ЭКОНОМИКА
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ

1965
том I, вып. 5

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЗАИМОДЕЙСТВИЯ ФАКТОРОВ

М. С. МИНЦ
(МОСКВА)

Целью построения математических моделей, как правило, является установление и количественное выражение связи между факторами или их группами для познания и рационального управления развитием того или иного процесса или для выявления причин, обусловивших тот или иной результат.

В настоящей статье рассмотрены следующие методы установления зависимости результативного признака (функции) от обусловивших его величину факторов: метод последовательной детализации первоначальной модели; метод одноэтапной многофакторной группировки; метод двухэтапной многофакторной группировки; метод приближенного определения формы связи парных и множественных статистических зависимостей.

При рассмотрении методов многофакторных группировок выявлено влияние количества учтенных аргументов на количество членов функции многих переменных и возможные ограничения целей исследований, которые позволяют упростить результативные уравнения. Здесь же показаны подходы использования методов многофакторной группировки для обобщения статистических данных.

Метод последовательной детализации первоначальной модели имеет важное методологическое значение для установления количественного влияния факторов-аргументов на результативный признак. Первоначальная модель основывается обычно на непосредственном описании (в общем или конкретном виде) экономической или физической модели, которая характеризует предмет или процесс в интересующих исследователя аспектах. Так, для того чтобы выявить причины, обусловившие уровень затрат на здание, представим первоначально его себестоимость y в виде суммы произведений количества конструктивных элементов $f_i (i = 1, 2, 3, \ldots, n)$ на себестоимость единицы каждого из них a_i

$$y = \sum_{i=1}^{n} f_i a_i. \tag{1}$$

Однако это уравнение не раскрывает, чем обусловлены величины f_i и a_i. Чтобы выяснить, от чего зависит уровень этих величин, их следует расшифровать, детализировать.

Проведем сначала детализацию величин оценок a_i. Для этого выразим a_i (например, себестоимость наружных стен) суммой произведений

$$a_i = \sum_{j=1}^{m} v_{ij} b_j, \tag{2}$$

где v_{ij} ($j = 1, 2, 3, \ldots, m$) — количество ресурсов j на единицу конструкций i, умноженное на себестоимость единицы каждого ресурса b_j, не зависимую от области применения a_i.

Подставив уравнение (2) в уравнение (1), получим

$$y = \sum_{i=1}^{n} f_i \sum_{j=1}^{m} v_{ij} b_j. \tag{3}$$

Но b_i можно также представить суммой произведений

$$b_j = \sum_{k=1}^{l} w_{ijk} c_k, \tag{4}$$

где w_{ijk} — количество ресурсов k ($k = 1, 2, 3, \ldots, l$) на единицу продуктов j, умноженное на себестоимость единицы каждого ресурса c_k.

Подставив уравнение (4) в уравнение (3), получим

$$y = \sum_{i=1}^{n} f_i \sum_{j=1}^{m} v_{ij} \sum_{k=1}^{l} w_{ijk} c_k.$$

Продолжение подобной расшифровки до некоторого заданного этапа приводит к выражению

$$y = \sum_{i=1}^{n} f_i \sum_{j=1}^{m} v_{ij} \sum_{k=1}^{l} w_{ijk} \ldots \sum_{p=1}^{s} u_{ijk\ldots rp}\, q_p, \tag{15}$$

где $u_{ijk\ldots rp}$ — количество ресурса p на единицу продукта r, учтенного на предпоследнем исследуемом этапе, а q_p — себестоимость единицы ресурса p, которая по условию задачи не подлежит расшифровке.

Рассмотренную выше последовательную расшифровку значений a_i, b_j, $c_k, \ldots, q_{p-1}$ назовем детализацией оценок.

Чтобы выявить влияние на себестоимость дома количеств ресурсов, следует расшифровать в уравнениях (1)—(5) значения f_i, v_{ij}, $w_{ijk}, \ldots, u_{ijk\ldots rp}$.

Такую расшифровку, в отличие от предыдущей, назовем детализацией количеств.

Например, при измерении f_i в потребительских единицах (в квадратных метрах стен, перегородок, перекрытий, окон, дверей и т. д.) эти величины будут зависеть только от объемно-планировочных параметров дома — длины, ширины, высоты этажа, числа этажей и т. д.

При этом, например, площадь наружных стен конкретного дома можно выразить

$$f_{\text{нс}} = (2A + 2B) h Э - F_{\text{пр}}, \tag{6}$$

где A — длина дома, B — его ширина, h — высота этажа, $Э$ — число этажей, $F_{\text{пр}}$ — площадь проемов в наружных стенах.

