

## Программа учебной дисциплины

Анализ и визуализация данных в психологии

Утверждена

Академическим советом ООП

Протокол №2.6-06.7/5 от 06.07.2018

Автор	Чумакова М.А., канд. психол. наук, mchumakova@hse.ru
Число кредитов	2
Контактная работа (час.)	32
Самостоятельная работа (час.)	44
Курс	3
Формат изучения дисциплины	без использования онлайн курса

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

Целями освоения дисциплины «Научно-исследовательский семинар Анализ и визуализация данных в психологии» являются:

- освоение основных методов многомерного статистического анализа данных психологических исследований;
- освоение практических навыков многомерной статистической обработки эмпирических данных с использованием компьютерных программ;
- освоение практических навыков визуализации данных.

В результате освоения дисциплины студент должен:

- Знать принципы, типовые задачи, методы и алгоритмы многомерной статистической обработки данных психологических исследований;
- Знать основные формы и способы визуализации данных;
- Уметь корректно выбирать и применять методы многомерного статистического анализа и визуализации данных для решения типовых задач статистической обработки результатов психологических исследований (с помощью компьютерных статистических пакетов);
- Иметь навыки представления результатов многомерного статистического анализа в научном тексте.

Настоящая дисциплина относится к циклу научно-исследовательских семинаров программы «Психология».

«Научно-исследовательский семинар Анализ и визуализация данных в психологии» является основой для изучения и закрепления знаний по следующим дисциплинам:

- Научно-исследовательский семинар (1, 2 курс)
- Математические методы в психологии
- Экспериментальная психология

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

- Знает принципы работы с информацией, ее обобщения и анализа
- Планирует и организует психологическое исследование под научным руководством
- Представляет результаты проведенного исследования в целостном тексте и публичном выступлении
- Умеет грамотно и критически анализировать чужие и свои исследования на предмет качества формального и содержательного дизайна, соответствия методики заявленному предмету, гипотезам и целям, а также правомерности сделанных выводов.
- Знает и умеет использовать основные понятия математической статистики при формализации и обработке данных психологических экспериментов, интерпретации полученных результатов

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

- Курсовая работа
- Выпускная квалификационная работа
- Научно-исследовательская деятельность

## II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1 Общее представление о многомерных методах статистического анализа данных

- Теоретические, эмпирические и статистические гипотезы.
- Квазиэкспериментальные исследования и статистические методы анализа данных
- Многомерные статистические методы анализа данных: область применения, цели, результаты
- Статистическая достоверность выводов на выборке относительно генеральной совокупности.
- Процедура статистического вывода: понятие статистического критерия, виды гипотез, понятия уровня значимости, ошибок I и II рода.
- Проблема статистической мощности анализа, её связь с методом проверки гипотез, размером эффекта, объёмом выборки, критерием принятия решений.

Тема 2 Визуализация данных

- Основные виды графического представления распределений: гистограммы, кривые плотности вероятности, box-plots.

- Диаграммы рассеяния и линии тренда как визуализация отношений между исследуемыми переменными
- Пакет ggplot2: визуализация данных

### Тема 3 Регрессионный анализ данных и Машинное обучение

- Цели и задачи регрессионного анализа данных
- Базовые математические основания регрессионного анализа
- Множественная регрессия
- Оценка качества регрессионных моделей
- Интерпретация результатов регрессионного анализа данных
- Реализация регрессионного анализа данных в программной среде R
- Нормативы представления результатов регрессионного анализа данных в научном отчете
- Основные принципы машинного обучения
- Реализация базовых моделей машинного обучения в программной среде R

### Тема 4 Эксплораторный факторный анализ данных

- Цели и задачи эксплораторного факторного анализа данных
- Базовые математические основания эксплораторного факторного анализа
- Оценка качества результатов эксплораторного факторного анализа
- Интерпретация результатов эксплораторного факторного анализа данных
- Реализация эксплораторного факторного анализа данных в программной среде R
- Нормативы представления результатов эксплораторного факторного анализа данных в научном отчете

### Тема 5 Конфирматорный факторный анализ данных

- Цели и задачи конфирматорного факторного анализа данных
- Базовые математические основания конфирматорного факторного анализа
- Оценка качества результатов конфирматорного факторного анализа
- Интерпретация результатов конфирматорного факторного анализа данных
- Реализация конфирматорного факторного анализа данных в программной среде R
- Нормативы представления результатов конфирматорного факторного анализа данных в научном отчете

### Тема 6 Кластерный анализ данных

- Цели и задачи кластерного анализа данных
- Базовые математические основания кластерного анализа
- Оценка качества результатов кластерного анализа
- Интерпретация результатов кластерного анализа данных

