9. МЕТОДЫ РЕГРЕССИОННОГО АНАЛИЗА

9.1. Сущность и задачи регрессионного анализа

Регрессионный анализ – совокупность статистических методов обработки экспериментальных данных, позволяющих в условии стохастической зависимости исследуемой величины от неслучайных или случайных переменных определять данную зависимость.

В дальнейшем будем рассматривать две модели регрессионного анализа (РА).

М о д е л ь 1 . В данной модели зависимая переменная $\hat{y}$ – случайная величина, а независимые переменные x_j, $j = \overline{1, k}$ – неслучайные, точно заданные переменные. Таким образом, модель 1 регрессионного анализа имеет вид (8.1.2).

М о д е л ь 2 . В данной модели как зависимая переменная, так и независимые переменные являются случайными величинами. Следовательно, модель 2 регрессионного анализа имеет вид (8.1.4).

В дальнейшем регрессионный анализ на основе модели 1 будем называть РА-1, а на основе модели 2 – РА-2. В некоторых источниках РА-2 принято объединять с корреляционным анализом. В данной брошюре РА-2 рассматривается как самостоятельный вид регрессионного анализа, при выполнении которого привлекаются методы корреляционного анализа. Так как РА-1 и РА-2 имеют много общего, то основное внимание уделяется методам РА-1, а для РА-2 показывается лишь специфика соответствующих методов анализа.

Сущность регрессионного анализа состоит в замене стохастической зависимости между переменными $\hat{y}$ и $\hat{x}_j$, $j = \overline{1, k}$ некоторой детерминированной зависимостью *f*, достаточно хорошо аппроксимирующей основные свойства исходной стохастической зависимости. В дальнейшем переменные $\hat{x}_j$, $j = \overline{1, k}$ будем обозначать также вектором $\hat{X}_{<k>}$. Иначе говоря, в процессе регрессионного анализа устанавливается аналитическая зависимость между некоторой характеристикой случайной величины $\hat{y}$ и независимыми переменными $\hat{X}_{<k>}$. Очевидно, что в данном случае возникает проблема выбора соответствующей характеристики случайной величины $\hat{y}$. В регрессионном анализе в качестве такой характеристики используется условное математическое ожидание

$$M[\hat{y} \mid x_1,\ x_2,..., x_k] = M[\hat{y} \mid \hat{x}_1 = x_1,\ \hat{x}_2 = x_2,..., \hat{x}_k = x_k]$$

случайной величины $\hat{y}$ при условии, что независимые переменные $\hat{X}_{<k>}$ приняли определённые значения $X_{<k>}$. Таким образом, сущность регрессионного анализа состоит в замене зависимостей вида (8.1.2) или (8.1.4) зависимостью вида

$$M[\hat{y} \mid x_1,\ x_2,...,x_k] = f(x_1, x_2,..., x_k). \qquad (9.1.1)$$

Выражение (9.1.1) называется ***регрессией***, именно это название и определило наименование методов, объединённых в регрессионном анализе.

Замена стохастической зависимости регрессионной определяет и ограниченность методов регрессионного анализа. Она состоит в том, что данные методы позволяют провести не всестороннее исследование того, как зависит $\hat{y}$ от $\hat{X}_{<k>}$, а лишь один аспект этой стохастической зависимости. Всесторонний анализ имел место, если бы, например, устанавливалась зависимость между законом распределения случайной величины $\hat{y}$ и переменными $\hat{X}_{<k>}$. Тем не менее, с практической точки зрения этот единственный аспект в большинстве случаев является наиболее существенным.

Можно провести классификацию видов регрессионного анализа.

По виду функции f в выражении (9.1.1) регрессионный анализ принято делить на ***линейный***, в котором указанная функция является линейной относительно оцениваемых параметров, т.е.

$$f(x_1, x_2,..., x_k) = \sum_{j=1}^{p} a_j \varphi_j\,, \qquad (9.1.2)$$

и ***нелинейный***, в котором она нелинейная относительно параметров a_j. В выражении (9.1.2) функции φ_j могут определяться одной, несколькими или всеми независимыми переменными.

По числу независимых переменных регрессионный анализ принято подразделять на ***однофакторный***, если имеет место только одна такая переменная, и ***многофакторный***, если число независимых переменных более одной.

Очевидно, что для установления зависимости (9.1.1) необходимо решить ряд задач, которые и составляют собственно регрессионный анализ. К их числу относятся:

1) выбор класса функций, в рамках которого определяется взаимосвязь между $\hat{y}$ и $\hat{X}_{<k>}$;

2) определение подходящих значений параметров a_j, определяющих конкретный вид функции;

3) оценка точности аппроксимации зависимости (8.1.2) или (8.1.4) функцией (9.1.1).

Необходимо отметить, что первая из перечисленных задач формально не решается методами регрессионного анализа. Иначе говоря, класс функции Ψ определяется на основе соображений, которые находятся вне рамок данных методов. Регрессионный анализ позволяет только оценить, насколько удачен этот выбор. При этом наилучшей оценкой зависимости $\hat{y}$ от $X_{<k>}$ в заданном классе Ψ является функция, реализующая минимум математического ожидания квадрата ошибки, т.е. величины

$$\varepsilon(X_{<k>}) = M\left[(\hat{y} - \tilde{y}(X_{<k>}))^2\right]. \qquad (9.1.3)$$

Оценка $\tilde{y}$ случайной величины $\hat{y}$, принадлежащая определённому классу функций Ψ и минимизирующая ошибку (9.1.3), называется средней квадратической регрессией $\hat{y}$ на $X_{<k>}$ класса Ψ.

Вместе с тем некоторые рекомендации по выбору класса функций Ψ могут быть сделаны на основе анализа совокупности результатов наблюдений, в частности, при построении выборочной кривой регрессии. Это можно сделать, по крайней мере, на качественном уровне.

9.2. Однофакторный регрессионный анализ

9.2.1. Модели однофакторного регрессионного комплекса

В однофакторном регрессионном анализе предполагается, что переменная $\hat{y}$ определяется только одной независимой переменной (одним фактором), следовательно, модели РА-1 и РА-2 имеют вид

$$\hat{y} = \hat{f}(x), \qquad \hat{y} = \hat{f}(\hat{x})$$

соответственно.

Данные модели могут быть представлены в несколько иной форме, а именно:

$$\hat{y} = f(x) + \hat{\varepsilon}, \qquad (9.2.1)$$

$$\hat{y} = f(\hat{x} + \hat{\delta}) + \hat{\varepsilon}, \qquad (9.2.2)$$

где $\hat{\varepsilon}$ – ошибка результата наблюдения; $\hat{\delta}$ – ошибка наблюдаемого значения фактора.

В ряде источников модели однофакторного регрессионного анализа именуются моделями парной регрессии.

Регрессионный комплекс, соответствующий модели (9.2.1), описывается следующим образом. Пусть проводится исследование некоторой системы, при этом выполняется n опытов. В результате фиксируется n значений параметра x, характеризующего воздействие среды на систему. При каждом значении x_i, $i = \overline{1, n}$ данного параметра фиксируется m значений наблюдаемого признака $\hat{y}$, который характеризует воздействие системы на среду (рис.9.1). Результаты наблюдений могут быть представлены в виде табл.9.1.

Рис.9.1. Взаимодействие системы и среды

Таблица 9.1

Представление результатов однофакторного эксперимента (модель РА-1)

Значения фактора	Наблюдаемые значения результата						$\bar{y}^*$
	1	2	...	j	...	m	
x_1	y_{11}	y_{12}	...	y_{1j}	...	y_{1m}	$\bar{y}_1^*$
x_2	y_{21}	y_{22}	...	y_{2j}	...	y_{2m}	$\bar{y}_2^*$
...	...	...	...	...	...	...	...
x_i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{im}	$\bar{y}_i^*$
...	...	...	...	...	...	...	...
x_n	y_{n1}	y_{n2}	...	y_{nj}	...	y_{nm}	$\bar{y}_n^*$

В данной таблице кроме результатов наблюдений приведены оценки $\bar{y}_i^*$ условных математических ожиданий результатов наблюдений для различных значений фактора x.

Графически результаты наблюдений могут быть изображены в виде поля корреляции, показанного на рис.9.2.

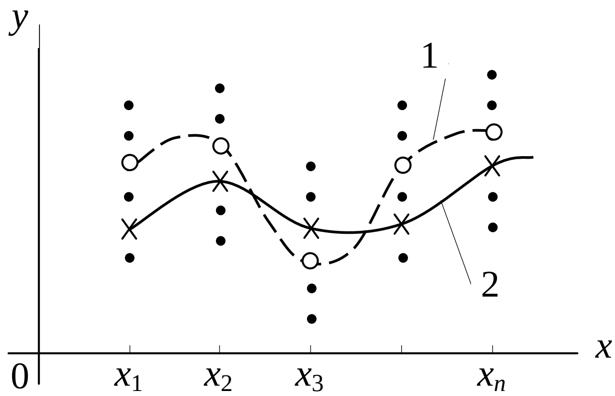

Рис.9.2. Поле корреляции и функции регрессии

На данном поле зачернённые точки соответствуют результатам наблюдений, крестики – значениям оценок математических ожиданий

$$\tilde{M}[\hat{y} \mid \hat{x} = x_i] = \bar{y}_i^*, \quad i = \overline{1, n},$$

а кружки – значениям математических ожиданий

$$M[\hat{y}\,|\,\hat{x}=x_i]=\bar{y}_i\,, \quad i=\overline{1,n}\,.$$

Предположим, что между $\hat{y}$ и x существует зависимость, которая может быть описана в виде

$$M[\hat{y}\,|\,\hat{x}=x_i]=f(x_i)\,, \quad i=\overline{1,n}\,.$$

Тогда кривая, проведённая через значения $M_{\hat{y}}(x_i)$ представляет собой *функцию регрессии генеральной совокупности* (кривая 1, рис.9.2). Так как каждая оценка $\widetilde{M}[\hat{y}\,|\,\hat{x}=x_i]$ условного математического ожидания $M[\hat{y}\,|\,\hat{x}=x_i]$ представляет собой несмещённую оценку, то кривая 2, рис.9.2, проходящая через значения $\widetilde{M}[\hat{y}\,|\,\hat{x}=x_i]$, будет одной из приемлемых оценок функции регрессии генеральной совокупности. Данная кривая называется *выборочной функцией регрессии* или *оценкой функции регрессии*.

Задачей регрессионного анализа является определение выборочной функции регрессии, наилучшим образом (в каком-либо смысле) соответствующей функции регрессии генеральной совокупности.

Для того чтобы данная задача была конструктивной, т.е. допускала решение, вводится ряд предположений, в рамках которых справедливо применение регрессионного анализа. Эти предположения состоят в следующем.

1. Величина x является неслучайной, т.е. задаётся или измеряется без ошибок.

2. Результаты наблюдений получены таким образом, что

$$M[\hat{y}_{ij}]=M_{\hat{y}_i}\,, \quad i=\overline{1,n}\,,\ j=\overline{1,m}\,. \qquad (9.2.3)$$

3. Для каждого x_i распределение величины $\hat{y}$ имеет постоянную дисперсию:

$$D_{\hat{y}_i}=D_{\hat{\varepsilon}_i}=\sigma_i^2=\sigma^2\,, \quad i=\overline{1,n}\,. \qquad (9.2.4)$$

Учитывая (9.2.4), для любого y_i можно записать

$$D[\widetilde{M}_{\hat{y}_i}]=\frac{\sigma^2}{m}\,, \quad i=\overline{1,n}\,. \qquad (9.2.5)$$

Так как $\widetilde{M}_{\hat{y}_i}$ является несмещённой оценкой $M_{\hat{y}_i}$, то ошибка

$$\hat{\varepsilon}_i=\widetilde{M}_{\hat{y}_i}-M_{\hat{y}_i}\,, \quad i=\overline{1,n} \qquad (9.2.6)$$

представляет собой случайную величину с математическим ожиданием

$$M_{\hat{\varepsilon}_i}=M[\widetilde{M}_{\hat{y}_i}]-M_{\hat{y}_i}=M_{\hat{y}_i}-M_{\hat{y}_i}=0$$

и дисперсией

$$D_{\hat{\varepsilon}_i}=D[\widetilde{M}_{\hat{y}_i}]=\frac{\sigma^2}{m}\,.$$

4. Величины $\hat{\varepsilon}_i$ и x_i являются стохастически независимыми, так как x_i являются детерминированными и, следовательно, справедливо равенство

$$K_{\hat{\varepsilon}_i x_i} = 0, \quad i = \overline{1, n},$$

где $K_{\hat{\varepsilon}_i x_i}$ - корреляционный момент $\hat{\varepsilon}_i$ и x_i.

5. Результаты наблюдений являются независимыми:

$$K_{\hat{y}_i \hat{y}_l} = K_{\hat{\varepsilon}_i \hat{\varepsilon}_l} = 0, \quad i \neq l, \quad i = \overline{1, n}, \quad l = \overline{1, n}.$$

6. Величины $\hat{y}_i$ и, следовательно, ошибки $\hat{\varepsilon}_i$ распределены по нормальному закону. Необходимо заметить, что отклонения от нормального закона встречаются часто, однако имеют существенное значение только в том случае, если они велики.

В большинстве практических случаев сбор данных или весьма затруднён, или связан с большими затратами. Поэтому нередко каждому значению фактора x соответствует только одно значение результата и тогда $m = 1$, $y_{ij} = y_i$. В связи с тем, что это значение извлекается случайным образом из генеральной совокупности, величина y_i является несмещённой оценкой величины $M[\hat{y} \mid \hat{x} = x_i]$. Учитывая данное обстоятельство, имеем

$$M_{\hat{\varepsilon}_i} = 0, \quad D_{\hat{y}_i} = D_{\hat{\varepsilon}_i} = \sigma^2, \quad i = \overline{1, n}.$$

Регрессионный комплекс, соответствующий модели (9.2.2), значительно отличается от рассмотренного и имеет следующие особенности.

