Лабораторна робота № 3

Мета роботи. Навчитися будувати лінійну багатофакторну економетричну модель класичним методом МНК за допомогою табличного редактора Microsoft Excel та виконувати економічний аналіз характеристик взаємозв'язку.

Приклад виконання лабораторної роботи

Завдання. За допомогою табличного редактора Microsoft Excel на основі середньостатистичних даних, дослідити залежність між тижневими витратами на харчування, загальними витратами та складом сім'ї (табл. 3.1).

Необхідно:

1. Виконати ідентифікацію змінних та специфікацію моделі.
2. Оцінити параметри моделі за класичним методом найменших квадратів (МНК).
3. Побудувати модель на основі покрокової регресії.
   - 3.1. Виконати нормалізацію змінних.
   - 3.2. Побудувати кореляційну матрицю.
   - 3.3. Оцінити параметри моделі на основі МНК методом покрокової регресії.
4. Побудувати базову таблицю дисперсійного аналізу (ANOVA − таблицю) та визначити дисперсії.
5. Визначити коефіцієнти детермінації, множинної кореляції та еластичності.
6. Визначити стандартні помилки оцінок параметрів моделі та оцінити їх.
7. Перевірити статистичну значущість оцінок параметрів моделі та визначити їх надійні інтервали.
8. Перевірити адекватність економетричної моделі.
9. Визначити точковий та інтервальний прогнози для заданих останніх значень незалежних змінних.
10. Побудувати графіки фактичних даних та лінію регресії.
11. Перевірити точність економетричної моделі за допомогою середньої відносної похибки апроксимації.
12. Побудувати економетричну модель за допомогою функції ЛИНЕЙН та інструменту ДАННЫЕ → АНАЛИЗ ДАННЫХ → РЕГРЕССИЯ.
13. Виконати економіко-математичний аналіз характеристик економетричної моделі.

Таблиця 3.1 – Вихідні дані

№ п/п	Витрати на харчування, у.г.о.	Загальні витрати, у.г.о.	Середня кількість членів сім'ї, люд.
1	20	45	1,5
2	32	75	1,6
3	48	125	1,9

4	65	223	1,8
5	45	92	3,4
6	64	146	3,6
7	79	227	3,5
8	104	358	5,5
9	68	135	5,4
10	93	218	5,4
11	117	331	5,3
12	145	490	8,5
13	91	175	8,3
14	131	205	8,1
15	167	468	7,3
16	195	749	8,4
	Прогнозні значення	*900*	*9,0*

Розв'язання.

1. *Ідентифікація змінних та специфікація моделі.*

Ідентифікуємо змінні економетричної моделі:

Y − вектор витрат на харчування (залежна змінна);

X_1 − вектор загальних витрат (незалежна змінна);

X_2 − вектор складу сім'ї (незалежна змінна);

u − вектор залишків (стохастична складова).

Загальний вигляд економетричної моделі:

$$Y = f\left(X_1, X_2, u\right).$$

Розглянемо специфікацію моделі у вигляді лінійної функції:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + u$$

та відповідне їй рівняння регресії:

$$\widehat{Y} = \widehat{a}_0 + \widehat{a}_1 X_1 + \widehat{a}_2 X_2,$$

де a_0, a_1, a_2 − невідомі параметри моделі; $\widehat{a}_0, \widehat{a}_1, \widehat{a}_2$ − їх оцінки;

$\widehat{Y}$ − теоретичне значення результативної ознаки;

$u = Y - \widehat{Y}$ − вектор залишків (стохастична складова).

За допомогою функції СЧЕТ (Y) визначимо обсяг вихідної сукупності $n = \text{СЧЕТ}\left(Y\right) = 16$. Для розрахунків будемо використовувати тільки перші 16 значень вихідних даних, оскільки останні значення незалежних змінних прогнозні: $x_{1\text{пр}} = 900$, $x_{2\text{пр}} = 9$.

2. *Оцінювання параметрів моделі класичними методом найменших квадратів (МНК).*

Для матричного рівняння лінійної багатофакторної моделі $Y = XA + U$ вектор оцінювання на основі МНК має вигляд:

$$A=\begin{pmatrix}\hat{a}_0\\ \hat{a}_1\\ \hat{a}_2\end{pmatrix}=\left(X^t\cdot X\right)^{-1}\cdot\left(X^t\cdot Y\right).$$

У даному операторі A матриця X характеризує всі незалежні змінні моделі. Оскільки економетрична модель має вільний член $\hat{a}_0$, для якого всі $x_i=1$, то матрицю X треба доповнити першим стовпцем, в якому всі шістнадцять членів є одиницями; Y − вектор залежної змінної; X^t − матриця, транспонована до матриці X .

Запишемо матрицю незалежних змінних X та вектор залежної змінної Y:

$$X\ \left\|\begin{matrix}1 & 45 & 1{,}5\\ 1 & 75 & 1{,}6\\ 1 & 125 & 1{,}9\\ 1 & 223 & 1{,}8\\ 1 & 92 & 3{,}4\\ 1 & 146 & 3{,}6\\ 1 & 227 & 3{,}5\\ 1 & 358 & 5{,}5\\ 1 & 135 & 5{,}4\\ 1 & 218 & 5{,}4\\ 1 & 331 & 5{,}3\\ 1 & 490 & 8{,}5\\ 1 & 175 & 8{,}3\\ 1 & 205 & 8{,}1\\ 1 & 468 & 7{,}3\\ 1 & 749 & 8{,}4\end{matrix}\right\|$$

$$Y\ \left\|\begin{matrix}20\\ 32\\ 48\\ 65\\ 45\\ 64\\ 79\\ 104\\ 68\\ 93\\ 117\\ 145\\ 91\\ 131\\ 167\\ 195\end{matrix}\right\|$$

Транспонувавши матрицю X (функція ТРАНСП), одержимо матрицю X^t:

$$
Xt \left\| \begin{array}{cccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 45 & 75 & 125 & 223 & 92 & 146 & 227 & 358 \\ 1{,}5 & 1{,}6 & 1{,}9 & 1{,}8 & 3{,}4 & 3{,}6 & 3{,}5 & 5{,}5 \end{array} \right.
$$

$$
\left. \begin{array}{cccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 135 & 218 & 331 & 490 & 175 & 205 & 468 & 749 \\ 5{,}4 & 5{,}4 & 5{,}3 & 8{,}5 & 8{,}3 & 8{,}1 & 7{,}3 & 8{,}4 \end{array} \right\|
$$

та виконаємо множення матриць X^t і X (функція МУМНОЖ):

$$
Xt*X \left\| \begin{array}{ccc} 16{,}00 & 4062{,}00 & 79{,}50 \\ 4062{,}00 & 1550562{,}00 & 25074{,}80 \\ 79{,}50 & 25074{,}80 & 495{,}69 \end{array} \right\|
$$

Визначимо матрицю помилок $C = \left(X^t \cdot X \right)^{-1}$, що обернена до матриці $X^t \cdot X$ (функція МОБР):

$$
\text{обр} \left\| \begin{array}{ccc} 0{,}308281 & -4{,}4188E-05 & -0{,}0472076 \\ -4{,}4E-05 & 3{,}5507E-06 & -0{,}0001725 \\ -0{,}04721 & -0{,}00017253 & 0{,}0183161 \end{array} \right\|
$$

потім помножимо матрицю X^t на матрицю Y (функція МУМНОЖ):

$$
Xt*Y \left\| \begin{array}{c} 1464{,}00 \\ 498666{,}00 \\ 8916{,}80 \end{array} \right\|
$$

Підставимо отримані значення оберненої матриці $\left(X^t \cdot X \right)^{-1}$ та добуток матриць $X^t \cdot Y$ у вектор оцінювання $\widehat{A}$ і визначимо оцінки параметрів моделі:

$$
\widehat{A} \left\| \begin{array}{c} 8{,}34760 \\ 0{,}16753 \\ 8{,}17526 \end{array} \right\|
$$

Таким чином, оцінки параметрів моделі будуть:

$\widehat{a}_0 = 8{,}34760$, $\widehat{a}_1 = 0{,}16753$, $\widehat{a}_2 = 8{,}17526$.

Звідси, економетрична модель витрат на харчування має вигляд:

$$
\widehat{Y} = 8{,}34760 + 0{,}16753 X_1 + 8{,}17526 X_2 .
$$

3. Побудова економетричної моделі на основі покрокової регресії.

3.1. Нормалізація змінних.

Обчислимо середні значення, дисперсії та стандартні відхилення залежної змінної Y та незалежних змінних $X1$, $X2$, використовуючи функції СРЗНАЧ, ДИСПР та СТАНДОТКЛОНП (рис. 3.1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	i	Yi	X1i	X2i		1	45	1,5		20		1	1	1	1	1	1
2	1	20	45	1,5		1	75	1,6		32	Xt	45	75	125	223	92	146
3	2	32	75	1,6		1	125	1,9		48		1,5	1,6	1,9	1,8	3,4	3,6
4	3	48	125	1,9		1	223	1,8		65							
5	4	65	223	1,8		1	92	3,4		45	Yt	20	32	48	65	45	64
6	5	45	92	3,4		1	146	3,6		64							
7	6	64	146	3,6	X	1	227	3,5	Y	79		16,00	4062,00	79,50			1464,00
8	7	79	227	3,5		1	358	5,5		104	Xt*X	4062,00	1550562,00	25074,80		Xt*Y	498666,00
9	8	104	358	5,5		1	135	5,4		68		79,50	25074,80	495,69			8916,80
10	9	68	135	5,4		1	218	5,4		93							
11	10	93	218	5,4		1	331	5,3		117		0,30828	-4,419E-05	-0,047208			
12	11	117	331	5,3		1	490	8,5		145	обр	-4E-05	3,551E-06	-0,000173			
13	12	145	490	8,5		1	175	8,3		91		-0,0472	-0,0001725	0,0183161			
14	13	91	175	8,3		1	205	8,1		131							
15	14	131	205	8,1		1	468	7,3		167		8,34760					
16	15	167	468	7,3		1	749	8,4		195	A	0,16753					
17	16	195	749	8,4								8,17526					
18	*прогн*		900	9,0			sig^2_Y	2282,3750		sig_Y	47,7742						
19	***n***	**16**					sig^2_X1	32457,6094		sig_X1	180,1600						
20	**sum**	1464	4062	79,5			sig^2_X2	6,2921		sig_X2	2,5084						
21	**среднее**	91,50	253,875	4,9688													

Рисунок 3.1 – Розрахункова таблиця

Потім за допомогою формул:

$$Y^* = \frac{Y - \bar{Y}}{\sigma_Y}, \qquad X_1^* = \frac{X_1 - \bar{X}_1}{\sigma_{X_1}}, \qquad X_2^* = \frac{X_2 - \bar{X}_2}{\sigma_{X_2}}$$

визначимо нормалізовані змінні, які розмістимо в окремій книзі табличного редактора Microsoft Excel (рис. 3.2, стовпчики A, B, C).

