

# Планирование эксперимента

Полный факторный эксперимент



# Основные определения

- Под экспериментом понимают **совокупность операций совершаемых над объектом исследования с целью получения информации о его свойствах.** Эксперимент, в котором исследователь по своему усмотрению может изменять условия его проведения, называется активным экспериментом. Если исследователь не может самостоятельно изменять условия его проведения, а лишь регистрирует их, то это пассивный эксперимент.
- Важнейшей задачей методов обработки полученной в ходе эксперимента информации является задача **построения математической модели изучаемого явления, процесса, объекта.** Ее можно использовать и при анализе процессов и при проектировании объектов. Можно получить хорошо аппроксимирующую математическую модель, если целенаправленно применяется активный эксперимент.



# Основные определения

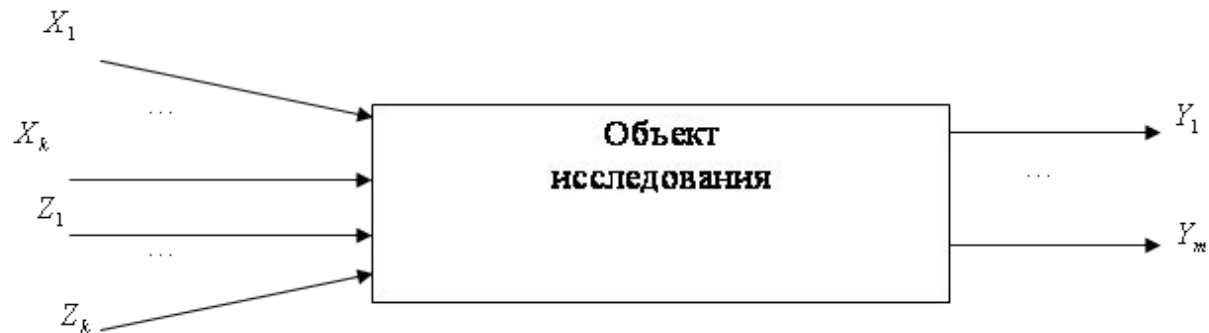
- Другой задачей обработки полученной в ходе эксперимента информации является **задача оптимизации, т.е. нахождения такой комбинации влияющих независимых переменных, при которой выбранный показатель оптимальности принимает экстремальное значение.**
- Опыт — это отдельная экспериментальная часть.
- План эксперимента — **совокупность данных определяющих число, условия и порядок проведения опытов.**



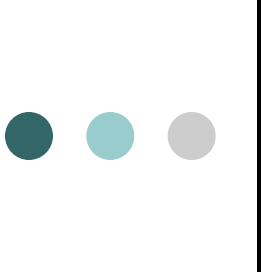
# Основные определения

- **Планирование эксперимента – это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.**
- **Задачи, для решения которых может использоваться планирование эксперимента, чрезвычайно разнообразны (выбор оптимального компонента смесей, повышение производительности действующих установок, повышение качества продукции и т.д.).**
- **Цель планирования эксперимента –при нахождение таких условий и правил проведения опытов которых удастся получить надежную и достоверную информацию об объекте с наименьшей затратой труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности.**

- При планировании эксперимента исследуемый объект представляется «черным ящиком», на который воздействуют факторы  $x$ .



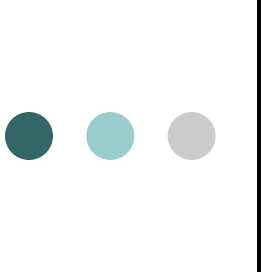
- Стрелки справа изображают численные характеристики целей исследования (параметры оптимизации). Для проведения эксперимента **необходимо иметь возможность воздействовать на поведение «черного ящика»**. Все способы такого воздействия мы называем факторами. Их также называют входами «черного ящика».



**Факторы должны быть совместимыми и независимыми. Совместимость предполагает допустимость любой комбинации факторов, а независимость - отсутствие между факторами корреляционной связи.**

К исследуемым параметрам предъявляют ряд требований. Они должны быть:

- ▣ **Управляемыми:** экспериментатор, выбрав нужное значение фактора, может его поддерживать постоянным в течение всего опыта;
- ▣ **Операциональными:** необходимо указывать последовательность действий, с помощью которых устанавливаются конкретные значения;
- ▣ **Точными:** степень точности определяется диапазоном изменения факторов;
- ▣ **Однозначными:** должны быть непосредственными воздействиями на объект.



