

Парная нелинейная регрессия

Теория и пример решения в Excel.

Данный способ применяется, в основном, в экономическом моделировании и прогнозировании. Его цель – пронаблюдать и выявить зависимости между двумя показателями.

Основными типами нелинейных регрессий являются:

- полиномиальные (квадратичная, кубическая);
- гиперболическая;
- степенная;
- показательная;
- логарифмическая.

Также могут применяться различные комбинации. Например, для аналитики временных рядов в банковской сфере, страховании, демографических исследованиях используют кривую Гомпцера, которая является разновидностью логарифмической регрессии.

В прогнозировании с помощью нелинейных регрессий главное выяснить коэффициент корреляции, который покажет нам есть ли тесная взаимосвязь между двумя параметрами или нет. Как правило, если коэффициент корреляции близок к 1, значит связь есть, и прогноз будет довольно точен. Ещё одним важным элементом нелинейных регрессий является средняя относительная ошибка (A), если она находится в промежутке $<8...10\%$, значит модель достаточно точна.

На этом, пожалуй, теоретический блок мы закончим и перейдём к практическим вычислениям.

У нас имеется таблица продаж автомобилей за промежуток 15 лет (обозначим его X), количество шагов измерений будет аргумент n , также имеется выручка за эти периоды (обозначим её Y), нам нужно спрогнозировать какова будет выручка в дальнейшем. Построим следующую таблицу (см. рис. 1).

Для исследования нам потребуется решить уравнение (зависимости Y от X): $y=ax^2+bx+c+e$. Это парная квадратичная регрессия. Применим в этом случае метод наименьших квадратов, для выяснения неизвестных аргументов - a, b, c. Он приведёт к системе алгебраических уравнений вида:

$$\begin{cases} \hat{a} \cdot \sum x^4 + \hat{b} \cdot \sum x^3 + \hat{c} \cdot \sum x^2 = \sum x^2 y \\ \hat{a} \cdot \sum x^3 + \hat{b} \cdot \sum x^2 + \hat{c} \cdot \sum x = \sum x \hat{y} \\ \hat{a} \cdot \sum x^2 + \hat{b} \cdot \sum x + \hat{c} \cdot n = \sum \hat{y} \end{cases}$$

	A	B	C
1	№	x	y
2	1	1,00	7,00
3	2	2,00	12,00
4	3	3,00	11,00
5	4	4,00	16,00
6	5	5,00	17,00
7	6	6,00	15,00
8	7	7,00	19,00
9	8	8,00	18,00
10	9	9,00	20,00
11	10	10,00	19,00
12	11	11,00	21,00
13	12	12,00	23,00
14	13	13,00	21,00
15	14	14,00	23,00
16	15	15,00	22,00
17	Сумм.		
18	Средн.		
19	n =	15	

Рис. 1

Для решения этой системы воспользуемся, к примеру, методом Крамера. Видим, что входящие в систему суммы являются коэффициентами при неизвестных. Для их вычисления добавим в таблицу несколько столбцов (D,E,F,G,H) и подпишем соответственно смыслу вычислений - в столбце D возведём x в квадрат, в E в куб, в F в 4 степень, в G перемножим показатели x и y, в H возведём x в квадрат и перемножим с y (см. рис. 2).

	A	B	C	D	E	F	G	H
1	№	x	y	x ²	x ³	x ⁴	x*y	x ² *y
2	1	1,00	7,00					
3	2	2,00	12,00					
4	3	3,00	11,00					
5	4	4,00	16,00					
6	5	5,00	17,00					
7	6	6,00	15,00					
8	7	7,00	19,00					
9	8	8,00	18,00					
10	9	9,00	20,00					
11	10	10,00	19,00					
12	11	11,00	21,00					
13	12	12,00	23,00					
14	13	13,00	21,00					
15	14	14,00	23,00					
16	15	15,00	22,00					
17	Сумм.							
18	Средн.							

Рис. 2

Получится заполненная нужными для решения уравнения таблица вида (см. рис. 3).

