

Министерство образования и науки Украины
Севастопольский национальный технический университет



**«ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ ВЗАИМОСВЯЗИ
СТАТИСТИЧЕСКИХ ДАННЫХ В СРЕДЕ GRETL 1.7.1.»**

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению лабораторной работы №2

по дисциплине «Эконометрия»

**для студентов специальности
6.050201 - «Менеджмент организаций»
дневной формы обучения**

Севастополь
2008



УДК 658

«Линейный регрессионный анализ взаимосвязи статистических данных в среде Gretl 1.7.1.» методические указания к выполнению лабораторной работы №2 по дисциплине «Эконометрия» для студентов специальности 8.050201 – «Менеджмент организаций» дневной формы обучения / Сост. А.В. Цуканов, Т.А. Кокодей. – Севастополь: Изд-во СевНТУ, 2008г. – 24 с.

Целью методических указаний является получение навыков применения регрессионного анализа данных с использованием средств, предоставляемых Gretl 1.7.1.

Методические указания утверждены на заседании кафедры менеджмента и экономико-математических методов, (протокол № 5 от 23.01.2008 г.).

Допущено учебно-методическим центром СевНТУ в качестве методических указаний.

Рецензент:

Фисун С.Н., канд. техн. наук, доцент кафедры «Кибернетика и вычислительная техника».

СОДЕРЖАНИЕ

1. Цель работы.....	4
2. Теоретические сведения о линейном регрессионном анализе.....	4
3. Описание средств системы Gretl для выполнения регрессионного анализа.....	6
3.1. Оценка параметров линейной регрессионной модели методом 1МНК (OLS) и проверка адекватности модели.....	6
3.2. Анализ выполнения предпосылок 1МНК.....	13
4. Порядок выполнения лабораторной работы	18
5. Содержание отчёта о выполнении лабораторной работы.....	18
Библиографический список.....	19
Приложение А. (справочное) Основные описательные статистики.....	20
Приложение Б. (справочное) Статистические таблицы в GRETL.....	21
Приложение В. (справочное) Построение графиков.....	23

1. ЦЕЛЬ РАБОТЫ

Целью данной работы является получение практических навыков регрессионного анализа в системе Gretl для автоматизированного поиска ранее неизвестных закономерностей в имеющихся в распоряжении менеджера данных с последующим использованием полученной информации для подготовки управленческих решений.

2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ О ЛИНЕЙНОМ РЕГРЕССИОННОМ АНАЛИЗЕ

Целью регрессионного анализа является оценка функциональной зависимости $y = f(x_1, x_2, \dots, x_n) + u$ результативного признака (y) от факторных ($x_1, x_2, \dots, x_n$). Формулы (1) и (2) представляют собой линейные модели парной и множественной регрессии соответственно.

$$y = f(x) + u = \alpha_0 + \alpha_1 \cdot x + u, \quad (1)$$

$$y = \alpha_0 + \alpha_1 \cdot x_1 + \dots + \alpha_n x_n + u, \quad (2)$$

где y — фактическое значение результативного признака;

x_i — признак-фактор;

α_i — параметр регрессионной модели;

u — случайная ошибка (остаток), характеризующая отклонения реального значения результативного признака от теоретического. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения.

Оценивание параметров линейной модели основан на обычном или одношаговом методе наименьших квадратов (1МНК или OLS – Ordinary Least Squares).

Этот метод позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака (y) от расчетных (теоретических) $\tilde{y}_x$ минимальна, формула (3).

$$\sum_i (y_i - \tilde{y}_{x_i})^2 \rightarrow \min, \quad (3)$$

Статистическое моделирование связи методом линейного регрессионного анализа осуществляется в 3 этапа:

а) Оценка параметров линейной регрессионной модели методом 1МНК

Вектор оценок параметров модели (2) определяется выражением (4).

$$\alpha = [X^T \cdot X]^{-1} \cdot X^T \cdot Y \quad (4)$$

б) Проверка адекватности регрессионной модели (проверки значимости индивидуальных оценок коэффициентов модели с помощью t-критерия Стьюдента и оценка значимости уравнения регрессии в целом с помощью F-критерия Фишера)

На первом шаге проверки адекватности (качества) модели оценивается существенность влияния каждой объясняющей переменной x_i , на зависимую переменную y , для этого необходимо оценить значимость полученных параметров α_i , используя t- критерий Стьюдента, формула (5). Значимость параметра определяется путём проверки нулевой гипотезы о равенстве его нулю (для выбранного уровня значимости).

$$t_{p_{\alpha_i}} = \frac{|\alpha_i|}{\sqrt{\sigma_{\alpha_i}^2}}, \quad (5)$$

где α_i - оценка i -го коэффициента модели, COEFFICIENT;

$\sigma_{\alpha_i}^2$ - оценка дисперсии параметра α_i , $\sqrt{\sigma_{\alpha_i}^2} = \text{STDERROR}$.

На втором шаге проверки адекватности модели оценивается её значимость (пригодность) в целом, используя показатели: F-критерий Фишера, формула (6), коэффициент детерминации R^2 , формула (7), (Unadjusted R^2 и Adjusted R^2), сумма квадратов остатков RSS Sum of squared residuals), стандартная ошибка регрессии (Standard error of residuals), информационные критерии (Akaike information criterion, Schwarz Bayesian criterion, Hannan-Quinn criterion).

Значимость регрессии проверяется путём проверки нулевой гипотезы о равенстве нулю всех параметров модели (для выбранного уровня значимости).

$$F_p = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1}, \quad (6)$$

где R^2 - коэффициент детерминации - часть вариации (дисперсии) зависимой переменной y , которая объясняется уравнением регрессии, UNADJUSTED R^2 .

$$R^2 = \frac{\mathbf{X}^T \cdot \mathbf{X}^T \cdot \mathbf{Y} - n \cdot \bar{Y}^2}{\mathbf{Y}^T \cdot \mathbf{Y} - n \cdot \bar{Y}^2}, \quad (7)$$

n - число наблюдений;

k – число коэффициентов факторов.

При анализе адекватности уравнения регрессии исследуемому процессу возможны следующие варианты:

- Построенная модель на основе ее проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений к осуществлению прогнозов.
- Модель по F-критерию Фишера адекватна, но часть коэффициентов регрессии незначима. В этом случае модель пригодна для принятия некоторых решений, но не для производства прогнозов.
- Модель по F-критерию Фишера адекватна, но все коэффициенты регрессии незначимы. Поэтому модель полностью считается неадекватной. На ее основе не

принимаются решения и не осуществляются прогнозы.

с) Анализ выполнения предпосылок 1МНК (условий Гаусса—Маркова).

Регрессионный анализ линейных функций, основанный на обычном или одношаговом методе наименьших квадратов (1МНК) должен удовлетворять четырем условиям Гаусса—Маркова:

1. Математическое ожидание случайной составляющей, $M(u_i)$ в любом наблюдении должно быть равно нулю.

2. Дисперсия случайной составляющей σ_u^2 должна быть постоянна для всех наблюдений. Дисперсия случайной составляющей σ_u^2 должна быть постоянна для всех наблюдений. Если это условие не соблюдается, то имеет место гетероскедастичность. Наличие гетероскедастичности можно определить при помощи теста Уайта, который позволяет проверить значимость регрессии квадратов остатков относительно комплекса переменных модели и их квадратов. При этом формулируется нулевая гипотеза о гомоскедастичности остатков (равенстве нулю всех коэффициентов модели).

3. Отсутствие систематической связи между значениями случайной составляющей u_i в любых двух наблюдениях. Отсутствие автокорреляции остатков.

Автокорреляция остатков означает наличие корреляции между остатками текущих и предыдущих (последующих) наблюдений. Оценить эту зависимость можно вычислив коэффициент корреляции между этими остатками по формуле (8).

$$r_{u_i u_j} = \frac{\overline{u_i u_j} - \bar{u}_i \cdot \bar{u}_j}{\sigma_{u_i} \sigma_{u_j}}. \quad (8)$$

4. Случайный характер остатков. Случайная составляющая должна быть распределена независимо от переменных u_i и x_i .

3. ОПИСАНИЕ СРЕДСТВ СИСТЕМЫ GRETL ДЛЯ ВЫПОЛНЕНИЯ РЕГРЕССИОННОГО АНАЛИЗА

3.1. Оценка параметров линейной регрессионной модели методом 1МНК (OLS) и проверка адекватности модели

В пакете программ GRETL параметры эконометрической модели можно оценить с применением метода наименьших квадратов, в частности, одношагового 1МНК (Ordinary Least Squares - OLS) для оценки линейных регрессионных моделей для срезов данных (cross-sectional data type).

Окно спецификации эконометрической модели, оцениваемой с применением 1МНК, вызывается функцией меню **Model\Ordinary Least Squares...** В этом окне Y (Dependent variable) выбирается при помощи кнопки

«Choose», а объясняющие переменные (Independent variable) – при помощи кнопки «Add».

Пример 1.

