Министерство образования и науки
Российской Федерации

Государственное образовательное учреждение
высшего профессионального образования
«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

В.И. Васянина, Ю.А. Жемчужникова,
О.И. Стебунова

ОБОБЩЕННАЯ ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ С ГЕТЕРОСКЕДАСТИЧНЫМИ ОСТАТКАМИ В ПАКЕТЕ STATISTICA

Методические указания
к семинарским занятиям, лабораторному практикуму, курсовым работам, дипломному проектированию и самостоятельной работе студентов

Рекомендовано к изданию Редакционно-издательским советом Государственного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет»

Оренбург
ИПК ГОУ ОГУ
2011

УДК 330: 519.862
ББК 65В631
В 19

Рецензент - доцент, кандидат экономических наук Е.С.Щукина

Васянина, В.И.
В19 Обобщенная линейная модель множественной регрессии с гетероскедастичными остатками в пакете Statistica: методические указания к лабораторному практикуму, семинарским занятиям, курсовой работе, дипломному проектированию и самостоятельной работе студентов / В.И. Васянина, Ю.А. Жемчужникова, О.И. Стебунова; Оренбургский гос. ун-т.– Оренбург: ОГУ, 2011. – 37 с.

Методические указания к семинарским занятиям, лабораторному практикуму, самостоятельной работе студентов, в том числе для выполнения расчетно-графических заданий, курсовых и дипломных работ, связанных с регрессионным анализом. Предназначены для специальности 080116 – Математические методы в экономике, направлений 231300 – Прикладная математика, 080500 – Бизнес-информатика и других экономических специальностей и направлений, изучающих дисциплины, использующие инструментарий регрессионного анализа.

Содержание

Введение

Для исследования регрессионных взаимосвязей между показателями в области экономики достаточно типична ситуация, связанная с неравноточностью измерений (наблюдений). Это связано с тем, что дисперсии регрессионных остатков, соответствующие значениям объясняющей переменной x_j, если они характеризуют объекты, различающиеся по своим масштабам, могут быть различными. Например, при исследовании зависимости среднедушевых сбережений от дохода, вариация среднедушевых сбережений в семьях с более высокими доходами, будет отличаться от вариации среднедушевых сбережений для семей с более низкими доходами, т.е. дисперсия регрессионных остатков не постоянна.

Игнорирование гетероскедастичности регрессионных остатков сказывается на свойствах оценок и может вести к недостоверным статистическим выводам. В связи с этим актуальными являются вопросы, связанные с выявлением гетероскедастичности, ее тестированием, способами устранения, либо уточнения результатов.

Цель работы заключается в выработке навыков исследования регрессионных моделей с гетероскедастичными остатками.

1 Теоретическая часть

1.1 Общая постановка задачи регрессионного анализа

Изучается регрессионная зависимость результативной переменной y от объясняющих переменных $x = (x_1, x_2, ..., x_k)^T$

$$\tilde{y} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k, \tag{1.1}$$

где $\tilde{y}$ – условное среднее значение результативной переменной y.

Результаты наблюдений результативной и объясняющих переменных для «*n*» объектов представлены вектором $\boxed{Y = (y_1, ..., y_n)^T}$ и матрицей X типа «объект-свойство» наблюденных значений признаков $x_1, ...x_k$:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ . & . & . & \dots & . \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}.$$

Связь между наблюденными значениями Y и X в данном случае имеет вид:

$$Y = X\beta + Z \tag{1.2}$$

где $\beta = (\beta_0 \beta_1 \dots \beta_k)^T$ - вектор коэффициентов линейной модели множественной регрессии (ЛММР);

$Z = (z_1, ..., z_n)^T$ - возможные значения случайного вектора ε - характеризующие отклонения наблюденных значений y_i от модельных значений $\tilde{y}_i$ для *i*–го объекта.

На y смотрим как на возможные значения случайной величины η, где η_i - случайная величина, для которой y_i - наблюденное значение на *i*–м объекте наблюдения $i = \overline{1, n}$.

Тогда выборочная модель имеет вид:

$$\eta_{1,n} = X\beta + \varepsilon \ , \qquad (1.3)$$

где $\eta_{1,n} = \left(\eta_1,\ \eta_2,\ \ldots,\eta_n\right)^T$ - случайный вектор, а (1.2) –реализация этой модели.

В рамках классической линейной модели множественной регрессии предполагается выполнение всех условий Гаусса-Маркова

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

2) ранг матрицы X равен "$к+1$" – среди признаков нет линейно зависимых;

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

4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2$, $\boxed{i = \overline{1,n}}$

5) $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0$, $\boxed{i \neq j}$, $\boxed{i = \overline{1,n}}\boxed{j = \overline{1,n}}$

4`) $\Sigma_\varepsilon = M(\bar{\varepsilon}\bar{\varepsilon}^T) = \sigma^2 E_n$.

Предположим, что нарушено 4-е условие Гаусса – Маркова, т.е. $D\varepsilon_i = M\varepsilon_i^2 = \sigma_{i^2}$, где $\boxed{i = \overline{1,n}}$, тогда ЛММР (1.3) является **обобщенной линейной моделью множественной регрессии с гетероскедастичными остатками.** Ковариационная матрица регрессионных остатков будет иметь вид:

$$\Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \sigma_n^2 \end{pmatrix} .$$

Условие 4`) можно записать в виде: $\Sigma_\varepsilon = M(\bar{\varepsilon}\bar{\varepsilon}^T) = \sigma_0^2 \Sigma_0$, где Σ_0 - некоторая симметричная положительно-определенная матрица с неравными элементами на главной диагонали.

1.2 ОМНК – оценки ОЛММР

Если игнорировать гетероскедастичность регрессионных остатков и оценить коэффициенты ЛММР обычным методом наименьших квадратов (МНК), то оценки коэффициентов регрессионной модели остаются несмещенными и состоятельными, при тех же условиях, что и в КЛММР (/1/). Однако оценка ковариационной матрицы $\hat{\Sigma}_{\hat{\beta}_{мнк}}$ является смещенной и таким образом, оценка $\hat{\beta}_{мнк}$ не является эффективной.

Для ОЛММР несмещенные, состоятельные и эффективные оценки $\hat{\beta}$ можно получить с помощью **обобщенного метода наименьших квадратов** (ОМНК)

$$\hat{\beta}_{ОМНК} = \left(X^T \Sigma_0^{-1} X\right)^{-1} X^T \Sigma_0^{-1} Y \,, \tag{1.4}$$

который требует знания матрицы Σ_0 с точностью до постоянного множителя, совпадающим с Σ_ε. Но для этого надо найти матрицу C, такую что выполняется соотношение $\Sigma_0 = CC^T$, (где C - квадратная, невырожденная, ортогональная матрица), и с помощью умножения правой и левой части (1.3) на C^{-1}:

$$C^{-1}\eta_{1,n} = C^{-1}X\beta + C^{-1}\varepsilon \,, \qquad \left(C^{-1}Y = C^{-1}X\beta + C^{-1}Z\right)$$

перейти к модели вида (1.5):

$$\eta_{1,n\,пр} = X_{пр}\beta + \varepsilon_{пр} \qquad \left(Y_{пр} = X_{пр}\beta + Z_{пр}\right) \tag{1.5}$$

в которой $\varepsilon_{пр}$ удовлетворяет условиям КЛММР.

Таким образом, оценка параметров β получена для КЛММР (1.5):

$$\hat{\beta}_{ОМНК} = \left(X_{пр}^{Т} X_{пр}\right)^{-1} X_{пр}^{Т} Y_{пр} \tag{1.6}$$

Несмещенная оценка остаточной дисперсии, ковариационной матрицы $\Sigma_{\hat{\beta}_{\text{ÎÌÍÊ}}}$ имеют вид:

$$\hat{S}^{2}_{\text{îñò}} = \frac{1}{n-k-1}\left(Y - X\hat{\beta}_{\text{ÎÌÍÊ}}\right)^{T}\Sigma_{0}^{-1}\left(Y - X\hat{\beta}_{\text{ÎÌÍÊ}}\right), \tag{1.7}$$

$$\hat{\Sigma}_{\hat{\beta}_{\text{ÎÌÍÊ}}} = \hat{S}^{2}_{\text{îñò}\ (\text{ÎÌÍÊ})}(X^{T}\Sigma_{0}^{-1}X)^{-1}. \tag{1.8}$$

Несмещенная оценка факторной дисперсии:

$$\hat{S}^{2}_{\text{ôàêò}} = \frac{1}{k}\left(X\hat{\beta}_{\text{ÎÌÍÊ}} - \overline{Y}\right)^{T}\Sigma_{0}^{-1}\left(X\hat{\beta}_{\text{ÎÌÍÊ}} - \overline{Y}\right) \tag{1.9}$$

Выборочный коэффициент детерминации определяется по формуле:

$$\hat{R}^{2}_{\text{ÎÌÍÊ}} = 1 - \frac{\left(Y - X\hat{\beta}_{\text{ÎÌÍÊ}}\right)^{T}\Sigma_{0}^{-1}\left(Y - X\hat{\beta}_{\text{ÎÌÍÊ}}\right)}{\left(Y-\overline{Y}\right)^{T}\Sigma_{0}^{-1}\left(Y-\overline{Y}\right)} \tag{1.10}$$

Для ОМНК-оценок коэффициентов ОЛММР $\hat{R}^{2}_{ОМНК}$ может принимать как значения больше 1, так и отрицательные значения и при анализе модели регрессии используется лишь как приближенная характеристика.

