

Домашнее задание

Deadline: 23.59 19 февраля 2022

Общая постановка задачи Задание выполняется на массиве CDAhw1.dta. Описание переменных представлено в этом [файле](#). Вам необходимо смоделировать участие в голосовании на выборах 4 декабря 2011 года. Используйте из предложенного массива в качестве отклика вопрос со следующей формулировкой: q29 — Принимали ли вы участие в выборах в Государственную Думу России 4 декабря этого года?

1. Самостоятельно выберите набор предикторов: хотя бы один из них должен быть с непрерывной шкалой, хотя бы один — дамми-переменная и хотя бы один — порядковая переменная (категорий от трех и более). Сохраните в массиве только те переменные, которые Вам нужны. Приведите описательные статистики переменных, если необходимо, выполните преобразования для последующего анализа данных и более удобной интерпретации.
2. Оцените логит-модель, приведите оценки коэффициентов β при всех переменных. Запишите спецификацию модели, используя латентную переменную. Поясните идею представления зависимой переменной как латентной: как можно проинтерпретировать данную латентную переменную. Дайте предварительную интерпретацию оценок коэффициентов в терминах взаимосвязи латентной переменной и предикторов.
3. Переоцените модель в допущении о стандартном нормальном распределении ошибок. Сравните оценки коэффициентов с соответствующими оценками, полученными в предыдущем пункте. Объясните своими словами соотношение между оценками в пробит- и логит-модели, обратившись к идее идентификации модели.
4. В этом и во всех последующих пунктах работайте с оценками логит-модели. Рассчитайте средние предельные эффекты (*marginal effect*) принадлежать группе голосовавших в выборах $y = 1$ и проинтерпретируйте их только для тех переменных, для которых вычисление средних предельных эффектов содержательно осмысленно.
5. Рассчитайте предсказанные вероятности для двух индивидов с заданными Вами характеристиками: рассчитайте разность вероятностей (*discrete change*), проверьте на статистическую значимость (используйте доверительные интервалы) и проинтерпретируйте полученное значение разности.
6. Проинтерпретируйте полученные оценки коэффициентов при предикторах в терминах отношения шансов.
7. Проверьте гипотезу о согласии наблюдаемых значений (данных) и модели посредством Hosmer-Lemeshow goodness-of-fit test. Кратко поясните, каким образом устроена статистика теста (общий смысл), сделайте вывод на основании полученных результатов.
8. Сохраните предсказанные вероятности участия в выборах 4 декабря 2011 г., задайте сами порог отсечения и представьте в качестве результата confusion matrix. По представленной таблице классификации рассчитайте ошибку первого рода, ошибку второго рода и мощность (запишите в явном виде, как рассчитываются

эти значения, объясните, что они содержательно показывают в контексте поставленной содержательной задачи в этом домашнем задании). Можно ли говорить о значимых различиях по сравнению с baseline ассигасу? Прокомментируйте результаты.

9. Сравните решение (выбранный Вами порог отсечения) с классификацией наблюдений в соответствии с оптимальным порогом, выбранным на основе минимизации ошибки классификации *Примечание: сравните по ряду мер: ассигасу, специфичность, чувствительность. Сделайте вывод: какая модель из представленных, на Ваш взгляд, наиболее удачная с точки зрения классификации?*
10. Сравните модели:
 - Предложите модель, вложенную в Вашу. Запишите ее спецификацию. Проверьте гипотезу об отсутствии различий между более и менее экономной моделью критерием отношения правдоподобия. Сделайте статистический и содержательный вывод.
 - Предложите самостоятельно две спецификации невложенных моделей. Протестируйте посредством информационных критериев, какая модель является предпочтительной.
 - Для порядковой переменной проверьте гипотезу о том, что коэффициенты при предикторах-категориях одновременно равны 0 (иными словами, нет статистически значимых различий в их эффекте по сравнению с базовой категорией). Используйте критерий Вальда. Сделайте статистический и содержательный вывод.