Для перевірки проведених розрахунків, визначимо значення нормалізованих змінних в матричному вигляді, застосувавши функцію НОРМАЛИЗАЦИЯ до вихідних даних:

$$\text{НОРМАЛИЗАЦИЯ}\left(Y, \bar{Y}, \sigma_Y\right),$$

$$\text{НОРМАЛИЗАЦИЯ}\left(X_1, \bar{X}_1, \sigma_{X_1}\right),$$

$$\text{НОРМАЛИЗАЦИЯ}\left(X_2, \bar{X}_2, \sigma_{X_2}\right)$$

порівняємо їх з раніше знайденими за формулами (рис. 3.2, стовпчики E, F, G).

L18

	A	B	C	D	E	F	G	H
1	Y*	X1*	X2*		^Y	^X1	^X2	
2	-1,49662	-1,15939	-1,38285		-1,49662	-1,15939	-1,38285	
3	-1,24544	-0,99287	-1,34298		-1,24544	-0,99287	-1,34298	
4	-0,91053	-0,71534	-1,22338		-0,91053	-0,71534	-1,22338	
5	-0,55469	-0,17138	-1,26325		-0,55469	-0,17138	-1,26325	
6	-0,97333	-0,89851	-0,62539		-0,97333	-0,89851	-0,62539	
7	-0,57562	-0,59877	-0,54566		-0,57562	-0,59877	-0,54566	
8	-0,26165	-0,14917	-0,58553		-0,26165	-0,14917	-0,58553	
9	0,26165	0,57796	0,21179		0,26165	0,57796	0,21179	
10	-0,49190	-0,65983	0,17192		-0,49190	-0,65983	0,17192	
11	0,03140	-0,19913	0,17192		0,03140	-0,19913	0,17192	
12	0,53376	0,42809	0,13206		0,53376	0,42809	0,13206	
13	1,11985	1,31064	1,40776		1,11985	1,31064	1,40776	
14	-0,01047	-0,43781	1,32803		-0,01047	-0,43781	1,32803	
15	0,82681	-0,27129	1,24830		0,82681	-0,27129	1,24830	
16	1,58035	1,18853	0,92937		1,58035	1,18853	0,92937	
17	2,16644	2,74825	1,36790		2,16644	2,74825	1,36790	
18								
19								
20								

специф | **нормализ**

Рисунок 3.2 – Нормалізовані змінні

3.2. Побудова кореляційної матриці.

Побудуємо кореляційну матрицю, тобто матрицю парних коефіцієнтів кореляції:

$$r_{YX_1} = \frac{\left(Y^*\right)^t X_1^*}{n} = 0,922164,$$

$$r_{YX_2} = \frac{\left(Y^*\right)^t X_2^*}{n} = 0,856654,$$

$$r_{X_1X_2} = \frac{\left(X_1^*\right)^t X_2^*}{n} = 0,676527.$$

Звідси кореляційна матриця:

$$r = \begin{pmatrix} 1 & 0,922164 & 0,856654 \\ 0,922164 & 1 & 0,676527 \\ 0,856654 & 0,676527 & 1 \end{pmatrix}.$$

Перевіримо знайдені елементи кореляційної матриці за допомогою статистичної функції КОРРЕЛ:

$$r_{YX_1} = \text{КОРРЕЛ}\left(Y^*,\ X_1^*\right) = 0,922164,$$

$$r_{YX_2} = \text{КОРРЕЛ}\left(Y^*,\ X_2^*\right) = 0,856654,$$

$$r_{X_1X_2} = \text{КОРРЕЛ}\left(X_1^*,\ X_2^*\right) = 0,676527.$$

Крім цього, кореляційну матрицю, автоматично можна отримати, використовуючи інструмент ДАННЫЕ→АНАЛИЗ ДАННЫХ →КОРРЕЛЯЦИЯ.

Розглянемо порядок побудови кореляційної матриці.

1. В книзі з вихідними даними на панелі інструментів вибираємо ДАННЫЕ→АНАЛИЗ ДАННЫХ→Корреляция (рис. 3.3).
2. У діалоговому вікні, що відкрилося, вказуємо діапазон вихідних даних разом з заголовками B1: D17, вибираємо режим **Метки в первой строке** та параметри виводу: **Выходной интервал** (рис. 3.4). Натискуємо клавішу ОК.
3. В результаті отримаємо окремі елементи кореляційної матриці у вигляді нижньої трикутної матриці (рис. 3.5).

Оскільки кореляційна матриця симетрична, то результати співпали.

