О. Л. Любимцева

Блочное планирование эксперимента и анализ данных

Учебное пособие

Нижний Новгород
2018

Министерство образования и науки российской федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Нижегородский государственный архитектурно-строительный университет»

О. Л. Любимцева

Блочное планирование эксперимента и анализ данных

Утверждено редакционно-издательским советом университета
в качестве учебного пособия

Нижний Новгород
ННГАСУ
2018

ББК 30.10
Л 93
УДК 519.8(075.8)

Печатается в авторской редакции

Рецензенты:

В. К. Вильданов – канд. физ-мат. наук, доцент кафедры математического моделирования экономических процессов ИЭП ННГУ им. Н. И Лобачевского

Н. Г. Чебочко – канд. физ-мат. наук, доцент кафедры алгебры, геометрии и дискретной математики ИТММ ННГУ им. Н.И. Лобачевского

Любимцева О. Л. Блочное планирование эксперимента и анализ данных [Текст]: учебн. пособие /О. Л. Любимцева; Нижегор. гос. архитектур.- строит. ун-т. – Н. Новгород: ННГАСУ, 2018. – 31с. ISBN 978-5-528-00276-7

Учебное пособие содержит теоретический материал, необходимый для овладения основами блочного планирования эксперимента и необходимым анализом данных в курсе лекций по дисциплине «Методология научных исследований». Рассмотрены теоретические вопросы и примеры решения задач.

Предназначено для студентов, обучающихся по направлению подготовки 27.03.01: «Стандартизация и метрология», профиль: «Стандартизация и сертификация».

ISBN 978-5-528-00276-7

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В курсе лекций по дисциплине «Математическая статистика» был рассмотрен факторный анализ. Простейшим, среди всех факторных планов, следует считать *однофакторный эксперимент*, который призван установить факт зависимости результата эксперимента от значения исследуемого единственного фактора, т.е. влияние других факторов предполагается несущественным. Однако во многих экспериментальных ситуациях возникает необходимость спланировать испытание так, чтобы можно было контролировать изменчивость, обусловленную посторонними источниками.

Под разбиением на блоки понимают расположение экспериментальных единиц в относительно однородных блоках таким образом, что внутри каждого блока ошибку эксперимента предполагают меньшей, чем можно было бы ожидать, если бы такое же число единиц было случайно отобрано в данную обработку. [2]

Эксперименты, в которых используются блоки, называются блочными рандомизированными планами. Хотя в таких схемах используются как обработки, так и блоки, основное внимание уделяется оценке разностей между обработками. Целью объединения условий в блоки является максимально возможное исключение изменчивости экспериментальной ошибки с тем, чтобы разности между обработками проявились как можно отчетливее. Блочные рандомизированные эксперименты оказываются более эффективными, чем полностью рандомизированные планы и, следовательно, позволяют получать более точные результаты.

Рандомизированные блочные планы наиболее часто встречаются при планировании экспериментов. Ситуации, в которых они должны применяться, весьма разнообразны, но при некотором опыте определяются достаточно легко. Экспериментальные установки или аппаратура часто обладают различными рабочими характеристиками и представляют собой типичный фактор группирования в блоки. Партии сырья, операторы, моменты времени также являются характерными источниками изменчивости в эксперименте, которые можно контролироваться систематическим образом при использовании блоков.

Рандомизированное полноблочное планирование

При изучении интересующего фактора можно пренебречь фактором, имеющим, на взгляд исследователя, не существенное влияние на первый. Однако, ошибка эксперимента будет отражать не только случайную ошибку, но и различия между уровнями второго фактора. Первостепенная задача − исключить из случайной ошибки данные различия. Это может обеспечить рандомизированный полноблочный план, где в качестве экспериментальных объектов будут выбраны уровни первого. Слово «полнота» означает, что каждый блок (уровень второго фактора) содержит все обработки. Внутри каждого блока порядок проверяемых образцов определяется случайным образом.

Рандомизированные полноблочные планы наиболее часто встречаются при планировании экспериментов. Экспериментальные установки или аппаратура часто обладают разными рабочими характеристиками и представляют собой типичный фактор группирования в блоки. Партии сырья, моменты времени так же являются характерными источниками изменчивости в эксперименте, который может контролироваться систематическим образом при использовании блоков.

<u>*Статистический анализ.*</u>

Пусть число обработок составляет a и число блоков равно b.На рис.1 приведен рандомизированный полноблочный план.

Блок 1	Блок 2	Блок b
y_{11}	y_{12}	y_{1b}
y_{21}	y_{22}	y_{2b}
y_{31}	y_{32}	y_{3b}
.	.	.
.	.	.
y_{a1}	y_{a2}	y_{ab}

Рис.1

Каждый блок содержит по одному наблюдению на каждую обработку, а порядок, в котором осуществляется выбор обработки внутри блока, определяется случайным образом. Рандомизация проводится только внутри блоков, поэтому говорят, что блоки представляют собой *ограничение на рандомизацию*.

Статистическая модель плана:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

$$i = 1,2,\ldots,a \quad j = 1, 2, \ldots, b$$

где μ - математическое ожидание общего среднего;

τ_i - эффект *i*- ой обработки;

β_j - эффект *j*- ого блока;

ε_{ij} - случайная ошибка, причем $\varepsilon_{ij} \sim \mathrm{N}\,(0, \sigma^2)$

Метод постоянных эффектов

Пусть обработки и блоки фиксированные факторы. Эффекты обработок и блоков определяются как отклонения от математического ожидания общего среднего, то есть $\sum_{i=1}^{a}\tau_i = 0$ и $\sum_{j=1}^{b}\beta_j = 0$.

Так как блоки представляют собой ограничение на рандомизацию, то рассматривается гипотеза относительно эффектов обработок:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1 : \tau_k \neq 0, \qquad \text{где} \quad 1 \le k \le a.$$

Введем следующие обозначения:

$y_{i\bullet} = \sum_{j=1}^{b} y_{ij}, \quad i = \overline{1,a}$ -сумма всех наблюдений для i обработки.

$y_{\bullet j} = \sum_{i=1}^{a} y_{ij}, \quad j = \overline{1,b}$ - сумма всех наблюдений в j блоке.

$y_{\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}, \quad \mathrm{i} = \overline{1,a}, \quad j = \overline{1,b}$.

