4. Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. Л. : Энергоатомиздат. Ленингр. отд-ние, 1985.

5. Кротова Е. И. Идентификация типа распределений результатов экспериментальных исследований // Изв. вузов. Химия и хим. технология. 1998. № 1. С. 57–59.

E. I. Krotova

METHOD OF ESTIMATION OF INFLUENCE OF ADDITIVE DISTURBANCE ON INPUT OF A RECEIVER OF SIGNALS WITH THE HELP OF IDENTIFICATION OF KINDS OF DISTRIBUTION LOWS

In the article a method of estimation of influence of additive disturbance with various laws of distribution on an input of the receiver of telecommunication system according to combined parameter of identification for signals with modulation by the minimal frequency shift is considered.

Keywords: information, law of distribution, identification, disturbance.

УДК 519.233.5

Н. Н. Щелканов

РОБАСТНЫЙ МЕТОД ПОСТРОЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ МЕЖДУ ДВУМЯ ФИЗИЧЕСКИМИ ВЕЛИЧИНАМИ С УЧЕТОМ ИХ СЛУЧАЙНЫХ ПОГРЕШНОСТЕЙ

Представлена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1 X$ для общего случая, когда разброс точек в корреляционной связи величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами. Все известные выражения для коэффициентов регрессии оказались частными случаями полученной формулы.

Ключевые слова: робастный метод, линейная регрессия, случайные погрешности.

При работе с разными массивами данных часто возникает необходимость нахождения коэффициентов линейной регрессии между двумя случайными физическими величинами. В большинстве случаев коэффициенты регрессии имеют конкретный физический смысл и для корректной интерпретации полученных результатов очень важно найти их значения наилучшим образом.

Существует несколько формул для определения коэффициентов регрессии [1–3], но не для всех формул есть общее понимание, в каких случаях их следует использовать. В настоящее время отсутствует единый подход к нахождению коэффициентов линейной регрессии для общего случая, когда разброс точек в корреляционной связи между двумя величинами обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами.

Постановка задачи. Рассмотрим две случайные физические величины X_0 и Y_0, между которыми существует статистическая корреляционная связь. Предположим, что эта связь может быть описана линейной зависимостью

$$Y_0 = K_0 + K_1 X_0. \tag{1}$$

Требуется найти коэффициенты регрессии K_0 и K_1, которые наилучшим образом отражают физическую взаимосвязь между ними.

Так как X_0 и Y_0 измеряются со случайными погрешностями, то на практике мы имеем дело с величинами X и Y, для которых уравнение регрессии запишется в виде

$$Y = K_0 + K_1 X. \tag{2}$$

Запись уравнений (1) и (2) с одинаковыми коэффициентами регрессии показывает, что последние не должны зависеть от случайных погрешностей измеренных величин X и Y. В дальнейшем будем говорить о нахождении только коэффициента регрессии K_1, так как K_0 вычисляется после нахождения K_1 по известной формуле

$$K_0 = \overline{Y} - K_1 \cdot \overline{X}, \tag{3}$$

где $\overline{X}$ и $\overline{Y}$ – средние значения X и Y.

Новый подход к нахождению коэффициента регрессии K_1. Этот подход заключается в следующих моментах:

– случайные величины X и Y нормируются на значения $\sqrt{\delta_X^2 + \delta_{X_0}^2}$ и $\sqrt{\delta_Y^2 + \delta_{Y_0}^2}$ соответственно, где δ_X и δ_Y – случайные среднеквадратические погрешности измерения X и Y для рассматриваемого массива данных; δ_{X_0} и δ_{Y_0} – некоторые величины, характеризующие разброс точек в корреляционной связи физических величин X_0 и Y_0 за счет неконтролируемых физических параметров;

– при нахождении коэффициента регрессии K_1 используется ортогональная среднеквадратическая регрессия, т. е. минимизируется сумма квадратов отклонений, перпендикулярных искомой прямой.

