

ОБ ОЦЕНКЕ КОЭФФИЦИЕНТОВ В УРАВНЕНИИ ЛИНЕЙНОЙ РЕГРЕССИИ, ОПИСЫВАЮЩЕМ ИЗМЕНЕНИЯ УРОВНЯ ВОДЫ В РУСЛЕ ГОРНОЙ РЕКИ МЗЫМТА
ON THE ESTIMATION OF COEFFICIENTS IN THE LINEAR REGRESSION EQUATION, WHICH DESCRIBES THE CHANGES IN WATER LEVEL IN THE MAINSTREAM OF THE MOUNTAIN RIVER MZYMTA

Данная работа посвящена оценке (методами регрессионного анализа) коэффициентов в уравнении линейной регрессии (математической модели), описывающем изменения уровня воды в русле горной реки Мзымта. Построена математическая модель изменений уровня воды в данной реке. На основе статистического материала о характеристиках паводка, имевшем место в октябре 2003 года в районе гидрологического поста этой реки, полученная модель была подвергнута проверке на адекватность экспериментальными данными.

This work is devoted to the evaluation (regression analysis) of the coefficients in the linear regression equation (mathematical model) that describes the changes in water level in the mainstream rock Mzymta. To construct a mathematical model of changes in water level in the river. On the basis of statistical data on the characteristics of the flood, which had a one hundred-ME in October 2003 in the hydrological position of the river, half-tained model was subjected to a check on the adequacy of the experimental data.

Многие реки Краснодарского края относятся к рекам горного типа. Реки такого типа обладают определенными гидрологическими особенностями, существенно отличающими их от рек равнинного и горно-равнинного типов. Поэтому методики, описывающие изменения уровня воды в руслах рек равнинного и горно-равнинного типов, не применимы (как показывают вычислительные эксперименты) для описания изменения уровня воды в руслах рек горного типа. В данной работе предложена методика измерения колебаний

уровня воды в русле реки Мзымта, относящейся к рекам этого типа. Точный прогноз повышения уровня воды позволит руководству ГУ «Краснодарский краевой центр по гидрометеорологии и мониторингу окружающей среды» (ГУ КЦГМС) заблаговременно предупреждать гражданское население о возможности возникновения паводковой ситуации на отдельных участках русла указанной реки.

Цель работы – методами регрессионного анализа разработать методику прогноза изменений уровня воды в русле одной из рек горного типа – реке Мзымта (Краснодарский край).

Для достижения поставленной цели были использованы статистические данные о метеоусловиях и уровне воды за июнь-август 2010 года, предоставленные ГУ КЦГМС.

Пусть значения уровня воды в реке Мзымта на момент времени t_i совпадают со значениями функции

$$y(t_i) = f(x_1(t_{i-3}), x_1(t_{i-2}), x_1(t_{i-1}), x_2(t_i)),$$

$$t_{i-3} < t_{i-2} < t_{i-1} < t_i, i = 3, 4, \dots, n, \quad (1)$$

где $y(t_i)$ – прогнозируемый уровень воды в створе данной реки Мзымта в районе гидрологического поста, $x_1(t_{i-3}), x_1(t_{i-2}), x_1(t_{i-1})$ – уровень воды в моменты $t_{i-3}, t_{i-2}, t_{i-1}$ соответственно, $x_2(t_i)$ – количество осадков, выпавших на момент t_i в окрестности русла.

Пусть функция (1) является линейной:

$$y(t_i) = b_1 + a_1 x_1(t_{i-3}) + a_2 x_1(t_{i-2}) + a_3 x_1(t_{i-1}) + a_4 x_2(t_i), \quad (2)$$

где a_1, a_2, a_3, a_4, b – некоторые постоянные.

На основе статистических данных методами регрессионного анализа были вычислены значения коэффициентов и свободного члена в (2) (см. табл. 1).