Пусть проводится исследование некоторой системы, при этом выполняется n опытов. В каждом опыте может быть зарегистрировано m значений фактора $\hat{x}$, характеризующего воздействие среды на систему, и столько же соответствующих значений результата $\hat{y}$, который является характеристикой воздействия системы на среду. Поскольку любая точка $(x_{ij}; y_{ij})$ случайным образом извлекается из генеральной совокупности, то её можно рассматривать как результат i-го опыта:

$$(x_{ij}; y_{ij}) = (x_i; y_i), \quad i = \overline{1, n}, \quad i = \overline{1, m}.$$

В этом случае результаты наблюдений могут быть представлены в виде табл.9.2.

Таблица 9.2

Представление результатов однофакторного эксперимента (модель РА-2)

Значения фактора	x_1	x_1	...	x_1	...	x_1
Значения результата	y_1	y_1	...	y_1	...	y_1

Очевидно, что нижняя строка табл.9.2 представляет собой одновременно и оценки условных математических ожиданий:

$$\tilde{M}[\hat{y} \mid \hat{x} = x_i] = \bar{y}^* = y_i, \quad i = \overline{1, n}.$$

Как и в модели РА-1 графически результаты наблюдений могут быть представлены в виде поля корреляции, рис.9.3.

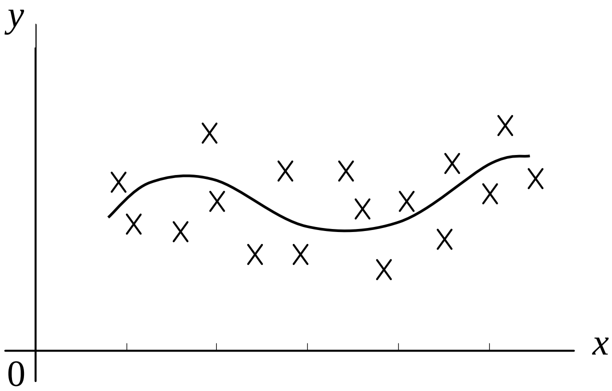

Рис.9.3. Поле корреляции и кривая регрессии

Из сравнения рис.9.2 и 9.3 вида разница между моделями РА-1 и РА-2. Если в первой модели результаты наблюдений рассеивались при определённых значениях фактора x, то во второй результаты располагаются произвольно по всему полю корреляции. Случайный характер значений фактора $\hat{x}$ определяет и особенности РА-2. Эти особенности проявляются как в системе предположений, в рамках которых выполняется РА-2, так и в методах, которые используются при его проведении. Для модели РА-2 справедливы предположения 2, 3, 5, 6.

Так как фактор является случайным, то в общем случае

$$K_{\hat{y}_i \hat{x}_i} = K_{\hat{\varepsilon}_i \hat{x}_i} \neq 0,$$

$$K_{\hat{y}_i \hat{x}_i} = M[(\hat{y}_i - M_{\hat{y}_i})(\hat{x}_i - M_{\hat{x}_i})], \quad i = \overline{1, n}.$$

Особенности методов, используемых в РА-2, рассмотрены ниже.

9.2.2. Построение уравнения регрессии

Пусть из каких-либо соображений выбран класс функций Ψ, которому принадлежит функция регрессии

$$y = f(x). \tag{9.2.6}$$

Эта функция определяется также и вектором числовых параметров

$$(a_0,\ a_1, \ldots, a_k)^{\mathrm{T}} = A_{<k+1>}. \tag{9.2.7}$$

Поэтому выражение (9.2.6) можно представить в виде

$$y = f(x;\ a_0,\ a_1, \ldots, a_k) = f(x; A_{<k+1>}). \tag{9.2.8}$$

Построение уравнения регрессии сводится к решению задачи оценивания параметров (9.2.7), которые называются коэффициентами регрессии. Эта задача может быть решена на основе ряда принципов, являющихся базовыми для статистических методов обработки данных. В практике исследований наиболее широкое применение имеет подход, опирающийся на принцип максимального правдоподобия и, в частности, подход, использующий метод наименьших квадратов.

В соответствии с данным методом задача сводится к получению подходящей оценки $\widetilde{A}_{<k+1>}$ вектора $A_{<k+1>}$, минимизирующей сумму квадратов отклонений (невязок) наблюдаемых значений результата от выборочной функции регрессии. Указанные невязки представляются выражением

$$e_i = y_i - \widetilde{M}_{\hat{y}_i}, \quad i = \overline{1, n}. \tag{9.2.9}$$

Следовательно, необходимо найти минимальное значение величины

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widetilde{M}_{\hat{y}_i})^2 . \tag{9.2.10}$$

Рассмотрим только случай, когда функция регрессии (9.2.8) является линейной относительно оцениваемых параметров:

$$y = f(x;\ a_0,\ a_1, ..., a_k) = a_0 f_0(x) + a_1 f_1(x) + ... + a_k f_k(x) = \sum_{j=0}^{k} a_j f_j(x) . \tag{9.2.11}$$

Тогда оценки математических ожиданий результата определяются из выражения

$$\widetilde{M}_{\hat{y}} = \widetilde{y} = \sum_{j=0}^{k} \widetilde{a}_j f_j(x) . \tag{9.2.12}$$

Принимая во внимание (9.2.12), соотношение (9.2.10) можно записать в виде

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} \widetilde{a}_j f_j(x) \right)^2 .$$

Следовательно, оценки $\widetilde{a}_0, \widetilde{a}_1 ..., \widetilde{a}_k$ должны быть таковы, чтобы выполнялось условие

$$V = \min_{\widetilde{a}_j \in \mathbf{R}} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} \widetilde{a}_j f_j(x) \right)^2 \right\} . \tag{9.2.13}$$

Из раздела 8 следует, что для нахождения минимума суммы квадратов невязок (9.2.13) необходимо составить систему нормальных уравнений вида (8.2.10). При условии, что используется функция регрессии (9.2.11), указанная система записывается следующим образом:

$$\begin{cases} \sum\limits_{i=1}^{n}\left(y_i-(\tilde{a}_0 f_0(x_i)+\tilde{a}_1 f_1(x_i)+...+\tilde{a}_k f_k(x_i))\right)f_0(x_i)=0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum\limits_{i=1}^{n}\left(y_i-(\tilde{a}_0 f_0(x_i)+\tilde{a}_1 f_1(x_i)+...+\tilde{a}_k f_k(x_i))\right)f_j(x_i)=0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum\limits_{i=1}^{n}\left(y_i-(\tilde{a}_0 f_0(x_i)+\tilde{a}_1 f_1(x_i)+...+\tilde{a}_k f_k(x_i))\right)f_k(x_i)=0 \end{cases} \qquad (9.2.14)$$

В уравнениях (9.2.14) учтено, что

$$\frac{\partial\left(\sum\limits_{i=1}^{n}\left(y_i-\sum\limits_{j=0}^{k}\tilde{a}_j f_j(x_i)\right)\right)^2}{\partial a_j}=2\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}\tilde{a}_j f_j(x_i)\right)\frac{\partial\left(\sum\limits_{i=1}^{n}\left(y_i-\sum\limits_{j=0}^{k}\tilde{a}_j f_j(x_i)\right)\right)}{\partial a_j}=$$

$$=2\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}\tilde{a}_j f_j(x_i)\right)(-f_j(x_i)=-2\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}\tilde{a}_j f_j(x_i)\right)f_j(x_i)=$$

$$=-2\sum_{i=1}^{n}\left(y_i-\left(\tilde{a}_0 f_0(x_i)+\tilde{a}_1 f_1(x_i)+...+\tilde{a}_k f_k(x_i)\right)\right)f_j(x_i). \qquad (9.2.15)$$

В выражении (9.2.15) использованы правила дифференцирования сложной функции многих переменных. Поскольку частная производная (9.2.15) приравнивается к нулю, имеем

$$-2\sum_{i=1}^{n}\left(y_i-(\tilde{a}_0 f_0(x_i)+\tilde{a}_1 f_1(x_i)+...+\tilde{a}_k f_k(x_i))\right)f_j(x_i)=0. \qquad (9.2.16)$$

Обе части уравнения (9.2.16) умножаем на –2 и, таким образом, получаем *j*-е уравнение системы (9.2.14):

$$\sum_{i=1}^{n}\left(y_i-\left(\tilde{a}_0 f_0(x_i)+\tilde{a}_1 f_1(x_i)+...+\tilde{a}_k f_k(x_i)\right)\right)f_j(x_i)=0.$$

Выполняем почленное суммирование в уравнениях (9.2.14), слагаемые, содержащие y_i переносим в правую часть, затем умножаем на –1 обе части каждого уравнения.

В результате получаем систему

$$\begin{cases} \tilde{a}_0 \sum_{i=1}^{n} f_0^2(x_i) + \tilde{a}_1 \sum_{i=1}^{n} f_1(x_i) f_0(x_i) + \ldots + \tilde{a}_k \sum_{i=1}^{n} f_k(x_i) f_0(x_i) = \sum_{i=1}^{n} y_i f_0(x_i) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \tilde{a}_0 \sum_{i=1}^{n} f_0(x_i) f_j(x_i) + \tilde{a}_1 \sum_{i=1}^{n} f_1(x_i) f_j(x_i) + \ldots + \tilde{a}_k \sum_{i=1}^{n} f_k(x_i) f_j(x_i) = \sum_{i=1}^{n} y_i f_j(x_i) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \tilde{a}_0 \sum_{i=1}^{n} f_0(x_i) f_k(x_i) + \tilde{a}_1 \sum_{i=1}^{n} f_1(x_i) f_k(x_i) + \ldots + \tilde{a}_k \sum_{i=1}^{n} f_k^2(x_i) = \sum_{i=1}^{n} y_i f_k(x_i) \end{cases} \tag{9.2.17}$$

Уравнения (9.2.17) представляют собой систему линейных уравнений относительно оценок $\tilde{a}_0,\ \tilde{a}_1 ..., \tilde{a}_k$. Следовательно, она решается любым из методов решения систем таких уравнений.

Например, оценки коэффициентов регрессии могут быть найдены по формулам Крамера:

$$\tilde{a}_0 = \frac{|A_0|}{|A|}; ...; \tilde{a}_j = \frac{|A_j|}{|A|}; ...; \tilde{a}_k = \frac{|A_k|}{|A|}, \tag{9.2.18}$$

где $|A|$ – определитель коэффициентов при неизвестных системы (9.2.17); $|A_j|$, $j = \overline{0,k}$ – определители, которые формируются на основе определителя $|A|$ путём замены j-го столбца столбцом свободных членов.

Таким образом, развёрнутый вид данных определителей будет следующим:

$$|A| = \begin{vmatrix} \sum_{i=1}^{n} f_0^2(x_i) & \sum_{i=1}^{n} f_1(x_i) f_0(x_i) & \cdots & \sum_{i=1}^{n} f_k(x_i) f_0(x_i) \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{i=1}^{n} f_0(x_i) f_j(x_i) & \sum_{i=1}^{n} f_1(x_i) f_j(x_i) & \cdots & \sum_{i=1}^{n} f_k(x_i) f_j(x_i) \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{i=1}^{n} f_0(x_i) f_k(x_i) & \sum_{i=1}^{n} f_1(x_i) f_k(x_i) & \cdots & \sum_{i=1}^{n} f_k^2(x_i) \end{vmatrix};$$

$$|A_0| = \begin{vmatrix} \sum_{i=1}^{n} y_i f_0(x_i) & \sum_{i=1}^{n} f_1(x_i) f_0(x_i) & \cdots & \sum_{i=1}^{n} f_k(x_i) f_0(x_i) \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{i=1}^{n} y_i f_j(x_i) & \sum_{i=1}^{n} f_1(x_i) f_j(x_i) & \cdots & \sum_{i=1}^{n} f_k(x_i) f_j(x_i) \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{i=1}^{n} y_i f_k(x_i) & \sum_{i=1}^{n} f_1(x_i) f_k(x_i) & \cdots & \sum_{i=1}^{n} f_k^2(x_i) \end{vmatrix};$$

$$|A_j| = \begin{vmatrix} \sum\limits_{i=1}^{n} f_0^2(x_i) & \cdots & \sum\limits_{i=1}^{n} y_i f_0(x_i) & \cdots & \sum\limits_{i=1}^{n} f_k(x_i) f_0(x_i) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{i=1}^{n} f_0(x_i) f_j(x_i) & \cdots & \sum\limits_{i=1}^{n} y_i f_j(x_i) & \cdots & \sum\limits_{i=1}^{n} f_k(x_i) f_j(x_i) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{i=1}^{n} f_0(x_i) f_k(x_i) & \cdots & \sum\limits_{i=1}^{n} y_i f_k(x_i) & \cdots & \sum\limits_{i=1}^{n} f_k^2(x_i) \end{vmatrix};$$

$$|A_k| = \begin{vmatrix} \sum\limits_{i=1}^{n} f_0^2(x_i) & \sum\limits_{i=1}^{n} f_1(x_i) f_0(x_i) & \cdots & \sum\limits_{i=1}^{n} y_i f_0(x_i) \\ \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{i=1}^{n} f_0(x_i) f_j(x_i) & \sum\limits_{i=1}^{n} f_1(x_i) f_j(x_i) & \cdots & \sum\limits_{i=1}^{n} y_i f_j(x_i) \\ \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{i=1}^{n} f_0(x_i) f_k(x_i) & \sum\limits_{i=1}^{n} f_1(x_i) f_k(x_i) & \cdots & \sum\limits_{i=1}^{n} y_i f_k(x_i) \end{vmatrix}.$$

Запишем систему уравнений (9.2.17) в матричной форме:

$$(F^{\mathsf{T}}_{[k+1;n]} F_{[n;k+1]}) \widetilde{A}_{<k+1>} = F^{\mathsf{T}}_{[k+1;n]} Y_{<n>}, \tag{9.2.19}$$

где

$$F = \begin{pmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ \cdots & \cdots & \cdots & \cdots \\ f_0(x_i) & f_1(x_i) & \cdots & f_k(x_i) \\ \cdots & \cdots & \cdots & \cdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_k(x_n) \end{pmatrix};$$

$$\widetilde{A} = (\widetilde{a}_0, \widetilde{a}_1, ..., \widetilde{a}_{k+1})^{\mathsf{T}}; \quad Y = (y_1, y_2, ..., y_n)^{\mathsf{T}}.$$

Умножим слева обе части матричного уравнения (9.2.19) на квадратную матрицу $(F^{\mathsf{T}} F)^{-1}_{[k+1]}$:

$$(F^{\mathsf{T}} F)^{-1} (F^{\mathsf{T}} F) \widetilde{A} = (F^{\mathsf{T}} F)^{-1} F^{\mathsf{T}} Y.$$

Далее учитываем, что

$$(F^{\mathsf{T}} F)^{-1} (F^{\mathsf{T}} F) = E, \quad E\widetilde{A} = \widetilde{A},$$

где $E = E_{[k+1]}$ – единичная матрица порядка $k+1$.

