

## Введение в теорию факторного планирования эксперимента

Если необходимо изучить влияние, например, количества углерода  $X$  на прочность стали  $Y$  проводят однофакторный эксперимент.

И чем больше различных значений примет  $X$ , тем более полно мы узнаем изучаемую зависимость  $Y(X)$ .

Допустим, что исследуем влияние на прочность стали  $Y$  количества углерода  $X_1$  и количества хрома  $X_2$ . Последовательно проведя 2 серии однофакторных экспериментов получим всего лишь 2 линии на двумерном экспериментальном поле – основная область возможных сочетаний факторов останется неисследованной

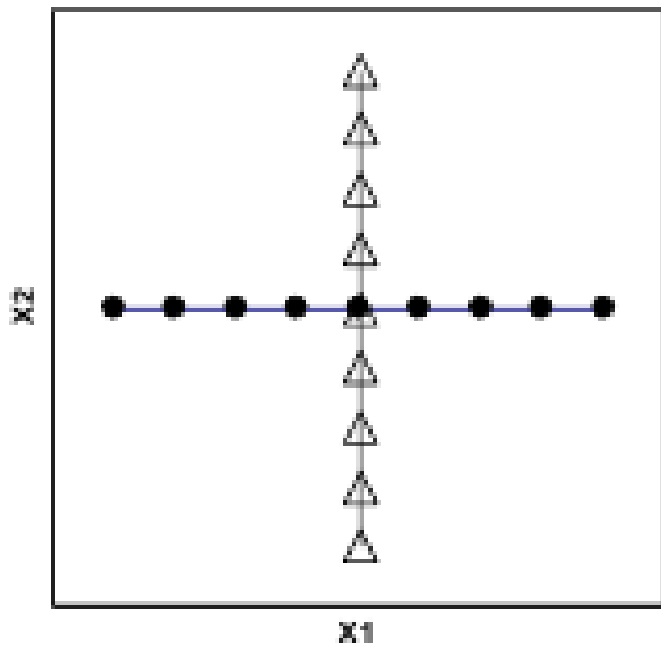


Схема эксперимента «крест»

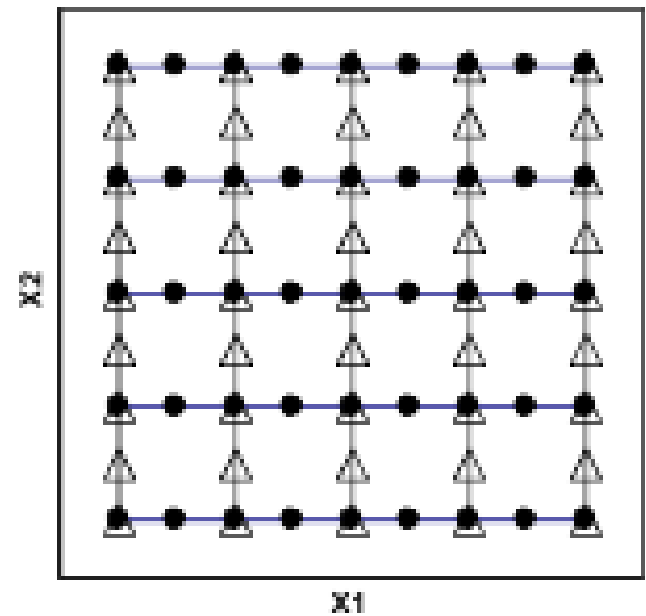


Схема эксперимента «решетка»

Попытка «заштриховать» всё поле эксперимента экспериментальными линиями (рис.) приведет к недопустимо высоким затратам по времени и по средствам. На рисунках треугольниками обозначены серии с варьированием  $X_2$  при постоянном  $X_1$ , точками – серии с варьированием  $X_1$  при постоянном  $X_2$ .

