СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Случайным называется **событие**, исход которого при определенном комплексе условий невозможно предсказать.

Случайная величина – это переменная, которая при заданных условиях принимает различные значения

Дискретная случайная величина принимает *только конечное (счётное) число* значений:

$$X\ (x_1, x_2, \ldots, x_n)$$

Вероятность случайной величины – числовая характеристика частоты появления некоторого значения случайной величины при заданных условиях :

$$P\,(X = x_{\mathrm{i}}) = p_{\mathrm{i}} = n_{\mathrm{i}} / \Sigma n_{\mathrm{i}} \qquad \Sigma p_{\mathrm{i}} = 1.$$

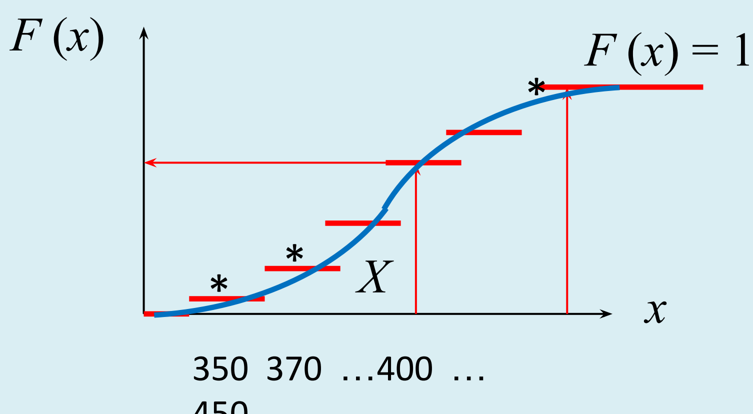
Функция $F\,(x)$, связывающая значения случайной величины с соответствующими им вероятностям, – ***закон распределения*** дискретной случайной величины.

Дискретные случайные величины X и Y считаются *независимыми*, если независимы при любых i и j события $X = x_i$ и $Y = y_{\mathrm{j}}$.

ФУНКЦИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

1. *Закона распределения* – а) таблица «$x_{\mathrm{i}} \div n_{\mathrm{i}}$» или ««$x_{\mathrm{i}} \div p_{\mathrm{i}}$»;
б) функция распределения $F\,(x) = P\,(X < x)$

$\sigma_{0,2}$	350	370	….	450
Число образцов 3	2	10	….	



ФУНКЦИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

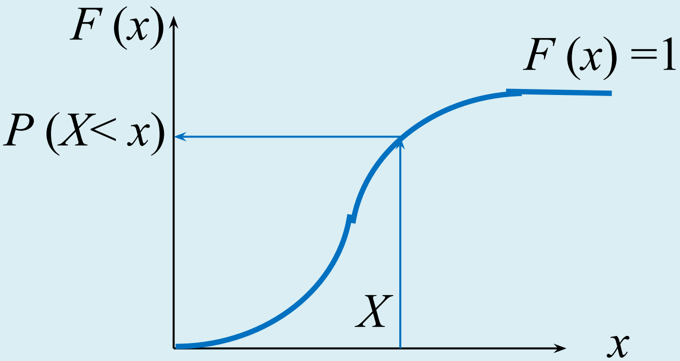
Функция распределения *F(x)* *определяет вероятность того, что случайная величина*

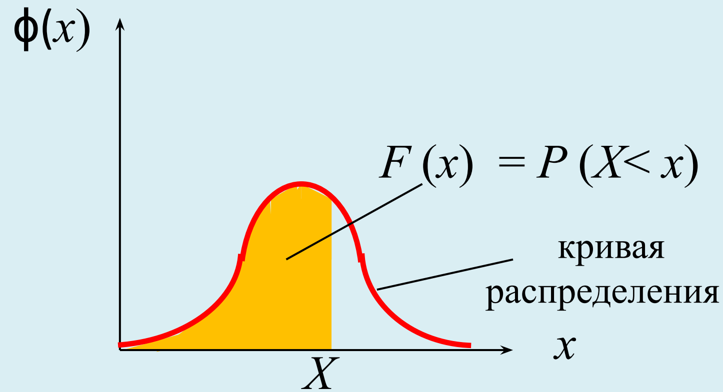
2. *Дифференциальный закона распределения или плотность вероятности:*

***X* принимает значение не больше, чем *x* :**

F(x)=P(X<x).

$$\phi(x) = F'(x)$$





Квантиль выборки – число x_p, ниже которого находится ***p***-я часть выборки.

$$X \le x_{0,7}$$

Характеристики случайной величины

1. ***Математическое ожидание М (Х)*** – это сумма произведений всех значений случайной величины на соответствующие им вероятности :

$$M(X) = \Sigma x_i p_i = \Sigma x_i n_i / n = (1/n)\, \Sigma x_i n_i$$

Пример Определить среднее значение предела текучести стали 40Х

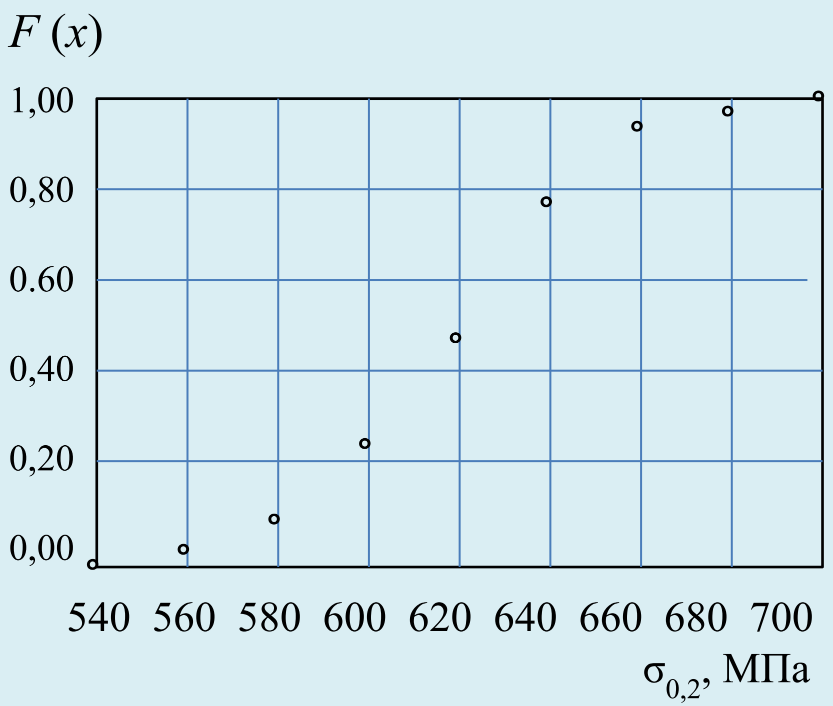
Значение $\sigma_{тi}$ в МПа	540	560	580	600	620	640	660	680	700
Количество образцов в пределах данного интервала	1	3	6	14	23	31	15	5	2
Вероятность p_i попадания значения $\sigma_т$ в данный интервал									
Суммарное число образцов									
Σp_i от i =1 до i = k									

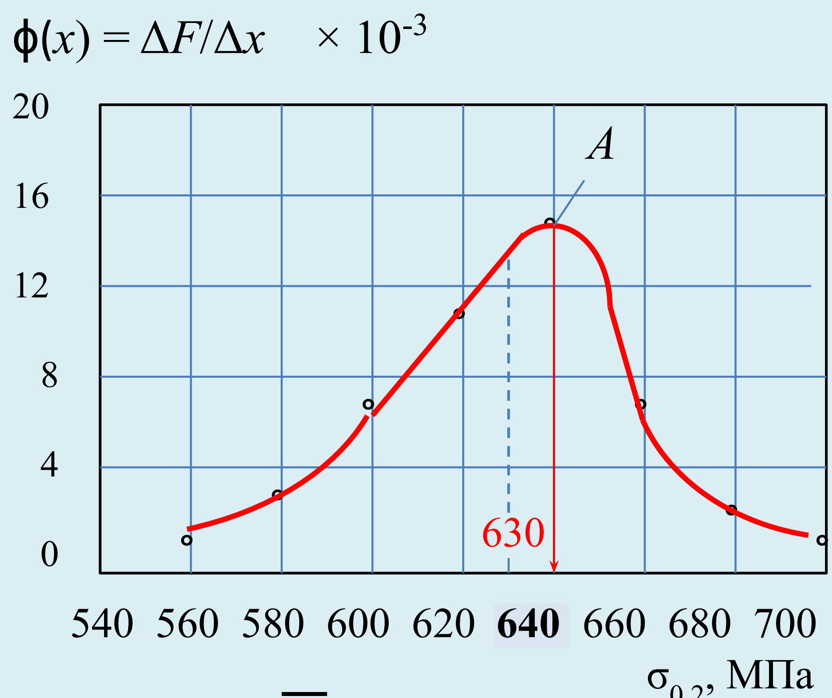
$$M(\sigma_{0,2}) = \overline{\sigma}_{0,2} = 630 \text{ МПа}$$

$$\mathbf{M} = x_{0,5}$$

Пример Нарисовать график $F(x)$ и $\phi(x)$

Значение $\sigma_{\tau i}$ в МПа	540	560	580	600	620	640	660	680	700
Количество образцов в пределах данного интервала	1	3	6	14	23	31	15	5	2
Вероятность p_i попадания значения σ_τ в данный интервал	0,01	0,03	0,06	0,14	0,23	0,31	0,15	0,05	0,02
Суммарное число образцов	1	4	10	24	47	78	93	98	100
Σp_i от $i = 1$ до $i = k$	0,01	0,04	0,10	0,24	0,47	0,78	0,93	0,98	1,00





$M(\sigma_{0,2}) = \overline{\sigma_{0,2}} = 630$ МПа

Мода – значение случайной величины X, соответствующее максимальному значению плотности вероятности . Здесь значение моды 640 МПа.

В данном случае плотность вероятности – несимметрична относительно $M(X)$

Характеристики случайной величины

1) *Математическое ожидание постоянной величины* М (a) = a .

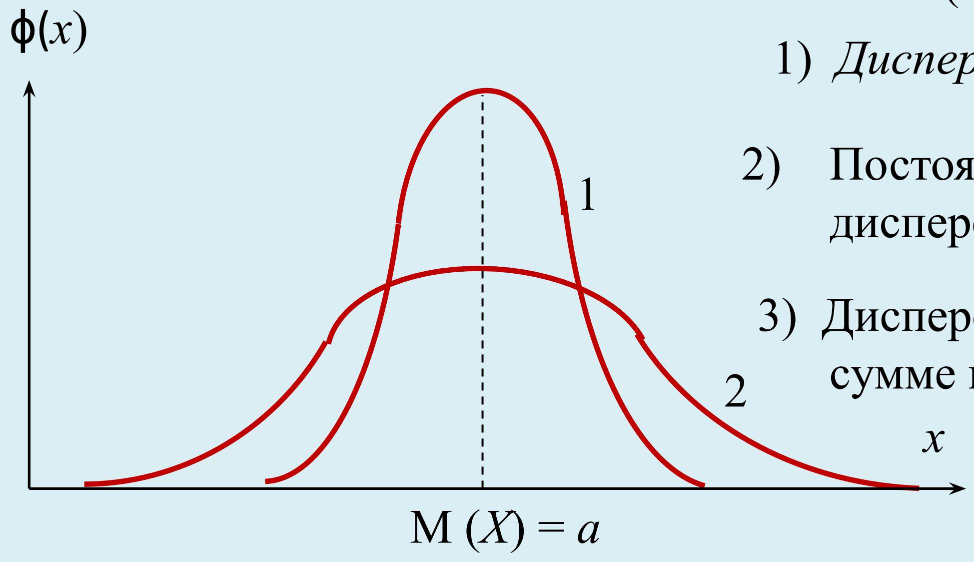
2) *Математическое ожидание* М (kX) = kМ (X), т.е. постоянный множитель можно вынести за символ математического ожидания.

3) *Математическое ожидание суммы* конечного числа случайных величин равно сумме их математических ожиданий: М ($X + Y + Z$) = М (X) + М (Y) + М (Z).

4) *Математическое ожидание произведения* конечного числа случайных величин равно произведению их математических ожиданий: М ($X\,Y\,Z$) = М (X) · М (Y) · М (Z).

2. ***Дисперсия D(X)*** как мера рассеяния значений случайной величины

$$D(X) = (1/n)\{[X - \mathrm{M}\,(X)]^2\} = \Sigma\,(x_i - \mathrm{a})^2 p_i$$



1) *Дисперсия постоянной величины* равна 0.

2) Постоянный множитель можно вынести за знак дисперсии, возведя его в квадрат : $D(kX) = k^2 D(X)$

3) Дисперсия суммы независимых случайных равна сумме их дисперсий: $D(X + Y) = D(X) + D(Y)$

Среднее квадратичное отклонение

$$\sigma(x) = \sqrt{D(X)}$$

Пример Определить дисперсию предела текучести стали 40Х

$$D(X) = (1/n)\{[X - \mathrm{M}(X)]^2\} = \Sigma (x_{\mathrm{i}} - \mathrm{a})^2 p_{\mathrm{i}}$$

Значение $\sigma_{\text{т}i}$ в МПа	540	560	580	600	620	640	660	680	700
Количество образцов в пределах данного интервала	1	3	6	14	23	31	15	5	2
Вероятность p_i попадания значения $\sigma_{\text{т}}$ в данный интервал									
Разность $[\sigma_{\mathrm{i}} - \mathrm{M}(\sigma_{0,2})]$									
$[\sigma_{\mathrm{i}} - \mathrm{M}(\sigma_{0,2})]^2$									
$[\sigma_{\mathrm{i}} - \mathrm{M}(\sigma_{0,2})]^2 p_{\mathrm{i}}$									

$\mathrm{M}(\sigma_{0,2}) = \sigma_{0,2} = 630$ МПа

$D(X) = 816$

Выборочная дисперсия

$S_{мв}^{2}(x) = [n/(n-1)]\, D(X) = 824$

Среднее квадратичное отклонение

$S_{мв}(x) = \sqrt{D(X)\,[(n/n-1)]} = 30{,}4$ МПа

Коэффициент вариации – отношение средне квадратичного отклонения к среднему значению случайной величины:

$\upsilon = S_{мв}(x) / \mathrm{M}(X) = 30{,}4/630 = 0{,}048.$

Критерий согласия Колмогорова – Смирнова

$$\lambda = D \cdot n^{1/2} \leq \lambda_{1-\alpha},$$

где D – максимальная разность вероятностей между экспериментальной и теоретической функцией распределения; n – количество испытанных образцов;
α – уровень значимости; $\lambda_{1-\alpha}$ – критерий Колмогорова:

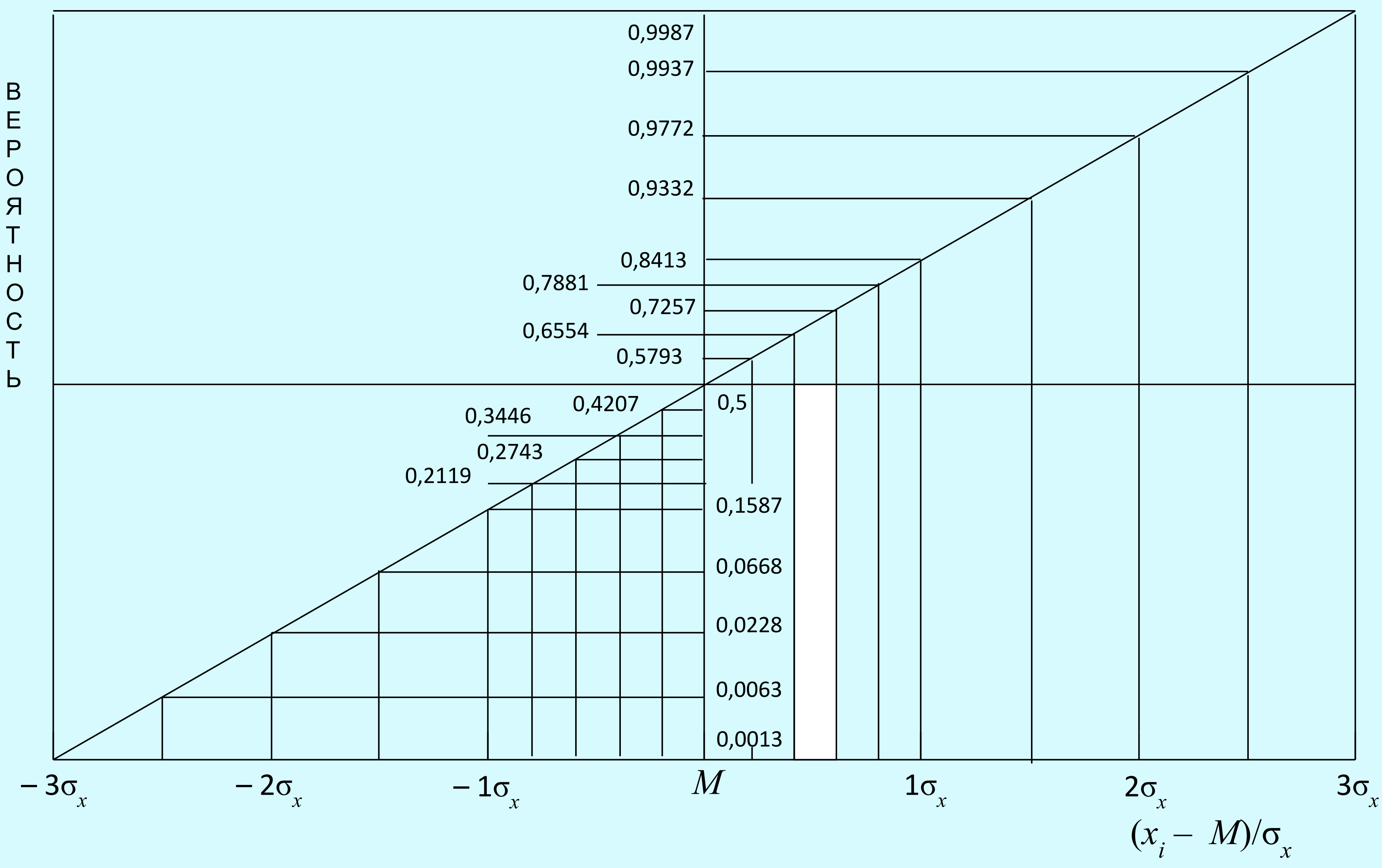
$\lambda_{1-\alpha}$ ……	0,33	0,37	0,44	0,57	0,64	0,83	1,07	1,22
α ………	0,9999	0,999	0,99	0,9	0,8	0,5	0,2	0,1.

Последовательность оценки :

– результаты эксперимента представляют в виде последовательности x_k возрастающих значений (от наименьшего до наибольшего);
– каждому последующему значению этой последовательности приписывается увеличение вероятности отказа на величину $\Delta q = 1/(n + 1)$;
для наименьшего значения при $k = 1$ вероятность отказа $q_1 = 1/(n + 1)$,
для i– ого образца $q_i = i/(n + 1)$;
– вычисляют значения M_x и σ_x;

1. результаты эксперимента x_i (p_i) представляют графически на специальную «вероятностную» сетку, например соответствующую нормальному распределению ; находят значение D, рассчитывают λ и принимают решение о соответствии экспериментального функции распределения теоретической функции.

2. определяют вероятность $p_{i\text{-}н} = 1 - q_i$ дл каждого x_i *согласно нормальному закону* [по $(x_i - M)/\sigma_x$],
затем вычисляют разность $(p_i - p_{i\text{-}н})$ и находят максимальное значение $(p_i - p_{i\text{-}н})_{\max} = D$,
рассчитывают λ и принимают решение о соответствии экспериментального функции распределения теоретической функции.


