

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧЕРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
ЛИПЕЦКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

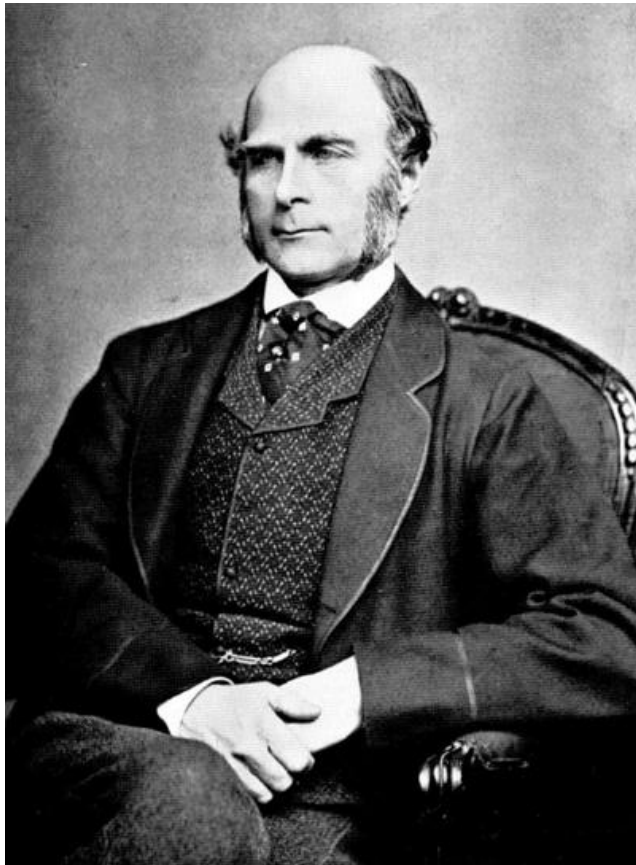
Кафедра Промышленной теплоэнергетики

Математическое моделирование

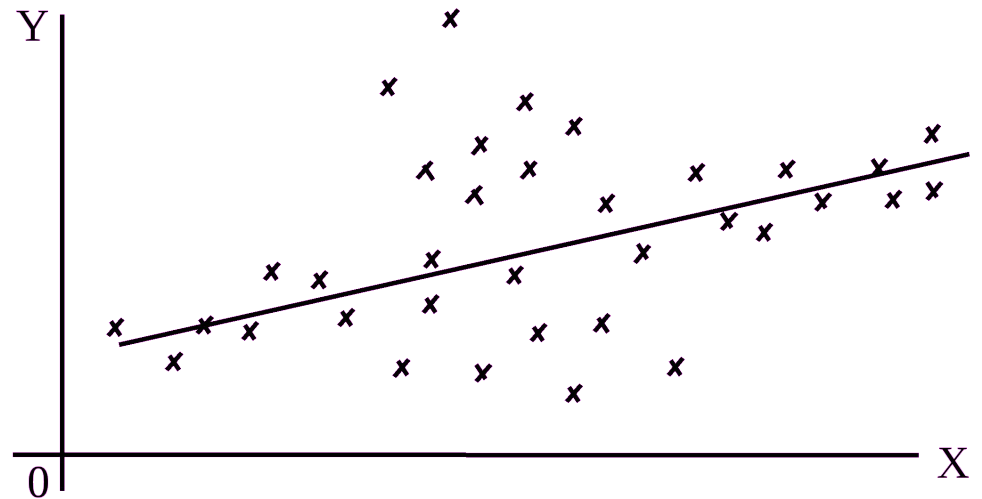
РЕГРЕССИОННЫЙ АНАЛИЗ

Липецк 2018

Регрессия – (от латинского слова «**regressio**» – обратное движение, возвращение) - математическое выражение, отражающее зависимость зависимой переменной y от независимых переменных x при условии, что это выражение будет иметь статистическую значимость.



Френсис Гальтон, английский исследователь



Регрессионный анализ – статистический метод исследования влияния одной или нескольких независимых переменных X_1, X_2, \dots, X_n на зависимую переменную Y .