Решение – провести отдельные эксперименты в точках, расположенных на границах, в углах и в центре исследуемой области. Это пример факторного планирования эксперимента

Факторное планирование эксперимента имеет цель: за минимальное количество экспериментов описать исследуемую область с достаточной для экспериментатора точностью.

Факторный эксперимент – мощное средство эмпирического изучения процессов, обеспечивающее точное математическое описание отклика системы при минимальном количестве экспериментов.

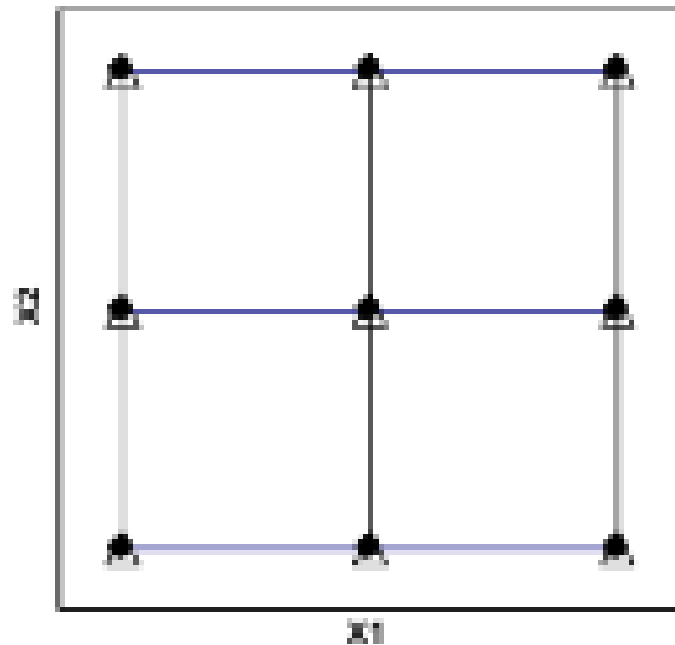


Схема полного факторного эксперимента (ПФЭ)  $3^2$

Не приводя строгих определений терминов, связанных с факторным планированием, опишем их упрощенно.

Фактор,  $X$  – величина, которую экспериментатор меняет (варьирует).

Отклик  $Y$  – величина, которую экспериментатор измеряет.

Факторное пространство – служит для мысленного расположения в нем экспериментальных точек. Количество измерений равно количеству факторов.

План полного факторного эксперимента ПФЭ обозначается  $m^k$  где  $m$  – число уровней варьирования факторов,  $k$  – число факторов.

Например, если 3 фактора варьируются на 2-х уровнях, то план ПФЭ обозначится  $2^3$  и будет состоять из 8 опытов на различных сочетаниях факторов. Очевидно, что план  $3^2$  состоит из 9 опытов. Планы  $2^k$  называют планами первого порядка, планы  $3^k$  – планы второго порядка.

Планы больших порядков используют редко – для повышения точности выгоднее повторить эксперимент, сузив диапазоны варьирования.