	A	B	C	D	E	F	G	H	I
1	№	x	y	x ²	x ³	x ⁴	x*y	x ² *y	y ²
2	1	1,00	7,00	1	1	1	7	7	49
3	2	2,00	12,00	4	8	16	24	48	144
4	3	3,00	11,00	9	27	81	33	99	121
5	4	4,00	16,00	16	64	256	64	256	256
6	5	5,00	17,00	25	125	625	85	425	289
7	6	6,00	15,00	36	216	1296	90	540	225
8	7	7,00	19,00	49	343	2401	133	931	361
9	8	8,00	18,00	64	512	4096	144	1152	324
10	9	9,00	20,00	81	729	6561	180	1620	400
11	10	10,00	19,00	100	1000	10000	190	1900	361
12	11	11,00	21,00	121	1331	14641	231	2541	441
13	12	12,00	23,00	144	1728	20736	276	3312	529
14	13	13,00	21,00	169	2197	28561	273	3549	441
15	14	14,00	23,00	196	2744	38416	322	4508	529
16	15	15,00	22,00	225	3375	50625	330	4950	484
17	Сумм.								
18	Средн.								

Рис. 3

Далее посчитаем суммы по каждому столбцу – воспользуемся Σ в программе Excel. (см. рис. 4)

	A	B	C	D	E	F	G	H	I
1	№	x	y	x2	x3	x4	x*y	x2*y	y2
2	1	1,00	7,00	1	1	1	7	7	49
3	2	2,00	12,00	4	8	16	24	48	144
4	3	3,00	11,00	9	27	81	33	99	121
5	4	4,00	16,00	16	64	256	64	256	256
6	5	5,00	17,00	25	125	625	85	425	289
7	6	6,00	15,00	36	216	1296	90	540	225
8	7	7,00	19,00	49	343	2401	133	931	361
9	8	8,00	18,00	64	512	4096	144	1152	324
10	9	9,00	20,00	81	729	6561	180	1620	400
11	10	10,00	19,00	100	1000	10000	190	1900	361
12	11	11,00	21,00	121	1331	14641	231	2541	441
13	12	12,00	23,00	144	1728	20736	276	3312	529
14	13	13,00	21,00	169	2197	28561	273	3549	441
15	14	14,00	23,00	196	2744	38416	322	4508	529
16	15	15,00	22,00	225	3375	50625	330	4950	484
17	Сумм.	120,00	264,00	1240,00	14400,00	178312,00	2382,00	25838,00	4954,00

Рис. 4

Сформируем матрицу A системы, состоящую из коэффициентов при неизвестных в левых частях уравнений. Поместим её в ячейку A22 и назовём " A ". Следуем той системе уравнений, которую мы избрали для решения регрессии.

	178312,00	14400,00	1240,00
A=	14400,00	1240,00	120,00
	1240,00	120,00	15

То есть, в ячейку B21 мы должны поместить сумму столбца, где возводили показатель X в четвёртую степень - F17. Просто сошлёмся на ячейку - " $=F17$ ". Далее нам необходима сумма столбца где возводили X в куб - E17, далее идём строго по системе. Таким образом, нам необходимо будет заполнить всю матрицу.

В соответствии с алгоритмом Крамера наберём матрицу $A1$, подобную A , в которой вместо элементов первого столбца должны размещаться

элементы правых частей уравнений системы. То есть сумма столбца X в квадрате умноженная на Y, сумма столбца XY и сумма столбца Y.

	25838,00	14400,00	1240,00
A1=	2382,00	1240,00	120,00
	264,00	120,00	15

Также нам понадобятся ещё две матрицы - назовём их A2 и A3 в которых второй и третий столбцы будут состоять из коэффициентов правых частей уравнений. Картина будет такова.

21		178312,00	14400,00	1240,00		178312,00	25838,00	1240,00
22	A=	14400,00	1240,00	120,00	A2=	14400,00	2382,00	120,00
23		1240,00	120,00	15		1240,00	264,00	15
24								
25		25838,00	14400,00	1240,00		178312,00	14400,00	25838,00
26	A1=	2382,00	1240,00	120,00	A3=	14400,00	1240,00	2382,00
27		264,00	120,00	15		1240,00	120,00	264,00

Следуя избранному алгоритму, нам нужно будет вычислить значения определителей (детерминантов, D) полученных матриц. Воспользуемся формулой МОПРЕД. Результаты разместим в ячейках J21:K24.