Окончательно получаем выражение для вычисления оценок искомых параметров:

$$\widetilde{A}_{<k+1>} = (F^{\mathsf{T}}_{[k+1;n]} F_{[n;k+1]})^{-1} F^{\mathsf{T}}_{[k+1;n]} Y_{<n>}. \tag{9.2.20}$$

Рассмотренный выше метод определения оценок коэффициентов регрессии без каких-либо изменений применим и в модели РА-2. Однако при этом необходимо учитывать то обстоятельство, что в модели РА-1 зависимость между $\hat{y}$ и x является односторонней. При рассмотрении же модели РА-2 исследователь сталкивается со взаимностью зависимости между $\hat{y}$ и $\hat{x}$. Поэтому в последнем случае правомерна формулировка задач двух типов.

1. Исследование зависимости $\hat{y}$ от $\hat{x}$ и построение уравнения регрессии $\hat{x}$ на $\hat{y}$:

$$\tilde{y} = \sum_{j=0}^{k} \tilde{a}_j f_j(x).$$

2. Исследование зависимости $\hat{x}$ на $\hat{y}$ и построение уравнения регрессии $\hat{y}$ на $\hat{x}$:

$$\tilde{x} = \sum_{j=0}^{k} \tilde{c}_j \varphi_j(y),$$

где $\tilde{c}_j$ – оценки коэффициентов регрессии.

9.2.3. Проверка адекватности уравнения регрессии

Под ***адекватностью*** уравнения регрессии понимается соответствие данного уравнения экспериментальным данным.

Адекватность уравнения определяется, прежде всего, правильным выбором класса Ψ функций регрессии.

Для проверки соответствия выбранного класса функций регрессии опытным данным рассмотрим очевидное тождество

$$\hat{y}_i - \bar{y}^* = (\hat{y}_i - \tilde{y}_i) - (\bar{y}^* - \tilde{y}_i),\ i = \overline{1, n}. \qquad (9.2.21)$$

Возведём обе части (9.2.21) в квадрат и найдём для них математические ожидания:

$$M[(\hat{y}_i - \bar{y}^*)^2] = M[((\hat{y}_i - \tilde{y}_i) - (\bar{y}^* - \tilde{y}_i))^2] =$$

$$= M[(\hat{y}_i - \tilde{y}_i)^2 - 2(\hat{y}_i - \tilde{y}_i)(\bar{y}^* - \tilde{y}_i) + (\bar{y}^* - \tilde{y}_i)^2] =$$

$$= M[(\hat{y}_i - \tilde{y}_i)^2] - 2M[(\hat{y}_i - \tilde{y}_i)(\bar{y}^* - \tilde{y}_i)] + M[(\bar{y}^* - \tilde{y}_i)^2] =$$

$$= M[(\hat{y}_i - \tilde{y}_i)^2] + M[(\bar{y}^* - \tilde{y}_i)^2],$$

так как

$$M[(\hat{y}_i - \tilde{y}_i)(\bar{y}^* - \tilde{y}_i)] = M[\hat{y}_i - \tilde{y}_i]M[\bar{y}^* - \tilde{y}_i] = 0.$$

В результате получим

$$M[(\hat{y}_i - \bar{y}^*)^2] = M[(\hat{y}_i - \tilde{y}_i)^2] + M[(\tilde{y}_i - \bar{y}^*)^2]$$

или

$$\sigma^2 = \sigma_1^2 + \sigma_2^2, \qquad (9.2.22)$$

где σ^2 – общая дисперсия результатов наблюдений (дисперсия выходной переменной), σ_1^2 – дисперсия, характеризующая рассеивание результатов наблюдений относительно регрессионной зависимости (остаточная дисперсия); σ_2^2 – дисперсия, характеризующая отклонение регрессионной зависимости относительно истинной.

Чем меньше рассеивание результатов относительно регрессионной зависимости, тем лучше последняя аппроксимирует истинную (но неизвестную) зависимость (удовлетворяет экспериментальным данным). Таким образом, чем больше отношение σ^2/σ_1^2, тем более адекватно уравнение регрессии. Поэтому выражение

$$\hat{\mathsf{F}} = \frac{\tilde{\sigma}^2}{\tilde{\sigma}_1^2}, \qquad (9.2.23)$$

принимается в качестве показателя согласованности при проверке нулевой гипотезы H_0 о соответствии выбранного класса функций регрессии экспериментальным данным. В связи с тем, что случайная величина (9.2.23) подчиняется закону распределения Фишера [1], критерием правильности гипотезы H_0 является выполнение неравенства

$$\mathsf{F} = \frac{\tilde{\sigma}^2}{\tilde{\sigma}_1^2} > \mathsf{F}_{(\alpha;\, n\text{-}1;\, n\text{-}k\text{-}1)}, \qquad (9.2.24)$$

где F – наблюдаемое (вычисленное) значение показателя согласованности гипотезы H_0; $\mathsf{F}_{(\alpha;\, n\text{-}1;\, n\text{-}k\text{-}1)}$ – критическое значение данного показателя при уровне значимости α и степенях свободы $f_1 = n - 1$, $f_2 = n - k - 1$. Критическое значение берётся по таблице критических точек распределения Фишера (приложение 5). Входами в таблицу являются величины α, f_1, f_2.

Оценки дисперсий $\tilde{\sigma}^2$ и $\tilde{\sigma}_1^2$ находятся по формулам:

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}; \qquad (9.2.25)$$

$$\tilde{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-k-1}. \qquad (9.2.26)$$

В случае невыполнения условия (9.2.24) принимается конкурирующая гипотеза H_1, т.е. возникает проблема выбора другого класса функций регрессии, более точно соответствующего экспериментальным данным. Проверка неравенства (9.2.24) называется проверкой адекватности уравнения регрессии по критерию Фишера.

9.2.4. Проверка значимости коэффициентов регрессии

Важным аспектом качества регрессионной зависимости является значимость коэффициентов регрессии a_j, $j=\overline{0,k}$. Поскольку оценки этих коэффициентов получены на основе случайной выборки, возникает задача проверки того, что их величина определяется именно видом зависимости между $\hat{y}$ и x, а не случайным характером выборки. Эта задача сводится к проверке гипотез о неравенстве нулю коэффициентов a_j:

$$H_{0j}\colon a_j \neq 0, \quad j=\overline{0,k}$$

при соответствующих конкурирующих гипотезах H_{1j}: $a_j = 0$.

В качестве показателя согласованности при проверке гипотез H_{0j} используется выражение

$$\hat{t}_j = \frac{|\tilde{a}_j|}{\tilde{\sigma}_{\tilde{a}_j}}, \quad j=\overline{0,k}, \tag{9.2.27}$$

где $|\tilde{a}_j|$ – модуль величины $\tilde{a}_j$; $\tilde{\sigma}_{\tilde{a}_j}$ – оценка среднего квадратического отклонения коэффициента $\tilde{a}_j$.

Получение оценок $\tilde{\sigma}_{\tilde{a}_j}$ осуществляется так же, как это было описано в п.п.8.3.2. Здесь учтём специфику рассматриваемого вида функций регрессии (9.2.11) и уточним данную процедуру.

В соответствии с выражением (8.3.14) можно написать:

$$K_{\tilde{A}[k+1]} = \tilde{\sigma}_1^2 (F^{\mathsf{T}}_{[k+1;\,n]} F_{[n;\,k+1]})^{-1}. \tag{9.2.28}$$

На главной диагонали корреляционной матрицы (9.2.28) вектора $\tilde{A}_{<k+1>}$ будут дисперсии оценок коэффициентов регрессии:

$$diag K_{\tilde{A}} = \left(\tilde{\sigma}^2_{\tilde{a}_0},\ \tilde{\sigma}^2_{\tilde{a}_1}, ..., \tilde{\sigma}^2_{\tilde{a}_k}\right). \tag{9.2.29}$$

Случайная величина (9.2.27) подчинена закону распределения Стьюдента с $f = n - k - 1$ степенями свободы [1, 4]. Поэтому критерием правильности гипотез H_{0j} является выполнение неравенств

$$t_j = \frac{|\tilde{a}_j|}{\tilde{\sigma}_{\tilde{a}_j}} > t_{(\alpha;\, n-k-1)},\ j=\overline{0,k}, \tag{9.2.30}$$

где t_j – наблюдаемое значение показателя согласованности гипотезы H_{0j}; $t_{(\alpha;\ n-k-1)}$ – критическое значение данного показателя при уровне значимости α и f степенях свободы.

Критическое значение $t_{(\alpha;\ n-k-1)}$ берётся из таблицы критических точек распределения Стьюдента (приложение 6). Очевидно, что входами в таблицу являются α и f.

Коэффициенты регрессии, для которых не выполняется условие (9.2.30), принимаются равными нулю. Проверка неравенств (9.2.30) называется проверкой значимости коэффициентов регрессии по критерию Стьюдента.

Следует отметить, что при решении практических задач встречаются случаи, когда после исключения незначимых коэффициентов в соответствии с критерием Стьюдента регрессионная зависимость становится неадекватной. Такие случаи вызывают сомнения в универсальности этого широко применяемого в настоящее время критерия. Поэтому после исключения незначимых коэффициентов проверку адекватности уравнения регрессии по критерию Фишера целесообразно повторить.

9.2.5. Примеры однофакторного регрессионного анализа

Из вышеизложенного следует, что однофакторный регрессионный анализ проводится в следующей последовательности.

1. Выбирается вид функции регрессии.

2. Составляется система нормальных уравнений.

3. Находится решение системы нормальных уравнений (определяются оценки коэффициентов регрессии).

4. Проверяется адекватность построенного уравнения регрессии экспериментальным данным.

5. Проверяется значимость коэффициентов регрессии.

6. Повторно проверяется адекватность уравнения регрессии после исключения незначимых коэффициентов.

П р и м е р 9.1. Для исследования зависимости выходного сигнала y системы от входного воздействия x проведены испытания, результаты которых сведены в табл.9.3.

Необходимо построить уравнение регрессии $y = f(x)$ в предположении, что оно является алгебраическим полиномом третьей степени. Расчёты произвести в скалярной форме.

Т а б л и ц а 9.3

Массив экспериментальных данных

x	–2	–1	0	1	3
y	5	4	5	2	–39

▼ 1. Класс функций Ψ задан в условии задачи – это полиномы третьей степени

$$y = a_0x^3 + a_1x^2 + a_2x + a_3. \qquad (9.2.31)$$

Они являются функциями вида (9.2.11). Для данного случая выражение (9.2.11) представляется как

$$y = a_0 f_0(x) + a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x), \qquad (9.2.32)$$

где

$$f_0(x) = x^3, \quad f_1(x) = x^2, \quad f_2(x) = x, \quad f_3(x) = 1. \qquad (9.2.33)$$

2. Для функции (9.2.32) и заданного количества опытов система нормальных уравнений (9.2.17) принимает вид

$$\begin{cases}\tilde{a}_0\sum_{i=1}^{5} f_0^2(x_i)+\tilde{a}_1\sum_{i=1}^{5} f_1(x_i)\,f_0(x_i)+\tilde{a}_2\sum_{i=1}^{5} f_2(x_i)\,f_0(x_i)+\tilde{a}_3\sum_{i=1}^{5} f_3(x_i)\,f_0(x_i)=\sum_{i=1}^{5} y_i\,f_0(x_i)\\ \tilde{a}_0\sum_{i=1}^{5} f_0(x_i)\,f_1(x_i)+\tilde{a}_1\sum_{i=1}^{5} f_1^2(x_i)+\tilde{a}_2\sum_{i=1}^{5} f_2(x_i)\,f_1(x_i)+\tilde{a}_3\sum_{i=1}^{5} f_3(x_i)\,f_1(x_i)=\sum_{i=1}^{5} y_i\,f_1(x_i)\\ \tilde{a}_0\sum_{i=1}^{5} f_0(x_i)\,f_2(x_i)+\tilde{a}_1\sum_{i=1}^{5} f_1(x_i)\,f_2(x_i)+\tilde{a}_2\sum_{i=1}^{5} f_2^2(x_i)+\tilde{a}_3\sum_{i=1}^{n} f_3(x_i)\,f_2(x_i)=\sum_{i=1}^{5} y_i\,f_2(x_i)\\ \tilde{a}_0\sum_{i=1}^{5} f_0(x_i)\,f_3(x_i)+\tilde{a}_1\sum_{i=1}^{5} f_1(x_i)\,f_3(x_i)+\tilde{a}_2\sum_{i=1}^{5} f_2(x_i)\,f_3(x_i)+\tilde{a}_3\sum_{i=1}^{5} f_3^2(x_i)=\sum_{i=1}^{5} y_i\,f_3(x_i)\end{cases}\qquad(9.2.34)$$

С учётом (9.2.33) система уравнений (9.2.34) преобразуется следующим образом:

$$\begin{cases}\tilde{a}_0\sum_{i=1}^{5}x_i^6+\tilde{a}_1\sum_{i=1}^{5}x_i^5+\tilde{a}_2\sum_{i=1}^{5}x_i^4+\tilde{a}_3\sum_{i=1}^{5}x_i^3=\sum_{i=1}^{5}y_ix_i^3\\ \tilde{a}_0\sum_{i=1}^{5}x_i^5+\tilde{a}_1\sum_{i=1}^{5}x_i^4+\tilde{a}_2\sum_{i=1}^{5}x_i^3+\tilde{a}_3\sum_{i=1}^{5}x_i^2=\sum_{i=1}^{5}y_ix_i^2\\ \tilde{a}_0\sum_{i=1}^{5}x_i^4+\tilde{a}_1\sum_{i=1}^{5}x_i^3+\tilde{a}_2\sum_{i=1}^{5}x_i^2+\tilde{a}_3\sum_{i=1}^{5}x_i=\sum_{i=1}^{5}y_ix_i\\ \tilde{a}_0\sum_{i=1}^{5}x_i^3+\tilde{a}_1\sum_{i=1}^{5}x_i^2+\tilde{a}_2\sum_{i=1}^{5}x_i+\tilde{a}_3n=\sum_{i=1}^{5}y_i\end{cases}\qquad(9.2.35)$$