1.3 Внешние признаки и тесты для проверки гипотезы о наличии/отсутствии гетероскедастичности

Выяснить, под влиянием какой объясняющей переменной появляется гетероскедастичность регрессионных остатков можно визуально.

Для этого строим МНК-оценки параметров модели регрессии, находим оценки регрессионных остатков $\hat{z}_i$ и изучаем характер изменения регрессионных остатков в зависимости от изменения анализируемой объясняющей переменной. Если по мере возрастания упорядоченной объясняющей переменной регрессионные остатки $|\hat{\varepsilon}_i|$ возрастают (или убывают), то на основе визуального анализа делаем предположение о наличии гетероскедастичности, порождаемой соответствующей переменной. Данная процедура проделывается для каждой из объясняющих переменных.

Предположение о наличии гетероскедастичности проверяется с помощью различных тестов.

1.3.1 Тест Голдфелда-Квандта

Этот тест применяется в том случае, если регрессионные остатки можно считать нормально распределенными случайными величинами. При выполнении данного теста будем считать, что дисперсия регрессионных остатков прямо или обратно пропорциональна значению объясняющей переменной (x_l), вариацией которых порождается гетероскедастичность.

Выдвигается гипотеза:

$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$ (нет гетероскедастичности)

$H_1 : \exists\, i \neq j : \sigma_i^2 \neq \sigma_j^2$ (есть гетероскедастичность)

Шаги теста:

1) Проранжировать в порядке возрастания значения объясняющей переменной, которая подозревается на порождение гетероскедастичности (x_l);

2) Упорядочить наблюдаемые значения результативного признака и объясняющих переменных в порядке возрастания объясняющей переменной (x_l);

3) Взять n′ первых наблюдаемых значений результативного признака (y′) и объясняющих переменных (X′) и n″ последних наблюдаемых значений, соответственно обозначив y″ и X″: $n' = n'' = \dfrac{n - 0.25n}{2}$;

4) Оцениваются уравнения регрессии y′ по n′ значениям и y″ по n″.

5) Вычисляется оценка регрессионных остатков $\overline{\hat{\varepsilon}'}$ и $\overline{\hat{\varepsilon}''}$ и их суммы квадратов отклонений: $Q' = (\overline{\hat{\varepsilon}'})^T \cdot \overline{\hat{\varepsilon}'}$ и $Q'' = (\overline{\hat{\varepsilon}''})^T \cdot \overline{\hat{\varepsilon}''}$.

6) Для проверки нулевой гипотезы строится статистика $F = \frac{\max\{Q_1;Q_2\}/n' - k - 1}{\min\{Q_1;Q_2\}/n' - k - 1}$, которая в случае справедливости нулевой гипотезы имеет закон распределения Фишера-Снедекора с числом степеней свободы $\nu_1 = n' - k - 1$, $\nu_2 = n'' - k - 1$.

Если $Q_1 < Q_2$, то наблюдается прямая зависимость между регрессионными остатками и объясняющей переменной (x_l), матрица

$$\hat{\Sigma}_0 = \begin{pmatrix} x_{1l}^2 & 0 & \cdots & 0 \\ 0 & x_{2l}^2 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & x_{nl}^2 \end{pmatrix}$$

Если $Q_1 > Q_2$, то наблюдается обратная зависимость между регрессионными остатками и объясняющей переменной (x_l), матрица

$$\hat{\Sigma}_0 = \begin{pmatrix} \frac{1}{x_{1l}^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_{2l}^2} & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_{nl}^2} \end{pmatrix}.$$

1.3.2 Тест Глейзера

В рамках рассмотренных ранее тестов, достаточно грубо оценивается характер вариации дисперсий. Более тонким в этом плане является тест Глейзера, который предполагает другие виды зависимостей между дисперсией регрессионных остатков и объясняющей переменной (x_l). Как и ранее, находятся МНК-оценки, находится оценка регрессионных остатков $\hat{z}_i$ и ищется зависимость абсолютных значений оценок регрессионных остатков от x_l:

$$|\hat{z}_i| = \alpha + \beta |x_{il}|^{\gamma} + \delta_i\,, \tag{1.11}$$

γ – параметр, который находится из промежутка $[-1;2]$

δ_i - случайная компонента, удовлетворяющая свойствам:

$$M(\delta_i, \delta_j) = \begin{cases} 0, \text{ если } i = j \\ \sigma_\delta^2, \text{ если } i \neq j \end{cases}$$

Варьируя γ оценивают модель регрессии. Если при оценивании есть значимые модели, то выбирают модель с наибольшим коэффициентом детерминации, если нет значимых - то делается вывод о гомоскедастичности регрессионных остатков.

Оценка матрицы Σ_0 для реализации ОМНК имеет вид:

$$\hat{\Sigma}_0 = \begin{pmatrix} \left(\hat{\alpha}+\hat{\beta}|x_{1j}|^{\gamma}\right)^2 & 0 & 0\ldots & 0 \\ 0 & \left(\hat{\alpha}+\hat{\beta}|x_{2j}|^{\gamma}\right)^2 & 0\ldots. & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \left(\hat{\alpha}+\hat{\beta}|x_{nj}|^{\gamma}\right)^2 \end{pmatrix}$$

Следует отметить, что кроме вышеописанных критериев могут использоваться тест ранговой корреляции Спирмена, тесты Бартлета, Бреуша-Пагана, Уайта и др. [1, 2, 4, 5,6].

1.4 Уточнение стандартных ошибок в форме Уайта и Невье-Веста

В тех случаях, когда основания подозревать гетероскедастичность есть, а способа выявить ее нет, т.к. она может формироваться под влиянием множества факторов, мы не можем выявить и оценить матрицу Σ_0 . Если использовать МНК-оценки, то как было отмечено ранее, они не являются эффективными, а смещенность ковариационной матрицы может привести к неверным статистическим выводам и неправильно характеризовать точность оценок. Рекомендуется необходимо уточнить стандартные ошибки коэффициентов модели, в форме, предложенной Уайтом и Невье - Вестом, которые приведены ниже.

Стандартные ошибки в форме Уайта вычисляются по формуле:

$$\hat{\Sigma}_b = n(X^T X)^{-1}\left(\frac{1}{n}\sum_{s=1}^{n}\hat{z}_s^2 x_s x_s^T\right)\cdot(X^T X)^{-1} \tag{1.12}$$

где x_s^T, s =1,..n. 1 ¿ k – векторы строки матрицы X
является состоятельной оценкой матрицы ковариаций оценок коэффициентов регрессии.

Для более сложного случая, когда в ковариационной матрице регрессионных остатков ненулевые элементы стоят не только на главной диагонали, но и на соседних диагоналях, отстоящих на главной не более чем на l, т.е w_{ij} =0, если $|i-j|>l$, рассчитываются стандартные ошибки в форме Невье – Веста:

$$\hat{\Sigma}_b = n(X^T X)^{-1}\left(\frac{1}{n}\sum_{s=1}^{n}\hat{z}_s^2 x_s x_s^T + \frac{1}{n}\sum_{j=1}^{l}\sum_{t=j+1}^{n} w_j \hat{z}_t \hat{z}_{t-j}(x_t x_{t-j}^T + x_{t-j} x_t^T)\right)\cdot(X^T X)^{-1} \tag{1.13}$$

Существует несколько способов выбора весовых коэффициентов w_j .

