

## Лабораторная работа № 3, 4

### 4. Регрессионный анализ

#### Теоретическая часть

##### 4.1. Основы регрессионного анализа

Регрессионный анализ является основным статистическим методом построения математических моделей объектов или явлений по экспериментальным данным. Эти модели связывают количественные переменные – результирующую и объясняющие. В различных областях используются разные названия этих переменных:

- выход, целевой, выходной или результирующий признак, отклик, результирующая, эндогенная или зависимая переменная;
- факторные, предсказывающие, экзогенные переменные, факторы, фактор-аргументы, предикторы и т.д.

Следует отметить, что определяемая в ходе анализа функция регрессии лишь формально устанавливает соответствие между переменными этих двух групп. В действительности они могут не состоять в причинно-следственных отношениях. Вследствие этого могут возникнуть так называемые нонсенс-регрессии (ложные, абсурдные). Следовательно, предварительно, на основе анализа проблемы, необходимо определить, какую из переменных рассматривать как результирующую, а какие – как объясняющие.

Рассмотрим общую схему регрессионного анализа. Пусть результирующая переменная  $Y$  связана с некоторыми объясняющими переменными  $x_1, \dots, x_k$  или в векторном виде  $\vec{x} = [x_1, \dots, x_k]^T$ . Связь является стохастической: значение  $y$  переменной  $Y$ , полученные в различных экспериментах при фиксированных значениях вектора  $\vec{x}$ , случайным образом колеблются вокруг некоторого неизвестного уровня  $\eta(\vec{x})$ :

$$Y \equiv Y(\vec{x}) = \eta(\vec{x}) + \varepsilon, \quad (1)$$

где  $\varepsilon$  определяет случайное отклонение результирующей переменной от величины  $\eta(\vec{x})$ . Случайные отклонения  $\varepsilon$  могут служить проявлением влияния неучтенных в векторе  $\vec{x}$  (и может быть, случайных) факторов, случайными ошибками измерения результирующей переменной и другими причинами, которые более подробно будут обсуждаться ниже. Среднее значение отклонений полагается равным нулю, поэтому математическое ожидание результирующей переменной совпадает со значением функции  $\eta(\vec{x})$ :

$$M\{Y(\vec{x})\} = \eta(\vec{x}) \quad (2)$$

Это уравнение называется *регрессией (уравнением регрессии)*, функция  $\eta(\vec{x})$  – *функцией регрессии*.

Существует большое количество типов регрессионных моделей, определяемых видом *функции регрессии*  $\eta(\vec{x})$ , которые, как правило, зависят не только от объясняющих переменных, но и от некоторых параметров  $\beta_1, \dots, \beta_m$ , которые удобно представить в виде векторов  $\vec{\beta} = [\beta_1, \dots, \beta_m]^T$ :

$$\eta(\vec{x}) = \eta(\vec{x}, \vec{\beta}). \quad (3)$$

Для таких функций регрессий задача их определения сводится к задаче оценки вектора параметров  $\vec{\beta}$  по экспериментальным данным. В зависимости от того, как эти параметры входят в функцию регрессии, модели делятся на *линейные* и *нелинейные* (по параметрам).

В основном результаты получены применительно к линейным регрессионным моделям, которые в общем виде записываются следующим образом:

$$Y(\vec{x}) = \eta(\vec{x}, \vec{\beta}) + \varepsilon = \sum_{j=1}^m \beta_j f_j(\vec{x}) + \varepsilon, \quad (4)$$

где  $f_j(\vec{x}) \equiv f_j(x_1, \dots, x_k)$  - некоторые известные функции объясняющих переменных, не включающие в себя неизвестные коэффициенты  $\beta_j$ . Функции  $f_j(\vec{x})$  называют *регрессорами*.

Эту модель можно представить в векторной форме:

$$Y \equiv Y(\vec{x}) = f^T \vec{\beta} + \varepsilon, \quad (5)$$

где  $f^T = [f_1(\vec{x}), \dots, f_m(\vec{x})]^T$  - вектор регрессоров. Эта модель описывает результат одного эксперимента.

