ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Кафедра «Управление качеством»

«Метод наименьших квадратов»

Методические указания

к проведению практических занятий
по дисциплине
«Планирование и организация эксперимента»

Автор:
Катаев В.С.

Ростов-на-Дону, 2013

Аннотация

Методические указания предназначены для проведения практических работ по дисциплине «Планирование и организация эксперимента с магистрами и студентами специальностей 221400, 200501, 200503, 220501 и других инженерных специальностей.

Цель работы – ознакомление студентов с методикой построения уравнений приближенной регрессии при изучении корреляционной зависимости с помощью метода наименьших квадратов.

Автор

Старший преподаватель кафедры «Управление качеством» Катаев Виктор Сергеевич

Оглавление

1. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Для характеристики формы связи при изучении корреляционной зависимости между выходным параметром и переменным фактором, при обработке результатов однофакторных экспериментов, используются уравнения приближенной регрессии [1]. Задача ставится следующим образом: по данной выборке объёма *n* найти уравнение приближенной регрессии и оценить допускаемую при этом ошибку. В качестве метода приближения обычно выбирают метод наименьших квадратов (МНК).

Суть метода заключается в том, что вид зависимости и значения коэффициентов описывающего ее уравнения должны обеспечивать минимальную сумму квадратов отклонений (Ф) ординат экспериментальных точек от ординат этой зависимости [2]:

$$\Phi = \sum_{i=1}^{n} (y_i - \tilde{y})^2 = min, \qquad (1.1)$$

где y_i – рассчитанное по уравнению регрессии значение выходного параметра, а $\tilde{y}$ – экспериментальное значение выходного параметра, полученное при том же значении переменного фактора x_i.

Задача определения коэффициентов уравнения регрессии по МНК сводится к определению минимума функции многих переменных [1]. Если:

$$y = f(x, b_0, b_1, \dots, b_R), \qquad (1.2)$$

и требуется выбрать коэффициенты $b_0, b_1, \dots, b_R$ таким образом, чтобы:

$$\Phi = \sum_{i=1}^{n} [\tilde{y}_i - f(x_i, b_0, b_1, \dots, b_R]^2 = \min \qquad (1.3)$$

то необходимым условием минимума $\Phi(b_0, b_1, \dots, b_R)$ будет являться выполнение равенств:

$$\frac{\partial \Phi}{\partial b_0} = 0, \qquad \frac{\partial \Phi}{\partial b_1} = 0, \dots, \frac{\partial \Phi}{\partial b_R} = 0 \qquad (1.4)$$

Т.е. минимум данной функции будет в точке, где её частные производные равны нулю.

Условие (1.4) можно записать в виде:

$$\begin{gathered}\sum_{i=1}^{n} 2[\tilde{y}_i - f(x_i, b_0, b_1, \dots, b_R)] \frac{\partial f(x_i)}{\partial b_0} = 0,\\ \sum_{i=1}^{n} 2[\tilde{y}_i - f(x_i, b_0, b_1, \dots, b_R)] \frac{\partial f(x_i)}{\partial b_1} = 0,\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots,\\ \sum_{i=1}^{n} 2[\tilde{y}_i - f(x_i, b_0, b_1, \dots, b_R)] \frac{\partial f(x_i)}{\partial b_R} = 0\end{gathered} \qquad (1.5)$$

После преобразования:

$$\begin{gathered}\sum_{i=1}^{n} \tilde{y}_i \frac{\partial f(x_i)}{\partial b_0} - \sum_{i=1}^{n} f(x_i, b_0, b_1, \dots, b_R) \frac{\partial f(x_i)}{\partial b_0} = 0,\\ \sum_{i=1}^{n} \tilde{y}_i \frac{\partial f(x_i)}{\partial b_1} - \sum_{i=1}^{n} f(x_i, b_0, b_1, \dots, b_R) \frac{\partial f(x_i)}{\partial b_1} = 0,\\ \sum_{i=1}^{n} \tilde{y}_i \frac{\partial f(x_i)}{\partial b_R} - \sum_{i=1}^{n} f(x_i, b_0, b_1, \dots, b_R) \frac{\partial f(x_i)}{\partial b_R} = 0,\end{gathered} \qquad (1.6)$$

Система уравнений (1.6) имеет столько же уравнений, сколько неизвестных коэффициентов $b_0, b_1, \dots, b_R$ входит в уравнение регрессии, и называется системой нормальных уравнений.

