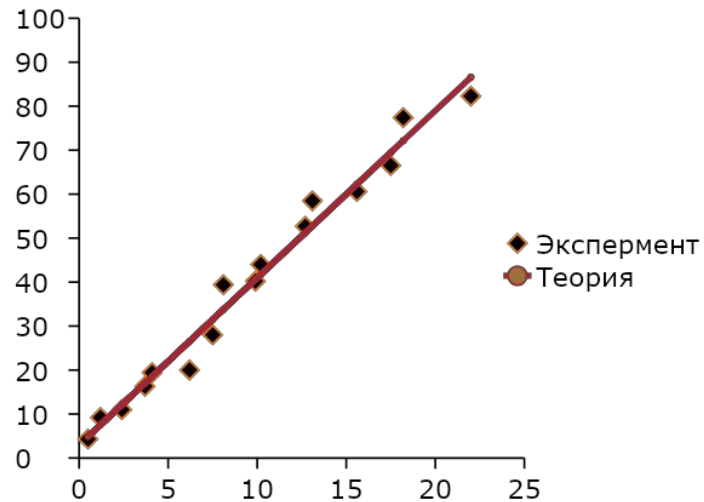


Линейная регрессия

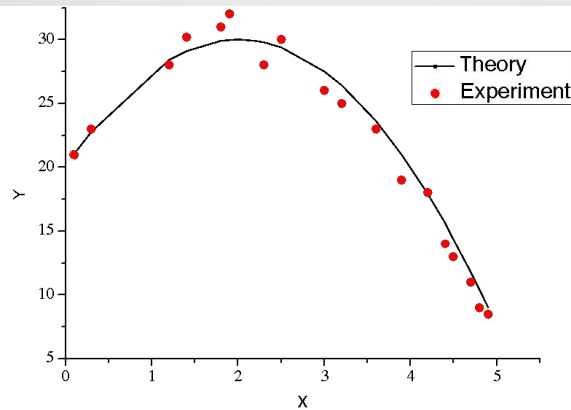
$y = a + b_1x_1 + b_2x_2 + \dots$ – уравнение регрессии