# Принятие решений перед планированием эксперимента

- При выборе области эксперимента прежде всего надо **оценить границы областей определения факторов**. При этом должны учитываться ограничения нескольких типов. Первый тип — **принципиальные ограничения для значений факторов, которые не могут быть нарушены ни при каких обстоятельствах**. Второй тип — **ограничения, связанные с технико-экономическими соображениями**. Третий тип — **определяется конкретными условиями проведения процесса**.
- **Оптимизация обычно начинается в условиях, когда объект уже подвергался некоторым исследованиям** информацию, содержащуюся в результатах предыдущих исследований называют априорной (т.е. полученной до начала эксперимента).
- Выбор экспериментальной области факторного



# Выбор основного уровня

- **Наилучшими условиями, определенными из анализа априорной информации, соответствует комбинация уровней факторов. Каждая комбинация является многомерной точкой в факторном пространстве. Ее можно рассматривать как исходную точку для построения плана эксперимента. Ее называют основным (нулевым) уровнем.**
- **Построение плана эксперимента сводится к выбору экспериментальных точек, симметричных относительно нулевого уровня.**



# Выбор интервалов варьирования

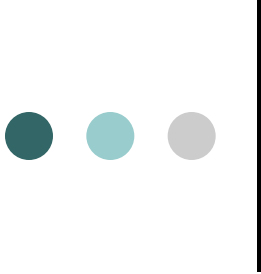
- Требуется исследовать влияние легирующих элементов (Cr – хрома, Nb – ниобия, W – вольфрама) на предел прочности литейного сплава ЖС6К. Номинальное содержание элементов: Cr=8,0%, Nb=1%, W=7,0%.
- Поставим ПФЭ при трех сериях опытов в точках: Cr=8,0±1,5%, Nb=1±1%, W=7,0±1,5%. Для стандартизации масштабов факторов условия проведения опыта сведем в таблицу

Характеристика плана	$z_{1(Nb)}$	$z_{2(W)}$	$z_{3(Cr)}$
Нулевой уровень	1%	7,0%	8,0%
Интервал варьирования	1%	1,5%	1,5%
Верхний уровень	2%	8,5%	9,5%
Нижний уровень	0%	5,5%	6,5%

План проведения экспериментов записывается в виде матрицы планирования, в которой в определенном порядке перечисляются различные комбинации факторов на двух уровнях. Например, в таблице приведена матрица планирования ПФЭ 2 для трех факторов:  $x_1$ ,  $x_2$ ,  $x_3$ . Знак «+» говорит о том, что во время опыта значение фактора устанавливают на верхнем уровне, а знак «-» показывает, что значение фактора устанавливают на нижнем уровне.

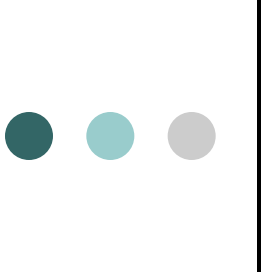
№ эксперимента	Факторы			Взаимодействия				Результаты опытов		
	$x_{1(Nb)}$	$x_{2(W)}$	$x_{3(Cr)}$	$x_1x_2$	$x_1x_3$	$x_2x_3$	$x_1x_2x_3$	$y_1$	$y_2$	$y_3$
1	+	+	+	+	+	+	+	511	555	545
2	-	+	+	-	-	+	-	429	542	448
3	+	-	+	-	+	-	-	460	408	440
4	-	-	+	+	-	-	+	394	430	370
5	+	+	-	+	-	-	-	722	646	678
6	-	+	-	-	+	-	+	603	600	606
7	+	-	-	-	-	+	+	595	588	605
8	-	-	-	+	+	+	-	476	522	478

- 
- При проведении экспериментов получают значения исследуемой величины  $y$  для каждого опыта (или серии опытов). Затем переходят к построению математической модели.
  - Под моделью понимается вид функции  $y = f(x_1, x_2, \dots, x_k)$ , которая связывает изучаемый параметр со значениями факторов, лежащих в интервале между верхним и нижним уровнями. **Эту функцию называют уравнением регрессии.** По накопленному разным исследователям опыту работы с различными моделями можно считать, что самыми простыми моделями являются алгебраические полиномы.



# Работу выполняем в следующем порядке:

- кодируем переменные;
- достраиваем матрицу планирования в кодированных переменных с учетом парных взаимодействий и дополняем столбцом средних значений отклика;
- вычисляем коэффициенты уравнения регрессии;
- проверяем вычисленные коэффициенты на значимость, предварительно определив дисперсию воспроизводимости, и получаем уравнение регрессии в кодированных переменных
- проверяем полученное уравнение на адекватность;
- проводим интерпретацию полученной модели;
- выписываем уравнение регрессии в натуральных переменных
- оптимизация параметров.

