

## Программа учебной дисциплины «Анализ данных»

Утверждена

Академическим советом ООП

Протокол № от «\_\_» \_\_\_\_ 20\_\_ г.

Автор	Меликян Алиса Валерьевна
Число кредитов	3
Контактная работа (час.)	60
Самостоятельная работа (час.)	54
Курс	1
Формат изучения дисциплины	Очный (full time)

### I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Основные цели освоения дисциплины «Анализ данных»:

- овладение студентами методов анализа структурированных и неструктурированных данных с использованием современных программных средств;
- формирование у студентов практических навыков работы с различными типами данных (статистические данные, тексты, мультимедиа) в программах SPSS и ATLAS.ti.

В результате освоения дисциплины студент должен:

#### **знать:**

- существующие возможности анализа структурированных и неструктурированных данных;
- методы количественного анализа данных;
- методы качественного контент-анализа данных;
- как осуществить предварительную подготовку данных для последующей работы с ними в программах по анализу количественных и качественных данных;
- как выбрать подходящий метод анализа в зависимости от типа данных и исследовательской задачи;
- как интерпретировать результаты анализа данных и представлять их в доступном для широкой аудитории виде.

#### **уметь:**

- осуществлять ввод данных и импорт данных в SPSS и ATLAS.ti из разных источников;
- осуществлять выбор подходящего метода анализа данных для проведения конкретного исследования в соответствии с целями, задачами, гипотезами и имеющимися в наличии данными;
- проводить анализ данных в программах SPSS и ATLAS.ti;
- экспортировать результаты анализа данных в другие программы;
- приводить результаты проведённого анализа к виду, доступному для представления широкой аудитории.

#### **владеть:**

- навыками подготовки данных для работы с ними в SPSS и ATLAS.ti;
- навыками анализа данных в SPSS и ATLAS.ti;
- навыками представления результатов анализа в презентациях и отчётах.

Изучение дисциплины «Анализ данных» базируется на следующих дисциплинах:

- математическая статистика и теория вероятностей;
- информатика в объёме средней школы.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- знать основы математической статистики;
- знать основы теории вероятностей;
- обладать навыками работы с компьютерными программами.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- Научно-исследовательский семинар.

## **II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ**

### **Тема 1. Возможности анализа количественных данных в статистическом пакете SPSS.**

О статистическом пакете SPSS. Интерфейс статистического пакета SPSS (редактор данных, окно вывода, разделы меню, панели инструментов открытие и сохранение файлов). Ввод, редактирование, модификация экспорт/импорт данных и результатов. Обзор доступных методов анализа данных.

#### **Тема 2. Описательный анализ данных.**

Частотный анализ. Графический анализ (гистограмма, ящичковая диаграмма, диаграмма «ствол-лист»). Подсчёт статистических характеристик (мода, медиана, среднее арифметическое, дисперсия и среднее квадратичное отклонение, стандартная ошибка среднего, доверительный интервал, квартили, межквартильная ширина, симметричность и заострённость распределения). Основные типы шкал и соответствующие им меры средней тенденции и меры разброса. Нормальное распределение, Z-стандартизация, тест Колмогорова-Смирнова. Работа с многовариантными вопросами.

#### **Тема 3. Исследование взаимосвязей между переменными. Проверка гипотез.**

Таблица сопряжённости. 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.

#### **Тема 4. Регрессионный анализ.**

Задачи регрессионного анализа. Простая линейная регрессия. Множественная регрессия. Логистическая регрессия. Нелинейная регрессия. Оценка качества модели. Анализ остатков. Мультиколлинеарность. Гетероскедастичность. Фиктивные переменные. Приближение с помощью кривых. Диагностика регрессионной модели.

#### **Тема 5. 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. Интерпретацию значимых факторов.

#### **Тема 6. Кластерный анализ.**

Иерархический кластерный анализ. Кластерный анализ методом k-средних. Сохранение переменной, идентифицирующей принадлежность наблюдения к кластеру. Содержательная характеристика кластеров.

#### **Тема 7. Контент-анализ как метод исследования.**

Типы контент-анализа и области их применения. Содержание основных процедур контент-анализа. Определение цели и задач контент-анализа. Этапы проведения контент-анализа. Обзор компьютерных программ, используемых для проведения контент-анализа.

