

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Локтионова Оксана Геннадьевна  
Должность: проректор по учебной работе  
Дата подписания: 12.10.2023 13:30:23  
Уникальный программный ключ:  
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

## МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Юго-Западный государственный университет»  
(ЮЗГУ)  
Кафедра вычислительной техники

УТВЕРЖДАЮ  
Проректор по учебной работе  
О.Г. Локтионова  
« 5 » 10 2023 г.



### МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ: ЛИНЕЙНАЯ РЕГРЕССИЯ

Методические указания по выполнению лабораторной работы  
для студентов направления подготовки 09.04.01 по дисциплине  
«Системы искусственного интеллекта»



## Постановка задачи

Парная (простая) линейная регрессия даёт нам правила, определяющие линию регрессии, которая лучше других предсказывает наиболее вероятные значения одной переменной на основании другой (переменных всего две).

Парная (простая) линейная регрессия представляет собой регрессию между двумя переменными –  $y$  и  $x$ :

$$Y_i = ax_i + b,$$

где  $Y_i$  – зависимая переменная,  $x_i$  – независимая переменная,  $b$  – константа, определяет точку пересечения прямой с осью  $Y$ ,  $a$  – угловой коэффициент, характеризует наклон прямой.

Коэффициент регрессии  $a$  показывает, на какую величину в среднем изменится результативный признак  $Y_i$ , если переменная  $X_i$  увеличится на единицу своего измерения.

Задача регрессионного анализа сводится к поиску коэффициентов  $a$  и  $b$ .

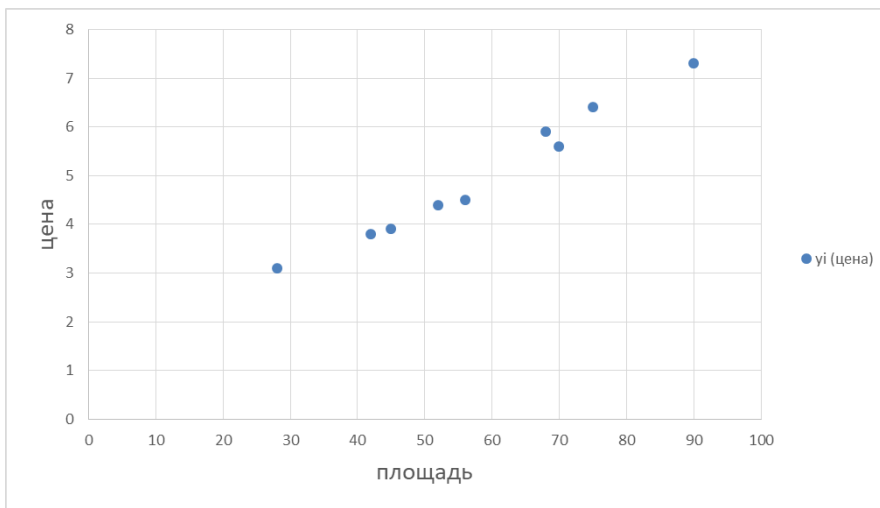
### Задание 1. Реализовать алгоритм линейной регрессии в MS Excel

#### Алгоритм решения в MS Excel 1 способ

##### 1. Формируем исходные данные в виде таблицы

	A	B	C
1			
2		$x_i$ (площадь)	$y_i$ (цена)
3		28	3,1
4		42	3,8
5		45	3,9
6		52	4,4
7		56	4,5
8		68	5,9
9		70	5,6
10		75	6,4
11		90	7,3

##### 2. Строим график по табличным данным (тип диаграммы - Точечная)



### 3. Составляем систему уравнений для расчета коэффициентов линейной зависимости $a$ и $b$ МНК (предварительно выполняем все необходимые вычисления в Excel)

