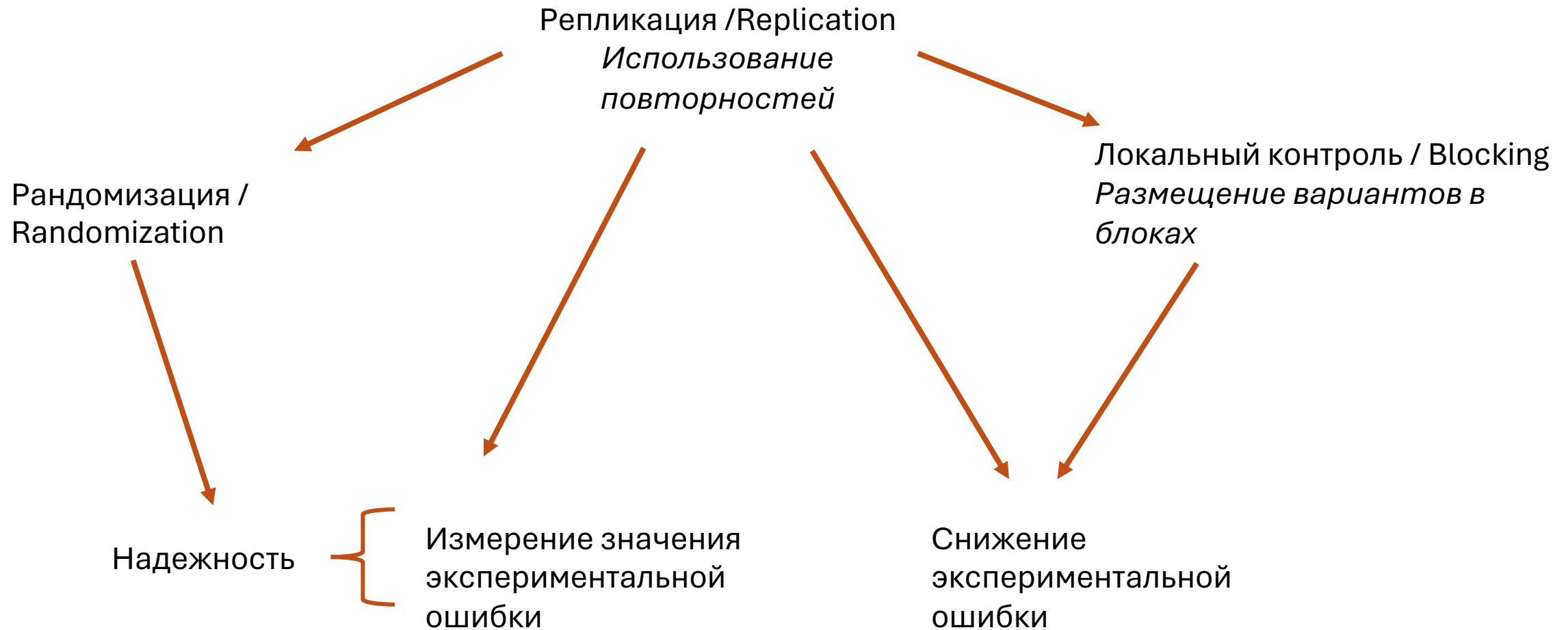


# Размещение вариантов в селекционных опытах. Ч 2

---

# Ключевые концепции в планировании опытов



# Пример

01 Рандомизация

02 Репликация

03 Разделение на блоки

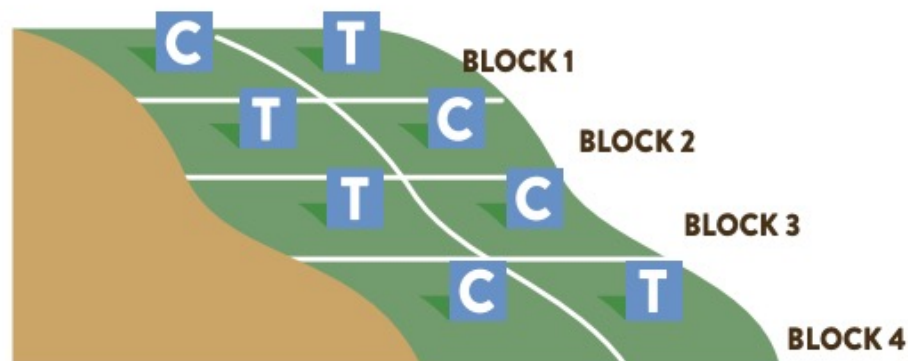


FIGURE 2a

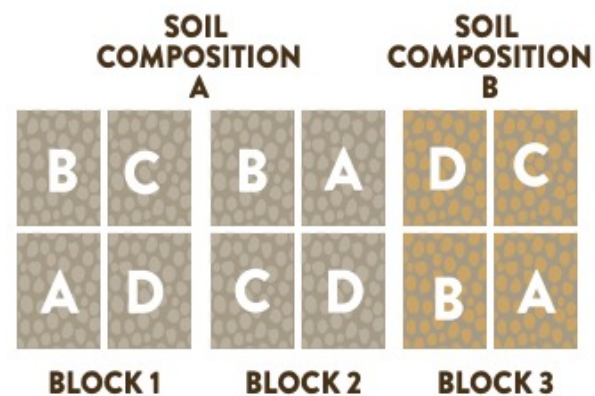


FIGURE 2b

<https://www.sare.org/wp-content/uploads/how-to-conduct-research-on-your-farm-or-ranch.pdf>

# Выбор дизайна эксперимента и способа статистической обработки данных

## Ранние поколения

- Наблюдения за отдельными растениями / рядами: без плана эксперимента и статистики
- Без повторностей / с частичной репликацией в одной или нескольких локациях
- 1-rep lattice, augmented designs (повторности для контроля), p-rep

Spatial adjustment (Moving Average, row-col, SpATS)

## Поздние поколения

- Испытания в нескольких локациях
- Блочные и lattice-дизайны (включая альфа-дизайн) с повторностями
- r-пер с частичной репликацией
- Без репликации (локация как повторность)
- Факторный эксперимент: split-plot

ANOVA или смешанные линейные модели

- С полным набором повторностей (*Fully replicated*)
- С частичным повторением (*P-rep*)
- Без повторения (*Unreplicated*)

С полным набором повторностей

# Полностью рандомизированный дизайн / Complete block design (CBD)

- Данный метод является наиболее базовым и простым в экспериментальной методологии.
- Варианты опыта распределяются случайным образом
- Этот подход наиболее эффективен, когда нет источников различий, способных исказить результаты (например, на малых и однородных участках).
- Основным преимуществом данного дизайна является его простота и легкость внедрения, однако недостатком становится отсутствие контроля над внешними факторами

Completely Randomized Design 4X12

	T-12	T-8	T-2	T-5	T-10	T-2	T-2	T-7	T-12	T-5	T-4	T-2
	T-3	T-3	T-3	T-4	T-4	T-9	T-1	T-11	T-7	T-6	T-8	T-9