Откроем встроенный набор данных attend.gdt на закладке Wooldridge (**File\Open Data\Sample File\attend.gdt**) и обратимся к функции **Model\Ordinary Least Squares**, чтобы построить линейную регрессионную модель, отражающую зависимость переменной Final (оценка за итоговый экзамен) от переменных attend (число посещённых занятий), termGPA (средний балл за семестр), priGPA (средний балл на начало семестра), ACT (оценка по вступительному тесту ACT), hwrtе (процент сданных домашних работ). В открывшемся окне спецификации модели выберем соответствующие зависимую и независимые переменные при помощи кнопок «Choose» и «Add», затем нажмём кнопку ОК (рисунок 1).

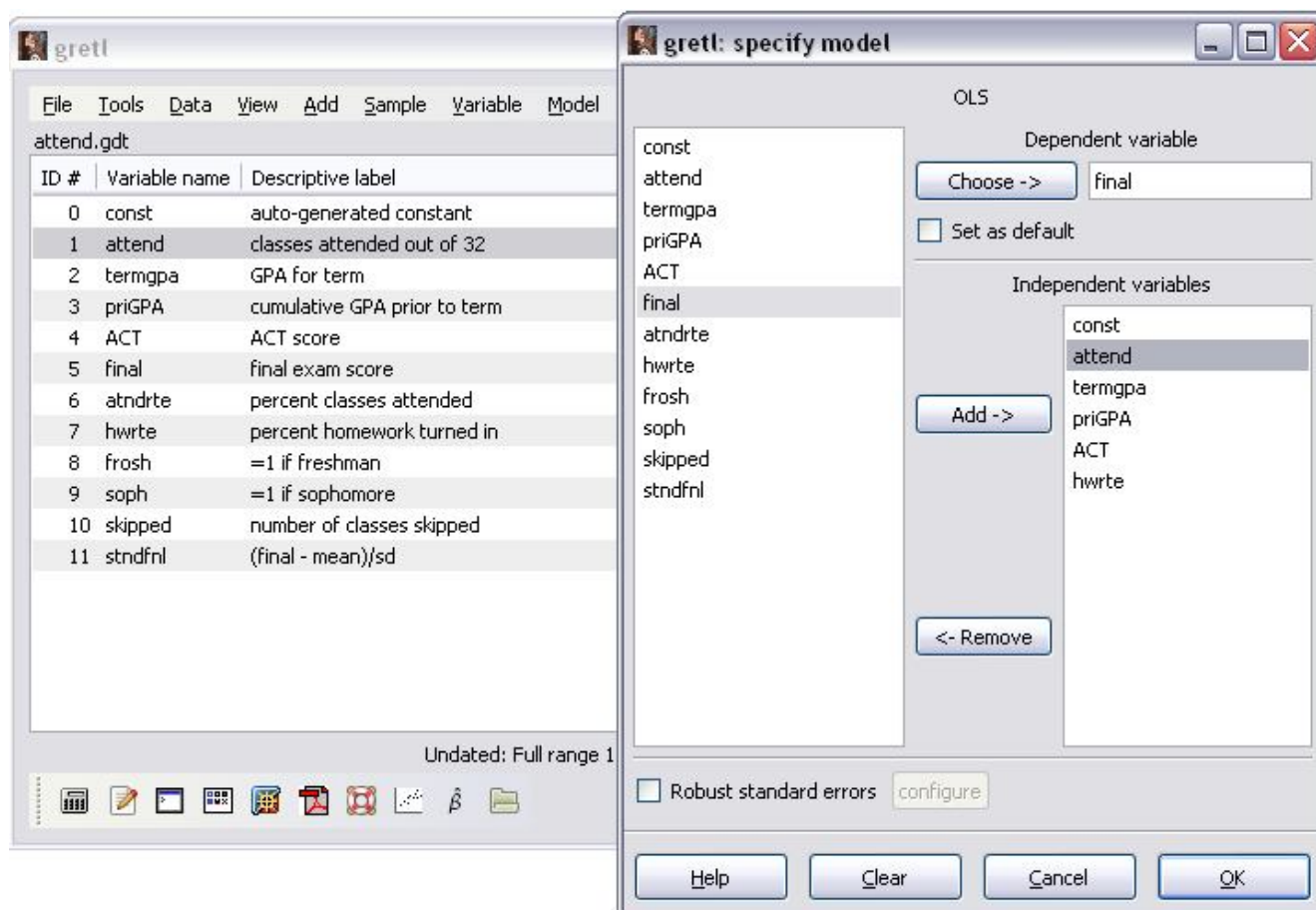
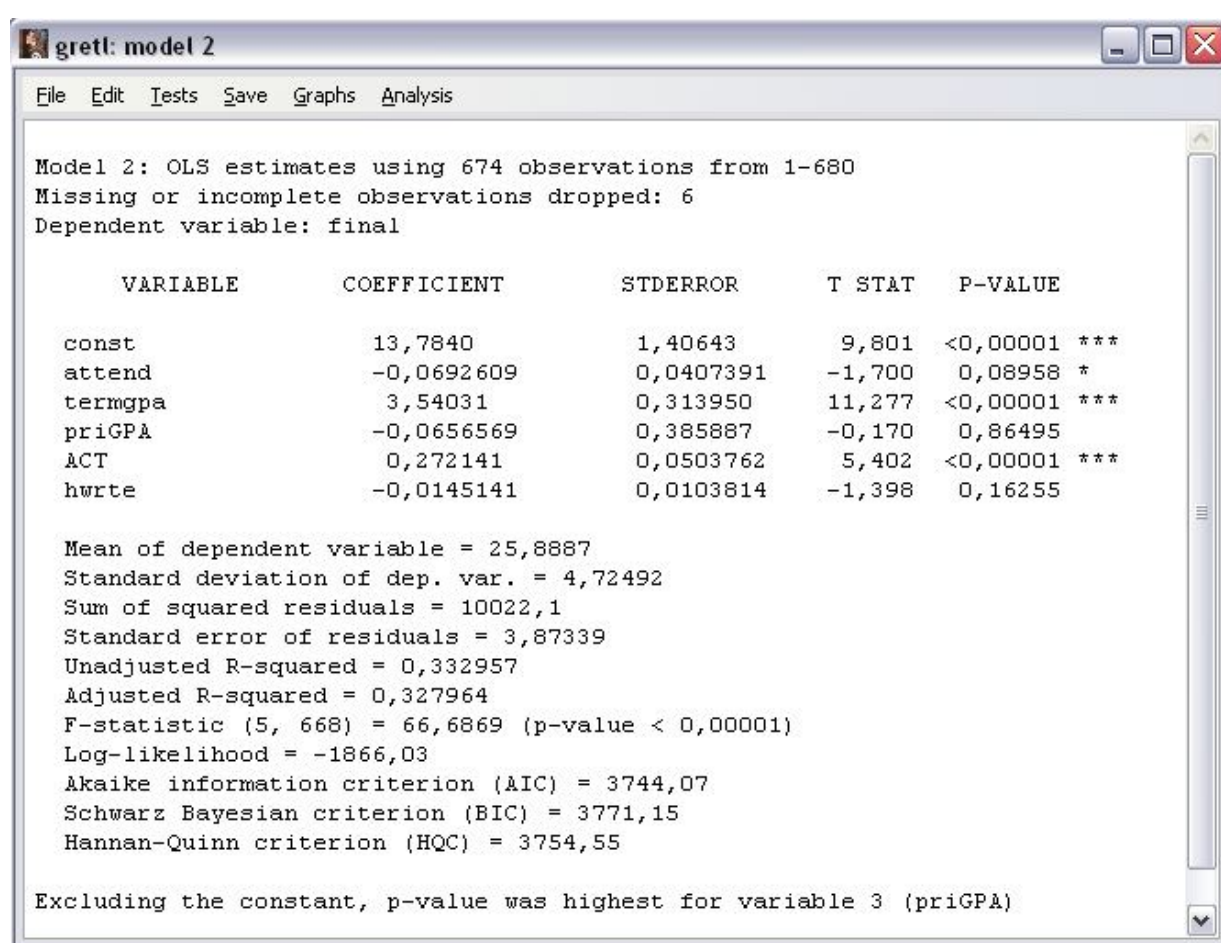


Рисунок 1 - Окно спецификации эконометрической модели, оцениваемой с применением 1МНК

Полученные результаты представлены на рисунке 2. По данным наблюдений attend.gdt была составлена модель (9).

$$\text{final} = 13,784 - 0,069 \cdot \text{attend} + 3,54 \cdot \text{termgpa} - 0,066 \cdot \text{priGPA} + 0,272 \cdot \text{Act} - 0,015 \cdot \text{hwrtе} + u. \quad (9)$$

Рассматривая значения параметров модели α_i для данной отдельной выборки, можно отметить существенность зависимости переменной final от переменных termgpa (сильная положительная связь, 3,54) и ACT (положительная связь 0,272), остальные коэффициенты имеют значения близкие к 0 и не оказывают существенного влияния на результирующий признак. Необходимо установить насколько вероятно, что зависимость, подобная найденной, подтвердится на данных другой выборки, извлеченной из той же самой генеральной совокупности, т.е. можно ли свойства данной выборки перенести на всю генеральную совокупность.