а) проводим необходимые расчеты в Excel

	A	B	C	D	E	F
1						
2		$x_i$ (площадь)	$y_i$ (цена)	$x_i^2$	$x_i * y_i$	$y_i^2$
3		28	3,1	784	86,8	9,61
4		42	3,8	1764	159,6	14,44
5		45	3,9	2025	175,5	15,21
6		52	4,4	2704	228,8	19,36
7		56	4,5	3136	252	20,25
8		68	5,9	4624	401,2	34,81
9		70	5,6	4900	392	31,36
10		75	6,4	5625	480	40,96
11		90	7,3	8100	657	53,29
12	Сумма	526	44,9	33662	2832,9	239,29

Предварительные расчеты

б) составляем систему уравнений для расчета коэффициентов  $a$  и  $b$

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases}$$

$$\begin{cases} a * 33662 + b * 526 = 2832,9 \\ a * 526 + b * 9 = 44,9 \end{cases}$$

Система уравнений для расчета коэффициентов  $a$  и  $b$

#### 4. Решаем систему уравнений матричным методом

Решение системы уравнений матричным методом			
Матрица коэффициентов при a и b		Столбец свободных членов	
33662	526	2832,9	
526	9	44,9	
Обратная матрица к матрице коэффициентов		Коэффициенты a и b	
0,00034244	-0,020013698	a=	0,0715
-0,020013698	1,280800548	b=	0,8111

Решение системы уравнений матричным методом

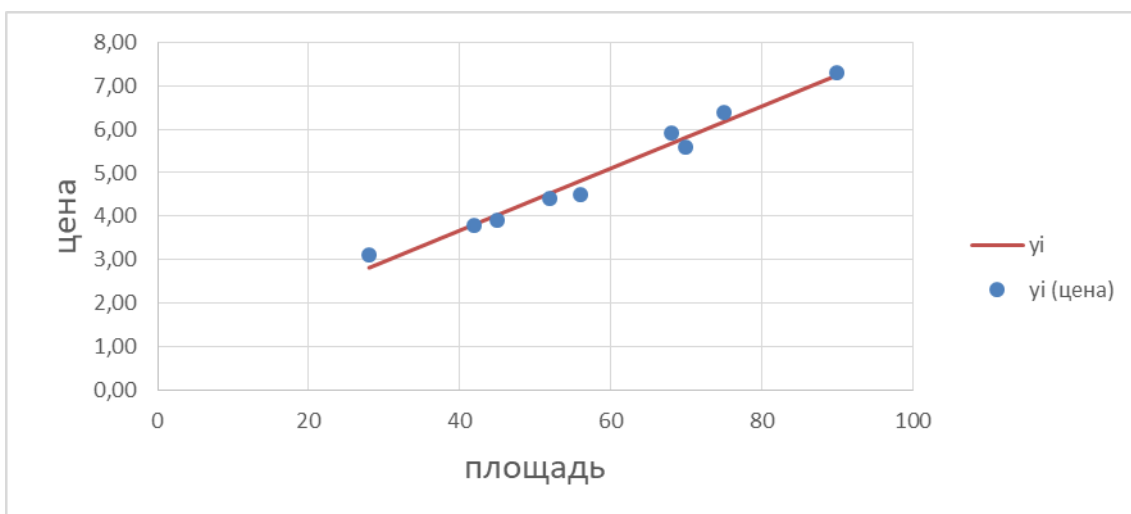
В результате решения системы уравнений получаем:

- коэффициенты  $a=0,0715$  и  $b=0,8111$ ;
- уравнение прямой  $y=0,0715x+0,8111$

#### 5. Рассчитываем столбец $y_i$ по формуле $y_i=0,0715x_i+0,8111$

	$x_i$ (площадь)	$y_i$ (цена)	$x_i^2$	$x_i \cdot y_i$	$y_i^2$	$y_i$
2						
3	28	3,1	784	86,8	9,61	2,81
4	42	3,8	1764	159,6	14,44	3,81
5	45	3,9	2025	175,5	15,21	4,03
6	52	4,4	2704	228,8	19,36	4,53
7	56	4,5	3136	252	20,25	4,81
8	68	5,9	4624	401,2	34,81	5,67
9	70	5,6	4900	392	31,36	5,81
10	75	6,4	5625	480	40,96	6,17
11	90	7,3	8100	657	53,29	7,24
12	Сумма	526	44,9	33662	2832,9	239,29

