

## Learning outcomes к экзаменационной работе

1. Кластерный анализ:
  - Иерархический кластерный анализ: параметры кластеризации, методы агломерации, виды расстояний; построение и интерпретация дендрограммы
  - Кластеризация методов k-средних: общая логика, отличия от иерархического кластерного анализа
  - Кластеризация неколичественных данных: методы сходства и связанные с ними расстояния
2. Метод наименьших квадратов: ключевой принцип. Выведение формул оценок для коэффициентов в модели: частный случай для парной регрессии. Уметь рассчитать МНК-оценки коэффициентов для парной линейной регрессии по заданным значениям отклика и предиктора
3. Условия верные по построению регрессионной модели (равенство суммы остатков нулю, нескоррелированность остатков и предиктора)
4. Множественная линейная регрессия. Ключевые предикторы и контрольные переменные. Каким требованиям должны удовлетворять контрольные переменные?
5. Интерпретация оценок коэффициентов в множественной линейной регрессии
6. Проверка значимости коэффициентов в множественной линейной регрессии
7. Построение доверительного интервала для коэффициента в регрессионной модели с последующей интерпретацией
8. Уметь рассчитать предсказанное значение зависимой переменной при условии заданного значения предиктора
9. Коэффициент детерминации: уметь рассчитать, интерпретировать, понимать ограничения
10. Проверка гипотезы о незначимости коэффициента детерминации
11. Условия Гаусса–Маркова. BLUE-оценки
12. ANOVA таблица для множественной линейной регрессии: уметь заполнить пропуски и сделать выводы на основе статистик
13. Запись спецификации регрессионной модели в векторно-матричном виде
14. Уметь по заданным значениям предикторов и значениям отклика получить вектор оценок коэффициентов регрессионной модели на основе формулы:  $(X^T X)^{-1} X^T y$
15. Эндогенность в регрессионной модели. Самое общее определение: эндогенность как проблема значимых пропущенных переменных
16. Последствия эндогенности для свойств оценок модели при применении классического МНК в качестве метода оценивания
17. Источники эндогенности:
  - значимые пропущенные переменные
  - selection bias (в частности, self-selection bias и non-response bias)
  - simultaneity problem

- post-treatment bias
- ошибки измерения

- Мультиколлинеарность: суть проблемы, примеры
- Откуда берется мультиколлинеарность?
- Последствия мультиколлинеарности
- Почему в случае строгой мультиколлинеарности оценки в модели не могут быть получены?
- Способы диагностики мультиколлинеарности:
  - исходная корреляционная матрица предикторов
  - визуализация: scatterplots
  - Variance inflation factor (*Примечание: VIF больше 10 указывает на сильную мультиколлинеарность, в качестве альтернативы можно использовать tolerance: меньше 0.1 указывает на сильную мультиколлинеарность*)
- Гетероскедастичность: определение, примеры
- Откуда берется гетероскедастичность?
- Последствия гетероскедастичности
- Способы выявления гетероскедастичности:
  - теоретические предпосылки
  - визуализация
  - 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)
- Нетипичные наблюдения:
  - Outliers – по зависимой переменной
  - Leverage – по предиктору
  - Влиятельные наблюдения
- Стьюдентизированные остатки: понимать, что используются для диагностики выбросов, как делать вывод по полученным значениям
- 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
- Мера Кука и мера DFBETA: в чем разница между этими мерами, как делать вывод на основе этих мер
- Что делать с нетипичными и влиятельными наблюдениями? Стоит ли их удалять?
- Различия между moderation effect и mediation effect. Примеры этих эффектов из исследований в области социальных наук: уметь разложить «цепочку» на independent variable – dependent variable – mediator и independent variable – dependent variable – moderator (condition), соответственно
- Переменные взаимодействия как способ проверки moderation effect. Правила построения спецификации линейной регрессионной модели с переменными взаимодействия

34. Интерпретация исходных коэффициентов в линейной регрессионной модели с переменными взаимодействия, а также интерпретация с помощью предельных эффектов (см. следующий пункт)
35. Предельный эффект: вычисление предельного эффекта по оценкам коэффициентов регрессионной модели, интерпретация
36. Визуализация результатов (график, демонстрирующий взаимосвязь предиктора-«условия» и предельного эффекта, и интерпретация данного графика: значения предельного эффекта и их значимость)
37. Центрирование (и другие возможные алгебраические преобразования переменных) в контексте регрессионного анализа с переменными взаимодействия: содержательный смысл данного преобразования, интерпретация коэффициентов при преобразованных предикторах
38. Мультиколлинеарность в контексте регрессионного анализа с переменными взаимодействия, или «не так страшен черт, как его малюют»
39. Сравнение подходов: включение переменных взаимодействия в регрессионную модель и оценивание регрессионных моделей на отдельных подвыборках, выделенных на основе значения предиктора-«условия» (moderator)
40. Метод главных компонент: постановка задачи, особенности метода и его геометрическая интерпретация
41. Алгоритм реализации метода главных компонент (с нахождением собственных значений и векторов); Информативность главных компонент
42. Выбор количества главных компонент: информативность и интерпретируемость, правило Кайзера и правило Кэттела, график каменистой осыпи
43. Интерпретация выдач и графиков R с результатами применения метода главных компонент
44. Различия факторный анализ VS МГК. Почему нельзя фактор отождествлять с индексом, получаемым в результате МГК?
45. Спецификация модели факторного анализа (ФА)
46. Разложение вариации наблюдаемого индикатора: общность и специфичность
47. Задачи разведывательного этапа ФА
48. Задачи подтверждающего (конфирматорного) этапа ФА
49. Допущения модели в рамках разведывательного ФА. Какие допущения можно ослабить в контексте подтверждающего ФА?
50. Идентификация модели ФА, уметь называть и определить (по количеству параметров и количеству исходных источников информации) примеры неидентифицируемой модели, точно идентифицируемой модели и идентифицируемой модели
51. Интерпретация факторной нагрузки в контексте стандартизированного решения:
  - (a) как корреляция между фактором и показателем
  - (b) квадрат факторной нагрузки как вариация наблюдаемого индикатора, объясненная фактором
52. Уметь на основе полученных значений MI (modification index) и EPC (expected parameter change) делать выводы об уточнении спецификации модели подтверждающего ФА
53. Уметь делать выводы о качестве моделей подтверждающего факторного анализа на основе мер SRMR, RMSEA, CFI

54. Понимать, зачем нужны информационные критерии (AIC, BIC) и что они показывают, как на основе них выбрать модель