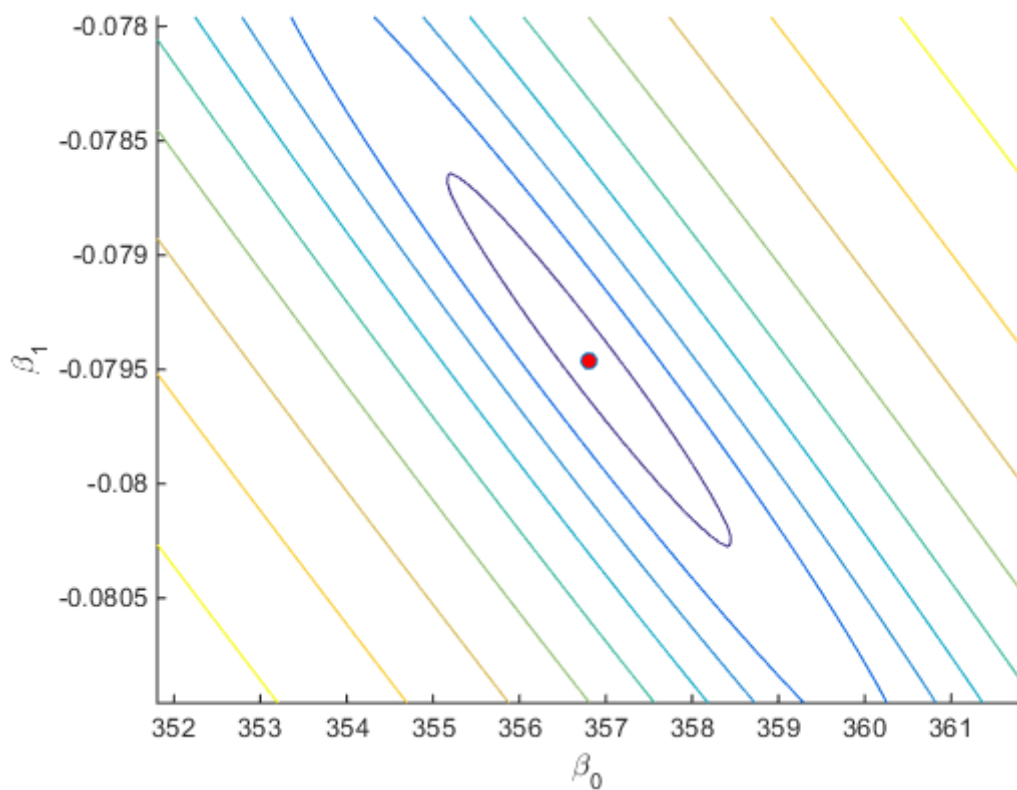


Практика 5. Критерий МНК как функция коэффициентов регрессии в трехмерном пространстве

1. Взять оценки коэффициентов для **линейной** модели QH-характеристики из прошлой работы ($\hat{\beta}_0, \hat{\beta}_1$) и в **данной работе считать их истинными** ($\beta_{ист}$). Сгенерировать выборку объема n по модели с полученным $\beta_{ист}$. Задаться дисперсией шума и объемом выборки от 10 до 20.
2. Рассчитать МНК-оценки коэффициентов регрессии на основе аналитического решения задачи минимизации функционала качества. Расчет сделать функцией regress и по формуле:
$$\hat{\beta} = (X^T X)^{-1} X^T Y$$
3. Построить трехмерный график (surf) и график изолиний (contour) функционала качества МНК $Q(\beta)$. Отобразить точкой аналитическое решение (plot3). Убедиться, что точка лежит точно в минимуме.
4. На графике изолиний (contour) необходимо обеспечить замкнутость линий уровня, (см. образец). Если это не удастся, можно прологарифмировать величину критерия $\log(J)$ перед вызовом contour (**сам критерий МНК трогать не надо!**).



Вопросы к защите

1. Объяснение невязок между моделью и данными как реализаций случайного шума. Насосный агрегат является объектом с детерминированным поведением (при одинаковом расходе будет один и тот же дифнапор). Однако, при сборе данных для одного заданного расхода могут получаться разные значения дифнапора насосного агрегата.
 - a. Указать конкретные причины, из-за которых появляется случайный разброс дифнапора для одного расхода. Привести по крайней мере 3 причины.
 - b. Пояснить, почему разброс носит случайный характер.
 - c. По какой причине разумно считать шум нормальной случайной величиной?
2. Вероятностная модель выборки до измерений. Имеется детерминированная составляющая в напоре, обусловленная физикой объекта. Имеется случайная составляющая (шум), обусловленный факторами, указанными в предыдущем пункте (а также другими факторами).

Записать вероятностную модель до измерений, указать в ней детерминированную составляющую отклика.

- a. Формы записи:
 - i. Скалярная, для каждого i -го элемента выборки
 - ii. Векторная, для каждого i -го элемента выборки
 1. Размерность вектора коэффициентов
 2. Размерность вектора факторов
 - iii. Векторная, для всех элементов выборки.
 1. Запись матрицы плана через вектор факторов
 2. **Для всех переменных указать размерности**
 - b. Нюансы
 - a. Вектор факторов
 - i. для задания линейной зависимости
 - ii. для задания полиномиальной зависимости
3. Восстановление коэффициентов регрессии по экспериментальным данным
- a. Понятие регрессионного остатка.
 - i. Изображение остатков на графике Q-N (эскиз)
 - ii. Формула расчета остатков в скалярном виде (для каждого i -го элемента)
 - iii. Формула расчета остатков векторном виде (для всех элементов выборки).
 - iv. Чем отличается остаток от шума
 - v. Как остатки характеризует качество модели?
 - b. Функционал метода наименьших квадратов
 - i. Как функционал качества характеризует остатки
 - ii. Записать функционал МНК в скалярном и векторном виде. Показать их эквивалентность.
 - iii. Может ли функционал быть отрицательным? Может ли он быть равным нулю. Если может, указать условия, при которых это возможно.
 - iv. В чем разница между $\hat{\beta}$, β , $\beta_{\text{ист}}$?
 - v. Графический вид функционала для линейной зависимости ($k=2$) (эскиз трехмерного графика, эскиз графика изолиний)
 - vi. Что происходит с остатками при достижении минимума функционала качества МНК?
 - vii. Что происходит с графиком регрессионной модели в осях исходных данных (Q, N) при достижении минимума функционала качества МНК?
 - viii. В какой точке значение функционала ниже: $\hat{\beta}$, $\beta_{\text{ист}}$?
 - c. Формальная постановка задачи поиска оценок коэффициентов регрессии $\hat{\beta}$ как задачи минимизации функционала (в векторном виде). Что дано, что требуется найти, за счет чего это достигается.
 - d. Аналитическое решение задачи поиска оптимальных оценок $\hat{\beta}$.
 - i. Что такое вектор-градиент?
 - ii. Вывод системы нормальных уравнений и ее решение
4. Коэффициент детерминации R^2
- a. Вероятностная модель выборки до измерений
 - b. Вероятностная модель тривиальной регрессионной модели
 - c. Понятие тривиальной (опорной) модели. Эквивалентность МНК тривиальной модели и прогноза по средневыборочному.
 - d. Формулы и содержательный смысл RSS , TSS

- e. Доказательство равенства $TSS = RSS + ESS$. Содержательный смысл ESS .
- f. Формула коэффициента детерминации R^2 . Чему соответствуют значения R^2 , равные 0%, 50%, 100%.
- g. Как на основе R^2 выбрать лучшую модель из нескольких моделей-кандидатов?