

8. СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

8.1. Статистический анализ и обработка данных

В предыдущих разделах были рассмотрены методы определения характеристик, описывающих свойства случайных объектов (величин, векторов, функций). Однако цель обработки экспериментальных данных в конечном счёте состоит в выявлении причинно-следственных связей, определяющих состояние и развитие изучаемого явления. Установление этих связей позволяет не только глубоко анализировать различные процессы, но и определять оптимальные пути управления ими.

Решение указанной задачи осуществляется с помощью ряда методов, объединяемых единым названием – **методы статистического анализа** экспериментальных данных. В число этих методов входят методы дисперсионного, корреляционного, регрессионного, компонентного и факторного анализов, метод наименьших квадратов.

Все эти методы целесообразно разделить на две группы:

- методы статического статистического анализа (ССА), в которых фактор времени в явном виде не учитывается;
- методы динамического статистического анализа (ДСА), в которых экспериментальные данные представляются в форме динамических или временных рядов.

Из всего многообразия методов статистического анализа ниже будут изложены широко распространённые методы – наименьших квадратов и регрессионного анализа. Они рассматриваются как методы ССА.

В методах ССА признак, характеризующий причины, принято называть **факторным признаком** или для краткости – **фактором**. Признак, характеризующий следствия, принято называть **результативным признаком** или для краткости – **результатом** (**результатом наблюдений**).

При получении и обработке данных предполагается, что результат наблюдения y зависит от одного или нескольких факторов x_1, x_2, \dots, x_m , и фиксируется по отношению к данным факторам. В процессе обработки решается ряд вопросов.

1. Справедливо ли предположение о зависимости результата y от факторов x_1, x_2, \dots, x_m ?
2. Как оценить степень этой зависимости?

3. Как выделить среди факторов наиболее существенные?
4. Нельзя ли сократить число факторов, используемых при анализе?

5. Какой вид имеет причинно-следственная зависимость между факторами и результатом?

Прежде чем приступить к рассмотрению данных вопросов, остановимся на процедуре формального представления причинно-следственных связей между результатом y и факторами x_1, x_2, \dots, x_m . Указанная процедура сводится к определению зависимости

$$y = f(x_1, x_2, \dots, x_m).$$

Подход к решению данной задачи различен в зависимости от свойств факторов $X_{<m>}$, функции f и, наконец, свойств результатов наблюдений. По этой причине постановка задачи анализа и методы её решения могут быть существенно различными. Для описания данной задачи будем обозначать символом f функциональную (детерминированную), а символом \hat{f} – стохастическую зависимость между $X_{<m>}$ и y .

Напомним, что детерминированная – это зависимость величины y или её некоторой характеристики, например математического ожидания, от факторов $X_{<m>}$. Зависимость закона распределения результата y от факторов $X_{<m>}$ является стохастической.

Тогда могут иметь место следующие виды зависимостей результата y от факторов $X_{<m>}$.

1. Функциональная зависимость от неслучайных факторов

$$\hat{y} = f(x_1, x_2, \dots, x_m) + \hat{\varepsilon}, \quad (8.1.1)$$

при которой случайный характер результата y обуславливается только ошибками $\hat{\varepsilon}$ при наблюдении данного результата.

2. Стохастическая зависимость от неслучайных факторов

$$\hat{y} = \hat{f}(x_1, x_2, \dots, x_m), \quad (8.1.2)$$

при которой случайный характер результата y обуславливается стохастическим характером зависимости \hat{f} .

3. Функциональная зависимость от случайных факторов

$$\hat{y} = f(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m), \quad (8.1.3)$$

при которой случайный характер результата y обуславливается случайным характером факторов.

4. Полная стохастическая зависимость

$$\hat{y} = \hat{f}(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m). \quad (8.1.4)$$

Решение задачи анализа для зависимостей типа (8.1.1) и (8.1.3) опирается на метод наименьших квадратов, а для (8.1.2) и (8.1.4) – на методы регрессионного и корреляционного анализов.

8.2. Сущность метода наименьших квадратов

Метод наименьших квадратов (МНК) получил широкое распространение при обработке экспериментальных данных в целях исследования различных функциональных зависимостей, определения параметров распределений и т.д.

