МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Методы статистического анализа и прогнозирования» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика (утв. приказом Минобрнауки России от 19.09.2017г. №916) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Ермолицкая М.З.

Кригер А.Б.

Утверждена на заседании научно-образовательный центр "искусственный интеллект" от 27.05.2026 , протокол № 5

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кригер А.Б.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1582918206
Номер транзакции	0000000000F73B8E
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Цели освоения учебной дисциплины:

- получение базовых знаний по актуальным разделам прикладной статистики;
- формирование умений и навыков применения методов статистического анализа для выявления закономерностей и извлечения знаний из данных, полученных в результате проведения эмпирических наблюдений и опросов;
- знакомство с технологией статистического анализа больших данных: прикладным программным обеспечением и статистическими информационными платформами

Задачи освоения дисциплины:

- изучение принципов работы с прикладными программами R и Rstudio, python;
- получение навыков создания наборов данных в Rstudio и/или python, конвертации данных в нужный тип;
- получение навыков выбора пакетов и функций необходимых для анализа данных;
- диагностика построенных моделей.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине		
			Код результата	Формулировка результата	
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-1 : Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.1к : Применяет и развивает знания в области математики, естественнонаучных и социально-экономических наук в профессиональной деятельности	РД3	Знание	базовых положений методик статистической проверки гипотез
			РД3	Умение	проводить диагностику статистических моделей
			РД3	Навык	работы со специальными статистическими пакетами и их инструментами
		ОПК-1.2к : Решает нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и	РД1	Знание	теоретических основ прикладной информатики
			РД1	Умение	применять информационные технологии в исследовательской и проектной деятельности
			РД1	Навык	выбора инструментальных средств для аналитической работы с учетом требований проекта

		профессиональных знаний	РД2	Знание	методик организации эмпирических наблюдений, методов статистического оценивания параметров выборок
			РД2	Умение	применять и развивать методы прикладной статистики
			РД2	Навык	внедрения и использования прикладного программного обеспечения, реализующего статистический анализ

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	**Формирование ценностей**	**Целевые ориентиры**
Формирование научного мировоззрения и культуры мышления		
Формирование осознания ценности научного мировоззрения и критического мышления	Гуманизм	Системное мышление
Формирование коммуникативных навыков и культуры общения		
Формирование навыков публичного выступления и презентации своих идей	Взаимопомощь и взаимоуважение	Умение работать в команде и взаимопомощь

2 Место дисциплины (модуля) в структуре ОПОП

Данная дисциплина относится обязательной части Блока 1 Дисциплины (модули).

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обуче-ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо-емкость	Объем контактной работы (час)						СРС	Форма аттес-тации
				(З.Е.)	Всего	Аудиторная			Внеауди-торная			
						лек.	прак.	лаб.	ПА	КСР		

09.04.03 Прикладная информатика	ОФО	М01.Б	1	5	47	10	36	0	1	0	133	Э

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Код результата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Данные: общие сведения. Работа с данными.	РД1, РД1, РД1	1	2	0	10	собеседование
2	Математическая статистика: общие сведения.	РД1, РД2	1	4	0	15	практическое задание
3	Выборочные характеристики и их свойства.	РД2, РД2, РД3	2	2	0	20	практическое задание
4	Сведения о статистическом оценивании параметров. Проверка гипотез	РД2, РД3	0	4	0	40	практическое задание
5	Модели временных рядов.	РД2, РД2, РД3, РД3	6	14	0	24	контрольная работа
6	Модели временных рядов	РД2, РД3, РД3	0	10	0	24	контрольная работа
Итого по таблице			**10**	**36**	**0**	**133**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Данные: общие сведения. Работа с данными.

Содержание темы: Типы данных: числовые, текстовые, изображения, категориальные, порядковые категориальные, временные ряды, панельные данные. Разница между временными рядами, объектами, набором объектов, панельными данными.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические задания, компьютерное моделирование, метод активного обучения – «конференция».

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 2 Математическая статистика: общие сведения.

Содержание темы: Роль в анализе данных, что описывают данные, какую информацию можно извлечь из них, цель использования, условия применения, ограничения и т.п.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические задания, компьютерное моделирование, метод активного обучения – «конференция».

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 3 Выборочные характеристики и их свойства.

Содержание темы: Оценка выборочных параметров. Описательная статистика (среднее, дисперсия, стандартное отклонение, мода, медиана, асимметрия, эксцесс и т.д.). Законы распределения (нормальное распределение и его свойства; распределения Пирсона, Стьюдента, Фишера, биноминальное распределение и др.). Проверка гипотез на нормальность распределения. Выборки: природа данных, смещенные и несмещенные выборки. Масштабирование, центрирование, нормализация / стандартизация данных. Логарифмирование данных выборок: возможности и проблемы. .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические задания, компьютерное моделирование, метод активного обучения – «конференция».

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 4 Сведения о статистическом оценивании параметров. Проверка гипотез.

Содержание темы: Параметрические, непараметрические и номинальные методы. Метод максимального правдоподобия: идея+ реализация. Критерии согласия. Дисперсионный анализ – однофакторный многофакторный. Проверка статистических гипотез на основе критериев: Колмагорова-Смирнова, Стьюдента, Манна-Уитни, Каскелла-Уоллиса. Сравнение двух групп независимых и зависимых выборок: критерий Стьюдента, критерий Тьюки. Корреляционный анализ (корреляции по Пирсону, Спирмену) .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические занятия, компьютерное моделирование.

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 5 Модели временных рядов.

Содержание темы: Временные ряды, формы представления данных для временных рядов Обобщенная модель временного ряда. Составляющие модели, методы моделирования. Тренды и периодические колебания Сглаживание временного ряда, фильтры. Исключение трендов. Строгие и эвристические методы – общие сведения. .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические занятия, компьютерное моделирование.

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 6 Модели временных рядов.

Содержание темы: Модели распределенных лагов – общие понятия. Стационарность временного ряда. Автокорреляционные функции. Единичные корни и коинтеграция. Тесты Дики-Фуллера. Общая методика построения модели временного ряда: модели авторегрессии – AR, модели скользящего среднего – MA. Модели ARMA, модели Бокса-Джекинса – ARIMA. Статистики Бокса, Льюинга, Акаике. Инфографика для задач временных рядов.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практические задания, компьютерное моделирование, метод активного обучения – «конференция».

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

Порядок выполнения практических заданий по теме «модели временные рядов»

1. Качественный анализ статистических данных – «гуманитарное описание»;
2. Оценка стационарности временного ряда.

2.1. Анализ графиков, ACF, PACF, трактовка результата.

2.2. Тестирование на единичные корни: DF, ADF. Трактовка результата.

2.3. Приведение временного ряда к стационарному виду: первая разность, вторая разность, дифференцирование. Обоснование выбора метода, трактовка результата.

3. Анализ трендов и периодических колебаний (нестационарной компоненты временного ряда)

3.1. Выявить устойчивость тенденции, тип колебаний.

3.2. Оценка необходимости выравнивания ряда. Выбор метода выравнивания: скользящего среднего, полиномиальная зависимость и т.п.;

3.3. Выбора типа тренда и обоснование выбора трендовой модели;

3.4. Построение модели, включающей периодические колебания – применение ряда Фурье, эвристические методы.

4. Моделирование стационарных временных рядов.

4.1. Построение AR модели. Оценка статистической значимости, обоснование возможных процедур оценки.

4.2. Построение MA модели. Оценка статистической значимости, обоснование возможных процедур оценки.

4.3. Построение смешанных моделей – ARMA. Применение алгоритма Бокса-Джекинса.

4.4. Статистики Бокса, Льюинга, Акаике. Формирование выводов о предпочтительной модели.

Методические рекомендации по организации самостоятельной работы

Перечень и тематика самостоятельных работ студентов по дисциплине

- Теоретическая подготовка к текущей аудиторной работе. Рекомендуется использовать материалы лекций, рекомендованную литературу и источники;
- Углубленное изучение инструментов Rstudio, алгоритмов математических пакетов Rstudio Рекомендуется примеры скриптов, документацию библиотек (пакетов), проанализировать приемы работы с Rstudio;
- Выполнение самостоятельного задания, доработка текущих аудиторных заданий, анализ результатов.
- Подготовка отчетов по текущим практическим заданиям.

Структура текущих отчетов (методы статистики)

Цель: формулировка цели текущей практической работы – 1-2 строки.

Ожидаемые результаты: краткое пояснение, что дает для исследования оценка характеристик выборок или применение теста, графика, модели – 2-3 строки.

Описание выборки: характер данных, способ получения данных, связь данных с изучаемым объектом, используемые шкалы и в т. ч. форматы представления, виды предварительной обработки (если использовались).

Источник данных: ссылка на источник

Числовые характеристики выборки: таблица с оценками числовых характеристик, скрин кода.

Графики, диаграммы. Пояснение выбора типа диаграммы.

Тест 1. Результаты: скрин программного кода, таблица с результатом, выводы по расчету с пояснениями и обоснованием дальнейших шагов.

……..

Тест N. Результаты: скрин программного кода, таблица с результатом, выводы по расчету с пояснениями и обоснованием дальнейших шагов.

Модель № k:

Общие характеристика модели. Формулы, проверяемые гипотезы, критерии, статистики.

Таблица с результатами оценки

variables	Parameters wish are associated with variables	p-value	comments

Общий вывод по полученным результатам

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Айвазян, С. А. Эконометрика - 2: продвинутый курс с приложениями в финансах : учебник / С. А. Айвазян, Д. Фантаццини. — Москва : Магистр : ИНФРА-М, 2024. — 944 с. - ISBN 978-5-9776-0333-1. - Текст : электронный. - URL: https://znanium.com/catalog/product/2121617 (Дата обращения - 05.09.2025)

2. Криволапов, С. Я. Анализ данных. Методы теории вероятностей и математической статистики на языке Python : учебное пособие / С.Я. Криволапов. — Москва : ИНФРА-М, 2025. — 678 с. — (Высшее образование). — DOI 10.12737/2034420. - ISBN 978-5-16-018616-0. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2034420 (дата обращения: 31.05.2026)

3. Миркин, Б. Г. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/583143 (дата обращения: 19.05.2026).

