

Learning outcomes к контрольной работе

1. Модель линейной регрессии: уравнение спецификации модели, базовые понятия: зависимая переменная (объясняемая переменная / отклик), независимая переменная (объясняющая переменная / предиктор), ошибки в регрессионной модели, остатки как оценки ошибок, параметры регрессионной модели (коэффициенты: константа и коэффициенты при предикторах)
2. Метод наименьших квадратов:
 - ключевой принцип
 - выведение оценок для случая парной регрессии (частный случай)
 - релевантная как для случая парной, так и множественной регрессии формула для получения вектора оценок коэффициентов: $(X^T X)^{-1} X^T y$. Выведение формулы. Уметь по заданным значениям предикторов и значениям отклика получить вектор оценок коэффициентов регрессионной модели
3. Условия верные по построению регрессионной модели (равенство суммы остатков нулю, нескоррелированность остатков и предикторов)
4. Теорема Гаусса – Маркова. Допущения об ошибках в регрессионной модели. Понятие BLUE-оценок
5. Теорема Фриша-Во-Ловелла (Frisch-Waugh-Lovell)
6. Интерпретация оценок коэффициентов в линейной регрессионной модели
7. Проверка значимости коэффициентов в линейной регрессионной модели
8. Построение доверительного интервала для коэффициента в линейной регрессионной модели с последующей интерпретацией
9. Разложение вариации. Коэффициент детерминации и проверка гипотезы о незначимости коэффициента детерминации
10. Уметь рассчитать предсказанное значение зависимой переменной при условии заданного значения предиктора
11. Спецификация множественной регрессионной модели: ключевые предикторы и контрольные переменные. Критерий «черного хода» для определения релевантных контрольных переменных (backdoor criterion). Способы, как можно «заблокировать» связь между переменными
12. Различия между модерацией и медиацией
13. Переменные взаимодействия как способ проверки совместного эффекта. Правила построения спецификации линейной регрессионной модели с переменными взаимодействия (см. статью Brambor et al. в разделе «Занятие 2» на странице курса)
14. Интерпретация исходных коэффициентов в линейной регрессионной модели с переменными взаимодействия, а также интерпретация с помощью предельных эффектов
15. Предельный эффект: определение, вычисление предельного эффекта по оценкам коэффициентов регрессионной модели, интерпретация
16. Вычисление стандартной ошибки предельного эффекта с помощью ковариационной матрицы оценок коэффициентов регрессионной модели
17. Визуализация результатов (график, демонстрирующий взаимосвязь предиктора-«условия» и предельного эффекта, и интерпретация данного графика: значения предельного эффекта и их значимость)

18. Центрирование (и другие возможные алгебраические преобразования переменных) в контексте регрессионного анализа с переменными взаимодействия: содержательный смысл данного преобразования, интерпретация коэффициентов при преобразованных предикторах
19. Сравнение подходов: включение переменных взаимодействия в регрессионную модель и оценивание регрессионных моделей на отдельных подвыборках, выделенных на основе значения предиктора-«условия» (модератора)
20. Модель разность разностей (difference-in-differences). Допущение о параллельности трендов. Уметь посчитать и проинтерпретировать оценки коэффициентов. Гипотетический исход (counterfactual) уметь рассчитать и проинтерпретировать полученное значение
21. Метод синтетического контроля: уметь посчитать веса для получения синтетического контроля, рассчитать средний эффект воздействия
22. Гетероскедастичность
 - Определение гетероскедастичности, примеры
 - Источники гетероскедастичности
 - Последствия гетероскедастичности
 - Способы диагностики гетероскедастичности:
 - (a) визуализация
 - (b) формальные тесты: тест Уайта (нулевая гипотеза и альтернатива, параметры во вспомогательной модели, вывод по p-value), тест Бреуша–Пагана как частный случай теста Уайта, тест Голдфелда–Квандта (нулевая гипотеза и альтернатива, статистика критерия, вывод)
 - Дисперсия $\hat{\beta}$ в условиях гомоскедастичности и отсутствия автокорреляции: уметь вывести формулу
 - Состоятельные в условиях гетероскедастичности стандартные ошибки HC3: знать общий вид вспомогательной матрицы весов
 - Обобщенный метод наименьших квадратов (ОМНК) и реализуемый обобщенный метод наименьших квадратов (РОМНК). Смысл данной стратегии в условиях гетероскедастичности, уметь объяснить, чем эта стратегия отличается от использования состоятельных в условиях гетероскедастичности стандартных ошибок. Уметь вывести формулу для расчета $\hat{\beta}$ в соответствии с (Р)ОМНК. Знать ограничения РОМНК. Модель со случайными эффектами и ее допущения. Ковариационная матрица ошибок модели со случайными эффектами
23. Нетипичные наблюдения:
 - Выбросы – по зависимой переменной
 - Leverage – по предиктору
 - Влиятельные наблюдения
24. Последствия наличия нетипичных и, в частности, влиятельных наблюдений в массиве
25. Стьюдинизированные остатки: понимать, что используются для диагностики выбросов, как делать вывод по полученным значениям
26. Матрица проекции (hat-matrix): определение потенциала влияния наблюдений, уметь получить матрицу проекции по заданным значениям x , уметь с помощью этой матрицы получить из y наблюдаемого значения \hat{y} предсказанного, знать свойства данной матрицы
27. Мера Кука и мера DFBETA: в чем разница между этими мерами, как делать вывод на основе этих мер

28. Что делать с нетипичными и влиятельными наблюдениями? Стоит ли их удалять?

29. Мультиколлинеарность

- Суть проблемы
- Строгая мультиколлинеарность. Невозможность получить оценки коэффициентов в условиях строгой мультиколлинеарности
- Нестрогая мультиколлинеарность и ее последствия
- Диагностики мультиколлинеарности. Показатель VIF (variance inflation factor): выводение формулы, связь VIF и оценки дисперсии коэффициента при предикторе, интерпретация VIF
- Мультиколлинеарность в контексте регрессионного анализа с переменными взаимодействия, или «не так страшен черт, как его малют»
- Метод главных компонент как способ перейти к ортогонализированному признаковому пространству: уметь полностью реализовывать процедуру МГК применительно к ковариационной / корреляционной матрице; уметь находить и интерпретировать собственные значения и собственные векторы; знать свойства ортогональной матрицы, использующейся для преобразования исходного признакового пространства; правило сохранения информации в рамках МГК; определение количества главных компонент, которые можно оставить в решении (критерий Кайзера, визуальный способ «график каменистой осьпи»)
- Регуляризация в линейной регрессии. Гребневая регрессия (ridge regression): основная идея, формула для оценивания параметров, преимущества и ограничения метода. Получение параметра регуляризации. Понимание принципа k-фолдовой кросс-валидации