

## Семинар 6.

## Тестирование гипотез.

**Сложение блочных матриц.** Две матрицы, разбитые на блоки одинаковым образом, можно складывать:

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] + \left[ \begin{array}{c|c} E & F \\ \hline G & H \end{array} \right] = \left[ \begin{array}{c|c} A+E & B+F \\ \hline C+G & D+H \end{array} \right].$$

**Умножение блочных матриц.** Если размеры блоков допускают операцию умножения, то:

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \cdot \left[ \begin{array}{c|c} E & F \\ \hline G & H \end{array} \right] = \left[ \begin{array}{c|c} AE+BG & AF+BH \\ \hline CE+DG & CF+DH \end{array} \right].$$

**Транспонирование блочных матриц.** Пусть матрица

$$M = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right].$$

Тогда

$$M^T = \left[ \begin{array}{c|c} A^T & C^T \\ \hline B^T & D^T \end{array} \right].$$

**Формула Фробениуса (блочное обращение).**

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]^{-1} = \left[ \begin{array}{c|c} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ \hline -H^{-1}CA^{-1} & H^{-1} \end{array} \right],$$

где  $A$  — невырожденная квадратная матрица размерности  $n \times n$ ,  $D$  — квадратная матрица размерности  $k \times k$ ,  $H = D - CA^{-1}B$ .

1. Рассмотрим модель множественной регрессии:

$$y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i.$$

Мы хотим проверить гипотезу следующего вида:

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0.$$

Покажите, что для проверки гипотезы  $H_0$  может быть использована статистика

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} \sim F(q, n-k).$$

2. Множественную регрессию с любым количеством регрессоров можно разбить на несколько шагов со вспомогательными регрессиями с меньшим числом регрессоров. Вместо непосредственного включения переменной  $x$  в качестве регрессора в модель можно сначала «очистить» от переменной  $x$  зависимую переменную  $y$  и остальные регрессоры, а затем оценить регрессию для «очищенных» переменных. Эту идею последовательных регрессий формализует теорема Фриша — Во — Ловелла.

**Теорема Фриша — Во — Ловелла (англ. Frisch–Waugh–Lovell theorem, FWL theorem)**

Рассмотрим два алгоритма.

Алгоритм  $A$ : оцениваем регрессию  $y$  на полный набор регрессоров  $X_1$  и  $X_2$  с помощью МНК:

$$\hat{y}_A = X_1 \hat{\beta}_1^A + X_2 \hat{\beta}_2^A.$$

Алгоритм  $B$ :

В1. Оцениваем регрессию  $y$  на часть регрессоров  $X_1$  с помощью МНК:

$$\hat{y}_B = X_1 \hat{\beta}_1^B.$$

В2. Оцениваем регрессию каждого столбца из матрицы  $X_2$  с помощью МНК:

$$\hat{X}_2^B = X_1 \hat{\beta}_1^B.$$

Уточним, что здесь  $\hat{\beta}_1^B$  — это не вектор, а целая матрица, в которой содержатся оценки регрессии каждого столбца из матрицы  $X_2$  на все регрессоры из матрицы  $X_1$ .

В3. Определяем «очищенные» переменные как остатки регрессий первых двух шагов,

$$y^* = y - \hat{y}_B, \quad X_2^* = X_2 - \hat{X}_2^B.$$

В4. Оцениваем регрессию для «очищенных переменных»

$$\hat{y}^* = X_2^* \hat{\beta}_2^B.$$

Алгоритмы  $A$  и  $B$  дают одинаковые оценки коэффициентов  $\hat{\beta}_2^A = \hat{\beta}_2^B$  и финальные векторы остатков  $e_A = y - \hat{y}_A = y^* - \hat{y}^* = e_B^*$ .

**Докажите теорему.**

3. Вместо того чтобы оценивать параметры  $\beta_1, \beta_2$  в модели

$$y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon$$

( $X_1, X_2$  —  $n \times k_1, n \times k_2$  матрицы соответственно,  $\beta_1, \beta_2$  — векторы размерности  $k_1, k_2$  соответственно), строятся МНК оценки этих параметров исходя из модели

$$y = X_1\beta_1^* + X_2^*\beta_2^* + \varepsilon^*,$$

где  $X_2^*$  — матрица остатков, полученных в результате регрессии каждого столбца матрицы  $X_2$  на  $X_1$ .

- (a) Покажите, что полученная таким образом оценка вектора  $\beta_1$  совпадает с оценкой, полученной в результате регрессии  $y$  только на  $X_1$ .
- (b) Найдите смещение оценки вектора  $\beta_1$ .
- (c) Покажите, что оценки вектора параметров  $\beta_2$  в обеих моделях совпадут.

4. Рассмотрим модель множественной регрессии:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i.$$

- (a) Что произойдет с МНК оценками коэффициентов модели, если добавить константу  $c_1$  к каждому наблюдению признака  $x_2$  и другую константу  $c_2$  к каждому наблюдению признака  $x_3$ ?
- (b) Что произойдет с МНК оценками коэффициентов множественной регрессии, если умножить зависимую переменную  $y$  на константу  $c$ ? А если на константу умножить какой-нибудь регрессор?