**Моделирование статистических зависимостей**

**Домашнее задание 1**

Максимум- 10 баллов

На проверку необходимо прислать два файла: с проведенными расчетами (ipynb/py) и описанием результатов (word или pdf).

Срок сдачи – до 23:59 30/11/2023

1. Сформулировать содержательно задачу исследования. Пояснить, как методы регрессионного анализа помогут ее решить (зависимая переменная в модели регрессии должна быть непрерывной величиной). Подобрать данные пространственной выборки для проведения регрессионного анализа. Ограничения: от 5 объясняющих переменных (из них допускается не более 20% бинарных переменных, все прочие - непрерывные), объем выборки от 100 объектов. (1 балл)

1. Провести первичный анализ данных. Обосновать выбор переменных, которые будут участвовать в анализе. Описать данные, визуализировать их. Предположить, каким образом независимые переменные должны оказывать влияние на объясняемую переменную из полученных предварительных результатов, общей логики, обзора соответствующей литературы. Провести корреляционный анализ, проверить наличие мультиколлинеарных признаков (в т.ч. с помощью VIFов). (1 балл)
2. Построить уравнение регрессии со значимыми коэффициентами, используя пошаговые алгоритмы регрессионного анализа. Рассмотреть возможность перехода от регрессии исходного признака к регрессии его логарифма (1 балл). Выбрать лучшую из полученных регрессионных моделей̆, основываясь на анализе значений коэффициентов детерминации, остаточных дисперсий и по результатам содержательного анализа моделей̆. (1 балл). На основе графиков выделить переменные, влияющие на гетероскедастичность ошибок. Провести тесты Бройша-Пагана и Уайта для проверки гомоскедастичности ошибок. В случае гетероскедастичности выполнить поправку Уайта. (1 балл). Провести анализ ошибок модели (1 балл)
3. Построить модель линейной регрессии с регуляризацией Ридж и применением кросс-валидации, сравнить полученную модель с лучшей моделью из пункта 3. Построить модель линейной регрессии с регуляризацией Лассо применением кросс-валидации, сравнить полученную модель с лучшей моделью из пункта 3 и моделью c регуляризацией Ридж. (1 балла). Сделать обоснованный выбор лучшей модели с регуляризацией, провести анализ ошибок этой модели (1 балл).
4. Выбрать наиболее удачную регрессионную модель. Аргументировать свой выбор, привести подробную интерпретацию модели. (2 балла)