4. Модели и методы предиктивной аналитики данных динамики показателей для доверительного искусственного интеллекта : монография / Н. Г. Ярушкина, В. С. Мошкин, Г. Ю. Гуськов [и др.] , под редакцией А. А. Романова. — Ульяновск : УлГТУ, 2025. — 216 с. — ISBN 978-5-9795-2539-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/518770 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.2 Дополнительная литература

1. Демидова, О. А. Эконометрика : учебник и практикум для вузов / О. А. Демидова, Д. И. Малахов. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 398 с. — (Высшее образование). — ISBN 978-5-534-20392-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/583223 (дата обращения: 19.05.2026).

2. Кремер, Н. Ш. Математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. — 5-е изд. — Москва : Издательство Юрайт, 2026. — 259 с. — (Высшее образование). — ISBN 978-5-534-01654-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/583827 (дата обращения: 19.05.2026).

3. Кремер, Н. Ш. Эконометрика : учебник и практикум для вузов / Н. Ш. Кремер, Б. А. Путко ; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва : Издательство Юрайт, 2026. — 308 с. — (Высшее образование). — ISBN 978-5-534-08710-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/582582 (дата обращения: 19.05.2026).

4. Статистика : учебник для вузов / под редакцией В. С. Мхитаряна. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 503 с. — (Высшее образование). — ISBN 978-5-534-18687-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/589650 (дата обращения: 19.05.2026).

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Образовательная платформа "ЮРАЙТ"
2. Электронно-библиотечная система "ZNANIUM.COM"
3. Электронно-библиотечная система "ЛАНЬ"
4. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/

5. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/

6. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Мультимедийный проектор Casio XJ-V2
- Облачный монитор 23" LG CAV42K
- Облачный монитор LG Electronics черный +клавиатура+мышь
- П/К:С/бл IRU Corp, прцессор Intel Core, мат/пл intel Soc-GA1156, опер/память NCP DDR3, Видео intel HDA, Жесткий дискSeagate ST320DM000, Оптич. прив. Lite-On IHAS124, корпусIRU Corp ATX, блок.пит.LinkW
- Усилитель-распределитель VGA/XGA Kramer VP-200

Программное обеспечение:

- Microsoft Windows 10 Professional RUS Upgrd
- Python
- RStudio

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-1 : Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.1к : Применяет и развивает знания в области математики, естественнонаучных и социально-экономических наук в профессиональной деятельности
		ОПК-1.2к : Решает нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и профессиональных знаний

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Компетенция ОПК-1 «Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте»

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ОПК-1.1к : Применяет и развивает знания в области математики, естественнонаучных и социально-экономических наук в профессиональной деятельности	РД 3	Знание	базовых положений методик статистической проверки гипотез	Сформировавшееся систематическое знание базовых положений методик статистической проверки гипотез
	РД 3	Умение	проводить диагностику статистических моделей	Сформировавшееся систематическое умение проводить диагностику статистических моделей
	РД 3	Навык	работы со специальными статистическими пакетами и их инструментами	Сформировавшееся систематическое владение навыками работы со специальными статистическими пакетами и их инструментами
ОПК-1.2к : Решает нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально	РД 1	Знание	теоретических основ прикладной информатики	Сформировавшееся систематическое знание теоретических основ прикладной информатики
	РД 1	Уме	применять информационные технологии в исследовательской и проектной деятельности	Сформировавшееся систематическое знание теоретически

-экономических и профессиональных знаний		ние		х основ прикладной информатики
	РД 1	Навык	выбора инструментальных средств для аналитической работы с учетом требований проекта	Сформировавшееся систематическое владение навыками выбора инструментальных средств для аналитической работы с учетом требований проекта
	РД 2	Знание	методик организации эмпирических наблюдений, методов статистического оценивания параметров выборок	Сформировавшееся систематическое знание методик организации эмпирических наблюдений, методов статистического оценивания параметров выборо
	РД 2	Умение	применять и развивать методы прикладной статистики	Сформировавшееся систематическое умение применять и развивать методы прикладной статистики
	РД 2	Навык	внедрения и использования прикладного программного обеспечения, реализующего статистический анализ	Сформировавшееся систематическое владение навыками внедрения и использования прикладного программного обеспечения, реализующего статистический анализ

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очная форма обучения				
РД1	Знание : теоретических основ прикладной информатики	1.1. Данные: общие сведения. Работа с данными.	Собеседование	Собеседование
РД1	Умение : применять информационные технологии в исследовательской и проектной деятельности	1.1. Данные: общие сведения. Работа с данными.	Практическая работа	Собеседование
		1.2. Математическая статистика: общие сведения.	Практическая работа	Собеседование
РД1	Навык : выбора инструментальных средств для аналитической работы с учетом требований проекта	1.1. Данные: общие сведения. Работа с данными.	Собеседование	Собеседование
РД2	Знание : методик организации эмпирических наблюдений, методов статистического оценивания параметров выборок	1.2. Математическая статистика: общие сведения.	Контрольная работа	Контрольная работа
		1.3. Выборочные характеристики и их свойства.	Контрольная работа	Контрольная работа

РД2	Умение : применять и развивать методы прикладной статистики	1.4. Сведения о статистическом оценивании параметров. Проверка гипотез	Практическая работа	Контрольная работа
		1.5. Модели временных рядов.	Практическая работа	Контрольная работа
РД2	Навык : внедрения и использования прикладного программного обеспечения, реализующего статистический анализ	1.3. Выборочные характеристики и их свойства.	Практическая работа	Контрольная работа
		1.5. Модели временных рядов.	Практическая работа	Контрольная работа
		1.6. Модели временных рядов	Практическая работа	Контрольная работа
РД3	Знание : базовых положений методик статистической проверки гипотез	1.3. Выборочные характеристики и их свойства.	Практическая работа	Контрольная работа
		1.4. Сведения о статистическом оценивании параметров. Проверка гипотез	Практическая работа	Контрольная работа
РД3	Умение : проводить диагностику статистических моделей	1.5. Модели временных рядов.	Контрольная работа	Контрольная работа
		1.6. Модели временных рядов	Контрольная работа	Контрольная работа
РД3	Навык : работы со специальными статистическими пакетами и их инструментами	1.5. Модели временных рядов.	Контрольная работа	Контрольная работа
		1.6. Модели временных рядов	Контрольная работа	Контрольная работа

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Вид учебной деятельности	Оценочное средство					
	Практические задания 1-4	Практические задания 5-10	Практические задания 11	Практические задания 12	Собеседование Перечень заданий для контрольной работы	Итого
Лекции	2	2	3	3		10
Практические занятия	15	15	20	10		60
Самостоятельная работа	3	3	2	2		10
ЭОС						
Промежуточная аттестация					20	20
Итого	20	20	25	15	20	100

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Примеры заданий для выполнения практических работ

Задание 1. Сбор данных предлагаемых источников в сети Интернет (Открытые данные РФ, Государственные органы статистического надзора). Хранение данных в программе R-Studio (python).

Задание 2. Работа с данными: определение размерности, выбор отдельных атрибутов, преобразование данных. Инфографика.

Задание 3. Статистическая обработка данных: подсчет описательных статистик. Законы распределения: генерация последовательностей, графическое представление.

Краткие методические указания

Собеседование проводится по вопросам выбранной темы . Преподаватель в устной форме задает вопросы студентам.

Шкала оценки

№	Баллы	Описание
5	19–20	Процент правильных и обоснованных ответов от 95% до 100%
4	16–18	Процент правильных и обоснованных ответов от 80 до 94%
3	13–15	Процент правильных ответов с помощью наводящих вопросов от 65 до 79%
2	9–12	Процент правильных ответов от 45 до 64%
1	0–8	Процент правильных ответов менее 45%

5.2 Примеры заданий для выполнения контрольных работ

Задание 9. Статистический анализ динамики.

Провести тестирование временного ряда на стационарность. Если гипотеза о стационарности отвергается, построить модели тренда и периодических колебаний.

Порядок выполнения работы:

1. Загрузить файл с динамическими рядами (динамика цен, социально-экономических показателей).
2. Сформировать структуру данных, соответствующую методу исследования и удобную для представления результатов.

3. Выбрать методы (тесты) для анализа устойчивости тенденции / стационарности временного ряда.

4. Провести тестирование стационарности временного ряда.

5. Для динамических рядов с устойчивыми трендами, построить соответствующие модели.

Краткие методические указания

Каждый студент по жребию выбирает данные для выполнения заданий 1 и 2. Время выполнения 60 минут. Невыполненная в установленный срок работа не оценивается. Работа сдается исключительно в электронном виде, с пояснительной запиской в произвольной форме

В состав материалов для проверки входят: исходный данные форматах табличного процессора или текстовом, скрипты, краткая пояснительная записка.

Шкала оценки

№	Баллы	Описание
5	9-10	Студент выполнил контрольную работу в соответствии с требованиями, демонстрируе т способность применять методы анализа и моделирования временных рядов, адекват ных задаче
4	7-8	Студент демонстрирует умения на среднем уровне. Демонстрирует способность демон стрирует способность применять методы анализа и моделирования временных рядов, адекватных задаче
3	5-6	Студент демонстрирует умения и навыки на базовом уровне.
2	0-4	Студент демонстрирует умения и навыки на уровне ниже базового.

5.3 Примеры заданий для выполнения контрольных работ

Задание 10

Качественный анализ статистических данных – «гуманитарное описание».

Оценка стационарности временного ряда.

Анализ графиков, расчеты ACF, PACF.

Тестирование на единичные корни.

Приведение временного ряда к стационарному виду.

Анализ трендов и периодических колебаний (нестационарной компоненты временного ряда)

Оценка устойчивость тенденции, типа колебаний.

Оценка необходимости выравнивания ряда. Выбор метода выравнивания: скользящего среднего, полиномиальная зависимость и т.п.;

Построение моделей тренда и моделей периодических колебаний.

