

**Фиктивные переменные. Критерий Чоу.**

1. Имеются ежеквартальные данные  $Y_t$  о расходах потребителей на газ и электричество в США (в млрд \$ в сопост. ценах 1972 года) за 1977-1982 годы

| Квартал\Год | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
|-------------|------|------|------|------|------|------|
| 1           | 7,33 | 7,65 | 7,96 | 7,74 | 8,04 | 8,26 |
| 2           | 4,7  | 4,92 | 5,01 | 5,1  | 5,27 | 5,51 |
| 3           | 5,1  | 5,15 | 5,05 | 5,67 | 5,51 | 5,41 |
| 4           | 5,46 | 5,55 | 5,59 | 5,92 | 6,04 | 5,83 |

Предполагается, что функция расходов на газ и электричество зависит от номера квартала. После введения фиктивных бинарных переменных  $D_2, D_3$  и  $D_4$ , соответствующих второму, третьему и четвёртому кварталам, были построены МНК-оценки параметров следующей регрессии

$$Y_t = a + bt + \delta_2 D_{t2} + \delta_3 D_{t3} + \delta_4 D_{t4} + \varepsilon_t$$

В таблице приведены оценки параметров и с.к.о. оценок

| параметр | $a$  | $b$   | $\delta_2$ | $\delta_3$ | $\delta_4$ |
|----------|------|-------|------------|------------|------------|
| оценка   | 7,5  | 0,03  | -2,78      | -2,58      | -2,19      |
| ско      | 0,09 | 0,005 | 0,09       | 0,1        | 0,1        |

- 1) Определите фиктивные переменные  $D_2, D_3$  и  $D_4$ ;
  - 2) Верно ли, что регрессионные зависимости для разных кварталов различны?
  - 3) Запишите регрессионные уравнения для  $Y_t$ , соответствующие первому, второму, третьему и четвёртому кварталам.
2. Имеются данные о 964 родах. Пусть  $Y$  – вес новорожденного,  $X$  – количество сигарет, которое будущая мать выкуривала в день. На основании этих данных по МНК была оценена регрессия

$$\hat{Y} = 3418 - 7,2X$$

Ско построенных оценок равны 14 и 2,1 соответственно.

$$R^2 = 0,012 \quad SS_{\text{случ}} = 158,6 \cdot 10^6.$$

Затем рассмотрели разбиение выборки на две подвыборки – матерей, рожаящих первенца (584 женщины), и матерей, рожаящих не первенца (380 женщин). Для каждой из этих подгрупп были построены линейные регрессии. Для первой подгруппы:

$$\hat{Y} = 3363 - 4,0X$$

Ско построенных оценок равны 18 и 2,8 соответственно;  $SS_{\text{случ}} = 91,2 \cdot 10^6$ .

Для второй подгруппы:

$$\hat{Y} = 3506 - 12,1X$$

Ско построенных оценок равны 23 и 3,1 соответственно;  $SS_{\text{случ}} = 63,5 \cdot 10^6$ .

В рамках этой задачи выясните:

- 1) Является ли первая построенная регрессия значимой?
- 2) Будет ли рассматриваемая регрессионная зависимость одинаковой для первой и второй подгрупп?
- 3) Следует ли объединять первую и вторую подвыборки в объединённую выборку?

Являются ли значимыми коэффициенты при X в построенных уравнениях регрессии?  
Сделайте выводы

### Домашнее задание

1. Пусть X – количество решённых (из 10 предложенных) задач на вступительном экзамене, а Y – количество решённых (из 7 предложенных) задач на экзамене по линейной алгебре на первом курсе. Для 12 выбранных студентов результаты представлены в таблице

|     |    |   |   |   |   |   |   |   |   |   |   |   |
|-----|----|---|---|---|---|---|---|---|---|---|---|---|
| X   | 10 | 6 | 8 | 8 | 6 | 7 | 6 | 7 | 9 | 6 | 5 | 7 |
| Y   | 6  | 4 | 4 | 5 | 4 | 7 | 3 | 4 | 7 | 3 | 2 | 3 |
| пол | м  | ж | м | ж | ж | м | ж | м | м | ж | м | ж |

- 1) Можно ли считать, что регрессионная модель

$$Y_i = a + bX_i + \varepsilon_i, i = 1, \dots, n,$$

одинакова для мужчин и женщин?

- 2) Вычислите нескорректированные и скорректированные коэффициенты детерминации для указанной выше модели и для модели с фиктивной переменной.

Сделайте выводы