Цель регрессионного анализа: определение связи между некоторой характеристикой Y наблюдаемого явления или объекта и величинами X_1, X_2, \dots, X_n , которые обуславливают, объясняют изменения Y .

Y – *зависимая переменная (отклик)*, описывает процесс или объект, который мы пытаемся предсказать или понять.

X_1, X_2, \dots, X_n – *влияющие переменные*, также называемые *факторами* (регрессорами), используются для моделирования и прогнозирования значений зависимых переменных.

Задачи регрессионного анализа:

- установка формы зависимости;
- подбор модели (уравнения) регрессии;
- оценка параметров модели.

Использование регрессионного анализа

- **Построение моделей**, объясняющих механизм влияния факторных признаков на результат;
- **Статистический прогноз** – вычисление значения результативной переменной для любых значений факторов;
- **Определение вклада** отдельных независимых переменных в вариацию зависимостей;
- **Восполнение пропусков** в данных.

Примеры применения регрессионного анализа

- Моделирование потоков миграции в зависимости от таких факторов, как средний уровень зарплат, наличие медицинских, образовательных учреждений, географического положения и т. д.;
- Моделирование автотранспортных происшествий как функции скорости, погодных и дорожных условий;
- Моделирование потерь от пожаров как функции таких переменных, как количество и расположение пожарных станций, время обработки вызова или цена собственности.

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

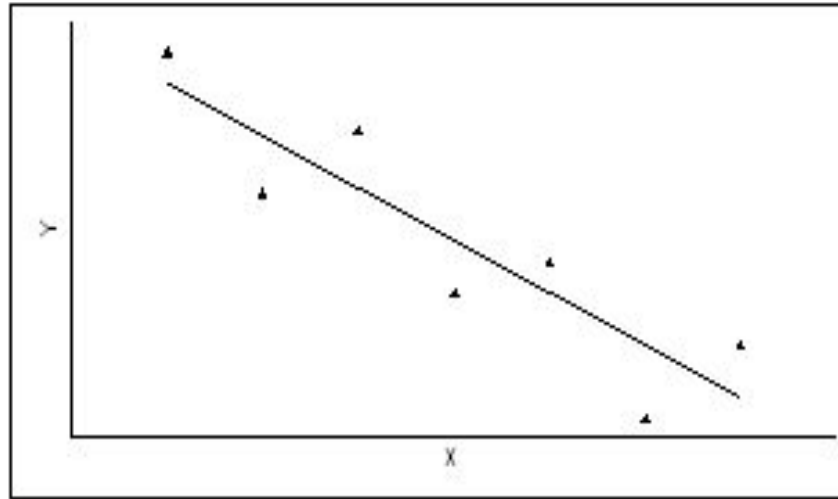
$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;

