

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

О.И. Стебунова, К.В. Пивоварова

СРЕДА РАЗРАБОТКИ «RSTUDIO» В ЭКОНОМЕТРИЧЕСКОМ МОДЕЛИРОВАНИИ

Методические указания

Рекомендовано к изданию редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательным программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.03.05 Бизнес-информатика, 38.04.01 Экономика

Оренбург
2018

УДК 330.4(076.5)

ББК 65в631я7

С79

Рецензент – кандидат экономических наук, доцент О.И. Бантикова

Стебунова О.И.

С79 Среда разработки «RStudio» в эконометрическом моделировании: методические указания/О.И. Стебунова, К.В. Пивоварова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018.– 37 с.

Методические указания содержат рекомендации по выполнению лабораторной работы в RStudio по дисциплине «Эконометрика», а также другим дисциплинам, требующим навыков моделирования взаимосвязей между социально-экономическими показателями.

Методические указания предназначены для обучающихся по направлениям подготовки 01.03.04 Прикладная математика, 38.03.05 Бизнес-информатика, 38.04.01 Экономика (профиль «Математические и инструментальные методы анализа социальных и экономических процессов»), а также для студентов других направлений с углубленной математической подготовкой, аспирантов и преподавателей, занимающихся моделированием социально-экономических процессов.

УДК 330.4(076.5)

ББК 65в631я7

© Стебунова О.И.,
Пивоварова К.В.

© ОГУ, 2018

Содержание

Введение.....	4
1 Теоретические аспекты построения квантильной модели регрессии	7
1.1 Квантильная модель регрессии: постановка задачи и оценка коэффициентов	7
1.2 Вопросы, выносимые на семинарские занятия по теме «Непараметрические методы оценивания эконометрических моделей».....	10
2 Описание лабораторной работы	10
3 Постановка задачи.....	11
4 Порядок выполнения лабораторной работы	11
4.1 Описание основного интерфейса RStudio	12
4.1 Ввод и первичная обработка данных в программе RStudio.....	15
4.2 Построение и исследование линейной модели регрессии в программе «RStudio».....	22
4.3 Построение модели квантильной регрессии	28
4.4 Описание команд для построения квантильной регрессии в кросс-платформенном пакете GRETL	32
5 Содержание письменного отчета	34
6 Вопросы к защите лабораторной работы	34
Список использованных источников	35
Приложение А	36
Приложение Б.....	37

Введение

В методических указаниях описывается подход к моделированию социально-экономических процессов на основе непараметрических методов оценивания, в частности квантильной регрессии. Непараметрические методы оценивания позволяют избежать ряд недостатков, связанных с наличием распределений отличных от нормального закона распределения, а также присутствием в выборке небольшой доли «выбросов», то есть наблюдений, вызванных каким-либо «шумом».

Цель методических указаний заключается в формировании навыков моделирования социально-экономических процессов на основе квантильной регрессии в среде разработки «RSTUDIO».

В теоретической части предлагаемых методических указаниях изложены основные вопросы построения квантильной модели регрессии. Приведен перечень теоретических вопросов по теме «Непараметрические методы оценивания эконометрических моделей», позволяющий студенту систематизировать свои знания и облегчить подготовку к семинарскому занятию. В практической части на конкретном примере описывается процесс оценки и исследования КЛММР и квантильной модели регрессии в программе RStudio, а также описание команд для оценки параметров модели в распространяемом профессиональном кросс-платформенном пакете GRET, приводится интерпретация и сравнение полученных результатов моделирования. В методических указаниях сформулирована постановка задачи, приведены требования к оформлению отчета и вопросы к защите лабораторной работы.

Методические указания направлены на формирование следующих результатов обучения:

- для 01.03.04 Прикладная математика, профиль «Применение математических методов к применению инженерных и экономических задач»: ОПК-1 (готовность к самостоятельной работе); ОПК-2 (способность

использовать современные математические методы и современные прикладные программные средства и осваивать современные технологии программирования); ПК-1 (способность использовать стандартные пакеты прикладных программ для решения практических задач на электронных вычислительных машинах); ПК-10 (готовность применять математический аппарат для решения поставленных задач, способностью применить соответствующую процессу математическую модель и проверить ее адекватность, провести анализ результатов моделирования, принять решение на основе полученных результатов); ПК-12 (способность самостоятельно изучать новые разделы фундаментальных наук).

- для 38.03.05 Бизнес-информатика профиль «Моделирование бизнес-систем»: ОПК-3 (способности работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных компьютерных сетях); ПК-2 (проведение исследования и анализа рынка информационных систем и информационно-коммуникативных технологий); ПК-17 (способности использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования); ПК-18 (способности использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования); ПК-19 (умение готовить научно-технические отчеты, презентации, научные публикации по результатам выполненных исследований);

- для 38.04.01 Экономика профиль «Математические и инструментальные методы анализа социальных и экономических процессов»: ОК-1 (способность к абстрактному мышлению, анализу, синтезу); ОПК-1 (готовность к коммуникации в устной и письменной формах на русском и иностранном языках для решения задач профессиональной деятельности); ПК-1 (способность обобщать и критически оценивать результаты, полученные отечественными и зарубежными исследователями, выявлять

перспективные направления, составлять программу исследований); ПК-3 (способность проводить самостоятельные исследования в соответствии с разработанной программой); ПК-4 (способность представлять результаты проведенного исследования научному сообществу в виде статьи или доклада); ПК-9 (способность анализировать и использовать различные источники информации для проведения экономических расчетов).

Использование предлагаемых методических указаний в учебном процессе позволит студенту в достаточной степени овладеть непараметрическими методами оценивания зависимости между социально-экономическими показателями, а также приобрести навыки моделирования в программе RStudio.

1 Теоретические аспекты построения квантильной модели регрессии

1.1 Квантильная модель регрессии: постановка задачи и оценка коэффициентов

Пусть на основе предварительного анализа установлено, что эндогенная переменная (результативный признак) y зависит от predetermined (объясняющих переменных) x_1, x_2, \dots, x_k . Результаты наблюдений результирующего признака и объясняющих переменных представлены вектором $Y_{n \times 1} = (y_1 \ y_2 \ \dots \ y_n)^T$ и матрицей X типа «объект-свойство»:

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

где y_i – наблюдаемое значение результативного признака для i -го объекта;

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

Задача заключается в построении модели, описывающей взаимосвязь между результативным признаком и объясняющими переменными. Для решения поставленной задачи предлагается использовать линейную модель регрессии, в том числе квантильную модель регрессии.

В регрессионной модели q -й квантилью случайной величины η с функцией распределения $F_\eta(x) = P(\eta < x)$ называется такое число, что данная случайная величина попадает в интервал левее его с вероятностью, не

превосходящей q . То есть, это такое число x_q , удовлетворяющая неравенствам $F_\eta(x_q) \leq q$ и $F_\eta(x_q + 0) \geq q$ или им эквивалентным $P(\eta < x_q) \leq q$ и $P(\eta > x_q) \leq 1 - q$. Если F_η - непрерывная и строго монотонная функция, то существует единственный квантиль x_q любого уроня $q \in (0,1)$, который однозначно определяется из уравнения $F_\eta(x_q) = q$, и, следовательно, выражается через функцию, обратную к функции распределения:

$$x_q = F_\eta^{-1}(q)$$

Для моделирования взаимосвязи между результативным признаком и объясняющими переменными предлагается использовать функцию условной квантили (1):

$$P(y_i < y(q) / x_i) = F_{\varepsilon_q}(y(q) - x_i^T \beta(q) / x_i). \quad (1)$$

Соотношение (1) в терминах регрессионного анализа может быть представлено в виде функции квантильной регрессии (2):

$$Quant_q(y / x_1, x_2, \dots, x_k) \equiv \tilde{y}(q) = \beta_0(q) + \beta_1(q)x_1 + \beta_2(q)x_2 + \dots + \beta_k(q)x_k, \quad (2)$$

или линейной модели квантильной регрессии вида (3):

$$Quant_q(y_i / x_{i1}, x_{i2}, \dots, x_{ik}) \equiv y_i(q) = \beta_0(q) + \beta_1(q)x_{i1} + \beta_2(q)x_{i2} + \dots + \beta_k(q)x_{ik} + \varepsilon_i(q), \quad (3)$$

Метод квантильной регрессии не обязательно привязан к линейным зависимостям вида (4). Его легко модифицировать путем ведения нелинейного преобразования. В частности, введение функции $g(\cdot)$, являющейся функцией связи, в соотношение

$$y = g(X\beta), \quad (4)$$

позволит его привести к виду:

$$Quant_\theta(y_i | x_i) = g^{-1}(x_i \beta_\theta), \quad (5)$$

где $\text{Quant}_\theta(y_i | x_i)$ – условная квантиль y_i для вероятности θ на векторе зависимых переменных x_i ;

β_0 – соответствующий вектор-столбец коэффициентов регрессии.