#### 6. Строим линию регрессии по столбцам $x_i$ (площадь) и $y_i$



#### 7. Рассчитываем коэффициенты корреляции $r_{xy}$ и детерминации $R^2$ (предварительно выполняем все необходимые вычисления в Excel)

	$x_i$ (площадь)	$y_i$ (цена)	$x_i^2$	$x_i \cdot y_i$	$y_i^2$	$y_i$	$(y_i - \bar{y}_{cp})^2$	$(y_i - \bar{y}_{cp})^2$
2								
3	28	3,1	784	86,8	9,61	2,81	3,567901235	4,736026847
4	42	3,8	1764	159,6	14,44	3,81	1,41345679	1,381772232
5	45	3,9	2025	175,5	15,21	4,03	1,185679012	0,923599673
6	52	4,4	2704	228,8	19,36	4,53	0,346790123	0,21221155
7	56	4,5	3136	252	20,25	4,81	0,239012346	0,030532221
8	68	5,9	4624	401,2	34,81	5,67	0,830123457	0,466562611
9	70	5,6	4900	392	31,36	5,81	0,37345679	0,682306814
10	75	6,4	5625	480	40,96	6,17	1,991234568	1,400507912
11	90	7,3	8100	657	53,29	7,24	5,341234568	5,088030547
12	Сумма	526	44,9	33662	2832,9	239,29	15,28888889	14,92155041
13	Среднее		4,99			4,99		

## Предварительные расчеты

Коэффициент корреляции  $r_{xy}$

$$r_{xy} = 0,9879$$

$$r_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[ \sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}}$$

Коэффициент детерминации  $R^2$

$$R^2 = 0,9759735$$

$$R^2 = \frac{SS_{regression}}{SS_{total}} = \frac{\sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$r_{xy} = \sqrt{R^2}$$

## Расчет коэффициентов

## 8. Результат выполнения работы

$x_i$ (площадь)	$y_i$ (цена)	$x_i^2$	$x_i \cdot y_i$	$y_i^2$	$y_i$	$(y_i - \bar{y}_{cp})^2$	$(y_i - \bar{y}_{cp})^2$
28	3,1	784	86,8	9,61	2,81	3,567901235	4,736026847
42	3,8	1764	159,6	14,44	3,81	1,41345679	1,381772232
45	3,9	2025	175,5	15,21	4,03	1,185679012	0,923599673
52	4,4	2704	228,8	19,36	4,53	0,346790123	0,21221155
56	4,5	3136	252	20,25	4,81	0,239012346	0,030532221
68	5,9	4624	401,2	34,81	5,67	0,830123457	0,466562611
70	5,6	4900	392	31,36	5,81	0,37345679	0,682306814
75	6,4	5625	480	40,96	6,17	1,991234568	1,400507912
90	7,3	8100	657	53,29	7,24	5,341234568	5,088030547
Сумма	526	44,9	33662	2832,9	239,29	15,28888889	14,92155041
Среднее		4,99			4,99		

**Формулы для расчета коэффициентов линейной зависимости а и b МНК**

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i \cdot y_i \\ a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases}$$

↓

$$\begin{cases} a \cdot 33662 + b \cdot 526 = 2832,9 \\ a \cdot 526 + b \cdot 9 = 44,9 \end{cases}$$

**Решение системы уравнений матричным методом**

Матрица коэффициентов при а и b	526
	9

Обратная матрица к матрице коэффициентов

0,00034244	-0,020013698
-0,020013698	1,280800548

Уравнение прямой  
 $y = 0,0715 \cdot x + 0,8111$

Столбец свободных членов

2832,9
44,9

Коэффициенты а и b

a = 0,0715
b = 0,8111

Коэффициент корреляции  $r_{xy}$

$$r_{xy} = 0,9879$$

$$r_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[ \sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}}$$