Аналогично, могут быть последовательно выражены остальные значения f_i, v_{ij}, $w_{ijk}, \ldots, u_{ijk\ldots rp}$ — в виде функций от соответствующих параметров, характеризующих объемно-планировочные конструктивные и технологические решения

$$f_i = \varphi_i(z_1, z_2, \ldots, z_б);$$

$$v_{ij} = \nu_{ij}(z_1, z_2, \ldots, z_б, \ldots, z_в);$$

$$w_{ijk} = \psi_{ijk}(z_1, z_2, \ldots, z_в, \ldots, z_д);$$

$$u_{ijk\ldots rp} = \omega_{ijk\ldots rp}(z_1, z_2, \ldots, z_я),$$

где $z_1, z_2, \ldots, z_я$ — факторы, определяющие величины f_i, v_{ij}, w_{ijk}, ..., $u_{ijk\ldots rp}$. Подставив эти функции в уравнение (5), получим более полное выражение влияния отдельных факторов на себестоимость жилого здания, а именно:

$$y = \sum_{i=1}^{n} \varphi_i(z_1, z_2, \ldots, z_б) \sum_{j=1}^{m} \nu_{ij}(z_1, z_2, \ldots, z_б, \ldots, z_в) \times$$

$$\times \sum_{k=1}^{l} \psi_{ijk}(z_1, z_2, \ldots, z_в, \ldots, z_д) \sum_{p=1}^{s} \omega_{ijk\ldots rp}(z_1, z_2, \ldots, z_д, \ldots, z_я). \quad (7)$$

Величины количеств ресурсов f_i, v_{ij}, w_{ijk}, ..., $u_{ijk\ldots rp}$ в уравнении (5) для конкретных объектов являются постоянными и могут быть измерены или учтены при производстве или рассчитаны при проектировании. Зависимость этих величин от обусловивших их факторов показана в уравнении (7).

Если этим уравнением выразить влияние отдельных факторов на затраты по определенному классу зданий и методов производства, то с его помощью можно решить задачу оптимизации проектных планировочных и конструктивных решений домов, а также методов заводского и строительного производства.

Вопросы использования этой методики для разнообразных экономических расчетов и исследований выходят за рамки настоящей статьи.

Здесь же необходимо отметить, что при сопоставлении конкретных объектов обычно ограничиваются одной расшифровкой оценок, а использование методики для оптимизации проектных решений зданий пока ограничивалось одной или двумя ступенями детализаций. Это обусловлено сложностью и трудоемкостью последовательной детализации первоначальной модели, а также громоздкостью результативных уравнений, что, в свою очередь, является следствием учета внутренних взаимосвязей факторов.

Во многих случаях допустимо не анализировать все внутренние промежуточные связи от исходных причин до результата. Например, не всегда обязательно знать, как влияет размер квартиры на количество стен, перегородок, перекрытий и т. д., а эти количества конструкций на себестоимость 1 *кв. м* площади дома.

Методы, которым посвящено дальнейшее изложение, основаны на изучении непосредственной зависимости конечного результата от исходных для исследователя факторов, минуя расшифровку взаимодействия промежуточных внутренних факторов.

Метод одноэтапной многофакторной группировки предусматривает одноэтапное определение формы связи и параметров уравнения. Он основан на изучении влияния каждого отдельного исходного фактора-аргумента на результативный признак при остальных исходных факторах, последовательно закрепляемых на разных уровнях.

Для иллюстрации выявим зависимость себестоимости 1 *кв. м* жилой площади от средней площади квартир в зданиях, этажности и секционности при предположении, что все остальные факторы полностью закреплены на некотором определенном уровне и благодаря этому их влияние на интересующий нас результат исключено.

Решим эту задачу сначала для трехсекционных домов. Для этого разработаем экспериментальные проекты 15 трехсекционных домов: пяти трехэтажных домов со средней площадью квартиры в здании в 30, 32, 34, 36 и 38 $м^2$, а также пяти четырехэтажных и пяти пятиэтажных домов с такими же средними площадями квартир.

Затем по всем 15 домам рассчитаем себестоимость 1 $м^2$ жилой площади. Результаты расчетов сведем в табл. 1. Оговорим, что данные, приведенные в этой таблице, условны и составлены таким образом, чтобы яснее выявить суть методики.

Таблица 1

Зависимость себестоимости 1 $м^2$ жилой площади y трехсекционных домов от средней жилой площади квартир и этажности здания (в руб.)

Количество секций в доме (x_3)	Этажность (x_2)	Средняя по дому жилая площадь квартир в $м^2$ (x_1)				
		30	32	34	36	38
3	3	107	103	99	95	91
3	4	106	101,5	97,0	92,5	88,0
3	5	105	100	95	90	85

Из этой таблицы видно, что связь между факторами по строкам и столбцам принята линейной. Проследим, как изменяется себестоимость 1 *кв. м* площади $y(x_1, x_2, x_3)$ в зависимости от изменения средней площади квартиры в 3, 4 и 5-этажных трехсекционных домах:

$$y(x_1, 3, 3) = 167-2{,}00\,x_1, \tag{8}$$

$$y(x_1, 4, 3) = 173{,}5-2{,}25\,x_1, \tag{9}$$

$$y(x_1, 5, 3) = 180{,}0-2{,}50\,x_1. \tag{10}$$

Очевидно, что изменение свободных членов и коэффициентов при неизвестных в этих уравнениях обусловлено только изменениями величин $\hat{x}_2$, т. е. этажности зданий.

