Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

ПОСТРОЕНИЕ И ИССЛЕДОВАНИЕ КЛАССИЧЕСКОЙ ЛИНЕЙНОЙ МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ В ППП STATISTICA

Под редакцией А.Г. Реннера

Рекомендовано к изданию Редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет» в качестве методических указаний для студентов, обучающихся по программам высшего профессионального образования по направлениям подготовки 231300 «Прикладная математика», 080500 «Бизнес-информатика», 080100 «Экономика» (общий профиль), специальности 080016 «Математические методы в экономике» и другим специальностям и направлениям подготовки

Оренбург
2011

УДК 330.4 (076)
ББК 65В631Я7
В 20

Рецензент – доцент, кандидат экономических наук Т.В. Леушина

Васянина, В.И.

В 20 Построение и исследование классической линейной модели множественной регрессии в ППП Statistica: методические указания / В.И. Васянина, Ю.А. Жемчужникова, О.И. Стебунова; под ред. А.Г. Реннера; Оренбургский гос. ун-т.– Оренбург: ОГУ. – 38 с.

Методические указания к семинарским занятиям, лабораторному практикуму, самостоятельной работе студентов, в том числе для выполнения расчетно-графических заданий, курсовых и дипломных работ, предполагающих проведение регрессионного анализа. Предназначены для студентов специальности 080016 – «Математические методы в экономике», направлений подготовки 231300 – «Прикладная математика», 080500 - «Бизнес-информатика», 080100 – «Экономика» (общий профиль) и других специальностей и направлений.

УДК 330.4 (076)
ББК 65В631Я7

Содержание

Введение

Принципиальная невозможность учесть влияние всех факторов, оказывающих влияние на исследуемую результативную переменную, в областях исследования с высокой степенью неопределенности, в частности, в экономике, приводит к необходимости моделировать зависимость последней от объясняющих переменных в форме функции регрессии.

Цель методических указаний – способствовать освоению таких категорий, как регрессионная зависимость (функция регрессии), модель регрессии и регрессионные остатки, способствовать закреплению знаний, умений, приобретению навыков реализации алгоритма оценки и исследования классической линейной модели множественной регрессии и содержательной интерпретации результатов.

1 Теоретическая часть

1.1 Постановка задачи регрессионного анализа

Ставится задача построения и исследования зависимости результирующего признака y от объясняющих переменных $x_0 = 1, x_1, x_2, ..., x_k$ на основе результатов наблюдений признаков на “n” объектах $O_1, O_2, ...O_n$, $n >> k$.

Результаты наблюдений результирующего признака и объясняющих переменных представлены вектором $Y_{n\times 1} = (y_1 \quad y_2 \quad ... \quad y_n)^T$ и матрицей X типа «объект-свойство»:

$$X_{n\times k+1} = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ ... & ... & ... & ... & ... \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix} = \{x_{ij}\}_{\substack{i=\overline{1,n} \\ j=\overline{0,k}}}$$

где y_i – наблюденное значение результативного признака для i-го объекта;
x_{ij} – значение j-го признака на i-м объекте наблюдения $i = \overline{1,n}$, $j = \overline{0,k}$; столбец из "1" можно считать столбцом "наблюденных" значений для признака $x_0 = 1$.

Поясним необходимость построения “специальной”, а именно, регрессионной зависимости следующим, несколько идеализированным способом: поскольку при анализе данных, характеризующих социально-экономические объекты, мы зачастую сталкиваемся с ситуацией (проиллюстрированной в двумерном случае на рисунке 1.1) состоящей в том, что объясняющие переменные для различных объектов могут принимать совпадающие значения, а значения результирующего признака для этих объектов различны, то не имеет смысла говорить о функциональной зависимости y от $x_1, x_2, ..., x_k$.

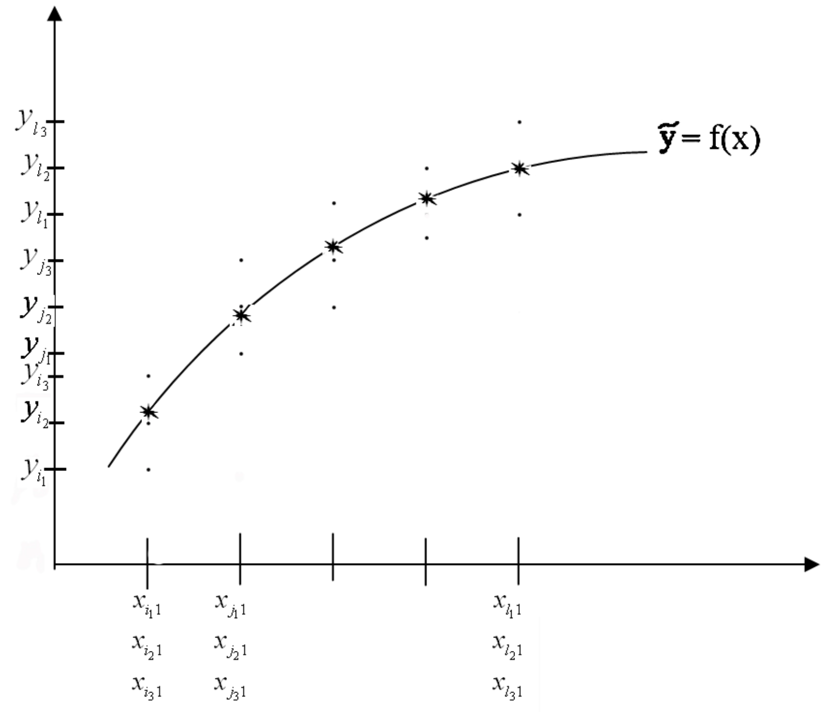


Рисунок 1.1 – Графическое изображение зависимости между результативным признаком y и объясняющей переменной x_1

Предлагается строить регрессионную зависимость (1.1):

$$\tilde{y}_{(x_1,x_2...x_k)} = f(x_1, x_2, ..., x_k), \tag{1.1}$$

где $\tilde{y}_{(x_1,x_2...x_k)}$ – условное среднее значение результативной переменной y для каждого фиксированного набора значений объясняющих переменных.

Рассмотрим достаточно общую, нелинейную относительно включенных в анализ объясняющих переменных, но линейную по оцениваемым параметрам, зависимость:

$$\tilde{y} = \beta_0\psi_0(x_1,x_2,...,x_k) + \beta_1\psi_1(x_1,x_2,...,x_k) + ... + \beta_k\psi_k(x_1,x_2,...,x_k) = \beta_0\psi_0(x) + \beta_1\psi_1(x) + ... + \beta_k\psi_k(x), \tag{1.2}$$

где $\psi_i(x\ ),\ i = \overline{0k}$ - линейно независимые базисные функции;

$\beta = (\beta_0\ \beta_1 ... \beta_k)^T$ - вектор коэффициентов функции регрессии.

Будем считать $\psi_0(x\ ) \equiv 1$.

Частным случаем (1.2) является линейная регрессионная зависимость:

$$\widetilde{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\,. \qquad (1.2а)$$

В векторном виде выражение (1.2) будет иметь вид:

$$\widetilde{y} = \beta^T \psi(x), \qquad (1.2б)$$

где $\psi = (\psi_0 \quad \psi_1 \quad \ldots. \quad \psi_k)^T$.

С учетом наблюденных значений результирующей переменной можем записать:

$$y_i = \beta_0 \psi_0 (x_{i1} x_{i2 \ldots} x_{ik}) + \beta_1 \psi_1 (x_{i1} x_{i2 \ldots} x_{ik}) + \ldots \beta_k \psi_k (x_{i1} x_{i2 \ldots} x_{ik}) + z_i\,, \quad i = \overline{1, n}$$

или

$$y_i = \beta^T \psi(x_i) + z_i\,, \quad i = \overline{1, n} \qquad (1.3)$$

где z_i - регрессионный остаток, характеризующий отклонения наблюденных значений y_i от модельных значений $\widetilde{y}_i$ для *i*–го объекта;

x_i - наблюденные значения объясняющих переменных *i*-ого объекта наблюдения (*i*-ая строчка матрицы X).

Регрессионная модель (1.3) в векторно-матричном виде будет иметь вид:

$$Y = \Psi\beta + Z\,, \qquad (1.3а)$$

где $\Psi_{n\times(k+1)} = \begin{pmatrix} \psi_0(x_{11}x_{12}...x_{1k}) & \psi_1(x_{11}x_{12}...x_{1k}) & .... & \psi_k(x_{11}x_{12}...x_{1k}) \\ \psi_0(x_{21}x_{22}...x_{2k}) & \psi_1(x_{21}x_{22}...x_{2k}) & .... & \psi_k(x_{21}x_{22}...x_{2k}) \\ .... & .... & ... & .... \\ \psi_0(x_{n1}x_{n2}...x_{nk}) & \psi_1(x_{n1}x_{n2}...x_{nk}) & ... & \psi_k(x_{n1}x_{n2}...x_{nk}) \end{pmatrix} = \begin{pmatrix} \psi_0(x_1) & \psi_1(x_1) & ... & \psi_k(x_1) \\ \psi_0(x_2) & \psi_1(x_2) & ... & \psi_k(x_2) \\ ... & .... & ... & .... \\ \psi_0(x_n) & \psi_1(x_n) & ... & \psi_k(x_n) \end{pmatrix}$,

$Z = (z_1,..., z_n)^T$ – вектор значений регрессионных остатков.