Результаты  $n$  экспериментов можно представить в виде:

$$\begin{aligned} y_1 &\equiv y(\vec{x}_1) = f_1^T \vec{\beta} + \varepsilon_1; \\ &\dots \\ y_n &\equiv y(\vec{x}_n) = f_n^T \vec{\beta} + \varepsilon_n, \end{aligned} \quad (6)$$

где  $y_i$  - значение результирующей переменной  $Y$ , полученное в  $i$ -м эксперименте;  $\vec{x}_i = [x_{i1}, \dots, x_{im}]^T$  - вектор объясняющих переменных в  $i$ -м эксперименте;  $f^T = [f_1(x_i), \dots, f_m(x_i)]^T$  - вектор регрессоров;  $\varepsilon_i$  -  $i$ -ое значение случайного отклонения.

Предполагается, что параметры  $\beta_j$  неизменны при проведении всех экспериментов. Величина  $x_{ij}$  ( $i=1, n$ ;  $j=1, m$ ) представляет собой значение  $j$ -й переменной, полученной  $i$ -м эксперименте.

Выражение (6) в векторно-матричной форме запишется следующим образом:

$$\vec{y} = F\vec{\beta} + \vec{\varepsilon}, \quad (7)$$

где  $\vec{y} = [y_1, \dots, y_n]^T$  - вектор откликов;

$$F = \begin{bmatrix} f_1^T \\ \dots \\ f_n^T \end{bmatrix} = \begin{bmatrix} f_1(x_1) \dots f_m(x_1) \\ \dots \\ f_1(x_n) \dots f_m(x_n) \end{bmatrix} - \text{матрица регрессоров};$$

$\vec{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_n]^T$  - вектор отклонений;  $\eta = [\eta(x_1), \dots, \eta(x_n)]^T$  - вектор регрессии.

Основной задачей регрессионного анализа является получение оценок параметров регрессии  $(\hat{\beta}_1, \dots, \hat{\beta}_m)$ , которые были бы оптимальными в определенном смысле.

Рассмотрим задачу оценивания параметров  $\beta_j$  ( $j=1, m$ ) по результатам экспериментов  $y_i$  ( $i=1, n$ ). Для оценивания могут использоваться различные методы. В классическом регрессионном анализе используется *метод наименьших квадратов* (МНК). В этом случае оценки минимизируют сумму квадратов *невязок оценивания*. *Невязками оценивания*, которые также называются *остатками*, являются разности

$$e_i = y_i - \hat{y}_i,$$

а вектором невязок - вектор  $\vec{e} = \vec{y} - \vec{\hat{y}}$ .

Остатки вызываются двумя причинами:

- отличием вектора оценок  $\vec{\hat{\beta}}$  от вектора истинных параметров  $\vec{\beta}$ ;
- наличием случайных возмущений  $\varepsilon$ :

$$\vec{e} = F(\vec{\beta} - \vec{\hat{\beta}}) + \vec{\varepsilon}. \quad (8)$$

Сумма квадратов остатков

$$Q_0 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = (\vec{y} - \vec{\hat{y}})^T (\vec{y} - \vec{\hat{y}}) = \vec{e}^T \vec{e} \quad (9)$$

выступает в МНК в качестве критерия качества оценки. Оценкой метода наименьших квадратов или МНК-оценкой называют вектор  $\vec{\hat{\beta}}$ , минимизирующий функционал (9). Для нахождения оценки перепишем (9) в виде

$$Q_0 = Q(\vec{\hat{\beta}}) = (\vec{y} - F\vec{\hat{\beta}})^T (\vec{y} - F\vec{\hat{\beta}}) = \vec{y}^T \vec{y} - 2\vec{\hat{\beta}}^T F^T \vec{y} + \vec{\hat{\beta}}^T F^T F \vec{\hat{\beta}},$$

продифференцируем его по  $\vec{\hat{\beta}}$  и приравняем нулю с учетом  $(\vec{\hat{\beta}}^T F^T \vec{y})^T = \vec{y}^T \vec{y} \vec{\hat{\beta}}$ . В результате получим уравнение для вычисления оценки:

$$\partial Q_0 / \partial \vec{\hat{\beta}} = -2F^T \vec{y} + 2F^T F \vec{\hat{\beta}} = 0.$$

Это уравнение можно преобразовать к виду

$$F^T F \vec{\hat{\beta}} = F^T \vec{y}. \quad (10)$$

Эта система линейных алгебраических уравнений (СЛАУ) является основой для получения МНК-оценок параметров регрессии и называется *системой нормальных уравнений*.