ВЕРОЯТНОСТЬ
0,9987
0,9937
0,9772
0,9332
0,8413
0,7881
0,7257
0,6554
0,5793
0,5
0,4207
0,3446
0,2743
0,2119
0,1587
0,0668
0,0228
0,0063
0,0013
$-3\sigma_x$
$-2\sigma_x$
$-1\sigma_x$
M
$1\sigma_x$
$2\sigma_x$
$3\sigma_x$
$(x_i - M)/\sigma_x$

Пример

При испытании серии из 24 деталей получены следующие значения долговечности в часах: 1456, 1376, 194, 955, 2066, 1704, 222, 1152, 1832, 362, 1938, 1562, 1179, 1746, 2088, 1258, 1169, 1685, 556, 1538, 1520, 1370, 1284, 1755. Требуется оценить соответствие распределения деталей нормальному закону.

Решени е

t_k, час	$p_k(t))$	p_k	$p_k - p_k(t))$

1 – в первом столбце представить значения долговечности по возрастанию от 194 до 2083 часов;

2 – во втором столбце записать вероятность каждого значения долговечности,
начиная от $q_1 = 1/(n + 1) = 1/(24 +1) = 0{,}04$;

3 –определить математическое ожидание *l* часа,
стандартное отклонение S = 538 часов;

4 –в третий столбец занести вероятности p_k каждого t_k, по нормальному закону распределения с параметрами M и S;

5 –в четвёртом столбце записать разность $p_k - p_k(t))$;

6 – максимальное значение этой разности D = 0,08727;
значение $\lambda = D \cdot n^{1/2} = 0.4275$;
$\lambda < \lambda_{1-\alpha} = 0{,}44$ при $\alpha = 0{,}99$.

$\lambda_{1-\alpha}$ … 0,33 0,37 0,44 0,57 0,64 0,83 1,07 1,22

α … 0,9999 0,999 0,99 0,9 0,8 0,5 0,2 0,1.

КРИТЕРИИ СОГЛАСИЯ теоретического и экспериментального распределений

Критерий согласия Пирсона (χ - квадрат)

Условия

– теоретический закон и его параметры неизвестны;

– параметры теоретического закона определяют по данным эксперимента.

Пусть случайная величина ***Y***, распределена по нормальному закону ***Y***⊂***N***(***a***, σ²).

Случайная величина $U = \frac{Y-a}{\sigma}$ = **χ** распределена по нормальному закону с параметрами ***M***(***U***)= 0 и σ(***U***)= 1, т.е. ***U***⊂ ***N***(0,1).

При ***n*** **≥**30 распределение χ² практически не отличается от нормального закона.

$$\chi^2 = U_1^2 + \ldots + U_n^2 = \sum_{1}^{N}\left(\frac{Y-a}{\sigma}\right)^2$$

, где *N – общее число испытанных образцов*

$\chi^2 \geq \chi^2_{\alpha,k}$ **–** гипотеза распределения по нормальному закону отвергается

***k* = *N* – 3** число степеней свободы

Статистические распределения принято оценивать по значениям их моментов.

Моменты для центрированных распределений *(найденных с исключением систематических составляющих) — центральные.*

Центральный момент *k -го порядка* *для дискретной случайной величины*

$$M_k = (1/n)\{[X - \mathrm{M}(X)]^k\} = \Sigma (x_i - a)^k p_i$$

При $k = 2$ $M_k = D(X) = (1/n)\{[X - \mathrm{M}(X)]^2\} = \Sigma (x_i - a)^2 p_i$ – дисперсия.

M_3 – асимметрия распределения; коэффициент асимметрии $A = M_3 / [\sigma(x)]^3$

$A < 0$ – левосторонняя; $A > 0$ – правосторонняя асимметрия

M_4 – эксцесс распределения; коэффициент эксцесса $E = M_4 / [\sigma(x)]^4 - 3$;

$E = 0$ для нормального закона распределения, а $M_4 / [\sigma(x)]^4 = 3$.

Закон распределения можно считать нормальным при условии:

$|A| \leq A_{1-p} = 3\sqrt{D(A)}$, $p = 0{,}05$ $\qquad$ $|E| \leq E_{1-p} = 5\sqrt{D(E)}$, $p = 0{,}05$

$D(A) = 6(n-1) / \{(n+1)(n+3)\}$ $\qquad$ $D(E) = 24\, n(n-2)(n-3) / \{(n+1)^2 (n+3)(n+5)\}$

СРАВНЕНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

Влияние испытательных машин и толщины листа сплава ПТ3В на малоцикловую долговечность

Лист 39 мм, $\sigma_{0,2} = 65$ кг/мм^2

Номер образца	Число циклов	Номер машины	Номер образца	Число циклов	Номер машины	Номер образца	Число циклов	Номер машины
1	2440	76	11	3455	76	21	2620	84
2	1920	78	12	2462	78	22	3020	76
3	2980	82	13	3180	82	23	2219	78
4	3333	84	14	3140	84	24	2360	82
5	3470	76	15	2890	84	25	2288	78
6	3292	15	16	4290	82	26	1890	82
7	2160	16	17	2887	78			
8	3879	13	18	3315	76			
9	2250	82	19	3240	82			
10	2440	84	20	3160	84			

Лист 130 мм, $\sigma_{0,2} = 64$ кг/мм^2

Номер образца	Число циклов	Номер машины	Номер образца	Число циклов	Номер машины	Номер образца	Число циклов	Номер машины
1	2960	16	11	3530	76	21	2400	16
2	7000 н	15	12	4355	76	22	3320	14
3	2310	14	13	5070	13	23	7250	13
4	3595	13	14	2490	78	24	1400	84
5	2750	84	15	4540	16	25	2700	14
6	4800	82	16	2435	84	26	2660	78
7	2440	16	17	3300	78	27	2860	16
8	3380	14	18	2350	16	28	3725	15
9	2490	16	19	2350	16	29	3305	76
10	2360	78	20	2710	15	30	2800	78

Необходимое число образцов при точности δ и доверительной вероятности p в случае нормального закона распределения

$$n \geq (z_{1-\alpha})^2 \, [\sigma(x) / \delta]^2$$

$z_{1-\alpha}$ – квантиль нормального распределения (при $n = \infty$):

p	0,9	0,95	0,99
$z_{1-\alpha/2}$	1,652	1,967	2,585

Необходимое число образцов

$\delta/\sigma(x)$	$p = 0{,}9$	$p = 0{,}95$	$p = 0{,}99$
1	5 (3)	7(4)	11(7)
0,5	13(11)	19(16)	31(27)
0,1	273 (271)	387(384)	668

Чёрным даны значения n *после* замены $z_{1-\alpha}$ на $t_{1-\alpha,\, n-1}$ (по Стьюденту)

Определите Двусторонний доверительный интервал

$$M \pm t_{1-\alpha,\, n-1} \; \sigma(x)/n^{1/2}$$

Проверка гипотез различия случайных величин

1. *t - критерий* (Стьюдента). Проверка гипотезы равенства $M(X)$ и $M(Y)$

$$t = \frac{M(X) - M(Y)}{\sqrt{(n_1 - 1)\sigma^2_X + (n_2 - 1)\sigma^2_Y}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Гипотеза равенства $M(X) = M(Y)$ не отвергается, если $t < t_{\alpha; k}$, где $k = n_1 + n_2 - 2$ (таб. 1.1.2.8)

2. *F - критерий* (Фишера). Проверка гипотезы равенства $D(X)$ и $D(Y)$

$$F = \sigma^2 (X) / \sigma^2 (Y)$$

Гипотезы равенства дисперсий $D(X) = D(Y)$ не отвергается, если $F < F_{\alpha; m1, m2}$, где $m_1 = n_1 - 1$; $m_1 = n_2 - 1$.
(таб. 1.1.2.10)

3. *Критерий Уилкоксона.* Проверка принадлежности двух выборок к одной и той же генеральной совокупности при неизвестном законе распределения

3.1. Значения $x_1, \ldots x_n$ и $y_1, \ldots y_m$ упорядочиваются вместе по возрастанию.

3.2. Определяется полное число инверсий u вида $y_j < x_i$ в этой последовательности

3.3. а. Гипотеза отвергается, если $| u - ½\ n_1 \cdot n_2 | > u_\alpha$, где u_α по табл. 1.1.2.11.

3.3. б. Гипотеза отвергается, если $| u - ½\ n_1 \cdot n_2 | > u_\alpha$, где $u_\alpha = z_\alpha [n_1 \cdot n_2 (n_1 + n_2 + 1)/ 12]^{1/2}$
z_α по табл. 1.1.2.6.2 из условия $\mathbf{2\Phi(z_\alpha) = 1 - \alpha}$

Оценка статистической значимости различия /сходства выборок значений чисел циклов до разрушения

Толщина 60 мм, $\sigma_{0,2}$ = 65.9 кг/мм^2.

$\sigma/\sigma_{0,2}$ = 1,0	$\sigma/\sigma_{0,2}$ = 0,82	$\sigma/\sigma_{0,2}$ = 0,7	$\sigma/\sigma_{0,2}$ = 0,6
624	1567	3144	7883
684	2322	3047	8000 н.р.
858	2744	1717	4265
		2822	
		4009	
		4012	
		2591	

Толщина 90 мм, $\sigma_{0,2}$ = 63 кг/мм^2

$\sigma/\sigma_{0,2}$ = 1,0	$\sigma/\sigma_{0,2}$ = 0,82	$\sigma/\sigma_{0,2}$ = 0,7	$\sigma/\sigma_{0,2}$ = 0,6
1095	1598	2232	3702
930	1660	2020	6086 н.р.
849	1783	2540	
		2360	

Толщина 130 мм, $\sigma_{0,2}$ = 64 кг/мм^2

$\sigma/\sigma_{0,2}$ = 1,0	$\sigma/\sigma_{0,2}$ = 0,82	$\sigma/\sigma_{0,2}$ = 0,7	$\sigma/\sigma_{0,2}$ = 0,6
995	1550	3620	3490
815	1830	3930	9550 н.р.
1005	2500	3340	7700
		3140	

При испытании серии из 24 образцов получены следующие значения долговечности:
1456, 1376, 194, 955, 2066, 1704, 222, 1152, 1832, 362, 1938, 1562, 1179, 1746, 2088, 1258, 1169, 1685, 556, 1538, 1520, 1370, 1284, 1755.

1. Оценить соответствие экспериментальной функции распределения нормальному закону.
2. Отбросив «худшие» значения, оцените соответствие нового распределения нормальному закону.
3. Сравните исходную выборку с новой, используя t– и F – критерии .

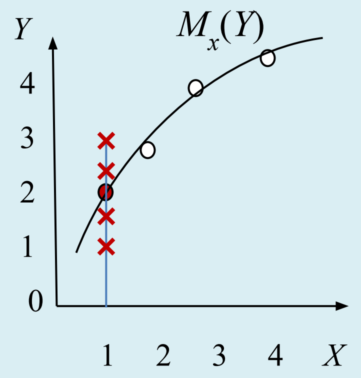
КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Детерминированные и стохастические (статистические) связи.

Корреляционная связь имеет место, если на изменение случайной величины X другая случайная величина Y реагирует изменением своего условного математического ожидания (M).

Условным математическим ожиданием $M_x(Y)$ случайной величины Y называется M этой величины, вычисленное при фиксированном значении другой случайной величины X.

Для дискретного распределения $\overline{Y}_x = M_x(Y) = \Sigma\, Y_i\, p(Y_i\,/X)$,
где $p(Y_i\,/X)$ – условная вероятность значения $Y = Y_i$ при условии $X = x$.



Для непрерывного распределения $\overline{Y}_x = M_x(Y) = \int_{-\infty}^{+\infty} Y\, p_{x/y}\, \mathrm{d}y$,

где $p_{x/y}$ – плотность условного распределения вероятности величины Y при $X = x$.

Уравнение $M_x(Y) = \mathrm{f}(X)$ – ***уравнение регрессии***, или корреляционное уравнение

Корреляционная связь случайных величин

Если есть связь между величинами X и Y, то какова форма этой связи?

Экспериментальные данные представляют в виде корреляционной таблицы
Корреляционная таблица –
а) по горизонтали записывают значения X,
б) по вертикали записывают значения Y,
в) в клетке с индексами (i, j) указывают частоту появления пары ($X_i Y_j$).

$Y \backslash X$	X_1	X_2	X_3	………	X_k	n_{yj}
Y_1	n_{11}	n_{12}	n_{13}	………	n_{1k}	n_{y1}
Y_2	n_{21}	n_{22}	n_{23}	………	n_{2k}	n_{y2}
…	………				…	….
Y_l	n_{l1}	n_{l2}	n_{l3}	………	n_{lk}	n_{yl}
n_{xi}	n_{x1}	n_{x2}	n_{x3}	………	n_{xk}	n

$$\text{Верт.:}\ n_{i1} + n_{i2} + \ldots + n_{il} = \sum_{J=1}^{l} n_{ij} = n_{xi};\ \text{гориз.:}\ n_{j1} + n_{j2} + \ldots + n_{jk} = \sum_{i=1}^{k} n_{ji} = n_{yj}$$

Общее число наблюдений

$$n = \sum_{i=1}^{k} \sum_{J=1}^{l} n_{ij} = \sum_{i=1}^{k} n_{xi} = \sum_{J=1}^{l} n_{yj}$$

Общая средняя арифметическая случайной величины:

$$\overline{X} = \sum X_i n_{xi} / n; \quad \overline{Y} = \sum Y_j n_{yj} / n.$$

Каждому значению X_i переменной X соответствует групповая средняя $\overline{Y}_i$.

Пример

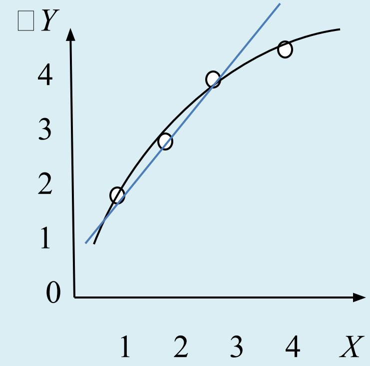
Y \ X	1	2	3	4
1	1	-	-	-
2	2	1	-	-
3	1	2	1	-
4	-	1	2	1
5	-	-	1	2

$x = X_1$ и $\overline{Y}_1 = (1 \cdot 1 + 2 \cdot 2 + 1 \cdot 3 + 0) / (1+2+1) = 2;$

$x = X_2$ и $\overline{Y}_2 = (2 \cdot 1 + 3 \cdot 2 + 4 \cdot 1 + 0) / (1+2+1) = 3;$

$x = X_3$ и $\overline{Y}_3 = (3 \cdot 1 + 4 \cdot 2 + 5 \cdot 1) / (1+2+1) = 4;$

$x = X_4$ и $\overline{Y}_4 = (4 \cdot 1 + 5 \cdot 2) / (1+2) = 4{,}66.$



ПОРЯДОК ВЫЧИСЛЕНИЯ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ

1. Вычисление генеральных средних

$$\overline{X} = \frac{X_1 n_{x1} + X_2 n_{x2} + \ldots + X_k n_{xk}}{n_{x1} + n_{x2} + \ldots + n_{xk}} = \frac{\sum^{k} X_i n_{xi}}{n}$$

$$\overline{Y} = \frac{Y_1 n_{y1} + Y_2 n_{y2} + \ldots + Y_l n_{ylk}}{n_{y1} + n_{y2} + \ldots + n_{yl}} = \frac{\sum^{l} Y_j n_{yj}}{n}$$

2. Вычисление ковариации величин X и Y (вычисление корреляционного момента)

$$К_{xy} = \frac{\sum\sum n_{ij}\,(X_i - \overline{X})\,(Y_j - \overline{Y})}{n} = \overline{XY} - \overline{X} \cdot Y;$$

Если X и Y независимы, то $К_{xy} = 0$.

Однако равенство $К_{xy} = 0$ не означает, что величины X и Y независимы; в этом случае между ними может быть другая, ***нелинейная связь***.

Значение корреляционного момента зависит от дисперсии случайной величины. Так, если между X и Y детерминированная связь, то $К_{xy}$ будет пропорционален дисперсии X.

3. Вычисление коэффициента корреляции

Для исключения влияния дисперсии применяется ***коэффициент корреляции***

$$\boldsymbol{r_{xy} = К_{xy} / (\sigma_x \sigma_y)}$$

$$r_{xy} = К_{xy} / (\sigma_x \sigma_y) = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sqrt{(\overline{X^2} - \overline{X}^2)(\overline{Y^2} - \overline{Y}^2)}};$$

4. Вычисление среднеквадратичного отклонения частного среднего от генерального среднего

$$\sigma_{\overline{Y}(x)} = \sqrt{\frac{\sum n_{xi} (\overline{Y}_{xi} - \overline{Y})^2}{n}}$$

5. Вычисление ***корреляционного отношения*** $\boldsymbol{\eta = \sigma_{\overline{Y}(x)} / \sigma_Y}$
– это отношение межгруппового среднего квадратичного отклонения переменной Y к общему среднему квадратичному отклонению этой переменной.

1. Если $\boldsymbol{r_{xy} = 1}$, то связь X и Y детерминированная: $\boldsymbol{Y = aX}$
2. Чем ближе r_{xy} к 1 (по модулю), тем ближе связь к прямолинейной.
3. Если $r_{xy} = 0$, то между X и Y нет прямолинейной корреляционной связи.
4. Если $\boldsymbol{\eta = 0}$, то между X и Y ***нет вообще никакой связи.***
5. Если $\eta \approx 1$, то между X и Y имеет место детерминированная связь.
6. Если $\eta \approx |r_{xy}|$, то связь между X и Y близка к линейной. Верно и обратное: если связь между X и Y линейная, то $\eta \approx |r_{xy}|$.
7. Если $|r_{xy}|(n-1)^{1/2} \geq 3$, то связь между случайными величинами X и Y достаточно вероятна.