Тогда уравнение линейной регрессии запишется в виде

$$\frac{Y}{\sqrt{\delta_Y^2+\delta_{Y_0}^2}} = K_0{}' + K_1{}' \cdot \frac{X}{\sqrt{\delta_X^2+\delta_{X_0}^2}}. \quad (4)$$

Здесь величины δ_{X_0} и δ_{Y_0} находятся из решения системы двух уравнений:

– первое уравнение имеет вид

$$\left|\rho_{X_0Y_0}\right| \cdot \sigma_{X_0} \cdot \sigma_{Y_0} = \sqrt{\sigma_{X_0}^2-\delta_{X_0}^2} \cdot \sqrt{\sigma_{Y_0}^2-\delta_{Y_0}^2}, \quad (5)$$

где $\sigma_{X_0}=\sqrt{\sigma_X^2-\delta_X^2}$ и $\sigma_{Y_0}=\sqrt{\sigma_Y^2-\delta_Y^2}$ – среднеквадратические отклонения величин X_0 и Y_0; σ_X и σ_Y – среднеквадратические отклонения величин X и Y; $\rho_{X_0Y_0}$ – коэффициент корреляции между X_0 и Y_0. Коэффициент корреляции $\rho_{X_0Y_0}$ находится из известного уравнения [1]:

$$\rho_{XY}\sigma_X\sigma_Y = \rho_{X_0Y_0}\ \sigma_{X_0}\ \sigma_{Y_0}, \quad (6)$$

где ρ_{XY} – коэффициент корреляции между X и Y. Заметим, что из уравнения (6) следует уравнение (5);

– второе уравнение запишем в виде

$$\frac{\delta_{X_0}}{\sigma_{X_0}} = \frac{\delta_{Y_0}}{\sigma_{Y_0}} \quad (7)$$

и назовем условием пропорциональности величин δ_{X_0}, δ_{Y_0} и σ_{X_0}, σ_{Y_0}. Введение величин σ_{X_0}, σ_{Y_0} и запись условия (7) являются ключевыми моментами, так как это позволяет получить обобщенное решение для коэффициентов линейной регрессии уравнения (2).

Результаты. После решения системы уравнений (5) и (7) имеем:

$$\delta_{X_0} = \sigma_X \cdot \sqrt{\left(1-\frac{\delta_X^2}{\sigma_X^2}\right)\cdot\left(1-\frac{\left|\rho_{XY}\right|}{\sqrt{\left(1-\delta_X^2/\sigma_X^2\right)\cdot\left(1-\delta_Y^2/\sigma_Y^2\right)}}\right)}, \quad (8)$$

$$\delta_{Y_0} = \sigma_Y \cdot \sqrt{\left(1-\frac{\delta_Y^2}{\sigma_Y^2}\right)\cdot\left(1-\frac{\left|\rho_{XY}\right|}{\sqrt{\left(1-\delta_X^2/\sigma_X^2\right)\cdot\left(1-\delta_Y^2/\sigma_Y^2\right)}}\right)}. \quad (9)$$

С учетом (8) и (9) найдем значения $\sqrt{\delta_X^2+\delta_{X_0}^2}$ и $\sqrt{\delta_Y^2+\delta_{Y_0}^2}$:

$$\sqrt{\delta_X^2+\delta_{X_0}^2} = \sigma_X \cdot A, \quad (10)$$

$$\sqrt{\delta_Y^2+\delta_{Y_0}^2} = \sigma_Y \cdot B, \quad (11)$$

где

$$A = \sqrt{1-\left|\rho_{X_0Y_0}\right|\cdot\left(1-\frac{\delta_X^2}{\sigma_X^2}\right)} = \sqrt{1-\left|\rho_{XY}\right|\cdot\sqrt{\frac{1-\delta_X^2/\sigma_X^2}{1-\delta_Y^2/\sigma_Y^2}}}; \quad (12)$$

$$B = \sqrt{1-\left|\rho_{X_0Y_0}\right|\cdot\left(1-\frac{\delta_Y^2}{\sigma_Y^2}\right)} = \sqrt{1-\left|\rho_{XY}\right|\cdot\sqrt{\frac{1-\delta_Y^2/\sigma_Y^2}{1-\delta_X^2/\sigma_X^2}}}. \quad (13)$$