Нелинейная регрессия

$y = a + b_1x_1 + b_2x_2 + \dots$ – уравнение регрессии

$$x_2 = x^2$$

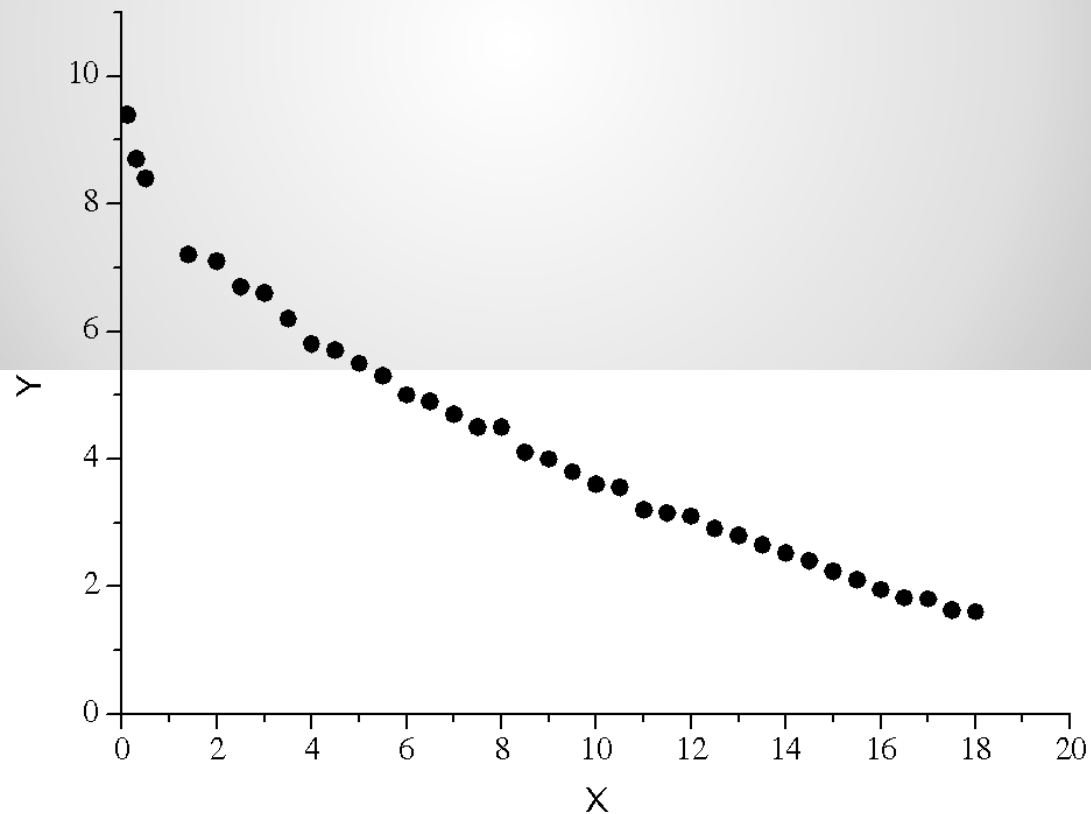


Нелинейная регрессия

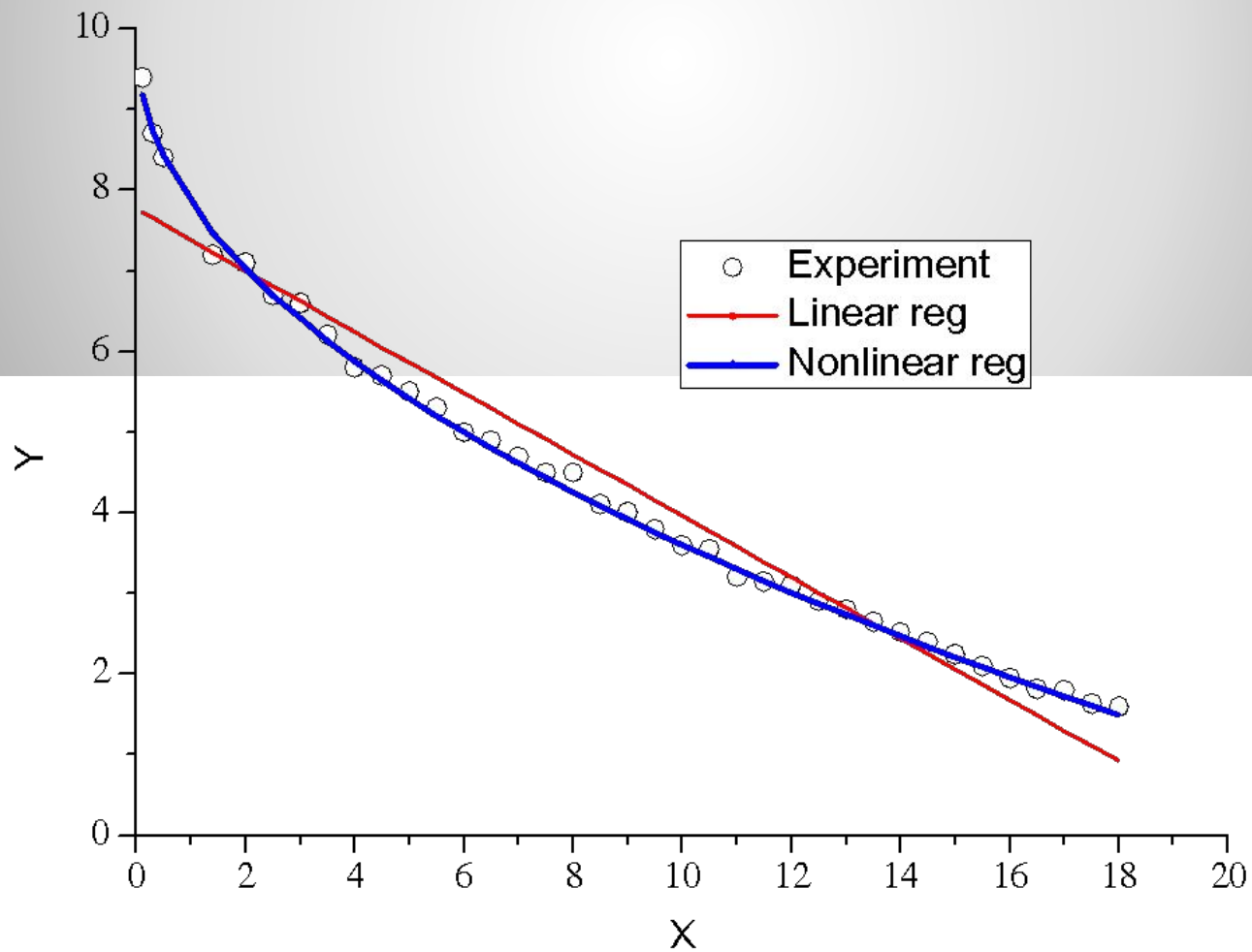
$$y = a + b_1x_1 + b_2x_2 + \dots - \text{уравнение регрессии}$$