Скорректированную сумму квадратов можно представить следующим образом:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \overline{y_{\bullet\bullet}}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\left[\left(\overline{y_{i\bullet}} - \overline{y_{\bullet\bullet}}\right) + \left(\overline{y_{\bullet j}} - \overline{y_{\bullet\bullet}}\right) + \left(y_{ij} - \overline{y_{i\bullet}} - \overline{y_{\bullet j}} + \overline{y_{\bullet\bullet}}\right)\right]^2,$$

где $\overline{y_{i\bullet}} = \dfrac{y_{i\bullet}}{b}, \qquad \overline{y_{\bullet j}} = \dfrac{y_{\bullet j}}{a}, \qquad \overline{y_{\bullet\bullet}} = \dfrac{y_{\bullet\bullet}}{N}$

Раскрывая скобки в правой части этого соотношения, получаем:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij}-\overline{y_{\bullet\bullet}}\right)^2 = b\sum_{i=1}^{a}\left(\overline{y_{i\bullet}}-\overline{y_{\bullet\bullet}}\right)^2 + a\sum_{j=1}^{b}\left(\overline{y_{\bullet j}}-\overline{y_{\bullet\bullet}}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij}-\overline{y_{i\bullet}}-\overline{y_{\bullet j}}+\overline{y_{\bullet\bullet}}\right)^2 +$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{y_{i\bullet}}-\overline{y_{\bullet\bullet}}\right)\left(\overline{y_{\bullet j}}-\overline{y_{\bullet\bullet}}\right)+2\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{y_{i\bullet}}-\overline{y_{\bullet\bullet}}\right)\left(y_{ij}-\overline{y_{i\bullet}}-\overline{y_{\bullet j}}+\overline{y_{\bullet\bullet}}\right)+2\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{y_{\bullet j}}-\overline{y_{\bullet\bullet}}\right)\left(y_{ij}-\overline{y_{i\bullet}}-\overline{y_{\bullet j}}+\overline{y_{\bullet\bullet}}\right)$$

Простые, но громоздкие вычисления показывают, что смешанные произведения обращаются в 0.

Отсюда следует:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij}-\overline{y_{\bullet\bullet}}\right)^2 = b\sum_{i=1}^{a}\left(\overline{y_{i\bullet}}-\overline{y_{\bullet\bullet}}\right)^2 + a\sum_{j=1}^{b}\left(\overline{y_{\bullet j}}-\overline{y_{\bullet\bullet}}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij}-\overline{y_{i\bullet}}-\overline{y_{\bullet j}}+\overline{y_{\bullet\bullet}}\right)^2$$

В символическом виде данное выражение можно записать как

$$SS_{общ} = SS_{обр} + SS_{k} + SS_{ош}$$

Число всех наблюдений $N= a*b$, следовательно, $SS_{общ}$ обладает N-1 степенью свободы. Так как a число обработок, а b число блоков, то $SS_{обр}$ и $SS_{бл}$ обладают $(a$-1) и $(b$-1) степенями свободы соответственно. Число степеней свободы для ошибки будет равно $(a$-1)(b-1) и определяется через разность степеней свободы общей суммы и сумм по обработкам и блокам.

Расчетные формулы для сумм квадратов:

$$SS_{общ} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{N}y_{\bullet\bullet}^2; \qquad SS_{обр} = \frac{1}{b}\sum_{i=1}^{a} y_{i\bullet}^2 - \frac{1}{N}y_{\bullet\bullet}^2;$$

$$SS_{бл} = \frac{1}{a}\sum_{j=1}^{b} y_{\bullet j}^2 - \frac{1}{N}y_{\bullet\bullet}^2; \qquad SS_{ош} = SS_{общ} - SS_{обр} - SS_{бл}.$$

Используя, допущение о нормальности ошибки, имеем: $\frac{SS_{обр}}{\sigma^2}, \frac{SS_{бл}}{\sigma^2}, \frac{SS_{ош}}{\sigma^2}$ независимые случайные величины, подчиняющиеся χ^2 распределению. Отношение суммы квадратов к соответствующему числу степеней свободы есть средний квадрат.

Для проверки гипотезы о равенстве эффектов обработок используется статистика $F_0 = \frac{MS_{обр}}{MS_{ош}}$, которая имеет F- распределение с (a-1) и (a-1)(b-1) степенями свободы при условии истинности H_0. Если $F_0 < F_{\alpha, a-1, (a-1)(b-1)}$, то гипотеза H_0 отвергается.

Замечание Определение целесообразности группировки в блоки, можно проверить с помощью гипотезы относительно отличий блоков друг от друга.

$$H_0: \beta_j = 0,\ 1 \le j \le b$$

$$H_1: \beta_k \ne 0,\ 1 \le k \le b$$

Для проверки гипотезы используется статистика $F_0 = \frac{MS_{\text{обр}}}{MS_{\text{ош}}}$, которая имеет F-распределение с (b-1) и (a-1)(b-1) степенями свободы при условии истинности H_0.

Результаты проверки гипотезы сводят в таблицу дисперсионного анализа.

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0
Обработки	$SS_{обр}$	*a-1*	$\frac{SS_{\text{обр}}}{a-1}$	$\frac{MS_{\text{обр}}}{MS_{\text{ош}}}$
Блоки	$SS_{бл}$	*b-1*	$\frac{SS_{\text{бл}}}{b-1}$	
Ошибка	$SS_{ош}$	*(a-1)(b-1)*	$\frac{SS_{\text{ош}}}{(a-1)(b-1)}$	
Сумма	$SS_{общ}$	*N-1*		

Пример. Необходимо проверить, приводят ли 4 различных типа острия к разным показаниям прибора для определения твердости. Измерения проводятся следующим образом: в образец металла вдавливается остриё, и по величине углубления судят о твердости образца. Экспериментатор решил провести по 4 наблюдения для каждого острия. Для проведения эксперимента выбирают 4 образца металла. Каждое остриё вдавливается в каждый из образцов металла один раз. Образцы метала представляют собой блоки. Порядок проверки для каждого из образцов металла был случайным. В результате эксперимента получены следующие данные:

Тип острия	**Образец металла (блок)**			
	1	**2**	**3**	**4**
1	9,3	9,4	9,6	10
2	9,4	9,3	9,8	9,9
3	9,2	9,4	9,5	9,7
4	9,7	9,6	10	10,2

Решение

$SS_{общ} = \sum_{i=1}^{4} \sum_{j=1}^{4} y_{ij}^2 - \frac{1}{N} y_{..}^2 = (9{,}3^2+9{,}4^2+9{,}2^2+\ldots+9{,}7^2+10{,}2^2) - \frac{1}{4*4} * (9{,}3 + 9{,}4 + +9{,}2 + \cdots + 9{,}7 + 10{,}2)^2 = 1483{,}54 - \frac{1}{16}*23716 = 1{,}29$

Тип острия	Образец металла (блок)				$y_{i.}$
	1	**2**	**3**	**4**	
1	9,3	9,4	9,6	10	38,3
2	9,4	9,3	9,8	9,9	38,4
3	9,2	9,4	9,5	9,7	37,8
4	9,7	9,6	10	10,2	39,5

$SS_{обр} = \frac{1}{b} \sum_{i=1}^{4} y_{i.}^2 - \frac{1}{N} y_{..}^2 = \frac{1}{4}(38{,}3^2 + 38{,}4^2 + 37{,}8^2 + 39{,}5^2) - \frac{1}{4*4} * (9{,}3 + 9{,}4 + +9{,}2 + \cdots + 9{,}7 + 10{,}2)^2 = 0{,}385$