Аналогичное положение имеем при рассмотрении влияния этажности на себестоимость 1 *кв. м* площади дома при неизменных секционности и площади квартир в доме:

$$y(30, x_2, 3) = 110 - x_2,$$

$$y(32, x_2, 3) = 107{,}5-1{,}5\,x_2,$$

$$y(34, x_2, 3) = 105-2x_2,$$

$$y(36, x_2, 3) = 102{,}5-2{,}5x_2,$$

$$y(38, x_2, 3) = 100-3x_2.$$

В общем виде влияние величины средней по дому площади квартиры на себестоимость 1 *кв. м* площади при закрепленных x_2 и x_3 (при $x_2 = \hat{x}_2$; $x_3 = x_3$) описывается уравнением:

$$y(x_1\check{x}_2, \check{x}_3) = a_0 + a_1x_1. \tag{11}$$

Выразим a_0 и a_1 этого уравнения в виде функций от этажности, исходя из их значений в уравнениях (7), (8) и (9):

$$a_0 = 147{,}5 + 6{,}5x_2; \tag{12}$$

$$a_1 = -1{,}25 - 0{,}25x_2. \tag{13}$$

Подставив в уравнение (11) уравнения (12)—(13), получим выражение, показывающее влияние на функцию y аргументов x_1 и x_2:

$$y(x_1, x_2, 3) = 147{,}5 + 6{,}5x_2 + (-1{,}25 - 0{,}25x_2)x_1 = \\ = 147{,}5 - 1{,}25x_1 + 6{,}5x_2 - 0{,}25x_1x_2. \quad (14)$$

Чтобы выявить зависимость себестоимости 1 *кв. м* жилой площади дома не только от этажности и средней площади квартир, но и числа секций в доме, необходимо полностью повторить расчеты, выполненные для трехсекционных зданий, также и для четырех- и пятисекционных.

Допустим, что в результате расчетов получены следующие уравнения: для трехсекционных домов:

$$y(x_1, x_2, 3) = 147{,}5 - 1{,}25x_1 + 6{,}5x_2 - 0{,}25x_1x_2; \quad (15)$$

для четырехсекционных домов:

$$y(x_1, x_2, 4) = 146{,}5 - 1{,}50x_1 + 6{,}0x_2 - 0{,}30x_1x_2; \quad (16)$$

для пятисекционных домов:

$$y(x_1, x_2, 5) = 145{,}5 - 1{,}75x_1 + 5{,}5x_2 - 0{,}35x_1, x_2. \quad (17)$$

В общем виде зависимость функции y от аргументов x_1 и x_2 при закрепленном x_3 описывается уравнением:

$$y(x_1x_2\check{x}_3) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2. \quad (18)$$

Свободные члены и коэффициенты при неизвестных уравнений (15)—(17) выразим в виде функций от числа секций в здании x_3:

$$a_0 = 150{,}5 - x_3, \quad (19)$$

$$a_1 = 1{,}00 - 0{,}25x_3, \quad (20)$$

$$a_2 = 8{,}0 - 0{,}5x_3, \quad (21)$$

$$a_3 = -0{,}10 - 0{,}05x_3. \quad (22)$$

Подставяя в уравнение (18) уравнения (19)—(22), выявим влияние трех аргументов x_1, x_2 и x_3 на функцию y:

$$y(x_1, x_2, x_3) = 150{,}5 - x_3 + (-1{,}0 - 0{,}25x_3)x_1 + (8{,}0 - 0{,}5x_3)x_2 + \\ + (-0{,}1 - 0{,}05x_1x_2) = 150{,}5 - x_1 + 8x_2 - x_3 - 0{,}1x_1x_2 - \\ - 0{,}25x_1x_3 - 0{,}5x_2x_3 - 0{,}05x_1x_2x_3. \quad (23)$$

Аналогичным образом можно найти зависимость функции от любого числа аргументов при любой (линейной и нелинейной) форме их связи по строкам и столбцам. Для этого удобно пользоваться формами таблиц, аналогичными табл. 2 и 3. Табл. 2 систематизирует порядок группировки, табл. 3 — порядок обобщения экспериментальных и статистических исходных данных в аналитические функции. Рассмотренный метод позволяет на основе экспериментальных или статистических данных одновременно установить:

1) форму связи между функцией и аргументами в виде уравнения, записанного в общем виде;

2) значения коэффициентов при неизвестных и свободного члена уравнения функции многих переменных.

Для установления влияния многих факторов на результативный признак при условии, что зависимость функциональна, необходимо следующее количество данных по изучаемым объектам:

$$Q = ck^p, \quad (24)$$

Таблица 2

x_4	x_3	x_2	x_1: a_1	x_1: a_2	x_1: a_3
d_1	c_1	b_1			
		b_2			
		b_3			
	c_2	b_1		×	
		b_2			
		b_3			
	c_3	b_1			
		b_2			
		b_3			
d_2	c_1	b_1			
		b_2		×	
		b_3			
	c_2	b_1		×	
		b_2	×	×	×
		b_3		×	
	c_3	b_1			
		b_2		×	
		b_3			
d_3	c_1	b_1			
		b_2			
		b_3			
	c_2	b_1			
		b_2		×	
		b_3			
	c_3	b_1			
		b_2			
		b_3			

где p — количество учтенных аргументов, k — количество точек, необходимое для установления формы связи между функцией и одним из ее аргументов (для прямой — 2 точки, для кривой второго порядка — 6 точек, третьего порядка — 9 точек и т. д.); c — количество однотипных испытаний, обеспечивающее в необходимых случаях повышение достоверности результатов. Например, при $c = 3$, $k = 3$, $p = 4$ $Q = 3 \cdot 3^4 = 243$.

В ряде случаев для получения необходимой информации целесообразно вместо обычной разработки экспериментальных проектов использовать методы математического моделирования производственных процессов на ЭВМ, в частности, вместо проектирования жилых домов, технологических производственных поточных линий, предприятий и их комплексов.

Нередко практически невозможно или экономически нецелесообразно обеспечивать наибольшую точность решения задач рассмотренного типа.

