigma^2_в$, равными 200, 400, 600, 800, 1000, 1200, 1400 и 1600. Каждый комплекс содержал по десять задач регрессии, состоящих из откликов, сформулированных из данного массива w, и факторов Xg из таблицы 1. Количество слагаемых n (в интервале (0,1)) составляло сто значений. Массивы с $\sigma^2_в$, равными 200 и 1000, создавались дважды (для контроля воспроизводимости результатов).

При решении этих задач определялись все характеристики, приведенные в таблице 3 и рабочее значение статистики Фишера (для вероятности 0,95)

$$F_{раб} = \frac{S^2_{ост}}{\sigma^2_в} \text{ или } F_{раб} = \frac{\sigma^2_в}{S^2_{ост}}, \qquad (7)$$

где в числитель помещалась большая по значению дисперсия. Число степеней свободы для остаточной дисперсии составляло $n-(k+1)$, те 90 – 46=44. Для $\sigma^2_в$ оно равнялось 89. Критическая граница интервала $F_{кр}$ распределения Фишера для первого отношения составила 1,51 и 1,56 – для второго.

Следует отметить, что наилучшим решением задач регрессии является такое, при котором $S^2_{ост}$ и $\sigma^2_в$ являются статистически неразличимыми, и поэтому неравенство $\sigma^2_в > S^2_{ост}$ невозможно. Поэтому второе отношение (7) означает, что решение «перева-

Таблица 2. Экспериментальные данные для массива с $s^2_в$=400

g	Yg7	Yg8	Yg9	Yg10	M{Yg}	$\sigma^2_в\{Yg\}$
1	2	3	4	5	6	7
1	221.80	195.20	255.90	223.70	231.03	402.94
6	155.00	144.20	158.40	155.10	162.38	400.64
12	131.90	113.10	157.70	136.50	138.94	401.00
18	115.00	148.10	133.60	123.30	125.46	400.21
24	99.80	112.60	117.30	121.00	126.61	398.70
30	118.90	122.90	172.70	160.40	140.85	400.47
36	119.50	145.50	132.80	80.00	121.98	399.32
42	133.80	129.80	119.10	116.80	124.00	403.41
48	140.80	109.60	109.90	77.80	116.39	403.86
54	125.10	127.50	114.20	149.00	136.77	400.89
60	157.90	205.40	222.80	162.50	172.91	402.60
66	93.50	120.00	115.60	110.10	109.91	398.47
72	161.10	96.80	172.10	102.50	136.29	402.89
78	273.60	243.70	197.80	236.50	249.87	400.20
84	282.70	222.20	301.70	221.60	254.42	399.59
90	273.30	253.00	273.70	261.10	264.32	404.44

Таблица 3. Характеристики уравнений с вектором откликов $M\{Yg\}$

Характеристика	$\sigma^2_в$=199,38	$\sigma^2_в$=400,67	$\sigma^2_в$=1000,06
Ост. дисперсия $S^2_{ост}$	0,085	0,106	0,368
Коррел. отнош. θ	0,99999	0,99999	0,99996
Интервал M{Yg}	0,82	0,92	1,70

лило» порог ($S^2_{ост} \approx \sigma^2_в$) и начало воспроизводить ошибку эксперимента в определении отклика *Yg*, приближаясь к экспериментальному облаку точек. Итак, имеем десять групп, каждая из которых имеет свою дисперсию $\sigma^2_в$ и десять задач, набранных в массиве откликов данной группы с объемом w=10000. Характеристики точности решений этих задач (с использованием уравнения (2), которое было математической моделью для задачи таблицы 1) представлены в таблице 4 в порядке возрастания $S^2_{ост}$.

В таблице полужирным шрифтом выделены «предельные» значения переменных: во-первых, значения, попавшие за критическую границу интервала распределения Фишера (таких значений четыре) или лежащие близь границы (это значения более 1,4 – таковых оказалось десять). Эти значения отмечены звездочкой. Значения $F_{раб}$, у которых отношения дисперсий $S^2_{ост}$ и $\sigma^2_в$ менее 1,1 (т. е. они «формально совпадают») подчеркнуты.

В таблице не помещены данные по двум векторам откликов: по вектору y2 и y9. По вектору y2 для $\sigma^2_в$, равной 1199,45, имеем отношение $F_{раб}$, равное **1,459***, а для группы с $\sigma^2_в$ 1600,59 – **1,407***. По вектору y9 для группы с $\sigma^2_в$, равной 1199,45, имеем «закритическое» значение $F_{раб}$ – **1,518***. В группе с $\sigma^2_в$, равной 1400,49 и 1600,59, имеем значение $F_{раб}$ в **1,037** и **1,012**.

