

## Семинар 9.

## 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. Метод главных компонент (Principal Component Analysis, PCA).
  - Вспомните МГК. Найдите первую главную компоненту  $Z_1$ .
  - Определим матрицу факторной нагрузки как  $A = C'\Lambda^{1/2}$ , где  $\Lambda$  — ковариационная матрица некоррелированных главных компонент. Покажите, что элемент  $(a_{ij})$  матрица  $A$  показывает удельный вес влияния пронормированной главной компоненты  $j$  на стандартизированный признак  $x_i$
3. Теоретическая регрессионная зависимость и выборочная корреляционная матрица центрированно-нормированных регрессоров имеют вид:

$$y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i,$$

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

- (a) Вычислите все главные компоненты (т.е. выразите их через линейные комбинации столбцов  $\tilde{X}$ ). Сколько главных компонент надо выбрать, чтобы они объясняли не менее 70% общей дисперсии?
- (b) Вычислите матрицу факторной нагрузки. Проинтерпретируйте полученные результаты.