

Практика 8. Регрессионные модели характеристик насосных агрегатов

Часть 1. Подготовка данных

1. Скачать с сайта курса данные по характеристикам насосных агрегатов (НА).

Данные представляют собой реальные обезличенные данные по напорным характеристикам (QH) и характеристикам мощности (η) насосных агрегатов нефтеперекачивающих станций. В одном файле содержатся данные по одному НА. В имени файла указывается номер нефтеперекачивающей станции (НПС) и номер агрегата. Каждый файл содержит 3 колонки Q, H, η .

Если данные по всем вариантам не выложены, нужно взять пример данных (он точно выложен) и начать отлаживать код на нем.

2. Для зачитки данных по агрегатам воспользуйтесь функцией (есть в архиве с данными):

```
function [Q, H, N] = read_pump_data(station_number, pump_number)
```

функция на вход принимает номер НПС и номер НА (возьмите их своего варианта), на выходе выдает Q, H, η .

Если данные не выложены, временно берите `station_number = 1` и `pump_number = 1`. По готовности данных, переделать под свой вариант.

Часть 2. QH-характеристика и характеристика КПД

1. Построить регрессионную модель с полиномом 0-го, 1-го, 2-го, 3-го порядков.
2. Построить модели на основе аппроксимации из литературы по гидравлике:

- a. Для QH-характеристики

$$H(Q) = a - bQ^2$$

- b. Для характеристики мощности

$$\eta(Q) = k_1Q - k_2Q^2$$

Для оценок коэффициентов моделей использовать как формулу $\hat{\beta} = (X^T X)^{-1} X^T Y$, так и функцию `regress`

3. Построить график $H(Q)$, на который вывести
 - a. исходные данные (вывести отдельными точками, например, кружками)
 - b. предсказания всех моделей из части 2 в диапазоне Q от нуля до $1.5 * \max(Q)$, (график вывести линиями)
 - c. легенду для графиков всех моделей (чтобы не путать модели друг с другом)

Отмасштабировать график по H в диапазоне от нуля до $\max(H)$, используя функцию `ylim`.

4. Построить график $\eta(Q)$, на который вывести
 - a. исходные данные (вывести отдельными точками, например, кружками)
 - b. предсказания всех моделей из части 2 в диапазоне Q от нуля до $1.5 * \max(Q)$, (график вывести линиями)
 - c. задать диапазон по кпд. от нуля
 - d. вывести легенду для графиков всех моделей (чтобы не путать модели друг с другом)

Отмасштабировать график по η в диапазоне от нуля до $\max(\eta)$, используя функцию `ylim`.

5. Для каждой модели рассчитать вектор регрессионных остатков $e = Y - X\hat{\beta}$
 - a. построить график регрессионных остатков как функции от расхода Q (вывести отдельными точками, например, кружками)
 - b. вывести гистограмму регрессионных остатков

Провести визуальный анализ графика регрессионных остатков и их гистограмм. Найти подозрительные на ваш взгляд особенности. По возможности обосновать свои подозрения.

6. Для каждой модели рассчитать (если рассказано на лекции):
 - a. несмещенную оценку дисперсии шума

b. коэффициент детерминации R^2

Вопросы к защите

1. Коэффициент детерминации R^2
 - a. Понятие тривиальной (опорной) модели. Эквивалентность МНК тривиальной модели и прогноза по средневыворочному.
 - b. Формулы и содержательный смысл RSS, TSS, ESS
 - c. Доказательство равенства $TSS = RSS + ESS$
 - d. Формула коэффициента детерминации R^2 . Чему соответствуют значения R^2 , равные 0%, 50%, 100%.
 - e. Как на основе R^2 выбрать лучшую модель и нескольких моделей-кандидатов?
2. Что можно сказать о возможности использования построенных моделей в работе?
 - a. Будут ли они отражать физику процесса на новых режимах работы?
 - b. Какова физика процесса?
 - c. Чтобы ответить на вопросы, надо указать, какие возможны новые режимы (а какие нет)?
 - d. Чтобы рассуждать о режимах, надо вспомнить, где установлен или может быть установлен насосный агрегат. Это не обязательно магистральный трубопровод.
3. Модель какого порядка лучше выбрать и почему?
4. Есть ли выбросы в ваших данных?
5. Ваши замечания по полученным регрессионным остаткам. Наблюдается ли тренд или колебательность?