- 
- Для каждого фактора находим центр, интервал варьирования и зависимость кодированной переменной  $x_i$  от натуральной  $z_i$  по формулам. Оформляем результаты в таблице

Характеристика плана	$z_{1(Nb)}$	$z_{2(W)}$	$z_{3(Cr)}$
Нулевой уровень	1%	7,0%	8,0%
Интервал варьирования	1%	1,5%	1,5%
Верхний уровень	2%	8,5%	9,5%
Нижний уровень	0%	5,5%	6,5%
Зависимость кодированной переменной от натуральной	$x_1 = \frac{z_1 - 1}{1}$	$x_2 = \frac{z_2 - 7}{1,5}$	$x_3 = \frac{z_3 - 8}{1,5}$

- Рассчитываем средние выборочные результатов для каждого эксперимента. Строим матрицу планирования с учетом всех взаимодействий и средних значений отклика

№ эксперимен та	Факторы			Взаимодействия				Результаты опытов			
	$x_{1(Nb)}$	$x_{2(W)}$	$x_{3(Cr)}$	$x_1x_2$	$x_1x_3$	$x_2x_3$	$x_1x_2x_3$	$y_1$	$y_2$	$y_3$	$уср$
1	+	+	+	+	+	+	+	511	555	545	537
2	-	+	+	-	-	+	-	429	542	448	473
3	+	-	+	-	+	-	-	460	408	440	436
4	-	-	+	+	-	-	+	394	430	370	398
5	+	+	-	+	-	-	-	722	646	678	682
6	-	+	-	-	+	-	+	603	600	606	603
7	+	-	-	-	-	+	+	595	588	605	596
8	-	-	-	+	+	+	-	473	520	480	491

- 
- Линейное уравнение регрессии относительно новых переменных имеет вид:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

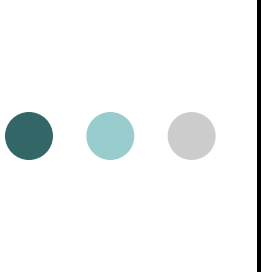
- Если требуется изучить влияние парных взаимодействий различных факторов на исследуемый параметр, то уравнение регрессии записывают в виде

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + \dots + b_{k-1}x_{k-1}x_k$$

- Или

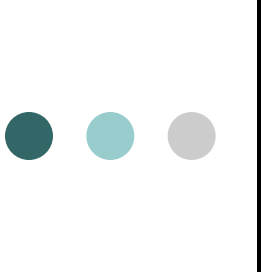
$$y = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i < j} b_{ij} x_i x_j$$

- Если надо учесть другие взаимодействия, то число слагаемых увеличивают.

- 
- Обычно проводят несколько серий опытов для каждого эксперимента. Это необходимо для проверки уравнения на адекватность.
  - Адекватность - это способность модели предсказывать результаты эксперимента в некоторой области с требуемой точностью. Результаты опытов в каждом  $j$ -ом эксперименте ( $j=1, \dots, n$ ) записывают в правые столбцы матрицы планирования. В последнем столбце записывают средние выборочные значения полученных результатов для каждой серии опытов (см. таблицу 2). Если каждый эксперимент повторяли  $m$  раз, то в матрице будет записано  $m$  столбцов  $y_1, y_2, \dots, y_m$ .
  - Если обозначить за  $y_{ji}$  значение результата, полученного в  $i$ -ом опыте ( $i=1, \dots, m$ ) для  $j$ -ого эксперимента ( $j=1, \dots, n$ ), то выборочное среднее для каждого эксперимента вычисляют по известной формуле:

$$\bar{y}_j = \frac{1}{m} \sum_{i=1}^m y_{ji}$$

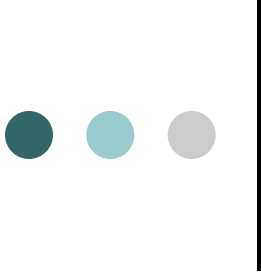


- 
- Коэффициенты уравнения регрессии находят с помощью метода наименьших квадратов.
  - Так как матрица планирования ПФЭ  $2k$  должна удовлетворять определенным требованиям (такие матрицы с заданными требованиями уже построены), то формулы, определяющие коэффициенты уравнения регрессии, достаточно просты:

