

**Министерство сельского хозяйства Российской Федерации
«Красноярский государственный аграрный университет»**

**СТАТИСТИЧЕСКИЕ МЕТОДЫ В АНАЛИЗЕ ДАННЫХ
АГРОПРОМЫШЛЕННОГО КОМПЛЕКСА**

**Методические указания
для самостоятельного изучения дисциплины**

Красноярск 2025

Рецензент
И.В. Сабодах, канд. физ.-мат. наук, доц.

Брит, А.А. Статистические методы в анализе данных агропромышленного комплекса: Метод. указания для самостоят. изучения дисциплины / Краснояр. гос. аграр. ун-т. - Красноярск, 2025. - 26 с.

Основной задачей методических указаний является оказание помощи магистрантам для самостоятельного изучения дисциплины «Статистические методы в анализе данных агропромышленного комплекса».

Предназначены для магистров очного отделения по программе 09.04.03 «Прикладная информатика».

Печатается по решению редакционно-издательского совета
Красноярского государственного аграрного университета

© Брит А.А., 2025

© ФГБОУ ВО «Красноярский государственный аграрный университет»,
2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ. КОМПЕТЕНЦИИ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ	5
2. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	7
3. КРИТЕРИИ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И ЗАЯВЛЕННЫХ КОМПЕТЕНЦИЙ	10
4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ	14
5. БИБЛИОГРАФИЧЕСКИЙ СПИСОК	15
6. ПРИМЕРЫ ЗАДАНИЙ	16
ГЛОССАРИЙ.....	20

ВВЕДЕНИЕ

Современный агропромышленный комплекс (АПК) сталкивается со множеством вызовов, связанных с необходимостью повышения эффективности производства, оптимизации использования ресурсов и обеспечения устойчивого развития. В условиях растущего объема данных, генерируемых в результате мониторинга полей, работы сельскохозяйственной техники, анализа почвы и климатических условий, становится критически важным применение современных статистических методов для обработки и интерпретации этой информации.

Дисциплина «Статистические методы в анализе данных агропромышленного комплекса» направлена на формирование у студентов навыков работы с массивами данных, их анализом и использованием для принятия обоснованных решений. Ключевые подходы и инструменты статистического анализа позволяют решать актуальные задачи АПК, такие как прогнозирование урожайности, оптимизация севооборота, управление ресурсами, контроль качества продукции и т.д.

В данном методическом пособии содержится программа, методические указания и примеры заданий по дисциплине.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ. КОМПЕТЕНЦИИ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ

Дисциплина Б1.В.04 «Статистические методы в анализе данных агропромышленного комплекса» является важной частью, формируемой участниками образовательных отношений, блока Б1 подготовки по программе 09.04.03 «Прикладная информатика», направленность «Прикладная информатика в агропромышленном комплексе» и относится к базовой части Б1. Изучение данной дисциплины основывается на знании курсов «Математика» или «Линейная алгебра», «Теория вероятностей и математическая статистика» для магистров, а также курса «Пакеты прикладных программ в научных исследованиях» для магистров. Освоение курса Б1.В.04 «Статистические методы в анализе данных агропромышленного комплекса» необходимо для изучения курса «Технологии обработки больших данных», а так же решения практико-ориентированных задач в различных областях.

Целью освоения дисциплины Б1.В.04 «Статистические методы в анализе данных агропромышленного комплекса» является развитие интеллекта магистрантов, способности к логическому и алгоритмическому мышлению; обучение основным статистическим методам, необходимым для анализа и моделирования процессов и явлений, при решении задач, возникающих в процессе практической деятельности.

Дисциплина нацелена на формирование универсальных компетенций:

УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий;

профессиональных компетенций выпускника:

ПК-1. Способность применять современные методы и инструментальные средства прикладной информатики для автоматизации и информатизации решения прикладных задач различных классов и создания ИС;

ПК-4 Способность принимать эффективные проектные решения в условиях неопределенности и риска;

ПК-6 Способность использовать информационные сервисы для автоматизации прикладных и информационных процессов.