gretl: model 2

File Edit Tests Save Graphs Analysis

Model 2: OLS estimates using 674 observations from 1-680
Missing or incomplete observations dropped: 6
Dependent variable: final

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	13,7840	1,40643	9,801	<0,00001 ***
attend	-0,0692609	0,0407391	-1,700	0,08958 *
termgpa	3,54031	0,313950	11,277	<0,00001 ***
priGPA	-0,0656569	0,385887	-0,170	0,86495
ACT	0,272141	0,0503762	5,402	<0,00001 ***
hwrt	-0,0145141	0,0103814	-1,398	0,16255

Mean of dependent variable = 25,8887
Standard deviation of dep. var. = 4,72492
Sum of squared residuals = 10022,1
Standard error of residuals = 3,87339
Unadjusted R-squared = 0,332957
Adjusted R-squared = 0,327964
F-statistic (5, 668) = 66,6869 (p-value < 0,00001)
Log-likelihood = -1866,03
Akaike information criterion (AIC) = 3744,07
Schwarz Bayesian criterion (BIC) = 3771,15
Hannan-Quinn criterion (HQC) = 3754,55

Excluding the constant, p-value was highest for variable 3 (priGPA)

Рисунок 2 - Окно результатов моделирования с применением 1МНК

Рассмотрим сущность показателей (таблица 1), используемых в таблице регрессии окна результатов моделирования (рисунок 2):

Таблица 1- Показатели таблицы регрессии

Variable -	независимая переменная, существенность влияния которой необходимо оценить.
Coefficient-	коэффициент (параметр) модели, α_i , значимость которого необходимо оценить.
Std. Error-	стандартная ошибка параметра модели, является оценкой среднеквадратичного отклонения параметра регрессии от его истинного значения, даёт общую оценку степени точности параметра.
T-STATISTIC-	расчётные значения t- критерия Стьюдента, отношение значения параметра α_i к STD.ERROR, формула (5).
P-value-	показывает вероятность того, что соответствующее значение критерия для генеральной совокупности может оказаться больше, чем расчётное значение по рассматриваемой выборке. Если p-value не превышает уровень значимости, то коэффициент является значимым и принимается альтернативная гипотеза.
Mean of dependent variable-	среднее значение зависимой переменной (y).
Standard deviation of dep. var.-	стандартное (среднеквадратическое) отклонение зависимой переменной (y) – корень квадратный из дисперсии, мера разброса данных.
Sum of squared residuals -	сумма квадратов остатков ($RSS = Y^T \cdot Y - \alpha^T \cdot X^T \cdot Y$), измеряет необъяснённую часть вариации зависимой переменной, используется как основная минимизируемая величина в 1МНК.
Standard error of residuals-	стандартная ошибка регрессии (среднеквадратическое отклонение ошибки), оценивает степень соответствия модели эмпирическим данным и качество оценивания; измеряет величину квадрата ошибки, приходящейся на одну степень свободы модели $(RSS^2/(n-k))^{1/2}$.
Unadjusted R ² -	нескорректированный коэффициент детерминации R^2 - показывает долю объяснённой (уравнением регрессии) дисперсии зависимой переменной y, формула (7).
Adjusted R ² -	скорректированный коэффициент детерминации, используемый при необходимости учёта количества наблюдений и оцениваемых параметров, чтобы обеспечить сопоставимость различных моделей.
F-statistic-	расчетное значение F-критерия Фишера, формула (6), отношение объяснённой суммы квадратов (в расчёте на одну независимую переменную) к остаточной сумме квадратов (в расчёте на одну степень свободы)
Log-likelihood -	логарифмическая функция правдоподобия. Функция правдоподобия – это плотность распределения y.
Akaike information criterion-	информационный критерий Акайке, анализирует правильность спецификации модели. Позволяет выбирать наилучшую модель из множества различных спецификаций.

Продолжение таблицы 1

Schwarz Bayesian criterion-	информационный Байесовский критерий Шварца, анализирует правильность спецификации модели, позволяет выбирать наилучшую модель из множества различных спецификаций.
Hannan-Quinn criterion-	информационный критерий Хеннана – Куинна, анализирует правильность спецификации модели
OLS	1МНК
Model 2: estimates using 674 observations from 1-680	модель2: использует для оценки 674 наблюдения из 680
Missing or incomplete observations dropped	число пропущенных наблюдений
Dependent variable	зависимая переменная (y)

Пример проверки адекватности регрессионной модели

1. Шаг Оценим существенность влияния каждой объясняющей переменной (attend, termGPA, priGPA, ACT, hwrt) согласно приведённому выше примеру, формула (9), на зависимую переменную final, для этого необходимо оценить значимость полученных параметров α_i (рисунок 2), используя t-критерий Стьюдента.

Сформулируем нулевую гипотезу о не значимости коэффициента ($\alpha_i = 0$, и лишь в силу случайных обстоятельств оказался равным проверяемой величине) и альтернативную – о значимости ($\alpha_i \neq 0$), а также выберем уровень значимости (1%, 5%, или 10% - максимально допустимая вероятность ошибочного принятия альтернативной гипотезы).

В оцениваемой модели (формула 9) (рисунок 2) существенные параметры при уровне значимости 1% обозначены ***, 5% - **, 10% - *. Обозначение звёздочками облегчает быстрое оценивание значимости параметров, в рассматриваемом примере существенными являются только константа const и коэффициенты при переменных termGPA и ACT (во всех трёх случаях вероятность ошибки при принятии гипотезы об их значимости P-VALUE=0,001%).

В последнем столбце представляется эмпирический уровень значимости P-VALUE (вероятность допустить ошибку при принятии альтернативной гипотезы, т.е. вероятность того, что значение t-критерия для генеральной совокупности превысит его расчётное значение по данной выборке), который позволяет проверить гипотезы о значимости каждого коэффициента и осуществить отбор существенных (P-value меньше выбранного уровня значимости) и наиболее слабых переменных модели (P-value больше выбранного уровня значимости). В рассмотренном примере самой слабой является переменная PriGPA - вероятность ошибки при принятии гипотезы о её значимости 86,5%, на которую также указывает сообщение в последней строке окна.

Значения столбца T- Stat (рисунок 2), представляющие собой отношение соответствующих величин в столбцах COEFFICIENT и STDERROR, показывают расчётные значения t- критерий Стьюдента. Согласно методу отбора объясняющих переменных a posteriori предполагается исключение переменных с минимальными (по модулю) значением t- критерия, в рассматриваемом случае - переменных attend, priGPA, и hwrtе.

2 Шаг. Оценим значимость (пригодность) модели (формула 9) в целом, используя показатели: F-критерий Фишера, коэффициент детерминации R^2 (Unadjusted R^2 и Adjusted R^2), сумма квадратов остатков (RSS, Sum of squared residuals), стандартная ошибка регрессии (Standard error of residuals), информационные критерии (Akaike information criterion, Schwarz Bayesian criterion, Hannan-Quinn criterion).

В рассматриваемом примере F-критерий Фишера F-statistic (5, 668) = 66,6869 для p-value < 0,00001. Поскольку p-value меньше выбранного уровня значимости (p=1%) принимается решение о принятии альтернативной гипотезы, т.е. об адекватности модели в целом. Однако $R^2 = 33,3\%$, что свидетельствует о невысоком уровне объяснения моделью фактических данных, однако согласно F-тесту, он может быть признан достаточно существенным.

Т.о. в результате анализа рассматриваемой модели на адекватность можно сделать вывод: модель по F-критерию Фишера адекватна, но три коэффициента регрессии (при переменных attend, priGPA, и hwrtе) незначимы. В этом случае модель пригодна для принятия некоторых решений относительно зависимости переменной final от переменных termgpa и АСТ, но не для производства прогнозов.

Пример 2. Согласно вышеизложенным рекомендациям исключим из полученной в Примере 1. модели, формула (9), переменные attend, priGPA, и hwrtе и повторим рассмотренную последовательность действий для получения линейной регрессионной модели, устанавливающей зависимость переменной FINAL от АСТ и termgpa. Получим скорректированную модель $final = 10,8 + 0,339 \text{ АСТ} + 2,87 \text{ termgpa} + u$ (рисунок 3), в которой все переменные существенны и модель в целом пригодна для практического использования (согласно рассмотренным выше критериям) для принятия решений и составления прогнозов.