$$b_0 = \frac{1}{n} \sum_{j=1}^n y_{ji}$$

$$b_i = \frac{1}{n} \sum_{j=1}^n x_{ji} \bar{y}_j, \quad i = \overline{1, k},$$

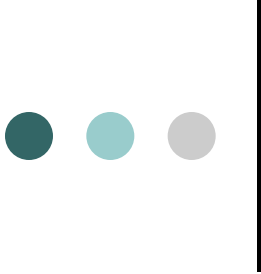
$$b_{r,p} = \frac{1}{n} \sum_{j=1}^n x_{jr} x_{jp} \bar{y}_j, \quad r < p, \quad r = \overline{1, k}, p = \overline{1, k},$$

- 
- Вычисляем коэффициенты уравнения регрессии. Составляем для наглядности таблицу, в которую заносим найденные коэффициенты уравнения регрессии.

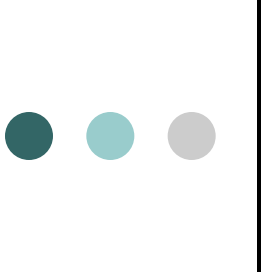
Свойства сплава	$b_0$	$b_1$	$b_2$	$b_3$	$b_{1,2}$	$b_{1,3}$	$b_{2,3}$	$b_{1,2,3}$
Предел прочности	527	35,75	46,75	-66	0	-10,25	-2,75	6,5

Записываем линейное уравнение регрессии относительно новых переменных:

$$y = 527 + 35,75x_1 + 46,75x_2 - 66x_3 - 10,25x_1x_3 - 2,75x_2x_3 + 6,5x_1x_2x_3$$

- 
- Полученные коэффициенты необходимо проверить на значимость. Это можно сделать с помощью критерия Стьюдента: если  $|b| > t_{кр} \cdot S_{коэф}$ , то  $b$  значим; если  $|b| < t_{кр} \cdot S_{коэф}$ , то  $b$  незначим и его полагают равным нулю в уравнении регрессии.
  - Критическую точку  $t_{кр}$  находят из таблиц распределения Стьюдента по числу степеней свободы  $n(m - 1)$  и с заданным уровнем значимости  $\alpha$  для случая двусторонней критической области.
  - Среднее квадратическое отклонение коэффициентов  $S_{коэф}$  зависит от дисперсии воспроизводимости результатов по всем проведенным опытам  $S^2\{y\}$  и вычисляется по формуле:

$$S_{коэф} = \sqrt{\frac{S^2\{y\}}{n \cdot m}}$$

- 
- Дисперсия воспроизводимости  $S^2_{\{y\}}$  характеризует ошибку всего эксперимента. В случае равномерного дублирования опытов (т.е. при одинаковом числе наблюдений в каждом эксперименте) для расчета  $S^2_{\{y\}}$  используют формулу:

$$S^2_{\{y\}} = \frac{1}{n \cdot (m - 1)} \sum_{j=1}^n \sum_{i=1}^m \left( y_{ji} - \bar{y}_j \right)^2,$$

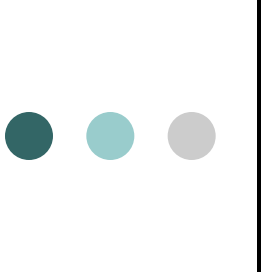
- где  $n$  - число экспериментов (число строк в матрице ПФЭ);
- $m$  - число опытов (наблюдений) в каждом эксперименте;
- $y_{ji}$  - результат отдельного  $i$ -го наблюдения в  $j$ -ом эксперименте;
- $\bar{y}_j$  - среднее выборочное значение наблюдений для  $j$ -ого эксперимента, которое определяется по формуле.

- Находим дисперсию воспроизводимости  $S^2_{\{y\}}$ . Для облегчения расчетов запишем формулу в другом виде

$$S^2_{\{y\}} = \frac{1}{n \cdot (m-1)} \sum_{j=1}^n \sum_{i=1}^m (y_{ji} - \bar{y}_j)^2 = \frac{1}{n} \sum_{j=1}^n \left( \frac{1}{m-1} \sum_{i=1}^m (y_{ji} - \bar{y}_j)^2 \right) = \frac{1}{n} \sum_{j=1}^n S_j^2,$$

- здесь внутренние суммы являются выборочными дисперсиями результатов опытов для  $j$ -го эксперимента ( $j=1, \dots, n$ ). Для удобства оформляем расчеты в виде таблицы