Для частного случая линейная модель множественной регрессии (ЛММР) будет иметь вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + z_i \ , \ i = \overline{1,n} .$$

В векторно-матричной форме ЛММР примет вид:

$$Y = X\beta + Z . \qquad (1.3б)$$

На $Z = (z_1,..., z_n)^T$ – можно смотреть как на возможные значения случайного вектора $\varepsilon = (\varepsilon_1,..., \varepsilon_n)^T$, на Y - как на возможные значения случайной априорной выборки $\eta_{1,n} = (\eta_1, \eta_2, ..., \eta_n)^T$, тогда выборочная модель будет иметь вид:

$$\eta_{1,n} = \Psi\beta + \varepsilon , \qquad (1.4)$$

а в частном случае:

$$\eta_{1,n} = X\beta + \varepsilon , \qquad (1.4а)$$

где (1.3а), (1.3 б) – реализация моделей (1.4), (1.4а).

Будем предполагать, что

1) $x_1,\ldots,x_к$ – детерминированные переменные;

2) rang Ψ = к+1;

3) $M\varepsilon_i = 0$, $i = \overline{1,n}$ - нет систематических ошибок в измерении у;

4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2$, $i = \overline{1,n}$ - гомоскедастичность регрессионных остатков (равноточные измерения);

5) $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0$, $i \neq j$, $i = \overline{1,n}$ $j = \overline{1,n}$ - условие некоррелированных регрессионных остатков.

Условия 4 – 5 можно переписать одним условием в векторной форме

4') $\sum_{\varepsilon} = M\bar{\varepsilon}\bar{\varepsilon}^T = \sigma^2 E_n$,

где E_n - единичная матрица (n) порядка

Условия (1 – 5) известны как условия Гаусса – Маркова. ЛММР (1.4), удовлетворяющая требованиям (1–5), называется классической “линейной” моделью множественной регрессии (К“Л”ММР).

Аналогичные требования, в которых rang X = к+1, будем предъявлять в модели (1.4а).

1.2 Оценка коэффициентов КЛММР и исследование статистических свойств МНК – оценок

Рассматривается реализация выборочной модели множественной регрессии (1.3а).

Оценку коэффициентов β модели регрессии будем искать исходя из критерия минимума суммы квадратов отклонений наблюденных значений y_i от "значений" функции регрессии (метод наименьших квадратов):

$$F = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \to \min$$

(См. задание №6 п.1.6)

Перейдем к оценке коэффициентов методом наименьших квадратов (МНК):

$$F = \sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2 = (Y - \widetilde{Y})^T (Y - \widetilde{Y}) = (Y - \Psi\beta)^T (Y - \Psi\beta) \to \min \tag{1.5}$$

Воспользовавшись необходимым условием существования экстремума, получим систему уравнений:

$$\frac{\partial F}{\partial \beta} = 2\Psi^T \Psi \beta - 2\Psi^T Y = 0\,. \tag{1.6}$$

В силу предположения о справедливости условий Гаусса-Маркова, в частности (rang Ψ =k+1), матрица $\Psi^T\Psi$ является невырожденной. Из невырожденности матрицы $\Psi^T\Psi$ вытекает существование обратной матрицы $(\Psi^T\Psi)^{-1}$, что позволяет записать решение системы в форме:

$$\widehat{\beta}_{\text{МНК}} = (\Psi^T\Psi)^{-1}\Psi^T Y\,, \tag{1.7}$$

Оценка коэффициентов реализации выборочной линейной модели множественной регрессии для частного случая будет представляться в виде:

$$\widehat{\beta}_{\text{МНК}} = (X^T X)^{-1} X^T Y \tag{1.7a}$$

Таким образом, оценку функции (уравнения) регрессии можно записать в виде:

$$\widehat{y} = \widehat{\beta}_0\psi_0(x) + \widehat{\beta}_1\psi_1(x) + \ldots + \widehat{\beta}_k\psi_k(x), \tag{1.8}$$

или в частном случае:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k \,. \tag{1.8a}$$

Рассмотрим свойства МНК-оценок, рассматривая $\hat{\beta}$, как функцию от $\eta_{1,n}$:

$$\hat{\beta}_{\text{МНК}}(\eta_{1,n}) = (\Psi^T \Psi)^{-1} \Psi^T \eta_{1,n} \,. \tag{1.9}$$

1) **Несмещенность**. МНК – оценка $\hat{\beta}_{(\text{ìíê})}$ является несмещенной оценкой вектора β:

$$\mathrm{M}\hat{\beta}_{(\text{ìíê})}(\eta_{1,n}) = \beta \,. \tag{1.10}$$

(См. задание №9 п.1.6)

2**) Эффективность.**

Ковариационная матрица вектора оценок $\hat{\beta}$ имеет вид:

$$\Sigma_{\hat{\beta}} = M[(\hat{\beta} - M\hat{\beta})(\hat{\beta} - M\hat{\beta})^T] = \sigma^2 (\Psi^T \Psi)^{-1}. \tag{1.11}$$

(См. задание №10 п.1.6)

Несмещенная оценка остаточной дисперсии σ^2 определяется по формуле:

$$\hat{S}_{ocm}{}^2 = \frac{(Y - \hat{Y})^T (Y - \hat{Y})}{n - k - 1} = \frac{Q_{ocm}}{n - k - 1} \,. \tag{1.12}$$

(См. задание №11 п.1.6)

Оценка ковариационной матрицы определяется по формуле:

$$\widehat{\Sigma}_{\widehat{\beta}} = \widehat{S}^2_{ост}(\Psi^T\Psi)^{-1}. \qquad (1.13)$$

3) Одно из д**остаточных условий состоятельности** оценок $\widehat{\beta}$ *и* $\widehat{S}^2$ заключается в [1]:

$$\lim_{n\to\infty} \lambda_{min} = \infty,$$

где $\lambda_{\min}$ - наименьшее собственное число матрицы $\Psi^T\Psi$.

1.3 Анализ вариации результативного признака. Выборочный коэффициент детерминации

В качестве характеристики качества функции регрессии используется коэффициент детерминации, который получается из тех соображений, что общая вариация (дисперсия) результативного признака складывается из вариации функции регрессии, обусловленной варьированием значений объясняющих переменных (факторной дисперсии) и из вариации случайной величины относительно функции регрессии (остаточной дисперсии).

Определим выборочную вариацию результативной переменной y величиной:

$$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \qquad (1.14)$$
$$= Q_{ост} + Q_{факт} + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}),$$

где $\bar{y} = \sum_{i=1}^{n} y_i / n$,

$\hat{y}_i$ - оценка модельных значений, полученная по (1.8);

$Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (Y - \bar{Y})^T \ (Y - \bar{Y})$ - полная вариация y_i относительно $\bar{y}$;

$Q_{ост} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} z_i^2 = (Y - \widehat{Y})^T (Y - \widehat{Y})$ - вариация регрессионных остатков;

$Q_{фак} = \sum(\hat{y}_i - \bar{y})^2 = (\hat{Y} - \bar{Y})^T (\hat{Y} - \bar{Y})$ - вариация относительно $\bar{y}$, объясняемая регрессией;

$Y = (y_1,...,y_n)^T$, $\bar{Y}_{n*1} = (\bar{y}, \bar{y},...,\bar{y})^T$, $\hat{Y} = \Psi\hat{\beta} = (\hat{y}_1 \quad \hat{y}_2 \quad ... \quad \hat{y}_n)^T$.

Покажем, что удвоенное произведение равно нулю:

$$\begin{pmatrix} \sum_{i=1}^{n}\psi_0(x_i)\cdot z_i \\ \sum_{i=1}^{n}\psi_1(x_i)\cdot z_i \\ ... \\ \sum_{i=1}^{n}\psi_k(x_i)\cdot z_i \end{pmatrix} = \begin{pmatrix} \psi_0(x_1) & \psi_0(x_2) & ... & \psi_0(x_n) \\ \psi_1(x_1) & \psi_1(x_2) & ... & \psi_1(x_n) \\ ... & .... & ... & .... \\ \psi_k(x_1) & \psi_k(x_2) & ... & \psi_k(x_n) \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \\ ... \\ z_n \end{pmatrix} = \Psi^T Z = \Psi^T(Y - \Psi\hat{\beta}) = \Psi^T Y - \Psi^T\Psi\hat{\beta} = ,$$

$$= \Psi^T Y - \Psi^T\Psi(\Psi^T\Psi)^{-1}\Psi^T Y = \Psi^T Y - \Psi^T Y = 0$$

то есть $\sum_{i=1}^{n}\psi_0(x_i)z_i = 0$, $\sum_{i=1}^{n}\psi_l(x_i)z_i = 0$, $l = \overline{0,k}$;

из чего следует, что

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} z_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} z_i(\hat{\beta}_0\psi_0(x_{i1}x_{i2...}x_{ik}) + \hat{\beta}_1\psi_1(x_{i1}x_{i2...}x_{ik}) + ...\hat{\beta}_k\psi_k(x_{i1}x_{i2...}x_{ik}) - \bar{y}) =$$
$$= \hat{\beta}_0\sum_{i=1}^{n} z_i\psi_0(x_i) + \hat{\beta}_1\sum_{i=1}^{n} z_i\psi_1(x_i) + ... + \hat{\beta}_k\sum_{i=1}^{n} z_i\psi_k(x_i) - \bar{y}\sum_{i=1}^{n} z_i = 0$$

Таким образом, $Q_{общ} = Q_{ост} + Q_{факт}$, следовательно,

$$\frac{Q_{фак}}{Q_{общ}} = 1 - \frac{Q_{ост}}{Q_{общ}} = \hat{R}^2, \tag{1.15}$$

$\hat{R}^2$ - оценка коэффициента детерминации.

