

**Задание на практическую работу по теме: "Математический инструментарий интеллектуального анализа данных"**

1. Ознакомьтесь с данными о спросе, производстве и ценах на продукцию предприятия, приведенными в таблице 1.1, их графическим изображением на диаграмме 1 и выводом, сделанным на основе графического анализа.

2. Самостоятельно сформируйте новые ряды исходных данных (показатели  $x$ ,  $y$ ,  $z$ ) и занесите их в таблицу 1.2. Показатель  $x$  будем считать результирующим, а показатели  $y$  и  $z$  - факторными (причинными). По мере занесения данных в таблицу 2, на диаграмме 2 будет появляться их графическое изображение. Сделайте вывод на основе графического анализа по образцу.

3. Рассчитайте (пользуясь таблицей 2.1. на листе 2) линейные коэффициенты корреляции между показателями  $x$  и  $y$ , а также между показателями  $x$  и  $z$ . Сделайте выводы о том, какая связь более сильная. Тот показатель ( $y$  или  $z$ ), у которого связь с показателем  $x$  более сильная, выберите для дальнейших расчетов.

4. Сделайте прогноз показателя  $x$  по двум вариантам уравнений тренда (линейному и квадратичному). Для расчета параметров уравнений тренда пользуйтесь вспомогательной таблицей на листе 3. Рассчитайте ошибки аппроксимации и сделайте вывод о том, какой прогноз более достоверен.

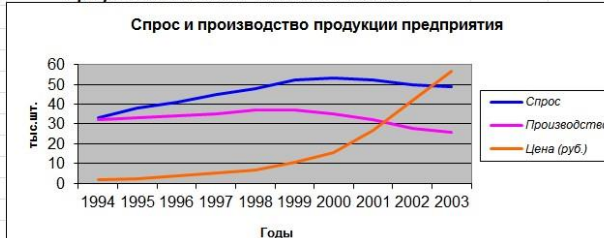
5. Рассчитайте параметры уравнения парной линейной регрессии, выражающей зависимость между показателем  $x$  и тем из двух показателей ( $y$  или  $z$ ), с которым связь показателя  $x$  более сильная. Пользуйтесь вспомогательными таблицами на листах 4 и 5. Рассчитайте ошибку аппроксимации и индекс детерминации. Сделайте вывод о том, насколько хорошо построенное уравнение отражает существующую зависимость.

6. Выполните (на листе 6) прогноз показателя, выбранного в п.3, по любому из уравнений тренда. Затем подставьте полученное прогнозное значение в уравнение регрессии между этим показателем и показателем  $x$  и рассчитайте прогноз спроса по уравнению регрессии.

**Таблица 1.1.**

Исходные данные (условный пример)			
Периоды времени	Спрос $x$	Произ-во $y$	Цена $z$
1994	33	32	2
1995	38	33	2,5
1996	41	34	3,7
1997	45	35	5,4
1998	48	37	6,8
1999	52	37	10,5
2000	53	35	15,6
2001	52	32	26,8
2002	50	28	42,2
2003	49	26	56,8

**Графический анализ исходных данных**



**Графический анализ**

**Примерный вывод (образец):** На графике видно, что цена на продукцию до конца 90-х годов равномерно растет. Затем наблюдается резкий рост цены, который продолжается вплоть до 2003г. Одновременно спрос на продукцию предприятия до конца 90-х годов развивается более быстрыми темпами, а потом начинает снижаться и постепенно стабилизируется. Производство вначале медленно растет, достигает максимальной величины в 1998-99 гг, а потом резко снижается и постепенно стабилизируется только к концу периода. Представляется, что имеется тесная обратная связь между ценой и спросом, и в то же время никакой связи между спросом и производством продукции предприятия не наблюдается. Можно предположить, что руководство предприятия планировало производство без всякого учета спроса. Для проверки необходимо рассчитать линейные коэффициенты корреляции между спросом (показателем  $x$ ) и производством (показателем  $y$ ), а также между спросом ( $x$ ) и ценой (показателем  $z$ ).

Таблица 1.2.