Тип острия	Образец металла (блок)			
	1	**2**	**3**	**4**
1	9,3	9,4	9,6	10
2	9,4	9,3	9,8	9,9
3	9,2	9,4	9,5	9,7
4	9,7	9,6	10	10,2
$y_{.j}$	37,6	37,7	38,9	39,8

$SS_{бл} = \frac{1}{a} \sum_{i=1}^{4} y_{.j}^2 - \frac{1}{N} y_{..}^2 = \frac{1}{4}(37{,}6^2 + 37{,}7^2 + 38{,}9^2 + 39{,}8^2) - \frac{1}{4*4} * (9{,}3 + 9{,}4 + +9{,}2 + \cdots + 9{,}7 + 10{,}2)^2 = 0{,}825$

$SS_{ош} = SS_{общ} - SS_{обр} - SS_{бл} = 1{,}29 - 0{,}385 - 0{,}825 = 0{,}08$

Строим таблицу ДА.

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0	$F_{кр}$ (0,05;3;9)
Обработки	0,385	3	0,128333	14,4375	3,862548
Блоки	0,825	3	0,275		
Ошибка	0,08	9	0,008889		
Сумма	1,29	15			

<u>Вывод:</u> $F_0 > F_{кр}$, следовательно гипотеза о том, что тип острия не влияет на показания твердости , отвергается. Кроме того, так как средний квадрат по блокам существенно превышает средний квадрат ошибки, то можно говорить, что разбиение на блоки целесообразно.

<u>Замечание</u>

1) Сколько именно блоков включать в план исследования? В общем случае число блоков должно выбираться достаточно большим, с тем чтобы обеспечить соответствующую мощность критерия. При увеличении числа блоков растет число степеней свободы ошибки, что повышает мощность критерия.

2) В тех случаях, когда разбиение на блоки не целесообразно (в соответствии с ДА), необходимо перейти к рассмотрению полностью рандомизированного плана. В противном случае, мощность критерия оказывается пониженной на b-1 степень свободы.

При проведении экспериментов по рандомизированному полноблочному плану, иногда оказывается, что в блоках недостает некоторых наблюдений. Это может быть вызвано невнимательностью, промахом экспериментатора или невозможностью снятия показаний. Недостающие данные приводят к новым проблемам при анализе, так как обработки уже не ортогональны блокам (каждый блок содержит все обработки).

Для решения данной проблемы существуют два подхода:

1. Рассматривается приближенный анализ, при этом находят оценку недостающего данного.
2. Рассматривается неполноблочный рандомизированный план.

<u>*Оценки недостающего данного*</u>

Обозначения:

x –недостающее данное;

$y'_{i.}$− сумма наблюдений в i-ой обработке без данного;

$y'_{.j}$ − сумма наблюдений в j-ом блоке без данного;

$y'_{..}$ − общая сумма всех наблюдений без данного.

Выбираем оценку таким образом, чтобы она давала наименьший вклад в $SS_{ош}$.

$SS_{ош} = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2$, с другой стороны:

$SS_{\text{ош}} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{b} y_{i.}^2 - \frac{1}{a} y_{.j}^2 + \frac{1}{ab} y_{..}^2 = x^2 - \frac{1}{b}(y'_{i.} + x)^2 - \frac{1}{a}(y'_{.j} + x)^2 +$ *1ab(y..′+x)2+R*, где R – это сумма всех слагаемых, не содержащих x.

Из условий минимизации $\frac{\partial SS_{\text{ош}}}{\partial x}$=0.

В результате: $x = \frac{a y'_{i.} + b y'_{.j} - y'_{..}}{(a-1)(b-1)}$.

Далее можно проводить ДА.

Замечание

1) Для определения оценки нескольких недостающих данных, необходимо равенство нулю частных производных по каждой переменной и решение системы уравнений.

2) Можно воспользоваться интерполяционным методом: находим значение одного них по представленной формуле, затем, используем его как полноценное наблюдение и определяем следующее и так далее.

3) Во всех случаях степень свободы ошибки уменьшается на число недостающих данных.

Латинский и Греко–латинский квадраты

Рандомизированный полноблочный план позволяет уменьшить остаточные ошибки эксперимента за счет изменчивости, обусловленной экспериментальными объектами. Существуют и другие виды планов, в основе которых лежит группирование в блоки. Рассмотрим наиболее популярные из них − латинский квадрат и греко-латинский квадрат.

а) *Латинский квадрат*

Рассмотрим следующую задачи: Экспериментатор исследует влияние пяти различных формул взрывчатой смеси, используемых при производстве динамита, на наблюдаемую силу взрыва. Смесь по каждой из формул изготавливается из партии сырья, объем которой позволяет проверить не более пяти формул. Смеси приготавливаются пятью различными операторами, которые могут существенно отличатся по квалификации и опыту.

План, позволяющий решить поставленную задачу, состоит в том, что необходимо проверить каждую формулу смеси в точности один раз в каждой партии сырья и по одному разу каждым из операторов. Таким образом, оказывается, что план эксперимента должен предусмотреть «усреднение» влияния двух внешних факторов: партий сырья и операторов. Поставленную задачу решает план «Латинский квадрат». В данном случае группирование в блоки происходят по двум направлениям, строки и столбцы. На строки и столбцы будут наложены ограничения на рандомизацию. Обработки основного, проверяемого фактора – формулы взрывчатой смеси, будут записаны латинскими буквами. Отсюда и название плана Латинский квадрат. Надо заметить, что партии сырья и операторы ортогональны обработкам, то есть, в каждом блоке встречаются все обработки только один раз.

В общем случае латинский квадрат представляет собой квадрат $p \times p$. Каждая из p^2 ячеек содержит одну из p латинских букв, соответствующим обработкам.

Примеры латинских квадратов:

4×4

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

5×5

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

Построим латинский квадрат для нашего примера.

Партии сырья	Операторы				
	1	2	3	4	5
1	A=24	B=20	C=19	D=24	E=24
2	B=17	C=24	D=30	E=24	A=36
3	C=18	D=38	E=26	A=27	B=21
4	D=26	E=31	A=26	B=23	C=22
5	E=22	A=30	B=20	C=29	D=31

Статистическая модель для Латинского квадрата

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \varepsilon_{ijk},$$

где $i = \overline{1,p}$; j =$\overline{1,p}$; k = $\overline{1,p}$.

Обозначения:

y_{ijk} – наблюдение в i строке, k столбце и j обработке.

μ – математическое ожидание общего среднего;

α_i – эффект i-ой строки;

τ_j – эффект j-ой обработки;

β_k – эффект k-го столбца;

ε_{ijk} – случайная ошибка, $\varepsilon_{ijk} \sim N(0; \sigma^2)$.