### Тема 7 Анализ латентных профилей

- Цели и задачи анализа латентных профилей
- Базовые математические основания анализа латентных профилей
- Оценка качества результатов анализа латентных профилей
- Интерпретация результатов анализа латентных профилей
- Реализация анализа латентных профилей в программной среде R
- Нормативы представления результатов конфирматорного факторного анализа данных в научном отчете

### III. ОЦЕНИВАНИЕ

По всем формам отчётности оценки ставятся по 10-бальной шкале, при этом текущие оценки могут не быть целочисленными, округление проводится до первой цифры после запятой. При выставлении накопленной, промежуточной и итоговой оценок производится округление до ближайшего целого числа (если дробная часть оценки равна 0.5, то округление производится в большую сторону).

**Домашнее задание 1** представляет собой тренировку навыков визуализации данных. Задание выполняется индивидуально. Студенты получают набора данных и образцы графиков, которые нужно построить. Оценка работы производится по десятибалльной шкале следующим образом: за каждый построенный график студент получает 1 балл, если график полностью соответствует образцу, или получает 0.5 балла, если график соответствует образцу частично.

**Домашнее задание 2** представляет собой описание и критическую оценку результатов регрессионного анализа из опубликованной научной статьи по психологии. Задание выполняется индивидуально, при этом выбранные статьи не должны повторяться среди студентов одного курса. Оценка работы производится по десятибалльной шкале и рассчитывается как сумма баллов по следующим критериям:

ОЦЕНКА	КРИТЕРИИ
«Отлично»: 10	Безупречное соблюдение всех критериев
«Отлично»: 9, 8	Оценка «9» может быть выставлена только при условии соответствия работы всем критериям при наличии незначительных неточностей. Оценка «8» может быть выставлена только при условии соответствия работы всем критериям, но при наличии незначительных содержательных неточностей
«Хорошо»: 7, 6	Оценка «7» может быть выставлена только при условии соответствия работы 4 критериям из 5 перечисленных. Оценка «6» может быть выставлена только при условии соответствия работы 4 критериям из 5 перечисленных и при наличии незначительных содержательных неточностей

«Удовлетворительно»: 5, 4	Оценка «5» может быть выставлена только при условии соответствия работы 3 критериям из 5 перечисленных Оценка «4» может быть выставлена только при условии соответствия работы 3 критериям из 5 перечисленных и при наличии незначительных содержательных неточностей
«Неудовлетворительно»: 3, 2, 1	Данные оценки выставляются в случае несоответствия работы большинству предъявляемых критериев

**Домашнее задание 3** представляет собой самостоятельный регрессионный анализ данных и описание полученных результатов. Задание выполняется индивидуально. Оценка работы производится по десятибалльной шкале по следующим критериям:

- Предложена корректная регрессионная модель для статистического анализа: в модели корректно определены и обоснованы предикторы и зависимая переменная
- Проведена оценка возможности применения регрессионного анализа для статистической проверки предложенной модели на имеющемся наборе данных
- Результаты корректно описаны с соблюдением норм представления статистических результатов применения регрессионного анализа
- Представлены графические подтверждения полученных результатов регрессионного анализа данных
- Представлен код в программной среде R, выполняющий описанный регрессионный анализ

**Домашнее задание 4** представляет собой факторный и кластерный анализ данных и описание полученных результатов. Задание выполняется в групповой форме. Оценка работы производится по десятибалльной шкале по следующим критериям:

- Проведена оценка возможности применения эксплораторного факторного анализа для обработки имеющегося набора данных
- Результаты корректно описаны с соблюдением норм представления статистических результатов применения эксплораторного факторного анализа
- Проведена проверка результатов эксплораторного факторного анализа средствами конфирматорного факторного анализа
- Результаты корректно описаны с соблюдением норм представления статистических результатов применения конфирматорного факторного анализа
- Представлены графические подтверждения полученных результатов эксплораторного и конфирматорного анализа данных
- Обоснованы параметры выделения групп для применения анализа латентных

профи- лей

- Результаты корректно описаны с соблюдением норм представления статистических результатов применения анализа латентных профилей
- Представлены графические подтверждения полученных результатов применения анализа латентных профилей
- Представлен код в программной среде R, выполняющий описанный статистический анализ данных
- 

