

Моделирование статистических зависимостей

Домашнее задание 1

Максимум- 10 баллов

Домашнее задание необходимо направить на почту a.a.egorov07@gmail.com. Допустимые форматы - R html или PDF файл с описание результатов + скрипт файл с кодом, на основе которого были получены результаты

о. Данные. Для выполнения домашнего задания предлагается использовать набор данных «education» из пакета “robustbase”. Эти данные использовались в работе (Chatterjee and Price, 1977) и отражают расходы на образование в 50 штатах.

Набор данных содержит следующие переменные:

State-штат

Region – регион (1-Северо-восток; 2-Север; 3- Юг; 4- Запад)

X_1 – количество жителей в расчете на тысячу жителей, проживающих в урбанизированных районах в

X_2 – доход на душу населения

X_3 – количество жителей в расчете на тысячу жителей в возрасте младше 18 лет

Y – государственные расходы на образование

1. Загрузить указанный набор данных в R, построить таблицу с описательными статистиками (0.5 балла)
2. Построить регрессионную модель $Y \sim X_1$, проинтерпретировать результаты (0.5 балла)
3. Построить график $Y \sim X_1$, на котором точками будут отмечены исходные данные и на котором будет представлена линия регрессии (одновременно) (1 балл)
4. Построить регрессионную модель $Y \sim X_1 + X_2 + X_3$, проинтерпретировать результаты (0.5 балла)
5. Какую дисперсию Y объясняет каждый из регрессоров в модели из п.4? (1 балл)
6. Построить регрессионную модель $Y \sim X_1 + X_2 + X_3 + \text{Region}$, проинтерпретировать результаты. Где в среднем выше расходы на образования в расчете на душу населения: На Западе или северо-востоке? На Юге или на Севере? Почему? (необходимо осуществить замену базового уровня) (2 балла)
7. Построить график прогноза и доверительных интервалов для него на основе модели и п.3. (1.5 балла)
8. Построить график Residuals vs Fitted для модели из п.6. Какой вывод вы можете сделать из этого графика? (0.5 балла)

9. Проверить наличие гетероскедастичности в модели из п.6 на основе теста Бреуша-Пагана (1 балл)

10. Получить устойчивую к гетероскедастичности ковариационную матрицу параметров модели из п.6. Рассчитать новые t -статистики. Сравнить результаты на основе устойчивой ковариационной матрицы с результатами оценки модели, полученными в п.6. (1.5 балла)