В связи с этим возникает необходимость в применении менее точных методов, позволяющих решить задачу при меньшем количестве исходных данных.

Метод двухэтапной многофакторной группировки предусматривает определение формы связи и последующий расчет параметров уравнения.

Предварительное определение формы связи существенно сокращает необходимое количество точек и, следовательно, объем эксперимента. На этом первом этапе устанавливается форма связи функции и каждого аргумента в отдельности, а затем на основе установленных зависимостей функции от каждого из аргументов устанавливается форма связи функции и

Таблица 3

Первый шаг	Второй шаг	Третий шаг	Четвертый шаг
$f_1(x_1) = y(x_1, b_1, c_1, d_1)$ $f_2(x_1) = y(x_1, b_2, c_1, d_1)$ $f_3(x_1) = y(x_1, b_3, c_1, d_1)$	$f_{28}(x_1, x_2) = y(x_1, x_2, c_1, d_1)$	$f_{37} = (x_1, x_2, x_3) = y(x_1, x_2, x_3, d_1)$	$f_{40} = y(x_1, x_2, x_3, x_4)$
$f_4(x_1) = y(x_1, b_1, c_2, d_1)$ $f_5(x_1) = y(x_1, b_2, c_2, d_1)$ $f_6(x_1) = y(x_1, b_3, c_2, d_1)$	$f_{29} = y(x_1, x_2, c_2, d_1)$		
$f_7(x_1) = y(x_1, b_1, c_3, d_1)$ $f_8(x_1) = y(x_1, b_2, c_3, d_1)$ $f_9(x_1) = y(x_1, b_3, c_3, d_1)$	$f_{30} = y(x_1, x_2, c_3, d_1)$		
$f_{10}(x_1) = y(x_1, b_1, c_1, d_2)$ $f_{11}(x_1) = y(x_1, b_2, c_1, d_2)$ $f_{12}(x_1) = y(x_1, b_3, c_1, d_2)$	$f_{31} = y(x_1, x_2, c_1, d_2)$	$f_{38} = y(x_1, x_2, x_3, d_2)$	
$f_{13}(x_1) = y(x_1, b_1, c_2, d_2)$ $f_{14}(x_1) = y(x_1, b_2, c_2, d_2)$ $f_{15}(x_1) = y(x_1, b_3, c_2, d_2)$	$f_{32} = y(x_1, x_2, c_2, d_2)$		
$f_{16}(x_1) = y(x_1, b_1, c_3, d_2)$ $f_{17}(x_1) = y(x_1, b_2, c_3, d_2)$ $f_{18}(x_1) = y(x_1, b_3, c_3, d_2)$	$f_{33} = y(x_1, x_2, c_3, d_2)$		
$f_{19}(x_1) = y(x_1, b_1, c_1, d_3)$ $f_{20}(x_1) = y(x_1, b_2, c_1, d_3)$ $f_{21}(x_1) = y(x_1, b_3, c_1, d_3)$	$f_{34} = y(x_1, x_2, c_1, d_3)$	$f_{39} = y(x_1, x_2, x_3, d_3)$	
$f_{22}(x_1) = y(x_1, b_1, c_2, d_3)$ $f_{23}(x_1) = y(x_1, b_2, c_2, d_3)$ $f_{24}(x_1) = y(x_1, b_3, c_2, d_3)$	$f_{35} = y(x_1, x_2, c_2, d_3)$		
$f_{25}(x_1) = y(x_1, b_1, c_3, d_3)$ $f_{26}(x_1) = y(x_1, b_2, c_2, d_3)$ $f_{27}(x_1) = y(x_1, b_3, c_3, d_3)$	$f_{36} = y(x_1, x_2, c_3, d_3)$		

всех аргументов. Вначале на основе экспериментальных данных составляются уравнения, выражающие зависимость функции от каждого ее аргумента в отдельности при остальных аргументах, закрепленных на среднем уровне, т. е. при $y = F(x_1, x_2, \ldots, x_i, \ldots, x_p)$ определяются уравнения

$$y_i = F(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{(i-1)}, x_i, \bar{x}_{(i+1)}, \ldots, \bar{x}_p) = S_i(x_i). \tag{25}$$

Выбор объектов, по которым должны быть определены все исходные данные, должен быть подчинен схеме расположения крестиков в табл. 2. В этом случае для определения всех парных зависимостей уравнения (25) потребуется минимальное количество экспериментальных объектов, которое равно:

$$Q_{\min} = [p(k-1) + 1]c. \tag{26}$$

Оно значительно меньше, чем при одновременном установлении формы связи и ее параметров.

При необходимости дополнительным закреплением аргументов на каком-либо другом уровне проверяется постоянство общего вида формы связи функции с каждым ее аргументом в отдельности в любом «сечении поверхности». Для этого потребуется почти такое же количество дополнительных точек. Если форма связи контрольных и основных уравнений различна, то надо найти такое общее их выражение, которое описывало бы любое из них. Полученные конкретные значения свободных членов и коэффициентов при неизвестных для решения поставленной задачи не представляют интереса и могут быть заменены буквенными обозначениями или числами, удобными в счете, например, единицами.

Если парные зависимости функции с каждым ее аргументом могут быть определены теоретически или на основе ранее проведенных исследований, то необходимость в перечисленных выше операциях отпадает.

