

Самостійна робота 9.3. Приклад розв'язання

На основі даних умови попередньої самостійної роботи 9.1 перевірити статистичну якість моделі, а саме:

- 1) перевірити адекватність економетричної моделі;
- 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

Оскільки статистика як метод дослідження має справу з даними, в яких досліджувані аналітиком закономірності створені різними випадковими факторами, тому більшість статистичних досліджень супроводжується перевіркою деяких припущень або гіпотез про джерела цих даних.

Основне перевірочне припущення називається нульовою гіпотезою і переважно формулюється як відсутність різниць, відсутність впливу фактору, рівність нулю значень вибірових характеристик тощо.

Друге перевірочне припущення (не завжди строго протилежне або обернене першому) називається конкуруючою або альтернативною гіпотезою.

При перевірці статистичних гіпотез можливі помилки (помилкові тлумачення) двох видів:

- можна відкинути нульову гіпотезу, коли вона насправді правильна (помилка першого роду);
- можна прийняти нульову гіпотезу, коли вона насправді не правильна (помилка другого роду).

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

1. Адекватність економетричної моделі статистичним даним оцінюється за F–критерієм Фішера.

В Ексель-файлі	7. Адекватність моделі за F-критерієм Фішера
----------------	--

Під адекватністю економетричної моделі розуміють відповідність моделі статистичним даним, за якими вона побудована.

Тестування регресійної моделі на адекватність за F–критерієм Фішера здійснюють для перевірки наявності (відсутності) зв'язку між факторами у такій послідовності:

1. Формулюють і розглядають гіпотези:

Тестування за критерієм Фішера дає змогу перевірити базову (нульову) гіпотезу ($H_0 : a_j = 0$ ($j=0, 1$)) проти альтернативної ($H_1 : a_j \neq 0$ ($j=0, 1$)) з метою отримання відповіді на питання щодо кращого шляху апроксимації даних – за середнім значенням $\hat{y}_i = \bar{y}$, чи за регресійною моделлю $\hat{y}_i = a_0 + a_1 \cdot x$, де $a_1 \neq 0$.

Розглянемо гіпотезу $H_0 : a_1=0$ проти альтернативної $H_1 : a_1 \neq 0$. Це рівнозначно тому, що перевіряється значущість коефіцієнту детермінації r^2 (гіпотеза $H_0 : r^2=0$, проти альтернативної $H_1 : r^2 \neq 0$). Скорочено це записують наступним чином (виокремлено жовтим):

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

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

Нульова гіпотеза означає в даному випадку наступне – між змінними x та y (вартістю ОЗ та прибутком підприємства не існує лінійного зв'язку). Тобто це гіпотеза про відсутність лінійного функціонального зв'язку.

Тобто нульова гіпотеза означає, що лінійний зв'язок між змінними відсутній. Якщо цю гіпотезу буде прийнято, то в моделі треба щось змінювати – наприклад, додати спостережень, або обрати інші змінні. Отже, якщо нульову гіпотезу буде прийнято, то з моделлю далі працювати не можна – її треба змінювати. Таким чином, нам треба щоб нульову гіпотезу було відхилено. Перевіримо це.

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

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

$$F_p = 18,4783 / 0,0322 = 573,6708$$

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

При перевірці значущості встановлюють значення її рівня α , який дає певну впевненість в тому, що помилкові висновки можуть бути нечасто. Рівень значущості виражає ймовірність того, що нульова гіпотеза H_0 відкидається тоді, коли вона насправді є правильною. Зрозуміло, що має сенс вибрати ймовірність якомога меншою.

В соціально-економічних дослідженнях рівень значущості α зазвичай приймають рівним 0,05, тоді довірча ймовірність $p = 1-\alpha=0,95$. Рівень значущості $\alpha=0,05$ означає, що в 5% випадків ми можемо помилитися, а в 95% випадків наші висновки правильні.

Отже беремо рівень значущості $\alpha=0,05$ або 5%.