Моделирование стационарных временных рядов.

Построение AR-моделей, MA-модели. Оценка статистической значимости, обоснование возможных процедур оценки.

Построение смешанных моделей – ARMA. Применение алгоритма Бокса-Джекинса. Оценка статистической значимости.

Выбор показателей оценки статистической значимости и сравнение мощности критериев.

Особенности оценивания динамических моделей

Краткие методические указания

На выполнение практического задания отводится не более четырех академических часов (включая затраты времени на СРС). После выполнения практического задания студент должен представить результаты выполнения в отчете, в соответствии со структурой представленной в РПД

Шкала оценки

№	Баллы	Описание
5	20-25	Студент демонстрирует умения на итоговом уровне: умеет свободно выполнять практ ические задания, предусмотренные программой, свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.

4	15-19	Студент демонстрирует умения на среднем уровне: освоил основные умения, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.
3	10-14	Студент демонстрирует умения и навыки на базовом уровне: допускаются значительные ошибки, проявляется отсутствие отдельных умений, навыков по дисциплинарной компетенции, испытываются значительные затруднения при оперировании умениями и при их переносе на новые ситуации.
2	6-9	Студент демонстрирует умения и навыки на уровне ниже базового: проявляется недостаточность умений и навыков.
1	0-5	Студентом проявляется полное или практически полное отсутствие умений и навыков, но присутствует на занятии и пытается выполнить задание.

5.4 Примеры заданий для выполнения практических работ

Задание 4. Проверка гипотез на нормальность: графический способ, критерии Шапиро-Уилка, тест Колмогорова-Смирнова в модификации Лиллиефорса, критерий хи-квадрат Пирсона.

Задание 5. Проведение корреляционного анализа. Подсчет коэффициентов корреляции Пирсона, Спирмена, Кендела.

Задание 6. Проверка гипотез: критерии Колмагорова-Смирнова, Стьюдента, Фишера, Тьюки.

Задание 7. Параметрический дисперсионный анализ: однофакторный, двухфакторный, многомерный.

Задание 8. Непараметрический анализ данных: критерий Манна-Уитни, двухвыборочный критерий Колмогорова-Смирнова, ранговый дисперсионный анализ Краскела-Уоллиса.

Краткие методические указания

Собеседование проводится по вопросам выбранной темы

Шкала оценки

№	Баллы	Описание
5	19–20	Процент правильных и обоснованных ответов от 95% до 100%
4	16–18	Процент правильных и обоснованных ответов от 80 до 94%
3	13–15	Процент правильных ответов с помощью наводящих вопросов от 65 до 79%
2	9–12	Процент правильных ответов от 45 до 64%
1	0–8	Процент правильных ответов менее 45%

5.5 Примерный перечень вопросов по темам

Контрольные вопросы для собеседования

Тема 1 Данные: общие сведения. Работа с данными.

1. Перечислите основные типы данных.
2. Дайте характеристику основным типам данных.
3. Определите разницу между временными рядами, объектами, набором объектов, панельными данными.
4. Для чего проводят анализ данных?

Тема 2. Выборочные характеристики и их свойства

1. Дайте определения основным описательным статистикам.
2. Перечислите характеристики и свойства нормального распределения.
3. Проверка статистических гипотез: этапы проведения, трактовка результата.
4. Какие критерии используются для проверки выборки на нормальность распределения?
5. Для чего используют масштабирование и центрирование данных?
6. Логарифмирование переменной: зачем это нужно и как интерпретировать.

1. Перечислите способы преобразования данных к нормальному виду распределению.
2. Для чего проводят первичную обработку данных?

Тема 3. Статистическое оценивание параметров выборок. Проверка гипотез.

1. Определите различия между параметрическими, непараметрическими и номинальными методами.
2. Основная идея метода максимального правдоподобия.
3. Основная идея корреляционного анализа.
4. Перечислите коэффициенты, определяющие степень зависимости между переменными.
5. Что показывает коэффициент корреляции Пирсона?
6. Какая гипотеза проверяется в критерии Колмагорова-Смирнова?
    1. Для чего используют критерий Стьюдента?
    2. Критерий Тьюки: в каких случаях используется, какая гипотеза проверяется.
    3. Основная идея дисперсионного анализа.
    4. Отличия многофакторного дисперсионного анализа от многомерного.
    5. Какие критерии относятся к непараметрическим методам анализа данных? Особенности их применения.
    6. В каких случаях применяют ранговый дисперсионный анализ Краскела-Уоллиса?

Краткие методические указания

Собеседование проводится по вопросам выбранной темы

Шкала оценки

№	Баллы	Описание
5	9-10	Процент правильных и обоснованных ответов от 95% до 100%
4	7-8	Процент правильных и обоснованных ответов от 80 до 94%
3	5-7	Процент правильных ответов с помощью наводящих вопросов от 65 до 79%
2	3-4	Процент правильных ответов от 45 до 64%
1	0–2	Процент правильных ответов менее 45%

Ключи для ФОС

Для дисциплины т «Методы статистического анализа и прогнозирования»

5.1 Примеры заданий для выполнения практических работ

Задание 2. Работа с данными: определение размерности, выбор отдельных атрибутов, преобразование данных. Инфографика

I. Обработка пропущенных значений.

Откройте встроенную таблицу данных mtcars (данные по автомобилям, размерность таблицы - 32×11). С помощью функции sum подсчитайте общее количество пропущенных значений.

Пример:

```
sum(is.na(trees))
```

Добавьте пропущенные значения в столбец mtcars$mpg в 2, 4 и 10 строках.

Пример:

```
dat$Girth[c(2,4,10)]=NA # назначим пропущенные значения во 2, 4 и 10 строках
sum(is.na(dat)) # подсчитать общее количество пропущенных значений
which(is.na(dat)) # найти местонахождение NA
dat$Girth[c(2,4,10)] # выведем значения в заданных строках
```

Удаление пропущенных значений. Удалите строки со значениями NA из таблицы с помощью функции na.omit {stats}.

Пример:

```
dat1=na.omit(dat) #
```

Если много пропущенных значений в столбце таблицы, можно удалить этот столбец.

Пример:

```
dat2=dat[,-1]
```

Замена пропущенных значений. Замените пропущенное значение во 2-ой строке средним значением, в 4-ой строке – медианой.

Пример:

```
dat$Girth[c(2,4,10)]=NA
dat$Girth[is.na(dat$Girth)] <- mean(dat$Girth, na.rm =T) # замена средним значением
dat$Girth[is.na(dat$Girth)]<-median(dat$Girth, na.rm=T)# замена медианным значением
dat$Girth [is.na(dat$Girth)] <-13# замена заданным числом
```

Сравните:

```
summary(trees$Girth)
summary(G1)
summary(G2)
summary(G3)
```

Прогнозирование пропущенных значений с помощью функции mice{mice}.

```
install.packages("mice")# загрузка пакета
library(mise)# активизация пакета
```

Пример:

```
m=mice(dat)# прогнозирование
complete(m)# выходные данные
```

II. Выбросы

Обнаружение выбросов и аномалий всегда актуально для исследователя. Если решаются задачи прогнозирования, то удаление нетипичных значений, как правило, повышает точность предсказаний, поскольку данные без аномалий представляют собой типичный объект. Кроме того, статистические характеристики чувствительны к наличию выбросов.

Поскольку множество статистических методов «буксуют» на выборках с выбросами, выбросы приходится обнаруживать (желательно автоматически) и исключать из выборки. Простейшие способы основаны на межквартильном расстоянии — например, всё, что не попадает в диапазон считается выбросами:

$$[(x_{25} - 1{,}5 \cdot (x_{75} - x_{25})),\ (x_{75} + 1{,}5 \cdot (x_{75} - x_{25}))],$$

Другой простой способ обнаружить выбросы – использовать функцию boxplot. Рассмотрим на примере.

Допустим у нас есть следующие данные:

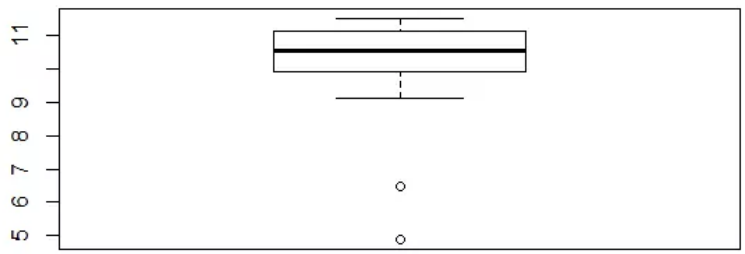
```
x <- c(
  4.27296076175273, 4.3253044335640, 4.55202839528403,
  4.33285651824668, 4.0128340310945, 4.07155452370293,
  4.04113475664987, 2.7753693269563, 2.49186430883914,
  2.39823431758359, 2.3789955162936, 2.14825292752989,
  2.19583315284264, 2.1947104626036, 2.14216783486566,
  2.28399128121205, 2.3048696257819, 2.23703873535593,
  2.28669486582313, 2.4163169738317, 2.03782758779637,
  2.61237703056071, 3.47620332697605)
y <- c(
  9.097731,9.512591,9.740439,9.910364,9.865059,9.935519,
  9.972640,9.920197,10.367693,10.680861,10.999012,11.246248,
  10.532816,11.144033,11.261961,11.400160,11.080695,11.173922,
  11.506877,11.105483,10.807685,6.489205,4.882802)
```

Выведите графическое отображение данных:

```
plot(x,y)
```

Скорее всего здесь есть выбросы. Если эти данные моделировать линейной регрессией, то эти точки очень сильно исказят результаты предсказания для множества остальных точек. Проверьте на выбросы:

```
boxplot(y)
```

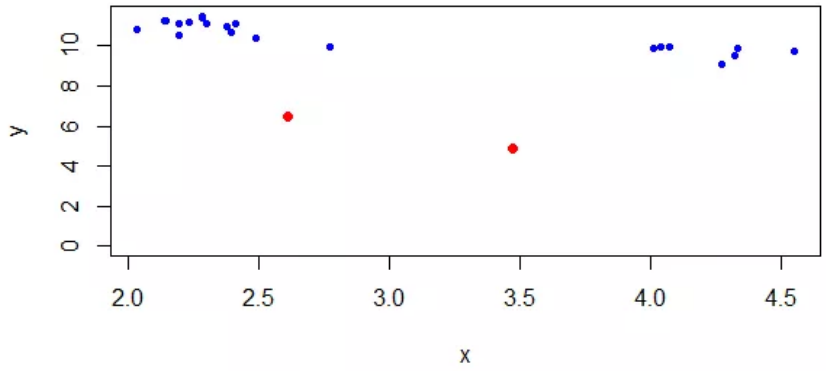


Присутствуют две точки выброса.