Поскольку в классическом регрессионном анализе полагают, что ранг матрицы регрессоров равен числу параметров функции регрессии:

- $rank F = m$  (т.е. имеет полный ранг), то матрица  $(F^T F)$  невырожденная и, следовательно, имеет обратную матрицу  $(F^T F)^{-1}$ .

Умножив правую и левую части системы (10) на  $(F^T F)^{-1}$ , получим явное выражение для МНК-оценки вектора параметров:

$$\vec{\hat{\beta}} = (F^T F)^{-1} F^T \vec{y}, \quad (11)$$

откуда получается выражение для оценки вектора регрессии:

$$\vec{\hat{\eta}} = F \vec{\hat{\beta}} = F (F^T F)^{-1} F^T \vec{y}. \quad (12)$$

Матрица  $G = F^T F$  называется *информационной*, а ей обратная  $S = (F^T F)^{-1}$  - *матрицей ошибок*.

## 4.2. Статистический анализ качества регрессионной модели

Полученные МНК-оценки коэффициентов регрессии (11) обеспечивают высокое качество полученной модели только при условии, что ее структура  $F\vec{\hat{\beta}}$  соответствует структуре истинной зависимости  $\eta_0 = F_0 \vec{\beta}_0$  между математическим ожиданием отклика и факторами. Однако на практике отсутствует априорная информация о структуре истинной модели и исследователь вынужден поочередно рассматривать различные виды регрессионных моделей и останавливаться на той, которая согласуется с экспериментальными данными. Такую модель называют *адекватной*. Она должна удовлетворять условию

$$M\{\vec{y}\} = F\vec{\beta}. \quad (13)$$

Адекватная модель не обязательно должна совпадать с истинной. Более того, адекватная модель не единственна – с помощью произвольного неособенного линейного преобразования от модели  $F\vec{\beta}$  можно перейти к другой адекватной модели  $F^* \vec{\beta}^*$ :  $F^* = FR$ ,  $\vec{\beta}^* = R^{-1} \vec{\beta}$  (здесь  $R$  – невырожденная матрица). Однако общим для всех адекватных моделей является то, что для каждой из них существует неособенное линейное преобразование  $R_F$ , приводящее ее к истинной модели:

$$R_F : F^* R_F = F_0; \quad R_F^{-1} \vec{\beta}^* = \vec{\beta}_0. \quad (14)$$

Ошибки в выборе структуры модели часто проявляются в следующих случаях:

- оцениваемая модель содержит больше параметров, чем истинная (так называемый *перебор параметров*);
- проверяемая модель содержит меньше параметров, чем истинная (*недобор параметров*).

Не приводя доказательств, отметим, что при недоборе параметров регрессии полученная оценка в общем случае является смещенной, т.е. даже компоненты вектора параметров, которые оцениваются, определяются с систематической погрешностью  $\Delta\beta$ .

При переборе параметров регрессии оценка  $\tilde{\beta}$  является несмещенной, т.е.  $M\{\tilde{\beta}\} = \beta_0$ . В то же время точность оценок при переборе параметров теряется.

Таким образом, следует отметить следующее:

- недобор параметров является более серьезным недостатком регрессионной модели, поскольку при этом оценки параметров регрессии получаются смещенными и не являются несостоятельными;
- при чрезмерном усложнении модели (переборе параметров) снижается эффективность оценивания.

При проверке адекватности модели традиционно последовательно исследуются два аспекта проблемы:

- соответствие выбранного класса функций регрессии истинной (эта подзадача часто также называется *проверкой адекватности*);
- проверка гипотезы о значимости параметров регрессии.

