$$y=\sum_{i=0}^{\infty} B_i (z-z_0)^{\frac{i-s}{m_1-s}} \quad (B_0 \neq 0, m_1 \in Z_+),\ x=\sum_{i=0}^{\infty} A_i (z-z_0)^{\frac{i-r}{m_1-s}}\ (A_0 \neq 0), \tag{5}$$

где r, s, m_1 - целые положительные числа.

Теорема 2 [2]. Если выполнены условия

$$p_1 \geq q_1+2,\ q_2=p_2+2,\ P_0(y) \text{ многочлен степени } p_2, \tag{6}$$

или

$$q_2 \geq p_2+2,\ p_1=q_1+2,\ Q_0(y) \text{ многочлен степени } q_2, \tag{7}$$

или $p_1 \geq q_1+2$, $q_2 \geq p_2+2$ и

$$l_i \leq l_0+iR,\ r_j \leq q_2-(k-j)R, \text{ где } R=\frac{q_2-l_0-1}{p_1-q_1-1+k}, \tag{8}$$

то система (1) имеет только алгеброидные решения со свойством (2).

Литература.

1. Грицук, Д.В. Существование полярных решений автономных полиномиальных систем двух дифференциальных уравнений / Грицук Д.В. // Материалы V международной научно-практической конференции молодых исследователей «Содружество наук. Барановичи - 2009» (21 – 22 мая 2009 года): В 2-х ч. – Барановичи: РИО БарГУ, 2009. – Ч. 2 – С. 71–73.
2. Грицук, Д.В. Условия алгеброидности решений автономных полиномиальных систем дифференциальных уравнений / Грицук Д.В., Климашевская И.Н. // Материалы республиканской научно-практической конференции «Современные проблемы математического моделирования и новые образовательные технологии в математике» (22 – 23 апреля 2009 года): В 2-х т. – Брест: альтернатива, 2009. – Т. 1 – С. 114–116.

УДК 519.23/.25

ОГРАНИЧЕНИЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ ПРИ ПОСТРОЕНИИ МОДЕЛЕЙ ЛИНЕЙНОЙ РЕГРЕССИИ

Дереченник С.С.
УО «Брестский государственный технический университет», г. Брест

Широкий круг задач обработки эмпирических данных имеет целью установление функциональной связи зависимой переменной (отклика) с одной или несколькими независимыми переменными (факторами). Например, в задаче простой линейной регрессии взаимосвязь между откликом y и фактором x находят в виде модели $y=ax+b$, при этом относительно имеющихся наблюдений (x_i,y_i), $i=\overline{1,n}$ полагают, что $y_i=ax_i+b+\varepsilon_i$, при этом случайные величины (остатки) $\varepsilon_1,..\varepsilon_n$, по условию гомоскедастичности, одинаково распределены и независимы.

Для подбора "наилучшей" прямой широко применяется известный метод наименьших квадратов ([1]), который заключается в такой оценке параметров $(\hat{a},\hat{b})$, которая минимизирует сумму квадратов величин e_i, называемых невязками:

$$\sum_i e_i^2=\sum_{i=1}^{n}(y_i-ax_i-b)^2 \to \min_{a,b}. \tag{1}$$

Расчет уравнения простой линейной регрессии не представляет сложности. Некоторые проблемы возникают в случае построения моделей линейной регрессии с нелинейной связью факторов и отклика. Ряд таких функциональных форм используется, например, в табличном процессоре Microsoft Excel – для построения полиномиальной, логарифмической, степенной и экспоненциальной линий тренда.

Термин "линейный" в названии "линейный регрессионный анализ" указывает на линейность модели относительно коэффициентов (параметров), но не самого фактора / факторов. Некоторые авторы в связи с этим утверждают, что с выборкой $\{x_i\}$ можно проделать любые функциональные преобразования и включить преобразованный фактор в линейное уравнение [2]. Действительно, таким методом удобно линеаризовать, например, полиномы $y = a_1 x_i + a_2 x_i^2 + \ldots + a_k x_i^k + b + \varepsilon_i$, сводя, тем самым, расчет параметров модели к стандартной задаче линейной регрессии. Важно при этом, что ошибка (остатки ε_i) уравнения остается аддитивной. Пример мультипликативного вхождения ошибки иллюстрируется следующей схемой линеаризации и обратного преобразования иной, например, степенной зависимости:

$$y = a x^b, \tag{2.1}$$

$$\lg y_i = \lg a + b \cdot \lg x_i + \varepsilon_i, \tag{2.2}$$

$$y_i = a \cdot x_i^b \cdot \varepsilon_i. \tag{2.3}$$

В подобных случаях регрессионные модели с линеаризованными функциональными зависимостями оказываются некорректными. Обычно рекомендуются два возможных выхода из такой ситуации.

