

3.2 Оценка значимости параметров линейной парной регрессии с помощью t -критерия Стьюдента

Пусть зависимость между переменными X и Y представляется в виде *теоретической* линейной регрессионной модели :

$$Y = A + B X + \varepsilon$$

По выборке ограниченного объема n можно построить *выборочное* уравнение регрессии:

$$\hat{y} = a + bx$$

где a и b - оценки параметров A и B .

Оценки a и b определяются на основе конечного числа статистических данных (выборки) и являются случайными величинами.

Правило проверки статистической значимости оценок a и b основывается на проверке статистических гипотез:

$$H_0: A = 0 \quad (H_1: A \neq 0)$$

$$H_0: B = 0 \quad (H_1: B \neq 0)$$

Невозможность отклонения нулевой гипотезы H_0 означает статистическую незначимость соответствующего коэффициента, и наоборот, отклонение нулевой гипотезы H_0 означает, что соответствующий коэффициент статистически значим.

Правило проверки статистической значимости коэффициента b

1. Определяется стандартная ошибка коэффициента регрессии b по формуле:

$$m_b = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2 / (n - 2)}{\sum (x_i - \bar{x})^2}}$$

2. Строится отношение: $t_b = \frac{b}{m_b}$,

которое называется t - статистикой и имеет распределение Стьюдента с числом степеней свободы $k = n - 2$

3. С помощью таблицы критических точек распределения Стьюдента определяется табличное значение t -критерия ($t_{табл}$) по заданному уровню значимости α и по числу степеней свободы $k = n - 2$.

4. Сравниваются фактическое t_b и табличное $t_{табл}$ значения t -критерия:

- ✓ Если $|t_b| \geq t_{табл}$, то H_0 отвергается, т.е. коэффициент b считается **статистически значимым** на уровне α , что указывает на наличие определенной линейной зависимости между Y и X .
- ✓ Если $|t_b| < t_{табл}$, то H_0 не отклоняется, и коэффициент b **статистически незначим** (признается случайная природа формирования коэффициента b).

5. Для коэффициента b строится доверительный интервал (интервальная оценка):

$$[b - t_{табл} \cdot m_b; b + t_{табл} \cdot m_b]$$

Если в границы доверительного интервала попадает 0, т.е. нижняя граница отрицательная, а верхняя положительная, то оцениваемый параметр принимается нулевым, т.к. признается статистически незначимым.

Правило проверки статистической значимости коэффициента a

1. Рассчитывается стандартная ошибка коэффициента a :

$$m_a = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2} \cdot \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$$

2. Определяется фактическое значение t -критерия Стьюдента: $t_a = a / m_a$

3. С помощью таблицы критических точек распределения Стьюдента определяется табличное значение t -критерия ($t_{табл}$) по заданному уровню значимости α и по числу степеней свободы $k = n - 2$.

4. Сравниваются фактическое и табличное значения t -критерия.

5. Определяется доверительный интервал для a :

$$[a - t_{табл} \cdot m_a; a + t_{табл} \cdot m_a]$$

Фактически доверительные интервалы определяют значения теоретических коэффициентов регрессии A и B , которые будут приемлемыми с надежностью $(1-\alpha)$ при найденных оценках a и b .

Доверительные интервалы содержат («накрывают») неизвестные коэффициенты A и B .