$j$	$y_1$	$y_2$	$y_3$	$\bar{y}_j$	$(y_{j1} - \bar{y}_j)^2$	$(y_{j2} - \bar{y}_j)^2$	$(y_{j3} - \bar{y}_j)^2$	$S_j^2$
1	511	555	545	537	676	324	64	532
2	429	542	448	473	1936	4761	625	3661
3	460	408	440	436	576	784	16	688
4	394	430	370	398	16	1024	784	912
5	722	646	678	682	1600	1296	16	1456
6	603	600	606	603	0	9	9	9
7	595	588	605	596	1	64	81	73
8	473	520	480	491	324	841	121	643

- 
- Суммируя элементы последнего столбца таблицы 2.5, получаем:

$$\sum_{j=1}^8 S_j^2 = 7974$$

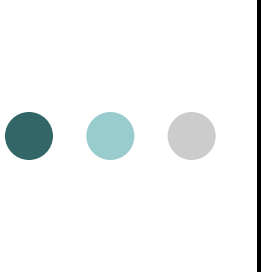
- Отсюда получаем дисперсию воспроизводимости:

$$S_{\{y\}}^2 = \frac{1}{8} \sum_{j=1}^8 S_j^2 = \frac{1}{8} S_j^2 = \frac{1}{8} \cdot 7974 = 996,75$$

- Определяем среднее квадратическое отклонение коэффициентов:

$$S_{\text{коэф}} = \sqrt{\frac{S_{\{y\}}^2}{n \cdot m}} = \sqrt{\frac{996,75}{8 \cdot 3}} = 6,45$$

- Из таблиц распределения Стьюдента по числу степеней свободы  $n(m-1)=8*2=16$  при уровне значимости  $\alpha=0,05$  находим  $t_{кр}=2,12$ . Следовательно,  $t_{кр} * S_{\text{коэф}} = 2,12 * 6,45 = 13,67$

- 
- Сравнивая полученное значение  $t_{кр} * S_{коэф} = 13,67$  с коэффициентами уравнения регрессии, представленными в таблице, видим, что все коэффициенты взаимодействия меньше по абсолютной величине 13,67. Следовательно, коэффициенты взаимодействия незначимы. Получаем уравнение регрессии в кодированных переменных:

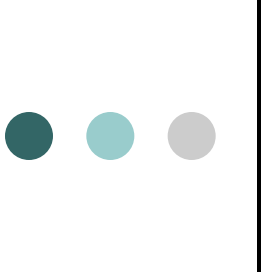
- $y = 527 + 35,75x_1 + 46,75x_2 - 66x_3$

- Проверка на адекватность полученного уравнения регрессии со значимыми коэффициентами осуществляется с помощью критерия Фишера: если  $F_{расч.} < F_{табл}$ , то уравнение адекватно, в противном случае - неадекватно.

- Расчетное значение критерия  $F_{расч}$  определяют по формуле:

$$F_{расч} = \frac{S_{ост}^2}{S_{\{y\}}^2}$$

- где  $S_{\{y\}}^2$  - дисперсия воспроизводимости, найденная по формуле, а  $S_{ост}^2$  - остаточная дисперсия (или дисперсия адекватности).



. Проверим полученное уравнение на адекватность по критерию Фишера. Так как дисперсия воспроизводимости найдена в предыдущем пункте, то для определения расчетного значения критерия  $F_{расч}$  необходимо вычислить остаточную дисперсию  $S^2_{ост}$ .

Для этого найдем значения изучаемого параметра по полученному уравнению регрессии  $y_j$  ( $j=1, \dots, 8$ ), подставляя +1 или -1 вместо  $x_i$  в соответствии с номером  $j$  эксперимента

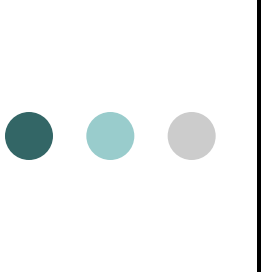
Остаточную дисперсию  $S^2_{ост}$  вычисляем по формуле:

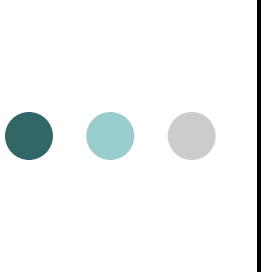
$$S^2_{ост} = \frac{3}{8-7} \sum_{j=1}^8 (\tilde{y}_j - \bar{y}_j)^2 = 3 \cdot [(543,5 - 537)^2 + (472 - 473)^2 + (450 - 436)^2 + (378,5 - 398)^2 \dots] = 3717$$