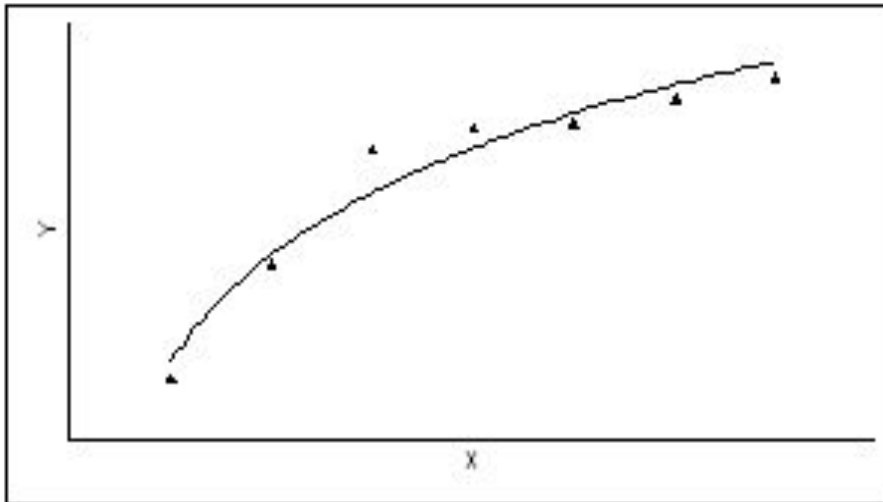
$$y_x = M_x(Y);$$

b_0, b_1, \dots, b_p - параметры функции регрессии φ

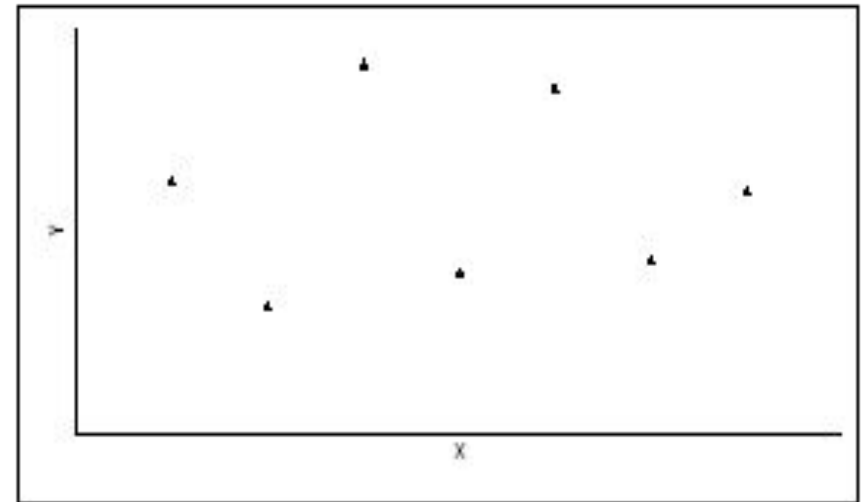
Формы связи



линейная регрессия



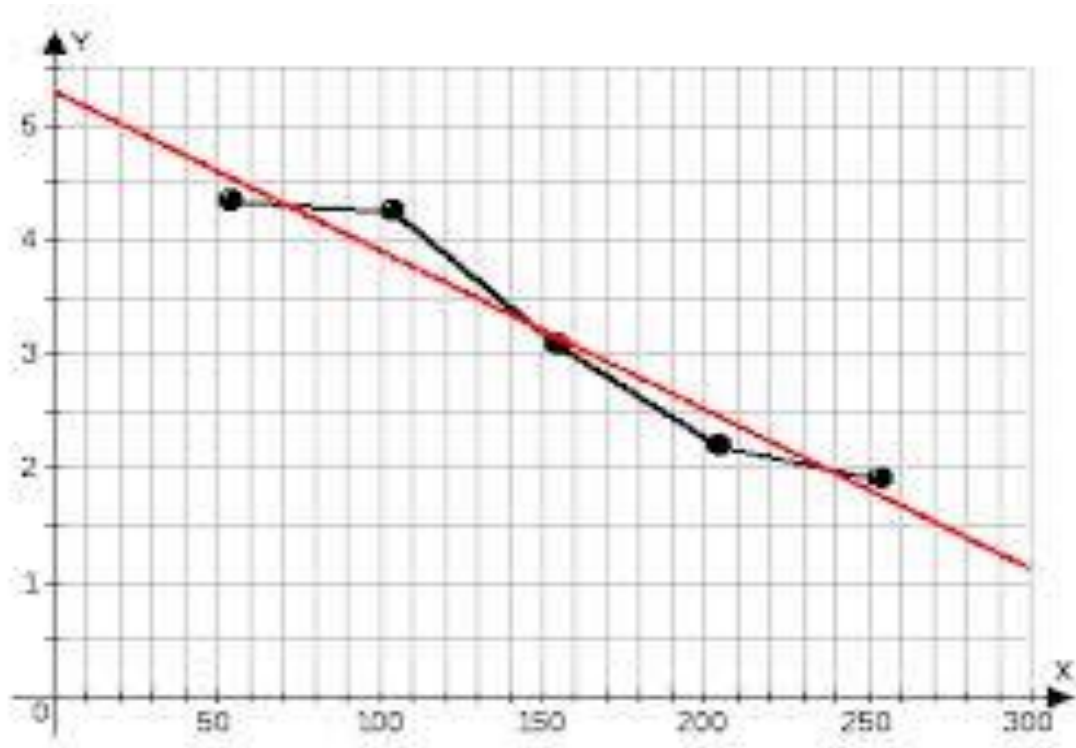
нелинейная регрессия



отсутствие взаимосвязи

Этапы регрессионного анализа

- Выявление наличия взаимосвязи между признаками;
- Определение формы связи;
- Определения силы (тесноты) и направления связи.



