

А. С. Малова



ОСНОВЫ ЭКОНОМЕТРИКИ в среде GRET

Учебное пособие

Линейная регрессионная модель

Тесты Фишера (Fisher test), Стьюдента (t-test), Чой (test Chow)

Проверка правильности спецификации модели (RESET test)

Интерпретация коэффициентов регрессии и прогнозирование

Оценка регрессии в логарифмах и интерпретация

Проверка линейных ограничений
на коэффициенты регрессии

И др.



ПРОСПЕКТ

Александра Малова

**Основы эконометрики в среде
GRETЛ. Учебное пособие**

«Проспект»

Малова А. С.

Основы эконометрики в среде GRETЛ. Учебное пособие /
А. С. Малова — «Проспект»,

ISBN 978-5-39-220234-8

Данное пособие представляет собой вспомогательный методический материал для работы в эконометрической среде GRETЛ. Оно предназначено студентам бакалавриата по направлениям «Экономика», «Бизнес-информатика», «Управление персоналом», «Менеджмент» для использования на практических и семинарских занятиях по курсу «Эконометрика (пространственные данные)», а также может использоваться любыми заинтересованными лицами в качестве краткого руководства по использованию GRETЛ. Пособие включает в себя обзор основных тем базового курса «Эконометрика», подробный разбор возможностей и функций эконометрического пакета GRETЛ, а также примеры практической реализации тех или иных методов. В данном издании для иллюстрации возможностей эконометрического пакета использовались примеры из учебника Jeffrey M. Wooldridge «Introductory Econometrics: A Modern Approach, 2nd edition». Все файлы с данными находятся в открытом доступе и могут быть свободно использованы. Пособие организовано таким образом, что читатель имеет возможность самостоятельно проделать все действия, необходимые для решения стоящей перед ним эконометрической задачи.

ISBN 978-5-39-220234-8

© Малова А. С.

© Проспект

Содержание

Введение	6
1. Линейная регрессионная модель	7
2. Оценка линейной регрессионной модели	8
3. Тест Фишера (Fisher test)	12
Конец ознакомительного фрагмента.	15

А. С. Малова
Основы эконометрики в среде GRETL
Учебное пособие



ebooks@prospekt.org

Введение

Цель данного пособия – познакомить читателя с основами проведения эконометрических исследований в среде **GRETL**. Основная аудитория данной книги – студенты бакалавриата, обучающиеся по направлениям «Экономика», «Бизнес-информатика», «Управление персоналом», «Менеджмент», однако она может быть полезна и студентам других направлений, а также представителям бизнес-сообщества, которые по роду своей деятельности столкнулись с необходимостью проведения эконометрических исследований. Данное учебное пособие – это попытка практического изложения основ эконометрики с минимальными теоретическими выкладками, при этом предполагается, что недостаток теоретических знаний должен быть восполнен читателем самостоятельно с помощью учебников по основам эконометрики. Для обеспечения связи практических навыков с теоретическими знаниями в области эконометрики ко всем рассматриваемым темам даются ссылки на литературу. При этом основная задача данного пособия – помочь читателю в освоении эконометрики, изложить некоторые технические аспекты проведения исследований с использованием среды **GRETL**. Почему именно **GRETL**? Данный эконометрический пакет является бесплатным программным продуктом, который, с одной стороны, доступен любому пользователю, а с другой – обладает достаточно обширными возможностями для анализа данных и проведения эмпирических исследований. Немаловажным является и то, что в **GRETL** имеется значительный пул данных из большинства классических зарубежных учебников по основам эконометрики, что позволит достаточно легко переключиться с простейших примеров, рассмотренных в данном пособии, на более сложные содержательные задачи и кейсы из учебников.

В данном пособии весь материал излагается с точки зрения практики – то есть все основные разделы курса эконометрики для бакалавриантов даны в примерах и задачах. Поскольку невозможно приобрести навык проведения эконометрических расчетов, только изучая учебник, предполагается, что читатель должен иметь возможность проделать все излагаемые действия на практике. С этой целью в пособии использовались данные из учебника J. M. Wooldridge «Basic econometrics», которые доступны в **GRETL**. Все наборы данных при первом обращении к ним в пособии обозначены ссылками и указателями на источник.