№ п/п	Функция	Линеаризующие преобразования			
		Преобразование переменных		Выражения для величин b_0 и b_1	
		y'	x'	b_0'	b_1'
1	$y = b_0 + b_1/x$	y	$1/x$	b_0	b_1
2	$y = 1/(b_0 + b_1x)$	$1/y$	x	b_0	b_1
3	$y = x/(b_0 + b_1x)$	x/y	x	b_0	b_1
4	$y = b_0b_1^x$	$\lg(y)$	x	$\lg(b_0)$	$\lg(b_1)$
5	$y = b_0 \cdot e^{b_1x}$	$\ln(y)$	x	$\ln(b_0)$	b_1
6	$y = 1/(b_0 + b_1e^{-x})$	$1/y$	e^{-x}	b_0	b_1
7	$y = b_0x^{b_1}$	$\lg(y)$	$\lg(x)$	$\lg(b_0)$	b_1
8	$y = b_0 + b_1 \lg(x)$	y	$\lg(x)$	b_0	b_1
9	$y = b_0/(b_1 + x)$	$1/y$	x	b_1/b_0	$1/b_0$
10	$y = b_0x/(b_1 + x)$	$1/y$	$1/x$	b_1/b_0	$1/b_0$
11	$y = b_0e^{b_1/x}$	$\ln(y)$	$1/x$	$\ln(b_0)$	b_1
12	$y = b_0 + b_1x^n$	y	x^n	b_0	b_1

Нелинейная регрессия



Нелинейная регрессия



Планирование эксперимента

Логические основы

Эксперимент – качественный и количественный

Изгиб проволоки

Планирование эксперимента

Логические основы

Пассивный эксперимент – накопление информации

«в режиме нормальной эксплуатации»

Много времени и затрат

Активный эксперимент – накопление информации

«в режиме покачивания», целенаправлено, планомерно изменять условия эксперимента

Быстро накапливать информацию

Основоположник Р. Фишер, 1930-е годы

Планирование эксперимента

Логические основы

Планирование эксперимента — наука о способах составления экономичных экспериментальных планов, которые позволяют извлекать наибольшее количество информации об объекте исследования, способах проведения эксперимента, способах обработки экспериментальных данных и их использования для оптимизации инженерных расчетов и оптимизации производственных процессов¶

Планирование эксперимента

Логические основы

Эксперимент – качественный и количественный

Изгиб проволоки

Фактор – переменная величина, по предположению влияющая на результаты эксперимента

Уровень фактора – фиксированное значение фактора относительно начала отсчета

Способность «вмешиваться» в фактор

Планирование эксперимента

Детерминистические системы

Три группы

контролируемые и управляемые – возможность регистрировать и управлять

(температура отжига, время отжига)

контролируемые, но неуправляемые – возможность регистрировать, но невозможно управлять

