

Семинар 8.

Мультиколлинеарность. Метод главных компонент.

1. Известно, что выборочная корреляция между переменными x и z равна 0.9.
 - (a) Найдите коэффициенты VIF для x и z в регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$.
 - (b) В каких пределах могут лежать коэффициенты VIF для x и z в регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \varepsilon_i$?
2. Теоретическая регрессионная зависимость и выборочная корреляционная матрица регрессоров имеют вид:

$$Y_i = \beta_1 + \beta_2 X_{i1} + \beta_3 X_{i2} + \beta_4 X_{i3} + \varepsilon_i,$$

$$\widehat{Corr}(X) = \begin{pmatrix} 1 & 0.95 & 0 \\ 0.95 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- (a) Найдите число обусловленности для матрицы $\tilde{X}'\tilde{X}$, где \tilde{X} — матрица центрированных и нормированных значений регрессоров. Какой вывод можно сделать о наличии мультиколлинеарности?
- (b) Вычислите все главные компоненты (т.е. выразите их через линейные комбинации столбцов \tilde{X}), объясняющие не менее 70% общей дисперсии.
- (c) Вычислите матрицу факторной нагрузки для отобранных на предыдущем шаге главных компонент. Проинтерпретируйте полученные результаты.
- (d) Выразите коэффициенты исходной регрессии через коэффициенты регрессии на главные компоненты, объясняющие не менее 70% общей дисперсии.