

Самостійна робота 9.3. (завдання та алгоритми)

На основі даних умови попередньої самостійної роботи 9.1 перевірити статистичну якість моделі.

1. Перевірити економетричну модель на адекватність за F -критерієм Фішера.

2. Перевірити статистичну значущість коефіцієнта кореляції та визначити його надійний інтервал.

3. Перевірити статистичну значущість оцінок параметрів моделі та визначити їх надійні інтервали.

Розв'язання

$$\hat{y} = -1,8994 + 1,2115 \cdot x$$

1. Перевірка адекватності економетричної моделі статистичним даним за F -критерієм Фішера.

1. Формулюють і розглядають гіпотези:

$H_0 : a_1 = 0$ – оцінка параметра незначуща;

$H_1 : a_1 \neq 0$ – оцінка параметра значуща.

2. Знаходимо розрахункове (фактичне) значення F -критерію Фішера за формулою:

$$F_p = \frac{MSE}{MSR} = \frac{n-m}{m-1} \cdot \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9.52)$$

3. Задаємо рівень значущості α та обчислюємо ступені свободи $k_1 = m - 1$, $k_2 = n - m$.

4. За таблицею розподілу функції Фішера з урахуванням значень α , k_1 , k_2 знаходимо табличне значення F_{α, k_1, k_2} (див. Додаток В в лекції).

5. Отримане розрахункове значення порівнюється з табличним. Перевіряємо виконання умови $F_p > F_{\alpha, k_1, k_2}$:

– якщо умова виконується, то вважається, що економетрична модель є адекватною з ймовірністю $p = 1 - \alpha$ існуючому зв'язку між ознаками x , y (лінійна форма зв'язку підтверджується);

– якщо умова не виконується, то з ймовірністю p розглянуту парну лінійну регресію не можна вважати адекватною (лінійна форма зв'язку не підтверджується).

2. Перевірка статистичної значущості коефіцієнта кореляції та визначення його надійного інтервалу.

В теорії перевірки гіпотез існує два взаємно доповнюючих один одного підходи:

- 1) перевірка значимості (значущості);
- 2) довірчі (надійні) інтервали.

1) Перевірка значимості.

Послідовність перевірки коефіцієнта кореляції Пірсона r_{xy} на значущість за t -критерієм Стьюдента:

1. Формулюють і розглядають гіпотези:

$H_0: r_{xy} = 0$ між змінними y і x немає зв'язку;

$H_1: r_{xy} \neq 0$ між змінними y і x є суттєвий зв'язок.

2. Обчислення фактичного (розрахункового, емпіричного) значення t -критерію за формулою:

$$t_{\text{розрах}} = \frac{|r_{xy}|}{\sigma_r} = r_{xy} \cdot \sqrt{\frac{n-m}{1-r_{xy}^2}} = r_{xy} \sqrt{\frac{k}{1-r_{xy}^2}} \quad (9.38)$$

3. Встановлення критичних (теоретичних, табличних) значень t -критерію з урахуванням рівня значущості α і ступеня свободи $k=n-m$ за таблицею t -розподілу (де n - обсяг вибіркової сукупності або кількість спостережень, m - кількість параметрів моделі).

Таблицю значень t -статистики наведено в *Додатку Б* до лекції - Тема 9.

4. Перевірка умови $|t_{\text{розрах}}| > t_{\alpha,k}$

- якщо умова виконується, то гіпотеза H_0 про нульове значення коефіцієнта кореляції у генеральній сукупності на рівні значущості α не підтверджується, відповідно коефіцієнт кореляції з ймовірністю $p=1-\alpha$ є статистично значущим. Тобто в генеральній сукупності між змінними y та x існує суттєвий лінійний зв'язок.

- якщо умова не виконується, то нульова гіпотеза H_0 на рівні значущості α приймається, коефіцієнт кореляції є статистично незначущим. Отже, в генеральній сукупності між змінними y та x не існує лінійного зв'язку.

2) Надійні інтервали.

Визначення надійних інтервалів для коефіцієнта парної кореляції (кореляційного відношення).