1. Наиболее простой w_j =1. Однако при таком выборе матрица может оказаться не положительно определенной.
2. $w_j = \frac{1}{l-1}$ (Барлетт)
3. $w_j = \begin{cases} 1-6\left(\frac{j}{l+1}\right)^2 + 6\left(\frac{j}{l+1}\right)^3, при 1 \le j \le \frac{l+1}{2} \\ 2\left(1-\frac{j}{l+1}\right)^2 при \frac{l+1}{2} < j \le l \end{cases}$ (Парзен)

1.5 Вопросы для практическо-семинарских занятий по теме «ОЛММР с гетероскедастичными остатками»

Группа А – базовые вопросы по лекционному материалу

1. Какими свойствами обладает МНК-оценка ОЛММР с гетероскедастичными остатками?
2. Как получить ОМНК – оценку вектора параметров β для ОЛММР?
3. Докажите статистические свойства ОМНК-оценок.
4. Приведите характеристики качества модели ОЛММР с гетероскедастичными остатками.
5. Как оценивается ковариационная матрица $\hat{\Sigma}_\beta$ в рамках КЛММР и в рамках ОЛММР с гетероскедастичными остатками?
6. Как проверить гипотезу о незначимости ОЛММР с гетероскедастичными остатками?
7. Как проверить гипотезу о незначимости отдельных коэффициентов ОЛММР гетероскедастичными остатками?
8. Назовите возможные причины, порождающие гетероскедастичность.
9. Перечислите последствия гетероскедастичности.
10. Как можно выявить гетероскедастичность графически?
11. Опишите алгоритм критерия Голдфелда-Квандта.
12. Когда применяется тест Глейзера. Опишите его.
13. В чем суть взвешенного МНК?
14. Для уравнения регрессии $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ имеет место следующее соотношение: $D\varepsilon_i = \sigma^2 x_i^2$. Какое преобразование можно предложить, чтобы устранить проблему гетероскедастичности?
15. Запишите стандартные ошибки в форме Уайта. Для чего они применяются?
16. В каких случаях производится коррекция стандартных ошибок в форме Невье-Веста?

17. Как осуществить точечный прогноз значения результативного показателя в условиях ОЛММР с гетероскедастичными остатками?
18. Опишите процедуру построения интервального прогноза значения результативного показателя в условиях ОЛММР с гетероскедастичными остатками.

Группа В – вопросы, требующие самостоятельной подготовки

1. Докажите, что МНК-оценки ОЛММР с гетероскедастичными остатками остаются несмещенными и состоятельными при тех же условиях, что и в классической модели [1, с.677].
2. Докажите, что МНК-оценки ОЛММР с гетероскедастичными остатками не являются эффективными [1, с.676].
3. Объясняете, почему в ОЛММР мы не можем утверждать, что коэффициент $\overset{¿}{R}^2_{ОМНК}$ не заключен в промежутке $[0\,;1]$ [1, с. 680].
4. Какие предположения делаются относительно дисперсии регрессионных остатков в тесте ранговой корреляции Спирмена при выявлении гетероскедастичности. [3, с. 231].
5. Приведите этапы реализации теста ранговой корреляции Спирмена для выявления гетероскедастичности. Какую структуру будет иметь матрица Σ_0 в случае гетероскедастичности регрессионных остатков? [3, с. 231].
6. В чем заключается тест Уайта на выявление гетероскедастичности? [2, с. 177; 6, с. 161]
7. Как проверить гипотезу об отсутствии гетероскедастичности с помощью критерия Бреуша – Пагана? Как реализуется в данном случае ОМНК в условиях ОЛММР с гетероскедастичными остатками? [2, с. 179; 4, с.96].
8. В каких случаях используется критерий Бартлетта? Какую структуру будет иметь матрица Σ_0 в случае гетероскедастичности регрессионных остатков? [2, с. 98].

2 Практическая часть

2.1 Содержание лабораторной работы

Выполнение лабораторной работы по теме «ОЛММР с гетероскедастичными остатками» состоит из следующих этапов:

- ознакомление с формулировкой задания к лабораторной работе и порядком её выполнения в пакетах прикладных программ;
- выполнение расчетов на компьютере по данным своего варианта;
- анализ полученных результатов;
- подготовка письменного отчета по лабораторной работе;
- защита лабораторной работы.

2.2 Задание к лабораторной работе

По данным Приложения А:

1) построить МНК-оценки коэффициентов линейной модели множественной регрессии;

2) исследовать регрессионные остатки на гетероскедастичность, используя различные тесты;

3) при необходимости построить ОМНК-оценки параметров регрессионной модели или найти несмещенную оценку ковариационной матрицы вектора оценок коэффициентов в форме Уайта и Невье-Веста.

2.3 Порядок выполнения лабораторной работы в пакете Statistica

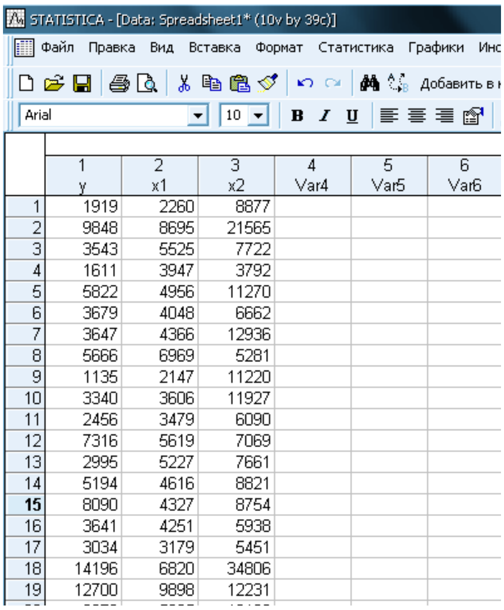
Ищется зависимость ввода в действие жилых домов на 1000 человек населения, кв.м. (y) от среднесписочной численности работников, человек (x_1) и инвестиции в основной капитал на душу населения, руб. (x_2):

$$\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Объектом исследования выступают города и районы Оренбургской области. Предметом исследования – взаимосвязи между вводом в действие жилых домов и указанными экономическими показателями.

Информационная база представлена данными о значениях соответствующих показателей для 39 городов и районов Оренбургской области за 2007 год.

Введем данные для анализа, как представлено на рисунке 2.1.

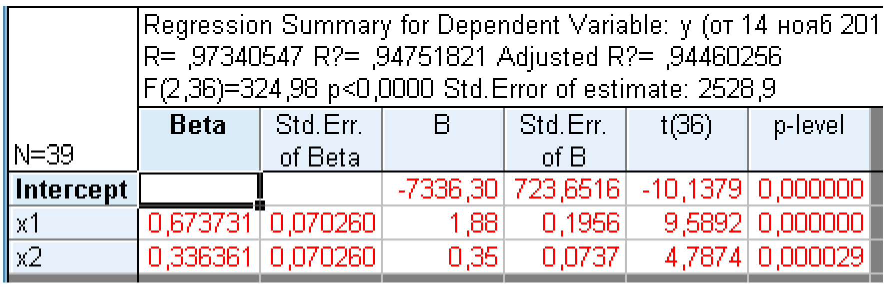


	1 y	2 x1	3 x2	4 Var4	5 Var5	6 Var6
1	1919	2260	8877			
2	9848	8695	21565			
3	3543	5525	7722			
4	1611	3947	3792			
5	5822	4956	11270			
6	3679	4048	6662			
7	3647	4366	12936			
8	5666	6969	5281			
9	1135	2147	11220			
10	3340	3606	11927			
11	2456	3479	6090			
12	7316	5619	7069			
13	2995	5227	7661			
14	5194	4616	8821			
15	8090	4327	8754			
16	3641	4251	5938			
17	3034	3179	5451			
18	14196	6820	34806			
19	12700	9898	12231			

Рисунок 2.1- Исходные данные для анализа

Для оценки параметров регрессионной модели воспользуемся методом пошаговой регрессии. Процедура построения уравнения множественной регрессии более подробно рассмотрена в лабораторной работе №1.

Результаты оценивания представлены на рисунке 2.2.



Regression Summary for Dependent Variable: y (от 14 нояб 201
R= ,97340547 R?= ,94751821 Adjusted R?= ,94460256
F(2,36)=324,98 p<0,0000 Std.Error of estimate: 2528,9

N=39	Beta	Std.Err. of Beta	B	Std.Err. of B	t(36)	p-level
Intercept			-7336,30	723,6516	-10,1379	0,000000
x1	0,673731	0,070260	1,88	0,1956	9,5892	0,000000
x2	0,336361	0,070260	0,35	0,0737	4,7874	0,000029

Рисунок 2.2 - Результаты оценивания параметров регрессионной модели

Далее можно приступить к исследованию остатков регрессионной модели. Остатки исследуются в специальном окне **Residuals analysis – Анализ остатков**. Для этого необходимо щелкнуть по кнопке **Residuals/assumptions/prediction – Остатки/распределение/предсказанные** в окне – **Multiple Regression.** Формально проверим тест на нормальный характер распределения регрессионных остатков, для этого в меню системы Statistica выберем пункт **Distribution Fitting**. Результаты исследования регрессионных остатков представлены на рисунке 2.3.