Представление условного математического ожидания, что характерно для классической регрессии, в виде условной квантили принципиально не меняет качество прогноза с точки зрения интерпретации. Так как при $q = 0,5$ мы имеем медиану, которая как и математическое ожидание является характеристикой положения, а для симметричных распределений они совпадают. Поэтому квантильная и классическая регрессия дают близкие по смыслу оценки, хотя и обладающие различными статистическими свойствами.

Если функция F_{ε_q} известна, то для оценки вектора коэффициентов $\hat{\beta}(q)$ могут использоваться различные подходы. Однако в данной модели распределение регрессионных остатков остается неопределенным, поэтому в основе метода оценивания лежит минимизация суммы взвешенных модулей регрессионных остатков:

$$\theta = \sum_{i=1}^n \rho_q(y_i - x_i^T \beta) \rightarrow \min, \quad (6)$$

где ρ_q понимается функция потерь $\rho_q(u) = u[q - I(u)]$,

$$I(u) = \begin{cases} 1, & \text{если } u < 0 \\ 0, & \text{если } u > 0 \end{cases} - \text{индикаторная функция.}$$

В результате задача (6) сводится к решению следующей задачи [2]:

$$\sum_{i: y_i \geq x_i^T \beta(q)} q \cdot |y_i - x_i^T \beta(q)| + \sum_{i: y_i < x_i^T \beta(q)} (1-q) \cdot |y_i - x_i^T \beta(q)| \rightarrow \min_{\beta}. \quad (7)$$

В литературе [2, 3] задачу (4) и (5) принято представлять как задачу линейного программирования, что имеет несколько важных следствий. Во-

первых, гарантируется, что оценка уравнения квантильной регрессии будет получена за конечное число итераций. Во-вторых, в отличие от классического случая, оценка вектора параметров устойчива к выбросам.

1.2 Вопросы, выносимые на семинарские занятия по теме «Непараметрические методы оценивания эконометрических моделей»

1 Непараметрические методы в эконометрическом исследовании: особенности применения в эконометрике.

2 Квантильная регрессия: постановка задачи, общий вид модели.

3 Что понимается под условной квантилью?

4 Квантильная регрессия: подходы к оценке коэффициентов квантильной регрессии.

5 Сформулируйте и реализуйте задачу линейного программирования для оценки коэффициентов квантильной регрессии.

6 Квантильная регрессия: статистическое исследование модели квантильной регрессии.

7 Примеры применения квантильной регрессии в моделировании социально-экономических процессах (уровень доходов, заработной платы и т.д.).

8 Опишите инструментальные средства для реализации непараметрических методов.

2 Описание лабораторной работы

Цель работы заключается в выработке навыков построения и исследования квантильной модели регрессии в программе RStudio.

Лабораторная работа включает в себя следующие этапы:

- постановку задачи;
- ознакомление с порядком выполнения работы;

- выполнение расчетов индивидуального задания на компьютере и анализ результатов;
- подготовку письменного отчета с выводами по работе;
- защиту лабораторной работы.

3 Постановка задачи

Предметную область исследования студент выбирает самостоятельно. В Приложении А представлен перечень ресурсов, представляющих открытый доступ к данным. Требуется провести исследование взаимосвязи между рассматриваемыми показателями:

- 1) оценить уравнение регрессии, описывающее взаимосвязь между результативным признаком и отобранными факторными показателями методом наименьших квадратов, провести её исследование;
- 2) построить квантильную регрессию, описывающую взаимосвязь между результативным признаком и отобранными факторными показателями, провести её исследование;
- 3) сравнить и проинтерпретировать полученные результаты.

4 Порядок выполнения лабораторной работы

На примере моделирования зависимости между расходами населения на жилье (*Rashod*), доходами (*Doxod*) и стоимостью на жилье (*Cena*) опишем принцип работы и основные этапы эконометрического моделирования в программе RStudio. Исходные данные представлены в Приложении В.

Итак, рассмотрим зависимость между рассмотренными признаками в виде:

$$\tilde{Ras} = \beta_0 + \beta_1 Doxod + \beta_2 Cena$$

На первоначальном этапе опишем основные этапы работы в программе RStudio.

4.1 Описание основного интерфейса RStudio

RStudio – свободная среда разработки программного обеспечения с открытым исходным кодом для языка программирования R, который предназначен для статистической обработки данных и работы с графикой. Данная среда является бесплатной. К основным достоинствам можно отнести большое количество бесплатных библиотек, содержащих готовые процедуры и функции для математической, эконометрической, статистической и графической обработки данных [5]. Среда получила популярность в области биоинформатики, хемометрики, эконометрики и других. Порядок установки RStudio состоит из двух этапов.

1. Сначала необходимо скачать и установить R консоль (<https://www.r-project.org/>).
2. Затем установить RStudio (<https://www.rstudio.com/products/rstudio/>).

RStudio имеет подробное встроенное англоязычное руководство, которое также доступно на официальном сайте <https://cran.r-project.org/>.

Для начала работы необходимо запустить RStudio в стартовом меню. На экране появится окно интерфейса RStudio, представленного на рисунке 1.

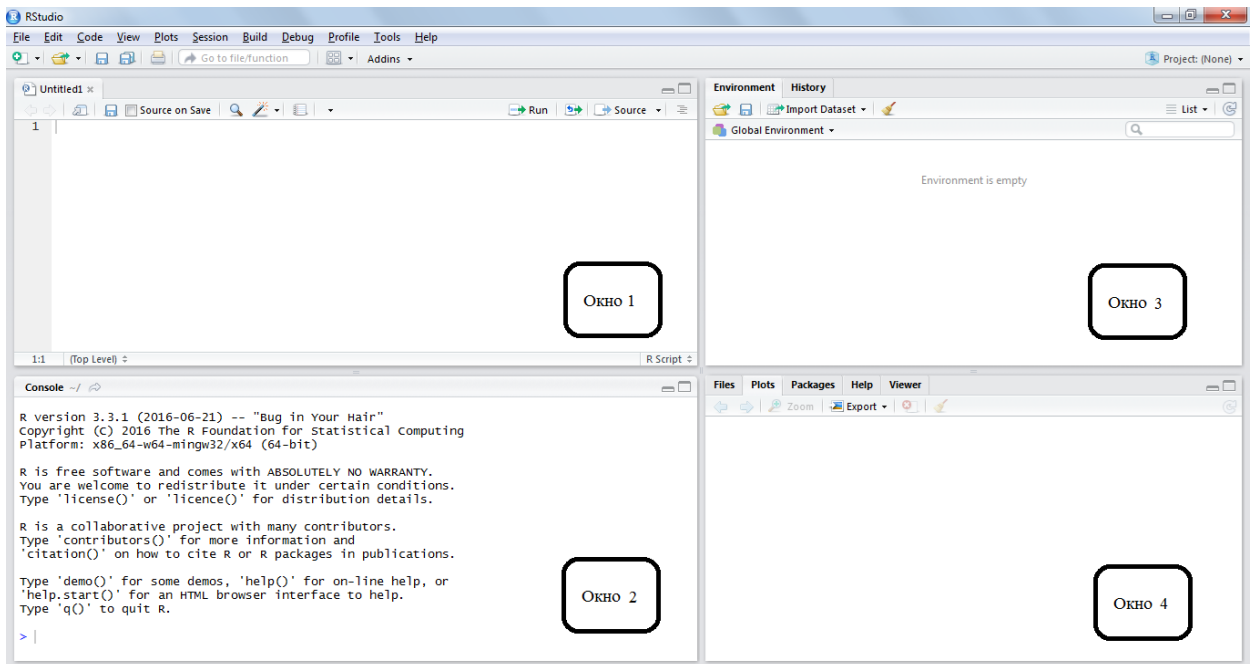


Рисунок 1- Пример интерфейса RStudio в операционной системе Windows

Окно 1 служит для создания и сохранения основного текста программы, которое будет доступно после сохранения. Для выполнения программного кода необходимо нажать на клавишу *Run*, либо поставить курсив на конец строки и нажать комбинацию клавиш *CTRL+ENTER*.

Окно 2 является консольным, то есть служит для выполнения промежуточных, зачастую проверочных действий, не вносящих изменений в основной код программы. Для выполнения какого-либо действия, необходимо набрать команду и нажать *ENTER*.

Окно 3 служит для отображения текущих значений переменных программного кода, а также исторических действий с ними.

Окно 4 является мультиокном со множеством вкладок. Наибольший интерес в рамках данной лабораторной работы представляют вкладки *Help*, *Plot*, служащие для отображения подсказок и справочных материалов, построенных графиков.

Текущая рабочая сессия в RStudio, сохраняемая в памяти компьютера и содержащая любые созданные пользователем объекты, называется рабочим

пространством. Для того чтобы указать рабочее пространство существуют два способа: прописать в окне 2 команду вручную (рисунок 2) или выбрать в меню *Session->Set Working Directory->Choose Directory*.

В конце каждой сессии можно сохранять рабочее пространство, и оно автоматически загрузится при следующем запуске программы. Используя функцию `getwd()`, можно узнать какая директория в данный момент является рабочей.

