

МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Маркетинг и менеджмент»

ЭКОНОМЕТРИКА

*Методические рекомендации к лабораторным работам
для студентов специальности 1-25 01 04 «Финансы и кредит»
очной и заочной форм обучения*



Могилев 2023

УДК 330.43(075.8)
ББК 65в6я73
Э40

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Маркетинг и менеджмент» «8» ноября 2023 г.,
протокол № 4

Составитель канд. физ.-мат. наук, доц. В. А. Ливинская

Рецензент канд. техн. наук, доц. Т. В. Пузанова

Методические рекомендации предназначены для приобретения практических навыков в решении эконометрических задач с помощью современных инструментов анализа данных студентами, обучающимися по специальности 1-25 01 04 «Финансы и кредит». Представлены темы, изучаемые студентами в курсе. Содержат образцы скриптов на языке R и описания возможностей применения MS Excel для решения задач.

Учебное издание

ЭКОНОМЕТРИКА

Ответственный за выпуск	А. В. Александров
Корректор	А. А. Подошевка
Компьютерная верстка	Е. В. Ковалевская

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. . Уч.- изд. л. . Тираж 26 экз. Заказ №

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 07.03.2019.
Пр-т Мира, 43, 212022, г. Могилев.

© Белорусско-Российский
университет, 2023

Содержание

Введение	4
Лабораторная работа № 1. Знакомство со средой для интеллектуального анализа данных R	5
Лабораторная работа № 2. Разведывательный анализ данных средствами R...	6
Лабораторная работа № 3. Проверка статистических гипотез о параметрах нормального распределения с применением MS Excel, R-Studio	8
Лабораторная работа № 4. Проведение множественного корреляционного анализа в MS Excel и R-Studio.....	9
Лабораторная работа № 5. Построение и эконометрический анализ двумерной регрессионной модели с использованием MS Excel, R-Studio	10
Лабораторная работа № 6. Расчет и экономическое обоснование параметров уравнения множественной регрессии с использованием MS Excel, R-Studio.	13
Лабораторная работа № 7. Проверка выполнения предпосылок МНК: наличие и устранение гетероскедастичности.	15
Лабораторная работа № 8. Проверка выполнения предпосылок МНК: наличие и устранение автокорреляции	17
Лабораторная работа № 9. Подбор наиболее адекватного уравнения нелинейной регрессии	18
Лабораторная работа № 10. Построение ANOVA- и ANCOVA-моделей.....	20
Лабораторная работа № 11. Моделирования тенденции временных рядов....	23
Лабораторная работа № 12. Моделирование сезонности различными способами	25
Лабораторная работа № 13. Экспоненциальное сглаживание	26
Список литературы	28

Введение

Цель преподавания дисциплины «Эконометрика» заключается в получении студентами навыков эконометрического моделирования социально-экономических процессов, с использованием современных инструментов для анализа данных MS Excel и среды R.

Цель методических рекомендаций – помочь студентам в самостоятельной подготовке и выполнении задания к лабораторным работам по дисциплине.

Порядок выполнения каждой лабораторной работы.

- 1 Изучить теоретические сведения.
- 2 Получить задание у преподавателя, выполнить в соответствии с заданным вариантом.
- 3 Сделать выводы по результатам выполнения задания.
- 4 Подготовить ответы на контрольные вопросы.
- 5 Оформить отчет.

Содержание отчета.

- 1 Цель работы.
- 2 Постановка задачи.
- 3 Результаты выполнения задания.
- 4 Выводы.
- 5 Ответы на контрольные вопросы.

Лабораторная работа № 1. Знакомство со средой для интеллектуального анализа данных R

Цель: ознакомиться с интерфейсом среды R-Studio; научиться работать в режиме консоли и путем написания скриптов, а также подключать внешние пакеты; изучить основные методы обработки статистических данных.

Методические указания

Для установки среды R и Rstudio на свой домашний компьютер воспользуйтесь инструкцией по ссылке, приведенной в сноске¹.

Директория – это место (папка), где находится ваш «проект». То есть там лежит скрипт, данные, картинки и прочее.

Зачастую, у каждого проекта своя директория, поэтому приходится часто их менять. Есть несколько способов это сделать.

1 Использовать функцию `setwd(~/Desktop/R)`, где в кавычках можно прописать путь к директории.

2 На панели R нажать кнопку `Session` и в выпадающем списке выбрать `Set Working Directory`. После этого можно выбрать нужную папку.

3 В нижнем правом окошке есть вся файловая система компьютера, где также можно установить нужную директорию. Нажав кнопку `More`, можно перейти в текущую директорию или установить новую.

С целью дальнейшего упрощения работы рекомендуется заранее установить некоторые основные пакеты с помощью функции `install.packages` набрав команду: `install.packages(«название пакета»)`, или, войдя во вкладку `Packages/Install Packages`, затем подключить нужный пакет. Для работы чаще всего используются следующие пакеты:

`psych` – описательная статистика;

`ggplot2` – графика;

`lmtest` – тестирование гипотез при построении линейных моделей;

`MASS` – поиск подходящих распределений;

`dplyr` – вычислительные операции с данными.

После загрузки всех необходимых пакетов нужно узнать, в какой директории вы находитесь. Это можно сделать с помощью функцию `getwd()`.

Импорт данных – это загрузка в среду R различных данных для последующей работы с ними. Важно понимать откуда мы берём данные. Обычно источником импорта выступают: обычные текстовые файлы (`.csv`, `.txt`), Excel-файлы (`.xlsx`, `.xls`), базы данных (SQL), интернет, статистические пакеты (SPSS, SAS, STATA).

Импорт файлов с расширениями `csv`, `txt` осуществляется с помощью функции `read.csv()`. Формат `csv` самый распространенный формат хранения данных в мире для анализа данных. Его можно получить, например, сохранив Excel-файл в формате `csv` (разделитель – запятая).