Самый простой способ – удалить эти точки. Функция boxplot не только рисует график, но и сохраняет все его параметры в объекте.

Теперь проверьте - те ли вообще точки нашли?

```
plot(x,y,col='blue', pch=20, ylim=c(0,max(y)))
points(outler$x, outler$y, col='red',pch=19)
```



Дальше очистите данные от выбросов, проверьте, что статистических выбросов не наблюдается и посмотрите на новые "чистые" данные.

```
x <- x[-ind]
y <- y[-ind]
```

III. Реализация критериев определения выбросов в статистическом пакете R

Тест Граббса — это статистический тест, который можно использовать для выявления наличия выбросов в наборе данных. Чтобы использовать этот тест, набор данных должен иметь примерно нормальное распределение и иметь не менее 7 наблюдений.

Для выполнения теста Граббса в R используется функция grubbs.test{ outliers}. Нулевая гипотеза: в данных нет выбросов.

Пример:

```
library(outliers)
data <- c(5, 14, 15, 15, 14, 13, 19, 17, 16, 20, 22, 8, 21, 28, 11, 9, 29, 40)
grubbs.test(data)
```

```
        Grubbs test for one outlier

data:  data
G = 2.65990, U = 0.55935, p-value = 0.02398
alternative hypothesis: highest value 40 is an outlier
```

Так как, полученное значение p-value = 0.02398<0.05, нулевую гипотезу отвергаем, выбросы присутствуют в данных. Делаем вывод, что максимальное значение 40 является выбросом.

Если бы вместо этого мы хотели проверить, является ли наименьшее значение «5» выбросом, то:

```
grubbs.test(data, opposite= TRUE )
```

Grubbs test for one outlier

data: data
G = 1.4879, U = 0.8621, p-value = 1
alternative hypothesis: lowest value 5 is an outlier

Нулевую гипотезу принимаем. Минимальное значение «5» не является выбросом.
Предположим, что у нас есть два больших значения на одном конце набора данных: 40 и 42. Чтобы проверить, являются ли оба этих значения выбросами, мы могли бы выполнить тест Граббса и указать, что type=20:
data <- c(5, 14, 15, 15, 14, 13, 19, 17, 16, 20, 22, 8, 21, 28, 11, 9, 29, 40, 42)
grubbs.test(data, type=20)

Grubbs test for two outliers

data: data
U = 0.38111, p-value = 0.01195
alternative hypothesis: highest values 40, 42 are outliers
Значение p теста составляет 0.01195. Поскольку это меньше 0,05, отклоняем нулевую гипотезу и делаем вывод, что значения 40 и 42 являются выбросами.

Иногда значения, которые отображаются как выбросы в наборах данных, являются просто опечатками, сделанными человеком при вводе данных. Вернитесь и убедитесь, что значение было введено правильно, прежде чем принимать какие-либо дальнейшие решения.
Если выброс окажется результатом опечатки или ошибки ввода данных, вы можете решить присвоить ему новое значение, например, среднее значение или медиану набора данных.
Если значение является истинным выбросом, вы можете удалить его, если оно окажет значительное влияние на общий анализ.

Фильтр Хэмпеля
Другой метод, известный как фильтр Хэмпеля, заключается в том, чтобы рассматривать как выбросы значения вне интервала, которые формируются медианным значением плюс-минус 3 медианы абсолютных отклонений (MAD):
I = [median - 3 * MAD; median + 3 * MAD]
В которых MAD — это медианное абсолютное отклонение и определяется как медиана абсолютных отклонений от медианы данных:

$$MAD = median(|X_i - \tilde{X}|)$$

Для этого метода сначала устанавливаем пределы интервала с помощью функций median() и mad() на примере:
x=rnorm(seq(1,500),23,1)
boxplot(x)
summary(x)
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.

19.81 22.36 23.04 23.04 23.72 25.96

lower_bound <- median(x) - 3 * mad(x, constant=1)
lower_bound
[1] 21.02341
upper_bound <- median(x) + 3 * mad(x, constant=1)
upper_bound [1] 25.06453

Все наблюдения меньше 21.02343 и больше 25.06453 будут рассматриваться как потенциальные выбросы. Номера строк наблюдений за пределами интервала затем могут быть извлечены с помощью функции which(). Согласно фильтру Хэмпеля, для переменной x есть 21 потенциальных выброса.

Задание 3. **Статистическая обработка данных: подсчет описательных статистик. Законы распределения: генерация последовательностей, графическое представление.**

1. Загрузите таблицу данных (datasets) NMES1988 с сайта:
https://vincentarelbundock.github.io/Rdatasets/datasets.html
2. Откройте таблицу данных в программе RStudio.

Пример:
data=read.csv("C:/Мои документы/Марина/ NMES1988.csv", sep=",", dec=".")

Фрейм данных содержит 4406 наблюдений по 19 переменным. Используем для анализа только следующие переменные:

visits - Количество посещений кабинета врача.
health - Фактор, указывающий на самооценку состояния здоровья; уровни "poor", "average"(эталонная категория), "excellent".
chronic - Количество хронических состояний.
adl - Фактор, указывающий, есть ли у человека состояние, ограничивающее повседневную деятельность ("limited") или нет ("normal").
region - Фактор - область проживания; уровни: northeast, midwest, west, other.
age - Возраст в годах (разделенный на 10).
gender - Фактор, указывающий на пол.
married – Фактор: человек женат?
school - Количество лет обучения.
income - доход в 10000 долларов США.
employed – Фактор: работает ли человек?
insurance – Фактор: застраховано ли физическое лицо частной страховкой?

1. Используйте функцию describe{psych} для расчета описательной статистики показателя visits, предварительно загрузив пакет.

Пример:
install.packages("psych") #загрузка пакета
library("psych") #активизация пакета

describe(mtcars$disp) # расчет описательных статистик

2. Рассчитайте значения коэффициентов крутости и асимметрии для показателя visits.

Пример:
install.packages("moments") # загрузка пакета
library(moments) # активизация пакета
kurtosis(mtcars$mpg) # коэффициентов крутости или эксцесс
skewness(mtcars$mpg) # коэффициентов асимметрии

3. Проверьте количественные данные на нормальность распределения (функция shapiro.test()).

   Пример:
shapiro.test(trees$Girth) # для одного показателя
sapply(trees,shapiro.test) # для таблицы

4. Постройте гистограмму показателя visits с кривой плотности распределения.

   Пример:
   X <- rnorm(n = 50, mean = 15, sd = 5)# создали нормально распределенную выборку
   hist(X, breaks = 20, freq = FALSE, col = "lightblue",
xlab = "Переменная X", ylab = "Плотность вероятности",
main = "Гистограмма, совмещенная с кривой плотности")
lines(density(X), col = "red", lwd = 2)

5. Рассчитайте количество мужчин и женщин в выборке. Рассчитайте среднее значение (mean) и стандартную ошибку среднего для показателя visits.

SE= sd(x)/ sqrt (length(x)),
т.е., стандартная ошибка среднего есть стандартное отклонение (sd), деленное на корень квадратный из количества элементов в выборке.

6. Рассчитайте медианное значение и квартили для показателя visits (summary).
7. Рассчитайте медианные значения и квартили для количественных данных отдельно для мужчин и женщин.

   Пример:
   tapply(data$visits, INDEX = data$gender, FUN= summary)

8. Постройте диаграмму размаха количества посещений по полу.

   boxplot(data$visits~data$gender, main="visits")

Диаграммы размахов, или "ящики с усами", получили свое название за характерный вид: точку или линию, соответствующую медиане или средней арифметической, окружает прямоугольник ("ящик"), длина которого соответствует одному из показателей разброса или точности оценки генерального параметра. Дополнительно от этого прямоугольника отходят "усы", также соответствующие по длине одному из показателей разброса или точности. Графики этого типа очень популярны, поскольку позволяют дать очень полную статистическую характеристику анализируемой совокупности. Кроме того, диаграммы размаха можно использовать для визуальной экспресс-оценки разницы между двумя и более

группами (например, между датами отбора проб, экспериментальными группами, участками пространства, и т.п.).
Верхний "ус" простирается от верхней границы "ящика" до наибольшего выборочного значения, находящегося в пределах расстояния 1.5*ИКР от этой границы. Аналогично, нижний "ус" простирается от нижней границы "ящика" до наименьшего выборочного значения, находящегося в пределах расстояния 1.5*ИКР от этой границы. Длину данного интервала (т.е. 1.5*ИКР) можно изменить при помощи аргумента range функции boxplot(). Наблюдения, находящиеся за пределами "усов", потенциально могут быть выбросами. Однако всегда следует внимательно относиться к такому рода нестандартным наблюдениям - они вполне могут оказаться "нормальными" для исследуемой совокупности, и поэтому не должны удаляться из анализа без дополнительного расследования причин их появления.

9. Рассчитайте медианные значения показателя visits по двум факторам, один из которых пол (gender).

   Пример:
   tapply(data$visits, INDEX = list(data$gender,data$health), FUN= median)

10. Создайте столбиковые диаграммы на основе медианных значений показателей visits по gender и region.