Для решения первой задачи используются методы дисперсионного анализа. Возможны несколько вариантов ее решения. Рассмотрим три возможных варианта его реализации.

Нулевая гипотеза  $H_0$  состоит в выполнении равенства  $F\tilde{\beta} = M\{\tilde{y}\}$ . Альтернативная гипотеза  $H_1$ :  $F\tilde{\beta} \neq M\{\tilde{y}\}$ . Гипотеза  $H_0$  проверяется при заданном уровне значимости  $\alpha$ . Вычисляется оценка дисперсии возмущений  $s^2$ :

$$s^2 = Q_0 / (n - m), \quad (15)$$

где  $Q_0 = (\tilde{y} - F\tilde{\beta})^T (\tilde{y} - F\tilde{\beta}) = \tilde{y}^T \tilde{y} - \tilde{\beta}^T F^T \tilde{y}$  - остаточная сумма квадратов, характеризующая разброс экспериментальных данных относительно оцененной функции регрессии.

В первом варианте проверки гипотезы используется известное априорное (точное) значение дисперсии  $\sigma_\varepsilon^2$ . Статистика

$$g = s^2 / \sigma_\varepsilon^2 \quad (16)$$

при справедливости нулевой гипотезы имеет  $\chi^2$ -распределение с  $\nu = n - m$  степенями свободы. Поэтому гипотеза  $H_0$  принимается, если

$$\chi_{\nu, \alpha/2}^2 \leq g \leq \chi_{\nu, 1-\alpha/2}^2,$$

в противном случае она отклоняется.

Во втором варианте проверки гипотезы полагается, что на основании  $N$  дополнительных экспериментов  $\{y'_1, \dots, y'_N\}$  при фиксированных значениях факторов определена независимая оценка дисперсии

$$s_1^2 = \left[ \sum_{i=1}^N \left( y'_i - \left( \sum_{j=1}^N y'_j \right) / n \right)^2 \right] / (N - 1).$$

При справедливости нулевой гипотезы статистика  $g = s^2/s_1^2$  имеет  $F$ -распределение с  $\nu_1 = n - m$  и  $\nu_2 = N - 1$  степенями свободы. Поэтому  $H_0$  принимается, если

$$g \leq F_{\nu_1, \nu_2, 1-\alpha/2},$$

в противном случае она отклоняется. Полагается, что  $s^2 \geq s_1^2$ , т.е.  $g \geq 1$ . В противном случае отношение изменяется на обратное:  $g = s_1^2/s^2$  и соответственно изменяется квантиль  $F$ -распределения:  $\nu_1 = N - 1$  и  $\nu_2 = n - m$ .

В третьем варианте проверки гипотезы вычисляется дисперсия случайной величины  $Y$  по результатам основного регрессионного эксперимента:

$$\hat{\sigma}_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{Y})^2,$$

где  $\bar{Y} = \left( \sum_{i=1}^n y_i \right) / n$ .

При справедливости нулевой гипотезы отношение  $g = s^2/\hat{\sigma}_Y^2$  имеет  $F$ -распределение с  $\nu_1 = n - m$  и  $\nu_2 = n - 1$  степенями свободы. Поэтому она принимается, если  $g \leq F_{\nu_1, \nu_2, 1-\alpha/2}$ . Здесь также предполагается, что  $g \geq 1$ .

Проверка гипотезы о значимости параметров регрессии (гипотезы о равенстве нулю соответствующего параметра) позволяет убрать из модели “лишние” параметры, что приводит к повышению точности оценки остальных. Рассматриваемые методы позволяют проверить, не отличаются ли полученные оценки коэффициентов от нуля только из-за случайных возмущений.

Проверка основывается на том, что, как отмечалось, оценки  $\hat{\beta}_i$  имеют гауссовское распределение с математическим ожиданием  $\beta_i$  и дисперсией  $D\{\beta_i\} = \sigma_\varepsilon^2 s_{ii}$ , где  $s_{ii}$  - элементы матрицы ошибок  $S$ . Оценкой дисперсии  $D\{\beta_i\}$  является величина  $s^2 s_{ii}$ .