Расчетное значение критерия Фишера  $F_{расч}$  определяем по формуле:

$$F_{расч} = \frac{S^2_{ост}}{S^2_{\{y\}}} = \frac{3717}{996,75} = 3,73$$



- 
- Табличное значение критерия  $F_{табл.}$  находим из таблиц критических точек распределения Фишера при уровне значимости  $\alpha=0,05$  по соответствующим степеням свободы  $k1 = n - r = 8 - 7 = 1$  и  $k2 = n(m-1) = 8-2 = 16$ ,  $F_{табл}=4,49$ .
  - Так как  $F_{расч}=3,73 < F_{таблj}= 4,49$ , то уравнение регрессии адекватно.
  - Проведем интерпретацию полученной модели
$$y = 527 + 35,75x_1 + 46,75x_2 - 66x_3$$
  - По уравнению видно, что наиболее сильное влияние оказывает фактор  $x_3$  – содержание хрома в сплаве, так как он имеет наибольший по абсолютной величине коэффициент.
  - Так как коэффициенты при  $x_1$  и  $x_2$  положительны, то с увеличением этих факторов увеличивается отклик, т. е. увеличивается прочность.

- 
- Выписываем уравнение регрессии в натуральных переменных, подставляя вместо  $x_i$  их выражения через  $z_i$ , которые берем из последнего столбца таблицы:

$$y = 527 + 35,75 + 46,75 - 66$$

Преобразовав это уравнение, окончательно получаем его вид в натуральных переменных:

$$y = 528,375 + 35,75z_1 + 70,125z_2 - 99z_3.$$



# Оптимизация параметров

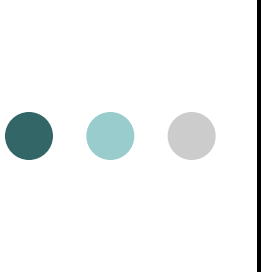
*Оптимизация* – процесс поиска максимума или минимума (поиск наилучшего значения параметра).

Оптимизация бывает двух типов:

- 1) оптимизация параметров, в процессе которой ищут такие значения параметров, при которых целевая функция имеет экстремальное значение при заданной структуре;
- 2) оптимизация структуры, когда ищется структура системы, при которой целевая функция имеет максимальное значение (функциональное преобразование при заданных параметрах)

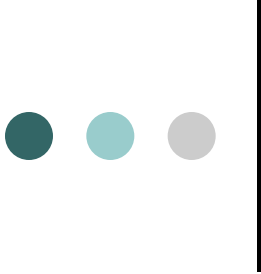
Сначала выбирается начальное значение  $x_1, x_2$  и  $x_3$ , затем интервалы варьирования и , составляется МПЭ.

Наименование величин	$z_{1(Nb)}$	$z_{2(W)}$	$z_{3(Cr)}$
$z_{0j}$	1%	7,0%	8,0%
$\Delta_j$	1%	1,5%	1,5%
$b_j$	35,75	70,125	99
$\Delta_j b_j$	35,75	105,1875	148,5
$e \Delta_j b_j$	1	1,5	1,5



№ эксперимента	Координаты точек крутого восхождения		
	$x_1$	$x_2$	$x_3$
9	2	8,5	6,5
10	3	10	5
11	4	11,5	3,5
12	5	13	2
13	6	14,5	0,5

Таблицу опыта обычно рассчитывают до наступления нереализуемого шага. В данном случае следующий шаг дает отрицательное значение  $z_3$ .

- 
- Проведем расчет для новых опытов крутого восхождения. Необходимо в первую очередь перевести натуральные значения таблицы в кодированные. Кодированные величины получаются с помощью известной формулы:

$$x_i = \frac{x_i - x_{io}}{I_i}$$

№ эксперимента	Координаты точек крутого восхождения			y
	$x_1$	$x_2$	$x_3$	
9	1	1	-1	675,5
10	2	2	-2	824
11	3	3	-3	972,5
12	4	4	-4	1121
13	5	5	-5	1269,5

- Расчеты (мысленные эксперименты) используются только для сокращения объема эксперимента. Параллельно с мысленными опытами проводится эксперимент, но он проводится не в каждой точке, и производится сравнение  $u_{расч}$  и  $u_{эксп}$ . Когда это расхождение становится значительным, то переходят к проведению эксперимента.