Пример

$Y \backslash X$	1	2	3	4	n_y
1	1	-	-	-	1
2	2	1	-	-	3
3	1	2	1	-	4
4	-	1	2	1	4
5	-	-	1	2	3
n_x	4	4	4	3	15

1. статистические средние

$\bar{X} = (1\cdot 4 + 2\cdot 4 + 3\cdot 4 + 4\cdot 3)/15 = 36/15 = 2{,}4$

$\bar{Y} = (1\cdot 1 + 2\cdot 3 + 3\cdot 4 + 4\cdot 4 + 5\cdot 3)/15 = 3{,}33$

2. средние квадратов

$\overline{X^2} = (1^2\cdot 4 + 2^2\cdot 4 + 3^2\cdot 4 + 4^2\cdot 3)/15 = 6{,}93$

$\overline{Y^2} = (1^2\cdot 1 + 2^2\cdot 3 + 3^2\cdot 4 + 4^2\cdot 4 + 5^2\cdot 3)/15 = 12{,}53$

3. ковариация $\overline{XY} = (X_1Y_1\,n_{11} + X_1Y_2\,n_{12} + \ldots)1/n =$
$= (1\cdot 1\cdot 1 + 1\cdot 2\cdot 2 + 1\cdot 3\cdot 1 + 2\cdot 2\cdot 1 + 2\cdot 3\cdot 2 +$
$+ 2\cdot 4\cdot 1 + 3\cdot 3\cdot 1 + 3\cdot 4\cdot 2 + 3\cdot 5\cdot 1 + 4\cdot 4\cdot 1 + 4\cdot 5\cdot 2)1/15$
$= 9{,}06$

4. корреляционный момент

$K_{xy} = \overline{XY} - \bar{X}\bar{Y} = 9{,}06 - 2{,}4\cdot 3{,}33 = 1{,}146$

5. среднеквадратичные отклонения

$\sigma_x = \sqrt{\overline{X^2} - (\bar{X})^2} = (6{,}93 - 2{,}4^2)^{1/2} = 1{,}08;$ $\sigma_y = \sqrt{\overline{Y^2} - (\bar{Y})^2} = (12{,}53 - 3{,}33^2)^{1/2} = 1{,}2$

6. коэффициент корреляции $r_{xy} = K_{xy}/(\sigma_x\,\sigma_y) = 1{,}146/(1{,}08\cdot 1{,}2) = 0{,}888$

7. вычисление средних $\bar{Y}_j$:

$X_1 = 1$ $\bar{Y}_1 = (1\cdot 1 + 2\cdot 2 + 3\cdot 1)/4 = 8/4 = 2;$

$X_2 = 2$ $\bar{Y}_2 = (1\cdot 0 + 2\cdot 1 + 3\cdot 2 + 4\cdot 1)/4 = 12/4 = 3;$

$X_3 = 3$ $\bar{Y}_3 = (3\cdot 1 + 4\cdot 2 + 5\cdot 1)/4 = 16/4 = 4;$

$X_4 = 4$ $\bar{Y}_4 = (4\cdot 1 + 5\cdot 2)/4 = 14/3 = 4{,}66$

8. среднеквадратичное отклонение локального среднего от генерального среднего

$$\sigma_{\bar{Y}(x)} = \sqrt{\frac{\sum n_{xi}(\bar{Y}_{xi} - \bar{Y})^2}{n}} = 0{,}99$$

9. корреляционное отношение $\eta = \sigma_{Y(x)}/\sigma_Y = 0{,}99/1{,}20 = 0{,}826.$

Вывод. Между X и Y линейная связь, это достоверно: $|r_{xy}|\sqrt{(n-1)} = 0{,}888\cdot 3{,}741 = 3{,}32 > 3.$

УРАВНЕНИЕ РЕГРЕССИИ

Линейное уравнение. Метод наименьших квадратов

Пусть согласно корреляционному анализу случайные величины связаны линейно:

$$Y = a\,X + b.$$

Требуется подобрать параметры уравнения регрессии a и b, чтобы суммарная квадратичная ошибка была наименьшей. Это метод наименьших квадратов.

Назовём «неувязкой» отклонение Y от $\square Y$, определяемого уравнением регрессии:

$$\varepsilon_I = \square Y_i - Y_i$$

Составим квадратичную форму из этих «неувязок» по всем возможным X

$$S(a,\, b) = \sum_i^N [(aX_i + b) - Y_i]^2$$

Минимизируем её по a и b: $\partial S / \partial a = 0$, $\partial S / \partial b = 0$

Решив систему двух уравнений, получим

$$a = К_{xy} / \sigma_x^{\,2} = r_{xy}\, \sigma_y / \sigma_x, \quad b = \square Y - \square X\, r_{xy}\, \sigma_y / \sigma_x.$$

Уравнение регрессии

$$y = \square Y + (X - \square X)\, r_{xy}\, \sigma_y / \sigma_x.$$

Это уравнение прямой через точки $\square X$ и $\square Y$ под углом к оси абсцисс, тангенс которого равен $r_{xy}\, \sigma_y / \sigma_x$.

Определение параметров прямой линии регрессии по сгруппированным данным многократных наблюдений

Пример. Y – процентное содержание углерода в стали, $X = \sigma_т / \sigma_в$

$X \backslash Y$	0,5	0,6	0,7	0,8	n_{yj}
0,5	0	2	0	8	10
0,6	0	4	2	9	15
0,7	2	12	3	1	18
0,8	21	14	0	0	35
0,9	1	0	0	0	1
n_y	24	32	5	18	n =79

1.Генеральные средние:

$\bar{X} = (0{,}5 \cdot 10 + 0{,}6 \cdot 15 + 0{,}7 \cdot 18 + 0{,}8 \cdot 35 + 0{,}9 \cdot 1)/79 = 0{,}703;$

$\bar{Y} = (0{,}5 \cdot 24 + 0{,}6 \cdot 32 + 0{,}7 \cdot 5 + 0{,}8 \cdot 18) / 79 = 0{,}622.$

2. Средние квадратов и ковариаций:

$\overline{X^2} = (0{,}5^2 \cdot 10 + 0{,}6^2 \cdot 15 + 0{,}7^2 \cdot 18 + 0{,}8^2 \cdot 35 + 0{,}9^2 \cdot 1)/ 79 = 0{,}505;$

$\overline{Y^2} = (0{,}5^2 \cdot 24 + 0{,}6^2 \cdot 32 + 0{,}7^2 \cdot 5 \;\; 0{,}8^2 \cdot 18) /79 = 0{,}398;$

$\overline{XY} = (0{,}5 \cdot 0{,}6 \cdot 2 + 0{,}5 \cdot 0{,}8 \cdot 8 + 0{,}6 \cdot 0{,}6 \cdot 4 + 0{,}6 \cdot 0{,}7 \cdot 2 + 0{,}6 \cdot 0{,}8 \cdot 9 + 0{,}7 \cdot 0{,}5 \cdot 2 + 0{,}7 \cdot 0{,}6 \cdot 12 + + 0{,}7 \cdot 0{,}7 \cdot 3 + 0{,}7 \cdot 0{,}8 \cdot 1 + 0{,}8 \cdot 0{,}5 \cdot 21 + 0{,}8 \cdot 0{,}6 \cdot 14 + 0{,}9 \cdot 0{,}5 \cdot 1) /79 = 0{,}427.$

3. Определение дисперсии и корреляционного момента:

$\sigma_x^2 = \overline{X^2} - (\bar{X})^2 = 0{,}505 - 0{,}703^2 = 0{,}012;\ \sigma_x = 0{,}11;$

$\sigma_y^2 = \overline{Y^2} - (\bar{Y})^2 = 0{,}398 - 0{,}622^2 = 0{,}011;\ \sigma_y = 0{,}105;$

$K_{xy} = \overline{XY} - \bar{X}\,\bar{Y} = 0{,}427 - 0{,}703 \cdot 0{,}622 = -0{,}01.$

4. Коэффициент корреляции

$r_{xy} = K_{xy} /(\sigma_x \sigma_y) = -0{,}01/ (0{,}11 \cdot 0{,}105) = = -0{,}867.$

5. Значение $|r_{xy}| \sqrt{(n-1)} = 0{,}867 \sqrt{79-1} = 7{,}66 > 3$, связь весьма вероятна

6. Уравнение регрессии $X - \bar{X} = (Y - \bar{Y})\, r_{xy}\, \sigma_x / \sigma_y$; $X - 0{,}703 = -(Y - 0{,}622) \cdot 0{,}908.$

Определение параметров нелинейной связи между величинами

$\overline{Y} = a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k$ – параболы различного порядка

$\overline{Y} = a / x + b$ – гипербола

$\overline{Y} = b \cdot a^x$ – показательная функция

Уравнение параболической регрессии *Y* на *X* (парабола 2-го порядка) $\overline{Y} = a_0 + a_1 x + a_2 x^2$

Квадратичная форма «неувязок» по всем значениям X $\quad S = \sum_{i=1}^{k} [(a_0 + a_1 X_i + a_2 X_i^2 - Y_i]^2$

Минимуму функции $S(a_0, a_1, a_2)$ соответствуют три уравнения

$$\partial S/\partial a_0 = 0;\ \partial S/\partial a_1 = 0;\ \partial S/\partial a_2 = 0$$

$$a_0 \sum_{i=1}^{k} n_{xi} + a_1 \sum_{i=1}^{k} n_{xi} X_i + a_2 \sum_{i=1}^{k} n_{xi} X_i^2 = \sum_{i=1}^{k} n_{xi} \overline{Y_i}$$

$$a_0 \sum n_{xi} X_i + a_1 \sum n_{xi} X_i^2 + a_2 \sum n_{xi} X_i^3 = \sum n_{xi} X_i \overline{Y_i}$$

$$a_0 \sum n_{xi} X_i^2 + a_1 \sum n_{xi} X_i^3 + a_2 \sum n_{xi} X_i^4 = \sum n_{xi} X_i^2 \overline{Y_i}$$

Пример. X – глубина орошения в *см*;
Y – урожайность в *ц* с гектара.

1. Вычисление Y_i :
$Y_1 = (10 \cdot 4 + 12 \cdot 1)/5 = 10,4$

X \ Y	10	12	14	16	n_{xj}
0	4	1	-	-	5
10	-	2	3	2	7
20	-	1	4	4	9
30	-	2	2	3	7
40	-	2	3	1	6
50	2	2	2	-	6
n_y	6	10	14	10	n =40

Таблица значений слагаемых уравнений

Y / X	10	12	14	16	n_{xi}
0	4	1	-	-	5
10	-	2	3	2	7
20	-	1	4	4	9
30	-	2	2	3	7
40	-	2	3	1	6
50	2	2	2	-	6
n_y	6	10	14	10	n =40

1. Вычисление Y_i :

$\overline{Y}_1 = (10 \cdot 4 + 12 \cdot 1)/5 = 10,4$

$\overline{Y}_2 = (12 \cdot 2 + 14 \cdot 3 + 16 \cdot 2)/7 = 14$

……………………………………

X	n_{xi}	$\overline{Y}_i$	$n_{xi} X_i$	$n_{xi} X_i^2$	$n_{xi} X_i^3$	$n_{xi} X_i^4$	$n_{xi} \overline{Y}_i$	$X_i \overline{Y}_i$	$X_i^2 \overline{Y}_i$
0	5	10,4	0	0	0	0	52	0	0
10	7	14	70	700	7000	70000	98	980	9800
20	9	132/9	180	3600	72000	1440000	132	2640	52800
30	7	100/7	210	6300	189000	5670000	100	3000	90000
40	6	41/3	240	9600	384000	15360000	82	3280	131200
50	6	12	300	150000	750000	37500000	72	3600	180000
Итого	40	-	1000	35200	1402000	60040000	536	13500	463800

$40\,a_0 + 1000\,a_1 + 35200\,a_2 = 536;$

$1000\,a_0 + 35200\,a_1 + 1402000\,a_2 = 13500;$

$35200 a_0 + 1402000\,a_1 + 60040000\,a_2 = 463800$

$a_0 + 25\,a_1 + 800\,a_2 = 13,4;$

$a_0 + 35,2\,a_1 + 1492\,a_2 = 13,5;$

$a_0 + 39,83\,a_1 + 1705,68\,a_2 = 13,18$

Результат: a_0 =10,9575; a_1= 0,2913; $a_2 \approx -0,0055$.

Уравнение $Y_x = 10,9575 + 0,2913x - 0,0055x^2$

Уравнение гиперболической регрессии Y на X

В уравнение $Y = a / x + b$ введём новую переменную $z = 1/ x$.

$$a\sum_{i=1}^{k} n_{xi} \cdot 1/X_i + b\sum_{i=1}^{k} n_{xi} = \sum_{i=1}^{k} n_{xi}\overline{Y_i}$$

$$a\sum n_{xi} \cdot 1/X_i^2 + b\sum n_{xi} \cdot 1/X_i = \sum n_{xi}\overline{Y_i} \cdot 1/X_i.$$

Пример. X – объём продукции;
Y – себестоимость единицы изделий.

X \ Y	100	110	120	130	n_{xi}
50	-	-	1	3	4
100	-	3	3	-	6
150	-	6	2	1	9
200	1	4	-	1	6
250	4	1	-	-	5
n_y	5	14	6	5	n =30

Уравнение регрессии Y на X в виде показательной функции $\overline{Y} = b \cdot a^x$

Логарифмируя, получим $\lg \overline{Y} = x \cdot \lg a + \lg b$ – линейная функция с коэффициентами $\lg a$ и $\lg b$

X_i	n_{xi}	$1/X_i$	$n_{xi} \cdot 1/X_i$	$n_{xi} \cdot 1/X_i^2$	$\overline{Y_i}$	$n_{xi}\overline{Y_i}$	$n_{xi}\overline{Y_i} \cdot 1/X_i$.
50	4	0,02	0,08	0,0016	127,5	510	10,2
100	6	0,01	0,06	0,0006	115	690	6,9
150	9	0,007	0,063	0,00042	114,4	1030	6,87
200	6	0,005	0,03	0,00015	111,7	670	3,35
250	5	0,004	0,02	0,00008	102	510	2,04
Итого	30	-	0,25	0,00283	-	3410	24,36

$0,25\ a + 30\ b = 3410;$
$0,00283\ a + 0,25\ b = 29,36;$

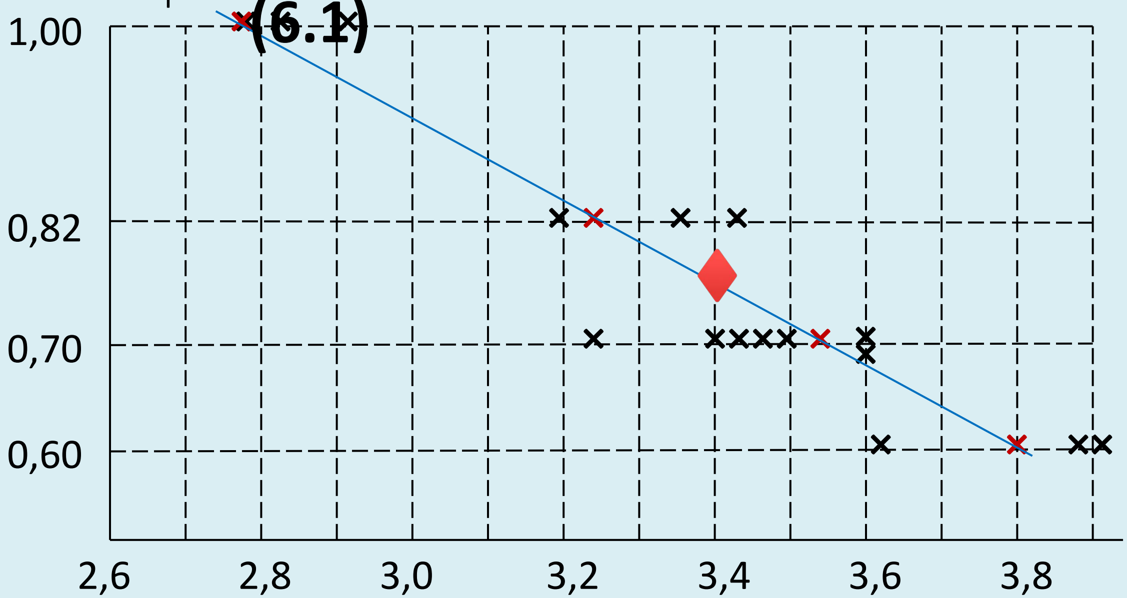
Коэффициенты

$a = 112,8;\ b = 103,2.$

Уравнение

$Y_x = 112,8 / x + 103,2.$

Сплав ПТ3В (6.1)



$\sigma/\sigma_т = C - m \lg N$

$f(X) \quad = \quad Y$

$\sigma/\sigma_т$	$\lg N$	N	$N_{эк}$
1,00	2,777	598	722
0,82	3,234	1713	2211
0,70	3,546	3517	3049
0,60	3,803	6347	6716

$\overline{X} = 0{,}76; \quad \overline{Y} = 3{,}392$

$X - \square X = (Y - \square Y)\, r_{xy}\, \sigma_x / \sigma_y ;$ $\lg N = 2{,}564(2{,}083 - \sigma/\sigma_т)$

$r_{xy} = -0{,}9232$

$\sigma_y = 0{,}3162; \quad \sigma_y^2 = 0{,}10$

$\sigma_x = 0{,}1336; \quad \sigma_x^2 = 0{,}0179$

$\sigma^2_{y} = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ – относительно среднего

$\sigma^2_{y\text{-}ост} = \dfrac{1}{n-l}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – относительно расчётного

$\sigma^2_{y\text{-}ост} = 0{,}017$

$\hat{y}_i$ – значение y по $f(X)$

$F = \sigma^2_y / \sigma^2_{y\text{-}ост} = 5{,}88 > F_{\alpha;\, m\,1,\, m2} = 3{,}36_{\alpha=0{,}01}$

$l = k + 1$; k – число коэффициентов в $f(X)$

$\rho_{xy} = [\,(\sigma^2_y - \sigma^2_{y\text{-}ост})/\sigma^2_y\,]^{1/2} = 0{,}91$ – выборочное корреляционное отношение

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА ПО ПЛАНИРОВАНИЮ ЭКСПЕРИМЕНТА

1. **Лунёв В.А.** Технический эксперимент. Изд-во Политехн. ун-та, 1985
2. **Востров В.Н., Кункин С.Н.. Кузнецов П.А**. Эксперимент- Планирование. СПБ.: Изд-во Политехн. ун-та, 2014.
3. **Налимов В.В., Голикова Т.И.** Логические основания планирования эксперимента. М.: Металлургия, 1981. 152 с.
4. **Кассандрова О.Н., Лебедев В.В.** Обработка результатов наблюдений. М.:Изд-во «Наука», 1970. 104 с.
5. **Новик Ф.С.** Математические методы планирования эксперимента. Учеб. пособие. М.: Изд-во МИСиС, 1979. 73 с.
6. **Пустыльник Е.И.** Статистические методы анализа и обработки наблюдений. М.:Наука, 1968. 285 с.
7. **Спирин Н.А. , Лавров В.В.** Методы планирования и обработки результатов инженерного эксперимента. Изд-во УГЕУ-УПИ, 2004. 257 с.
8. **Астахова Л.Г.** Математическая теория планирования эксперимента. Учеб. пособие. Владикавказ : Изд-во Северо-Кавказского горно-металлургического института. 2013. 96 с.

Теория и методы инженерного эксперимента: Курс лекций/**Н.Г.Бойко, Т.А. Устименко.** Донецк, ДонНТУ, 2009г. 158с.

Методы планирования и математической обработки результатов инженерного эксперимента: Конспект лекций/**Н.А. Спирин, В.В. Лавров**. Екатеринбург: УГТУ-УПИ, 2004г. 257 с.

При **пассивном эксперименте** существуют только факторы в виде входных **контролируемых**, но **неуправляемых** переменных, и экспериментатор находится в положении пассивного наблюдателя.

Задача планирования в этом случае сводится к оптимальной организации сбора информации (выбор количества и частоты измерений, выбор метода обработки результатов измерений). Социологические, явления природы, исследования в астрономии.

Наиболее часто целью пассивного эксперимента является построение математической модели объекта:

-в виде детерминированных функций для хорошо организованного объекта,

- в виде статистической модели для плохо организованного или диффузного объекта.

В **однофакторном пассивном** эксперименте выполняют n пар измерений в дискретные моменты времени единственного входного параметра x и соответствующих значений выходного параметра y. Аналитическая зависимость математического ожидания y от значения x называется *регрессионной*.

Критерии выбора аппроксимирующей функции: простота, удобство пользования, обеспечение требуемой точности аппроксимации, адекватность. (См. выше)

Многофакторный пассивный эксперимент дает n значений выходного параметра y объекта, соответствующих измерениям k совокупностей значений выходных параметров:

$$
\begin{array}{ll}
x_{11}, x_{12}, \ldots, x_{1k}; & y_1 \\
x_{21}, x_{22}, \ldots, x_{2k}; & y_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots & \ldots \\
x_{n1}, x_{n2}, \ldots, x_{nk}., & y_n
\end{array}
$$

где x_{ij} – значение i входного параметра ($i = 1, 2, \ldots, n$) в j–м измерении ($j = 1,2,\ldots,k$).