21		178312,00	14400,00	1240,00		178312,00	25838,00	1240,00		D=	17326400
22	A=	14400,00	1240,00	120,00	A2=	14400,00	2382,00	120,00		D1=	-1285200
23		1240,00	120,00	15		1240,00	264,00	15		D2=	37270800
24										D3=	113021440
25		25838,00	14400,00	1240,00		178312,00	14400,00	25838,00			
26	A1=	2382,00	1240,00	120,00	A3=	14400,00	1240,00	2382,00			
27		264,00	120,00	15		1240,00	120,00	264,00			

Расчёт коэффициентов уравнения по Крамеру будем производить в ячейках напротив соответствующих детерминантов по формуле: **a** (в ячейке M22) - " $=K22/K21$ "; **b** (в ячейке M23) - " $=K23/K21$ "; **c** (в ячейке M24) - " $=K24/K21$ ".

21		178312,00	14400,00	1240,00		178312,00	25838,00	1240,00		D=	17326400		
22	A=	14400,00	1240,00	120,00	A2=	14400,00	2382,00	120,00		D1=	-1285200	a=	-0,074
23		1240,00	120,00	15		1240,00	264,00	15		D2=	37270800	b=	2,151
24										D3=	113021440	c=	6,523
25		25838,00	14400,00	1240,00		178312,00	14400,00	25838,00					
26	A1=	2382,00	1240,00	120,00	A3=	14400,00	1240,00	2382,00					
27		264,00	120,00	15		1240,00	120,00	264,00					

Получим наше искомое уравнение парной квадратичной регрессии:

$$y = -0,074x^2 + 2,151x + 6,523$$

Оценим тесноту линейной связи индексом корреляции.

$$\rho = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Для вычисления добавим в таблицу дополнительный столбец J (назовём его y^*). Расчёта будет следующей (согласно полученному нами уравнению регрессии) - " $m \cdot B^2 + M \cdot B + M^2$ ". Поместим её в ячейку J2. Останется протянуть вниз маркер автозаполнения до ячейки J16.

	A	B	C	D	E	F	G	H	I	J
1	№	x	y	x ²	x ³	x ⁴	x*y	x ² *y	y ²	y*
2	1	1,00	7,00	1	1	1	7	7	49	8,60
3	2	2,00	12,00	4	8	16	24	48	144	10,53
4	3	3,00	11,00	9	27	81	33	99	121	12,31
5	4	4,00	16,00	16	64	256	64	256	256	13,94
6	5	5,00	17,00	25	125	625	85	425	289	15,42
7	6	6,00	15,00	36	216	1296	90	540	225	16,76
8	7	7,00	19,00	49	343	2401	133	931	361	17,95
9	8	8,00	18,00	64	512	4096	144	1152	324	18,98
10	9	9,00	20,00	81	729	6561	180	1620	400	19,87
11	10	10,00	19,00	100	1000	10000	190	1900	361	20,62
12	11	11,00	21,00	121	1331	14641	231	2541	441	21,21
13	12	12,00	23,00	144	1728	20736	276	3312	529	21,65
14	13	13,00	21,00	169	2197	28561	273	3549	441	21,95
15	14	14,00	23,00	196	2744	38416	322	4508	529	22,10
16	15	15,00	22,00	225	3375	50625	330	4950	484	22,10
17	Сумм.	120,00	264,00	1240,00	14400,00	178312,00	2382,00	25838,00	4954,00	264,00
18	Средн.	8,00	17,60							

Для вычисления сумм $(Y - \bar{Y})^2$ добавим в таблицу столбцы K и L с соответствующими формулами. Среднее по столбцу Y посчитаем с помощью функции СРЗНАЧ.

