

**Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент математики
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ

Проректор по учебной и
методической работе

_____ Е.А. Каменева

28 декабря 2023 г.

Камалетдинов А.Ш.

Анализ данных на языке R

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки
38.03.01 - Экономика,
ОП «Корпоративные финансы»,
профиль: «Корпоративные финансы и инвестиции»

*Рекомендовано Ученым советом
Факультета информационных технологий и анализа больших данных
(протокол № 38 от 23.11.2023 г.)*

*Одобрено Советом учебно-научного Департамента математики
(протокол № 05 от 23.10.2023 г.)*

Москва 2023

СОДЕРЖАНИЕ

1. Наименование дисциплины.....	2
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.....	2
3. Место дисциплины в структуре образовательной программы.....	4
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся.....	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	5
5.1. Содержание дисциплины.....	5
5.2. Учебно-тематический план.....	6
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	8
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	8
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю.....	9
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по данной дисциплине.....	10
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....	15
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	16
10. Методические указания для обучающихся по освоению дисциплины.....	18
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	19
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	20
Содержание семинаров, практических занятий.....	7

1. Наименование дисциплины

«Анализ данных на языке R».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Дисциплина «Анализ данных на языке R» обеспечивает формирование компетенций: ПКН-3, ПКН-6, ПКП-2

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-3	Способность осуществлять сбор, обработку и статистический анализ данных, применять математические методы для решения стандартных профессиональных финансово экономических задач, интерпретировать полученные результаты	1. Проводит сбор, обработку и статистический анализ данных для решения финансово-экономических задач.	Знать технологию импорта данных разных форматов для последующей обработки в R Уметь осуществлять поиск, сбор и первичную подготовку данных для экономических исследований
		2. Формулирует математические постановки финансово-экономических задач, переходит от экономических постановок задач к математическим моделям.	Знать основные принципы спецификации экономико-математических моделей Уметь формализовать постановки прикладных задач из сфер экономики и финансов
		3. Системно подходит к выбору математических методов и информационных технологий для решения конкретных финансово-экономических задач в профессиональной области.	Знать основные понятия теории вероятностей и математической статистики; методики расчетов, используемые при анализе данных; вероятностные и статистические методы Уметь использовать инструменты описательной статистики и визуализации данных языка R, вероятностные и статистические методы для решения профессиональных задач
		4. Анализирует результаты исследования математических моделей финансово-экономических задач и де-	Знать основные принципы программирования алгоритмов статистической обработки данных

		дает на их основании количественные и качественные выводы и рекомендации по принятию финансово-экономических решений.	в R(RStudio) для анализа и прогноза внутри- и внешнеэкономических процессов Уметь применять технические и инструментальные средства для решения прикладных экономических задач и подготовки обоснований управленческих решений
ПKN-6	Способность предлагать решения профессиональных задач в меняющихся финансово-экономических условиях	1. Понимает содержание и логику проведения анализа деятельности экономического субъекта, приемы обоснования оперативных, тактических и стратегических управленческих решений.	Знать методы анализа деятельности экономического субъекта Уметь использовать R(RStudio) в статистическом анализе данных для обоснования тактических и стратегических управленческих решений
		2. Предлагает варианты решения профессиональных задач в условиях неопределенности.	Знать основные принципы спецификации экономико-математических моделей, учитывающих неопределенность Уметь формализовать постановки прикладных задач из сфер экономики и финансов с учетом неопределенности
ПКП-2	Способность решать финансово-экономические задачи, проводить расчеты с использованием современных технических средств и информационных технологий в корпоративных финансах	1. Проводить необходимые для решения финансово-экономических задач, расчеты показателей с использованием современных технических средств и информационных технологий в корпоративных финансах.	Знать методы построения точечных и интервальных оценок параметров распределений случайных величин, методику проверки статистических гипотез. Уметь строить точечные и интервальные оценки финансово-экономических показателей как параметров распределений случайных величин, проверять статистические гипотезы

