

1 Домашнее задание

На занятии мы обсуждали регрессию на категориальные признаки, эффекты взаимодействия и полиномиальные модели. Был разобран пример с ростом зубов морских свинок (`toothgrowth`) и с абитуриентами университетов.

Задание 1.1. Для данных `teengambling` (сколько денег подростки из разных социальных слоев спускают на азартные игры) провести анализ (аналогично анализу для `toothgrowth`) и определить зависимость между полом, доходом, статусом, образованием и тратами на игры. Проинтерпретировать результат. Завершить анализ графиком, который окончательно прояснит происходящее.

Hint: В исходных данных много признаков, но далеко не все они значимые.

Задание 1.2. Повторить анализ для данных по абитуриентам, при этом рассмотреть частные университеты. Сравнить модели. Построить модель для всех университетов, используя эффекты взаимодействия с фактором `PPIND`, используя признаки, отобранные при анализе видов университетов отдельно. Проинтерпретировать коэффициенты или объяснить, почему это сделать невозможно.

2 Дополнительно

Задание 2.1. Провести анализ для данных `infmort` (статистика по детской смертности по разным странам в зависимости от региона, дохода и того, является ли страна экспортером нефти). Нужно построить зависимость детской смертности от всего остального и проинтерпретировать результаты. Разумеется, хороший график не помешает.

Hint: Во-первых, тут данных мало, а признаков много, поэтому не имеет смысла строить сложную модель с большим количеством перекрестных влияний. Во-вторых, страны — это не морские свинки, мы про каждую что-то знаем, поэтому при анализе и интерпретации можно и нужно пользоваться информацией о конкретных странах и регионах.

Задание 2.2. Данные `twins` взяты из статьи Cyril Burt, 1966: “The genetic determination of differences in intelligence: A study of monozygotic twins reared apart”. Данные содержат сведения об IQ разлученных однойцевых (идентичных) близнецов, один из которых выращен биологическими, а другой приемными родителями. Также известен социальный класс биологических родителей. Нужно построить зависимость интеллекта приемного ребенка от интеллекта его близнеца и социального класса его биологических родителей и проинтерпретировать результат.

Кстати, известно, что Бёрт был обвинен в фальсификации. А Вы видите что-нибудь подозрительное в этих данных? Разумеется, правильный график не будет лишним.