МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

МИНИСТЕРСТВО НАУКИ, ВЫСШЕГО ОБРАЗОВАНИЯ И ИННОВАЦИЙ КЫРГЫЗСКОЙ РЕСПУБЛИКИ

МОО ВО Кыргызско-Российский Славянский университет
имени первого Президента Российской Федерации Б.Н. Ельцина

Высшая школа Цифровых технологий

Фонд

оценочных средств

по дисциплине **«Статистическая обработка данных»**

Уровень высшего образования

МАГИСТРАТУРА

Направление подготовки

09.04.04 - РФ, 710400 - КР Программная инженерия

(код и наименование направления подготовки)

"Основы анализа и разработки приложений с большими объемами распределенных данных" (профиля) образовательной программы)

Квалификация

магистр

Фонд оценочных средств предназначен для контроля знаний обучающихся по направлению подготовки 09.04.04 – РФ, 710400 - КР «Программная инженерия» по дисциплине «Статистическая обработка данных».

Фонд оценочных средств рассмотрен и утвержден
на заседании Высшей школы Цифровых технологий протокол №1 от 02.09.26

Директор Высшей школы Цифровых технологий:

к.т.н., доцент Хмелева И.В.

Руководитель образовательной программы:

д.т.н., профессор Лыченко Н.М.

Исполнители (разработчики):

доцент Осмонов М.С.

д.т.н., профессор Лыченко Н.М.

СОГЛАСОВАНО:
Зам. директора по учебной работе
Политехнического института КРСУ Комарцов Н.М.

Раздел 1. Перечень компетенций и этапы их формирования

ПК-4:.

Планируемые результаты обучения воспроизведены по п. 3 РПД (обобщённая формулировка компетенции и дисциплинарные результаты).

Компетенция	Планируемые результаты обучения	Оценочные средства
ПК-4 Способен разрабатывать программное обеспечение разного уровня сложности и руководить его разработкой с использованием систем конфигурационного управления	ЗНАТЬ: методы создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов; возможности, основные понятия, задачи и методы статистической обработки данных; статистики и критерии оценки характеристик случайных величин.	Блок A: устный опрос; теоретические вопросы при защите практических работ; вопросы зачёта (Блок D).
	УМЕТЬ: использовать методы создания программного обеспечения для анализа и обработки информации; выбирать методы статистического исследования экспериментальных данных, рассчитывать доверительные интервалы, определять степень регрессионного полинома, рассчитывать интерполяционные полиномы, проверять статистические гипотезы, проводить дисперсионный и корреляционный анализ, анализ временных рядов.	Блок B: практические задания № 1–3, задания по расчетам и анализу выборочных данных; Блок D: практические задания зачёта.
	ВЛАДЕТЬ: методами и средствами разработки программного обеспечения для анализа и обработки информации с применением современных компьютерных технологий; навыками систематизации, математического анализа и интерпретации количественной информации в R/RStudio.	Блок C: программные реализации в R, защита электронных отчётов, самостоятельная работа с результатами эксперимента; Блок D.

Раздел 2. Технологическая карта дисциплины

Курс 1, семестр 2. Количество ЗЕ — 3. Отчётность — зачёт с оценкой.

Модуль дисциплины	Контроль	Форма контроля	Мин., балл	Макс., балл	Неделя
1. Предмет и методы статистического	Текущий	Посещаемость, активность, устный опрос, самостоятельная работа по основам R	4	8	26

анализа данных. Представление исходных данных	Рубежный	Защита практической работы № 1 «Знакомство со средой RStudio»	8	12	26
2. Статистические методы обработки и анализа экспериментальных данных	Текущий	Посещаемость, активность, устный опрос, самостоятельная работа по методам описательной статистики	5	10	31
	Рубежный	Защита практической работы № 2 «Распределения и проверка гипотез»	9	15	31
3. Оценивание неизвестных параметров распределения. Статистические критерии	Текущий	Посещаемость, активность, устный опрос, самостоятельная работа по оцениванию и регрессии	5	10	37
	Рубежный	Защита практической работы № 3 «Регрессионный анализ»	9	15	37
ВСЕГО за семестр			40	70	37
Промежуточный контроль		Письменный зачёт с оценкой (теория — 10 баллов, практика — 20 баллов)	20	30	—
СЕМЕСТРОВЫЙ РЕЙТИНГ			60	100	

Раздел 3. Типовые контрольные задания и иные оценочные материалы

Блок А. Задания репродуктивного уровня (теоретический контроль)

Вопросы используются при устном опросе, собеседовании и защите практических работ. По тематике вопросы сгруппированы по трём разделам РПД.