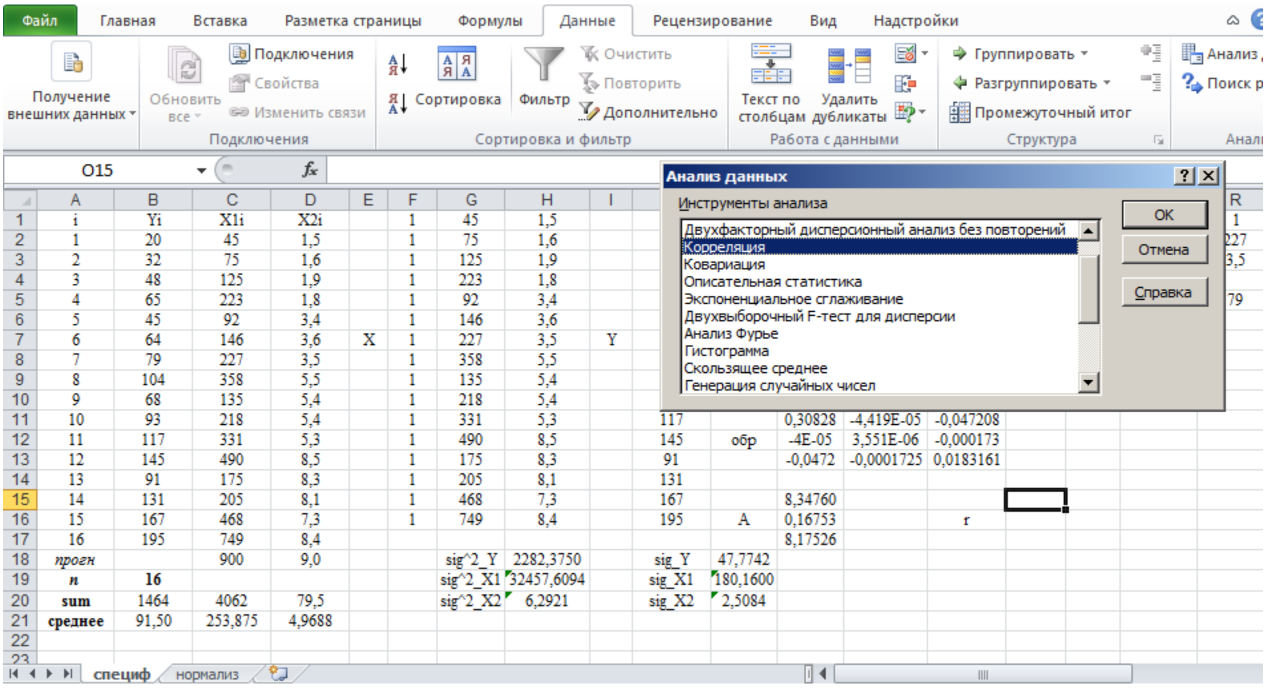


Рисунок 3.3 – Застосування ДАННЫЕ→АНАЛИЗ ДАННЫХ→Корреляция

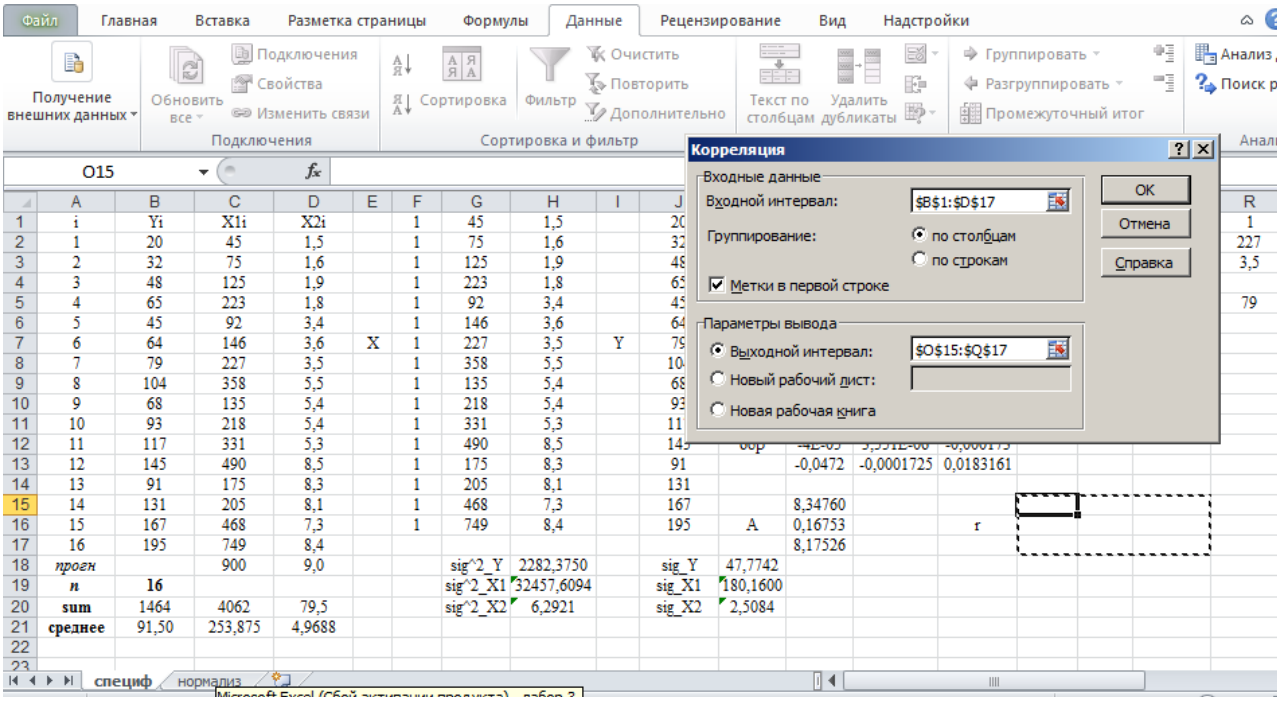


Рисунок 3.4 – Застосування **Корреляция**

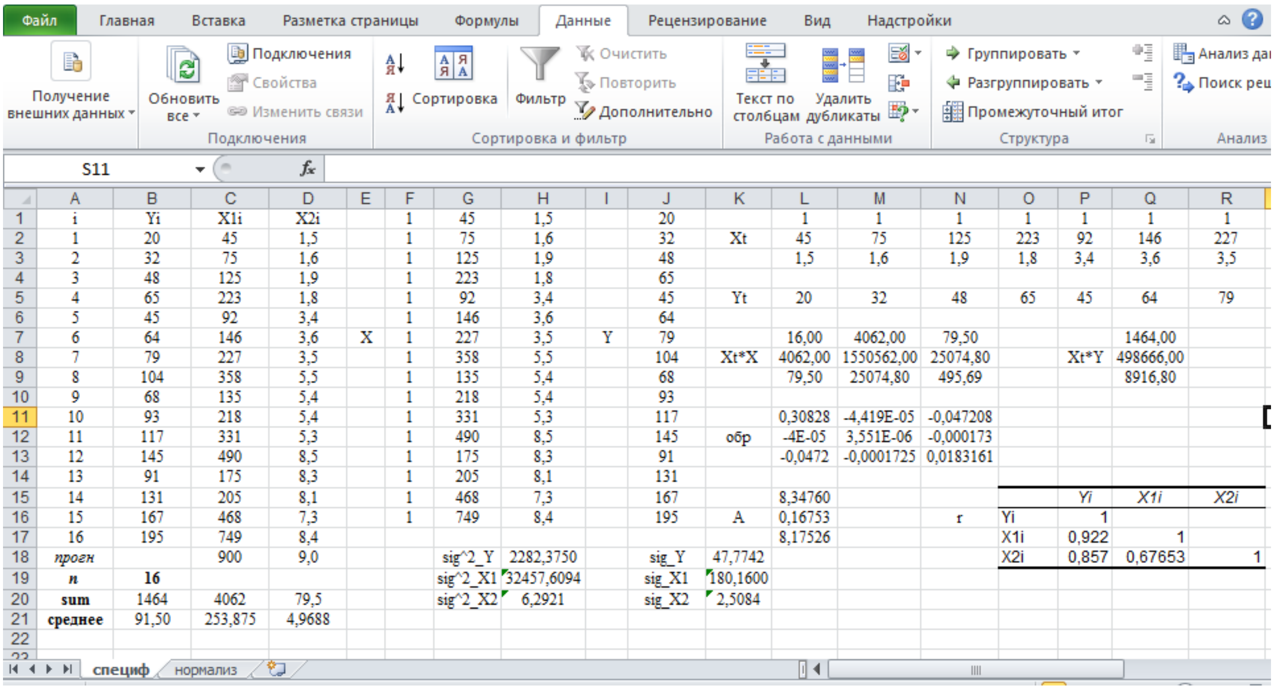


Рисунок 3.5 – Результат застосування **Корреляция**

3.3. Оцінювання параметрів моделі методом МНК на основі покрокової регресії.

Побудуємо економетричну модель для нормалізованих даних:

$$Y^* = \hat{b}_1 X_1^* + \hat{b}_2 X_2^*,$$

де Y^*, X_1^*, X_2^* – нормалізовані змінні.

1 етап. Порівнюючи абсолютні значення r_{YX_i}, вибираємо $\max\left\{\left|r_{YX_i}\right|\right\}$, що вказує на ту незалежну змінну, яка найтісніше зв'язана з *Y*. Враховуючи, що

$$\max\left\{ r_{YX_1} = 0{,}922164;\ r_{YX_2} = 0{,}856654\right\} = 0{,}922164,$$

то на першому етапі потрібно ввести змінну X_1^*, тобто побудувати економетричну модель виду:

$$\hat{Y}^* = \hat{b}_1 X_1^*.$$

Застосуємо функцію ЛИНЕЙН$\left(Y^*;\ X_1^*; 0; 1\right)$ до нормалізованих даних Y^* та X_1^* (рис. 3.6, табл. 3.2).

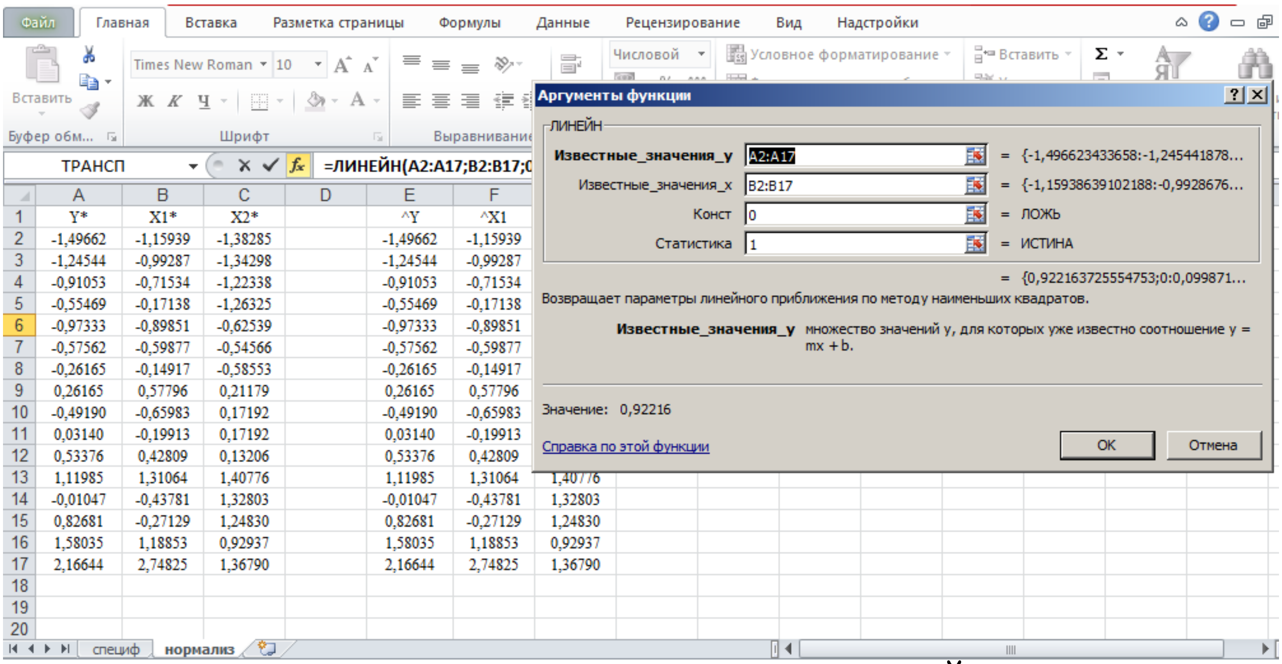


Рисунок 3.6 – Застосування функції ЛИНЕЙН до нормалізованих даних Y^* та X_1^*

Таблиця 3.2 – Застосування функції ЛИНЕЙН до нормалізованих даних Y^* та X_1^*

$\hat{a}_1$	$\hat{a}_0$
$S_{\hat{a}_1}$	$S_{\hat{a}_0}$
R^2	σ_u
F	$k = n - m$
$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$	$SSE = \sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$

0,92216	0
0,09987	#Н/Д
0,85039	0,399485
85,25795	15
13,60617	2,39383

2 етап. Серед тих значень r_{YX_i}, що залишилися, вибирається $\max\left\{\left|r_{YX_i}\right|\right\}$ і в модель вводиться наступна незалежна змінна і т.д.

На другому етапі включимо в економетричну модель змінну X_2^*, в результаті чого модель набуде наступного вигляду:

$$\hat{Y}^* = \hat{b}_1 X_1^* + \hat{b}_2 X_2^*.$$

Застосуємо функцію ЛИНЕЙН$\left(Y^*;\, X_1^*;\, X_2^*;\, 0;\, 1\right)$ до нормалізованих даних Y^*, X_1^* та X_2^* (рис. 3.7, табл. 3.3).

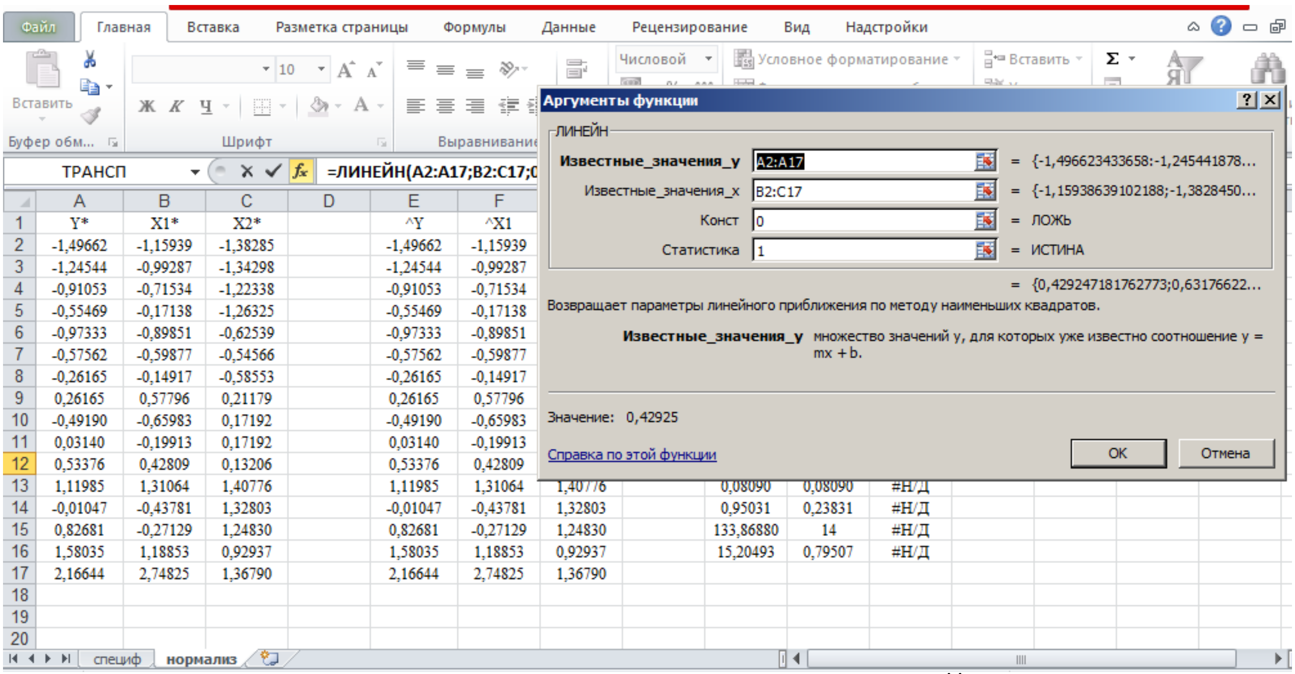


Рисунок 3.7 – Застосування функції ЛИНЕЙН до нормалізованих даних Y^*, X_1^* та X_2^*

Таблиця 3.2 – Застосування функції ЛИНЕЙН до нормалізованих даних Y^*, X_1^* та X_2^*

$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_0$		0,42925	0,63177	0
$S_{\hat{\beta}_2}$	$S_{\hat{\beta}_1}$	$S_{\hat{\beta}_0}$		0,08090	0,08090	#Н/Д
R^2	σ_u			0,95031	0,23831	#Н/Д
F	$k=n-m$			133,86880	14	#Н/Д
$SSR=\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2$	$SSE=\sum_{i=1}^{n}\left(\hat{Y}_i-Y_i\right)^2$			15,20493	0,79507	#Н/Д

Перехід до економетричної моделі, в якій змінні виражені в абсолютних значеннях.

Виконаємо перехід до економетричної моделі, в якій змінні виражені в абсолютних значеннях, використовуючи одержані результати:

$$\hat{a}_1=\hat{\beta}_1\frac{\sigma_Y}{\sigma_{X_1}}=0,63177\cdot\frac{47,7742}{180,1600}=0,1675296,$$

$$\hat{a}_2=\hat{\beta}_2\frac{\sigma_Y}{\sigma_{X_2}}=0,42925\cdot\frac{47,7742}{2,5084}=8,1752581,$$

$$\hat{a}_0=\bar{Y}-\hat{a}_1\bar{X}_1-\hat{a}_2\bar{X}_2=$$

$$=91,50-0,1675296\cdot 253,875-8,1752581\cdot 4,9688=8,3476001.$$

Таким чином, економетрична модель витрат на харчування має вигляд:

$$\hat{Y}=8,34760+0,16753X_1+8,17526X_2.$$

Теоретичні регресійні значення залежної змінної $\widehat{Y}$ можна отримати на основі одержаної моделі, підставивши в неї фактичні значення незалежних змінних X_1 та X_2 (рис. 3.8, стовпчик N2:N17) або перемноживши матриці X та $\widehat{A}$, тобто $\widehat{Y} = X\widehat{A}$ (рис. 3.8, стовпчик O2:O17).