¹ https://bdemeshev.github.io/installation/r/R_installation.html

Проверка того, что все данные импортировались нормально, осуществляется с помощью функции `str()`.

Экспорт данных – это сохранение данных с требуемым расширением на свой компьютер. Для экспорта данных применяют функции: `write.csv(df, «table_car.csv»)` или `write.xlsx(df, «table_car.xlsx»)` или `write.html(df, «table_car.html»)`, где `df` – имя фрейма данных в среде R.

Задание

- 1 Инсталлировать пакеты для ввода данных в различных форматах.
- 2 Изучить возможность получать справочную информацию по необходимым пакетам.
- 3 Создать свою директорию.
- 4 Загрузить данные для своего варианта, полученные у преподавателя в переменную.
- 5 Получить справочную информацию по своим данным, просмотреть их содержимое.
- 6 Проверить, есть ли среди данных пропуски, удалить с помощью соответствующей команды.
- 7 Изучить типы данных.
- 8 Выполнить экспорт очищенных данных в MS Excel.
- 9 Составить отчет.

Контрольные вопросы

- 1 Какие типы данных используются в среде R?
- 2 Как начать работу в R-Studio?
- 3 Как получить справочную информацию по пакетам и встроенным в среду R датасетам?
- 4 Как импортировать в среду R файлы с различными расширениями?
- 5 Как экспортировать из среды R файлы с различными расширениями?

Лабораторная работа № 2. Разведывательный анализ данных средствами R

Цель: провести простой анализ данных, состоящий из описательной статистики и визуализации.

Методические указания

Разведочный анализ данных основан на построении визуализаций и вычислении характеристик описательной статистики, представленных в таблице 1.

Таблица 1 – Параметры описательной статистики

Функция	Описание
mean()	Среднее значение
median()	Медиана
var()	Дисперсия
sd()	Стандартное отклонение
min()	Минимальное значение
max()	Максимальное значение
quantile()	Квантили (по умолчанию рассчитывается максимальное и минимальное значения, а также квантили)
summary()	Сводка по параметрам описательной статистики для всех переменных набора

Среди данных могут быть пропуски, которые не позволят рассчитать статистические параметры для столбцов количественных переменных с пропусками без специальных настроек.

Функция `is.na()` возвращает набор данных (вектор, таблицу и т. п.), заполненный значениями `true` (если значение отсутствует) и `false` (при его наличии). Так как `true` маркируется 1, а `false` маркируется нулем, то функция `sum()`, примененная к такому преобразованному набору данных, дает число пропусков.

Для построения графиков используют высокоуровневые функции, доступные без подключения специализированных пакетов для построения визуализаций (таблица 2).

Таблица 2 – Функции для построения графиков

Функция	Описание
plot()	Общая функция для построения графиков. В зависимости от передаваемых данных и настройки параметров, определяемая ей визуализация может быть и точечным графиком, и столбчатой диаграммой
boxplot()	График «ящик с усами»
hist()	Гистограмма
barplot()	Столбчатая диаграмма
scatterplot()	Диаграмма рассеяния (точечная)
pie()	Круговая диаграмма

Для построения гистограммы с помощью библиотеки `ggplot2` воспользоваться [3].

Команда `Glimpse(f)` (данная функция из ранее уставленного пакета «`dplyr`») позволяет посмотреть краткие сведения о наборе данных. Вывод на экран описательной статистики осуществляется также с помощью функции «`describe`», относящейся к пакету «`psych`».

Задание

1 Импортировать данные, полученные у преподавателя, согласно своему варианту.

2 Создать новую переменную-вектор, в которой будут 1, если значение в исходном векторе больше среднего, и -1, если значение переменной меньше среднего, и 0, если значение равно среднему с команды `ifelse`.

3 Вывести описательную статистику.

4 Построить графики абсолютных частот и плотности распределения с помощью базового пакета и библиотеки `ggplot2`.

5 Оформить отчет.

Контрольные вопросы

1 Как создать количественную переменную в среде R?

2 Как создать вектор качественных переменных в среде R?

3 Как рассчитать показатели описательной статистики?

4 Как построить гистограмму и график плотности распределения в среде R?

Лабораторная работа № 3. Проверка статистических гипотез о параметрах нормального распределения с применением MS Excel, R-Studio

Цель: научиться формулировать статистические гипотезы и оценивать вероятности ошибок принятия гипотезы, проверять статистические гипотезы согласно алгоритму, делать выводы по результатам статистической проверки, использовать возможности Excel и среды R для проверки гипотез.

Методические указания

Алгоритм проверки статистических гипотез.

1 По выборочным данным формулируют основную H_0 и альтернативную H_1 гипотезы.

2 Задают уровень значимости α (0,05 или 0,01).

3 В зависимости от H_0 определяют статистический критерий K , имеющий известное распределение.

4 По выборке и формуле критерия K рассчитывают наблюдаемое значение критерия $K_{набл}$.

5 В зависимости от вида H_1 определяют вид критической области W и критические точки по соответствующим таблицам для распределения критерия K .

6 По результатам проверки принадлежности $K_{набл}$ к критической области делают вывод о принятии или отклонении гипотезы H_0 . Формулируют общий вывод, исходя из поставленной задачи.

Для освоения навыков проверки статистических гипотез в MS Excel воспользоваться материалом в [2].

Для освоения навыков проверки статистических гипотез в среде R воспользоваться материалом [3].

Задание

- 1 Получить данные у преподавателя согласно своему варианту.
- 2 Описать исходную совокупность с помощью показателей дескриптивной статистики.
- 3 Построить гистограммы для количественных признаков и ящик с усами для исходного набора данных, выдвинуть предположение о функции распределения.
- 4 Проверить гипотезу о нормальном распределении с помощью различных критериев.
- 5 Протестировать гипотезу об однородности и о различии в типичном значении выбранного признака выборки согласно варианту.
- 6 Оформить отчет.