Линия регрессии – линия, характеризующаяся тем, что сумма квадратов расстояний от точек на диаграмме до этой линии минимальна.

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

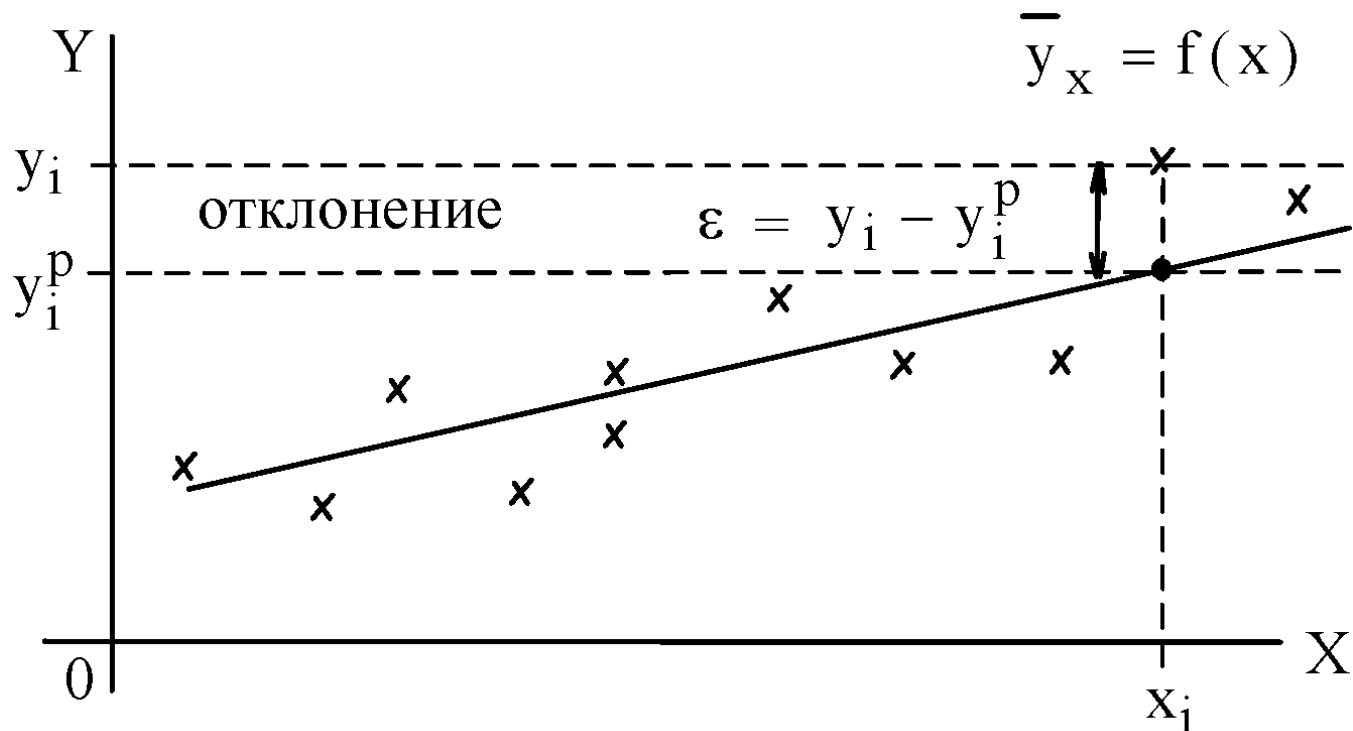
Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