## Исходные данные

Периоды времени	Показатель x	Показатель y	Показатель z
1994	7	11	20
1995	13	24	18
1996	20	30	19
1997	28	42	15
1998	33	47	25
1999	53	48	29
2000	64	63	30
2001	72	75	28
2002	85	73	35
2003	95	88	45

## Графический анализ исходных данных

## Спрос и производство продукции предприятия



**Вывод:** На графике видно, что спрос и объём производства растут на протяжении всего анализируемого периода. Причём до 1999 объём производства (показатель y) превышал показатель спроса, а после 2001 спрос превысил показатель производства. Показатель цены начал активно расти лишь с 2001 года.

## Расчет линейного коэффициента корреляции между спросом и ценой

Исходные данные		Вспомогательные расчеты				
x	z	$x - \bar{x}$ x-x'	$z - \bar{z}$ z-z'	$(x - \bar{x})^2$	$(z - \bar{z})^2$	
7	20	-40,00	-6,40	1600,00	40,96	256,00
13	18	-34,00	-8,40	1156,00	70,56	285,60
20	19	-27,00	-7,40	729,00	54,76	199,80
28	15	-19,00	-11,40	361,00	129,96	216,60
33	25	-14,00	-1,40	196,00	1,96	19,60
53	29	6,00	2,60	36,00	6,76	15,60
64	30	17,00	3,60	289,00	12,96	61,20
72	28	25,00	1,60	625,00	2,56	40,00
85	35	38,00	8,60	1444,00	73,96	326,80
95	45	48,00	18,60	2304,00	345,96	892,80
Σ	470,00	264,00	0,00	8740,00	740,40	2314,00
$\bar{x} =$	$\bar{z} =$	Линейный коэффициент корреляции				С исп. функ. КОРРЕЛ
47,00	26,40	$r(x,z) = 0,91$				0,91

**Вывод :** Коэффициент корреляции равен 0,91. Значит связь

$$(x - \bar{x}) * (z - \bar{z})$$

между двумя показателями  $x$  и  $y$  более тесная. Для более точного выражения количественной зависимости между показателем  $x$  и  $y$ , необходимо построить уравнение регрессии.

$$M(x^2) = \frac{7^2 + 13^2 + 20^2 + 28^2 + 33^2 + 53^2 + 64^2 + 72^2 + 85^2 + 95^2}{10} = 3083$$

$$\sigma(X) = \sqrt{3083 - 2209} \approx 30$$

$$M(Y)^2 = 50,1^2 = 2510$$

$$M(Y^2) = \frac{11^2 + 24^2 + 30^2 + 42^2 + 47^2 + 48^2 + 63^2 + 75^2 + 73^2 + 88^2}{10} = 3054$$

$$\sigma(Y) = \sqrt{3054 - 2510} = 23 \quad \text{cov}(X; Y) = M(XY) - M(X)M(Y) = 671$$

$$r(X, Y) = \frac{\text{cov}(X; Y)}{\sigma(X)\sigma(Y)} = \frac{671}{23 \cdot 30} = 0,97$$

$$\bar{x} = \sum_{i=1}^n x_i / n = \frac{7 + 13 + 20 + 28 + 33 + 53 + 64 + 72 + 85 + 95}{10} = \frac{470}{10} = 47 = M(X)$$

$$\bar{y} = \sum_{i=1}^n y_i / n = \frac{11 + 24 + 30 + 42 + 47 + 48 + 63 + 75 + 73 + 88}{10} = \frac{501}{10} = 50,1 = M(Y)$$

$$r(x, y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{M(X^2) - M(X)^2}$$

$$M(X)^2 = 47^2 = 2209$$

$$\bar{z} = \sum_{i=1}^n z_i / n = \frac{20 + 18 + 19 + 15 + 25 + 29 + 30 + 28 + 35 + 45}{10} = 26,4 = M(Z)$$

$$\sigma(Z) = \sqrt{M(Z^2) - M(Z)^2}$$

$$M(Z^2) = 771$$

$$M(Z)^2 = 697$$

$$\sigma(Z) = \sqrt{771 - 697} = 8,6$$

$$\text{cov}(X; Z) = M(XZ) - M(X) \cdot M(Z) = 231$$

$$r(X; Z) = \frac{\text{cov}(X; Z)}{\sigma(X)\sigma(Z)} = \frac{231}{8,6 \cdot 30} = 0,90$$

