Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования Санкт-Петербургский горный университет

Кафедра высшей математики

СПЕЦИАЛЬНЫЕ ГЛАВЫ МАТЕМАТИКИ

РАСЧЕТ НАИЛУЧШИХ ЛИНЕЙНЫХ ОЦЕНОК КОЭФФИЦИЕНТОВ РЕГРЕССИИ И СТАТИСТИЧЕСКИЙ АНАЛИЗ УРАВНЕНИЯ РЕГРЕССИИ

Методические указания к самостоятельным работам для студентов магистратуры направления 27.04.01

САНКТ-ПЕТЕРБУРГ
2021

УДК 519.2.06(073)

СПЕЦИАЛЬНЫЕ ГЛАВЫ МАТЕМАТИКИ. **Расчет наилучших линейных оценок коэффициентов регрессии и статистический анализ уравнения регрессии:** Методические указания к самостоятельным работам / Санкт-Петербургский горный университет. Сост.: *Л.М. Могилева*, *М.Б. Шабаева*. СПб, 2021. 31 с.

Методические указания разработаны в соответствии с требованиями федерального государственного образовательного стандарта высшего образования.

Содержат основные сведения о методах регрессионного анализа, примеры решения задач, варианты индивидуальных домашних заданий.

Предназначены для студентов магистратуры направления 27.04.01 «Стандартизация и метрология», изучающих дисциплину «Специальные главы математики».

Научный редактор проф. *А.П. Господариков*

Рецензент вед. научн. сотрудник, проф. *Н.В. Смородина* (ПОМИ РАН)

ВВЕДЕНИЕ

Регрессионные модели – один из основных классов математических моделей. Они применяются для более полного понимания сущности происходящих процессов и их анализа. Модель, построенная на основе уже имеющихся значений объясняющих переменных (*факторов*), может быть использована для прогноза зависимой переменной (*отклика*) в будущем или для других значений объясняющих переменных.

В регрессионных моделях зависимая (объясняемая) переменная (отклик) y представляется в виде функции $\eta\left(\vec{x};\vec{\theta}\right)=\eta\left(x^{(1)},x^{(2)},...,x^{(k)};\theta_1,\theta_2,...,\theta_m\right)$, где $x^{(1)},x^{(2)},...,x^{(k)}$ – независимые (объясняющие) переменные (факторы), а $\theta_1,\theta_2,...,\theta_m$ – параметры. В зависимости от вида функции $\eta(\vec{x};\vec{\theta})$ модели делятся на линейные и нелинейные: в случае *линейной* модели параметры входят в регрессионную функцию линейно, в противном случае имеем *нелинейную* модель. Например, $\eta(x;a,b)=a+bx$, $\eta(x;a,b,c)=a+bx+cx^2$, $\eta(x;a,b)=a\sin x+b\cos x$ – линейные (по параметрам) модели, $\eta(x;a,b)=\exp(a+bx)$, $\eta(x;a,b)=1/(a+bx)$ – нелинейные модели. Если число независимых переменных (факторов) $k=1$, то это простая регрессия, если $k>1$ – множественная.

1. ПОДГОНКА КРИВОЙ. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Пусть дан набор двух переменных $\{(x_j,y_j)\}_{j=1}^{n}$; пары (x_j, y_j) можно отобразить точками на плоскости Oxy (рис. 1). Цель – подобрать функцию $y=\eta\left(x,\hat{\vec{\theta}}\right)$ из параметрического семейства функций $\eta(x;\vec{\theta})$, «наилучшим» способом описывающую зависимость y от x (например, найти «наилучшую» линейную функцию

$y = \hat{a} + \hat{b}x)$. В качестве меры близости функций $\eta(x;\vec{\theta})$ от набора наблюдений $\{(x_j, y_j)\}_{j=1}^{n}$ можно взять сумму квадратов отклонений

$$\Delta = \sum_{j=1}^{n} e_j^{\ 2} = \sum_{j=1}^{n} \left(y_j - \hat{y}_j\right)^2 = \sum_{j=1}^{n} \left(y_j - \eta(x_j;\vec{\theta})\right)^2 = \sum_{j=1}^{n} \left(y_j - a - bx_j\right)^2 .$$

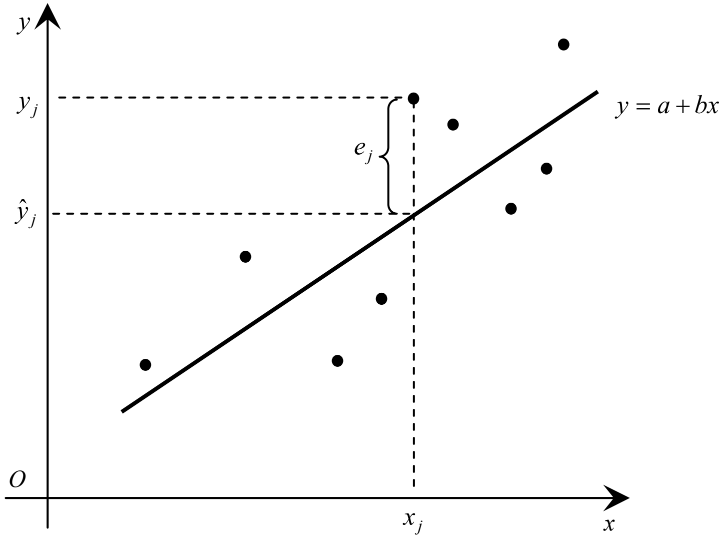

Рис. 1. Подбор “наилучшей” линейной функции

Запишем необходимые условия экстремума:

$$\begin{cases} \dfrac{\partial \Delta}{\partial a} = 0; \\ \dfrac{\partial \Delta}{\partial b} = 0, \end{cases}$$

т.е.

$$\begin{cases} -2\sum_{j=1}^{n}\left(y_j - a - bx_j\right) = 0; \\ -2\sum_{j=1}^{n} x_j\left(y_j - a - bx_j\right) = 0; \end{cases} \quad \Rightarrow \quad \begin{cases} \sum_{j=1}^{n}\left(y_j - a - bx_j\right) = 0; \\ \sum_{j=1}^{n} x_j\left(y_j - a - bx_j\right) = 0. \end{cases} \qquad (1)$$

Раскроем скобки и получим так называемую *стандартную форму нормальных уравнений*:

$$\begin{cases} an + b\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j; \\ a\sum_{j=1}^{n} x_j + b\sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} x_j y_j. \end{cases}$$

Решив систему методом Крамера, получим:

$$\begin{cases} \hat{b} = \dfrac{n\sum_{j=1}^{n} x_j y_j - \left(\sum_{j=1}^{n} x_j\right)\cdot\left(\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2}; \\ \hat{a} = \dfrac{1}{n}\sum_{j=1}^{n} y_j - \dfrac{1}{n}\sum_{j=1}^{n} x_j \cdot \hat{b}. \end{cases} \qquad (2)$$

2. ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Рассмотрим n-мерное векторное пространство $\mathbb{R}^n$ со стандартным скалярным произведением:

$$(\vec{x}, \vec{y}) = \vec{x}^T \cdot \vec{y} = \sum_{j=1}^{n} x_j y_j,$$

где $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$, $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$, $\vec{x}^T$ – вектор-строка.

