Семинар 4.
Множественная регрессия.

1. Рассмотрим классическую линейную модель регрессии
$$y = X\beta + \varepsilon.$$
Найдите:
   (a) $\mathrm{Cov}(e, \hat{\beta})$;
   (b) $\mathrm{Cov}(e, y)$;
   (c) $\mathrm{Cov}(e, \hat{y})$.

2. Пусть регрессионная модель $Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$, $i = 1, \ldots, n$, задана в матричном виде при помощи уравнения $y = X\beta + \varepsilon$, где $\beta = \begin{pmatrix}\beta_1 & \beta_2 & \beta_3\end{pmatrix}'$. Известно, что $\varepsilon \sim N(0, \sigma_\varepsilon^2 \cdot \mathrm{I}_n)$. Известно также, что:
$$y = \begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}, X = \begin{pmatrix}1 & 0 & 0\\1 & 0 & 0\\1 & 0 & 0\\1 & 1 & 0\\1 & 1 & 1\end{pmatrix}.$$
Для удобства расчётов ниже приведены матрицы:
$$X'X = \begin{pmatrix}5 & 2 & 1\\2 & 2 & 1\\1 & 1 & 1\end{pmatrix} \text{ и } (X'X)^{-1} = \begin{pmatrix}1/3 & -1/3 & 0\\-1/3 & 4/3 & -1\\0 & -1 & 2\end{pmatrix}.$$
   (a) Рассчитайте при помощи метода наименьших квадратов оценку для вектора неизвестных коэффициентов.
   (b) Рассчитайте несмещенную оценку для неизвестного параметра σ_ε^2 регрессионной модели.
   (c) Рассчитайте $\widehat{\mathrm{Var}}(\hat{\beta})$, оценку для ковариационной матрицы МНК-оценки $\widehat{\beta}$ вектора коэффициентов β.
   (d) Рассчитайте TSS, RSS и ESS.
   (e) Сформулируйте основную и альтернативную гипотезы, которые соответствуют тесту на значимость переменной X_2 в уравнении регрессии.
   (f) Протестируйте на значимость переменную X_2 в уравнении регрессии на уровне значимости 10%:

i. Приведите формулу для тестовой статистики.

ii. Укажите распределение тестовой статистики при верной H_0.

iii. Вычислите наблюдаемое значение тестовой статистики.

iv. Укажите границы области, где основная гипотеза не отвергается.

v. Сделайте статистический вывод о значимости переменной X_2.

(g) Найдите $p-value$, соответствующее наблюдаемому значению тестовой статистики (t_{obs}) из предыдущего пункта. На основе полученного p-value сделайте вывод о значимости переменной X_2.

(h) Постройте 90%-ый доверительный интервал для оценки коэфициента β_2.

3. По n наблюдениям Вася оценил исходную модель:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

Для надежности Вася стандартизировал переменные, т.е. перешёл к $y_i^* = (y_i - \bar{y})/\hat{\sigma}_y$ и $x_i^* = (x_i - \bar{x})/\hat{\sigma}_x$. Затем Вася оценил ещё две модели:

$$y_i^* = \beta_1' + \beta_2' x_i^* + \varepsilon_i'$$

и

$$y_i^* = \beta_2'' x_i^* + \varepsilon_i''.$$

В решении можно считать $\hat{\sigma}_x$ и $\hat{\sigma}_y$ известными.

(a) Найдите $\hat{\beta}_1'$.

(b) Как связаны между собой $\hat{\beta}_2$, $\hat{\beta}_2'$ и $\hat{\beta}_2''$?

(c) Как связаны между собой e_i, e_i' и e_i''?

(d) Как связаны между собой $\widehat{\text{Var}}\left(\hat{\beta}_2\right)$, $\widehat{\text{Var}}\left(\hat{\beta}_2'\right)$ и $\widehat{\text{Var}}\left(\hat{\beta}_2''\right)$?

(e) Как выглядит матрица $\widehat{\text{Var}}\left(\hat{\beta}'\right)$?

(f) Как связаны между собой t-статистики $t_{\hat{\beta}_2}$, $t_{\hat{\beta}_2'}$ и $t_{\hat{\beta}_2''}$?

(g) Как связаны между собой R^2, $R^{2\prime}$ и $R^{2\prime\prime}$?

(h) В нескольких предложениях прокомментируйте последствия перехода к стандартизированным переменным.