Таблица 4. Результаты расчета задач с разной дисперсией $s^2_в$

$\sigma^2_в$	показатели	y1	y3	y4	y5	y6	y7	y8	y10
199,38	$S^2_{ост}$	164,99	186,25	200,64	213,71	215,93	230,35	246,10	269,18
	θ	0,9840	0,9815	0,9790	0,9783	0,9790	0,9770	0,9756	0,9719
	Int	36,09	38,34	39,79	41,07	41,28	42,64	44,67	46,09
	$F_{раб}$	1,208	**1,070**	**1,006**	**1,072**	**1,083**	1,155	1,234	1,350
199,74	$S^2_{ост}$	132,24	158,59	160,84	198,93	198,95	223,76	229,49	234,56
	θ	0,9863	0,9825	0,9857	0,9807	0,9788	0,9772	0.9777	0,9786
	Int	27,13	35,38	36,63	39,62	39,63	42,02	42,56	43,03
	$F_{раб}$	**1,510***	1,259	1,242	**1,004**	**1,004**	1,120	1,150	1,174
400,67	$S^2_{ост}$	341,33	379,22	397,75	406,00	460,76	468,84	482,22	559,13
	θ	0,9709	0,9623	0,9654	0,9608	0,9585	0,9563	0,9557	0,9490
	Int	51,90	54,71	56,03	56,61	60,30	60,83	61,92	66,43
	$F_{раб}$	1,174	**1,057**	**1,007**	**1,015**	1,150	1,170	1,203	1,365
600,29	$S^2_{ост}$	452,47	499,36	528,55	542,70	607,13	653,47	694,04	759,51
	θ	0,9636	0,9538	0,9519	0,9550	0,9501	0,9498	0,9448	0,9364
	Int	59,76	62,78	64,59	65,45	69,22	71,82	74,01	77,42
	$F_{раб}$	1,327	1,202	1,136	1,106	**1,011**	**1,089**	1,156	1,265
799,78	$S^2_{ост}$	681,18	699,39	754,96	825,87	899,08	904,67	911,42	1147,5
	θ	0,9436	0,9519	0,9351	0,9226	0,9239	0,9354	0,9356	0,9057
	Int	73,32	74,30	77,19	80,74	84,24	84,50	84,81	95,17
	$F_{раб}$	1,174	1,143	**1,059**	**1,033**	1,124	1,131	1,140	**1,435***
1000,06	$S^2_{ост}$	586,65	782,85	877,76	939,17	975,21	992,82	996,26	1163,5
	θ	0,9530	0,9348	0,9314	0,9266	0,9240	0,9275	0,9089	0,9073
	Int	68,05	78,60	83,23	86,10	87,73	88,52	88,67	95,83
	$F_{раб}$	**1,705***	1,278	1,139	**1,065**	**1,025**	**1,007**	**1,004**	1,163
999,87	$S^2_{ост}$	693,69	920,94	962,87	983,20	1052,8	1053,7	1116,1	1174,8
	θ	0,9406	0,9223	0,9170	0,9204	0,9157	0,9155	0,9087	0,9232
	Int	73,99	85,26	87,18	88,09	91,15	91,20	93,86	96,29
	$F_{раб}$	**1,441***	**1,086**	**1,038**	**1,017**	**1,053**	**1,054**	1,116	1,175
1199,45	$S^2_{ост}$	812,33	831,96	1016,4	1023,4	1262,2	1268,9	1520,3	2054,9
	θ	0,9390	0,9381	0,9247	0,9100	0,9062	0,9180	0,8806	0,8620
	Int	80,07	81,03	89,57	89,88	99,81	100,07	109,54	127,33
	$F_{раб}$	**1,477***	**1,442***	1,180	1,172	**1,052**	**1,058**	1,267	**1,713***
1400,49	$S^2_{ост}$	1017,4	1139,0	1223,9	1229,5	1256,0	1351,3	1404,6	1566,2
	θ	0,9199	0,9267	0,9249	0,9169	0,8947	0,9009	0,9111	0,8835
	Int	89,61	94,81	98,28	98,51	99,57	103,27	105,29	111,18
	$F_{раб}$	1,376	1,230	1,144	1,139	1,115	**1,036**	**1,003**	1,118
1600,59	$S^2_{ост}$	1050,2	1393,6	1437,0	1467,3	1483,6	1543,7	1597,4	2378,9
	θ	0,9410	0,9130	0,8926	0,9122	0,8829	0,9075	0,8851	0,8489
	Int	91,04	104,88	106,50	107,61	108,21	110,38	112,28	137,03
	$F_{раб}$	**1,524***	1,148	1,114	**1,091**	**1,079**	**1,024**	**1,002**	**1,486***