Выборочный коэффициент детерминации характеризует долю общей вариации результативного признака *y*, обусловленную влиянием объясняющих переменных, включенных в модель.

Из (1.15) следует, что $0 \le \hat{R}^2 \le 1$.

1.4 Проверка гипотезы об адекватности линейной модели выборочным данным

Дальнейшее изучение свойств оценок КЛММР проводится при дополнительном предположении о нормальном характере распределения регрессионных остатков: $\varepsilon_i \in N(0,\sigma^2)$, $i=\overline{1,n}$, $\varepsilon \in N(0,\sigma^2 E_n)$, которое должно быть проверено после оценки коэффициентов линейной модели множественной регрессии.

Для проверки значимости построенного уравнения регрессии выдвигаются гипотезы:

H_0: $\beta_1=\beta_2=\ldots=\beta_к=0$ (линейная модель множественной регрессии неадекватна выборочным данным); (вариация ни одной из объясняющих переменных не оказывает влияние на вариацию результативного признака)

H_1: $\exists j \in [\overline{1,\kappa}]:\ \beta_j \neq 0$ (ЛММР адекватна выборочным данным); (вариация хотя бы одной из объясняющих переменных оказывает влияние на вариацию результативного признака).

Для проверки гипотезы H_0 используется статистика:

$$F=\frac{\widehat{R}^2/k}{(1-\widehat{R}^2)/(n-k-1)}=\frac{Q_{факт}/k}{Q_{ост}/(n-k-1)}, \tag{1.16}$$

которая в случае справедливости H_0 имеет, распределение Фишера – Снедекора с числом степеней свободы $\nu_1=k$ и $\nu_2=n-k-1$.

(См. задание 18 п.1.6)

1.5 Проверка гипотез о значимости коэффициентов КЛММР

В случае, если нулевая гипотеза о незначимости уравнения регрессии отвергнута, проверяем гипотезы о значимости коэффициентов уравнения регрессии. Выдвигаются гипотезы:

H_0: $\beta_j=0$ (коэффициент β_j незначимо отличен от нуля); (объясняющая переменная x_j не оказывает влияние на результативный признак)

H_1: $\beta_j \neq 0$ (коэффициент β_j – значимо отличен от нуля); (объясняющая переменная x_j оказывает влияние на результативный признак)

Для проверки таких гипотез H_0 строятся статистики

$$t_j = \frac{\hat{\beta}_j}{\hat{S}_{\hat{\beta}_j}}, \quad j = 1,2,...,k, \quad S_{\beta_j} = \hat{S}_{\text{ост}} \sqrt{[(\Psi^T \Psi)^{-1}]_{j+1\,j+1}}\,, \qquad (1.17)$$

которые в случае справедливости H_0, имеют распределение Стьюдента с $\nu = n - k - 1$ степенями свободы.

1.6 Вопросы и задания, выносимые на практическо-семинарские занятия

1. Записать нелинейную относительно включенных в анализ объясняющих переменных, но линейную по оцениваемым параметрам зависимость результативного признака от объясняющих переменных.
2. Записать выборочную модель множественной регрессии и ее реализацию.
3. Что такое регрессионный остаток? Чем обусловлено его наличие в модели?
4. Сформулируйте условия Гаусса-Маркова.

5. Какая модель называется классической линейной моделью множественной регрессии (КЛММР)?

6. Какие методы существуют для оценивания коэффициентов в рамках КЛММР? В чем их суть? [1, С. 632-637].

7. Выведите формулу для нахождения МНК-оценки параметров β.

8. Какими свойствами обладают МНК-оценки КЛММР?

9. Докажите свойство несмещенности МНК-оценки коэффициентов КЛММР [4, С. 48-54].

10. Выведите формулу для ковариационной матрицы вектора МНК-оценок КЛММР [1, С. 644-646].

11. Доказать, что $\widehat{S}_{ост}{}^2 = \frac{(Y-\widehat{Y})^T(Y-\widehat{Y})}{n-k-1} = \frac{Q_{ост}}{n-k-1}$ есть несмещенная оценка остаточной дисперсии σ^2 [4, С. 48-54].

12. Доказать справедливость разложения:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\widehat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\widehat{y}_i)^2$$

13. Запишите формулу выборочного коэффициента детерминации.

14. Что характеризует коэффициент детерминации в регрессионном анализе? В каких пределах он изменяется?

15. Как проверить гипотезу об адекватности КЛММР?

16. Доказать, что при $\beta_1=\beta_2=\ldots=\beta_к=0$ статистика $F = \frac{Q_{факт}/k}{Q_{ост}/(n-k-1)}$ имеет F-распределение с числом степеней свободы $\nu_1 = k$ и $\nu_{21} = n-k-1$ [4, С. 54-55].

17. Как проверить гипотезу о значимости отдельных коэффициентов КЛММР?

18. Постройте доверительные интервалы для значимых коэффициентов КЛММР.[4, С.58-59],[6, С.76]

19. Запишите доверительные интервалы для $\tilde{y}_n$ и $\tilde{y}_{n+1}$ [4, С. 58-60]

20. По данным $n=15$ фирм исследована зависимость прибыли y от числа работающих x вида $\tilde{y}=\hat{\beta}_0+\hat{\beta}_1 x$. Была получена оценка остаточной дисперсии $\hat{S}^2=2.2$ и обратная матрица $\left(X^T X\right)^{-1}=\begin{pmatrix} 0.31 & -0.03 \\ -0.03 & 0.05 \end{pmatrix}$

Определить, чему равна дисперсия оценки коэффициента регрессии $\hat{S}_{b_1}$.

21. Уравнению регрессии $\tilde{y}=3.57-0.63x_1-1.78x_2$ соответствует множественный коэффициент корреляции $\hat{R}_{y/x_1x_2}=0{,}79$. Какая доля вариации результативного показателя y (в %) объясняется вошедшими в уравнение регрессии переменными x_1 и x_2?

22. При исследовании зависимости себестоимости продукции *"у"* от объема выпуска *X₁* и других факторов (всего вместе *"к")* по данным *"n"* обследованных предприятий получена оценка уравнения регрессии $\hat{y}$. Определить с доверительной вероятностью γ на какую величину максимально может измениться себестоимость продукции *у*, если объем производства *X₁* увеличить на 1 при неизменных значениях других факторов:

$$\hat{y}=2{,}88-\underset{(0{,}052)}{0{,}72}\,x_1+\ldots;\quad n=20,\quad \kappa=2;\quad \gamma=0{,}95$$

P.S. $t=\dfrac{в_j-\beta_j}{S_{в_j}}\in t(1-\gamma, n-\kappa-1)$

а) 0,65 б) -0,83 в) –0,052 г) –0,72

23. Какое условие относится к условиям гомоскедастичности в линейной модели множественной регрессии

а) $M\varepsilon_i\varepsilon_j=\begin{cases} 0, & i\neq j,\quad i=\overline{1,n},\quad j=\overline{1,n} \\ \sigma^2, & i=j, \end{cases}$

б) $M\varepsilon_i\varepsilon_j=\begin{cases} 1, & i\neq j,\quad i=\overline{1,n},\quad j=\overline{1,n} \\ \sigma^2, & i=j, \end{cases}$

в) $M\varepsilon_i\varepsilon_j = \begin{cases} 0 & i \neq j, \quad i = \overline{1,n}, \quad j = \overline{1,n} \\ \sigma_i^2 & i = j, \end{cases}$

г) $M\varepsilon_i\varepsilon_j = \begin{cases} 1 & i \neq j, \quad i = \overline{1,n}, \quad j = \overline{1,n} \\ 0 & i = j, \end{cases}$

24. Модель $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ называют классической линейной моделью множественной регрессии, если выполняются следующие условия:

а) $\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \textit{неслучайные переменные}; \\ rang\ X = k+1 << n; \\ M\varepsilon_i \neq 0,\ i = 1,2,...,n; \\ M(\varepsilon_i\varepsilon_j) = \begin{cases} \sigma^2 \textit{ при } i = j \\ 0 \textit{ при } i \neq j \end{cases} \quad i = \overline{1,n}; \quad j = \overline{1,n} \end{cases}$

б) $\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \textit{неслучайные переменные}; \\ rang\ X = k+1 << n \\ M\varepsilon_i = 0,\ i = 1,2,...,n; \\ M(\varepsilon_i\varepsilon_j) = \begin{cases} \sigma^2 \textit{ при } i = j \\ 0 \textit{ при } i \neq j \end{cases} \quad i = \overline{1,n}; \quad j = \overline{1,n} \end{cases}$

в) $\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \textit{неслучайные переменные}; \\ rang\ X = k+1 << n \\ M\varepsilon_i = 0,\ i = 1,2,...,n; \\ M(\varepsilon_i\varepsilon_j) = \begin{cases} \sigma_i^2 \textit{ при } i = j \\ 0 \textit{ при } i \neq j \end{cases} \quad i = \overline{1,n}; \quad j = \overline{1,n} \end{cases}$

г) $\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \textit{неслучайные переменные}; \\ rang\ X = k+1 << n \\ M\varepsilon_i = 0,\ i = 1,2,...,n; \\ \sum_{\varepsilon} = \sigma^2 \sum_0, \textit{ где } \sum_0 \neq E_n \end{cases}$

25. Для изучения рынка жилья в городе по данным о 43 коттеджах, получена следующая оценка уравнения регрессии:

$$\hat{y} = \underset{(8,2)}{21.1} - \underset{(3,8)}{6.2} x_1 + \underset{(0,83)}{3.57} x_2 + \underset{(0,59)}{1.17} x_3, \quad \hat{R}^2 = 0.7$$

где у –цена объекта, тыс.долл.;

x_1 - расстояние до центра города, км;

x_2 - полезная площадь объекта, кв.м;

x_3 - число комнат в квартире;

Модель оказалась значимой. Укажите факторы, оказывающие существенное влияние на цену объекта на заданном уровне значимости α=0,05.

