

**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет компьютерных наук
Департамент программной инженерии

**Рабочая программа дисциплины по выбору
общеуниверситетского пула (МАГОЛЕГО)
"Анализ данных"**

для уровня подготовки - магистратура

Разработчик программы: Меликян А.В. amelikyan@hse.ru

Одобрена на заседании комиссии

«__»_____ 2018 г.

Председатель комиссии _____

Утверждена «__»_____ 2018 г.

Руководитель Методического центра ДООП

«__»_____ 2018 г.

Серова А.В. _____

Москва, 2018

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения департамента-разработчика программы.



1. Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов, изучающих дисциплину «Анализ данных».

2. Цели освоения дисциплины

Основные цели освоения дисциплины:

- формирование у слушателей целостного представления о методах анализа структурированных и неструктурированных данных с использованием современных программных средств;

- формирование практических навыков работы с различными типами данных (статистические данные, тексты, мультимедиа) в программах SPSS и ATLAS.ti.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать
 - существующие возможности анализа структурированных и неструктурированных данных;
 - методы количественных 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.
- Иметь навыки (приобрести опыт)
 - подготовки данных для работы с ними в SPSS и ATLAS.ti;
 - анализа данных в SPSS и ATLAS.ti;
 - представления результатов анализа в презентациях и отчётах.



4. Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к пулу дисциплин МАГОЛЕГО. Трудоемкость дисциплины - 3 кредита. Всего 106 академических часов, в том числе 24 лекций и 28 семинаров. Дисциплина проводится в 3-м и 4-м модулях.

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

- Знания основ математической статистики и теории вероятностей;
- Владение английским языком на базовом уровне.

5. Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Семинары	
1	Возможности анализа количественных данных в программе SPSS.	3	2	1	
2	Описательный анализ данных.	11	3	2	6
3	Исследование взаимосвязей между переменными. Проверка гипотез.	18	4	4	10
4	Регрессионный анализ.	21	5	6	10
5	Факторный анализ.	15	3	4	8
6	Кластерный анализ.	15	3	4	8
7	Контент-анализ как метод исследования.	11	2	3	6
8	Структурирование и кодирование текстов, изображений, видео и аудио материалов в программе ATLAS.ti.	12	2	4	6
Итого:		106	24	28	54

6. Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год		Параметры
		3	4	
Текущий	Домашнее задание		+	Провести анализ данных и подготовить отчет по результатам.
Текущий	Контрольная работа	+		Письменная работа
Итоговый	Экзамен		+	Письменная работа

6.1. Критерии оценки знаний, навыков

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.

7. Содержание дисциплины

Раздел 1. Возможности анализа количественных данных в статистическом пакете SPSS.

♦ Содержание раздела:

- 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.

Основная литература

Arthur Griffith (2010). SPSS For Dummies (2nd Edition). Wiley Publishing, Inc.

Kremelberg D. Practical statistics: a quick and easy guide to IBM SPSS Statistics, Stata, and other statistical software. Los Angeles [etc.]: Sage Publications, 2011.

Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «Диа-СофтЮП», 2002.

Andy Field (2005). Discovering Statistics Using SPSS (2nd edition). London: Sage.

Pete Greasley (2008). Quantitative Data Analysis Using SPSS. An Introduction for Health & Social Science. New York: Open University Press.

Robert Ho (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Chapman & Hall/CRC Taylor & Francis Group.

Раздел 2. Описательный анализ данных.

♦ Содержание раздела:

- 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, дове-



рительный интервал, квартили, межквартильная широта, 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.

Основная литература

Arthur Griffith (2010). SPSS For Dummies (2nd Edition). Wiley Publishing, Inc.

Kremelberg D. Practical statistics: a quick and easy guide to IBM SPSS Statistics, Stata, and other statistical software. Los Angeles [etc.]: Sage Publications, 2011.

Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «ДиасофтЮП», 2002.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS - М.: Изд.дом ГУ ВШЭ, 2006.

Andy Field (2005). Discovering Statistics Using SPSS (2nd edition). London: Sage.

Pete Greasley (2008). Quantitative Data Analysis Using SPSS. An Introduction for Health & Social Science. New York: Open University Press.

Robert Ho (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Chapman & Hall/CRC Taylor & Francis Group.

Раздел 3. Исследование взаимосвязей между переменными. Проверка гипотез.

♦ Содержание раздела:

- 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.

Основная литература

Kremelberg D. Practical statistics: a quick and easy guide to IBM SPSS Statistics, Stata, and other statistical software. Los Angeles [etc.]: Sage Publications, 2011.



Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «Диа-СофтЮП», 2002.

Таганов Д.Н. SPSS: Статистический анализ в маркетинговых исследованиях. – СПб.: Питер, 2005

Andy Field (2005). *Discovering Statistics Using SPSS* (2nd edition). London: Sage.

Pete Greasley (2008). *Quantitative Data Analysis Using SPSS. An Introduction for Health & Social Science*. New York: Open University Press.

Robert Ho (2006). *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. Chapman & Hall/CRC Taylor & Francis Group.

Раздел 4. Регрессионный анализ

◆ **Содержание раздела:**

- 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.