Для того, чтобы найти уравнение зависимости функции от всех ее аргументов, необходимо произвести ряд последовательных подстановок (как это было показано в уравнениях (11) — (14); (18) — (23)): в уравнение, выражающее в общем виде зависимость функции от первого аргумента, подставить вместо свободного члена и коэффициентов при неизвестном уравнения, характеризующие зависимость этих параметров от следующего второго аргумента; в полученное уравнение функции двух переменных вместо свободного члена и коэффициентов при неизвестных подставить уравнения, выражающие зависимость этих параметров от третьего аргумента, и т. д.

При небольшом количестве изучаемых объектов (меньшем, чем $Q = k^p$) нельзя определить функциональную численную зависимость параметров a_i уравнения, содержащего l учтенных аргументов из p, от $(l+1)$-го аргумента при $l \leqslant p-1$.

Однако для многих функций нескольких аргументов x_i $(i = 1, 2, \ldots \ldots, l, \ldots, p)$ допустимо принять общий вид формы связи параметров a_i уравнения $y = f(x_1, x_2, \ldots, x_l)$ при l аргументах таким же, как самой функции y с $(l+1)$-м аргументом.

Исходя из этой предпосылки, можно на основе частных парных зависимостей функции с каждым из ее аргументов установить общий вид формы связи функции со всеми аргументами. Для этого достаточно произвести последовательную подстановку вместо параметров уравнения, выражающего в общем виде частную зависимость функции от первого аргумента, уравнение частной зависимости функции от второго аргумента, а в полученное уравнение функции двух аргументов вместо его параметров — уравнение частной зависимости функции от третьего аргумента и т. д.

Приведенные положения поясним следующим примером: пусть требуется определить общий вид уравнения зависимости функции y от двух переменных x_1 и x_2.

Предположим, что частная парная зависимость между функцией и первым аргументом при закрепленном действии второго аргумента описывается уравнением:

$$y(x_1, \check{x}_2) = a_0 + a_1 x_1, \tag{27}$$

а частная парная зависимость между функцией и вторым аргументом при закрепленном действии первого аргумента описывается уравнением:

$$y(\check{x}_1, y_2) = a_2 + a_3 / x_2. \tag{28}$$

Принимаем форму связи зависимости как a_0, так и a_1 с x_2 такой же, как между y и x_2.

Тогда

$$a_0 = a_2 + a_3 / x_2; \tag{29}$$

$$a_1 = a_2 + a_3 / x_2. \tag{30}$$

Подставляем уравнения (29) и (30) в уравнение (27):

$$y(x_1x_2) = (a_2 + a_3 / x_3) + (a_2 + a_3 / x_3)x_1 = a_2 + a_2x_1 + a_3 / x_2 + a_3x_1 / x_2. \quad (31)$$

После замены индексов получим:

$$y = b_0 + b_1x_1 + b_2 / x_2 + b_3x_1 / x_2. \quad (32)$$

Это и есть искомый результат.

Аналогичным образом можно установить связь функции с третьим аргументом. Для этого следует вместо параметров b_0, b_1, b_2, b_3 уравнения (32) подставить уравнения, выражающие зависимость результативного признака со следующим — третьим аргументом.

Определить форму связи часто можно, составляя произведения из уравнений, каждое из которых выражает в общем виде частную зависимость функции только от одного аргумента. В результате умножения устанавливается приближенная или точная форма связи между функцией и всеми учтенными аргументами. Воспользуемся для иллюстрации этого положения предыдущим примером.

Перемножим уравнения (27) и (28):

$$y(x_1x_2) = (a_0 + a_1x_1)(a_2 + a_3 / x_3) = a_0a_2 + a_1a_2x_1 + a_0a_3 / a_2 + a_1a_3x_1 / x_2.$$

Обозначив $a_0a_2 = b_0$, $a_1a_2 = b_1$, $a_0a_3 = b_2$, $a_1a_3 = b_3$, получим следующее уравнение:

$$y(x_1x_2) = a_0 + a_1(x_1) + b_2 / x_2 + b_3x_1 / x_2.$$

Оно было получено ранее методом подстановки (см. уравнение (32)). Аналогично по частным парным зависимостям можно установить форму связи между функцией и любым количеством аргументов.

Рассчитанные по рекомендуемой методике уравнения совпадут или дадут хорошее приближение к уравнениям, характеризующим реальное взаимодействие факторов, если последние описывают кривые на плоскости или поверхности (гиперповерхности) достаточно простой формы. Это было проверено на многих примерах, часть из которых приведена в настоящей статье. В тех же случаях, когда при изменениях уровня закрепляемых аргументов меняется общий вид формы связи функции с изменяемым фактором, не исключена возможность расхождений между уравнением поверхности, которое получено по ограниченному количеству «сечений», и уравнением поверхности, выявленным по существенно большему количеству «сечений».

Второй этап решения задачи заключается в определении величины параметров уравнения, записанного в общем виде.

При известной форме связи и функциональной зависимости количество точек (объектов) должно быть равным количеству параметров в уравнении, выражающем зависимость функции от ее аргументов. Так, положение прямой на плоскости определяется всего двумя точками, плоскости в трехмерном пространстве — тремя точками и т. д. При нелинейных зависимостях минимальное количество точек может быть бо́льшим, но, конечно, гораздо меньшим, чем это необходимо при неизвестной форме связи.