Содержание дисциплины Б1.В.04 «Статистические методы в анализе данных агропромышленного комплекса» охватывает круг во-

просов, связанных с изучением следующих разделов: Основы статистических методов, Статистический анализ данных.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, лабораторные занятия, самостоятельная работа магистранта.

Программой дисциплины предусмотрены следующие виды контроля:

текущий контроль успеваемости в форме опросов и выполнения упражнений на занятиях и промежуточных аттестаций, контроль в форме тестирований, итоговый контроль в форме экзамена. Мониторинг познавательной деятельности магистрантов проводится на основе бально-рейтинговой системы.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единиц, 144 часа. Программой дисциплины предусмотрены лекционные (28 часов), лабораторные (28 часов) занятия и (52 часа) самостоятельной работы магистранта.

2. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Содержание дисциплины представлено в виде тематического плана:

Модуль 1. Основы статистических методов

Модульная единица 1.1 Средства сбора данных

Этапы и правила сбора данных. Статистическое наблюдение: понятие, виды, формы и способы. Статистическая сводка. Группировка. Стратификация.

Модульная единица 1.2 Средства предоставления данных

Статистические ряды распределения. Табличное представление данных. Графики, диаграммы, гистограммы и их виды.

Модульная единица 1.3 Основы математической статистики

Выборка и генеральная совокупность. Виды абсолютных и относительных величин. Средние величины: понятие, виды и способы расчета. Понятие вариации и показатели ее размера. Виды распределений. Нормальное распределение. Проверка распределения на нормальность. Преобразование распределения к нормальному. Статистические оценки параметров распределения. Несмещенные, эффективные и состоятельные оценки. Генеральная и выборочная средние. Оценка генеральной средней по выборочной средней. Устойчивость выборочных средних. Групповая и общие средние. Отклонение от общей средней и его свойства. Генеральная и выборочная дисперсии. Групповая, внутригрупповая, межгрупповая и общая дисперсии. Сложение дисперсий. Оценка генеральной дисперсии по исправленной выборочной. Точность оценки, доверительная вероятность. Доверительный интервал. Оценка точности измерений.

Модуль 2. Статистический анализ данных

Модульная единица 2.1 Статистическое изучение взаимосвязи признаков

Понятие, цель, задачи корреляционного анализа. Возникновение и развитие корреляционного анализа. Понятия функциональной, статистической и корреляционной зависимости. Корреляционное поле. Виды зависимости экономических переменных. Типы связи экономических переменных: линейные и нелинейные. Меры линейной связи переменных: парный, частный и множественный коэффициенты корреляции. И их свойства. Проверка статистических гипотез для оценки значимости корреляции. Определение доверительного интервала.

Ранговые коэффициенты корреляции: коэффициенты Спирмена и Кендалла. Корреляционное отношение.

Понятие, цель, задачи регрессионного анализа. Возникновение и развитие регрессионного анализа. Понятия регрессионного анализа: эндогенные и экзогенные переменные. Основные предпосылки регрессионного анализа (теорема Гаусса-Маркова). Уравнение регрессии. Геометрическая интерпретация регрессии. Метод наименьших квадратов. Статистические свойства оценок параметров. Стандартизованные коэффициенты регрессии. Коэффициенты эластичности. Коэффициент детерминации. Ошибка аппроксимации. Анализ вариации результирующего показателя. Проверка значимости уравнения регрессии и коэффициентов уравнения регрессии. Оценка качества регрессионной зависимости. Построение точечных и интервальных прогнозов. Построение доверительного интервала. Нелинейные модели и линеаризующие преобразования.

Понятие мультиколлинеарности. Признаки и причины мультиколлинеарности. Методы устранения мультиколлинеарности. Понятие гомоскедастичности случайных остатков. Понятие гетероскедастичности случайных остатков. Геометрическая интерпретация. Методы, позволяющие выявить наличие гетероскедастичности случайных остатков.