Основная литература

Kremelberg D. *Practical statistics: a quick and easy guide to IBM SPSS Statistics, Stata, and other statistical software*. Los Angeles [etc.]: Sage Publications, 2011.

Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «Диа-СофтЮП», 2002.

Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 2. С. 53–72.

Таганов Д.Н. SPSS: Статистический анализ в маркетинговых исследованиях. – СПб.: Питер, 2005

Andy Field (2005). *Discovering Statistics Using SPSS* (2nd edition). London: Sage.



Robert Ho (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Chapman & Hall/CRC Taylor & Francis Group.

Раздел 5. Факторный анализ

◆ Содержание раздела:

- 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 значений факторов.

Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «Диа-СофтЮП», 2002.

Andy Field (2005). Discovering Statistics Using SPSS (2nd edition). London: Sage.

Robert Ho (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Chapman & Hall/CRC Taylor & Francis Group.

Раздел 6. Кластерный анализ

◆ Содержание раздела:

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

Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «Диа-СофтЮП», 2002.

Таганов Д.Н. SPSS: Статистический анализ в маркетинговых исследованиях. – СПб.: Питер, 2005

Раздел 7. Контент-анализ как метод исследования.

◆ Содержание раздела:

- 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.

Основная литература

Аверьянов, Л. Я. Контент-анализ. М. КноРус, 2009. - 451 с.

Шалак, В.И. Современный контент-анализ. ОМЕГА-Л, 2009. - 272 с.

Дополнительная литература

Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. Sage, Thousand Oaks, CA, 2nd edition.

Lewins A., Silver Ch. (2010) Using Software in Qualitative Research. A Step-by-Step Guide, London, Sage

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

♦ Содержание раздела:

- 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.

Основная литература

ATLAS.ti 7 User Guide and Reference, 2013

http://www.atlasti.com/uploads/media/atlasti_v7_manual_201312.pdf

Saldana, J. The coding manual for qualitative researchers. Los Angeles [etc.] SAGE Publications, 2013. - 303 с.

Дополнительная литература

Friese, S. Qualitative data analysis with ATLAS.ti. Los Angeles [etc.] SAGE Publications, 2012. - 274 с.



8. Образовательные технологии

Работа с реальными базами данных, документами и материалами. Обсуждение результатов анализа данных и актуальных вопросов по темам курса.

9. Оценочные средства для текущего контроля и аттестации студента

9.1. Вопросы для оценки качества освоения дисциплины

- 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.

9.2. 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.



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

10. Порядок формирования оценок по дисциплине

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 бальной 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 бальной 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 за итоговый контроль в форме экзамена выставляется по следующей формуле, где $O_{экзамен}$ – оценка за работу, выполненную на письменном экзамене:

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

Способ округления накопленной оценки итогового контроля в форме экзамена: в пользу студента.

На передаче студенту 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 за итоговый контроль в форме экзамена, получаемая на передаче, выставляется по формуле

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

В диплом выставляет результирующая оценка по учебной дисциплине, которая формируется по следующей формуле:

$$O_{дисциплина} = O_{итоговый}$$

Способ округления результирующей оценки по учебной дисциплине: в пользу студента.

В диплом ставится оценка за итоговый контроль, которая является результирующей оценкой по учебной дисциплине.

11. Учебно-методическое и информационное обеспечение дисциплины

11.1. Основная литература

Аверьянов, Л. Я. Контент-анализ. М. КноРус, 2009. - 451 с.



ATLAS.ti 7 User Guide and Reference, 2013
http://www.atlasti.com/uploads/media/atlasti_v7_manual_201312.pdf

Arthur Griffith (2010). SPSS For Dummies (2nd Edition). Wiley Publishing, Inc.

Friese, S. Qualitative data analysis with ATLAS.ti. Los Angeles [etc.] SAGE Publications, 2012. - 274 с.

Kremelberg D. Practical statistics: a quick and easy guide to IBM SPSS Statistics, Stata, and other statistical software. Los Angeles [etc.]: Sage Publications, 2011.

Lewins A., Silver Ch. (2010) Using Software in Qualitative Research. A Step-by-Step Guide, London, Sage

Saldana, J. The coding manual for qualitative researchers. Los Angeles [etc.] SAGE Publications, 2013. - 303 с.

11.2. Дополнительная литература

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./ Ахим Бююль, Петер Цефель – СПб.: ООО «ДиаСофтЮП», 2002.

Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 2. С. 53–72.

Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS - М.: Изд.дом ГУ ВШЭ, 2006.

Таганов Д.Н. SPSS: Статистический анализ в маркетинговых исследованиях. – СПб.: Питер, 2005

Шалак, В.И. Современный контент-анализ. ОМЕГА-Л, 2009. - 272 с.

Andy Field (2005). Discovering Statistics Using SPSS (2nd edition). London: Sage.

Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. Sage, Thousand Oaks, CA, 2nd edition.

Pete Greasley (2008). Quantitative Data Analysis Using SPSS. An Introduction for Health & Social Science. New York: Open University Press.

Robert Ho (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Chapman & Hall/CRC Taylor & Francis Group.

11.3. Программные средства

Для успешного освоения дисциплины, студент использует программы для обработки и анализа данных SPSS и ATLAS.ti.

Автор программы: _____/Меликян А.В./ amelikyan@hse.ru