Модель «Экзаменационная оценка – характеристики студента» дисциплины Высшая математика

Y = 0,143 ПА+ 0,08 ПЯ – 0,078 ДЕ +0,082 СО +0,056 СП – 0,037 НП + 0,034 НА - 0,038 ОБ + 0,045УВ –
– 0,129 СЛ -0,03 ШК + + 0,01 РА –0,178 ИП - 0,101 В

R = 0,979

Экзаменационные оценки	2,75	4	2,6	2,75	3,33	2,75	2,6	4	2,8	3	2,67	5	2,75	2,6	2,75	2,75	2,5	2,5
Расчётные оценки	2,65	4,07	2,66	2,73	3,56	2,67	2,63	4,04	2,65	2,98	2,73	4,87	2,78	2,73	2,75	2,73	2,55	2,65

Модель «Экзаменационная оценка – характеристики студента» дисциплины Теория машин и механизмов

R = 0,989

Экзаменационные оценки	3	5	3	4	3	4	3	5	3	3	4	5	3	2,5	3	3	4	3	3
Расчётные оценки	2,87	5,04	2,96	4,16	3,07	3,98	3,15	5,07	2,99	2,99	3,82	5,02	2,99	2,65	2,95	3,02	4,05	3,24	3,19

коэфф.корр.	эффективн	понять др.	Исполнит.	Воспитан.	тв.воля	ответств.	рац.реш.	Професс.	Честн
самодис	-0,1303	-0,21718	0,016471	-0,22348	0,333489	0,436265	0,23684	-0,34184	-0,25497
Доброс.	-0,0319	0,225735	-0,2556	0,179935	**-0,278**	0,049555	0,161032	0,144003	0,050719
воспитан	0,0769	0,103881	-0,23209	**0,62**	-0,14515	-0,43	-0,09684	-0,30949	**0,63**
целеустр	0,099229	0,274409	0,337362	-0,39826	-0,35751	0,138381	0,199596	0,06129	-0,21349
Убеж.в пол.	-0,28396	0,017555	0,043315	0,04621	0,374606	-0,11485	**0,55**	-0,45183	0,162288
общител	0,14757	**0,44**	0,311957	0,005046	-0,40509	-0,04006	-0,0947	0,291181	-0,28875
прогноз послед	-0,11751	0,28943	-0,13913	-0,05419	- 0,44	0,27358	0,023087	-0,14986	-0,14619
воля	0,108041	-0,11191	0,095192	-0,16105	0,381931	0,26846	-0,34348	-0,03255	-0,21442
эффект	-0,07082	0,196312	-0,25396	-0,25396	-0,22419	-0,45	0,044499	0,132051	0,183946
работосп	0,090871	-0,07368	**-0,42038**	0,127584	-0,25212	-0,03564	-0,1566	**0,54**	**0,45**
решител	**0,48**	- **0,51**	-0,28044	-0,35817	0,22819	0,188245	-0,22364	**0,48**	-0,20614
самостоя	-0,37799	-0,43344	-0,29748	-0,14786	**0,58**	0,084738	0,214624	-0,06707	-0,05313
лог. мыш	0,004722	-0,15925	- **0,50**	-0,38738	0,120767	0,165558	**0,397056**	0,110511	-0,09882
ответств	0,27732	-0,20809	0,398252	**-0,38568**	0,180734	0,398834	- **0,45**	0,286402	-0,30633
твёрд. мн	-0,16528	0,192462	0,253133	0,317657	0,22678	-0,25485	0,093204	- **0,57**	-0,03748
быстр. переор.	-0,30444	0,148266	0,390405	0,014184	0,032939	-0,11212	0,21315	-0,05424	0,014014

коэфф.корр.	эффективн	понять др.	Исполнит.	Воспитан.	тв.воля	рац.реш.	Професс.	Честн
воспитан	0,0769	0,103881	-0,23209	**0,62**	-0,14515	-0,09684	-0,30949	**0,63**
Убеж.в пол.	-0,28396	0,017555	0,043315	0,04621	0,374606	**0,55**	-0,45183	0,162288
общител	0,14757	**0,44**	0,311957	0,005046	-0,40509	-0,0947	0,291181	-0,28875
прогноз послед	-0,11751	0,28943	-0,13913	-0,05419	- 0,44	0,023087	-0,14986	-0,14619
работосп	0,090871	-0,07368	-**0,42038**	0,127584	-0,25212	-0,1566	**0,54**	**0,45**
решител	**0,48**	- **0,51**	-0,28044	-0,35817	0,22819	-0,22364	**0,48**	-0,20614
самостоя	-0,37799	-0,43344	-0,29748	-0,14786	**0,58**	0,214624	-0,06707	-0,05313
лог.мыш	0,004722	-0,15925	- **0,50**	-0,38738	0,120767	**0,397056**	0,110511	-0,09882
ответств	0,27732	-0,20809	0,398252	-**0,38568**	0,180734	- **0,45**	0,286402	-0,30633
твёрд.мн	-0,16528	0,192462	0,253133	0,317657	0,22678	0,093204	- **0,57**	-0,03748
выдержк	0,010675	-0,11813	**0,45**	0,177578	0,136061	-0,14084	-0,30595	-0,12897

ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

При решении задач оценки влияния технологических или эксплуатационных условий приходится учитывать большое число предполагаемых значимых факторов. Кроме того, каждый фактор в эксперименте может принимать одно и более значений, называемых уровнями.

Число возможных состояний N равно числу уровней p, возведённому в степень k, т. е. $N = p^k$, где k - *число факторов.*

При $k = 10$ и $p = 3$ число состояний $3^{10} = 59049$.
Задача постановки такого эксперимента практически невыполнима.

Планирование эксперимента

– это процедура выбора ***числа опытов*** и условий ***их проведения***, необходимых для решения поставленной задачи с ***требуемой точностью***:

- стремление к минимизации общего числа опытов;
- одновременное варьирование всеми переменными по специальным правилам;
- использование специального математического аппарата, формализующего действия экспериментатора;
- выбор чёткой стратегии, позволяющей принимать обоснованные решения;
- минимизация ошибки (неопределённости) эксперимента

(англ. статистик Рональду Фишеру в конце 1920-ых)

Пример взвешивания трёх объектов

Традиционная схема взвешивания | Планирование эксперимента при взвешивании

Номер опыта	*a*	*b*	*c*	Результат взвешивания	Номер опыта	*a*	*b*	*c*	Результат взвешивания
1	-1	-1	-1	y_0	1	+1	-1	-1	y_1
2	+1	-1	-1	y_1	2	-1	+1	-1	y_2
3	-1	+1	-1	y_2	3	-1	-1	+1	y_3
4	-1	-1	+1	y_3	4	+1	+1	+1	y_4

Примечание.

1. «-1» означает отсутствие груза на весах, «+1» - наличие груза на весах.
2. *a, b ,c* – взвешиваемые грузы.

1. Вес груза *a* при ТрСВ равен $A = y_1 - y_0$; дисперсия результата $\sigma^2\{ A \} = \sigma^2\{ y_1 - y_0\} = 2\sigma^2\{ y\}$.

2. Вес каждого груза по схеме ПЭ определяется по формулам:

$A = ½\,(y_1 - y_2 - y_3 + y_4)$; $B = ½\,(-y_1 + y_2 - y_3 + y_4)$; $C = ½\,(-y_1 - y_2 + y_3 + y_4)$.

Дисперсия взвешивания $\sigma^2\{ A \} = \sigma^2\{½\,(y_1 - y_2 - y_3 + y_4)\} = 4\sigma^2\{ y\}/4 = \sigma^2\{ y\}$, т.е. в 2 раза меньше!!!

Задачи, решаемые при ПЭ:

- поиск оптимальных решений,
- построение интерполяционных функций,
- выбор существенных факторов,
- выбор наиболее приемлемых гипотез.

Основные требования к эксперименту:

- *воспроизводимость результатов;*
- *управляемость* эксперимента.

ОСНОВЫ ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Применение ТПЭ основано на ряде допущений:

1. Результаты, полученные путем усреднения повторных опытов в каждой точке плана, представляют собой независимые, нормально распределенные случайные величины.

2. Ошибка определения значения функции отклика обусловлена в основном влиянием на результат неучтенных или случайных факторов, а не погрешности измерений.

3. Дисперсии среднего значения функции отклика в различных точках равны друг другу (выборочные оценки дисперсии однородны).

Экспериментатор, применяющий методы планирования эксперимента, должен уметь формулировать свою задачу в терминах «**черного ящика**»:

1. Входы «черного ящика» называются **факторами**. Каждый фактор может принимать некоторое определенное число различных значений, называемых **уровнями**.

2. Сочетание определенных уровней всех факторов определяет возможное состояние «черного ящика» и условия одного из возможных опытов.

3. Причиной нарушения воспроизводимости эксперимента являются неуправляемые факторы.

4. Если требование воспроизводимости результатов активного эксперимента не выполняется, приходится обращаться к активно–пассивному эксперименту.

Этапы планирования эксперимента:

- постановка задачи (определение цели эксперимента, выяснение исходной ситуации, оценка допустимых затрат времени и средств, установление типа задачи);
- сбор априорной информации (изучение литературы, опрос специалистов и т.п.);
- выбор способа решения и стратегии его реализации (установление типа модели, выявление возможных влияющих факторов, выявление выходных параметров, выбор целевых функций, создание необходимых нестандартных технических средств, формулировка статистических задач, выбор или разработка алгоритмов программ обработки экспериментальных данных).

Основными **концепциями современного подхода к организации эксперимента** являются *рандомизация*, *многофакторность* и *автоматизация*.

Концепция рандомизации. План эксперимента составляется таким образом, чтобы **сделать случайными** в пространстве и во времени систематически действующие **мешающие факторы**; тогда эти факторы рассматриваются как случайные величины.

Принцип многофакторности. При изучении объектов с несколькими факторами опыты ставят так, чтобы варьировать все управляемые факторы в отличие от традиционного подхода, при котором влияние каждого фактора изучается отдельно.

Все ранее указанные задачи ТПЭ так или иначе связаны с определением **функции отклика** – уравнения, связывающего искомую величину и управляемые факторы,

Функция отклика задаётся экспериментатором в форме математической модели.

Математическая модель – это уравнение, связывающее параметры (факторы) задачи и искомую функцию.

Уравнение (отклик) в общем случае имеет вид $Y = \phi\,(X_1, X_2, \ldots, X_k)$,

где ϕ – *функция отклика*, X_k – параметры *(фактор)* задачи; Y – *параметр оптимизации*, *выходной* параметр: k - *число факторов*.

ТРЕБОВАНИЯ к искомой функции *Y* *(отклик на действие факторов):*

– определена количественно; количеством может быть *ранг, квалитет, класс*;

– обладает однозначностью в статистическом смысле, т.е. заданному набору значений факторов должно соответствовать определённое значение параметра оптимизации (с точностью до ошибки эксперимента);

– обладает полнотой, т.е. достаточно полно характеризует облик технологического параметра (большой универсальностью обладают обобщённые параметры оптимизации, которые строятся как функции нескольких частных параметров оптимизации).

Нужно стремиться к уменьшению числа параметров.

Между парой параметров X_1 и X_2 вычислить коэффициент корреляции по результатам N опытов:

$$r_{12} = \frac{\Sigma(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\Sigma(X_1 - \bar{X}_1)^2\ \Sigma(X_2 - \bar{X}_2)^2}}$$

1. Если экспериментально найденное значение $r_{12} < r_{12}^{кр}$, то статистическая связь Y_1 и Y_2 маловероятна

Критическое значение коэффициента парной корреляции при α = 0,05

Число степеней свободы $f_1 = N - 2$	$r_{12-кр}$	Число степеней свободы $f_1 = N - 2$	$r_{12-кр}$
1	0,997	10	0,576
2	0,950	15	0,482
3	0,878	20	0,423
5	0,754	50	0,217

2. Если $r_{12} \approx 1$ (≈детерминированная связь), то один из параметров можно изъять из рассмотрения.

Требования к факторы:

- *управляемые*;
- *операционные* (указана последовательность операций, обеспечивающих получение данного уровня фактора).

Требования к совокупности факторов:

- *совместимость;*
- *независимость.*

ВЫБОР МОДЕЛИ

Если параметр оптимизации можно описать полиномом, то для двух параметров возможны следующие модели:

- полином 0-го порядка $Y = b_0$
- полином 1-го порядка $Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2$; ($Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{12}X_1X_2$ – теор. модель)
- полином 2-го порядка $Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 + b_{11}X_1^2 + b_{22}X_2^2$

Минимизировать число опытов – **предпочтительнее выбор полинома наименьшей степени** , а в дальнейшем для хорошего предсказывания направления улучшения параметра оптимизации использовать полиномы более высокого порядка, позволяющие более точно указать *направление градиента*.

ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

1. Определить область значений факторов (область планирования эксперимента).
2. Выбирать основной (нулевой) уровень каждого фактора.
3. Назначить интервал варьирования каждого фактора (число, прибавление которого к основному уровню даёт верхний, а вычитание – нижний уровни значения фактора).

Пример

Натуральные значения x……….. 1 2 3 4 5

Кодированные значения X …….. –1 –0,5 0 +0,5 +1. (Здесь основной уровень $x_0 = 3$, интервал варьирования $I = 2$)

4. Нормировать уровни факторов безразмерными значениями, принимающими на границах интервала варьирования значения ± 1.

Для **линейной математической модели** достаточно варьировать **факторы на двух уровнях.**

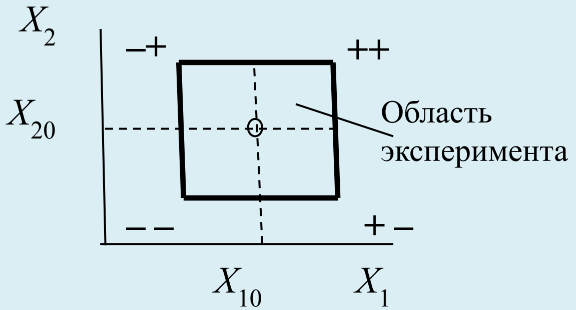
Если число факторов равно k, то требуемое число опытов $N = 2^k$.

В ***полном факторном эксперименте*** реализуются **все возможные сочетания уровней факторов**.

Минимальное число уровней факторов должно быть на единицу больше порядка уравнения.

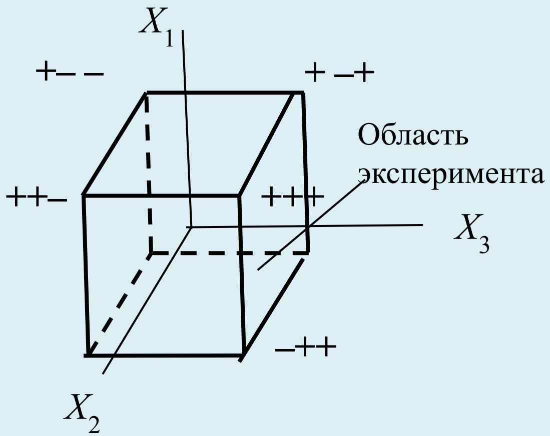
Матрица планирования эксперимента 2^k при $k = 2$

Номер опыта	X_1	X_2	Y
1	-1	-1	Y_1
2	+1	-1	Y_2
3	-1	+1	Y_3
4	+1	+1	Y_4



Матрица планирования эксперимента 2^k при $k = 3$

Номер опыта	X_1	X_2	X_3	Y
1	-1	-1	+1	Y_1
2	-1	+1	-1	Y_2
3	+1	-1	-1	Y_3
4	+1	+1	+1	Y_4
5	-1	-1	-1	Y_5
6	-1	+1	+1	Y_6
7	+1	-1	+1	Y_7
8	+1	+1	-1	Y_8



Свойства полного факторного эксперимента

1. $\sum\limits_{u=1}^{N} X_{iu} = 0$ – условие ортогональности к столбцу из единиц;

2. $\sum\limits_{u=1}^{N} X_{iu} X_{ju} = 0$ – условие парной ортогональности столбцов;

3. $\sum\limits_{u=1}^{N} X_{iu}^2 = N$ – условие нормировки.

ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ РЕГРЕССИИ $Y = b_0 + b_1X_1 + \ldots + b_kX_k$

Y_i, вычисленное в i – строке плана, умножим на столбец X_{ji};(+1 и -1) сложим эти произведения по всем строкам плана и получим: $\Sigma X_{ji} Y_i = \Sigma b_0 X_{ji} + b_1 \Sigma X_{1i} X_{ji} + \ldots + b_k \Sigma X_{k\,i} X_{ji}$.

Согласно свойствам ПФЭ $\Sigma X_{j\mathrm{i}} X_{ji} = 0$, $\Sigma X_{j\mathrm{i}} = 0$ и $\Sigma X_{j\mathrm{i}}^2 = N$, имеем $b_i = (1/N) \Sigma X_{ji} Y_i$.

Для нахождения b_0 достаточно сложить по всем строкам плана значения Y_i:

$\Sigma Y_i = b_0 N + b_1 \Sigma X_{1i} + \ldots + b_k \Sigma X_{k\,\mathrm{i}}$. Учитывая $\Sigma X_{j\mathrm{i}} = 0$, получим $b_0 = (1/N) \Sigma Y_i$.

Зависимости $b_i = (1/N) \Sigma X_{ji} Y_i$ для вычисления коэффициентов согласуются с методом наименьших квадратов и соответствует уравнению регрессии в регрессионном анализе.

Для линейного отклика неувязка в узловых точках $\varepsilon = (a_0 + a_1 X_{1\,i} + a_2 X_{2\,i} - Y_i)$.
Согласно методу наименьших квадратов $S(a_0, a_1, a_2) = \Sigma (a_0 + a_1 X_{1\,i} + a_2 X_{2\,i} - Y_i)^2 \rightarrow \min$.

Задача сводится к решению системы уравнений:

$$
\begin{aligned}
&a_0 N + a_1 \Sigma X_{1i} + a_2 \Sigma X_{2i} = \Sigma Y_i \\
&a_0 \Sigma X_{1i} + a_1 \Sigma X_{1i}^2 + a_2 \Sigma X_{2i} X_{1i} = \Sigma Y_i X_{1i} \\
&a_0 \Sigma X_{2i} + a_1 \Sigma X_{1i} X_{2i} + a_2 \Sigma X_{2i}^2 = \Sigma Y_i X_{2i}
\end{aligned}
$$

Учитывая, что $\Sigma X_{ji} = 0$ и $\Sigma X_{ji}^2 = N$, получим: $a_0 = (1/N) \Sigma Y_i$; $a_1 = (1/N) \Sigma Y_i X_{1i}$; $a_2 = (1/N) \Sigma Y_i X_{2i}$.

Из корреляционного анализа следует, что уравнение линейной регрессии $Y = a_0 + a_1 X \rightarrow Y - \bar{Y} = r_{xy} (X - \bar{X}) \sigma_y / \sigma_x$.

Следовательно, $a_1 = r_{xy} \sigma_y / \sigma_x = (\overline{XY} - \bar{X}\,\bar{Y}) / (\overline{X^2} - \bar{X}^2)$.

В силу свойств векторов-столбцов таблицы плана $(\bar{X} = 0,\ \Sigma X_i = 0,\ \overline{X^2} = N)$ $a_1 = (1/N) \Sigma Y_i X_{1i}$.

Далее $a_0 = \bar{Y} - (r_{xy} \sigma_y / \sigma_x) \bar{X}$, т.е. $a_0 = \bar{Y} = (1/N) \Sigma Y_i$.

Число экспериментов должно быть не менее числа искомых коэффициентов

Пример.

Нужно оценить коэффициенты регрессии линейного уравнения $Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3$

План эксперимента для линейной модели с тремя независимыми переменными

Номер опыта	Матрица коэффициентов				Результат эксперимента
		План эксперимента			
	X_0	X_1	X_2	X_3	
1	+1	+1	-1	-1	Y_1
2	+1	-1	+1	-1	Y_2
3	+1	-1	-1	+1	Y_3
4	+1	+1	+1	+1	Y_4

Это **насыщенный план**: число опытов = числу оцениваемых параметров (оцениваемые a_0, a_1, a_2, a_3).