3. Оценки коэффициентов уравнения регрессии (т.е. решение системы линейных уравнений (9.2.35)) находим по формулам (9.2.18):

$$\tilde{a}_0=\frac{|A_0|}{|A|};\quad \tilde{a}_1=\frac{|A_1|}{|A|};\quad \tilde{a}_2=\frac{|A_2|}{|A|};\quad \tilde{a}_3=\frac{|A_3|}{|A|}.\qquad(9.2.36)$$

Развёрнутый вид определителей в соотношениях (9.2.36):

$$|A|=\begin{vmatrix}\sum_{i=1}^{5}x_i^6 & \sum_{i=1}^{5}x_i^5 & \sum_{i=1}^{5}x_i^4 & \sum_{i=1}^{5}x_i^3\\ \sum_{i=1}^{5}x_i^5 & \sum_{i=1}^{5}x_i^4 & \sum_{i=1}^{5}x_i^3 & \sum_{i=1}^{5}x_i^2\\ \sum_{i=1}^{5}x_i^4 & \sum_{i=1}^{5}x_i^3 & \sum_{i=1}^{5}x_i^2 & \sum_{i=1}^{5}x_i\\ \sum_{i=1}^{5}x_i^3 & \sum_{i=1}^{5}x_i^2 & \sum_{i=1}^{5}x_i & n\end{vmatrix};\qquad |A_0|=\begin{vmatrix}\sum_{i=1}^{5}y_ix_i^3 & \sum_{i=1}^{5}x_i^5 & \sum_{i=1}^{5}x_i^4 & \sum_{i=1}^{5}x_i^3\\ \sum_{i=1}^{5}y_ix_i^2 & \sum_{i=1}^{5}x_i^4 & \sum_{i=1}^{5}x_i^3 & \sum_{i=1}^{5}x_i^2\\ \sum_{i=1}^{5}y_ix_i & \sum_{i=1}^{5}x_i^3 & \sum_{i=1}^{5}x_i^2 & \sum_{i=1}^{5}x_i\\ \sum_{i=1}^{5}y_i & \sum_{i=1}^{5}x_i^2 & \sum_{i=1}^{5}x_i & n\end{vmatrix};$$

$$|A_1|=\begin{vmatrix} \sum\limits_{i=1}^{5} x_i^6 & \sum\limits_{i=1}^{5} y_i x_i^3 & \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 \\ \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} y_i x_i^2 & \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 \\ \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} y_i x_i & \sum\limits_{i=1}^{5} x_i^2 & \sum\limits_{i=1}^{5} x_i \\ \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} y_i & \sum\limits_{i=1}^{5} x_i & n \end{vmatrix}; \quad |A_2|=\begin{vmatrix} \sum\limits_{i=1}^{5} x_i^6 & \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} y_i x_i^3 & \sum\limits_{i=1}^{5} x_i^3 \\ \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} y_i x_i^2 & \sum\limits_{i=1}^{5} x_i^2 \\ \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} y_i x_i & \sum\limits_{i=1}^{5} x_i \\ \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 & \sum\limits_{i=1}^{5} y_i & n \end{vmatrix};$$

$$|A_3|=\begin{vmatrix} \sum\limits_{i=1}^{5} x_i^6 & \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} y_i x_i^3 \\ \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} y_i x_i^2 \\ \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 & \sum\limits_{i=1}^{5} y_i x_i \\ \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 & \sum\limits_{i=1}^{5} x_i & \sum\limits_{i=1}^{5} y_i \end{vmatrix}.$$

Составляем расчётную таблицу 9.4 для вычисления коэффициентов системы линейных уравнений (9.2.35).

Т а б л и ц а 9.4

Расчётная таблица

x_i	y_i	x_i^2	x_i^3	x_i^4	x_i^5	x_i^6	$y_i x_i$	$y_i x_i^2$	$y_i x_i^3$
1	2	3	4	5	6	7	8	9	10
–2	5	4	–8	16	–32	64	–10	20	–40
–1	4	1	–1	1	–1	1	–4	4	–4
0	5	0	0	0	0	0	0	0	0
1	2	1	1	1	1	1	2	2	2
3	–39	9	27	81	243	729	–117	–351	–1053
$\sum\limits_{i=1}^{5} x_i =$ =1	$\sum\limits_{i=1}^{5} y_i =$ = –23	$\sum\limits_{i=1}^{5} x_i^2 =$ =15	$\sum\limits_{i=1}^{5} x_i^3 =$ =19	$\sum\limits_{i=1}^{5} x_i^4 =$ =99	$\sum\limits_{i=1}^{5} x_i^5 =$ =211	$\sum\limits_{i=1}^{5} x_i^6 =$ =795	$\sum\limits_{i=1}^{5} y_i x_i =$ = –129	$\sum\limits_{i=1}^{5} y_i x_i^2 =$ = –325	$\sum\limits_{i=1}^{5} y_i x_i^3 =$ = –1095

Получаем систему уравнений (9.2.35) с числовыми значениями коэффициентов при неизвестных:

$$\begin{cases} 795\tilde{a}_0 + 211\tilde{a}_1 + 99\tilde{a}_2 + 19\tilde{a}_3 = -1095 \\ 211\tilde{a}_0 + 99\tilde{a}_1 + 19\tilde{a}_2 + 15\tilde{a}_3 = -325 \\ 99\tilde{a}_0 + 19\tilde{a}_1 + 15\tilde{a}_2 + \tilde{a}_3 = -19 \\ 19\tilde{a}_0 + 15\tilde{a}_1 + \tilde{a}_2 + 5\tilde{a}_3 = -23. \end{cases}$$

Определитель $|A|$ четвёртого порядка вычисляем разложением по первой строке:

$$|A|=\begin{vmatrix} 795 & 211 & 99 & 19 \\ 211 & 99 & 19 & 15 \\ 99 & 19 & 15 & 1 \\ 19 & 15 & 1 & 5 \end{vmatrix}=795(-1)^2\begin{vmatrix} 99 & 19 & 15 \\ 19 & 15 & 1 \\ 15 & 1 & 5 \end{vmatrix}+$$

$$+211(-1)^3\begin{vmatrix} 211 & 19 & 15 \\ 99 & 15 & 1 \\ 19 & 1 & 5 \end{vmatrix}+99(-1)^4\begin{vmatrix} 211 & 99 & 15 \\ 99 & 19 & 1 \\ 19 & 15 & 5 \end{vmatrix}+19(-1)^5\begin{vmatrix} 211 & 99 & 19 \\ 99 & 19 & 15 \\ 19 & 15 & 1 \end{vmatrix}=$$

$$=795\cdot 2716-211\cdot 3780+99(-13384)-19(-3696)=106848.$$

При разложении определителя $|A|$ получены четыре определителя третьего порядка, которые вычисляем также разложением по первой строке:

$$\begin{vmatrix} 99 & 19 & 15 \\ 19 & 15 & 1 \\ 15 & 1 & 5 \end{vmatrix}=99(-1)^2\begin{vmatrix} 15 & 1 \\ 1 & 5 \end{vmatrix}+19(-1)^3\begin{vmatrix} 19 & 1 \\ 15 & 5 \end{vmatrix}+15(-1)^4\begin{vmatrix} 19 & 15 \\ 15 & 1 \end{vmatrix}=$$

$$=99\cdot 74-19\cdot 80+15(-206)=2716;$$

$$\begin{vmatrix} 211 & 19 & 15 \\ 99 & 15 & 1 \\ 19 & 1 & 5 \end{vmatrix}=211(-1)^2\begin{vmatrix} 15 & 1 \\ 1 & 5 \end{vmatrix}+19(-1)^3\begin{vmatrix} 99 & 1 \\ 19 & 5 \end{vmatrix}+15(-1)^4\begin{vmatrix} 99 & 15 \\ 19 & 1 \end{vmatrix}=$$

$$=211\cdot 74-19\cdot 476+15(-186)=3780;$$

$$\begin{vmatrix} 211 & 99 & 15 \\ 99 & 19 & 1 \\ 19 & 15 & 5 \end{vmatrix}=211(-1)^2\begin{vmatrix} 19 & 1 \\ 15 & 5 \end{vmatrix}+99(-1)^3\begin{vmatrix} 99 & 1 \\ 19 & 5 \end{vmatrix}+15(-1)^4\begin{vmatrix} 99 & 19 \\ 19 & 15 \end{vmatrix}=$$

$$=211\cdot 80-99\cdot 476+15\cdot 1124=-13384;$$

$$\begin{vmatrix} 211 & 99 & 19 \\ 99 & 19 & 15 \\ 19 & 15 & 1 \end{vmatrix}=211(-1)^2\begin{vmatrix} 19 & 15 \\ 15 & 1 \end{vmatrix}-99(-1)^3\begin{vmatrix} 99 & 15 \\ 19 & 1 \end{vmatrix}+19(-1)^4\begin{vmatrix} 99 & 19 \\ 19 & 15 \end{vmatrix}=$$

$$=211(-206)+99(-186)+19\cdot 1124=-3696.$$

Аналогично находим

$$|A_0|=\begin{vmatrix} -1095 & 211 & 99 & 19 \\ -325 & 99 & 19 & 15 \\ -129 & 19 & 15 & 1 \\ -23 & 15 & 1 & 5 \end{vmatrix}=-103992\,;$$

$$|A_1| = \begin{vmatrix} 795 & -1095 & 99 & 19 \\ 211 & -325 & 19 & 15 \\ 99 & -129 & 15 & 1 \\ 19 & -23 & 1 & 5 \end{vmatrix} = -209016;$$

$$|A_2| = \begin{vmatrix} 795 & 211 & -1095 & 19 \\ 211 & 99 & -325 & 15 \\ 99 & 19 & -129 & 1 \\ 19 & 15 & -23 & 5 \end{vmatrix} = -3216;$$

$$|A_3| = \begin{vmatrix} 795 & 211 & 99 & -1095 \\ 211 & 99 & 19 & -325 \\ 99 & 19 & 15 & -129 \\ 19 & 15 & 1 & -23 \end{vmatrix} = 531360.$$

По формулам (9.2.36) вычисляем оценки коэффициентов регрессии:

$$\tilde{a}_0 = \frac{-103992}{106848} = -0{,}97; \quad \tilde{a}_1 = \frac{-209016}{106848} = -1{,}96;$$

$$\tilde{a}_2 = \frac{-3216}{106848} = -0{,}03; \quad \tilde{a}_3 = \frac{531360}{106848} = 4{,}97.$$

Получили уравнение регрессии

$$\tilde{y} = -0{,}97x^3 - 1{,}96x^2 - 0{,}03x + 4{,}97.$$

4. Проверяем адекватность регрессионной зависимости экспериментальным данным. Для этого необходимо вычислить оценки дисперсий (9.2.25) и (9.2.26). Составляем табл.9.5.

Таблица 9.5

Расчётная таблица

x_i	y_i	$y_i - \bar{y}^*$	$(y_i - \bar{y}^*)^2$	$\tilde{y}_i$	$y_i - \tilde{y}_i$	$(y_i - \tilde{y}_i)^2$
–2	5	9,6	92,16	4,95	0,05	0,0025
–1	4	8,6	73,96	4,01	–0,01	0,0001
0	5	9,6	92,16	4,97	0,03	0,0009
1	2	6,6	43,56	2,01	–0,01	0,0001
3	–39	–34,4	1183	–38,95	–0,05	0,0025
$\sum_{i=1}^{5}(y_i - \bar{y}^*)^2 = 1485$				$\sum_{i=1}^{5}(y_i - \tilde{y}_i)^2 = 0{,}0061$		

Оценка математического ожидания выходной переменной y, которая используется при расчётах в табл.9.5, найдена по формуле (5.1.1):

$$\bar{y}^* = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{1}{5}(5+4+5+2-39) = -4{,}6\,.$$

Необходимые данные для вычисления оценок дисперсий σ^2 и σ_1^2 берём из табл.9.5:

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}^*)^2}{n-1} = \frac{1485}{4} = 371{,}25;$$

$$\tilde{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n-k-1} = \frac{0{,}0061}{5-4} = 0{,}0061\,.$$

Наблюдаемое значение показателя согласованности (9.2.23):

$$\mathsf{F} = \frac{\tilde{\sigma}^2}{\tilde{\sigma}_1^2} = \frac{371{,}25}{0{,}0061} = 60860\,.$$

Критическое значение показателя согласованности при уровне значимости $\alpha = 0{,}01$ и степенях свободы $f_1 = n - 1 = 4$, $f_2 = n - k - 1 = 1$, находим в приложении 5:

$$\mathsf{F}_{(0{,}01;\,4;\,1)} = 5625.$$

Очевидно, что неравенство (9.2.24) выполняется, т.е. $\mathsf{F} > \mathsf{F}_{(0{,}01;\,4;\,1)}$. Следовательно, нулевая гипотеза H_0 о соответствии функции регрессии вида (9.2.31) экспериментальным данным принимается.

5. Проверяем значимость коэффициентов уравнения регрессии.

