

Семинар 8.  
Фиктивные переменные. Тест Чоу.

1. Рассмотрим следующую регрессионную модель зависимости логарифмированной платы  $\ln(W)$  от уровня образования  $Edu$ , опыта работы  $Exp$ ,  $Exp^2$  и уровня образования родителей  $Fedu$  и  $Medu$ :

$$\widehat{\ln(W_i)} = \hat{\beta}_1 + \hat{\beta}_2 Edu_i + \hat{\beta}_3 Exp_i + \hat{\beta}_4 Exp_i^2 + \hat{\beta}_5 Fedu_i + \hat{\beta}_6 Medu_i.$$

Модель регрессии была отдельно оценена по выборкам из 35 мужчин и 23 женщин, и были получены остаточные суммы квадратов  $RSS_1 = 34.4$  и  $RSS_2 = 23.4$ , соответственно. Остаточная сумма квадратов в регрессии, оцененной по объединенной выборке, равна 70.3. Протестируйте на 5% уровне значимости гипотезу об отсутствии дискриминации в оплате труда между мужчинами и женщинами.

2. (Универсиада по эконометрике, МГУ, 2018 год, отборочный этап). В некоторой отрасли заработка 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  $W_i$  — заработка 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,  $\varepsilon_i$  — случайные ошибки. Предполагается, что все предпосылки 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:

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

- (а) 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 заработка 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. (компьютерный практикум). В файле *Chow\_2.xls* содержатся данные об экономике Баккардии в период с 1 квартала 2015 года по 4 квартал 2022 года. Показатели выражены в миллиардах баккардийских крон 2015 года.

- a) Оцените следующую модель регрессии:

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 D2_t + \beta_4 D3_t + \beta_5 D4_t + \varepsilon_t,$$

где  $C_t$  — конечное потребление в момент времени  $t$ ,  $Y_t$  — конечное потребление в момент времени  $t$ ,  $Dj_t$  — дамми переменная на квартал ( $j = 2, 3, 4$ ). Проинтерпретируйте полученные результаты.

- b) На уровне значимости 5% проверьте гипотезу о наличии сезонности. Сформулируйте нулевую и альтернативную гипотезы.
- c) Оцените модель в следующем виде:

$$C_t = \beta_1 Y_t + \beta_2 D1_t + \beta_3 D2_t + \beta_4 D3_t + \beta_5 D4_t + \varepsilon_t.$$

Сравните полученные результаты с предыдущим пунктом.

- d) Попробуйте улучшить модель, включив в нее переменные взаимодействия:

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 D2_t + \beta_4 D3_t + \beta_5 D4_t + \beta_6 (Y * D2) + \beta_7 (Y * D3) + \beta_8 (Y * D4) + \varepsilon_t.$$

Проинтерпретируйте полученные результаты.