Свойства насыщенного плана:

1. $\sum_{u=1}^{N} X_{iu} = 0$ – условие ортогональности к столбцу из единиц;
2. $\sum_{u=1}^{N} X_{iu} X_{ju} = 0$ – условие попарной ортогональности столбцов;
3. $\sum_{u=1}^{N} X_{iu}^2 = N$ – условие нормировки.

Все коэффициенты регрессии независимы друг от друга и определяются формулами: $a_i = (1/N) \sum_{u=1}^{N} X_{iu} Y_u$ $(i = 0, 1,2,\ldots n)$

с дисперсией $\sigma^2\{a_i\} = \sigma^2\{y\}/N$.

В случае эксперимента с 3 независимыми переменными каждый коэффициент определяется по двум точкам и в этом случае дисперсия $\sigma^2\{y\}/2$.

В случае 15 коэффициентов регрессии выигрыш в дисперсии 8-кратный.

Число опытов N для оценки коэффициентов линейной модели $Y = a_0 + \sum_{i=1}^{k} a_i X_i$ равно числу коэффициентов,

!!! т.е. $N = k + 1$.

План эксперимента для линейной модели с тремя независимыми переменными

Номер опыта	Матрица коэффициентов				Результат эксперимента
	X_0	План эксперимента			
		X_1	X_2	X_3	
1	+1	+1	-1	-1	80
2	+1	-1	+1	-1	65
3	+1	-1	-1	+1	45
4	+1	+1	+1	+1	90

$a_0 = (80+65+45+90)/4 = 70{,}0$; $a_1 = (+80 - 65 - 45 + 90)/4 = 15$;
$a_2 = (-80 + 65 - 45 + 90)/4 = 7{,}5$; $a_3 = (- 80 - 65 + 45 + 90)/4 = -2{,}5$.

Модельное уравнение $Y = 70 + 15X_1 + 7{,}5\,X_2 - 2{,}5X_3$.

Для получения уравнения в значениях действительных параметров в модельное уравнение вместо кодированного параметра X_1 вставить $(x_1 - x_{01})/I_1$, где x_0 – основной уровень x_1, а I – интервал варьирования x_1. Остальные также.

Уравнение $Y = 70 + 15(x_1 - x_{01})/I_1 + 7{,}5\,(x_2 - x_{02})/I_2 - 2{,}5(x_3 - x_{03})/I_3$.

Насыщенный план не позволяет оценить дисперсию и значимость коэффициентов полученного уравнения.

ПРЕДСТАВЛЕНИЕ О ХОРОШЕМ ЭКСПЕРИМЕНТЕ

1. **Насыщенный план** обладает свойством ***ротатабельности***: дисперсия оценки модели не зависит от угла, проведённого из центра эксперимента. Величина $1/\sigma^2\{y\}$ одинакова для всех эквидистантных точек.

2. **Насыщенный план** обладает свойством ***ортогональности***: план обеспечивает независимость оценки коэффициентов регрессии, а также доверительных границ для оценок коэффициентов регрессии.

Неортогональность плана приводит к взаимной зависимости коэффициентов регрессии, соответственно, один и тот же процесс будет представлен различными параметрами в модели.

3. При проведении эксперимента опыты, заданные матрицей планирования, желательно ***рандомизировать***, т.е. обеспечить случайный порядок их проведения, например с помощью таблиц случайных чисел.

Рандомизация нивелирует систематические воздействия неконтролируемых факторов.

4. Задание вида модели должно предшествовать эксперименту.

Обычно стремятся на первом этапе использовать линейную модель. Имеется несколько способов проверки пригодности линейной математической модели, т.е. её адекватности.

Часто имеют место взаимодействие между факторами/.

Полный факторный эксперимент позволяет количественно оценить эффект взаимодействия факторов

План эксперимента для линейной модели $Y = a_0 + a_1X_1 + a_2X_2 + a_{12}X_1X_2$

Номер опыта	Матрица коэффициентов				Результат эксперимента
	X_0	X_1	X_2	X_1X_2	
1	+1	+1	-1	-1	Y_1
2	+1	-1	+1	-1	Y_2
3	+1	-1	-1	+1	Y_3
4	+1	+1	+1	+1	Y_4

Для вычисления коэффициента a_{12} нужно умножить Y_i на $X_{1i}X_{2i}$ и сложить по всем строкам матрицы опытов:

$$\Sigma X_{1i}X_{2i}Y_i = a_0\Sigma X_{0i}X_{1i}X_{2i} + a_1\Sigma X_{1i}^2X_{2i} + a_2\Sigma X_{1i}X_{2i}^2 + a_{12}N;$$

$$a_{12} = (1/N)\Sigma X_{1i}X_{2i}Y_i$$

Пример

Номер опыта	X_0	X_1	X_2	$X_1 X_2$	Y
1	+1	-1	-1	+1	95
2	+1	+1	-1	-1	90
3	+1	-1	+1	-1	85
4	+1	+1	+1	+1	82

a_0 = (95+90+85+82)/4 = 88,0; a_1 = (-95 +90 -85 +82)/4 =- 2,0;

a_2 = (-95 -90 +85 +82)/4 = -4,5; a_{12} = (95 – 90 – 85 + 82)/4 = 0,5.

a_{12} близок нулю – корреляции нет ? Взаимодействие мало ?

Модельное уравнение $Y = 88 + 2X_1 - 4{,}5\,X_2 + 0{,}5X_1\,X_2$.

Для получения уравнения в значениях действительных параметров в модельное уравнение вместо кодированного параметра X_1 вставить $(x_1 - x_0)/I$, где x_0 – основной уровень x_1, а I – интервал варьирования x_1. Остальные также.

План полного факторного эксперимента типа 2^3 с учётом возможных взаимодействий

Номер опыта	X_0	X_1	X_2	X_3	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	$X_1X_2 X_3$	Y
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

Коэффициенты уравнения регрессии независимы друг от друга в случае ортогональности столбцов.

Это справедливо только для модели с линейными эффектами.

Технологические примеры применения теории планирования эксперимента

Пример 1. Физический эксперимент «Технология продукта»

На выход продукта *A* влияют температура *t*, давление *p*, продолжительность процесса τ.
Определить выход *Y* продукта *A* как функцию параметров *t*, *p* и τ.
Диапазон изменения параметров:
t – 250 …350°С; *p* – 3,0 … 5,0 атм; τ – 90 … 150 сек. Число параллельных опытов γ =2.

РЕШЕНИЕ

1. Выбираем основной (нулевой) уровень параметров: t_0 = 300°С; *p* = 4 атм; τ= 120 с.

2. Задаём интервал варьирования: Δ *t* = 50°С; Δ *p* = 1 атм; Δτ= 30 с.

Верхний уровень параметров: $t_0 + \Delta t$; $p + \Delta p$; $\tau + \Delta\tau$.
Нижний уровень параметров: $t_0 - \Delta t$; $p - \Delta p$; $\tau - \Delta\tau$.

3. Составим таблицу входных параметров

Факторы X_i	Значения факторов		
	- 1	0	+1
X_1 – температура	250	300	350
X_2 – время	90	120	150
X_3 – давление	3	4	5

4. Запишем матрицу полного факторного эксперимента 2^3

№ опыта	Температура		Продолжительность		Давление	
	X_1		X_2		X_3	
1	250	-	90	-	3	-
2	350	+	90	-	3	-
3	250	-	150	+	3	-
4	350	+	150	+	3	-
5	250	-	90	-	5	+
6	350	+	90	-	5	+
7	250	-	150	+	5	+
8	350	+	150	+	5	+

Значения некоторых коэффициентов могут быть признаны незначимыми и отброшены. В новой модели будет меньшее число членов $k_{м}$ по сравнению с исходной.

Новая модель проверяется на адекватность сравнением расчётных значений $Y_{u\text{- расч.}}$ по новой модели и экспериментального значения $Y_{u\text{- эксп.}}$ Адекватности проверяется по F – критерию:

$$\boldsymbol{F_{расч.} = S^2_{неад} \,/\, S_y^{\,2} \le F_{\alpha;\, f1;\, f2}},$$

где $F_{\alpha;\, f1;\, f2}$ – значение критерия распределения Фишера;

f_2 – число степеней свободы новой модели, $f_2 = N - k_{м}$;

$$S^2_{неад} = [\Sigma\, (Y_{u\text{- расч.}} - Y_{u\text{- эксп.}})^2] \,/\, (N - k_{\,м}),$$

где u – номер строки таблицы плана;

$Y_{u\text{- расч}}$ – расчётное (по модели) значение Y по значениям переменных u –ой строки;

$Y_{u\text{- эксп}}$ – экспериментальное значение Y u –ой строки.

***Пример*. Эксперимент по плану ПФЭ 2^3 , каждый опыт повторялся по три раза.**

Факторы, влияющие на прочность склейки *y (кг/2,5 см²)*:

x_1 – количество наносимого клея (г/см²), ; $x_{1\,min} = 0{,}02$; $x_{1\,max} = 0{,}06$;
x_2 – время активации клеевой пленки (сек), ; $x_{2\,min} = 60$; $x_{2\,max} = 300$;
x_3 – давление прессования при склеивании (кгс/см²) ; $x_{3\,min} = 2$; $x_{3\,max} = 8$.

Требуется построить уравнение регрессии (проверить полученную модель на адекватность и произвести ее интерпретацию

№ опыта	Факторы			Результаты опытов		
	X1	*X2*	*X3*	*Y1*	*Y2*	*Y3*
1	+	+	+	7,4	8,4	6,4
2	–	+	+	8,6	7	7,8
3	+	–	+	12,3	9	9,3
4	–	–	+	5,8	5,8	5,7
5	+	+	–	18,8	17	15,2
6	–	+	–	8,4	8,4	6
7	+	–	–	11,8	7	9,4
8	–	–	–	10,5	7,8	8,1

$$Y = \mathbf{b_0} + \mathbf{b_1}X_1 + \mathbf{b_2}X_2 + \mathbf{b_3}X_3 + \mathbf{b_{12}}X_1 X_2 + \mathbf{b_{13}}X_1X_3 + \mathbf{b_{23}}X_2X_3 + \mathbf{b_{123}}X_1 X_2X_3$$

1. Кодирование факторов

Факторы	Верхний уровень	Нижний уровень	Центр x_{0i}	Интервал варьирования I_i	Зависимость кодированной переменной от натуральной
x_1	*0,06*	*0,02*	*0,04*	*0,02*	$X_1 = (x_1 - 0{,}04)/0{,}02$
x_2	300	60	180	120	$X_2 = (x_2 - 180)/120$
x_3	8	2	5	3	$X_3 = (x_3 - 5)/3$

2. Определение средних выборочных значений $\bar{y}_i$

$\overline{y1} = 1/3\ (7{,}4 + 8{,}4 + 6{,}4) = 7{,}4$

$\overline{y2} = 7{,}8$

$\overline{y3} = 10{,}2$

$\overline{y4} = 5{,}77$

$\overline{y5} = 17$

$\overline{y6} = 7{,}6$

$\overline{y7} = 9{,}4$

$\overline{y8} = 8{,}8$

3. Матрица планирования для обработки

№	Факторы			Взаимодействия				Результаты опытов			Значения $\overline{y}_i$
	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$	y_1	y_2	y_3	
1	+	+	+	+	+	+	+	7,4	8,4	6,4	7,4
2	–	+	+	-	-	+	-	8,6	7	7,8	7,8
3	+	-	+	-	+	-	-	12,3	9	9,3	10,2
4	–	-	+	+	-	-	+	5,8	5,8	5,7	5,77
5	+	+	-	+	-	-	-	18,8	17	15,2	17
6	–	+	-	-	+	-	+	8,4	8,4	6	7,6
7	+	-	-	-	-	+	+	11,8	7	9,4	9,4
8	–	-	-	+	+	+	-	10,5	7,8	8,1	8,8

4. Определение коэффициентов уравнения регрессии

b_0 = 1/8 $\Sigma \overline{y}_i$ = 1/8 (7,4 + 7,8 + 10,2 + 5,77 + 17 + 7,6 + 9,4 + 8,8) = 9,25;

b_1 = 1/8 $\Sigma(X_1 \cdot \overline{y}_i)$= 1/8 (7,4 – 7,8 + 10,2 – 5,77 + 17 – 7,6 + 9,4 –-8,8) = 1,75;

b_2 = 1/8 $\Sigma(X_2 \cdot \overline{y}_i)$= 1/8 (7,4 +7,8 – 10,2 – 5,77 + 17 +7,6 – 9,4 – 8,8) = 0.70 и т.д.;

b_0 = 9,25; b_1 = 1,75; b_2 = 0,70; b_3 = - 1,45; $b_{1\text{-}2}$ = 0,5; $b_{1\text{-}3}$ = - 0,75; $b_{2\text{-}3}$ = - 0,90; $b_{1\text{-}2\text{-}3}$ = - 1,70.

$$Y = \mathbf{b_0} + \mathbf{b_1}X_1 + \mathbf{b_2}X_2 + \mathbf{b_3}X_3 + \mathbf{b_{12}}X_1X_2 + \mathbf{b_{13}}X_1X_3 + \mathbf{b_{23}}X_2X_3 + \mathbf{b_{123}}X_1X_2X_3$$

$$Y = \mathbf{b_0} + \mathbf{b_1}(x_1 - 0{,}04)/0{,}02 + \mathbf{b_2}(x_2 - 180)/120 + \mathbf{b_3}(x_3 - 5)/3 + \mathbf{b_{12}}[(x_1 - 0{,}04)/0{,}02][(x_2 - 180)/120] + \mathbf{b_{13}}X_1X_3 + \mathbf{b_{23}}X_2X_3 + \mathbf{b_{123}}X_1X_2X_3$$

Коэффициенты уравнения регрессии

b_0	b_1	b_2	b_3	$b_{1,2}$	$b_{1,3}$	$b_{2,3}$	$b_{1,2,3}$	
9,25	1,75	0,7	-1,45		0,5	- 0,75	$\boxed{-0{,}9}$	- 1,7

5. Определение выборочных дисперсий и дисперсию

j	y_1	y_2	y_3	y_i	$(y_1 - \overline{y_i})^2$	$(y_2 - \overline{y_i})^2$	$(y_3 - \overline{y_i})^2$	S_i^2
1	7,4	8,4	6,4	7,4	0	1	1	1
2	8,6	7	7,8	7,8	0,64	0,64	0	0,64
3	12,3	9	9,3	10,2	1,21	1,44	0,81	1,73
4	5,8	5,8	5,7	5,77	0,0009	0,0009	0,0049	0,003
5	18,8	17	15,2	17	3,24	0	3,24	3,24
6	8,4	8,4	6	7,6	0,64	0,64	2,56	1,92
7	11,8	7	9,4	9,4	5,76	5,76	0	5,76
8	10,5	7,8	8,1	8,8	2,89	1	0,49	2,19

Дисперсия воспризводимости

$S^2_{\{y\}} = (1/n)\, \Sigma\, S_i^2 = 1/8\,(1 + 0{,}64 + 1{,}73 + 0{,}003 + 3{,}24 + 1{,}92 + 5{,}76 + 2{,}19) = 2{,}06$

Среднее квадратичное отклонение коэффициентов:

$S_{\text{коэф}} = \sqrt{S^2_{\{y\}}/(\gamma N)} = \sqrt{2{,}06/(3 \cdot 8)} = 0{,}293$; *γ – число опытов в каждом эксперименте,*
N – число экспериментов; дисперсия ≡ $1/\gamma^{1/2}$

По критерию Стьюдента при числе степеней свободы $(\gamma - 1)\, N = 2 \cdot 8 = 16$ и $\alpha = 0{,}05$
$t_{кр} = 2{,}12$ и $t_{кр} S_{\text{коэф}} = 2{,}12 \cdot 0{,}293 = 0{,}52$. Коэффициенты $b_* < t_{кр} S_{\text{коэф}} = 0{,}52$ незначимы

6. Проверка уравнения на адекватность

$$Y = \boldsymbol{b_0} + \boldsymbol{b_1}X_1 + \boldsymbol{b_1}X_2 + \boldsymbol{b_3}X_3 + \boldsymbol{b_{13}}X_1X_3 + \boldsymbol{b_{23}}X_2X_3 + \boldsymbol{b_{123}}X_1X_2X_3$$

По критерию Фишера при $F_{расч} < F_{табл}$ уравнение считается адекватным.

$F_{расч} = S^2_{ост} / S^2_{\{y\}}$, где остаточная дисперсия $S^2_{ост} = [\gamma/(N-r)] \sum (\bar{y}_i - y_i)^2$, где $\bar{y}_i$ – *значени* y_i, вычисленное по уравнению регресси

r – число значимых коэффициентов уравнения регрессии

y = 9,25 + 1,75X₁ + 0,7X₂ – 1,45X₃ – 0,75X₁X₃ – 0,9X₂X₃ – 1,7X₁X₂X₃, здесь r = 7

$\bar{y}_1 = 9{,}25 + 1{,}75 + 0{,}7 - 1{,}45 - 0{,}75 - 0{,}9 - 1{,}7 = 6.9$, *здесь все X с (+)*

$\bar{y}_2 = 9{,}25 - 1{,}75 + 0{,}7 - 1{,}45 + 0{,}75 - 0{,}9 + 1{,}7 = 8.3$, *здесь все X* (–1, +1,+1, –1, –1, +1, –1)

$\bar{y}_3 = 10{,}7$; $\bar{y}_4 = 5{,}3$; $\bar{y}_5 = 16{,}3$; $\bar{y}_6 = 8{,}1$; $\bar{y}_7 = 9{,}85$; $\bar{y}_8 = 8{,}3$

$S^2_{ост} = [\gamma/(N-r)] \sum (\bar{y}_i - y_i)^2 = [3/(8-7)]\,[(6{,}9 - 7{,}4)^2 + (8{,}3 - 7{,}8)^2 + \ldots] = 5{,}77$

$F_{расч} = 5{,}77/2{,}06 = 2{,}8$

$F_{табл} = 4{,}49$ при $\alpha = 0{,}05$; $k_1 = (n - r) = 8 - 7 = 1$; $k_2 = (\gamma - 1)\, n = 16$

Так как $F_{расч} < F_{табл}$, регрессионное уравнение считается адекватным

ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ (ДФЭ)

Для полного двухуровневого факторного эксперимента (ПФЭ) согласно модели первой степени объекта $N = 2^k$:

$k = 2 \ldots N = 2^2 = 4$; число всех коэффициентов $\sum \frac{k!}{n!\,(k-n)!} = 1 + 2 + 1 = N$;

$k = 3 \ldots N = 2^3 = 8$; число всех коэффициентов $1 + 3 + 3 + 1 = 8$;

$k = 4 \ldots N = 2^4 = 16$; $k = 5 \ldots N = 2^5 = 32$; $k = 6 \ldots N = 2^6 = 64$; …. –

При $k = 2$ число коэффициентов взаимодействия факторов равно 1;

$k = 3$ … КВФ = 4; $k = 4$ … КВФ = 9; $k = 5$ … КВФ = 26; $k = 6$ … КВФ = 57; …

Если есть основания считать некоторые взаимодействия незначимыми, то можно существенно сократить число экспериментов.

