Семинар 2.
Модель множественной регрессии.
Тестирование гипотез.

1. Рассмотрим классическую линейную модель регрессии

$$y = X\beta + \varepsilon.$$

Найдите:

(a) $\mathrm{Cov}(\hat{\beta}, y)$;

(b) $\mathrm{Cov}(\hat{\beta}, \hat{y})$;

(c) $\mathrm{Cov}(e, y)$;

(d) $\mathrm{Cov}(e, \hat{y})$;

(e) $\mathrm{Cov}(e, \bar{y})$

Решение:

(a)

$$\hat{\beta} = (X^T X)^{-1} X^T y, \quad y = X\beta + \varepsilon.$$

$$\mathrm{Cov}(\hat{\beta}, y) = \mathrm{Cov}\big((X^T X)^{-1} X^T y,\ y\big).$$

Используем свойство: $\mathrm{Cov}(Ay, y) = A\,\mathrm{Cov}(y) = \sigma^2 A$ при A постоянной матрице.

$$A = (X^T X)^{-1} X^T.$$

$$\mathrm{Cov}(\hat{\beta}, y) = \sigma^2 (X^T X)^{-1} X^T.$$

(b)

$$\hat{y} = X\hat{\beta}.$$

$$\mathrm{Cov}(\hat{\beta}, \hat{y}) = \mathrm{Cov}\big(\hat{\beta}, X\hat{\beta}\big).$$

Так как X неслучайна:

$$\mathrm{Cov}(\hat{\beta}, X\hat{\beta}) = \mathrm{Cov}(\hat{\beta}) X^T.$$

Известно:

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}.$$

Тогда:

$$\mathrm{Cov}(\hat{\beta}, \hat{y}) = \sigma^2 (X^T X)^{-1} X^T.$$

Заметим, что (a) и (b) совпадают.

(c)

$$e = My, \quad M = I - P.$$

$$\operatorname{Cov}(e, y) = \operatorname{Cov}(My, y) = M \operatorname{Cov}(y) = \sigma^2 M.$$

Так как $M = I - P$ симметрична и идемпотентна.

(d)

$$\hat{y} = Py.$$

$$\operatorname{Cov}(e, \hat{y}) = \operatorname{Cov}(My, Py) = M \operatorname{Cov}(y) P^T = \sigma^2 MP.$$

Но $MP = (I - P)P = P - P^2 = 0$, так как $P^2 = P$. Следовательно:

$$\operatorname{Cov}(e, \hat{y}) = 0.$$

(e) Здесь $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{1}{n}\mathbf{1}^T y$, где $\mathbf{1}$ — вектор из единиц длины n.

$$\operatorname{Cov}(e, \bar{y}) = \operatorname{Cov}\left(My, \frac{1}{n}\mathbf{1}^T y\right) = M \operatorname{Cov}(y) \left(\frac{1}{n}\mathbf{1}\right) = \sigma^2 M \left(\frac{1}{n}\mathbf{1}\right) = \frac{\sigma^2}{n} M\mathbf{1}.$$

В общем случае $M\mathbf{1} \neq 0$, если вектор $\mathbf{1}$ не лежит в пространстве столбцов X (т.е. если в регрессии нет константы). Если же константа есть, то $\mathbf{1}$ — один из столбцов X, и тогда $P\mathbf{1} = \mathbf{1}$, значит $M\mathbf{1} = 0$, и ковариация нулевая. В общем случае:

$$\operatorname{Cov}(e, \bar{y}) = \frac{\sigma^2}{n}(I - P)\mathbf{1}.$$

2. Рассмотрим нормальную классическую линейную модель множественной регрессии $y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$ с неслучайными регрессорами. Дополнительно известно, что на самом деле $\beta_2 = ... = \beta_k = 0$.

   (a) Найдите $\mathbb{E}(R^2)$.

   (b) Найдите $\mathbb{E}(R^2_{adj.})$.