Модуль 1. Предмет и методы статистического анализа данных. Представление исходных данных

1. Предмет, основные задачи и этапы статистической обработки данных. Разведочный анализ и статистический вывод.
2. Генеральная совокупность и выборка. Требования к репрезентативности; наблюдение и переменные.
3. Качественные, количественные, дискретные и непрерывные признаки. Вариационный и статистический ряды.
4. Абсолютные, относительные и накопленные частоты. Таблицы распределений.
5. Эмпирическая функция распределения, гистограмма и полигон частот.
6. Числовые характеристики выборки: среднее, медиана, мода, дисперсия, стандартное отклонение.
7. Базовые типы и структуры данных языка R: вектор, матрица, список, фактор и data.frame.
8. Основные возможности RStudio, работа со скриптами, функциями, пакетами и графиками.

Модуль 2. Статистические методы обработки и анализа экспериментальных данных

1. Меры положения, меры разброса, асимметрия, эксцесс. Выбросы и робастные оценки.
2. Центрирование, нормирование, стандартизация данных. Преобразования данных.
3. Нормальное распределение, его свойства и признаки отклонения от нормальности.
4. Нулевая и альтернативная гипотезы. Уровень значимости и p-value.
5. Ошибки первого и второго рода; мощность критерия.
6. Критерий хи-квадрат Пирсона: условия применения, расчёт и интерпретация.
7. Критерии Колмогорова–Смирнова и Шапиро–Уилка.
8. Корреляция Пирсона, ранговые корреляции Спирмена и Кендалла; интерпретация коэффициентов.

Модуль 3. Оценивание неизвестных параметров распределения. Статистические критерии

1. Точечные и интервальные оценки. Несмещённость, состоятельность и эффективность оценок.
2. Стандартная ошибка и доверительные интервалы для характеристик генеральной совокупности.
3. Метод моментов и идея метода наименьших квадратов.
4. Простая линейная регрессия: модель, коэффициенты, остатки, коэффициент детерминации.
5. Полиномиальная регрессия: выбор степени, проверка адекватности модели.
6. Проверка значимости параметров регрессии. Интерпретация результатов.
7. Параметрические и непараметрические критерии; критерии Стьюдента, Фишера, Манна–Уитни, Уилкоксона.
8. Дисперсионный анализ; проверка статистических гипотез о средних значениях.

Блоки B и C. Практические и практико-ориентированные задания

Выполнение — на языке R в RStudio, итог — электронный отчёт и демонстрация работоспособности программного решения.

Практическая работа № 1. Знакомство с языком R и средой RStudio

1. Создать проект RStudio; составить и сохранить исполняемый R-скрипт, содержащий операции ввода и вывода данных.
2. Создать примеры структур: вектор, матрица, массив, список, фактор и data.frame. Проверить размерности и типы элементов.
3. Продемонстрировать работу конструкций if/else, for, while, repeat на небольших вычислительных примерах.
4. Написать собственную функцию, содержащую аргументы и возвращаемое значение; продемонстрировать вызов функции.
5. Установить или подключить пакет расширения, получить список установленных пакетов и справку по функции.
6. Построить минимум два графика исходного вектора (например, точечный и гистограмму), снабдить осями и заголовками.

Ожидаемые результаты: R-скрипт, краткая характеристика структур данных, таблица результатов работы функций, два графика и электронный отчёт.

