

Использование качественных параметров в уравнениях регрессии

Меньших В.В., email: menshikh@list.ru
Морозова В. О., email: leryxa.28@mail.ru

Воронежский институт МВД России

***Аннотация.** В данной работе был предложен метод, позволяющий использовать описания лингвистически заданных значений показателей в моделях множественной регрессии. В статье были рассмотрены методы, позволяющие заменять лингвистические оценки экзогенных и эндогенных показателей специальным образом построенными нечёткими множествами, использовать их при регрессионном моделировании и осуществлять обратный переход к лингвистическим оценкам.*

***Ключевые слова:** статистические данные, уравнения регрессии, экзогенные показатели, эндогенные показатели, нечеткие множества, количественные и качественные показатели.*

Введение

Одним из самых эффективных инструментов прогнозирования и исследования объектов, сложных систем и явлений на основе использования статистических данных являются модели множественной регрессии:

$$Y = f(X_1, X_2, \dots, X_m) + \varepsilon, \quad (1)$$

позволяющие по значениям нескольких известных (экзогенных) показателей X_1, X_2, \dots, X_m определять значение неизвестного (эндогенного) показателя Y [1, 2].

Достаточно хорошо изучены вопросы отбора существенных для прогнозирования экзогенных показателей, выбора вида модели

регрессии, т. е. функциональной зависимости f и оценки значимости результатов моделирования с помощью модели (1) [2].

Вместе с тем, в целом ряде предметных областей значения отдельных показателей (как экзогенных, так и эндогенных), приводимых в официальной статистике, заданы не количественно, а описаны лингвистически (например, качество «абсолютно недопустимо», «низкое», «среднее», «высокое», «абсолютно приемлемо»). Такие оценки очень часто используются в организационных и организационно-технических системах, в частности в процессе принятия решений в системах критического применения (например, правоохранительных органах) [3].

Указанное обстоятельство делает невозможным непосредственное использование регрессионного моделирования в указанных ситуациях, т. к. регрессионные модели могут быть построены только на основе числовых значений показателей. Вопросы количественного описания лингвистически описанных параметров описаны в [4, 5], где представлены методы, позволяющие заменять лингвистические оценки специальным образом построенными нечёткими множествами.

В данной работе описывается метод, позволяющий использовать указанные описания лингвистически заданных значений показателей в моделях множественной регрессии и, если это необходимо, осуществлять обратный переход к лингвистическим оценкам.

1. Описание метода

Для решения описанной задачи предлагается заменять лингвистические значения экзогенных и эндогенных показателей их количественными оценками, определяемыми медианными значениями a функций принадлежности $\mu(x)$ соответствующих этим значениям нечётких множеств. Данные оценки находятся как решения следующего уравнения [5]:

$$\int_{-\infty}^d \mu(x) dx = \int_d^{+\infty} \mu(x) dx \quad (2)$$

Сложность данной задачи зависит от вида функции $\mu(x)$. Если эндогенный показатель y первоначально описывался лингвистически, то после использования регрессионной модели его значение будет

описывать количественно и поэтому необходимо осуществить обратный переход от количественного представления к лингвистическому. Для решения этой задачи предлагается использовать следующий алгоритм.

Пусть z_1, z_2, \dots, z_s – возможные лингвистические значения показателя Y ;

$d(z_1), d(z_2), \dots, d(z_s)$ – медианные значения функций принадлежности соответствующих им нечётких множеств.

Если в результате регрессионного анализа определено, что значение показателя Y приняло значение d , то в качестве лингвистической оценки этого показателя следует принять:

$$z_k = A_{\min} (d - d(z_i)), \quad (3)$$

где $i = 1, 2, \dots, s$.

Таким образом, решение данной задачи сводится к нахождению медианных оценок функций принадлежности нечётких множеств, описывающих лингвистически заданные оценки эндогенного показателя. Обратимся к решению данной задачи.

2. Описания качественных значений показателей

Рассмотрим наиболее распространённый подход к описанию качественных значений показателей, заключающийся в использовании L-R нечётких множеств, т. е. множеств с функциями принадлежности треугольного вида.

Предполагается, что $\mu(x)$ линейно возрастает на промежутке $[a, c]$ и линейно убывает на промежутке $[c, b]$. Любой из указанных промежутков может оказаться вырожденным (иметь нулевую длину), но не оба промежутка одновременно.