Существует широкий класс задач, в которых МНК является оптимальным методом обработки данных. В других классах задач использование МНК часто оправдывается алгоритмической простотой его реализации ценой небольших потерь в оптимальности получаемого результата. Для нелинейных задач статистического анализа данных зачастую невозможно использование каких-либо других методов, кроме МНК.

Эти и другие причины объясняют широкое распространение МНК при статистическом анализе экспериментальных данных, в частности, при выявлении функциональных зависимостей. Исторически МНК возник значительно раньше других методов обработки данных. Вероятностное обоснование МНК дано К. Гауссом в начале XIX в. и А.А. Марковым в начале XX в.

Предположим, что требуется определить компоненты вектора

$$A_{<k>} = (a_1, a_2, \dots, a_k)^T,$$

который в общем случае не поддается непосредственному наблюдению. Однако можно наблюдать вектор

$$Y_{<n>} = (y_1, y_2, \dots, y_n)^T,$$

функционально связанный с искомым вектором $A_{<k>}$:

$$Y_{<n>} = F_{<n>}(t; A_{<k>}). \quad (8.2.1)$$

При этом соотношение размерностей векторов A и Y может быть произвольным. В частном случае A может быть скалярной величиной, а Y – вектором, и наоборот.

В общем случае вектор-функция F является нелинейной. Схема оценивания, в которой по наблюдениям в некоторые моменты времени t_i , $i = \overline{1, N}$ одного набора параметров (в данном случае компонентов вектора Y) необходимо оценить компоненты другого набора параметров (компоненты вектора A), функционально связанного с первым, называется схемой косвенных наблюдений.

Процесс наблюдения всегда сопровождается ошибками. Наблюдаемое значение функции (8.2.1) в момент времени t_i отклоняется от теоретического вследствие случайных факторов. Следовательно, результат наблюдения всегда представляет собой реализацию случайной величины. В общем случае ошибка наблюдения нелинейным образом связана с наблюдаемой функцией.

На практике часто удаётся путём линеаризации уравнений модели (8.2.1) относительно случайных ошибок свести уравнения к форме, когда случайные ошибки входят аддитивно или мультипликативно (см. §1.1). Однако наиболее простым и самым распространённым типом связи ошибок наблюдения и наблюдаемых величин является линейная аддитивная связь, когда модель наблюдения может быть представлена уравнениями

$$\hat{Y}_{<n>i} = F_{<n>i}(t_i; A_{<k>}) + \hat{E}_{<n>i}, \quad i = \overline{1, N}, \quad (8.2.2)$$

где \hat{E}_i – вектор аддитивной ошибки в i -й момент наблюдения.

В дальнейшем будем рассматривать эту схему наблюдения. Уравнения типа (8.2.2) называются **уравнениями наблюдения**. Поскольку компоненты вектора \hat{E}_i являются случайными неизвестными наблюдателю величинами, то для поиска оценок вектора $A_{<k>}$ используется уравнение вида

$$\hat{Y}_{<n>i} = F_{<n>i}(t_i; A_{<k>}), \quad (8.2.3)$$

которое может оказаться и несовместным, поскольку отражает наблюдаемый процесс приближённо. Поэтому уравнение (8.2.3) принято называть **условным**.

Если моменты времени $t_i, i = \overline{1, N}$ представляют собой известные и в данной задаче фиксированные величины, то фактически вектор-функция F является функцией только вектора A . Поэтому в дальнейшем в число аргументов будем включать моменты времени t_i тогда, когда они либо неизвестны, либо известны с ошибкой. С учётом сказанного уравнение наблюдения запишется в виде

$$\hat{Y}_{<n>} = F_{<n>}(A_{<k>}) + \hat{E}_{<n>}. \quad (8.2.4)$$

В соответствии с методом наименьших квадратов оценки компонентов вектора A отыскиваются на основе минимизации суммы квадратов отклонений между Y и F :

$$\hat{V}(\tilde{A}) = \min_{A \in \mathbf{R}^k} \{(\hat{Y} - F(A))^T (\hat{Y} - F(A))\}, \quad (8.2.5)$$

где \tilde{A} – вектор, представляющий собой решение задачи (8.2.5); \mathbf{R}^k – k -мерное вещественное пространство.