    Примеры:
    medians = tapply(data$visits, list(data$gender,data$insurance), median)
    barplot(medians, beside = TRUE, col = topo.colors(2),
legend.text = rownames(means), xlab = "Страховка", ylab="Медиана", main= "Кол-ва посещений")
Рассмотрим пример построения диаграммы, где столбцы сгруппированы в соответствии с уровнями какого-либо фактора.
barplot(medians, beside = FALSE, col = topo.colors(4), xlab = "Страховка", legend.text = rownames(means),main="Медиана посещений")# столбиковая диаграмма с накоплением
Можно добавить «усы» на график:
se=tapply(data$visits, list(data$gender,data$insurance), SE);se
b=barplot(means, beside = TRUE, col = topo.colors(2),ylim=c(0,10),
legend.text = rownames(means), xlab = "Страховка", main= "Среднее кол-во посещений")
arrows(b, means+se, b, means-se, angle = 90, code = 3, length = 0.05)

11. Рассчитайте медианное значение возраста по категориям показателя married по всей выборке и отдельно для мужчин и женщин. Визуализируйте результат с помощью функции barplot.

Задание 4. Применение статистических критериев

1. Откройте таблицу данных Uni.csv (результаты по анализу крови у мужчин и женщин, проживающих в 4-х районах РФ).

Uni=read.csv("C:/Мои документы/Uni.csv", sep=";", dec=",")
View(Uni)

2. В таблице Uni удалите строки с пропущенными значениями (функция na.omit).
3. С помощью критерия Шапиро-Уилка (функция shapiro.test) проверьте на нормальность распределения показатель TTG.
4. С помощью критерия Колмогорова-Смирнова (ks.test) проверьте принадлежность показателя Age к нормальному виду распределения.
5. Для группы показателей Ferritin, VitD, B12, Kreatinin, Cholesterin проверку на нормальность проведите следующим образом:

sapply(data.frame(Uni$Ferritin,Uni$VitD,Uni$B12,Uni$Kreatinin,Uni$Cholesterin), shapiro.test)

6. Постройте гистограммы с наложением кривой плотности распределения для количественных показателей.
7. Рассчитайте медианные значения и соответствующие доверительные интервалы $[Q_1,Q_3]$ для всех количественных показателей по всей выборке.
8. Рассчитайте медианные значения количественных показателей по категориям показателей Sex и District (по двум факторам одновременно).
9. Постройте столбиковые диаграммы для количественных показателей по категориям показателей Sex и District (по двум факторам одновременно).
10. Преобразуйте показатель Age в категориальный согласно классификации ВОЗ:

18-44 – молодость;
45-59 – средний возраст;
60-74 – пожилой человек;
75-90 – старческий возраст;
90+ – долгожитель.

Пример:
Age1=cut(x, breaks=c(0,10,50), labels=c(“A”,”B”))
Вектор значений от 0 до 50 разбиваем на два интервала, присваиваем имя A и B.

11. Рассчитайте количество человек в каждой возрастной группе, сколько среди них мужчин и женщин. Визуализируйте результаты.
12. Создайте вектор со средними значениями продолжительности жизни по республикам Кавказа (включая Адыгею):
    rosstat <-c(79.42, 75.83, 74.16, 73.91, 73.82, 73.06, 72.01)
13. Проверьте вектор на нормальность распределения.
14. Запустите одновыборочный t-тест, указав в параметре *mu* - среднее значение ожидаемого возраста жизни по России равное 70.93. Опишите полученные результаты.
15. Нужно узнать, отличается ли урожайность картофеля на севере и на юге региона. Для этого, собрали данные с 40 фермерских хозяйств: 20 из которых располагались на севере и сформировали выборку "North", а остальные 20 - на юге, сформировав выборку "South".

North <- c(122, 150, 136, 129, 169, 158, 132, 162, 143, 179, 139, 193, 155, 160, 165, 149, 173, 173, 141, 166)
South <- c(170, 163, 178, 150, 166, 142, 157, 149, 151, 164, 163, 161, 159, 139, 180, 155, 144, 139, 151, 160)

16. Проверьте данные на нормальность распределения.

17. Проведите двувыборочный t-тест для сравнения двух независимых выборок North и South.

Опишите полученные результаты.

18. Объедините выборки в таблицу и постройте диаграмму размаха.
19. Для проверки когнитивных способностей группе людей были выданы логические задания до и после прохождения курса обучения. Нужно с помощью критерия Стьюдента для зависимых выборок проверить значимость различий результатов (время выполнения заданий).

do=c(25,23,28,29,35,31,24,24,38,26,20)
posle=c(22,25,23,22,30,27,20,19,32,25,19)
Опишите полученные результаты.

5.2 Примеры заданий для выполнения контрольных работ

Задание 9. Статистический анализ динамики.
Анализ экономической динамики: исследовать динамику цен на продукты питания: молоко, растительное масло и чёрный чай в РФ. Провести анализ тенденции, устойчивости и стационарности, построить адекватную модель.

3.1 Корреляция
- Молоко: Спирмен: 0.999, Кендалл: 0.995

> cor.test(year, milk, method = "spearman")

Spearman's rank correlation rho
data: year and milk
S = 2, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9994527

> cor.test(year, milk, method = "kendall")

Kendall's rank correlation tau

data: year and milk
T = 377, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.994709

- Масло: Спирмен: 0.975, Кендалл: 0.884
> cor.test(year, oil, method = "spearman")
Spearman's rank correlation rho

data: year and oil
S = 90, p-value = 8.66e-09
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9753695

> cor.test(year, oil, method = "kendall")
Kendall's rank correlation tau

data: year and oil
T = 356, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.8835979

- Чёрный чай: Спирмен: 0.989, Кендалл: 0.963
> cor.test(year, tea, method = "spearman")

Spearman's rank correlation rho

data: year and tea
S = 40, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9890531

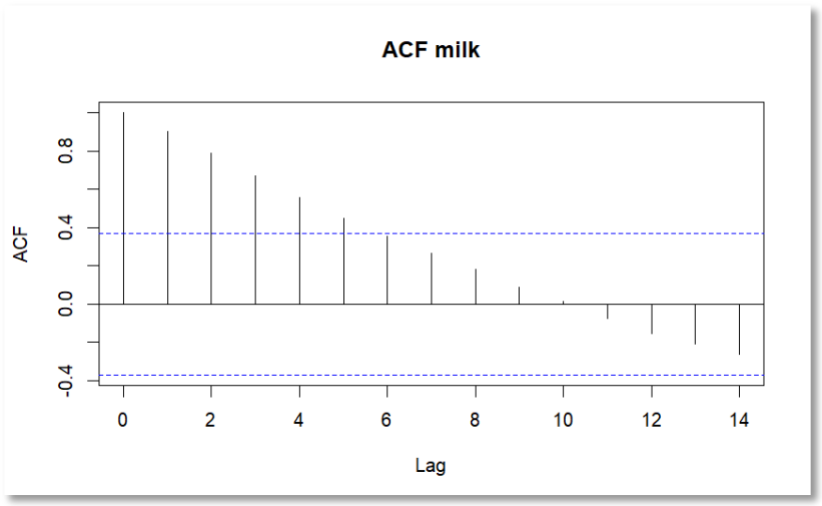
> cor.test(year, tea, method = "kendall")
Kendall's rank correlation tau

data: year and tea
T = 371, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.962963

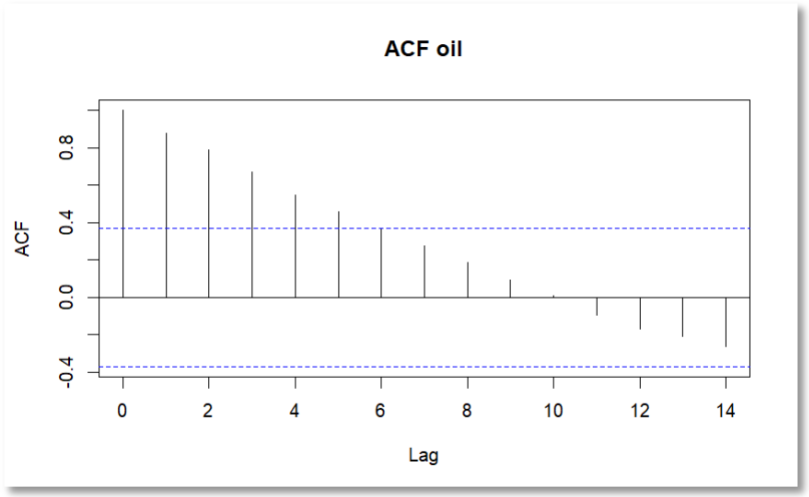
Все три продукта демонстрируют устойчивый рост цен по времени.

1. **ACF (Автокорреляционная функция)**

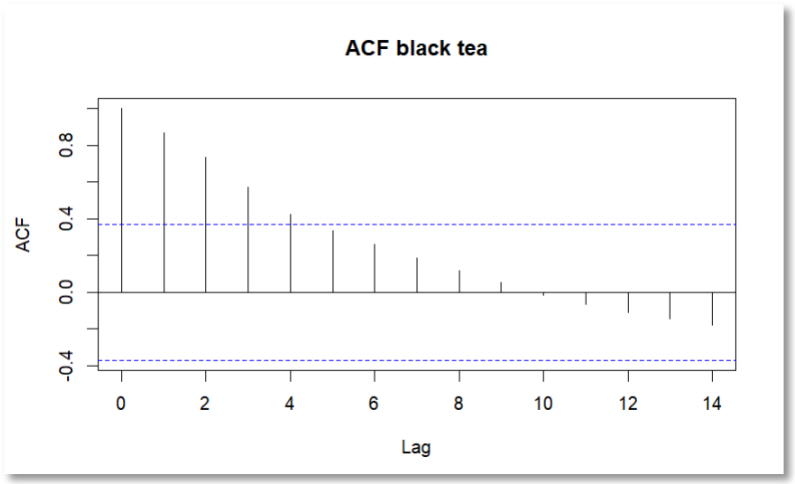
acf(milk, lag.max = 14, main = "ACF milk")



acf(oil, lag.max = 14, main = "ACF oil")



acf(tea, lag.max = 14, main = "ACF black tea")



ACF для всех трёх рядов показывает медленное убывание автокорреляций, что указывает на наличие тренда и нестационарность.