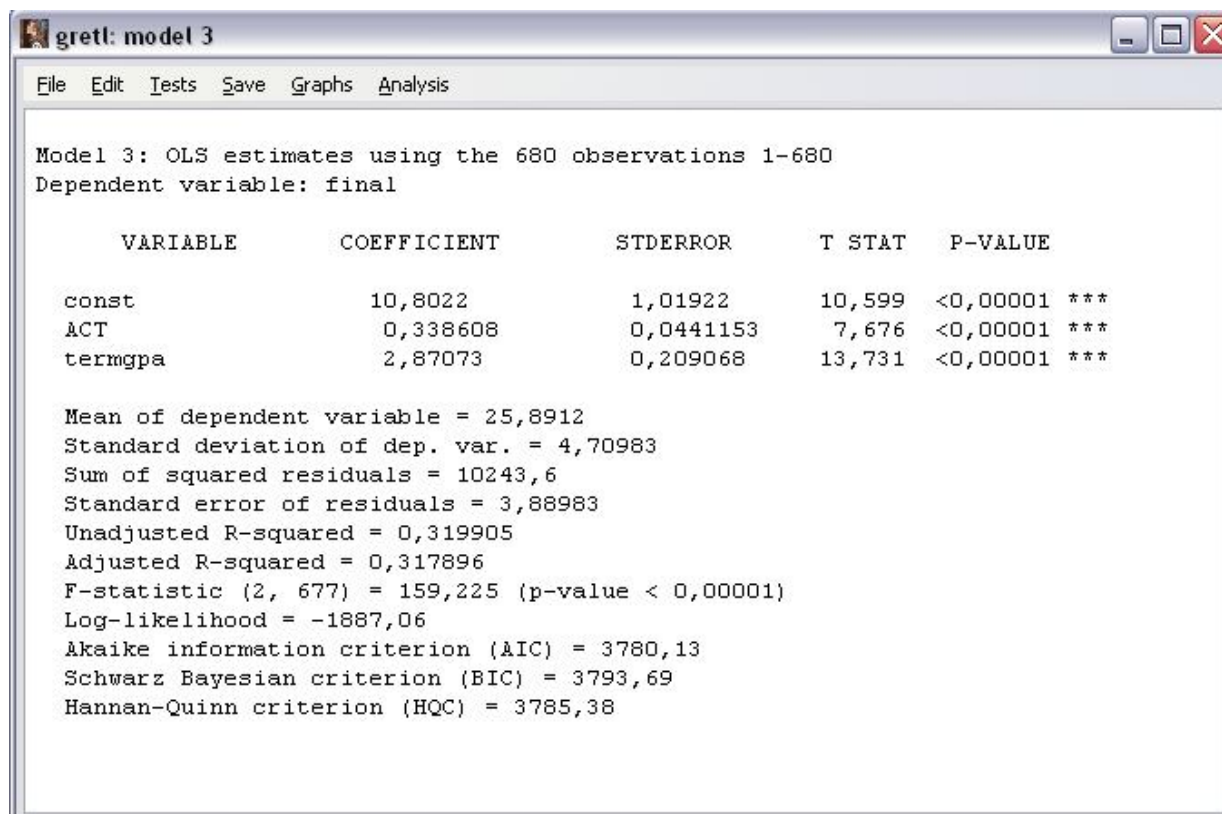


Рисунок 3 - Окно результатов моделирования с применением 1МНК, скорректированная модель

Сохраним значения остатков данной модели как отдельную переменную RESIDUALS набора attend.gdt при помощи функции **Save\Residuals** окна результатов моделирования (рисунок 3). После нажатия кнопки ОК диалогового окна, данная переменная добавится в список переменных рассматриваемого набора данных attend.gdt. Аналогичным образом сохраним модельные значения результативного признака (final) как FITSfinal при помощи функции **Save\Fitted Values**.

Пример построения графика регрессионной модели

Для графического отражения фактических и модельных данных рассмотренного Примера 2. необходимо обратиться к команде **Graphs\ Fitted, Actual Plot\ Against ACT and Termgpa** окна результатов моделирования (рисунок 3).

Получим графическое изображение фактических и модельных данных (рисунок 4). Левой кнопкой мыши возможно вращать данное изображение для удобства его просмотра.

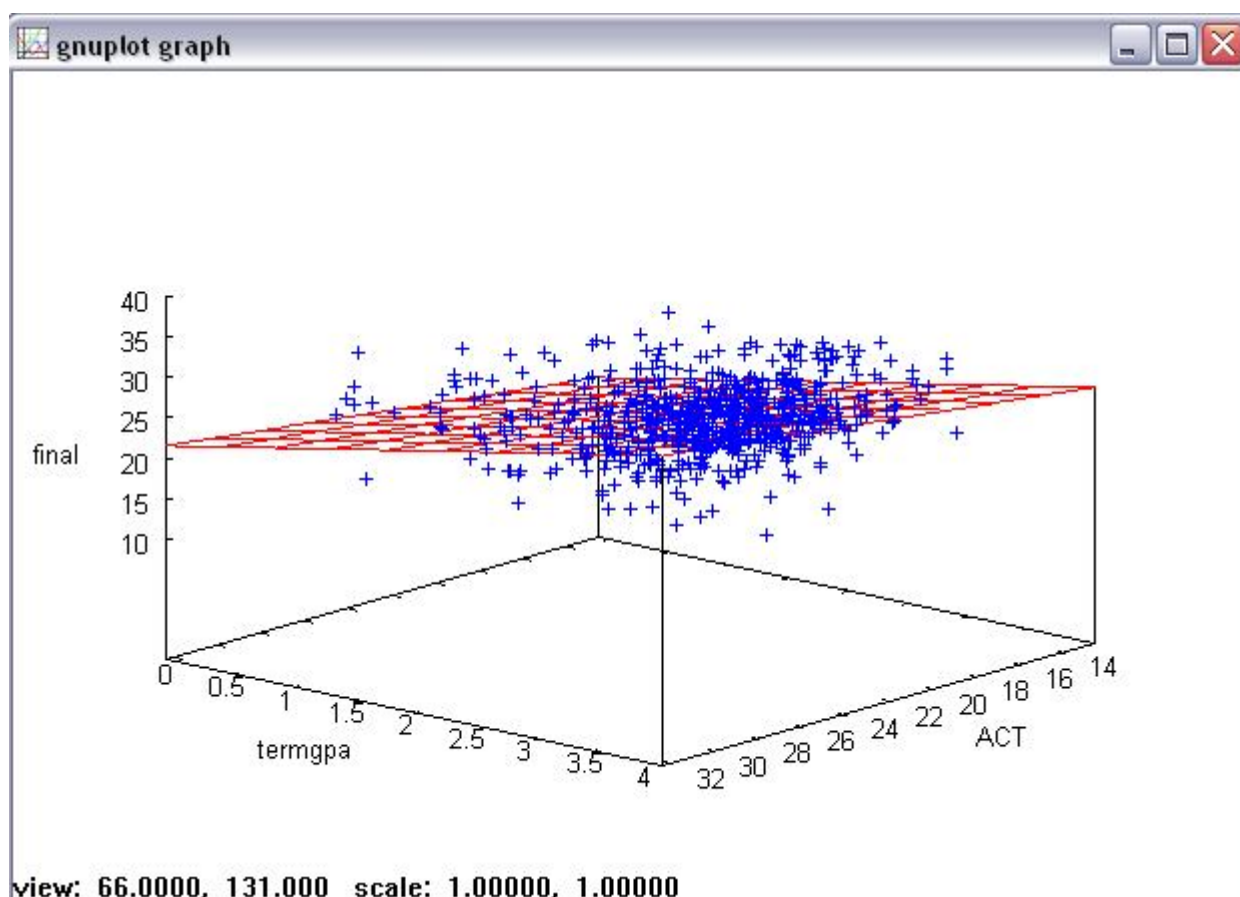


Рисунок 4 - Исходные данные и график функции
 $final = 10,8 + 0,339ACT + 2,87termgpa + u$

3.2. Анализ выполнения предпосылок 1МНК

Проверим условия Гаусса—Маркова при помощи инструментария GRETL для данных примера 2.:

1. Нулевая средняя величина (математическое ожидание) остатков, $M(u_i) = 0$.

Для проверки данного утверждения выберем щелчком мыши ранее созданную переменную RESIDUALS в списке переменных стартового экрана и обратимся к функции **View\ Summary Statistics** (рисунок 5), в открывшемся окне среднее значение остатков (mean) равна 0.

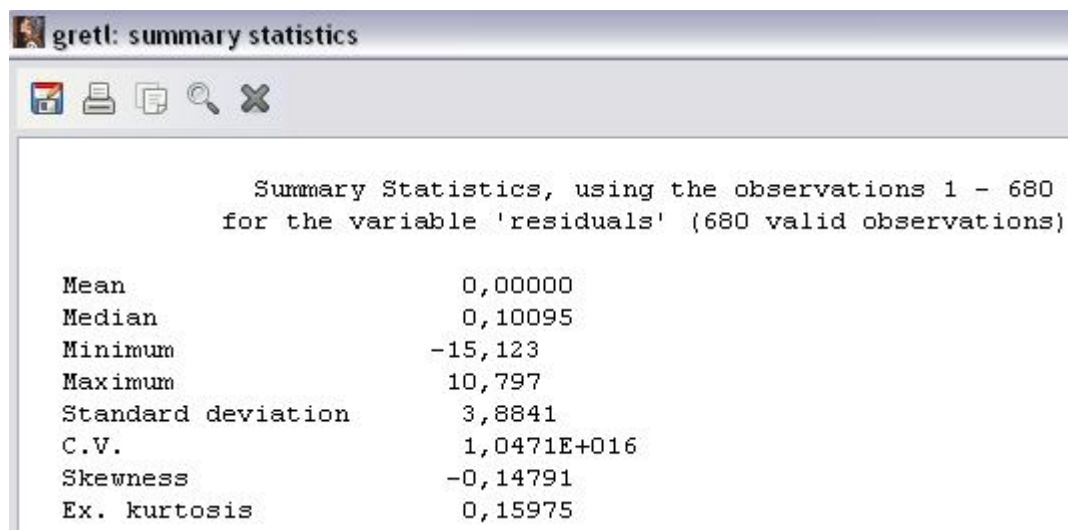


Рисунок 5 - Общая статистика для переменной RESIDUALS

2. Проверка условия гомоскедастичности остатков:

Проверку можно выполнить в окне текущей модели (рисунок 3), для чего в меню следует выбрать Tests\ heteroskedasticity. Окно результатов в этом случае имеет вид, представленный на рисунке 6. Значение P-value = 0,734603 больше уровня значимости 0,01 свидетельствует о том, что нулевую гипотезу следует принять и условие гомоскедастичности остатков выполняется.

