ИЗВЕСТИЯ
ТОМСКОГО ОРДЕНА ОКТЯБРЬСКОЙ РЕВОЛЮЦИИ И ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ ПОЛИТЕХНИЧЕСКОГО ИНСТИТУТА им. С. С. КИРОВА

Том 221 1976

ВЛИЯНИЕ АППАРАТУРНЫХ ПОГРЕШНОСТЕЙ НА ТОЧНОСТЬ РЕГРЕССИОННЫХ ОЦЕНОК ПРИ МНОГОПАРАМЕТРОВОМ КОНТРОЛЕ

Н. М. ОСКОРБИН, В. С. ПЛОТНИКОВ

(Представлена научно-техническим семинаром кафедры информационно-измерительной техники)

В приборе многопараметрового контроля качества изделий измеряемый параметр h определяется по совокупности информативных переменных $x_1, x_2, \ldots, x_n$.

При разработке такого прибора могут быть выделены предпроектный этап и этап проектирования. На предпроектном этапе находится математическая модель взаимосвязи контролируемого параметра h с информативными переменными $X_i (i = 1, 2, \ldots, n)$ и путем анализа модели решается вопрос о принципиальной возможности построения прибора. Затем проектируется прибор, реализующий найденную модель и удовлетворяющий заданию по точности, надежности и т. д.

В большинстве случаев математическая модель может быть взята линейной

$$\hat{h}_1 = \sum_{i=1}^{n} c_i X_i + c_0, \tag{1}$$

где $\hat{h}_1$ — оценка измеряемого параметра (зависимой переменной);
c_i, c_0 — соответственно весовые коэффициенты и свободный член уравнения.

Построение математической модели (1) можно осуществить с использованием метода наименьших квадратов [1]. При определении коэффициентов выражения (1) метод наименьших квадратов приводит к решению системы уравнений

$$c_i D[X_i] + \sum_{\substack{j=1 \\ j \neq i}}^{n} c_j R_{ij} = R_{ih}, \quad i = 1, 2, \ldots, n, \tag{2}$$

где $D[x_i]$ — дисперсии независимых переменных;
R_{ij} — корреляционные моменты между величинами X_i и X_j;
R_{ih} — корреляционные моменты между независимыми и зависимой переменной.

Свободный член уравнения (1) может быть найден из уравнения

$$c_0 = \bar{h} - \sum_{i=1}^{n} c_i \bar{X}_i, \tag{3}$$

где $\bar{h}$, $\bar{X}_i$ — математические ожидания переменных h и X_i соответственно.

Дисперсия ошибки построенной модели находится из выражения

$$D[\varepsilon_1] = M[h - \hat{h}_1]^2 = D[h] - \sum_{i=1}^{n} c_i R_{ih}, \tag{4}$$

где M — знак математического ожидания;

$D[h]$ — дисперсия зависимостей переменной.

Дисперсии, математические ожидания и корреляционные моменты в формулах (2), (3), (4) находят по экспериментальным данным, измеренным лабораторной аппаратурой на предпроектном этапе, причем как значения интересующего параметра h измерены с ошибкой ε_h, так и каждая из информативных переменных регистрируется с ошибкой $\varepsilon_i (i = 1, 2,..., n)$. Очевидно, что ошибки наблюдения h и X_i снижают точность математической модели. Прибор производит оценку h по совокупности информативных переменных X_i, причем каждая из них также регистрируется с некоторой ошибкой ε_i^*.

Основной целью настоящей работы является определение степени влияния ошибок измерения величин X_i и h на точность контроля. Вначале исследуется влияние аппаратурных ошибок на точность регрессионного уравнения и находится истинная модель, т. е. модель, в которой исключено влияние ошибок измерения. Затем определяются условия по точности каналов прибора, а также исследуются пути снижения влияния погрешностей каналов на точность контроля. При анализе влияния аппаратурных погрешностей будем находить приращение дисперсии ошибки регрессионного уравнения, а все выводы делать при следующих допущениях:

1) размер выборки бесконечно большой;

2) математические ожидания и дисперсии ошибок измерения не изменяются во времени;

3) случайные ошибки линейно независимы как от величин h и X_i, так и между собой.