3. Ниже представлены результаты МНК–оценивания двух регрессий, часть из которых не сохранилась. Утерянные оценки коэффициентов регрессий заменены символами.

$$\text{Модель 1: } \hat{y}_i = \hat{\beta}_1 + \underset{(2)}{\hat{\beta}_2} x_i$$

$$\text{Модель 2: } \hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 x_i + \underset{(2)}{10} w_i, R^2 = 0.8$$

Оценивание проводилось по 103 наблюдениям. В скобках под оценками коэффициентов указаны их стандартные ошибки. Восстановите значение оценки коэффициента $\hat{\beta}_2$ первой регрессии?

4. Рассмотрим модель $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$. При оценке модели по 24 наблюдениям оказалось, что $RSS = 15$, $\sum(y_i - \bar{Y} - x_{i3} + \bar{x_3})^2 = 20$. На уровне значимости 1% протестируйте гипотезу

$$H_0 : \begin{cases} \beta_2 + \beta_3 + \beta_4 = 1 \\ \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}.$$

Решение:
Заметим, что в основной гипотезе есть линейно зависимые ограничения, оставим только линейно независимые:

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}$$

Ограниченная модель имеет вид:

$$y_i = \beta_1 + x_{i3} + \varepsilon_i$$

Переносим x_{3i} в левую часть, и получим оценку коэффициента β_1:

$$\hat{\beta}_1 = \bar{y} - \bar{x}_3$$

Теперь можно найти RSS_R:

$$RSS_R = \sum_{i=1}^{24}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{24}(y_i - \bar{y} + \bar{x}_3 - x_{3i})^2 = 20$$

Осталось найти значение F-статистики, которая при верной H_0 имеет распределение $F_{3,20}$:

$$F_{obs} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n - k_{UR})} = \frac{(20-15)/3}{15/(24-4)} = 20/9$$

Так как $F_{obs} < F_{3,20;0.99} = 4.94$, оснований отвергать нулевую гипотезу нет.

5. Рассмотрим модель $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, где

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}.$$

Случайные ошибки ε_i независимы и нормально распределены с $\mathbb{E}(\varepsilon) = 0$ и $\mathbb{V}\mathrm{ar}(\varepsilon) = \sigma^2 I$.

Для удобства расчётов даны матрицы: $X^T X$, $(X^T X)^{-1}$ и $X^T y$:

$$X^T X = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (X^T X)^{-1} = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 \\ 0 & -0.5 & 1.5 \end{pmatrix}, \quad X^T y = \begin{pmatrix} 15 \\ 12 \\ 5 \end{pmatrix}.$$

(a) Определите n и k.

(b) Вычислите МНК оценку вектора β.

(c) Найдите $\hat{\sigma}^2$, $\mathbb{E}(\hat{\sigma}^2)$.

(d) Найдите $\mathbb{V}\mathrm{ar}(\varepsilon_1)$, $\mathbb{V}\mathrm{ar}(\beta_1)$, $\mathbb{V}\mathrm{ar}(\hat{\beta}_1)$, $\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_1)$, $\mathbb{E}(\hat{\beta}_1^2) - \beta_1^2$;

(e) Найдите $\mathbb{C}\mathrm{ov}(\hat{\beta}_2, \hat{\beta}_3)$, $\widehat{\mathbb{C}\mathrm{ov}}(\hat{\beta}_2, \hat{\beta}_3)$, $\mathbb{V}\mathrm{ar}(\hat{\beta}_2 - \hat{\beta}_3)$, $\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_2 - \hat{\beta}_3)$;

(f) Найдите $\mathbb{V}\mathrm{ar}(\beta_2 - \beta_3)$, $\mathbb{C}\mathrm{orr}(\hat{\beta}_2, \hat{\beta}_3)$, $\widehat{\mathbb{C}\mathrm{orr}}(\hat{\beta}_2, \hat{\beta}_3)$.

Решение:

(a) Число наблюдений $n = 5$. Число регрессоров, включая свободный член равно $k = 3$.

