**1.2 Линейная регрессия**

Статическая характеристика ОУ представляет собой функцию, отображаемую графически или описываемую аналитически. Получить ее можно, как правило, путем проведения эксперимента. Для этого на вход объекта последовательно подают входные ступенчатые воздействия *x*1, *x*2, *x*3… через равные промежутки времени Δ*t*, и после каждого из них при окончании переходного процесса фиксируют соответствующие им установившиеся значения *y*1, *y*2, *y*3… переходной характеристики ОУ (рис. 1.2.1, а).

Так как зачастую объект подвержен действию различного рода помех вследствие изменения характеристик оборудования, условий окружающей среды, неточностью измерений параметров *x* и *y* датчиками и т. д., то результатом съема информации является массив точек, который аппроксимируется функцией *y*(*x*) (рис. 1.2.1, б).

Для обработки данных такого эксперимента наибольшее распространение получил метод наименьших квадратов, являющийся основой корреляционного и регрессионного анализов. При этом используется частный вид уравнений для функции *y*(*x*) вида [16]:

 (1.2.1)

где *x*1, *x*2,…, *xm* – факторы эксперимента;

*m* – число факторов;

*f*(*x*1), *f*(*x*2),…, *f*(*xm*) – различные функциональные преобразования над аргументом *x*;

*b*1, *b*2,…, *bm* – коэффициенты линейной регрессии;

*y*р – расчетные значения зависимой переменной *y*.

Метод наименьших квадратов позволяет вычислить коэффициенты *b*1, *b*2,…, *bm*, при которых получается минимальная сумма квадратов отклонений расчетных значений *i*-го эксперимента *y*р*i* от фактических *yi*:

 (1.2.2)

*n* – число экспериментов.

*x*1

Δ*t*

*t*

*x*3

*x*2

*y*

*y*1

*y*2

*y*3

*t*

*x*

*y*

*yn*

*y*1

*y*2

*y*3

*yn*

*…*

*…*

*x*1

*x*2

*x*3

*xn*

*xn*

*x*

а) б)

Рис. 1.2.1. Принцип получения статической характеристики ОУ: а) подача плановых воздействий; б) – аппроксимация статической характеристики ОУ по экспериментальным точкам

Подставляя *y*р из уравнения (1.2.1) в выражение (1.2.2), получим:

 (1.2.3)

Если вычислить частные производные по параметрам *b*1, *b*2,…, *bm* и затем приравнять их к нулю, то можно найти условия, при которых функция (1.2.2) будет минимальна. Получается система из *m*+1 уравнений и *m*+1 неизвестных, которая решается относительно *b*1, *b*2,…, *bm*:

 (1.2.4)

Вектор зависимой переменной *y* имеет размерность *n*×1, вектор коэффициентов *b* –(*m+*1)×1, а матрица факторов *x* – *n*×*m*:

 (1.2.5)

Помимо коэффициентов линейной регрессии в статистической обработке информации используется термин «доверительный интервал». Он представляет собой предельные значения величины, находящейся в данном интервале с заданной вероятностью, которую для инженерных расчетов обычно выбирают равной 0,95 [16].

Для проверки наличия связи между зависимой переменной и выбранными факторами, используют понятие коэффициента детерминации *R*2. Это так называемая доля дисперсии (мера разброса случайной величины относительно ее среднего значения) зависимой переменной, которую можно объяснить полученной зависимостью. Данный коэффициент находится из выражения:

 (1.2.6)

Значение *R*2 обычно получается в пределах от 0 до 1. При *R*2 > 0,8 регрессионная модель достаточно точна. Если *R*2 = 1, то зависимость является не статистической, а функциональной. В случае *R*2 = 0 считается, что связи между переменными *x* и *y* либо нет, либо необходимо ввести в рассмотрение дополнительные факторы *x*.

Отрицательный коэффициент *R*2 показывает, что модель крайне неадекватна.