Модульная единица 2.2 Статистическое изучение динамических процессов

Понятие временного ряда и его отличие от случайной выборки. Составляющие временного ряда. Понятие тренда. Проверка гипотезы о неизменности среднего значения временного ряда. Процедуры аналитического выравнивания (сглаживания) временного ряда. Метод последовательных разностей.

Стационарные временные ряды и их характеристики. Понятие автокорреляции. Автокорреляционная функция. Временной лаг. Коэффициент автокорреляции. Коррелограмма.

Гетероскедастичность пространственной выборки. Искажение характеристик точности МНК-оценок. Проверка гипотезы о наличии (отсутствии) автокоррелированности регрессионных остатков. Положительная и отрицательная автокорреляция.

Построение множественной регрессионной модели по временным рядам. Модели рядов, содержащие сезонную компоненту.

Модульная единица 2.3 Классификация многомерных наблюдений

Особенности задач многомерной классификации. Основные понятия и определения кластерного анализа. Расстояния между объектами и меры близости групп. Дендрограммы. Итерационные методы классификации. Метод k-средних. Классификация с обучением. Дискриминантный анализ. Функции потерь и вероятность неправильной классификации.

3. КРИТЕРИИ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И ЗАЯВЛЕННЫХ КОМПЕТЕНЦИЙ

Текущая аттестация магистрантов производится в дискретные временные интервалы лектором и преподавателем, ведущими лабораторные занятия по дисциплине в следующих формах: тестирование; опрос; выполнение заданий. Отдельно оцениваются личностные качества магистранта (аккуратность, исполнительность, инициативность) – работа на занятиях, своевременная сдача тестов, выполнению заданий, посещение занятий.

Оценка знаний, умений, навыков и заявленных компетенций магистрантов проводится с использованием рейтинговой системы. Для получения допуска к экзамену магистранту необходимо набрать 40-60 баллов, в том числе по модулям:

Таблица 2. – Рейтинговый балл

Дисциплинарные модули (ДМ)	Количество академических часов	Рейтинговый балл
М ₁	40	20
М ₂	68	40
Экзамен	36	40
Итого часов	144	100

Промежуточный контроль по результатам семестра по дисциплине проходит в форме итогового тестирования.

Вопросы к экзамену

1. Этапы и правила сбора данных.
2. Статистическое наблюдение: понятие, виды, формы и способы.
3. Статистическая сводка. Группировка. Стратификация.
4. Табличное представление данных. Графики, диаграммы, гистограммы и их виды.
5. Выборка и генеральная совокупность.
6. Виды абсолютных и относительных величин. Средние величины: понятие, виды и способы расчета.
- Понятие вариации и показатели ее размера.
8. Виды распределений. Нормальное распределение. Проверка распределения на нормальность. Преобразование распределения к нормальному.

9. Статистические оценки параметров распределения. Несмещенные, эффективные и состоятельные оценки.
10. Генеральная и выборочная средние. Оценка генеральной средней по выборочной средней.
11. Устойчивость выборочных средних. Групповая и общие средние. Отклонение от общей средней и его свойства.
12. Генеральная и выборочная дисперсии. Групповая, внутригрупповая, межгрупповая и общая дисперсии.
13. Сложение дисперсий. Оценка генеральной дисперсии по исправленной выборочной.
14. Точность оценки, доверительная вероятность. Доверительный интервал. Оценка точности измерений.
15. Корреляционный анализ: понятие, задачи, применение.
16. Корреляционный анализ в «сильных» шкалах: корреляционное поле, типы корреляционной зависимости, формы корреляционной зависимости.
17. Парный коэффициент корреляции и его свойства. Проверка значимости парного коэффициента корреляции. Интервальная оценка парного коэффициента корреляции.
18. Ранговая корреляция: понятие, ранговые коэффициенты корреляции.
19. Частный коэффициент корреляции: понятие, сущность, проверка значимости. Множественный коэффициент корреляции. Корреляционное отношение: понятие и свойства.
20. Регрессионный анализ: понятие, задачи, применение, виды уравнений, понятие парной регрессии, виды уравнений, метод наименьших квадратов.
21. Оценка качества построенной модели: коэффициент детерминации, средняя ошибка аппроксимации, коэффициент эластичности. Понятие, формула, значение. Оценка качества уравнения регрессии.
22. Регрессионный анализ: понятие множественной регрессии, виды уравнений, Метод наименьших квадратов.
23. Мультиколлинеарность факторов.
24. Гомоскедастичность/гетероскедастичность случайных остатков.
25. Временные ряды: понятие, сущность, задачи, виды моделей, этапы построения модели.