Практическая работа № 2. Исследование законов распределения и проверка гипотез

1. На индивидуальной выборке экспериментальных данных реализовать выявление грубых ошибок измерения: сравнение нормированного отклонения с критическим значением t-распределения при $\alpha = 0{,}05$; альтернативно — сравнение с заданным коэффициентом K ($K = 4$ при $6 < n \leq 100$; $K = 4{,}5$ при $100 < n \leq 1000$; $K = 5$ при $n > 1000$).
2. Вычислить минимум и максимум, сформировать вариационный ряд, задать число интервалов группировки, построить таблицу абсолютных и относительных частот.
3. Построить гистограмму, полигон частот и эмпирическую функцию распределения.
4. Рассчитать выборочные начальные моменты, смещённую и несмещённую дисперсии, стандартное отклонение, коэффициенты асимметрии и эксцесса, стандартизованные значения.
5. Проверить гипотезу о нормальности распределения критерием χ^2 Пирсона при $\alpha = 0{,}05$ и обосновать статистическое решение с учётом числа степеней свободы и выбора групп.
6. Подготовить отчёт с кодом, графиками, результатами критериев и интерпретацией выводов.

Практическая работа № 3. Регрессионный анализ

1. По экспериментальным данным построить полиномиальную регрессионную модель степени $m = 1$.
2. Оценить коэффициенты модели методом наименьших квадратов и вычислить прогнозные значения и остатки.
3. Проверить адекватность модели. При отклонении гипотезы последовательно увеличивать степень полинома до получения обоснованного результата (при ограничении $m \leq n - 2$, указанном в РПД).
4. Для линейной регрессионной модели проверить значимость параметров, исключать незначимые параметры и повторять оценивание и проверку адекватности.
5. Построить график исходных наблюдений и график регрессионной модели; вычислить и интерпретировать коэффициент детерминации.
6. Составить электронный отчёт, продемонстрировать код и объяснить выбранную степень модели и результаты проверок.

Ожидаемые результаты: R-скрипт, таблицы коэффициентов и остатков, диагностические графики, проверка гипотез и вывод об адекватности модели.

Дополнительные типовые задачи для оценки умений и владений

1. Для выборки по индивидуальному варианту построить вариационный ряд, эмпирическую функцию распределения, гистограмму, полигон частот и рассчитать описательные статистики.
2. Для двух выборок сформулировать H_0 и H_1, подобрать статистический критерий с учётом предпосылок, выполнить проверку в R и обосновать вывод.
3. Для двух количественных признаков вычислить коэффициент корреляции, построить поле рассеяния и интерпретировать силу и направление связи.
4. По набору экспериментальных пар (x, y) оценить линейную и квадратичную регрессию, проверить адекватность и сравнить результаты.
5. На выборке по варианту выбрать число кластеров, выполнить кластеризацию средствами R, визуализировать результат и оценить интерпретируемость групп.
6. Построить базовое дерево решений для классификации наблюдений, разделить данные на обучающую и контрольную части и оценить качество предсказаний.

Блок D. Промежуточный контроль — зачёт с оценкой

Письменное задание сочетает теоретические вопросы и практические задачи, как предусмотрено п. 8 РПД. Формат, предлагаемый в настоящем ФОС: три кратких теоретических вопроса (в совокупности до 10 баллов) и два практических задания (в совокупности до 20 баллов).

Перечень вопросов для проверки уровня ЗНАТЬ

1. Этапы статистического исследования. Генеральная и выборочная совокупности.
2. Статистический ряд, эмпирическая функция распределения и средства визуализации.
3. Выборочное среднее, дисперсия, асимметрия и эксцесс.
4. Грубые ошибки измерения и подходы к выявлению выбросов.
5. Точечные оценки и свойства статистических оценок.
6. Доверительные интервалы и стандартная ошибка оценки.
7. Задача проверки статистических гипотез; уровень значимости и мощность критерия.
8. Критерии согласия Пирсона и Колмогорова–Смирнова.
9. Проверка нормальности распределения и практическая интерпретация результата.
10. Парная корреляция и регрессионная связь.
11. Метод наименьших квадратов и коэффициент детерминации.
12. Оценка адекватности и значимости параметров регрессии.
13. Дисперсионный анализ и его предпосылки.
14. Обучение с учителем и без учителя; классификация и прогнозирование.
15. Метод опорных векторов, дерево решений: назначение и ограничения.
16. Кластерный анализ: иерархические и итерационные методы.
17. Анализ взаимосвязей между переменными; ассоциативные правила.
18. Использование R и RStudio для анализа экспериментальных данных.