Під терміном *ступінь вільності (свободи)* в економетрії розуміють число, яке показує, скільки незалежних елементів інформації зі змінних y_i ($i=1, \dots, n$) потрібно для розрахунку даної суми квадратів. В кореляційному аналізі в конспекті лекцій ми розглядали рівняння, яке пов'язує суми (формули (9.18), (9.19)):

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9.18)$$

Або вираз (9.18) у скороченому вигляді:

$$TSS = ESS + RSS \quad (9.19)$$

Кожна із зазначених сум пов'язана зі ступенями вільності: для загальної суми квадратів TSS потрібно $(n-1)$ незалежних чисел, тобто ступенів вільності; для регресійної суми квадратів ESS ($k_1=m-1$) ступенів вільності; для залишкової суми квадратів RSS ($k_2=n-m$) ступенів вільності.

Отже ступені свободи – це кількість спостережень, вільних від параметрів.

Обчислимо ступені вільності:

$$k_1=m-1=2-1=1$$

$$k_2=n-m=10-2=8$$

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).

Таблиця F – розподілу для ймовірності $p = 0,95$

Число ступенів свободи $k_1 = m$													
k_2	1	2	3	4	5	6	7	8	9	10	11	12	13
1	162	200	216	225	230	234	237	239	241	242	243	244	245
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,20	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55
8	5,33	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26
9	5,12	4,26	3,88	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28

Також F -критерій Фішера можна знайти за допомогою вбудованої в Ексель статистичної функції – назви функцій наведені нижче в табл. 1 для різних версій офісів.

Таблиця 1

Функції в MS Excel для пошуку F -критерію Фішера

Функція	$F.OBR.PX$	2010
	$F.PASPOBR$	2007 та більш ранні версії
	$F.INV.RT$	Англомова версія

Інтерактивне вікно функції $F.INV.RT$ та приклад його заповнення наведено на рис. 1.

Аргументи функції

F.INV.RT

Імовірність	0,05	↑	= 0,05
Супінь_свободи1	1	↑	= 1
Супінь_свободи2	8	↑	= 8

= 5,317655072

Повертає обернене значення (правобічного) F-розподілу ймовірності: якщо $p = F.DIST.RT(x,...)$, то $F.INV.RT(p,...) = x$.

Імовірність імовірність, пов'язана з F-інтегральним розподілом, число в інтервалі від 0 до 1 включно.

Значення: 5,317655072

[Довідка з цієї функції](#)

OK Скасувати

Рис. 1. Синтаксис функції F.INV.RT

В Excel всі функції «захиті» від зворотної ймовірності, тобто від ймовірності помилки. Тому для ймовірності 95% ми будемо брати зворотню від неї $1-0,95=0,05$ або 5% помилки.

Отже, $F_{a,k1,k2}=5,32$

5. Отримане розрахункове значення порівнюється з табличним. Перевіряємо виконання умови $F_p > F_{a,k1,k2}$. Якщо ця умова виконується, то гіпотезу H_0 відхиляємо.

Отже:

- якщо умова виконується, то нульову гіпотезу відхиляємо, оскільки в цьому випадку 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).

Оскільки в нашому випадку умова $F_p > F_{a,k1,k2}$ виконується ($573 > 5,32$), то це означає, що нульова гіпотеза про відсутність лінійного зв'язку між змінними x та y (вартістю ОЗ та прибутком підприємства) відхиляємо. Тобто між змінними існує лінійний зв'язок. Отже, з ймовірністю 95% можемо сказати, що економетрична модель є адекватною і з 5% ймовірністю ми можемо помилятися.

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

В Ексель-файлі	5. 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.

Для статистичного висновку про наявність або відсутність кореляційного зв'язку між досліджуваними змінними необхідно провести перевірку рівня значущості вибіркового коефіцієнта кореляції. Використаний критерій для розв'язку задач такого типу ґрунтується на розподілі різних статистик і називається критерієм значущості.

