

Learning outcomes к экзаменационной работе №1

1. Модель линейной регрессии: уравнение спецификации модели, базовые понятия: зависимая 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: $(X^T X)^{-1} X^T y$, 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 предиктора)
4. Допущения об ошибках в регрессионной модели
5. Теорема Фриша-Во-Ловелла (Frisch-Waugh-Lovell)
6. Интерпретация оценок коэффициентов в регрессионной модели
7. Проверка значимости коэффициентов в регрессионной модели
8. Построение доверительного интервала для коэффициента в 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 moderation effect. Правила построения спецификации линейной регрессионной модели с переменными взаимодействия
14. Интерпретация исходных коэффициентов в линейной регрессионной модели с переменными взаимодействия, а также интерпретация с помощью предельных эффектов (см. следующий пункт)
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» (moderator)
20. Мультиколлинеарность
 - Суть проблемы
 - Строгая мультиколлинеарность. Невозможность получить оценки коэффициентов в условиях строгой мультиколлинеарности
 - Сильная мультиколлинеарность и ее последствия
 - Диагностики мультиколлинеарности. Показатель VIF (variance inflation factor)
 - Мультиколлинеарность в контексте регрессионного анализа с переменными взаимодействия, или «не так страшен черт, как его малюют»
 - Метод главных компонент как способ перейти к ортогонализированному признаковому пространству
 - Регуляризация в линейной регрессии. Гребневая регрессия (ridge regression)
21. Гетероскедастичность
 - Определение гетероскедастичности, примеры
 - Условная вариация и условное математическое ожидание
 - Источники гетероскедастичности
 - Способы диагностики гетероскедастичности:
 - (а) визуализация
 - (б) 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)
 - Робастные стандартные ошибки
22. Нетипичные наблюдения:
 - Outliers – по зависимой переменной
 - Leverage – по предиктору
 - Влиятельные наблюдения
23. Стюдентизированные остатки: понимать, что используются для диагностики выбросов, как делать вывод по полученным значениям
24. Hat-matrix: определение потенциала влияния наблюдений, уметь получить матрицу проекции по заданным значениям x , уметь с помощью hat-matrix получить из y наблюдаемого значения y предсказанного, знать свойства hat-matrix
25. Мера Кука и мера DFBETA: в чем разница между этими мерами, как делать вывод на основе этих мер
26. Что делать с нетипичными и влиятельными наблюдениями? Стоит ли их удалять?