Обозначим

$$\vec{q} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}, \qquad \hat{\vec{y}} = a \cdot \vec{q} + b \cdot \vec{x} = \begin{pmatrix} a + bx_1 \\ a + bx_2 \\ \dots \\ a + bx_n \end{pmatrix}, \qquad \vec{e} = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = \vec{y} - \hat{\vec{y}}.$$

Здесь a и b – числовые коэффициенты, $\hat{\vec{y}}$ – вектор, лежащий в плоскости π, натянутой на векторы $\vec{q}$ и $\vec{x}$ (рис. 2).

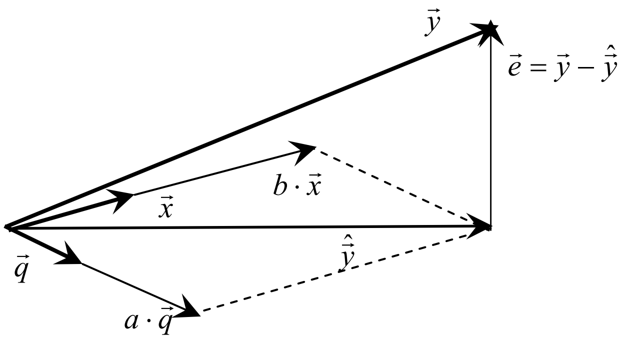

Рис. 2. Геометрический подход к методу наименьших квадратов

ор $\vec{e}$ имел наименьшую длину (чтобы «наилучшим» образом аппроксимировать вектор $\vec{y}$ вектором $\hat{\vec{y}}$, лежащим в плоскости π).

Ясно, что решением является вектор $\hat{\vec{y}}$, для которого $\vec{e} \perp \pi$, т.е. $\vec{e} \perp \vec{x}$ и $\vec{e} \perp \vec{q}$. Условием перпендикулярности векторов является равенство нулю их скалярного произведения:

$$\begin{cases} \vec{q}^{\,T} \cdot \vec{e} = 0; \\ \vec{x}^{\,T} \cdot \vec{e} = 0, \end{cases}$$

т.е. $\begin{cases} \sum_{j=1}^{n} e_j = 0; \\ \sum_{j=1}^{n} x_j e_j = 0 \end{cases}$ или $\begin{cases} \sum_{j=1}^{n} \left(y_j - a - bx_j\right) = 0; \\ \sum_{j=1}^{n} x_j \left(y_j - a - bx_j\right) = 0. \end{cases}$

Таким образом, получили те же условия (см. (1)), из которых находятся "наилучшие" оценки метода наименьших квадратов.

3. МАТРИЧНАЯ ФОРМА ЗАПИСИ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Введем обозначения: $G = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \ldots & \ldots \\ 1 & x_n \end{pmatrix}$ – матрица размера $n \times 2$,

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}, \ \vec{\theta} = \begin{pmatrix} a \\ b \end{pmatrix}, \ \vec{e} = \vec{y} - G\vec{\theta} = \begin{pmatrix} y_1 - (a + bx_1) \\ y_2 - (a + bx_2) \\ \ldots \\ y_n - (a + bx_n) \end{pmatrix}.$$

Из условия $\vec{e} \perp \pi$ следует, что $G^T \cdot \vec{e} = 0$, т.е. $G^T \cdot (\vec{y} - G\vec{\theta}) = 0 \ \Rightarrow \ G^T \vec{y} - G^T G\vec{\theta} = 0 \ \Rightarrow \ G^T G\vec{\theta} = G^T \vec{y}$, откуда получили

$$\hat{\vec{\theta}} = (G^T G)^{-1} G^T \vec{y}\,. \qquad (3)$$

Можно убедиться, что полученное выражение (3) для оценки параметров совпадает с предыдущей оценкой (2):

$$\hat{\vec{\theta}} = (G^T G)^{-1} G^T \vec{y} =$$

$$= \left[\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \right]^{-1} \cdot \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum_{j=1}^{n} x_j \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_{j=1}^{n} y_j \\ \sum_{j=1}^{n} x_j y_j \end{pmatrix} =$$

$$= \frac{1}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2} \begin{pmatrix} \sum_{j=1}^{n} x_j^2 & -\sum_{j=1}^{n} x_j \\ -\sum_{j=1}^{n} x_j & n \end{pmatrix} \cdot \begin{pmatrix} \sum_{j=1}^{n} y_j \\ \sum_{j=1}^{n} x_j y_j \end{pmatrix} =$$

$$= \frac{1}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2} \begin{pmatrix} \left(\sum_{j=1}^{n} x_j^2\right) \cdot \sum_{j=1}^{n} y_j - \left(\sum_{j=1}^{n} x_j\right) \cdot \sum_{j=1}^{n} x_j y_j \\ -\left(\sum_{j=1}^{n} x_j\right) \cdot \left(\sum_{j=1}^{n} y_j\right) + n\sum_{j=1}^{n} x_j y_j \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}.$$

4. МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

Ранее нас интересовало лишь качество подгонки кривой. Теперь рассмотрим статистические свойства данных.

В модели парной линейной регрессии зависимость между переменными представляется в виде

$$y = a + bx + \varepsilon\ , \qquad (4)$$

где x - неслучайная (детерминированная) величина (объясняющая переменная); ε - случайная величина (случайное возмущение или отклонение); y - тоже случайная величина, это объясняемая (зависимая) переменная.

Наличие в равенстве (4) случайного слагаемого ε (так называемой *ошибки*) объясняется тем, что отклик y зависит не только от x, но и от других, неучтенных в уравнении факторов, а также от неизбежных неточностей измерений. Таким образом, можно считать, что ε - случайная величина с некоторой функцией распределения, которой соответствует функция распределения случайной величины y.

Уравнения для отдельных наблюдений зависимой переменной y записываются в виде

$$y_j = a + bx_j + \varepsilon_j\ .$$

Относительно ошибок ε_j, $j = 1,...,n$, в регрессионных моделях принимают следующие предположения (*условия Гаусса – Маркова*):

1. Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю, т.е.

$$M(\varepsilon_j) = 0\ .$$

2. Дисперсия случайного члена должна быть постоянной для всех наблюдений, т.е.

$$D(\varepsilon_j) = M(\varepsilon_j{}^2) = \sigma^2\ .$$

3. Случайные члены должны быть некоррелированы (или даже статистически независимы между собой), т.е.

$$M(\varepsilon_j \cdot \varepsilon_{j'}) = 0 \text{ при } i \neq j\ .$$

4. x – детерминированная величина; векторы $(x_1, x_2, ..., x_n)^T$ и $\vec{q} = (1,1,...,1)^T$ неколлинеарны, т.е. измерения зависимой переменной проводились более чем в одной точке.

5. Ошибки ε_j имеют нормальное распределение (с $M(\varepsilon_j) = 0$ и неизвестной дисперсией $D(\varepsilon_j) = \sigma^2$).

Предположение о нормальности распределения случайной составляющей является следствием центральной предельной теоремы. Оно необходимо для проверки значимости параметров регрессии и их интервального оценивания.