Вернемся к рассмотренному выше примеру (см. табл. 1), но предположим упрощенную форму связи y и x_1 и x_2 в виде уравнения плоскости:

$$y = a_0 + a_1x_1 + a_2x_2. \quad (33)$$

Для определения параметров этого уравнения возьмем в табл. 1 три значения y из расположенных в средней строке и среднем столбце, например: $y_1 = 106$ (при $x_1 = 30$; $x_2 = 4$); $y_2 = 97$ (при $x_1 = 34$; $x_2 = 4$); $y_3 = 95$ (при $x_1 = 34$, $x_2 = 5$) и составим систему уравнений:

$$\begin{aligned} 106 &= a_0 + a_1 30 + a_2 4, \\ 97 &= a_0 + a_1 34 + a_2 4, \\ 95 &= a_0 + a_1 34 + a_3 5. \end{aligned} \tag{34}$$

Решив эту систему, получим следующее уравнение плоскости:

$$y = 181{,}5 - 2{,}25x_1 - 2x_1, \tag{35}$$

на основе которого можно заново рассчитать 7 значений y из 15 (табл. 1), а именно те из них, которые расположены в средней строке и в среднем столбце этой таблицы. Для того чтобы определить значения всех 15 точек, форма связи должна быть принята в виде уравнения поверхности:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2. \tag{36}$$

Для определения параметров a_i этого уравнения потребуется решить систему из четырех уравнений на основе всего четырех экспериментальных объектов.

Рассмотрим возможные *упрощения многофакторных моделей,* полученных методами группировки, и влияние количества учтенных аргументов на количество членов (слагаемых правой части) многочлена, образованного перемножением множителей, каждый из которых выражает частную парную зависимость функции с одним из аргументов для следующих двух простейших случаев.

1) В первом случае

$$y = (a_1 + b_1x_1^{d_1})(a_2 + b_2x_2^{d_2})\ldots(a_m + b_m + x_m^{d_m}). \tag{37}$$

Количество слагаемых правой части уравнения $y = f(x_1, x_2, \ldots, x_m)$ составит:

$$q = 2^m. \tag{38}$$

Например, для функции пяти аргументов $q = 2^5 = 32$, а само уравнение при всех $d_i = 1$ будет иметь следующий вид:

$$\begin{aligned} y = {} & a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_1x_2 + \\ & + a_7x_1x_3 + a_8x_1x_4 + a_9x_1x_5 + a_{10}x_2x_3 + a_{11}x_2x_4 + \\ & + a_{12}x_2x_5 + a_{13}x_3x_4 + a_{14}x_3x_5 + a_{15}x_4x_5 + \\ & + a_{16}x_1x_2x_3 + a_{17}x_1x_2x_4 + a_{18}x_1x_2x_5 + \\ & + a_{19}x_1x_3x_4 + a_{20}x_1x_3x_5 + a_{21}x_1x_4x_5 + \\ & + a_{22}x_2x_3x_4 + a_{23}x_2x_3x_5 + a_{24}x_2x_4x_5 + \\ & + a_{25}x_3x_4x_5 + a_{26}x_1x_2x_3x_4 + a_{27}x_1x_2x_3x_5 + \\ & + a_{28}x_1x_2x_4x_5 + a_{29}x_1x_3x_4x_5 + a_{30}x_2x_3x_4x_5 + \\ & + a_{31}x_1x_2x_3x_4x_5. \end{aligned} \tag{39}$$

2) Во втором случае

$$\begin{aligned} y = {} & (a_1 + b_1x_1^{d_1})(a_2 + b_2x_2^{d_2}), \ldots, (a_m + b_mx_m^{d_m}) \times \\ & \times (a_{m+1} + b_{m+1}x_{m+1} + c_{m+1}x_{m+1}^{d_{m+1}})(a_{m+1} + b_{m+2}x_{m+2} + \\ & + c_{m+2}x_{m+2}^{d_{m+2}}), \ldots, (a_{m+n} + b_{m+n}x_{m+n} + c_{m+n}x_{m+n}^{d_{m+n}}). \end{aligned} \tag{40}$$

В этом более общем случае

$$q = 2^m \cdot 3^n, \tag{41}$$

где $(m + n) = p$ — число учтенных аргументов.

Таким образом, с увеличением количества факторов-аргументов количество слагаемых правой части многочлена, описывающего функцию многих переменных, растет в геометрической прогрессии и тем интенсивнее, чем сложнее форма связи функции с каждым из аргументов в отдельности. Например, если зависимость функции только одного из пяти аргументов имеет вид $y = a_0 + a_1x_1 + a_2x_1^{d_1}$, а от каждого из остальных аргументов $y = b_i + b_ix_i^{d_i}$ $(i = 1, 2, 3, 4)$, то $q = 2^4 \cdot 3^1 = 48$, что существенно больше количества слагаемых в уравнении (39).

Возможности учета значительного количества факторов при рассмотренных формах связи обычно ограничиваются недостатком исходных данных, которых должно быть не меньше, чем слагаемых правой части уравнения многочлена, или лимитируются объемом памяти электронных вычислительных машин. Кроме того, результат в виде чрезмерно громоздкого уравнения, как правило, неприемлем для практики. Поэтому необходимо дальнейшее упрощение задачи.

Ограничение количества членов уравнения функции многих переменных может быть достигнуто способом ограничения целей исследования и, следовательно, исключения произведений аргументов; полного исключения из уравнения несущественных факторов, а также комбинацией этих двух способов. Рассмотрим первый из них, так как второй не нуждается в пояснениях.