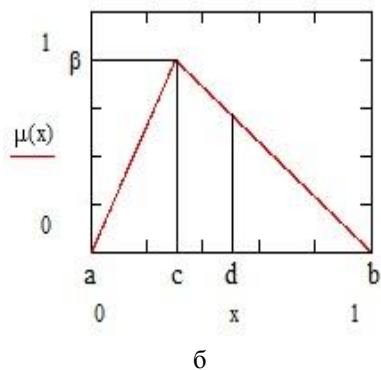
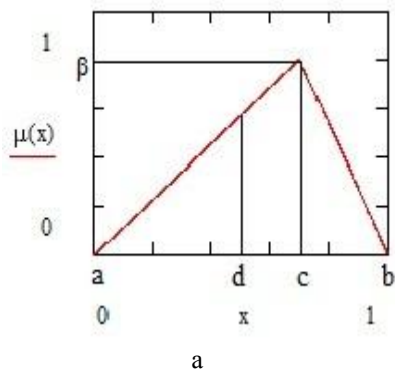
Аналитически функцию принадлежности такого нечёткого множества можно описать следующим образом:

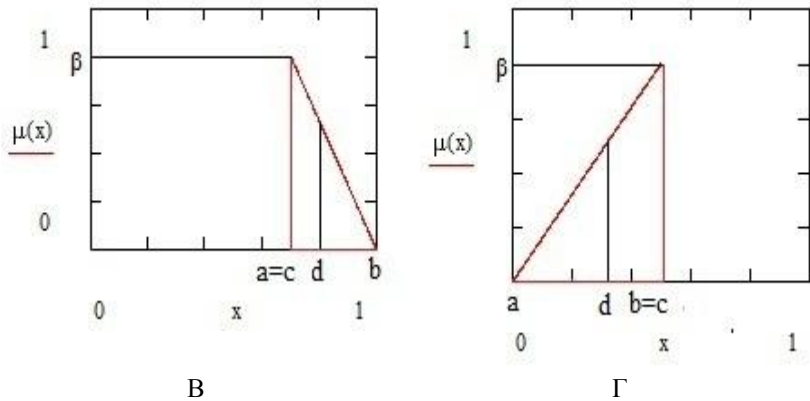
$$\mu(x) = \begin{cases} 0, & \text{если } x \leq a; \\ \frac{\beta}{c-a}(x-a) & \text{если } a < x \leq c; \\ \frac{\beta}{c-b}(x-b) & \text{если } c < x \leq b; \\ 0, & \text{если } x > b, \end{cases} \quad (4)$$

где $\beta \leq 1$ – максимальное значение функции принадлежности нечёткого множества.

Как следует из анализа возможных представлений нечётких оценок лингвистических оценок показателей с помощью методов, описанных в [4], на рис. 1 возможные четыре ситуации:

2. $c - a \geq b - c > 0$;
3. $0 < c - a \leq b - c$;
4. $a = c$;
5. $b = c$.





a – при $c - a \geq b - c > 0$, b – при $0 < c - a \leq b - c$, $в$ – при $a = c$, $г$ – при $b = c$

Рис. 1. Множество с треугольными функциями принадлежности

Обратимся к решению уравнения (2) для каждого из описанных случаев, т. е. к нахождению медианных оценок функций принадлежности нечётких оценок.

3. Медианные оценки показателей

Приведём результаты численного решения уравнения (2) для каждого из перечисленных результатов.

1. при $c - a \geq b - c > 0$ медианное значение

$$d = \frac{4a + \sqrt{8(b-a)(c-a)}}{4};$$

2. при $0 < c - a \leq b - c$ медианное значение

$$d = b - \sqrt{\frac{1}{2}(b-a)(b-c)};$$

3. при $a = c$ медианное значение $d = b(1 - \frac{\sqrt{2}}{2}) + \frac{a\sqrt{2}}{2};$

4. при $b = c$ медианное значение $d = a(1 - \frac{\sqrt{2}}{2}) + \frac{\sqrt{2}b}{2}$.

4. Численный пример

Рассмотрим фрагмент уголовной статистики [3], использующий показатели:

γ - уровень преступности (низкий, средний, высокий);

x_1 - численность населения;

x_2 - количество преступлений совершенных в 2020 году.