Регрессионная модель в предположении справедливости условий Гаусса-Маркова (1 - 4) называется *классической (гауссовской) регрессионной моделью*. Если в предпосылки регрессионной модели включено условие нормальности распределения ошибки, то модель называется *классической нормальной регрессионной моделью.*

Таким образом, в гауссовской линейной регрессионной модели предполагается:

1) все опыты были проведены независимо друг от друга, т.е. случайности, вызвавшие отклонение отклика от закономерности в j - м опыте, не оказывали влияния на подобные отклонения в других опытах;

2) статистическая природа случайных составляющих оставалась неизменной во всех опытах.

Таким образом, справедлива

теорема (Гаусса-Маркова). Если регрессионная модель (4) удовлетворяет условиям (1 - 4), то оценки $\hat{a}$ и $\hat{b}$, полученные по методу наименьших квадратов, имеют наименьшую дисперсию в классе всех линейных (по y_j) несмещенных оценок.

5. МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

5.1. Классическая линейная модель множественной регрессии

Рассмотрим модель с m параметрами θ_i, линейно входящими в регрессионную функцию $\eta(\vec{x}, \vec{\theta})$. Эта модель является обобщением предыдущей модели. Здесь допускается возможность проведения *нескольких измерений в одной и той же точке*, а также наличие *нескольких факторов*, влияющих на отклик.

Уравнение регрессии, характеризующее функциональную зависимость среднего значения y от объясняющих переменных $\vec{x} = \left(x^{(1)}, x^{(2)}, ..., x^{(k)}\right)$, имеет вид

$$M\left(y_j \middle| \vec{x}_j\right) = \eta\left(\vec{x}_j; \vec{\theta}\right) = \vec{\theta}^T g\left(\vec{x}_j\right), \qquad j = 1, ..., n .$$

Здесь в левой части – значение исследуемой величины y_j при значениях контролируемых переменных, определяемых координатами вектора $\vec{x}$ (в j - й точке); $\eta\left(\vec{x}; \vec{\theta}\right)$ – функция регрессии, известная с точностью до конечного числа *линейно* входящих параметров; $g\left(\vec{x}\right) = \left(g_1\left(\vec{x}\right), g_2\left(\vec{x}\right), ..., g_m\left(\vec{x}\right)\right)^T$ – вектор известных базисных функций, которые предполагаются непрерывными на ограниченном замкнутом множестве $\mathfrak{X}$ в $\mathbb{R}^k$;

m – число неизвестных и подлежащих оценке параметров;

n – число *различных* точек наблюдений; k – размерность множества $\mathfrak{X}$.

Другими словами, имеем

$$\eta\left(\vec{x}; \vec{\theta}\right) = \theta_1 g_1\left(\vec{x}\right) + \theta_2 g_2\left(\vec{x}\right) + ... + \theta_m g_m\left(\vec{x}\right) = \theta^T g\left(\vec{x}\right),$$

где $\vec{x} = \left(x^{(1)}, x^{(2)}, ..., x^{(k)}\right) \in \mathfrak{X}$.

Относительно случайной ошибки ε_j $(j = 1, ..., n)$ предполагается выполнение следующих условий:

1. $M(\varepsilon_j) = 0$;

2. $M(\varepsilon_j \cdot \varepsilon_{j'}) = \sigma_j^{\ 2} \cdot \delta_{j,j'}$, где $\delta_{j,j'} = \begin{cases} 1 \text{ при } j = j' \\ 0 \text{ при } j \neq j' \end{cases}$ – символ Кронекера;

3. $\sigma_j^{\ 2} = \sigma^2 / r_j$, где r_j – число наблюдений в точке $\vec{x}_j$.

$N = \sum\limits_{j=1}^{n} r_j$ – общее число наблюдений.

В матричной форме линейная модель имеет вид

$$Y = G\vec{\theta} + \varepsilon ,$$

где $Y = (y_1, y_2, ..., y_n)^T$ – вектор результатов наблюдений функции регрессии $\eta(\vec{x};\vec{\theta})$; $\vec{\theta} = (\theta_1, \theta_2, .., \theta_m)^T$ – вектор параметров регрессионной функции; $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$ – вектор ошибок наблюдений, $M(Y) = G\vec{\theta}$ − математическое ожидание вектора Y, $G = \left(g_1(\vec{x}_j), g_2(\vec{x}_j), ..., g_m(\vec{x}_j)\right)_{j=1}^m$ − матрица значений базисных функций в точках наблюдений. Вектор Y имеет многомерное нормальное распределение, а ковариационная матрица вектора Y $D(Y) = \sigma^2 R^{-1}$:

$$Y \sim N(G\vec{\theta}\,; \sigma^2 R^{-1}), \qquad R = \begin{pmatrix} r_1 & 0 & ... & 0 \\ 0 & r_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & r_n \end{pmatrix}.$$

Если хотя бы одно из чисел $r_j > 1$, то $Y = (y_{1\text{ср}}, y_{2\text{ср}}, ..., y_{n\text{ср}})^T$, где $y_{j\text{ср}} = \dfrac{1}{r_j}\sum_{j'=1}^{r_j} y_{jj'}$ – среднее из значений отклика, вычисленных в точке $\vec{x}_j$.

Рассмотрим только модель полного ранга, в которой

$$\operatorname{rank} G = m \ \left(\Rightarrow \ \det(G^T G) \neq 0\right).$$

В этом случае, согласно теореме Гаусса-Маркова, вектор параметров $\vec{\theta}$ допускает наилучшую линейную оценку:

$$\hat{\vec{\theta}} = \left(G^T R G\right)^{-1} G^T R Y\,, \tag{5}$$

а ковариационная матрица оценок имеет вид

$$D(\hat{\vec{\theta}}) = \sigma^2 \left(G^T R G\right)^{-1}. \tag{6}$$

Несмещенной оценкой $\hat{\sigma}^2$ дисперсии σ^2 (в предположении адекватности модели) служит величина:

$$s^2 = \hat{\sigma}^2 = \Delta / (n - m), \tag{7}$$

где сумма квадратов остатков равна

$$\Delta = \left(Y - G\hat{\vec{\theta}}\right)^T R \left(Y - G\hat{\vec{\theta}}\right), \tag{8}$$

где $(n-m)$ − число степеней свободы.

Допущения о виде распределения ошибки ε :

$$\varepsilon \sim N(0; \sigma^2 R^{-1}) \quad \Rightarrow \quad Y \sim N\left(G\vec{\theta}; \sigma^2 R^{-1}\right).$$

5.2. Статистический анализ регрессионной модели

Об *адекватности* регрессионной модели в целом судят по качеству согласия предсказанных по ней значений с результатами наблюдений. *Значимость* отдельных коэффициентов регрессионной модели говорит о том, что их оценки получились отличными от нуля не только из-за вклада случайных ошибок, т.е. что соответствующие базисные функции нельзя исключить из предложенной регрессионной модели. Для исследования точности полученных оценок строят *доверительные интервалы* для параметров регрессии. Можно строить доверительные интервалы и для значений отклика – это так называемые *доверительные коридоры*; они покажут, насколько значение $\hat{Y}$, предсказанное по модели, отличается от своего математического ожидания η, т.е. от «истинной» величины отклика.