	T-12	T-11	T-10	T-9	T-10	T-6	T-9	T-7	T-10	T-1	T-1	T-8
	T-11	T-12	T-1	T-6	T-3	T-11	T-5	T-5	T-4	T-6	T-7	T-8
ROWS	COLUMNS											

*Пример:*

*Число вариантов = 12 (число сортов)*

*Число повторностей = 4*

# Полностью рандомизированный блочный дизайн / Randomize Complete Block Design (RCBD)

- Этот дизайн полезен, когда экспериментальные единицы могут быть сгруппированы в однородные блоки, и варианты опыта случайным образом распределяются внутри каждого блока
- Он подходит, когда внешние источники различий могут быть контролированы путем организации блоков
- Преимуществом этого дизайна является его способность контролировать внешние источники вариации
- Однако его недостаток заключается в ограниченной способности тестировать большое количество вариантов (до 20 вариантов)
- Различия между блоками могут быть «отделены» от случайной ошибки в ходе дисперсионного анализа
- Условия внутри блока должны быть однородными
- Условия между блоками могут быть различными («эффект блока»)

Randomized Complete Block Design 9X6

	G-16	G-17	G-4	G-3	G-8	G-14
	G-5	G-15	G-13	G-6	G-1	G-2
	G-9	G-18	G-10	G-7	G-11	G-12
ROWS	G-15	G-7	G-10	G-8	G-3	G-9
	G-14	G-11	G-1	G-4	G-17	G-18
	G-6	G-12	G-2	G-16	G-13	G-5
	G-16	G-12	G-9	G-18	G-8	G-7
	G-11	G-5	G-4	G-13	G-1	G-17
	G-15	G-14	G-3	G-10	G-2	G-6
	COLUMNS					

Пример:

Число вариантов = 18 (число сортов)

Число повторностей = 3

## ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	7.997	2.6657	<u>30.013</u>	9.5e-09 ***
Cultivar	9	4.257	0.4730	<u>5.326</u>	0.00033 ***
Residuals	27	2.398	0.0888		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Фактор «блока» объяснил больше различий в результатах, чем фактор «сорта»

$$LSD = t \times \sqrt{\frac{2 * MSE}{n}}$$

n - = количество наблюдений, использованных для расчета каждого среднего значения (4 повторности!).