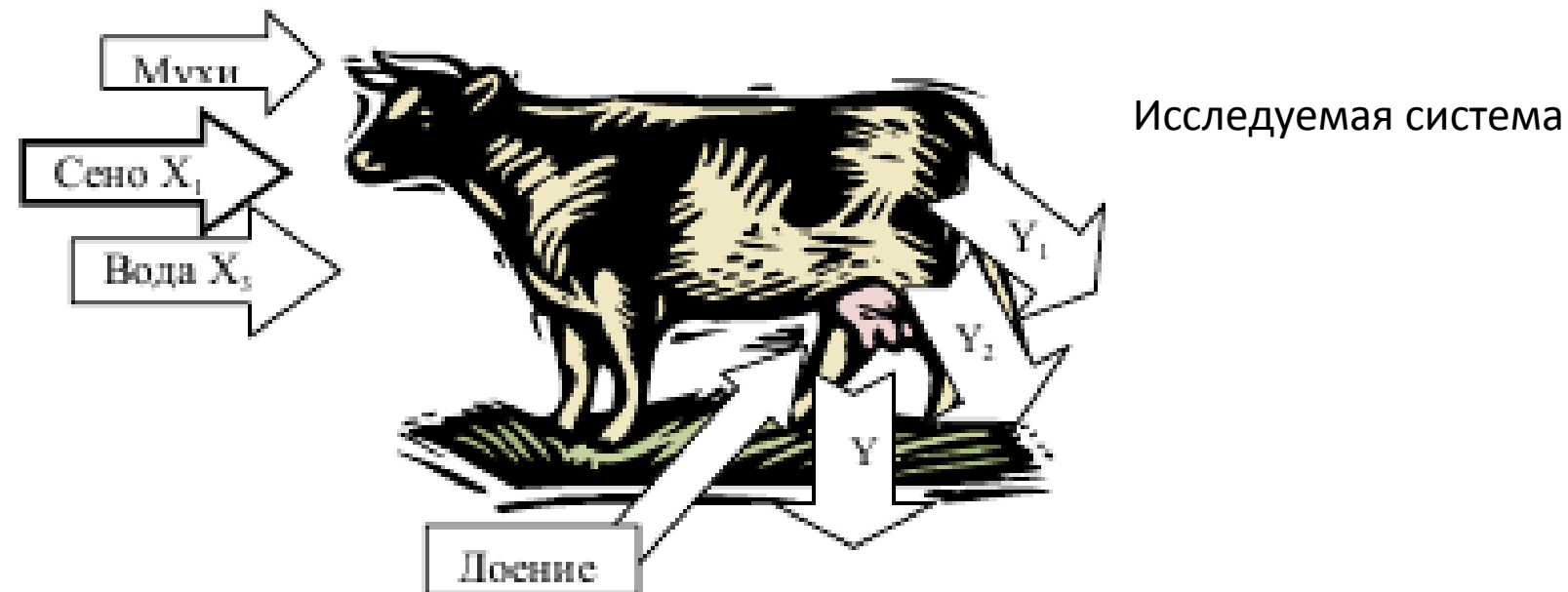
## Пример

Гипотеза: Чтобы корова меньше ела и давала больше молока – ее надо меньше кормить и чаще доить.

### 1. Постановка задачи.

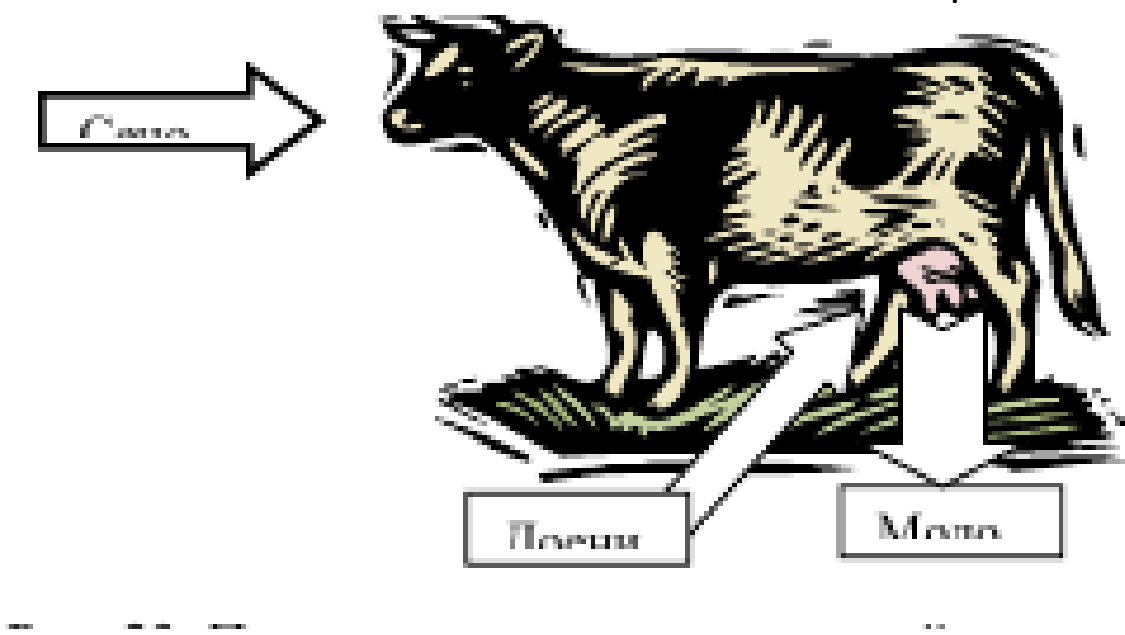
Корова (рис.) представляет собой систему. Система имеет на входе контролируемые воздействия (варьируемые факторы)  $X_1$  и  $X_2$  и неконтролируемые воздействия (случайные факторы), например  $X_3$ ,  $X_4$ , ... .