Индивидуальные доверительные интервалы для параметров θ_i $(i=1,\dots,m)$:

$$\left(\hat{\theta}_i - p_i \cdot s \cdot t_{\text{кр}}(\alpha; n-m); \hat{\theta}_i + p_i \cdot s \cdot t_{\text{кр}}(\alpha; n-m)\right), \tag{9}$$

где $p_i = \sqrt{\left[(G^T R G)^{-1}\right]_{ii}}$, $\left[(G^T R G)^{-1}\right]_{ii}$ – i-й диагональный элемент матрицы $(G^T R G)^{-1}$;

$$s = \sqrt{\left(Y - G\hat{\vec{\theta}}\right)^T R \left(Y - G\hat{\vec{\theta}}\right) / (n-m)}; \tag{10}$$

$t_{\text{кр}}(\alpha; n-m) = t_{n-m; 1-\frac{\alpha}{2}}$ – квантиль уровня $1-\dfrac{\alpha}{2}$ распределения Стьюдента t_{n-m}; $\hat{\theta}_i$ – наилучшая линейная оценка параметра θ_i.

Для ***проверки статистической значимости параметра*** θ_i достаточно проверить, накрывает ли доверительный интервал число 0: если да, то с вероятностью $1-\alpha$ параметр θ_i следует признать незначимым; если нет, то гипотезу о незначимости θ_i отвергаем (т.к. это способ проверки гипотезы $H_0: \theta_i = 0$).

Для функции регрессии $\eta(\vec{x}; \vec{\theta}) = \vec{\theta}^T g(\vec{x})$ $1-\alpha$ – ***доверительный коридор ошибок***:

$$\eta(\vec{x}; \vec{\theta}) \in \Big(g^T(\vec{x})\hat{\vec{\theta}} - s\sqrt{g^T(\vec{x})\left(G^T R G\right)^{-1} g(\vec{x})} \cdot t_{\text{кр}}(\alpha; n-m);$$

$$g^T(\vec{x})\hat{\vec{\theta}} + s\sqrt{g^T(\vec{x})\left(G^T R G\right)^{-1} g(\vec{x})} \cdot t_{\text{кр}}(\alpha; n-m)\Big). \tag{11}$$

По результатам расчетов строится изображение $1-\alpha$ – доверительного коридора ошибок (только для $k=1$).

Для ***проверки статистической гипотезы об адекватности регрессионной и истинной модели в целом*** разработан критерий Фишера:

$$F = \frac{s^2_{\text{ад}}}{s^2\{y\}}, \tag{12}$$

где дисперсия адекватности равна

$$s^2_{\text{ад}} = s^2 = \frac{\Delta}{n-m} = \frac{\left(Y - G\hat{\vec{\theta}}\right)^T R\left(Y - G\hat{\vec{\theta}}\right)}{n-m}, \tag{13}$$

а дисперсия воспроизводимости равна

$$s^2\{y\} = \frac{\sum\limits_{j=1}^{n}\sum\limits_{j'=1}^{r_j}\left(y_{jj'} - y_{j\,\text{ср}}\right)^2}{N-n}. \tag{14}$$

По таблице критических точек распределения Фишера-Снедекора F_{ν_1,ν_2} находится значение $F_{\text{кр}}(\alpha;\nu_1,\nu_2)$,

где $\nu_1 = n - m -$ число степеней свободы для числителя;

$\nu_2 = N - n -$ число степеней свободы для знаменателя.

Если $F_{\text{набл}} \le F_{\text{кр}}$, то регрессионную модель с вероятностью $1-\alpha$ можно считать адекватной; в противном случае гипотеза об адекватности отвергается.

6. МОДЕЛИРОВАНИЕ НОРМАЛЬНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Рассмотрим моделирование значений независимых нормальных случайных величин (с произвольными параметрами), исходя из так называемых таблиц равномерно распределенных случайных чисел, в которых даны значения независимых равномерно распределенных на отрезке [0;1] случайных величин.

Будем опираться на *центральную предельную теорему*, согласно которой, если случайная величина *X* представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то эта случайная величина *X* имеет распределение, близкое к нормальному. Как известно, для случайной величины *Y*, распределенной равномерно на отрезке [*a*;*b*], математическое ожидание и дисперсия равны соответственно:

$$M(Y) = \frac{a+b}{2} \quad , \quad D(Y) = \frac{(b-a)^2}{12}.$$

Пусть $R_j \ (j = 1, ..., L) -$ независимые равномерно распределенные на отрезке [0;1] случайные величины, тогда имеем

$$M(R_j) = \frac{0+1}{2} = \frac{1}{2}, \quad D(R_j) = \frac{(1-0)^2}{12} = \frac{1}{12}.$$

Для случайной величины $\sum_{j=1}^{L} R_j$ получим:

$$M\left(\sum_{j=1}^{L} R_j\right) = \sum_{j=1}^{L} M(R_j) = \frac{L}{2}; \quad D\left(\sum_{j=1}^{L} R_j\right) = \sum_{j=1}^{L} D\left(R_j\right) = \frac{L}{12}.$$

Следовательно, случайная величина $U_L = \dfrac{\sum_{j=1}^{L} R_j - \dfrac{L}{2}}{\sqrt{L/12}}$ имеет стандартное нормальное распределение при $L \to \infty$, а случайная величина $V_L = \sigma U_L + a -$ нормальное распределение с $M(V_L) = a$ и $D(V_L) = \sigma^2$.

Для практических целей достаточно взять *L*=12, тогда имеем

$$U = \sum_{j=1}^{12} R_j - 6\,. \qquad (15)$$

Например, выпишем из таблицы подряд 12 чисел: 0,10; 0,09; 0,73; 0,25; 0,33; 0,76; 0,52; 0,01; 0,35; 0,86; 0,34; 0,67; тогда $u_1 = (0{,}10 + 0{,}09 + 0{,}73 + 0{,}25 + 0{,}33 + 0{,}76 + 0{,}52 + 0{,}01 + 0{,}35 +$ $+0{,}86 + 0{,}34 + 0{,}67) - 6 = -0{,}9$ – значение стандартной нормальной случайной величины U_1; аналогично получим

$$u_2 = (0{,}35 + 0{,}48 + 0{,}76 +$$
$$+0{,}80 + 0{,}95 + 0{,}90 + 0{,}91 + 0{,}17 + 0{,}37 + 0{,}54 + 0{,}20 + 0{,}48) - 6 -$$

значение стандартной нормальной случайной величины U_2, причем случайные величины U_1 и U_2 независимы.

Для проверки результатов моделирования можно получить оценки: значение $\overline{u}_в = \frac{1}{12}\left(u_1 + u_2 + \ldots + u_{12}\right)$ должно быть близко к 0, $d_в = \frac{1}{12}\sum_{i=1}^{12}\left(u_i - \overline{u}_в\right)^2$ должно быть близко к 1.