26. Временные ряды: понятие автокорреляции уровней ряда, формула; понятие тренда, виды.
27. Временные ряды: метод наименьших квадратов, методы устранения тенденции.
28. Временные ряды: критерий Дарбина-Уотсона, коэффициент автокорреляции остатков.
29. Временные ряды: модели с распределенным лагом.
30. Временные ряды: оценка параметров модели с распределенным лагом, расчет параметров модели с распределенным лагом методом Алмон, модели авторегрессии.
31. Особенности задач многомерной классификации.
32. Основные понятия и определения кластерного анализа.
33. Расстояния между объектами и меры близости групп. Дендрограммы.
34. Итерационные методы классификации. Метод k-средних.
35. Классификация с обучением. Дискриминантный анализ.
36. Функции потерь и вероятность неправильной классификации.

Таблица 3. - Рейтинг план

Модуль	Максимально возможный балл по видам работ				ИТОГО
	Текущая работа			Аттестация	
	опрос	задания	Тестирование по модулям	Экзамен	
М ₁	5	10	5	-	20
М ₂	10	20	10	-	40
Экз	-	-	-	40	40
ИТОГО	15	30	15	40	100

Студенту, не набравшему требуемое минимальное количество баллов (<60), дается две недели после окончания календарного модуля для того, чтобы добрать необходимое количества баллов.

Баллы, полученные за экзамен, суммируются с баллами, полученными в течение соответствующего семестра.

Экзаменационная академическая оценка устанавливается в соответствии со следующей бальной шкалой:

- 60 – 72 балла – 3 (удовлетворительно);
 73 – 86 баллов – 4 (хорошо);

87 – 100 баллов – 5 (отлично).

Магистрант, не набравший 60 баллов (минимальное количество) приходит на пересдачу в соответствии с графиком ликвидации задолженностей http://www.kgau.ru/new/news/news/2017/grafik_lz.pdf.

4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Самостоятельная работа магистрантов (СРС) организуется с целью развития навыков работы с учебной и научной литературой, выработки способности вести научно-исследовательскую работу, а также для систематического изучения дисциплины.

– Реализуются следующие формы организации самостоятельной работы магистрантов:

- работа над теоретическим материалом, прочитанным на лекциях;
- подготовка к лекционным и практическим занятиям;
- самотестирование по контрольным вопросам (тестам);
- самостоятельная работа с обучающими программами в компьютерных классах и в домашних условиях (система Moodle).

Таблица 4. - Перечень вопросов для самостоятельного изучения

№ п/п	№ модуля и модульной единицы	Перечень рассматриваемых вопросов для самостоятельного изучения	Кол-во часов
Подготовка отдельных тем дисциплины			20
1.	Модуль 1. Случайные события		
	Темы: Стратификация. Темы: Точность оценки, доверительная вероятность. Доверительный интервал. Оценка точности измерений.		4
2.	Модуль 2. Случайные величины		
	Темы: Определение доверительного интервала. Ранговые коэффициенты корреляции: коэффициенты Спирмена и Кендалла. Корреляционное отношение. Темы: Понятие мультиколлинеарности. Признаки и причины мультиколлинеарности. Методы устранения мультиколлинеарности. Темы: Построение множественной регрессионной модели по временным рядам. Модели рядов, содержащие сезонную компоненту. Темы: Итерационные методы классификации. Метод k-средних. Функции потерь и вероятность неправильной классификации.		2
Подготовка к тестированию			4
Подготовка к Л, ПЗ, ЛР			14
Выполнение заданий			14
ИТОГО			52