а) расстояние до центра города;

б) полезная площадь объекта;

в) число комнат в квартире;

г) расстояние до центра города, и полезная площадь объекта.

P.S. $t_{кр}(0{,}05{,}39) = 2{,}022$; $t_{кр}(0{,}05{,}40) = 2{,}021$;

$t_{кр}(0{,}05{,}41) = 2{,}020$; $t_{кр}(0{,}05{,}47) = 2{,}012$.

26. По выборке объемом $"n"$ произведена оценка параметров линейной модели множественной регрессии $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i$ методом наименьших квадратов. Модель окажется значимой при заданном уровне значимости α= 0,05, n = 35; k = 4, если наблюденное значение статистики F оказалось равным:

а) 1,68;

б) 2,14;

в) 2,58;

г) 2,82.

P.S. $F_{кр}(0{,}05;4;30) = 2{,}6896$; $F_{кр}(0{,}05;4;31) = 2{,}6787$;

$F_{кр}(0{,}05;4;35) = 2{,}6414$; $F_{кр}(0{,}05;5;30) = 2{,}5335$;

2 Практическая часть

2.1 Описание лабораторной работы

Лабораторная работа включает в себя следующие этапы:

- постановку задачи;
- ознакомление с порядком выполнения работы;
- выполнение расчетов индивидуальных задач на компьютере и анализ результатов;
- подготовку письменного отчета с выводами по работе;
- защиту лабораторной работы.

2.2 Задание к лабораторной работе

На основе показателей, характеризующих социально-экономическое развитие городов и районов Оренбургской области (Приложение А), провести регрессионный анализ:

- построить МНК-оценки коэффициентов линейной модели множественной регрессии;
- проверить значимость уравнения регрессии и значимость коэффициентов уравнения регрессии;
- для значимых коэффициентов уравнения регрессии простроить доверительные интервалы;
- провести экономический анализ результатов.

2.3 Порядок выполнения работы

Ищется зависимость ожидаемой продолжительности жизни мужчин, число лет (у) от ряда факторов:

x_1 – общий коэффициент рождаемости (на 1000 человек);

x_2 – общий коэффициент смертности (на 1000 человек)

x_3 – уровень брачности населения (на 1000 человек);

x_4 – уровень разводимости (на 1000 человек);

x_5 – коэффициент младенческой смертности (на 1000 родившихся живыми);

x_6 – соотношение денежного дохода и прожиточного минимума, (%);

x_7 – соотношении средней оплаты труда и прожиточного минимума трудоспособного населения, (%);

x_8 – численности населения с денежными доходами ниже прожиточного минимума (в % от численности населения);

x_9 –число зарегистрированных преступлений (на 100000 человек).

Зависимость будем искать в виде:

$$\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9$$

Объектом исследования выступают города и районы Оренбургской области. Предметом исследования – взаимосвязи между ожидаемой продолжительностью жизни мужчин и указанными показателями.

Информационная база представлена данными о значениях соответствующих показателей для 48 городов и районов Оренбургской области.

Запуск ППП Statistica и подготовка данных.

Запустить ППП Statistica. После запуска на экране откроется основное окно системы Statistica, представленное на рисунке 2.1.

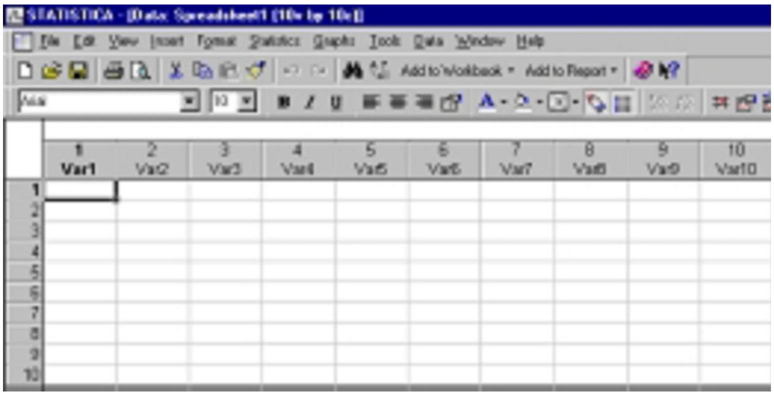


Рисунок 2.1 – Стартовое окно пакета Statistica

Стандартный вид исходной таблицы содержит 10 строк (10 cases) и 10 столбцов (10 variables). Так как исходная информация может быть представлена произвольного размера, то возникает необходимость в корректировке размерности таблицы. Если необходимо увеличить число столбцов, то в меню Insert, выбираем Add Variables, если необходимо изменить число строк, то –Add Cases. При этом откроется меню возможных операций со столбцами (строками).

Далее необходимо ввести данные для проведения регрессионного анализа. Если исходная информация уже имеется, то следует открыть нужный файл – для этого используется кнопка **Open Data** – **Открыть данные**. Окно с частью данных для анализа представлено на рисунке 2.2.

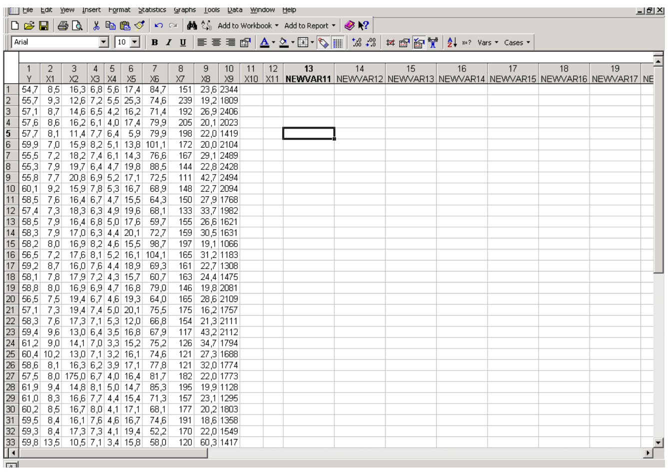


Рисунок 2.2– Исходные данные

Для построения уравнения множественной регрессии в меню системы открыть **Statistics - Критерии** и выбрать в появившемся меню строку **Multiple Regression – Множественная регрессия**:

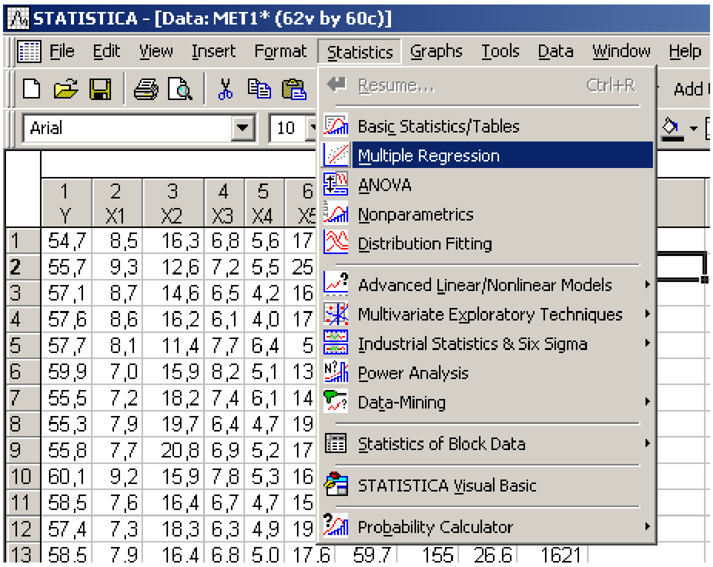


Рисунок 2.3– Выбор пункта меню для проведения регрессионного анализа

На экране появится окно:

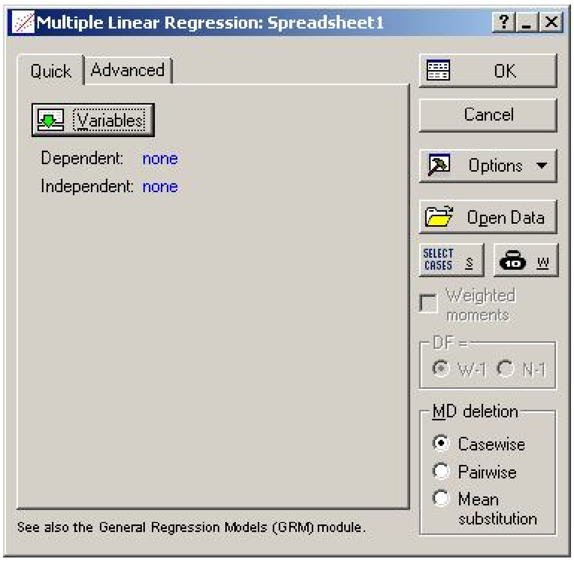


Рисунок 2.4 – Окно выбора переменных

Далее необходимо выбрать зависимую (результирующую, объясненную) и независимые (объясняющие) переменные для анализа.