## Тема 8. Структурирование и кодирование текстов, изображений, видео и аудио материалов в программе ATLAS.ti.

Рабочая среда ATLAS.ti. Структура и содержание пользовательского справочника по программе ATLAS.ti. Создание и редактирование проекта в ATLAS.ti. Структурирование и кодирование текстов, изображений, видео и аудио материалов. Создание сетей отношений между кодами. Выявление тенденций и взаимосвязей, проверка гипотез. Функции "Coding", "Quoting", "Memos". Визуализация результатов анализа с использованием "Network Views". Формулировка результатов анализа и подготовка отчёта.

### III. ОЦЕНИВАНИЕ

Предусмотрены следующие формы контроля знаний студентов:

Тип контроля	Форма контроля	1 год		Параметры
		3	4	
Текущий	Домашнее задание		*	Провести анализ данных и подготовить отчёт по результатам.
Текущий	Контрольная работа	*		Работа на компьютере.
Итоговый	Экзамен		+	Работа на компьютере. Оценка результатов работы в течение 3-х дней.

На текущем контроле в конце 3-го модуля обучения студент должен продемонстрировать навыки анализа структурированных данных с использованием программы SPSS, по темам, пройденным в течение 3-го модуля.

На итоговом контроле студент должен продемонстрировать навыки самостоятельного поиска подходящего метода анализа данных разного типа, которые могут быть применимы для решения поставленного исследовательского вопроса, интерпретации и представления результатов анализа, формулировки выводов на основе проведённого анализа данных.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

Преподаватель оценивает работу студентов на семинарских занятиях: активность при ответе на вопросы преподавателя, правильность выполнения заданий на семинарах. Оценки за работу на семинарских занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на семинарских занятиях определяется перед итоговым контролем -  $O_{аудиторная}$ .

Преподаватель оценивает самостоятельную работу студентов: правильность выполнения домашнего задания. Накопленная оценка по 10-ти балльной шкале за домашнее задание определяется перед итоговым контролем –  $O_{дом. задание}$ .

Преподаватель осуществляет текущий контроль знаний студентов в форме контрольной работы в конце 3-го модуля –  $O_{к/р}$ .

Способ округления накопленной оценки текущего контроля: в пользу студента.

Результирующая оценка выставляется по следующей формуле:

$$O_{результ.} = 0,3 \cdot O_{экзамен} + 0,3 \cdot O_{к/р} + 0,2 \cdot O_{дом. задание} + 0,2 \cdot O_{аудиторная},$$

где  $O_{экзамен}$  – оценка за работу, выполненную на письменном экзамене:

Способ округления результирующей оценки: в пользу студента.

На пересдаче студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль.

На экзамене студент может получить дополнительную практическую задачу, которая оценивается в 1 балл. Таким образом, результирующая оценка, получаемая на пересдаче, выставляется по формуле

$$O_{результ.} = (0,3 \cdot O_{экзамен} + 0,3 \cdot O_{к/р} + 0,2 \cdot O_{дом. задание} + 0,2 \cdot O_{аудиторная}) + O_{доп. вопрос}$$

#### IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

**Вопросы для оценки качества освоения дисциплины в рамках текущего контроля студентов:**

- 1) В чём отличие между количественными и качественными данными, структурированными и неструктурированными данными? Приведите примеры.
- 2) Что такое исследовательская гипотеза? Приведите примеры.
- 3) Какие существуют коэффициенты корреляции и в чём их отличия?
- 4) Каковы цели регрессионного анализа? Каковы требования к данным для проведения регрессионного анализа?
- 5) Каковы цели дисперсионного анализа данных?
- 6) Каковы цели факторного анализа? Каковы требования к данным для проведения факторного анализа?
- 7) Каковы цели кластерного анализа? Каковы требования к данным для проведения кластерного анализа?
- 8) Что такое контент-анализ и какова его область применения?
- 9) Перечислите типы контент-анализа и их особенности.
- 10) Каковы основные этапы контент-анализа?
- 11) Какие компьютерные программы используются для проведения контент-анализа?
- 12) В чём особенности, преимущества и недостатки применения программы ATLAS.ti для качественного анализа данных?
- 13) Что такое кодирование и каковы особенности его применения для контент-анализа?
- 14) Что такое герменевтическая единица в ATLAS.ti?
- 15) Опишите процесс кодирования данных в ATLAS.ti.