Например, при $k = 3$ вместо ПФЭ 2^3 ($N = 2^3 = 8$) можно взять ДФЭ 2^{3-1} ($N = 4$)

Схема определение коэффициентов остаётся прежней. НО … при этом *утрачивается независимость* *экспериментальных значений коэффициентов модели*

Матрица ПФЭ

№	X_0	X_1	X_2	X_1X_2	Y
1	+	–	–	+	Y_1
2	+	+	–	–	Y_2
3	+	–	+	–	Y_3
4	+	+	+	+	Y_4

Здесь все коэффициенты определены однозначно:

1) все столбцы взаимно ортогональны;

2) Знак (X_1X_2) = (знак X_1) $\cdot$ (знак X_1);

3) Число КВФ = N = 4

Матрица ПФЭ 2^3

№	X_0	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$	Y
1	+	–	–	+	+	–	–	+	Y_1
2	+	+	–	–	–	–	+	+	Y_2
3	+	–	+	–	–	+	–	+	Y_3
4	+	+	+	+	+	+	+	+	Y_4
5	+	–	–	–	+	+	+	–	Y_5
6	+	+	–	+	–	+	–	–	Y_6
7	+	–	+	+	–	–	+	–	Y_7
8	+	+	+	–	+	–	–	–	Y_8

Матрица ПФЭ 2^3 →ДФЭ 2^{3-1}

При $X_1X_2X_3 = +1$ $X_3 = X_1X_2$

При $X_1X_2X_3 = -1$ $X_3 = -X_1X_2$

Произведения столбцов матриц, равные +1 или – 1, называются ***определяющими контрастами***

По результатам опытов вычисляются все коэффициенты кроме коэффициента при ОК

Матрица ДФЭ 2^{3-1} при ОК = +1

№	X_0	X_1	X_2	X_3	Y
1	+	–	–	+	Y_1
2	+	+	–	–	Y_2
3	+	–	+	–	Y_3
4	+	+	+	+	Y_4

Матрица ДФЭ 2^{3-1} при ОК = – 1

№	X_0	X_1	X_2	X_3	Y
1	+	–	–	–	Y_1
2	+	+	–	+	Y_2
3	+	–	+	+	Y_3
4	+	+	+	–	Y_4

Оценки коэффициентов будут смешанными:

$B_1 \to \beta_1 + \beta_{23}$

$B_2 \to \beta_2 + \beta_{13}$

$B_3 \to \beta_3 + \beta_{13}$

Здесь вместо X_1X_2 матрицы **ПФЭ 2^2** введён третий X_3 фактор

Матрицы **ДФЭ 2^{3-1}** называются полурепликами матрицы **ПФЭ 2^3**

ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ (ДФЭ)

В случаях использования только <u>линейные приближения функции отклика</u>, количество опытов МОЖНО сократить, используя для планирования так называемые *регулярные дробные реплики ПФЭ, содержащие подходящее число опытов и сохраняющие основные свойства матрицы планирования.*

Построение регулярной дробной реплики или проведение *дробного факторного эксперимента (ДФЭ) типа 2^{k-p} предусматривает отбор из множества k факторов*

k – p основных факторов, для которых строится план ПФЭ.

Этот план дополняется p столбцами, которые соответствуют остальным

Матрица ДФЭ типа 2^{k-p} содержит k + 1 столбец и N = 2^{k-p}

Кол-во факторов	Дробная реплика	ДФЭ	Количество опытов	
			Для ДФЭ	Для ПФЭ
3	**1/2 от 2^3**	**2^{3-1}**	4	8
4	**1/2 от 2^4**	**2^{4-1}**	8	16
5	**1/4 от 2^5**	**2^{5-2}**	8	32
6	**1/8 от 2^6**	**2^{6-3}**	8	64
7	**1/16 от 2^7**			
5	**1/2 от 2^5**			
6	**1/4 от 2^6**			
7	**1/8 от 2^7**			
8	**1/16 от 2^8**			

Планы типа 2^{k-p} *являются ортогональными для моделей с взаимодействиями.* Поэтому для вычисления оценок коэффициентов получаются простые

При выборе **ДФЭ 2^{4-1}** возможны 8 вариантов**:**

$X_4 = X_1X_2$; $X_4 = -X_1X_2$; $X_4 = X_2X_3$; $X_4 = -X_2X_3$;

$X_4 = X_1X_3$; $X_4 = -X_1X_3$;

$X_4 = X_1X_2X_3$; $X_4 = -X_1X_2X_3$

Последние два с максимальной разрешающей способностью (максимальное число факторов в определяющем контрасте)

Виды планов в зависимости от оптимизации дисперсии

1. ***D-оптимальный*** план: минимизируется объём эллипсоида рассеяния ошибок параметров уравнения регрессии; детерминант матрицы ($\mathbf{X}^{\mathsf{T}}\mathbf{X}$) будет максимальным, а детерминант ковариационной матрицы $(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}$ – минимальным.
2. ***E- оптимальный*** план: минимизируется максимальную ось эллипсоида рассеивания; ковариационная матрица с минимальным значением максимального характеристического числа.
3. ***A- оптимальный*** план: минимизируется среднюю дисперсию оценок коэффициентов регрессии, эллипсоид рассеяния с наименьшей суммой квадратов длин осей; ковариационные матрицы имеют наименьшие значения следа.

Нет плана, который отвечал бы хотя бы нескольким важнейшим критериям. Нужно компромиссное решение.

Оптимальные насыщенные планы первого порядка при условии минимизации наибольшей дисперсии оценок коэффициентов модели:

1) насыщенные планы **дробного факторного эксперимента.** так называемые планы Плакетта-Бермана;
2) симплекс планы первого порядка. (симплекс – выпуклая оболочка из n точек m – мерного пространства при $n \le m+1$)

1.Число опытов $\boldsymbol{N}$ в планах первой группы кратно 4.

2.Каждый фактор варьируется на двух уровнях +1 (+) и – 1 (–).

3.Насыщенные дробные реплики при числе факторов
$\boldsymbol{k}=3$ (2^{3-1}, $N = 4$); $\boldsymbol{k}=7$ (2^{7-4}, $N=8$); $\boldsymbol{k}=15$ (2^{15-11}, $N = 16$) и т.д.

k	N	Комбинация знаков
3	4	+ – +
7	8	+ + + – + – –
11	12	+ + – + + + – – – + –
15	16	+ + + + – + – + + – – + – – –
19	20	+ + – – + + + + – + – + – – – – + + –
23	24	+ + + + + – + – + + – – + + – – + – + – – – –
27	28	3 плана
31	32	– – – – + – + – + + + – + + – – – + + + + + – – + + – + – – +
35	36	– + – + + + – – – + + + + + – + + + – – + – – – – + – + – + + – – + –

Правило составления плана

1) вторую и последующие строки плана получают сдвигом всех элементов предыдущей строки на одну позицию вправо (или влево);

2) последний элемент предыдущей строки переносится на первое место в новой строке (или наоборот, если сдвиг элементов выполнялся влево).

3) к полученной таблице размером $k \times k$ добавляют строку минусов;

4) при выполнении обработки результатов эксперимента вводят (слева от X_1) ещё один столбец фиктивной переменной X_0, состоящий только из плюсов; получается квадратная таблица размерности $(k + 1)$.

План Плакетта – Бермана для *k* = 11

Номер образца	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+	+	+	-	+	+	+	-	-	-	+	-
2	+	-	+	+	-	+	+	+	-	-	-	+
3	+	+	-	+	+	-	+	+	+	-	-	-
4	+	-	+	-	+	+	-	+	+	+	-	-
5	+											
6	+											
7	+											
8	+											
9	+											
10	+											
11	+											
12	+	–	–	–	–	–	–	–	–	–	–	–

ОЦЕНКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ И ДОСТОВЕРНОСТИ ВЫБРАННОГО ОТКЛИКА ДФЭ при $\gamma > 1$

1. Дисперсия ошибок воспроизводимости опытных данных

Под ***ошибкой*** воспроизводимости понимается ошибка опыта, т.е. насколько точно воспроизводится результат опыта при повторном его проведении в идентичных условиях.

Вычисление дисперсии ошибок производится по формуле

$$S^2 = S_E^2 / (f_2 \cdot \gamma),$$

где S^2 – дисперсия воспроизводимости выходного параметра;

S_E^2 – суммарная построчная дисперсия ($S_E^2 = \Sigma S_i^2$);

S_i^2 – дисперсия i- той строки плана;

γ – число параллельных опытов;

f_2 – число степеней свободы, $f_2 = N(\gamma - 1)$.

Проверка воспроизводимости опытов производится по соответствию критерию Кохрена:

$$G = S_{n\,\max}^2 / S_E^2 \le G_\alpha \quad \text{при } (\gamma - 1) \text{ и } N,$$

где $S_{n\,\max}^2$ – максимальная величина построчечной дисперсии;

G_α – критическое значение критерия Кохрена.

Значения G_α в зависимости от числа степеней свободы ($\gamma - 1$) и числа опытов N при $\alpha = 0{,}05$

N \ $\gamma - 1$	1	2	3	4	5	10	20
2	0,99	0,97	0,94	0,90	0,87	0,78	0,50
3	0,96	0,87	0,79	0,75	0,70	0,60	0,33
5	0,85	0,68	0,59	0,54	0,50	0,41	0,20
10	0,60	0,44	0,37	0,33	0,30	0,23	0,10
20	0,39	0,27	0,22	0,19	0,17	0,13	0,05

2.Значимость коэффициентов уравнения регрессии

По критерию **Стьюдента** $|a_i; a_{ij}| \geq t_{кр}\sqrt{\overline{S_i}^2}$, где $S_i^{\,2} = S^2/N$;

$t_{кр}$ – параметр Стьюдента при $f_2 = N(\gamma - 1)$ и принятом значении α.

Значение $t_{кр}$ при α = 0,01; 0,05; 0,10

Число степеней свободы $f_2 = N(\gamma - 1)$	Вероятность промаха α		
	0,01	0,05	0,10
1	63,66	12,7	6,31
2	9,93	4.30	2,92
3	5,84	3,18	2,35
4	4,60	2,78	2,13
5	2,76	2,57	2,02
6	3,71	2,45	1,94
7	3,50	2,37	1,86
8	3,36	2,31	1,83
9	3,25	2,26	1,81
10	3,17	2,23	1,80

3. Проверка уравнения регрессии на адекватность

По критерию **Фишера** сравнивают отношение дисперсии адекватности к дисперсии . воспроизводимости:

$$F = (S_D^2/f_1) : (S_E^2/f_2) \le F_{кр},$$

где $S_D^2 = \gamma\, S_R^2,\ S_R^2 = \Sigma(Y_n - \bar{Y}_n)^2 = \Sigma\, S_{Rn}^2,$

S_{Rn} – дисперсия отклонения опытного среднего значения Y от расчётного по уравнению регрессии $\bar{Y}$;

S_R^2 – суммарная дисперсия этого отклонения,

$f_1 = N - q$, $f_2 = N(\gamma - 1)$, q – число коэффициентов в уравнении регрессии

Значение $F_{кр}$ при $\alpha = 0{,}05$

f_2 \ f_1	1	2	3	4	5	8
1	161	200	216	225	230	238
2	18,5	19	19,2	19,2	19,3	19,4
3	10,1	9,55	9,28	9,12	9,01	8,85
5	6,61	5,79	5,41	5,19	5,05	4,82
8	5,32	4,46	4,07	3,84	3,69	3,44
10	4,96	4,10	3,71	3,48	3,33	3,07
15	4,54	3,68	3,29	3,06	2,90	2,64
20	4,35	3,49	3,10	2,87	2,71	2,39

Общая стратегия планирования эксперимента

1. Выбор определяющих параметров.
2. Группировка в безразмерные комплексы.
3. Определение границ изменения
4. Составление полного факторного плана
5. Выбор математической модели
6. Оценка составляющих модели по критериям

5. Составим полную таблицу плана с учётом парного и тройного взаимодействия

№ опыта	X_0	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$
1	+	+	+	+	-	+	-	-
2	+	-	+	+	+	-	+	-
3	+	-	-	+	+	+	-	+
4	+	+	-	-	+	+	+	-
5	+	-	+	-	-	+	+	+
6	+	+	-	+	-	-	+	+
7	+	+	+	-	+	-	-	+
8	+	-	-	-	-	-	-	-

№ опыта	X_0	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$
9	+	-	-	-	+	-	+	+
10	+	+	-	-	-	+	-	+
11	+	+	+	-	-	-	+	-
12	+	-	+	+	-	-	-	+
13	+	+	-	+	+	-	-	-
14	+	-	+	-	+	+	-	-
15	+	-	-	+	-	+	+	-
16	+	+	+	+	+	+	+	+

Внимание.

Составление двух планов обусловлено наличием парного и тройного взаимодействия.

Это повышает вероятность оценить линейные эффекты независимо от парных взаимодействий и проверить адекватность линейной модели.

Во второй серии эксперимента все знаки факторов противоположные.

Этот метод называется **методом перевала**

6. Порядок обработки результатов измерений

Наблюдения проводились дважды ($\gamma = 2$) при каждом сочетании определяющих факторов.

Определение построчных дисперсий

№ опыта	X_1	X_2	$X_{3\ldots}$	Y_1	$Y_{2\ldots}$	$\overline{Y}$	S_i^2
1				24,0	23,2	23,6	0,32
2				30,1	30,7	30,4	0,18
3				31,5	31,9	31,7	0,08
4				40,0	40,6	40,3	0,18
5				28,4	28,6	28,5	0,02
6				31,7	32,7	32,2	0,50
7				26,9	26,7	26,8	0,02
8				30,0	31,0	30,5	0,50

Суммарная построчная дисперсия $\boldsymbol{S_E^2 = \Sigma\, S_i^2 = 1{,}80}$

$\boldsymbol{S_i^2}$ – строчная дисперсия, для первой строки $\boldsymbol{S_1^2} = (23{,}2 – 23{,}6)^2 + (24{,}0 – 23{,}6)^2 = 0{,}32$.

Дисперсия воспроизводимости выходного параметра $\boldsymbol{S^2 = S_E^2/(f_2 \cdot \gamma)} = 1{,}80/(8 \cdot 2) = 0{,}1125$,

при $\boldsymbol{\gamma = 2}$ и $\boldsymbol{f_2 = N(\gamma – 1)} = 8\ (2- 1) = 8$.

Максимальная построчная дисперсия $\boldsymbol{S_{n\ \max}^2} = 0{,}50$.

Параметр Кохрена в этом случае равен $\boldsymbol{G = S_{n\ \max}^2 / S_E^2} = 0{,}5/1{,}80 = 0{,}278$

Критическое значение параметра Кохрена при $(\gamma – 1) = 1$ и $N = 8$ равно $\approx 0{,}68$.

$G = 0{,}278 < G_{кр} = 0{,}68$ – ***воспроизводимость опытов*** достаточная.

7. Вычисление коэффициентов уравнения регрессии

№ опыта	$X_1 \overline{Y}$	$X_2 \overline{Y}$	$X_3 \overline{Y}$	$X_1 X_2 \overline{Y}$	$X_1 X_3 \overline{Y}$	$X_2 X_3 \overline{Y}$	$X_1 X_2 X_3 \overline{Y}$	$\overline{Y}$
1	- 23,6	- 23,6	- 23,6	+ 23,6	+ 23,6	+ 23,6	- 23,6	23,6
2	+ 30,4	- 30,4	- 30,4	- 30,4	- 30,4	+ 30,4	+ 30,4	30,4
3	- 31,7	+31,7	-31,7	-31,7	+31,7	-31,7	+ 31,7	31,7
4	+40,3	+40,3	- 40,3	+40,3	-40,3	-40,3	-40,3	40,3
5	- 28,5	- 28,5	+ 28,5	+28,5	-28,5	- 28,5	+28,5	28,5
6	+32,2	-32,2	+ 32,2	- 32,2	+ 32,2	-32,2	- 32,2	32,2
7	- 26,8	+ 26,8	+ 26,8	- 26,8	- 26,8	+ 26,8	- 26,8	26,8
8	+ 30,5	+ 30,5	+ 30,5	+ 30,5	+ 30,5	+ 30,5	+ 30,5	30,5
Сумма	22,8	14,6	- 8	1,8	- 8	- 21,4	- 1,8	244

Коэффициенты уравнения регрессии:

$a_0 = \Sigma Y / N = 244/ 8 = 30{,}5$; $a_j = \Sigma X_{ji} \overline{Y} / N$; $a_{jk} = \Sigma X_{ji} X_{ki} \overline{Y} / N$; $a_{jkm} = \Sigma X_{ji} X_{ki} X_{mi} \overline{Y} / N$;

$a_1 = 22{,}8/8 = 2{,}85$; $a_2 = 14{,}6/8 = 1{,}83$; $a_3 = -8/8 = -1$;

$a_{12} = 1{,}8/8 = 0{,}23$; $a_{13} = -8/8 = -1$; $a_{23} = -21{,}4/8 = -2{,}68$; $a_{123} = -1{,}8/8 = -0{,}23$.

Уравнение регрессии:

$$Y = 30{,}5 + 2{,}85X_1 + 1{,}83\,X_2 - X_2 + 0{,}23\,X_1 X_2 - X_1 X_3 - 2{,}68\,X_2 X_3 - 0{,}23\,X_1 X_2 X_3.$$

8. Оценка значимости коэффициентов уравнения регрессии

По критерию Стьюдента $|a_i; a_{ij}| \geq t_{кр}\sqrt{S_i^{\,2}}$, где $S_i^{\,2} = S^{\,2}/N$;

Дисперсия коэффициентов уравнения регрессии $S_i^{\,2} = S_{ij}^{\,2} = S^{\,2}/N = 0{,}1125/8 = 0{,}014$.

При $f_2 = N(\gamma - 1) = 8\ (2 - 1) = 8$ и $\alpha = 0{,}05$ значение $t_{кр} = 1{,}86$ и $|\ t_{кр}\sqrt{S_i^{\,2}} = 1{,}86\sqrt{0{,}014} = 0.223$.

Все коэффициенты значимы, хотя a_{12} и a_{123} почти в пределах ошибки. Осталось $q = 6$.

9. Проверка уравнения регрессии на адекватность

По критерию Фишера $F = (S_D^{\,2}/f_1) : (S_E^{\,2}/f_2) \leq F_{кр}$, где $S_D^{\,2} = \gamma\, S_R^{\,2}$, $S_R^{\,2} = \Sigma(Y_n - \hat{Y}_n)^2 = \Sigma\, S_{Rn}^{\,2}$,

S_{Rn} – дисперсия отклонения опытного среднего значения Y от расчётного по уравнению регрессии Y;

$S_R^{\,2}$ – суммарная дисперсия этого отклонения,

$f_1 = N - q$, $f_2 = N(\gamma - 1)$, q – число коэффициентов в уравнении регрессии

№ опыта	a_0	a_1X_1	a_2X_2	a_3X_3	$a_{13}X_1X_3$	$a_{23}X_2X_3$	Y_n	$\overline{Y}_n$	$(Y_n - \overline{Y}_n)^2$
1	30,5	- 2,85	- 1,83	1	- 1	- 2,68	22,9	23,6	0,49
2	30,5	2,85	- 1,83	1	1	- 2,68	31,1	30,4	0,49
3	30,5	- 2,85	1,83	1	- 1	2,68	31,9	31,7	0,04
4	30,5	2,85	1,83	1	1	2,68	40,1	40,3	0,04
5	30,5	- 2,85	- 1,83	- 1	1	2,68	28,7	28,7	0
6	30,5	2,85	- 1,83	- 1	- 1	2,68	31,9	32,2	0,09
7	30,5	- 2,85	1,83	- 1	1	- 2,68	27,0	26,8	0,04
8	30,5	2,85	1,83	- 1	- 1	- 2,68	30,2	30,5	0,09

$S_R^{\,2} = \Sigma(Y_n - \hat{Y}_n)^2 = \Sigma\, S_{Rn}^{\,2} = 1{,}28$

$S_D^{\,2} = \gamma\, S_R^{\,2} = 2 \cdot 1{,}28 = 2{,}56$; $S_E^{\,2} = 1{,}8$; $f_1 = N - q = 8 - 6 = 2$; $f_2 = N(\gamma - 1), = 8$; $F = (2{,}56/2) / (1{,}8/8) = 5{,}867 > F_{кр}, = 4{,}46$.

