

Семинар 2.

Модель множественной регрессии. Тестирование гипотез.

1. Используя матрицы $N = X(X'X)^{-1}X'$ и $N_1 = \vec{1}(\vec{1}'\vec{1})^{-1}\vec{1}'$, запишите TSS , RSS и ESS в матричной форме.

Примечание: $\vec{1}$ — вектор размерности $n \times 1$, состоящий из единиц.

2. (Универсиада по эконометрике, МГУ, 2018 год). В некоторой отрасли заработная плата работника следующим образом зависит от его опыта работы и пола:

$$\ln W_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i D_i + \varepsilon_i,$$

где W_i — заработная плата i -го работника в рублях в месяц, X_i — стаж i -го работника в годах, D_i — бинарная переменная, равная единице для женщин и нулю для мужчин, ε_i — случайные ошибки. Предполагается, что все предположения классической линейной модели множественной регрессии выполнены.

Оценка параметров модели при помощи МНК на основе данных о десяти тысячах наблюдений позволила получить следующие результаты:

$$\widehat{\ln W_i} = 10 + 60X_i - 3X_i^2 - 12X_i D_i.$$

Оценка ковариационной матрицы вектора оценок коэффициентов имеет вид:

$$\widehat{Var}(\hat{\beta}) = \begin{pmatrix} 51 & -20 & 2 & 0 \\ -20 & 9 & -1 & 0 \\ 2 & -1 & 0.1 & 0 \\ 0 & 0 & 0 & 0.4 \end{pmatrix}.$$

- (a) Опираясь на полученные оценки параметров, изобразите на одном рисунке графики логарифмов заработной платы типичного работника и типичной работницы в зависимости от их стажа. Интерпретируйте полученный результат.
- (b) Аналитик Афанасий предполагает, что женщины в данной отрасли достигают максимума своей производительности при стаже, равном 10 годам. Соответственно и их заработная плата максимальна именно в этот момент. Сформулируйте (в терминах коэффициентов модели) гипотезу, которая соответствует предположению Афанасия, и проверьте её при уровне значимости 5%.
- (c) Аналитик Евгения утверждает, следующее: «Если обозначить X_M^* — стаж работы, при котором зарплата мужчины максимальна, а X_W^* — стаж, при котором зарплата женщины максимальна, то окажется, что $X_M^* = X_W^* + 1$.

То есть женщины достигают пика своей зарплаты на год раньше мужчин». Сформулируйте (в терминах коэффициентов модели) гипотезу, которая соответствует утверждению Евгении, и проверьте её на уровне значимости 5%.

3. Всего имеется 100 наблюдений. Для первых 50 наблюдений

$$X'X = \begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}, X'y = \begin{pmatrix} 300 & 2000 \end{pmatrix}', y'y = 2100.$$

По последним 50 наблюдениям:

$$X'X = \begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}, X'y = \begin{pmatrix} 300 & 2200 \end{pmatrix}', y'y = 2500.$$

По первым 50 наблюдениям оценивается модель $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, по последним 50 наблюдениям оценивается модель $y_i = \gamma_1 + \gamma_2 x_i + \varepsilon_i$. Предположим, что во всех 100 наблюдениях ε_i независимы и нормальны $N(0; \sigma^2)$. На уровне значимости 5% проверьте гипотезу $H_0 : \beta = \gamma$.