Используя (10) и (11), уравнение линейной регрессии (4) запишем в виде

$$\frac{Y}{\sigma_Y \cdot B} = K_0{}' + K_1{}' \cdot \frac{X}{\sigma_X \cdot A}. \quad (14)$$

Уравнение (14) легко привести к виду (2):

$$Y = K_0{}' \cdot \sigma_Y \cdot B + K_1{}' \cdot \frac{\sigma_Y \cdot B}{\sigma_X \cdot A} \cdot X = K_0 + K_1 \cdot X, \quad (15)$$

где

$$K_0 = K_0{}' \cdot A \cdot \sigma_Y \cdot B; \quad (16)$$

$$K_1 = K_1{}' \cdot \frac{\sigma_Y \cdot B}{\sigma_X \cdot A}. \quad (17)$$

Применяя ортогональную среднеквадратическую регрессию к уравнению (14) и используя соотношение (17), получим выражение для искомого коэффициента регрессии:

$$K_1 = \frac{\sigma_Y}{\sigma_X}\cdot\frac{B}{A}\cdot\frac{1}{2\cdot\rho_{XY}} \times \left\{\left(\frac{A}{B}-\frac{B}{A}\right)+\sqrt{\left(\frac{A}{B}-\frac{B}{A}\right)^2+4\cdot\rho_{XY}^2}\right\}, \quad (18)$$

где A и B определяются выражениями (12) и (13).

Впервые формула (18) была представлена в [4], а подробно описана в [5].

Анализ полученных результатов. Выражение (18) позволяет устанавливать однозначную связь между величинами X и Y и определять условия использования известных типов линейной регрессии.

Покажем, что все известные аналитические выражения для коэффициента регрессии K_1 уравнения (2) являются частными случаями формулы (18).

Так, для случая когда разброс точек в корреляционной связи X и Y обусловлен только их случайными погрешностями, т. е. $\rho_{X_0Y_0} = 1$, получим выражение для коэффициента регрессии K_1, приведенное в [1]:

$$K_1 = \frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+\sqrt{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)^2+4\cdot\rho_{XY}^2}\right\}. \quad (19)$$

При $\rho_{X0Y0} = 1$, $\delta_X = 0$ и $\delta_Y \neq 0$ имеем

$$K_1 = \lim_{\delta x\to 0}\frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+\left(-\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}+\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)\sqrt{1+4\cdot\rho_{XY}^2\cdot\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}\right)^2}\right\}.$$

Разлагая выражение под квадратным корнем в ряд Маклорена [6] и оставляя первые два члена, получим

$$K_1=\lim_{\delta x\to 0}\frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+\right.$$
$$\left.+\left(-\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}+\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)\left[1+2\cdot\rho_{XY}^2\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}\right)^2\right]\right\}= \quad (20)$$
$$=\frac{\sigma_Y}{\sigma_X}\cdot\rho_{XY}.$$

Это известная формула для коэффициента K_1 уравнения прямой регрессии $Y = K_0 + K_1 X$, которая находится путем минимизации суммы квадратов отклонений вдоль оси Y от искомой прямой [2].

При $\rho_{X_0Y_0} = 1$, $\delta_Y = 0$ и $\delta_X \neq 0$

$$K_1=\lim_{\delta Y\to 0}\frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+\right.$$
$$\left.+\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)\sqrt{1+4\cdot\rho_{XY}^2\cdot\left(\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)^2}\right\}.$$

Проведя процедуру разложения выражения под квадратным корнем в ряд Маклорена [6] и оставляя первые два члена, получим

$$K_1=\lim_{\delta Y\to 0}\frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+\right.$$
$$\left.+\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)\left[1+2\cdot\rho_{XY}^2\left(\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)^2\right]\right\}. \quad (21)$$
$$=\frac{\sigma_Y}{\sigma_X}\cdot\frac{1}{\rho_{XY}}.$$

Формула (21) – также известная формула для коэффициента $1/K_1^*$ уравнения обратной регрессии $X = K_0^* + K_1^* \cdot Y$, которая получается путем минимизации суммы квадратов отклонений вдоль оси X от искомой прямой [2].