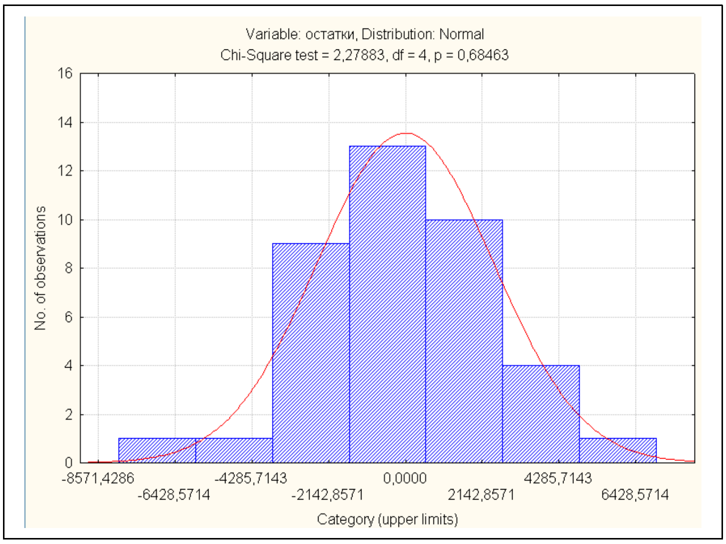


Рисунок 2.3 - Гистограмма распределения регрессионных остатков

Результаты формальной проверки гипотезы о нормальном характере распределения регрессионных остатков позволяют ее не отвергнуть, и есть смыл проводить дальнейший анализ построенного уравнения множественной регрессии.

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = \underset{(723,65)}{-7336,3} + \underset{(0,196)}{1,88}\, x_1 + \underset{(0,074)}{0,35}\, x_2 \quad (2.1)$$

Как видно из отчета, коэффициент при переменной x_1 и x_2 значимо отличаются от нуля. Регрессионная модель адекватна экспериментальным данным.

Поскольку построили значимую регрессионную модель, то следующим этапом является исследование регрессионных остатков на наличие/отсутствие гетероскедастичности.

Проверим наличие/отсутствие гетероскедастичности по переменной x_1.

Наличие гетероскедастичности можно предположить по графику зависимости остатков $|\hat{z}_i|$ от упорядоченных по возрастанию значений той объясняющей переменной, вариацией которой возможно порождается гетероскедастичность. Для построения графика воспользуемся MS Excel. Для этого из ППП «Statistica» в MS Excel скопируем столбцы регрессионных остатков $\hat{z}_i$ и значений объясняющей переменной x_1.

Упорядочим регрессионные остатки $\hat{z}_i$ по возрастанию значений x_1 с помощью команды **«Данные» - «Сортировка»**. В появившемся окне выбираем сортировать по «x_1», затем по «остаткам» (рисунок 2.4).

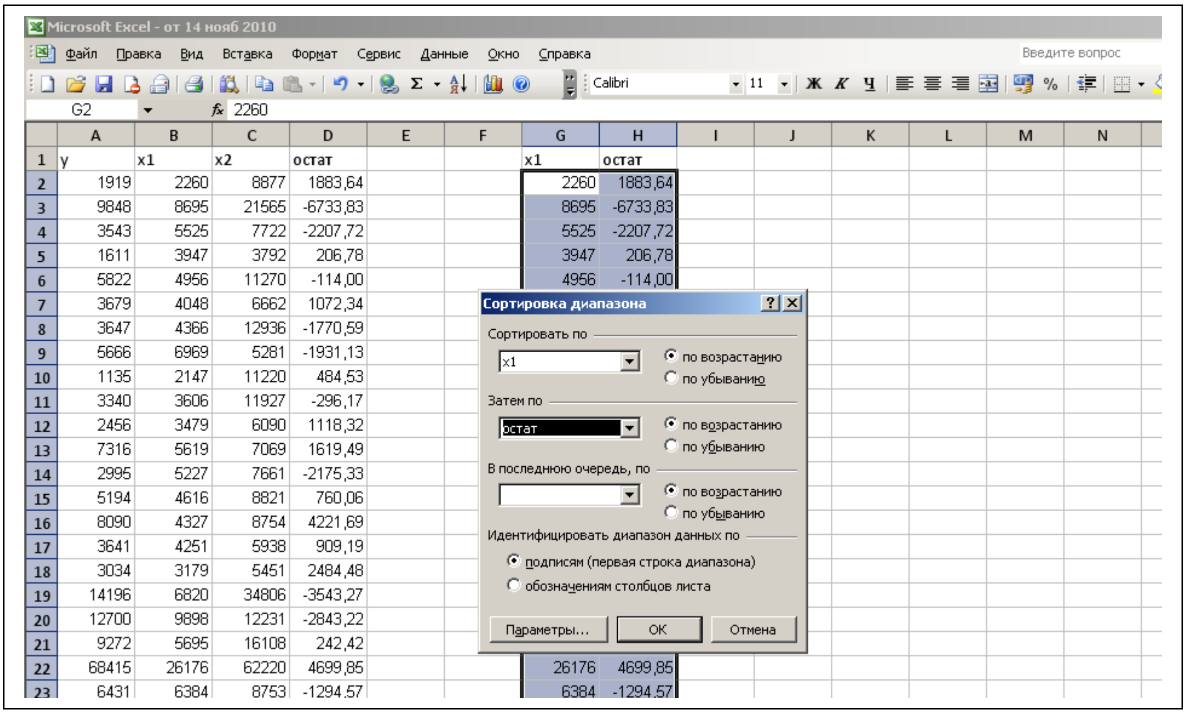


Рисунок 2.4 – Сортировка регрессионных остатков $\hat{z}_i$ по возрастанию значений объясняющей переменной x_1

Используя **Мастер функций**, категорию **Математические**, функцию **ABS** найдем модули регрессионных остатков. Построим график зависимости остатков /

e_i / от упорядоченных по возрастанию значений в пункте «Мастер диаграмм» (рисунок 2.5).

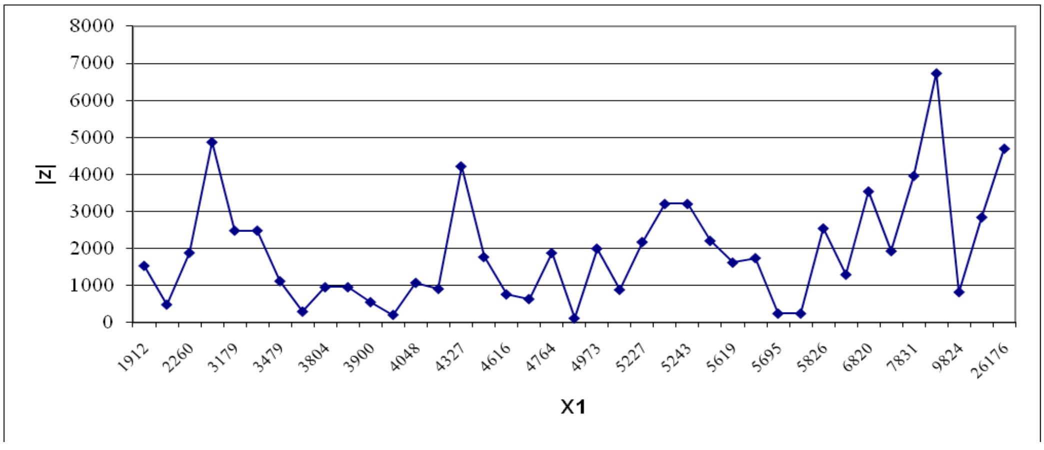


Рисунок 2.5 – График зависимости модуля значений регрессионных остатков $|\hat{z}_i|$ и значений объясняющей переменной x_1

На графике видно, что при увеличении значений объясняющей переменной, модули регрессионных остатков имеют тенденцию к росту. Следовательно, можно заподозрить гетероскедастичность по переменной x_1.

Выявим гетероскедастичность помощью различных критериев (тестов).

Тест ранговой корреляции Спирмена

Для определения коэффициента ранговой корреляции Спирмена в меню системы открыть **Statistics - Критерии** и выбрать в появившемся меню строку **Nonparametrics – Непараметрические критерии.** На экране откроется окно.

