Санкт-Петербургский Политехнический Университет Петра Великого

Физико-механический институт

Кафедра прикладной математики и вычислительной физики

Отчёт по лабораторной работе №4 по дисциплине

«Многомерный статистический анализ»

Выполнил студент гр. 5030102/90401: Веселый Д. В.

Преподаватель: Павлова Л. В.

Санкт-Петербург

2023

# Постановка задачи

По данным химического эксперимента построить и исследовать регрессионную модель

# 1. Регрессия и различные оценки

## 1.1 Построение

Пусть – вектор наблюдений, – вектор целевых переменных, - вектор оценок .

Тогда МНК-оценка

* Оценка Дисперсии , где – размерности вектора наблюдений
* Выборочная матрица Ковариаций
* Корреляция для МНК-оценки a– точная! «крышечка» лишняя
* Стандартные ошибки
* Коэффициент детерминации (смещённый и несмещённый)

*Изображение выглядит как текст

Автоматически созданное описание*

*Изображение выглядит как текст

Автоматически созданное описание*

*Изображение выглядит как текст

Автоматически созданное описание*

*Изображение выглядит как текст

Автоматически созданное описание*

## 1.2 Результаты и выводы

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описание

Подпись и комментарии к рисунку

Гистограмма остатков показывает, что регрессия довольно точно строит зависимость от параметров регрессии, поскольку значения разностей содержатся в интервале (-4.5, 3,5). При этом наибольшее число элементов попадает в интервал примерно от 0.5 до 1.5, что как раз и показывает малое отклонение построенной модели от истинных значений (а в бОльшие по модулю интервалы попадает меньшее число элементов)

Что можно отметить, так это довольно заметную коррелированность 2 и 4 параметра регрессии (в матрице корреляций для строк 2 и 4 довольно близкие по модулю к 1 значения). Кроме того,  
стандартная ошибка на 1 параметр заметно ниже, чем у всех остальных.

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описание

Коэффициент детерминации почти равен единице, что указывает на то, что модель регрессии сильно лучше модели среднего. Помимо этого, визуально регрессия достаточно хорошо совпадает с истинными значениями целевой переменной.

# 2. Доверительные интервалы области и проверка стат. гипотезы

## 2.1 Построение и результаты

Доверительный интервал для

Доверительная область в соответствии с принципом Тьюки:

Строим с уровнем значимости

Изображение выглядит как текст

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Видим, что коэффициенты попадают в доверительные интервалы.

Что можно сказать о точности интервальных оценок?

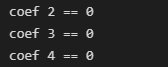
Разбросы интервала и области достаточно большие. По ним довольно сложно “прикинуть” хорошее значение параметров.

Теперь сделаем проверку гипотезы о равенстве нулю коэффициентов:

По формуле выше принимаем гипотезу о равенстве нулю i-го коэффициента

Изображение выглядит как текст

Автоматически созданное описание



Оказалось, что можно сделать равными нулю 2, 3, 4 коэффициенты. Далее:

Изображение выглядит как диаграмма

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Видим, что, оставив один коэффициент (все-таки оставив один признак!), оценка дисперсии регрессии увеличилась в два раза, однако визуально график ухудшился несильно (для наглядности – приведите здесь и исходный график оцененной регрессии).

Нет интервальной оценки!

После того, как она будет построена, сравните точность оценивания коэфф. для двух разных моделей. Будет повод вспомнить о «проклятии размерности»…

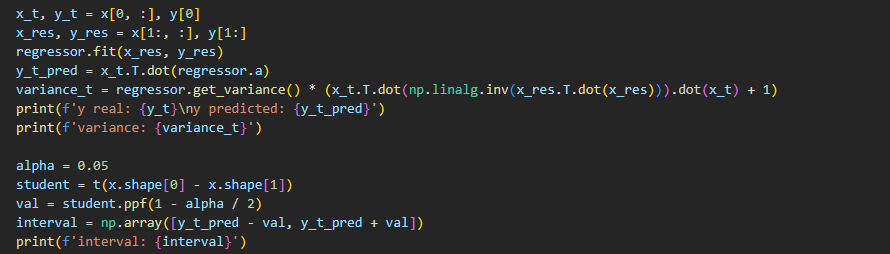
Для случая одного признака интервалы стали заметно меньше, чем для 4 признаков. Скорее всего, это и объясняется тем, что мы уменьшили число признаков.

# 3. Прогнозы

Убираем одно наблюдение и строим прогноз:

Оценка Дисперсии для прогноза: ( + 1)

Интервал:

**

Изображение выглядит как текст

Автоматически созданное описание

Видим, что прогноз достаточно близок к истинному значению, правда реальное значение не попало в доверительный интервал.

# 4. Выводы

В результате проделанной работы можно сделать вывод о том, что регрессия позволяет довольно точно построить зависимость целевого признака от наблюдений. Кроме того, с помощью проверки линейных гипотез можно уменьшить вычислительную нагрузку, убрав некоторые параметры у наблюдений.