Перед тем как начать осваивать основы эконометрики в среде **GRETL**, необходимо скачать и установить на свой компьютер сам статистический пакет. Он доступен по ссылке <http://GRETL.sourceforge.net/>. Вся информация о том, как установить **GRETL**, приводится на сайте, поэтому нет нужды в подробном изложении, стоит лишь сказать, что программа имеет версию как под ОС Windows, так и под Mac OS, а также что библиотеки данных должны быть установлены отдельно, для этого нужно перейти по ссылке http://GRETL.sourceforge.net/GRETL_data.html.

Удачи в проведении интересных, содержательных и полезных эконометрических исследований!

1. Линейная регрессионная модель

Для начала введем некоторые обозначения. Предположим, что некоторая величина Y зависит от величин X_1, X_2, \dots, X_{k-1} . Введем понятие регрессионного уравнения – это уравнение вида $Y_t = \beta_0 + \beta_1 \cdot X_{t1} + \beta_2 \cdot X_{t2} + \dots + \beta_{k-1} \cdot X_{tk-1} + \varepsilon_t$, где $t \in 1 \dots n$. Через n обозначим число наблюдений, по которым строится регрессия, k – число регрессоров в модели, ε_t – случайная величина, которая носит название ошибки регрессии.

Модель такого вида называется классической линейной регрессионной моделью (ЛРМ) в случае, если выполняются следующие предпосылки:

1. $Y_t = \beta_0 + \beta_1 \cdot X_{t1} + \beta_2 \cdot X_{t2} + \dots + \beta_{k-1} \cdot X_{tk-1} + \varepsilon_t, t \in 1 \dots n$ – линейная спецификация модели, где $\beta_0, \beta_1 \dots \beta_{k-1}$ – коэффициенты модели, которые подлежат определению, $\varepsilon_t, t \in 1 \dots n$ – ошибки модели.

2. $X_{t1}, X_{t2}, \dots, X_{tk-1}, t \in 1 \dots n$ – детерминированные величины.

3. $E(\varepsilon_t) = 0$ – математическое ожидание ошибок равно нулю, $E(\varepsilon_t^2) = \sigma^2$, дисперсия ошибок не зависит от номера наблюдения.

4. $E(\varepsilon_t, \varepsilon_s) = 0, t \neq s$ – совместное математическое ожидание ошибок разных наблюдений равно нулю.

5. Если выполняется дополнительная предпосылка о нормальном распределении ошибок $\varepsilon_t \sim N(0, \sigma^2)$, то классическая линейная регрессионная модель называется нормальной линейной регрессионной моделью (НЛРМ).

Подробнее о предпосылках линейной регрессионной модели можно прочесть в [2, 3].

2. Оценка линейной регрессионной модели

Рассмотрим множественную линейную регрессию

$$wage_t = \beta_0 + \beta_1 \cdot educ_t + \beta_2 \cdot exper_t + \beta_3 \cdot tenure_t + \varepsilon_t, t \in 1 \dots n,$$

где $wage_t$ – средний уровень заработной платы в час в долларах, $educ_t$ – образование в годах, $exper_t$ – общий стаж работы в годах, $tenure_t$ – опыт работы у текущего работодателя, в годах, ε_t – ошибка регрессии, n – число наблюдений [файл с данными *wage1.gdt*].

Для того чтобы оценить предложенную модель по методу наименьших квадратов (МНК), используем команду меню *Модель – Метод наименьших квадратов*.

В появившемся диалоговом окне в поле *Зависимая переменная* помещаем переменную $wage_t$ (для этого выделяем ее курсором в списке переменных и нажимаем на стрелку, соответствующую окну *Зависимая переменная*. Данный способ перемещения переменных справедлив для всех операций с диалоговыми окнами).

Для дальнейшего удобства можно поставить галочку в окошке *Установить по умолчанию*. Это делается для того, чтобы при изменении спецификации исследуемой модели зависимая переменная не менялась. В окно *Регрессоры* отправляем регрессоры модели – это переменные $educ_t$, $exper_t$, $tenure_t$.