Лабораторная работа № 4. Проведение множественного корреляционного анализа в MS Excel и R-Studio

Цель: овладеть методикой проведения парного и множественного корреляционного анализа в MS Excel и в среде R-Studio.

Методические указания

В MS Excel для проведения корреляционного анализа существует два способа [2].

Способ 1: определение корреляции через Мастер функций.

Одним из способов, с помощью которого можно провести корреляционный анализ, является использование функции КОРРЕЛ. Сама функция имеет общий вид КОРРЕЛ(массив1; массив2).

Способ 2: вычисление корреляции с помощью Пакета анализа (рисунок 1).

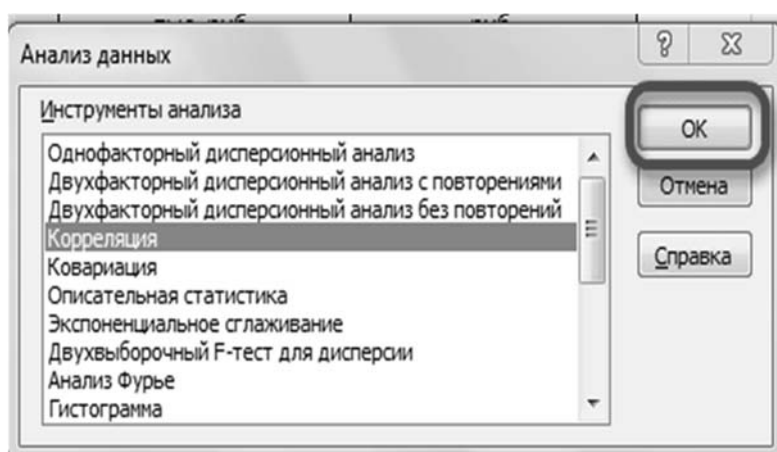


Рисунок 1 – Построение корреляционной матрицы в MS Excel

В среде R существует четыре распространенных способа создания корреляционной матрицы (таблица 3).

Таблица 3 – Корреляционный анализ в R

Библиотека	Функция	Описание
Базовый R	cor(df)	Для получения простой матрицы коэффициентов корреляции
Hmisc	rcorr(as.matrix(df))	Для получения р-значений коэффициентов корреляции
corrplot	corrplot(cor(df))	Для визуализации матрицы корреляции
ggcorrplot	ggcorrplot(cor(df))	Для визуализации матрицы корреляции
Базовый R	cor.test(d\$x, d\$y)	Расчет коэффициента и проверка его значимости в среде R

Задание

- 1 Загрузить набор данных для своего варианта, ознакомиться с его содержимым.
- 2 Построить график корреляционного поля для каждого фактора.
- 3 Провести корреляционный анализ с помощью возможностей MS Excel и R.
- 4 Проверить гипотезу о силе линейной связи между признаками.
- 5 Составить отчет.

Контрольные вопросы

- 1 Что такое корреляция признаков?
- 2 Как графически определить наличие/отсутствие связи между признаками?
- 3 Тесноту какого типа связи можно оценить с помощью коэффициента корреляции Пирсона?
- 4 Как вычисляется линейный коэффициент парной корреляции r ?
- 5 Как осуществляется оценка статистической значимости линейного коэффициента парной корреляции r ?
- 6 Где можно посмотреть критическое значение для проверки коэффициента на значимость?
- 7 Когда и зачем строится доверительный интервал для линейного коэффициента парной корреляции?

Лабораторная работа № 5. Построение и эконометрический анализ двумерной регрессионной модели с использованием MS Excel, R-Studio

Цель: изучить возможности построения парной линейной регрессии.

Методические указания

Методы регрессионного анализа рассчитаны, главным образом, на случай устойчивого нормального распределения, в котором изменения от опыта к опыту проявляются лишь в виде независимых испытаний.

Уравнение регрессии – функция, позволяющая по величине изменения одного коррелируемого признака определить среднюю величину другого признака. Выделим основные этапы регрессионного анализа.

Первый этап. Предположение. На этом этапе происходит выбор формы связи между переменными (модель).

Второй этап. Параметризация – происходит оценка значений параметра в выбранной формуле статистической связи. Форма связи (функция) линейная, нелинейная.

Третий этап. Проверка надёжности полученных оценок. На этом этапе осуществляются следующие тесты: F-тест (проверка статистической значимости выбранной формы связи), t-тест (проверка статистической значимости найденных числовых значений параметра). В результате анализа статистических данных, выбора и построения модели последовательно выполняются все три этапа.

В MS Excel данный анализ выполняется с помощью вкладки Анализ данных (рисунок 2). Подробная инструкция может быть изучена в [2].

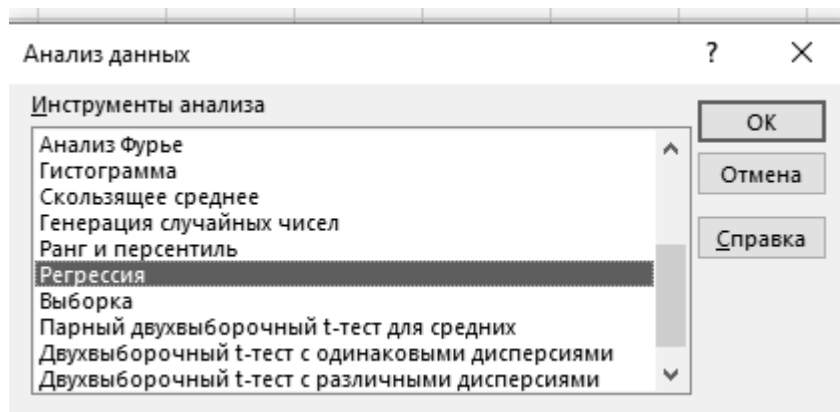


Рисунок 2 – Регрессионный анализ в MS Excel

Рассмотрим построение парной линейной регрессии в среде R (символ # означает начало комментария).

```
# Загрузка необходимых библиотек
library(ggplot2)
library(ggpubr)
# Загрузка данных
data <- read.csv("data.csv")
# Парная линейная регрессия
linear_model<-lm(dependent_variable ~ independent_variable, data=data)
# Визуализация результатов
summary(linear_model)
# Создание графика для визуализации результатов
ggplot(data, aes(x=independent_variable, y=dependent_variable)) +
  geom_point() +
```

```
geom_smooth(method="lm", se=FALSE) +
labs(x="Independent Variable", y="Dependent Variable") +
theme_minimal()
```

Результат анализа модели линейной регрессии отображается с помощью функции `summary()`. Для визуализации результатов используется `ggplot2`, где `geom_point()` показывает исходные данные, а `geom_smooth()` строит линию регрессии, отражающую модель.