Модель *аддитивна*, то есть взаимодействия между строками, столбцами и обработками отсутствует.

Так как блоки представляют собой ограничение на рандомизацию, то рассматривается гипотеза относительно эффектов обработок:

$$H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$$

$$H_1 : \tau_k \neq 0, \qquad \text{где} \quad 1 \leq k \leq a.$$

Для проверки статистической гипотезы используется дисперсионный анализ.

$$SS_{\text{общ}} = SS_{\text{стр}} + SS_{\text{ст}} + SS_{\text{обр}} + SS_{\text{ош}}$$

Таблица ДА для латинского квадрата:

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0
Обработки	$SS_{обр}$	*p-1*	$\frac{SS_{\text{обр}}}{p-1}$	$\frac{MS_{\text{обр}}}{MS_{\text{ош}}}$
Строки	$SS_{стр}$	*p-1*	$\frac{SS_{\text{стр}}}{p-1}$	
Столбцы	$SS_{ст}$	*p-1*	$\frac{SS_{\text{ст}}}{p-1}$	
Ошибка	$SS_{ош}$	*(p-2)(p-1)*	$\frac{SS_{\text{ош}}}{(p-2)(p-1)}$	
Сумма	$SS_{общ}$	P^2*-1*		

$SS_{\text{обр}} = \frac{1}{p} * \sum_{j=1}^{p} y_{.j.}^2 - \frac{1}{N} * y_{...}^2$

$SS_{\text{стр}} = \frac{1}{p} * \sum_{i=1}^{p} y_{i..}^2 - \frac{1}{N} * y_{...}^2$

$SS_{\text{ст}} = \frac{1}{p} * \sum_{k=1}^{p} y_{..k}^2 - \frac{1}{N} * y_{...}^2$

$SS_{\text{общ}} = \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} y_{ijk}^2 - \frac{1}{N} * y_{...}^2$

$SS_{\text{ош}} = SS_{\text{общ}} - SS_{\text{обр}} - SS_{\text{стр}} - SS_{\text{ст}}.$

Проведем дисперсионный анализ для ранее сформулированной задачи.

Партии сырья	Операторы					$y_{i..}$
	1	2	3	4	5	
1	A=24	B=20	C=19	D=24	E=24	111
2	B=17	C=24	D=30	E=24	A=36	131
3	C=18	D=38	E=26	A=27	B=21	130
4	D=26	E=31	A=26	B=23	C=22	128
5	E=22	A=30	B=20	C=29	D=31	132
$y_{..k}$	107	143	121	127	134	$y_{...}$=632

$SS_{\text{общ}}$ = 675,04; $SS_{\text{стр}}$ = 61,04; $SS_{\text{ст}}$ = 147,84;

$SS_{\text{обр}} = \frac{1}{p} * \sum_{j=1}^{p} y_{.j.}^2 - \frac{1}{N} * y_{...}^2 = \frac{1}{5} * (143^2 + 101^2 + 112^2 + 149^2 + 127^2) - \frac{1}{25} * 632^2$= 324,84.

Определим суммы по обработкам:

Латинская буква	Сумма по обработке
A	24+36+27+26+30 = 143
B	20+17+21+23+20 = 101
C	19+24+18+22+29 = 112
D	24+30+38+26+31 = 149
E	24+24+26+31+22 = 127

$SS_{\text{ош}}$ = 675,04 - 324,84 – 61,04 – 147,84 = 138,32.

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0	$F_{\text{кр}\ (0{,}05;4;12)}$
Обработки (Формулы)	324,84	4	81,96	7,1104	3,259166727
Строки (Партии сырья)	61,04	4	15,26		
Столбцы (Операторы)	147,84	4	36,96		
Ошибка	138,32	12	11,53		
Сумма	675,04	24			

Сравним значение статистики с критической точкой:
7,1104 > 3,259166727.

Вывод: формула взрывчатого вещества влияет на силу взрыва.

Замечание

1. Можно проверить гипотезы о влиянии партий сырья и операторов на силу взрыва. Однако, так как строки и столбцы представляют собой ограничение на рандомизацию, то проверка может оказаться нестрогой.
2. Латинский квадрат, в котором буквы первой строки и первого столбца расположены в алфавитном порядке, называется *стандартным*.
3. Для латинского квадрата p×p оценку пропущенного значения можно получить по формуле $y_{ijk} = \frac{p(y'_{i..}+y'_{.j.}+y'_{..k})-2y'_{...}}{(p-2)(p-1)}$.
4. Применение латинского квадрата полезно в тех случаях, когда необходимо рассмотреть три фактора, при чем взаимодействия между ними отсутствует.

б) *Греко−латинский квадрат*

Пусть есть два Латинских квадрата p×p. В первом Латинском квадрате обработки обозначены латинскими буквами, а во втором – греческими. Наложим один квадрат на другой. Если при наложении каждая латинская буква «встречается» с каждой греческой буквой только один раз, то образующийся при этом квадрат называется *Греко−латинским квадратом*.

Строка	Столбец			
	1	2	3	4
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

Греко−латинские квадраты могут использоваться при планировании экспериментов для систематического контроля трех источников, мешающих неоднородности, то есть для группирования в блоки по трем направлениям. Таким образом, план позволяет при p^2 наблюдениях четыре фактора, причем каждый из них на p уровнях. Греко−латинские квадраты существуют для всех $p \geq 3$, кроме $p = 6$.

Статистическая модель для Греко−латинского квадрата

$$y_{ijkl} = \mu + \theta_i + \tau_j + \omega_k + \varphi_l + \varepsilon_{ijkl},$$

где $i = \overline{1,p}$; $j = \overline{1,p}$; $k = \overline{1,p}$; $l = \overline{1,p}$

Обозначения:

y_{ijkl} – наблюдение в i строке, *l* столбце, j латинской букве и k греческой.

μ – математическое ожидание общего среднего;

θ_i – эффект i-ой строки;

τ_j – эффект j-ой латинской буквы;

ω_k – эффект k-ой греческой буквы;

φ_l – эффект l столбца;

ε_{ijkl} – случайная ошибка, $\varepsilon_{ijkl} \sim N(0; \sigma^2)$.

Дисперсионный анализ проводится аналогично анализу для Латинского квадрата.