При изучении зависимости выходного параметра от одного переменного фактора необходимо построить эмпирическую линию регрессии для определения вида уравнения регрессии [1]. Для этого весь диапазон изменения *x* на поле корреляции разбивается на k равных интервалов Δx. Все точки, попавшие в данный интервал Δx_j, относят к его середине x_j. Для этого подсчитывают частные средние $\bar{y}_j$ для каждого интервала:

$$\bar{y}_j = \frac{\sum_{i=1}^{n_j} y_{ji}}{n_j}, \qquad (1.7)$$

где y_{ji} – экспериментальные значения выходного параметра, попавшие в интервал Δx_j, а n_j – количество значений y_{ji}.

Затем последовательно соединяют точки $(x_j;\ \bar{y}_j)$ отрезками прямой. Полученная ломаная называется эмпирической линией регрессии y по x. По виду эмпирической линии регрессии можно подобрать уравнение регрессии $y = f(x)$.

Для линейной зависимости $y = ax + b$ условие (1.4) будет иметь вид:

$$\begin{cases} \dfrac{\partial \Phi}{\partial b} = \sum_{i=1}^{n}[y_i - (ax_i + b)] = 0 \\ \dfrac{\partial \Phi}{\partial a} = \sum_{i=1}^{n}[y_i - (ax_i + b)]x_i = 0 \end{cases} \tag{1.8}$$

Для определения коэффициентов a и b линейного уравнения будем иметь систему линейных уравнений (1.9):

$$\begin{cases} bn + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n}(x_i \times y_i) \end{cases} \tag{1.9}$$

Решение системы уравнений (1.9) относительно a и b дает следующие формулы для их расчета:

$$b = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x)^2} \tag{1.10}$$

$$a = \frac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \tag{1.11}$$

Аналогичным образом, с помощью МНК можно получить формулы для расчета коэффициентов нелинейных зависимостей (1.12) – (1.18) [2]:

Логарифмическая зависимость $y = a\ln x + b$, $x_i > 0$, x ≠ 0

$$\begin{cases} a = \dfrac{\sum y_i \ln x_i - \dfrac{1}{n}\sum y_i \sum \ln x_i}{\sum(\ln x_i)^2 - \dfrac{1}{n}\sum(\ln x_i)^2} \\ b = \dfrac{\sum y_i a \sum \ln x_i}{n} \end{cases} \tag{1.12}$$

Экспоненциальная функция $y = be^{ax}$, все y_i и $x_i > 0$, $y_i \neq 0$

$$\begin{cases} a = \dfrac{\sum(\ln y_i)\, x_i - \dfrac{1}{n}\sum x_i \sum \ln y_i}{\sum x_i^2 - \dfrac{1}{n}\sum(x_i)^2} \\ b = exp\left[\dfrac{\sum \ln y_i - a\sum x_i}{n}\right] \end{cases} \tag{1.13}$$

Степенная функция $y = ax^b$, $x_i \neq 0$, $y_i \neq 0$, все y_i и $x_i > 0$

$$\begin{cases} b = \dfrac{\sum \ln x_i \ln y_i - \frac{1}{n}\sum \ln x_i \sum \ln y_i}{\sum (\ln x_i)^2 - \frac{1}{n}\left[\sum \ln x_i\right]^2} \\ a = exp\left[\dfrac{\sum \ln y_i - b\sum \ln x_i}{n}\right] \end{cases} \quad (1.14)$$