Рассчитанные по модели значения можно получить командой:

```
fitted(linear_model)
# базовый пакет stats
```

Для того чтобы построить прогноз по полученной модели, нужно задать значения регрессора и поместить их в новый `data.frame` данных:

```
nd <- data.frame(x=c(40,60))
```

Строим прогноз функцией `predict`:

```
predict(linear_model, nd)
```

Задание

1 Изучить возможность применения MS Excel, среды R для проведения регрессионного анализа двух признаков.

2 Для эндогенной переменной Y (согласно варианту) и экзогенной переменной X постройте поле корреляции и сформулируйте гипотезу о форме связи между Y и X .

3 Рассчитайте оценки параметров уравнения парной линейной регрессии.

4 Оцените тесноту связи между Y и X с помощью выборочного коэффициента корреляции. Проверьте значимость коэффициента корреляции ($\alpha = 0,05$).

5 Рассчитайте выборочный коэффициент детерминации. Сделайте экономический вывод.

6 Проверьте значимость оценки коэффициента регрессии с помощью критерия Стьюдента при уровне значимости $\alpha = 0,05$.

7 Постройте доверительный интервал для коэффициента регрессии. Дайте экономическую интерпретацию.

8 Составьте таблицу дисперсионного анализа. Оцените с помощью F-критерия Фишера – Снедекора значимость уравнения линейной регрессии ($\alpha = 0,05$).

9 Рассчитайте интегральный показатель результативного признака, если факторный признак снизится на 10 % относительно последнего периода. Постройте доверительный интервал для прогнозного значения эндогенной переменной. Сделайте экономический вывод.

10 Определить среднюю ошибку аппроксимации.

11 На поле корреляции постройте линию регрессии.

Контрольные вопросы

1 С какой целью в линейном регрессионном анализе применяется метод наименьших квадратов?

2 Как изменится доверительный интервал для выборочного значения \hat{y} при различных значениях аргумента x ?

3 Что такое коэффициент детерминации?

4 Опишите процедуру проверки гипотез относительно коэффициентов линейной регрессии. Кстати, какое распределение имеет используемая при этом статистика?

5 Как проверить значимость (качество) уравнения регрессии? Какая статистика используется при этом?

Лабораторная работа № 6. Расчет и экономическое обоснование параметров уравнения множественной регрессии с использованием MS Excel, R-Studio

Цель: научиться оценивать линейное уравнение множественной регрессии и проверять его качество.

Методические указания

Множественная регрессия – это метод статистического анализа, который используется для прогнозирования зависимой переменной на основе нескольких независимых переменных.

Оценка коэффициентов уравнения множественной регрессии в MS Excel и в среде R осуществляется аналогично тому, как она производится с помощью тех же функций, которые применяются для парной регрессии (см. лабораторную работу № 5).

Проверка качества уравнения множественной регрессии заключается в следующих действиях:

1) проверка значимости всех α_j ;

2) проверка общего качества уравнения регрессии с помощью коэффициента множественной детерминации R^2 ;

3) проверка свойств данных, выполнение которых предполагалось при оценивании уравнений. Ошибки ε_j распределены по нормальному закону с постоянной дисперсией ($\sigma^2 = \text{const}$).

Коэффициент множественной детерминации показывает долю вариации результирующего признака, обусловленного вариацией факторных признаков.

Проверка значимости коэффициента детерминации осуществляется с помощью F-распределения.

Основной синтаксис для соответствия модели множественной линейной регрессии в среде R выглядит следующим образом:

```
model<-lm(response_variable ~ predictor_variable1 + predictor_variable2 + ...,
data = df)
```

В регрессионном анализе мультиколлинеарность возникает, когда две или более переменных-предикторов сильно коррелируют друг с другом, так что они не предоставляют уникальной или независимой информации в регрессионной модели.

Если степень корреляции между предикторными переменными достаточно высока, это может вызвать проблемы при подгонке и интерпретации регрессионной модели.

Самый простой способ обнаружить мультиколлинеарность в регрессионной модели – это вычислить показатель, известный как коэффициент инфляции дисперсии, часто сокращаемый как VIF.

VIF измеряет силу корреляции между предикторными переменными в модели. Он принимает значение от 1 до положительной бесконечности.

Используют следующие эмпирические правила для интерпретации значений VIF, представленные в таблице 4.

Таблица 4 – Принятие решения на основании коэффициента инфляции дисперсии

Значение VIF	Вывод
VIF = 1	Нет корреляции между данной предикторной переменной и любыми другими предикторными переменными в модели
$1 < VIF < 5$	Существует умеренная корреляция между данной предикторной переменной и другими предикторными переменными в модели
VIF > 5	Существует серьезная корреляция между данной предикторной переменной и другими предикторными переменными в модели

Чтобы вычислить VIF для каждой переменной-предиктора в модели, мы можем использовать функцию `vif(model)` из пакета `car`.

Задание

1 Построить уравнение множественной регрессии согласно варианту, полученному у преподавателя. Пояснить смысл параметров уравнения.

2 Проверить значимость коэффициентов уравнения.

3 Дать оценку полученного уравнения с помощью общего F-критерия Фишера.