Для задания переменных воспользуемся кнопкой **Variables** – **Переменные** из панели **Multiple Regression – Множественная регрессия** (рисунок 2.4).

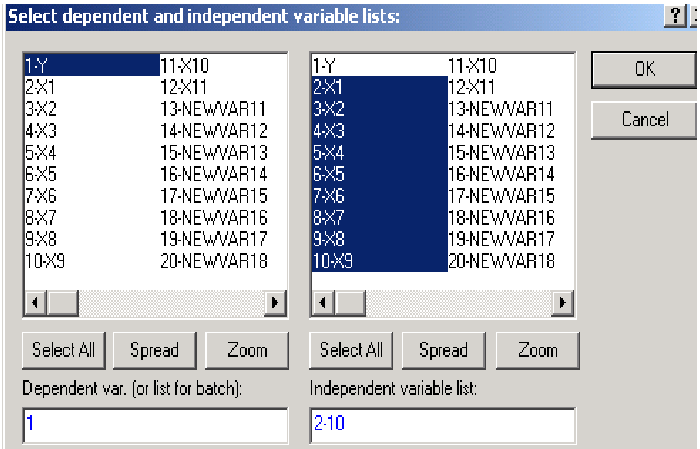


Рисунок 2.5 – Выбор зависимой и независимых переменных для проведения регрессионного анализа

В окне **Select dependent and independent variable list – Выбор зависимой переменной и списка независимых переменных**, выделяя имя переменной в левой части окна, производится выбор зависимой переменной **Dependent**. В правой части окна выбираем независимую переменную (**Independent)**. Выбор нескольких несмежных переменных производят при нажатой клавише **CTRL**. После выбора переменных необходимо щелкнуть на кнопке **OK**, вновь окажемся в панели модуля Множественная регрессия.

Нажатие на кнопку **Advanced** позволяет перейти к окну функциональных возможностей модуля Множественная регрессия.

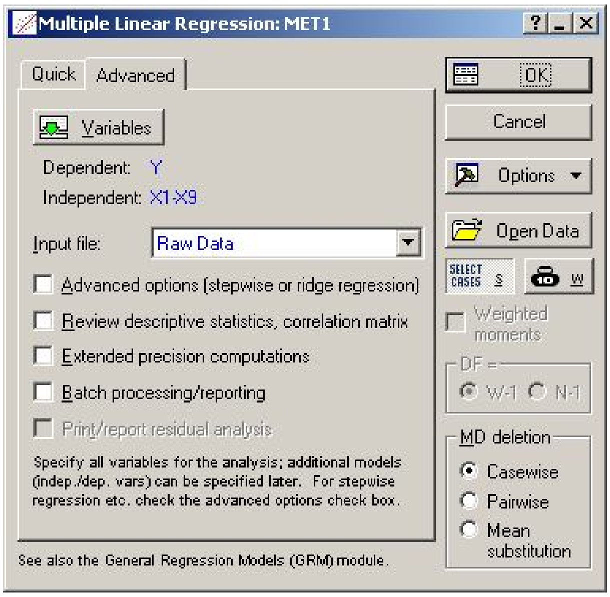


Рисунок 2.6 – Модуль множественная регрессия

Строка **Input file** определяет тип входной информации. Если входная информация представляет собой массив исходных данных, следует оставить **Raw Data** (необработанные данные). В поле окна **MD deletion** можно задать правило обработки пропущенных данных. Установка флажка в поле **Advanced options** позволит перейти к диалоговому окну **Model Defenition,** открывающему возможность выбора метода анализа, среди которых методы пошаговой регрессии и гребневой. Установка флажка в поле **Review descriptive statistics, correlations matrix** позволит провести предварительный анализ исходных переменных и построить корреляционную матрицу, анализ которой дает возможность сделать важные выводы о структуре связей между выбранными переменными. Установка флажка в поле **Extended precision computations** позволит выбрать метод расчета с расширенной точностью. После определения всей необходимой информации для построения модели, щелкните по кнопке **OK** в правом углу окна. Результаты расчетов приведены в виде отчета на рисунке 2.7.

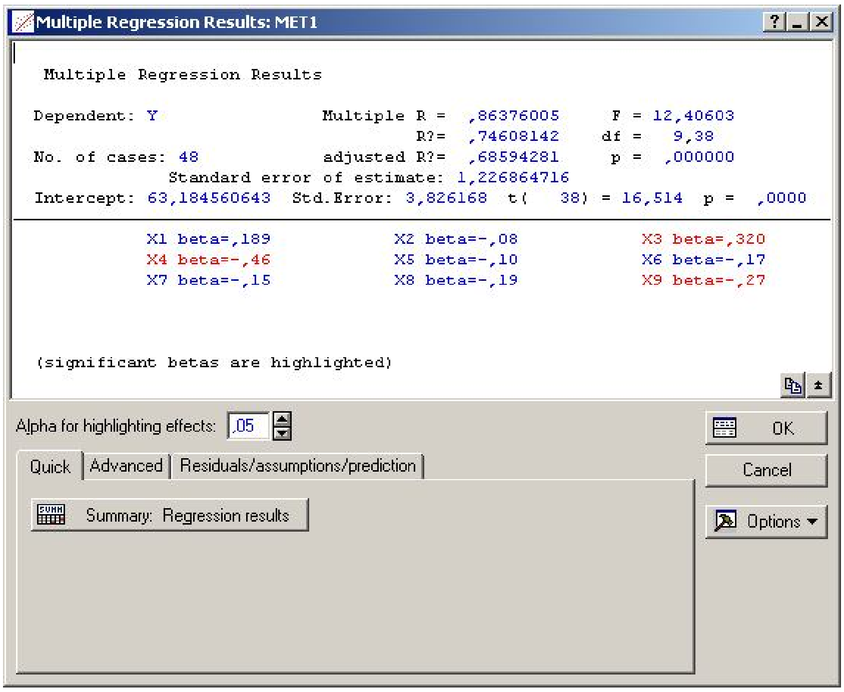


Рисунок 2.7 – Окно с результатами вычислений

В верхней информационной части окна результатов представлены основные характеристики построенной модели, а нижняя – содержит кнопки доступа к дополнительной информации, позволяющей провести исчерпывающий анализ модели, дать интерпретацию вычисленным параметрам и оценить адекватность модели исходным данным.

Рассмотрим содержание информационной части окна.

В левой части окна приводится имя зависимой переменной (**Dependent**) и число наблюдений, по которым построено уравнение регрессии (**No. Of Cases**).

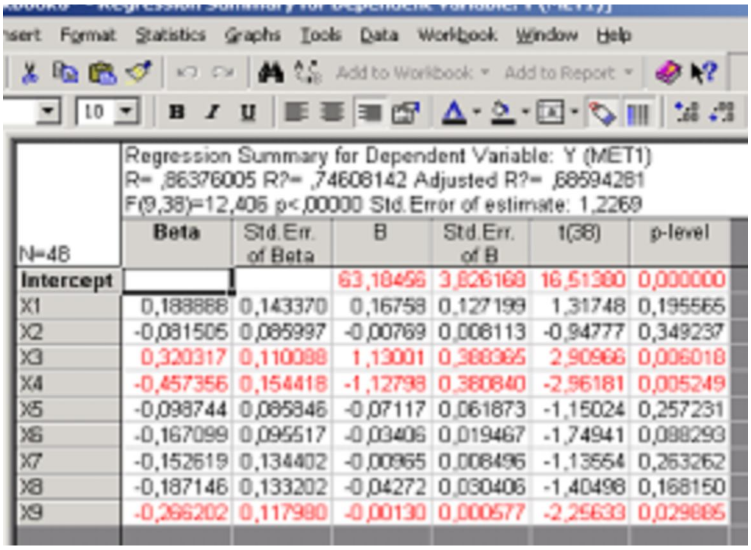
В правой части окна приводится оценка коэффициента множественной корреляции (**Multiple R**) и значение квадрата этого коэффициента (**R^2**) – коэффициента детерминации, несмещенная оценка R^2 (**Adjusted R^2**)служит скорректированный на потерю степеней свободы коэффициент множественной детерминации (**Adjusted R^2**), значение **F–критерий** [1].

Также в верхней части окна результатов анализа приводится оценка свободного члена уравнения регрессии (**Intercept**), стандартная ошибка (среднеквадратическое отклонение) этой оценки (**Std. Error**), значение t- критерия и уровень значимости, используемые для проверки гипотезы о равенстве нулю свободного члена.

Standard Error of estimate является оценкой $\sqrt{S^2_{ост}}$, где $S^2{}_{ост}$ – несмещенная оценка остаточной дисперсии.

Во второй части информационного окна подсвечены оценки значимых регрессионных коэффициентов (речь в данном случае идет о нормированных оценках: Beta**-** коэффициентах).

Более подробную информацию получим после нажатия на кнопку **Regression summary** (рисунок 2.8).