(b) МНК-оценка вектора β равна $\hat{\beta} = (X^T X)^{-1} X^T y$. Тогда

$$\hat{\beta} = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 \\ 0 & -0.5 & 1.5 \end{pmatrix} \times \begin{pmatrix} 15 \\ 12 \\ 5 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 2 \\ 1.5 \end{pmatrix}$$

(c) Несмещенная оценка дисперсии случайной ошибки равна $\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{RSS}{5-3}$.
Вычислим RSS. Знаем, что $RSS = y^T (I - X(X^T X)^{-1} X^T) y = 1$.

Тогда $\hat{\sigma}^2 = 1/2$.

Так как по построению оценка $\hat{\sigma}^2$ несмещённая, то $\mathbb{E}(\hat{\sigma}^2 \mid X) = \sigma^2$.

(d)

$$\mathbb{V}\mathrm{ar}(\varepsilon_1) = \sigma^2$$

$$\mathbb{V}\mathrm{ar}(\beta_1) = 0,$$

$$\mathbb{V}\mathrm{ar}(\hat\beta_1) = \sigma^2 (X^T X)^{-1}_{(1,1)} = 0.5\sigma^2$$

$$\widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_1) = \hat\sigma^2 (X^T X)^{-1}_{(1,1)} = 0.5\hat\sigma^2 = 0.5\frac{1}{5-3} = 0.25$$

Так как оценки МНК являются несмещёнными, то $\mathbb{E}(\hat\beta) = \beta$, значит:

$$\mathbb{E}(\hat\beta_1) - \beta_1^2 = \mathbb{E}(\hat\beta_1) - (\mathbb{E}(\hat\beta_1))^2 = \widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_1) = 0.25$$

(e)

$$\mathbb{C}\mathrm{ov}(\hat\beta_2, \hat\beta_3) = \sigma^2 (X^T X)^{-1}_{(2,3)} = \sigma^2 \cdot \left(-\frac{1}{2}\right)$$

$$\widehat{\mathbb{C}\mathrm{ov}}(\hat\beta_2, \hat\beta_3) = \widehat{\mathbb{V}\mathrm{ar}}(\hat\beta)_{(2,3)} = \hat\sigma^2 (X^T X)^{-1}_{(2,3)} = \frac{1}{2} \cdot \left(-\frac{1}{2}\right) = -\frac{1}{4}$$

$$\mathbb{V}\mathrm{ar}(\hat\beta_2 - \hat\beta_3) = \mathbb{V}\mathrm{ar}(\hat\beta_2) + \mathbb{V}\mathrm{ar}(\hat\beta_3) - 2\,\mathbb{C}\mathrm{ov}(\hat\beta_2, \hat\beta_3) =$$

$$= \sigma^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} - 2(X'X)^{-1}_{(2,3)} = \sigma^2(1 + 1.5 - 2 \cdot (-0.5)) = 3.5\sigma^2$$

$$\widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_2 - \hat\beta_3) = \widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_2) + \widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_3) + 2\,\widehat{\mathbb{C}\mathrm{ov}}(\hat\beta_2, \hat\beta_3) =$$

$$= \hat\sigma^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)}) = \frac{1}{2} \cdot 3.5 = 1.75$$

(f)

$$\mathbb{V}\mathrm{ar}(\beta_2 - \beta_3) = 0$$

$$\mathbb{C}\mathrm{orr}(\hat\beta_2, \hat\beta_3) = \frac{\mathbb{C}\mathrm{ov}(\hat\beta_2, \hat\beta_3)}{\sqrt{\mathbb{V}\mathrm{ar}(\hat\beta_2)\,\mathbb{V}\mathrm{ar}(\hat\beta_3)}} = \frac{-0.5}{\sqrt{1 \cdot 1.5}} = -\frac{1}{\sqrt{6}}$$

$$\widehat{\mathbb{C}\mathrm{orr}}(\beta_2, \beta_3) = \frac{\widehat{\mathbb{C}\mathrm{ov}}(\hat\beta_2, \hat\beta_3)}{\sqrt{\widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_2)\,\widehat{\mathbb{V}\mathrm{ar}}(\hat\beta_3)}} = \frac{-0.5}{\sqrt{1 \cdot 1.5}} = -\frac{1}{\sqrt{6}}$$