		2. Предлагает эффективные решения по реализации финансово-экономических задач.	<u>Знать</u> основные понятия и методы теории вероятностей и прикладной статистики, необходимые анализа финансово-экономических задач и процессов <u>Уметь</u> решать социально-экономические задачи вероятностными и статистическими методами
--	--	--	---

3. Место дисциплины в структуре образовательной программы

Дисциплина «Анализ данных на языке R» относится к Циклу профиля (элективный), по направлению подготовки 38.03.01 – Экономика, ОП «Корпоративные финансы», профиль: «Корпоративные финансы и инвестиции».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Таблица 1

Вид учебной работы по дисциплине	Всего (в з./е. и часах)	Семестр 6 (в часах)
Общая трудоемкость дисциплины	3/108	108
<i>Контактная работа</i> <i>- Аудиторные занятия</i>	34	34
Лекции	16	16
Семинары, практические занятия	18	18
<i>Самостоятельная работа</i>	74	74
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	зачет	зачет

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

1. Работа в среде R Studio. Назначение и основные возможности R. Интерфейс среды разработки Rstudio. Работа в консоли. Написание скрипта. Работа с историей. Рабочие директории и рабочие пространства. Установка дополнительных пакетов. Данные в экономике: объекты и признаки. Разведочный анализ и подтверждающий анализ. Программные средства анализа больших данных.

2. Данные и признаки. Понятие данных. Классификация признаков. Типы данных в R, импорт и экспорт данных в R. Импорт данных из csv-файлов. Пакет `xlsx`. Сохранение результатов работы. Формат `rds`.

3. Разведочный анализ в R. Объекты, признаки и таблицы. Типы признаков в экономике и управлении: интервальные, порядковые, ранговые, дихотомические. Построение описательной статистики, её элементы: среднее, стандартное отклонение, коэффициент асимметрии, коэффициент эксцесса, медиана, верхний и нижний квартиль, стандартная ошибка. Типы и назначение диаграмм. Пакет `ggplot2`. Динамическая визуализация. Пакет `plotly`. Формулировка гипотез.

4. Интервальные оценки характеристик в R. Понятие доверительного интервала, построение доверительных интервалов. Доверительные интервалы для параметров распределения, уровень надёжности. Построение доверительных интервалов для среднего и дисперсии, для вероятности события, для прогноза.

5. Изучение взаимосвязи между признаками. Корреляционный анализ, ковариация, корреляция, доверительный интервал для коэффициента корреляции. Ковариация и корреляция в R, коэффициенты корреляции для разных шкал. Основы регрессионного анализа в R.

6. Проверка статистических гипотез. Критерии согласия. Параметрические и непараметрические критерии. Критерии Стьюдента и Манна-Уитни. Дисперсионный анализ и критерий Краскела-Уоллиса. Post hoc тесты.

7. Многомерные статистические методы. Факторный анализ и метод главных компонент. Методы многомерной классификации: классификация с обучением и без обучения. Кластерный анализ. Дискриминантный анализ.

5.2. Учебно-тематический план

Таблица 2

п/п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоятельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1.	Работа в среде R Studio	14	4	2	2	10	Решение задач на практических занятиях. Самостоятельная работа. Опрос.
2.	Данные и признаки	14	4	2	2	10	Решение задач на практических занятиях. Самостоятельная работа. Опрос.
3.	Разведочный анализ в R	14	4	2	2	10	Решение задач на практических занятиях. Самостоятельная работа. Опрос.
4.	Интервальные оценки характеристик в R.	13	3	1	2	10	Решение задач на практических занятиях. Самостоятельная работа. Опрос.
5.	Изучение взаимосвязи между признаками.	13	3	1	2	10	Решение задач на практических занятиях. Самостоятельная работа. Опрос.
6.	Проверка статистических гипотез.	20	8	4	4	12	Решение задач на практических занятиях. Самостоятельная работа. Опрос.