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	x	y	x ²	x ³	x ⁴	x*y	x ² *y	y ²	y*	(y-y*) ²	(y-ycp.) ²
2	1	1,00	7,00	1	1	1	7	7	49	8,60	2,56	112,36
3	2	2,00	12,00	4	8	16	24	48	144	10,53	2,17	31,36
4	3	3,00	11,00	9	27	81	33	99	121	12,31	1,71	43,56
5	4	4,00	16,00	16	64	256	64	256	256	13,94	4,24	2,56
6	5	5,00	17,00	25	125	625	85	425	289	15,42	2,48	0,36
7	6	6,00	15,00	36	216	1296	90	540	225	16,76	3,10	6,76
8	7	7,00	19,00	49	343	2401	133	931	361	17,95	1,11	1,96
9	8	8,00	18,00	64	512	4096	144	1152	324	18,98	0,97	0,16
10	9	9,00	20,00	81	729	6561	180	1620	400	19,87	0,02	5,76
11	10	10,00	19,00	100	1000	10000	190	1900	361	20,62	2,61	1,96
12	11	11,00	21,00	121	1331	14641	231	2541	441	21,21	0,04	11,56
13	12	12,00	23,00	144	1728	20736	276	3312	529	21,65	1,81	29,16
14	13	13,00	21,00	169	2197	28561	273	3549	441	21,95	0,91	11,56
15	14	14,00	23,00	196	2744	38416	322	4508	529	22,10	0,81	29,16
16	15	15,00	22,00	225	3375	50625	330	4950	484	22,10	0,01	19,36
17	Сумм.	120,00	264,00	1240,00	14400,00	178312,00	2382,00	25838,00	4954,00	264,00	24,55	307,60
18	Средн.	8,00	17,60									

В ячейке K25 разместим формулу подсчёта индекса корреляции - " $\text{=КОРЕНЬ}(1-(K17/L17))$ ".

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	№	x	y	x2	x3	x4	x*y	x2*y	y2	y*	(y-y*)^2	(y-ycp.)^2	abs(A)
2	1	1,00	7,00	1	1	1	7	7	49	8,60	2,56	112,36	22,86%
3	2	2,00	12,00	4	8	16	24	48	144	10,53	2,17	31,36	12,26%
4	3	3,00	11,00	9	27	81	33	99	121	12,31	1,71	43,56	11,90%
5	4	4,00	16,00	16	64	256	64	256	256	13,94	4,24	2,56	12,87%
6	5	5,00	17,00	25	125	625	85	425	289	15,42	2,48	0,36	9,27%
7	6	6,00	15,00	36	216	1296	90	540	225	16,76	3,10	6,76	11,73%
8	7	7,00	19,00	49	343	2401	133	931	361	17,95	1,11	1,96	5,55%
9	8	8,00	18,00	64	512	4096	144	1152	324	18,98	0,97	0,16	5,47%
10	9	9,00	20,00	81	729	6561	180	1620	400	19,87	0,02	5,76	0,63%
11	10	10,00	19,00	100	1000	10000	190	1900	361	20,62	2,61	1,96	8,51%
12	11	11,00	21,00	121	1331	14641	231	2541	441	21,21	0,04	11,56	1,00%
13	12	12,00	23,00	144	1728	20736	276	3312	529	21,65	1,81	29,16	5,85%
14	13	13,00	21,00	169	2197	28561	273	3549	441	21,95	0,91	11,56	4,53%
15	14	14,00	23,00	196	2744	38416	322	4508	529	22,10	0,81	29,16	3,91%
16	15	15,00	22,00	225	3375	50625	330	4950	484	22,10	0,01	19,36	0,45%
17	Сумм.	120,00	264,00	1240,00	14400,00	178312,00	2382,00	25838,00	4954,00	264,00	24,55	307,60	
18	Средн.	8,00	17,60										7,79%
19	n =	15											
20	Квадратичная регрессия: $y=a*x^2+b*x+c$												
21		178312,00	14400,00	1240,00		178312,00	25838,00	1240,00		D=	17326400		
22	A=	14400,00	1240,00	120,00	A2=	14400,00	2382,00	120,00		D1=	-1285200	a=	-0,074
23		1240,00	120,00	15		1240,00	264,00	15		D2=	37270800	b=	2,151
24										D3=	113021440	c=	6,523
25		25838,00	14400,00	1240,00		178312,00	14400,00	25838,00		ro=	0,959		
26	A1=	2382,00	1240,00	120,00	A3=	14400,00	1240,00	2382,00					
27		264,00	120,00	15		1240,00	120,00	264,00					

Видим, что значение 0,959 очень близко к 1, значит между продажами и годами есть тесная нелинейная связь.