Regression Summary for Dependent Variable: Y (MET1)
R= ,86376005 R?= ,74608142 Adjusted R?= ,68594281
F(9,38)=12,406 p<,00000 Std.Error of estimate: 1,2269

N=48	Beta	Std.Err. of Beta	B	Std.Err. of B	t(38)	p-level
Intercept			63,18456	3,826168	16,51380	0,000000
X1	0,188888	0,143370	0,16758	0,127198	1,31748	0,195565
X2	-0,081505	0,085997	-0,00769	0,008113	-0,94777	0,349237
X3	0,320317	0,110088	1,13001	0,388365	2,90966	0,006018
X4	-0,457356	0,154418	-1,12798	0,380840	-2,96181	0,005249
X5	-0,098744	0,085846	-0,07117	0,061873	-1,15024	0,257231
X6	-0,167099	0,095517	-0,03406	0,019467	-1,74941	0,088293
X7	-0,152619	0,134402	-0,00965	0,008496	-1,13554	0,263262
X8	-0,187146	0,133202	-0,04272	0,030406	-1,40498	0,168150
X9	-0,266202	0,117980	-0,00130	0,000577	-2,25633	0,029886

Рисунок 2.8 – Результаты оценивания параметров линейной модели множественной регрессии

В данном окне модуля представлены оценки параметров модели (**B**- обычные оценки и **Beta**- нормированные оценки), оценки их стандартных ошибок (**St. Error**) и уровни значимости (**p-level))** t-критерий Стьюдента [1-3].

Далее можно приступить к исследованию остатков регрессионной модели. Остатки исследуются в специальном окне **Residuals analysis – Анализ остатков**. В нем приведен широкий набор статистических и визуальных методов исследования остатков модели. Для этого необходимо щелкнуть мышкой по кнопке

Residuals/assumptions/prediction – Остатки/распределение/предсказанные в окне рисунка 2.7 (рисунок 2.9).

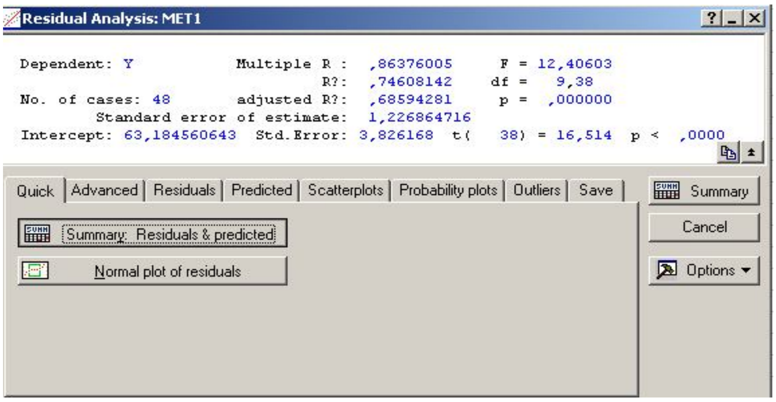


Рисунок 2.9 – Окно для анализа регрессионных остатков

Информация о значениях остатков может быть получена нажатием на кнопку **Summary: Residuals & predicted** (рисунок 2.10).

Predicted & Residual Values (MET1)
Dependent variable: Y

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
1	54,70000	56,21295	-1,51295	-1,53777	-1,23318	0,562058	8,88513	-1,91483	0,0511
2	55,70000	56,75719	-1,05719	-1,24996	-0,86170	0,768218	17,44862	-1,73902	0,0788
3	57,10000	57,42084	-0,32084	-0,89900	-0,26151	0,591881	9,95973	-0,41817	0,0027
4	57,60000	57,45398	0,14602	-0,88148	0,11902	0,568884	9,12621	0,18602	0,0005
5	57,70000	58,09875	-0,39875	-0,54050	-0,32502	0,870725	22,69454	-0,80345	0,0216
6	59,90000	58,07185	1,82816	-0,55473	1,49010	0,631415	11,46983	2,48685	0,1088
7	55,50000	56,01296	-0,51296	-1,64353	-0,41811	0,522596	7,54863	-0,62666	0,0047
8	55,30000	56,34161	-1,04161	-1,46973	-0,84900	0,578350	9,46531	-1,33922	0,0265
9	55,80000	56,42005	-0,62005	-1,42825	-0,50539	0,551672	8,52394	-0,77719	0,0081
10	60,10000	58,78236	1,31764	-0,17900	1,07399	0,523481	7,57752	1,61092	0,0314
11	58,50000	58,36900	0,13100	-0,39759	0,10678	0,350604	2,85914	0,14265	0,0001
12	57,40000	56,84306	0,55694	-1,20455	0,45396	0,461144	5,66097	0,64857	0,0039
13	58,50000	58,39967	0,10033	-0,38137	0,08177	0,426808	4,70897	0,11414	0,0001
14	58,30000	57,66798	0,63202	-0,76831	0,51515	0,442326	5,13011	0,72645	0,0046
15	58,20000	59,90450	-1,70449	0,41442	-1,38931	0,567311	9,07037	-2,16807	0,0668
16	56,50000	58,38821	-1,88821	-0,38743	-1,53905	0,685727	13,70358	-2,74608	0,1565
17	59,20000	60,21419	-1,01419	0,57820	-0,82665	0,356279	2,98440	-1,10759	0,0069
18	58,10000	59,92093	-1,82093	0,42311	-1,48422	0,422034	4,58243	-2,06533	0,0335
19	58,80000	58,04236	0,75764	-0,57033	0,61754	0,400556	4,03077	0,84803	0,0051
20	56,50000	57,56338	-1,06338	-0,82362	-0,86675	0,360187	3,07182	-1,16368	0,0078
21	57,10000	58,31242	-1,21242	-0,42751	-0,98823	0,392317	3,82677	-1,35052	0,0124
22	58,30000	58,09826	0,20174	-0,54076	0,16443	0,499889	6,82364	0,24190	0,0006
23	59,40000	58,74685	0,65315	-0,19777	0,53238	0,509359	7,12210	0,78918	0,0071
24	61,20000	60,09682	1,10318	0,51613	0,89919	0,510295	7,15189	1,33396	0,0205
25	60,40000	60,99086	-0,59086	0,98892	-0,48160	0,489708	6,50906	-0,70284	0,0052
26	58,60000	58,31413	0,28587	-0,42661	0,23301	0,477503	6,14045	0,33691	0,0011
27	57,50000	57,28628	0,21372	-0,97016	0,17420	1,224763	45,85995	62,43660	258,1066
28	61,90000	60,00879	1,89121	0,46958	1,54150	0,433881	4,89906	2,16155	0,0388
29	61,00000	60,47494	0,52506	0,71609	0,42797	0,368717	3,26596	0,57720	0,0020
30	60,20000	60,44295	-0,24295	0,69917	-0,19803	0,583557	9,65422	-0,31399	0,0015

Рисунок 2.10 – Наблюденные значения, оценки модельных значений результативного признака, оценки регрессионных остатков

Для проведения теста на нормальный характер распределения регрессионных остатков, скопируем столбец **Residual** в окно с исходными данными. Затем в меню системы **Statistica** выберем пункт **Distribution Fitting**. На экране появится окно:

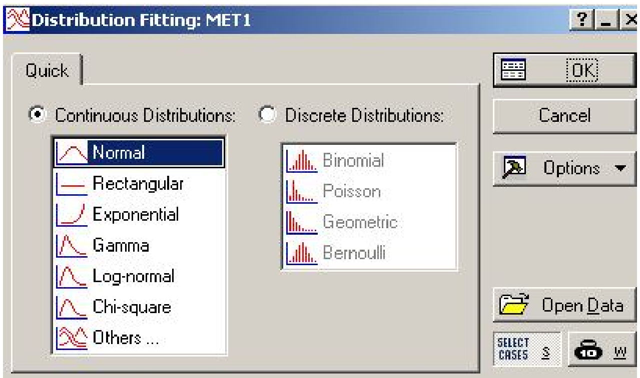


Рисунок 2.11 – Выбор вида распределения регрессионных остатков

В появившемся окне выберем распределение **Normal – Нормальное** и щелкнем по кнопке **OK.** После чего на экране появится окно:

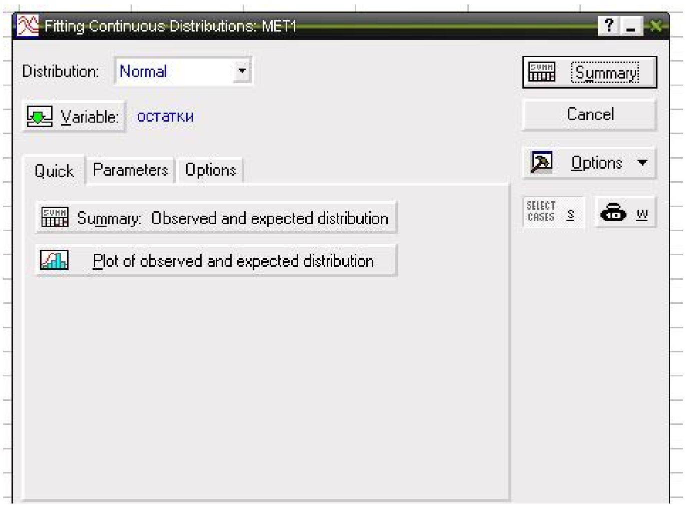


Рисунок 2.12 – Выбор пунктов для построения гистограммы регрессионных остатков

В данном окне сначала необходимо выбрать переменные, используя кнопку **Variable.** Кроме того, в данном модуле, используя кнопку **Parameters – Параметры,** можно изменить количество интервалов, верхнюю и нижнюю границы интервалов и т.д. Для получения графика нормального распределения, нажмем по кнопке **Plot of observed and expected distribution.**

На экране появится окно, содержащее гистограмму распределения, значение χ^2 – критерия, степени свободы, значимость нулевой гипотезы [1].