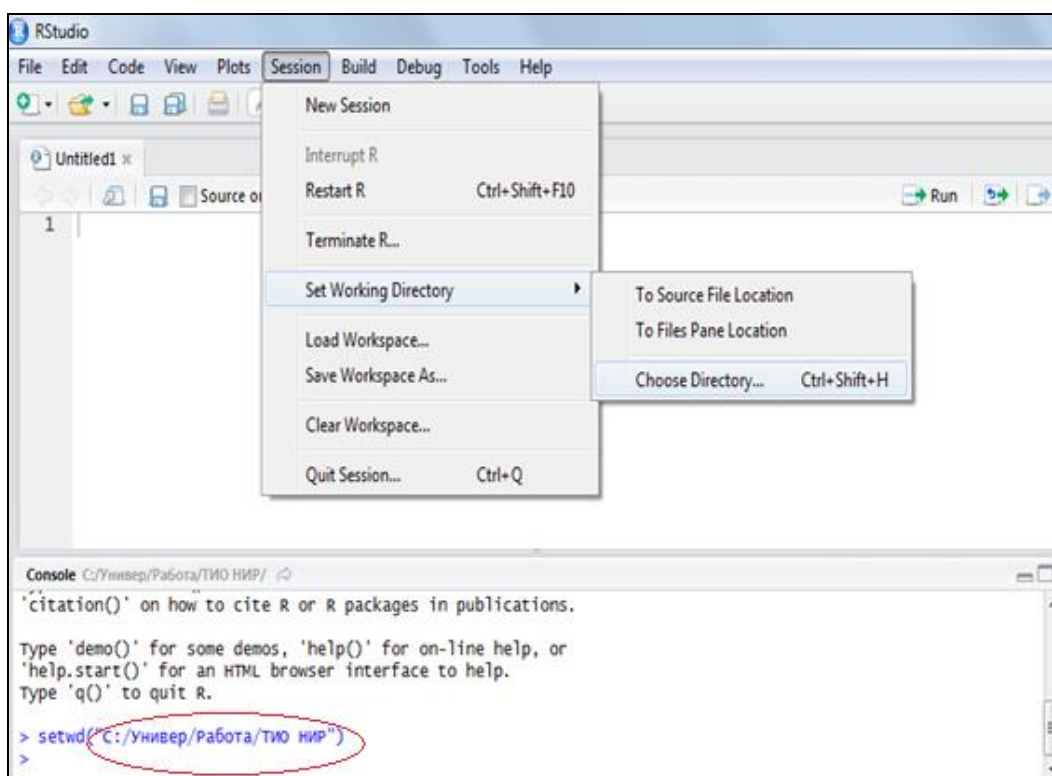


Рисунок 2 – Рабочее пространство в программе «RStudio»

Если появляется необходимость обратиться к файлу, который находится не в рабочей директории, нужно написать полный путь к нему, например “C:/Users/1/Desktop/data.csv”.

В таблице 1 представлены основные команды для управления рабочим пространством.

Таблица 1 – Основные функции, использующиеся для управления рабочим пространством в программе «RStudio»

Функция	Описание
<code>getwd()</code>	Вывести на экран название текущей рабочей директории
<code>setwd("моя_директория")</code>	Назначить <i>моя_директория</i> текущей рабочей директорией
<code>ls()</code>	Вывести на экран список объектов в текущем рабочем пространстве
<code>rm("список_объектов")</code>	Удалить один или несколько объектов
<code>help(options)</code>	Справка о возможных опциях
<code>history(#)</code>	Вывести на экран последние # команд (по умолчанию)
<code>save.image("мой_файл")</code>	Сохранить рабочее пространство в файл <i>мой_файл</i> (по умолчанию .Rdata)
<code>q()</code>	Выйти из программы

4.1 Ввод и первичная обработка данных в программе RStudio

Под набором данных понимают прямоугольный массив данных, в которых ряды соответствуют наблюдениям, а столбцы – признакам [5, 6]:

- `c()` - создание вектора;

`matrix(вектор, nrow = число_строк, ncol = число_столбцов,`

`byrow = логическое_значение, dim.names = list(текст_вектор_названия_строк, текст_вектор_названия_столбцов))` - создание

матриц;

- `data.frame(col1,col2,col3,...)` - создание таблицы данных. Название каждому столбцу можно присвоить при помощи функции `names()`.

Проиллюстрируем создание таблицы данных о расходах на жилье, дохода и ценах при помощи примера программного кода, который должен быть записан в окне 1 (рисунок 3).

```

Rashod <- c(23569, 38696, 47698, 75404, 119833, 180088, 255219, 352235, 572315, 834006, 1194732, 838826, 1104545, 14
Doxod <- c(1776032, 2908069, 3814655, 5325795, 6831036, 8900529, 10976249, 13818065, 17190065, 21311452, 25244047,
Time <- c(1998:2012)
Cena <- c(5050, 6999, 8678, 10567, 12939, 16320, 20810, 25394, 36221, 47482, 52504, 47715, 48144, 43686, 48163)
Data <- as.data.frame(cbind(Rashod, Doxod, Time, Cena))
# read data

```

Data			
Data		15 obs. of 4 variab...	
df		15 obs. of 4 variab...	
Values			
Cena	num [1:15]	5050 6999 ..	
Doxod	num [1:15]	1776032 29...	
fn		"data.csv"	
Rashod	num [1:15]	23569 3869...	
Time	int [1:15]	1998 1999 ..	

Рисунок 3 – Программный код «Создание таблицы данных».

Исходные данные также могут напрямую считываться из файла, например Excel. Создание программного кода «Импорт данных» осуществляется в окне 1. Пример записи программного кода представлен на рисунке 4. Данный пример позволяет корректно считывать исходные данные в виде матрицы объект-свойство, а также автоматическими названиями признаков и определением типа переменных, подсчетом объектов для анализа.

```

6 # read data
7 fn <- "data.csv"
8 df <- data.frame(read.csv(file=fn,
9                       head=TRUE,
10                      sep=";",
11                      dec = ".",
12                      stringsAsFactors=FALSE,
13                      na.strings = "NA"))
14

```

Рисунок 4 – Программный код «Импорт данных».

На рисунке 4 в 6 строке показан пример создания строки-комментария, для этого необходимо поставить знак «#». Если необходимо закомментировать несколько строчек кода, то следуют выделить текст и нажать комбинацию клавиш *CTRL+SHIFT+C*.

В 7 строке кода создана текстовая переменная *fn* (file name), которой присваивается имя и расширения файла для считывания информации.

В 8 строке кода создается переменная `df` (data frame) в которую посредством функций `as.data.frame()` и `read.csv()` считываются исходные данные.

Функция `as.data.frame()` означает, что входную информацию следует понимать как матрицу объект-свойство, в чем можно убедиться, посмотрев на отображение переменной `df` в окне 3.

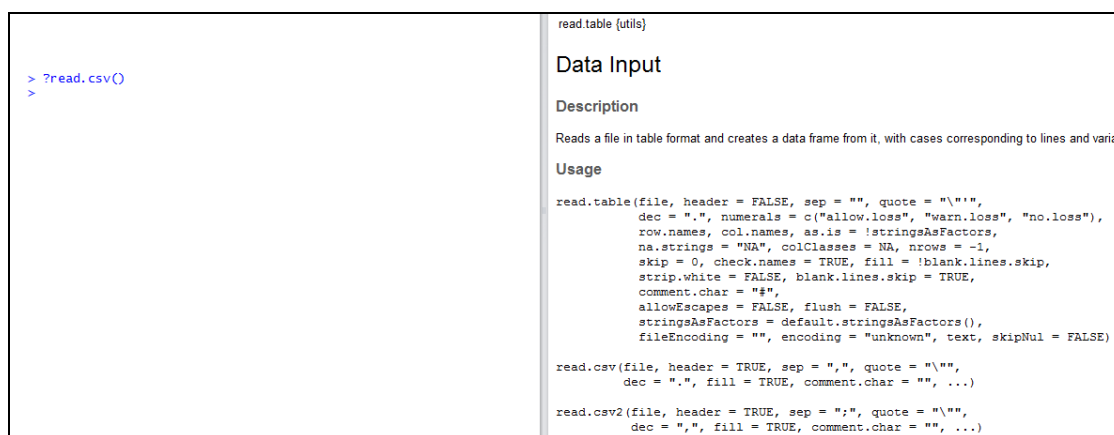
Функция `read.csv()` имеет несколько входных параметров:

9 строка кода – передача имени файла; считывать данные, учитывая, что в первой строке содержатся названия столбцов;

10 строка кода – объекты разделены между собой «;»;

11 строка кода – целая и десятичная части числа разделены «!»