## Сравнения средних значений – НСР!! (или LSD)

```
> print(LSD.test(model1, "Cultivar"))
```

\$statistics

MSerror	Df	Mean	CV	t.value	LSD
0.0888163	27	6.4345	4.631605	2.051831	0.4323872

Минимальная средняя разница, которую мы должны учитывать между особями (генотипами) в выборке, которую мы анализируем - 0,432 тонны / га

\$groups

	Yield	groups
7	6.9925	a
8	6.7950	ab
1	6.6175	abc
10	6.5450	bc
5	6.4775	bcd
9	6.3700	bcd
3	6.3425	cd
6	6.3300	cd
2	6.1075	de
4	5.7675	e

Разность средних значений урожайности между генотипами 7 и 9 составляет:  
6.9925 т/га - 6.3700 т/га = 0.6225 т/га.

Теперь сравним эту разность с LSD:  
•0.6225 т/га > 0.4323872 т/га (LSD)

Разность между средними урожаями генотипов 7 и 9 больше значения LSD, следовательно, мы можем заключить, что средний урожай этих двух генотипов статистически значительно различается.



## Важно!

- НСР - пост hoc тест для сравнения средних значений ! (один из самых простых, что делает его доступным даже для тех, кто не имеет глубоких знаний в статистике)
- НСР позволяет сравнивать все пары обработок **после того, как основная ANOVA показала наличие значимых различий**
- НСР может быть использован в различных дизайнах экспериментов, включая RCBD и альфа-дизайн
- **Подходит для «предварительно не запланированных сравнений»:** в селекционных исследованиях, как правило, заранее не планируются конкретные пары сравнений, и НСР предоставляет метод, который позволяет свободно выбирать сравнения после проведения эксперимента.
- Хорошо работает на небольших наборах данных
- Когда проводим большое количество сравнений – работает плохо, так как это увеличивает общую вероятность совершения ошибки первого рода (т.е., неправильно выявляет различие, когда его нет). В таких случаях можно использовать более строгие методы множественного тестирования (тест Тьюки или Бонферрони).
- **Не используем НСР для оценки достоверности опыта!!!**

# **Пакет agricolae – все (ну почти), что нужно для анализа данных агрономических испытаний**

<https://cran.r-project.org/web/packages/agricolae/vignettes/tutorial.pdf>

# Неполный блочный дизайн / Incomplete Blocks Design (IBD)

- Когда невозможно распределить все варианты в эксперименте по полным блокам (ранние стадии селекционной программы – очень много образцов!)
- План неполного блочного дизайна размера  $(v, k, r)$  представляет собой расположение  $v$  вариантов, распределенных в неполных блоках размера  $k$
- Варианты произвольно распределяются в каждом неполном блоке, но не каждый вариант встречается в каждом блоке
- Неполный блочный дизайн более эффективен, чем RCBD, при большом количестве тестируемых вариантов ( $> 20$ )

Alpha Lattice Design Field Layout 15X16

48	79	19	69	60	65	51	17	73	64	56	38	46	5	32	74
70	61	71	10	23	41	67	26	13	8	20	11	75	28	3	50
16	49	59	53	18	30	68	36	44	25	37	34	39	14	4	77
31	76	63	35	40	12	72	24	42	33	66	43	58	7	54	57
80	29	6	9	15	55	22	2	52	62	45	1	21	27	78	47
67	19	13	14	55	24	33	26	23	46	73	6	11	53	71	9
70	44	16	74	32	47	61	66	1	15	17	42	39	56	48	52
65	57	68	64	21	63	72	23	75	31	29	43	20	50	49	38
54	76	10	51	7	34	69	30	12	59	58	41	17	80	62	2
77	35	18	36	79	40	37	8	45	3	22	78	60	5	28	4
67	45	3	17	52	6	49	18	54	44	56	33	9	30	28	43
1	32	13	71	48	2	41	4	73	37	64	59	47	29	78	23
26	16	68	53	51	80	46	61	63	34	79	11	20	31	66	69
35	36	8	65	40	39	10	12	5	70	15	77	55	14	57	58
7	19	76	60	75	72	24	50	25	21	42	27	62	38	22	74

Пример:

$(v, k, r) = (80, 5, 3)$

Число вариантов  $v=80$  (число сортов)

Размер неполного блока  $k=5$  вариантов

Число повторностей = 3

# Сбалансированный неполный блочный дизайн

Каждый вариант опыта появляется в одном блоке с каждым другим вариантом опыта одинаковое количество раз

- $t$  = количество вариантов опыта
- $k$  = количество единиц в блоке (размер блока)
- $b$  = общее количество блоков в эксперименте
- $r$  = количество повторений каждого варианта опыта
- $\lambda$  = количество раз, когда варианты опыта встречаются вместе в одном блоке

$$\lambda = \frac{r(k-1)}{t-1}$$

Все пары вариантов опыта сравниваются с одинаковой степенью точности, несмотря на то, что различия между блоками могут быть значительными.