Случайные факторы не учитываем – полагаем систему детерминированной. Из выходных характеристик системы,  $Y$ ,  $Y_1$ ,  $Y_2$ , ... в соответствии с целями исследования для контроля выбираем  $Y$



В рамках принятой модели (рис. исследуем зависимость количества молока в сутки  $Y$  от количества корма  $X_1$  и числа доений  $X_2$ .

Принятая модель исследуемой системы



## Планирование и обработка результатов эксперимента

План эксперимента  $2^2$ .

Технически возможные пределы изменения факторов. Количество корма от 0 до 100 кг. Количество доений от 1 до 10.

Пределы варьирования факторов не должны превышать технически возможных и выбираются на усмотрение экспериментатора.

С учетом гуманного отношения к животным принимаем пределы варьирования:  $X_1 = 10 \dots 70$  кг,  $X_2 = 2 \dots 5$  шт. (табл. 1).

В планах первого порядка два уровня варьирования факторов, верхний, обозначаемый «+» или «1» и нижний, обозначаемый «-» или «-1». При этом от натуральных значений факторов (X) переходят к кодированным (x) и оформляют в виде таблицы.

Следующая таблица (табл. 2) – матрица эксперимента – состоит из уровней варьирования факторов, взаимодействий и отклика.

Столбцы взаимодействий получаются перемножением соответствующих кодированных значений факторов.

# Уровни и интервалы варьирования факторов

Факторы	д. изм.	Кодовое обозначение	Интервал варьирования	Натуральные уровни соответствующие кодированным	
				1	-1
$X_1$ – количество корма	г	$x_1$	60	70	10
$X_2$ – число досний	т.	$x_2$	3	5	2

## Матрица эксперимента

№ опыта	$x_1$	$x_2$	$x_1 x_2$	Y
1	1	1	1	12
2	-1	1	-1	2
3	1	-1	-1	17
4	-1	-1	1	6

Для описания результатов планов первого порядка используют полиномы первого порядка, в данном случае:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 .$$

Коэффициенты при кодированных факторах дают информацию о влиянии факторов или из сочетаний на отклик.



В данном случае применив надстройку Excel «Регрессия» получаем:

Вывод итогов

Регрессионная статистика	
Множественный R	1
R-квадрат	1
Нормированный R-квадрат	65535
Стандартная ошибка	0
Наблюдения	4

$a_0$	$a_1$	$a_2$	$a_{12}$
9,24999	5,24999	-2,25	-0,25
7	9		

Это говорит о том, что повышение количества корма сильно влияет на количество молока, а вот повышения частоты доений – немного снижает количество молока, совместное влияние факторов не выражено. Таким образом – чтобы было больше молока, корову надо больше кормить и реже доить.