Если на каком-либо этапе возникают трудности с атрибутами функций, то необходимо воспользоваться встроенным подсказчиком. Для этого выполнить команду: «?*имя_функции*» или поставить курсор на функцию и нажать F1. После этого в нижнем правом поле появится полное описание функции. Пример подсказки функции `read.csv()` приведен на рисунке 5. Корректнее и удобнее, если подсказчик вызывать из консольного окна 2.



```
read.table {utils}

Data Input
Description
Reads a file in table format and creates a data frame from it, with cases corresponding to lines and variab
Usage
read.table(file, header = FALSE, sep = ",", quote = "\"",
  dec = ".", numerals = c("allow.loss", "warn.loss", "no.loss"),
  row.names, col.names, as.is = !stringsAsFactors,
  na.strings = "NA", colClasses = NA, nrows = -1,
  skip = 0, check.names = TRUE, fill = !blank.lines.skip,
  strip.white = FALSE, blank.lines.skip = TRUE,
  comment.char = "#",
  allowEscapes = FALSE, flush = FALSE,
  stringsAsFactors = default.stringsAsFactors(),
  fileEncoding = "", encoding = "unknown", text, skipNul = FALSE)

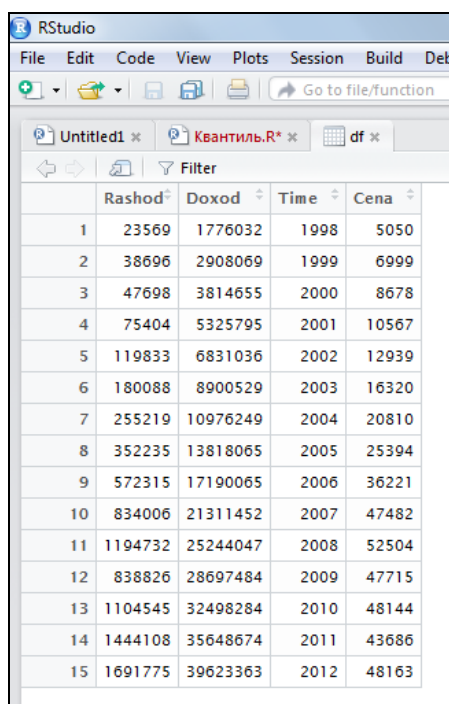
read.csv(file, header = TRUE, sep = ",", quote = "\"",
  dec = ".", fill = TRUE, comment.char = "#", ...)

read.csv2(file, header = TRUE, sep = ":", quote = "\"",
  dec = ":", fill = TRUE, comment.char = "#", ...)
```

Рисунок 5 – Вызов описания функции `read.csv()`

Для того, чтобы посмотреть исходные данные можно ввести команду `View(df)` в окне 2 или нажать на клавишу просмотр напротив переменной в окне 3. Таблица с исходными статистическими данными о расходах

населения на жилье (Rashod), стоимостью жилья (Cena) и доходами населения (Doход) представлена на рисунке 6.



	Rashod	Doход	Time	Cena
1	23569	1776032	1998	5050
2	38696	2908069	1999	6999
3	47698	3814655	2000	8678
4	75404	5325795	2001	10567
5	119833	6831036	2002	12939
6	180088	8900529	2003	16320
7	255219	10976249	2004	20810
8	352235	13818065	2005	25394
9	572315	17190065	2006	36221
10	834006	21311452	2007	47482
11	1194732	25244047	2008	52504
12	838826	28697484	2009	47715
13	1104545	32498284	2010	48144
14	1444108	35648674	2011	43686
15	1691775	39623363	2012	48163

Рисунок 6 – Вид импортированных исходных статистических данных.

В «RStudio» существует множество встроенных процедур для первичной обработки данных. Так например, для выявления аномальных наблюдений используется функция *boxplot()*:

```
boxplot(df$Rashod, main = "Расход")  
boxplot(df$Doход, main = "Доход")  
boxplot(df$Cena, main = "Цена")
```

Результаты построения «ящика с усами» для признака Rashod представлен на рисунке 7.

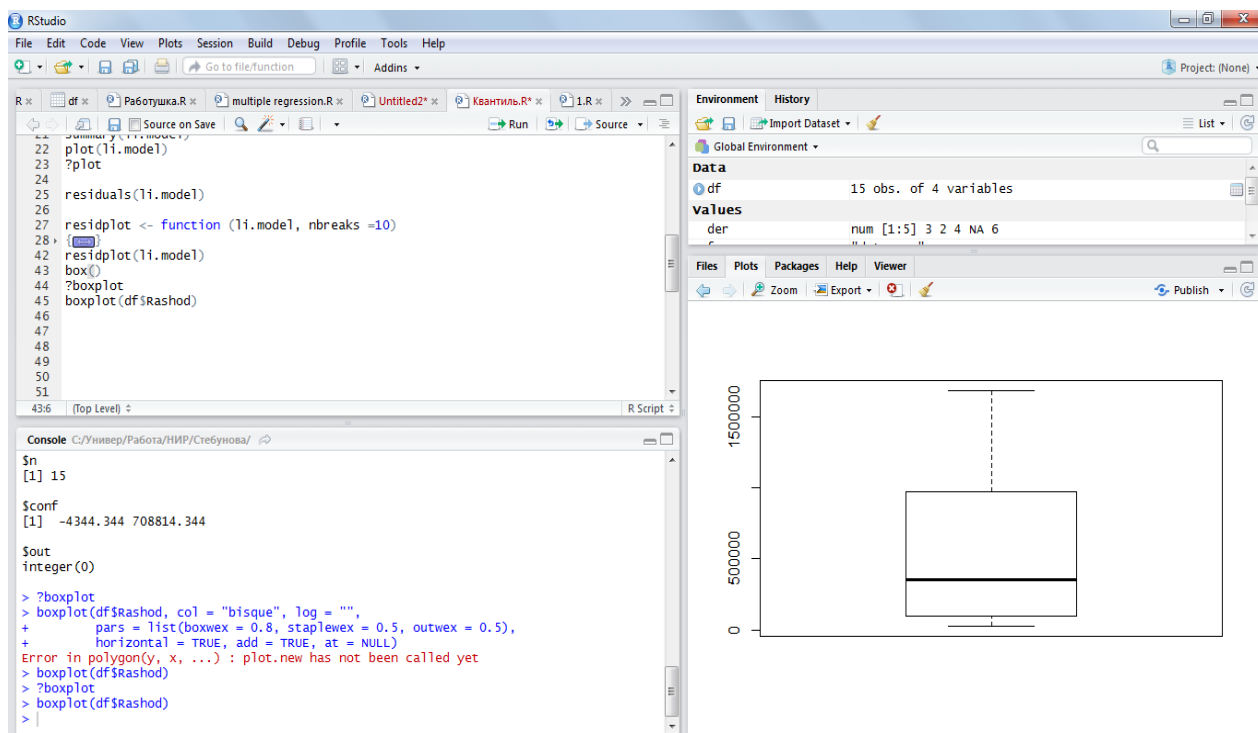


Рисунок 7 – Результаты построения графика «Ящик с усами» для признака *Rashod*.

Из рисунка 7 видно, что аномальных наблюдений не наблюдается, в противном случае были бы отмечены наблюдения, выходящие за внешние границы «ящика с усами» [1]. Однако, также можно заметить, что центр тяжести рассмотренного показателя смещен к нижней границе.

Для построения диаграмм различного типа, используется следующий набор функций:

- *plot* (имя переменной, имя переменной) - открывает окно графики и создает диаграмму рассеяния;
- *abline*() - добавляет регрессионную прямую;
- *title* ("название диаграммы") - добавляет название диаграммы;
- *detach* ("название таблицы данных") - удаляет таблицу данных из пути поиска.

Так, например, программный код для создания диаграммы рассеяния показателей *Dohod* и *Rashod* представлен на рисунке 8. Данный код необходимо записать в окне 2 (консоли) рабочей директории.

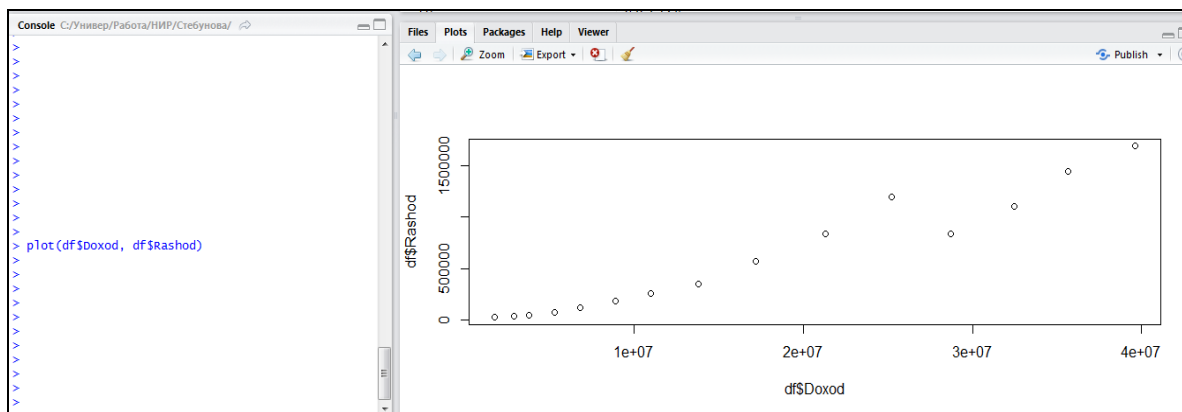


Рисунок 8 – Команда для создания диаграммы рассеивания признаков

Для сохранения диаграммы в формате PDF при помощи кода используется функция `pdf("название.pdf")`. Способ сохранения диаграмм при помощи графического интерфейса различается в зависимости от операционной системы. Под Windows при активированном графическом устройстве нужно выбрать в меню *File (Файл) → Save as (Сохранить как)*, а затем в появившемся диалоге выбрать нужный формат графического файла и директорию для сохранения.