**ЗАДАНИЕ 4.РАСЧЕТ ПАРАМЕТРОВ ЛИНЕЙНОГО И КВАДРАТИЧЕСКОГО ТРЕНДА  
(ДЛЯ ПОКАЗАТЕЛЯ X)**

**Исходные данные**

**Вспомогательные расчеты**

Периоды времени	X	Условное обозначение времени			xt	xt <sup>2</sup>
		t	t <sup>2</sup>	t <sup>4</sup>		
1994	7	-5	25	625	-35	175
1995	13	-4	16	256	-52	208
1996	20	-3	9	81	-60	180
1997	28	-2	4	16	-56	112
1998	33	-1	1	1	-33	33
1999	53	1	1	1	53	53
2000	64	2	4	16	128	256
2001	72	3	9	81	216	648
2002	85	4	16	256	340	1360
2003	95	5	25	625	475	2375
<b>Σ</b>	<b>470</b>	<b>0</b>	<b>110</b>	<b>1958</b>	<b>976</b>	<b>5400</b>

**Расчет параметров линейного и квадратического тренда**

Линейный тренд x		Квадратический тренд x	
$x^{\wedge} = a_0 + a_1 \cdot t$		$x^{\wedge\wedge} = b_0 + b_1 \cdot t + b_2 \cdot t^2$	
a <sub>0</sub> =	47	b <sub>0</sub> =	43,62
a <sub>1</sub> =	8,87	b <sub>1</sub> =	8,87
		b <sub>2</sub> =	0,31

Периоды времени	Исходные данные		Расчетные данные			
	X	t	x <sup>^</sup>	x <sup>^^</sup>	(x <sup>^</sup> - x) <sup>2</sup>	(x <sup>^^</sup> - x) <sup>2</sup>
1985	7	-5	2,64	6,94	19,04	0,00
1986	13	-4	11,51	13,05	2,22	0,00
1987	20	-3	20,38	19,77	0,15	0,05

1988	28	-2	29,25	27,10	1,57	0,81
1989	33	-1	38,13	35,05	26,29	4,21
1990	53	1	55,87	52,49	8,25	0,26
1991	64	2	64,75	62,59	0,56	1,98
1992	72	3	73,62	73,00	2,62	1,01
1993	85	4	82,49	84,03	6,30	0,94
1994	95	5	91,36	95,67	13,22	0,45
$\Sigma$	470	0	470,00	469,69	80,22	9,72

□

#### Ошибки аппроксимации для разных уравнений тренда

Вид уравнения тренда	Ошибка	
$\hat{x} = a_0 + a_1 \cdot t$	□□□□	2,83
$\hat{x} = b_0 + b_1 \cdot t + b_2 \cdot t^2$	□□□□	0,99

#### Расчет прогнозных значений по тренду

Вид уравнения тренда	Прогноз	Ошибка
$\hat{x} = a_0 + a_1 \cdot t$	100,24	2,83
$\hat{x} = b_0 + b_1 \cdot t + b_2 \cdot t^2$	107,92	0,99

ВЫВОД:	
(укажите, какое из прогнозных значений Вы считаете более достоверным и почему)	Достоверным можно считать прогнозное значение, рассчитанное по квадратичному уравнению тренда, так как ошибка аппроксимации наименьшая, тем самым квадратичное уравнение тренда более точно описывает динамику, чем линейное уравнение тренда.



Расчет параметров линейного и квадратичного трендов.

$$a_0 = \frac{\sum_{i=1}^n x_i y_i}{n} \quad a_1 = \frac{\sum_{i=1}^n x_i^2}{n}$$

Расчет параметров линейного и квадратичного трендов.

$$a_0 = \frac{\sum x_i}{n}; \quad a_1 = \frac{\sum x_i \cdot t}{\sum t^2} \quad - \text{ для линейного тренда}$$

$$b_0 = \frac{\sum t^4 \sum x - \sum t^2 \sum x t^2}{n \sum t^4 - \sum t^2 \sum t^2}; \quad b_1 = \frac{\sum x_i \cdot t}{\sum t^2}; \quad b_2 = \frac{n \sum x_i t^2 - \sum t^2 \sum x}{n \sum t^4 - \sum t^2 \sum t^2}$$

$$\sum x_i = 470 \quad \sum x_i t = 976 \quad \sum t^2 = 110 \quad \sum t^4 = 1958 \quad \sum x_i t^2 = 5400$$