Таблица 1

Значения коэффициентов и свободного члена в регрессионной модели (2)

b	– 3,990373807
a_1	0,958433081
a_2	0,683938674
a_3	– 0,628201975

a_4	- 0,001390261
-------	---------------

Согласно данным, приведенным в табл. 1, функция (2) (регрессионная модель) имеет вид:

$$y(t_i) = - 3,990374 + 0,95843x_1(t_{i-3}) + 0,683939x_1(t_{i-2}) - 0,628201x_1(t_{i-1}) - 0,00139x_2(t_i). \quad (3)$$

Точность оценки прогноза значений $y(t_i)$ определяется числовыми значениями характеристик регрессионной модели (3), приведенными в табл. 2,

Таблица 2

Характеристики регрессионной модели (2)

R^2	0,993972
Нормированное значение R^2	0,993692
Стандартная ошибка σ	1,627712
Количество наблюдений n	91
F – критерий Фишера	3545,484

коэффициента детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4)$$

и значением

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1}, \quad (5)$$

где n – размер выборки (количество проведенных замеров), k – число коэффициентов модели (3), y_i – наблюдаемое значение $y(t)$, \tilde{y}_i – значение $y(t)$ вычисленное с помощью уравнения регрессии (3) в момент t_i , $i = 3, \dots, n$, \bar{y} – среднее значение $y(t)$, вычисленное на основе статистических данных. Если F превышает некоторое критическое значение $F_{кр}$, определяемое по таблице значений F -критерия Фишера при заданных n , k и заданной доверительной вероятности α , а величина R^2 близка к 1, то значение $y(t_i)$, вычисленное по

формуле (3), с вероятностью, близкой к 1, совпадает с экспериментальными данными. Согласно статистическим данным и допущениям, указанным выше, $F_{кр} = 1,4170883451$, $\alpha = 0,95$, $n = 91$, $k = 4$, $R^2 = 0,996982$, $F = 3545,48$. Т.к. $F > F_{кр}$ ($3545,48 > 1,4170883451$, т.е. F во много раз превышает $F_{кр}$), то с вероятностью 0,95 при стандартной ошибке $\sigma = 1,627712$ значения $y(t_i)$ совпадают с реальными значениями $y(t_i)$, $i = 3, 4, \dots, n$.

На рис. 1 приведены графики значений $y(t)$, построенные на основе статистических данных и значений регрессионной модели (3). Из этих графиков видно, что прогнозируемые с помощью (3) значения $y(t_i)$ несущественно отличаются от статистических данных.