7. ЧИСЛЕННЫЙ ПРИМЕР

Прежде, чем рассмотреть численный пример, который будет служить эталоном выполнения заданий студентами, отметим следующее. В реальных задачах, требующих применения аппарата регрессионного анализа, истинные значения параметров θ_i неизвестны (в противном случае нет необходимости строить оценки для них). Поэтому нельзя судить о том, насколько хорошие результаты дает рассмотренный метод. Для исследования качества оценок можно поставить искусственный контролируемый эксперимент:

1) зададим «истинные» значения параметров $\theta_{i\,\text{ист}}$, где i=1,…,m;

2) зададим *план эксперимента*, т.е. набор значений x_j (j=1,…,n) и числа наблюдений r_j в точках x_j;

3) используем процесс генерации случайных ошибок $\varepsilon_{jj'}$ $(j=1,...,n;\ j'=1,...,r_j)$;

4) вычислим $y_{jj'}=\eta\left(x_j;\theta_{i\,\text{ист}}\right)+\varepsilon_{jj'}$, сознательно испортив истинные значения отклика;

5) применим аппарат регрессионного анализа для получения оценок $\hat{\theta}_i$ параметров θ_i.

Этот эксперимент позволит сравнить $\hat{\theta}_i$ с $\theta_{i\,\text{ист}}$.

Рассмотрим пример.

I. Исходные данные.

$k=1$ (число независимых переменных); $\mathfrak{X}=[-1;1]$;

$m=2$ (число оцениваемых параметров);

$\eta(x;\theta)=\theta_1+\theta_2 x$; $\theta_{1\,\text{ист}}=1$, $\theta_{2\,\text{ист}}=2$; $n=5$ (число различных точек наблюдений).

План эксперимента:

x_j	-1	-0,5	0	0,5	1
r_j	1	2	3	2	1

$N=\sum_{j=1}^{n} r_j=1+2+3+2+1=9-$ общее число точек наблюдений с учетом повторений; $\sigma=0,7;\ \alpha=0,05$.

II. Моделирование результатов наблюдений.

Воспользовавшись таблицами равномерно распределенных случайных чисел (см. [1], прил. 9), получим:

$R_s^{(1)}$: 0,98; 0,52; 0,01; 0,77; 0,67; 0,14; 0,90; 0,56; 0,86; 0,07; 0,22; 0,10;

$R_s^{(2)}$: 0,94; 0,05; 0,58; 0,60; 0,97; 0,09; 0,34; 0,33; 0,11; 0,80; 0,50; 0,54;

$R_s^{(3)}$: 0,31; 0,39; 0,80; 0,82; 0,77; 0,32; 0,50; 0,72; 0,56; 0,82; 0,48; 0,29;

$R_s^{(4)}$: 0,40; 0,52; 0,42; 0,01; 0,83; 0,45; 0,29; 0,96; 0,34; 0,06; 0,28; 0,89;

$R_s^{(5)}$: 0,80; 0,83; 0,13; 0,74; 0,67; 0,00; 0,78; 0,18; 0,47; 0,54; 0,06; 0,10;

$R_s^{(6)}$: 0,88; 0,68; 0,54; 0,02; 0,00; 0,86; 0,50; 0,75; 0,84; 0,01; 0,36; 0,76;

$R_s^{(7)}$: 0,66; 0,79; 0,51; 0,90; 0,36; 0,47; 0,64; 0,93; 0,99; 0,59; 0,46; 0,73;

$R_s^{(8)}$: 0,48; 0,87; 0,51; 0,76; 0,49; 0,69; 0,91; 0,82; 0,60; 0,89; 0,28; 0,93;

$R_s^{(9)}$: 0,78; 0,56; 0,13; 0,68; 0,65; 0,48; 0,11; 0,76; 0,74; 0,17; 0,46; 0,85.

Согласно (15), получим:

$$u_1 = \sum_{s=1}^{12} R_s^{(1)} - 6 = (0{,}98+0{,}52+0{,}01+0{,}77+0{,}67+0{,}14+0{,}90+0{,}56+0{,}86+$$
$$+0{,}07+0{,}22+0{,}10) - 6 = 5{,}80 - 6 = -0{,}20;$$

Аналогично имеем: $u_2 = \sum_{s=1}^{12} R_s^{(2)} - 6 = 5{,}85 - 6 = -0{,}15;$

$u_3 = 6{,}78 - 6 = 0{,}78;$ $u_4 = -0{,}55;$ $u_5 = -0{,}70;$ $u_6 = 0{,}20;$

$u_7 = 2{,}03;$ $u_8 = 2{,}23;$ $u_9 = 0{,}37.$

Т.к. $r_1 = 1$, то $\varepsilon_{11} = \sigma u_1 = 0{,}7 \cdot (-0{,}20) = -0{,}14$.

Т.к. $r_2 = 2$, то $\varepsilon_{21} = \sigma u_2 = 0{,}7 \cdot (-0{,}15) = -0{,}105$; $\varepsilon_{22} = \sigma u_3 = 0{,}7 \cdot 0{,}78 = 0{,}546$.

Т.к. $r_3 = 3$, то $\varepsilon_{31} = \sigma u_4 = 0{,}7 \cdot (-0{,}55) = -0{,}385$; $\varepsilon_{32} = \sigma u_5 = 0{,}7 \cdot (-0{,}70) = -0{,}49$; $\varepsilon_{33} = \sigma u_6 = 0{,}7 \cdot 0{,}20 = 0{,}14$.

Т.к. $r_4 = 2$, то $\varepsilon_{41} = \sigma u_7 = 0,7 \cdot 2,03 = 1,421$; $\varepsilon_{42} = \sigma u_8 = 0,7 \cdot 2,23 = 1,561$.

Т.к. $r_5 = 1$, то $\varepsilon_{51} = \sigma u_9 = 0,7 \cdot 0,37 = 0,259$.

$\eta(x;\theta_{\text{ист}}) = 1 + 2x \;\Rightarrow\; \eta(-1;\theta_{\text{ист}}) = -1; \quad \eta(-0,5;\theta_{\text{ист}}) = 0;$

$\eta(0;\theta_{\text{ист}}) = 1; \quad \eta(0,5;\theta_{\text{ист}}) = 2; \quad \eta(1;\theta_{\text{ист}}) = 3.$

Тогда получим:

$$y_{1\,\text{ср}} = y_{11} = -1 + \varepsilon_{11} = -1,14;$$

$$\left.\begin{aligned} y_{21} &= 0 + \varepsilon_{21} = -0,105 \\ y_{22} &= 0 + \varepsilon_{22} = 0,546 \end{aligned}\right\} \Rightarrow \quad y_{2\,\text{ср}} = \frac{0,441}{2} = 0,2205;$$

$$\left.\begin{aligned} y_{31} &= 1 + \varepsilon_{31} = 0,615 \\ y_{32} &= 1 + \varepsilon_{32} = 0,510 \\ y_{33} &= 1 + \varepsilon_{33} = 1,440 \end{aligned}\right\} \Rightarrow \quad y_{3\,\text{ср}} = \frac{2,265}{3} = 0,755;$$

$$\left.\begin{aligned} y_{41} &= 2 + \varepsilon_{41} = 3,421 \\ y_{42} &= 2 + \varepsilon_{42} = 3,561 \end{aligned}\right\} \Rightarrow \quad y_{4\,\text{ср}} = \frac{6,982}{2} = 3,491;$$

$$y_{5\,\text{ср}} = y_{51} = 3 + \varepsilon_{51} = 3,259.$$

Таким образом, $Y = (-1,14;\; 0,2205;\; 0,755;\; 3,491;\; 3,259)^T$.