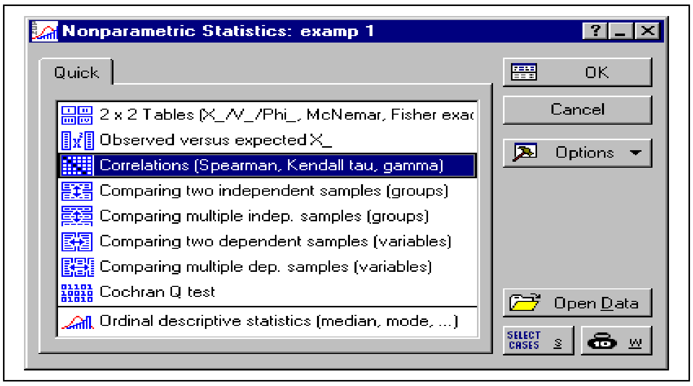


Рисунок 2.6 – Выбор пунктов меню для расчета коэффициента ранговой корреляции Спирмена

Выбираем пункт **Correlaitions (Spearmam, …) – Корреляция (Спирмен, …)**, далее в открывшемся окне выбираем переменные, между которыми необходимо рассчитать данный коэффициент (в нашем случае между регрессионными остатками и значениями объясняющей переменной).

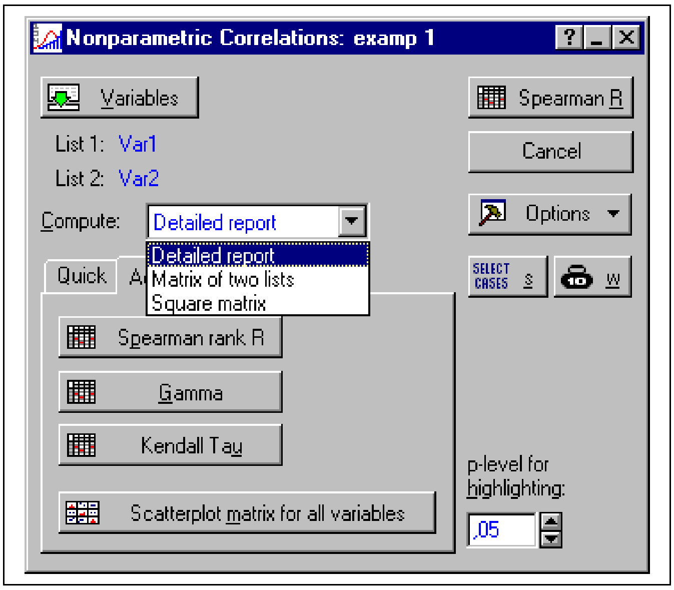


Рисунок 2.7 – Выбор пунктов для расчета коэффициента ранговой корреляции Спирмена

После нажатия на кнопку **Spearman rank R** программа произведет расчеты (рисунок 2.8).

	Spearman Rank Order Correlations (от 14 MD pairwise deleted Marked correlations are significant at p <,			
Pair of Variables	**Valid N**	Spearman R	t(N-2)	p-level
x1 & остатки	39	-0,461725	-3,16628	0,003088

Рисунок 2.8 – Результаты оценивания теста ранговой корреляции Спирмена

Во втором столбце данного окна определяется оценка коэффициента ранговой корреляции Спирмена, в третьем – значении статистики, с помощью которой проверяется нулевая гипотеза об отсутствии гетероскедастичности. В данном случае нулевая гипотеза отвергается, то есть можно сделать вывод о наличии гетероскедастичности.

Тест Голдфелда-Квандта

Для проверки на гетероскедастичность с помощью теста Голдфелда-Квандта упорядочим данные (y и x_2) по возрастанию независимой переменной x_1.

Для сортировки данных выделим переменную x_1 и используем на панели инструментов команду «**Cases**» - «**Sort Cases**» (рисунок 2.9).

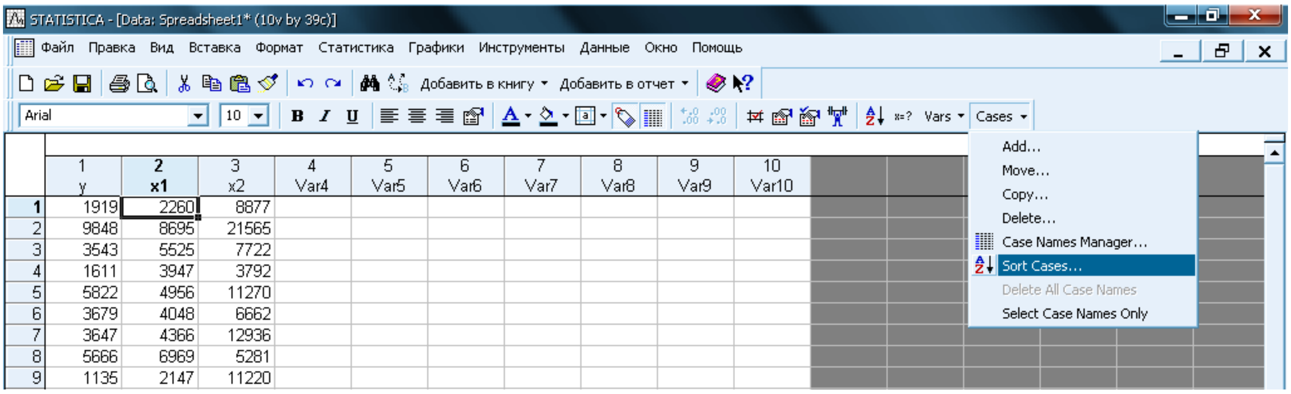


Рисунок 2.9 – Сортировка данных по значениям объясняющей переменной x_1.

В появившемся окне **Sort Option** нажимаем **ОК** и получаем упорядоченные значения *y* и x_2 по переменной x_1.

Возьмем n′ первых наблюдаемых значений результативного признака (y′) и объясняющих переменных x_1', x_2' и n″ последних наблюдаемых значений, соответственно обозначив y″ и x_1'', x_2''. $n' = n'' = \frac{39 - 0.25*39}{2} \approx 15$.

Сначала необходимо построить оценки уравнений регрессии для первых 15 объектов, а затем - для 15 последних.

Для этого в окне **Multiple Regression** укажем в качестве независимой переменной – x_1 и x_2, а в качестве зависимой – *y*. При задании параметров в данном окне следует осуществить отбор того подмножества наблюдений, которое будет участвовать в расчетах, используя для этого кнопку **Select Cases**.

После нажатия этой кнопки откроется диалоговое окно **Cases Selection Conditions,** в котором следует задавать условия отбора наблюдений.

При построении оценки уравнения регрессии по первым 15 наблюдениям в строке этого окна **include if** (включать если) укажем неравенство v0<=15, т.е. с 1 по 15 наблюдение (рисунок 2.10).

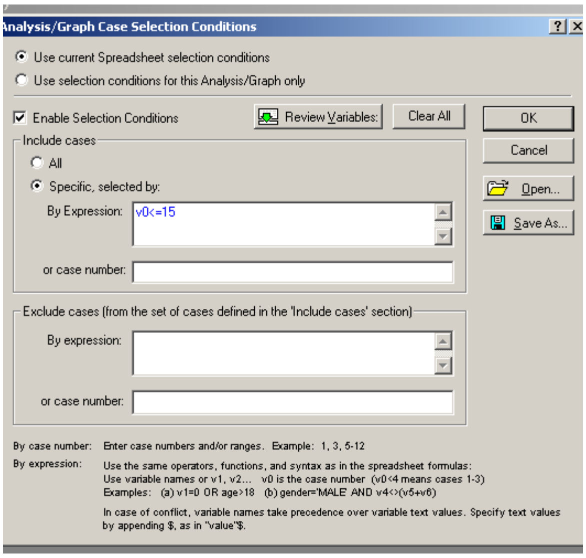


Рисунок 2.10 – Выбор подмножества наблюдений для оценки уравнения регрессии

В нижней части окна результатов регрессионного анализа нажмем на кнопку **ANOVA** (рисунок 2.11).

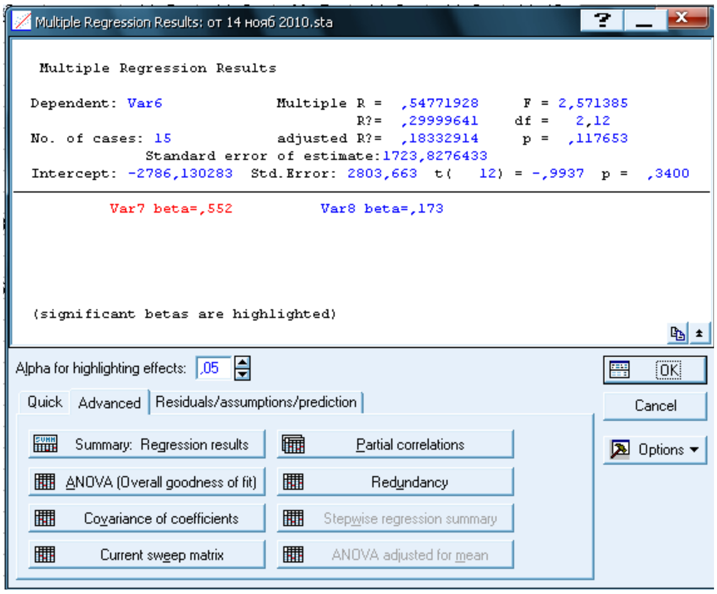


Рисунок 2.11 – Результаты оценивания для подмножества, состоящего из 15 первых наблюдений

На рисунке 2.12 представлена таблица с результатами дисперсионного анализа. Требуемое для теста Голдфелда-Квандта значение суммы квадратов остатков находится на пересечении строки **Residual** (остаточная) и столбца **Sums of Squares** (сумма квадратов).