4 Рассчитать частные коэффициенты эластичности для значимых факторов.

5 Ранжировать факторы по силе воздействия на результативный фактор.

6 Рассчитать абсолютную, относительную и среднюю ошибки прогноза.

7 Получить прогнозное значение результата при значении важнейшей объясняющей переменной, равной максимальному наблюдаемому значению,

увеличенному на 10 %, и при значении второй объясняющей переменной, равной минимальному наблюдаемому значению, уменьшенному на 15 %.

8 Получить интервальное предсказание значения y с надежностью 0,95.

9 Составить отчет.

Контрольные вопросы

1 Как оценивается модель множественной регрессии в матричном виде?

2 Как проверяется качество регрессионной модели?

3 Что такое мультиколлинеарность, как ее определить и всегда ли следует с ней бороться?

4 Как интерпретируются коэффициенты уравнения множественной регрессии?

5 Что такое скорректированный коэффициент детерминации?

Лабораторная работа № 7. Проверка выполнения предпосылок МНК: наличие и устранение гетероскедастичности

Цель: научиться оценивать наличие эффекта гетероскедастичности и устранять его.

Методические указания

Прежде чем приступить к проверке выходных данных модели, необходимо сначала проверить, выполняются ли условия Гаусса – Маркова (в MS Excel не реализована возможность выполнять тесты для их проверки, вычисления следует проводить в ручном режиме, по формулам). В среде R данные тесты выполняются довольно просто. Рассмотрим эти условия.

1 *Распределение невязок модели должно быть примерно нормальным.*

Для проверки выполняется ли это предположение, можно создать простую гистограмму остатков с помощью функции `hist(residuals(model), col = «steelblue»)` и сопоставить ее с гистограммой нормально-распределенной случайной величины. Также провести тест Шапиро – Уилка (`shapiro.test()`).

2 *Дисперсия остатков должна быть постоянной для всех наблюдений.*

Это предпочтительное условие известно как гомоскедастичность. Нарушение этого предположения известно как гетероскедастичность.

Графическое обнаружение гетероскедастичности.

Оценивается исходная регрессия и из неё получают остатки с помощью функции `residuals(model)`.

Чтобы проверить, соблюдается ли предположение о гомоскедастичности остатков, можно создать *график зависимости теоретических значений от остатка:*

```
plot(fitted(model), residuals(model))
```

```
abline(h = 0, lty = 2)
```

В идеале желательно, чтобы остатки были одинаково разбросаны при каждом подобранном значении.

Можно построить график зависимости вектора квадрата нормализованных остатков модели от переменной, предположительно повинной в гетроскедастичности (vari):

```
model.st.resid <- rstandard(model)
# получаем стандартизированные остатки модели
```

Построим график соответствующей зависимости

```
qplot(vari,abs(m1.st.resid),data=h, alpha=0.05)
```

В дальнейших тестах используются следующие гипотезы:

H_0 (нулевая гипотеза): в остатках нет гетероскедастичности;

H_A (альтернативная гипотеза): остатки гетероскедастичны.

Для проверки нулевой гипотезы разработаны следующие тесты, представленные в таблице 5.

Таблица 5 – Тестирование гетероскедастичности

Тест	Функция	Примечание
<i>Бройша – Пагана</i>	library(bstats) library(lmtest) bptest(model)	Требуется нормальность остатков
<i>Голдфельда – Квандта</i>	library(lmtest) h2 <- h[order(h\$totsp),] # сменим порядок строк в табличке h m2 <- lm(price~totsp,data=h2) gqtest(m2, fraction=0.2) # проведем GQ тест, # выкинув посередине 20 % наблюдений	Требуется нормальность остатков
Уайта	library(bstats) white.test(model)	Не требуется нормальность остатков

Задание

1 Получить вектор ошибок в лабораторной работе № 6 и протестировать его на нарушение предпосылок МНК (гетероскедастичность) различными тестами в среде R и в MS Excel.

2 Составить отчет.

Контрольные вопросы

1 В чем суть гетероскедастичности?

2 Каковы последствия гетероскедастичности?

3 Какими методами можно обнаружить гетероскедастичность?

4 Как можно устранить гетероскедастичность?

Лабораторная работа № 8. Проверка выполнения предпосылок МНК: наличие и устранение автокорреляции

Цель: научиться оценивать наличие эффекта автокорреляции и устранение ее влияния.

Методические указания

Наличие автокорреляции остатков является нарушением одной из предпосылок МНК, поэтому при обнаружении автокорреляции необходимо выявить причину и устранить ее.

Существует несколько методов обнаружения автокорреляции:

- графический метод;
- тестирование гипотез.

Графический метод обнаружения автокорреляции остатков заключается в построении диагностической диаграммы – точечной графической зависимости между прогнозируемыми значениями результирующего показателя и эмпирическими случайными отклонениями (в случае парной регрессии можно также рассматривать зависимость случайных отклонений от значений фактора).

Для проверки этого допущения используются следующие гипотезы:

H_0 (нулевая гипотеза): между остатками нет корреляции;

H_A (альтернативная гипотеза): остатки автокоррелированы.

В MS Excel не реализована возможность выполнять данный тест, вычисления следует проводить в ручном режиме. В среде R это реализовано с помощью различных библиотек (таблица 6).

Таблица 6 – Тестирование автокорреляции

Тест	Функция	Примечание
<i>Дарбина – Уотсона</i>	library(car) durbinWatsonTest(model) library(lmtest) dwtest (formula)	Обязательна нормальность и гомоскедастичность случайных возмущений, formula – спецификация модели, которую нужно оценить
<i>Бреуша – Годфри</i>	library(lmtest) bgttest (formula, order = 1, order.by = NULL, type = c("Chisq", "F")). Параметры функции: formula – спецификация модели order – порядок авторегрессии; type = c("Chisq", "F") – выбор статистики	

Чтобы выполнить тест Дарбина – Уотсона, сначала нужно подобрать модель линейной регрессии.