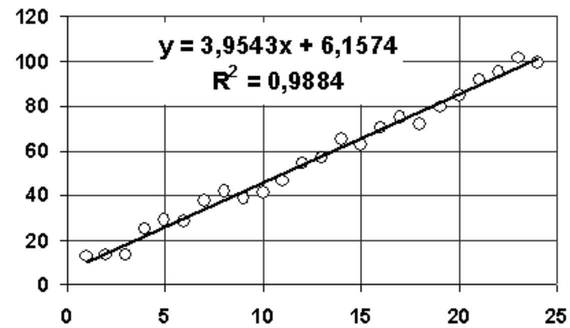
1) Обратное преобразование вида (2.3) не выполняется, а в качестве окончательного и самодостаточного принимается линеаризованное уравнение (2.2), в рамках которого выполняется не только оценка параметров $\left(\hat{a}, \hat{b}\right)$, но и весь последующий анализ данных. В этом случае остатки ε_i аддитивны, но уже зависят от величины x, т.е. условие их гомоскедастичности не выполнено.

2) Преобразования (2.2) и (2.3) вообще не осуществляются, оценка параметров $\left(\hat{a}, \hat{b}\right)$ выполняется с использованием итерационных вычислительных процедур нелинейной оптимизации, например, методом Ньютона-Рафсона. Данный подход реализуется в стандартных пакетах статистических программ.

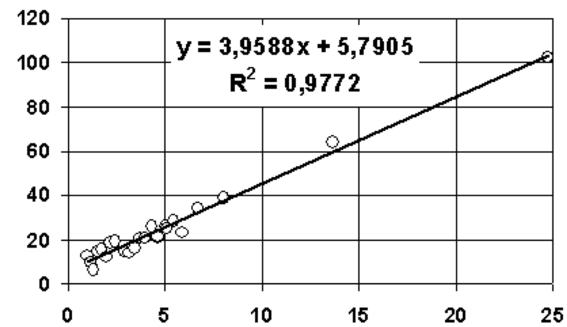
Мы полагаем, что допустимость функциональных преобразований фактора x определяется также видом числовой шкалы, в которой представлены результаты его измерения. Известно, что операции логарифмирования, возведения в дробную степень, использования в качестве логарифма либо показателя степени неприменимы к результатам, представленным в шкале интервалов, разностей либо отношений (имеющим физическую единицу измерения), но возможны над результатами измерений в абсолютной шкале [2]. Поэтому отсутствие размерности у фактора x также является необходимым условием корректности линеаризации произвольной функциональной зависимости $y = f(x)$.

Линеаризация применяется, например, в задачах исследования вероятностных распределений экстремальных метеорологических величин, в частности, годовых максимумов снеговой нагрузки [3]. При этом эмпирическое распределение в области больших значений ("хвостовой" части распределения) проверяется на принадлежность к одному из характерных типов. Распределению Гумбеля (тип I) соответствует линейная аппроксимация нагрузки (отклика) вида $y = ax + b$, где $x = \ln(-\ln F(y))$. Данные действия допустимы, т.к. ошибка отклика остается аддитивной, а двойное логарифмирование выполняется над безразмерной вероятностью. Это сопровождается неравномерной группировкой точек наблюдений (x_i, y_i).

На рисунке 1 представлены результаты расчета методом наименьших квадратов условной линейной модели регрессии с равномерным и неравномерным расположением точек наблюдений. Точки наблюдений соответствуют исходному уравнению $y_i = 4x_i + 5 + \varepsilon_i$, последовательные значения $\varepsilon_i \in [-5;5]$, $i = 1,\ldots,24$ равномерно распределенной случайной ошибки идентичны для обоих вариантов. Усредненные по серии из 50 экспериментов (50 наборам $\{\varepsilon_i\}$) результаты сведены в таблицу.



