

**Программа учебной дисциплины
«Анализ данных»**

Утверждена
Академическим советом ОП
Протокол № _____ от _____ .20 _____

Разработчик	Меликян Алиса Валерьевна, Старший преподаватель, Департамент программной инженерии
Число кредитов	3
Контактная работа (час.)	52
Самостоятельная работа (час.)	62
Курс, Образовательная программа	1 (М) курс, МАГОЛЕГО
Формат изучения дисциплины	Без использования онлайн курса

1. Цель, результаты освоения дисциплины и пререквизиты

Цели:

1. Овладение современными методами анализа данных и навыками работы с данными с использованием современных программных средств для проведения исследований.

Планируемые результаты обучения (ПРО):

1. Уметь провести описательный анализ данных
2. Уметь исследовать и дать количественную оценку взаимосвязям между переменными разного типа
3. Уметь построить модель линейной регрессии и описать её
4. Уметь построить модель логистической регрессии и описать её
5. Уметь провести факторный анализ данных
6. Уметь провести кластерный анализ данных
7. Уметь осуществлять выбор подходящего метода анализа данных для проведения конкретного исследования в соответствии с целями, задачами, гипотезами и имеющимися в наличии данными

Пререквизиты:

1. Знание основ математической статистики и теории вероятностей
2. Обладание навыками работы с компьютерными программами

2. Содержание учебной дисциплины

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	лк		
	см		
	онл/ср		
Введение в анализ данных	1	№: 1, 7.	АР, КР.
	0		
	0		
Описательный анализ данных	3	№: 1, 2, 3, 4, 5, 7.	АР, КР.

	4		
	8		
Исследование взаимосвязей между переменными	4	№: 2, 3, 4, 7.	АР, КР.
	6		
	8		
Линейная регрессия	4	№: 2, 3, 7.	АР, КР.
	4		
	14		
Логистическая регрессия	4	№: 4, 7.	АР.
	4		
	10		
Факторный анализ	4	№: 5, 7.	АР.
	4		
	10		
Кластерный анализ	4	№: 7.	АР.
	6		
	12		
Часов по видам учебных занятий:	24		
	28		
	62		
Итого часов:	114		

Содержание разделов дисциплины:

1. Введение в анализ данных

Обзор методов анализа данных. Обзор статистических пакетов. Источники данных. Ввод, редактирование, модификация экспорт/импорт данных и результатов.

2. Описательный анализ данных

Частотный анализ. Графический анализ. Описательная статистика. Оценка распределения значений переменной. Основные типы шкал и соответствующие им меры средней тенденции и меры разброса. Нормальное распределение, Z-стандартизация, тест Колмогорова-Смирнова. Работа с многовариантными вопросами.

3. Исследование взаимосвязей между переменными

Таблица сопряженности. Формулировка гипотез. Этапы проверки гипотез. Уровень значимости и ошибка первого рода. Тест Хи-квадрат. Построение диаграммы рассеяния. Парные коэффициенты корреляции (Пирсона, Кендалла, Спирмана). Частные корреляции. Сравнение средних (t-тест для независимых и зависимых выборок, однофакторный дисперсионный анализ). Непараметрические тесты.

4. Линейная регрессия

Задачи регрессионного анализа. Множественная линейная регрессия. Оценка качества модели. Анализ остатков. Мультиколлинеарность. Гетероскедастичность. Фиктивные переменные. Диагностика регрессионной модели.

5. Логистическая регрессия

Особенности модели логистической регрессии. Оценка качества модели. Предельные эффекты. Анализ остатков. Мультиколлинеарность. Гетероскедастичность. Фиктивные переменные. Диагностика регрессионной модели.

6. Факторный анализ

Порядок выполнения факторного анализа. Оценка пригодности исходных данных для проведения факторного анализа. Метод главных компонент. Факторные нагрузки. Вращение осей. Сохранение факторов в виде новых переменных в файле данных. Интерпретация значений факторов.

7. Кластерный анализ

Иерархический кластерный анализ. Кластерный анализ методом к-средних. Содержательная характеристика кластеров.