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	3	130,75	43,58333	#ЧИСЛО!	#ЧИСЛО!
Остаток	0	0	65535		
Итого	3	130,75			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	9,25	0	65535	#ЧИСЛО!	9,25	9,25	9,25	9,25
Переменная X 1	5,25	0	65535	#ЧИСЛО!	5,25	5,25	5,25	5,25
Переменная X 2	-2,25	0	65535	#ЧИСЛО!	-2,25	-2,25	-2,25	-2,25
Переменная X 3	-0,25	0	65535	#ЧИСЛО!	-0,25	-0,25	-0,25	-0,25

## Анализ полученных результатов

Вывод несколько противоречит сложившимся представлениям – если корову совсем не доить то, скорее всего, молока не будет, да и бесконечно увеличивать кормление тоже нецелесообразно.

Предполагаем, что существует оптимальное количество корма и числа доений, соответствующее максимуму молока.

Проведем уточняющий эксперимент 32 сузив диапазоны варьирования и перейдя в предполагаемую оптимальную область.

Поместим центр плана в точку с  $x_1 = 1$  ( $X_1 = 70$  кг) и  $x_2 = -1$  ( $X_2 = 2$  шт.).

Уровни и интервалы варьирования факторов

Факторы	д. изм.	Кодовое обозначение	Интервал варьирования	Натуральные уровни соответствующие кодированным		
				1		-1
$X_1$ – количество корма	г	$x_1$	20	9 0	0	50
$X_2$ – число доений	т.	$x_2$	1	3		1

Кодированные значения факторов  $x$  связаны с натуральными  $X$  через диапазон варьирования  $e$  и натуральное значение фактора в центре плана  $X_0$ :

$$x = \frac{X - X_0}{e}$$

В факторном планировании часто при переходе от одного плана к другому стараются использовать данные предыдущего плана – в данном случае опыт 5 плана второго порядка можно не проводить, т.к. он соответствует опыту 3 плана второго порядка.

Для описания результатов плана второго порядка применяют полиномы второй степени, в данном случае:

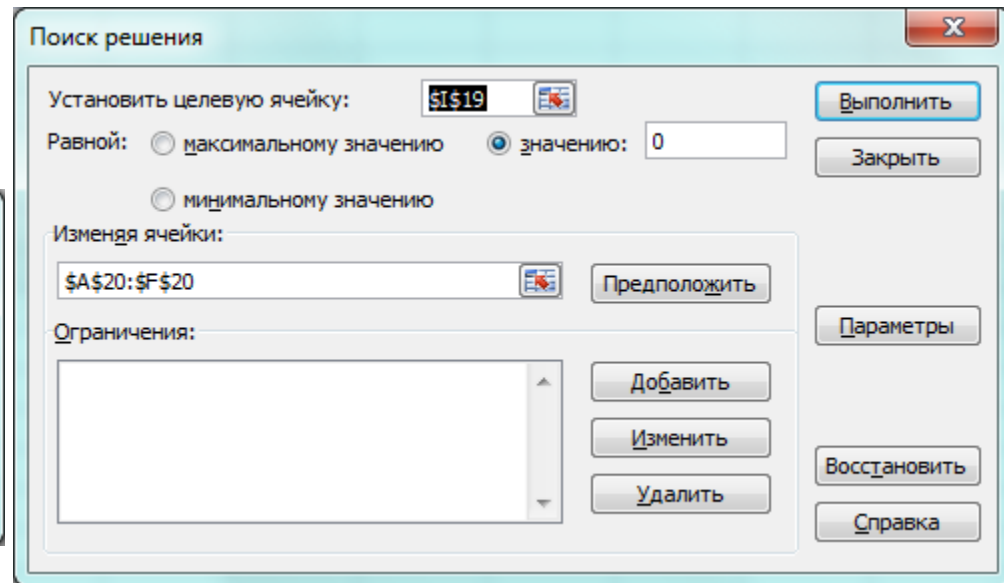
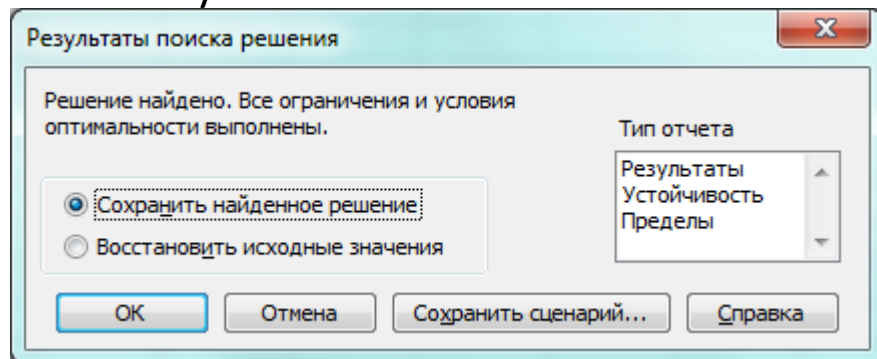
$$Y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_{12} + a_{22}x_{22}$$