$$a_0 = \frac{470}{10} = 47; \quad a_1 = \frac{976}{110} = 8,87$$

$$b_0 = \frac{1958 \cdot 470 - 110 \cdot 5400}{10 \cdot 1958 - 110 \cdot 110} = \frac{326260}{7480} = 43,62; \quad b_1 = \frac{976}{110} = 8,87;$$

$$b_2 = \frac{10 \cdot 5400 - 110 \cdot 470}{10 \cdot 1958 - 110 \cdot 110} = \frac{2300}{7480} = 0,31$$

$$x^1 = a_0 + a_1 t; \quad \sum x^1 = 470; \quad \sum (x^1 - x)^2 = 80,22;$$

$$x^{11} = b_0 + b_1 t + b_2 t^2; \quad \sum x^{11} = 469,69; \quad \sum (x^{11} - x)^2 = 9,72.$$

$$\sigma_1 = \sqrt{\frac{\sum (x^1 - x)^2}{n}} = \sqrt{8,02} = 2,83; \quad \sigma_2 = \sqrt{\frac{\sum (x^{11} - x)^2}{n}} = \sqrt{9,72/10} = 0,99$$

$$x^1(6) = 47 + 8,87 \cdot 6 = 100,24$$

$$x^{11}(6) = 43,62 + 8,87 \cdot 6 + 0,31 \cdot 6^2 = 107,92$$

### Расчет параметров парной линейной регрессии (задание 5)

Вспомогательная таблица для расчета параметров уравнения парной линейной регрессии ( $X = k_0 + k_1 \cdot y$ )

данные		Вспомогательные расчеты			
		Расчет параметров		Расчет ошибки(□)	
		y <sup>2</sup>	y*x	x <sub>z</sub> =k <sub>0</sub> +k <sub>1</sub> *y	(x - xy) <sup>2</sup>
x	y				
7	11	121	77,00	-1,22	67,57
13	24	576	312,00	14,81	3,28
20	30	900	600,00	22,21	4,89
28	42	1764	1176,00	37,01	81,19
33	47	2209	1551,00	43,18	103,57

	53	48	2304	2544,00	44,41	73,79
	64	63	3969	4032,00	62,91	1,19
	72	75	5625	5400,00	77,71	32,58
	85	73	5329	6205,00	75,24	95,23
	95	88	7744	8360,00	93,74	1,59
<b>Σ</b>	470	501	30541	30257,00	470,00	464,88

Параметры регрессии	
<b>k<sub>0</sub> =</b>	<b>-14,8</b>
<b>k<sub>1</sub> =</b>	<b>1,23</b>

Ошибка аппроксимации	
<b>σ<sub>с</sub></b>	<b>6,82</b>

**ВЫВОД:** построенное линейное уравнение регрессии отражает существующую зависимость с ошибкой 6,82. Это говорит о том, что уравнение регрессии достоверно оценивает взаимосвязь показателей x и y.

*Расчет параметров линейной регрессии.*

$$\sum x = 470 \quad \sum y = 501 \quad \sum y^2 = 30541 \quad \sum yx = 30257$$

$$\bar{x} = 47 \quad \bar{y} = 50,1 \quad \bar{y}x = 3025,7 \quad \bar{y}^2 = 3054,1 \quad \bar{y}^2 = 2510,01$$

$$k_1 = \frac{3025,7 - 47 \cdot 50,1}{3054,1 - 2510,01} = \frac{671}{544,09} = 1,23$$

$$k_0 = \bar{x} - k_1 \bar{y} = 47 - 1,23 \cdot 50,1 = -14,62$$

$$\hat{x} = k_0 + k_1 \cdot y; \quad \sum \hat{x} = 470; \quad \sum (x - \hat{x})^2 = 464,88$$

$$\sigma_s = \sqrt{\frac{\sum (x - \hat{x})^2}{n}} = 6,82$$

**ЗАДАНИЕ 6.** Расчет прогноза цены по тренду и прогноза результирующего показателя x по регрессии )

1. Скопируйте на этот лист формулы для расчета параметров тренда, из таблицы, построенной на листе 3, заменив исходный ряд (x) на ряд значений выбранного факторного показателя (y или z), и рассчитайте параметры уравнений тренда для расчета прогнозного значения этого показателя.