Вычисленное значение DW -критерия сравнивается с критическими точками dl и du , которые выбираются из таблицы Дарбина – Уотсона при заданном уровне значимости α в зависимости от числа наблюдений (объема выборки) n и числа факторов в регрессионной модели m (в парной регрессионной модели $m = 1$).

Вывод о наличии автокорреляции делается в зависимости от принадлежности критерия DW одному из интервалов:

- при $0 \leq DW \leq dl$ наблюдается положительная автокорреляция остатков;
- при $du \leq DW \leq (4 - du)$ автокорреляция остатков отсутствует;
- при $(4 - dl) \leq DW \leq 4$ наблюдается отрицательная автокорреляция остатков;
- оставшиеся интервалы являются областями неопределенности, однозначно сделать вывод о наличии или отсутствии автокорреляции остатков в этом случае невозможно.

Если нулевая гипотеза отвергается и делается заключение, что в остатках присутствует автокорреляция, то возможно несколько различных вариантов поведения.

1 Для положительной последовательной корреляции рассмотреть возможность добавления в модель лагов зависимой и/или независимой переменной.

2 Для отрицательной последовательной корреляции убедиться, что ни одна из переменных не является сверхизменчивой.

3 Для сезонной корреляции рассмотреть возможность добавления в модель сезонных фиктивных переменных.

4 При обнаружении автокорреляции остатков необходимо выяснить причины ее возникновения и предложить способ устранения.

Задание

1 Получить вектор ошибок в лабораторной работе № 6 и протестировать его на нарушение предпосылок МНК (автокорреляция) различными тестами в среде R и в MS Excel.

2 Составить отчет.

Лабораторная работа № 9. Подбор наиболее адекватного уравнения нелинейной регрессии

Цель: научиться подбирать лучшую из возможных нелинейных зависимостей в парной регрессии по критерию минимальной остаточной дисперсии.

Методические указания

Различают два класса нелинейных регрессий:

- 1) регрессии нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;
- 2) регрессии нелинейные по оцениваемым параметрам.

Если модель внутренне нелинейна по параметрам, то для оценки параметров используются итеративные процедуры, успешность которых зависит от вида уравнений и особенностей итеративной процедуры. Модели внутренне нелинейные по параметрам, могут иметь место в эконометрических исследованиях, однако большее распространение получили модели, приводимые к линейному виду. Решение такого типа моделей реализовано в стандартных пакетах прикладных программ.

Для реализации в среде R модели нелинейной регрессии можно воспользоваться следующим кодом:

```
# Загрузка необходимых библиотек
library(ggplot2)
library(ggpubr)
# Загрузка данных
data <- read.csv("data.csv")

# Определение исходной функции
model_function <- function(x, a, b) {
  y <- a * x^b
  return(y)
}
# Нелинейная регрессия
nonlinear_model <-
nls(dependent_variable~model_function(independent_variable, a, b),
data=data, start=list(a=1, b=1))
# Визуализация результатов
summary(nonlinear_model)
# Создание графика для визуализации результатов
ggplot(data, aes(x=independent_variable, y=dependent_variable)) +
  geom_point() +
  stat_smooth(method="nls", formula=y~model_function(x, a, b),
method.args=list(start=c(a=1, b=1)), se=FALSE) +
  labs(x="Independent Variable", y="Dependent Variable") +
  theme_minimal()
```

В данном примере предполагается, что у вас есть файл data.csv с двумя переменными (dependent_variable и independent_variable). Вы можете заменить названия переменных в коде на свои собственные.

В результате анализа функции nls() возвращается статистика результата, которую можно отобразить с помощью функции summary(). Для визуализации результатов используется ggplot2, где функция geom_point() показывает исходные данные, а функция stat_smooth() строит кривую, соответствующую модели нелинейной регрессии.

Сопоставив для различных моделей среднеквадратические ошибки (RMSE), делается выбор в пользу той модели, у которой ошибка меньше.

Задание

1 На основании данных, полученным у преподавателя, оценить несколько нелинейных моделей и выбрать наилучшую.

2 Составить отчет.

Контрольные вопросы

1 Перечислите основные виды нелинейных моделей в регрессионном анализе.

2 Какова общая формула для нахождения коэффициента эластичности?

3 Как классический МНК применяется к нелинейным моделям регрессии?

4 Опишите метод линеаризации для сведения нелинейной модели к линейной для различных видов моделей.

Лабораторная работа № 10. Построение ANOVA- и ANCOVA-моделей

Цель: формирование у студентов представления о моделях дисперсионного и ковариационного анализа.

Методические указания

Однофакторный ANOVA (однофакторный дисперсионный анализ) – это метод статистического анализа данных, который используется для определения наличия статистически значимых различий между двумя или более группами по одной независимой переменной.

Входными данными для однофакторного ANOVA являются значения зависимой переменной и групповой фактор, на основе которых проводится анализ. Фактор может быть любой номинальной или порядковой переменной, которая разделяет выборку на группы (в простом случае, это может быть пол, возраст, уровень образования и т. д.). Зависимая переменная – это та переменная, которую мы хотим сравнить в различных группах.

Однофакторный ANOVA проверяет нулевую гипотезу о том, что среднее значение зависимой переменной одинаково во всех группах. Если p -значение меньше заданного уровня значимости (обычно 0,05), тогда можно сделать вывод о том, что средние значения по группам различаются статистически значимо друг от друга. Кроме того, однофакторный ANOVA дает множество других статистических показателей, включая среднее значение, стандартное отклонение, диапазон, размах, медиану, аномальные значения и т. д.

