

Подборка задач №5.

Мультиколлинеарность и гетероскедастичность.

1. Какая(ие) из предпосылок ТГМ нарушаются при наличии строгой мультиколлинеарности в данных?
2. Будут ли МНК-оценки несмещенными, эффективными при наличии квази-мультиколлинеарности? Объясните свой ответ.
3. Какие последствия мультиколлинеарности (нестрогой) для оценивания модели регрессии с помощью МНК Вам известны?
4. Запишите оптимизационную задачу построения первой главной компоненты и решите её.
5. Дайте определение матрицы факторной нагрузки в методе главных компонент. Чем является элемент a_{ij} ?
6. Теоретическая регрессионная зависимость и выборочная корреляционная матрица стандартизированных регрессоров X имеют вид:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i,$$

где

- y — оценка студента за курс по 100-балльной шкале;
- x_1 — количество часов самостоятельной подготовки в неделю
- x_2 — посещаемость лекций (в процентах)
- x_3 — уровень базовых знаний перед началом курса (тестовый балл)

$$\hat{R}(X) = \begin{pmatrix} 1 & 0.8 & 0.1 \\ 0.8 & 1 & 0.2 \\ 0.1 & 0.2 & 1 \end{pmatrix}.$$

- (a) Вычислите все главные компоненты. Сколько главных компонент надо выбрать, чтобы они объясняли не менее 75% общей дисперсии?
 - (b) Вычислите матрицу факторной нагрузки. Проинтерпретируйте полученные результаты.
 - (c) Дайте содержательную интерпретацию построенным главным компонентам.
7. При нарушении предпосылки о гомоскедастичности ошибок будут ли

- (a) несмещенными МНК-оценки коэффициентов модели;
- (b) эффективными МНК-оценки коэффициентов модели;
- (c) несмещенными дисперсии оценок коэффициентов?

В каждом пункте обоснуйте математически свой ответ.

8. Для множественной модели регрессии $y = X\beta + \varepsilon$ с гетероскедастичными независимыми ошибками $\varepsilon \sim (0, \Omega)$ запишите и решите оптимизационную задачу метода ОМНК (GLS). Является ли GLS-оценка несмещенной и эффективной? Обоснуйте свой ответ математически.
9. В чем заключается отличие FGLS-оценки от GLS-оценки?
10. Рассмотрим модель множественной регрессии вида

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

где X — $n \times k$ детерминированная матрица с рангом равным k , $\mathbb{E}(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \Omega$. Обозначим через $\hat{\beta}_{OLS}$ и $\hat{\beta}_{GLS}$ МНК-оценку и ОМНК-оценку для вектора параметров β соответственно. Покажите, что если столбцы матрицы X являются собственными векторами ковариационной матрицы Ω , то

- (a) аналитические выражения для $\hat{\beta}_{OLS}$ и $\hat{\beta}_{GLS}$ совпадут;
 - (b) аналитические выражения для ковариационных матриц оценок $\hat{\beta}_{OLS}$ и $\hat{\beta}_{GLS}$ также совпадут.
11. Пусть задана модель $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, где ε_i — независимые случайные величины с $\mathbb{E}(\varepsilon_i) = 0$ и $\text{Var}(\varepsilon_i) = \sigma^2$. Найдите наиболее эффективную оценку неизвестного параметра β_2 в классе линейных по y_i и несмещенных оценок.
 12. Рассмотрим линейную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ по 50 наблюдениям. При оценивании с помощью МНК были получены результаты:

$$\hat{\beta}_1 = 1.21, \quad \hat{\beta}_2 = 1.11, \quad \hat{\beta}_3 = 3.15, \quad R^2 = 0.72.$$

Оценена также вспомогательная регрессия:

$$e_i^2 = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i + u_i.$$

Результаты оценивания следующие:

$$\hat{\delta}_1 = 1.50, \quad \hat{\delta}_2 = -2.18, \quad \hat{\delta}_3 = 0.23, \quad \hat{\delta}_4 = 1.87, \quad \hat{\delta}_5 = -0.56, \quad \hat{\delta}_6 = -0.09, \quad R_{aux}^2 = 0.36.$$

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. С помощью теста Уайта протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

13. Известно, что после деления каждого уравнения регрессии $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ на x_i гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\text{Var}(\varepsilon_i)$?