5. БИБЛИОГРАФИЧЕСКИЙ СПИСОК

5.1. Карта обеспеченности литературой

1. Статистические методы оценки качества продукции : учебное пособие / М. Ю. Полянчикова, Н. И. Егорова, А. Н. Воронцова, А. А. Кожевникова. – Волгоградский государственный технический университет. – 2019. – 128с.

2. Анализ данных: учебное пособие / В.С. Мхитрян и др. – Юрайт – 2021. – 448с.

3. Статистическая обработка данных агрономических исследований в программе «Statistica»: учебное пособие / Р.Р. Усманов – Российский государственный аграрный университет. — МСХА им. К. А. Тимирязева. – 2020. – 177с.

5.2. Научные журналы

1. Прикладная эконометрика - <http://appliedeconometrics.cemi.rssi.ru/>

2. Сборник трудов «Управление большими системами» - <http://ubs.mtas.ru/about/>

3. Моделирование и анализ данных - <https://psyjournals.ru/mad/>

5.3. Информационно-телекоммуникационные ресурсы сети ИНТЕРНЕТ

Электронные библиотечные системы:

1. Электронная библиотечная система «Лань» e.lanbook.com

2. Электронная библиотечная система «Юрайт» <https://urait.ru/>

Электронные библиотеки

3. Научная электронная библиотека eLIBRARY.RU elibrary.ru

4. Научная библиотека Красноярского ГАУ www.kgau.ru/new/biblioteka

Информационные справочные системы

5. Справочная правовая система «КонсультантПлюс» <http://www.consultant.ru>

6. Информационно-правовой портал «Гарант»: <http://www.garant.ru>

Научные базы данных и профессиональные сайты

7. Русскоязычный сайт международного издательства Elsevier www.elsevier.com (Списки журналов Scopus, Списки журналов ScienceDirect)

6. ПРИМЕРЫ ЗАДАНИЙ

1. Описательная статистика

Задание: На основе предоставленных данных об урожайности пшеницы на 20 полях (в ц/га) выполните следующие действия:

Рассчитайте среднее значение, медиану, дисперсию и стандартное отклонение.

Постройте гистограмму распределения урожайности.

Сделайте выводы о характере распределения данных.

Данные:

45, 48, 50, 52, 55, 47, 49, 51, 53, 54, 46, 50, 52, 56, 58, 49, 51, 53, 55, 57.

2. Корреляционный анализ

Задание: Используя данные о количестве осадков (мм) и урожайности кукурузы (ц/га) за 10 лет, выполните следующие действия:

Постройте диаграмму рассеяния.

Рассчитайте коэффициент корреляции Пирсона.

Сделайте вывод о наличии и силе связи между осадками и урожайностью.

Данные:

Год	Осадки (мм)	Урожайность (ц/га)
1	300	60
2	350	65
3	400	70
4	380	68
5	420	72
6	390	69
7	410	71
8	430	74
9	440	75
10	450	76

3. Регрессионный анализ

Задание: На основе данных о внесении удобрений (кг/га) и урожайности картофеля (ц/га) постройте линейную регрессионную модель.

Определите уравнение регрессии.

Оцените качество модели с помощью коэффициента детерминации.

Спрогнозируйте урожайность при внесении 150 кг/га удобрений.
Данные:

Удобрения (кг/га)	Урожайность (ц/га)
50	200
100	250
150	300
200	350
250	400
300	450
350	500
400	550
450	600
500	650

4. Дисперсионный анализ (ANOVA)

Задание: Сравните урожайность трех сортов пшеницы (в ц/га) с использованием однофакторного дисперсионного анализа.

Сформулируйте нулевую и альтернативную гипотезы.

Проведите анализ и сделайте вывод о наличии статистически значимых различий между сортами.

