Санкт-Петербургский Политехнический Университет Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

Отчёт по лабораторной работе №4 по дисциплине

«Многомерный статистический анализ»

Выполнил студент гр. 5030102/90401: Реутов А. С.

Преподаватель: к.ф.-м.н., доцент, Павлова Л. В.

Санкт-Петербург

2023

Оглавление

[Постановка задачи 3](#_Toc131943178)

[Построение оценок 3](#_Toc131943179)

[Доверительные интервалы 6](#_Toc131943180)

[Проверка гипотез 7](#_Toc131943181)

[Прогноз 9](#_Toc131943182)

[Вывод 10](#_Toc131943183)

# Постановка задачи

По результату химического эксперимента, в котором реагируют вещество с веществом и получается вещество , построить и проанализировать регрессионную модель.

# Построение оценок

Имеется матрица признаков Х, целевой вектор y, тогда **МНК-оценка параметров «а»** выражается следующим образом:

Пусть оценка для вектора y, тогда **оценка дисперсии :**

,

где n – количество наблюдений, m – число признаков

**Оценка матрицы ковариаций:**

**Стандартная ошибка оценки -го коэффициента:**

**Матрица корреляций corr(a):**

**Код программы:**

**Изображение выглядит как текст

Автоматически созданное описание**

**Результаты:**

**Изображение выглядит как текст

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание**

По корреляционной матрице заметна сильная корреляция между вторым и четвертым параметром.

**Построим гистограмму остатков:**

**Изображение выглядит как диаграмма

Автоматически созданное описание**

Рисунок Гистограмма остатков

Гистограмма остатков показывает, насколько хорошо модель линейной регрессии соответствует данным. Как видно по гистограмме, остатки распределены нормально. Это говорит о том, что модель линейной регрессии хорошо соответствует данным. Это, скорее, говорит о том, что вектор сл.отклонений **ε** ~ N(0,… ), т.е. дальнейшая работа – проверка гипотез, построение интервальных оценок и пр. – правомочна.

Построим график регрессии и оцененной регрессии:

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок Регрессия и ее оценка

Коэффициент детерминации :

Где

Несмещенный коэффициент детерминации

Результаты:

Изображение выглядит как текст

Автоматически созданное описание

Коэффициент детерминации довольно высокий, что означает, что модель регрессии сильно лучше модели среднего.

# Доверительные интервалы

Построим индивидуальные доверительные интервалы для коэффициентов регрессии и для них же – обобщенную доверительную область на основе принципа Тьюки.

Индивидуальные доверительные интервалы для :

Где уровень значимости

Получим следующие интервалы:

Изображение выглядит как текст

Автоматически созданное описание

Рисунок Индивидуальные доверительные интервалы

Доверительная область:

Результаты:

Изображение выглядит как текст

Автоматически созданное описание

Рисунок Доверительная область (на основе принципа Тьюки)

В обоих случаях коэффициенты попадают в доверительные интервалы. Точность интервальных оценок небольшая и по ним довольно сложно определить хорошее значение (например, для четвертого параметра).

# Проверка гипотез

Проверим гипотезу о равенстве отдельных коэффициентов нулю.

Область принятия гипотезы:

В результате получим булевый вектор:



Примем второй и четвертый коэффициенты равными нулю и посмотрим, что получится (принимаем их равными нулю на основе того, что как раз эти коэффициенты сильно коррелируют друг с другом могут вносить мультиколлинеарность в модель, что может приводить к ее нестабильности и менее точным прогнозам).

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок

Видно, что, убрав два признака, модель регрессии не сильно ухудшилась (хотя вообще визуально сложно понять ухудшилась ли она).

Посмотрим на оценки:

Изображение выглядит как текст

Автоматически созданное описание

Оценка дисперсии увеличилась немного больше чем в два раза.

Интервалы:

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Проверим гипотезу

Нулевая гипотеза утверждает, что набор независимых переменных X не оказывает статистического влияния на переменную y

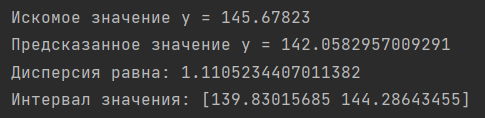
По результатам была принята альтернативная гипотеза, которая говорит о том, что существует хотя бы один фактор, который значимо влияет на зависимую переменную.

# Прогноз

Для простоты уберем первое наблюдение, тогда:

Оценка дисперсии для прогноза: ( + 1), X – исходная матрица без

Получим:



Предсказанное значение близко к искомому, но истинное значение не попало в доверительный интервал.

# Вывод

В результате проведенной работы было установлено, что линейная регрессия может быть использована для построения зависимости целевого признака от наблюдений с высокой точностью. При использовании гипотезы о равенстве некоторых параметров нулю, можно улучшить производительность вычислений и сократить размерность модели. Однако, не следует убирать параметры основываясь только на результате проверки гипотезы – необходимо дополнительно провести анализ корреляции между параметрами и на основе этого делать выводы о целесообразности их сокращения.