Для реализации в среде R модели дисперсионного анализа (ANOVA) можно воспользоваться следующим кодом:

```
# Загрузка необходимых библиотек
library(ggplot2)
library(ggpubr)
# Загрузка данных
data <- read.csv("data.csv")
# ANOVA анализ
anova_model<-aov(dependent_variable ~ independent_variable, data=data)
# Отображение результатов анализа
summary(anova_model)
# Создание графика для визуализации результатов
ggplot(data, aes(x=independent_variable, y=dependent_variable)) +
  geom_boxplot() +
  labs(x="Independent Variable", y="Dependent Variable") +
  theme_minimal()
```

Код позволяет построить модель ANOVA, отобразить результаты анализа с помощью функции `summary()` и построить график с использованием `ggplot2`. График показывает средние значения зависимой переменной для каждой группы независимой переменной.

ANCOVA-модель помогает анализировать влияние независимой переменной (`independent_variable`) на зависимую переменную (`dependent_variable`), учитывая влияние возможной контрольной переменной (`covariate`). Код позволяет построить модель ANCOVA, отобразить результаты анализа с помощью функции `summary()` и построить график с использованием `ggplot2`.

```
# Загрузка необходимых библиотек
library(ggplot2)
library(ggpubr)
library(car)
# Загрузка данных
data <- read.csv("data.csv")
# ANCOVA анализ
ancova_model<-lm(dependent_variable ~ independent_variable + covariate,
  data=data)
# Отображение результатов анализа
summary(ancova_model)
# Создание графика для визуализации результатов
ggplot(data, aes(x=independent_variable, y=dependent_variable,
  color=factor(covariate)))
+ geom_point() +
```

```
stat_smooth(method="lm", se=FALSE) +
labs(x="Independent Variable", y="Dependent Variable") +
theme_minimal()
```

В данном примере предполагается, что у вас есть файл data.csv с двумя непрерывными переменными (dependent_variable и independent_variable) и одной категориальной переменной (covariate). Можно заменить названия переменных в коде на свои собственные.

Задание

1 Добавить (домыслить) правдоподобную категориальную переменную к данным того же варианта задания «Парная регрессия», разбив наблюдения в соответствии с уровнями сопутствующей качественной переменной (два уровня в нечетных вариантах и три – в чётных).

2 Ввести в модель нужное число дихотомических фиктивных переменных, оценить параметры модели и записать оценки уравнений регрессии для каждого уровня качественной переменной отдельно.

3 По данным для какого-либо уровня отдельно оценить уравнение регрессии. Сравнить результаты моделирования с таковыми в п. 2. Сделать выводы.

4 Отразить в отчете описание фиктивных переменных, оцененное уравнение регрессии с фиктивными переменными, математическое ожидание результативной переменной при различных значениях фиктивной переменной.

Контрольные вопросы

1 В чем заключается сущность дисперсионного анализа?

2 Назовите основные этапы проверки гипотезы о равенстве выборочных средних вариантов по критерию Фишера.

3 Сформулируйте нулевую и альтернативную гипотезы, которые выдвигаются при дисперсионном анализе.

4 В каком случае возникает необходимость вводить фиктивную переменную? Приведите примеры.

5 Как вводится фиктивная переменная, если у признака больше двух альтернатив?

6 Как интерпретируются коэффициенты при фиктивных переменных?

7 Когда используется тест Чоу? Как он связан с применением фиктивных переменных?

Лабораторная работа № 11. Моделирования тенденции временных рядов

Цель: изучить методы статистического моделирования временных рядов; приобрести основные навыки работы с пакетами в среде R для прогнозирования.

Методические указания

Экстраполяционное прогнозирование экономических процессов, представленных одномерными временными рядами, сводится к выполнению следующих основных этапов:

- 1) предварительный анализ данных (сглаживание с помощью скользящей средней);
- 2) построение моделей: формирование набора аппроксимирующих функций (кривых роста) и численное оценивание параметров моделей;
- 3) проверка адекватности моделей и оценка их точности;
- 4) выбор лучшей модели;
- 5) расчет точечного и интервального прогнозов.

Для проведения анализа одномерного временного ряда в среде R рекомендуется придерживаться следующего алгоритма.

1 Преобразовать исходные данные в формат временного ряда, используя различные настройки временного сегмента.

2 Выполнить сглаживание с помощью различных скользящих средних – выбрать наиболее оптимальную из них по критерию минимума ошибки.

3 Построить кореллограмму для определения структуры ряда с использованием функции `acsf()`.

4 В зависимости от вывода о структуре ряда подобрать наилучший тренд.

Для сглаживания временного ряда с помощью скользящих средних необходимо воспользоваться пакетами `library(zoo)`, `library(TTR)`, `library(quantmod)`.

Степень тесноты статистической связи между уровнями временного ряда, сдвинутыми на τ единиц времени, определяется величиной коэффициента корреляции $r(\tau)$, т. к. $r(\tau)$ измеряет тесноту связи между уровнями одного и того же временного ряда, поэтому его принято называть коэффициентом автокорреляции. При этом длину временного смещения называют обычно лагом (τ).

Автокорреляция – это мера корреляции между значениями ряда с разницей во времени. Если временной ряд имеет высокую автокорреляцию, это означает, что значения ряда в разные периоды времени имеют сильную связь между собой. Автокорреляцию можно вычислить с помощью функции корреляции Пирсона, которая вычисляет корреляцию между двумя переменными. Для временных рядов это означает вычисление корреляции между значениями ряда в разные периоды времени. Если автокорреляция является значимой, то это может означать наличие тренда, цикличности или сезонности в ряде.

Поэтому при оценке «качества» модели проверяют, удовлетворяет ли остаточная последовательность следующим свойствам:

- независимость уровней: их случайность (с помощью критерия серий и критерия поворотных точек) и соответствие нормальному закону распределения;
- проверка равенства математического ожидания уровней ряда остатков нулю (критерий Стьюдента);
- наличие (отсутствие) автокорреляции в отклонениях от модели роста проверяют с помощью критерия Дарбина – Уотсона.

Оценка точности модели осуществляется с помощью расчета максимальной по абсолютной величине, средней по модулю и максимальной относительной ошибке. Лучшей по точности считается та модель, у которой все перечисленные характеристики имеют меньшую величину.