	x1	x2	Y
1	1	1	11
2	0	1	13
3	-1	1	14
4	1	0	15
5	0	0	17
6	-1	0	18
7	1	-1	9
8	0	-1	11
9	-1	-1	12

Построим в Excel таблицу на основе таблицы 13. Справа от столбца Y с экспериментальными данными располагаем столбец Yp где с помощью формул строим выражение  $\sum (Y - Y_p)^2$  положив в качестве начальных значений всех коэффициентов ноль (для этого нужно для каждого коэффициента выбрать ячейку в Excel и записать в нее «0» ).

		x1	x2	x1x2	x1^2	x2^2	Y	Yp	(Y-Yp)^2
	1	1	1	1	1	1	11	0	121
	2	0	1	0	0	1	13	0	169
	3	-1	1	-1	1	1	14	0	196
	4	1	0	0	1	0	15	0	225
	5	0	0	0	0	0	17	0	289
	6	-1	0	0	1	0	18	0	324
	7	1	-1	-1	1	1	9	0	81
	8	0	-1	0	0	1	11	0	121
	9	-1	-1	1	1	1	12	0	144
a0	a1	a2	a12	a11	a22	1670			
0	0	0	0	0	0				

Применив надстройку «Поиск решения», положив значение целевой ячейки «0» (см. рис. ), получаем:



