Мультиколлинеарность

Теоретическая мультиколлинеарность данных – явление, наблюдаемое при нарушении условий теоремы Гаусса – Маркова об отсутствии точной линейной связи между регрессорами. При наличии теоретической мультиколлинеарности однозначное нахождение оценов МНК коэффициентов регрессии невозможно.

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K + \varepsilon ,$$

Теоретическая мультиколлинеарность: Rank(X) < k + 1

Ex.1.

$$\ln wage = \beta_1 + \beta_2 S + \beta_3 MALE + \beta_4 FEMALE + \ldots + \varepsilon ,$$

$$FEMALE + MALE = i$$

Ex.2.

$$\ln price = \beta_1 + \beta_2 livsq + \beta_3 nonlivsq + \beta_4 totsq + \ldots + \varepsilon ,$$

$$livsq + nonlivsq = totsq$$

Пример теоретической мультиколлинеарности

Ex.3.

$$\Pr ice = \beta_1 + \beta_2 D_I + \beta_3 D_{II} + \beta_4 D_{III} + \beta_5 D_{IV} + \ldots + \varepsilon,$$

$$D_I + D_{II} + D_{III} + D_{IV} = i$$

Dummy trap

Квазимультиколлинеарность

При работе с реальными данными часто имеет место квазимультиколлинеарность, когда между регрессорами существует почти линейная зависимость.

$$Y = X\beta + \varepsilon$$

теоретическая мультиколлинеарность :

$$\mathrm{rang}X < k + 1$$

$$\Rightarrow \mathrm{rang}(X'X) < k + 1 \Rightarrow \det(X'X) = 0$$

кваимультиколлинеарность :

$$\det(X'X) \approx 0$$

8

Признаки мультиколлинеарности

- Небольшие изменения в данных приводят к значительным изменениям в оценках коэффициентов регрессии.

- Многие коэффициенты по-отдельности не значимы, хотя в целом регрессия адекватная, R^2 может быть достаточно высоким.

- Оценки коэффициентов регрессии (обычно незначимых) могут иметь "неправильный" знак (с экономической точки зрения).

Индикаторы мультиколлинеарности

- В корреляционной матрице факторов встречаются элементы, по модулю близкие к 1.

- Достаточно большое значение (больше 6) VIF – variance inflation factor хотя бы для одного фактора

$$VIF(X_j) = \frac{1}{1 - R_j^2},$$

где R_j^2 – коэффициент множественной детерминации регрессора X_j на все остальные регрессоры.

Методы борьбы с мультиколлинеарностью

•Переспецификация модели (функциональные преобразования переменных)

•Исключение одной или нескольких объясняющих переменных

•Метод главных компонент

•Использование ridge (гребневых), LASSO и т.п. оценок параметров