7.	Многомерные статистические методы.	20	8	4	4	12	Решение задач на практических занятиях. Самостоятельная работа. Опрос.
В целом по дисциплине		108	34	16	18	74	Согласно учебному плану: контрольная работа
Итого в %		100	31	47	53	69	

*объем контактной работы в очно-заочной/заочной формах обучения и индивидуальных учебных планах определяется соответствующими учебными планами. Темы, реализуемые в виде контактной работы, определяются преподавателем самостоятельно, исходя из уровня их сложности.

Содержание семинаров, практических занятий

Таблица 3

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Работа в среде R Studio	Исследование функциональных возможностей среды разработки. Работа в консоли и в редакторе кода. История команд. Создание и удаление объектов. Загрузка и подключение библиотек. Экспорт и импорт данных. Сохранение настроек. <i>Рекомендуемые источники: (8.2, 8.3, 9.20)</i>	Опрос. Проверка самостоятельной работы. Решение задач в интерактивной форме.
Данные и признаки	Данные и признаки экономических объектов. Форматы данных в R. Работа с массивами. Работа с фреймами. Особенности тиббл-фреймов <i>Рекомендуемые источники: (8.1, 8.2, 8.3, 9.20)</i>	Опрос. Проверка самостоятельной работы. Решение задач в интерактивной форме
Разведочный анализ в R	Визуализации формы распределения. Группировка данных с целью визуализации и описательной статистики. <i>Рекомендуемые источники: (8.2, 8.3, 9.21)</i>	Решение задач в интерактивной форме, проверка самостоятельной работы
Интервальные оценки характеристик в R	Изучение функций R оценки доверительных интервалов для доли, математического ожидания, дисперсии. <i>Рекомендуемые источники: (8.1, 8.2.)</i>	Решение задач в интерактивной форме, проверка самостоятельной работы
Изучение взаимосвязи между признаками	Построение и визуализация корреляционной матрицы в R, визуализация попарных диаграмм рассеяния на фасеточном графике. <i>Рекомендуемые источники: (8.2, 9.21)</i>	Решение задач в интерактивной форме, проверка самостоятельной работы

Проверка статистических гипотез.	Инструменты R для проверки гипотез согласия, о значении математического ожидания и дисперсии, о равенстве математических ожиданий и дисперсии. <i>Рекомендуемые источники: (8.1, 8.2.)</i>	Решение задач в интерактивной форме, проверка самостоятельной работы
Многомерные статистические методы.	Метод главных компонент в R, кластерный анализ в R. <i>Рекомендуемые источники: (8.1)</i>	Решение задач в интерактивной форме, проверка самостоятельной работы

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Таблица 4

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Работа в среде R Studio	Способы и функционал импорта данных в R Studio.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к занятию.
Данные и признаки	Поиск наров данных, актуальных для профессиональной области в соответствии с профилем подготовки. Самостоятельный сбор данных.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к занятию.
Разведочный анализ в R	Распределение модифицированных признаков. Логнормальное распределение в экономике.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Интервальные оценки характеристик в R	Различия между оценками доверительных интервалов при помощи prop.test и BinomCI	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Изучение взаимосвязи между признаками	Корреляция порядковых признаков. Коэффициент корреляции рангов	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение

		домашних заданий к каждому занятию
Проверка статистических гипотез.	Незначимость коэффициентов корреляции.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Многомерные статистические методы	Факторный анализ.	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерный вариант контрольной работы

На основании набора данных, соответствующего номеру варианта, во всех работах требуется выполнить:

1. Предварительную обработку данных

Построить гистограммы для всех количественных признаков, столбиковые диаграммы – для качественных. Для указанных признаков построить гистограммы логарифмов указанных признаков. Для указанных признаков удалить выбросы. Для указанных пар признаков построить диаграммы рассеяния. Провести разбиение данных на поднаборы по указанному качественному признаку. Все задания выполнить средствами R.

2. Построение доверительных интервалов

Получить точечные оценки математических ожиданий и дисперсий всех количественных признаков. Оценки долей всех вариантов качественных признаков. Построить доверительные интервалы с надежностью γ для математических ожиданий и дисперсий количественных признаков, а также – для доли одной из вариантов любого качественного признака. Все задания выполнить средствами R.