Для вычисления наблюдаемых значений показателя согласованности (9.2.30) необходимо найти диагональные элементы корреляционной матрицы (9.2.28). В рассматриваемом примере матрица F представляется следующим образом:

$$F = \begin{pmatrix} f_0(x_1) & f_1(x_1) & f_2(x_1) & f_3(x_1) \\ f_0(x_2) & f_1(x_2) & f_2(x_2) & f_3(x_2) \\ f_0(x_3) & f_1(x_3) & f_2(x_3) & f_3(x_3) \\ f_0(x_4) & f_1(x_4) & f_2(x_4) & f_3(x_4) \\ f_0(x_5) & f_1(x_5) & f_2(x_5) & f_3(x_5) \end{pmatrix} = \begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ x_3^3 & x_3^2 & x_3 & 1 \\ x_4^3 & x_4^2 & x_4 & 1 \\ x_5^3 & x_5^2 & x_5 & 1 \end{pmatrix}. \quad (9.2.37)$$

Находим произведение транспонированной матрицы F на исходную:

$$F_{[5;4]}^{\mathsf{T}} F_{[4;5]} = \begin{pmatrix} x_1^3 & x_2^3 & x_3^3 & x_4^3 & x_5^3 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 & x_5^2 \\ x_1 & x_2 & x_3 & x_4 & x_5 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ x_3^3 & x_3^2 & x_3 & 1 \\ x_4^3 & x_4^2 & x_4 & 1 \\ x_5^3 & x_5^2 & x_5 & 1 \end{pmatrix} =$$

$$=\begin{pmatrix} \sum\limits_{i=1}^{5} x_i^6 & \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 \\ \sum\limits_{i=1}^{5} x_i^5 & \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 \\ \sum\limits_{i=1}^{5} x_i^4 & \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 & \sum\limits_{i=1}^{5} x_i \\ \sum\limits_{i=1}^{5} x_i^3 & \sum\limits_{i=1}^{5} x_i^2 & \sum\limits_{i=1}^{5} x_i & 5 \end{pmatrix} = \begin{pmatrix} 795 & 211 & 99 & 19 \\ 211 & 99 & 19 & 15 \\ 99 & 19 & 15 & 1 \\ 19 & 15 & 1 & 5 \end{pmatrix}. \quad (9.2.38)$$

Далее требуется найти элементы главной диагонали матрицы $(F^{\mathsf{T}}F)^{-1}$:

$$diag(F_{[5;4]}^{\mathsf{T}}F_{[4;5]})^{-1} = \left(\frac{A_{11}}{|F^{\mathsf{T}}F|}, \frac{A_{22}}{|F^{\mathsf{T}}F|}, \frac{A_{33}}{|F^{\mathsf{T}}F|}, \frac{A_{44}}{|F^{\mathsf{T}}F|} \right),$$

где $|F^{\mathsf{T}}F|$ – определитель матрицы $F^{\mathsf{T}}F$; A_{ii}, $i=\overline{1,4}$ – алгебраические дополнения элементов главной диагонали этой же матрицы.

Вычисляем определитель и алгебраические дополнения:

$$|F^{\mathsf{T}}F| = \begin{vmatrix} 795 & 211 & 99 & 19 \\ 211 & 99 & 19 & 15 \\ 99 & 19 & 15 & 1 \\ 19 & 15 & 1 & 5 \end{vmatrix} = 106848;$$

$$A_{11} = (-1)^2 \begin{vmatrix} 99 & 19 & 15 \\ 19 & 15 & 1 \\ 15 & 1 & 5 \end{vmatrix} = 2716; \qquad A_{22} = (-1)^4 \begin{vmatrix} 795 & 99 & 19 \\ 99 & 15 & 1 \\ 19 & 1 & 5 \end{vmatrix} = 8172;$$

$$A_{33} = (-1)^6 \begin{vmatrix} 795 & 211 & 19 \\ 211 & 99 & 1 \\ 19 & 15 & 5 \end{vmatrix} = 76576; \qquad A_{44} = (-1)^8 \begin{vmatrix} 795 & 211 & 99 \\ 211 & 99 & 19 \\ 99 & 19 & 15 \end{vmatrix} = 49248.$$

Таким образом, имеем

$$diag(F^{\mathsf{T}}F)^{-1} = \left(\frac{2716}{106848}; \frac{8172}{106848}; \frac{76576}{106848}; \frac{49248}{106848} \right) =$$

$$= (0{,}025;\ 0{,}076;\ 0{,}717;\ 9{,}461).$$

Главная диагональ (9.2.29) корреляционной матрицы вектора оценок коэффициентов регрессии:

$$diagK_{\tilde{A}[4]} = \left(\tilde{\sigma}_{\tilde{a}_0}^2,\ \tilde{\sigma}_{\tilde{a}_1}^2,\ \tilde{\sigma}_{\tilde{a}_2}^2,\ \tilde{\sigma}_{\tilde{a}_3}^2 \right) = (0{,}0061 \cdot 0{,}025;\ \ 0{,}061 \cdot 0{,}076;$$

$$0{,}0061 \cdot 0{,}717;\ \ 0{,}0061 \cdot 0{,}461) = (1{,}5 \cdot 10^{-4};\ \ 4{,}6 \cdot 10^{-4};\ \ 4{,}3 \cdot 10^{-4};\ \ 28{,}1 \cdot 10^{-4})..$$

Вычисляем оценки средних квадратических отклонений коэффициентов $\tilde{a}_j$, $j=\overline{0,3}$:

$$\sigma_{\tilde{a}_0} = \sqrt{1{,}5 \cdot 10^{-4}} = 0{,}012; \qquad \sigma_{\tilde{a}_1} = \sqrt{4{,}6 \cdot 10^{-4}} = 0{,}021;$$

$$\sigma_{\tilde{a}_2} = \sqrt{43{,}7 \cdot 10^{-4}} = 0{,}066; \qquad \sigma_{\tilde{a}_3} = \sqrt{28{,}1 \cdot 10^{-4}} = 0{,}053.$$

Находим наблюдаемые значения показателя (9.2.27):

$$t_0 = \frac{0{,}97}{0{,}012} = 80{,}8; \quad t_1 = \frac{1{,}96}{0{,}021} = 93{,}3; \quad t_2 = \frac{0{,}03}{0{,}066} = 0{,}45; \quad t_3 = \frac{4{,}97}{0{,}053} = 93{,}8.$$

Критическое значение показателя согласованности при уровне значимости $\alpha = 0{,}01$ и одной степени свободы $f = 1$ находим в приложении 6: $t_{(0{,}01;\,1)} = 63{,}7$.

Проверяем условие (9.2.30) и получаем:

$$t_0 > t_{(0{,}01;\,1)}; \quad t_1 > t_{(0{,}01;\,1)}; \quad t_2 < t_{(0{,}01;\,1)}; \quad t_3 > t_{(0{,}01;\,1)}.$$

На основе приведённых неравенств делаем вывод, что коэффициенты a_0, a_1 и a_3 являются значимыми, а коэффициент a_2 принимаем равным нулю.

Окончательный вид уравнения регрессии

$$\tilde{y} = -0{,}97x^3 - 1{,}96x^2 + 4{,}97.$$

6. Проверяем адекватность последнего уравнения по критерию Фишера. Составляем табл.9.6.

Таблица 9.6

Расчётная таблица

x_i	y_i	$\tilde{y}_i$	$y_i - \tilde{y}_i$	$(y_i - \tilde{y}_i)^2$
–2	5	4,89	0,11	0,0121
–1	4	3,98	0,02	0,0004
0	5	4,97	0,03	0,0009
1	2	2,04	–0,04	0,0016
3	–39	–38,86	–0,14	0,0196
$\sum_{i=1}^{5}(y_i - \tilde{y}_i)^2 = 0{,}0346$				

Вычисляем оценку остаточной дисперсии с учётом результата, полученного в табл.9.6:

$$\tilde{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n - k - 1} = \frac{0{,}0346}{1} = 0{,}0346.$$

Общая дисперсия остаётся прежней. Поэтому наблюдаемое значение показателя согласованности (9.2.33) будет следующим:

$$\mathsf{F} = \frac{\tilde{\sigma}^2}{\tilde{\sigma}_1^2} = \frac{371{,}25}{0{,}0346} = 10730.$$

Очевидно, что при критическом значении показателя согласованности $\mathsf{F}_{(0{,}01;\,4;\,1)} = 5625$ неравенство (9.2.24) выполняется. Таким образом, повторная проверка адекватности по критерию Фишера подтверждает справедливость гипотезы о соответствии функции регрессии вида (9.2.31) экспериментальным данным.

▲

<u>П р и м е р 9.2</u>. В условиях примера 9.1 построить уравнение регрессии, но расчёты произвести в матричной форме.

▼ Матричное уравнение (9.2.19) для рассматриваемого примера имеет вид

$$(F^{\mathrm{T}}_{[4;5]}F_{[5;4]})\widetilde{A}_{<4>} = F^{\mathrm{T}}_{[4;5]}Y_{<5>}F^{\mathrm{T}}F\,, \qquad (9.2.39)$$

где матрицы F и $F^{\mathrm{T}}F$ представлены выражениями (9.2.37), (9.2.38) соответственно;

$$\widetilde{A}_{<4>} = (\tilde{a}_0,\ \tilde{a}_1,\ \tilde{a}_2,\ \tilde{a}_3)^{\mathrm{T}}\,; \quad Y_{<5>} = (y_1,\ y_2,\ y_3,\ y_4,\ y_5)^{\mathrm{T}} = (5;\ 4;\ 5;\ 2;\ -39)^{\mathrm{T}}\,.$$

Выражение (9.2.20) для вычисления оценок коэффициентов регрессии представляется в виде

$$\widetilde{A}_{<4>} = (F^{\mathrm{T}}_{[4;5]}F_{[5;4]})^{-1}F^{\mathrm{T}}_{[4;5]}Y_{<5>}\,. \qquad (9.2.40)$$

Вычисляем обратную матрицу $(F^{\mathrm{T}}F)^{-1}$ для матрицы

$$F^{\mathrm{T}}_{[4;5]}F_{[5;4]} = \begin{pmatrix} 795 & 211 & 99 & 19 \\ 211 & 99 & 19 & 15 \\ 99 & 19 & 15 & 1 \\ 19 & 15 & 1 & 5 \end{pmatrix}.$$

Алгебраические дополнения элементов матрицы $F^{\mathrm{T}}F$:

$A_{11} = 2716$; $A_{21} = -3780$; $A_{31} = -13384$; $A_{41} = 3696$;
$A_{12} = -3780$; $A_{22} = 8172$; $A_{32} = 15480$; $A_{42} = -13248$;
$A_{13} = -13384$; $A_{23} = 126304$; $A_{33} = 76576$; $A_{43} = -10896$;
$A_{14} = 3696$; $A_{24} = -13248$; $A_{34} = -10896$; $A_{44} = 49248$.

Определитель указанной матрицы $|F^{\mathrm{T}}F| = 106848$.

В результате получили обратную матрицу

$$(F^{\mathrm{T}}F)^{-1} = \frac{1}{|F^{\mathrm{T}}F|} = \begin{pmatrix} A_{11} & A_{21} & A_{31} & A_{41} \\ A_{12} & A_{22} & A_{32} & A_{42} \\ A_{13} & A_{23} & A_{33} & A_{43} \\ A_{14} & A_{24} & A_{34} & A_{44} \end{pmatrix} =$$

$$= \frac{1}{106848}\begin{pmatrix} 2716 & -3780 & -13384 & 3696 \\ -3780 & 8172 & 15480 & -13248 \\ -13384 & 15480 & 76576 & -10896 \\ 3696 & -13248 & -10896 & 49248 \end{pmatrix}.$$

Далее вычисляем матрицу в правой части уравнения (9.2.39):

$$F^{\mathsf{T}}_{[5;4]}Y_{<5>}=\begin{pmatrix} x_1^3 & x_2^3 & x_3^3 & x_4^3 & x_5^3 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 & x_5^2 \\ x_1 & x_2 & x_3 & x_4 & x_5 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}=$$

$$=\begin{pmatrix} -8 & -1 & 0 & 1 & 27 \\ 4 & 1 & 0 & 1 & 9 \\ -2 & -1 & 0 & 1 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 5 \\ 4 \\ 5 \\ 2 \\ -39 \end{pmatrix}=\begin{pmatrix} -1095 \\ -325 \\ -129 \\ -23 \end{pmatrix}.$$

По формуле (9.2.40) находим оценки коэффициентов регрессии

$$\tilde{A}_{<4>}=\begin{pmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \tilde{a}_2 \\ \tilde{a}_3 \end{pmatrix}=\frac{1}{106848}\begin{pmatrix} 2716 & -3780 & -13384 & 3696 \\ -3780 & 8172 & 15480 & -13248 \\ -13384 & 15480 & 76576 & -10896 \\ 3696 & -13248 & -10896 & 49248 \end{pmatrix}\begin{pmatrix} -1095 \\ -325 \\ -129 \\ -23 \end{pmatrix}=$$

$$=\frac{1}{106848}\begin{pmatrix} -103992 \\ -209016 \\ -3216 \\ 531360 \end{pmatrix}=\begin{pmatrix} -0{,}97 \\ -1{,}96 \\ -0{,}03 \\ 4{,}97 \end{pmatrix}.$$

Так же, как и в примере 9.1 при расчётах в скалярной форме, получили уравнение регрессии

$$\tilde{y}=-0{,}97x^3-1{,}96x^2-0{,}03x+4{,}97\,.$$

Проверка адекватности уравнения экспериментальным данным и значимость коэффициентов регрессии выполняется аналогично тому, как это сделано в примере 9.1.

▲

9.3. Многофакторный линейный регрессионный анализ

9.3.1. Модели многофакторного линейного регрессионного анализа

В § 9.2 рассматривались модели однофакторного регрессионного анализа, линейные относительно коэффициентов регрессии. В то же время они могут быть нелинейными относительно независимой переменной (фактора), в примерах 9.1 и 9.2 рассматривалась именно такая нелинейная модель.

В настоящем подпараграфе рассмотрим модели многофакторного (множественного) регрессионного анализа, являющиеся линейными как относительно коэффициентов регрессии, так и относительно факторов.

Модель РА-1 определяется выражением

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_k x_k , \qquad (9.3.1)$$

а модель РА-2 – выражением

$$\hat{y} = b_0 + b_1 \hat{x}_1 + b_2 \hat{x}_2 + ... + b_k \hat{x}_k . \qquad (9.3.2)$$

Условное математическое ожидание (9.1.1) результата наблюдения представляется как

$$M[\hat{y} \mid X_{<k>}] = b_0 + b_1 x_1 + b_2 x_2 + ... + b_k x_k .$$

Для модели РА-1 экспериментальные данные могут быть представлены табл.9.7.