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1 : \tau_k \neq 0, \qquad \text{где} \quad 1 \leq k \leq a.$$

Формулы для вычисления скорректированных сумм

$SS_{\text{лат}} = \frac{1}{p} * \sum_{j=1}^{p} y_{.j..}^2 - \frac{1}{N} * y_{....}^2$ $\qquad SS_{\text{стр}} = \frac{1}{p} * \sum_{i=1}^{p} y_{i...}^2 - \frac{1}{N} * y_{....}^2$

$SS_{\text{гр}} = \frac{1}{p} * \sum_{k=1}^{p} y_{..k.}^2 - \frac{1}{N} y_{....}^2$ $\qquad SS_{\text{ст}} = \frac{1}{p} * \sum_{l=1}^{p} y_{...l}^2 - \frac{1}{N} * y_{....}^2$

$SS_{\text{общ}} = \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} y_{ijkl}^2 - \frac{1}{N} * y_{...}^2$

$SS_{\text{ош}} = SS_{\text{общ}} - SS_{\text{лат}} - SS_{\text{гр}} - SS_{\text{стр}} - SS_{\text{ст}}.$

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0
Латинские	$SS_{лат}$	*p-1*	$\frac{SS_{\text{лат}}}{p-1}$	$\frac{MS_{\text{лат}}}{MS_{\text{ош}}}$
Греческие	$SS_{гр}$	*p-1*	$\frac{SS_{\text{гр}}}{p-1}$	
Строки	$SS_{стр}$	*p-1*	$\frac{SS_{\text{стр}}}{p-1}$	
Столбцы	$SS_{ст}$	*p-1*	$\frac{SS_{\text{ст}}}{p-1}$	
Ошибка	$SS_{ош}$	*(p-3)(p-1)*	$\frac{SS_{\text{ош}}}{(p-3)(p-1)}$	
Сумма	$SS_{общ}$	p^2*-1*		

<u>Пример</u> Предположим, что при исследовании формул взрывчатого вещества может оказаться важным еще один фактор – испытательная установка. Пусть таких установок пять и обозначим их греческими буквами.

Партии сырья	Операторы					$y_{i...}$
	1	2	3	4	5	
1	Aα=24	Bγ=20	Cε=19	Dβ=24	Eδ=24	111
2	Bβ=17	Cδ=24	Dα=30	Eγ=24	Aε=36	131
3	Cγ=18	Dε=38	Eβ=26	Aδ=27	Bα=21	130
4	Dδ=26	Eα=31	Aγ=26	Bε=23	Cβ=22	128
5	Eε=22	Aβ=30	Bδ=20	Cα=29	Dγ=31	132
$y_{...l}$	107	143	121	127	134	$y_{....}$=632

Латинская буква	Сумма по латинским

A	24+36+27+26+30 = 143
B	20+17+21+23+20 = 101
C	19+24+18+22+29 = 112
D	24+30+38+26+31 = 149
E	24+24+26+31+22 = 127

Греческие	Сумма по греческим

буква	
α	24+31+30+29+21 = 135
β	17+30+26+24+22 = 119
γ	18+20+26+24+31 = 119
δ	26+24+20+27+24 = 121
ε	22+38+19+23+36 = 138

$SS_{\text{лат}} = 327{,}84$; $SS_{\text{гр}} = 69{,}44$;

$SS_{\text{стр}} = 61{,}04$; $SS_{\text{ст}} = 147{,}84$;

$SS_{\text{общ}} = 675{,}04$;

$SS_{\text{ош}} = 68{,}88$.

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0	$F_{\text{кр}\,(0,05;4;12)}$
Латинские (Формулы)	324,84	4	81,96	9,5192	3,837853
Греческие (Установки)	69,44	4	17,36		
Строки (Партии сырья)	61,04	4	15,26		
Столбцы (Операторы)	147,84	4	36,96		
Ошибка	68,88	8	8,61		
Сумма	675,04	24			

Вывод: формула взрывчатого вещества влияет на силу взрыва.

Замечание: Можно проверить гипотезы о влиянии партий сырья, операторов и установок на силу взрыва. Однако, так как строки и столбцы представляют собой ограничение на рандомизацию, то проверка может оказаться нестрогой.

Рандомизированные неполноблочные планы

В некоторых экспериментах с рандомизированными блочными планами может оказаться, что нельзя проверить все обработки в каждом из блоков. Ситуации такого рода обычно объясняется либо нехваткой средств или экспериментальной аппаратуры, либо реальными размерами блока. Такие планы известны под названием *рандомизированные неполноблочные планы*. Рассмотрим далее некоторые из них.

1) Сбалансированный неполноблочный план

<u>*Сбалансированный неполноблочный план*</u> – это такой неполноблочный план, в котором <u>все возможные пары</u> обработок встречаются одинаковое число раз. Если сравнения обработок между собой является одинаково важными, то обработки в каждом блоке должны выбираться сбалансированно, то есть так, чтобы каждая пара обработок встречалась столько же раз, сколько и любая другая.

Пусть a – число обработок, b – число блоков, k – число обработок в каждом блоке, каждая обработка встречается в эксперименте r раз (или производится r реплик). Тогда общее число наблюдений составляет N = a*r =b*k.

Если a = b, то план называется *симметричным*.

Каждая пара обработок в одном и том же блоке встречается $\lambda = \frac{r(k-1)}{a-1}$ раз, причем параметр λ должен быть целым.

Статистическая модель плана:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$
$$i = 1,2,\ldots, a \quad j = 1, 2, \ldots, b$$

где μ - математическое ожидание общего среднего;

τ_i - эффект *i*- ой обработки;

β_j - эффект *j*- ого блока;

ε_{ij} - случайная ошибка, причем $\varepsilon_{ij} \sim N(0, \sigma^2)$

Так как блоки представляют собой ограничение на рандомизацию, то рассматривается гипотеза относительно эффектов обработок:

$H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$

$H_1: \tau_k \neq 0,$ где $1 \le k \le a$.

Общая изменчивость данных, выраженная общей скорректированной суммой квадратов может быть представлена в виде разбиения

$$SS_{\text{общ}} = SS_{\text{обр(испр)}} + SS_{\text{бл}} + +SS_{\text{ош}}$$

В сумме квадратов для обработок введена поправка для разделения эффектов блоков и обработок. Исправленная сумма квадратов для обработок определяется выражением:

$$SS_{\text{обр(испр)}} = \frac{k}{\lambda a} \sum_{i=1}^{a} Q_i^2,$$

где Q_i – исправленная сумма наблюдений для i – ой обработки.

$$Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij}y_{.j},\ i = \overline{1,a},$$

причем, n_{ij}= 1, если i – ая обработка встречается в j – ом блоке

n_{ij}= 0, если i – ая обработка не встречается в j – ом блоке.

<u>Замечание:</u> Сумма исправленных сумм по обработкам всегда равна 0.

$$SS_{общ} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{N}y_{\bullet\bullet}^2; \qquad SS_{бл} = \frac{1}{k}\sum_{j=1}^{b} y_{\bullet j}^2 - \frac{1}{N}y_{\bullet\bullet}^2;$$

$$SS_{ош} = SS_{общ} - SS_{обр(испр)} - SS_{бл}.$$

При проверке гипотезы о равенстве эффектов обработок должна использоваться статистика : $F_0 = \frac{MS_{обр(испр)}}{MS_{ош}}$.