2. **Линейные модели**

Молоко: $R^2 = 0.956$, рост ~2.12 руб/год

> milk_model <- lm(milk ~ year)
> summary(milk_model)

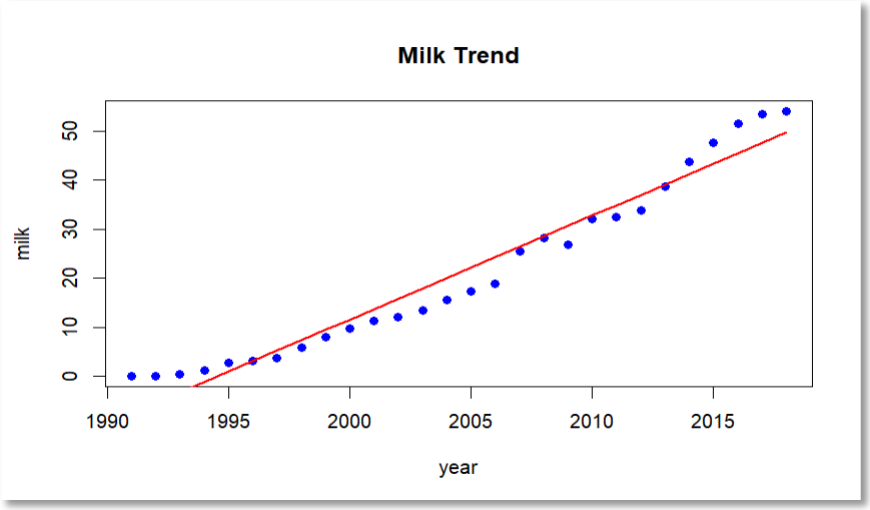
Call:
lm(formula = milk ~ year)

Residuals:
Min 1Q Median 3Q Max
-5.5185 -2.5485 -0.8885 2.8278 7.5223
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.228e+03 1.787e+02 -23.66 <2e-16 ***
year 2.120e+00 8.916e-02 23.78 <2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.811 on 26 degrees of freedom
Multiple R-squared: 0.956, Adjusted R-squared: 0.9543
F-statistic: 565.4 on 1 and 26 DF, p-value: < 2.2e-16

График трендов
> plot(year, milk, col = "blue", pch = 16, main = "Milk Trend")
> lines(year, milk_model$fitted.values, col = "red", lwd = 2)



Масло: $R^2 = 0.936$, рост ~4.01 руб/год

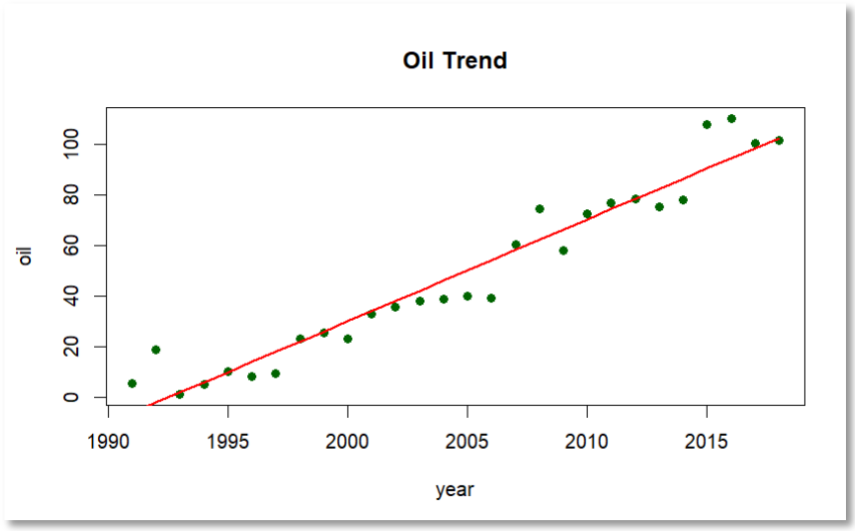
> oil_model <- lm(oil ~ year)
> summary(oil_model)

Call:
lm(formula = oil ~ year)