Осталось оценить качество подгонки полученного квадратичного уравнения регрессии (индекс детерминации). Он рассчитывается по формуле

квадрата индекса корреляции. То есть формула в ячейке K26 будет очень проста - " $=K25*K25$ ".

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	№	x	y	x ²	x ³	x ⁴	x*y	x ² *y	y ²	y*	(y-y*) ²	(y-ycp.) ²	abs(A)
2	1	1,00	7,00	1	1	1	7	7	49	8,60	2,56	112,36	22,86%
3	2	2,00	12,00	4	8	16	24	48	144	10,53	2,17	31,36	12,26%
4	3	3,00	11,00	9	27	81	33	99	121	12,31	1,71	43,56	11,90%
5	4	4,00	16,00	16	64	256	64	256	256	13,94	4,24	2,56	12,87%
6	5	5,00	17,00	25	125	625	85	425	289	15,42	2,48	0,36	9,27%
7	6	6,00	15,00	36	216	1296	90	540	225	16,76	3,10	6,76	11,73%
8	7	7,00	19,00	49	343	2401	133	931	361	17,95	1,11	1,96	5,55%
9	8	8,00	18,00	64	512	4096	144	1152	324	18,98	0,97	0,16	5,47%
10	9	9,00	20,00	81	729	6561	180	1620	400	19,87	0,02	5,76	0,63%
11	10	10,00	19,00	100	1000	10000	190	1900	361	20,62	2,61	1,96	8,51%
12	11	11,00	21,00	121	1331	14641	231	2541	441	21,21	0,04	11,56	1,00%
13	12	12,00	23,00	144	1728	20736	276	3312	529	21,65	1,81	29,16	5,85%
14	13	13,00	21,00	169	2197	28561	273	3549	441	21,95	0,91	11,56	4,53%
15	14	14,00	23,00	196	2744	38416	322	4508	529	22,10	0,81	29,16	3,91%
16	15	15,00	22,00	225	3375	50625	330	4950	484	22,10	0,01	19,36	0,45%
17	Сумм.	120,00	264,00	1240,00	14400,00	178312,00	2382,00	25838,00	4954,00	264,00	24,55	307,60	
18	Средн.	8,00	17,60										7,79%
19	n =	15											
20	Квадратичная регрессия: $y=a*x^2+b*x+c+e$												
21		178312,00	14400,00	1240,00		178312,00	25838,00	1240,00		D=	17326400		
22	A=	14400,00	1240,00	120,00	A2=	14400,00	2382,00	120,00		D1=	-1285200	a=	-0,074
23		1240,00	120,00	15		1240,00	264,00	15		D2=	37270800	b=	2,151
24										D3=	113021440	c=	6,523
25		25838,00	14400,00	1240,00		178312,00	14400,00	25838,00		ro=	0,959		
26	A1=	2382,00	1240,00	120,00	A3=	14400,00	1240,00	2382,00		R2=	0,920		
27		264,00	120,00	15		1240,00	120,00	264,00					

Коэффициент 0,920 близок к 1, что свидетельствует о высоком качестве подгонки.

Последним действием будет вычисление относительной ошибки. Добавим столбец и внесём туда формулу: " $=ABS((C2-J2)/C2)$ ", ABS - модуль, абсолютное значение. Протянем маркером вниз и в ячейке M18 выведем среднее значение (СРЗНАЧ), назначим ячейкам процентный формат. Полученный результат - 7,79% находится в пределах допустимых значений ошибки <8...10%. Значит вычисления достаточно точны.

Если возникнет необходимость, по полученным значениям мы можем построить график

Задание.

Задача на практику выполнить программную реализовать предложенного в теории подхода.