Всі розрахунки для зручності краще розмістити в окремій книзі (рис. 3.9).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Y*	X1*	X2*		^Y	^X1	^X2							теор ^Y	
2	-1,49662	-1,15939	-1,38285		-1,49662	-1,15939	-1,38285		r_YX1	0,922164		0,922164		28,149321	28,149321
3	-1,24544	-0,99287	-1,34298		-1,24544	-0,99287	-1,34298		r_YX2	0,856654		0,856654		33,992736	33,992736
4	-0,91053	-0,71534	-1,22338		-0,91053	-0,71534	-1,22338		r_X1X2	0,676527		0,676527		44,821795	44,821795
5	-0,55469	-0,17138	-1,26325		-0,55469	-0,17138	-1,26325							60,422173	60,422173
6	-0,97333	-0,89851	-0,62539		-0,97333	-0,89851	-0,62539		0,92216	0				51,556204	51,556204
7	-0,57562	-0,59877	-0,54566		-0,57562	-0,59877	-0,54566		0,09987	#Н/Д		^a_0	8,3476001	62,237856	62,237856
8	-0,26165	-0,14917	-0,58553		-0,26165	-0,14917	-0,58553		0,85039	0,399485		^a_1	0,1675296	74,990231	74,990231
9	0,26165	0,57796	0,21179		0,26165	0,57796	0,21179		85,25795	15		^a_2	8,1752581	113,28713	113,28713
10	-0,49190	-0,65983	0,17192		-0,49190	-0,65983	0,17192		13,60617	2,39383				75,110495	75,110495
11	0,03140	-0,19913	0,17192		0,03140	-0,19913	0,17192							89,015454	89,015454
12	0,53376	0,42809	0,13206		0,53376	0,42809	0,13206		0,42925	0,63177	0			107,12878	107,12878
13	1,11985	1,31064	1,40776		1,11985	1,31064	1,40776		0,08090	0,08090	#Н/Д			159,92682	159,92682
14	-0,01047	-0,43781	1,32803		-0,01047	-0,43781	1,32803		0,95031	0,23831	#Н/Д			105,51993	105,51993
15	0,82681	-0,27129	1,24830		0,82681	-0,27129	1,24830		133,86880	14	#Н/Д			108,91077	108,91077
16	1,58035	1,18853	0,92937		1,58035	1,18853	0,92937		15,20493	0,79507	#Н/Д			146,43085	146,43085
17	2,16644	2,74825	1,36790		2,16644	2,74825	1,36790							202,49947	202,49947

Рисунок 3.8 – Теоретичні регресійні значення залежної змінної $\widehat{Y}$

	A	B	C	D	E	F	G	H
1	i	Yi	^Y	u^2=(^Y-Y)^2	(Y-_Y)^2	(^Y-_Y)^2	mod(Y-^Y)/Y*100	
2	1	20	28,1493209	66,41143058	5112,25	4013,308547	40,74660433	
3	2	32	33,9927357	3,970995697	3540,25	3307,085444	6,227299162	
4	3	48	44,8217949	10,10098751	1892,25	2178,854829	6,621260576	
5	4	65	60,4221734	20,95649679	702,25	965,831309	7,042810223	
6	5	45	51,5562041	42,98381274	2162,25	1595,506828	14,56934254	
7	6	64	62,2378561	3,10515126	756,25	856,2730679	2,753349904	
8	7	79	74,9902307	16,07824988	156,25	272,5724825	5,075657348	
9	8	104	113,287129	86,25076753	156,25	474,6789959	8,929931859	
10	9	68	75,1104947	50,55913469	552,25	268,6158844	10,45660983	
11	10	93	89,0154544	15,87660362	2,25	6,172966824	4,284457631	
12	11	117	107,128777	97,44103639	650,25	244,2586818	8,436942426	
13	12	145	159,926815	222,8098157	2862,25	4682,229055	10,2943554	
14	13	91	105,519929	210,8283275	0,25	196,5583989	15,95596553	
15	14	131	108,910766	487,9342559	1560,25	303,1347749	16,86201064	
16	15	167	146,430854	423,0897832	5700,25	3017,398678	12,31685413	
17	16	195	202,499465	56,24197545	10712,25	12320,88123	3,845879493	
18	**sum**	**1464**	**1464**	**1814,6388**	**36518**	**34703,3612**	**174,4193**	
19	**среднее**	**91,50**	**91,50**					

Рисунок 3.9 – Розрахункова таблиця

4. Побудова базової таблиці дисперсійного аналізу (ANOVA – таблиці) та визначення дисперсій.

Використовуючи результати розрахункової таблиці (рис. 3.9) або за допомогою статистичних функцій редактора Excel (табл. 3.3, рис. 3.10), визначимо основні елементи ANOVA − таблиці, в якій $n = 16$ − обсяг сукупності, $m = 3$ − кількість параметрів моделі ($\hat{a}_0, \hat{a}_1, \hat{a}_2$).

Таблиця 3.3 – Основні елементи базової таблиці дисперсійного аналізу

позначення	*формула*	*статистична функція*
SSR	$SSR = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 =$	=КВАДРОТКЛ($\hat{Y}$)
SSE	$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n}\left(\hat{y}_i - y\right)^2$	=СУММКВРАЗН($Y;\hat{Y}$)
MSE	$\frac{SSE}{n-2}$	$\frac{SSE}{n-2}$
SST	$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$	=КВАДРОТКЛ(Y)
MST	$\frac{SST}{n-1}$	$\frac{SST}{n-1}$

O13

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	i	Yi	^Y	u^2=(^Y-Y)^2	(Y-_Y)^2	(^Y-_Y)^2	mod(Y-^Y)/Y*100							
2	1	20	28,1493209	66,41143058	5112,25	4013,308547	40,74660433	*n*	16	*SSR*	34703,361	*MSR*	17351,681	
3	2	32	33,9927357	3,970995697	3540,25	3307,085444	6,227299162	*m*	3	*SSE*	1814,6388	*MSE*	139,5876	11,81472
4	3	48	44,8217949	10,10098751	1892,25	2178,854829	6,621260576			*SST*	36518	*MST*	2434,5333	
5	4	65	60,4221734	20,95649679	702,25	965,831309	7,042810223	*k_1*	2					
6	5	45	51,5562041	42,98381274	2162,25	1595,506828	14,56934254	*k_2*	13	проверка	36518			
7	6	64	62,2378561	3,10515126	756,25	856,2730679	2,753349904							
8	7	79	74,9902307	16,07824988	156,25	272,5724825	5,075657348							
9	8	104	113,287129	86,25076753	156,25	474,6789959	8,929931859							
10	9	68	75,1104947	50,55913469	552,25	268,6158844	10,45660983							
11	10	93	89,0154544	15,87660362	2,25	6,172966824	4,284457631							
12	11	117	107,128777	97,44103639	650,25	244,2586818	8,436942426							
13	12	145	159,926815	222,8098157	2862,25	4682,229055	10,2943554							
14	13	91	105,519929	210,8283275	0,25	196,5583989	15,95596553							
15	14	131	108,910766	487,9342559	1560,25	303,1347749	16,86201064							
16	15	167	146,430854	423,0897832	5700,25	3017,398678	12,31685413							
17	16	195	202,499465	56,24197545	10712,25	12320,88123	3,845879493							
18	**sum**	**1464**	**1464**	**1814,6388**	**36518**	**34703,3612**	**174,4193**							
19	**среднее**	**91,50**	**91,50**											
20														

специф | нормализ | **розрах табл AVONA**

Рисунок 3.10 – Розрахунок ANOVA − таблиці

Отже, отримаємо ANOVA − таблицю (табл. 3.4).

Таблиця 3.4 – ANOVA – таблиця

Джерело варіації	Ступені свободи	Сума квадратів	Дисперсії (середні квадрати)
Регресії	$k_1 = m - 1 = 2$	$SSR = 34703{,}3612$	$\hat{\sigma}_r^2 = MSR = 17351{,}681$
Залишків	$k_2 = n - m = 13$	$SSE = 1814{,}63882$	$\hat{\sigma}_u^2 = MSE = 139{,}5876$
Загальної змінної	$n - 1 = 15$	$SST = 36518$	$\hat{\sigma}_y^2 = MST = 2434{,}5333$

Перевіримо розрахунки, відповідно до теореми додавання дисперсій:

$$SSR + SSE = SST,$$

$$34703{,}3612 + 1814{,}63882 = 36518.$$

Чим менша стандартна помилка залишків $\hat{\sigma}_u$ (в нашому випадку $\hat{\sigma}_u = 11{,}81472$), тим краще підібрана функція.

5. Визначення коефіцієнтів детермінації, множинної кореляції та еластичності.

Для подальшої роботи використовуємо результати розрахункової таблиці (рис. 3.9) та ANOVA – таблиці (табл. 3.4).

Коефіцієнт детермінації без урахування числа ступенів свободи:

$$R^2 = \frac{SSR}{SST} = \frac{34703{,}3612}{36518} = 0{,}950308.$$

Перевіримо отримане значення, використовуючи результати розрахунків методу покрокової регресії:

$$R^2 = \hat{\beta}_1 r_{YX_1} + \hat{\beta}_2 r_{YX_1} = 0{,}63177 \cdot 0{,}922164 +$$

$$+0{,}42925 \cdot 0{,}856654 = 0{,}950308.$$

Коефіцієнт детермінації є певною мірою універсальною характеристикою ступеня щільності статистичного зв’язку. Значення коефіцієнта детермінації $R^2 = 0{,}950308$ свідчить про те, що варіація витрат на харчування на 95,03% визначається варіацією загальних затрат і середньої кількості членів сім’ї.

Скорегований коефіцієнт детермінації з урахуванням числа ступенів свободи (формула Тейла):

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-m} = 0{,}9426635$$

або

$$\bar{R}^2 = 1 - \frac{MSE}{MST} = 0{,}9426635.$$

Якщо потрібно порівняти значення коефіцієнтів детермінації в різних моделях, то необхідно коригувати його з урахуванням кількості факторів *X*, які входять в різні моделі, тобто зменшити вплив залежності коефіцієнта детермінації від кількості факторів. Перевага віддається моделі з більшим значенням скоригованого коефіцієнта детермінації.

Коефіцієнт множинної кореляції:

$$R = \sqrt{R^2} = \sqrt{0{,}950308} = 0{,}9748376.$$

Коефіцієнт множинної кореляції є мірою лінійного зв’язку залежної змінної *Y* з незалежними змінними X_1, X_2. Його значення $R = 0{,}9748376$ характеризує достатньо сильний зв’язок між відповідними соціально-економічними показниками (додаток А).

Оскільки коефіцієнт множинної кореляції наближається до одиниці, то це свідчить, що зв'язок між витратами на харчування, загальними затратами та складом сім'ї є дуже тісним.