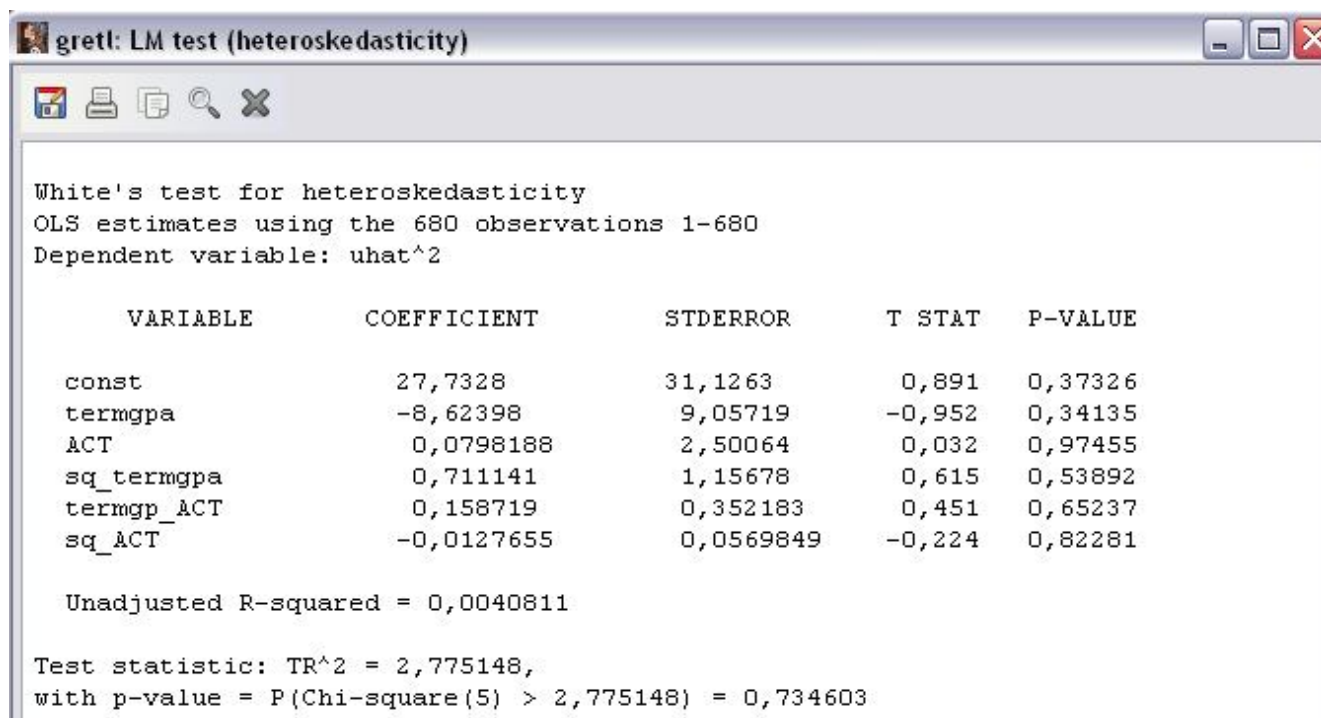


Рисунок 6 - Тест Уайта на гетероскедастичность остатков

3. Отсутствие систематической связи между значениями случайной составляющей u_i в любых двух наблюдениях (отсутствие автокорреляции остатков).

Определим наличие автокорреляции статков рассматриваемой модели.

Экспортируем ряд значений созданной в Примере 2. переменной Residuals (остатки модели) в файл Residuals.csv (File\Export data\ CSV..., поставив флажок comma (,) в разделе decimal point character). Создадим в файле Residuals.csv новую переменную Residuals1, которая отличается на один лаг от переменной Residuals (длина рядов сокращается на одно наблюдение), затем сохраним файл в формате Residuals.xls. Создадим новый набор данных в Gretl (File\New dataset) и импортируем в него данные из файла Residuals.xls (File\Open Data\Import\Excel), ответив «no» на вопрос о смене типа данных.

Рассчитаем коэффициент корреляции между данными переменными, обратившись к функции **View\Correlation matrix**, выбрав переменные Residuals и Residuals1. Получим коэффициент -0,1488, свидетельствующий о незначительной корреляции (корреляция считается сильной, если ее коэффициент выше |0,6|).

4. Случайная составляющая должна быть распределена независимо от переменных x и y (случайный характер остатков).

Для проверки строится график зависимости остатков u_i от теоретических значений результативного признака y и x .

Способом, аналогичным описанному выше, построим парную регрессию ошибки RESIDUALS от модельных значений результативного признака FitsFINAL (рисунок 7). В результате получим нулевое значение коэффициента и единичное значение p -value, а также расположение остатков на графике в виде горизонтальной полосы, что свидетельствует об отсутствии данной зависимости и о случайном характере остатков.

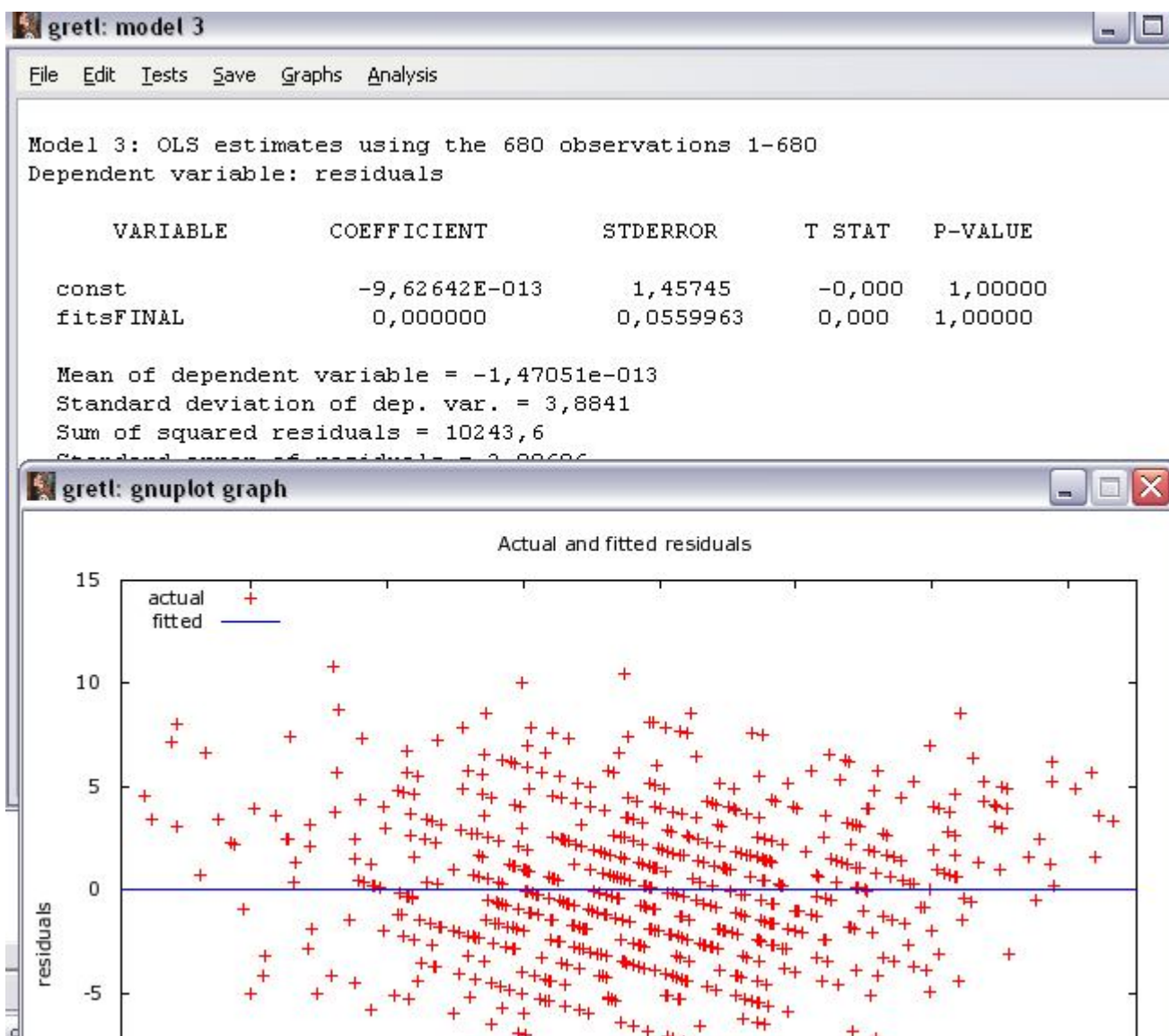


Рисунок 7 - Проверка случайного характера остатков

Проверку зависимости остатков от переменных termgpa и ACT можно осуществить из окна модели $\text{final} = 10,8 + 0,339\text{ACT} + 2,87\text{termgpa} + u$ (рисунок 3), построив соответствующие графики Graphs\Residual Plot\Against termgpa (Against ACT)(рисунок 8, 9). На полученных графиках остатки также расположены в виде горизонтальных полос, что свидетельствует об отсутствии соответствующих зависимостей.