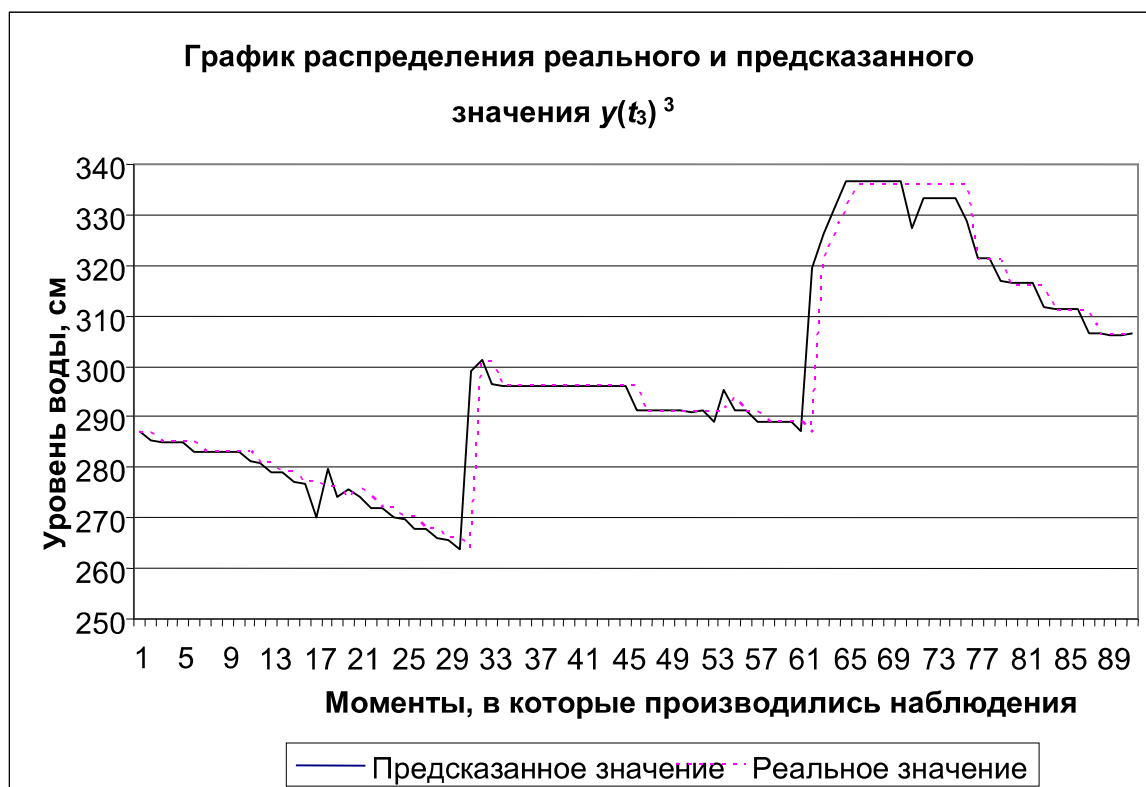


Рис. 1. Графики значений $y(t)$, построенные на основе статистических данных регрессионной модели (3)

Пример. В районе гидрологического поста на реке Мзымта (на данной реке один гидрологический пост) при уровне воды, равном 280 см, как считает ГУ КЦГМС, будет наблюдаться неблагоприятное явление (НЯ) (незначительный паводок, который не принесет значительного экономического

ущерба краю). При уровне воды, равном 300 см, будет наблюдаться опасное явление (ОЯ), т.е. сильный (значительный) паводок, который может принести существенный ущерб.

Пусть $n = 3$, $t_0 = 8$, $t_1 = 20$, $t_2 = 32$, $t_3 = 44$. Статистические данные, взятые из журнала записи ежедневной информации на гидропосту за октябрь 2003 года, приведены в табл. 3.

Таблица 3

Значения стат. данных из журнала записи ежедневной информации за октябрь 2003 года

уровень воды в момент $t_0 = 8$ (см)	уровень воды в момент $t_1 = 20$ (см)	уровень воды в момент $t_2 = 32$ (см)	кол-во осадков в момент $t_3 = 44$ (мм)
211	247	251	43

Подставив данные, приведенные в табл. 3, в (3), получим значение $y(t_3) = 280,8795$ (см). Таким образом, в момент $t_3 = 44$ должен был наблюдаться незначительный паводок. Статистические данные подтверждают такой прогноз: 25 октября 2003 года на реке Мзымта действительно наблюдались неблагоприятное явление (незначительный паводок).

Замечание. Момент $t_0 = 8$ соответствует 8 ч. МСК 24.10.2003 г., $t_1 = 20 - 20$ ч. МСК 24.10.2003 г., $t_2 = 32 - 8$ ч. МСК 25.10.2003 г., $t_3 = 44 - 20$ ч. МСК 25.10.2003 г.

Литература

1. Гусаров В.М. Теория статистики, М.: ЮНИТИ, 2001.
2. Ивченко Г.И., Медведев Ю.И. Математическая статистика: Учеб. пособие для втузов. М.: Высш. шк., 1984.
3. Крамер Г. Математические методы статистики. М.: Мир, 1975.
4. Общая теория статистики: Учебник / Т.В. Рябушкин, М.Р. Ефимова, Ипатова, Н.И. Яковлева. М.: Финансы и статистика, 1981.
5. Статистическое моделирование и прогнозирование под. ред. Гранборга. Москва: Финансы и статистика, 2000.

КЛЮЧЕВЫЕ СЛОВА:

Математическое моделирование, регрессионный анализ, прогнозирование, метод наименьших квадратов, паводковые ситуации, критерий Фишера.

Mathematical modeling, regression analysis, forecasting, least squares method, floods situation, Fisher's exact test.