3. Оценивание

- **АР**, Не блокирующее, Работа на семинаре
Задания, выполняемые на компьютере во время семинаров. Оцениваются по шкале от 0 до 10.
- **КР**, Не блокирующее, Контрольная работа
Письменная контрольная работа проводится в конце 3-го модуля с целью оценки результатов освоения пройденного материала. Оцениваются по шкале от 0 до 10.
- **ДЗ**, Не блокирующее, Домашнее задание
Домашнее задание выполняется в группе из 2-3 человек и сдаётся в конце 3-го модуля. Представляет собой исследовательский проект по сбору и анализу данных. Предусматривает подготовку аналитического отчёта по результатам анализа. Оцениваются по шкале от 0 до 10.
- **Э**, Не блокирующее, Экзамен (письменный)
Письменный экзамен выполняется с целью оценки результатов освоения пройденного материала. Оцениваются по шкале от 0 до 10.

Формула округления: Стандартное арифметическое округление

Шкала оценки: Десятибалльная

Вид формулы оценивания: Линейная

Формула оценивания:

Окончательная оценка = Округление($0,3*Э + 0,3*КР + 0,2*ДЗ + 0,2*АР$)

4. Примеры оценочных средств

Примеры заданий экзамена:

Оцените распределение значений переменной.

Оцените взаимосвязь между переменными на основе таблицы сопряжённости и теста Хи-квадрат.

Проведите корреляционный анализ данных.

Постройте модель множественной линейной регрессии. Оцените параметры качества модели. Охарактеризуйте взаимосвязи между предикторами и зависимой переменной.

Проведите факторный анализ данных и интерпретируйте его результаты.

Проведите кластерный анализ данных и интерпретируйте его результаты.

5. Ресурсы

5.1. Рекомендуемая основная литература

П/п	Наименование
1	Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS, Крыштановский, А. О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов / А. О. Крыштановский. – М.: ГУ-ВШЭ, 2006. – 281 с. – (Сер. "Учебники Высшей школы экономики"). - ISBN 5-7598-0373-
2	Доугерти К. Введение в эконометрику, Доугерти, К. Введение в эконометрику: учебник для вузов / К. Доугерти; Пер. с англ. О. О. Замкова, и др.. – Изд. 3-е. – М.: ИНФРА-М, 2010. – 465 с. – (Сер. "Университетский учебник"). - ISBN 978-5-16-003640-3.

5.2. Рекомендуемая дополнительная литература

п/п	Наименование
1	Вербик М. Путеводитель по современной эконометрике, Вербик, М. Путеводитель по современной эконометрике: учеб. пособие для вузов / М. Вербик; Пер. с англ. В. А. Банникова; Науч. ред., предисл. С. А. Айвазяна. – М.: Научная книга, 2008. – 615 с. – (Б-ка Солев). - НП. - ISBN 5-913930-35-4.
2	Борзых Д. А., Б. Б. Демешев Эконометрика в задачах и упражнениях, Борзых, Д. А. Эконометрика в задачах и упражнениях / Д. А. Борзых, Б. Б. Демешев. – Изд. 2-е, перераб. и существенно доп. – М.: УРСС: ЛЕНАНД, 2017. – 293 с. - ISBN 9785971040620.
3	Mirkin B. Core concepts in data analysis: summarization, correlation and visualization, Mirkin, B. Core concepts in data analysis: summarization, correlation and visualization / B. Mirkin. – London [etc.]: Springer, 2011. pp. 390. ISBN 978-0-85729-286-5.

5.3. Программное обеспечение

п/п	Наименование	Условия доступа/скачивания
1	Microsoft Windows 7 Professional RUS Microsoft Windows 8.1 Professional RUS Microsoft Windows 10	Из внутренней сети университета (договор)
2	Microsoft Office Professional Plus 2010	Из внутренней сети университета (договор)
3	Stata	Из внутренней сети университета (договор)

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

п/п	Наименование	Условия доступа/скачивания
Профессиональные базы данных, информационно-справочные системы		
1	Электронно-библиотечная система Юрайт	URL: https://biblio-online.ru/
2	OECD iLibrary	Из внутренней сети университета (договор)
3	World Bank e-Library	Из внутренней сети университета (договор)
4	Единый архив экономических и социологических данных	URL: https://openedu.ru/
Интернет-ресурсы (электронные образовательные ресурсы)		
1	Открытое образование	URL: https://openedu.ru/

5.5. Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных 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.

Учебные аудитории для семинарских и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

Компьютерные классы оборудованы ПЭВМ с доступом в Интернет, операционными системами и программным обеспечением, необходимыми для освоения дисциплины. При необходимости допускается замена оборудования его виртуальными аналогами.

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.