III. Вычисление оценок параметров

Наилучшие линейные оценки параметров $\hat{\theta}_1$ и $\hat{\theta}_2$ вычисляются по формуле (5) $\hat{\theta} = \left(G^T R G\right)^{-1} G^T R Y$, где

$$G=\left(g_1(x_j),g_2(x_j)\right)_{j=1}^{5}=\begin{pmatrix}1 & x_1\\ 1 & x_2\\ 1 & x_3\\ 1 & x_4\\ 1 & x_5\end{pmatrix}=\begin{pmatrix}1 & -1\\ 1 & -0,5\\ 1 & 0\\ 1 & 0,5\\ 1 & 1\end{pmatrix},$$

$$R=\begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 2 & 0 & 0 & 0\\ 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & 2 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}\Rightarrow \qquad G^T RG=$$

$$=\begin{pmatrix}1 & 1 & 1 & 1 & 1\\ -1 & -0,5 & 0 & 0,5 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 2 & 0 & 0 & 0\\ 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & 2 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & -1\\ 1 & -0,5\\ 1 & 0\\ 1 & 0,5\\ 1 & 1\end{pmatrix}=$$

$$=\begin{pmatrix}1 & 2 & 3 & 2 & 1\\ -1 & -1 & 0 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & -1\\ 1 & -0,5\\ 1 & 0\\ 1 & 0,5\\ 1 & 1\end{pmatrix}=\begin{pmatrix}9 & 0\\ 0 & 3\end{pmatrix}.$$

Следовательно, $\left(G^T RG\right)^{-1}=\begin{pmatrix}1/9 & 0\\ 0 & 1/3\end{pmatrix}$.

$$\hat{\theta} = \left(G^T RG\right)^{-1} G^T RY =$$

$$= \begin{pmatrix} 1/9 & 0 \\ 0 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ -1 & -1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1,14 \\ 0,2205 \\ 0,755 \\ 3,491 \\ 3,259 \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{9} & \frac{2}{9} & \frac{3}{9} & \frac{2}{9} & \frac{1}{9} \\ \frac{-1}{3} & \frac{-1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} -1,14 \\ 0,2205 \\ 0,755 \\ 3,491 \\ 3,259 \end{pmatrix} = \begin{pmatrix} 1,3119 \\ 2,5565 \end{pmatrix}.$$

Т.е. получим $\hat{\eta}(x) = 1,3119 + 2,5565x$.

Согласно (6), ковариационная матрица вектора оценок $\hat{\theta}$:

$$D\left(\hat{\theta}\right) = \sigma^2 \left(G^T RG\right)^{-1} = 0,49 \begin{pmatrix} 1/9 & 0 \\ 0 & 1/3 \end{pmatrix} = \begin{pmatrix} 0,054 & 0 \\ 0 & 0,163 \end{pmatrix}.$$

Для расчета несмещенной оценки дисперсии по формуле (7) вычислим значения:

$$Y - G\hat{\theta} = \begin{pmatrix} -1,14 \\ 0,2205 \\ 0,755 \\ 3,491 \\ 3,259 \end{pmatrix} - \begin{pmatrix} 1 & -1 \\ 1 & -0,5 \\ 1 & 0 \\ 1 & 0,5 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1,3119 \\ 2,5565 \end{pmatrix} =$$

$$
= \begin{pmatrix} -1{,}14 \\ 0{,}2205 \\ 0{,}755 \\ 3{,}491 \\ 3{,}259 \end{pmatrix} - \begin{pmatrix} -1{,}2446 \\ 0{,}0336 \\ 1{,}3119 \\ 2{,}5901 \\ 3{,}8684 \end{pmatrix} = \begin{pmatrix} 0{,}1046 \\ 0{,}1869 \\ -0{,}5569 \\ 0{,}9009 \\ -0{,}6094 \end{pmatrix};
$$

$$
\Delta = \left(Y - G\hat{\theta}\right)^T R\left(Y - G\hat{\theta}\right) =
$$

$$
= 0{,}1046^2 + 2 \cdot 0{,}1869^2 + 3 \cdot \left(-0{,}5569\right)^2 + 2 \cdot 0{,}9009^2 +
$$

$$
+ \left(-0{,}6094\right)^2 = 3{,}0058;
$$

$$
s^2 = \Delta / \left(n - m\right) = 3{,}0058 / 3 = 1{,}002; \;\; s = 1{,}001.
$$

IV. Статистический анализ уравнения регрессии

Построим доверительные интервалы для параметров регрессионной функции в соответствии с выражением (9).

Для рассмотренного примера имеем:

$$
\hat{\theta}_1 = 1{,}3119, \; s = 1{,}001, \; p_1 = \sqrt{1/9} = 1/3, \; t_{кр}\left(0{,}05;3\right) = 3{,}18.
$$

Таким образом, имеем следующий *доверительный интервал для* θ_1:

$$
\left(\hat{\theta}_1 - p_1 \cdot s \cdot t_{\text{кр}}\left(\alpha; n-m\right); \hat{\theta}_1 + p_1 \cdot s \cdot t_{\text{кр}}\left(\alpha; n-m\right)\right) =
$$

$$
= \left(1{,}3119 - 1{,}001 \cdot \frac{1}{3} \cdot 3{,}18; 1{,}3119 + 1{,}001 \cdot \frac{1}{3} \cdot 3{,}18\right) =
$$

$$
= \left(0{,}2508; 2{,}3730\right).
$$

$$
\hat{\theta}_2 = 2{,}5565, \; s = 1{,}001, \; p_2 = \sqrt{1/3} = 0{,}5774, \; t_{\text{кр}}\left(0{,}05;3\right) = 3{,}18.
$$

Т.е. *доверительный интервал для* θ_2 имеет вид:

$$\left(\hat{\theta}_2 - p_2 \cdot s \cdot t_{\text{кр}}(\alpha; n-m); \hat{\theta}_2 + p_2 \cdot s \cdot t_{\text{кр}}(\alpha; n-m)\right) =$$

$$= (2{,}5565 - 1{,}001 \cdot 0{,}5774 \cdot 3{,}18; 2{,}5565 + 1{,}001 \cdot 0{,}5774 \cdot 3{,}18) =$$

$$= (0{,}7185; 4{,}3945).$$

Отметим, для нахождения критических значений $t_{\text{кр}}$ можно воспользоваться таблицами из прил. 6 [1].

Проверим *гипотезу об адекватности регрессионной модели*, пользуясь критерием Фишера (12).