Основные соотношения собраны в таблицу дисперсионного анализа

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0
Обработки (исправленные)	$SS_{обр(испр)}$	a-1	$\frac{SS_{обр(испр)}}{a-1}$	$\frac{MS_{обр(испр)}}{MS_{ош}}$
Блоки	$SS_{бл}$	b-1	$\frac{SS_{бл}}{b-1}$	
Ошибка	$SS_{ош}$	N-a-b+1	$\frac{SS_{ош}}{N-a-b+1}$	
Сумма	$SS_{общ}$	N-1		

<u>Пример</u> Инженер-химик считает, что время протекания некоторого химического процесса зависит от вида применяемого катализатора. В процедуру эксперимента входит выбор партии сырья, загрузка опытной установки, проведение отдельного цикла работы опытной установки с каждым из катализаторов и определение продолжительности реакции. На эффективность применения катализатора может влиять изменчивость сырья по партиям. Инженер решил использовать партии в качестве блоков. Объем партии позволяет провести только 3 цикла работы. Результаты измерений приведены в таблице.

Обработка (катализаторы)	Блок (партии сырья)				$y_{i.}$
	1	2	3	4	
1	73	74		71	**218**
2		75	67	72	**214**

3	73	75	68		**216**
4	75		72	75	**222**
$y_{.j}$	**221**	**224**	**207**	**218**	***y.. = 870***

<u>Решение:</u> a = 4; b = 4; k = 3; r = 3; N = 12.

$\lambda = \frac{3(3-1)}{4-1} = 2.$

$SS_{общ} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{N}y_{..}^2 = 81;\ \ SS_{бл} = \frac{1}{k}\sum_{j=1}^{b} y_{.j}^2 - \frac{1}{N}y_{..}^2 = 55.$

Определим исправленные суммы по обработкам.

$Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij} y_{.j},\ i = \overline{1, a},$

$Q_1 = y_{1.} - (1/3) \cdot (1 \cdot y_{.1} + 1 \cdot y_{.2} + 0 \cdot y_{.3} + 1 \cdot y_{.4}) = 218 - (1/3) \cdot (221+224+218) = -3$

$Q_2 = y_{2.} - (1/3) \cdot (0 \cdot y_{.1} + 1 \cdot y_{.2} + 1 \cdot y_{.3} + 1 \cdot y_{.4}) = 214 - (1/3) \cdot (224+207+218) =$
$= -7/3$

Q3 = 216 – (1/3)*(221+224+207) = -4/3

Q4 = 222 – (1/3)*(221+207+218) = 20/3.

Проверим правильность найденных значений:

-3 + (-7/3) + (-4/3) + 20/3 = -20/3+20/3 = 0.

$SS_{обр(испр)} = \frac{k}{\lambda a}\sum_{i=1}^{a} Q_i^2 = \frac{3}{2*4}\sum_{i=1}^{4} Q_i^2 = \frac{3}{8}(9 + \frac{49}{9} + \frac{16}{9} + \frac{400}{9}) = 22{,}75.$

$SS_{ош} = SS_{общ} - SS_{обр(испр)} - SS_{бл} = 81 - 55 - 22{,}75 = 3{,}25.$

Результаты дисперсионного анализа приведены в таблице

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0	$F_{кр\ (0,05;3;5)}$
Обработки (исправленные)	22,75	3	7,58	11,66	5,41
Блоки	55	3			
Ошибка	3,25	5	0,65		
Сумма	81	11			

<u>Вывод</u>: Так как $F_0 > F_{(0,05;3;5)}$, то применяемые катализаторы оказывают значимое влияние на время протекания процесса.

2)Частично сбалансированный неполноблочный план

Для сбалансированного неполноблочного плана характерно то, что параметр λ должен быть целым. Однако, это может привести к слишком большому числу или размеру блоков. Поэтому, для уменьшения числа блоков исследователь может использовать частично сбалансированный план, в котором одни пары обработок встречаются λ_1 раз, другие λ_2 раз, … и оставшиеся пары λ_m раз.

Пары обработок, которые встречаются λ_i раз, называются i-ассоциированными и говорят, что план содержит m ассоциированных классов.

Например,

Блок	Комбинации обработок		
1	1	2	3
2	3	4	5
3	2	5	6
4	1	2	4
5	3	4	6
6	1	5	6

Пара обработок (1,2) встречается дважды, а пара (2,5) только один раз. Проверив все пары обработок, придем к выводу, что в плане есть только 1-ассоциированный класс ($\lambda_1 = 2$) и 2-ассоциированный класс ($\lambda_2 = 1$).

Если две обработки являются i-ассоциированными, план внутриблочного анализа с двумя ассоциированными классами (разница между блоками исключается):

1. a – число обработок; b – число блоков. Каждый блок содержит k наблюдений и каждая обработка встречается в r блоках.
2. Пара обработок, являющихся i-ассоциированными, встречается в λ_i блоках i = 1,2.
3. У каждой обработки ровно n_i i – ассоциированных обработок, i = 1,2. Число n_i не зависит от выбранной обработки.

4. Если две обработки являются i-ассоциированными, то число обработок, j-ассоциированных с одной из них и k-ассоциированных с другой равно p_{jk}^{i} (i, j, k = 1,2).

Рассмотрим правило определения величины p_{jk}^{i}.

В условиях ранее указанного примера построим 1− ассоциированную и 2 – ассоциированную матрицы. Для построения выберем пару обработок и укажем, какие обработки 1− ассоциированы с ними и 2 – ассоциированы.

Например, пара (3,4) 1− ассоциированных.

<table>
<tr><td rowspan="2">Обработка 3</td><td colspan="4">Обработка 4</td></tr>
<tr><td colspan="2">1 − ассоциирована</td><td colspan="2">2 − ассоциирована</td></tr>
<tr><td>1 − ассоциирована</td><td>4
3</td><td>Нет совпаде-ний (0)</td><td>4
1,2,5,6</td><td>Нет совпа-дений (0)</td></tr>
<tr><td>2 − ассоциирована</td><td>1,2,5,6
3</td><td>Нет совпаде-ний (0)</td><td>1,2,5,6
1,2,5,6</td><td>Совпадений (4)</td></tr>
</table>

В результате матрица $p_{jk}^{1} = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$.

Аналогично построим матрицу для 2 − ассоциированной пары.

<table>
<tr><td rowspan="2">Обработка 5</td><td colspan="4">Обработка 4</td></tr>
<tr><td colspan="2">1 − ассоциирована</td><td colspan="2">2 − ассоциирована</td></tr>
<tr><td>1 − ассоциирована</td><td>6
3</td><td>Нет совпаде-ний (0)</td><td>6
1,2,5,6</td><td>Совпадений (1)</td></tr>
<tr><td>2 − ассоциирована</td><td>1,2,3,4
3</td><td>Совпадений (1)</td><td>1,2,3,4
1,2,5,6</td><td>Совпадений (2)</td></tr>
</table>

В результате матрица $p_{jk}^{2} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$.