Интервальные прогнозы строятся на основе точечных прогнозов. *Доверительным интервалом* называется такой интервал, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значение прогнозируемого показателя. Ширина интервала зависит от качества модели, т. е. степени ее близости к фактическим данным, числа наблюдений, горизонта прогнозирования и выбранного пользователем уровня вероятности.

Задание

- 1 Самостоятельно выбрать данные для анализа за период с 2010 г. по 2020 г. по месяцам с официального сайта Госкомстата России (<http://www.gks.ru>).
- 2 По данному временному ряду за период с 2010 г. по 2019 г.:
 - определить структуру ряда с выявлением типа процесса по его коррелограмме;
 - оценить форму кривой выравнивания одним из приемов;
 - получить расчетные коэффициенты (параметры) модели;
 - оценить качество модели.
- 3 Получить прогноз на 2020 г. по месяцам и оценить точность, сопоставляя с реальными данными.

Контрольные вопросы

- 1 Назовите известные вам модели, описывающие долговременную тенденцию временного ряда.
- 2 Опишите суть метода наименьших квадратов.
- 3 Как строится линейная модель временного ряда?
- 4 Как определяется структура временного ряда?
- 5 Как оценивается качество модели временного ряда?
- 6 Как осуществляется прогнозирование по модели тренда?

Лабораторная работа № 12. Моделирование сезонности различными способами

Цель: научиться определять структуру ряда с помощью автокорреляционной функции, различать аддитивную и мультипликативную модели сезонности, определять индексы сезонности в каждом из вариантов сезонности, использовать полученную информацию для прогноза тенденции временного ряда.

Методические указания

Существует несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания.

Простейший подход – расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда. Общий вид аддитивной модели следующий:

$$Y = T + S + E .$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент. Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E .$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой T , сезонной S и случайной E компонент. Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T , S и E для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

- 1 Выравнивание исходного ряда методом скользящей средней.
- 2 Расчет значений сезонной компоненты S .
- 3 Устранение сезонной компоненты из исходных уровней ряда и получение выравненных данных ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели.
- 4 Аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда.

5 Расчет полученных по модели значений $(T + S)$ или $(T \cdot S)$.

6 Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

Задание

Получить у преподавателя данные о значении спроса на продукцию за каждый квартал в течение пяти лет. Требуется построить поквартальный прогноз спроса для следующего года.

Контрольные вопросы

- 1 Дайте определение сезонных колебаний.
- 2 Какие факторы вызывают сезонные изменения в экономических процессах?
- 3 В чем отличие циклической компоненты временного ряда от сезонной?
- 4 Как графически определить наличие сезонных колебаний во временном ряде?
- 5 Какие показатели используются для измерения сезонности?
- 6 Что такое сезонная волна?
- 7 Какими моделями возможно представление сезонной компоненты в тренд-сезонных экономических процессах?

Лабораторная работа № 13. Экспоненциальное сглаживание

Цель: научиться строить прогнозы с применением метода экспоненциального сглаживания

Методические указания

Кроме метода скользящего среднего, для устранения колебаний в динамическом ряде используется метод экспоненциального сглаживания. Каждое сглаженное значение рассчитывается путём сочетания предыдущего сглаженного значения и текущего значения временного ряда. В этом случае текущее значение временного ряда взвешивается с учётом сглаживающей константы. Расчет производится по формуле

$$S_t = a \cdot y_t + (1 - a) \cdot S_{t-1},$$

где S_t – значение экспоненциальной средней в момент времени t ;

a – параметр сглаживания, $a = \text{const}$, $0 < a < 1$;

y_t – текущее значение временного ряда;

S_{t-1} – предыдущее значение экспоненциальной средней.

Для выполнения экспоненциального сглаживания в среде R можно использовать различные пакеты и библиотеки, такие как `forecast`, `expsmooth`, `stlab`, `stats` и другие.

В качестве примера, рассмотрим экспоненциальное сглаживание с использованием функции `filter()` из пакета `stats`:

```
# Загрузка данных
data(AirPassengers)
plot(AirPassengers)
# Экспоненциальное сглаживание
h <- 0.2 # Параметр сглаживания
fit <- filter(AirPassengers, h = h, method = "exponential")
lines(fit, col = "red", lty = 3)
```

Здесь загрузили данные `AirPassengers` и применили к ним экспоненциальное сглаживание с параметром $h = 0,2$. Результат сглаживания отображается на графике.

Можно использовать другие функции для экспоненциального сглаживания из других пакетов, например, `exponentialSmooth()` из библиотеки `stats`, или `ets()` из пакета `forecast`.

Задание

Требуется построить поквартальный прогноз спроса (см. исходные данные из лабораторной работы № 12) для следующего года методами:

- экспоненциального сглаживания;
- Холта (с учетом тренда);
- Винтера (с учетом тренда и сезонных колебаний).

Контрольные вопросы

- 1 Назначение метода экспоненциального сглаживания.
- 2 Как влияет увеличение и уменьшение интервала сглаживания на характер прогнозной оценки?
- 3 От каких параметров и каким образом зависит величина ошибки прогноза при экспоненциальном сглаживании?

Список литературы

1 **Атчаде, М. Н.** Учебное пособие по дисциплине «Статистика. Статистический анализ: реализация в программе R» / М. Н. Атчаде, И. И. Елисеева. – Санкт-Петербург : СПбГЭУ, 2017. – 80 с.

2 **Козлов, А. Ю.** Статистический анализ данных в MS Excel : учебное пособие / А. Ю. Козлов, В. С. Мхитарян, В. Ф. Шишов. – Москва : ИНФРА-М, 2023. – 320 с.

3 **Лонг, Дж. Д.** Книга рецептов: Проверенные рецепты для статистики, анализа и визуализации данных : пер. с англ. / Дж. Д. Лонг, П. Титор. – Москва: ДМК Пресс, 2020. – 510 с.