b_0, b_1, \dots, b_n – параметры функции регрессии φ



Парная линейная регрессионная модель

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

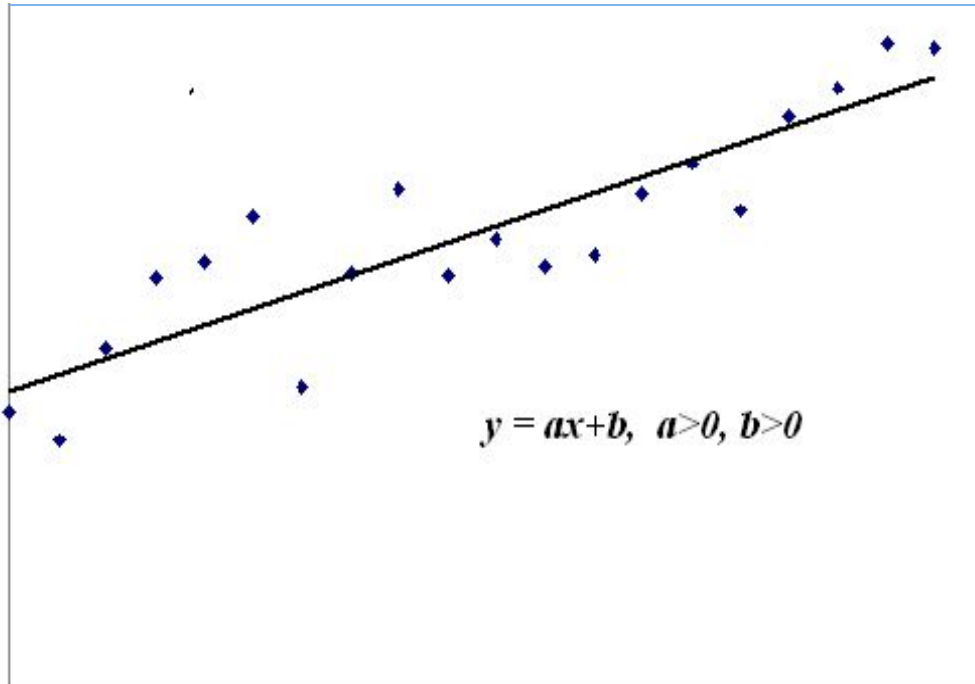
где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$$y_x = M_x(Y);$$



В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$$y_x = M_x(Y);$$

b_0, b_1, \dots, b_p – параметры функции регрессии φ

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$$y_x = M_x(Y);$$

b_0, b_1, \dots, b_p – параметры функции регрессии φ

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$$y_x = M_x(Y);$$

b_0, b_1, \dots, b_p – параметры функции регрессии φ

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$$y_x = M_x(Y);$$

b_0, b_1, \dots, b_p – параметры функции регрессии φ

Коэффициент регрессии

Смысл коэффициента регрессии:

- в общем случае коэффициент регрессии a показывает как в среднем изменится *результативный признак* (Y), если *факторный признак* (X) увеличится на единицу.

Свойства коэффициента регрессии

В регрессионном анализе изучаются модели вида:

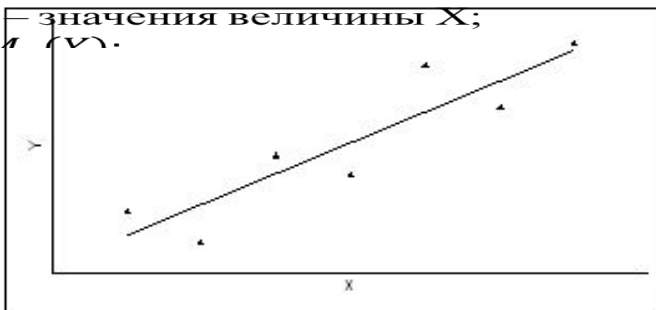
$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

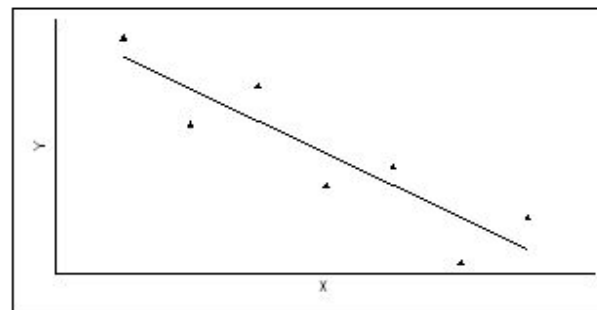
Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;
 $y = M(Y)$.