**Примеры заданий итогового контроля**

- 1) Проанализировать основные характеристики распределения (мода, смещение и т.д.);
- 2) Определить наличие выбросов и экстремумов;
- 3) Оценить близость распределения к нормальному;
- 4) Построить таблицу сопряженности и определить наличие зависимости между переменными, тесноту и направление связи. Найти факторы, наиболее полно объясняющие наблюдаемые связи между переменными.
- 5) Построить уравнение регрессии и оценить адекватность модели.

- 6) Сформировать группы наблюдений с помощью кластерного анализа.
- 7) Сформировать факторы для выбранных переменных.
- 8) Провести контент-анализ текстов, изображений видео и аудио материалов в ATLAS.ti. Интерпретировать результаты и сформулировать выводы.

### **Пример домашнего задания**

Необходимо сформировать файл с данными в SPSS, содержащий не менее 10 переменных с разными типами шкал и не менее 100 наблюдений. Данные могут быть взяты из любых открытых источников, например, <http://www.hse.ru/rlms>, <http://sophist.hse.ru/>.

Нужно провести мини-исследование, в котором необходимо применить все знания и навыки, полученные в ходе курса:

- Провести описательный анализ данных (частотные таблицы, статистические показатели, графики),
- Рассчитать новые переменные на основе существующих,
- Исследовать взаимосвязи (таблицы сопряженности, корреляции),
- Построить регрессионные модели и оценить их.
- Провести дисперсионный, факторный и кластерный анализы.

По результатам анализа нужно подготовить доклад. В докладе должны быть приведены исследовательские вопросы и гипотезы, а также основные результаты исследования.

## **V. РЕСУРСЫ**

### **V.1 Основная литература**

1. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «ДиасофтЮП», 2002.
2. Введение в эконометрику: учебник для вузов / К. Доугерти; Пер. с англ. О. О. Замкова, Е. Н. Лукаша, О. Ю. Шибалкина; Науч. ред. пер. О. О. Замков. – Изд. 3-е. – М.: ИНФРА-М, 2009. – 464 с. – (Сер. "Университетский учебник"). ISBN 978-5-16-003640

### **V.2 Дополнительная литература**

1. Крыштановский, А. О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов / А. О. Крыштановский. – М.: ГУ-ВШЭ, 2006. – 281 с. ISBN 5-7598-0373-5.
2. Таганов, Д. SPSS: Статистический анализ в маркетинговых исследованиях / Д. Таганов. – СПб.: Питер, 2005. – 190 с. ISBN 5-469-00640-9.
3. Наследов, А. SPSS 19: профессиональный статистический анализ данных / А. Наследов. – СПб.: Питер, 2011. – 399 с. - НП. - ISBN 978-5-459-00344-4.
4. ATLAS.ti 7 User Guide and Reference, 2013 [Электронный ресурс]. – Режим доступа: [http://www.atlasti.com/uploads/media/atlasti\\_v7\\_manual\\_201312.pdf](http://www.atlasti.com/uploads/media/atlasti_v7_manual_201312.pdf)
5. Connolly Paul. Quantitative Data Analysis in Education : A Critical Introduction Using SPSS [Электронный ресурс] Routledge, 2007. ProQuest Ebook Central. Режим доступа: <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=324865>.

### V.3 Программное обеспечение

п/п	Наименование	Условия доступа
1.	IBM SPSS Statistics Premium Campus Edition	<i>Из внутренней сети университета (договор)</i>
2.	ATLAS.ti	<i>Из внутренней сети университета (договор)</i>

### V.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

п/п	Наименование	Условия доступа
<b><i>Профессиональные базы данных, информационно-справочные системы</i></b>		
1.	OECD iLibrary	<i>Из внутренней сети университета (договор)</i>
2.	World Bank e-Library	<i>Из внутренней сети университета (договор)</i>
3.	Единый архив экономических и социологических данных	URL: <a href="http://sophist.hse.ru">sophist.hse.ru</a>
<b><i>Интернет-ресурсы (электронные образовательные ресурсы)</i></b>		
1.	Открытое образование	URL: <a href="https://openedu.ru/">https://openedu.ru/</a>

### V.5 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных 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.

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