Статистическая модель плана:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

$$i = 1,2,\dots, a \qquad j = 1, 2, \dots, b$$

где μ - математическое ожидание общего среднего;

τ_i - эффект *i*- ой обработки;

β_j - эффект *j*- ого блока;

ε_{ij} - случайная ошибка, причем $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Так как блоки представляют собой ограничение на рандомизацию, то рассматривается гипотеза относительно эффектов обработок:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1 : \tau_i \neq 0, \qquad \text{где} \quad 1 \leq i \leq a\,.$$

Формулы для вычисления скорректированных сумм:

1. Исправленная сумма квадратов по i обработке.

a) $Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij} y_{.j}$;

b) $S_1(Q_i) = \sum Q_s$, s и i - 1 – ассоциированы;

c) $\Delta = \frac{1}{k^2}\{(rk - r + \lambda_1)(rk - r + \lambda_2) + (\lambda_1 - \lambda_2)[r(k-1)(\,p_{12}^1 - \,p_{12}^2) +$

+λ2p121–λ1 p122]

d) $c_1 = \frac{1}{k\Delta}[\lambda_1(rk - r + \lambda_2) + (\lambda_1 - \lambda_2)(\lambda_2 p_{12}^1 - \lambda_1\, p_{12}^2)]$

e) $c_2 = \frac{1}{k\Delta}[\lambda_2(rk - r + \lambda_1) + (\lambda_1 - \lambda_2)(\lambda_2 p_{12}^1 - \lambda_1\, p_{12}^2)]$

f) Оценка эффекта i–ой обработки:

$\widehat{\tau_i} = \frac{1}{r(k-1)}[(k - c_2)Q_i + (c_1 - c_2)S_1(Q_i)]$

k) $SS_{\text{обр(испр)}} = \sum_{i=1}^{a} \widehat{\tau_i}\, Q_i$.

2. Исправленная сумма квадратов по блокам.

$$SS_{\text{бл}} = \frac{1}{k}\sum_{j=1}^{b} y_{.j}^2 - \frac{1}{bk} y_{..}^2\ .$$

3. Общая сумма

$$SS_{\text{общ}} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{bk} y_{..}^2\ .$$

При проверке гипотезы о равенстве эффектов обработок будет использоваться статистика : $F_0 = \frac{MS_{\text{обр(испр)}}}{MS_{\text{ош}}}$.

Основные соотношения собраны в таблицу дисперсионного анализа

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0
Обработки (испр)	$SS_{обр(испр)}$	*a-1*	$\frac{SS_{\text{обр(испр)}}}{a-1}$	$\frac{MS_{\text{обр(испр)}}}{MS_{\text{ош}}}$
Блоки	$SS_{бл}$	*b-1*	$\frac{SS_{\text{бл}}}{b-1}$	

Ошибка	$SS_{ош}$	$bk\text{-}b\text{-}a\text{+}1$	$\dfrac{SS_{ош}}{bk-b-a+1}$	
Сумма	$SS_{общ}$	$bk\text{-}1$		

3) Квадрат Юдена

Существуют неполные латинские квадраты, в которых число столбцов не совпадает с числом строк и обработок. Большинство таких планов были разработаны Юденом, поэтому соответствующие квадраты носят его имя.

Строка	Столбец			
	1	2	3	4
1	A	B	C	D
2	B	C	D	E
3	C	D	E	A
4	D	E	A	B
5	E	A	B	C

Квадрат Юдена представляет собой, в общем случае, латинский квадрат, в котором не хватает хотя бы одного столбца (или строки, или диагонали). При произвольном выбрасывании более одного столбца из латинского квадрата может нарушиться его сбалансированность. В общем случае квадрат Юдена соответствует симметричному сбалансированному неполноблочному плану (число обработок совпадает с числом блоков), при чем строки представляют собой блоки, а каждая обработка встречается в каждом столбце или каждом «положении» блока только один раз.

Статистическая модель плана:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \varepsilon_{ijk},$$

i = 1,2,…, b j = 1, 2, …, a

где μ - математическое ожидание общего среднего;

α_i - эффект i-го блока

τ_j - эффект j- ой обработки;

β_k - эффект k- ого «положения»;

ε_{ijk} - случайная ошибка, причем $\varepsilon_{ijk} \sim N(0, \sigma^2)$

Так как «положения» встречаются в каждом блоке и с каждой обработкой только один раз, то они ортогональны блокам и обработкам. Анализ при исследовании квадрата Юдена аналогичен анализу сбалансированного неполноблочного плана, но дополнительно желательно найти сумму квадратов для «положений».

Пример Инженер по организации производства исследует влияние пяти уровней освещенности на частость появления дефектов при некоторо сборочной операции. В этом эксперименте время рассматривается как фактор, поэтому инженер решает использовать пять дней как блоки. Однако в помещении, где проводится эксперимент, четыре различных рабочих места, что является потенциальным источником изменчивости. Для анализа ситуации инженер решает применить квадрат Юдена с пятью строками (дни или блоки), четырьмя столбцами (рабочее место) и пятью обработками (уровни освещенности). В таблице приведены кодированные данные.

День	Рабочее место			
	1	**2**	**3**	**4**
1	A=3	B=1	C=-2	D=0
2	B=0	C=0	D=-1	E=7
3	C=-1	D=0	E=5	A=3
4	D=-1	E=6	A=4	B=0
5	E=5	A=2	B=1	C=-1

Решение

Проведем предварительные расчеты.

День	Рабочее место				$y_{i..}$	Суммы по обработкам $y_{.j.}$
	1	**2**	**3**	**4**		
1	A=3	B=1	C=-2	D=0	2	$y_{.1.}$ =12 (A)
2	B=0	C=0	D=-1	E=7	6	$y_{.2.}$ = 2 (B)

3	C=-1	D=0	E=5	A=3	7	$y_{.3.}$ = -4 (C)
4	D=-1	E=6	A=4	B=0	9	$y_{.4.}$ = -2 (D)
5	E=5	A=2	B=1	C=-1	7	$y_{.5.}$ = 23 (E)
$\boldsymbol{y_{..h}}$	6	9	7	9	$y_{...}$ =31	

Определим значения коэффициентов для нашей задачи:

a = b = 5; r = k =4 и λ = $\frac{4*(4-1)}{5-1}$ = 3.

Найдем исправленные суммы для обработок.

$Q_j = y_{.j.} - \frac{1}{k}\sum_{i=1}^{b} n_{ij}y_{i..},\ j = \overline{1,a},$

$Q_1 = y_{.1.} - \frac{1}{4}\sum_{i=1}^{5} n_{ij}y_{i..} = 12 - \frac{1}{4} * (2 + 7 + 9 + 7) = \frac{23}{4}$

(Обработка A встречается в 1 день, 3, 4 и 5.)

