

**Программа учебной дисциплины
«Анализ и визуализация данных»**

Утверждена
Академическим советом ОП
Протокол № 6 от 26. 08. 2019 г.

Разработчики	Деркач Денис Александрович, Доцент, Департамент больших данных и информационного поиска Кантонистова Елена Олеговна, Приглашенный преподаватель, Департамент больших данных и информационного поиска Ананьева Марина Евгеньевна, Приглашенный преподаватель, Департамент больших данных и информационного поиска
Число кредитов	4
Контактная работа (час.)	40
Самостоятельная работа (час.)	112
Курс, Образовательная программа	2 (Б) курс, Иностранные языки и межкультурные коммуникации
Формат изучения дисциплины	С использованием онлайн курса

1. Цель, результаты освоения дисциплины и пререквизиты

Цели:

1. ознакомление студентов с основами науки о данных
2. формирование у студентов практических навыков работы с данными, решения прикладных задач анализа данных по специальности и визуализации данных
3. формирование у студентов навыков поиска информации

Планируемые результаты:

1. Знать основы работы с MS Excel
2. Находить данные под конкретный анализ
3. Строить разведочный анализ данных
4. Применять регрессионный анализ
5. Интерпретировать регрессионный анализ
6. Формулировать статистические гипотезы
7. Проверять статистические гипотезы
8. Проводить дисперсионный анализ
9. Выполнять полный цикл анализа данных

Пререквизиты:

1. Цифровая грамотность

2. Содержание учебной дисциплины

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	лк		
	см		
	онл/ср		
Простейшие функции MS Excel	0	<ul style="list-style-type: none"> Знать основы работы с MS Excel 	КР1, КР2, ЭКЗ, ДЗ1,
	0		
	24		
Введение в анализ данных	6	<ul style="list-style-type: none"> Находить данные под конкретный анализ Строить разведочный анализ данных 	КР2, ПР, ЭКЗ, Т1, Т2, Т3, ДЗ1,
	6		
	24		
Регрессионный анализ	4	<ul style="list-style-type: none"> Применять регрессионный анализ Интерпретировать регрессионный анализ Проводить дисперсионный анализ 	КР3, ЭКЗ, Т4, Т5, Т8, ДЗ1,
	4		
	20		
Проверка статистических гипотез	6	<ul style="list-style-type: none"> Формулировать статистические гипотезы Проверять статистические гипотезы Проводить дисперсионный анализ 	КР3, ЭКЗ, Т6, Т7, Т8, ДЗ1,
	8		
	28		
Построение анализа данных	2	<ul style="list-style-type: none"> Находить данные под конкретный анализ Выполнять полный цикл анализа данных 	ПР, ЭКЗ, Т1, Т2, Т3,
	4		
	16		
Часов по видам учебных занятий:	18		
	22		
	112		
Итого часов:	152		

Содержание разделов дисциплины:

1. Простейшие функции MS Excel

Основные навыки работы с MS Excel, открытие документов, редактирование таблиц, создание простейших графиков. Курс проходит

онлайн на основе курса "Excel essentials", доступного на сайте coursera:
<https://ru.coursera.org/learn/excel-essentials>

2. **Введение в анализ данных**

Постановка задачи анализа данных. Основные применения анализа данных в лингвистике. Поиск данных. Разведочный анализ данных.

3. **Регрессионный анализ**

Регрессионный анализ. Построение и интерпретация. Интервальные оценки и погрешность измерений.

4. **Проверка статистических гипотез**

Постановка задачи проверки гипотезы. Подбор тестовых статистик. Критерии согласия, двухвыборочные критерии. Дисперсионный анализ. Проверка множественных гипотез. Коррекция Бонферрони.

5. **Построение анализа данных**

Основы анализа данных. Цикл анализа данных. Базовые понятия.

3. Оценивание

ID	Статус	Тип	Проверяемые результаты
КР1	Не блокирующее	Оценка онлайн курса	<ul style="list-style-type: none"> Знать основы работы с MS Excel
КР2	Не блокирующее	Контрольная работа	<ul style="list-style-type: none"> Знать основы работы с MS Excel Строить разведочный анализ данных
КР3	Не блокирующее	Контрольная работа	<ul style="list-style-type: none"> Применять регрессионный анализ Интерпретировать регрессионный анализ Формулировать статистические гипотезы Проверять статистические гипотезы
ПР	Не блокирующее	Командный проект	<ul style="list-style-type: none"> Находить данные под конкретный анализ Выполнять полный цикл анализа данных
ЭКЗ	Не блокирующее	Экзамен (письменный)	<ul style="list-style-type: none"> Знать основы работы с MS Excel Строить разведочный анализ данных Применять регрессионный анализ Интерпретировать регрессионный анализ Формулировать статистические гипотезы Проверять статистические гипотезы Проводить дисперсионный анализ