Уравнения (14), (23) и (39) и аналогичные им позволяют определить достоверные значения функции y при любых допустимых значениях аргументов x_i от $\min x_i$ до $\max x_i$.

Уравнение (14) позволяет воспроизвести все значения табл. 1.

В то же время уравнение (35), в котором на одно слагаемое меньше, чем в уравнении (14), позволяет воспроизвести значения функции, записанные только в одной (средней) строке и в одном (среднем) столбце этой таблицы.

Это означает, что уравнение (35) решает более скромную задачу: определяет значения функции y при изменении x_1 от $\min x_1$ до $\max x_2$, когда x_2 закреплен на среднем уровне (в результате получаем значения функции по средней строке табл. 1), а затем определяем значения функции y при изменениях x_2 от $\min x_2$ до $\max x_2$, но при x_1, закрепленном на среднем уровне (в результате получаем величины функции по среднему столбцу значений y табл. 1). Расчет остальных величин функции по формуле (35) может рассматриваться лишь в качестве приближенного.

Можно утверждать, что, исключая из многочлена слагаемые с произведениями аргументов и тем самым сводя их количество до величины лишь на единицу большей количества аргументов, мы ограничиваем круг исследования определением значения функции при последовательном изменении действительных значений каждого из факторов-аргументов x_i, при закрепленных на данном, обычно, среднем уровне остальных факторах — $x_1, x_2, \ldots, x_{(i-1)}, \ldots, x_{i+1}, \ldots, x_p$.

Когда учитываются немногие факторы (до 3—5), результаты значительно обедняются при небольшой экономии вычислительных операций. С увеличением количества факторов указанное ограничение целей исследования существенно упрощает вычисления, но в такой же мере обедняет результаты. Поэтому целесообразно принимать наиболее простые формы связи, прежде всего, для ориентировочного выявления значимости каждого из факторов. Это позволит отразить в модели существенные факторы в упрощенной форме, наиболее существенные — детально, а от учета несущественных — отказаться вовсе.

Поясним это положение следующим примером.

Пусть изучаемое явление описывается уравнением (39), которое является функцией пяти аргументов и представляет собой многочлен с 32 слагаемыми, т. е. с 32 искомыми параметрами.

Предположим, что имеющаяся в нашем распоряжении вычислительная техника не позволяет справиться с таким уравнением, так как допускает решение системы только с семью неизвестными. Требуется найти такое уравнение с семью неизвестными параметрами, которое минимально обедняет результаты вследствие упрощения формы связи.

Примем в первом приближении форму связи в виде многочлена, в котором $q = p + 1$, т. е. $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5$. Допустим, что в результате решения уравнения и его анализа установлено, что факторы x_1 и x_2 влияют на y существеннее, чем x_3, x_4 и x_5. При этом условии требованию задачи наилучшим образом удовлетворяет следующее уравнение:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_3 + a_5x_4 + a_6x_5. \qquad (42)$$

По уравнению (42) можно точно определить значения y при изменении x_1 от $\min x_1$ до $\max x_1$, когда x_2 закреплен на любом уровне от $\min x_2$ до $\max x_2$ и x_3, x_4 и x_5 — только на одном, например среднем уровне при изменении x_2 от $\min x_2$ до $\max x_2$, когда x_1 закреплен на любом и x_3, x_4 и x_5 — только на среднем уровне, а также при изменении x_3, или x_4, или x_5, когда остальные аргументы закреплены только на среднем уровне. Закрепление несущественных факторов именно на среднем уровне уменьшает отклонения результативного уравнения от более точного по сравнению с теми отклонениями, которые получились бы при расчетах, основанных на закреплении несущественных факторов на любом другом уровне.

Рассмотрим отдельные *особенности статистических корреляционных зависимостей.* До сих пор рассматривались функциональные зависимости. Во многих случаях при обработке статистических данных способом многофакторной группировки определенным конкретным значениям аргументов соответствует множество случайных значений результативного признака (зависимости этого типа называют статистическими). В этом случае отдельные экспериментальные и даже групповые средние значения y при данных x_i, как правило, не будут соответствовать значениям $\tilde{y}_i$, вычисленным из уравнения, предполагающего функциональную зависимость между учтенными факторами. В подобной ситуации необходимо из множества возможных уравнений выбрать такое, которое давало бы наиболее вероятные расчетные значения результативного признака $\tilde{y}$ при данных значениях факторов переменных: x_1, x_2, $\ldots$, x_p. При этом статистическая закономерность выражает вероятную зависимость результативного признака об учтенных и сопутствующих им, неучтенных уравнением регрессии факторов. При большом статистическом материале выявление зависимости результативного признака от влияющих на него факторов осуществимо на основе многофакторной группировки исходных данных, их осреднения в каждой однородной группе и выравнивания этих осредненных значений. Полученные в результате этого систематизированные по факторам-аргументам и выравненные значения результативного признака представляют собой статистическую зависимость, обобщенную наиболее вероятными ее представителями и тем приведенную к «условно-функциональной», регрессивной зависимости, но заданной пока еще в табличной форме. Переход от табличной формы выражения такой зависимости к алгебраическому ее выражению может быть произведен так же, как и при обычной функциональной зависимости.