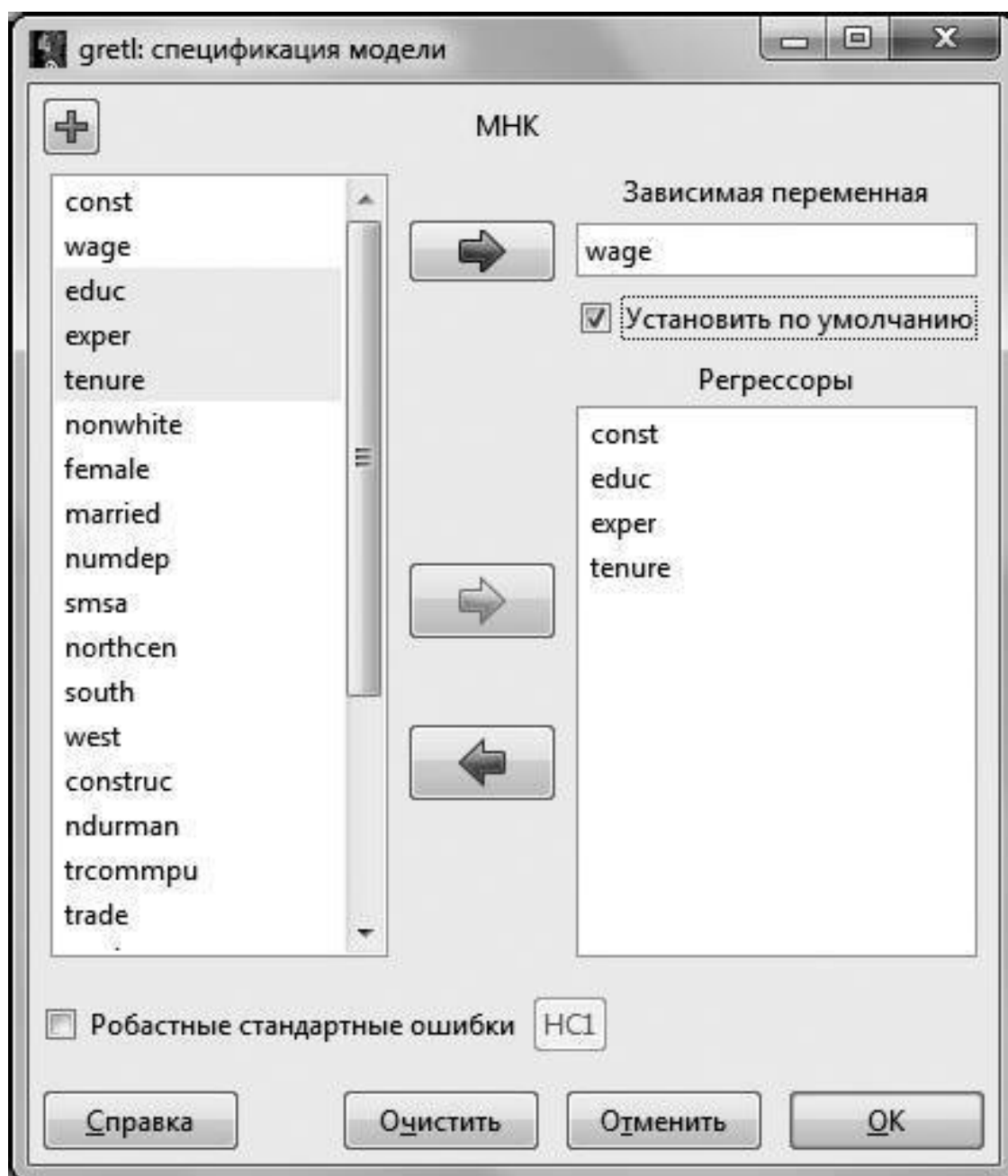


Рис. 2.1

После этого нажимаем *ОК*. В результате коэффициенты модели были оценены методом наименьших квадратов. Результат оценки представлен на рис. 2.2.

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	-2,87273	0,728964	-3,941	9,22e-05	***
educ	0,598965	0,0512835	11,68	3,68e-028	***
exper	0,0223395	0,0120568	1,853	0,0645	*
tenure	0,169269	0,0216446	7,820	2,93e-014	***
Среднее зав. перемен	5,896103	Ст. откл. зав. перемен	3,693086		
Сумма кв. остатков	4966,303	Ст. ошибка модели	3,084476		
R-квадрат	0,306422	Испр. R-квадрат	0,302436		
F(3, 522)	76,87317	P-значение (F)	3,41e-41		
Лог. правдоподобие	-1336,831	Крит. Акаике	2681,662		
Крит. Шварца	2698,723	Крит. Хеннана-Куинна	2688,342		

Рис. 2.2

Для того чтобы понимать, какие результаты позволяет получить **GRETL**, разберем информацию, представленную на распечатке по строкам сверху вниз.