Часто вместо минимизации квадратичной функции (8.2.5) для оценивания вектора A используют минимизацию квадратичной функции более общего вида:

$$\hat{V}(\tilde{A}) = \min_{A \in \mathbf{R}^k} \{(\hat{Y} - F(A))^T Q_{[n]} (\hat{Y} - F(A))\}, \quad (8.2.6)$$

где $Q_{[n]}$ – неотрицательно определённая симметричная матрица, которая называется **весовой**.

Очевидно, что задача (8.2.5) является частным случаем задачи (8.2.6), если в качестве весовой выбрать единичную матрицу.

Показатель качества оценивания (8.2.5) в скалярной форме имеет вид

$$\hat{V}(\tilde{A}) = \min_{A \in \mathbf{R}^k} \left\{ \left(\sum_{i=1}^n (\hat{y}_i - f_i(A))^2 \right) \right\}, \quad (8.2.7)$$

где \hat{y}_i и f_i , $i = \overline{1, n}$ – компоненты вектора Y и вектор-функции F соответственно.

Скалярная форма показателя (8.2.6):

$$\hat{V}(\tilde{A}) = \min_{A \in \mathbf{R}^k} \left\{ \left(\sum_{i,j=1}^n (\hat{y}_i - f_i(A)) q_{ij} (\hat{y}_j - f_j(A)) \right) \right\}, \quad (8.2.8)$$

Для сокращения записей, а также упрощения некоторых выкладок в последующем будем использовать в основном векторно-матричную запись показателя качества оценивания вектора A .

Необходимое условие минимума функции (8.2.5) или (8.2.6) состоит в том, что её частные производные по всем компонентам a_i вектора A должны быть равны нулю:

$$\frac{\partial \hat{V}}{\partial a_i} = 0, \quad i = \overline{1, k}. \quad (8.2.9)$$

Развёрнутый вид условий (8.2.9) в скалярной форме представляется следующей системой уравнений:

$$\left\{ \begin{array}{l} \frac{\partial \hat{V}}{\partial a_1} = \sum_{i=1}^n (\hat{y}_i - f_i(a_1, a_2, \dots, a_k)) \frac{\partial f_i(a_1, a_2, \dots, a_k)}{\partial a_1} = 0 \\ \dots\dots\dots \\ \frac{\partial \hat{V}}{\partial a_j} = \sum_{i=1}^n (\hat{y}_i - f_i(a_1, a_2, \dots, a_k)) \frac{\partial f_i(a_1, a_2, \dots, a_k)}{\partial a_j} = 0 \\ \dots\dots\dots \\ \frac{\partial \hat{V}}{\partial a_k} = \sum_{i=1}^n (\hat{y}_i - f_i(a_1, a_2, \dots, a_k)) \frac{\partial f_i(a_1, a_2, \dots, a_k)}{\partial a_n} = 0 \end{array} \right. \quad (8.2.10)$$

Система уравнений (8.2.10) в теории МНК называется системой **нормальных уравнений**. Её решение, т.е. вектор \tilde{A} будем в дальнейшем называть МНК-оценкой.

Пример 8.1. Требуется оценить скалярную величину a , для которой уравнение наблюдения имеет вид

$$\hat{y}_i = a + \hat{\varepsilon}_i, \quad i = \overline{1, n}.$$

▼ Предположим, что оценку параметра a необходимо отыскивать минимизацией суммы квадратов отклонений

$$\hat{V}(a) = \sum_{i=1}^n (\hat{y}_i - a)^2.$$

Тогда необходимое условие минимума этой суммы квадратов в соответствии с (8.2.9) запишется в виде

$$\frac{\partial \hat{V}}{\partial a} = -2 \sum_{i=1}^n (\hat{y}_i - a) = \sum_{i=1}^n (y_i - a) = 0.$$

Оценка искомой величины

$$\tilde{a} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i.$$

Следует заметить, что при решении данной задачи обоснование выбора функции $V(a)$ отсутствовало, хотя ранее в § 2.4 указывалось, что оптимальный выбор данной функции диктуется условиями задачи.

Замечание приведено в связи с тем, что в дальнейшем будет даваться и иное решение этой же задачи. ▲

В общем случае система нормальных уравнений нелинейная относительно искомых параметров, а искомые параметры – компоненты вектора A – выражаются нелинейным образом через компоненты вектора наблюдения \hat{Y} .