Данные:

Сорт А	Сорт В	Сорт С
50	55	60
52	54	58
48	56	59
51	57	61
49	53	62
50	55	60
52	54	58
48	56	59
51	57	61
49	53	62

5. Кластерный анализ

Задание: На основе данных о содержании питательных веществ в почве (азот, фосфор, калий) на 15 полях выполните кластерный анализ.

Используя метод k-средних, разделите поля на 3 кластера.

Визуализируйте результаты на графике.

Интерпретируйте полученные кластеры.

Данные:

Поле	Азот	Фосфор	Калий
1	30	20	50
2	35	25	55
3	40	30	60
4	45	35	65
5	50	40	70
6	55	45	75
7	60	50	80
8	65	55	85
9	70	60	90
10	75	65	95

6. Прогнозирование временных рядов

Задание: На основе данных об урожайности пшеницы (ц/га) за последние 10 лет постройте модель прогнозирования на следующие 2 года.

Используйте метод скользящего среднего для сглаживания данных.

Постройте график исходных данных и прогноза.

Оцените точность модели.

Данные:

Год	Урожайность
2013	50
2014	52
2015	54
2016	56
2017	58
2018	60

Год	Урожайность
2019	62
2020	64
2021	66
2022	68

7. Статистическое тестирование гипотез

Задание: Проверьте гипотезу о равенстве средней урожайности двух сортов пшеницы (в ц/га) при уровне значимости 0,05.

Сформулируйте нулевую и альтернативную гипотезы.

Проведите t-тест для независимых выборок.

Сделайте вывод о наличии статистически значимых различий.

Данные:

Сорт А	Сорт В
50	55
52	54
48	56
51	57
49	53
50	55
52	54
48	56
51	57
49	53

ГЛОССАРИЙ

Модуль 1. Основы статистических методов

Модульная единица 1.1. Средства сбора данных

1. **Статистическое наблюдение** — процесс сбора данных о явлениях и процессах, происходящих в генеральной совокупности.
2. **Сплошное наблюдение** — сбор данных по всем единицам генеральной совокупности.
3. **Выборочное наблюдение** — сбор данных по части единиц генеральной совокупности.
4. **Статистическая сводка** — систематизация и обобщение данных, полученных в результате статистического наблюдения.
5. **Группировка** — разделение данных на группы по определенным признакам для анализа.
6. **Стратификация** — метод разделения данных на слои (страты) с однородными характеристиками.
7. **Статистический ряд** — упорядоченное представление данных, показывающее частоту встречаемости значений.
8. **Атрибутивная группировка** — группировка данных по качественным признакам.
9. **Количественная группировка** — группировка данных по количественным признакам.
10. **Интервальный ряд** — группировка данных по интервалам значений.

Модульная единица 1.2. Средства предоставления данных

11. **Статистическая таблица** — форма представления данных в виде строк и столбцов.
12. **График** — визуальное представление данных с использованием линий, точек или других элементов.
13. **Диаграмма** — графическое представление данных с использованием столбцов, кругов или других фигур.
14. **Гистограмма** — график, показывающий распределение данных в виде столбцов, где высота соответствует частоте значений.
15. **Полигон частот** — график, построенный путем соединения точек, соответствующих частотам интервалов.
16. **Круговая диаграмма** — график, где данные представлены в виде секторов круга.

17. **Столбчатая диаграмма** — график, где данные представлены в виде вертикальных или горизонтальных столбцов.
18. **Линейный график** — график, где данные представлены в виде линии, соединяющей точки.
19. **Точечная диаграмма** — график, где данные представлены в виде точек на плоскости.
20. **Картограмма** — графическое представление данных на географической карте.