Таблица 5. Связь корреляционного отношения и дисперсии $s^2_в$

$\sigma^2_в$	Корреляционное отношение	$\sigma^2_в$	Корреляционное отношение
199,38	от 0,9815 до 0,9790	1000,06	от 0,9266 до 0,9089
199,74	от 0,9807 до 0,9788	999,87	от 0,9204 до 0,9155
400,67	от 0,9623 до 0,9608	1199,45	от 0,9180 до 0,9062
600,29	от 0,9501 до 0,9498	1400,49	от 0,9111 до 0,8940
799,78	от 0,9351 до 0,9226	1600,59	от 0,9075 до 0,8851

Как видим, в целом величины $S^2_{ост}$ и $\sigma^2_в$ являются статистически неразличимыми за исключением четырех значений из ста, что достаточно близко к теоретическому – для вероятности в 0,95. Отметим также, что близкие значения $S^2_{ост}$ и $\sigma^2_в$, в целом располагаются в центральной части строк таблицы, но с повышением дисперсии $\sigma^2_в$ имеют явную тенденцию к сдвигу вправо. Число таких значений колеблется от 2 до 5, т. е. достаточно велико. Однако, снимая с объекта в ходе эксперимента только один вектор откликов, всегда имеем возможность ошибки, попадая либо в левый, либо в правый край таблицы 4. Главным ориентиром является дисперсия воспроизводимости $\sigma^2_в$, но для реальных объектов исследования она известна далеко не всегда. Определенную информацию по поводу правильности решения несут и другие характеристики полинома – корреляционное отношение θ и интервальная оценка для *M{Yg}*. Посмотрим, например, как меняется корреляционное отношение θ у «формально совпадающих» значений $S^2_{ост}$ с изменением $\sigma^2_в$ – смотрите таблицу 5.

Таким образом, чем больше остаточная дисперсия $\sigma^2_в$, тем ниже должно быть корреляционное отношение; чрезмерное приближение его к единице будет свидетельствовать о том, что результаты расчета воспроизводят ошибку эксперимента.

Результаты настоящей работы позволяют сделать следующие выводы.

1. Разработана методика моделирования многофакторного эксперимента, включающая:

– эмпирический подбор требуемого количества векторных аргументов-факторов; парные коэффициенты корреляции которых меньше |±1| и вектора откликов *Yg*, образующих в совокупности задачу регрессии:

– решение задачи с использованием алгебраического степенного полинома, выбранного в качестве истинной математической модели табличной функции (в данной работе принято уравнение (2). Полученный вектор откликов и принимается как функция данной задачи;

– генерируем массив случайных величин с нормальным распределением объемом w=10000; в нормированной форме математическое ожидание его должно быть близко нулю, а дисперсия – единице.

– при соблюдении последнего условия переводим массив в ненормированную форму по условию $Yg[i]= \sigma^2_в \times U[i]+M\{Yg\}$, где *M{Yg}* есть значение табличной функции на данной g-строке таблицы, а дисперсия воспроизводимости $\sigma^2_в$ принимается по условиям эксперимента. Извлекая из массива выборку в q элементов (по числу факторов – в данной работе q=90), получаем **таблицу экспериментальных данных**.

2. Проведено сто параллельных экспериментов:

– из десяти задач регрессии, сформированных из одного и того же массива с заданной дисперсией воспроизводимости.;

– из десяти групп задач (по десяти задач в каждой) для дисперсий воспроизводимости $\sigma^2_в$, равных 200, 200, 400, 600, 800, 1000, 1000, 1200, 1400, 1600. Результаты представлены в таблицах 4 и 5.

Список использованной литературы:

1. Иванов А.З., Круг Г.К., Филаретов Г.Ф. Статистические методы в инженерных исследованиях. Регрессионный анализ. – М.:МЭИ, 1977.-243 с., ил.
2. Шашков В.Б., Табилов А.Р., Жарова Е.С., Жаров А.М. Искусственное синтезирование задач многофакторной и многостепенной регрессии в учебных целях. В сб. «Современные информационные технологии в науке, образовании и практике».– Оренбург, ИПК ОГУ, 2002. – с. 5-11.

04.12.06 г.