Критерии бальной оценки различных форм текущего контроля успеваемости

Критерии бальной оценки различных форм текущего контроля успеваемости содержится в соответствующих методических рекомендациях Департамента математики Факультета информационных технологий и анализа больших данных.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по данной дисциплине

Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине содержится в разделе «2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, знаний и умений

Таблица 5

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
Способность осуществлять сбор, обработку и статистический анализ данных, применять математические методы для решения стандартных профессиональных финансово-экономических задач, интерпретировать полученные результаты (ПКН-3)	1. Проводит сбор, обработку и статистический анализ данных для решения финансово-экономических задач.	<u>Знать</u> технологию импорта данных разных форматов для последующей обработки в R <u>Уметь</u> осуществлять поиск, сбор и первичную подготовку данных для экономических исследований	С помощью средств языка R импортируйте набор данных об экономических объектах, представленный в формате .csv, удалите пропуски, охарактеризуйте типы признаков набора.
	2. Формулирует математические постановки финансово-экономических задач, переходит от экономических постановок задач к математическим моделям.	<u>Знать</u> основные принципы спецификации экономико-математических моделей <u>Уметь</u> формализовать постановки прикладных задач из сфер экономики и финансов	На основании выдвинутых предположений о представленных в наборе данных характеристиках объектов сформулируйте статистические гипотезы. В чем заключаются ошибки 1 и 2 родов для каждой гипотезы?

	3. Системно подходит к выбору математических методов и информационных технологий для решения конкретных финансово-экономических задач в профессиональной области.	Знать основные понятия теории вероятностей и математической статистики; методики расчетов, используемые при анализе данных; вероятностные и статистические методы Уметь использовать инструменты описательной статистики и визуализации данных языка R, вероятностные и статистические методы для решения профессиональных задач	С помощью средств Языка R импортируйте набор данных об экономических объектах, представленный в формате .csv, разбейте его на поднаборы по указанному преподавателем качественному признаку, выполните описательную статистику поднаборов данных
	4. Анализирует результаты исследования математических моделей финансово-экономических задач и делает на их основании количественные и качественные выводы и рекомендации по принятию финансово-экономических решений.	Знать основные принципы программирования алгоритмов статистической обработки данных в R(RStudio) для анализа и прогноза внутри- и внешнеэкономических процессов Уметь применять технические и инструментальные средства для решения прикладных экономических задач и подготовки обоснований управленческих решений	На основе предложенного набора данных осуществите группировку внешних и внутренних факторов методом главных компонент, интерпретируйте полученные компоненты
Способность предлагать решения профессиональных задач в меняющихся финансово-экономических условиях (ПКН-6)	1. Понимает содержание и логику проведения анализа деятельности экономического субъекта, приемы обоснования оперативных, тактических и стратегических управленческих решений.	Знать методы анализа деятельности экономического субъекта Уметь использовать R(RStudio) в статистическом анализе данных для обоснования тактических и стратегических управленческих решений	На основе предложенного набора данных проверьте гипотезы об однородности распределений данных количественного признака, сгруппированного по категориям качественного признака. Если это отвечает результатам проверки предыдущей гипотезы, проверьте гипотезу о равенстве математических ожиданий для разных пар поднаборов значений количественного признака