Закончим данный параграф описанием основных функций для расчета выборочных числовых характеристик (таблица 2).

Таблица 2 – Статистические функции для вычисления числовых характеристик

Функция	Описание
<code>mean("имя вектора")</code>	Выборочное значение среднего арифметического
<code>median("имя вектора")</code>	Выборочное значение медианы
<code>sd("имя вектора")</code>	Выборочное значение среднего квадратического отклонения
<code>var("имя вектора")</code>	Выборочное значение дисперсии
<code>quantiel("имя вектора", probs)</code>	Выборочное значение квантили, <i>probs</i> - числовой вектор с указанием вероятностей
<code>range("имя вектора")</code>	Выборочное значение размаха вариации
<code>min("имя вектора")</code>	Минимальное значение
<code>cor()</code>	Оценка корреляции

Программный код для вычисления оценок выборочных характеристик для признака «Rashod» и запись их в новую переменную типа list представлен на рисунке 9.

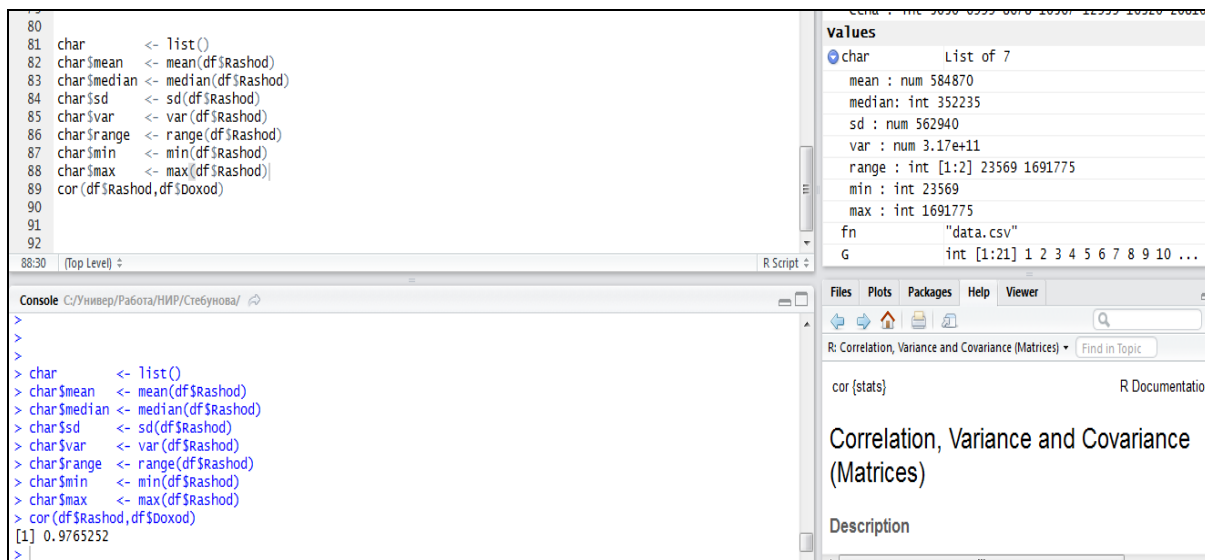


Рисунок 9 – Программный код «Описательная статистика»

Как видно из рисунка 9, выборочная средняя для признака «Rashod» составила 584870 тыс.руб., оценка среднего квадратического отклонения 562940 тыс.руб и т.д.

Кроме того, в базовой версии реализована функция `summary()`, которая позволяет вычислить описательные статистики (рисунок 10).

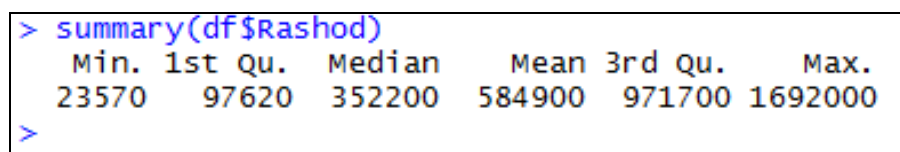


Рисунок 10 – Вычисление описательных статистик при помощи команды `summary()`

Также функции для вычисления описательных статистик реализованы в дополнительных пакетах - *Hmisc*, *pastecs* и *psych*.

Для построения функции (плотности) распределения, а также генерации случайных чисел используется следующий набор функций:

[*dpqr*]сокращенное_название_распределения()

d - плотность распределения вероятностей;

p - функция распределения;

q – функция, определяющая квантили;

r – генератор случайных чисел.

4.2 Построение и исследование линейной модели регрессии в программе «RStudio»

Для оценивания регрессионных моделей в «RStudio» используется функция:

$$myfit <- lm(formula, data),$$

где *formula* описывает вид модели, которую используют для аппроксимации функции регрессии;

data - таблица с исходными данными

В таблице 3 представлены основные функции для построения линейной регрессионной модели.

Таблица 3 – Основные функции для построения регрессионных моделей

Функция	Описание
<i>summary()</i>	Выводит детальную информацию о построенной информации
<i>coefficients()</i>	Перечисляет оценки коэффициентов регрессии
<i>confit()</i>	Вычисляет доверительные интервалы для коэффициентов регрессии
<i>fitted()</i>	Выводит модельные значения результативного признака
<i>residuals()</i>	Выводит значения регрессионных остатков
<i>anova()</i>	Создает таблицу дисперсионного анализа для построенной модели
<i>vcov()</i>	Выводит оценку ковариационной матрицы для коэффициентов модели
<i>AIC()</i>	Вычисляет информационный критерий Акаике

<code>plot()</code>	Создает диаграмму рассеяния по исходным данным
<code>predict()</code>	Использует построенную модель для прогнозирования результативного признака для нового набора данных

Программный код для построения линейной модели, описывающей зависимость между расходами, доходами и ценами на жилье ($\tilde{Ras} = \beta_0 + \beta_1 Doxod + \beta_2 Cena$) представлен на рисунке 11.

```

15 # линейная модель
16 li.model <- lm(df$Rashod ~ df$Doxod + df$Cena, df)
17 summary(li.model)
18 plot(li.model)

```

Рисунок 11 – Программный код для оценивания линейной модели регрессии

Для отображения результатов оценки линейной модели регрессии, следует воспользоваться функцией `summary()`, которую следует записать в окне 1. Результаты представлены на рисунке 12.

```

> summary(li.model)

Call:
lm(formula = df$Rashod ~ df$Doxod + df$Cena, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-255736  -63071   -4199    64888   237637

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.542e+05  6.637e+04  -2.324  0.0385 *
df$Doxod     4.149e-02  7.238e-03   5.733 9.42e-05 ***
df$Cena      1.217e+00  5.120e+00   0.238  0.8162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 130700 on 12 degrees of freedom
Multiple R-squared:  0.9538,    Adjusted R-squared:  0.9461
F-statistic: 123.9 on 2 and 12 DF,  p-value: 9.7e-09

```

Рисунок 12 – Результаты оценивания линейной модели регрессии, описывающей зависимость между расходами на жилье и доходом населения, ценами на жилье.

Таким образом, оценка линейной функции регрессии имеет вид:

$$Ra\hat{s}hod = -154237 + 0,0415 Doxod - 1,217Cena$$

(663706)
(0,0072)
(0,12)

Как известно, следующим этапом эконометрического моделирования является проверка гипотезы о нормальном характере распределения регрессионных остатков.

Программный код для построения гистограммы остатков регрессионной модели и проверки на нормальный закон распределения представлен на рисунке 13.