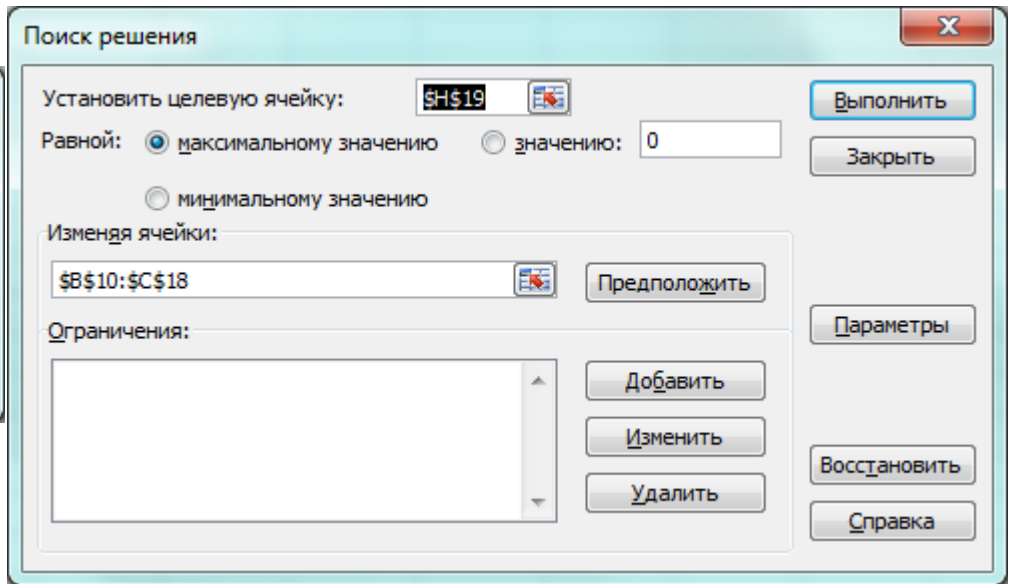
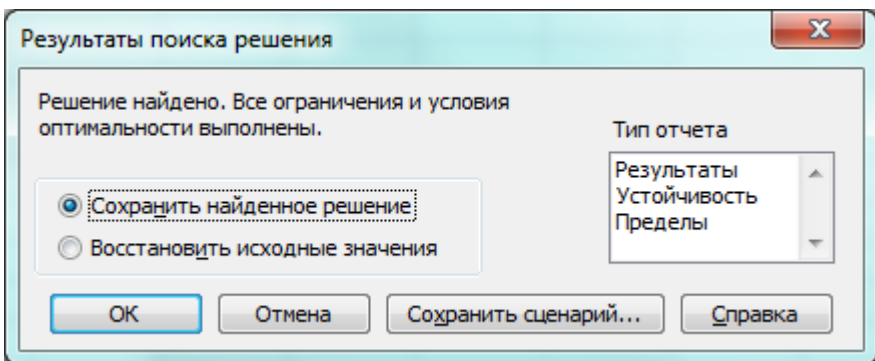
	x1	x2	x1x2	x1^2	x2^2	Y	Yp	(Y-Yp)^2
1	1	1	1	1	1	11	11,00001	6,92E-11
2	0	1	0	0	1	13	13	4,82E-14
3	-1	1	-1	1	1	14	14	2,31E-12
4	1	0	0	1	0	15	14,99999	4,99E-11
5	0	0	0	0	0	17	16,99998	2,43E-10
6	-1	0	0	1	0	18	17,99998	2,85E-10
7	1	-1	-1	1	1	9	9,000003	1,21E-11
8	0	-1	0	0	1	11	10,99999	2,53E-11
9	-1	-1	1	1	1	12	11,99999	3,97E-11
a0	a1	a2	a12	a11	a22			7,26E-10
16,99998	-1,5	1,000002	1,53E-08	-0,5	-4,99999			

Разные знаки при квадратичных и линейных коэффициентах указывают, что возможно, оптимум лежит внутри исследованной области.

Найти искомый оптимум можно с использованием надстройки «Поиск решения» положив в качестве целевой функции с найденными коэффициентами и изменяя ячейки x1 и x2.

Результат:

$$Y=a_0+a_1x_1+a_2x_2+a_{12}x_1x_2+a_{11}x_{12}+a_{22}x_{22}$$



	x1	x2	x1x2	x1^2	x2^2	Y	Yp
1	-1,50001	0,100001	-0,15	2,250018	0,01	11	18,17499
2	-1,50001	0,100001	-0,15	2,250018	0,01	13	18,17499
3	-1,50001	0,100001	-0,15	2,250019	0,01	14	18,17499
4	-1,50001	0,1	-0,15	2,250018	0,01	15	18,17499
5	-1,50001	0,1	-0,15	2,250018	0,01	17	18,17499
6	-1,50001	0,1	-0,15	2,250019	0,01	18	18,17499
7	-1,50001	0,1	-0,15	2,250018	0,01	9	18,17499
8	-1,50001	0,1	-0,15	2,250018	0,01	11	18,17499
9	-1,50001	0,1	-0,15	2,250019	0,01	12	18,17499
a0	a1	a2	a12	a11	a22		163,5749
16,99998	-1,5	1,000002	1,53E-08	-0,5	-4,99999		

Перейдя от кодированных значений к натуральным по получаем, что максимальный суточный надой 18,2 л возможен при 2-х разовом доении и кормлении в объеме 40 кг в день.

Оптимум по сену лежит за пределами диапазона варьирования и, строго говоря, нуждается в дополнительной экспериментальной проверке.



Вопросы статистической обработки при планировании и обработке результатов факторного эксперимента в данном примере не рассмотрены – для лучшего понимания основных принципов факторного планирования.