Коэффициент детерминации  $R^2$

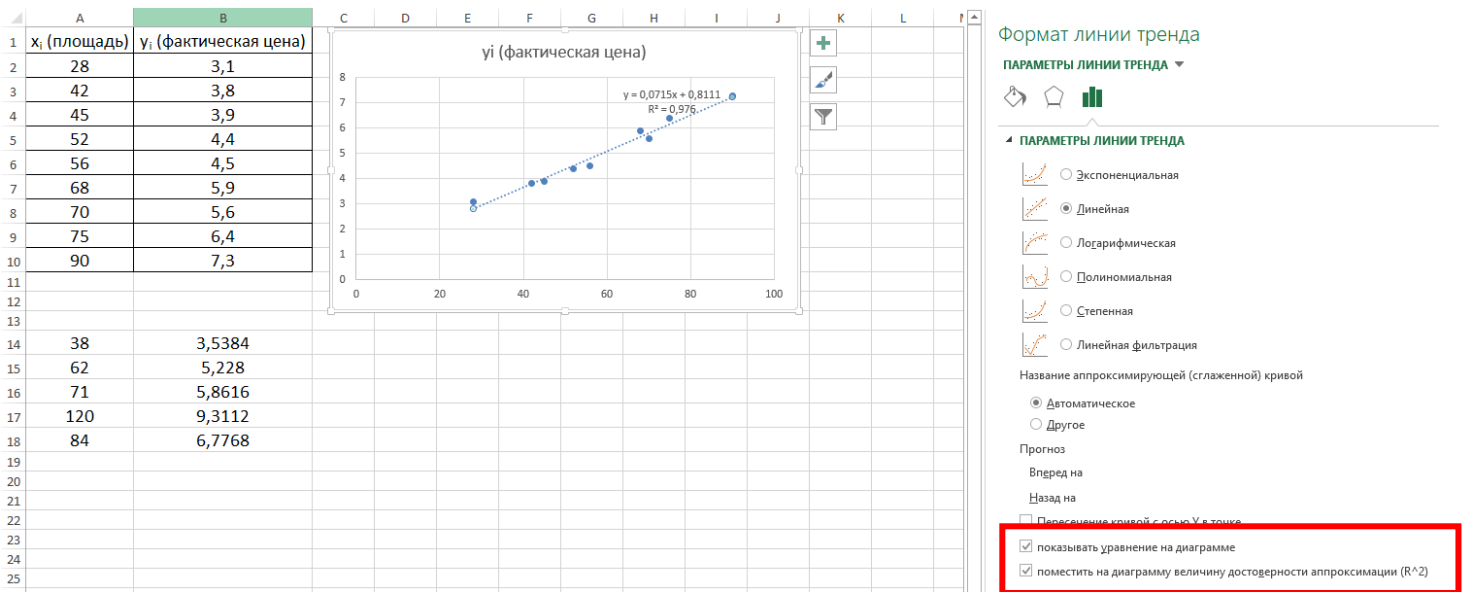
$$R^2 = \frac{SS_{regression}}{SS_{total}} = \frac{\sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$R^2 = 0,9759735$$

$$r_{xy} = \sqrt{R^2}$$

## 2 способ (построение линии тренда)

Для построения линии тренда надо выделить ряд данных и в контекстном меню выбрать команду «Добавить линию тренда».



### Задание

Для следующих данных подобрать линейную регрессионную модель.

<b>Площадь, кв. м</b>	<b>Стоимость, млн. руб.</b>
30	2,8
100	7,0
46	4,9
69	6,5
84	6,7
77	7,2
54	5,9
84	7,4
66	6,0
93	6,7
33	1,9
65	6,9
44	3,5
54	5,3
61	6,0
67	6,1
89	7,8
62	5,6
41	3,6
92	8,4
70	7,4
45	5,1
35	3,7
68	6,7
65	5,5



По полученной модели сделать прогноз о стоимости квартир площадью 38 кв. м, 50 кв. м, 110 кв. м.

### **Контрольные вопросы**

1. Приведите достоинства линейной регрессии как алгоритма машинного обучения.
2. Перечислите области применения линейной регрессии.
3. В чем заключается цель алгоритма линейной регрессии.
4. Назовите способы получения регрессионной модели.
5. Назовите, как применяются методы машинного интерфейса в интеллектуальных системах.