Analysis of Variance; DV: Var6 (от 14 нояб 2С

Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	15282159	2	7641079	2,571385	0,117653
Residual	35658981	12	2971582		
Total	50941140				

Рисунок 2.12 – Результаты дисперсионного анализа для 15 первых наблюдений

При оценивании коэффициентов модели по 15 последним (с 25 по 39 наблюдение) укажем неравенство v0>=25. Символы v0 в логических операциях определяют номер строки (рисунок 2.13).

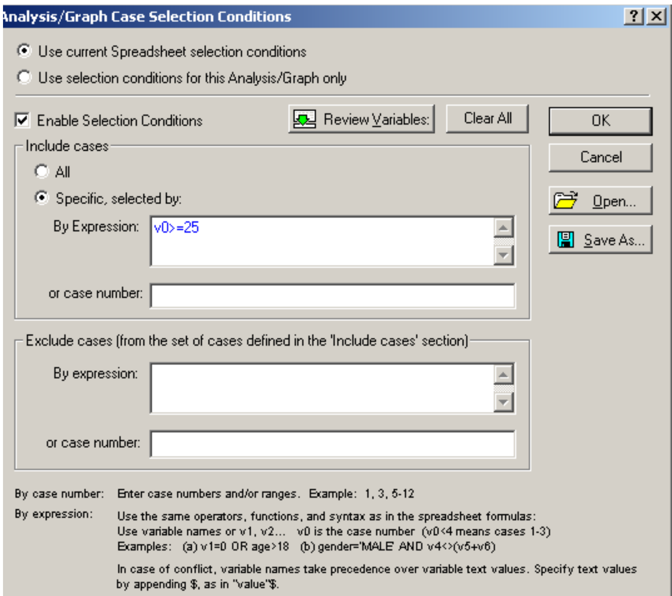


Рисунок 2.13 – Результаты оценивания для подмножества, состоящего из 15 последних наблюдений

После задания всех необходимых параметров, произведем вычисления.

Effect	Analysis of Variance; DV: Var10 (от 14 нояб 2010.sta) Sums of Squares	df	Mean Squares	F	p-level
Regress.	3,376896E+09	2	1,688448E+09	190,5045	0,000000
Residual	1,063564E+08	12	8,863034E+06		
Total	3,483253E+09				

Рисунок 2.14 – Результаты дисперсионного анализа для 15 последних наблюдений

Рассчитаем $F_{набл} = \frac{Q''}{Q'} = \frac{1{,}06 * 10^8}{35658981} = 2{,}98$. Критическое значение определим в окне Excel, используя **Мастер функций**, категорию **Статистические**, функцию **FРАСПОБР** (рисунок 2.15).

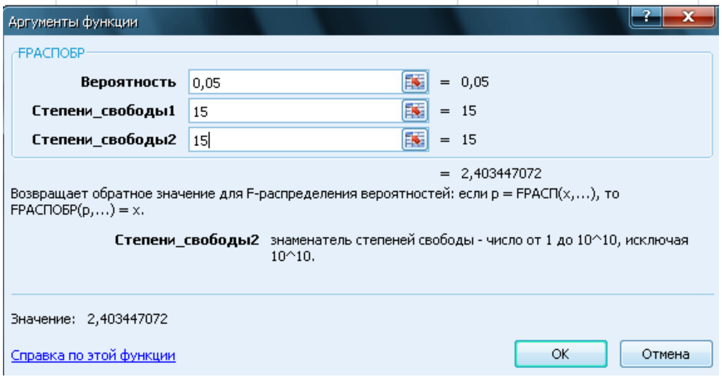


Рисунок 2.15 – Расчет критического значения для F- статистики

Так как $F_{набл.} > F_{крит}$, следовательно нулевая гипотеза об отсутствии гетероскедастичности отклоняется.

Тест Глейзера

В тех случаях, когда хотим установить более точный характер поведения σ_i, целесообразно использовать тест Глейзера. Оценив регрессионные остатки исходной модели, будем строить модель:

$$|\hat{z}_i| = \alpha + \beta |x_{il}|^{\gamma} + \delta_i, \quad (2.2)$$

Перебирая γ в промежутке от –1 до 2 оценим регрессионную модель вида (2.2). Отбираются только значимые модели, поскольку в случае отклонения нулевой гипотезы (H_0: β=0 при альтернативной H_1: β≠0), гипотеза об отсутствии гетероскедастичности не принимается.

В нашем случае, подбирая γ в промежутке от –1 до 2 были оценены уравнения с использованием модуля **Множественная регрессия**. Результаты представлены в обобщенном виде в таблице 2.1.

Таблица 2.1 - Результаты оценивания регрессионной модели вида (2.2)

γ	b_0	S_{b_0}	b_1	S_{b_1}	$\hat{R}^2$	F
-0.5	4146,59	1100,11	-151618	73415,09	0,103	4,26
0.5	-321,849	860,93	31,223	11,556	0,1647	7,3
0,7	**446,78**	**585,5**	**3,646**	**1,336**	**0,1675**	**7,44**
1	1049,398	393,56	0,1577	0,058	0,1645	7,28
1.5	1508,377	278,52	0,000885	0,0003	0,148	7,3

Статистически значимые оценки были получены для всех представленных уравнений. Максимальный $\hat{R}^2$ соответствует значению γ =0,7 (рисунок 2.16).

Regression Summary for Dependent Variable: /e/ (от 14 нояб 201
R= ,40933887 R?= ,16755831 Adjusted R?= ,14505989
F(1,37)=7,4476 p<,00966 Std.Error of estimate: 1388,4

N=39	Beta	Std.Err. of Beta	B	Std.Err. of B	t(37)	p-level
Intercept			446,7815	585,5698	0,762986	0,450309
NewVar5	0,409339	0,149995	3,6470	1,3364	2,729021	0,009664

Рисунок 2.16 – Результаты оценки регрессионной модели, соответствующей значению γ =0,7

Таким образом, наилучшая аппроксимация $|\hat{z}_i| = 446,78 + 3,647\, x_{i1}^{0.7}$. Оценка матрицы Σ_0 имеет вид:

$$\hat{\Sigma}_0 = \begin{pmatrix} \left(\hat{\alpha}+\hat{\beta}|x_{11}|^{0,7}\right)^2 & 0 & 0\ldots & 0 \\ 0 & \left(\hat{\alpha}+\hat{\beta}|x_{21}|^{0,7}\right)^2 & 0\ldots. & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \left(\hat{\alpha}+\hat{\beta}|x_{n1}|^{0.7}\right)^2 \end{pmatrix} \tag{2.3}$$

Аналогично, исследуя модель на гетероскедастичность по переменной x_2, выявили по всем критериям отсутствие гетероскедастичности.

Для определения вектора-оценок коэффициентов уравнения регрессии воспользуемся функциональными возможностями Mathcad. Для формирования матрицы $\hat{\Sigma}_0$, у которой на главной диагонали стоят значения $446,78+3,647x_{i1}^{0.7}$, скопируем в Mathcad столбец матрицы $X=\{x_1\}$. Обозначим эту матрицу-столбец «*K*» (рисунок 2.17). Присвоим матрице $\hat{\Sigma}$ функцию **identity**(39), которая создает единичную квадратную матрицу размером $n\times n$, в данном случае 39×39. Ниже введем символ *i* и присвоим ему изменение от 0 до 38 с помощью команды **Rang Variable** (рис. 2.2.8). Припишем матрице $\hat{\Sigma}$ нижний индекс i,j и зададим структуру матрицы $\hat{\Sigma}_0$, в данном случае $446,78+3,647(K_i)^{0.7}$. Далее можем получить матрицу $\hat{\Sigma}_0$ размерности 39×39 (рис. 2.17).