Требуемый результат исследования – оптимальное сочетание факторов – достигнут за 12 опытов

## Контрольные вопросы

1. В чем цель корреляционного анализа?
1. Что такое коэффициент корреляции?
2. Для чего используется t-статистика Стьюдента?
3. Какими способами можно определить коэффициент корреляции в MS Excel?
4. В чем цель регрессионного анализа?
5. Опишите уравнение линейной регрессии.
6. Какими способами можно найти модель регрессии в MS Excel? Коротко опишите эти способы.
7. В чем задача прогнозирования данных?
8. Какими способами осуществить прогнозирование в MS Excel?
9. Что обозначает план эксперимента 34?
10. Как подключить надстройку «Поиск решения»?
11. Для чего выполняют кодирование переменных при планировании и обработке результатов эксперимента?
12. Что такое «целевая ячейка»?
13. Для чего используются относительные, абсолютные и смешанные ссылки в формулах?
14. Полный факторный план какого порядка целесообразно применить при 8 факторном эксперименте?
15. Чем отличаются уравнения регрессии в описании планов первого и второго порядка?
16. Для чего выполняется переход от натуральных размерных значений факторов к кодированным безразмерным?

Порядок выполнения задания

- 1. Перед выполнением задания\_1 изучить теоретическую часть практикума и ответить на контрольные вопросы.
- 2. Открыть новую книгу Excel и сохранить под именем «Статфункции.xls».
- 3. В книге выполнить задание со следующими условиями:  
Имеются данные по двум экономическим показателям X и Y:

Цена (X)	995	983	1001	1012	1011	1017	978	997	1010	989	900	1100	5000
Спрос (Y)	122	144	114	100	100	90	150	130	95	155	?	?	?

Необходимо:

- ☐ вычислить коэффициент корреляции;
- ☐ построить корреляционное поле (диаграмму) на отдельном листе;
- ☐ построить регрессионную модель (с использованием функции ЛИНЕЙН);
- ☐ спрогнозировать значение Y для 3-х новых значений X с помощью функции ПРЕДСКАЗ.

Все действия (в том числе форматирование таблицы) необходимо выполнять, опираясь на образец.

4. На диаграмме разместить линию тренда с уравнением регрессии и оформить их как показано в образце. Дополнить диаграмму спрогнозированными данными (кроме последнего значения цены 5000).
5. Используя инструмент «Регрессия» на отдельном листе построить регрессионную модель с учетом новых спрогнозированных значений. Записать на листе уравнение регрессии на основании данных из «Вывода итогов».
6. Представить файл с выполненной работой преподавателю для проверки.
7. Перед выполнением задания\_2 изучите теоретическую часть работы (1.2) и ответьте на контрольные вопросы.
8. Создать книгу MS Excel с названием «Корова». Лист 1 озаглавить «2-2» и воспроизвести на нем пример плана первого порядка. Лист 2 озаглавить «3-2» и воспроизвести на нем пример плана второго порядка и поиск оптимальных значений. Скопировать лист «2-2» на новый лист «доить-не-кормить» и путем подбора результатов эксперимента подтвердить проверяемую гипотезу: «Чтобы корова меньше ела и давала больше молока ее надо меньше кормить и чаще доить».
9. В п.3 примера принято решение, в результате которого для достижения цели понадобилось 12 опытов. Предложить вариант решения с достижением цели за 9 опытов.

Создать книгу «Корова2» и провести в ней расчеты по новому варианту с описанием проводимых действий в отчете по работе по аналогии с примером.

10. Оформить отчет по работе, содержащий: цель работы, описание действий, выводы по работе.

Требования к оформлению, процедура защиты

Отчет по данной работе должен содержать распечатку каждого листа книги «Статфункции.xls». При защите необходимо дать требуемые пояснения к содержанию каждого листа книги, продемонстрировать выполнение работы в файле книги «Статфункции.xls» и ответить на контрольный вопрос