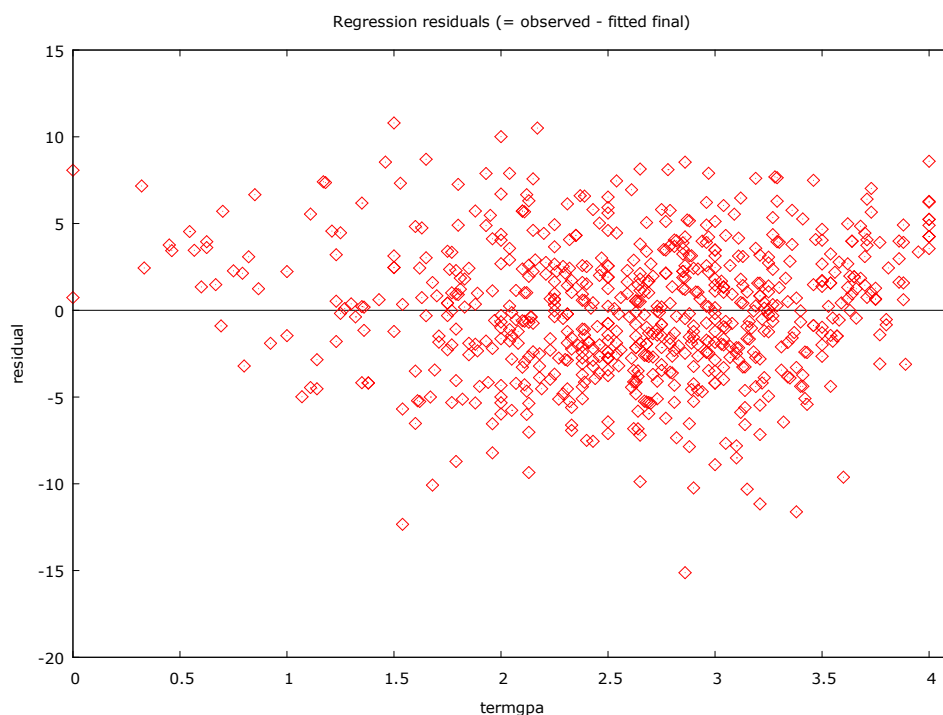


Рисунок 8 - График зависимости остатков от переменной termrgpa

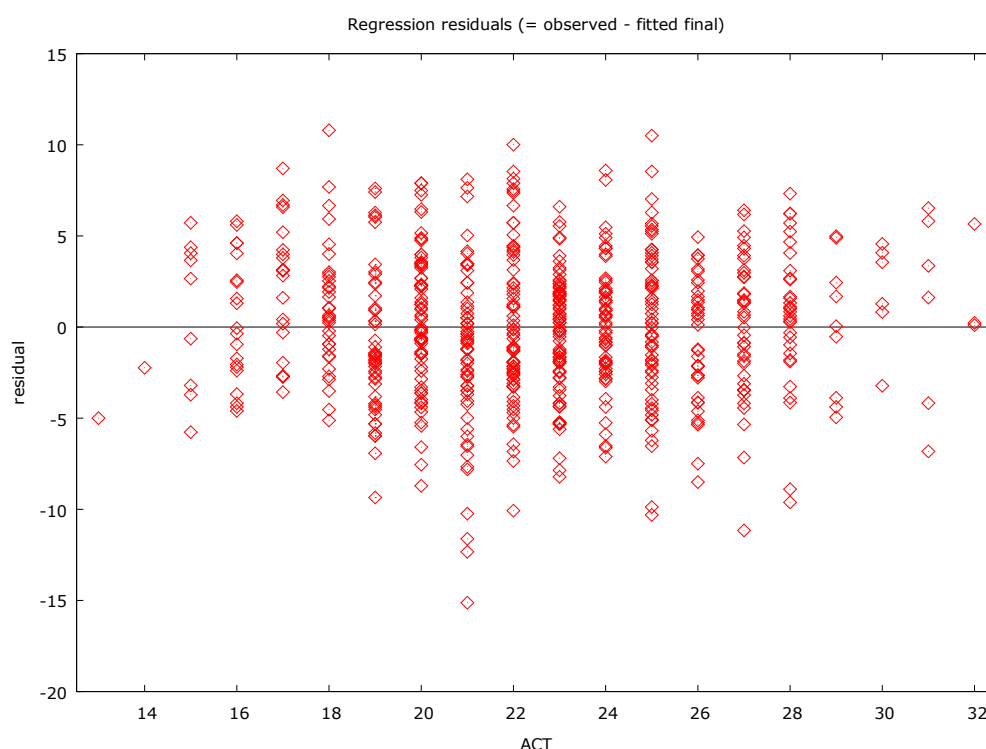


Рисунок 9 - График зависимости остатков от переменной АСТ

Из вышесказанного можно установить, что выполняются все предпосылки для применения МНК для определения параметров рассматриваемой модели (полученной в примере 2.). Построенная модель $\text{final} = 10,8 + 0,339\text{АСТ} + 2,87\text{termrgpa} + u$ на основе ее проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы (в

результате проверки по t -критерию Стьюдента). Такая модель может быть использована для принятия решений и осуществления прогнозов.

4. ПОРЯДОК ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ

1. Открыть набор данных **File\Open Data\ Sample File** (закладка **Wooldridge**) в соответствии с номером варианта (рисунок 10). Для каждого варианта также указаны номера переменных ID# для y , x_1 , x_2 , информацию о которых можно просмотреть, обратившись к команде **Data\Print Description**.

2. Провести оценку параметров линейной регрессионной модели $y = \alpha_0 + \alpha_1 \cdot x_1 + \alpha_2 x_2 + u$ методом 1МНК.

3. Оценить адекватность регрессионной модели в целом и значимость её отдельных параметров

4. Проверить были ли все предпосылки к тому, чтобы применять 1МНК и линейное уравнение регрессии к исходным данным.

№ Вар.	Файл	ID# для	Y	X1	X2
1	401k Pension plan generosity and participation	3	4	6	
2	401ksubs Wealth, income and pension plans	2	5	6	
3	fish Daily data from Fulton fish market	4	2	3	
4	affairs Incidence of extra-marital affairs	11	4	7	
5	consump Consumption, income and interest rate	9	7	2	
6	earns Earnings, productivity and hours (macro)	1	2	3	
7	ceosal2 Firm performance and CEO salary	1	6	3	
8	cps91 Wages and their determinants, 1991 (large)	3	1	4	
9	beautv Physical attractiveness and earnings	1	2	5	
10	cps78_85 Wages and demographics, 1978 and 1985	8	1	9	

Рисунок 10 - Варианты заданий: название файла на закладке Wooldridge, номера варианта и переменных.

5. СОДЕРЖАНИЕ ОТЧЕТА О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ

- 1) Название и цель работы.
- 2) Постановка задачи.
- 3) Этапы выполнения задачи в Gretl.
- 4) Выводы.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Себер Дж. Линейный регрессионный анализ/ Дж. Себер. - М.: Мир, 1980-200с.
2. Демиденко Е. З. Линейная и нелинейная регрессия/ Е.З. Демиденко.- М.: Финансы и статистика, 1981 -320с.
3. Куфель Т. Эконометрика. Решение задач с применением пакета программ GRETL / Т. Куфель. - М.: Горячая линия – Телеком, 2007. – 200с.
4. Using gretl for Principles of Econometrics, 3rd Edition Version 1.01 Lee C. Adkins Professor of Economics Oklahoma State University // <http://www.learneconometrics.com/gretl.html>

ПРИЛОЖЕНИЕ А

(справочное)

Основные описательные статистики

Основные статистики для описания переменных можно получить различными способами. Например, можно отметить несколько переменных и вызвать функцию **View\Summary Statistics** или выбрать в контекстном меню, вызванном нажатием правой кнопкой мыши на выбранной переменной, команду **Summary statistics**.