			<ul style="list-style-type: none"> • Выполнять полный цикл анализа данных
T1	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Находить данные под конкретный анализ • Строить разведочный анализ данных
T2	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Находить данные под конкретный анализ • Строить разведочный анализ данных
T3	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Находить данные под конкретный анализ • Строить разведочный анализ данных
T4	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Интерпретировать регрессионный анализ
T5	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Применять регрессионный анализ • Интерпретировать регрессионный анализ
T6	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Формулировать статистические гипотезы
T7	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Формулировать статистические гипотезы • Проверять статистические гипотезы
T8	Не блокирующее	Компьютерное тестирование	<ul style="list-style-type: none"> • Формулировать статистические гипотезы • Проверять статистические гипотезы • Проводить дисперсионный анализ
Д31	Не блокирующее	Домашнее задание	<ul style="list-style-type: none"> • Знать основы работы с MS Excel • Строить разведочный анализ данных • Применять регрессионный анализ • Интерпретировать регрессионный анализ • Формулировать статистические гипотезы • Проверять статистические гипотезы

- | | | | |
|--|--|--|--|
| | | | <ul style="list-style-type: none">• Проводить дисперсионный анализ |
|--|--|--|--|

Формула округления: Стандартное арифметическое округление

Вид формулы оценивания: Линейная

Формула оценивания:

$$O_{\text{тест}} = 1/8*(O_{\text{T1}}+O_{\text{T2}}+O_{\text{T3}}+O_{\text{T4}}+O_{\text{T5}}+O_{\text{T6}}+O_{\text{T7}}+O_{\text{T8}})$$

$$O_{\text{практ}} = 1/2*(O_{\text{Практ1}}+O_{\text{Практ2}})$$

$$O_{\text{КР}} = 1/3*(O_{\text{КР1}}+O_{\text{КР2}}+O_{\text{КР3}})$$

$$O_{\text{ДЗ}} = O_{\text{ДЗ1}}$$

$$O_{\text{Проект}} = O_{\text{ПР1}}$$

$$O_{\text{накопл}} =$$

$$\text{Округление}(0.25*O_{\text{Проект}}+0.2*O_{\text{тест}}+0.15*O_{\text{Практ}}+0.2*O_{\text{КР}}+0.2*O_{\text{ДЗ}})$$

$$O_{\text{итог}} = \text{Округление}(0.8*O_{\text{накопл}} + 0.2*O_{\text{экз}})$$

Онлайн курс проходит исключительно с корпоративного почтового адреса студента. Подключение студентов к онлайн курсу на платформе Coursera (<https://coursera.org/>) производит Дирекция по онлайн обучению НИУ ВШЭ по заявке администратора учебного офиса образовательной программы. Скрытая сессия для студентов ВШЭ автоматически появляется в личном аккаунте на платформе. Регистрироваться на открытую сессию для всех желающих слушателей нельзя.

Если студент проходит онлайн курс в открытой сессии для всех желающих слушателей или с личного почтового адреса, результат прохождения онлайн курса не засчитывается.

Академическая этика

Преподаватель оставляет за собой право устроить устную защиту любой из форм контроля. Если при проверке работ (текущий и итоговый контроль) установлен факт нарушения академической этики, студент получает оценку «0» за данную работу. Работа студента, предоставившего свою работу для списывания, также аннулируется, к обоим студентам применяется дисциплинарное взыскание.

В случае нарушения правил проведения экзамена студент удаляется с экзамена с оценкой «0». К нарушениям правил проведения экзамена относятся: коммуникация с другими студентами во время выполнения работы, использование социальных сетей/телефона во время экзамена (с любой целью), списывание.

4. Примеры оценочных средств

Примеры вопросов итоговой аттестации:

1. Как считается среднее выборки?

2. Сформулируйте нулевую и альтернативную гипотезы для проверки среднего.
3. Постройте скользящее среднее по заданной выборке.

5. Ресурсы

5.1. Рекомендуемая основная литература

п/п	Наименование
1	<i>Евсеев, Е. А.</i> Эконометрика : учебное пособие для бакалавриата и специалитета / Е. А. Евсеев, В. М. Буре. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019. — 186 с. — (Бакалавр и специалист). — ISBN 978-5-534-10752-4. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://www.biblio-online.ru/bcode/431441 (дата обращения: 27.08.2019).
2	<i>Мхитарян, В. С.</i> Анализ данных в MS Excel : основные сведения о MS Excel, статистические таблицы и графики, статистические функции, пакет анализа (анализ данных) : учеб. пособие для вузов, М.: КУРС, 2018. – 364 с. - ISBN 978-5-906923-26-4.

5.2. Рекомендуемая дополнительная литература

п/п	Наименование
1	<i>Яковлев, В. Б.</i> Статистика. Расчеты в Microsoft Excel : учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019. — 353 с. — (Университеты России). — ISBN 978-5-534-01672-7. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://www.biblio-online.ru/bcode/437852 (дата обращения: 27.08.2019).

5.3. Программное обеспечение

п/п	Наименование	Условия доступа/скачивания
1	MS Excel	<i>Из внутренней сети университета (договор)</i>

5.4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

Не требуется.

5.5. Материально-техническое обеспечение дисциплины

Учебные аудитории для занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

6. Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида, могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

6.1.1. *для лиц с нарушениями зрения:* в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); в печатной форме на языке Брайля; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

6.1.2. *для лиц с нарушениями слуха:* в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

6.1.3. *для лиц с нарушениями опорно-двигательного аппарата:* в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.