Частинні коефіцієнти еластичності:

$$E_{Y/X_1} = \frac{\partial \hat{Y}}{\partial X_1} : \frac{\bar{Y}}{\bar{X}_1} = \hat{a}_1 \cdot \frac{\bar{X}_1}{\bar{Y}} = 0{,}16753 \cdot \frac{253{,}875}{91{,}50} = 0{,}4648261,$$

$$E_{Y/X_2} = \frac{\partial \hat{Y}}{\partial X_2} : \frac{\bar{Y}}{\bar{X}_2} = \hat{a}_2 \cdot \frac{\bar{X}_2}{\bar{Y}} = 8{,}17526 \cdot \frac{4{,}9688}{91{,}50} = 0{,}4439433.$$

Обчислені частинні коефіцієнти еластичності показують, що коли загальні витрати збільшуються на 1%, то витрати на харчування підвищуються на 0,4648261 за умови, що решта факторів сталі, а якщо склад сім'ї збільшується на 1%, то витрати на харчування підвищуються на 0,4439433 за умови, що решта факторів сталі.

Загальна еластичність:

$$A = \sum_{i=1}^{n} E_{Y/X_i} = 0{,}4648261 + 0{,}4439433 = 0{,}9087694.$$

Загальна еластичність показує, що коли всі враховані фактори збільшуються одночасно на 1%, то витрати на харчування збільшуються на 0,9087694.

Всі розрахунки зробимо за допомогою статистичних функцій редактора Excel (рис. 3.11).

N16

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	i	Yi	^Y	u^2=(^Y-Y)^2	(Y-_Y)^2	(^Y-_Y)^2	mod(Y-^Y)/Y*100							
2	1	20	28,1493209	66,41143058	5112,25	4013,308547	40,74660433	*n*	16	*SSR*	34703,361	*MSR*	17351,681	
3	2	32	33,9927357	3,970995697	3540,25	3307,085444	6,227299162	*m*	3	*SSE*	1814,6388	*MSE*	139,5876	11,81472
4	3	48	44,8217949	10,10098751	1892,25	2178,854829	6,621260576			*SST*	36518	*MST*	2434,5333	
5	4	65	60,4221734	20,95649679	702,25	965,831309	7,042810223	*k*_1	2					
6	5	45	51,5562041	42,98381274	2162,25	1595,506828	14,56934254	*k*_2	13	проверка	36518			
7	6	64	62,2378561	3,10515126	756,25	856,2730679	2,753349904							
8	7	79	74,9902307	16,07824988	156,25	272,5724825	5,075657348			R^2	0,950308	R	0,9748376	
9	8	104	113,287129	86,25076753	156,25	474,6789959	8,929931859				0,950308			
10	9	68	75,1104947	50,55913469	552,25	268,6158844	10,45660983			скориг R^2	0,9426635			
11	10	93	89,0154544	15,87660362	2,25	6,172966824	4,284457631				0,9426635			
12	11	117	107,128777	97,44103639	650,25	244,2586818	8,436942426							
13	12	145	159,926815	222,8098157	2862,25	4682,229055	10,2943554			E_Y/X_1	0,4648261			
14	13	91	105,519929	210,8283275	0,25	196,5583989	15,95596553			E_Y/X_2	0,4439433			
15	14	131	108,910766	487,9342559	1560,25	303,1347749	16,86201064			A	0,9087694			
16	15	167	146,430854	423,0897832	5700,25	3017,398678	12,31685413							
17	16	195	202,499465	56,24197545	10712,25	12320,88123	3,845879493							
18	**sum**	**1464**	**1464**	**1814,6388**	**36518**	**34703,3612**	**174,4193**							
19	**среднее**	**91,50**	**91,50**											
20														

специф | нормализ | розрах табл AVONA

Рисунок 3.11 – Визначення коефіцієнтів детермінації, множинної кореляції та еластичності

6. *Визначення стандартних помилок оцінок параметрів моделі та їх оцінка*

Використовуючи дисперсію залишків та матрицю помилок, визначимо ***матрицю коваріацій оцінок параметрів моделі*** (дисперсійно-коваріаційну матрицю):

$$\operatorname{cov}\left(\widehat{A}\right)=\widehat{\sigma}_u^2\left(X^T X\right)^{-1}=\begin{Vmatrix} 43{,}0321930 & -0{,}0061681 & -6{,}5895907 \\ -0{,}0061681 & 0{,}0004956 & -0{,}0240828 \\ -6{,}5895907 & -0{,}0240828 & 2{,}5567001 \end{Vmatrix}$$

Діагональні елементи цієї матриці характеризують дисперсії оцінок параметрів моделі: $\widehat{\sigma}_{\widehat{a}_0}^2=43{,}0321930$, $\widehat{\sigma}_{\widehat{a}_1}^2=0{,}0004956$, $\widehat{\sigma}_{\widehat{a}_2}^2=2{,}5567001$. Інші елементи даної матриці визначають рівень коваріації між оцінками параметрів моделі. Знак «мінус» перед оцінками коваріацій $\widehat{\sigma}_{\widehat{a}_j \widehat{a}_k}$ вказує на те, що зі збільшенням однієї оцінки параметрів інша зменшується в середньому та навпаки.

Визначимо стандартні помилки оцінок параметрів, використовуючи матрицю коваріацій:

$$S_{\widehat{a}_0}=\sqrt{\widehat{\sigma}_{\widehat{a}_0}^2}=6{,}559893,$$

$$S_{\widehat{a}_1}=\sqrt{\widehat{\sigma}_{\widehat{a}_1}^2}=0{,}022263,$$

$$S_{\widehat{a}_2}=\sqrt{\widehat{\sigma}_{\widehat{a}_2}^2}=1{,}598968.$$

Стандартні похибки характеризують середні лінійні коливання оцінок параметрів моделі навколо свого математичного сподівання. Чим менші ці похибки, тим стійкіші оцінки параметрів моделі.

Остаточні висновки відносно стійкості оцінок можна зробити лише тоді, коли порівняти їх з абсолютними значеннями оцінок параметрів моделі.

Вибіркова оцінка параметрів $\widehat{A}$ називається *незміщеною*, якщо вона задовольняє рівність $M\left(\widehat{A}\right)=A$.

Незміщеність – це мінімальна вимога, яка ставиться до оцінок параметра $\widehat{A}$. Про наявність зміщенності оцінки можна стверджувати, коли її стандартна похибка перевищує 10% від абсолютного значення оцінки. ***Порівняємо стандартні помилки оцінок параметрів моделі зі значеннями цих оцінок***:

$$\frac{S_{\widehat{a}_0}}{|\widehat{a}_0|}\cdot 100\%=78{,}58\%,$$

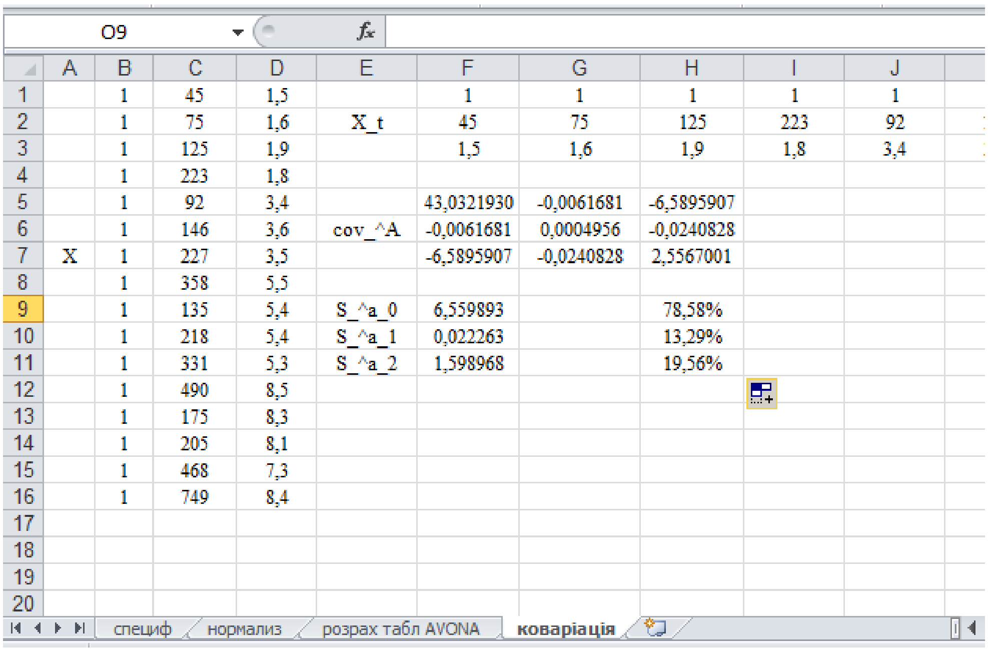
$$\frac{S_{\widehat{a}_1}}{|\widehat{a}_1|}\cdot 100\%=13{,}29\%,$$

$$\frac{S_{\widehat{a}_2}}{|\widehat{a}_2|}\cdot 100\%=19{,}56\%.$$

Так, співвідношення стандартної помилки й абсолютного значення параметра $\widehat{a}_0$ становить 78,58%, параметра $\widehat{a}_1$ – 13,29%, параметра $\widehat{a}_2$ – 19,56%.

Перше й третє співвідношення свідчать про те, що оцінки параметрів моделі $\widehat{a}_0$ і $\widehat{a}_2$ можуть мати зміщення, яке зумовлене невеликою сукупністю спостережень $(n = 16)$, а друге співвідношення показує незначну незміщеність оцінки параметра $\hat{a}_1$.

Всі розрахунки зробимо за допомогою статистичних функцій редактора Excel (рис. 3.12).



	A	B	C	D	E	F	G	H	I	J
1		1	45	1,5		1	1	1	1	1
2		1	75	1,6	X_t	45	75	125	223	92
3		1	125	1,9		1,5	1,6	1,9	1,8	3,4
4		1	223	1,8						
5		1	92	3,4		43,0321930	-0,0061681	-6,5895907		
6		1	146	3,6	cov_^A	-0,0061681	0,0004956	-0,0240828		
7	X	1	227	3,5		-6,5895907	-0,0240828	2,5567001		
8		1	358	5,5						
9		1	135	5,4	S_^a_0	6,559893		78,58%		
10		1	218	5,4	S_^a_1	0,022263		13,29%		
11		1	331	5,3	S_^a_2	1,598968		19,56%		
12		1	490	8,5						
13		1	175	8,3						
14		1	205	8,1						
15		1	468	7,3						
16		1	749	8,4						
17										
18										
19										
20										

Рисунок 3.12 – Матриця коваріації

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

Перевіримо статистичну значущість оцінок параметрів моделі t −статистикою Стьюдента.

Розглянемо гіпотези:

$H_0 : \hat{a}_j = 0 \; (j = 0, 1, 2)$ – оцінка параметра незначуща;

$H_A : \hat{a}_j \neq 0 \; (j = 0, 1, 2)$ – оцінка параметра значуща.