$Q_2 = y_{.2.} - \frac{1}{4}\sum_{i=1}^{5} n_{ij}y_{i..} = 2 - \frac{1}{4} * (2 + 6 + 9 + 7) = \frac{-16}{4}$

$Q_3 = y_{.3.} - \frac{1}{4}\sum_{i=1}^{5} n_{ij}y_{i..} = -4 - \frac{1}{4} * (2 + 6 + 7 + 7) = \frac{-38}{4}$

$Q_4 = y_{.4.} - \frac{1}{4}\sum_{i=1}^{5} n_{ij}y_{i..} = -2 - \frac{1}{4} * (2 + 6 + 7 + 9) = \frac{-32}{4}$

$Q_5 = y_{.5.} - \frac{1}{4}\sum_{i=1}^{5} n_{ij}y_{i..} = 23 - \frac{1}{4} * (6 + 7 + 9 + 7) = \frac{63}{4}$

Проверяем условие: сумма исправленных сумм по обработкам должна быть равна 0.

$\frac{23}{4} + \frac{-16}{4} + \frac{-38}{4} + \frac{-32}{4} + \frac{63}{4} = \frac{86}{4} - \frac{86}{4} = 0$.

Находим исправленную скорректированную сумму по обработкам.

$SS_{\text{обр(испр)}} = \frac{k}{\lambda a}\sum_{j=1}^{a} Q_j^2 = \frac{4}{3*5}\sum_{j=1}^{5} Q_j^2 = \frac{4}{15}(\left(\frac{23}{4}\right)^2 + \left(\frac{-16}{4}\right)^2 + \left(\frac{-38}{4}\right)^2 + \left(\frac{-32}{4}\right)^2 + \left(\frac{63}{4}\right)^2)$ = 120,37.

Для построения дисперсионного анализа необходимо рассмотреть исправленные суммы по блокам.

$Q_i' = y_{i..} - \frac{1}{r}\sum_{j=1}^{a} n_{ij}y_{.j.},\ i = \overline{1,b},$

$Q_1' = y_{1..} - \frac{1}{4}\sum_{j=1}^{5} n_{ij}y_{.j.} = 2 - \frac{1}{4} * (12 + 2 - 4 - 2) = 0$

$Q_2' = y_{2..} - \frac{1}{4}\sum_{j=1}^{5} n_{ij}y_{.j.} = 6 - \frac{1}{4} * (2 - 4 - 2 + 23) = \frac{5}{4}$

$$Q'_3 = y_{3..} - \frac{1}{4}\sum_{j=1}^{5} n_{ij}y_{.j.} = 7 - \frac{1}{4} * (-4 - 2 + 23 + 12) = -\frac{1}{4}$$

$$Q'_4 = y_{4..} - \frac{1}{4}\sum_{j=1}^{5} n_{ij}y_{.j.} = 9 - \frac{1}{4} * (-2 + 23 + 12 + 2) = \frac{1}{4}$$

$$Q'_5 = y_{5..} - \frac{1}{4}\sum_{j=1}^{5} n_{ij}y_{.j.} = 7 - \frac{1}{4} * (23 + 12 + 2 - 4) = -\frac{5}{4}$$

Проверяем условие:

$$0 + \frac{5}{4} + \frac{-1}{4} + \frac{1}{4} + \frac{-5}{4} = 0\ .$$

Вычисляем исправленную скорректированную сумму квадратов по блокам и составляем таблицу дисперсионного анализа.

$$SS_{\text{бл(испр)}} = \frac{r}{\lambda b}\sum_{i=1}^{b} Q_i^2 = \frac{4}{3*5}\sum_{i=1}^{5} Q_i^2 = \frac{4}{15}((0)^2 + \left(\frac{5}{4}\right)^2 + \left(\frac{-1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + $$
$$+ \left(\frac{-5}{4}\right)^2) = 0.87$$

Полный дисперсионный анализ приведен в таблице.

Источник изменчивости	Сумма квадратов	Степень свободы	Средний квадрат	Статистика F_0	$F_{\text{кр}\ (0,05;4;8)}$
Уровень освещенности	120,37	4	30,09	19,41	3,84
Дни	0,87	4	0,22	0,14	3,84
Рабочее место	1,35	3	0,45	0,29	4,07
Ошибка	12,36	8	1,55		
Сумма	134,95	19			

<u>Вывод:</u> Так как $F_0 > F_{(0,05;4;8)}$, то уровень освещенности оказывают значимое влияние на частость появления дефектов. Кроме того, дни недели и расположение рабочего времени не оказывает влияние, следовательно, разбиение на блоки в данном эксперименте нецелесообразно. Проведем исследование с помощью однофакторного анализа.

Уровень освещенности	Наблюдение			
	1	**2**	**3**	**4**
A	3	2	4	3
B	0	1	1	0
C	-1	0	-2	-1
D	-1	0	-1	0
E	5	6	5	7

Результат обработки в MS Excel:

Однофакторный дисперсионный анализ					
ИТОГИ					
Группы	Счет	Сумма	Среднее	Дисперсия	
A	4	12	3	0,666667	
B	4	2	0,5	0,333333	
C	4	-4	-1	0,666667	
D	4	-2	-0,5	0,333333	
E	4	23	5,75	0,916667	
ДИСПЕРСИОННЫЙ АНАЛИЗ					
Источник вариации	SS	df	MS	F	F критическое
Уровень освещенности	126,2	4	31,55	54,08571	3,055568276
Ошибка	8,75	15	0,583333		
Сумма	134,95	19			

Вывод: Так как $F_0 > F_{(0,05;4;15)}$, то уровень освещенности оказывают значимое влияние на частость появления дефектов.

СПИСОК ЛИТЕРАТУРЫ

1. Монтгомери, Д.К. Планирование эксперимента и анализ данных. − Л:, Судостроение, 1980г., −384с.
2. Ллойд Э., Ледерман У. (ред.). Справочник по прикладной статистике. Том 1. − М.: Финансы и статистика, 1989. - 510 с.
3. Хикс Ч. Основные принципы планирования эксперимента. − М.: Мир. 1967, - 406 с.
4. Р 50.1.040-2002 «Статистические методы. Планирование экспериментов. Термины и определения».

Любимцева Ольга Львовна

Блочное планирование эксперимента и анализ данных

Учебное пособие

Подписано в печать ___ Формат 60х90 1/8. Бумага газетная. Печать трафаретная.
Уч.-изд. л. 3,5. Усл. Печ. 3,8. Тираж 300 экз. Заказ № _____
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Нижегородский государственный архитектурно-строительный университет»
603950, Нижний Новгород, ул. Ильинская, 65.
Полиграфический центр ННГАСУ, 603950, Н.Новгород, Ильинская, 65
http://www. nngasu.ru, srec@nngasu.ru