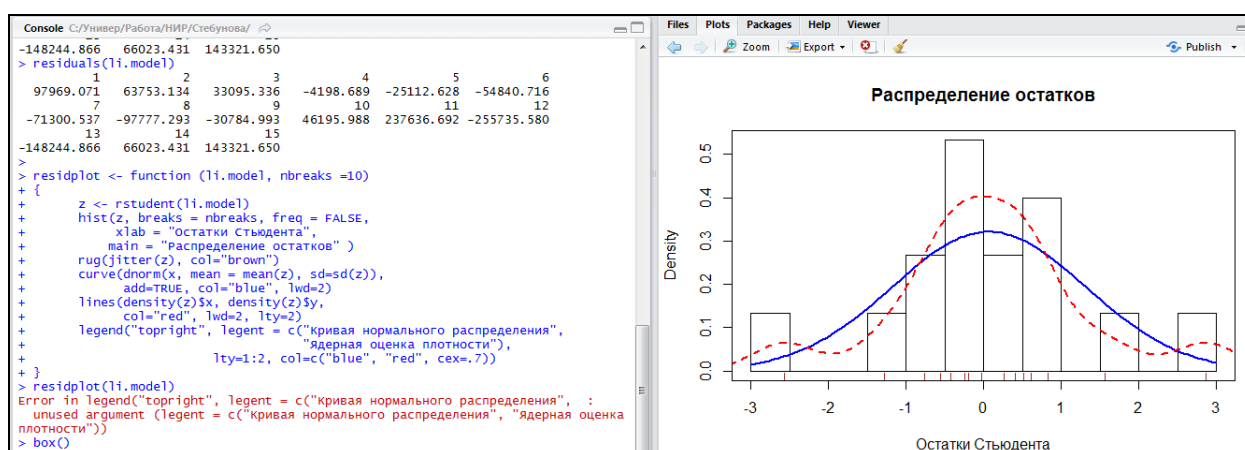


Рисунок 13 – Результаты исследования остатков регрессионной модели

Для проверки гипотезы о нормальном характере распределения регрессионных остатков, используя критерий Пирсона, необходимо предварительно установить (`install.packages("nortest")`) и подключить (`library(nortest)`) пакет *nortest*, а затем вызвать функцию `pearson.test()`. В данной библиотеке содержатся и другие критерия (тесты).

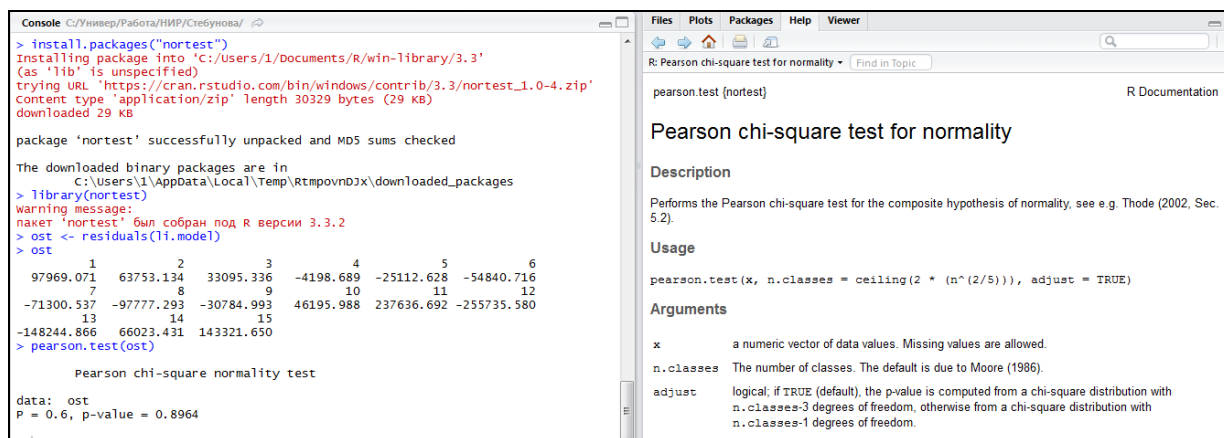


Рисунок 14 – Результаты проверки гипотезы о нормальном законе распределения регрессионных остатков

Как видно из рисунка, распределение регрессионных остатков соответствует нормальному закону распределения, $p\text{-value} = 0,8964$. Следовательно, можно сделать вывод о значимости модели в целом (рисунок 12). Однако, значимое влияние на уровень расходов на жилье оказывает только ценовой признак (коэффициенты при признаке *Doxod* не значим).

В RStudio существует возможность совместной проверки выполнения всех требований, предъявляемых к КЛММР. Для этого рекомендуется воспользоваться встроенной функцией *gvlna()* из одноименного пакета. В качестве параметра в функцию подается результат оценивания линейной модели регрессии, который записан у нас в *li.model*. Программная реализация представлена на рисунке 15.

```

> gvmodal <- gvlma(li.model)
> summary(gvmodal)

Call:
lm(formula = df$Rashod ~ df$Doxod + df$Cena, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-255736 -63071  -4199   64888  237637

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.542e+05  6.637e+04  -2.324  0.0385 *
df$Doxod     4.149e-02  7.238e-03   5.733 9.42e-05 ***
df$Cena      1.217e+00  5.120e+00   0.238  0.8162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 130700 on 12 degrees of freedom
Multiple R-squared:  0.9538,    Adjusted R-squared:  0.9461
F-statistic: 123.9 on 2 and 12 DF,  p-value: 9.7e-09

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance = 0.05

Call:
gvlma(x = li.model)

              value p-value              Decision
Global stat    9.655575 0.04665 Assumptions NOT satisfied!
Skewness       0.053556 0.81699  Assumptions acceptable.
Kurtosis       0.004456 0.94678  Assumptions acceptable.
Link Function  6.083329 0.01365 Assumptions NOT satisfied!
Heteroscedasticity 3.514233 0.06084  Assumptions acceptable.
> |

```

Рисунок 15 – Результаты проверки требований КЛММР

В результате реализации теста *Global Stat* следуют, что модель не удовлетворяет требованиям КЛММР ($p\text{-value} = 0.047$). Далее рассмотрена процедура проверки регрессионных остатков на отсутствие гетероскедастичности и автокорреляции.

Для проверки остатков на автокорреляцию и гетероскедастичность рекомендуется установить пакет «*car*» и воспользоваться встроенными функциями *durbinWatsonTest()* и *ncvTest()*. Программная реализация проверки гипотезы о отсутствии автокорреляции ($H_0: \rho = 0$) с помощью критерия Дарбина-Уотсона представлена на рисунке 16.

```

138 # первичная установка (только во время первого использования)
139 install.packages("car")
140
141 # подключение пакета
142 library(car)
143
144 # оценивание линейной модели
145 li.model <- lm(df$Rashod ~ df$Doxod + df$Cena, df)
146
147 # реализация критерия Дарбина-Уотсона
148 durbinwatsonTest(li.model)

```

147:38 (Top Level) ⚡

Console C:/Универ/Работа/НИР/Стебунова/ ↻

```

> # первичная установка (только во время первого использования)
> install.packages("car")
Error in install.packages : updating loaded packages
>
> # подключение пакета
> library(car)
>
> # оценивание линейной модели
> li.model <- lm(df$Rashod ~ df$Doxod + df$Cena, df)
>
> # реализация критерия Дарбина-Уотсона
> durbinwatsonTest(li.model)
lag Autocorrelation D-w Statistic p-value
1      0.04861808      1.755664  0.258
Alternative hypothesis: rho != 0

```

Рисунок 16 – Программный код реализации критерия Дарбина-Уотсона

На рисунке 16 представлена оценка коэффициента автокорреляции первого порядка ($\hat{\rho} = 0,048$), значение критерия Дарбина-Уотсона ($DW=1,756$, $p\text{-value} = 0.258$), что позволяет сделать вывод об отсутствии автокорреляции.

На рисунке 17 представлен пример проверки гипотезы об отсутствии гетероскедастичности с помощью теста Уайта.

```

150 # Проверка на гомоскедастичность
151 ncvTest(li.model)
152

```

136:1 (Top Level) ⚡

Console C:/Универ/Работа/НИР/Стебунова/ ↻

```

> ncvTest(li.model)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 3.988573    Df = 1    p = 0.04580984
> |

```

Рисунок 17 – Программный код проверки гипотезы об отсутствии гетероскедастичности

Как видно, нулевая гипотеза отвергается, следовательно делаем вывод о наличии гетероскедастичности, что связано с неоднородность выборочных данных. Так как при формировании выборки не учитывалось различие в доходах наимее и наиболее обеспеченного населения.

В RStudio существует возможность реализации пошагового алгоритма реализации МНК с помощью пакета MASS. Код программы и пример реализации представлен на рисунке 18. Слева первоначальная модель регрессии по всем параметрам, справа, процедура удаление наименее значимых параметров.

```

> wateruse.lm<-lm(df$Rashod~ df$Doxod + df$Cena,data=df) # Fit full model
> summary(wateruse.lm)
Call:
lm(formula = df$Rashod ~ df$Doxod + df$Cena, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-255736  -63071  -4199   64888  237637

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.542e+05  6.637e+04  -2.324  0.0385 *
df$Doxod     4.149e-02  7.238e-03  5.733  9.42e-05 ***
df$Cena      1.217e+00  5.120e+00  0.238  0.8162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 130700 on 12 degrees of freedom
Multiple R-squared:  0.9538,    Adjusted R-squared:  0.9461
F-statistic: 123.9 on 2 and 12 DF,  p-value: 9.7e-09

> wateruse.lm.back<-stepAIC(wateruse.lm,trace=FALSE)
> summary(wateruse.lm.back)
Call:
lm(formula = df$Rashod ~ df$Doxod, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-251284  -64219  -7735   60925  253414

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.463e+05  5.529e+04  -2.647  0.0201 *
df$Doxod     4.308e-02  2.636e-03  16.346  4.76e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 125800 on 13 degrees of freedom
Multiple R-squared:  0.9536,    Adjusted R-squared:  0.95
F-statistic: 267.2 on 1 and 13 DF,  p-value: 4.762e-10

```

Рисунок 18 – Программный код реализации пошаговой регрессии с исключением переменных

Если необходимо осуществить пошаговую регрессию с включением переменных, то необходимо первоначально оценить модель с минимальным количеством параметров, а затем с помощью настроек функции *stepAIC(...,direction="forward")* добавить направление «назад».