Процедура перевірки значущості починається з формулювання нульової гіпотези H_0 . У загальному випадку вона полягає в тому, що між параметром вибірки і параметром генеральної сукупності немає ніяких суттєвих різниць. Альтернативна гіпотеза H_1 полягає в тому, що між цими параметрами є суттєві різниці. Наприклад, при перевірці наявності кореляції в генеральній сукупності нульова гіпотеза полягає в тому, що істинний коефіцієнт кореляції рівний нулю ($H_0: \rho = 0$). Якщо в результаті перевірки виявиться, що нульова гіпотеза неприйнятна, то вибіркового коефіцієнта кореляції r_{xy} значно відрізняється від нуля (нульова гіпотеза відкидається і приймається альтернативна H_1). Іншими словами, припущення відносно некорельованості випадкових змінних в генеральній сукупності треба признати необґрунтованим. І навпаки, якщо на основі критерію значущості нульова гіпотеза приймається, тобто r_{xy} міститься в допустимій зоні випадкового розсіяння, тоді немає підстави вважати сумнівним припущення відносно некорельованості змінних у генеральній сукупності.

Перевіряючи значущість коефіцієнта парної кореляції, встановлюють наявність або відсутність кореляційного зв'язку між досліджуваними явищами. При відсутності зв'язку коефіцієнт кореляції генеральної сукупності рівний нулю ($\rho = 0$). Процедура перевірки гіпотези починається з формулювання нульової та альтернативної гіпотез:

H_0 : різниця між вибіровим коефіцієнтом кореляції r і $\rho = 0$ є незначною;

H_1 : різниця між r і $\rho = 0$ значна і, як наслідок, між змінними y і x є суттєвий зв'язок.

Вибірковий коефіцієнт кореляції, розрахований на основі вибірових даних, є точковою оцінкою коефіцієнта кореляції, і в свою чергу, є випадковою величиною. Тому, доцільно виконати перевірку гіпотези про відсутність кореляційного зв'язку, тобто перевіряється нульова гіпотеза $H_0: r_{xy}=0$ і альтернативна гіпотеза $H_1: r_{xy} \neq 0$.

Отже, у випадках незначного обсягу вибірки або при $|r| \leq 0,5$ необхідно здійснити статистичне тестування лінійного коефіцієнта кореляції Пірсона r_{xy} (для лінійних моделей) або кореляційного відношення η_{xy} (для нелінійних моделей) на

значущість за t -критерієм Стьюдента (псевдонім англійського статистика У.С. Госсета).

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

1. Формулюють і розглядають гіпотези:

Статистичні гіпотези сформулюємо наступним чином. Перевіримо нульову гіпотезу $H_0: r_{xy}=0$ (яка фактично означає, що між змінними y і x немає зв'язку) проти альтернативної гіпотези $H_1: r_{xy} \neq 0$ (між змінними y і x є суттєвий зв'язок).

$$H_0: r_{xy} = 0;$$

$$H_1: r_{xy} \neq 0$$

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

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

$$t_{\text{розрах}}=23,95$$

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).

Обираємо рівень значущості $\alpha=0,05$ або 5%.

Ступінь свободи $k=n-m=10-2$.

За таблицею t -критерію Стьюдента знаходимо критичне (теоретичне) значення.

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

Додаток Б

Критичні значення t – критерію Стьюдента ($t_{\text{табл}}$)

Ступінь свободи	Рівень значущості α та довірна ймовірність P					
	$\alpha = 0,40$	$\alpha = 0,20$	$\alpha = 0,10$	$\alpha = 0,05$ Υ	$\alpha = 0,02$	$\alpha = 0,01$
	$P = 0,60$	$P = 0,80$	$P = 0,90$	$P = 0,95$	$P = 0,98$	$P = 0,99$
1	1,376	3,078	6,314	12,706	31,821	63,657
2	1,061	1,886	2,920	4,303	6,965	9,925
3	0,978	1,638	2,353	3,182	4,541	5,841
4	0,941	1,533	2,132	2,776	3,747	4,604
5	0,920	1,476	2,015	2,571	3,365	4,032
6	0,906	1,440	1,943	2,447	3,143	3,707
7	0,896	1,415	1,895	2,365	2,998	3,499
8✓	0,889	1,397	1,860	2,306	2,896	3,355
9	0,883	1,383	1,833	2,262	2,821	3,250
10	0,879	1,372	1,812	2,228	2,764	3,169
11	0,876	1,363	1,796	2,201	2,718	3,106
12	0,873	1,356	1,782	2,179	2,681	3,055
13	0,870	1,350	1,771	2,160	2,650	3,010

В нашому випадку t -критичне (табличне)=2,306

Його можна визначити як за допомогою таблиць, так і за допомогою вбудованих функцій *Excel* статистичних функцій – назви функцій наведені нижче в табл. 2 для різних версій офісів.