В первой строке указывается метод оценки и количество наблюдений, по которым производилась оценка. Достаточно часто случается, что количество наблюдений, по которым производилась оценка, не совпадает с числом наблюдений в исходной выборке, даже если она не была ограничена. Это может быть связано, например, с наличием пропусков в данных.

Вторая строка напоминает нам о том, какая переменная была выбрана в качестве зависимой.

После двух первых строк следуют подтаблицы непосредственно с результатами оценивания. В первой подтаблице указаны регрессоры, включенные в модель, напротив каждого из них указывается его коэффициент (столбец *Коэффициенты*), стандартная ошибка оценки коэффициента (столбец *Ст. ошибка*), значение статистики Стьюдента для коэффициента (столбец *t-статистика*) и вероятность ошибки I рода (столбец *P-значение*). Стоит отметить, что константа тоже является регрессором, и для нее также рассчитываются все указанные характеристики.

По распечатке, представленной на рис. 2.2, мы можем выписать получившееся уравнение регрессии:

$$\widehat{wage}_i = -2,87 + 0,6 \cdot educ_i + 0,02 \cdot exper_i + 0,17 \cdot tenure_i$$

Аналогично можно получить оцененное уравнение и в **GRETL**, для этого выбираем в меню регрессии *Файл – Просмотреть как уравнение*.

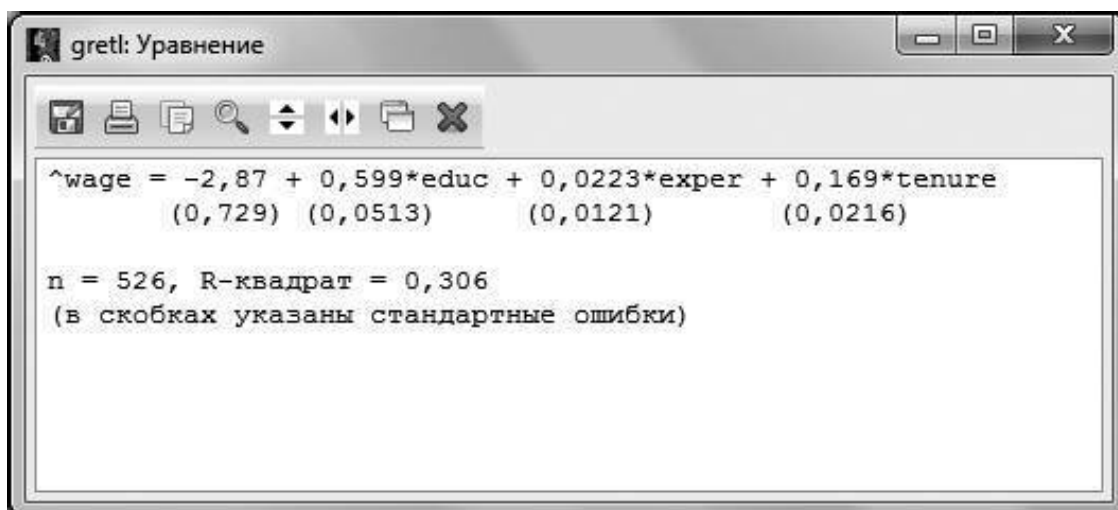


Рис. 2.3

Однако для того, чтобы иметь возможность дать интерпретацию коэффициентам регрессии и строить прогнозы, необходимо проверить, является ли полученная модель адекватной.

Для этого, в свою очередь, необходимо провести ряд эконометрических тестов, а именно проверить значимость регрессии в целом, значимость отдельных коэффициентов регрессии, оценить качество полученного регрессионного уравнения. Вообще говоря, перед проверкой значимости и качества уравнения необходимо провести тесты на выполнение основных предпосылок линейной регрессионной модели (гомоскедастичность, отсутствие автокорреляции). На данном этапе мы будем считать эти тесты проведенными и вернемся к вопросам выполнения предпосылок ЛРМ позднее.

3. Тест Фишера (Fisher test)

Для начала проверим гипотезу о незначимости регрессии в целом. Тест позволит понять, является ли построенная модель адекватной с точки зрения статистики. Для этой цели воспользуемся тестом Фишера [3].