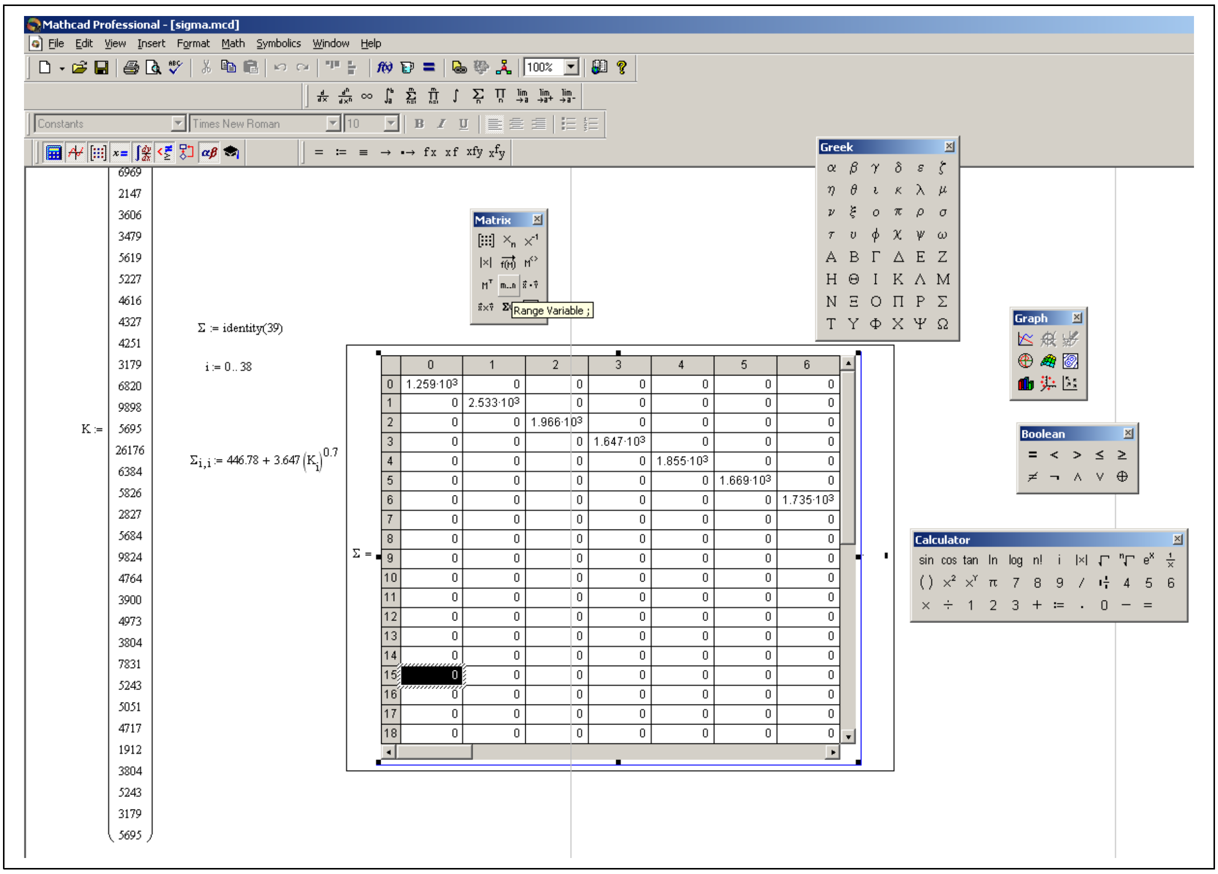


Рисунок 2.17 -Формирование матрицы $\hat{\Sigma}_0$

Для того, чтобы ввести в Mathcad матрицы X и Y, сохраним их в текстовом формате, затем в Mathcad меню **Insert** выбираем пункт **Components** (рисунок 2.18).

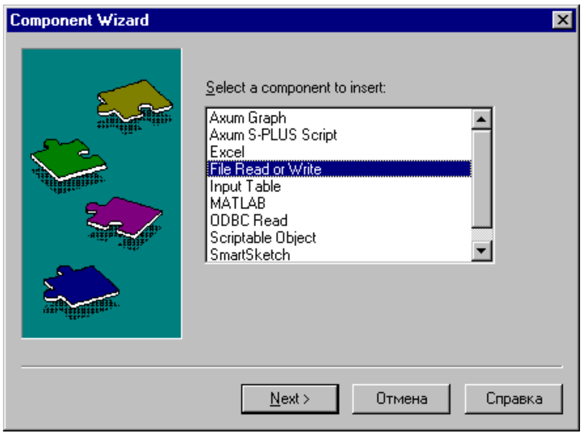


Рисунок 2.18 – Выбор пунктов меню для импортирования данных из MS Excel

В появившемся окне находим пункт **File Read or Word**. В окне **File Read or Word** нажимаем на кнопку **Browse - Обзор** и открываем текстовый файл, в котором сохранили матрицы X, Y. Выбрав нужный файл, нажимаем на кнопку Готово. В появившемся окне полученной матрице присваиваем имя, например X, и соответственно, получаем матрицы X и Y (рисунок 2.19). Используя функциональные возможности Mathcad вычисляем вектор-оценок коэффициентов уравнения регрессии, $\hat{S}_{b_j}$, критерии, с помощью которых проверяется значимость модели в целом и отдельных коэффициентов.

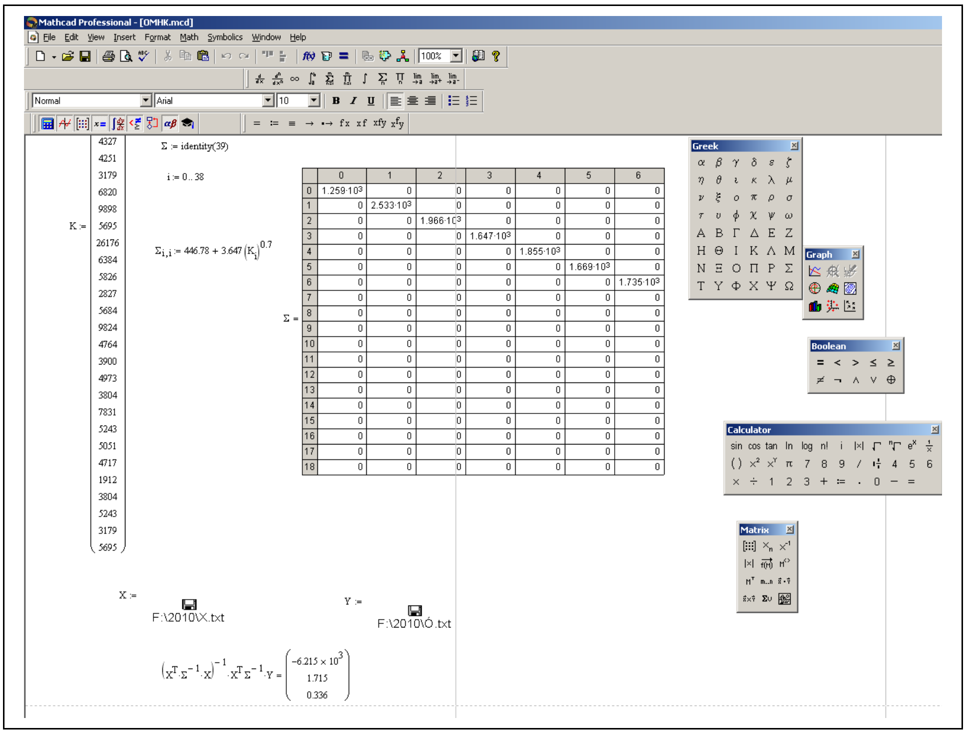


Рисунок 2.19 -Оценивание коэффициентов уравнения регрессии ОМНК

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y}_{ОМНК} = \underset{(761,77)}{-6215} + \underset{(0,197)}{1,715\,x_1} + \underset{(0,072)}{0,336\,x_2} \tag{2.4}$$

Проверим на значимость уравнение регрессии с помощью статистики Фишера – Снедекора.

$Q_{факт}=1,007*10^6$; $Q_{ост}=1,063*10^5$

$F= 170,51$; $F_{крит} = 3,259$.

Так как $F_{набл.}>F_{крит}$, следовательно регрессионная модель адекватна экспериментальным данным.