а



б

Рисунок – Регрессионные модели с равномерной (а) и неравномерной (б) группировкой точек наблюдений

Таблица 1 – Средние оценки параметров и коэффициент детерминации модели линейной регрессии $y = ax + b$

Модель с равномерной группировкой наблюдений		
параметр $\hat{a}$	параметр $\hat{b}$	R-квадрат
3,998±0,003	5,16±0,72	0,9897
Модель с равномерной группировкой наблюдений		
параметр $\hat{a}$	параметр $\hat{b}$	R-квадрат
3,995±0,007	5,16±0,35	0,9802

Нами установлено, что при неравномерном расположении точек наблюдений, соответствие найденных оценок параметров модели исходным коэффициентам, в среднем, несколько ухудшается. Значение коэффициента детерминации R-квадрат при таком расположении закономерно (в каждом эксперименте) ниже. По-видимому, это связано с существенным смещением значений выборочных средних $\bar{x}$ и $\bar{y}$, которые используются для расчета оценок $(\hat{a}, \hat{b})$ и коэффициента R^2. Поэтому стандартный метод наименьших квадратов, основанный на "точечном" анализе невязок согласно соотношению (1), может оказаться не вполне корректным в случае неравномерного расположения точек наблюдений, например, при нелинейном преобразовании шкалы фактора. Данный вывод косвенно подтверждается сведениями из раздела "О типах линий тренда" справочного руководства табличного процессора MS Excel, где указано (без каких-либо дополнительных пояснений), что "отображаемое вместе с линией тренда значение величины R-квадрат не является корректным".

Нами предлагается в качестве меры близости линии регрессии к эмпирическим точкам наблюдений использовать интегральную квадратичную оценку – функционал вида:

$$J = \int_{x_{\min}}^{x_{\max}} \left[y^*(x) - (ax+b) \right]^2 dx \to \min_{a,b}, \tag{3}$$

где $y^*(x)$ – некоторое кусочно-линейное приближение искомой функции, заданной точечными отсчетами (x_i, y_i).

Подходящим приближением в общем случае может являться ломаная, а в задачах вероятностного типа – ступенчатая функция. В этих случаях функционал вычислим аналитически как для простой линейной регрессии, так и для некоторых ее нелинейных форм (например, экспоненциальной). Поэтому возможна также и аналитическая его минимизация, что позволит получить оценки параметров регрессии в замкнутой форме, без необходимости использования численных итерационных процедур. Аналогичный подход применим также для нахождения интегрального коэффициента детерминации.

Таким образом, минимизация интегральной квадратичной оценки при построении моделей регрессии нечувствительна к неравномерности расположения (группировки) точек наблюдения, не требует линеаризации некоторых нелинейных форм зависимости отклика и фактора, и, по-видимому, более корректна в отношении оценки коэффициента детерминации.

Литература

1. Корн, Г. Справочник по математике (для научных работников и инженеров) / Г. Корн, Т. Корн. – М.: Наука, 1978. – 832 с.

2. Шитиков, В.К. Количественная гидроэкология: методы системной идентификации / В.К. Шитиков, Г.С. Розенберг, Т.Д. Зинченко – Тольятти, ИЭВБ РАН, 2003. – 463 с.

3. Тур, В.В. Нормирование снеговых нагрузок для территории Республики Беларусь / Тур В.В., Валуев В.В., Дереченник С.С. [и др.] // Строительная наука и техника. – 2008. – № 2 (17). – С. 27–45.

УДК 517.512.2

АППРОМАКСИМАЦИОННЫЕ МЕТОДЫ ВЫЧИСЛЕНИЯ ИНТЕГРАЛОВ

Дыбко А.В., Дацык В.Т.

УО «Брестский государственный университет им. А.С.Пушкина», г. Брест

Пусть требуется вычислить определённый интеграл $\int_a^b f(x)dx$, где $f(x)$ есть некоторая заданная на промежутке $[a,b]$ непрерывная функция. Мы имеем много примеров вычисления подобных интегралов, либо с помощью первообразной, если она выражается в конечном виде, либо же – минуя первообразную – с помощью различных приёмов, большей частью искусственных. Нужно отметить, однако, всем этим исчерпывается лишь довольно узкий класс интегралов; за его пределами обычно прибегают к различным методам приближённого вычисления.