# Сбалансированный неполный блочный дизайн

- Условия, что бы достичь «баланса» ,  $r = \lambda(t-1)/(k-1)$ , где:
  - $t$  = количество вариантов опыта
  - $k$  = количество единиц в блоке (размер блока)
  - $b$  = общее количество блоков в эксперименте
  - $r$  = количество повторностей каждого варианта опыта
  - $\lambda$  = количество раз, когда варианты опыта встречаются вместе в одном блоке
  - $N$  = общее количество экспериментальных единиц
- $\lambda$  должно быть целым числом;  $N = b \cdot k = r \cdot t$
- Пример:  $t = 10$  вариантов опыта с размером блока  $k = 4$   
 $r = 6, b = 15, \lambda = 2$   
-> Следовательно,  $N = b \cdot k = r \cdot t = 6 \cdot 10 = 60$

В селекции растений минимальное количество повторностей, необходимое для достижения «баланса», часто слишком велико, чтобы это было возможным реализовать на практике

# Частично сбалансированный неполный блочный дизайн / Partially Balanced Incomplete Block Designs

- Различные пары вариантов опыта встречаются в одних и тех же блоках неодинаковое количество раз или некоторые пары вариантов опыта никогда не встречаются вместе в одном блоке.
  - Статистический анализ более сложен.
- Часто встречается в селекционных испытаниях из-за большого количества вариантов.

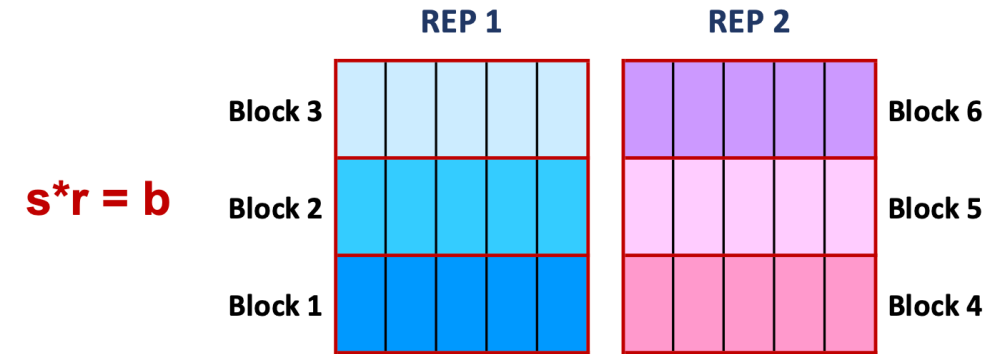
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		1	2	3
a	a	a	a	a	a	a	a	a	a	b	c	b	b	b		a	b	c
b	b	b	b	b	d	b	c	c	c	c	d	d	c	c		d	e	f
c	c	c	d	e	e	d	d	e	d	e	e	e	d	d		b	c	a
d	e	f	e	f	f	f	e	f	f	f	f	f	f	e		c	f	d
<b>Balanced</b>															<b>Unbalanced</b>			
Total Size (N) = 60															Total Size (N) = 12			
Treatments (t) = 6															Treatments (t) = 6			
Treatment Size (n) = 10															Treatment Size (n) = 2			
Replication (r) = 10															Replication (r) = 2			
Block Size (s) = 4															Block Size (s) = 4			
Blocks (b) = 15															Blocks (b) = 3			
<b>Balanced and Unbalanced Incomplete Block Designs: Constrained by Six Treatments and a Block Size of Four</b>																		

<https://pbgworks.org/index.php?q=node/1534>

# Resolvable Incomplete Block Designs

Блоки группируются таким образом, что каждая группа блоков составляет одну полную повторность варианта опыта:

- "блок" = неполный блок = "суб-блок"
- повторность = "суперблок"
- Минимизация риска потери данных: в случае потери данных в одном блоке, остальные блоки могут оставаться нетронутыми, что позволяет сохранить ценность и достоверность общего эксперимента.
- Требуется особый подход в рандомизации, для того, чтобы обеспечить появление пар вариантов опытов в одном и том же неполном подблоке с одинаковой частотой



$t$  (количество вариантов опыта) = 15

$k$  (количество единиц в блоке, размер блока) = 5

$b$  (общее количество блоков в эксперименте) = 6

$r$  (количество повторений каждого варианта опыта) = 2

$s$  (количество блоков в каждом полном повторении) = 3

**Resolvable (разрешимый):** Этот термин означает, что схема опыта может быть разделена на меньшие, независимые блоки или группы, которые можно анализировать отдельно. В контексте IBD, это означает, что эксперимент в рамках одной большой схемы можно разделить на более мелкие подгруппы без потери статистической эффективности.