8.3. Метод наименьших квадратов при линейной связи наблюдаемых и оцениваемых параметров

8.3.1. Линейная модель наблюдения

Рассматриваемый ниже метод наименьших квадратов при линейной модели наблюдения получил название схемы Гаусса–Маркова.

Пусть наблюдаемые параметры и параметры искомой функциональной зависимости связаны линейным уравнением, а ошибки наблюдения аддитивны, причём имеют равные нулю математические ожидания:

$$\hat{Y}_{<n>} = X_{[n;k]} A_{<k>} + \hat{E}_{<n>}, \quad (8.3.1)$$

$$M[\hat{E}_{<n>}] = 0, \quad (8.3.2)$$

где $X_{[n;k]}$ – прямоугольная матрица, называемая матрицей наблюдения; $\hat{Y}_{<n>}$ и $\hat{E}_{<n>}$, как и ранее, соответственно случайные векторы наблюдения и ошибок наблюдения.

На основании равенства (8.3.2) можно записать

$$M[\hat{Y}_{<n>}] = X_{[n;k]} A_{<k>}. \quad (8.3.3)$$

В модели Гаусса–Маркова предполагается, что относительно вектора \hat{E} известна некоторая дополнительная информация. Рассмотрим возможные варианты использования данной информации.

Один из наиболее простых случаев тот, когда известно, что наблюдения некоррелированы и равноточны. В этом варианте корреляционная матрица вектора \hat{E} или, что то же самое, вектора \hat{Y} выражается формулой

$$K_{\hat{E}[n]} = \sigma^2 E_{[n]}, \quad (8.3.4)$$

где σ^2 – дисперсия наблюдения; $E_{[n]}$ – единичная матрица.

Дисперсия наблюдения может быть и неизвестной, тогда она подлежит оценке наряду с компонентами вектора A . При неизвестной дисперсии σ^2 не представляется возможным получить какие-либо характеристики точности оценивания.

Более общим является случай, когда наблюдения коррелированы и равноточны, однако для них известна только нормированная корреляционная матрица $G_{[n]}$, а корреляционная матрица вектора \hat{E} имеет вид

$$K_{\hat{E}[n]} = \sigma^2 G_{[n]}, \quad (8.3.5)$$

причём σ^2 – в общем случае неизвестная дисперсия наблюдения.

Покажем, что модель наблюдения (8.3.3), (8.3.4) легко сводится к модели (8.3.3), (8.3.5). Из линейной алгебры известно, что любая симметричная положительно-определённая матрица (а матрица G является таковой) может быть представлена в виде

$$G_{[n]} = D_{[n]} D_{[n]}^T, \quad (8.3.6)$$

где $D_{[n]}$ – невырожденная матрица.

Произведём замену переменных по формуле

$$\hat{Z} = D^{-1} \hat{Y}, \quad (8.3.7)$$

тогда

$$\hat{Y} = D \hat{Z}.$$

Корреляционная матрица вектора \hat{Z} вычисляется следующим образом:

$$\begin{aligned} K_{\hat{Z}[n]} &= M((\hat{Z} - M_{\hat{Z}})(\hat{Z} - M_{\hat{Z}})^T) = \\ &= M(D^{-1}(\hat{Y} - M_{\hat{Y}})(\hat{Y} - M_{\hat{Y}})^T D^{-1T}) = \\ &= D^{-1} K_{\hat{E}} D^{-1T} = \sigma^2 D^{-1} D D^T D^{-1T} = \sigma^2 E_{[n]}. \end{aligned}$$

Получили выражение, аналогичное выражению (8.3.4).

Условное уравнение (8.3.3) с учётом (8.3.7) приобретает вид

$$M_{\hat{Z}} = D^{-1} X A = \bar{X} A.$$

Таким образом, модель наблюдения (8.3.3), (8.3.5) с помощью невырожденного линейного преобразования сводится к соотношениям

$$\hat{Z} = \bar{X}A + \hat{E}, \quad M[\hat{E}] = 0, \quad K_{\hat{E}} = \sigma^2 E.$$

Это и есть исходная модель наблюдения.

Если наблюдения коррелированы и неравноточны, то корреляционная матрица ошибок наблюдения

$$K_{\hat{E}[n]} = R_{[n]}, \quad (8.3.8)$$

где $R_{[n]}$ – известная симметричная положительно-определённая матрица, которая, как и матрица G в формуле (8.3.5), может быть представлена в виде произведения двух невырожденных квадратных матриц аналогично равенству (8.3.6). Это означает, что преобразованием, аналогичным преобразованию (8.3.7), модель наблюдения (8.3.3), (8.3.8) сводится к исходной модели.