Нулевая гипотеза  $H_0$ :  $\beta_i = 0$ . Альтернативная гипотеза  $H_1$ :  $\beta_i \neq 0$ . Проверка производится при заданном уровне вероятности ошибочного признания значимым коэффициента, в действительности равного нулю. Используется статистика  $g = \hat{\beta}_i / \sqrt{s^2 s_{ii}}$ , которая при справедливости гипотезы  $H_0$  имеет  $t$ -распределение Стьюдента с  $\nu_1 = n - m$  степенями свободы.

Гипотеза  $H_0$  принимается, если  $|g| \leq t_{\nu, 1-\alpha/2}$ , в противном случае коэффициент  $\beta_i$  считается значимым.

Незначимые коэффициенты целесообразно отбросить и вновь решить регрессионную задачу. Однако после этого необходимо снова проверить гипотезу об адекватности новой модели. Если она не подтвердила адекватности, то следует вернуться к прежней модели.

## Практическая часть

### Задание

#### 4.3. Полиномиальная регрессия

Оценить параметры полиномиальной регрессии по экспериментальным данным, приведенным в таблице 1. Вариант получить у преподавателя.

Полиномиальная регрессия представляется в виде:

$$Y(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_m X^m$$

**I. На первом шаге** представить функциональную связь между экспериментальными данными в виде полиномиальной регрессионной модели со степенью  $m=1$ .

**II. Вычислить оценки параметров регрессии  $\tilde{\beta}$  методом наименьших квадратов:** составить систему нормальных уравнений, решить ее методом Гаусса с полным упорядочением.

**III. Проверить гипотезу об адекватности выбранной модели:**

**1) Вычислить оценку дисперсии возмущений**

$s^2 = \frac{Q_0}{n-m}$ , где  $Q_0 = (y - F\tilde{\beta})^T (y - F\tilde{\beta})$  – остаточная сумма квадратов, характеризующая разброс экспериментальных данных относительно оцененной функции регрессии,  $F$  – матрица регрессоров,  $\tilde{\beta}$  – вектор оценки параметров регрессии, полученный при решении системы уравнений.

**2) Вычислить дисперсию случайной величины  $Y$  по результатам основного регрессионного эксперимента:**

$$\hat{\sigma}_Y^2 = \frac{1}{n-1} \sum (y_i - \bar{Y})^2, \text{ где } \bar{Y} = \frac{\sum_{i=1}^n y_i}{n};$$

**3) Вычислить статистику  $g = \frac{s^2}{\hat{\sigma}_Y^2}$  и если  $g \leq F_{v1, v2, 1-\alpha/2}$ , то гипотеза об адекватности модели принимается.**

$F_{v1, v2, 1-\alpha/2}$  – квантиль  $F$ -распределения Фишера с  $v1 = n-m$  и  $v2 = n-1$  степенями свободы и уровнем значимости, равным  $\alpha/2$  ( $q = 1 - \alpha/2$ );

**IV. Если гипотеза об адекватности не принимается,** то необходимо увеличить степень полинома на единицу ( $m=m+1$ ) и повторять пункты **II**, **III** и **IV** до тех пор, пока гипотеза об адекватности модели не станет справедливой (до  $m \leq n-2$ ).

**4.3.1. Реализация алгоритма LU-разложения методом Гаусса с полным упорядочением:**

Развернутая форма представления системы линейных алгебраических уравнений (СЛАУ)  $Ax = b$ :

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

При использовании методов  $LU$ -разложения матрица  $A$  представляется в виде произведения нижней треугольной матрицы  $L$  и верхней треугольной матрицы  $U$ :  $A = LU$ . алгоритмы различаются способом определения матриц  $L$  и  $U$  по матрице  $A$ .

Среди методов  $LU$ -разложения известен метод Гаусса. Его идея состоит в определении матрицы  $C = L^{-1}$  последовательно за  $(m-1)$  шагов вычислений с тем, чтобы исходную систему привести к эквивалентной системе с верхней треугольной матрицей:

$$CAx = Cb \Rightarrow L^{-1}LUx = Cb \Rightarrow Ux = b'$$

Этап перехода от исходной системы к эквивалентной называется прямым ходом (прямым исключением).