(состав примесей)

неконтролируемые – не регистрируются, а иногда, даже не подозреваются

возможность их существования

Планирование эксперимента

Логические основы

Отклик – наблюдаемая переменная, по предположению, зависящая от факторов (число перегибов)

Функция отклика – зависимость математического ожидания отклика от факторов

Истинный вид функции до эксперимента, как правило, неизвестен, поэтому часто задаются

$$y = b_0 + \sum_i b_i x_i + \sum_{i,j} b_{ij} x_i x_j + \dots$$

Планирование эксперимента

Логические основы

Нахождение уравнения регрессии методом планирования экспериментов состоит из следующих этапов:

- выбор основных факторов и их уровней;
- планирование и проведение собственно эксперимента;
- определение коэффициентов уравнения регрессии;
- статистический анализ результатов эксперимента.

Планирование эксперимента Логические основы

Планирование первого порядка

На первой стадии планирование обычно применяют полином первой степени. Для 3-х факторной задачи

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

$b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3$ - парное взаимодействие

Планирование эксперимента

Логические основы

В качестве факторов можно выбирать только **контролируемые** и **управляемые** переменные.

Важнейшее условие – управляемость процессов по каждому фактору, т.е. возможность **независимого** изменения каждого из этих факторов и поддержания его на данном уровне в ходе всего процесса

Для каждого из этих факторов необходимо указать **интервал** изменения параметров.

Планирование эксперимента

Логические основы

В качестве факторов можно выбирать только **контролируемые** и **управляемые** переменные.

Важнейшее условие – управляемость процессов по каждому фактору, т.е. возможность **независимого** изменения каждого из этих факторов и поддержания его на данном уровне в ходе всего процесса

Для каждого из этих факторов необходимо указать **интервал** изменения параметров.

Планирование эксперимента

Интервалом варьирования факторов называется некоторое число (каждое для соответствующего фактора), прибавление которого к основному уровню дает верхний, а вычитание – нижний пределы. Для упрощения записи условий эксперимента и обработки экспериментальных данных масштабы по осям выбираются так, чтобы верхний уровень составлял +1, нижний -1, а основной – 0.

Для факторов с непрерывной областью определения это достигается с помощью преобразования (кодирования) факторов:


$$X_i = \frac{x_i - x_{i0}}{\Delta x_i}.$$

В теории планирования экспериментов показано, что минимально необходимое число уровней факторов на единицу больше порядка уравнения.

Таблица полного факторного эксперимента

Номер опыта	План								Результат y_i
	x_0	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$	
1	+1	-1	-1	-1	+1	+1	+1	-1	y_1
2	+1	+1	-1	-1	-1	-1	+1	+1	y_2
3	+1	-1	+1	-1	-1	+1	-1	+1	y_3
4	+1	-1	-1	+1	+1	-1	+1	+1	y_4
5	+1	+1	+1	-1	+1	-1	-1	-1	y_5
6	+1	+1	-1	+1	-1	+1	-1	-1	y_6
7	+1	-1	+1	+1	-1	-1	-1	-1	y_7
8	+1	+1	+1	+1	+1	+1	+1	+1	y_8

для b_0



Планирование эксперимента

Полный факторный эксперимент

Матрица ПФЭ обладает следующими свойствами:

1) свойство симметричности: алгебраическая сумма элементов вектор-столбца каждого фактора равна нулю (за исключением столбца, соответствующего свободному члену):

$$\sum_{j=1}^n X_{ij} = 0,$$

где i – номер фактора; j – номер опыта;

2) свойство нормирования: сумма квадратов элементов каждого столбца равна числу опытов:

$$\sum_{j=1}^n X_{ij}^2 = n;$$

3) свойство ортогональности: скалярное произведение всех вектор-столбцов (сумма почленных произведений элементов любых двух вектор-столбцов матрицы) равно нулю:

$$\sum_{j=1}^n X_{ij} X_{uj} = 0, \quad i \neq u.$$

Планирование эксперимента

Полный факторный эксперимент

Поскольку результаты наблюдений отклика носят случайный характер, приходится в каждой точке плана проводить не один, а m^* параллельных опытов (обычно $m^*=2\div 4$), осреднение результатов которых, как уже отмечалось, дает возможность уменьшить погрешности оценки истинного значения отклика в $\sqrt{m^*}$ раз.

