

Семинар 5.

Множественная регрессия.

Стандартизированные данные. Тестирование гипотез.

1. 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 $\widehat{Var}(\hat{\beta}_2)$, $\widehat{Var}(\hat{\beta}'_2)$ и $\widehat{Var}(\hat{\beta}''_2)$?

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

(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 , $R^{2'}$ и $R^{2''}$?

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

2. Рассмотрим модель $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$. При оценке модели по 24 наблюдениям оказалось, что $RSS = 15$, $\sum (y_i - \bar{y} - x_{i3} + \bar{x}_3)^2 = 20$. На уровне значимости 1% протестируйте гипотезу

$$H_0 : \begin{cases} \beta_2 + \beta_3 + \beta_4 = 1 \\ \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}.$$

Решение:

Заметим, что в основной гипотезе есть линейно зависимые ограничения, оставим только линейно независимые:

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}$$

Ограниченная модель имеет вид:

$$y_i = \beta_1 + x_{i3} + \varepsilon_i$$

Переносим x_{3i} в левую часть, и получим оценку коэффициента β_1 :

$$\hat{\beta}_1 = \bar{y} - \bar{x}_3$$

Теперь можно найти RSS_R :

$$RSS_R = \sum_{i=1}^{24} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{24} (y_i - \bar{y} + \bar{x}_3 - x_{3i})^2 = 20$$

Осталось найти значение F-статистики, которая при верной H_0 имеет распределение $F_{3,20}$:

$$F_{obs} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n - k_{UR})} = \frac{(20 - 15)/3}{15/(24 - 4)} = 20/9$$

Так как $F_{obs} < F_{3,20;0.99} = 4.94$, оснований отвергать нулевую гипотезу нет.