Таблица 9.7

Представление результатов многофакторного эксперимента (модель РА-1)

Опыты	Факторы						Результаты наблюдений						Средние значения результатов наблюдений $\bar{y}^*$
	x_1	x_2	…	x_j	…	x_k	y_1	y_2	…	y_s	…	y_m	
1	x_{11}	x_{12}	…	x_{1j}	…	x_{1k}	y_{11}	y_{12}	…	y_{1s}	…	y_{1m}	$\bar{y}_1^*$
2	x_{21}	x_{22}	…	x_{2j}	…	x_{2k}	y_{21}	y_{22}	…	y_{2s}	…	y_{2m}	$\bar{y}_2^*$
…	…	…	…	…	…	…	…	…	…	…	…	…	…
i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{ik}	y_{i1}	y_{i2}	…	y_{is}	…	y_{im}	$\bar{y}_i^*$
…	…	…	…	…	…	…	…	…	…	…	…	…	…
n	x_{n1}	x_{n2}	…	x_{nj}	…	x_{nk}	y_{n1}	y_{n2}	…	y_{ns}	…	y_{nm}	$\bar{y}_n^*$

В данной таблице значения факторов $X_{<k>i}$, $i = \overline{1, n}$ являются фиксированными, что даёт возможность получить при данных значениях m результатов наблюдений. В частном случае может быть $m = 1$. При построении регрессионных зависимостей используется среднее значение результатов.

Для модели РА-2 экспериментальные данные представляются табл.9.8. В данной таблице имеют место системы случайных величин $(\hat{y}_i ; \hat{X}_{<k>i})$, $i = \overline{1, n}$, т.е. каждому сочетанию значений факторов $\hat{X}_{<k>}$ соответствует одно значение результата $\hat{y}$.

В рассматриваемом случае линейного регрессионного анализа методы построения моделей вида (9.3.1) и (9.3.2) одинаковы. Это связано с тем, что математические ожидания ошибок наблюдений равны нулю. Предположения, рассмотренные в п.п.9.2.1, остаются в силе.

Таблица 9.8

Представление результатов многофакторного эксперимента (модель РА-2)

Опыты	Факторы						Результаты наблюдений
	x_1	x_2	…	x_j	…	x_k	
1	x_{11}	x_{12}	…	x_{1j}	…	x_{1k}	y_1
2	x_{21}	x_{22}	…	x_{2j}	…	x_{2k}	y_2
…	…	…	…	…	…	…	…
i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{ik}	y_i
…	…	…	…	…	…	…	…
n	x_{n1}	x_{n2}	…	x_{nj}	…	x_{nk}	y_n

9.3.2. Построение уравнения множественной регрессии

Задача построения уравнения регрессии сводится к оцениванию коэффициентов

$$(b_0,\ b_1,\ b_2,...,b_k)^{\mathrm{т}} = B_{<k+1>} \tag{9.3.3}$$

в выражении

$$y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n = b_0 + \sum_{j=1}^{k} b_jx_j\ . \tag{9.3.4}$$

Поскольку значения факторов могут иметь различный порядок, то для упрощения вычислений целесообразно использовать их центрированные значения

$$\dot{x}_{ij} = x_{ij} - \bar{x}_j^{\,*},\quad i = \overline{1,\,n},\quad j = \overline{1,\,k}, \tag{9.3.5}$$

где $\bar{x}_j^{\,*} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_{ij}$.

Кроме этого вводится n-мерный единичный вектор

$$\dot{X}_{<n>} = (1,\,1,...,1)^{\mathrm{т}}\,,$$

что необходимо для оценки свободного члена уравнения регрессии. Матрица центрированных значений факторов при этом имеет вид

$$\dot{X}_{[n;\,k+1]} = \begin{pmatrix} 1 & \dot{x}_{11} & \dot{x}_{12} & \cdots & \dot{x}_{1j} & \cdots & \dot{x}_{1k} \\ 1 & \dot{x}_{21} & \dot{x}_{22} & \cdots & \dot{x}_{2j} & \cdots & \dot{x}_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & \dot{x}_{i1} & \dot{x}_{i2} & \cdots & \dot{x}_{ij} & \cdots & \dot{x}_{ik} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & \dot{x}_{n1} & \dot{x}_{n2} & \cdots & \dot{x}_{nj} & \cdots & \dot{x}_{nk} \end{pmatrix}. \tag{9.3.6}$$

Из вышеизложенного следует, что коэффициенты регрессии оцениваются в уравнении

$$y = b_0 x_0 + b_1 \dot{x}_1 + b_2 \dot{x}_2 + \ldots + b_n \dot{x}_n = \sum_{j=0}^{k} b_j \dot{x}_j \,. \qquad (9.3.7)$$

После вычисления данных оценок возврат к уравнению вида (9.3.4) осуществляется подстановкой (9.3.5). При этом значение коэффициента b_0 изменяется.

Подходящую оценку $\widetilde{B}_{<k+1>}$ вектора $B_{<k+1>} = (b_0,\ b_1,\ b_2,\ldots,b_k)^{\mathrm{T}}$ находим методом наименьших квадратов. Так же, как и в случае однофакторной регрессии, необходимо минимизировать сумму квадратов невязок (9.2.10). В данной формуле оценки $\widetilde{M}_{\hat{y}_i}$ математических ожиданий результата находятся из соотношения

$$\widetilde{y} = \widetilde{M}_{\hat{y}} = \sum_{j=0}^{k} \widetilde{b}_j \dot{x}_j \,. \qquad (9.3.8)$$

Учитывая (9.3.8), выражение (9.2.10) принимает вид

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} \widetilde{b}_j \dot{x}_j \right)^2 . \qquad (9.3.9)$$

Таким образом, найденные оценки $\widetilde{b}_0,\ \widetilde{b}_1,\ldots,\widetilde{b}_k$ должны удовлетворять условию

$$V = \min_{\widetilde{b}_j \in \mathbf{R}} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} \widetilde{b}_j \dot{x}_j \right)^2 \right\}. \qquad (9.3.10)$$

Для определения минимума (9.3.10) функции (9.3.9) составляем систему нормальных уравнений вида (8.2.10):

$$\begin{cases} \sum_{i=1}^{n} \left(y_i - (\widetilde{b}_0 \dot{x}_0 + \widetilde{b}_1 \dot{x}_1 + \ldots + \widetilde{b}_k \dot{x}_k) \right) \dot{x}_0 = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_{i=1}^{n} \left(y_i - (\widetilde{b}_0 \dot{x}_0 + \widetilde{b}_1 \dot{x}_1 + \ldots + \widetilde{b}_k \dot{x}_k) \right) \dot{x}_j = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_{i=1}^{n} \left(y_i - (\widetilde{b}_0 \dot{x}_0 + \widetilde{b}_1 \dot{x}_1 + \ldots + \widetilde{b}_k \dot{x}_k) \right) \dot{x}_k = 0 \end{cases} . \qquad (9.3.11)$$

Уравнения (9.3.11) получены исходя из того, что частная производная квадратичной функции (9.3.9) по переменной $\widetilde{b}_j$ находится следующим образом:

$$\frac{\partial\left(\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}\tilde{b}_j\dot{x}_j\right)\right)^2}{\partial b_j}=2\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}\tilde{b}_j\dot{x}_j\right)\frac{\partial\left(\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}\tilde{b}_j\dot{x}_j\right)\right)}{\partial b_j}=$$

$$=2\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}b_j\dot{x}_j\right)(-\dot{x}_j)=-2\sum_{i=1}^{n}\left(y_i-\sum_{j=0}^{k}b_j\dot{x}_j\right)\dot{x}_j= \qquad (9.3.12)$$

$$=-2\sum_{i=1}^{n}(y_i-(b_0\dot{x}_0+b_1\dot{x}_1+b_2\dot{x}_2+\ldots+b_k\dot{x}_k))\dot{x}_j.$$

Частную производную (9.3.12) приравниваем к нулю и обе части полученного уравнения умножаем на –2. В результате имеем j-е уравнение системы (9.3.11). Далее выполняем почленное суммирование в уравнениях рассматриваемой системы, переносим в правую часть слагаемые, содержащие y_i, а затем умножаем на –1 обе части каждого уравнения. Указанная последовательность операции приводит к эквивалентной системе уравнений

$$\begin{cases}\tilde{b}_0\sum_{i=1}^{n}\dot{x}_{i0}^2+\tilde{b}_1\sum_{i=1}^{n}\dot{x}_{i1}\dot{x}_{i0}+\ldots+\tilde{b}_k\sum_{i=1}^{n}\dot{x}_{ik}\dot{x}_{i0}=\sum_{i=1}^{n}y_i\dot{x}_{i0}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \tilde{b}_0\sum_{i=1}^{n}\dot{x}_{i0}\dot{x}_{ij}+\tilde{b}_1\sum_{i=1}^{n}\dot{x}_{i1}\dot{x}_{ij}+\ldots+\tilde{b}_k\sum_{i=1}^{n}\dot{x}_{ik}\dot{x}_{ij}=\sum_{i=1}^{n}y_i\dot{x}_{ij}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \tilde{b}_0\sum_{i=1}^{n}\dot{x}_{i0}\dot{x}_{ik}+\tilde{b}_1\sum_{i=1}^{n}\dot{x}_{i1}\dot{x}_{ik}+\ldots+\tilde{b}_k\sum_{i=1}^{n}\dot{x}_{ik}^2=\sum_{i=1}^{n}y_i\dot{x}_{ik}.\end{cases}$$

Учитывая, что $\dot{x}_{i0}$ = 1, окончательно получаем

$$\begin{cases}\tilde{b}_0 n+\tilde{b}_1\sum_{i=1}^{n}\dot{x}_{i1}+\ldots+\tilde{b}_k\sum_{i=1}^{n}\dot{x}_{ik}=\sum_{i=1}^{n}y_i\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \tilde{b}_0\sum_{i=1}^{n}\dot{x}_{ij}+\tilde{b}_1\sum_{i=1}^{n}\dot{x}_{i1}\dot{x}_{ij}+\ldots+\tilde{b}_k\sum_{i=1}^{n}\dot{x}_{ik}\dot{x}_{ij}=\sum_{i=1}^{n}y_i\dot{x}_{ij}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \tilde{b}_0\sum_{i=1}^{n}\dot{x}_{ik}+\tilde{b}_1\sum_{i=1}^{n}\dot{x}_{i1}\dot{x}_{ik}+\ldots+\tilde{b}_k\sum_{i=1}^{n}\dot{x}_{ik}^2=\sum_{i=1}^{n}y_i\dot{x}_{ik}.\end{cases} \qquad (9.3.13)$$

Система (9.3.13) является системой линейных уравнений относительно оценок коэффициентов регрессии $\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_k$. Данные оценки находим по формулам Крамера:

$$\tilde{b}_0 = \frac{|B_0|}{|B|}, ..., \tilde{b}_j = \frac{|B_j|}{|B|}, ..., \tilde{b}_k = \frac{|B_k|}{|B|}, \qquad (9.3.14)$$

где $|B|$ – определитель коэффициентов при неизвестных системы уравнений (9.3.13); $|B_j|$, $j = \overline{0, k}$ – определители, полученные из определителя $|B|$ заменой j-го столбца столбцом свободных членов.

Развёрнутый вид данных определителей:

$$|B| = \begin{vmatrix} n & \sum_{i=1}^{n} \dot{x}_{i1} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \dot{x}_{ij} & \sum_{i=1}^{n} \dot{x}_{i1}\dot{x}_{ij} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik}\dot{x}_{ij} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \dot{x}_{ik} & \sum_{i=1}^{n} \dot{x}_{i1}\dot{x}_{ik} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik}^2 \end{vmatrix}; \quad |B_0| = \begin{vmatrix} \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} \dot{x}_{i1} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} y_i\dot{x}_{ij} & \sum_{i=1}^{n} \dot{x}_{i1}\dot{x}_{ij} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik}\dot{x}_{ij} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} y_i\dot{x}_{ik} & \sum_{i=1}^{n} \dot{x}_{i1}\dot{x}_{ik} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik}^2 \end{vmatrix};$$

$$|B_j| = \begin{vmatrix} n & \cdots & \sum_{i=1}^{n} y_i & \cdots & \sum_{i=1}^{n} \dot{x}_{ik} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \dot{x}_{ij} & \cdots & \sum_{i=1}^{n} y_i\dot{x}_{ij} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik}\dot{x}_{ij} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \dot{x}_{ik} & \cdots & \sum_{i=1}^{n} y_i\dot{x}_{ik} & \cdots & \sum_{i=1}^{n} \dot{x}_{ik}^2 \end{vmatrix}; \quad |B_k| = \begin{vmatrix} n & \sum_{i=1}^{n} \dot{x}_{i1} & \cdots & \sum_{i=1}^{n} y_i \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \dot{x}_{ij} & \sum_{i=1}^{n} \dot{x}_{i1}\dot{x}_{ij} & \cdots & \sum_{i=1}^{n} y_i\dot{x}_{ij} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \dot{x}_{ik} & \sum_{i=1}^{n} \dot{x}_{i1}\dot{x}_{ik} & \cdots & \sum_{i=1}^{n} y_i\dot{x}_{ik} \end{vmatrix}.$$

Система уравнений (9.2.13) может быть записана в виде матричного уравнения

$$\left(\dot{X}^{\mathsf{T}}_{[k+1;\,n]}\dot{X}_{[n;\,k+1]}\right)\tilde{B}_{<k+1>} = \dot{X}^{\mathsf{T}}_{[k+1;\,n]}Y_{<n>}, \qquad (9.3.15)$$

где
$$\dot{X} = \begin{pmatrix} 1 & \dot{x}_{11} & \cdots & \dot{x}_{1k} \\ \cdots & \vdots & \cdots & \cdots \\ 1 & \dot{x}_{i1} & \cdots & \dot{x}_{ik} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & \dot{x}_{n1} & \cdots & \dot{x}_{nk} \end{pmatrix},$$

$$\tilde{B} = (\tilde{b}_0,\ \tilde{b}_1, ..., \tilde{b}_k)^{\mathsf{T}}; \quad Y = (y_1,\ y_2, ..., y_n)^{\mathsf{T}}.$$

Для получения вектора $\tilde{B}$ умножаем обе части матричного уравнения (9.3.15) на $(\dot{X}^{\mathsf{T}}\dot{X})^{-1}$ слева:

$$(\dot{X}^{\mathsf{T}}\dot{X})^{-1}(\dot{X}^{\mathsf{T}}\dot{X})\tilde{B} = (\dot{X}^{\mathsf{T}}\dot{X})^{-1}\dot{X}^{\mathsf{T}}Y,$$

отсюда

$$E\tilde{B} = (\dot{X}^{\mathsf{T}}\dot{X})^{-1}\dot{X}^{\mathsf{T}}Y$$

или

$$\widetilde{B}_{<k+1>} = \left(\dot{X}^{\mathsf{T}}_{[k+1;\, n]} \dot{X}_{[n;\, k+1]}\right)^{-1} \dot{X}^{\mathsf{T}}_{[k+1;\, n]} Y_{<n>}, \qquad (9.3.16)$$

где $E = E_{[k+1]}$ – единичная матрица порядка $(k+1)$.

