Подборка задач №4.

Фиктивные переменные. Тест Чоу. Безусловное прогнозирование.

Во всех задачах предполагается, что предпосылки теоремы Гаусса-Маркова выполнены, а ошибки имеют нормальное распределение.

1. (из problem set 3) Рассмотрим регрессию

$$y_i = \beta_1 + \beta_2 d_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

где d — некоторая фиктивная переменная. Пусть $\overline{y}_0$ — среднее значение переменной y по n_0 наблюдениям, для которых $d = 0$, и $\overline{y}_1$ — среднее значение переменной y по n_1 наблюдениям, для которых $d = 1$ ($n_0 + n_1 = n$). Найдите $\mathbb{V}\mathrm{ar}(\hat{\beta}_1)$, $\mathbb{V}\mathrm{ar}(\hat{\beta}_2)$.

2. (из problem set 3) На основе квартальных данных с 1971 по 1976 г. с помощью метода наименьших квадратов получено следующее уравнение:

$$y_i = \underset{(2.14)}{1.12} - \underset{(0.0034)}{0.0098}x_{i1} - \underset{(3.42)}{5.62}x_{i2} + \underset{(0.009)}{0.044}x_{i3}$$

где в скобках указаны стандартные ошибки, $ESS = 110.32$, $RSS = 21.43$.

   (a) Проверьте значимость каждого из коэффициентов.

   (b) Найдите коэффициент детерминации.

   (c) Протестируйте значимость регрессии в целом.

   (d) Когда в уравнение были добавлены три фиктивные переменные, соответствующие трем первым кварталам года, величина ESS выросла до 118.20. Проверьте гипотезу о наличии сезонности, сформулировав необходимые предположения о виде этой сезонности.

   (e) Для той же исходной модели были раздельно проведены две регрессии на основе данных: 1-й квартал 1971 г. — 1-й квартал 1975 г. и 2-й квартал 1975 г. — 4-й квартал 1976 г. соответственно и получены следующие значения сумм квадратов остатков:

$$RSS_1 = 12.25, \quad RSS_2 = 2.32.$$

   Проверьте гипотезу о том, что между 1-м и 2-м кварталами 1975 г. произошло структурное изменение.

3. (из problem set 3) Рассмотрим следующую функцию спроса с сезонными переменными $SPRING$ (весна), $SUMMER$ (лето), $FALL$ (осень):

$$\ln Q_i = \beta_1 + \beta_2 \cdot \ln P_i + \beta_3 \cdot SPRING_i + \beta_4 \cdot SUMMER_i + \beta_5 \cdot FALL_i + \varepsilon_i.$$

Данная функция спроса была оценена с помощью МНК по 20 наблюдениям. Известно, что $R^2 = 0.37$.

(a) Напишите спецификацию регрессии с ограничениями для проверки статистической гипотезы $H_0 : \beta_3 = \beta_5$. Дайте интерпретацию проверяемой гипотезе.

(b) Пусть для регрессии с ограничениями (restricted model) был вычислен коэффициент $R_R^2 = 0.23$. На уровне значимости 5% проверьте нулевую гипотезу.

4. (из problem set 3) Для города и для деревни рассматриваются две модели парной регресии. Двадцать наблюдений для города дали следующие результаты:

$$X^T X = \begin{pmatrix} 20 & 20 \\ 20 & 25 \end{pmatrix},\ X^T y = \begin{pmatrix} 10 \\ 20 \end{pmatrix},\ y^T y = 30,$$

а десять наблюдений для деревни –

$$X^T X = \begin{pmatrix} 10 & 10 \\ 10 & 20 \end{pmatrix},\ X^T y = \begin{pmatrix} 8 \\ 20 \end{pmatrix},\ y^T y = 24$$

На 5%-ом уровне значимости проверьте гипотезу о том, что эти две модели совпадают.

5. Рассмотрим регрессионную модель $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \ldots, n$. Предположим, что параметр β_2 известен. Предложите способ прогноза величины y_{n+1} (при заданном x_{n+1}) и найдите дисперсию ошибки прогноза.

6. Рассмотрим модель $y = X\beta + \varepsilon$, где $\mathbb{E}(\varepsilon) = 0$ и $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2 I$.

(a) Вспомните или найдите $\mathrm{Var}(\hat{y})$ и $\mathrm{Var}(y)$.

(b) В каких пределах может лежать произвольный элемент матрицы-шляпницы $P = X(X^T X)^{-1} X^T$?

(c) Вениамин оценил множественную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$. А после этого его попросили построить прогноз y для $x = 1$ и $z = 2$. Вениамин неожиданно обнаружил, что в его данных в 42-м наблюдении как раз $x_{42} = 1$, а $z_{42} = 2$. Что ему лучше взять в качестве прогноза, y_{42} или $\hat{y}_{42}$?

7. Эконометресса Ксения строит регрессию y на X, а эконометрист Иван использует QR-разложение для матрицы X, то есть представляет X в виде $X = QR$.

Другими словами, Иван переходит к ортонормированному базису в пространстве столбцов матрицы $X : Q^T Q = I$, R – верхнетреугольная обратимая матрица. Затем Иван строит регрессию y на Q.

Верно ли, что Ксения и Иван получат одинаковые

(a) оценки коэффициентов?

(b) прогнозы $\hat{y}$?

(c) ковариационные матрицы прогнозов $\mathrm{Var}(\hat{y})$?

8. Всегда ли доверительный интервал для суммы коэффициентов регрессии $\beta_1+\beta_2$ шире каждого доверительного интервала для β_1 и β_2? Если да, то почему?