Таблиця 2

Функції в *Excel* для пошуку *t*-критерію Стьюдента

Функція	СТЮДЕНТ.ОБР.2Х	2010
	T.INV.2T	Англomовна версія
	СТЮДЕНТРАСПОБР	2007 та більш ранні версії

Інтерактивне вікно функції *T.INV.2T* та приклад його заповнення наведено на рис. 2.

Рис. 2. Синтаксис функції *T.INV.2T*

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

– якщо умова виконується, то нуль-гіпотеза відкидається – r_{xy} є статистично значущий і між змінними y і x є суттєвий зв'язок ;

– якщо умова не виконується, то нуль-гіпотеза приймається, r_{xy} є статистично незначущим.

Розраховане за формулою (9.38) значення $|t_{\text{розрах}}|$ порівнюють з його критичним (теоретичним, табличним) значенням при прийнятому рівні значущості α та числі ступенів свободи $k=n-m$:

– 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.

В нашому випадку умова $|t_{\text{розрах}}| > t_{\alpha,k}$ ($23,95 > 2,306$) справджується, тому нульова гіпотеза не підтверджується з ймовірністю 95%. Тобто в генеральній сукупності між змінними y та x існує суттєвий лінійний зв'язок.

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 .

$$\sigma_r = \frac{1 - r^2}{\sqrt{n - m}} = \frac{1 - r^2}{\sqrt{k}} = \frac{1 - 0,9931^2}{\sqrt{10 - 2}} = 0,004863$$

Тепер можемо визначити граничну помилку коефіцієнту кореляції Δr :

$$\Delta r = \pm t_{\alpha, k} \cdot \sigma_r = 2,306 \cdot 0,004863 = 0,01121$$

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

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

$$0,9931 - 0,01121 \leq r_{xy} \leq 0,9931 + 0,01121$$

$$0,98189 \leq r_{xy} \leq 1,00431$$

Таким чином, коефіцієнт кореляції r_{xy} попадає в інтервал від 0,98 до 1. Оскільки нуль в цей інтервал не входить, то це означає, що з ймовірністю 95% коефіцієнт кореляції не дорівнюватиме 0. Отже коефіцієнт кореляції є статистично значущим та між змінними x та y існує суттєвий лінійний зв'язок.

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

В Ексель-файлі	6. Перевірка статистичної значущості оцінок параметрів моделі та визначення їх надійних інтервалів.
----------------	---

Перевірити гіпотезу про статистичну значущість оцінок параметрів моделі можна здійснити двома способами (так само, як і коефіцієнт кореляції Пірсона): 1) перевіркою на значущість та 2) перевірка інтервалів надійності.

1) Перевірка значимості (значущості).

Послідовність тестування за t-критерієм Стьюдента:

1. Формують 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 $k=n-m$ 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.

Проведемо перевірку значущості для параметру a_0 :

1. Формулюємо та розглядаємо гіпотези:

$H_0 : a_0=0$ – оцінка параметра незначуща;

$H_1 : a_0 \neq 0$ – оцінка параметра значуща.