# Lattice (решётчатые) дизайны являются известным типом **resolvable** неполных блочных дизайнов

- **Square Lattice (квадратная решетка):** количество вариантов  $v$  представляет собой полный квадрат, а размер блока  $k$  представляет собой квадратный корень из этого числа. Пример  $(v, k, r)=(100, 10, 2)$
- **Ractangular Lattice (прямоугольная решетка) :** *Количество вариантов  $v$  равно  $k^*(k+1)$ , размещенных в блоках размером  $k$  единиц.* Этот тип дизайна является полезным дополнением к square lattice. Пример  $(v, k, r)=(56, 7, 2)$
- **Alpha Lattices (Альфа-дизайн):** является одной из самых гибких экспериментальных схем в селекции растений и сельском хозяйстве. Так же известен как обобщенная решетка (generalized lattice). Пример  $(v, k, r)=(80, 5, 3)$



# Альфа-дизайн

Пример:

$(v, k, r) = (80, 5, 3)$

Число вариантов  $v=80$

(число сортов)

Размер неполного блока

$k=5$  вариантов

Число повторностей  $= 3$

Паттерсон и Уильямс  
(1976) описали новый  
способ построения  
циклических,  
разрешимых неполных  
блочных дизайнов.

Alpha Lattice Design Field Layout 15X16

ROWS	48	79	19	69	60	65	51	17	73	64	56	38	46	5	32	74
	70	61	71	10	23	41	67	26	13	8	20	11	75	28	3	50
	16	49	59	53	18	30	68	36	44	25	37	34	39	14	4	77
	31	76	63	35	40	12	72	24	42	33	66	43	58	7	54	57
	80	29	6	9	15	55	22	2	52	62	45	1	21	27	78	47
	67	19	13	14	55	24	33	26	35	46	73	6	11	53	71	9
	70	44	16	74	32	47	61	66	1	15	17	42	39	56	48	52
	65	57	68	64	21	63	72	23	75	31	29	43	20	50	49	38
	54	76	10	51	7	34	69	30	12	59	58	41	27	80	62	2
	77	35	18	36	79	40	37	8	45	3	22	78	60	5	28	4
	67	45	3	17	52	6	49	18	54	44	56	33	9	30	28	43
	1	32	11	71	48	2	41	4	73	37	64	59	47	29	78	23
	26	16	68	53	51	80	46	61	63	34	79	13	20	31	66	69
	35	36	8	65	40	39	10	12	5	70	15	77	55	14	57	58
	7	19	76	60	75	72	24	50	35	21	42	27	62	38	22	74
COLUMNS																

# Как провести рандомизацию?

```
library(FielDHub)  
run_app(FielDHub)
```

<https://github.com/DidierMurilloF/FielDHub>

Что внутри?

[https://didiermurillof.github.io/FielDHub/articles/alpha\\_lattice.html#print-alpha-object](https://didiermurillof.github.io/FielDHub/articles/alpha_lattice.html#print-alpha-object)

# Линейная модель для lattice дизайнов

$$Y_{ijl} = \mu + \tau_i + \gamma_j + \rho_{l(j)} + \varepsilon_{ijl}$$

Варианты опыта (генотипы)

$i = 1, 2, \dots, t$

Повторности

$j = 1, 2, \dots, r$

Неполные блоки

$l = 1, 2, \dots, s$

Случайная ошибка

# ANOVA

Можно рассчитать два показателя ошибки:

- $E_b$  – Ошибка блока =  $SSB/r(k-1)$
- $E_e$  – Экспериментальная ошибка =  $SSE/((k-1)(rk-k-1))$

Источник	df	SS	MS
Итог	$rk^2-1$	SSTot	
Повторности	$r-1$	SSR	
Варианты опыта	$k^2-1$	SST	
Блок	$r(k-1)$	SSB	$E_b$
Эксперимент. ош.	$(k-1)(rk-k-1)$	SSE	$E_e$

Unadj ( модель не учитывает влияние других факторов)

Adj (модель учитывает различия, которые могут возникнуть, например, из-за специфических условий для каждого блока).

# Пример

Данные: <https://kwstat.github.io/agridat/reference/burgueno.alpha.html>

## Format

A data frame with 48 observations on the following 6 variables.

<b>rep</b>	rep, 3 levels
<b>block</b>	block, 12 levels
<b>row</b>	row
<b>col</b>	column
<b>gen</b>	genotype, 16 levels
<b>yield</b>	yield

## Details

A field experiment with 3 reps, 4 blocks per rep, laid out as an alpha design.