По этим причинам в данном подразделе детально рассматривается только наиболее простая модель наблюдения (8.3.3), (8.3.4), а в конце его с помощью преобразования типа (8.3.6) получаются аналогичные результаты для схем наблюдения с корреляционными матрицами ошибок (8.3.5) и (8.3.8).

8.3.2. Нормальные уравнения и оценки наименьших квадратов

Для линейной модели наблюдения (8.3.1) квадратичная функция (8.2.6), при минимизации которой отыскиваются оценки компонентов вектора A , будет иметь вид

$$\hat{V} = (\hat{Y} - XA)^T Q(\hat{Y} - XA), \quad (8.3.9)$$

а нормальные уравнения (8.2.9) – вид

$$X^T Q X A - X^T Q Y = 0. \quad (8.3.10)$$

Можно показать, что система нормальных уравнений (8.3.10) всегда совместна [10]. Будем считать, что матрица X имеет ранг k (предполагается, что $n \geq k$), а матрица весов Q – невырожденная. Тогда матрица $X^T Q X$ будет невырожденной, а потому из равенства (8.3.10) можно получить выражение для оценки вектора A :

$$\tilde{A} = (X^T Q X)^{-1} X^T Q \hat{Y}. \quad (8.3.11)$$

Если весовая матрица единичная ($Q = E$), то вместо соотношения (8.3.11) получим равенство

$$\tilde{A} = (X^T X)^{-1} X^T \hat{Y}. \quad (8.3.12)$$

Если относительно вектора ошибок ничего не известно, то ничего нельзя сказать и о свойствах оценок (8.3.11) или (8.3.12). Если же соот-

ношение (8.3.2) выполняется, то оценки (8.3.11), (8.3.12) являются несмещёнными. Действительно

$$\begin{aligned} M_{\tilde{A}} &= M[(X^T Q X)^{-1} X^T Q \hat{Y}] = M[(X^T Q X)^{-1} X^T Q (X A + \hat{E})] = \\ &= M[(X^T Q X)^{-1} (X^T Q X) A] + M[(X^T Q X)^{-1} X^T Q \hat{E}] = \\ &= M_A + (X^T Q X)^{-1} X^T Q M_{\hat{E}} = M_A = A, \end{aligned}$$

поскольку $M_{\hat{E}} = 0$.

Пусть корреляционная матрица вектора ошибок наблюдений имеет вид (8.3.4). Вычислим корреляционную матрицу вектора оценок \tilde{A} . Обозначим

$$X^T Q X = S_{[n]}; \quad (X^T Q X)^{-1} = C_{[n]}.$$

В этих обозначениях формула для МНК-оценки (8.3.11) переписывается в виде

$$\tilde{A} = C X^T Q \hat{Y} = C \hat{Z},$$

где $\hat{Z} = X^T Q \hat{Y}$.

Тогда корреляционная матрица вектора оценок \tilde{A} вычисляется по формуле

$$\begin{aligned} K_{\tilde{A}[k]} &= M[(C X^T Q \hat{Y} - C X^T Q M_{\hat{Y}})(C X^T Q \hat{Y} - C X^T Q M_{\hat{Y}})^T] = \\ &= C X^T Q M[(\hat{Y} - M_{\hat{Y}})(\hat{Y} - M_{\hat{Y}})^T] Q^T X C^T = \\ &= C X^T Q K_{\hat{Y}[n]} Q^T X C^T = \sigma^2 C X^T Q Q^T X C^T. \end{aligned} \quad (8.3.13)$$

В процессе преобразований в выражении (8.3.13) учтено, что

$$(C X^T Q)^T = Q^T X C^T, \quad K_{\hat{Y}[n]} = \sigma^2 E_{[n]}.$$

Если $Q = E$, то

$$K_{\tilde{A}[k]} = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} = \sigma^2 C. \quad (8.3.14)$$

На практике наиболее часто используется вариант МНК-оценивания, при котором в качестве весовой матрицы Q выбирается матрица $K_{\hat{Y}}^{-1}$, поскольку в таком случае МНК-оценка получается эффективной. Доказательство этого факта можно найти, например, в [10]. Для такого варианта корреляционная матрица оценки вычисляется по формуле

$$K_{\tilde{A}} = (X^T K_{\hat{Y}}^{-1} X)^{-1},$$

которая при $K_{\hat{Y}} = \sigma^2 E$ совпадает с выражением (8.3.14).