4.3 Построение модели квантильной регрессии

Естественно, что для влияния доходов домохозяйств на их расходы на покупку жилья происходит не только «в среднем», но различается для наименее и наиболее обеспеченных, поэтому целесообразнее использовать квантильную регрессию вида [2]:

$$Quant_q(Rashod_i / Dohod_i, Cena_i) \equiv \tilde{y}_i(q) = \beta_0(q) + \beta_1(q)Dohod_i + \beta_2Cena_i + z_i(q)$$

Для реализации квантильной регрессии используется пакет «*quantreg*» и встроенная функция *rq()*. Параметр *tau* обычно задается квантиль в интервале от 0 до 1, однако в RStudio представляется возможным также задать параметр *tau* как вектор значений в интервале от 0 до 1, таким образом, существует возможность с помощью одной команды оценивать нескольких регрессий с разным уровнем квантили. Для представления такой последовательности используется функция *seq(min, max, by)*, в качестве параметров которой задаются границы отрезков и шаг разбиения. На рисунке 19 представлена реализация программного кода для построения квантильных регрессий.

```
library(quantreg)
q25 <- rq(df$Rashod~df$Dohod + df$Cena|,data=df, tau = 0.25)
summary(q25)

q50 <- rq(df$Rashod~df$Dohod + df$Cena,data=df, tau = 0.50)
summary(q50)

q75 <- rq(df$Rashod~df$Dohod + df$Cena,data=df, tau = 0.75)
summary(q75)

q2575 <- rq(df$Rashod~df$Dohod + df$Cena,data=df, tau = c(0.25,0.75))
summary(q2575)

qA11 <- rq(df$Rashod~df$Dohod + df$Cena,data=df, tau = seq(0.05, 0.95, by = 0.05))
summary(qA11)
plot(qA11)
```

Рисунок 19 – Программный код для построения модели квантильной регрессии для разных уровней квантилей

Для вывода результата оценивания используется команда `summary()`. На рисунке 20 представлены результаты построения квантильной регрессии для уровня 0.05.

```
> q05 <- rq(df$Rashod~df$Doxod + df$Cena,data=df, tau = 0.05)
> summary(q05)

Call: rq(formula = df$Rashod ~ df$Doxod + df$Cena, tau = 0.05, data = df)

tau: [1] 0.05

Coefficients:
              coefficients  lower bd      upper bd
(Intercept) -6.128853e+04 -1.797693e+308  1.797693e+308
df$Doxod      7.268000e-02 -1.797693e+308  1.797693e+308
df$Cena      -2.484518e+01 -1.797693e+308  1.797693e+308
```

Рисунок 20 – Программный код и результаты построения квантильной регрессии на уровне 0,05.

Для того чтобы проанализировать значения коэффициентов различных уровней квантилей необходимо оценить регрессию, в которой параметр *tau* задан вектором значений, как показано на рисунке 19, функция (*qAll*). График значений коэффициентов каждой переменной для различного уровня квантилей представлен на рисунке 21.

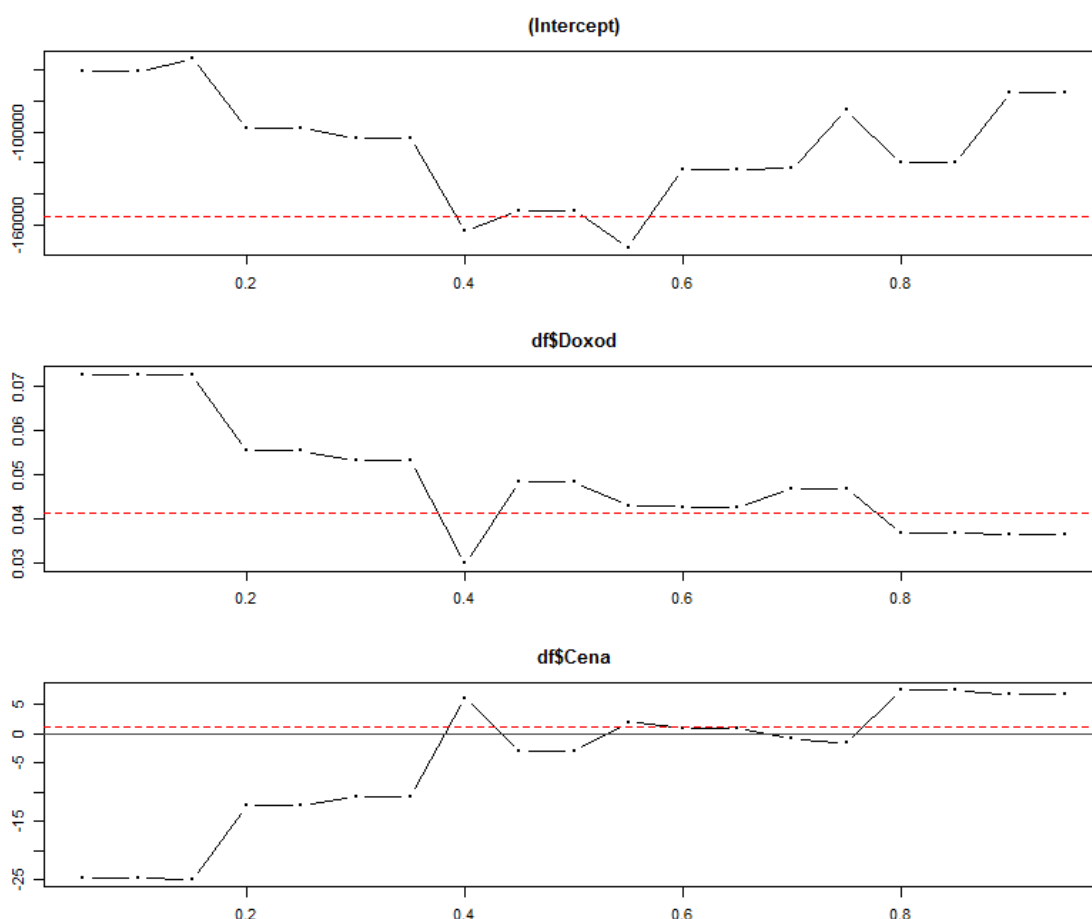


Рисунок 21 – Значения коэффициентов переменных при различных уровнях квантилей

Представим результаты оценивания эконометрических моделей для разных уровней квантилей (таблица 4), и сравним полученные результаты.

Таким образом, анализ расходов на покупку жилой недвижимости с помощью квантильной регрессии позволяет увидеть чувствительность расходов от того, к какому уровню дохода принадлежит домохозяйство. В результате моделирования выяснено, что эластичность спроса на покупку жилья по мере увеличения доходов снижается, а влияние средних цен уменьшается с ростом дохода. Рассмотренный пример показывает, что квантильная регрессия представляет собой новый аппарат построения оценок коэффициентов эконометрической модели с учетом предпочтений исследователя относительно ее положения по отношению к наблюдаемым значениям.

Таблица 4 – Результаты оценки эконометрической модели расходов на покупку жилья от располагаемых доходов и средних цен 1 кв.м. жилой недвижимости

Метод оценки	Оценка модели
МНК	$Ra\hat{s}hod = -154237 + 0,0415 Doxod - 1,217 Cena$ (663706) (0,0072) (0,12)
q=0,05	$Ra\hat{s}hod(0,05) = -61288,5 + 0,073 Doxod - 24,85 Cena$ (101476) (0,0010) (0,007)
q=0,25	$Ra\hat{s}hod(0,25) = -97919,1 + 0,053 Doxod - 12,42 Cena$ (115639) (0,0013) (0,892)
q=0,5	$Ra\hat{s}hod(0,5) = -150542,0 + 0,048 Doxod - 3,04 Cena$ (633136) (0,0069) (0,87)
q=0,75	$Rashod(0,75) = -85967,9 + 0,0469 Doxod - 1,68 Cena$ (622089) (0,007) (0,79)
q=0,95	$Ra\hat{s}hod(0,95) = -74690,5 + 0,366 Doxod - 1,59 Cena$ (257799) (0,003) (0,98)

4.4 Описание команд для построения квантильной регрессии в кросс-платформенном пакете GRETЛ

Также квантильная регрессия может быть реализована в программном средстве GRETЛ. Для чего в основном меню программы необходимо выбрать модуль *Модель* и рассмотреть *Робастные методы*, среди которых выбрать квантильную регрессию (рисунок 22).