The plot size is not given.

Electronic version of the data obtained from CropStat software.

Used with permission of Juan Burgueno.

k=4 – размер неполного блока

# Вариант 1

```
> modelPBIB <- with(dat, PBIB.test(block, gen, rep, yield, k=4, group=TRUE, console=TRUE))
```

ANALYSIS PBIB: yield

Class level information

block : 12

gen : 16

Number of observations: 48

Стоит в модели PBIB по умолчанию

Estimation Method: Residual (restricted) maximum likelihood

Parameter Estimates

Variance

block:rep 86895.65

rep 200856.45

Residual 133217.38

← Только для случайных факторов (нет генотипов)

← Наиболее важный источник дисперсии (rep)

Fit Statistics

AIC 539.6454

BIC 567.4944

-2 Res Log Likelihood -250.8227

## Analysis of the Partially Balanced Incomplete Block Design

### Description

Analysis of variance PBIB and comparison mean adjusted. Applied to resolvable designs: Lattices and alpha design.

### Usage

```
PBIB.test(block, trt, replication, y, k, method=c("REML", "ML", "VC"),  
test = c("lsd", "tukey"), alpha=0.05, console=FALSE, group=TRUE)
```

Если экспериментатор хочет повысить мощность эксперимента - то есть уменьшить вероятность ошибки второго рода (не обнаружить эффект, когда он на самом деле есть) - ему следует увеличить количество повторений (reps), а не блоков внутри повторений.

## Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gen	15	5496795	366453	2.7508	0.01663 *
Residuals	21	2797565	133217		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Coefficient of variation: 27.4 %

yield Means: 1330.854

Генетика имеет  
значение!

Comparison test lsd

Treatments with the same letter are not significantly different.

	yield.adj	groups
G11	1946.4414	a
G15	1850.4626	a
G05	1771.3774	ab
G10	1685.1556	abc
G04	1607.2719	abc
G12	1556.0055	abcd
G03	1401.8365	abcde
G16	1361.7628	abcde
G09	1315.2203	abcde
G14	1313.4362	abcde
G13	1156.1650	bcdef
G08	1065.7424	cdef
G06	938.9526	def
G01	907.9782	def
G07	792.1634	ef
G02	623.6949	f

Сравнение  
скорректированных  
средних значений



# Из нашей модели можем "экстрагировать" дополнительную информацию – например, скорректированные средние

```
> modelPBIB$means
```

	yield	yield.adj	SE	r	std	Min	Max	Q25	Q50	Q75
G01	993.3333	907.9782	356.6432	3	503.8594	419	1361	809.5	1200	1280.5
G02	698.3333	623.6949	355.8377	3	619.2999	228	1400	347.5	467	933.5
G03	1488.0000	1401.8365	356.6432	3	405.7043	1058	1864	1300.0	1542	1703.0
G04	1818.6667	1607.2719	355.8377	3	309.5421	1481	2089	1683.5	1886	1987.5
G05	1864.6667	1771.3774	355.8377	3	1164.3386	867	3144	1225.0	1583	2363.5
G06	1096.6667	938.9526	356.6432	3	418.1391	767	1567	861.5	956	1261.5
G07	673.3333	792.1634	355.8377	3	293.1149	453	1006	507.0	561	783.5
G08	1049.3333	1065.7424	355.8377	3	399.0292	625	1417	865.5	1106	1261.5
G09	1279.6667	1315.2203	356.6432	3	163.0654	1125	1450	1194.5	1264	1357.0
G10	1609.0000	1685.1556	355.8377	3	687.5551	847	2183	1322.0	1797	1990.0
G11	1806.6667	1946.4414	356.6432	3	856.1053	1086	2753	1333.5	1581	2167.0
G12	1477.0000	1556.0055	356.6432	3	875.1280	500	2189	1121.0	1742	1965.5
G13	1020.3333	1156.1650	355.8377	3	712.4987	336	1758	651.5	967	1362.5
G14	1209.3333	1313.4362	356.6432	3	209.9389	967	1336	1146.0	1325	1330.5
G15	1879.6667	1850.4626	356.6432	3	574.1301	1447	2531	1554.0	1661	2096.0
G16	1329.6667	1361.7628	355.8377	3	1113.1677	383	2556	716.5	1050	1803.0

# Вариант 2

```
> modelPBIB2 <- with(dat, PBIB.test(block, gen, rep, yield, k=4, console = TRUE, method =c("VC"),test = "lsd", alpha = 0.05, group = TRUE))
```