В качестве нечётких множеств рис. 2, описывающих лингвистические оценки используем следующие функции:

$$\mu_1(x) = \begin{cases} \frac{1}{-0,3}(x - 0,3) & \text{если } 0 < x \leq 0,3; \\ 0, & \text{если } x > 0,3. \end{cases} \quad (5)$$

$$\mu_2(x) = \begin{cases} 0, & \text{если } x \leq 0,2; \\ \frac{1}{0,5 - 0,2}(x - 0,2) & \text{если } 0,2 < x \leq 0,5; \\ \frac{1}{0,5 - 0,8}(x - 0,8) & \text{если } 0,5 < x \leq 0,8; \\ 0, & \text{если } x > 0,8. \end{cases} \quad (6)$$

$$\mu_3(x) = \begin{cases} 0, & \text{если } x \leq 0,7; \\ \frac{1}{1 - 0,7}(x - 0,7) & \text{если } 0,7 < x \leq 1. \end{cases} \quad (7)$$

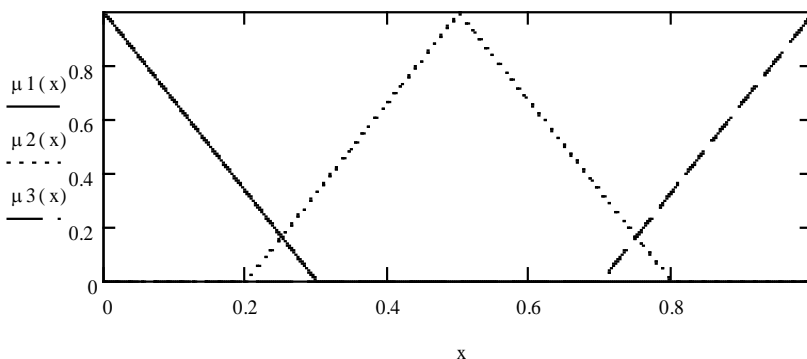


Рис. 2. Нечёткие множества, описывающие лингвистические оценки

Как видно, нечеткие числа зацеплены друг за друга. Это отражает тот факт, что нет резкого разделения между соседними оценками, и переход от одной лингвистической переменной к другой происходит постепенно.

Медианные оценки для указанных нечётких множеств имеют следующие значения, полученные в п.4:

$$d(z_1) = 0,088 ;$$

$$d(z_2) = 0,5;$$

$$d(z_3) = 0,912$$

На основе регрессионного анализа, получено следующее уравнение линейной множественной регрессии:

$$Y = 1340,3 - 0,13 \cdot X_1 + 0,014 \cdot X_2 + \varepsilon.$$

Указанное уравнение, применённое к данным Воронежской области и после решения оптимизационной задачи (3), позволят сделать вывод, что в Воронежской области средний уровень преступности, что соответствует известным статистическим данным [3]. При этом, если показатель x_2 увеличится на 30%, то уровень станет высоким, а если показатель x_2 уменьшится на 50%, то уровень станет низким.

Полученный вывод показывает, что применение разработанного метода может быть полезно при осуществлении аналитической работы и принятии управленческих решений в системах критического применения.

Заключение

Таким образом, в данной работе был предложен метод, позволяющий использовать описания лингвистически заданных значений показателей в моделях множественной регрессии.

Разработанный метод может быть усовершенствован на основе замены скалярной оценки функции принадлежности (её медианного значения) векторной оценкой. Однако в этом случае значительно возрастает вычислительная сложность решения задачи.

Список литературы

2. Елисеева И. И. Эконометрика : учебник для бакалавриата и магистратуры / И. И. Елисеева – М. : Издательство Юрайт, 2017. — 449 с.

3. Кремер, Н. Ш. Эконометрика : учебник для студентов вуза / Н. Ш. Кремер, Б. А. Путко ; под ред. Н. Ш. Кремера. - 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2010.- 125 с.

4. Сведения о состоянии преступности в Российской Федерации за январь-декабрь 2020 года: [Электронный ресурс] // аналитический отчет ФКУ ГИАЦ МВД России. – Режим доступа: <https://xn--b1aew.xn--p1ai/reports/item/21933965/>.

5. Меньших В.В. Правовая статистика: методы и модели : учебное пособие / О. Ю. Данилова, В. В. Меньших, С. В. Синегубов. – В. : Воронежский институт МВД России, 2018. - 302 с.

6. Борисов А. Н. Принятие решений на основе нечетких моделей : Примеры использования / А. Н. Борисов, О. А. Крумберг, И. П. Федоров – Р. : 1990. - 184 с.

7. Пономарев А. С. Нечеткие множества в задачах автоматизированного управления и принятия решений. – Х. : НТУ «ХПИ», 2005. -258 с.