Рассмотренный метод построения уравнения регрессии применим как для модели РА-1, так и РА-2.

9.3.3. Проверка адекватности уравнения множественной регрессии

Подобно однофакторному регрессионному анализу, проверка адекватности экспериментальным данным уравнения множественной регрессии производится на основании анализа отношения дисперсий. В качестве показателя согласованности нулевой гипотезы H_0 об адекватности уравнения принимается выражение, аналогичное выражению (9.2.23):

$$\hat{F} = \frac{\widetilde{\sigma}^2}{\widetilde{\sigma}_1^2}, \qquad (9.3.17)$$

где $\widetilde{\sigma}^2$ – оценка общей дисперсии наблюдаемой переменной; $\widetilde{\sigma}_1^2$ – оценка остаточной дисперсии.

Оценки дисперсий, входящих в соотношение (9.3.17), определяются по формулам

$$\widetilde{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y}^*)^2}{n-1}, \qquad (9.3.18)$$

$$\widetilde{\sigma}_1^2 = \frac{\sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{k} \widetilde{b}_j \dot{x}_{ij} \right)^2}{n-k-1} = \frac{\sum_{i=1}^{n} (y_i - \widetilde{y}_i)^2}{n-k-1}, \qquad (9.3.19)$$

где $\bar{y}^*$ – оценка математического ожидания наблюдаемой переменной; $n-1$ – число степеней свободы дисперсии σ^2; $n-k-1$ – число степеней свободы дисперсии σ_1^2.

Случайная величина (9.3.17) подчинена закону распределения Фишера [1]. Следовательно, для принятия нулевой гипотезы необходимо и достаточно, чтобы выполнялось условие

$$\mathsf{F} = \frac{\widetilde{\sigma}^2}{\widetilde{\sigma}_1^2} > \mathsf{F}_{(\alpha;\, n-1;\, n-k-1)}, \qquad (9.3.20)$$

где F –наблюдаемое значение показателя согласованности гипотезы H_0; $\mathsf{F}_{(\alpha;\, n-1;\, n-k-1)}$ – критическое значение данного показателя при уровне значимости α и степенях свободы $f_1 = n-1$, $f_2 = n-k-1$. Критическое значение показателя берётся по таблице критических точек распределения Фишера (приложение 5).

При невыполнении условия (9.3.20) принимается конкурирующая гипотеза H_1 о том, что линейная модель (9.3.1) или (9.3.2) неадекватна экспериментальным данным.

9.3.4. Селекция факторов

Уравнение регрессии связывает наблюдаемую переменную с совокупностью k факторов. Можно предположить (выдвинуть гипотезу), что какая-то часть этих факторов не оказывает существенного влияния на величину переменной y. Такие факторы без ущерба для точности могут быть исключены из уравнения. Возникает задача выявления таких факторов. Значимость любого фактора определяется, прежде всего, величиной коэффициента регрессии при данном факторе. Следовательно, задача селекции факторов сводится к проверке значимости коэффициентов регрессии. Такая задача решалась в однофакторном регрессионного анализе (пп. 9.2.4).

Следуя представленной в указанном подпараграфе схеме, необходимо проверить нулевые гипотезы о неравенстве нулю коэффициентов b_j:

$$H_{0j}\colon b_j \neq 0, \quad j = \overline{0, k}.$$

Конкурирующие гипотезы состоят в предположении о равенстве нулю коэффициентов:

$$H_{1j}\colon b_j = 0.$$

В качестве показателя согласованности при проверке гипотез H_{0j} используется выражение

$$\hat{t}_j = \frac{|\tilde{b}_j|}{\tilde{\sigma}_{\tilde{b}_j}}, \quad j = \overline{0, k}, \tag{9.3.21}$$

где $|\tilde{b}_j|$ – модуль величины $\tilde{b}_j$; $\tilde{\sigma}_{\tilde{b}_j}$ – оценка среднего квадратического отклонения коэффициента $\tilde{b}_j$.

Известно [1], что случайная величина (9.3.21) подчинена закону распределения Стьюдента.

Для того, чтобы выражение (9.3.21) можно было использовать практически, необходимо иметь методику вычисления величин $\tilde{\sigma}_{\tilde{b}_j}$. Она аналогична представленной в пп.9.2.4.

Рассматриваем корреляционную матрицу вектора $\tilde{B}_{<k+1>}$:

$$K_{\tilde{B}[k+1]} = \tilde{\sigma}_1^2 (\dot{X}^{\mathsf{T}}_{[k+1;\,n]} \dot{X})^{-1}. \tag{9.3.22}$$

Элементы главной диагонали матрицы (9.3.22) и являются дисперсиями оценок коэффициентов регрессии:

$$diagK_{\tilde{B}} = (\tilde{\sigma}^2_{\tilde{b}_0},\ \tilde{\sigma}^2_{\tilde{b}_1},\ldots,\tilde{\sigma}^2_{\tilde{b}_k}).$$

Проверяются условия

$$t_j = \frac{|\tilde{b}_j|}{\tilde{\sigma}_{\tilde{b}_j}} > t_{(\alpha;n-k-1)}, \quad j = \overline{0,k}, \qquad (9.3.23)$$

где t_j – вычисленное значение показателя согласованности гипотезы H_{0j}; $t_{(\alpha;n-k-1)}$ – критическое значение данного показателя при уровне значимости α и $f_1 = n - k - 1$ степенях свободы. Критическое значение $t_{(\alpha;n-k-1)}$ берётся из таблицы критических точек распределения Стьюдента (приложение 6).

Коэффициенты регрессии, для которых условие (9.3.23) не выполняется, принимаются равными нулю. Следовательно, соответствующие им факторы являются незначимыми. Проверка условий (9.3.23) называется селекцией факторов по критерию Стьюдента.

9.3.5. Пример многофакторного линейного регрессионного анализа

Из вышеизложенного очевидно, что многофакторный линейный регрессионный анализ проводится в следующей последовательности.

1. Определяется количество слагаемых в линейной функции регрессии вида (9.3.1) или (9.3.2) в зависимости от числа факторов.

2. Выполняется центрирование факторов.

3. Составляется система нормальных уравнений.

4. Находится решение системы нормальных уравнений (определяются оценки коэффициентов регрессии).

5. Проверяется адекватность построенного уравнения регрессии экспериментальным данным.

6. Производится селекция факторов.

7. Повторно проверяется адекватность уравнения регрессии после исключения незначимых факторов.

Пример 9.3. Для исследования зависимости выходного сигнала y системы от входного воздействия $X_{<2>} = (x_1, x_2)^{\mathrm{T}}$, проведены испытания, результаты которых сведены в табл.9.9.

Необходимо построить регрессионную зависимость y от $X_{<2>}$ в предположении, что она является линейным алгебраическим полиномом.

Таблица 9.9

Массив экспериментальных данных

x_1	–0,5	0	0,8	0,4	0,5	0,6
x_2	–3	–1	2	0,5	1,5	6
y	–15,1	–1	19,9	9,5	16,5	47,9

▼ 1. В рассматриваемой задаче функция регрессии представляет собой линейный алгебраический полином от двух независимых переменных

$$y = b_0 + b_1x_1 + b_2x_2.$$

2. Составляем таблицу экспериментальных данных с центрированными значениями факторов.

Предварительно вычисляем средние значения факторов:

$$\bar{x}_1^* = \frac{1}{6}\sum_{i=1}^{6} x_{i1} = \frac{1}{6}(-0,5+0+0,8+0,4+0,5+0,6) = 0,3;$$

$$\bar{x}_2^* = \frac{1}{6}\sum_{i=1}^{6} x_{i2} = 1;$$

Используя формулы (9.3.5), производим центрирование факторов, результаты заносятся в таб.9.10.

Т а б л и ц а 9.10

Массив экспериментальных данных с центрированными значениями факторов

$\dot{x}_1$	–0,8	–0,3	0,5	0,1	0,2	0,3
$\dot{x}_2$	–4	–2	1	–0,5	0,5	5
y	–15,1	–1	19,9	9,5	16,5	47,9

Таким образом, вначале коэффициенты регрессии оцениваем в выражении функции (9.3.7), которая для данной задачи принимает вид

$$y = b_0 + b_1\dot{x}_1 + b_2\dot{x}_2 \qquad (9.3.24)$$

3. Система нормальных уравнений представляется следующим образом:

$$\begin{cases} 6\tilde{b}_0 + \tilde{b}_1\sum_{i=1}^{6}\dot{x}_{i1} + \tilde{b}_2\sum_{i=1}^{6}\dot{x}_{i2} = \sum_{i=1}^{2} y_1 \\ \tilde{b}_0\sum_{i=1}^{6}\dot{x}_{i1} + \tilde{b}_1\sum_{i=1}^{6}\dot{x}_{i1}^2 + \tilde{b}_2\sum_{i=1}^{6}\dot{x}_{i1}\dot{x}_{i2} = \sum_{i=1}^{6} y_1\dot{x}_{i1} \\ \tilde{b}_0\sum_{i=1}^{6}\dot{x}_{i2} + \tilde{b}_1\sum_{i=1}^{6}\dot{x}_{i1}\dot{x}_{i2} + \tilde{b}_2\sum_{i=1}^{6}\dot{x}_{i2}^2 = \sum_{i=1}^{6} y_1\dot{x}_{i2}. \end{cases} \qquad (9.3.25)$$

4. Представим и решим систему (9.3.25) в матричной форме.

Матричное уравнение, эквивалентное данной системе, принимает вид

$$\left(\dot{X}^{\mathsf{T}}_{[3;6]}\dot{X}_{[6;3]}\right)\tilde{B}_{<3>} = \dot{X}^{\mathsf{T}}_{[3;6]}Y_{<6>}, \qquad (9.3.26)$$

где
$$\dot{X}=\begin{pmatrix}1 & \dot{x}_{11} & \dot{x}_{12}\\ 1 & \dot{x}_{21} & \dot{x}_{22}\\ 1 & \dot{x}_{31} & \dot{x}_{23}\\ 1 & \dot{x}_{41} & \dot{x}_{24}\\ 1 & \dot{x}_{51} & \dot{x}_{25}\\ 1 & \dot{x}_{61} & \dot{x}_{26}\end{pmatrix}=\begin{pmatrix}1 & -0{,}8 & -4\\ 1 & -0{,}3 & -2\\ 1 & 0{,}5 & 1\\ 1 & 0{,}1 & -0{,}5\\ 1 & 0{,}2 & 0{,}5\\ 1 & 0{,}3 & 5\end{pmatrix};$$
$$\widetilde{B}=(\widetilde{b}_0,\ \widetilde{b}_1,\ \widetilde{b}_2)^{\mathsf{T}};$$
$$Y=(y_1,\ y_2,\ y_3,\ y_4,\ y_5,\ y_6)^{\mathsf{T}}=(-15{,}1;\ -1;\ 19{,}9;\ 9{,}5;\ 16{,}5;\ 47{,}9)^{\mathsf{T}}.$$

Выражение (9.3.16) для вычисления оценок коэффициентов регрессии представляется равенством
$$\widetilde{B}_{<3>}=\left(\dot{X}^{\mathsf{T}}_{[3;6]}\dot{X}_{[6;3]}\right)^{-1}\dot{X}^{\mathsf{T}}_{[3;6]}Y_{<6>}. \qquad (9.3.27)$$

Вычисляем матрицу $\dot{X}^{\mathsf{T}}\dot{X}$:
$$\dot{X}^{\mathsf{T}}\dot{X}=\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1\\ -0{,}8 & -0{,}3 & 0{,}5 & 0{,}1 & 0{,}2 & 0{,}3\\ -4 & -2 & 1 & -0{,}5 & 0{,}5 & 5\end{pmatrix}\begin{pmatrix}1 & -0{,}8 & -4\\ 1 & -0{,}3 & -2\\ 1 & 0{,}5 & 1\\ 1 & 0{,}1 & -0{,}5\\ 1 & 0{,}2 & 0{,}5\\ 1 & 0{,}3 & 5\end{pmatrix}=$$
$$=\begin{pmatrix}6 & 0 & 0\\ 0 & 1{,}12 & 5{,}85\\ 0 & 5{,}85 & 46{,}5\end{pmatrix}.$$

Для полученной матрицы находим обратную матрицу:
$$\left(\dot{X}^{\mathsf{T}}\dot{X}\right)^{-1}=\frac{1}{107}\begin{pmatrix}17{,}9 & 0 & 0\\ 0 & 279 & -35{,}1\\ 0 & -35{,}1 & 6{,}72\end{pmatrix}. \qquad (9.3.28)$$

Далее находим матрицу в правой части уравнения (9.3.26):
$$\dot{X}^{\mathsf{T}}Y=\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1\\ -0{,}8 & -0{,}3 & 0{,}5 & 0{,}1 & 0{,}2 & 0{,}3\\ -4 & -2 & 1 & -0{,}5 & 0{,}5 & 5\end{pmatrix}\begin{pmatrix}-15{,}1\\ -1\\ 19{,}9\\ 9{,}5\\ 16{,}5\\ 47{,}9\end{pmatrix}=\begin{pmatrix}77{,}7\\ 40{,}9\\ 321\end{pmatrix}.$$

По формуле (9.3.27) вычисляем оценку вектора коэффициентов регрессии

$$\widetilde{B} = \frac{1}{107}\begin{pmatrix} 17{,}8 & 0 & 0 \\ 0 & 279 & -35{,}1 \\ 0 & -35{,}1 & 6{,}72 \end{pmatrix}\begin{pmatrix} 77{,}7 \\ 40{,}9 \\ 321 \end{pmatrix} = \begin{pmatrix} 12{,}9 \\ 1{,}35 \\ 6{,}74 \end{pmatrix}.$$

Получим следующее уравнение регрессии:

$$\widetilde{y} = 12{,}9 + 1{,}35\dot{x}_1 + 6{,}74\dot{x}_2. \qquad (9.3.29)$$

5. Проверяем адекватность уравнения (9.3.29) экспериментальным данным.

