

## Консультация 20.10.2023

**Задача 1.** Модель порождающая данные имеет вид  $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ . Ошибки независимы и их дисперсия равна  $Var(\varepsilon_i) = \sigma^2$ . Регрессоры детерминированы. Для оценки дисперсии  $\sigma^2$  используется формула  $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$ . Является ли  $\hat{\sigma}^2$  несмещенной оценкой  $\sigma^2$ ? Если оценка смещена, то что можно сказать о знаке смещения?

**Задача 2.** Рассмотрим регрессию вида

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

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

**Задача 3.** Вместо того чтобы оценивать параметры  $\beta_1, \beta_2$  в модели

$$y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon,$$

( $X_1, X_2$  —  $n \times k_1$ ,  $n \times k_2$  матрицы соответственно,  $\beta_1, \beta_2$  — векторы размерности  $k_1, k_2$  соответственно), строятся МНК-оценки этих параметров, исходя из модели

$$y = X_1^* \beta_1 + X_2 \beta_2 + \varepsilon^*,$$

где  $X_1^*$  — матрица остатков, полученных в результате регрессии каждого столбца матрицы  $X_1$  на  $X_2$ .

- (а) Покажите, что полученная таким образом оценка вектора  $\beta_2$  совпадает с оценкой, полученной в результате регрессии  $y$  только на  $X_2$ .
- (б) Найдите смещение оценки вектора  $\beta_2$ .

**Задача 4.** Ниже представлены результаты МНК-оценивания двух регрессий, часть из которых не сохранилась. Утерянные оценки коэффициентов регрессий заменены символами.

$$\text{Модель 1: } \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

(2)

$$\text{Модель 2: } \hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 x_i + 10w_i, R^2 = 0.8$$

(2)

Оценивание проводилось по 103 наблюдениям. В скобках под оценками коэффициентов указаны их стандартные ошибки. Восстановите значение оценки коэффициента  $\hat{\beta}_2$  первой регрессии?

Зависимая переменная: логарифм среднего дохода $\ln(AHE)$			
Регрессор	(1)	(2)	(3)
College	1,05 (0,16)	1,00 (0,10)	0,96 (0,11)
Age		0,04 (0,01)	0,05 (0,01)
Northeast			0,04 (0,01)
Midwest			0,06 (0,02)
South			-0,03 (0,01)
Константа	2,96 (0,03)	0,82 (0,16)	0,83 (0,15)
$R^2$	0,12	0,58	0,60
Число наблюдений	1000	1000	1000

**Задача 5.** Вопросы этого задания связаны с анализом результатов оценивания трех моделей регрессии, представленных в таблице 1:

- AHE — средний доход (долл. в час),
  - College — переменная, равная единице, если работник закончил колледж, и равная нулю в противном случае,
  - Age — возраст в годах,
  - Northeast — переменная, равная единице, если работник из северо-восточного региона, и равная нулю в противном случае,
  - Midwest — переменная, равная единице, если работник со Среднего Запада, и равная нулю в противном случае,
  - South — переменная, равная единице, если работник из южного региона, и равная нулю в противном случае,
  - West — переменная, равная единице, если работник из западного региона, и равная нулю в противном случае (в выборке есть работники только из 4 перечисленных регионов).
- (a) Существуют ли статистически значимые региональные особенности на рассматриваемом рынке труда?
- (b) Существуют ли статистически значимые региональные особенности на рассматриваемом рынке труда? На основе результатов расчетов из предыдущего пункта проверьте соответствующую гипотезу при уровне значимости 5% и ответьте на вопрос.
- (c) Кто получает более высокую заработную плату при прочих равных условиях: работники из северо-восточного или из южного регионов? На сколько процентов?
- (d) В соответствии с результатами модели №2 на сколько процентов зарплата выпускника колледжа выше по сравнению с работником такого же возраста, который не заканчивал колледж?

**Задача 6.** Рассматривается уравнение  $y_i = \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \varepsilon_i$ , для которого выполнены все предпосылки классической линейной модели множественной регрессии. Параметры уравнения оцениваются при помощи МНК. Имеются следующие данные о 1000 наблюдениях соответствующих переменных:

$$\sum_{i=1}^{1000} x_i^{(1)} x_i^{(2)} = \sum_{i=1}^{1000} \left(x_i^{(2)}\right)^2 = 1, \sum_{i=1}^{1000} \left(x_i^{(1)}\right)^2 = 3,$$

$$\sum_{i=1}^{1000} x_i^{(1)} y_i = 30, \sum_{i=1}^{1000} x_i^{(2)} y_i = 20.$$

- (а) Вычислите оценку разности коэффициентов  $\beta_1 - \beta_2$ . Ответ в виде числа.
- (б) Пусть также известно, что сумма квадратов остатков в оцененной регрессии оказалась равна 5988. Постройте 99-процентный доверительный интервал для разности  $\beta_1 - \beta_2$ .

### Список источников

1. Универсиада по эконометрике — 2018. <https://new.universiade-ecm.com>.