Пример 1. Откроем файл `example1.gdt`, выделим переменные X1 и X2. Получим основные описательные статистики для данных переменных: **View\Summary Statistics**. Окно результатов имеет следующий вид (рисунок А.1)

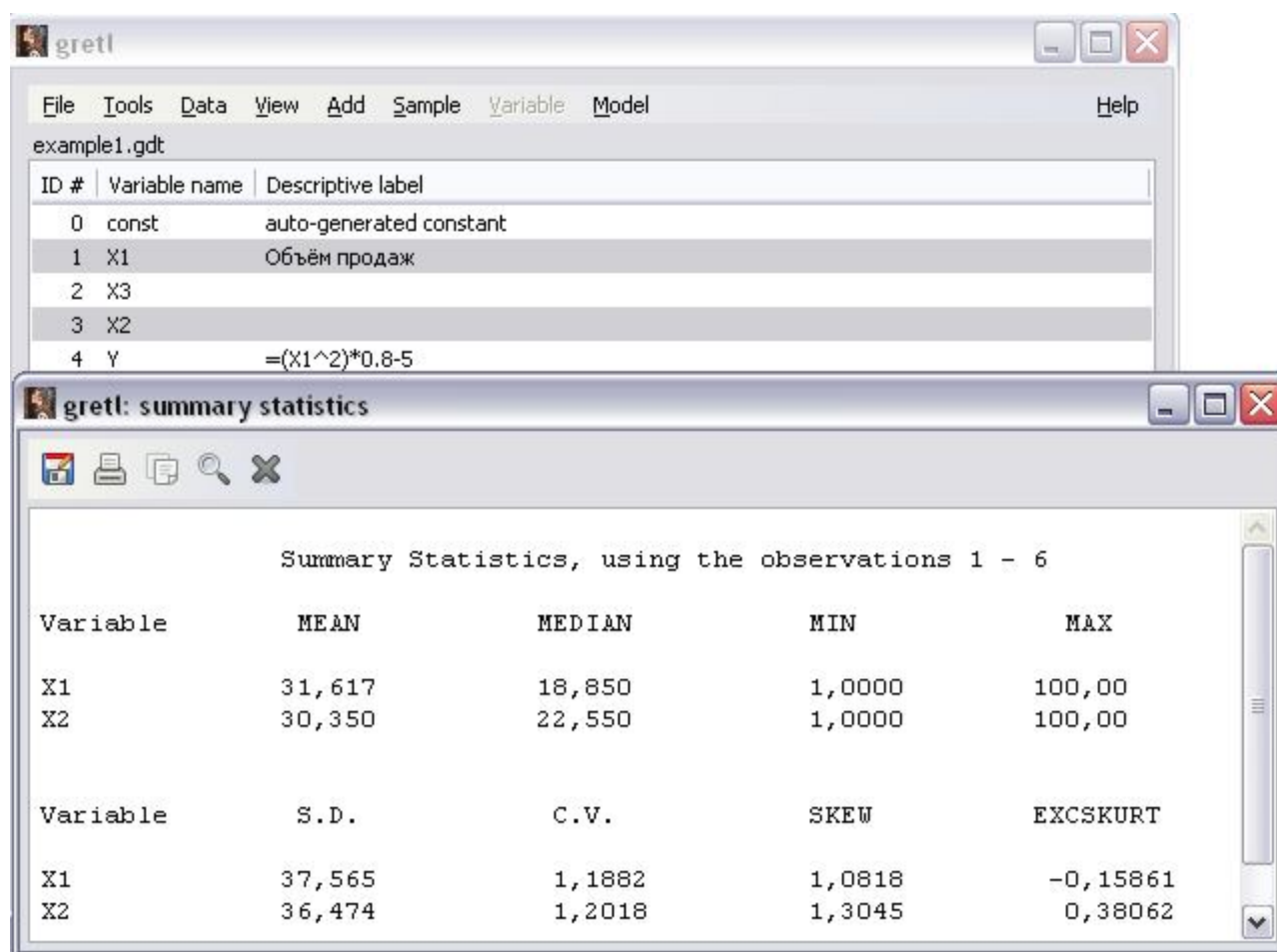


Рисунок А.1 - Основные описательные статистики переменных X1 и X2

В окне результатов представляются:

- среднее арифметическое (**mean**),
- медиана (**median**),
- минимальное (**min**) и максимальное (**max**) значения,
- среднеквадратическое отклонение (**S.D.**),
- коэффициент изменчивости (вариации) (**C.V.= S.D./mean**),
- коэффициент асимметрии (**SKEW=центральный момент третьего порядка/S.D.³**) –

величина, характеризующая асимметрию распределения данной случайной величины (степень асимметричности плотности вероятности относительно оси, проходящей через ее центр тяжести)

Центральный момент третьего порядка определяется как математическое ожидание куба разности случайной величины (X_1) и её математического ожидания.

- **коэффициент концентрации** ($EXCSKURT = \text{центральный момент четвёртого порядка} / S.D.^4 - 3$) - **коэффициент эксцесса** (kurtosis) выборки случайных данных X_1 характеризует степень сглаженности плотности вероятности в окрестности главного максимума. Он показывает, насколько острую вершину имеет плотность вероятности по сравнению с нормальным распределением. Если коэффициент эксцесса больше нуля, то распределение имеет более острую вершину, чем распределение Гаусса, если меньше нуля, то более плоскую.

ПРИЛОЖЕНИЕ Б (справочное) Статистические таблицы в GRETl

Gretl предусматривает непосредственный доступ к статистическим таблицам. Пакет Gretl содержит встроенные статистические таблицы для следующих распределений: нормального, t-распределения Стьюдента, F-распределения Фишера, хи-квадрат, Пуассона, биномиального и распределения Дарбина-Уотсона.

- **Tools\Statistical tables** - с помощью данной функции можно получить критические значения для вышеперечисленных критериев путём указания параметров (числа степеней свободы, стандартного отклонения, среднего и т.д.) и уровня значимости (максимально допустимой вероятности ошибочного принятия альтернативной гипотезы).

- **Tools\P-value finder** позволяет найти вероятность того, что значение критерия будет превышать расчётное (p-value).

- **Tools\ Distribution Graphs** выводит графики вышеперечисленных распределений

Пример 2.

1. Выберем закладку t- распределение Стьюдента, введём число степеней свободы ($5 = \text{число наблюдений} - 1$). Предусматривая уровень значимости 5%, введём его половинное значение (0.025), поскольку критерий двусторонний и нажмём ОК (рисунок Б.1).

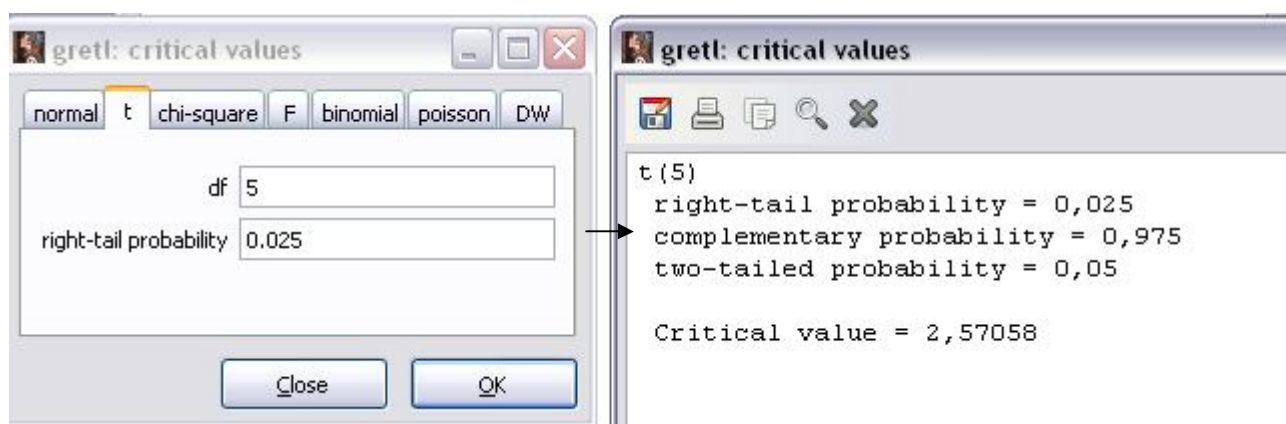


Рисунок Б.1 - Нахождение критических значений критерия по заданным уровню значимости и числу степеней свободы

Critical value – верхнее критическое значение t-критерия 2,57 (верхняя граница области принятия гипотезы и критической области отвержения гипотезы H_0).

Two –tailed probability - вероятность того, что значение t-критерия будет по модулю больше критического значения 2,57, т.е. будет находиться в критической области, где нулевая гипотеза H_0 отвергается.

Right –tail probability - вероятность того, что значение t-критерия будет больше критического значения 2,57.

2. Функция **Tools\ P-value finder** позволяет найти вероятность того, что значение критерия будет превышать указанное (расчётное). Введём число степеней свободы (df) 5, а расчётное значение критерия (value=3) превышающее критическое (гипотеза H_0 отвергается), тогда в результате получим значение p-value=0,03 или 3% (рисунок Б.2). Т.о. если p-value меньше выбранного уровня значимости, то нулевая гипотеза H_0 отвергается, а если больше - принимается.