Влияние ошибок измерения при формировании выборки на точность регрессионных оценок

Измеренные значения величин X_i и h можно представить в следующем виде:

$$Y_i = X_i + \varepsilon_i, \quad i = 1, 2, ..., n;$$
$$z = h + \varepsilon_h. \tag{5}$$

Ошибки измерения соответствующих переменных имеют как систематические, так и случайные составляющие, причем

$$M[\varepsilon_i] = \bar{\varepsilon}_i, \quad M[\varepsilon_i - \bar{\varepsilon}_i]^2 = D[\varepsilon_i], \quad i = 1, 2, ..., n;$$
$$M[\varepsilon_h] = \bar{\varepsilon}_h, \quad M[\varepsilon_h - \bar{\varepsilon}_h]^2 = D[\varepsilon_h].$$

Для измеренных значений переменных можно составить регрессионное уравнение:

$$\hat{h}_2 = \sum_{i=1}^{n} b_i Y_i + b_0. \tag{6}$$

Ввиду того, что $D[Y_i] = D[X_i] + D[\varepsilon_i]$, $i = 1, 2, \ldots, n$,
а корреляционные моменты соответствующих величин не зависят от ошибок измерения [2], например,

$$M[(z - \bar{Z})(Y_i - \bar{Y}_i)] = M[(h - \bar{h})(X_i - \bar{X}_i)] = R_{ih},$$

то система уравнений для определения коэффициентов b_i $(i = 1, 2, ..., n)$ имеет вид

$$b_i \{D[X_i] + D[\varepsilon_i]\} + \sum_{\substack{j=1 \\ j \neq i}}^{n} b_j R_{ij} = R_{ih}, \quad i = 1, 2, ..., n, \qquad (7)$$

а

$$b_0 = \bar{z} - \sum_{i=1}^{n} b_i \bar{Y}_i.$$

Анализ приведенных выражений показывает, что систематические ошибки измерения могут вызывать лишь смещение оценки h_2 и не влияют на ее разброс. Смещение оценки h_2 обусловлено лишь систематической составляющей ошибки измерения зависимой переменной и может быть найдено из выражения

$$M[\hat{h}_2 - \bar{h}] = M\left[\sum_{i=1}^{n} b_i Y_i + \bar{z} - \sum_{i=1}^{n} b_i \bar{Y}_i - \bar{h}\right] = \bar{\varepsilon}_h.$$

Случайная составляющая ошибки измерения зависимой переменной ε_h, как это следует из системы уравнений (7), не вызывает изменения весовых коэффициентов, а значит и не оказывает влияния на точность оценки h. Обычно остаточная дисперсия уравнения регрессии (6) определяется из выражения

$$D[\varepsilon_2] = D[z] - \sum_{i=1}^{n} b_i R_{ih} = D[h] + D[\varepsilon_h] - \sum_{i=1}^{n} b_i R_{ih} \qquad (8),$$

и используется в регрессионном анализе для статистических оценок уравнения, например для определения доверительных интервалов найденных коэффициентов. Однако при использовании этого уравнения для оценки параметра h, остаточная дисперсия будет завышенной на величину $D[\varepsilon_h]$, так как оцениваться будет истинное значение переменной h, а не ее измеренное значение z.

Случайные составляющие ошибок регистрации независимых переменных $X_i (i = 1, 2, \ldots, n)$ оказывают влияние как на величину весовых коэффициентов, так и на остаточную дисперсию. Приближенные выражения для оценки приращений весовых коэффициентов и остаточной дисперсии обусловленных этими ошибками приведены в [2]. Точную формулу для расчета приращения остаточной дисперсии, обусловленного ошибками измерения всех переменных, можно найти следующим образом:

$$D[\varepsilon_2/\varepsilon_i, \varepsilon_h] = M[\hat{h}_2 - \hat{h}_1 - \bar{\varepsilon}_h]^2 = M\left[\sum_{i=1}^{n} b_i (Y_i - \bar{Y}_i) - \sum_{i=1}^{n} c_i (X_i - \bar{X}_i)\right]^2.$$

Несложные преобразования этого выражения, проведенные с учетом (2), (5), (7), приводят к окончательной формуле

$$D[\varepsilon_2/\varepsilon_i, \varepsilon_h] = \sum_{i=1}^{n} b_i c_i D[\varepsilon_i]. \qquad (9)$$

Из выражения (9) непосредственно следует ранее сделанный вывод о невлиянии ошибок измерения зависимой переменной на точность регрессионных оценок.