	2. Предлагает варианты решения профессиональных задач в условиях неопределенности.	<u>Знать</u> основные принципы спецификации экономико-математических моделей, учитывающих неопределенность <u>Уметь</u> формализовать постановки прикладных задач из сфер экономики и финансов с учетом неопределенности	На основании набора данных об изменении цен трех активов за последний год, определите наименее рискованный. Наиболее рискованный и Парето оптимальный актив. Получив точечные оценки риска средствами R
Способность решать финансово-экономические задачи, проводить расчеты с использованием современных технических средств и информационных технологий в корпоративных финансах (ПКП-2)	1. Проводить необходимые для решения финансово-экономических задач, расчеты показателей с использованием современных технических средств и информационных технологий в корпоративных финансах	<u>Знать</u> методы построения точечных и интервальных оценок параметров распределений случайных величин, методику проверки статистических гипотез. <u>Уметь</u> строить точечные и интервальные оценки финансово-экономических показателей как параметров распределений случайных величин, проверять статистические гипотезы	На основании предложенного набора данных дайте описательную статистику указанных преподавателем поднаборов количественных признаков, постройте доверительные интервалы для математических ожиданий и дисперсий количественных признаков. Задания выполните, используя наиболее подходящие средства языка R
	2. Предлагает эффективные решения по реализации финансово-экономических задач.	<u>Знать</u> основные понятия и методы теории вероятностей и прикладной статистики, необходимые анализа финансово-экономических задач и процессов. <u>Уметь</u> решать социально-экономические задачи вероятностными и статистическими Методами	На основании данных об изменении цен трех активов. проверить гипотезы о равенстве ожидаемых доходностей и уровней риска всех возможных пар активов

Примерные теоретические вопросы для подготовки к зачету

1. Типы данных в R.
2. Средства описательной статистики в R.
3. Средства визуализации данных в R.
4. Построение доверительных интервалов для математического ожидания, дисперсии и генеральной доли в R.
5. Статистические гипотезы. Понятие статистической гипотезы. Виды статистических гипотез: параметрические и непараметрические, простые и сложные.
6. Проверка гипотезы о равенстве математического ожидания гипотетическому значению. Проверка гипотезы о равенстве двух математических ожиданий в R.
7. Проверка гипотезы о равенстве дисперсии гипотетическому значению. Проверка гипотезы о равенстве двух дисперсий в R.
8. Проверка гипотезы о равенстве вероятности события гипотетическому значению. Проверка гипотезы о равенстве двух вероятностей в R.
9. Проверка гипотез о незначимости коэффициента корреляции в R.
10. Критерии согласия проверка гипотезы о нормальном распределении в R.
11. Проверка об однородности распределений в R.
12. Непараметрические критерии. Проверка гипотез на малых выборках. Критерий знаков.
13. Распределение Вилкоксона и его критические границы. Применение критерия в R.
14. Ранговая корреляция. Коэффициент ранговой корреляции Спирмена. Коэффициент ранговой корреляции Кендалла. Вычисление корреляционной матрицы в R.
15. Метод главных компонент в R.
16. Классификация без обучения: кластерный анализ в R.

Примерные задания для подготовки к зачету
(все задания выполняются с применением языка R)

Задача 1. По заданной двумерной выборке сгруппируйте данные по значениям атрибутивного признака, постройте описательную статистику для каждой группы. Укажите выборочное среднее, стандартное отклонение, медиану, верхнюю и нижнюю границу нормы, количество выбросов. Постройте для каждой группы гистограмму и ящик с усами.

Задача 2. По данным многомерной выборки, содержащей сведения о поле, возрасте, месте проживания и уровне доходов жителей региона N (приводятся данные)

- 1) проверьте гипотезу о равенстве распределения мужчин и женщин по возрасту, используя критерий однородности Колмогорова-Смирнова;
- 2) проверьте гипотезу о равенстве математических ожиданий доходов респондентов, проживающих в городской и в сельской местностях.

Задача 3. В таблице представлены адреса отелей и количество отзывов, оставленных на сайте, через который осуществлялось бронирование. Проверьте гипотезу о равенстве вероятностей получения отзыва каждым из представленных отелей.

Адрес отеля	Общее количество отзывов
s Gravesandestraat 55 Oost 1092 AA Amsterdam Netherlands	568215
1 15 Templeton Place Earl s Court Kensington and Chelsea London SW5 9NB United Kingdom	1036346
1 2 Serjeant s Inn Fleet Street City of London London EC4Y 1LL United Kingdom	2715903
1 3 Queens Garden Westminster Borough London W2 3BA United Kingdom	7752600
1 3 Rue d Argentine 16th arr 75116 Paris France	5985
1 5 Passage Ruelle 18th arr 75018 Paris France	3743
1 8 Russell Square Camden London WC1B 5BE United Kingdom	4488750
1 Addington Street Lambeth London SE1 7RY United Kingdom	13598091
1 Aldwych Westminster Borough London WC2B 4BZ United Kingdom	35224

Задача 4. Представьте данные, полученные из глобальной сети Интернет в

формате csv, в виде фрейма данных. Сколько пропущенных значений (N/A) содержит признак, размещенный в первом столбце? Каков объем многомерной выборки после удаления пропусков?