Вычислим значения:

$$s_{\text{ад}}^2 = s^2 = \frac{\Delta}{n-m} = 1{,}002 \text{ (см. выше)};$$

$$s^2\{y\} = \frac{\sum_{j=1}^{n} \sum_{j'=1}^{r_j} \left(y_{jj'} - y_{j\text{ср}}\right)^2}{N-n} = \frac{1}{9-5}\Big[0^2 + (-0{,}105 - 0{,}2205)^2 +$$

$$+ (0{,}546 - 0{,}2205)^2 + (0{,}615 - 0{,}755)^2 + (0{,}510 - 0{,}755)^2 +$$

$$+ (1{,}440 - 0{,}755)^2 + (3{,}421 - 3{,}491)^2 + (3{,}561 - 3{,}491)^2 + 0^2\Big] =$$

$$= 0{,}77055 / 4 = 0{,}1926;$$

$$F_{\text{набл}} = \frac{s_{\text{ад}}^2}{s^2\{y\}} = \frac{1{,}002}{0{,}1926} = 5{,}202.$$

По таблице критических точек распределения Фишера (прил. 7 [1]) находим:

$$F_{\text{кр}}(\alpha; \nu_1; \nu_2) = F_{\text{кр}}(\alpha; n-m; N-n) = F_{\text{кр}}(0{,}05; 3; 4) = 16{,}69.$$

Поскольку $F_{\text{набл}} < F_{\text{кр}}$, то с вероятностью $\gamma = 1-\alpha = 0,95$ можно утверждать, что модель адекватна.

Построим $1-\alpha = 0{,}95$-*доверительный коридор ошибок*.

В соответствии с формулой (11) нижняя граница этого коридора имеет вид

$$\hat{\eta}_{\text{н}}(x) = g^T(x)\hat{\theta} - s\sqrt{g^T(x)\left(G^T RG\right)^{-1} g(x)} \cdot t_{\text{кр}}(\alpha; n-m) =$$

$$= (1; x)\begin{pmatrix} 1{,}3119 \\ 2{,}5565 \end{pmatrix} - 1{,}001\sqrt{(1; x)\begin{pmatrix} 1/9 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}} \cdot t_{\text{кр}}(0{,}05; 3) =$$

$$= 1{,}3119 + 2{,}5565x - 1{,}001\sqrt{\frac{1}{9} + \frac{1}{3}x^2} \cdot 3{,}18 =$$

$$= 1{,}3119 + 2{,}5565x - 1{,}061\sqrt{1+3x^2}\ ,$$

а верхняя – вид

$$\hat{\eta}_{\text{в}}(x) = g^T(x)\hat{\theta} + s\sqrt{g^T(x)\left(G^T RG\right)^{-1} g(x)} \cdot t_{\text{кр}}(\alpha; n-m) =$$

$$= (1; x)\begin{pmatrix} 1{,}3119 \\ 2{,}5565 \end{pmatrix} + 1{,}001\sqrt{(1; x)\begin{pmatrix} 1/9 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}} \cdot t_{\text{кр}}(0{,}05; 3) =$$

$$= 1{,}3119 + 2{,}5565x + 1{,}061\sqrt{1+3x^2}\ .$$

По результатам расчетов на рис. 3 изображены функции $\hat{\eta}(x)$, $\hat{\eta}_{\text{н}}(x)$, $\hat{\eta}_{\text{в}}(x)$, а также отмечены крестиками средние значения результатов наблюдений, вычисленные в каждой точке плана эксперимента.

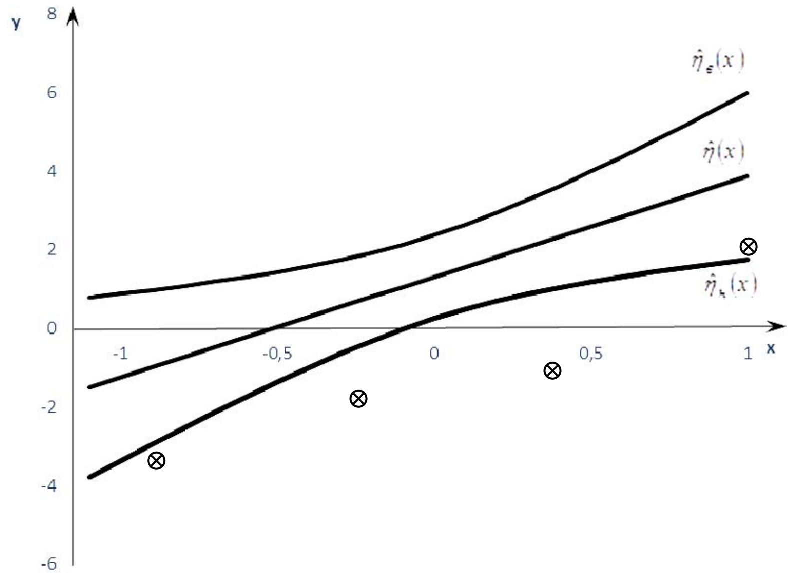

Рис.3. Доверительный коридор для функции регрессии

8. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

I. Содержание отчета.

1. Исходные данные, соответствующие взятому варианту:

– размерность области определения функции регрессии k;

– тип области определения функции регрессии $\mathfrak{X}$;

– базисные функции регрессии $\{g_i(x)\}_{i=1}^m$;

– линейно параметризованная модель функции регрессии

$\eta(x;\theta) = \theta_1 g_1(x) + \theta_2 g_2(x) + ... + \theta_m g_m(x) = \theta^T g(x),$ где

$\theta = (\theta_1, \theta_2, ..., \theta_m)^T$, $g(x) = (g_1(x), g_2(x), ..., g_m(x))^T$,

m – число неизвестных параметров;

– вектор истинных значений параметров $\theta = (\theta_1, \theta_2, ..., \theta_m)^T$;
– план эксперимента:

x_j	x_1	x_2	…	x_n
r_j	r_1	r_2	…	r_n

т.е. набор значений x_j (j=1,…,n) и числа наблюдений r_j в точках x_j ; $r_1 + r_2 + ... + r_m = N -$ общее число наблюдений;

– закон распределения ошибок ε_j и результатов наблюдений y_j в схеме Гаусса-Маркова:

$$y_j = \eta(x_j; \theta) + \varepsilon_j \ (j=1,\dots,n);$$

$$\varepsilon_j \sim N\left(0; \sigma_j^2\right), \quad \sigma_j^2 = \sigma^2 / r_j, \ \ y_j \sim N\left(\eta\left(x_j; \theta\right); \sigma_j^2\right).$$

– уровень значимости α при проверке статистических гипотез.

2. Практическая реализация задания, включающая следующие основные процедуры линейного регрессионного анализа:

1) построение наилучших линейных оценок $\hat{\theta}_i$ параметров θ_i регрессионной функции $\eta(x, \theta)$;

2) расчет ковариационной матрицы $D(\hat{\theta})$ оценок параметров и несмещенной оценки дисперсии $\hat{\sigma}^2$;

3) построение индивидуальных доверительных интервалов для параметров θ_i регрессионной модели;

4) проверка статистических гипотез о статистической значимости параметров θ_i при заданном уровне значимости α по критерию Стьюдента;

5) проверка статистической гипотезы об адекватности регрессионной модели $\eta(x; \theta)$ по критерию Фишера;

6) определение и графическое изображение (при $k=1$) $(1-\alpha)$-доверительного коридора ошибок для функции регрессии $\eta(x;\theta)$.

II. Варианты заданий.

1. Общие исходные данные для всех вариантов.

