Подборка задач №2.

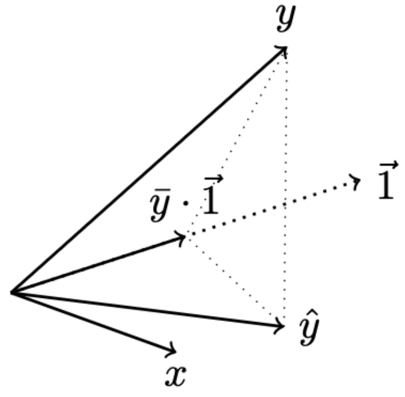
Вопросы по лекции.

1. Геометрическая интерпретация МНК для множественной регрессии.
2. Сформулируйте предпосылки классической линейной модели множественной регрессии.
3. Сформулируйте теорему Гаусса–Маркова для модели множественной регрессии.

Задачи.

Примечание: Регрессор x во всех задачах блоках предполагается детерминированным.

1. Рассмотрим модель регрессии $y_i = \beta x_i + \varepsilon_i$, $\mathbb{E}(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2, \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$. При каких значениях параметров c_i несмещённая оценка
$$\hat{\beta} = \sum_{i=1}^{n} c_i y_i / \sum_{i=1}^{n} c_i x_i$$
имеет наименьшую дисперсию?
2. Рассмотрим модель $y_i = \mu + \varepsilon_i$, где $\mathbb{E}(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$ и $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$. При каких значениях параметров c_i несмещённая оценка $\hat{\mu} = \sum_{i=1}^{n} c_i y_i$ имеет наименьшую дисперсию?
3. Найдите на картинке все перпендикулярные векторы. Найдите на картинке все прямоугольные треугольники. Сформулируйте для них теоремы Пифагора.



4. Покажите на картинке из предыдущего задания TSS, ESS, RSS, R^2, $\widehat{Corr}(\hat{y}, y)$, $\widehat{Cov}(\hat{y}, y)$.

5. Покажите, что матрицы-проекторы $P = X(X^TX)^{-1}X^T)$ и $M = I_n - P$ симметричные и идемпотентые.

6. Рассмотрим матрицы-проекторы $P = X(X^TX)^{-1}X^T$ и $\pi = \vec{i}(\vec{i}^T\vec{i})^{-1}\vec{i}^T$, где $\vec{i}$ — вектор размерности $n \times 1$, состоящий из единиц. Будет ли матрица $(P - \pi)$ идемпотентной? Рассмотрите все случаи.

7. Выведите несмещённую оценку для дисперсии случайной ошибки в модели множественной регрессии, используя матричное представление.

8. 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.

9. Используя матричное представление, вычислите $\mathrm{Cov}(e, \hat{\beta})$, $\mathrm{Cov}(\hat{\beta}, y)$, $\mathrm{Cov}(\hat{\beta}, \hat{y})$, $\mathrm{Cov}(e, y)$, $\mathrm{Cov}(e, \hat{y})$.

10. Эконометрист Илья оценил методом наименьших квадратов модель 1, $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, а затем модель 2, $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \varepsilon_i$. Сравните полученные ESS, RSS, TSS и R^2.

11. Рассмотрим оценку вида $\widetilde{\beta} = \left(X'X + rD\right)^{-1} X'y$ (ридж-регрессия (ridge regression)) для вектора коэффициентов регрессионного уравнения $y = X\beta + \varepsilon$, где D — диагональная $k \times k$ матрица, состоящая из диагональных элементов матрицы $X'X$.

    (a) Найдите математическое ожидание, матрицу ковариаций и матрицу среднеквадратичных отклонений оценки $\widetilde{\beta}$ $\left(\mathrm{MSE}(\hat{\beta}) = \mathbb{E}\left(\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\right)\right)\right.$.

    (b) Покажите, что существует $r > 0$ такое, что $\mathrm{Var}(\widetilde{\beta}) < \mathrm{Var}(\widehat{\beta})$, где $\widehat{\beta}$ — оценка метода наименьших квадратов.

    (c) Можно ли найти такое $r > 0$, что оценка $\widetilde{\beta}$ более эффективна, чем оценка метода наименьших квадратов $\widehat{\beta}$ (т.е. для всех $j = 1, \ldots, k$, $\mathrm{MSE}\left(\widetilde{\beta}_j\right) < \mathrm{MSE}\left(\widehat{\beta}_j\right)\big)$?

Список использованных источников

1. Катышев П.К., Магнус Я.Р., Пересецкий А.А., Головань С.В. Сборник задач к начальному курсу эконометрики: Учеб. пособие. — 4-е изд., перераб. и доп. — М.: Дело, 2007. — 368 с.