Residuals:
```
    Min      1Q  Median      3Q     Max
-14.8629  -6.9247  -0.7562   2.0468  20.9298
```
Coefficients:
```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7998.775    413.002  -19.37   <2e-16 ***
year            4.014      0.206   19.48   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 8.807 on 26 degrees of freedom
Multiple R-squared: 0.9359, Adjusted R-squared: 0.9334
F-statistic: 379.6 on 1 and 26 DF, p-value: < 2.2e-16

График трендов
```
> plot(year, oil, col = "darkgreen", pch = 16, main = "Oil Trend")
> lines(year, oil_model$fitted.values, col = "red", lwd = 2)
```



Чёрный чай: $R^2 = 0.853$, рост ~26.74 руб/год

3. Линейные модели
```
> tea_model <- lm(tea ~ year)
> summary(tea_model)
```

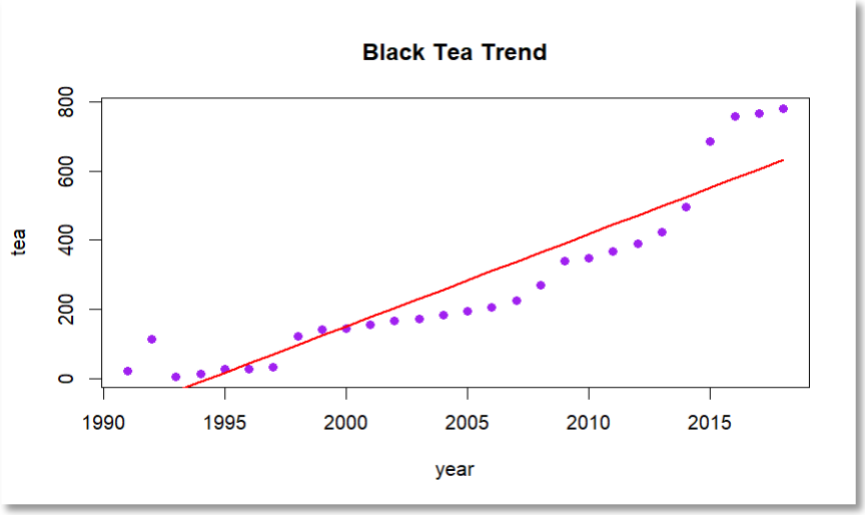
Call:
lm(formula = tea ~ year)
Residuals:
```
    Min      1Q  Median      3Q     Max
-112.76  -74.44  -24.62   29.40  181.13
```

Coefficients:
```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -53331.588   4367.957  -12.21 2.85e-12 ***
```

year 26.741 2.179 12.27 2.55e-12 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 93.14 on 26 degrees of freedom
Multiple R-squared: 0.8528, Adjusted R-squared: 0.8471
F-statistic: 150.6 on 1 and 26 DF, p-value: 2.547e-12

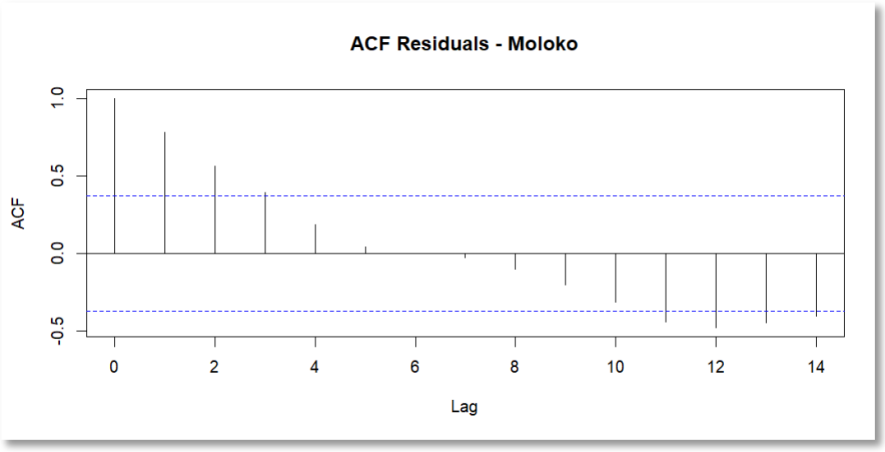
> # График трендов
> plot(year, tea, col = "purple", pch = 16, main = "Black Tea Trend")
> lines(year, tea_model$fitted.values, col = "red", lwd = 2)



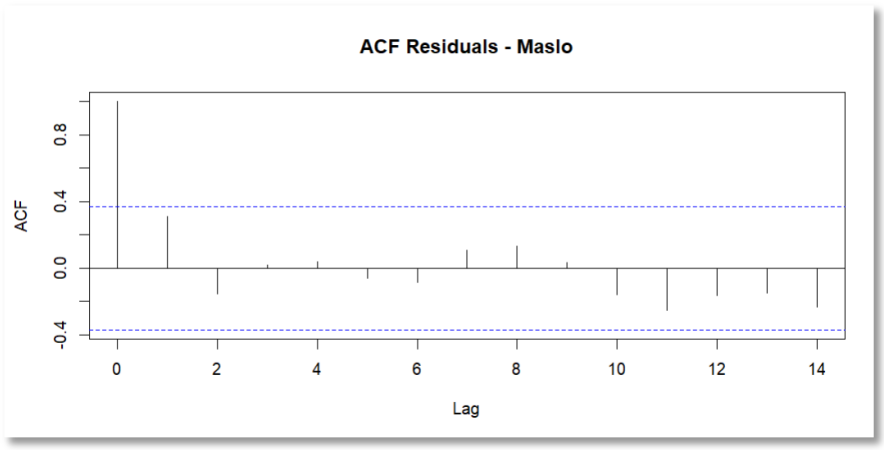
Все модели значимы, коэффициенты интерпретируемы, тенденция устойчива.

3. ACF остатков

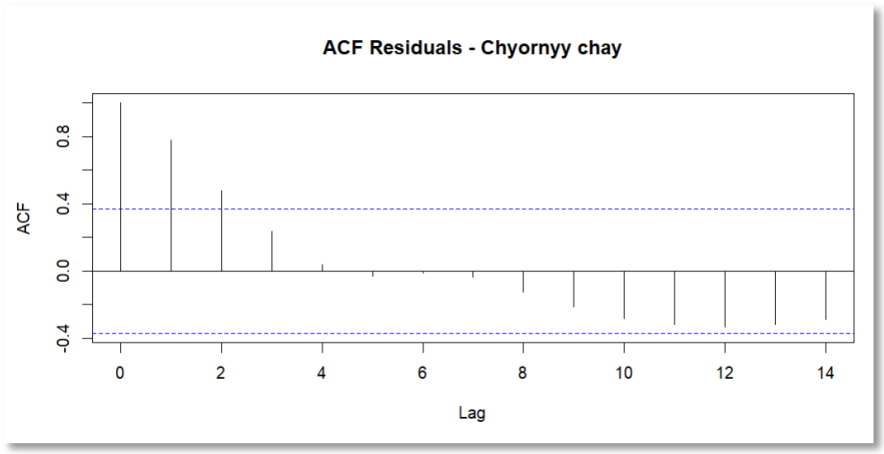
> acf(milk_model$residuals, lag.max = 14, main = "ACF Residuals - Milk")



> acf(oil_model$residuals, lag.max = 14, main = "ACF Residuals - Oil")



> acf(tea_model$residuals, lag.max = 14, main = "ACF Residuals - Black Tea")



Графики ACF остатков показывают отсутствие значимых автокорреляций, что подтверждает адекватность моделей.

4. Тест Дики-Фуллера (ADF)

- Молоко: p-value = 0.987
> adf.test(milk, alternative = "stationary")

Augmented Dickey-Fuller Test
data: milk
Dickey-Fuller = -1.4139, Lag order = 3, p-value = 0.7968
alternative hypothesis: stationary

- Масло: p-value = 0.999
> adf.test(oil, alternative = "stationary")

Augmented Dickey-Fuller Test
data: oil
Dickey-Fuller = -2.819, Lag order = 3, p-value = 0.2592
alternative hypothesis: stationary

- Чёрный чай: p-value = 0.998
> adf.test(tea, alternative = "stationary")

Augmented Dickey-Fuller Test
data: tea
Dickey-Fuller = -0.30624, Lag order = 3, p-value = 0.9836
alternative hypothesis: stationary

Все ряды нестационарные. Даже с учетом тренда, p-value >> 0.05.

5. **Скользящее среднее остатков и тест ADF**

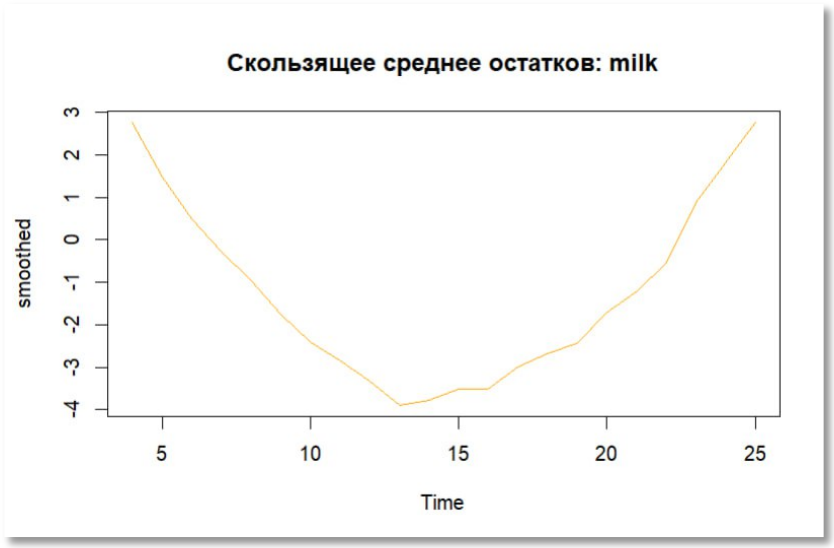
Делаем скользящее среднее остатков, чтобы:

- *Проверить, устранил ли тренд нестационарность,*
- *Увидеть скрытые циклы или структуры,*
- *Повысить эффективность ADF-теста на устойчивость поведения временного ряда.*

Молоко (milk):
ADF-статистика: -0.727
p-value: 0.840
Скользящее среднее не является стационарным

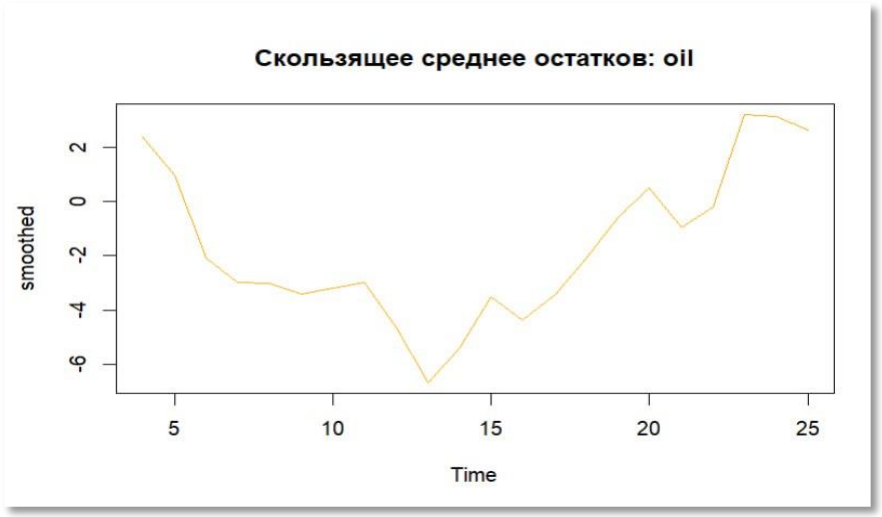
> analyze_smoothed_resid(milk, "Moloko")
[1] "ADF test for smoothed residuals: Moloko"

Augmented Dickey-Fuller Test
data: smoothed
Dickey-Fuller = 0.31727, Lag order = 2, p-value = 0.99
alternative hypothesis: stationary



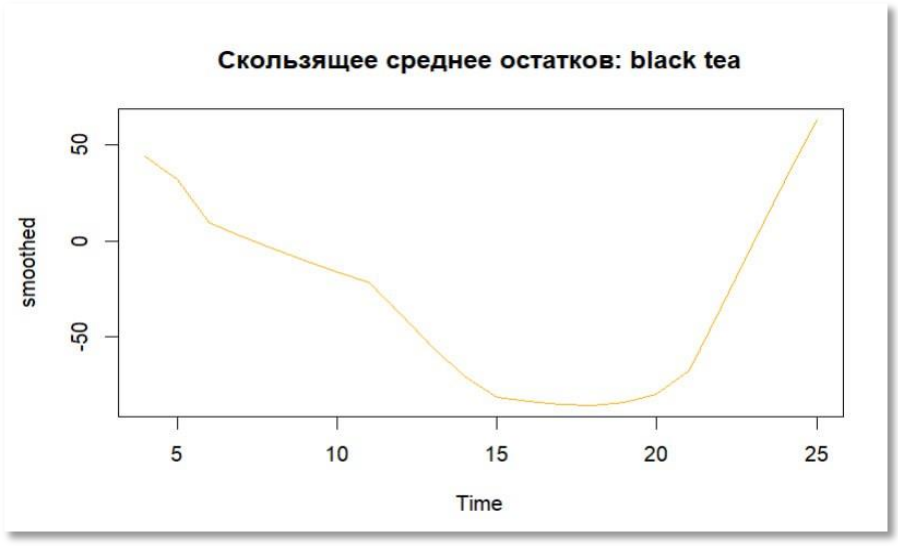
Масло (oil):
ADF-статистика: -0.050
p-value: 0.954

Скользящее среднее не является стационарным
> analyze_smoothed_resid(oil, "Maslo")
[1] "ADF test for smoothed residuals: Maslo"
	Augmented Dickey-Fuller Test
data: smoothed
Dickey-Fuller = -1.6206, Lag order = 2, p-value = 0.7169
alternative hypothesis: stationary



Чёрный чай (black tea):
ADF-статистика: -3.459
p-value: 0.99
Скользящее среднее не является стационарным
> analyze_smoothed_resid(tea, "Chyornyy chay")
[1] "ADF test for smoothed residuals: Chyornyy chay"

	Augmented Dickey-Fuller Test
data: smoothed
Dickey-Fuller = 0.73106, Lag order = 2, p-value = 0.99
alternative hypothesis: stationary



Вывод по пункту 6:

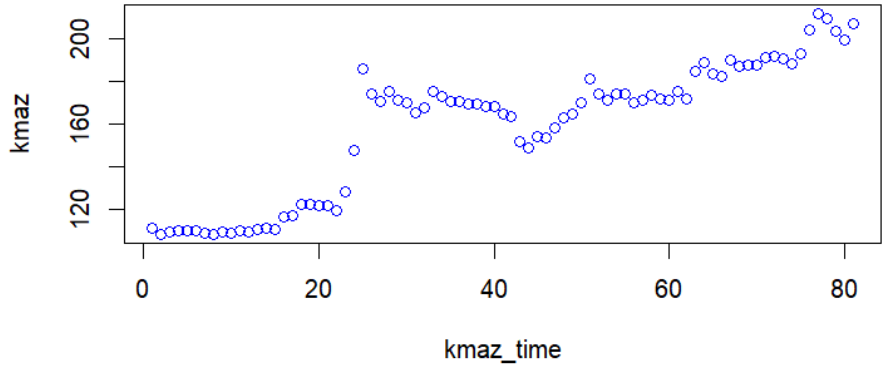
- Для **молока и масла** скользящее среднее не помогло устранить нестационарность.
- Для **чёрного чая** — скользящее среднее **оказалось стационарным**, что может указывать на более выраженную краткосрочную стабильность в изменениях тренда.

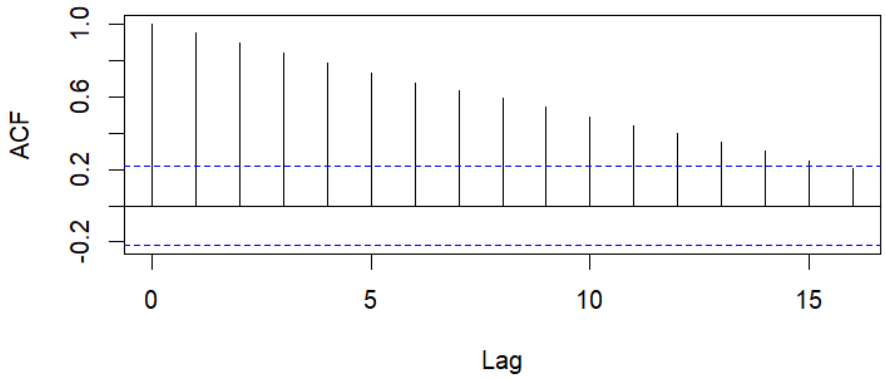
Задание 10. Модели динамики стационарных временных рядов

Анализ финансовых данных – временного ряда «Close» акций KMAZ

Коэффициент корреляции Спирмена = 0.8966845 – есть тенденция.

Коэффициент корреляции Кендалла = 0.7367455 – колебания периодические.





На графике временного ряда можно заметить, что тенденция, хоть и есть, но она искажена. На кореллограмме можно заметить, что автокорреляционная функция медленно убывает, значит можно предположить, что ряд нестационарный.

Проведя базовый и расширенный тест Дики-Фулера, можно убедиться в нестационарности Ряда.