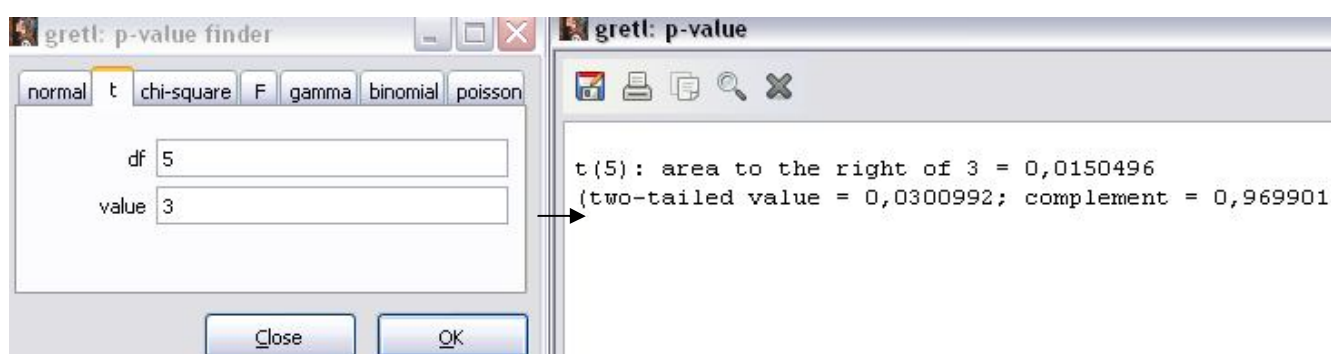


Рисунок Б.2 - Нахождение P-value по расчётному значению критерия и числу степеней свободы

3. Используя функцию **Tools\ Distribution Graphs** построим график t-распределения для df= 5

ПРИЛОЖЕНИЕ В

(справочное)

Построение графиков

Пакет программ GRETЛ обладает обширными возможностями построения графиков. Наиболее разнообразен перечень возможных графиков для данных в форме временных рядов. Графики строятся с применением группы функций **View\Graph specified vars\...**

Функция **View\Graph specified vars\ 3D plot** позволяет создавать трёхмерные графики и при помощи мыши вращать изображение для получения необходимого ракурса.

Функция **View\Multiple Graphs\X-Y scatters...** позволяет создавать окна, содержащие несколько графиков. В данном окне можно создавать от двух до шести графиков. Окно с несколькими графиками может иметь только одну категорию X или Y и не более шести остальных категорий.

Пример. 3.

Построим трёхмерный график, отражающий зависимость между переменными Y, X1 и X2 (файл example1.gdt): **View\Graph specified vars\ 3D plot**. Используем кнопку Choose для выбора соответствующих переменных и нажмём кнопку ОК (рисунок В.1). Для наилучшего рассмотрения графика развернём его при помощи мыши на 90 градусов в горизонтальной плоскости (рисунок В.2).

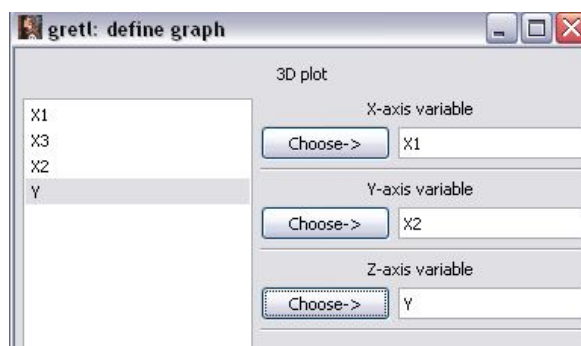


Рисунок В.1 - Построение трёхмерных графиков

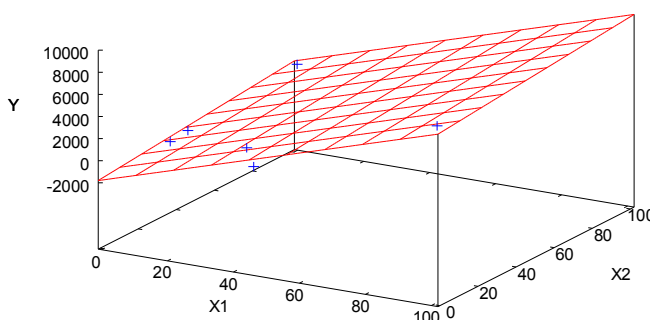


Рисунок В.2 - Графическое отражение зависимости между Y, X1 и X2

Построим диаграммы рассеяния X_1 - Y , X_2 - Y и X_3 - Y , обратившись к функции View\Multiple Graphs\X-Y scatters в открывшемся окне выберем значения соответствующих переменных кнопками Choose и Add и нажмём кнопку ОК (рисунок В.3). Полученные графики показаны на Рисунке В.4.

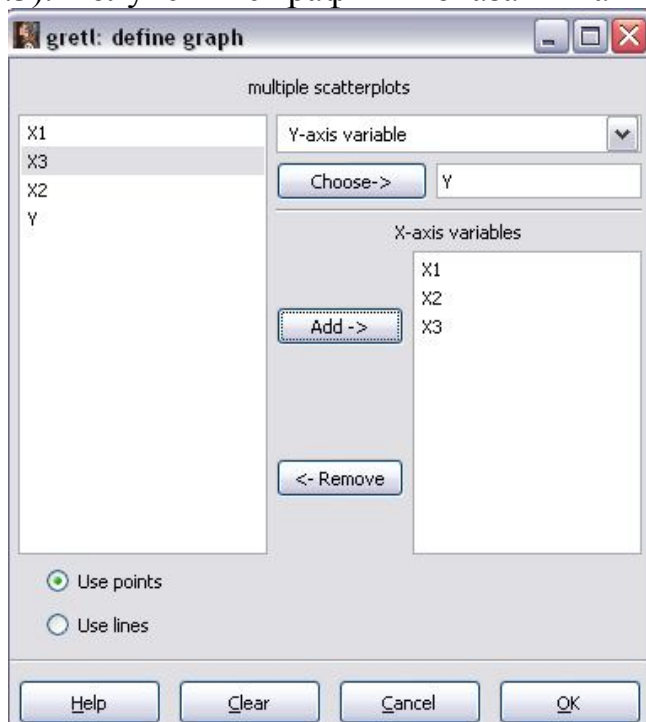


Рисунок В.3 - Построение нескольких графиков в одном окне

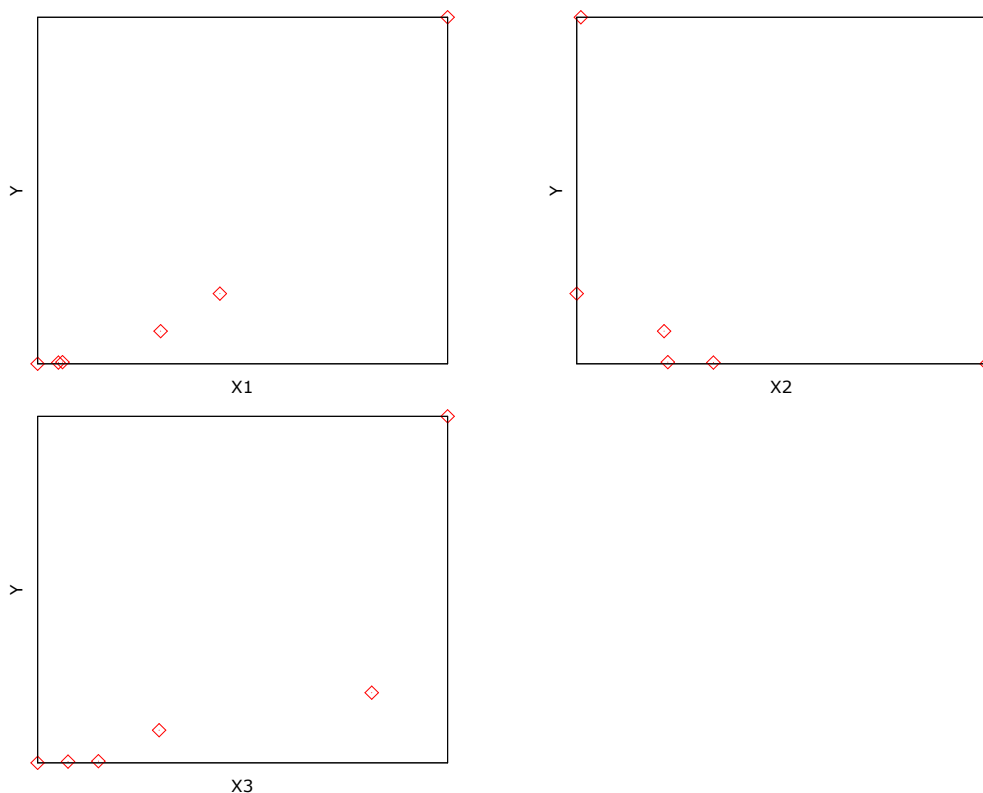


Рисунок В.4 - Диаграммы рассеяния Y - X_1 , Y - X_2 и Y - X_3

Заказ № _____ от « _____ » _____ 2008. Тираж _____ экз.

Изд-во СевНТУ