Задача 5. Дана трехмерная выборка показателей X , Y , Z предприятия ABC. По данным выборки вычислите коэффициенты корреляции Пирсона для каждой пары признаков, проверьте гипотезы о незначимости каждого из коэффициентов корреляции.

Задача 6. По результатам опроса в рамках исследования состояния малого предпринимательства ответы респондентов на определенный вопрос анкеты представлены в виде выборки (задается выборка). Используя Excel или R (RStudio), очистите выборку от пропусков, обозначенных как "NA", и ответьте на следующие ниже вопросы.

1. Определите объем исходной выборки.
2. Определите количество пропущенных данных "NA" в исходной выборке.
3. Определите объем очищенной от "NA" выборки.
4. Определите количество различных вариантов ответов респондентов, встречающиеся в очищенной выборке.
5. Определите количество респондентов, которые дали ответ (приводится конкретная варианта).
6. Определите долю респондентов, которые дали ответ (приводится конкретная варианта).
7. Определите левую границу 0,95-доверительного интервала для истинной доли ответов (приводится конкретная варианта).
8. Введите правую границу 0,95-доверительного интервала для истинной доли ответов (приводится конкретная варианта).

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Соловьев, В. И. Анализ данных в экономике: теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel: учебник для направления бакалавриата "Экономика и управление" / В. И. Соловьев; Финуниверситет. – Москва : Кнорус, 2019. - 498 с. - Текст : непосредственный. - То же. – 2023. - ЭБС BOOK.ru. - URL: <https://book.ru/book/946789> (дата обращения: 10.10.2023). — Текст : электронный.

2. Криволапов, С. Я. Статистические вычисления на платформе Jupyter Notebook с использованием Python: учебник / С. Я. Криволапов. — Москва : КноРус, 2022. — 431 с. — ЭБС BOOK.RU. - URL:<https://book.ru/book/943660> (дата обращения: 10.10.2023). — Текст : электронный.

3. Маркова, С. В. Анализ данных на языке R. : учебник и практикум / С. В. Маркова. — Москва : КноРус, 2023. — 216 с. — ЭБС BOOK.RU. — URL: <https://book.ru/book/948838> (дата обращения: 10.10.2023). — Текст : электронный.

Дополнительная литература:

4. Глебов, В. И. Анализ данных в экономике. Сборник задач: учебник / В. И. Глебов, С. Я. Криволапов. — Москва : КноРус, 2022. — 578 с. — ЭБС BOOK.ru. — URL: <https://book.ru/book/943011> (дата обращения: 10.10.2023). — Текст : электронный.

5. Калинина, В. Н. Анализ данных: Компьютерный практикум: учебное пособие / В. Н. Калинина, В. И. Соловьев. — Москва : КНОРУС, 2017. — 166 с. — (Бакалавриат). — Текст : непосредственный. — То же 2022. — ЭБС BOOK.ru. — URL: <https://www.book.ru/book/942681> (дата обращения: 10.10.2023). — Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Информационно-образовательный портал Финансового университета при Правительстве Российской Федерации <http://org.fa.ru/>
2. Сайт Департамента математики Факультета информационных технологий и анализа больших данных. <http://www.fa.ru/org/faculty/pmit/dm/Pages/Home.aspx>
3. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
4. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
5. Электронно-библиотечная система «Университетская библиотека ОНАЙН» <http://biblioclub.ru/>
6. Электронно-библиотечная система Znanium <http://www.znaniy.com>
7. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
8. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
9. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
10. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
11. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
12. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
13. Национальная электронная библиотека <http://нэб.рф/>
14. Массовый открытый онлайн-курс/специализация «Машинное обучение и анализ данных» / МФТИ и Яндекс. – <https://www.coursera.org/specializations/machine-learning-data-analysis>
15. Массовый открытый онлайн-курс/специализация “Recommender Systems”/ niversity of Minnesota – <https://www.coursera.org/specializations/recommender-systems>