ВЫВОД. Принятая модель не адекватна физическому эксперименту.

Планирование отсеивающего эксперимента по выбору значимых факторов без повторных опытов (при $\gamma = 1$).

1. Адекватность можно проверить, если часть коэффициентов окажется незначимой. В этом случае появится необходимая степень свободы.

Для проверки адекватности модели необходима некоторая степень свободы. Насыщенные планы не имеют её и НЕ могут быть проверены на адекватность.

2. Можно заранее ввести фиктивные факторы, так чтобы вместе с реальными факторами они образовали насыщенный план.

Примечание. Эффекты фиктивных факторов были бы равны нулю, если бы все опыты проводились абсолютно точно, что невозможно. Поэтому фиктивные факторы можно использовать для оценки значимости коэффициентов модели по критерию Стьюдента. Статистически значимыми признаются эффекты, равные или превышающие свои доверительные интервалы:

$$|a_i| \geq t_{\alpha, f1} \cdot S_a,$$

где a_i – коэффициент регрессии при x_i; $t_{\alpha, f1}$ – значение критерия распределения Стьюдента;

f_1 – число степеней свободы при расчёте дисперсии опыта S_y^2, $f_1 = N - k - 1$;

S_a^2 – дисперсия оценок коэффициентов, $S_a^2 = S_y^2 / N$;

$$S_y^2 = \left(\sum_{J=1}^{N-k-1} b_j^2\right) N / (N - k - 1),$$ где k – число реальных факторов;

b_j – коэффициент регрессии при j – том фиктивном факторе.

Определение параметров модели по данным эксперимента, выполненного по насыщенному плану

Провести выбор значимых факторов, влияющих на отбеливание чугуна по результатам отсеивающего эксперимента, выполненного согласно насыщенного плана с фиктивными факторами. Число факторов $k = 15$, число опытов $N = 20$.

<u>Значения факторов отсеивающего эксперимента</u>

Фак-торы	Содержание, %					Тип шихты	Способ науглероживания	Время расплавления, час	Температура расплавления, °C	Максимальная температура, °C	Время выдержки, мин	Тип модификатора	Температура расплавления, °C	Время модифицирования, мин	Кол-во модификатора, %	Фиктивные факторы			
	Mn	Si	C	S	P														
Верхний уровень (+1)	0,6	3	3,5	0,05	0,05	Уг. Ст.	Кус-ки	2,5	1450	1600	20	SiCa	1450	15	0,6	-	-	-	-
Нижний																			

Опыт	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	Отбел, %
1	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	30
2	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	10
3	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	30
4	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	18
5	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	40
6	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	40
7	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	22
8	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	36
9	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	31
10	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	30
11	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	21
12	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	31
13	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	32
14	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	45
15	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	55
16	+	+	+	+	-	+	-	+	-+	-	-	-	+	+	-	+	+	-	-	17
17	-	+	+	+	+	-	+	-	-	-	-	-	-	+	+	-	+	+	-	50
18	-	-	+	+	+	+	-	+	+	+	-	-	-	-	+	+	-	+	+	30
19	+	-	-	+	+	+	+	-	-	-	+	-	-	-	-	+	+	-	+	10
20	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	15
a_i	-0,75	2,95	2,85	-0,35	2,65	-0.75	1,15	-5,25	-1 95	1,55	-2,85	5,35	-0,45	3,65	5,95	-0,95	0,05	0,55	1,25	

$a_0 = 29,65$

Пример

Число опытов *N* = 20; число факторов 19, из них число реальных *k* =15 и 4 фиктивных.

$$S_y^2 = \frac{20\,(0{,}95^2 + 0{,}05^2 + 0{,}55^2 + 1{,}25^2)}{20 - 15 - 1} = 13{,}85;\ S_y = 3{,}722.$$

Дисперсия оценок коэффициентов $S_a^2 = S_y^2 / N. = 13{,}85/20 = 0{,}693;\ S_a = 0{,}832.$

$t_{\alpha, f1}$ в данном эксперименте определим при $f_1 = N - k - 1 = 20 - 15 - 1 = 4$ и $\alpha = 0{,}05$; критерий $t_{0{,}05;\,4} = 2{,}78.$
Таким образом, значимы коэффициенты регрессии, значения которых $S_a \cdot t_{\alpha, f1} = 0{,}832 \cdot 2{,}78 = 2{,}313$ и выше.
Остальные отбросим. Теперь в модели число членов $k_м$ меньше исходного

Пусть по 20 опытам получена модель с 9 коэффициентами:

$$Y = Y_0 + a_2 X_2 + a_3 X_3 + a_5 X_5 + a_8 X_8 + a_{11} X_{11} + a_{12} X_{12} + a_{14} X_{14} + a_{15} X_{15} =$$
$$= 29{,}66 + 2{,}95\, X_2 + 2{,}85 X_3 + 2{,}65 X_5 - 5{,}25 X_8 - 2{,}85 X_{11} + 5{,}35 X_{12} + 3{,}65\, X_{14} + 5{,}95 X_{15}$$

Для каждого опыта рассчитываем значение $Y_{u\text{- расч}}$, определяет $Y_{u\text{- расч.}} - Y_{u\text{- эксп}}$.

Пусть $\sum_{u=1}^{N} (Y_{u\text{- расч.}} - Y_{u\text{- эксп.}})^2 = 235$; $S^2_{неад} = 235/\,(20 - 9) = 21{,}36$; $f_2 = 11$, $\alpha = 0{,}05$

$F_{\alpha;\, f1;\, f2} = F_{0{,}5;\,11} = 5{,}93$. $F_{расч.} = S^2_{неад} \,/\, S_y^2 = 21{,}36/\,13{,}85 = 1{,}54 < F_{0{,}5;\,11} = 5{,}93.$

Модель можно считать адекватной.

Критерии оптимальности и типы планов. Параметр оптимизации

Для некоторых планов важную роль играет свойство *композиционности.*

Композиционные планы для построения полиномов более второго порядка получают
добавлением некоторых точек к планам формирования линейных функций.

Построение планов производится с использованием каталогов планов или с использованием методов планирования экспериментов, что является непростой задачей.

Требования к параметру оптимизации

- параметр оптимизации – это количественный признак, по которому оптимизируют процесс; *задаётся числом* (измеряемой величиной или субъективно рангом).

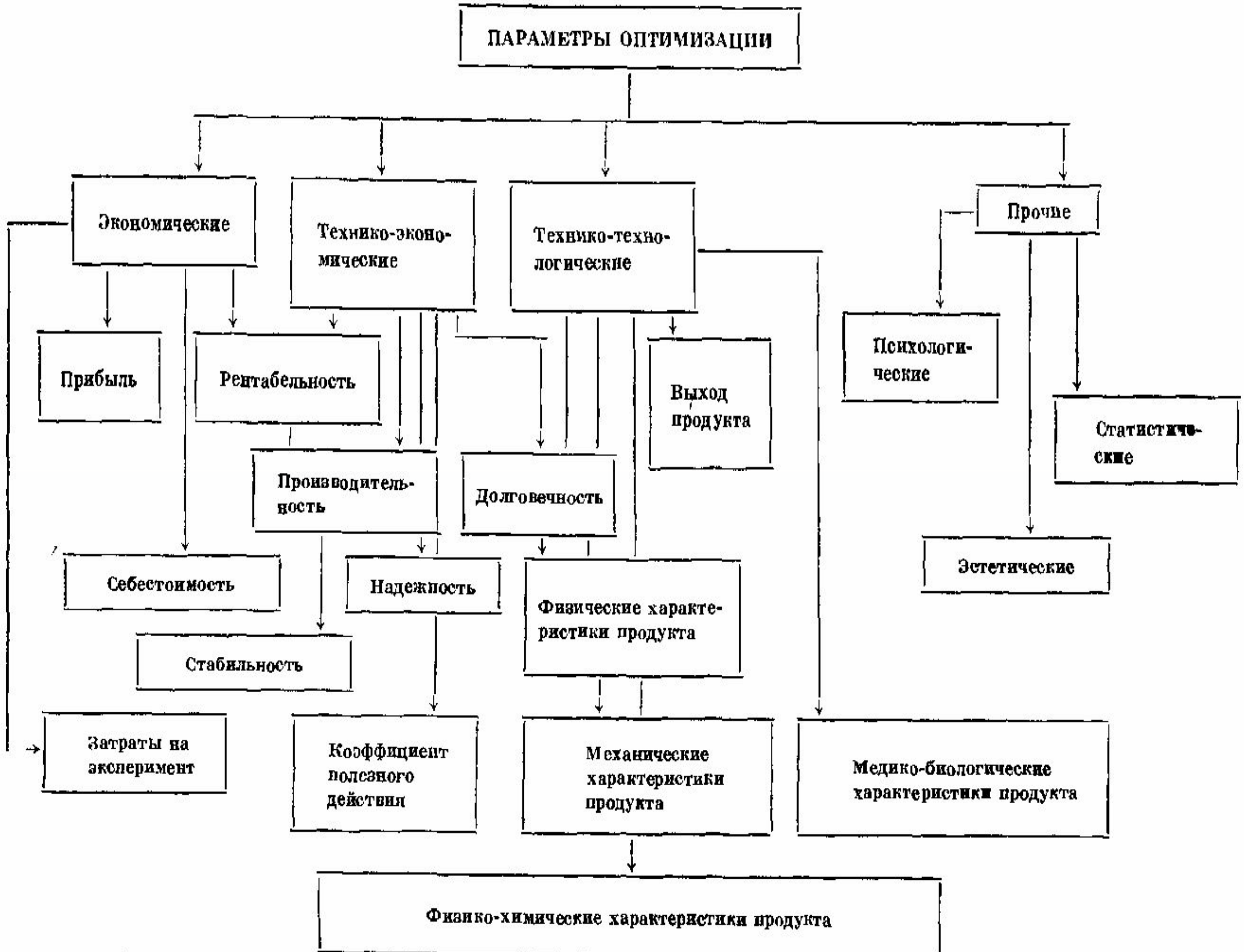
- параметр оптимизации должен выражаться *одним числом.*

- однозначность параметра в статистическом смысле, т.е. заданному набору значений факторов должно соответствовать одно с точностью до ошибки эксперимента значение параметра оптимизации.

- функционально эффективный параметр оптимизации, должен быть достаточно эффективным и в статистическом смысле, т.е. определяемым с достаточной точностью.

- параметр оптимизации должен удовлетворять требованию *универсальности* или *полноты, т.е.* всесторонне характеризовать объект; полнота обеспечивается применением обобщенного параметра оптимизации, составленного из частных.

- параметр оптимизации имел *физический смысл, был простым и легко вычисляемым.*


ПАРАМЕТРЫ ОПТИМИЗАЦИИ
Экономические
Технико-эконо-мические
Технико-техно-логические
Прочие
Прибыль
Рентабельность
Выход продукта
Психологи-ческие
Статистиче-ские
Производитель-ность
Долговечность
Себестоимость
Надежность
Физические характе-ристики продукта
Эстетические
Стабильность
Затраты на эксперимент
Коэффициент полезного действия
Механические характеристики продукта
Медико-биологические характеристики продукта
Физико-химические характеристики продукта

Задачи с несколькими выходными параметрами

Математические модели можно построить с учётом каждого из параметров, но одновременно оптимизировать несколько функций невозможно.

Из многих выходных параметров выбирается один в качестве параметра оптимизации, а остальные служат ограничениями.

Всегда полезно исследовать возможность уменьшения числа выходных параметров. Для этого можно воспользоваться корреляционным анализом.

ГРАДИЕНТНЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

Градиент указывает направление наибольшего возрастания функции. В точке экстремума градиент равен нулю.

Задача оптимизации заключается в нахождение экстремума функции отклика в области допустимых значений параметров. Такой эксперимент называется экстремальным.

Решение задачи оптимизации

- определение критерия эффективности системы, цели исследования, изучения сущности исследуемого процесса, анализа имеющихся ресурсов, возможности проведения экспериментов с изучаемым объектом в необходимом диапазоне изменения *существенных* (менее 15) факторов;

- определение диапазона и характера изменения (непрерывность или дискретность) каждой переменной, начальной области планирования эксперимента и интервалов варьирования факторов;

- факторы должны быть управляемыми, т.е. поддерживаться постоянными в течение каждого опыта; обеспечиваться независимость изменения каждого фактора.

Требования, предъявляемые к факторам при планировании эксперимента

- должны быть *управляемыми;*

- *точность замера факторов должна быть возможно более высокой.*

- должны быть *однозначны.*

- *совместимыми,* т.е. все их комбинации осуществимы и безопасны

- должны быть *независимыми, т.е. возможность* установления фактора на любом уровне вне зависимости от уровней других факторов.

– отсутствие корреляции между факторами; достаточно, чтобы возможная связь была *нелинейной*.

Требования к модели плана при решении оптимизационной задачи

–способность предсказывать направление дальнейших опытов с требуемой точностью;

– адекватность модели. предсказанное с помощью модели значение отклика не должно отличаться от фактического больше, чем на некоторую заранее заданную величину;

– простота модели; предпочитают алгебраические полиномы, или степенные ряды (отрезки степенных рядов), начиная с минимальной степени полинома.

Всегда существует такая окрестность любой точки (точнее, почти любой точки),
в которой линейная модель адекватная. Размер такой области заранее не известен,
но может быть установлен экспериментально.

Поиск линейной модели продолжается пошагово до выхода в «почти стационарную»
область. Для уточнения оптимума в этой области потребуется полином более

Способы градиентной оптимизации

1. Метод *покоординатного подъема или* ***метод Гаусса – Зейделя****.*

из некоторой точки по координате v_1 *до тех пор, пока не станет равной нулю соответствующая производная* $\partial f(V) / \partial v_1 = 0$.

Все остальные координаты (аргументы функции) сохраняют постоянное значение.

После этого подъем начинается по другой координате.

Процесс заканчивается, когда все частные производные будут равны нулю

2. Метод *наискорейшего подъема* (***метод крутого восхождения***).

Движение осуществляется в направлении градиента ***grad y = f*** (x_1, x_2)., определённого в исходной точке, далее подъем в этом направлении осуществляется до тех пор, пока производная $df(V) / dV$ в этом направлении не обратится в нуль. После этого снова определяют градиент и осуществляют по нему подъем до нулевого значения производной и т.д.

Модификация этого метода предусматривает вычисление градиента в каждой новой точке траектории перемещения.

Одна из основных проблем применения градиентного метода поиска заключается в выборе величины каждого дискретного шага. Шаги могут быть постоянными или переменными. Второй вариант в реализации алгоритма более сложный, но обычно требует меньшего количества итераций.

Если функция отклика является линейной, то шаг выбирается исходя из эвристических предположений исследователя о виде функции отклика.

Пусть в окрестности точки M_0, как центра плана, поставлен ПФЭ 2^2
Координаты отдельных опытов соответствуют точкам 1, 2, 3, 4.
По ПФЭ найти коэффициенты линейного уравнения регрессии: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$.
После этого можно найти градиент $\mathbf{grad}\ y = (\partial y/\partial x_1)\vec{i} + (\partial y/\partial x_2)\vec{j} = b_1\vec{i} + b_2\vec{j}$

Для движения по градиенту необходимо изменять факторы пропорционально их коэффициентам регрессии в сторону, соответствующую знакам коэффициентов.
В процессе поиска двигаться в этом направлении, пока не будет найден локальный максимум (т.M_1). После чего находят направление градиента, осуществляя ПФЭ.
И далее процедура повторяется.

1. Планирование и постановка ПФЭ (или ДФЭ) в окрестности точки (M_0). Расчет коэффициентов линейной регрессии; определение направления градиента.

2. Расчет произведений $b_i \Delta x_i$, где Δx_i- интервал варьирования факторов при ПФЭ

7. В окрестности локального экстремума ставят новую серию опытов (ПФЭ или ДФЭ), определяют новые значения коэффициентов уравнения регрессии и новое направление градиента. Процедура повторяется до достижения нового локального экстремума и т.д., вплоть до определения окрестности координат максимума функции отклика, которая
носит название **почти стационарной области**.
Признаком достижения этой области является статистическая незначимость коэффициентов. В этой области становятся значимыми эффекты взаимодействия и квадратичные эффекты. Здесь требуется переходить от ДФЭ к ПФЭ и к планам второго порядка.

Бокс и Уилсон предложили в 1951 г. **модификацию метода крутого восхождения**.

1. На начальном этапе поиска применяют линейные полиномы для описания функции отклика.
2. Значение градиента оценивается в начальной точке, после чего пошаговое движение по градиенту продолжается до попадания в частный оптимум (до тех пор, пока значение функции отклика возрастает при переходе от точки к точке).
3. В точке частного оптимума с помощью факторного эксперимента снова определяется градиент. И пошаговое движение начинается по новому направлению.
4. Процесс продолжается до попадания в область глобального экстремума.
5. Но эта область не может быть адекватно описана линейным уравнением; переходят к более точному описанию поверхности отклика на основе полиномов второго порядка. Построение плана для формирования полинома второй степени производится путем добавления некоторых точек к "ядру", уже сформированному для линейного приближения (такие планы называют *композиционных).*
6. *В целом метод Бокса – Уилсона во многих случаях требует меньшего количества опытов возможно при несколько большем числе шагов.*

Градиентные методы не обеспечивают гарантированного нахождения глобального оптимума.

Если эксперимент проводится на реальном объекте и требует больших затрат ресурсов, то поиск значений параметров может завершиться при получении удовлетворительных, а не оптимальных, значений функции отклика.

Симплекс-планирование

Позволяет без предварительного изучения влияния факторов найти область оптимума.

Этот метод относится к безградиентным метода поиска оптимума.

Для этого используется специальный план эксперимента в виде **симплекса**.

Симплекс – простейший выпуклый многогранник, образованный ***к+1*** вершинами в ***к***-мерном пространстве, которые соединены между собой прямыми линиями: симплекс ***к=2*** – треугольник, ***к=3*** – тетраэдр и т.д. Симплекс называется **правильным**,
если все расстояния между его вершинами (ребра) равны.

Координаты вершин симплекса являются значениями факторов в отдельных опытах.

1. Строится исходный симплекс, проводятся опыты в его вершинах результатов.
2. Выбирается вершина с наименьшим значением функции отклика.
3. Ставится опыт в новой точке, являющейся зеркальным отображением точки с наихудшим (минимальным) результатом.

Процесс повторяется до тех пор, пока не будет найдена стационарная область.

Для окончания процесса используются следующие критерии:

1 – разность значений функции отклика в вершинах симплекса становится меньше ранее заданной. Это означает вход в область вблизи оптимума или области «плато»;

2 - отражение любой из вершин симплекса после однократного «качания» приводит к возврату в исходное положение.

3 – циклическое движение симплекса вокруг одной из его вершин на протяжении более, чем нескольких шагов, т.е. циркулирует вокруг области оптимума.

В случаях 2 и 3 рекомендуется уменьшать размеры симплекса, т.е. расстояние между вершинами, для уточнения координаты оптимума.

Данный метод прост, но работает не достаточно быстро.