Для получения КЛММР, можно умножив правую и левую часть ЛММР слева на матрицу C^{-1} , в данном случае, имеющую вид:

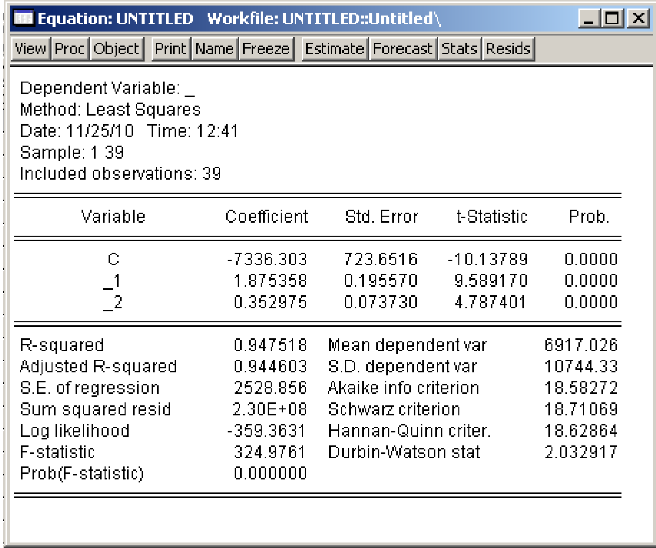
$$C^{-1}=\begin{pmatrix} \frac{1}{\hat{\alpha}+\hat{\beta}|x_{11}|^{0.7}} & 0 & 0\ldots & 0 \\ 0 & \frac{1}{\hat{\alpha}+\hat{\beta}|x_{21}|^{0.7}} & 0\ldots. & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{\hat{\alpha}+\hat{\beta}|x_{n1}|^{0.7}} \end{pmatrix},$$

получить преобразованную модель и оценить коэффициенты обычным МНК.

2.4 Уточнение стандартных ошибок в форме Уайта и Невье-Веста в пакете Eviews

Если гетероскедастичность порождается вариацией не одной, а несколькими переменными, то в этом случае используют МНК для оценки коэффициентов регрессионной модели, а стандартные ошибки коэффициентов модели уточняют с помощью стандартных ошибок в форме Уайта и Невье –Веста, расчет которых производится в пакете Eviews **.**

Для построения линейной модели множественной регрессии методом наименьших квадратов используется команда следующего формата: LS «эндогенная переменная», «константа», «список экзогенных переменных» или с пунктов меню **QUICK/ESTIMATE EQUATION.** Результаты оценивания представлены на рисунке 2.20.



Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: _
Method: Least Squares
Date: 11/25/10 Time: 12:41
Sample: 1 39
Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7336.303	723.6516	-10.13789	0.0000
_1	1.875358	0.195570	9.589170	0.0000
_2	0.352975	0.073730	4.787401	0.0000

R-squared	0.947518	Mean dependent var	6917.026
Adjusted R-squared	0.944603	S.D. dependent var	10744.33
S.E. of regression	2528.856	Akaike info criterion	18.58272
Sum squared resid	2.30E+08	Schwarz criterion	18.71069
Log likelihood	-359.3631	Hannan-Quinn criter.	18.62864
F-statistic	324.9761	Durbin-Watson stat	2.032917
Prob(F-statistic)	0.000000		

Рисунок 2.20 – Результаты оценивания параметров линейной модели множественной регрессии

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = \underset{(723,65)}{-7336{,}3} + \underset{(0,196)}{1{,}88}\, x_1 + \underset{(0,074)}{0{,}35}\, x_2 \qquad (2.5)$$

Когда гетероскедастичность присутствует, проблему можно решить, сделав коррекцию стандартных ошибок в форме Уайта. Для коррекции стандартных ошибок в форме Уайта, выберем в текущем окне пункты **Proc/Specify/Estimate…** Появится форма оценки регрессии, где необходимо нажать кнопку **Options**. В появившемся окне отметить **Heteroskedasticity** и указать вид коррекции – **White (**рисунок 2.21).

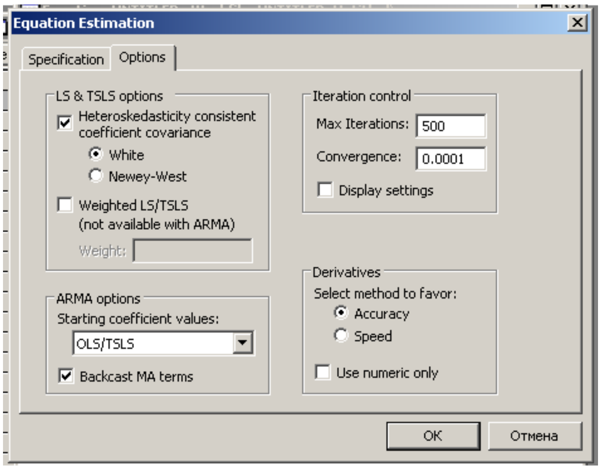


Рисунок 2.21 – Выбор вида коррекции стандартных ошибок

После нажатия на кнопку **ОК** на экране появится окно, содержащее результаты коррекции (рисунок 2.22).

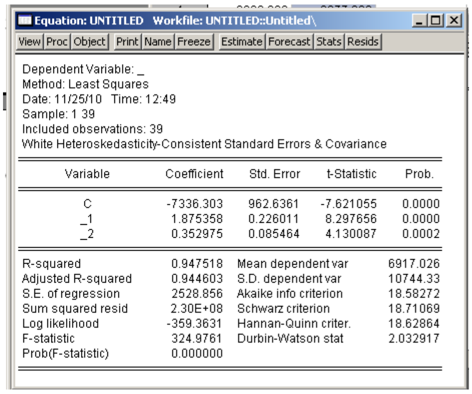


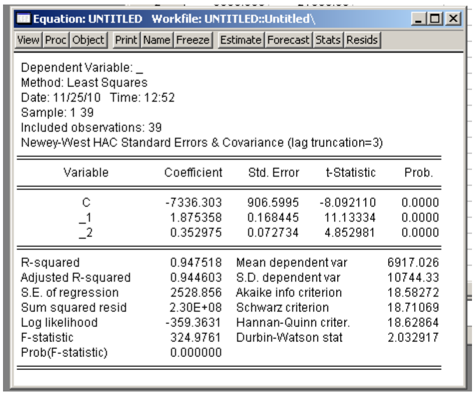
Рисунок 2.22 – Результаты коррекции стандартных ошибок коэффициентов регрессионной модели в форме Уайта

Оценка уравнения регрессии имеет вид:

$$\hat{y} = \underset{(962,64)}{-7336,3} + \underset{(0,226)}{1,88}\, x_1 + \underset{(0,854)}{0,35}\, x_2 \qquad (2.6)$$

Если сравнить результаты оценки модели (2.5) и (2.6) можно заметить, что оценки параметров модели остались прежними, а изменились лишь стандартные ошибки коэффициентов.

Пакет Eviews позволяет также производить коррекцию стандартных ошибок в форме Невье–Веста.


Equation: UNTITLED Workfile: UNTITLED::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: _
Method: Least Squares
Date: 11/25/10 Time: 12:52
Sample: 1 39
Included observations: 39
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7336.303	906.5995	-8.092110	0.0000
_1	1.875358	0.168445	11.13334	0.0000
_2	0.352975	0.072734	4.852981	0.0000

R-squared	0.947518	Mean dependent var	6917.026
Adjusted R-squared	0.944603	S.D. dependent var	10744.33
S.E. of regression	2528.856	Akaike info criterion	18.58272
Sum squared resid	2.30E+08	Schwarz criterion	18.71069
Log likelihood	-359.3631	Hannan-Quinn criter.	18.62864
F-statistic	324.9761	Durbin-Watson stat	2.032917
Prob(F-statistic)	0.000000		

Рисунок 2.23 – Результаты коррекции стандартных ошибок коэффициентов регрессионной модели в форме Невье–Веста.

2.5 Содержание письменного отчета

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

1) постановку задачи с исходными данными для анализа;
2) краткое изложение теории;
3) результаты компьютерной обработки данных;
4) анализ полученных результатов;
5) содержательная интерпретация полученных результатов.

2.6 Вопросы к защите лабораторной работы

1. Сформулируйте постановку задачи лабораторной работы.
2. Дайте определение ОЛММР с гетероскедастичными остатками.
3. Приведите аргументы в пользу применяемых Вами методов для выявления гетероскедастичности.
4. Какие тесты для выявления гетероскедастичности Вы использовали? Опишите их алгоритм.

5. Какую структуру имеет матрица Σ_0 при реализации ОМНК в условиях гетероскедастичности, выявленной с помощью различных тестов?
6. Какие необходимо сделать преобразования для получения КЛММР с помощью матрицы *C*? Какую структуру имеет матрица *C*?

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 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.

3 Доугерти, К. Введение в эконометрику: учебник для вузов/ К. Доугерти. – М.:ИНФРА-М, 1999. – 402 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 Кремер, Н. Ш. Эконометрика: учебник для вузов / под ред. проф. Н. Ш. Кремер. - М. : ЮНИТИ-ДАНА, 2003. - 311 с.

Приложение А

(обязательное)

Исходные данные для анализа

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

Номер объекта	Административно-территориальные 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

Продолжение таблицы А.1

1	2	3	4	5	6	7	8	9	10
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
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	Административно-территориальные 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
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

Продолжение таблицы А.2

1	2	3	4	5	6	7	8	9	10
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