При известной форме связи результативного признака от учтенных факторов и количестве изучаемых объектов, существенно превышающем количество независимых переменных, включая и их произведения, но $Q < ck^p$, можно определить параметры уравнения регрессии аппроксимированием функции. Для этого нужно отыскать такие величины параметров a_i теоретического уравнения регрессии, при которых вычисленные значения функции минимально отличались бы от исходных значений результативного признака. К этому результату приводит, например, минимизация суммы квадратов отклонений фактических значений результативного признака по каждому из объектов y_i от их значений, вычисленных по уравнению регрессии $\tilde{y}_i$, т. е.:

$$z = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \min, \tag{43}$$

где $\tilde{y}_i = f(x_1, x_2, \ldots, x_p)$.

Этот метод, известный под названием метода наименьших квадратов, чаще всего применяют в корреляционном анализе.

Необходимо подчеркнуть, что способом наименьших квадратов могут быть определены параметры не только линейных, но и нелинейных уравнений множественной регрессии, представленных в виде многочлена нескольких аргументов x_i, каждый из которых может иметь произвольную, но конкретную степень d_i, если параметры этих независимых переменных a_i имеют первую степень. Большинство экономических явлений и процессов характеризуется нелинейной зависимостью результативного признака от влияющих на него факторов. Обоснованный выбор формы связи является непременным условием успешного применения множественной корреляции.

Рассматриваемые *методы приближенного определения формы связи множественной статистической зависимости* основаны на предположении, что зависимость функции многих аргументов может быть установлена на основе частных парных зависимостей. При этом формы связи частных парных зависимостей могут быть установлены на основе сокращенной многофакторной группировки в соответствии со схемой, показанной в табл. 2 (клетки, отмеченные крестиками).

Во многих случаях можно найти приближенное выражение формы связи частных парных зависимостей функции от каждого из аргументов корректировкой эмпирических линий парных регрессий.

Парные эмпирические линии регрессии более или менее правильно выражают тип зависимости функции от данного фактора и сопутствующих ему неучтенных факторов, если при изменениях его групповых средних средние значения других факторов-аргументов, рассчитанные по этим же группам объектов, остаются неизменными. Необходимость в корректировке возникает, если это условие не соблюдается. При этом форма эмпирической линии парной регрессии может существенно отличаться от линии парной частной регрессии из-за незакрепленного влияния других факторов. Однако для такой корректировки необходимо, прежде всего, «подключить» к эмпирическим данным некоторую экономическую (техническую) гипотезу о взаимосвязи рассматриваемых факторов, т. е. составить и решить уравнение множественной регрессии следующего вида:

$$y = a_0 + a_1 x_1^{d_1} + a_2 x_2^{d_2} + \ldots + a_c x_c^{d_c} + \ldots + a_i x_i^{d_i} + \ldots + a_p x_p^{d_p},$$

где $x_1, x_2, \ldots, x_p$ — учтенные факторы; $d_1, d_2, \ldots, d_p$ — конкретные степени аргументов, установленные теоретическим путем или по здравому смыслу.

Корректировка каждой из групповых средних результативного признака $\hat{y}_{ij}$ при группировке объектов по величине признака x_i выполняется по формуле:

$$\hat{y}_{ij} = y_{ij} + \sum_{i=1}^{p-1} (\bar{x}_c^{d_c} - x_{cij}^{d_c})\, a_c,$$

где y_{ij} — групповая средняя результативного признака в интервале j $(j = 1, 2, 3, \ldots, h)$ при группировке объектов по величине признака x_i; $\bar{x}_c$ — средняя арифметическая величина фактора x_c по всей совокупности объектов; x_{cij} — групповая средняя аргумента x_c при группировке объектов по величинам признака x_i в интервале j.

Количество корректировок для выравнивания одной парной эмпирической линии регрессии равно количеству групп h, на которые разделена совокупность объектов по признаку x_i. Общее количество выравниваний равно количеству учтенных аргументов. Корректировка приводит к уточнению конфигурации эмпирических линий парных регрессий, что позволяет уточнить форму связи функции с каждым аргументом, а следовательно, форму связи функции от всех учтенных аргументов. После первого цикла корректировки может быть выполнен второй, на основе нового уточненного уравнения множественной регрессии.

Вопрос о форме связи при выборе типа уравнения множественной регрессии может быть снят ограничением колебания изменения значений учтенных аргументов — разбиением совокупности рассматриваемых объектов на однородные группы. Внутри каждой из этих групп допустимо аппроксимировать нелинейные зависимости линейными.

При исследовании статистических зависимостей, помимо вопросов, связанных с установлением формы связи и величин ее параметров, возникает необходимость в определении отклонений действительных значений результативного признака от расчетных, определении корреляционного отношения, коэффициентов частной корреляции, вероятных ошибок коэффициентов регрессии и коэффициентов корреляции и ряда других показателей. Методы их расчета подробно изложены в литературе, посвященной математической статистике и специально корреляционному анализу.

Рассмотренные в настоящей статье методы были использованы автором для анализа экономичности проектных решений жилых зданий, изучения влияния отдельных факторов на отчетную и плановую себестоимость крупнопанельных конструкций и в некоторых других задачах.

Представляется возможным использовать результаты расчетов для построения целевых функций и функций ограничения при формулировке экстремальных задач, для установления норм плановых прибылей и плановых дотаций, расчетов плановой себестоимости продукции отрасли и т. д., а также для планирования и обобщения экспериментальных исследований.

Поступила в редакцию
8 II 1965