Типовые практические задания зачёта

1. Для ряда наблюдений x = (8, 9, 10, 10, 12, 13, 15, 16) вычислить среднее, медиану, несмещённую дисперсию и стандартное отклонение; построить эмпирическую функцию распределения в R.
2. Для выборки x = (4, 5, 5, 6, 7, 8, 8, 9, 10, 12) построить таблицу частот, гистограмму и полигон; описать форму распределения.
3. По данным x = (1, 2, 3, 4, 5), y = (2.1, 3.8, 6.2, 7.9, 10.3) оценить уравнение линейной регрессии и R^2, построить график модели и остатков.
4. Для двух независимых выборок x = (14, 15, 16, 16, 17) и y = (12, 13, 13, 14, 15) сформулировать проверяемую гипотезу о равенстве средних, выбрать критерий при $\alpha = 0{,}05$, перечислить проверяемые предпосылки и объяснить результат.
5. Для пары признаков x = (1, 3, 4, 5, 7), y = (2, 5, 7, 9, 13) определить коэффициент линейной корреляции Пирсона и построить диаграмму рассеяния.
6. В RStudio создать data.frame с тремя переменными, обработать пропуски согласно указанному правилу, провести стандартизацию признаков и построить сравнительные графики.

Примеры билетов промежуточного контроля

БИЛЕТ № 1

1. Вопросы для проверки «ЗНАТЬ»: 1) Этапы статистического исследования. Генеральная и выборочная совокупности. 5) Точечные оценки и свойства статистических оценок. 7) Задача проверки статистических гипотез; уровень значимости и мощность критерия.

2. Задания для проверки «УМЕТЬ» и «ВЛАДЕТЬ»: задание 1 и задание 3 из перечня типовых практических заданий зачёта. Каждое задание — до 10 баллов.

БИЛЕТ № 2

1. Вопросы для проверки «ЗНАТЬ»: 2) Статистический ряд, эмпирическая функция распределения и средства визуализации. 10) Парная корреляция и регрессионная связь. 12) Оценка адекватности и значимости параметров регрессии.

2. Задания для проверки «УМЕТЬ» и «ВЛАДЕТЬ»: задание 2 и задание 4 из перечня типовых практических заданий зачёта. Каждое задание — до 10 баллов.

БИЛЕТ № 3

1. Вопросы для проверки «ЗНАТЬ»: 3) Выборочное среднее, дисперсия, асимметрия и эксцесс. 6) Доверительные интервалы и стандартная ошибка оценки. 11) Метод наименьших квадратов и коэффициент детерминации.

2. Задания для проверки «УМЕТЬ» и «ВЛАДЕТЬ»: задание 3 и задание 5 из перечня типовых практических заданий зачёта. Каждое задание — до 10 баллов.

Раздел 4. Методические материалы, процедуры и критерии оценивания

Текущий контроль проводится в ходе семестра (посещаемость, активность, ответы при опросе и самостоятельная работа). Рубежный контроль — защита электронных отчётов по практическим работам № 1–3. Промежуточная аттестация проводится письменно. Устное собеседование при защите оценивается в составе баллов за практическую работу, а не как отдельный дополнительный вид контроля.