$N = 12$ – общее число наблюдеий;

σ^2 – дисперсия единичного наблюдения (задается преподавателем);

$\alpha = 0,05$ – уровень значимости;

$k = 1$ – размерность области определения $\mathfrak{X}$;

$\mathfrak{X} = [-1;1]$ – область определения функции регрессии $\eta(x;\theta)$;

$\eta(x;\theta) = \theta^T g(x) = \sum_{i=1}^{m} \theta_i g_i(x) = \sum_{i=1}^{m} \theta_i x^{i-1}$ – функция регрессии;

$\theta = (\theta_1, \theta_2, ..., \theta_m)^T$ – вектор неизвестных и подлежащих оценке параметров;

$\theta_{\text{ист}} = (\theta_{1\,\text{ист}}, \theta_{2\,\text{ист}}, ..., \theta_{m\,\text{ист}})^T$ – вектор истинных значений параметров (используется только при моделировании результатов наблюдений);

$m = 2$ – число параметров регрессионной функции.

2. Индивидуальные исходные данные.

<table>
<tr><th>№ варианта</th><th>n</th><th>$(\theta_{i\ \text{ист}})_{i=1}^{m}$</th><th><table><tr><td>$x_j$</td><td>$x_1$</td><td>$x_2$</td><td>…</td><td>$r_n$</td></tr><tr><td>$r_j$</td><td>$r_1$</td><td>$r_2$</td><td>…</td><td>$r_n$</td></tr></table></th></tr>
<tr><td>1</td><td>3</td><td>2,1
1,5</td><td><table><tr><td>-1</td><td>0</td><td>1</td></tr><tr><td>4</td><td>4</td><td>4</td></tr></table></td></tr>
<tr><td>2</td><td>3</td><td>4,2
-1,8</td><td><table><tr><td>-0,5</td><td>0</td><td>0,5</td></tr><tr><td>3</td><td>6</td><td>3</td></tr></table></td></tr>
<tr><td>3</td><td>4</td><td>0,4
3,2</td><td><table><tr><td>-1</td><td>-0,5</td><td>0,5</td><td>1</td></tr><tr><td>3</td><td>3</td><td>3</td><td>3</td></tr></table></td></tr>
<tr><td>4</td><td>4</td><td>2,7
-1,1</td><td><table><tr><td>-0,9</td><td>-0,2</td><td>0,2</td><td>0,9</td></tr><tr><td>5</td><td>1</td><td>1</td><td>5</td></tr></table></td></tr>
<tr><td>5</td><td>4</td><td>-1,7
2,8</td><td><table><tr><td>-0,7</td><td>-0,3</td><td>0,3</td><td>0,7</td></tr><tr><td>1</td><td>5</td><td>5</td><td>1</td></tr></table></td></tr>
<tr><td>6</td><td>4</td><td>2,0
0,8</td><td><table><tr><td>-3/4</td><td>-1/4</td><td>1/4</td><td>3/4</td></tr><tr><td>2</td><td>4</td><td>4</td><td>2</td></tr></table></td></tr>
<tr><td>7</td><td>3</td><td>-5,2
2,6</td><td><table><tr><td>-0,8</td><td>0</td><td>0,8</td></tr><tr><td>2</td><td>8</td><td>2</td></tr></table></td></tr>
<tr><td>8</td><td>4</td><td>3,6
-2,4</td><td><table><tr><td>-1</td><td>-0,25</td><td>0,25</td><td>1</td></tr><tr><td>3</td><td>3</td><td>3</td><td>3</td></tr></table></td></tr>
<tr><td>9</td><td>4</td><td>0,4
3,2</td><td><table><tr><td>-1</td><td>-0,4</td><td>0,4</td><td>1</td></tr><tr><td>4</td><td>2</td><td>2</td><td>4</td></tr></table></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th>№
варианта</th><th>n</th><th>$\left(\theta_{i\,\text{ист}}\right)_{i=1}^{m}$</th><th><table><tr><td>$x_j$</td><td>$x_1$</td><td>$x_2$</td><td>…</td><td>$r_n$</td></tr><tr><td>$r_j$</td><td>$r_1$</td><td>$r_2$</td><td>…</td><td>$r_n$</td></tr></table></th></tr>
<tr><td>10</td><td>4</td><td>-2,5
1,3</td><td><table><tr><td>-0,9</td><td>-0,4</td><td>0,4</td><td>0,9</td></tr><tr><td>1</td><td>5</td><td>5</td><td>1</td></tr></table></td></tr>
<tr><td>11</td><td>5</td><td>2,3
-1,6</td><td><table><tr><td>-3/4</td><td>-1/4</td><td>0</td><td>1/4</td><td>3/4</td></tr><tr><td>2</td><td>2</td><td>4</td><td>2</td><td>2</td></tr></table></td></tr>
<tr><td>12</td><td>4</td><td>-2,4
3,0</td><td><table><tr><td>-0,6</td><td>-0,3</td><td>0,3</td><td>0,6</td></tr><tr><td>1</td><td>5</td><td>5</td><td>1</td></tr></table></td></tr>
<tr><td>13</td><td>3</td><td>-1,8
4,2</td><td><table><tr><td>-0,5</td><td>0</td><td>0,5</td></tr><tr><td>5</td><td>2</td><td>5</td></tr></table></td></tr>
<tr><td>14</td><td>3</td><td>2,0
-1,6</td><td><table><tr><td>-0,8</td><td>0</td><td>0,8</td></tr><tr><td>4</td><td>4</td><td>4</td></tr></table></td></tr>
</table>

РЕКОМЕНДАТЕЛЬНЫЙ БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Гмурман В.Е.* Теория вероятностей и математическая статистика. – М.: Высшая школа, 2006.

2. *Себер Дж.* Линейный регрессионный анализ. – М.: Мир, 1980.

3. *Денисов В.И.* Математическое обеспечение системы ЭВМ-экспериментатор. – М.: Наука, 1977.

4. Статистические методы в инженерных исследованиях / Под ред. Г. К. Круга. – М: Высшая школа, 1983.

5. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.

6. *Доугерти К.* Введение в эконометрику. – М.: ИНФРА-М, 2004.

7. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс. – М.: Дело, 2001.

8. *Айвазян С.А., Мхитарян В.С.* Практикум по прикладной статистике и эконометрике. – М.: МЭСИ, 2003.

СОДЕРЖАНИЕ

СПЕЦИАЛЬНЫЕ ГЛАВЫ МАТЕМАТИКИ

РАСЧЕТ НАИЛУЧШИХ ЛИНЕЙНЫХ ОЦЕНОК КОЭФФИЦИЕНТОВ РЕГРЕССИИ И СТАТИСТИЧЕСКИЙ АНАЛИЗ УРАВНЕНИЯ РЕГРЕССИИ

Методические указания к самостоятельным работам для студентов магистратуры направления 27.04.01

Сост. *Л.М. Могилева, М.Б. Шабаева*

Печатается с оригинал-макета, подготовленного кафедрой высшей математики

Ответственный за выпуск *Л.М. Могилева*
Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 02.09.2021. Формат 60×84/16.
Усл. печ. л. 1,8. Усл.кр.-отт. 1,8. Уч.-изд.л. 1,5. Тираж 50 экз. Заказ 778.

Санкт-Петербургский горный университет
РИЦ Санкт-Петербургского горного университета
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2