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1m} \\ 0 & u_{22} & u_{23} & \dots & u_{2m} \\ 0 & 0 & u_{33} & \dots & u_{3m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & u_{mm} \end{bmatrix} * \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b'_1 \\ \vdots \\ \vdots \\ \vdots \\ b'_m \end{bmatrix}$$

Тогда, последнее уравнение  $u_{mm} * x_m = b'_m$  содержит единственный неизвестный аргумент  $x_m$  и может быть решено относительно него:

$$x_m = b'_m / u_{mm}$$

Из предпоследнего уравнения можно определить  $(m-1)$ -й неизвестный аргумент:

$$x_{m-1} = \frac{b'_{m-1} - u_{(m-1)m} * x_m}{u_{(m-1)(m-1)}}$$

Таким образом, на каждой итерации можно определять из очередного уравнения значение очередной неизвестной:

$$x_i = \frac{b' - \sum_{j=i+1}^m (u_{ij} * x_j)}{u_{ii}}, \text{ где } i = \overline{m-1, 1}$$

Эта процедура называется обратным ходом (обратной подстановкой).

Алгоритм  $LU$ -разложения методом Гаусса с полным упорядочением является модификацией метода Гаусса. Для пояснения сути этого метода используем расширенную матрицу системы:

$$A = \begin{bmatrix} a_{11} \dots a_{1q} & \dots & a_{1m} & b'_1 \\ \dots & \dots & \dots & \dots \\ a_{p1} \dots a_{pq} & \dots & \dots & b'_p \\ \dots & \dots & \dots & \dots \\ a_{m1} \dots \dots & \dots & a_{mm} & b'_m \end{bmatrix}$$

1. Среди элементов матрицы  $a_{ij}$  выбирается наибольший по модулю, называемый главным элементом. Пусть главным элементом матрицы является элемент  $a_{pq}$ . Строка с номером  $p$ , содержащая главный элемент, называется главной строкой, а соответствующее уравнение системы - главным уравнением.

2. Вычисляются множители  $M_i = \frac{a_{iq}}{a_{pq}}$  для всех  $i \neq p$ . Затем матрица преобразуется следующим образом: из каждой  $i$ -й строки вычитается главная строка, умноженная на  $M_i$ . В результате получается матрица, у которой все элементы  $q$ -го столбца, кроме  $a_{pq}$ , равны нулю.

3. Отбрасывая этот столбец и главную строку, получают новую матрицу **A1** с меньшим на единицу числом строк и столбцов.

4. Над матрицей **A1** повторяются те же операции (см. п. 1-3) и получают матрицу **A2**.

5. Пункты 1-4 повторяются до тех пор, пока не получится матрица-строка, содержащая коэффициент перед последним неизвестным и свободный член уравнения.

6. Все главные строки, начиная с последней, объединяются.

7. Следует изменить порядок следования неизвестных в векторе  $x$  так, чтобы матрица системы стала треугольной.

8. Определить значения элементов вектора  $x$ .

#### 4.4. Линейная регрессия

Случайная величина  $Y$  зависит от ряда явлений, характеризуемых признаками  $X1, X2, X3, \dots, X_m$ . Каждый из этих признаков описывается рядом значений (см. таблицу 2).

Оценить параметры линейной регрессии  $\beta_i$  ( $i = 1, \dots, m$ ) по результатам экспериментов, приведенных в таблице.

**1. Линейная регрессионная модель представляется в виде:**

$$Y_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_m X_{tm} + \varepsilon_t,$$

где  $t = 1, \dots, n$  - номер наблюдения;  
 $\beta_1, \beta_2, \dots, \beta_m$  - параметры, которые необходимо оценить;  
 $X_{t1}, \dots, X_{tm}$  - значение объясняющих переменных (факторов) для  $t$ -го наблюдения;  
 $\varepsilon_t$  - случайное отклонение.