положительная линейная регрессия



отрицательная линейная регрессия

Коэффициент детерминации

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

b_0, b_1, \dots, b_p – параметры функции регрессии φ .

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

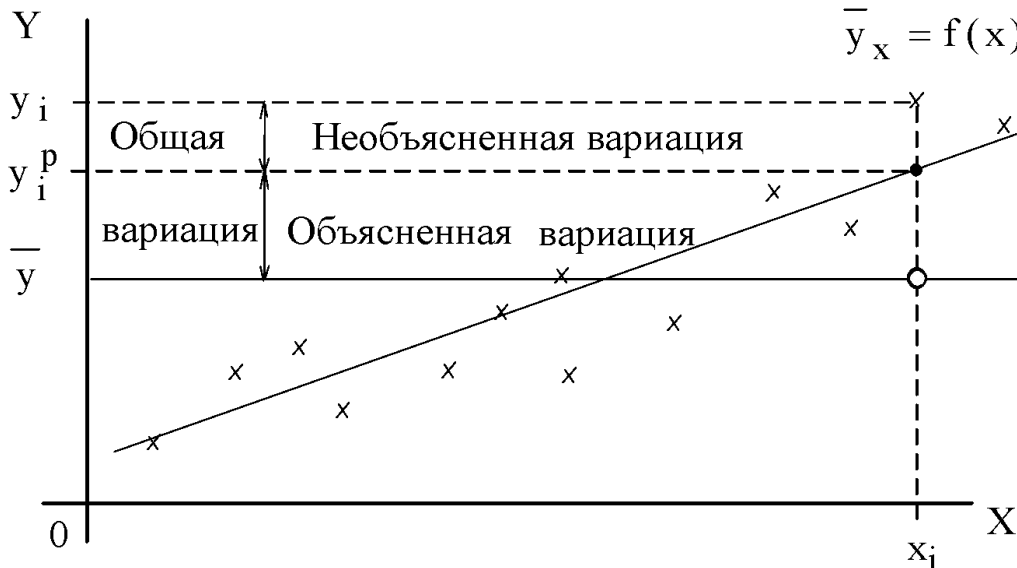
Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

b_0, b_1, \dots, b_p – параметры функции регрессии φ .



Графическая интерпретация коэффициента детерминации

Нелинейная регрессия

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

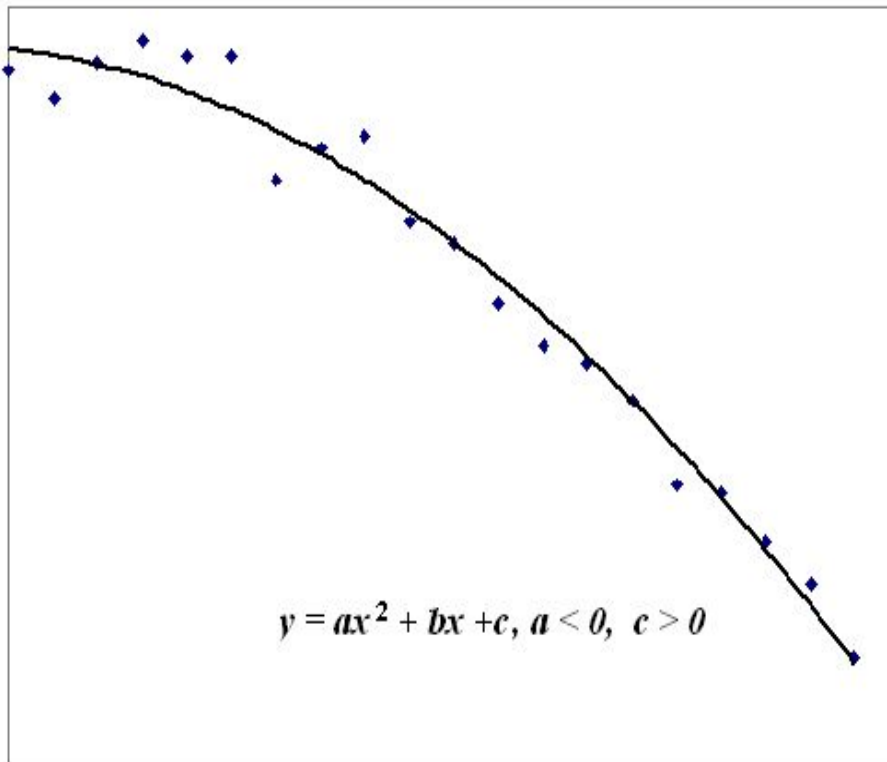
где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$y_x = M_x(Y)$.