2. Обчислюємо розрахункове значення t -критерію Стюдента для параметру a_0 за формулою:

$$t_{a_0} = \frac{|a_0|}{\sigma_{a_0}}$$

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

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

Розрахунки починаємо знизу (формули 9.45)

$$\sigma_u = \text{SQRT}(0,2577 / (10-2)) = 0,1795$$

Наступною обрахуємо σ_{a_0} (стандартну помилку оцінки параметру a_0) за формулою (9.43):

$$\sigma_{a_0} = 0,1795 \cdot \text{SQRT}(181,51 / (10 \cdot 12,589)) = 0,2155$$

Тепер можемо вже обрахувати розрахункове значення t -критерію Стюдента для параметру a_0 :

$$t_{a_0} = \text{ABS}(-1,8994) / 0,2155 = 8,8138$$

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

Задаємо рівень значущості $\alpha=0,05$ або 5%.

Визначаємо ступінь свободи $k=n-m=10-2$

Отже, табличне (критичне) значення t -критерію Стюдента необхідне для розрахунку збігається зі значенням t -критерію, визначеному при аналізі значимості коефіцієнту кореляції r_{xy} і відповідно дорівнює 2,306.

Його також можна визначити як за допомогою таблиць, так і за допомогою вбудованих функцій Ексель (див. табл. 2). Нижче цю таблицю продубльовано.

Функція	СТЮДЕНТ.ОБР.2Х	2010
	T.INV.2T	Англomовна версія
	СТЮДЕНТРАСПОБР	2007 та більш ранні версії

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

Умова справджується ($8,8138 > 2,306$), що означає, що нульова гіпотеза відкидається і параметр a_0 генеральної сукупності є статистично значущим.

Проведемо перевірку значущості для параметру a_1 :

1. Формулюємо та розглядаємо гіпотези:

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

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

2. Обчислюємо розрахункове значення t -критерію Стюдента для параметру a_0 за формулою:

$$t_{a_1} = \frac{|a_1|}{\sigma_{a_1}}$$

де 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) вже визначали для параметру a_0 . Для параметру a_1 він такий самий:

$$\sigma_u = \text{SQRT}(0,2577 / (10-2)) = 0,1795$$

Наступною обрахуємо σ_{a_1} (стандартну помилку оцінки параметру a_1) за формулою (9.44):

$$\sigma_{a_1} = 0,1795 / \text{SQRT}(12,589) = 0,0506$$

Тепер можемо вже обрахувати розрахункове значення t -критерію Стюдента для параметру a_1 :

$$t_{a1} = \text{ABS}(1,2115) / 0,0506 = 23,95$$

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

Вони такі самі як для оцінки параметру a_0 .

Отже, табличне (критичне) значення t -критерію Стюдента дорівнює 2,306.

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

Умова справджується ($23,95 > 2,306$), що означає, що нульова гіпотеза відкидається і параметр a_1 генеральної сукупності є статистично значущим.

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 $y=a_0+a_1 \cdot x$. Спочатку знаходять граничні помилки оцінок параметрів:

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

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:

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

Знайдемо граничну помилку для параметру a_0 :

$$\Delta a_0 = 2,306 \cdot 0,2155 = 0,4970$$

Межі довірчих інтервалів становлять:

$$\begin{aligned} -1,8994 - 0,4970 &\leq a_0 \leq -1,8994 + 0,4970 \\ -2,3964 &\leq a_0 \leq -1,4025 \end{aligned}$$

Таким чином, оцінка параметру a_0 попадає в інтервал від -2,3964 до -1,4025. Оскільки нуль в цей інтервал не входить, то це означає, що з ймовірністю 95% оцінка параметру a_0 не дорівнюватиме 0. Отже параметр a_0 є статистично значущим.

Знайдемо граничну помилку для параметру a_1 :

$$\Delta a_1 = 2,306 \cdot 0,0506 = 0,1166$$

Межі довірчих інтервалів становлять:

$$\begin{aligned} 1,2115 - 0,1166 &\leq a_1 \leq 1,2115 + 0,1166 \\ 1,0949 &\leq a_1 \leq 1,3282 \end{aligned}$$

Таким чином, оцінка параметру a_1 попадає в інтервал від 1,0949 до 1,3282. Оскільки нуль в цей інтервал не входить, то це означає, що з ймовірністю 95% оцінка параметру a_1 не дорівнюватиме 0. Отже параметр a_0 є статистично значущим та між змінними x та y існує суттєвий лінійний зв'язок.