При $\rho_{X_0Y_0} = 1$ и $\delta_X = \delta_Y \neq 0$ для коэффициента K_1 уравнения ортогональной регрессии $Y = K_0 + K_1 X$ получим формулу

$$K_1=\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}-\frac{\sigma_X}{\sigma_Y}\right)+\sqrt{\left(\frac{\sigma_Y}{\sigma_X}-\frac{\sigma_X}{\sigma_Y}\right)^2+4\cdot\rho_{XY}^2}\right\}, \quad (22)$$

которая определяется путем минимизации суммы квадратов отклонений, перпендикулярных искомой прямой [3].

Если для массива данных выполняется соотношение $\dfrac{\delta_X}{\sigma_X}=\dfrac{\delta_Y}{\sigma_Y}$, то из выражения (18) вытекает простая формула для коэффициента регрессии:

$$K_1=\frac{\sigma_Y}{\sigma_X}. \quad (23)$$

Так как соотношение $\dfrac{\delta_X}{\sigma_X}=\dfrac{\delta_Y}{\sigma_Y}$ выполняется для большинства экспериментальных данных, то формулу (23) можно рекомендовать к использованию при отсутствии информации о величинах случайных погрешностей X и Y. Заметим, что эта формула представляет собой среднее геометрическое формул (20) и (21).

Диапазон изменчивости коэффициента регрессии. Для случая когда разброс точек в корреляционной связи величин X и Y обусловлен только их случайными погрешностями, т. е. $\rho_{X_0Y_0} = 1$, коэффициент регрессии будет изменяться в следующих пределах:

$$\frac{\sigma_Y}{\sigma_X}\cdot\left|\rho_{XY}\right|\le\left|K_1\right|\le\frac{\sigma_Y}{\sigma_X}\cdot\frac{1}{\left|\rho_{XY}\right|}, \quad (24)$$

а при $\rho_{X_0Y_0} < 1$ – в пределах

$$\frac{\sigma_Y}{\sigma_X}\cdot\left|\rho_{XY}\right|<\left|K_1\right|<\frac{\sigma_Y}{\sigma_X}\cdot\frac{1}{\left|\rho_{XY}\right|}. \quad (25)$$

Из выражений (24), (25) следует, что коэффициенты для прямой и обратной регрессий принимают соответственно минимальное и максимальное значения.

Кратко сформулируем основные выводы:

– получена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1X$ при условии, что разброс точек в корреляционной связи случайных величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами;

– все известные выражения для коэффициентов регрессии являются частными случаями полученной формулы. Определены условия использования данных выражений;

– обобщенная формула позволяет получать робастные, достоверные и физически корректные коэффициенты регрессии. Эта формула представляет интерес для специалистов, занимающихся обработкой разных массивов данных и может быть использована для их корректной физической интерпретации, независимо от области знания.

Библиографические ссылки

1. Кендалл М., Стьюарт А. Статистические выводы и связи. М. : Наука, 1973.

2. Зайдель А. Н. Погрешности измерений физических величин. Л. : Наука. Ленингр. отд-ние, 1985.

3. Крамер Г. Математические методы статистики. М. : Мир, 1975.

4. Щелканов Н. Н. Построение регрессионной зависимости между аэрозольными оптическими толщами атмосферы с учетом их случайных погрешностей // Аэрозоли Сибири : тез. докл. II заседания раб. группы проекта / Ин-т оптики атмосферы Сиб. отд-ния Рос. акад. наук. Томск, 1995. С. 16.