В каждой серии экспериментов их последовательность рандомизируется, т.е. с помощью таблиц случайных чисел определяется случайная последовательность реализации экспериментов. Рандомизация дает возможность свести эффект некоторого случайного фактора к случайной погрешности. Это позволяет в определенной степени исключить предвзятость и субъективизм исследователя.

Планирование эксперимента

Полный факторный эксперимент

Определение коэффициентов уравнения регрессии

Воспользуемся свойствами ПФЭ для определения коэффициентов уравнения регрессии методом наименьших квадратов $\hat{y} = b_0 + b_1x_1 + b_2x_2$.

$$\Phi = \sum_{j=1}^n (y_j - \hat{y}_j)^2 \rightarrow \min;_{b_i}$$

- (симметричности) $b_0 \sum X_{1j} = 0$;

- (нормирования) $b_1 \sum X_{1j}^2 = nb_1$;

- (ортогональности) $b_2 \sum X_{1j}X_{2j} = 0$;

$$\frac{\partial \Phi}{\partial b_1} = 2 \sum_{j=1}^n (y_j - b_0 - b_1X_{1j} - b_2X_{2j})X_{1j} = 0;$$

$$\sum_{j=1}^n y_j X_{1j} - b_0 \sum_{j=1}^n X_{1j} - b_1 \sum_{j=1}^n X_{1j}^2 - b_2 \sum_{j=1}^n X_{1j}X_{2j} = 0.$$

Планирование эксперимента

Полный факторный эксперимент

$$b_1 = \frac{\sum_{j=1}^n y_j X_{1j}}{n}; \quad b_2 = \frac{\sum_{j=1}^n y_j X_{2j}}{n}; \quad b_0 = \frac{\sum_{j=1}^n y_j X_{0j}}{n}.$$

Планирование эксперимента

Полный факторный эксперимент

$$b_1 = \frac{\sum_{j=1}^n y_j X_{1j}}{n}; \quad b_2 = \frac{\sum_{j=1}^n y_j X_{2j}}{n}; \quad b_0 = \frac{\sum_{j=1}^n y_j X_{0j}}{n}.$$

Следовательно, любые коэффициенты уравнения регрессии определяются скалярным произведением столбца y на соответствующий столбец X .

Можно показать, что аналогичным образом определяются коэффициенты, если в уравнении регрессии учитываются линейные взаимодействия (двойные, тройные):

$$b_{12} = \frac{\sum_{j=1}^n y_j (X_1 X_2)_j}{n}; \quad b_{123} = \frac{\sum_{j=1}^n y_j (X_1 X_2 X_3)_j}{n} \text{ и т.д.}$$

Планирование эксперимента

Полный факторный эксперимент

Следует обратить особое внимание на то, что все линейные коэффициенты независимы, так как в формулы для их расчета входят свои одноименные переменные. Поэтому каждый коэффициент характеризует роль соответствующей переменной в процессе или силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает этот фактор. Если коэффициент имеет знак плюс, то с увеличением значения фактора отклик увеличивается, а если минус – уменьшается.

Планирование эксперимента Полный факторный эксперимент

В результате определения уравнения регрессии может получиться так, что один (или несколько) коэффициентов не очень большие и окажутся незначимыми. Факторы, имеющие коэффициенты, незначимо отличающиеся от нуля, могут быть выведены из состава уравнения, так как их влияние на параметры отклика будет отнесено к ошибке эксперимента. Учитывая ортогональность плана, оставшиеся коэффициенты уравнения регрессии можно не пересчитывать. При отсутствии ортогональности плана эксперимента все коэффициенты необходимо пересчитывать заново.

Планирование эксперимента

Дробный факторный эксперимент