**2. Вычислить оценки параметров регрессии  $\hat{\beta}$ :**

- составить систему нормальных уравнений;
- решить систему линейных уравнений методом Гаусса с полным упорядочением.

**3. Проверить гипотезу об адекватности выбранной модели:**

(см. предыдущую лаб. работу).



#### 4. Проверить гипотезу о значимости параметров регрессии:

- вычислить статистику  $g = \frac{\hat{\beta}_i}{\sqrt{s^2 s_{ii}}}$ , где  $s_{ii}$  - элементы матрицы ошибок **S**;
- вычислитель квантиль t-распределения с  $\nu = n - m$  степенями свободы при заданном уровне вероятности  $p = 1 - \alpha / 2$ ;
- гипотеза  $H_0$ , что  $\beta_i = 0$  принимается, если  $|g| \leq t_{\nu, 1-\alpha/2}$ , в противном случае  $\beta_i$  считается значимым.

#### 5. Незначимые параметры следует отбросить и вновь решить регрессионную задачу (см. пункт 2), а затем следует выполнить пункт 3.

Если гипотеза об адекватности не подтверждается, то следует вернуться к предыдущей модели.

#### 6. Пункты 4 и 5 повторять, пока все параметры не будут просмотрены на значимость.

Таблица 1

I		II		III		IV		V	
x	y	x	y	x	y	x	y	x	y
1.0	7.190	0.00	4.588	-5.0	17.663	2.0	0.190	0.000	2.994
1.2	7.466	0.25	4.197	-4.5	16.206	2.2	0.228	0.025	3.004
1.4	7.912	0.50	4.129	-4.0	16.360	2.4	0.246	0.050	3.012
1.6	8.495	0.75	3.066	-3.5	15.772	2.6	0.269	0.075	3.018
1.8	8.727	1.00	3.111	-3.0	14.087	2.8	0.300	0.100	3.017
2.0	8.842	1.25	2.305	-2.5	13.684	3.0	0.279	0.125	3.026
2.2	9.673	1.50	2.348	-2.0	13.080	3.2	0.334	0.150	3.033
2.4	9.622	1.75	1.414	-1.5	11.589	3.4	0.345	0.175	3.034
2.6	10.090	2.00	0.926	-1.0	11.050	3.6	0.365	0.200	3.037
2.8	10.744	2.25	0.659	-0.5	10.815	3.8	0.387	0.225	3.047
3.0	10.988	2.50	0.067	0.0	9.802	4.0	0.414	0.250	3.048
3.2	11.028	2.75	-0.818	0.5	8.603	4.2	0.441	0.275	3.055
3.4	11.897	3.00	-1.014	1.0	8.666	4.4	0.438	0.300	3.059
3.6	12.311	3.25	-1.389	1.5	7.927	4.6	0.456	0.325	3.065
3.8	12.737	3.50	-1.748	2.0	7.518	4.8	0.471	0.350	3.067
4.0	13.054	3.75	-2.831	2.5	5.723	5.0	0.513	0.375	3.076
4.2	13.166	4.00	-2.854	3.0	5.009	5.2	0.528	0.400	3.081
4.4	13.956	4.25	-3.215	3.5	5.112	5.4	0.562	0.425	3.080
4.6	13.949	4.50	-3.981	4.0	3.986	5.6	0.573	0.450	3.092
4.8	14.562	4.75	-4.299	4.5	2.822	5.8	0.584	0.475	3.096
5.0	15.312	5.00	-4.843	5.0	2.519	6.0	0.597	0.500	3.105

Таблица 1 (продолжение)