5. Щелканов Н. Н. Обобщенный метод построения линейной регрессии и его применение для построения однопараметрических моделей аэрозольного ослабления // Оптика атмосферы и океана. 2005. Т. 18. № 1–2. С. 86–90.

6. Кудрявцев В. А., Демидович Б. П. Краткий курс высшей математики. М. : Наука, 1975.

N. N. Shchelkanov

ROBUST METHOD FOR CONSTRUCTION OF LINEAR REGRESSION BETWEEN TWO PHYSICAL PARAMETERS IN VIEW OF THEIR RANDOM ERRORS

In the article the author presents a generalized equation for determination of the regression coefficients of the linear equation $Y = K_0 + K_1 X$ for a general case, when spread of points in correlation between X and Y is caused both by random measurement errors and by uncontrollable physical factors. All the known equations for the regression coefficients appeared to be particular cases of the equation obtained.

Keywords: robust method, linear regression, random errors.

УДК 519.6

В. М. Белолипецкий, С. Н. Генова

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ГИДРОЛЕДОТЕРМИЧЕСКОГО РЕЖИМА РЕКИ АНГАРЫ

Рассматривается компьютерное моделирование гидроледотермического режима реки Ангары выше и ниже плотины строящейся Богучанской ГЭС.

Ключевые слова: гидроледотермика, ГЭС, длина полыньи.

Зарегулирование плотинами гидроузлов вносит существенные изменения в природные условия прилегающих районов. Изменяются температурные и скоростные режимы реки как выше, так и ниже гидроузла. Смена температурного режима оказывает влияние на развитие речной флоры и фауны, а понижение температуры в летний период приводит к уменьшению самоочищающей способности реки. Для полной оценки экологического воздействия строящейся Богучанской ГЭС необходимо знать прогноз гидротермического и ледового режима реки выше и ниже плотины и выработать рекомендации по уменьшению негативного влияния строительства ГЭС.

Для выполнения прогнозных расчетов используются разработанные в Институте вычислительного моделирования Сибирского отделения Российской академии наук математические модели и вычислительные программы, которые применялись для исследования гидроледотермических процессов в бьефах Красноярской, Туруханской и Средне-Енисейской ГЭС [1].

Математические модели гидротермических процессов в водоемах строятся на основе уравнений механики жидкости и теплопереноса с учетом специфики рассматриваемых задач. Для описания гидротермического режима водохранилища используются уравнения гидротермодинамики в приближении пограничного слоя с учетом коэффициента вертикального турбулентного обмена, который определяется по формуле Обухова. В модели учитываются тепловые потоки через свободную поверхность и напряжение трения ветра. Для численного решения уравнения теплопроводности применяются неявные разностные схемы. Полученные системы линейных алгебраических уравнений решаются методом прогонки. Для медленных течений в водохранилищах большой протяженности задача сводится к двумерной модели течения путем применения оператора осреднения по ширине. Скорости течения находятся из решения задачи для функции тока.

Для исследования неустановившихся течений в нижнем бьефе ГЭС используется одномерная постановка соответствующих гидравлических задач, основанных на использовании классических уравнений Сен-Венана. Математическая модель температурного режима водного потока основана на упрощенном одномерном уравнении переноса для средней по сечению температуры воды, учитывающем теплообмен с атмосферой. При наличии притоков температура воды определяется как средневзвешенная.

При изучении ледотермических режимов открытых водотоков основной интерес представляет определение изменений во времени положения створа кромки ледяного покрова (ЛП), толщины ледяного покрова по длине водотока, расхода и общего количества образующейся шуги. При этом различают три основных ледотермических режима в каналах и нижних бьефах ГЭС:

– перемещение кромки льда вверх по течению (наступление) имеет место, когда кромка ЛП расположена ниже створа нулевой изотермы;

– отступление кромки ЛП (температура воды у кромки положительна);

– кромка ЛП не перемещается.

Математическое моделирование ледотермического режима реки основывается на методе расщепления по физическим процессам: исходная задача разделяется на фрагменты, каждый из которых описывает отдель-