16. Массовый открытый онлайн-курс/специализация “Machine Learning”/Stanford University - <https://www.coursera.org/learn/machine-learning/home/welcome>

17. Профессиональный ресурс по машинному обучению. - <https://stackoverflow.com> (<https://stackoverflow.com/>)

18. Профессиональный ресурс по машинному обучению. - <https://stackexchange.com> (<https://stackexchange.com/>)

19. Платформа для соревнований по машинному обучению. – www.kaggle.com (<http://www.kaggle.com/>)

20. Платформа для создания и размещения образовательных материалов. – <https://welcome.stepik.org/ru>, курс «Анализ данных в R, части 1 и 2»

21. Платформа для создания и размещения образовательных материалов. – <https://welcome.stepik.org/ru>, курс «Графика в R»

10. Методические указания для обучающихся по освоению дисциплины

Студентам при подготовке следует использовать нормативные документы Финансового университета, Методические рекомендации по планированию и организации внеаудиторной самостоятельной работы студентов по образовательным программам бакалавриата и магистратуры в Финансовом университете, утвержденные приказом Финуниверситета от 11.05.2021 г. № 1040 (см. сайт Финансового Университета: на главной странице раздел «Наш университет»; далее «Единая правовая база Финуниверситета»), использовать методические рекомендации департамента.

Самостоятельная работа студентов проходит аудиторно и внеаудиторно. Организации самостоятельной работы служит учебно-тематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного изучения.

Домашние задания следует выполнять регулярно при подготовке к практическим занятиям. В большинстве своем задания являются типовыми, и образцы их

решения содержатся в рекомендованных пособиях, в материале лекций и практических занятий. Если то или иное задание вызвало затруднение, необходимо обратиться к преподавателю на консультации или ближайшем практическом занятии. Регулярность в выполнении домашних заданий — важный фактор освоения дисциплины. Даже небольшие отклонения от графика могут спровоцировать серьезное отставание и в дальнейшем — риск получения неудовлетворительных оценок в ходе текущей и промежуточной аттестации. Для выполнения домашних заданий следует завести отдельную тетрадь. Контроль выполнения домашних заданий осуществляется в ходе практических занятий в процессе выборочного собеседования.

Домашняя контрольная работа является одной из основных форм текущего контроля самостоятельной работы студентов по дисциплине «Анализ данных на языке R». Примерное время их выполнения составляет 4 часа. Каждый вариант домашней контрольной работы (ДКР) содержит несколько задач, выполняя которые студент демонстрирует умение решать типовые эконометрические задачи и проводить типовые расчеты на компьютере. Сроки выполнения ДКР указываются в учебно-тематическом плане изучения дисциплины. Конкретные сроки сдачи ДКР устанавливаются преподавателем. Оценка за ДКР выставляется по итогам проверки отчета и устного собеседования по работе. Эта оценка является существенной компонентой оценки самостоятельной работы студента в течение семестра.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11. 1. Комплект лицензионного программного обеспечения

1. Windows, Microsoft Office.
2. Антивирус Kaspersky

11.2. Современные профессиональные базы данных и информационные справочные системы

1. Информационно-правовая система «Гарант»

2. Информационно-правовая система «Консультант Плюс»

3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

4. Система комплексного раскрытия информации «СКРИН» -
<http://www.skrin.ru/>

11.3. Сертифицированные программные и аппаратные средства защиты информации

Не используются

11.4. Электронная таблица Calc LibreOffice

11.5. Прикладной [программный пакет](#) для [эконометрического моделирования](#) Gretl: <http://gretl.sourceforge.net/>

11.6. Программная среда R.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Практические занятия по дисциплине проходят в компьютерном классе.