В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);

X – фактор (неслучайная независимая переменная);

ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p).$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

b_0, b_1, \dots, b_p – параметры функции регрессии φ

Множественная регрессия

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);
 X – фактор (неслучайная независимая переменная);
 ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

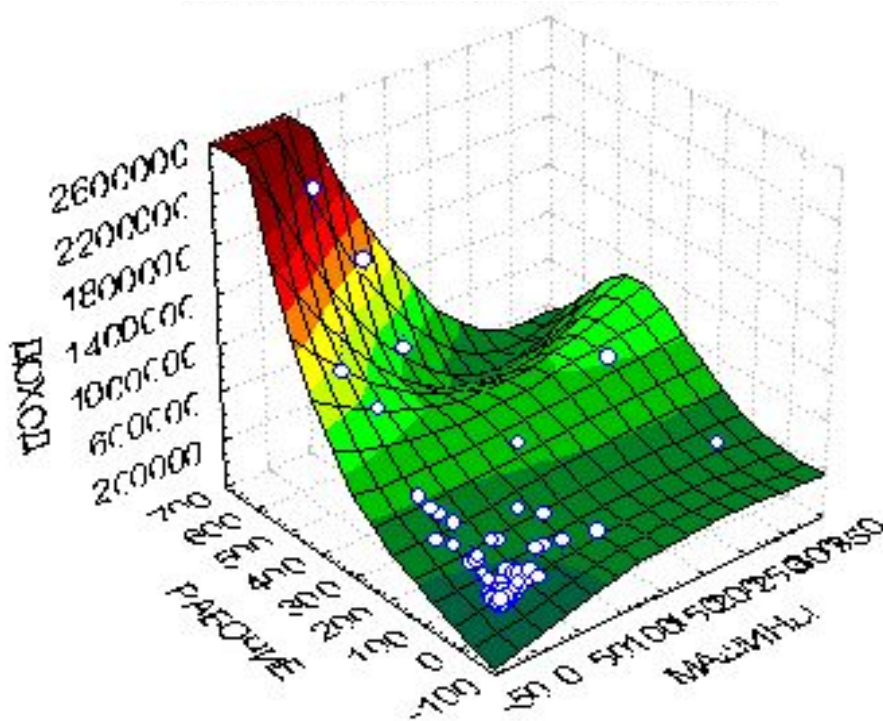
Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