Дробно-линейная функция $y = x/(ax + b)$, $y_i \neq 0$, $x_i \neq 0$

$$\begin{cases} a = \dfrac{\sum \frac{x_i^2}{y_i} - \frac{1}{n}\sum x_i \sum \frac{x_i}{y_i}}{\sum x_i^2 - \frac{1}{n}\sum x_i^2} \\ b = \dfrac{\sum \frac{x_i}{y_i} - a\sum x_i}{n} \end{cases} \quad (1.15)$$

Гиперболическая функция $y = a/x + b$, $x_i > 0$

$$\begin{cases} b = \dfrac{\sum x_i^2 y_i - \frac{1}{n}\sum x_i \sum x_i^2}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2} \\ a = \dfrac{\sum x_i^2 - b\sum x_i}{n} \end{cases} \quad (1.16)$$

Дробно-рациональная функция $y = 1/(ax + b)$, $y_i \neq 0$

$$\begin{cases} a = \dfrac{\sum \frac{x_i}{y_i} - \frac{1}{n}\sum x_i \sum \frac{1}{y_i}}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2} \\ b = \dfrac{\sum \frac{1}{y_i} - a\sum x_i}{n} \end{cases} \quad (1.17)$$

Квадратичная (параболическая) функция $y = ax^2 + bx + c$

$$\begin{cases} cn + b\sum x_i + a\sum x_i^2 = \sum y_i \\ c\sum x_i + b\sum x_i^2 + a\sum x_i^3 = \sum y_i x_i \\ c\sum x_i^2 + b\sum x_i^3 + a\sum x_i^4 = \sum x_i^2 y_i \end{cases} \quad (1.18)$$

Точность описания корреляционной связи между параметром выхода и переменным фактором нагляднее всего характеризует средняя погрешность аппроксимации ($\bar{\delta}, \%$), которая рассчитывается по следующей формуле:

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \times 100\%\right\} \quad (1.19)$$

Очевидно, что лучшей зависимостью для описания связи между x и y будет та, которая обеспечивает минимальную среднюю погрешность аппроксимации $\bar{\delta} \to min$.

2. МОДЕЛЬНЫЙ ПРИМЕР

Приведена зависимость напряжённости электростатического поля (у) в процессе гидромониторной промывки ёмкости после нефтепродукта от диэлектрической проницаемости моющей жидкости (х). Результаты эксперимента представлены в таблице 2.1.

Таблица 2.1 – Результаты эксперимента

№ п/п	X	Y	№ п/п	X	Y	№ п/п	X	Y
1	50	32	11	55	54	21	62	72
2	51	35	12	56	50	22	63	71,5
3	52	29	13	57	60	23	63,5	73
4	52,5	31,5	14	58	52	24	64	75
5	53	30	15	59	65	25	64,5	74
6	54	35	16	60	69	26	65	77
7	54,5	38	17	61	68,5	27	66	75

Построим эмпирическую линию регрессии, чтобы определить вид уравнения.

Для этого разобьем рассматриваемый диапазон значений x на n равных интервалов. Для определения n числа интервалов воспользуемся формулой Стерджеса:

$$n = 1 + 3{,}322 lgN = 1 + 3{,}322 lg21 = 5{,}39 \quad (2.1)$$

Округляем полученный результат до 5.

Ширина интервала h будет равна

$$\frac{X_{max} - X_{min}}{h} = \frac{66 - 50}{5} = 3{,}2 \quad (2.2)$$

Таким образом получим следующие границы и частные средние значения интервалов (таблица 2.2).

Таблица 2.2 – Границы и частные средние значения интервалов

Интервал	Границы интервала	Частные средние $\bar{y}_j = \frac{\sum_{i=1}^{n_j} y_{ji}}{n_j}$
1	[50;53,2]	31,5
2	(53,2;56,4]	44,25
3	(56,4;59,6]	59
4	(59,6;62,8]	69,8
5	(62,8;66]	74,25

Соединим полученные точки отрезками прямой (рисунок 2.1).