ОЦЕНКА	КРИТЕРИИ
«Отлично»: 10	Безупречное соблюдение всех критериев
«Отлично»: 9, 8	Оценка «9» может быть выставлена только при условии соответствия работы всем критериям при наличии незначительных неточностей. Оценка «8» может быть выставлена только при условии соответствия работы всем критериям, но при наличии незначительных содержательных неточностей
«Хорошо»: 7, 6	Оценка «7» может быть выставлена только при условии соответствия работы 7 или 8 критериям из 9 перечисленных Оценка «6» может быть выставлена только при условии соответствия работы 7 или 8 критериям из 9 перечисленных и при наличии незначительных содержательных неточностей
«Удовлетворительно»: 5, 4	Оценка «5» может быть выставлена только при условии соответствия работы 5 или 6 критериям из 9 перечисленных Оценка «4» может быть выставлена только при условии соответствия работы 5 или 6 критериям из 9 перечисленных и при наличии незначительных содержательных неточностей
«Неудовлетворительно»: 3, 2, 1	Данные оценки выставляются в случае несоответствия работы большинству предъявляемых критериев

Преподаватель оценивает работу студентов на семинарских занятиях: активность на семинарских занятиях, участие в ролевых играх, дискуссиях, работу по разбору кейсов.

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 промежуточным или итоговым контролем – О аудиторная.

Преподаватель оценивает самостоятельную работу студентов: результаты домашних

заданий 1 и 2.

Оценки за самостоятельную работу студента преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за самостоятельную работу определяется перед промежуточным или итоговым контролем –  $O_{\text{сам. работа}}$

$$O_{\text{сам.раб}} = 0.5 * O_{\text{дз1}} + 0.5 * O_{\text{дз2}}$$

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

$$O_{\text{накопленная}} = 0.6 * O_{\text{текущий}} + 0.3 * O_{\text{сам.раб}} + 0.1 * O_{\text{аудиторная}}$$

где  $O_{\text{текущий}}$  рассчитывается как взвешенная сумма всех форм текущего контроля, предусмотренных в РУП

$$O_{\text{текущий}} = 0.5 * O_{\text{дз3}} + 0.5 * O_{\text{дз4}}$$

Итоговая оценка за дисциплину рассчитывается следующим образом:

$$O_{\text{итог}} = 0.6 * O_{\text{накопленная}} + 0.4 * O_{\text{экзамен}}$$

Способ округления результирующей оценки по учебной дисциплине: арифметический (например, оценка 4,4 округляется до 4, а оценка 4,5 до 5).

В случае если накопленная оценка студента превышает 8 баллов (без предварительного округления), ему может быть выставлена в ведомость оценка «отлично» без прохождения итогового испытания. Отличная оценка выставляется исходя из следующего принципа: накопленная оценка от 8 до 8,49 – в ведомость выставляется оценка 8, накопленная оценка от 8,5 до 9,49 – в ведомость выставляется оценка 9, накопленная оценка от 9,5 – в ведомость выставляется оценка 10.

Условия пересдач

На пересдаче студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль. Первая пересдача проводится преподавателем, отвечающим за чтение дисциплины на факультете. При выставлении результирующей оценки учитывается накопленная оценка за текущий контроль, самостоятельную и аудиторную работу

Вторая пересдача проводится в присутствии комиссии, включающей не менее трех преподавателей, при выставлении результирующей оценки учитывается накопленная оценка за текущий контроль, самостоятельную и аудиторную работу

#### IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

**Тематика заданий для оценки самостоятельной работы студентов и для текущего контроля**

Домашнее задание 1 представляет собой тренировку навыков визуализации данных. Задание выполняется индивидуально. Студенты получают набора данных и образцы графиков, которые

нужно построить. Формат отчета о выполнении задания – предоставления файла с кодом R.

Домашнее задание 2 представляет собой анализ статьи. Задание выполняется индивидуально, при этом выбранные статьи не должны повторяться среди студентов одного курса. В качестве статьи может быть выбрана работа на любую тему, однако в ней должен быть представлен регрессионный анализ эмпирических данных. Подходящими являются статьи на русском либо английском языке, опубликованные в научных журналах, входящих хотя бы в один из списков: WoS, Scopus, ВАК, опубликованные не ранее 2000 г. Необходимо представить письменный отчет, в котором описаны результаты регрессионного анализа, представленного в статье, и проведена критическая оценка качества этого анализа и его интерпретации авторами статьи. К работе должен быть приложен полный текст анализируемой статьи.

Структура отчета:

- Краткое описание гипотезы исследования, выборки и измеренных в исследовании переменных
- Формулирования перечня гипотез, проверяемых в статье средствами регрессионного анализа
- Краткое описание результатов регрессионного анализа в соответствии с выделенным перечнем гипотез
- Оценка возможностей применения регрессионного анализа к данным, полученным в анализируемом исследовании
- Оценка соответствия применения регрессионного анализа целям и гипотезам, заявленным в статье
- Оценка корректности представления результатов регрессионного анализа в исследовании
- Оценка корректности авторских выводов из результатов регрессионного анализа данных

Домашнее задание 3 представляет собой самостоятельный регрессионный анализ данных и описание полученных результатов. Задание выполняется индивидуально. Необходимо представить письменный отчет, в котором описаны результаты выполненного регрессионного анализа.