Предварительно вычисляем оценки дисперсий (9.3.18) и (9.3.19). Для этого составляем табл.9.11.

Таблица 9.11

Расчётная таблица

$\dot{x}_1$	$\dot{x}_2$	y_i	$y_i - \bar{y}^*$	$(y_i - \bar{y}^*)^2$	$\widetilde{y}_i$	$y_i - \widetilde{y}_i$	$(y_i - \widetilde{y}_i)^2$
–0,8	–4	–15,1	–28	784	–15,2	0,01	0,0001
–0,3	–2	–1	–13,9	193	–0,98	–0,02	0,0004
0,5	1	19,9	7	49	20,3	–0,40	0,16
0,1	–0,5	9,5	–3,4	11,6	9,66	0,16	0,0256
0,2	0,5	16,5	3,6	13	16,6	–0,10	0,01
0,3	5	47,9	35	1225	47	0,90	0,81
$\sum_{i=1}^{6}(y_i - \bar{y}^*)^2 = 2716$					$\sum_{i=1}^{6}(y_i - \widetilde{y}_i)^2 = 1{,}01$		

Оценка математического ожидания выходной переменной y, используемая при расчётах в табл.9.11, найдена по формуле (5.1.1):

$$\bar{y}^* = \frac{1}{6}\sum_{i=1}^{6} y_i = \frac{1}{6}(-15{,}1 - 1 + 19{,}9 + 9{,}5 + 16{,}5 + 47{,}9) = 12{,}9.$$

Необходимые данные для вычисления оценок дисперсий берём из табл.9.11:

$$\widetilde{\sigma}^2 = \frac{\sum_{i=1}^{6}(y_i - \bar{y})^2}{6-1} = \frac{2716}{5} \approx 543,$$

$$\widetilde{\sigma}_1^2 = \frac{\sum_{i=1}^{6}(y_i - \widetilde{y}_i)^2}{6-2-1} = \frac{1{,}01}{3} = 0{,}33, \qquad (9.3.30)$$

Наблюдаемое значение показателя согласованности (9.3.17):

$$\mathsf{F} = \frac{\widetilde{\sigma}^2}{\widetilde{\sigma}_1^2} = \frac{543}{0{,}33} = 1645$$

Для отыскания критического значения показателя согласованности при уровне значимости $\alpha = 0{,}01$ и степенях свободы $f_1 = n - 1 = 5$, $f_2 = n - k - 1 = 3$ используем приложение 5 и получаем $F_{(0{,}01;5;3)} = 28{,}24$.

Поскольку неравенство (9.3.20) выполняется ($F > F_{(0{,}01;5;3)}$), нулевую гипотезу об адекватности функции регрессии вида (9.3.24) экспериментальным данным принимаем.

6. Выполняем селекцию факторов. Для этого находим элементы главной диагонали корреляционной матрицы (9.3.22). Учитывая выражения (9.3.28) и (9.3.30), имеем

$$diagK_{\tilde{B}} = (\tilde{\sigma}^2_{\tilde{b}_0},\ \tilde{\sigma}^2_{\tilde{b}_1},\ \tilde{\sigma}^2_{\tilde{b}_2}) = (0{,}078;\ 1{,}224;\ 0{,}030)\,.$$

Оценки средних квадратических отклонений коэффициентов $\tilde{b}_j$, $j = \overline{0,\,2}$ принимают значения:

$$\tilde{\sigma}_{\tilde{b}_0} = 0{,}279; \qquad \tilde{\sigma}_{\tilde{b}_1} = 1{,}106; \qquad \tilde{\sigma}_{\tilde{b}_2} = 0{,}173\,.$$

Для каждого фактора находим наблюдаемое значение показателя согласованности (9.3.21):

$$t_0 = \frac{|\tilde{b}_0|}{\tilde{\sigma}_{\tilde{b}_0}} = \frac{12{,}9}{0{,}279} = 46{,}2; \quad t_1 = \frac{|\tilde{b}_1|}{\tilde{\sigma}_{\tilde{b}_1}} = \frac{1{,}35}{1{,}106} = 1{,}22; \quad t_2 = \frac{|\tilde{b}_2|}{\tilde{\sigma}_{\tilde{b}_2}} = \frac{6{,}74}{0{,}173} = 38{,}9\,.$$

Для числа степеней свободы $f = n - k - 1 = 3$ и уровня значимости $\alpha = 0{,}01$ критическое значение показателя согласованности $t_{(0{,}01;3)} = 5{,}84$. Следовательно,

$$t_0 > t_{(0{,}01;3)}, \qquad t_1 < t_{(0{,}01;3)}, \qquad t_3 > t_{(0{,}01;3)}.$$

В отношении фактора $\dot{x}_1$ принимаем конкурирующую гипотезу о его незначимости. Тогда в правой части выражения (9.3.24) второе слагаемое приравниваем к нулю. Поскольку в исходной матрице $\dot{X}$ исключается второй столбец, оценки коэффициентов регрессии b_0 и b_2 необходимо пересчитать.

Пересчёт выполняем в том же порядке, который приведён выше (сохраняем ту же нумерацию пунктов).

1. Функция регрессии в данном случае представляет собой линейный алгебраический полином от одной независимой переменной

$$y = b_0 + b_2 x_2. \qquad (9.3.31)$$

2. Составляем табл.9.12 экспериментальных данных с центрированными значениями фактора x_2.

Таблица 9.12

Массив экспериментальных данных с центрированными значениями фактора

$\dot{x}_2$	–4	–2	1	–0,5	0,5	5
y	–15,1	–1	19,9	9,5	16,5	47,9

Коэффициенты регрессии предварительно оцениваем в уравнении

$$y = b_0 + b_2 \dot{x}_2 . \qquad (9.3.32)$$

3. Система нормальных уравнений принимает вид

$$\begin{cases} 6\tilde{b}_0 + \tilde{b}_2 \sum_{i=1}^{6} \dot{x}_{i2} = \sum_{i=1}^{6} y_1 \\ \tilde{b}_0 \sum_{i=1}^{6} \dot{x}_{i2} + \tilde{b}_2 \sum_{i=1}^{6} \dot{x}_{i2}^2 = \sum_{i=1}^{6} y_1 \dot{x}_{i2}. \end{cases} \qquad (9.3.33)$$

4. Матричное уравнение, эквивалентное системе (9.3.33), представляется как

$$\left(\dot{X}^{\mathsf{т}}_{[2;6]} \dot{X}_{[6;2]}\right) \tilde{B}_{<2>} = \dot{X}^{\mathsf{т}}_{[2;6]} Y_{<6>}, \qquad (9.3.34)$$

где

$$\dot{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \dot{x}_{12} & \dot{x}_{22} & \dot{x}_{32} & \dot{x}_{42} & \dot{x}_{52} & \dot{x}_{62} \end{pmatrix}^{\mathsf{т}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -4 & -2 & 1 & -0{,}5 & 0{,}5 & 5 \end{pmatrix}^{\mathsf{т}};$$

$$\tilde{B} = (\tilde{b}_0,\ \tilde{b}_2)^{\mathsf{т}};$$

$$Y = (y_1,\ y_2,\ y_3,\ y_4,\ y_5,\ y_6)^{\mathsf{т}} = (-15{,}1;\ -1;\ 19{,}9;\ 9{,}5;\ 16{,}5;\ 47{,}9)^{\mathsf{т}}.$$

Оценки коэффициентов регрессии определяются равенством

$$\tilde{B}_{<2>} = \left(\dot{X}^{\mathsf{т}}_{[2;6]} \dot{X}_{[6;2]}\right)^{-1} \dot{X}^{\mathsf{т}}_{[2;6]} Y_{<6>}. \qquad (9.3.35)$$

Вычисляем матрицу $\dot{X}^{\mathsf{т}} \dot{X}$:

$$\dot{X}^{\mathsf{т}} \dot{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -4 & -2 & 1 & -0{,}5 & 0{,}5 & 5 \end{pmatrix} \begin{pmatrix} 1 & -4 \\ 1 & -2 \\ 1 & 1 \\ 1 & -0{,}5 \\ 1 & 0{,}5 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 46{,}5 \end{pmatrix}.$$

Находим обратную матрицу:

$$\left(\dot{X}^{\mathsf{т}} \dot{X}\right)^{-1} = \frac{1}{279} \begin{pmatrix} 45{,}5 & 0 \\ 0 & 6 \end{pmatrix}..$$

Матрица в правой части уравнения (9.3.33) есть не что иное, как вектор-столбец с двумя компонентами:

$$\dot{X}^{\mathsf{T}}Y=\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1\\ -4 & -2 & 1 & -0{,}5 & 0{,}5 & 5\end{pmatrix}\begin{pmatrix}-15{,}1\\ -1\\ 19{,}9\\ 9{,}5\\ 16{,}5\\ 47{,}5\end{pmatrix}=\begin{pmatrix}77{,}7\\ 321\end{pmatrix}.$$

Оценку вектора коэффициентов регрессии находим по формуле (9.3.35):

$$\tilde{B}=\frac{1}{279}\begin{pmatrix}46{,}5 & 0\\ 0 & 6\end{pmatrix}\begin{pmatrix}77{,}7\\ 321\end{pmatrix}=\begin{pmatrix}12{,}9\\ 6{,}90\end{pmatrix}.$$

Получим уравнение регрессии с одним фактором

$$\tilde{y}=12{,}9+6{,}9\dot{x}_2. \qquad (9.3.36)$$

5. Проверяем адекватность уравнения (9.3.36) экспериментальным данным.

Для вычисления оценки дисперсии (9.3.19) составляем расчётную табл.9.13. Оценка дисперсии (9.3.18) остаётся прежней.

Таблица 9.13

Расчётная таблица

$\dot{x}_{i2}$	y_i	$\tilde{y}_i$	$y_i-\tilde{y}_i$	$(y_i-\tilde{y}_i)^2$
–4	–15,1	–14,7	–0,4	0,16
–2	–1	–0,9	–0,1	0,01
1	19,9	19,8	0,1	0,01
–0,5	9,5	9,4	0,1	0,01
0,5	16,5	16,3	0,2	0,04
5	47,9	47,4	0,5	0,25
$\sum_{i=1}^{6}(y_i-\tilde{y}_i)^2=0{,}48$				

Получаем оценку остаточной дисперсии:

$$\tilde{\sigma}_1^2=\frac{\sum_{i=1}^{6}(y_i-\tilde{y}_i)^2}{6-1-1}=\frac{0{,}48}{4}=0{,}12.$$

Показатель согласованности (9.3.17) принимает значение

$$\mathsf{F}=\frac{\tilde{\sigma}^2}{\tilde{\sigma}_1^2}=\frac{543}{0{,}12}=4525.$$

Критическое значение данного показателя при $\alpha=0{,}01$, $f_1=n-1=5$, $f_2=n-k-1=4$ составляет $\mathsf{F}_{(0{,}01;5;4)}=15{,}52$. Поскольку имеет место неравенство $\mathsf{F}>\mathsf{F}_{(0{,}01;5;4)}$, нулевая гипотеза об адекватности функции регрессии (9.3.31) экспериментальным данным принимается.

6. Выполняем селекцию факторов. Главная диагональ корреляционной матрицы (9.3.22) с учётом выражений (9.3.28) и (9.3.30) принимает вид

$$diagK_{\tilde{B}} = (\tilde{\sigma}^2_{\tilde{b}_0},\ \tilde{\sigma}^2_{\tilde{b}_2}) = (0{,}020;\ 0{,}003)\,.$$

Из полученного результата следует, что

$$\tilde{\sigma}_{\tilde{b}_0} = 0{,}14; \qquad \tilde{\sigma}_{\tilde{b}_2} = 0{,}05\,.$$

Наблюдаемые значения показателя согласованности (9.3.21) для факторов $\dot{x}_0$ и $\dot{x}_2$:

$$t_0 = \frac{|\tilde{b}_0|}{\tilde{\sigma}_{\tilde{b}_0}} = \frac{12{,}9}{0{,}14} = 92{,}1; \qquad t_2 = \frac{|\tilde{b}_2|}{\tilde{\sigma}_{\tilde{b}_2}} = \frac{6{,}9}{0{,}05} = 138\,.$$

Находим критическое значение данного показателя в приложении 6, оно составляет $t_{(0,01;4)} = 4{,}6$. Таким образом,

$$t_0 > t_{(0,01;\,4)}, \qquad t_2 > t_{(0,01;4)}.$$

Принимаем нулевую гипотезу о значимости факторов $\dot{x}_0$ и $\dot{x}_2$ в уравнении (9.3.36).

Переходим к уравнению вида (9.3.31) с нецентрированными факторами:

$$\tilde{y} = \tilde{b}_0 + \tilde{b}_2(x_2 - \bar{x})\,,$$

т.е.

$$\tilde{y} = 12{,}9 + 6{,}9(x_2 - 1) = 12{,}9 + 6{,}9x_2 - 6{,}9 = 6 + 6{,}9x_2$$

или, окончательно

$$\tilde{y} = 6 + 6{,}9x_2\,.$$

▲