АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Б1.В.ДВ.01.24.07 Анализ и описание данных

Автор:
К.соц.н, доцент кафедры теоретической социологии и эпистемологии Астахова А.С.

Код и наименование направления подготовки, профиля:

50.03.01 Искусства и гуманитарные науки

«Филология и перевод» (Liberal Arts)

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Цель освоения дисциплины: сформировать способность использовать основы социологических знаний и применять методы социологического исследования для решения прикладных задач в различных сферах деятельности

План курса:

№ 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 статистических пакетах. Синтаксис. 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 анализ
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	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. Оценка и интерпретация регрессионных коэффициентов в моделях множественной регрессии. Эффекты взаимодействия. Коэффициент детерминации R^2. F-статистика и F-тест. Допущения моделей множественной регрессии и диагностика моделей. Гетероскедастичность. Нелинейные связи. Статистические выбросы. Мультиколлинеарность. Принципы построения регрессионных моделей. Модель линейной вероятности. Логистическая регрессия. 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. Контрфактуальная модель Неймана-Рубина. Проблема пропущенных переменных. Проблема обратной причинно-следственной зависимости. Возможные решения: регрессии с фиксированными эффектами, инструментальные переменные, мэтчинг
Тема 4	Факторный анализ. Кластерный анализ	Модель факторного анализа как модель латентных переменных. Различные подходы к определению числа факторов. Процент объясненной дисперсии как показатель качества факторной модели. Индивидуальные значения факторов. Сохранение факторов как новых переменных. Вращение матрицы факторных нагрузок. Ортогональные и неортогональные методы вращения. Иерархический агломеративный кластерный анализ. Кластерный анализ методом k-средних. Проблемы выбора меры расстояния и формы кластера. Проблема устойчивости кластеризации. Методы оценки устойчивости. Описание и интерпретация результатов кластеризации.
Тема 5	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.

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)
Тема 6	Методы анализа качественных данных. Интерпретация данных.	Описание и интерпретация данных: различие в количественных и качественных исследованиях. Виды данных в качественном исследовании: текстовые данные, иконическая информация. Подходы к работе с текстовыми и визуальными данными: кодирование, категоризация конденсация смыслов, интерпретация. Дискурс-анализ как тактика и способ качественного анализа данных: преимущества и ограничения. Виды дискурс-анализа (критический, исторический и др.).

Формы текущего контроля и промежуточной аттестации:

В ходе реализации дисциплины используются следующие методы текущего контроля и успеваемости обучающихся:
– при проведении занятий семинарского типа: опрос, тестирование.

Промежуточная аттестация проводится в 7 семестре **устный зачет с оценкой, состоящий из ответов на 2 вопроса**, **а также завершается презентацией и публичной защитой в формате, выбранном руководителем майнора (в зависимости от специфики майнора), в данном майноре – это курсовая работа**.

Основная литература:

1. Стружкин, Н. П. Базы данных: проектирование : учебник для академического бакалавриата / Н. П. Стружкин, В. В. Годин. — М. : Издательство Юрайт, 2017. — 477 с. Режим доступа: https://www.biblio-online.ru/book/BF8DDE6E-054D-4BB4-A6FA-2E9898529E96.