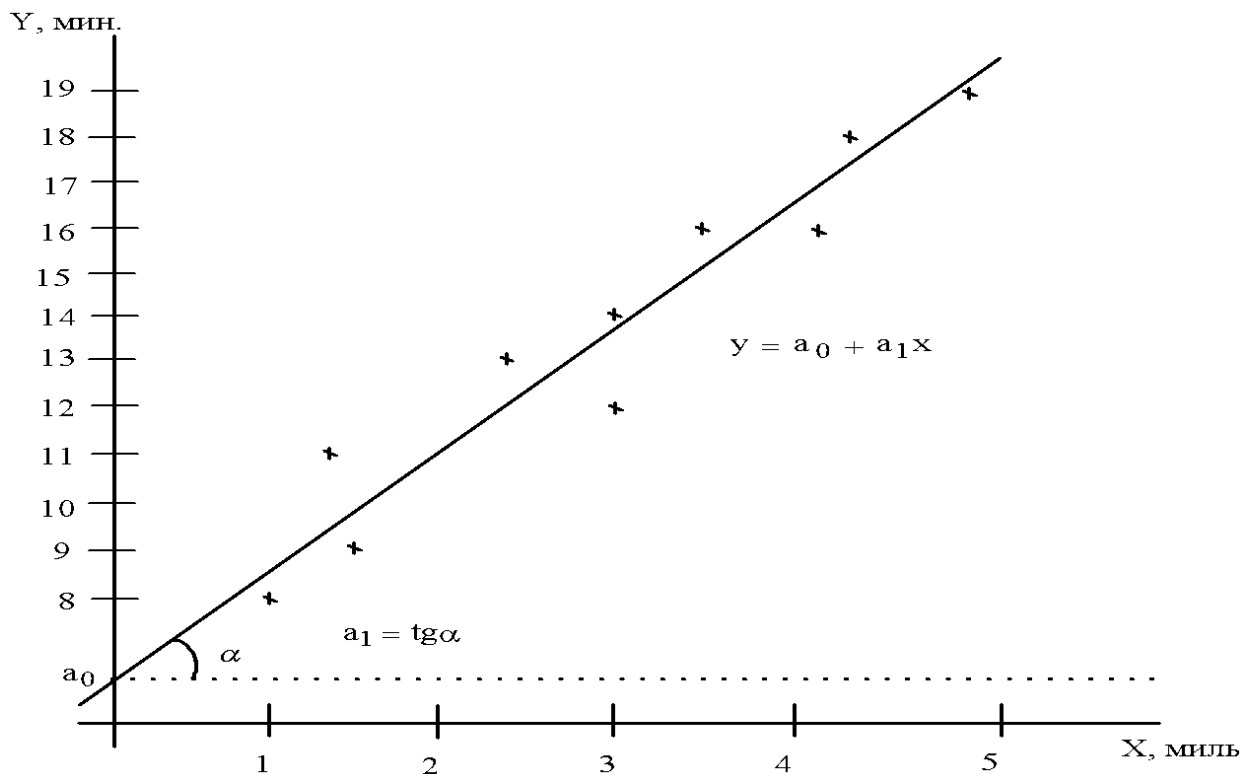
b_0, b_1, \dots, b_p – параметры функции регрессии φ



Смысл коэффициента регрессии в уравнении множественной регрессии состоит в том, то он показывает как в среднем изменится значение *результативного фактора*, если соответствующий *факторный признак* увеличить на единицу при *фиксированных значениях всех остальных факторов*

Пример. В механическом цехе анализируется структура себестоимости продукции и доля покупных комплектующих. Было отмечено, что стоимость комплектующих зависит от времени их поставки. В качестве наиболее важного фактора, влияющего на время поставки, выбрано пройденное расстояние. Провести регрессионный анализ данных о поставках:

Расстояние, миль	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Время, мин	16	13	19	18	12	11	8	14	9	16



№	x_i	y_i
1	3,5	16
2	2,4	13
3	4,9	19
4	4,2	18
5	3,0	12
6	1,3	11
7	1,0	8
8	3,0	14
9	1,5	9
10	4,1	16
Σ	28,9	136

x_i^2	$x_i y_i$
12,25	56,00
5,76	31,20
24,01	93,10
17,64	75,60
9,00	36,00
1,69	14,30
1,00	8,00
9,00	42,00
2,25	13,50
16,81	65,60
99,41	435,30

y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
15,22	2,63	5,76
12,30	1,70	0,36
18,95	28,59	29,16
17,09	12,15	19,36
13,89	0,08	2,56
9,37	17,88	6,76
8,57	25,27	31,36
13,89	0,09	0,16
9,90	13,67	21,16
16,82	10,36	5,76
—	112,42	122,40

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);

X – фактор (неслучайная независимая переменная);

ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;

$$y_x = M_x(Y);$$

b_0, b_1, \dots, b_p – параметры функции регрессии φ

В регрессионном анализе изучаются модели вида:

$$Y = \varphi(X) + \varepsilon,$$

где Y – результирующий признак (отклик, случайная зависимая переменная);

X – фактор (неслучайная независимая переменная);

ε – случайная переменная, характеризующая отклонение фактора X от линии регрессии (остаточная переменная).

Уравнение регрессии в общем виде выглядит так:

$$y_x = \varphi(x, b_0, b_1, \dots, b_p),$$

Где x – значения величины X ;

$y_x = M_x(Y)$;

b_0, b_1, \dots, b_p - параметры функции регрессии φ