Модульная единица 1.3. Основы математической статистики

21. **Генеральная совокупность** — полное множество объектов или явлений, обладающих общими свойствами, которые изучаются.
22. **Выборка** — часть генеральной совокупности, отобранная для анализа.
23. **Абсолютные величины** — количественные показатели, выраженные в натуральных или стоимостных единицах.
24. **Относительные величины** — показатели, выражающие соотношение между абсолютными величинами (например, проценты, коэффициенты).
25. **Среднее арифметическое** — сумма всех значений выборки, деленная на их количество.
26. **Медиана** — значение, которое делит упорядоченную выборку на две равные части.
27. **Мода** — наиболее часто встречающееся значение в выборке.
28. **Вариация** — разброс значений в данных, измеряемый с помощью дисперсии, стандартного отклонения и других показателей.
29. **Дисперсия** — мера разброса данных, показывающая, насколько значения отклоняются от среднего.
30. **Стандартное отклонение** — квадратный корень из дисперсии, используемый для измерения разброса данных.
31. **Нормальное распределение** — симметричное распределение данных, где большинство значений сосредоточено вокруг среднего.
32. **Проверка на нормальность** — анализ данных для определения их соответствия нормальному распределению.

- 33. **Преобразование распределения** — изменение формы распределения данных для приведения к нормальному виду.
- 34. **Статистическая оценка** — значение, вычисленное по выборке и используемое для оценки параметров генеральной совокупности.
- 35. **Несмещенная оценка** — оценка, математическое ожидание которой равно оцениваемому параметру.
- 36. **Эффективная оценка** — оценка с минимальной дисперсией среди всех возможных оценок.
- 37. **Состоятельная оценка** — оценка, которая с увеличением объема выборки стремится к истинному значению параметра.
- 38. **Генеральная средняя** — среднее значение признака в генеральной совокупности.
- 39. **Выборочная средняя** — среднее значение признака в выборке.
- 40. **Групповая средняя** — среднее значение признака в отдельной группе данных.
- 41. **Общая средняя** — среднее значение признака для всей совокупности данных.
- 42. **Отклонение от общей средней** — разница между значением признака и общей средней.
- 43. **Генеральная дисперсия** — дисперсия признака в генеральной совокупности.
- 44. **Выборочная дисперсия** — дисперсия признака в выборке.
- 45. **Групповая дисперсия** — дисперсия признака в отдельной группе данных.
- 46. **Внутригрупповая дисперсия** — дисперсия внутри групп данных.
- 47. **Межгрупповая дисперсия** — дисперсия между группами данных.
- 48. **Общая дисперсия** — сумма внутригрупповой и межгрупповой дисперсий.
- 49. **Доверительный интервал** — диапазон значений, в котором с заданной вероятностью находится параметр генеральной совокупности.
- 50. **Доверительная вероятность** — вероятность того, что доверительный интервал содержит истинное значение параметра.

Модуль 2. Статистический анализ данных

Модульная единица 2.1. Статистическое изучение взаимосвязи признаков

1. **Корреляционный анализ** — метод изучения взаимосвязи между переменными.
2. **Функциональная зависимость** — связь, при которой каждому значению одной переменной соответствует строго определенное значение другой.
3. **Статистическая зависимость** — связь, при которой изменение одной переменной влияет на распределение другой.
4. **Корреляционная зависимость** — частный случай статистической зависимости, при которой изменение одной переменной приводит к изменению средней величины другой.
5. **Корреляционное поле** — графическое представление взаимосвязи между двумя переменными в виде точек на плоскости.
6. **Линейная связь** — зависимость, при которой изменение одной переменной пропорционально изменению другой.
7. **Нелинейная связь** — зависимость, при которой изменение одной переменной не пропорционально изменению другой.
8. **Парный коэффициент корреляции** — мера линейной связи между двумя переменными.
9. **Частный коэффициент корреляции** — мера связи между двумя переменными при исключении влияния других переменных.
10. **Множественный коэффициент корреляции** — мера связи между одной переменной и несколькими другими.
11. **Коэффициент Спирмена** — ранговый коэффициент корреляции, используемый для оценки нелинейной связи.
12. **Коэффициент Кендалла** — ранговый коэффициент корреляции, основанный на сравнении пар наблюдений.
13. **Корреляционное отношение** — мера связи, учитывающая как линейные, так и нелинейные зависимости.
14. **Регрессионный анализ** — метод моделирования зависимости одной переменной от другой (или нескольких).
15. **Эндогенные переменные** — зависимые переменные в регрессионной модели.
16. **Экзогенные переменные** — независимые переменные в регрессионной модели.