Когда величина σ^2 известна, формулы (8.3.13), (8.3.14) позволяют отыскать корреляционную матрицу вектора \tilde{A} . Из данных формул следует, что даже при некоррелированных равноточных наблюдениях компоненты вектора оценок оказываются коррелированными. Для их некорре-

лированности необходима ещё ортогональность столбцов матрицы наблюдений X (при $Q = E$) или столбцов матрицы $CX^T Q$ в общем случае.

Если дисперсия наблюдений σ^2 неизвестна, её необходимо оценить наряду с компонентами вектора A , иначе невозможно отыскать корреляционную матрицу вектора \tilde{A} , которая даёт характеристики точности оценивания.

Покажем, каким образом можно получить оценку величины σ^2 . В математической статистике доказывается, что остаточная сумма квадратов

$$\hat{V}(\tilde{A}) = (\hat{Y} - X\tilde{A})^T (\hat{Y} - X\tilde{A}) \quad (8.3.15)$$

(\tilde{A} – МНК-оценка при линейной модели наблюдения) имеет закон распределения $\sigma^2 \chi_{n-k}^2$, если вектор ошибок наблюдения \hat{E} характеризуется корреляционной матрицей $\sigma^2 E$ (n – размерность вектора \hat{Y} , k – число оцениваемых параметров).

В скалярной форме выражение (8.3.15) имеет вид

$$V(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_k) = \sum_{i=1}^n \left(y_i - \sum_{j=1}^k x_{ij} \tilde{a}_j \right)^2.$$

На основании свойств $\sigma^2 \chi_{n-k}^2$ -распределения получается, что

$$M[V(\tilde{A})] = (n - k)\sigma^2,$$

а потому оценку для величины σ^2 можно приближённо вычислять по формуле

$$\tilde{\sigma}^2 = \frac{V(\tilde{A})}{n - k}. \quad (8.3.16)$$

Отметим, что это оценка несмещённая.

Без доказательства укажем, что МНК-оценки не всегда получаются эффективными. Ранее уже отмечалось со ссылкой на работу [10], что свойством эффективности обладают МНК-оценки для моделей наблюдения (8.3.3), (8.3.4) при $Q = E$, (8.3.3), (8.3.5) при $Q = G^{-1}$ и (8.3.3), (8.3.8) при $Q = R^{-1}$. Это следует, в частности, для нормального распределения вектора ошибок \hat{E} из эффективности оценок, получаемых по методу максимального правдоподобия, поскольку МНК и ММП-оценки совпадают при указанном способе выбора весовых матриц. При всех остальных способах выбора матрицы весов МНК-оценки неэффективны. Однако это не означает, что варианты выбора матрицы весов Q , приводящие к неэффективным оценкам, нецелесообразны. Различные соображения могут привести к выбору матрицы весов в другой форме.

Пример 8.2. Известно, что величина a – постоянная, схема оценивания имеет вид

$$\hat{y}_i = a + \hat{\varepsilon}_i, \quad i = \overline{1, n},$$

где \hat{y}_i – наблюдаемая величина, $\hat{\varepsilon}_i$ – ошибка наблюдения. Требуется по результатам наблюдения определить оценку величины a с использованием МНК, а также получить характеристики точности оценивания при известных моментных характеристиках ошибки $\hat{\varepsilon}_i$:

$$M[\hat{\varepsilon}_i] = 0, \quad M[\hat{\varepsilon}_i^2] = \sigma^2, \quad i = \overline{1, n}.$$

Данная задача совпадает с задачей из примера 8.1, однако там ничего не говорилось о статистических свойствах ошибок измерения $\hat{\varepsilon}_i$.