Истинную модель (1) можно получить из модели (6), если будут известны математические ожидания и дисперсии ошибок измерения величин h и $X_i (i = 1, 2, \ldots, n)$. При этом истинные весовые коэффициенты находятся из системы уравнений:

$$c_i \{D[Y_i] - D[\varepsilon_i]\} + \sum_{\substack{j=1 \\ j \neq i}}^{n} c_j R_{ij} = R_{ih}, \quad i = 1, 2, \ldots, n, \tag{10}$$

$$c_0 = \bar{z} - \bar{\varepsilon}_h - \sum_{i=1}^{n} c_i (\bar{Y}_i - \bar{\varepsilon}_i).$$

Остаточная дисперсия истинного уравнения определяется выражением

$$D[\varepsilon_1] = D]z] - D[\varepsilon_h] - \sum_{i=1}^{n} c_i R_{ih}. \tag{11}$$

Если математические ожидания и дисперсии аппаратурных погрешностей остаются постоянными как при определении коэффициентов уравнения регрессии, так и при оценке измеряемой величины h, то можно сделать следующие выводы:

1) случайная составляющая ошибки измерения зависимой переменной и систематические составляющие ошибок измерения информативных переменных не оказывают влияния на точность оценки;

2) систематическая составляющая ошибки измерения зависимой переменной может быть учтена введением соответствующей поправки;

3) влияние случайных составляющих ошибок измерения информативных переменных следует учитывать при проектировании многопараметрового прибора.

Задание точности каналов при проектировании прибора многопараметрового контроля

Проектирование прибора многопараметрового контроля можно начинать, если имеется запас по точности истинного уравнения регрессии, т. е. выполняется условие

$$\Delta D[\varepsilon] = D[\varepsilon_n] - D[\varepsilon_1] > 0,$$

где $D[\varepsilon_n]$ — задание на точность проектируемого прибора.
Возможные погрешности каналов проектируемого прибора можно характеризовать величинами $D[\varepsilon^*_i]$ $(i = 1, 2, \ldots, n)$. Тогда величина остаточной дисперсии $D[\varepsilon^*_i]$ может быть найдена по выражению (11) путем замены весовых коэффициентов c_i $(i = 1, 2, \ldots, n)$ на весовые коэффициенты b^*_i, полученные из системы уравнений

$$b^*_i \{D[X_i] + D[\varepsilon^*_i]\} + \sum_{\substack{j=1 \\ j \neq i}}^{n} b^*_j R_{ij} = R_{ih}, \quad i = 1, 2, \ldots, n,$$

а оставшийся запас по точности должен удовлетворять условию

$$\Delta D[\varepsilon] = D[\varepsilon_n] - D[\varepsilon^*] > D_0, \tag{12}$$

где D_0 — суммарная дисперсия погрешностей решающего устройства, выходного индикатора и других узлов, погрешности которых не учитываются в данной методике.

Одним из возможных путей задания допустимой погрешности каналов может быть путь последовательных уточнений, при котором вначале задаются дисперсии каналов $D[\varepsilon^*_i]$, исходя из возможностей выбранных вариантов схем и веса ошибок отдельных каналов в ошибках прибора, а затем проверяется выполнимость условия (12). Если условие (12) не выполняется, то возможны следующие пути повышения точности прибора:

1) нахождение нового состава информативных переменных;
2) снижение величины D_0;
3) увеличение точности каналов;
4) введение новых каналов коррекции ошибок измерения информативных переменных.

Выбор пути повышения точности прибора определяется конкретными условиями проектирования.

Особый интерес представляют каналы коррекции ошибок измерения информативных переменных, введение которых отличается от нахождения нового состава информативных переменных тем, что в истинном уравнении регрессии от них не уменьшается величина остаточной дисперсии. Переменными, служащими для коррекции, могут быть сигналы пропорциональные как факторам, вызывающим ошибки какого-либо канала, так и некоторым параметрам контролируемого изделия, а отличительным свойством этих переменных является равенство нулю их весовых коэффициентов в истинном уравнении регрессии. Способность корректирующих переменных уменьшать влияние ошибок каналов объясняется свойствами уравнения регрессии. Уменьшение остаточной дисперсии прибора за счет введения канала коррекции ошибок измерения информативных переменных может быть найдено как разность дисперсий, обусловленных ошибками каналов без коррекции и с коррекцией:

$$\Delta D[\varepsilon] = \sum_{i=1}^{n} c_i b_i^* D[\varepsilon_i^*] - \sum_{i=1}^{n+} c_i (b_i^* + \Delta b_i^*) D[\varepsilon_i^*] = \sum_{i=1}^{n} c_i \Delta b_i^* D[\varepsilon_i^*],$$

где Δb_i^* $(i = 1, 2, \ldots, n, n+1)$ — приращения весовых коэффициентов при введении канала коррекции.

В практических случаях можно найти некоторые независимые переменные, которые увеличивают как запас по точности истинного уравнения регрессии, так и обладают корректирующими свойствами.

ЛИТЕРАТУРА

1. Ю. В. Линник. Метод наименьших квадратов и основы теории обработки наблюдений. М., Физматгиз, 1962.

2. В. П. Бородюк, Э. К. Лецкий. Статистическое описание промышленных объектов. М., «Энергия», 1971.