```
> adfTest(kmaz,lags = 1)

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: 1.24
  P VALUE:
    0.9426
```

```
> adfTest(kmaz,lags = 1,type = c('ct'))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -2.551
  P VALUE:
    0.3501
```

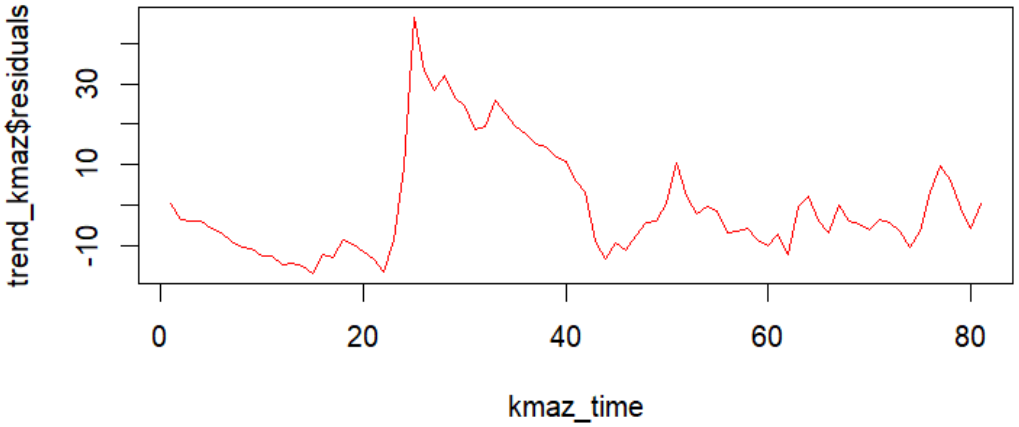
Оценка линейного тренда:

```
lm(formula = kmaz ~ kmaz_time, method = "qr")
Residuals:
    Min      1Q  Median      3Q     Max
-16.980  -9.090  -4.098   5.711  46.517

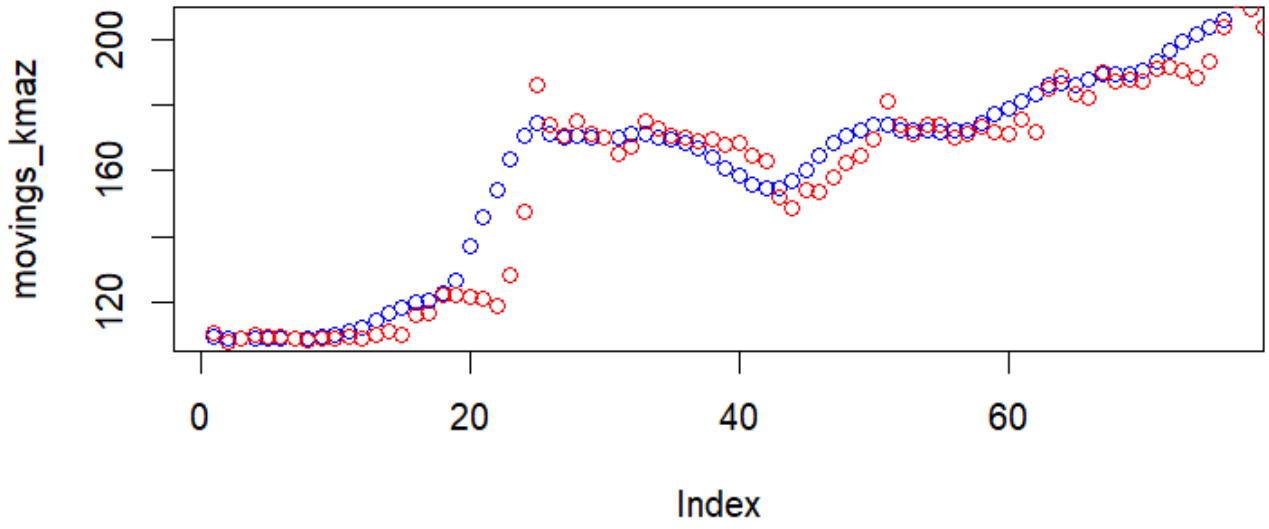
Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 109.47481    3.03076   36.12   <2e-16 ***
kmaz_time     1.20034    0.06421   18.69   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.51 on 79 degrees of freedom
Multiple R-squared:  0.8156, Adjusted R-squared:  0.8133
F-statistic: 349.4 on 1 and 79 DF,  p-value: < 2.2e-16
```

Тренд отражает закономерность на 82%. Чтобы понять, насколько модель линейного тренда хороша, можем проанализировать её остатки.



Тест Дики-Фулера (p = 0.01096) позволяет отвергнуть нулевую гипотезу о наличии единичного корня. То есть, что большинство ошибок (остатков) стационарны, т.е. не имеют тенденцию и колебания. Чтобы уточнить модель тренда, можно применить метод скользящего сглаживания.



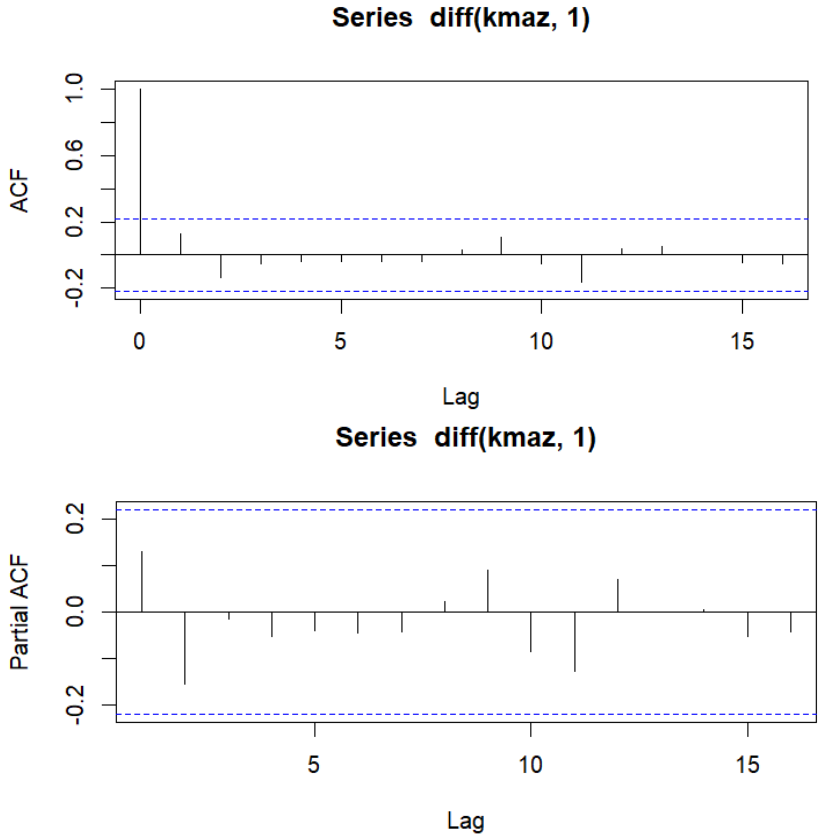
Исходные данные – красный, сглаженные данные - синий

Далее, приводим ряд к стационарному виду – к первой разности. Проверив её расширенным тестом Дики-Фулера, можно сказать, что скорость изменения цены стационарна.

> adfTest(diff(kmaz, 1),lags = 1,type = c('ct'))# «Первая разность» - она же скорость – стационарна

Title:
 Augmented Dickey-Fuller Test
Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -6.636
  P VALUE:
    0.01

Теперь мы можем строить ARMA модель. Сначала проанализируем обычную и частную автокорреляционную функции:



Анализ кореллограммы частной автокорреляционной функции позволяет предположить, что показатель AR = 1 или 2. Анализ кореллограммы автокорреляционной функции позволяет предположить, что показатель MA = 0.

После численных экспериментов, можно сказать, что из всех моделей, модель (1, 0, 2) наиболее адекватна.

arima(x = diff(kmaz, 1), order = c(1, 0, 2))

Coefficients:

	ar1	ma1	ma2	intercept
	0.8434	-0.7491	-0.2509	1.2044
s.e.	0.0715	0.1332	0.1288	0.1915

sigma^2 estimated as 37.42: log likelihood = -259.43, aic = 528.87

Модель	$t\alpha_1$	$t\beta_1$	$t\beta_1$	**AIC**
(1,0,1)	11,8	5,6	1,9	528.87