

**АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ****Б1.В.ДВ.6.1 Введение в анализ данных с использованием SPSS**

Автор: Кандидат социологических наук Д.М. Рогозин

Код и наименование направления подготовки, профиля: «Фундаментальная социология»,  
39.04.01 Социология

Форма обучения: очная

**Цель освоения дисциплины:**

- изучение и практическое освоение базовых методов статистического анализа данных в социальных науках,
- изучение и практическое освоение компьютерных программ, применяемых для статистического анализа данных,
- приобретение понимания специфики работы с количественными данными в социальных науках, понимания типов задач, которые могут быть решены с помощью статистических методов.

**План курса:**

Таблица 1.

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)
Тема 1	Статистический анализ данных в социологии. Статистические гипотезы и их тестирование. Дисперсионный анализ	Источники данных для статистического анализа: опрос, перепись, агрегированная статистика. Основные методы анализа. Обзор содержания курса. Основы работы в SPSS. Синтаксис SPSS. Описательная статистика: основные процедуры. Понятие статистической гипотезы. Тестирование гипотез. Тесты на равенство средних и пропорций. Доверительные интервалы и их интерпретация. Ошибки типа I и типа II. Понятие статистической мощности. Корректная интерпретация статистической значимости. Величина статистических эффектов и статистическая значимость. Дисперсионный
Тема 2	Таблицы сопряженности и их анализ и корреляция	Таблицы сопряженности. Критерий хи-квадрат. Отношения шансов (преобладания). Логлинейные модели. Применение логлинейных моделей в исследованиях социальной мобильности. Коэффициент корреляции Пирсона как мера связи между двумя интервальными переменными. Другие коэффициенты корреляции в зависимости от используемых шкал.

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)
Тема 3	Регрессионный анализ	<p>Парная линейная регрессия. Связь между корреляцией и регрессией. Оценка и интерпретация регрессионных коэффициентов и стандартных ошибок. Статистическая значимость коэффициентов. Регрессии с фиктивными переменными. Понятие статистического контроля.</p> <p>Оценка и интерпретация регрессионных коэффициентов в моделях множественной регрессии. Эффекты взаимодействия. Коэффициент детерминации <math>R^2</math>. F-статистика и F-тест. Допущения моделей множественной регрессии и диагностика моделей.</p> <p>Гетероскедастичность. Нелинейные связи.</p> <p>Статистические выбросы. Мультиколлинеарность.</p> <p>Принципы построения регрессионных моделей. Модель линейной вероятности. Логистическая регрессия.</p> <p>Интерпретация коэффициентов логистической регрессии. Шансы и отношения шансов. Предсказанные вероятности. Мультиномиальная логистическая регрессия. Презентация эффектов в линейной и логистической регрессии. Идентификация причинно-следственных связей и регрессионный анализ.</p> <p>Контрфактуальная модель Неймана-Рубина. Проблема пропущенных переменных. Проблема обратной причинно-следственной зависимости. Возможные решения: регрессии с фиксированными эффектами, инструментальные переменные, мэтчинг.</p>
Тема 4	Факторный анализ.	<p>Модель факторного анализа как модель латентных переменных. Различные подходы к определению числа факторов. Процент объясненной дисперсии как показатель качества факторной модели.</p> <p>Индивидуальные значения факторов. Сохранение факторов как новых переменных. Вращение матрицы факторных нагрузок. Ортогональные и неортогональные методы вращения.</p>
Тема 5	Кластерный анализ	<p>Иерархический агломеративный кластерный анализ.</p> <p>Кластерный анализ методом k-средних. Проблемы выбора меры расстояния и формы кластера. Проблема устойчивости кластеризации. Методы оценки устойчивости. Описание и интерпретация результатов кластеризации.</p>

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)
Тема 6	Модели деревьев решений.	Особенности работы алгоритмов CHAID и CRT. Ошибки классификации, определение понятия риска. Цена ошибки классификации (Misclassification Cost). Возможность задавать различные значения цены ошибки классификации для разных типов ошибок, интерпретация соответствующих моделей. Проверка качества модели, способы решения проблемы излишней подгонки дерева: кросс-проверка, V-кратная кросс-проверка, отсечение ветвей.

#### **Формы текущего контроля и промежуточной аттестации:**

При проведении занятий лекционного типа: контроль посещаемости

При проведении занятий семинарского типа: контрольные работы

При контроле результатов самостоятельной работы студентов: презентация темы эссе.

Форма промежуточной аттестации: зачет с оценкой в форме эссе

#### **Основная литература:**

1. Федин Ф.О. Анализ данных. Часть 1. Подготовка данных к анализу [Электронный ресурс]: учебное пособие / Ф.О. Федин, Ф.Ф. Федин. — Электрон. текстовые данные. — М. : Московский городской педагогический университет, 2012. — 204 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/26444.html>
2. Федин Ф.О. Анализ данных. Часть 2. Инструменты Data Mining [Электронный ресурс]: учебное пособие / Ф.О. Федин, Ф.Ф. Федин. — Электрон. текстовые данные. — М. : Московский городской педагогический университет, 2012. — 308 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/26445.html>