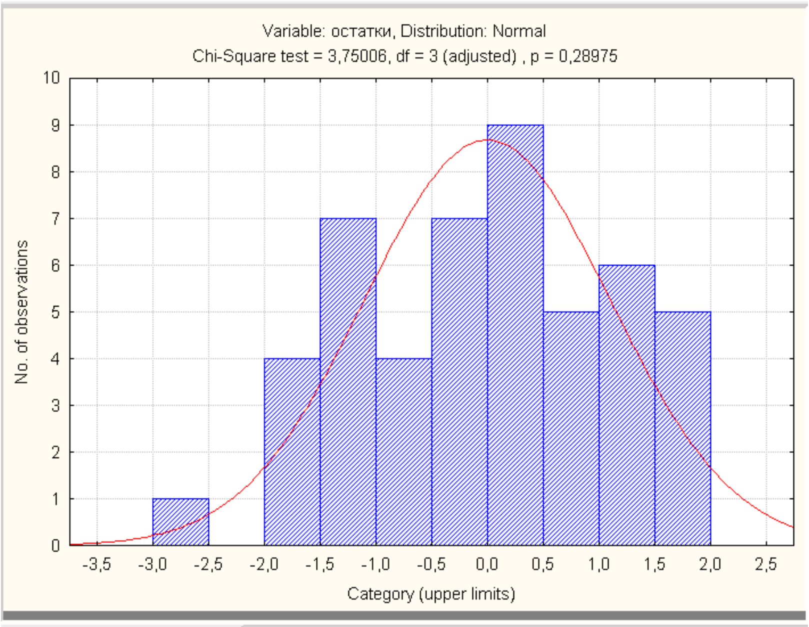


Рисунок 2.13 – График распределения регрессионных остатков

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что распределение регрессионных остатков не отличаются от нормального, так как значимость нулевой гипотезы (p=0,29).

Так как регрессионные остатки имеют нормальное распределение, то есть смысл проводить дальнейший анализ построенного уравнения множественной регрессии.

Итак, вернемся к окну **Multiple Regression Results -** Результаты множественной регрессии:

Regression Summary for Dependent Variable: Y (METT)
R= ,86376005 R?= ,74608142 Adjusted R?= ,68594281
F(9,38)=12,406 p<,00000 Std.Error of estimate: 1,2269

N=48	Beta	Std.Err. of Beta	B	Std.Err. of B	t(38)	p-level
Intercept			63,18456	3,826168	16,51380	0,000000
X1	0,188888	0,143370	0,16758	0,127199	1,31748	0,195565
X2	-0,081505	0,085997	-0,00769	0,008113	-0,94777	0,349237
X3	0,320317	0,110088	1,13001	0,388365	2,90966	0,006018
X4	-0,457356	0,154418	-1,12798	0,380840	-2,96181	0,005249
X5	-0,098744	0,085846	-0,07117	0,061873	-1,15024	0,257231
X6	-0,167099	0,095517	-0,03406	0,019467	-1,74941	0,088293
X7	-0,152619	0,134402	-0,00965	0,008496	-1,13554	0,263262
X8	-0,187146	0,133202	-0,04272	0,030406	-1,40498	0,168150
X9	-0,266202	0,117980	-0,00130	0,000577	-2,25633	0,029885

Рисунок 2.14 - Результаты оценивания параметров линейной модели множественной регрессии

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y}=63{,}18+0{,}17X_1-0{,}008X_2+1{,}13X_3-1{,}13X_4-0{,}07X_5-0{,}034X_6-0{,}0097X_7-0{,}04X_8-0{,}001X_9$$

(0,13) (0,008) (0,39) (0,38) (0,06) (0,02) (0,008) (0,03) (0,005)

Как видно из отчета, уравнение регрессии значимо, т.е. модель адекватна экспериментальным данным, значимыми оказались только коэффициенты при переменных X_3, X_4, X_9. Согласно полученной модели, на ожидаемую продолжительность жизни значимое влияние оказывают уровень брачности населения, уровень разводимости и число зарегистрированных преступлений: при увеличении уровня брачности ожидаемая продолжительность жизни мужчин будет увеличиваться; при увеличении уровня разводимости ожидаемая продолжительность жизни мужчин будет уменьшаться; при увеличении числа зарегистрированных преступлений ожидаемая продолжительность жизни мужчин будет уменьшаться.

Так как среди коэффициентов уравнения регрессии много незначимых, то требуется исследовать модель на наличие мультиколлинеарности.

2.4 Содержание письменного отчета

Отчет должен быть оформлен на листах формата А4 с титульным листом, оформленным соответствующим образом и содержать следующее:

1) постановку задачи с вариантом выборок;

2) краткое изложение теории по методам построения и исследования линейных моделей множественной регрессии;

3) результаты компьютерной обработки данных;

4) анализ полученных результатов;

5) выводы по полученным результатам.

2.5 Вопросы к защите лабораторной работы

1) Сформулируйте постановку задачи лабораторной работы.

2) Запишите результаты наблюдений в виде вектора Y и матрицы X типа «объект-свойство».

3) Каким методом были оценены коэффициенты ЛММР?

4) В чем суть метода наименьших квадратов оценки коэффициентов КЛММР?

5) Какими свойствами обладают МНК-оценки КЛММР?

6) В чем суть метода максимального правдоподобия оценки коэффициентов КЛММР?

7) Чему равна оценка коэффициента детерминации в лабораторной работе? Что она характеризует?

8) При каком дополнительном предположении относительно регрессионных остатков исследуется значимость модели, ее коэффициентов, осуществляется интервальное оценивание?

9) Как проверить гипотезу об адекватности модели регрессии выборочным данным? Как проверить гипотезу о значимости коэффициентов?

10) Дайте интерпретацию коэффициентов уравнения регрессии.

Список использованных источников

1 Айвазян, С.А. Прикладная статистика и основы эконометрики: учебник для вузов/ С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.

2 Большаков, А.А. Методы обработки многомерных данных и временных рядов: учебное пособие для вузов / А.А. Большаков, Р.Н. Каримов – М.: Горячая линия – Телеком, 2007. – 522 с.

3 Магнус, Я.Р. Эконометрика. Начальный курс: учебник/ Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 57 с.

4 Мхитарян, В.С. Эконометрика: учебник / под ред. В.С. Мхитаряна. – М: Проспект, 2009.-384 с.

5 Тихомиров, Н.П. Эконометрика: учебник/ Н.П. Тихомиров, Е.Ю. Дорохина. – М.: Издательство «Экзамен», 2003. – 512 с.

6 Чураков, Е.П. Математические методы обработки экспериментальных данных в экономике: учеб.пособие/ Е.П. Чураков – М.: Финансы и статистика, 2004. – 240с.:ил.

Приложение А

(обязательное)

Исходные данные для анализа

Таблица А.1 - Значения социально-экономических показателей, характеризующих города и районы Оренбургской области

Номер объекта	Муниципальные образования	X1	X2	X3	X4	X5	X6	X7	X8
1	2	3	4	5	6	7	8	9	10
1	Абдулинский	716,59	0	50	71,8	0,00	21,90	53,09	16,64
2	Адамовский	4791,44	10296,08	16,7	46,2	0,00	8,42	61,90	22,39
3	Акбулакский	5677,90	1478,16	52,9	47,3	0,18	11,80	62,56	21,02
4	Александровский	1571,20	377,53	45,5	0	0,00	14,55	60,55	20,02
5	Асекеевский	3704,46	642,03	40,9	57,5	0,54	13,58	58,68	18,11
6	Беляевский	3304,59	341,69	72,7	22,9	0,22	17,10	61,23	19,72
7	Бугурусланский	4367,39	261,84	58,3	28,7	0,00	14,97	59,99	17,51
8	Бузулукский	2127,96	1111,62	52,6	72,8	0,36	10,76	58,65	17,74
9	Гайский	13657,15	0,00	16,7	15,1	0,00	15,65	59,70	20,32
10	Грачевский	2252,99	1385,21	55,6	73	0,13	10,93	60,46	18,07
11	Домбаровский	2242,38	508,48	81,8	25,6	0,00	6,73	62,41	22,86
12	Илекский	2803,27	505,37	42,9	4,9	3,53	11,41	59,78	19,15
13	Кваркенский	1984,05	3094,73	41,2	35,6	1,49	10,29	60,23	21,36
14	Красногвардейский	3618,35	1314,39	25	8	0,44	11,35	60,34	20,93
15	Кувандыкский	2438,19	0,00	52,9	33,3	0,00	14,47	59,02	20,38
16	Курманаевский	2074,29	0,00	81,2	57,9	1,58	21,11	60,02	17,32
17	Матвеевский	2172,78	102,99	11,1	8,3	0,00	14,66	58,83	18,09
18	Новоорский	10893,40	82540,63	61,9	8,4	0,00	12,86	61,61	19,81
19	Новосергиевский	5723,31	4935,74	40,5	20,6	0,16	11,94	59,04	19,02
20	Октябрьский	4967,20	444,28	31,2	46,7	0,13	12,81	60,50	18,03
21	Оренбургский	20071,10	25359,07	31,1	5,8	0,01	4,83	63,93	18,41
22	Первомайский	1795,32	3312,16	17,6	18,4	1,09	18,87	62,33	22,06
23	Переволоцкий	3561,15	86,88	30,8	0,8	0,32	14,63	60,62	18,92
24	Пономаревский	2217,02	184,32	81,8	52,2	0,00	16,39	57,82	16,93
25	Сакмарский	4551,40	374,80	14,3	1,4	0,03	9,11	63,10	18,20
26	Саракташский	3384,80	3525,57	40	13,6	0,00	2,30	59,67	18,21
27	Светлинский	3775,83	12159,95	64,3	12,6	0,43	10,12	61,36	21,16
28	Северный	2264,20	0,00	21,4	58,8	0,00	7,12	58,67	17,78
29	Соль-Илецкий	1047,46	358,37	81	43,1	0,88	8,19	59,74	22,86
30	Сорочинский	2833,94	13,55	33,3	40,9	0,00	26,12	56,48	20,25
31	Ташлинский	6881,07	5509,45	55,6	12,7	1,08	8,45	61,57	20,67
32	Тоцкий	1755,21	159,03	54,2	13	0,04	16,18	72,49	13,84
33	Тюльганский	3196,66	1403,25	50	14,2	2,48	15,21	62,25	18,85
34	Шарлыкский	3649,02	299,66	26,3	43,6	0,05	12,45	57,57	16,84
35	Ясненский	7148,83	0,00	100	84,4	0,63	34,94	60,79	24,15
36	Абдулино	2784,39	3277,23	25	27,4	0,02	7,49	64,93	19,16
37	Бугуруслан	4229,97	191924,71	38,9	20,6	0,05	15,35	62,18	18,16
38	Бузулук	61679,53	240951,01	27,7	0,3	0,09	6,19	64,27	15,81
39	Гай	27338,48	106449,61	10	7,8	0,04	1,82	65,93	15,24
40	Кувандык	2012,36	20786,78	27,3	0,7	0,00	7,73	63,44	17,12
41	Медногорск	11170,01	27319,93	31,2	1,5	0,00	18,05	63,00	16,65