**ПРИМЕЧАНИЕ:** Ниже в таблице для определенности указан показатель z, но это может быть и y

y	Условное обозначение времени			y*t	y*t <sup>2</sup>
	t	t <sup>2</sup>	t <sup>4</sup>		
11	-5	25	625	-55	275

	24	-4	16	256	-96	384
	30	-3	9	81	-90	270
	42	-2	4	16	-84	168
	47	-1	1	1	-47	47
	48	1	1	1	48	48
	63	2	4	16	126	252
	75	3	9	81	225	675
	73	4	16	256	292	1168
	88	5	25	625	440	2200
Σ	501	0	110	1958	759	5487

Линейный тренд z			Квадратический тренд z		
$y^{\wedge} = u_0 + u_1 \cdot t$			$y^{\wedge\wedge} = w_0 + w_1 \cdot t + w_2 \cdot t^2$		
$u_0 =$	50,10		$w_0 =$	50,45	
$u_1 =$	6,90		$w_1 =$	6,90	
			$w_2 =$	-0,03	

2) Рассчитайте две ошибки аппроксимации, по аналогии с тем, как это делалось для линейного и квадратичного тренда показателя x.

Исходные данные			Расчетные данные		
y	t	$y^{\wedge}$	$y^{\wedge\wedge}$	$(y^{\wedge} - y)^2$	$(y^{\wedge\wedge} - y)^2$
11	-5	15,60	15,15	21,16	17,23
24	-4	22,50	22,34	2,25	2,76
30	-3	29,40	29,46	0,36	0,29
42	-2	36,30	36,52	32,49	29,98
47	-1	43,20	43,52	14,44	12,10
48	1	57,00	57,32	81,00	86,88
63	2	63,90	64,12	0,81	1,26
75	3	70,80	70,86	17,64	17,11
73	4	77,70	77,54	22,09	20,61
88	5	84,60	84,15	11,56	14,82
Σ	501,00	0,00	501,00	203,80	203,03

#### Ошибки аппроксимации для разных уравнений тренда

Вид уравнения тренда	Ошибка	
$y^{\wedge} = u_0 + u_1 \cdot t$	Σ	4,5144 2
$y^{\wedge\wedge} = w_0 + w_1 \cdot t + w_2 \cdot t^2$	Σ	4,5058 8

Расчет прогнозных значений показателя по тренду



Вид уравнения тренда	Прогноз	Ошибка
$y^{\wedge} = u_0 + u_1 \cdot t$	91,50	4,51
$y^{\wedge\wedge} = w_0 + w_1 \cdot t + w_2 \cdot t^2$	90,70	4,51

д)

3. Определите, какое из прогнозных значений показателя более достоверно, и подставьте его в уравнение регрессии, построенное на листе 4.

#### Расчет прогнозных значений показателя x по регрессии

Вид уравнения регрессии	Прогноз	Ошибка
$x^{\wedge} = k_0 + k_1 \cdot y$	96,76	11,33

Расчет прогноза по регрессии

$$\sum y_t = 501 \quad \sum t^2 = 110 \quad \sum t^3 = 1958 \quad \sum y \cdot t = 759 \quad \sum y \cdot t^2 = 5487$$

$$b_0 = \frac{1958 \cdot 501 - 110 \cdot 5487}{10 \cdot 1958 - 110 \cdot 110} = 50,45; \quad b_1 = \frac{501 \cdot 759}{110} = 6,9;$$

$$b_2 = \frac{10 \cdot 5487 - 110 \cdot 501}{10 \cdot 1958 - 110 \cdot 110} = -0,03 \quad y^{\wedge} = b_0 + b_1 \cdot t + b_2 \cdot t^2 = 90,7$$

$$\sigma_b = \sqrt{\frac{(y^{\wedge} - y)^2}{n}} = 4,505 \approx 4,51$$

$$x^{\wedge} = k_0 + k_1 \cdot y^{\wedge} = -14,62 + 1,23 \cdot 90,7 = 96,76$$

$$\sigma_{\text{итог}} = \sigma_b + \sigma_y = 6,82 + 4,51 = 11,33$$

#### Вывод:

Итоговая ошибка прогноза по регрессии составила 11,33. Эффективнее всего для исходных данных использовать квадратичное уравнение тренда для определения прогнозного значения параметра y, которое мы затем подставили в уравнение регрессии.