Визначимо фактичні значення t −статистики (рис. 3.13):

$$t_0 = \frac{|\hat{a}_0|}{S_{\hat{a}_0}} = 1{,}2725208, \qquad t_1 = \frac{|\hat{a}_1|}{S_{\hat{a}_1}} = 7{,}5250698, \qquad t_2 = \frac{|\hat{a}_2|}{S_{\hat{a}_2}} = 5{,}1128326.$$

Порівняємо знайдені значення з табличним значенням t− статистики при ступені свободи $k = k_2 = 13$ і рівні значущості $\alpha = 0{,}05$:

$$t_{\text{табл}} = \text{СТЬЮДЕНТ.ОБР.2Х}(0{,}05;13) = 2{,}16037.$$

Оскільки $t_1 > t_{\text{табл}}$ і $t_2 > t_{\text{табл}}$, то гіпотезу H_0 для параметрів $\hat{a}_1$ і $\hat{a}_2$ відхиляємо на користь альтернативної і параметри моделі статично значущі, тобто мають значний вплив на залежну змінну *Y*. Оскільки $t_0 < t_{\text{табл}}$, то гіпотезу H_0 для параметру $\hat{a}_0$ приймаємо і параметр моделі статично незначущий.

M11

	A	B	C	D	E	F	G	H	I	J	K	L
1		1	45	1,5		1	1	1	1	1	1	1
2		1	75	1,6	X_t	45	75	125	223	92	146	227
3		1	125	1,9		1,5	1,6	1,9	1,8	3,4	3,6	3,5
4		1	223	1,8								
5		1	92	3,4		43,0321930	-0,0061681	-6,5895907				
6		1	146	3,6	cov_^A	-0,0061681	0,0004956	-0,0240828				
7	X	1	227	3,5		-6,5895907	-0,0240828	2,5567001				
8		1	358	5,5								
9		1	135	5,4	S_^a_0	6,559893		78,58%		t_^a_0	1,2725208	
10		1	218	5,4	S_^a_1	0,022263		13,29%		t_^a_1	7,5250698	
11		1	331	5,3	S_^a_2	1,598968		19,56%		t_^a_2	5,1128326	
12		1	490	8,5								
13		1	175	8,3								
14		1	205	8,1								
15		1	468	7,3								
16		1	749	8,4								
17												
18												
19												
20												

специф | нормализ | розрах табл AVONA | **коваріація**

Рисунок 3.13 – Фактичні значення t –статистики

Визначимо інтервали надійності для оцінок параметрів моделі за t–розподілом при рівні значущості $\alpha = 0,05$. Оскільки $t_{\text{табл}} = 2,16037$, то відповідно:

– для параметра $\hat{a}_0 = 8,34760$ маємо:

$$\hat{a}_0 - \Delta\hat{a}_0 \le \hat{a}_0 \le \hat{a}_0 + \Delta\hat{a}_0,$$

$$\hat{a}_0 - tS_{\hat{a}_0} \le \hat{a}_0 \le \hat{a}_0 + tS_{\hat{a}_0},$$

$$8,34760 - 2,16037 \cdot 6,559893 \le \hat{a}_0 \le 8,34760 + 2,16037 \cdot 6,559893,$$

$$-5,82419 \le \hat{a}_0 \le 22,51939;$$

– для параметра $\hat{a}_1 = 0,16753$ маємо:

$$\hat{a}_1 - \Delta\hat{a}_1 \le \hat{a}_1 \le \hat{a}_1 + \Delta\hat{a}_1,$$

$$\hat{a}_1 - tS_{\hat{a}_1} \le \hat{a}_1 \le \hat{a}_1 + tS_{\hat{a}_1},$$

$$0,16753 - 2,16037 \cdot 0,022263 \le \hat{a}_1 \le 0,16753 + 2,16037 \cdot 0,022263,$$

$$0,11943 \le \hat{a}_1 \le 0,21563;$$

– для параметра $\hat{a}_2 = 8,17526$ маємо:

$$\hat{a}_2 - \Delta\hat{a}_2 \le \hat{a}_2 \le \hat{a}_2 + \Delta\hat{a}_2,$$

$$\hat{a}_2 - tS_{\hat{a}_2} \le \hat{a}_2 \le \hat{a}_2 + tS_{\hat{a}_2},$$

$$8,17526 - 2,16037 \cdot 1,598968 \le \hat{a}_2 \le 8,17526 + 2,16037 \cdot 1,598968,$$

$$4,72090 \le \hat{a}_2 \le 11,62962;$$

8. Перевірка адекватності економетричної моделі.

Перевіримо за допомогою F−критерію Фішера адекватність економетричної моделі фактичним даним, тобто гіпотезу про значущість зв’язку між незалежними та залежною змінними.

Розглянемо гіпотезу $H_0 : R^2 = 0$ проти альтернативної $H_A : R^2 > 0$. Це рівнозначне тому, що перевіряється значущість водночас усіх параметрів моделі (гіпотеза $H_0 : \hat{a}_0 = \hat{a}_1 = \hat{a}_2 = 0$ проти альтернативної $H_A : \hat{a}_0 \neq 0, \hat{a}_1 \neq 0$, $\hat{a}_2 \neq 0$).

Використовуючи попередні результати, маємо:

$$F_{\text{факт}} = \frac{MSR}{MSE} = 124,30675.$$

Табличне значення для заданого рівня значущості $\alpha = 0,05$ та числа ступенів свободи $k_1 = m - 1 = 2$ і $k_2 = n - 3 = 13$ можна взяти з таблиці (додаток Е): $F_{\text{табл}} = F(0,05;\ 2;\ 13) = 3,81$, або, використавши функцію редактора Excel:

$$F_{\text{табл}} = \text{F.ОБР.ПХ}(\alpha;\ k_1;\ k_2) = 3,805565.$$

Оскільки $F_{\text{факт}} > F_{\text{табл}}$, то нульову гіпотезу відхиляємо і з заданою ймовірністю $p = 0,95$ економетричну модель вважаємо адекватною фактичним даним, тобто гіпотеза про значущість зв'язку між незалежною та залежною змінними підтверджується.

9. *Визначення точкового та інтервальних прогнозів.*

Для визначення точкової оцінки прогнозу підставимо задані прогнозні значення $X_{1\text{прогн}} = 900$ і $X_{2\text{прогн}} = 9,0$ (табл. 1.1) у рівняння моделі $\hat{Y} = 8,34760 + 0,16753 X_1 + 8,17526 X_2$:

$$\hat{Y} = 8,34760 + 0,16753 \cdot 900 + 8,17526 \cdot 9,0 = 232,70159.$$

Значення $\hat{Y}_{\text{прогн}}$ можна розглядати як точкову оцінку прогнозного значення математичного сподівання та індивідуального значення витрат на харчування, коли відомі загальні витрати $X_{1\text{прогн}} = 900$ та розмір сім'ї $X_{2\text{прогн}} = 9,0$.

Визначимо прогнозний інтервал математичного сподівання $M(\hat{Y}_{\text{прогн}})$. Обчислимо дисперсію прогнозу:

$$\hat{\sigma}^2_{\hat{Y}_{\text{прогн}}} = X^T_{\text{прогн}} \operatorname{cov}(\hat{A}) X_{\text{прогн}} = 131,73333.$$

Тоді стандартна похибка прогнозу:

$$\hat{\sigma}_{\hat{Y}_{\text{прогн}}} = \sqrt{\hat{\sigma}^2_{\hat{Y}_{\text{прогн}}}} = \sqrt{131,73333} = 11,477514.$$

Отже, прогнозний інтервал математичного сподівання $M(\hat{Y}_{\text{прогн}})$ при $t_{\text{табл}} = 2,16037$:

$$\hat{Y}_{\text{прогн}} - \Delta\hat{Y}_{\text{прогн}} \leq M(\hat{Y}_{\text{прогн}}) \leq \hat{Y}_{\text{прогн}} + \Delta\hat{Y}_{\text{прогн}};$$

$$\hat{Y}_{\text{прогн}} - t\hat{\sigma}_{\hat{Y}_{\text{прогн}}} \leq M(\hat{Y}_{\text{прогн}}) \leq \hat{Y}_{\text{прогн}} + t\hat{\sigma}_{\hat{Y}_{\text{прогн}}};$$

$$232,70159 - 2,16037 \cdot 11,477514 \le M\left(\widehat{Y}_{\text{прогн}}\right) \le 232,70159 + 2,16037 \cdot 11,477514;$$

$$207,90593 \le M\left(\widehat{Y}_{\text{прогн}}\right) \le 257,49726.$$

Визначимо прогнозний інтервал індивідуального значення $\widehat{Y}_{\text{прогн}}$:

– дисперсія прогнозу індивідуального значення $\widehat{Y}_{\text{прогн}}$:

$$\widehat{\sigma}^2_{\widehat{Y}_{\text{прогн}(i)}} = \widehat{\sigma}^2_u + \widehat{\sigma}^2_{\widehat{Y}_{\text{прогн}}} = 139,5876 + 131,73333 = 271,32093;$$

– стандартна похибка прогнозу індивідуального значення $\widehat{Y}_{\text{прогн}}$:

$$\sigma_{\widehat{Y}_{\text{прогн}(i)}} = \sqrt{\widehat{\sigma}^2_{\widehat{Y}_{\text{прогн}(i)}}} = \sqrt{271,32093} = 16,471822.$$

Тоді ***інтервальний прогноз індивідуального значення*** $\widehat{Y}_{\text{прогн}}$ при $t_{\text{табл}} = 2,16037$ матиме вигляд:

$$\widehat{Y}_{\text{прогн}} - \Delta\widehat{Y}_{\text{прогн}(i)} \le \widehat{Y}_{\text{прогн}} \le \widehat{Y}_{\text{прогн}} + \Delta\widehat{Y}_{\text{прогн}(i)};$$

$$\widehat{Y}_{\text{прогн}} - t\sigma_{\widehat{Y}_{\text{прогн}(i)}} \le \widehat{Y}_{\text{прогн}} \le \widehat{Y}_{\text{прогн}} + t\sigma_{\widehat{Y}_{\text{прогн}(i)}};$$

$$232,70159 - 2,16037 \cdot 16,471822 \le \widehat{Y}_{\text{прогн}} \le 232,70159 + 2,16037 \cdot 16,471822;$$

$$197,11639 \le \widehat{Y}_{\text{прогн}} \le 268,2868.$$

Отже, з ймовірністю $p = 0,95$ $(\alpha = 0,05)$ прогноз математичного сподівання $M\left(\widehat{Y}_{\text{прогн}}\right)$ потрапляє в інтервал $\left[207,90593; 257,49726\right]$, а прогноз індивідуального значення – в інтервал $\left[197,11639;\ 268,2868\right]$.

Економічна інтерпретація: якщо у прогнозному періоді загальні витрати мають рівень 900 одиниць, а розмір сім’ї дорівнює 9, то середні витрати на харчування потрапляють в інтервал:

$$207,90593 \le M\left(\widehat{Y}_{\text{прогн}}\right) \le 257,49726.$$

Водночас окреме (індивідуальне) значення витрат на харчування містяться в ширшому інтервалі:

$$197,11639 \le \widehat{Y}_{\text{прогн}} \le 268,2868.$$

10. Побудова графіків фактичних даних та лінії регресії.

Для візуального порівняння фактичних та регресійних значень залежної змінної, за допомогою Мастердиаграмм, побудуємо графіки фактичних даних та лінію регресії (рис. 3.14).