17. **Теорема Гаусса-Маркова** — теорема, определяющая условия, при которых оценки метода наименьших квадратов являются наилучшими.
18. **Уравнение регрессии** — математическое выражение, описывающее зависимость между переменными.
19. **Метод наименьших квадратов (МНК)** — метод оценки параметров регрессии, минимизирующий сумму квадратов отклонений.
20. **Стандартизованные коэффициенты регрессии** — коэффициенты, выраженные в стандартных отклонениях для сравнения влияния переменных.
21. **Коэффициент эластичности** — показатель, отражающий процентное изменение зависимой переменной при изменении независимой на 1%.
22. **Коэффициент детерминации (R^2)** — мера, показывающая долю дисперсии зависимой переменной, объясненную моделью.
23. **Ошибка аппроксимации** — разница между наблюдаемыми и предсказанными значениями.
24. **Мультиколлинеарность** — сильная корреляция между независимыми переменными в регрессионной модели.
25. **Гомоскедастичность** — свойство случайных остатков, при котором их дисперсия постоянна.
26. **Гетероскедастичность** — свойство случайных остатков, при котором их дисперсия непостоянна.

Модульная единица 2.2. Статистическое изучение динамических процессов

27. **Временной ряд** — последовательность данных, измеренных в последовательные моменты времени.
28. **Тренд** — долгосрочная тенденция изменения временного ряда.
29. **Сезонная компонента** — периодические колебания временного ряда, связанные с временами года или другими циклами.
30. **Стационарный временной ряд** — ряд, статистические свойства которого не зависят от времени.
31. **Автокорреляция** — корреляция между значениями временного ряда, разделенными определенным интервалом времени.

- 32. **Автокорреляционная функция** — функция, показывающая зависимость автокорреляции от временного лага.
- 33. **Временной лаг** — интервал времени между наблюдениями временного ряда.
- 34. **Коэффициент автокорреляции** — мера связи между значениями временного ряда, разделенными лагом.
- 35. **Коррелограмма** — график автокорреляционной функции.
- 36. **Метод последовательных разностей** — метод сглаживания временного ряда путем вычисления разностей между соседними значениями.
- 37. **Положительная автокорреляция** — ситуация, когда высокие значения ряда следуют за высокими, а низкие — за низкими.
- 38. **Отрицательная автокорреляция** — ситуация, когда высокие значения ряда следуют за низкими, и наоборот.

Модульная единица 2.3. Классификация многомерных наблюдений

- 39. **Кластерный анализ** — метод группировки объектов на основе их схожести по определенным признакам.
- 40. **Расстояние между объектами** — мера различия между объектами в многомерном пространстве.
- 41. **Меры близости групп** — показатели, используемые для оценки схожести групп объектов.
- 42. **Дендрограмма** — графическое представление иерархической структуры кластеров.
- 43. **Метод k-средних** — итерационный метод кластеризации, основанный на минимизации расстояний до центроидов.
- 44. **Классификация с обучением** — метод классификации, использующий заранее определенные классы.
- 45. **Дискриминантный анализ** — метод классификации, основанный на разделении объектов по дискриминантным функциям.
- 46. **Функции потерь** — показатели, оценивающие ошибки классификации.
- 47. **Вероятность неправильной классификации** — вероятность того, что объект будет отнесен к неправильному классу.
- 48. **Многомерная классификация** — процесс разделения объектов на группы в многомерном пространстве признаков.

49. **Итерационные методы классификации** — методы, которые последовательно улучшают результаты классификации.
50. **Центроид** — средняя точка кластера, используемая в методах кластеризации.