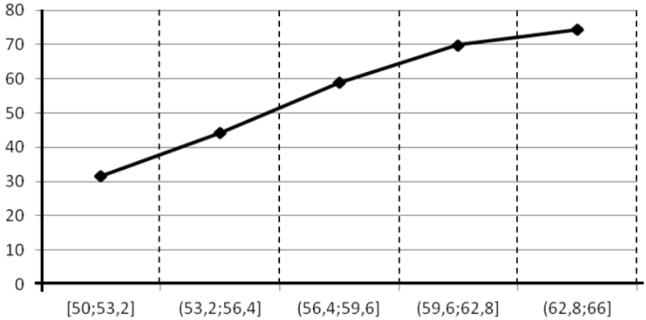

Рисунок 2.1 – Эмпирическая линия регрессии

Исходя из полученного графика, можно предположить, что данная зависимость имеет линейный, либо степенной вид. Соответственно, необходимо найти коэффициенты для двух уравнений регрессии: $y = ax + b$ и $y = ax^b$.

Линейное уравнение $y = ax + b$.

Коэффициенты линейного уравнения определяются по формулам (1.10) и (1.11.)

Промежуточные расчёты представлены в таблице 2.3

Таблица 2.3 – Промежуточные расчёты

Величина	Значение
$\sum X$	1221
$\sum У$	1166,5
$\sum X^2$	71515
$\sum XУ$	69589,25
$(\sum X)^2$	1490841

$$b = \frac{1166{,}5 \times 71515 - 1221 \times 69589{,}25}{21 \times 71515 - 1490841} = -140{,}9$$

$$a = \frac{21 \times 69589{,}25 - 1221 \times 1166{,}5}{21 \times 71515 - 1490841} = 3{,}38$$

Степенное уравнение $y = ax^b$

Коэффициенты данного уравнения будут определяться по формуле (1.14). a=22,296; b=0,3954.

Проверим среднюю погрешность аппроксимации $(\bar{\delta}, \%)$ для каждого из построенных уравнений регрессии. Расчёт проводится по формуле (1.19).

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \times 100\%\right\}$$

$\bar{\delta}_{\text{лин.}} = 9{,}83\%$; $\bar{\delta}_{\text{степен.}} = 126{,}66\%$

Исходя из значений средней погрешности аппроксимации можно сделать вывод, что уравнение регрессии линейного вида точнее описывает эмпирическую зависимость.

3. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Используя данные из таблицы 3.1 построить уравнения регрессии, описывающие зависимость Y от X.

Таблица 3.1 – Индивидуальные задания.

№п/п	X	1	2	3	4	5
		Y	Y	Y	Y	Y
1	50	4,3	150	25	7,8	1,9
2	51	3,5	155	52	6,5	1,5
3	52	4,7	157	40	8,65	2,1
4	52,5	4	158	44	7,25	1,8
5	53	4,4	160	38	8,15	2
6	54	4,5	163	53	8,3	2
7	54,5	5	165	30	9	2,1
8	55	4	166	60	7,4	1,8
9	56	4,5	170	47	8,2	1,9
10	57	4,9	172	51	9,1	2,1
11	58	4	175	40	7,6	1,8
12	59	4,6	178	62	8,6	2
13	60	4,6	181	72	8,8	2
14	61	5	184	75	9,5	2,2
15	62	5,1	187	79	7,8	1,8
16	63	4,5	190	84	8,3	2
17	63,5	5	192	64	9,6	2,3
18	64	4,2	193	81	8	1,9
19	64,5	4,6	194	69	9	2,1
20	65	4,7	196	79	9,1	2,2
21	66	5,1	199	80	10	2,3
22	67,2	4,2	202	82	8,1	2
23	68	4,65	205	91	9,1	2,4
24	69	5,1	208	95	9,9	2,5
25	69,5	4,2	209	93	8,4	2,4
26	70	4,8	211	97	9,4	2,6

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Ахназарова С.Л., Кафаров В.В. Методы оптимизации эксперимента в химической технологии: учебное пособие – М.: Высшая школа, 1985 – 327 с.
2. Квеско Н.Г. Методы и средства исследований: учебное пособие. – Томск: Изд-во Томского политехнического университета, 2010 – 112 с.