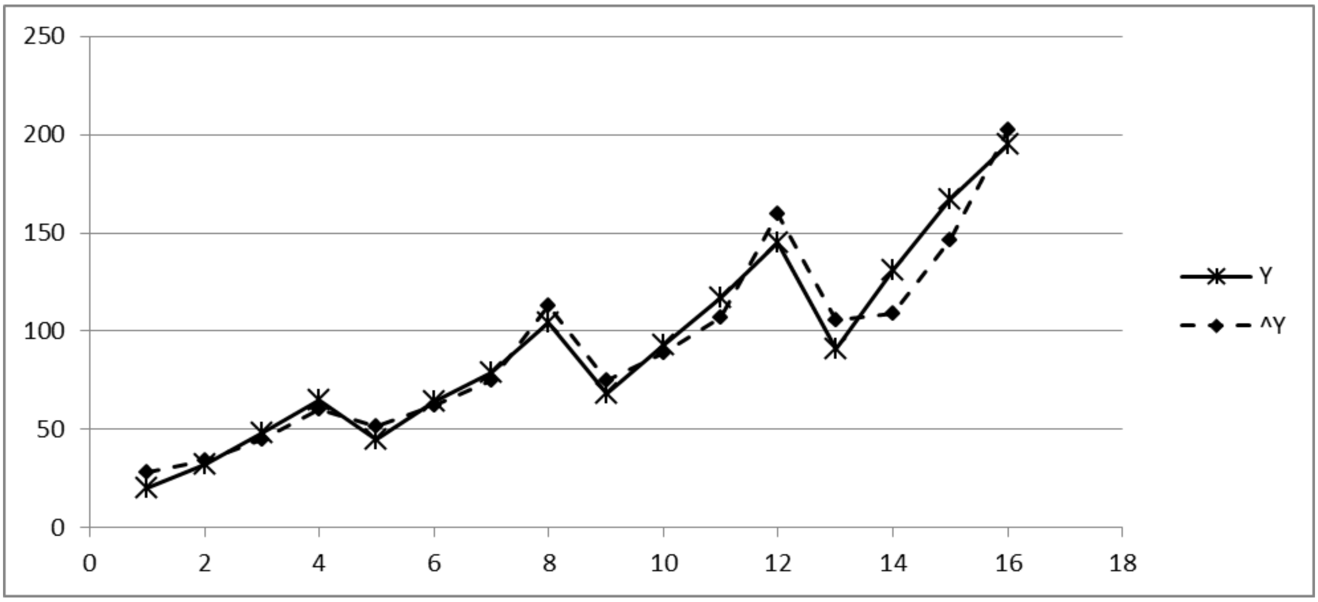


Рисунок 3.14 – Графіки фактичних даних та лінії регресії

11. Перевірка точності економетричної моделі за допомогою середньої відносної похибки апроксимації.

Визначимо ***абсолютну середню відносну похибку апроксимації*** $\bar{\varepsilon}$. Використовуючи розрахунки (рис. 3.11), маємо

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% = \frac{1}{16} \cdot 174{,}4193 = 10{,}90\% < 15\%,$$

що свідчить про досить добру якість моделі (додаток Г).

При порівнянні різних моделей, побудованих на одній статистичній базі, перевагу надають моделям з меншою похибкою апроксимації $\bar{\varepsilon}$.

12. Побудова економетричної моделі за допомогою функції ЛИНЕЙН *та інструменту* ДАННЫЕ→АНАЛИЗ ДАННЫХ→РЕГРЕССИЯ.

Перевіримо наші результати, застосувавши функцію ЛИНЕЙН (Y;X;1;1) до вихідних даних (рис. 3.15, табл. 3.5).

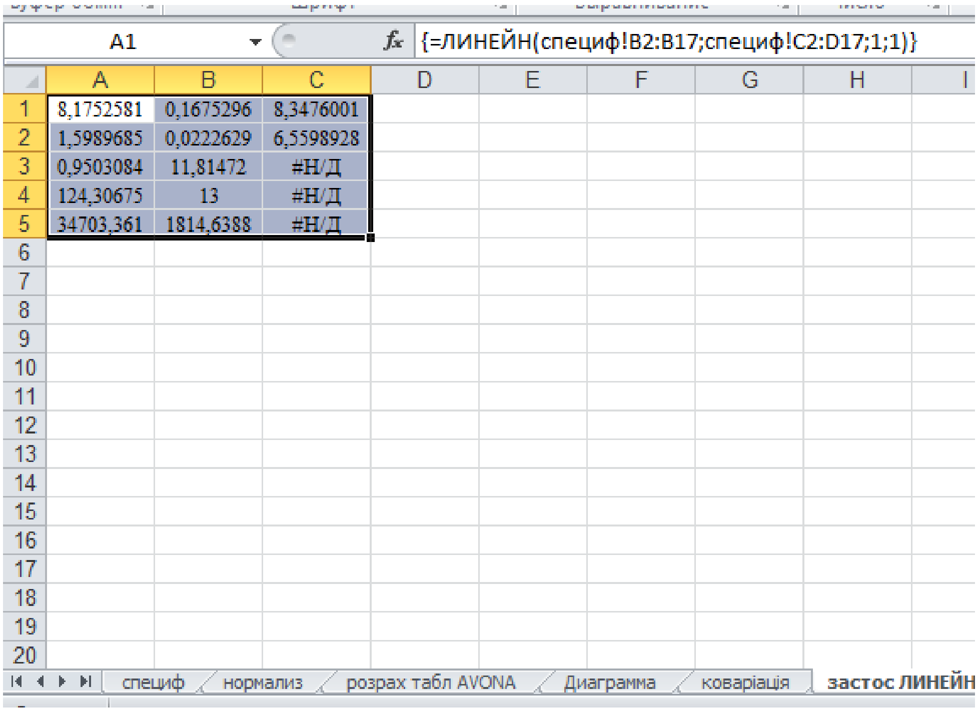


Рисунок 3.15 – Побудова економетричної моделі за допомогою функції ЛИНЕЙН

Таблиця 3.5 – Побудова економетричної моделі за допомогою функції ЛИНЕЙН

$\hat{a}_2$	$\hat{a}_1$	$\hat{a}_0$
$S_{\hat{a}_2}$	$S_{\hat{a}_1}$	$S_{\hat{a}_0}$
R^2	σ_u	
F	$k = n - m$	
$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$	$SSE = \sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$	

8,17526	0,16753	8,34760
1,59897	0,02226	6,55989
0,95031	11,81472	#Н/Д
124,30675	13	#Н/Д
34703,36118	1814,63882	#Н/Д

Отже, економетрична модель витрат на харчування має вигляд:

$$\hat{Y} = 8{,}34760 + 0{,}16753 X_1 + 8{,}17526 X_2 .$$

Але найбільш повне та ґрунтовне дослідження моделі можна отримати за допомогою інструменту ДАННЫЕ→АНАЛИЗ ДАННЫХ→РЕГРЕССИЯ. Розглянемо порядок побудови моделі за допомогою цього інструменту:

1. На панелі інструментів вибираємо ДАННЫЕ→АНАЛИЗ ДАННЫХ (рис. 3.16).
2. У вікні ***Анализ данных*** серед інструментів аналізу вибираємо ***Регрессия***. Натискуємо клавішу ОК (рис. 3.16).

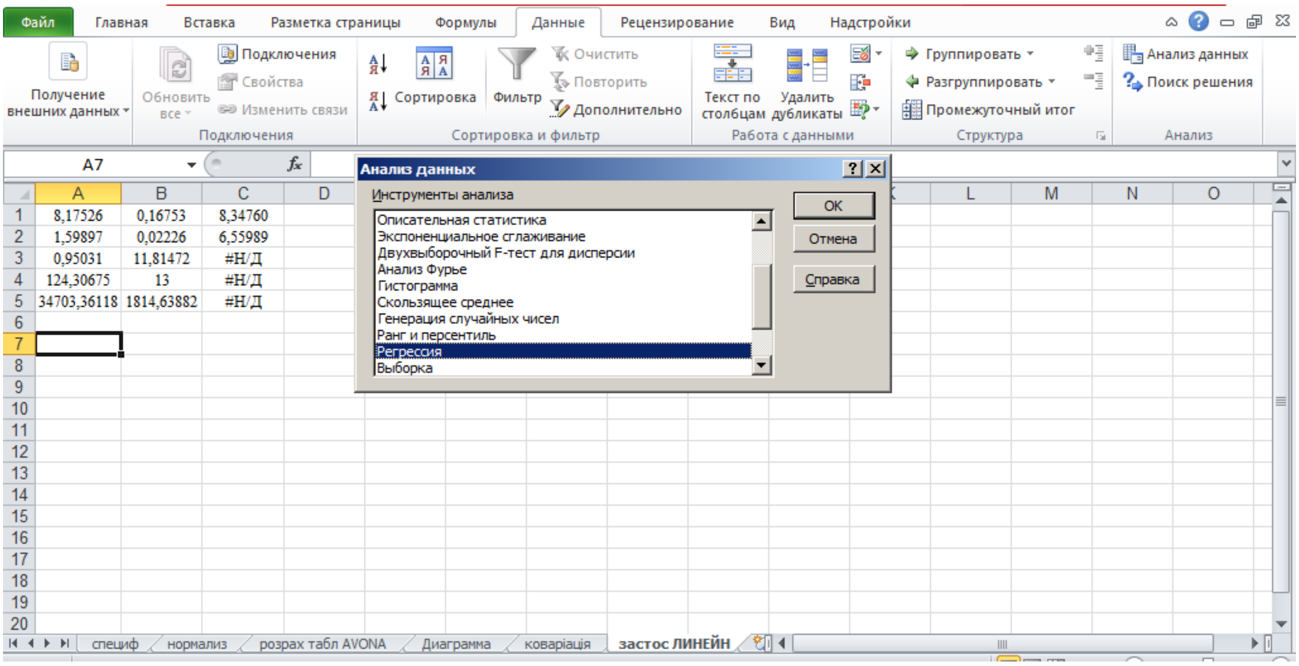


Рисунок 3.16 – Вибір ДАННЫЕ→АНАЛИЗ ДАННЫХ

3. У відкритому діалоговому вікні вказуємо діапазони вихідних даних разом з заголовками E1:E17 та F1:G17, вибираємо режим Метки, рівень надійності 95%, вказуємо виведення залишків та параметри виводу (новий робочий лист ***Регрессия***). Натискуємо клавішу ОК (рис. 3.17).