VI		VII		VIII		IX		X	
<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
10	61.53	0.00	0.539	2.50	3.538	0.50	4.178	0.000	1.525
15	70.80	0.75	1.758	3.25	4.982	0.55	4.181	0.075	1.448
20	71.05	1.50	3.557	4.00	6.101	0.60	4.497	0.150	1.301
25	86.21	2.25	5.843	4.75	7.269	0.65	4.779	0.225	1.204
30	86.54	3.00	6.541	5.50	9.526	0.70	4.962	0.300	1.090
35	91.95	3.75	10.687	6.25	9.226	0.75	5.402	0.375	1.024
40	102.44	4.50	10.866	7.00	11.117	0.80	5.541	0.450	0.894
45	109.23	5.25	12.488	7.75	11.589	0.85	5.864	0.525	0.650
50	115.56	6.00	15.414	8.50	13.847	0.90	5.881	0.600	0.609
55	125.26	6.75	17.420	9.25	13.752	0.95	6.303	0.675	0.414
60	126.35	7.50	17.010	10.00	16.087	1.00	6.346	0.750	0.363
65	138.10	8.25	21.069	10.75	16.989	1.05	6.721	0.825	0.279
70	140.17	9.00	22.861	11.50	17.316	1.10	7.062	0.900	0.065
75	149.14	9.75	23.871	12.25	19.670	1.15	7.132	0.975	0.085
80	156.95	10.50	24.990	13.00	19.849	1.20	7.566	1.050	-0.055
85	161.53	11.25	28.548	13.75	21.478	1.25	7.702	1.125	-0.210
90	166.69	12.00	30.679	14.50	22.547	1.30	7.841	1.200	-0.299
95	172.72	12.75	31.546	15.25	23.634	1.35	8.295	1.275	-0.285
100	179.69	13.50	32.900	16.00	24.541	1.40	8.568	1.350	-0.561
105	184.44	14.25	34.477	16.75	25.110	1.45	8.818	1.425	-0.612
110	193.05	15.00	36.757	17.50	26.950	1.50	8.963	1.500	-0.781

Таблица 2

Номер варианта	Переменные	Номер варианта	Переменные	Номер варианта	Переменные
1	X1,X2,X4,X5	6	X1,X2,X6,X10	11	X2,X3,X10,X11
2	X1,X2,X6,X7	7	X1,X2,X4,X7	12	X2,X3,X4,X10
3	X1,X2,X8,X9	8	X2,X3,X4,X5	13	X2,X3,X6,X11
4	X1,X2,X10,X11	9	X2,X3,X6,X7	14	X2,X3,X6,X8
5	X1,X2,X3,X8	10	X2,X3,X8,X9	15	X1,X3,X5,X9

X1	X2	X3	X4	X5	X6
55	3,9	30	5	28	124
100	2,6	47	8, 2	121	87
93	5,3	37	12	146	74
20	4,1	12,4	7,9	52	141
20	3,7	4,3	6,5	72	134
72	3,6	28	5,4	38	120
85	6,9	48	11	83	72
65	3	18	9,5	92	156
67	3,5	39	8,8	91	91
73	1,7	40	10,9	73.	106
88	6,8	35	8,1	138	73
83	1	24	8,8	99	108
21	3,8	36	9,8	55	140
98	5	38	10,3	89	77

99	3,3	31	9,6	87	102
89	0,4	26	8,95	103	72
84	2,2	27	9,6	169	118
61	4,2	19,2	7,2	10	191
98	3,1	44	7,4	123	77
46	4,1	23,5	6,7	20	134

X7	X8	X9	X10	X11
44,5	84,98	20,4	3,2	14,4
32,5	30,58	71,4	8, 5	11,6
33,9	38,42	78,7	9,2	56,1
38,8	60,34	12,1	3,3	16,4
34,4	60,22	10,9	3,2	13,5
43,6	60,79	20,4	5,4	22,4
41	29,82	79,7	8,3	65,5
36,4	70,57	17,3	5,4	27,8
17,9	34,51	69,7	7,1	62,3
32,1	64,73	24,5	6	39,8
38,1	36,63	76,2	8,6	56,9
41,5	32,84	44,4	5,7	37,4
55	62,64	11,3	3,5	18,6
36,7	34,07	79,2	6,7	54,4
15,8	39,27	57	6,7	64,2
40,9	28,46	54,8	7,3	22,6
49,4	30,27	72,1	8,5	46
38,1	69,04	13,4	3,3	7,9
27,6	25,42	79,9	10,2	25,4
33,2	53,13	11,2	3,4	17