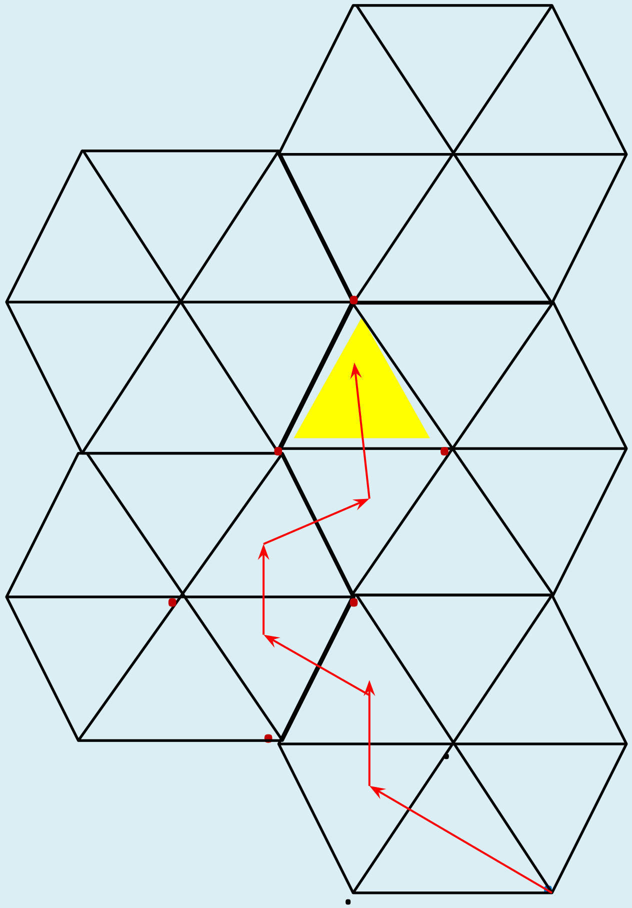
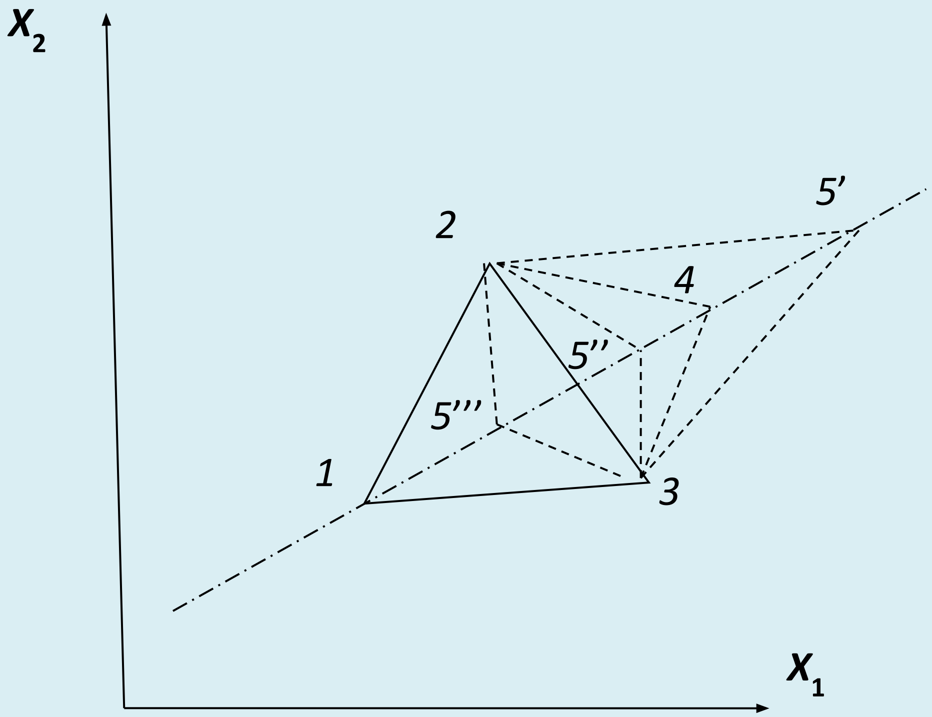


Схема поиска области оптимума симплексным методом

Метод деформируемого симплекса (многогранника).

Ускорение достигается за счет того, что отражение осуществляется не на постоянную величину.



Точка 4 очередного опыта соответствует нормальному отражению наихудшей вершины 1, Точки 5′, 5″, 5‴ – последующих опытов для случаев, соответственно, растяжения, сжатия и отрицательного сжатия многогранника.

ПЛАНЫ ДЛЯ ОПИСАНИЯ ПОВЕРХНОСТИ ОТКЛИКА

1. Композиционные планы

Применение линейных планов совместно с методом градиентного поиска оптимума позволяет достичь окрестностей точки оптимума. Поиск оптимального решения в этой области требует перехода от линейных моделей к моделям более высокого порядка – как минимум к полиномам второй степени

Требуется план, в котором каждая переменная принимает хотя бы три различных значения. ПФЭ типа 3^k *имеют большую избыточность. Так, при* $k = 3$ *количество точек*

27 при количестве оцениваемых коэффициентов – 10.

Планирование рационально осуществлять путем добавления специально подобранных точек к “ядру”, образованному планированием для линейного приближения.

Такие планы называют *композиционными (последовательными), они позволяют использовать информацию, полученную в результате реализации линейного плана.*

В центральных композиционных планов (ЦКП) используют в качестве ядра полный факторный эксперимент или минимально возможные регулярные дробные реплики типа 2^{k-p}*. В качестве дробной реплики применяют такую, в которой два любых парных взаимодействия по модулю не равны друг другу.*

Центральный композиционный план второго порядка называют планом Бокса, если его ядром является ПФЭ 2^k *или регулярная реплика типа* 2^{k-p}*, для которой парные взаимодействия не равны по модулю линейным факторам и не равны между собой.*

Ядром плана Бокса при $k < 5$ *является ПФЭ, а при* $k > 5$ *может быть ДФЭ.*

Центральный композиционный план второго порядка называют планом Хартли, если его ядром является регулярная реплика типа 2^{k-p}*, в которой некоторые парные взаимодействия равны по модулю линейным факторам.*

Планы Хартли более экономичны по числу опытов, чем планы Бокса, но уступают им по точности оценивания коэффициентов, кроме того, их нельзя сделать ни ортогональными,
ни ротатабельными. Такой план не позволяет получить раздельные оценки соответствующих коэффициентов.

2. Ортогональные центральные композиционные планы

В планах *ЦКП* Бокса *второго порядка* к ядру, построенному на основе ПФЭ или ДФЭ, добавляется одна точка в центре плана с координатами (0, 0, ..., 0) и 2*k* *"звездных" точек с координатами* (±γ, 0, ..., 0), ..., (0, 0, ..., ±γ).
Общее количество точек плана $N = N_0 + 2k + 1$*, где* N_0 *– количество точек ядра плана.*

Ядро ЦКП при *k* =2		Дополнительные точки	
x1	*x2*	*x1*	*x2*
+	+	γ	0
–	+	– γ	0
+	–	0	γ
–	–	0	– γ
		0	0

Аналогично строятся ЦКП для произвольного числа факторов, при этом каждый фактор варьируется на пяти уровнях: – γ; – 1; 0; 1; γ.

ЦКП второго порядка для трех переменных

	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
План 2^3	+	-	-	-	+	+	+	+	+	+
	+	+	-	-	-	-	+	+	+	+
	+	-	+	-	-	+	-	+	+	+
	+	+	+	-	+	-	-	+	+	+
	+	-	-	+	+	-	-	+	+	+
	+	+	-	+	-	+	-	+	+	+
	+	-	+	+	-	-	+	+	+	+
	+	+	+	+	+	+	+	+	+	+
Звёзд. план	+	$-\gamma$	0	0	0	0	0	γ^2	0	0
	+	γ	0	0	0	0	0	γ^2	0	0
	+	0	$-\gamma$	0	0	0	0	0	γ^2	0
	+	0	γ	0	0	0	0	0	γ^2	0
	+	0	0	$-\gamma$	0	0	0	0	0	γ^2
	+	0	0	γ	0	0	0	0	0	γ^2
Центр плана	+	0	0	0	0	0	0	0	0	0

Для устранения асимметрии и нарушений ортогональности ЦКП Бокса преобразуют квадратичные параметры и специальным образом выбирают величину плеча γ.

1. Для обеспечения симметричности в столбцах квадратичных параметров вводится поправка $(-c)= -(N_0+2\gamma^2)/N$

2. Для обеспечения ортогональности принимают $\gamma= \{[(N N_0)^{1/2} - N_0]/2\}^{1/2}$.

Значения γ, обеспечивающие ортогональность, например, для ядер 2^2, 2^3, 2^4, 2^{5-1}, 1,

x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
+	-	-	-	+	+	+	1-c	1-c	1-c
+	+	-	-	-	-	+	1-c	1-c	1-c
+	-	+	-	-	+	-	1-c	1-c	1-c
+	+	+	-	+	-	-	1-c	1-c	1-c
+	-	-	+	+	-	-	1-c	1-c	1-c
+	+	-	+	-	+	-	1-c	1-c	1-c
+	-	+	+	-	-	+	1-c	1-c	1-c
+	+	+	+	+	+	+	1-c	1-c	1-c
+	-γ	0	0	0	0	0	γ^2-c	-c	-c
+	γ	0	0	0	0	0	γ^2-c	-c	-c
+	0	-γ	0	0	0	0	-c	γ^2-c	-c
+	0	γ	0	0	0	0	-c	γ^2-c	-c
+	0	0	-γ	0	0	0	-c	-c	γ^2-c
+	0	0	γ	0	0	0	-c	-c	γ^2-c
+	0	0	0	0	0	0	-c	-c	-c

3. Ротатабельные центральные композиционные планы

При описании поверхности отклика в окрестностях точки оптимума более значимой является оценка дисперсии уравнения в целом, а не оценка дисперсии отдельных коэффициентов полинома.

Путем специального подбора звездного плеча □ ЦКП Бокса можно сделать ротатабельным

Точки ротатабельного ЦКП Бокса второго порядка располагают на концентрических гиперсферах:

- первая гиперсфера может быть вырожденной, т. е. представлять собой центральную точку плана, ее радиус равен 0; обычно это используется на практике;

-вторая гиперсфера соответствует вписанному в нее кубу, выбранному в качестве ядра плана; *ядро представляет собой ПФЭ вида $2k$ или ДФЭ вида $2k – p$, причем должно соблюдаться условие $(k – p)/4 > ¾$; с учетом ограничений на ЦКП Бокса, если $k \geq 5$, то в качестве ядра можно использовать полуреплику, если $k \geq 8$, ядром может служить чет- верть реплика;*

- третья гиперсфера имеет радиус, равный $2^{k/4}$ для ядра в виде ПФЭ , и радиус, равный $2^{(k - p)/4}$ для ядра в виде ДФЭ.

Таким образом, каждый фактор в ротатабельном ЦКП Бокса варьируется на пяти уровнях.

В некоторых случаях радиусы второй и третьей гиперсферы совпадают:

при $n = 2$ радиус $\rho_2 = 2^{1/2}$, $\rho_3 = 2^{2/4} = 2^{1/2}$;

при $n = 8$ и $p = 2$ радиус $\rho_2 = 8^{1/2} = 2^{3/2}$, $\rho_3 = 2^{(8 – 2)/4} = 2^{3/2}$.

Количество точек плана второго порядка для *n* -трехфакторного эксперимента

n	α	$N_α$	N_0	N_c	N
2	1,414	4	5	4	13
3	1,682	6	6	8	20
4	2	8	7	16	31

№ опыта	x_0	x_1	x_2	x_3	x_{12}	x_{22}	x_{32}	x_1x_2	x_1x_3	x_2x_3
	z_0	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
1	+1	-1	-1	-1	+1	+1	+1	+1	+1	+1
2	+1	+1	-1	-1	+1	+1	+1	-1	-1	+1
3	+1	-1	+1	-1	+1	+1	+1	-1	+1	-1
4	+1	+1	+1	-1	+1	+1	+1	+1	-1	-1
5	+1	-1	-1	+1	+1	+1	+1	+1	-1	-1
6	+1	+1	-1	+1	+1	+1	+1	-1	+1	-1
7	+1	-1	+1	+1	+1	+1	+1	-1	-1	+1
8	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
9	+1	-1,682	0	0	2,828	0	0	0	0	0
10	+1	+1,682	0	0	2,828	0	0	0	0	0
11	+1	0	-1,682	0	0	2,828	0	0	0	0
12	+1	0	+1,682	0	0	2,828	0	0	0	0
13	+1	0	0	-1,682	0	0	2,828	0	0	0
14	+1	0	0	+1,682	0	0	2,828	0	0	0
15	+1	0	0	0	0	0	0	0	0	0
16	+1	0	0	0	0	0	0	0	0	0
17	+1	0	0	0	0	0	0	0	0	0
18	+1	0	0	0	0	0	0	0	0	0
19	+1	0	0	0	0	0	0	0	0	0
20	+1	0	0	0	0	0	0	0	0	0

Пример

Построить матрицу ротатабельного ЦКП Бокса второго порядка для 3-х факторов.

Решение.

1. Ядро плана ПФЭ вида 2^3 ; радиус соответствующей гиперсферы $\rho_2 = 3^{1/2} = 1{,}732$.
2. Звездные точки располагаются на гиперсфере с радиусом $\rho_3 = 2^{3/4} = 1{,}682$ и имеют координаты (±1,682; 0; 0), (0; ± 1,682; 0), (0; 0; ± 1,682).
3. Матрица планирования включает три гиперсферы, в которой γ = 1,682.
4. План содержит 15 точек и является ненасыщенным – количество оцениваемых коэффициентов 10.

В табл. приведены минимально необходимые сведения для составления рассмотренного вида ротатабельных ЦКП.

Количество факторов	Число точек ПФЭ	Число звездных точек	Значение γ
2	4	4	1,414
3	8	6	1,682
4	16	8	2,000
5	32	10	2,378
5, полуреплика	16	10	2,000
6	64	12	2,828
6, полуреплика	32	12	2,378
7	128	14	3,364
7, полуреплика	64	14	2,828

Пример.

Уравнение Аррениуса $k = k_0 \exp - E/RT$. Обычно переменная T изменяется не более чем на 10-15% от среднего, в этом случае информационная матрица для линеаризованной модели близка к вырождению.
Коэффициенты корреляции оценок параметров могут доходить до 0,97 и 0,98.

Рекомендуется **репараметризация**, т.е. переход к новой модели:

$k = k_0(\overline{T}) \exp [-(1/T - 1/\overline{T}) E/R]$, где $k_0(\overline{T}) = k_0 \exp - E/R\overline{T}$, $\overline{T}$ – среднее значение температуры.

Параметры $k_0(\overline{T})$ и E оказываются закоррелированными слабее, коэффициент корреляции удаётся снизить до 0,5, резко сужаются совместные доверительные границы.

В 1951 г. Бокс и Уилсон предложили подход к решению подобных задач:

1. Экспериментатор ставит последовательно небольшие серии опытов, в каждой из которых одновременно варьируются по определенным правилам все факторы.
2. Серии организуются так, чтобы после математической обработки предыдущей можно было выбрать условия проведения (т.е. спланировать) следующую серию.

Двухпараметрическая нелинейная модель

$$Y = a_0 X_0 + a_1 X_1 + a_2 X_2 + a_{12} X_1 X_2 + a_{11} X_1^2 + a_{22} X_{22}^2.$$

Если в этой модели коэффициенты регрессии для эффектов взаимодействия не равны нулю, то оценка параметров будет смешанной:

□ $Y / N = \Sigma Y_i / N = a_0 + a_{11} + a_{22}$, (в этом случае вместо a_0 получаем $a_0 + a_{11} + a_{22}$).

Если модель нелинейная, то для оценки параметров по результатам наблюдений методом наименьших квадратов надо линеаризовать нелинейную по параметрам функцию, разлагая её в ряд Тейлора в окрестности некоторой точки. Для этого необходимо знать не только вид функции, но и хотя бы грубые оценки её параметров, которые уточняются в процессе эксперимента.

Естественно, должен измениться и оптимальный план эксперимента. На каждом шаге ставится некоторое число опытов, происходит затем переоценка коэффициентов и изменение плана. Это ***последовательная*** стратегия.

Выделить те факторы, которыми можно управлять, и записать предполагаемые соотношения.

- Выбор модели связан с глубоким знанием объекта исследования и области пространства независимых переменных, где будет ставиться эксперимент (многомерный куб, шар, правильный симплекс или несимметричная область).
- Выбор плана эксперимента, оптимального для данной модели, совершенно от объекта исследования не зависит.
- При интерпретации модели исследователь снова обращается к физической реальности.

Виды планов

1. ***D-оптимальный*** план: минимизируется объём эллипсоида рассеяния ошибок параметров уравнения регрессии; детерминант матрицы ($\mathbf{X}^{\mathrm{T}}\mathbf{X}$) будет максимальным, а детерминант ковариационной матрицы $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ – минимальным.

2. ***E- оптимальный*** план: минимизируется максимальную ось эллипсоида рассеивания; ковариационная матрица с минимальным значением максимального характеристического числа.

3. ***A- оптимальный*** план: минимизируется среднюю дисперсию оценок коэффициентов регрессии, эллипсоид рассеяния с наименьшей суммой квадратов длин осей; ковариационные матрицы имеют наименьшие значения следа.

Нет плана, который отвечал бы хотя бы нескольким важнейшим критериям. Нужно компромиссное решение.

Критерии оптимальности и типы планов. Параметр оптимизации

Две основные группы критериев:

-критерии, связанные с ошибками оценок коэффициентов,

- критерии, связанные с ошибкой оценки поверхности отклика.[

Критерии первой группы - для задач оптимизации, выделения доминирующих (наиболее значимых) параметров на начальных этапах решения оптимизационных задач или для выявления несущественных параметров в задачах определения закономерности функционирования объекта.

Пространственное расположение, форма и размер эллипсоида рассеяния ошибок полностью зависят от плана эксперимента.

Критерию *D-оптимальности соответствует минимальный объем эллипсоида*
рассеяния ошибок (минимум произведения всех дисперсий коэффициентов полинома);
эффекты факторов максимально независимы друг от друга; минимизируется ожидаемая ошибка предсказания функции отклика. *Используется при поиске оптимума функции отклика.*

Критерию *A- оптимальности* соответствует план с минимальной суммарной дисперсией всех коэффициентов. *Используется при поиске оптимума функции отклика, построение плана проще, чем при D-оптимальности*

Критерию *E-оптимальности* – план, в котором наибольшая дисперсия коэффициентов будет минимальна. *Используется при изучении влияния отдельных факторов.*

Критерии второй группы используются при решении задач описания поверхности
отклика, определения ограничений на значения параметров.
Основным является критерий *G-оптимальности*, которому соответствует план с минимальным значением наибольшей ошибки в описании функции отклика.

Ортогональным называется план, для которого выполняется условие парной ортогональности столбцов матрицы планирования. При ортогональном планировании коэффициенты полинома определяются независимо друг от друга – вычеркивание или добавление слагаемых в функции отклика не изменяет значения остальных коэффициентов полинома. Эллипсоид рассеяния ориентирован в пространстве так, что
направления его осей совпадают с направлениями координат пространства параметров.

Ротатабельным называется план, для которого обеспечивает при любом направлении от центра эксперимента равнозначность точности оценки функции отклика (постоянство
дисперсии предсказания) на <u>равных расстояниях от центра эксперимента</u>.
Важно при решении задач поиска оптимальных значений параметров на основе градиентного метода. До начала экспериментов направление градиента неизвестно.

По соотношению между количеством оцениваемых неизвестных параметров модели и количеством точек плана эксперимента все планы подразделяются:

- *ненасыщенные – количество параметров меньше числа точек плана;*
- *насыщенные* – обе величины одинаковы;
- *сверхнасыщенные – количество параметров больше числа* точек плана.

Метод наименьших квадратов применяют только при ненасыщенном и насыщенном планировании.