Продолжение таблице А.1

1	2	3	4	5	6	7	8	9	10
42	Новотроицк	29743,64	217430,62	39,5	14,8	0,00	13,14	60,46	15,03
43	Оренбург	21460,65	8736,67	22,3	6,7	0,01	25,37	64,76	15,29
44	Орск	4301,33	139154,85	28,8	14,1	0,00	2,96	66,55	15,09
45	Соль-Илецк	4401,00	12593,97	42,9	6,4	0,00	0,00	63,42	16,20
46	Сорочинск	3446,14	315863,20	12,5	3,3	0,02	14,08	63,03	20,16
47	Ясный	3539,32	29399,98	50	46,3	0,00	7,82	63,26	18,17

Таблица А.2 - Значения социально-экономических показателей, характеризующих города и районы Оренбургской области

Номер объекта	Муниципальные образования	X9	X10	X11	X12	X13	X14	X15	X16
1	2	3	4	5	6	7	8	9	10
1	Абдулинский	751	0	31,35	2226,11	0,00	5158,03	329,71	99,63
2	Адамовский	2910	355	2,25	135701,34	-0,02	5908,71	2008,04	144,78
3	Акбулакский	1357	263	0,11	-8567,61	0,61	4379,21	1458,21	142,39
4	Александровский	969	26	0,17	-36522,68	0,00	6962,00	1821,81	140,83
5	Асекеевский	1643	141	0,37	17280,55	3,50	4529,49	2005,23	124,18
6	Беляевский	1502	173	0,17	-23702,21	0,00	5330,26	1583,71	131,50
7	Бугурусланский	2158	43	0,07	2327,56	-3,36	6830,25	1283,09	150,45
8	Бузулукский	1829	574	0,77	-20227,66	2,15	3813,11	1556,33	189,27
9	Гайский	1622	67	0,15	90494,93	0,00	5260,96	1543,57	132,57
10	Грачевский	1306	90	0,22	-21387,79	-4,28	5562,88	2376,39	179,24
11	Домбаровский	716	21	0,60	-66252,50	-13,30	4790,24	1855,85	158,35
12	Илекский	2098	151	2,94	-38968,18	-7,22	4117,78	1780,37	113,90
13	Кваркенский	1904	319	1,83	95392,98	9,08	4916,49	1746,81	137,01
14	Красногвардейский	814	215	0,92	-6880,28	-20,36	5483,97	1738,92	169,90
15	Кувандыкский	1529	50	0,73	-12601,50	0,00	2805,11	660,66	102,55
16	Курманаевский	1223	40	0,08	-19203,87	0,00	5175,18	1614,86	183,10
17	Матвеевский	1223	51	0,12	27154,79	0,00	8012,55	1479,75	114,97
18	Новоорский	531	1468	1,11	-88359,71	-20,64	6883,78	2791,04	219,82
19	Новосергиевский	2747	998	1,52	53771,06	57,54	12916,69	2447,89	164,51
20	Октябрьский	2019	221	0,07	6046,84	0,00	7530,85	2109,25	166,12
21	Оренбургский	2965	1984	6,90	222587,21	40,10	10051,88	9987,58	414,60
22	Первомайский	1023	161	0,11	11834,97	2,66	4820,43	1600,04	191,82
23	Переволоцкий	1548	110	2,12	5089,93	0,00	7200,79	2466,00	149,14
24	Пономаревский	609	27	17,11	-4358,73	0,00	9429,51	1663,79	153,87
25	Сакмарский	1415	714	2,07	47042,61	-18,18	5460,64	2252,94	171,96
26	Саракташский	2855	554	0,39	18636,05	-17,28	7163,74	2442,46	148,52
27	Светлинский	1261	739	0,28	-31576,96	1,10	5828,43	3104,97	173,23
28	Северный	842	55	0,29	12573,25	0,00	10527,40	2041,39	164,51
29	Соль-Илецкий	2160	238	1,42	-27755,42	2,87	9624,39	749,79	87,13
30	Сорочинский	2366	611	0,23	-41927,34	0,00	3258,68	992,49	125,95
31	Ташлинский	3706	448	0,43	19211,06	5,86	8003,63	1932,97	113,73

Продолжение таблицы А.2

1	2	3	4	5	6	7	8	9	10
32	Тоцкий	913	163	0,17	2703,44	-13,48	5710,41	1623,23	155,14
33	Тюльганский	1395	260	0,14	-1805,00	-5,70	6922,20	2492,35	135,65
34	Шарлыкский	1386	122	2,09	29131,25	0,00	8597,17	1848,44	140,58
35	Ясненский	565	0	3,86	-70126,16	0,00	5709,72	2688,39	108,26
36	Абдулино	0	1062	10,65	28860,67	3,23	20528,50	5245,52	253,70
37	Бугуруслан	0	3504	0,76	93608,09	3,43	12009,56	6175,02	254,24
38	Бузулук	0	12002	6,22	8735783,87	41,62	16533,55	8125,77	362,54
39	Гай	0	8180	0,38	1526864,98	4,32	10133,53	7199,35	316,32
40	Кувандык	0	2192	0,35	27775,53	1,38	11328,86	7323,76	192,27
41	Медногорск	0	4033	2,18	321986,13	6,97	10895,88	5295,27	248,44
42	Новотроицк	199	24413	13,08	1476312,51	18,23	13505,43	6902,37	306,62
43	Оренбург	840	52066	26,49	144294,97	8,97	57813,57	10336,03	284,29
44	Орск	0	24492	0,47	710226,74	2,28	15867,96	6455,67	339,72
45	Соль-Илецк	0	1035	0,15	29769,64	6,72	14853,78	7020,26	221,38
46	Сорочинск	0	879	0,85	40383,38	23,39	14838,68	7599,74	241,49
47	Ясный	0	4084	0,87	282356,73	34,75	9427,45	7508,34	263,32

Таблица А.3 – Наименование показателей

Обозначения	Наименование показателя
X1	Объем инвестиций в основной капитал на душу населения, рублей
X2	Объем промышленной продукции на душу населения, рублей
X3	Удельный вес убыточных предприятий и организаций, в процентах от общего числа предприятий
X4	Просроченная кредиторская задолженность предприятий, в процентах от общей задолженности
X5	Задолженность организаций по заработной плате, в процентах от общего фонда заработной платы
X6	Уровень безработицы, в процентах от населения в трудоспособном возрасте
X7	Доля населения в трудоспособном возрасте в общей численности населения, в процентах
X8	Доля лиц моложе трудоспособного возраста, в общей численности населения, в процентах
X9	Среднегодовая численность работников, занятых в сельскохозяйственном производстве, человек
X10	Среднегодовая численность работников, занятых в промышленности, человек
X11	Число зарегистрированных иностранных рабочих, в промилле от численности населения в трудоспособном возрасте
X12	Сальдированный финансовый результат (прибыль минус убыток) на одно предприятие, рублей
X13	Уровень рентабельности реализованной продукции сельского хозяйства в сельскохозяйственных организациях, в процентах
X14	Оборот розничной торговли на душу населения, рублей
X15	Объем платных услуг на душу населения, рублей
X16	Соотношение среднемесячной номинальной начисленной заработной платы работников с величиной прожиточного минимум, в процентах

Таблица А.4 – Варианты заданий

Номер варианта	Результативный признак, (обозначить Y)	Номера факторных признаков, *X*
1	X1	4,6,10,11,14
2	X1	5,10,11,14,15
3	X1	2,10,11,13,14
4	X1	6,7,10,12,15
5	X1	4,5,6,10,15
6	X1	3,10,11,12,15
7	X1	2,12,13,14,15
8	X1	2,9,11,14,15
9	X1	3,510,12,13
10	X1	4,5,14,15,16
11	X2	3,12,13,14,15
12	X2	4,7,11,12,13
13	X2	4,10,12,14,16
14	X2	1,9,13,15,16
15	X2	9,10,12,14,16
16	X2	9,10,13,15,16
17	X2	1,4,6,7,15
18	X3	1,4,6,8,13
19	X4	3,6,7,15,16
20	X4	2,3,6,15,16