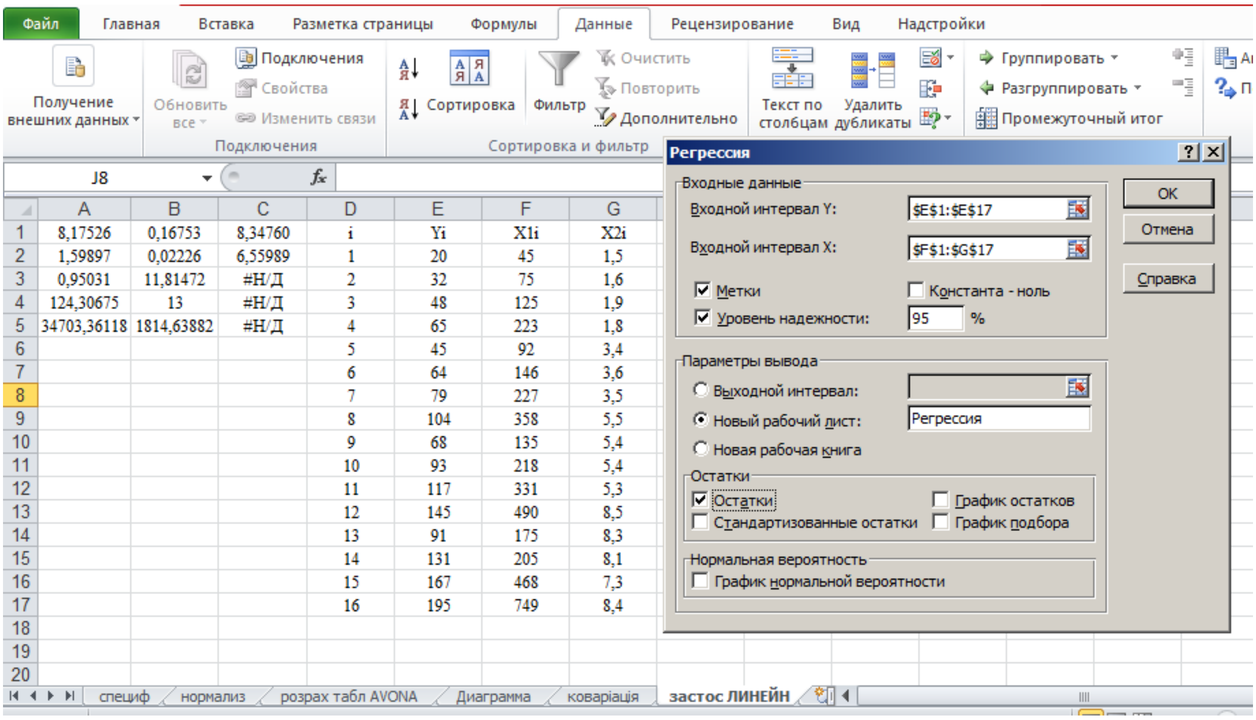


Рисунок 3.17 – Інструмент РЕГРЕССИЯ

4. В результаті отримаємо оцінки параметрів регресії та її детальний аналіз (табл. 3.18).

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественны	0,974837614
R-квадрат	0,950308373
Нормированны	0,942663508
Стандартная оц	11,81471971
Наблюдения	16

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	34703,36118	17351,68059	124,3067461	3,35615E-09
Остаток	13	1814,638824	139,5876019		
Итого	15	36518			

	Коэффициенты	*Стандартная оши*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	8,347600109	6,559892761	1,272520819	0,225477464	-5,824186602	22,51938682	-5,824186602	22,51938682
X1i	0,167529635	0,022262868	7,525069768	4,34162E-06	0,119433633	0,215625637	0,119433633	0,215625637
X2i	8,175258118	1,598968458	5,112832637	0,000199212	4,720896779	11,62961946	4,720896779	11,62961946

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*
1	28,14932087	-8,149320866
2	33,99273573	-1,992735732
3	44,82179492	3,178205076
4	60,42217335	4,577826645
5	51,55620414	-6,556204141
6	62,23785606	1,762143938
7	74,9902307	4,009769305
8	113,2871291	-9,287129133
9	75,11049469	-7,110494687
10	89,0154544	3,984545597
11	107,1287774	9,871222639
12	159,9268153	-14,92681532
13	105,5199286	-14,51992863
14	108,9107661	22,08923394
15	146,4308536	20,56914639
16	202,499465	-7,499465011

розрах табл AVONA / Диаграмма / коваріація / Регрессия / застос ЛИНЕЙН

Готово

Рисунок 3.18 – Оцінки параметрів регресії та її детальний аналіз за допомогою інструменту РЕГРЕССИЯ

Побудова економетричної моделі на основі стандартної процедури РЕГРЕССИЯ дає найбільшу кількість характеристик взаємозв'язку змінних моделі, що дає можливість перевірити попередні наші дослідження. Розглянемо результати.

Регресійна статистика (рис. 3.18).

$R = 0,9748376$ – множинний коефіцієнт кореляції;

$R^2 = 0,950308$ – коефіцієнт детермінації без урахування числа ступенів свободи;

$\bar{R}^2 = 0,942664$ – коефіцієнт детермінації з урахуванням числа ступенів свободи (формула Амемія);

$S_u = 11,81472$ – стандартна похибка (виправлене середнє квадратичне відхилення залишків);

$n = 16$ – кількість спостережень.

Дисперсійний аналіз містить 5 стовпчиків (рис. 3.18).

Перший df – ступені вільності для суми квадратів відхилень:
$k_1 = m - 1 = 2$ – регресії;
$k_2 = n - m = 13$ – залишків;
$k = n - 1 = 15$ – залежної змінної.

Другий SS – суми квадратів відхилень:

$SSR = \sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y} \right)^2 = 34703{,}36118$ – регресії;

$SSE = \sum_{i=1}^{n} \left(\widehat{Y}_i - Y_i \right)^2 = 1814{,}638824$ – залишків;

$SST = \sum_{i=1}^{n} \left(Y_i - \overline{Y} \right)^2 = 36518$ – залежної змінної.

Третій MS – виправлені дисперсії:

$MSR = \widehat{\sigma}_p^2 = \dfrac{\sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y} \right)^2}{m-1} = 17351{,}68059$ – регресії;

$MSE = \widehat{\sigma}_u^2 = \dfrac{\sum_{i=1}^{n} \left(\widehat{Y}_i - Y_i \right)^2}{n-m} = 139{,}5876019$ – залишків.

Четвертий – $F-$ критерій Фішера з 95% довіри:

$$F_{\text{факт}} = \frac{MSR}{MSE} = 124{,}3067461.$$

П'ятий – рівень значущості $F-$ критерію

$$\alpha = 3{,}36 \cdot E - 0{,}9 = 3{,}35615E - 09.$$

Оскільки даний показник менше 0,05, то побудована регресійна модель адекватна фактичним даним.

Оцінки параметрів моделі та їх значущість (рис. 3.18). Цей блок результатів містить 9 стовпчиків.

Перший і *другий* – назва та рівень оцінок параметрів моделі:
Y – переріз – $\widehat{a}_0 = 8{,}347600109$;
змінна X_1 – $\widehat{a}_1 = 0{,}167529635$;
змінна X_2 – $\widehat{a}_2 = 8{,}175258118$.

Третій стовпець – стандартні похибки оцінок параметрів моделі:
$S_{\widehat{a}_0} = 6{,}559892761$;
$S_{\widehat{a}_1} = 0{,}022262868$;
$S_{\widehat{a}_2} = 1{,}598968458$.

Четвертий стовпець – t –статистика:
$t_{\widehat{a}_0} = 1{,}272520819$;
$t_{\widehat{a}_1} = 7{,}525069768$;

$t_{\hat{a}_2} = 5{,}112832637$.

П'ятий стовпець – рівень значущості:

$\alpha_0 = 0{,}225477464$;

$\alpha_1 = 4{,}34162E - 06$;

$\alpha_2 = 0{,}000199212$.

Якщо рівень значущості менший за 0,05, то з ймовірністю 0, 95 можна стверджувати, що оцінені параметри – достовірні. Звідси параметри $\hat{a}_1$ і $\hat{a}_2$ – достовірні, а параметр $\hat{a}_0$ – недостовірний.

Шостий та *сьомий* стовпці – з імовірністю 0,95 визначають інтервали надійності оцінок параметрів моделі. Якщо число нуль не попадає в жодний надійний інтервал, то з 95% впевненістю можна стверджувати, що отримана модель придатна для даного економічного процесу.

Виведення залишків (рис. 3.18).

У цьому блоці результатів виводяться розрахункові значення змінної та залишків, які визначаються як відхилення розрахункових значень залежної змінної від фактичних.

Результати, отримані за допомогою інструменту ДАННЫЕ→АНАЛИЗ ДАННЫХ→РЕГРЕССИЯ, повністю підтвердили раніше проведені дослідження та доповнили їх.

13. Економіко-математичний аналіз характеристик економетричної моделі.

Модель отримана на основі нормалізованих даних методом покрокової регресії:

$$\hat{Y}^* = 0{,}63177 X_1^* + 0{,}42925 X_2^*.$$

1. Перш за все звернемо увагу на відсутність вільного члена в даній економетричній моделі. Це пов'язано з тим, що всі змінні нормалізовані та мають одну й ту саму одиницю виміру.

2. Параметри рівняння характеризують граничну зміну залежної змінної, якщо незалежна збільшиться на величину свого середньоквадратичного відхилення σ_{X_j}.

Так, якщо X_1^* збільшиться на σ_{X_1}, то Y^* – на $0{,}63177\sigma_Y$ при незмінній величині X_2^*; якщо X_2^* збільшиться на σ_{X_2}, то Y^* – на $0{,}42925\sigma_Y$ при незмінній величині фактора X_1^*.

3. Оскільки всі змінні нормалізовані (мають однакові одиниці вимірювання – стандартні відхилення), то оцінки параметрів $\hat{\beta}_j$ першої економетричної моделі показують порівняльну силу впливу кожної незалежної змінної на залежну: чим більше за модулем значення оцінки параметра $\hat{\beta}_j$, тим сильніше впливає j -та змінна на результат.

Так як $\widehat{\beta}_1 = |0,631766| > \widehat{\beta}_2 = |0,429247|$, то загальні витрати сильніше впливають на витрати на харчування, ніж склад сім'ї.

Модель отримана на основі покрокової регресії в абсолютних значеннях:

$$\widehat{Y} = 8,34760 + 0,16753 X_1 + 8,17526 X_2 .$$

1. В даній економетричній моделі, яка характеризує зв'язок витрат на харчування з загальними витратами та складом сім'ї, коли кожен економічний показник має свою початкову одиницю виміру, є вільний член. Його рівень залежить від початку відліку змінних, а також від одиниць виміру кожної змінної моделі.

2. Коефіцієнт детермінації $R^2 = 0,950308$ свідчить, що на 95,03% варіація витрат на харчування залежить від варіації загальних витрат і складу сім'ї.

3. Коефіцієнт множинної кореляції $r = 0,9748376$ свідчить про досить сильний зв'язок між витратами на харчування, загальними затратами та складом сім'ї (додаток А).

4. Оцінка параметра $\widehat{a}_1$ характеризує граничну зміну величини витрат на харчування залежно від зміни загальних затрат на одиницю. Тобто, коли за всіх однакових умов загальні затрати сім'ї зростуть на одиницю, то витрати на харчування в них збільшаться на 0,1675 одиниці при незмінному складі сім'ї.

Оцінка $\widehat{a}_2$ характеризує граничне зростання витрат на харчування при збільшенні сім'ї на одного члена. Якщо склад сім'ї збагатиться ще одним членом, то витрати на харчування зростуть на 8,1722 одиниці при незмінній величині доходу.

5. Сумарна еластичність показує, що коли всі враховані фактори збільшуються одночасно на 1%, то витрати на харчування збільшуються на 0,908769.

6. Згідно з $F-$ критерієм, з надійністю 0,95 економетричну модель $\widehat{Y} = 8,34760 + 0,16753 X_1 + 8,17526 X_2$ можна вважати адекватною фактичним експериментальним даним і на підставі прийнятої моделі проводити економічний аналіз та знаходити значення прогнозу.

7. Оскільки $\overline{\varepsilon} = 10,90\% < 15\%$, це свідчить про досить добру якість прогнозу (додаток Г).