Тест Фишера (Fisher test) позволяет проверить незначимость регрессии в целом, то есть установить, равны ли коэффициенты одновременно при всех регрессорах нулю. Если коэффициенты признаются равными нулю, регрессия считается незначимой, если коэффициент хотя бы при одном регрессоре отличен от нуля, регрессия значима.

Гипотезы	Расчетная статистика	Правило принятия решения (гипотеза H_0 отвергается, если):
$H_0 : \beta_1 = \dots = \beta_{k-1} = 0$ H_1 : как минимум один из коэффициентов отличен от нуля	$F_{\text{расчет}} = \frac{RSS / (k-1)}{ESS / (n-k)}$ <p>где RSS — сумма квадратов остатков, обусловленная регрессией;</p> <p>ESS — сумма квадратов остатков, обусловленная ошибками</p>	$F_{\text{расчет}} > F_{\alpha}(k-1; n-k)$ <p>где n — число наблюдений;</p> <p>k — число регрессоров;</p> <p>α — уровень значимости</p>

Сформулируем гипотезы для проверки незначимости регрессии в целом в рассматриваемом примере [файл с данными *wage1.gdt*] модели $wage_t = \beta_0 + \beta_1 \cdot educ_t + \beta_2 \cdot exper_t + \beta_3 \cdot tenure_t + \varepsilon_t, t \in 1 \dots n$:

$$H_0 : \beta_{\text{exper}} = \beta_{\text{tenure}} = \beta_{\text{educ}} = 0$$

H_1 : как минимум один из коэффициентов отличен от нуля.

Для принятия решения о том, какую гипотезу нужно отвергнуть, построим F -статистику. Для этого нам должны быть известны (помимо уже имеющихся параметров n — объем выборки и k — число регрессоров в модели) величины RSS и ESS . В явном виде в распечатке на рис. 2.2 дано значение ESS — сумма квадратов остатков, которая составляет $ESS = 4966,3$, а также из распечатки известен коэффициент детерминации $R^2 = 0,31$ (подробнее о коэффициенте детерминации и его интерпретации можно прочесть в § 7).

Если вспомнить, что $R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$, а $TSS = ESS + RSS$, то можно путем простых алгебраических преобразований найти необходимую нам величину RSS . При этом $RSS = \frac{R^2 \cdot ESS}{1 - R^2} = 2194,1$.

Отсюда можно вычислить $F_{расчет} = \frac{RSS/k-1}{ESS/n-k} = \frac{2231,2/(4-1)}{4966,3/(526-4)} \approx 76,8$. Критическое значение F -статистики возьмем на уровне значимости 5 %: $F_{\alpha}(k-1; n-k) = F_{0,05}(4-1; 526-4) = 2,6$ (чтобы получить это значение, в основном меню **GRETL** нужно выбрать *Инструменты – Критические значения – Фишера* и ввести необходимое число степеней свободы и правостороннюю вероятность либо посмотреть в статистических таблицах распределения Фишера для уровня значимости 5 %, например в [7]).