Estimation Method: Variances component model

## Fit Statistics

AIC 719.5941  
BIC 771.9877

## Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
<u>rep</u>	2	7389007	3694504	<u>27.3490</u>	1.422e-06	***
gen.unadj	15	7104523	473635	3.5061	0.004373	**
block/rep	9	3358291	373143	2.7622	0.026438	*
Residual	21	2836838	135088			

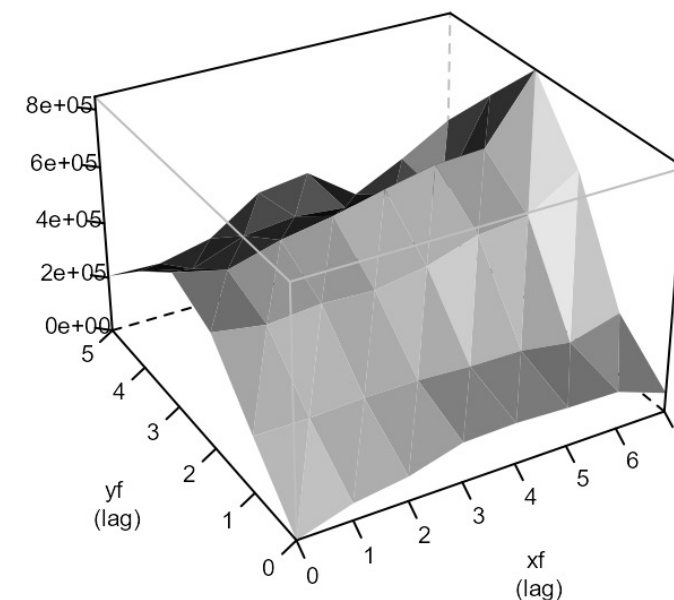
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Coefficient of variation: 27.6 %

yield Means: 1330.854

## burgueno.alpha



## Comparison test lsd

Treatments with the same letter are not significantly different.

	yield.adj	groups
G11	1947.0775	a
G15	1850.2966	ab
G05	1770.9804	ab
G10	1685.5600	abc
G04	1606.2940	abcd
G12	1556.4111	abcde
G03	1401.4580	abcdef
G16	1361.8316	abcdef
G09	1315.4165	abcdef
G14	1313.9578	abcdef
G13	1156.7639	bcdefg
G08	1065.7947	cdefg
G06	938.1734	defg
G01	907.5425	efg
G07	792.6846	fg
G02	623.4241	g

> modelPBIB2\$means

	yield	yield.adj	SE	r	std	Min	Max	Q25	Q50	Q75
G01	993.3333	907.5425	237.6623	3	503.8594	419	1361	809.5	1200	1280.5
G02	698.3333	623.4241	236.4227	3	619.2999	228	1400	347.5	467	933.5
G03	1488.0000	1401.4580	237.6623	3	405.7043	1058	1864	1300.0	1542	1703.0
G04	1818.6667	1606.2940	236.4227	3	309.5421	1481	2089	1683.5	1886	1987.5
G05	1864.6667	1770.9804	236.4227	3	1164.3386	867	3144	1225.0	1583	2363.5
G06	1096.6667	938.1734	237.6623	3	418.1391	767	1567	861.5	956	1261.5
G07	673.3333	792.6846	236.4227	3	293.1149	453	1006	507.0	561	783.5
G08	1049.3333	1065.7947	236.4227	3	399.0292	625	1417	865.5	1106	1261.5
G09	1279.6667	1315.4165	237.6623	3	163.0654	1125	1450	1194.5	1264	1357.0
G10	1609.0000	1685.5600	236.4227	3	687.5551	847	2183	1322.0	1797	1990.0
G11	1806.6667	1947.0775	237.6623	3	856.1053	1086	2753	1333.5	1581	2167.0
G12	1477.0000	1556.4111	237.6623	3	875.1280	500	2189	1121.0	1742	1965.5
G13	1020.3333	1156.7639	236.4227	3	712.4987	336	1758	651.5	967	1362.5
G14	1209.3333	1313.9578	237.6623	3	209.9389	967	1336	1146.0	1325	1330.5
G15	1879.6667	1850.2966	237.6623	3	574.1301	1447	2531	1554.0	1661	2096.0
G16	1329.6667	1361.8316	236.4227	3	1113.1677	383	2556	716.5	1050	1803.0

> |

# А если не корректировать средние значения?

```
> m1 <- lm(yield ~ 0 + gen + rep + rep:block, dat)
```

```
> anova(m1)
```

Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gen	16	92120818	5757551	42.6209	2.016e-12 ***
rep	2	7389007	3694504	27.3490	1.422e-06 ***
rep:block	9	3358291	373143	2.7622	0.02644 *
Residuals	21	2836838	135088		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