Структура отчета:

- Описание и обоснование регрессионной модели, проверяемой в ходе статистического анализа данных
- Оценка возможности применения регрессионного анализа для статистической проверки предложенной модели на имеющемся наборе данных
- Описание результатов применения регрессионного анализа
- Графическое подтверждение полученных результатов регрессионного анализа данных
- Интерпретация полученных статистических результатов



В качестве приложения к отчету в электронном виде предоставляется код в программной среде R, выполняющий описанный регрессионный анализ

Домашнее задание 4 представляет собой факторный и кластерный анализ данных и описание полученных результатов. Задание выполняется в групповом формате (группы до 5 человек). Необходимо представить письменный отчет, в котором описаны результаты выполненного анализа.

Структура отчета:

- Оценка возможности применения эксплораторного факторного анализа для обработки имеющегося набора данных
- Описание результатов применения эксплораторного факторного анализа
- Интерпретация результатов эксплораторного факторного анализа
- Проверка результатов эксплораторного факторного анализа средствами конфирматорного факторного анализа
- Описание результатов применения конфирматорного факторного анализа
- Интерпретация результатов конфирматорного факторного анализа
- Графическое подтверждение полученных результатов эксплораторного и конфирматорного факторного анализа данных
- Описание параметров выделения групп для применения анализа латентных профилей
- Описание результатов применения анализа латентных профилей
- Интерпретация результатов применения анализа латентных профилей
- Графическое подтверждение полученных результатов анализа латентных профилей

В качестве приложения к отчету в электронном виде предоставляется код в программной среде R, выполняющий описанный статистический анализ данных

Вопросы для оценки качества освоения дисциплины

1. Цели и задачи регрессионного анализа данных. Базовые математические основания регрессионного анализа
2. Множественная регрессия. Оценка качества регрессионных моделей
3. Нормативы представления результатов регрессионного анализа данных в научном отчете
4. Цели и задачи эксплораторного факторного анализа данных. Базовые математические основания эксплораторного факторного анализа
5. Оценка качества результатов эксплораторного факторного анализа. Интерпретация

результатов эксплораторного факторного анализа данных

6. Нормативы представления результатов эксплораторного факторного анализа данных в научном отчете
7. Цели и задачи конфирматорного факторного анализа данных. Базовые математические основания конфирматорного факторного анализа
8. Оценка качества результатов конфирматорного факторного анализа
9. Интерпретация результатов конфирматорного факторного анализа данных. Нормативы представления результатов конфирматорного факторного анализа данных в научном отчете
10. Цели и задачи кластерного анализа данных. Базовые математические основания кластерного анализа
11. Оценка качества результатов кластерного анализа. Интерпретация результатов кластерного анализа данных
12. Цели и задачи анализа латентных профилей. Базовые математические основания анализа латентных профилей
13. Оценка качества результатов анализа латентных профилей. Интерпретация результатов анализа латентных профилей

## **V. РЕСУРСЫ**

### **1. Основная литература**

Корнилова Т.В. Экспериментальная психология: Теория и методы: Учебник для вузов. М.: Аспект Пресс, 2005.

Наследов А.Д. Математические методы психологического исследования: Анализ и интерпретация данных. Учебное пособие. СПб: Речь, 2006. 392 с.

### **2. Дополнительная литература**

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>

Chris Fraley, Adrian E. Raftery, T. Brendan Murphy, and Luca Scrucca (2012) mclust Version 4 for R:

### 3. Программное обеспечение

№ п/п	Наименование	Условия доступа
1.	Microsoft Windows 7 Professional RUS Microsoft Windows 10 Microsoft Windows 8.1 Professional RUS	<i>Из внутренней сети университета (договор)</i>
2.	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>

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

№ п/п	Наименование	Условия доступа
<i>Профессиональные базы данных, информационно-справочные системы</i>		
1.	Консультант Плюс	<i>Из внутренней сети университета (договор)</i>
2.	Электронно-библиотечная система Юрайт	URL: <a href="https://biblio-online.ru/">https://biblio-online.ru/</a>
<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>		
1.	Открытое образование	URL: <a href="https://openedu.ru/">https://openedu.ru/</a>

### 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 лабораторных и самостоятельных занятий по дисциплине оснащены персональными компьютерами, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.