▼ Случайную величину \hat{y} представим в виде вектора

$$\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^\top.$$

Роль вектора оцениваемых параметров играет скалярная величина a , т.е. $k = 1$. Поэтому матрица X , имеющая размерность $n \times k$, представляет собой вектор-столбец, состоящий из единиц:

$$X = (1, 1, \dots, 1)^\top$$

В качестве матрицы весов Q выбираем единичную матрицу $E_{[n]}$. МНК-оценку скалярной величины a вычисляем по формуле (8.3.12). Предварительно найдём матрицы $(X^\top X)^{-1}$ и $X^\top Y$:

$$X^\top X = (1, 1, \dots, 1)(1, 1, \dots, 1)^\top = n;$$

$$(X^\top X)^{-1} = \frac{1}{n};$$

$$X^\top Y = (1, 1, \dots, 1)(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^\top = \sum_{i=1}^n \hat{y}_i.$$

Окончательно получим

$$\tilde{a} = (X^\top X)^{-1} X^\top \hat{Y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i.$$

Заметим, что решение совпадает с полученным в примере 8.1.

Эта оценка является несмещённой и эффективной. Определим точность оценки \tilde{a} . На основании формулы (8.3.14) получим

$$C = (X^\top X)^{-1} = \frac{1}{n},$$

$$K_{\tilde{a}} = \frac{\sigma^2}{n}. \quad (8.3.17)$$

Если бы величина σ^2 не была априори известной, то её оценку можно было бы получить на основании формулы (8.3.16):

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \tilde{a})^2}{n-1}.$$

Затем можно найти дисперсию оценки $K_{\tilde{a}}$ по формуле (8.3.17) с заменой σ^2 на $\tilde{\sigma}^2$.

Данный пример фактически указывает, каким образом следует оценивать математическое ожидание случайной величины и каковы при этом точечные характеристики этих оценок. Напомним, что ранее тот же результат для оценки математического ожидания был получен с помощью предельных теорем.



Пример 8.3. Рассмотрим задачу, аналогичную приведённой в примере 8.2, с тем отличием, что независимые измерения величины a производятся с различной от эксперимента к эксперименту точностью, которая характеризуется дисперсией σ_i^2 , $i = \overline{1, n}$.

▼ Итак, имеем схему неравноточных наблюдений:

$$\hat{y}_i = a + \hat{\varepsilon}_i; \quad \hat{Y}_{<n>} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T; \quad \text{diag} K_{\hat{Y}_{[n]}} = (\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2).$$

Корреляционная матрица $K_{\hat{Y}}$ взята в диагональной форму в силу независимости измерений.

Данную задачу сводим к предыдущей, используя преобразования (8.3.6), (8.3.7). При этом

$$K_{\hat{Y}_{[n]}} = D_{[n]} D_{[n]}^T,$$

где D – диагональная матрица с диагональю

$$\text{diag} D_{[n]} = (\sigma_1, \sigma_2, \dots, \sigma_n).$$

Тогда формула (8.3.7) приводит к новому вектору

$$\hat{Z} = D^{-1} \hat{Y} = \left(\frac{\hat{y}_1}{\sigma_1}, \frac{\hat{y}_2}{\sigma_2}, \dots, \frac{\hat{y}_n}{\sigma_n} \right)^T.$$

Легко убедиться, что корреляционная матрица вектора \hat{Z} получается при этом единичной, а задача сводится к предыдущей. Уравнение (8.3.3) записывается в виде

$$\hat{Z} = D^{-1} X A = \bar{X} A,$$

где $\bar{X} = D^{-1} X = \left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_n} \right)^T.$

Далее находим, как и в примере 8.2:

$$\bar{X}^T \bar{X} = \sum_{i=1}^n \frac{1}{\sigma_i^2}; \quad (\bar{X}^T \bar{X})^{-1} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}; \quad \bar{X}^T \hat{Z} = \bar{X}^T D^{-1} \hat{Y} = \sum_{i=1}^n \frac{\hat{y}_i}{\sigma_i^2};$$

$$\tilde{a} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T \hat{Z} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \sum_{j=1}^n \frac{\hat{y}_j}{\sigma_j^2}.$$

Данная оценка является несмещённой и эффективной, причём

$$K_{\tilde{a}} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}.$$



Сравнивая полученное решение с решением задачи 8.1 или 8.2, видим, что учёт неравноточности измерений приводит к иному решению. Возвращаясь к замечанию, указанному в примере 8.1, можно опять же подчеркнуть, что учёт статистических свойств ошибок измерений приводит к необходимости иного выбора функции $V(a)$, чем это было сделано в примере 8.1 или 8.2.