4.1. Распределение баллов

Вид и объект оценивания	Минимум	Максимум
Модуль 1: текущий контроль / защита практической работы № 1	4 / 8	8 / 12
Модуль 2: текущий контроль / защита практической работы № 2	5 / 9	10 / 15
Модуль 3: текущий контроль / защита практической работы № 3	5 / 9	10 / 15
Итого текущий и рубежный контроль за семестр	40	70
Промежуточный контроль: теория / практика	7 / 13	10 / 20
Итого зачёт с оценкой	20	30
Общий итог	60	100

4.2. Критерии текущего контроля

На каждом модуле текущий контроль оценивается по полноте самостоятельной работы, регулярности участия в занятиях и корректности устных ответов. Баллы текущего контроля начисляются в пределах, указанных в технологической карте; посещаемость без участия в выполнении заданий не заменяет проверку знаний.

Рекомендованная шкала для текущего контроля: 85–100% от максимума — самостоятельные точные ответы и выполненные задания; 70–84% — существенные знания при отдельных

неточностях; 60–69% — базовое понимание и выполнение основной части; 31–59% — существенные пробелы; 0–30% — отсутствие результата. При выставлении итоговых баллов применяются пределы соответствующего контрольного мероприятия.

4.3. Критерии защиты практических работ

Внутри каждого максимального балла за защиту практической работы оценка складывается из следующих долей: корректность программной реализации и вычислений — 50%; полнота и воспроизводимость результатов — 25%; оформление и интерпретация отчёта — 15%; ответы на вопросы преподавателя — 10%.

Доля от максимального балла	Характеристика выполнения
85–100%	Программа работает корректно, расчёты и графики проверяемы, отчёт полон, выводы обоснованы.
70–84%	Основное решение верно, присутствуют малосущественные ошибки или неполнота интерпретации.
60–69%	Выполнены ключевые шаги, но имеются недочёты реализации/оформления и недостаточно убедительные выводы.
31–59%	Выполнена лишь часть задания, имеются существенные вычислительные или методические ошибки.
0–30%	Отсутствует рабочее решение или отсутствуют результаты выполнения.

4.4. Оценивание зачёта с оценкой

Теоретическая часть — до 10 баллов суммарно за три вопроса: полнота понимания и правильное применение понятий (0–6), аргументированность и логичность изложения (0–2), корректность примеров и интерпретации (0–2). Практическая часть — два задания по 10 баллов: выбор метода (0–2), правильный расчёт/код (0–5), проверка и интерпретация результата (0–3).

Для минимального зачётного результата рекомендуется набрать не менее 7 баллов за теорию и не менее 13 за практические задания (в совокупности 20 баллов); итог определяется суммарным рейтингом за дисциплину.

Суммарный рейтинг	Итоговая оценка
85–100 баллов	Отлично
70–84 балла	Хорошо
60–69 баллов	Удовлетворительно
Менее 60 баллов	Неудовлетворительно

Раздел 5. Методические указания обучающимся

5.1. Организация самостоятельной работы

Перед занятием необходимо изучить лекционный материал, литературу и предоставленные задания, подготовить исходные данные для индивидуального варианта. Методические материалы и задания размещаются преподавателем в Google Classroom, как указано в РПД.

Работы выполняются преимущественно в R/RStudio. Исходный код должен содержать достаточные комментарии для воспроизведения расчётов; численные результаты, таблицы и графики сохраняются в электронном отчёте.

5.2. Выполнение и защита практических работ

1. Уточнить исходные данные своего варианта и используемые критерии, сформулировать цель работы и критерий правильности результата.
2. Разработать программу на языке R, проверить её выполнение и сохранить скрипты/файлы данных в понятной структуре.
3. Представить расчёты, графики, промежуточные результаты и статистическую интерпретацию, указать уровень значимости для проверяемых гипотез.
4. Подготовить электронный отчёт с целью, исходными данными, методикой, реализацией, результатами и выводами.
5. Представить отчёт преподавателю и ответить на вопросы по теории и программной реализации. Сроки проверок определяются технологической картой.

5.3. Подготовка к зачёту с оценкой

Повторить теоретические вопросы блоков A и D; проверить умение самостоятельно выбирать статистические методы, строить графики, вычислять числовые характеристики выборки, проверять статистические гипотезы и оценивать регрессионные модели. В письменном ответе последовательно указать постановку задачи, метод решения, вычисления и интерпретацию.