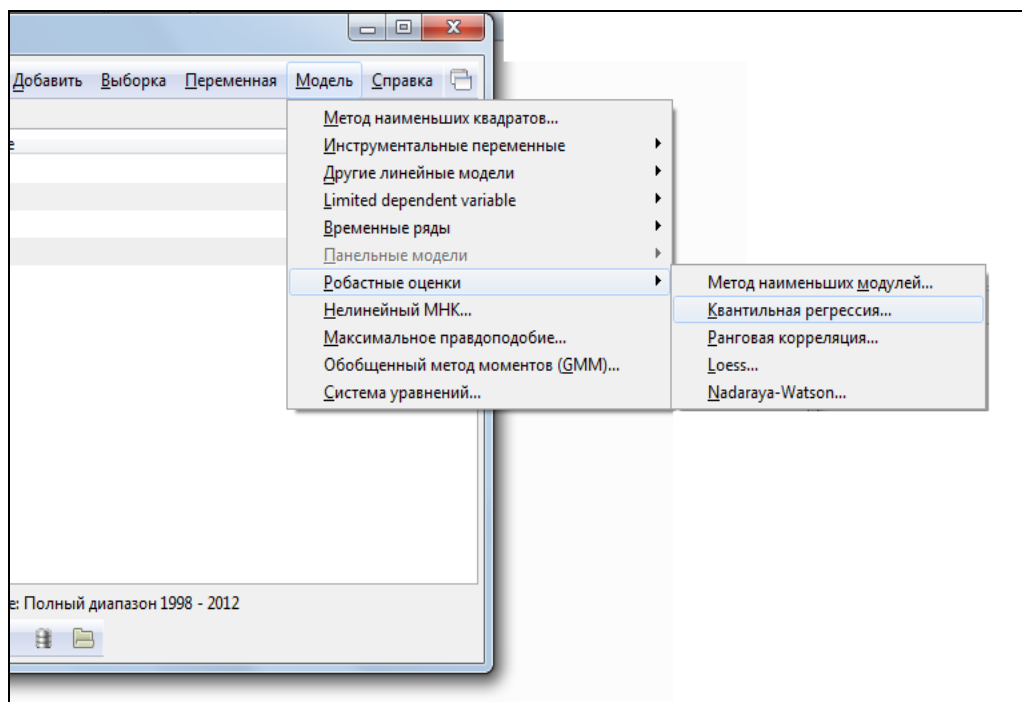


Рисунок 22 – Построение квантильной регрессии в ППП GRETL

На рисунке 23 представлены результаты моделирования для различных уровней квантилей.

The screenshot shows the GRETL software interface displaying the results of a quantile regression model. The window title is 'gretl: модель 1'. The menu bar includes 'Файл', 'Правка', 'Тесты', 'Сохранить', 'Графики', 'Анализ', and 'LaTeX'. The main text reads: 'Модель 1: Квантильная оценка', 'Зависимая переменная: Rashod', and 'Асимптотические стандартные ошибки считаются независимыми и одинаково распределенными'. Below this is a table of results for the variables 'const', 'Doход', and 'Цена' across different quantiles (tau).

	tau	Коэффициент	Ст. ошибка	t-статистика
const	0,050	-61288,5	101,476	-603,969
	0,250	-97919,1	11563,9	-8,46768
	0,500	-150542,	63313,6	-2,37773
	0,750	-85967,9	62208,9	-1,38192
	0,950	-74690,9	25779,9	-2,89726
Doход	0,050	0,0726754	1,10664e-005	6567,23
	0,250	0,0553932	0,00126108	43,9250
	0,500	0,0484597	0,00690459	7,01847
	0,750	0,0469045	0,00678412	6,91387
	0,950	0,0365590	0,00281139	13,0039
Цена	0,050	-24,8452	0,00782892	-3173,51
	0,250	-12,4152	0,892154	-13,9160
	0,500	-3,04154	4,88466	-0,622672
	0,750	-1,67704	4,79943	-0,349424
	0,950	6,59998	1,98892	3,31838
Медиана зав. перемен		352235,0	Ст. откл. зав. перемен	562939,6

Рисунок 23 – Результаты оценивания квантильной регрессии

5 Содержание письменного отчета

Отчет по лабораторной работе оформляется на листах формата А4 и должен иметь следующую структуру:

- 1) титульный лист;
- 2) задание на лабораторную работу;
- 2) краткие теоретические сведения, необходимые для решения поставленных задач;
- 3) постановка задачи и математические модели, применяемые для исследования;
- 4) результаты применения ППП для решения задач;
- 5) анализ полученных результатов и выводы.

6 Вопросы к защите лабораторной работы

1. Охарактеризуйте квантильную регрессию: постановка задачи, общий вид модели.
2. Что такое условная квантиль? И ее применении в эконометрическом моделировании.
3. Опишите подходы к оценке коэффициентов квантильной регрессии.
4. Сформулируйте и реализуйте задачу линейного программирования для оценки коэффициентов квантильной регрессии.
5. Квантильная регрессия: статистическое исследование модели квантильной регрессии.
6. Примеры применения квантильной регрессии в моделировании социально-экономических процессах (уровень доходов, заработной платы и т.д.).

Список использованных источников

1. Стебунова О.И. О робастном оценивании в эконометрическом моделировании /О.И. Стебунова //Университетский комплекс как региональный центр образования, науки и культуры. Материалы Всероссийской научно-методической конференции; Оренбургский гос. ун-т. - Оренбург: ООО ИПК «Университет», 2014. – С.116-120.
2. Koenker R. Regression Quantiles /R. Koenker, G.Bassett Jr.// *Econometrica*. – 1978. - №1. – P. 33-50.
3. Стебунова О.И. К вопросу о применении квантильной регрессии в исследованиях экономических процессов/ О.И. Стебунова//Математические методы и модели в исследовании государственных и корпоративных финансов и финансовых рынков: сборник материалов Всероссийской научно-практической конференции (10-11 декабря 2015г., г.Уфа.) – Уфа: Аэтерна.- 2015.- С.111-114.
4. Abduramanov R., The method of quantile regression, a new approach to actuarial mathematics/R/ Abduramanov, A. Kudryavtsev A. // 11th International Congress “Insurance: Mathematics and Economics” July 10-12, 2007, Piraeus, Greece. Book of Abstracts. P. 56–57.
5. Роберт И. Кабаков R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588с.
6. Мастицкий С.Э. Статистический анализ и визуализация данных с помощью R/ С.Э. Мастицкий, В.К. Шитиков. – М.:ДМК Пресс, 2015. – 496 с.

Приложение А

(справочное)

Список источников статистических данных

При выполнении лабораторной работы студентам рекомендуется использовать статистические данные, соответствующие их учебным и научным интересам. Ниже представлен краткий список ресурсов, предоставляющих открытый доступ к данным.

1. **Федеральная служба государственной статистики РФ** предоставляет статистическую информацию на уровне РФ, федеральных округов и субъектов РФ. – Режим доступа: <http://www.gks.ru>

2. **Инвестиционная компаний «ФИНАМ»** предоставляет информацию о котировках акций, ценах на финансовые инструменты, сырьевые ресурсы и т. д. – Режим доступа: www.finam.ru

3. **Единый архив экономических и социальных данных Высшей Школы Экономики.** – Режим доступа: <http://sophist.hse.ru>

4. **Евростат** предоставляет информацию по широкому спектру показателей, характеризующих население, экономику и финансы, социальные условия большинства европейских стран. – Режим доступа: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes>

5. Перечень открытых источников статистических данных на сайте, поддерживаемым университетом Бристоля. – Режим доступа: http://www.economicsnetwork.ac.uk/links/data_free

6. Анатольев С. Где найти данные в сети?/ С. Анатольев, А. Цыплаков// Квантиль. – 2009. – №6. – С.59-71. – Режим доступа: <http://quantile.ru/06/06-AT.pdf>

Приложение Б (обязательное)

Исходные данные для эконометрического моделирования

Таблица Б.1 – Статистические данные о расходах на жилье, доходах и ценах на жилье

Период (Time)	Расходы на жилье, руб. (Rashod)	Среднедушевые доходы, руб. (Doход)	Средняя цена на жилиую недвижимость, руб. (Cena)
1997	12332,6	1656405	4980
1998	23569	1776032	5050
1999	38696	2908069	6999
2000	47698	3814655	8678
2001	75404	5325795	10567
2002	119833	6831036	12939
2003	180088	8900529	16320
2004	255219	10976249	20810
2005	352235	13818065	25394
2006	572315	17190065	36221
2007	834006	21311452	47482
2008	1194732	25244047	52504
2009	838826	28697484	47715
2010	1104545	32498284	48144
2011	1444108	35648674	43686
2012	1691775	39623363	48163