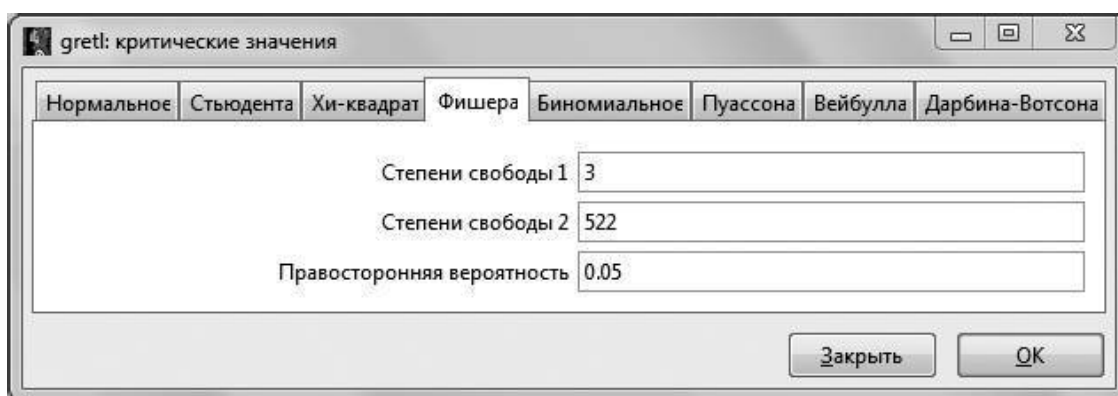


Рис. 3.1

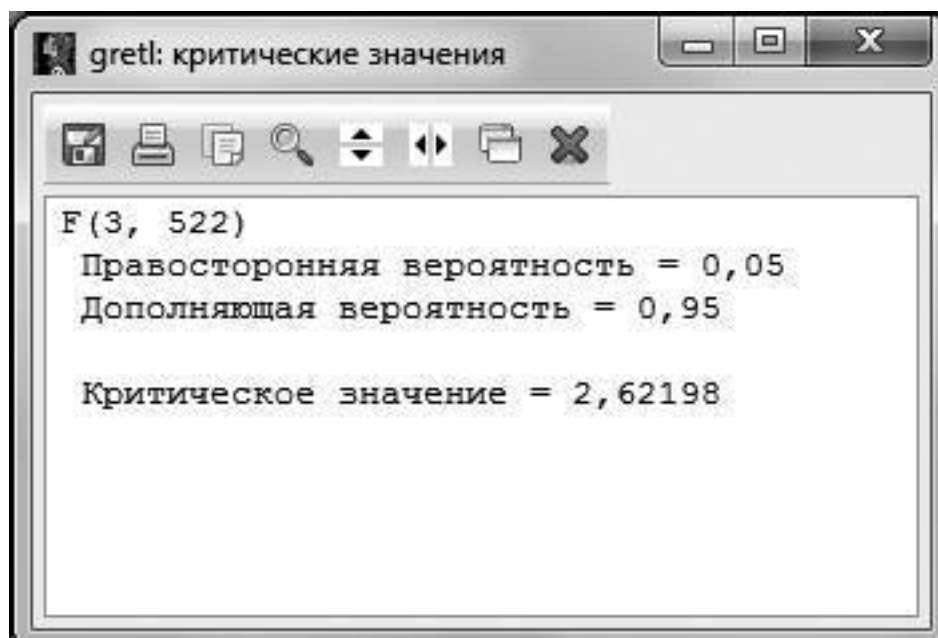


Рис. 3.2

Уровень значимости, на котором принимается решение о том, какую гипотезу не отвергать, остается на усмотрение исследователя. Как правило, если нет представления, какой именно уровень значимости брать, предлагается выбирать 5 %. В случаях работы с маленьким по объему выборками (от 30 до 100 наблюдений) предлагается брать уровень значимости 10 %. Для больших выборок (более 1000 наблюдений) можно взять уровень значимости 1 %. В

нашем случае объем выборки средний (526 наблюдений, эта информация дана в первой строке распечатки на рис. 2.2.), поэтому можно было принять $\alpha = 0,05$.

Сравниваем расчетное значение F -статистики с критическим $F_{\alpha}(k-1; n-k)$, то есть $78,2 > 2,6$. Следовательно, можно сделать вывод, что гипотеза H_0 незначимости регрессии в целом отвергается.

Тест Фишера можно провести также в полуавтоматическом режиме и в автоматическом режиме. Полуавтоматический режим состоит в том, что нам не нужно вручную вычислять значение расчетной F -статистики, оно дано в распечатке на рис. 2.2. В этом случае нужно лишь выяснить критическое значение F -статистики и сравнить расчетное значение с критическим.

В автоматическом режиме нужно также воспользоваться распечаткой **GRETL** и посмотреть на p -значение статистики Фишера на рис. 2.2 (в распечатке p -значение (F)). В p -значении содержится вероятность ошибки I рода. Таким образом, p -значение (F) для теста Фишера – это вероятность ошибки I рода при тестировании гипотезы $H_0: \beta_1 = \dots = \beta_{k-1} = 0$. По существу это вероятность ошибиться, отвергнув гипотезу H_0 . Для принятия решения, можно ли отвергнуть гипотезу H_0 , нужно сравнить p -значение с заданным уровнем значимости α . Уровень значимости задает вероятность ошибки I рода, то есть, грубо говоря, какую долю ошибок мы готовы себе позволить, отвергнув гипотезу H_0 . Если p -значение меньше принятого уровня значимости, то маловероятно, что мы ошибемся, отвергая гипотезу H_0 в ситуации, когда p -значение больше уровня значимости, вероятна ошибка в случае отклонения нулевой гипотезы, поэтому ее стоит принять. Отсюда можно сделать вывод, что p -

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.