Для цього необхідно визначити граничну помилку коефіцієнту кореляції Δr .

$$\Delta r = \pm t_{\alpha, k} \cdot \sigma_r \quad (9.39)$$

де $t_{\alpha, k}$ – нормовані (табличні) значення t -критерію Стьюдента, визначені з рівнем значущості α та числом ступенів вільності k ;

σ_r – дисперсія коефіцієнта парної кореляції, що визначається за формулою:

$$\sigma_r = \frac{1 - r^2}{\sqrt{n - m}} = \frac{1 - r^2}{\sqrt{k}} \quad (9.40)$$

Отже, довірчі межі коефіцієнта парної кореляції (кореляційного відношення) визначаються через оцінку дисперсії σ_r .

Таким чином, з довірчою ймовірністю $p=1-\alpha$ оцінка показника тісноти зв'язку ρ для генеральної сукупності буде знаходитись у межах:

$$\rho = r \pm t_{\alpha, k} \cdot \sigma_r \quad (9.41 \text{ A})$$

або

$$r - t_{\alpha, k} \cdot \sigma_r \leq \rho \leq r + t_{\alpha, k} \cdot \sigma_r \quad (9.41 \text{ Б})$$

3. Перевірка статистичної значущості оцінок параметрів моделі та визначення їх надійних інтервалів.

1) Перевірка значимості (значущості).

Послідовність тестування за t-критерієм Стьюдента:

1. Формулюють і розглядають гіпотези:

$H_0 : a_j=0$ ($j=0, 1$) – оцінка параметра незначуща;

$H_1 : a_j \neq 0$ ($j=0, 1$) – оцінка параметра значуща.

2. Обчислюють розрахункові значення t -критерію Стьюдента для параметрів a_0 та a_1 за формулами:

$$t_{a_0} = \frac{|a_0|}{\sigma_{a_0}}; \quad t_{a_1} = \frac{|a_1|}{\sigma_{a_1}}. \quad (9.42)$$

де σ_{a_0} – це стандартна помилка оцінки параметру a_0

$$\sigma_{a_0} = \sigma_u \cdot \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (9.43)$$

σ_{a_1} – це стандартна помилка оцінки параметру a_1

$$\sigma_{a_i} = \frac{\sigma_u}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (9.44)$$

σ_u – корінь квадратний з дисперсії залишків (середньоквадратичне відхилення, середня похибка), що визначається за формулою:

$$\sigma_u = \sqrt{\frac{\sum_{i=1}^n u_i^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}}. \quad (9.45)$$

3. Встановлення критичних (табличних) значень t -критерію з урахуванням рівня значущості α і ступеня свободи $n-m$ за таблицею t -розподілу.

4. Перевірка умови $t_{a0}, t_{a1} \geq t_{\alpha, k}$

- якщо умова виконується, то t_{a0}, t_{a1} потрапляє у критичну область, нуль-гіпотеза відкидається і параметри a_0 та a_1 генеральної сукупності значущі;

- якщо умова не виконується, то t_{a0}, t_{a1} потрапляє в область значень, які відкидають, нуль-гіпотеза приймається, відповідні параметри є статистично незначущими.

2) Надійні інтервали.

Визначення інтервалів надійності для оцінок параметрів моделі за t -розподілом при рівні значущості 0,05.

Оцінка значущості довірчих границь параметрів a_0 та a_1 лінійної (лінеалізованої) функції $y = a_0 + a_1 \times x$. Спочатку знаходять граничні помилки коефіцієнтів регресії:

$$\Delta a_i = \pm t_{\alpha, k} \cdot \sigma_{a_i} \quad (9.46)$$

де $t_{\alpha, k}$ – нормовані (табличні) значення t -критерію Стюдента, визначені з рівнем значущості α та числом ступенів вільності k ;

σ_{a_i} – визначається за формулами (9.43) та (9.44).

Тоді, межі довірчих інтервалів становлять:

$$\tilde{a}_0 = a_0 \pm t_{\alpha, k} \cdot \sigma_{a_0}; \quad \tilde{a}_1 = a_1 \pm t_{\alpha, k} \cdot \sigma_{a_1} \quad (9.47)$$

де $t_{\alpha, k}$ – табличне значення t – критерію Стюдента при рівні значущості α та $k = n - m - 1$ степенях свободи.

Рівняння регресії для генеральної сукупності можна записати у вигляді:

$$\tilde{y} = (a_0 \pm t_{\alpha, k} \cdot \sigma_{a_0}) + (a_1 \pm t_{\alpha, k} \cdot \sigma_{a_1}) \cdot x \quad (9.48)$$