\$groups

	yield	groups
G15	1879.6667	a
G05	1864.6667	a
G04	1818.6667	ab
G11	1806.6667	ab
G10	1609.0000	abc
G03	1488.0000	abc
G12	1477.0000	abc
G16	1329.6667	abc
G09	1279.6667	abcd
G14	1209.3333	bcd
G06	1096.6667	cd
G08	1049.3333	cd
G13	1020.3333	cd
G01	993.3333	cd
G02	698.3333	d
G07	673.3333	d

С варианта 2

Comparison test lsd

Treatments with the same letter are not significantly

	yield.adj	groups
G11	1947.0775	a
G15	1850.2966	ab
G05	1770.9804	ab
G10	1685.5600	abc
G04	1606.2940	abcd
G12	1556.4111	abcde
G03	1401.4580	abcdef
G16	1361.8316	abcdef
G09	1315.4165	abcdef
G14	1313.9578	abcdef
G13	1156.7639	bcdefg
G08	1065.7947	cdefg
G06	938.1734	defg
G01	907.5425	efg
G07	792.6846	fg
G02	623.4241	g

**Без использования повторностей – augmented design**

# Особенности

- Впервые представлен в публикации Federer, Walter T. (1956)\*
- **Контрольные образцы (эталонные сорта)** повторяются в соответствии со стандартными правилами дизайна эксперимента
- **Новые варианты опыта (генотипы)** не повторяются или представлены в меньшем количестве повторностей, чем контрольные образцы – они «дополняют» стандартный дизайн.

8	5	112	49	119	109	6	26	28	48	59	3	70	47	4	23	25
7	17	82	20	95	58	22	16	C1	92	88	117	12	107	105	15	100
6	38	C2	108	93	56	39	114	62	106	64	2	52	89	C1	110	31
5	103	19	72	86	18	33	101	75	87	32	120	60	71	40	C2	11
4	98	7	54	66	65	34	51	104	35	81	C2	96	14	37	63	50
3	113	53	43	68	42	1	29	C1	116	C1	85	46	73	30	36	84
2	77	74	57	C2	90	97	115	118	76	80	83	24	13	78	91	99
1	45	94	9	102	10	111	79	69	8	27	61	67	55	44	41	21
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

\*<https://ecommons.cornell.edu/items/22e171a6-71d7-422a-9de5-9d7cc372ccd0>

# Когда используется в селекции растений

## **Ранние поколения:**

- Ограниченное количество семян
- Ограничены земельные (и другие) ресурсы при этом
  - нужно оценить как можно больше генотипов
  - нужно провести переоценку отобранных вариантов в последующие сезоны
- Сложно поддерживать однородность блоков при сравнении очень большого количества генотипов
  - Необходим механизм для корректировки неоднородности поля

## **Адаптация к разнообразным условиям окружающей среды является основной целью программы:**

- Совместные селекционные программы:
  - Участники совместных селекционных программ (например, сельхозтоваропроизводители) будут выращивать по одной повторности
  - Может быть невозможно разместить все варианты
- Исследование производственных систем:
  - Необходимо оценить перспективные генотипы в максимально возможном количестве разнообразных условий окружающей среды

# Преимущества

- Нереплицированные дизайны могут эффективно использовать ограниченные ресурсы:
  - Позволяют оценить больше генотипов
  - Протестировать в большем количестве условий окружающей среды
- При таком подходе требуется меньше контрольных участков, чем для дизайнов с систематическим повторением одного и того же контроля.
- Обеспечивает оценку стандартной ошибки, которая может быть использована для сравнений:
  - Между новыми генотипами
  - Между новыми генотипами и контрольными образцами
- Наблюдения за новыми генотипами могут быть скорректированы с учетом неоднородности поля (размещение в блоках)
- Гибкость – блоки могут быть разного размера



# Недостатки

- На производство и обработку контрольных участков тратятся значительные ресурсы.
- Относительно **немного степеней свободы для экспериментальной ошибки**, что снижает способность обнаруживать различия между вариантами опыта.
- Нереплицированные эксперименты по своей природе не такие точные, как с репликацией, независимо от сложности дизайна.

# Варианты размещения образцов

Можно выбрать «базовый» способ размещения контрольных образцов на экспериментальные единицы.

- Однонаправленное размещение в блоках:
  - RCBD
  - Неполные блочные дизайны (например, Lattice Design)
- Двухнаправленное размещение в блоках:
  - Латинский квадрат (Полные блоки)
  - Квадрат Юдена (Неполные блоки)
  - Row-column дизайны
- Пример:
  - Контрольные образцы присутствуют один раз в каждом блоке
  - Новые варианты присутствуют один раз в эксперименте (одна повторность)
- Такой дизайн можно назвать **ARCB**D (Дополненный Рандомизированный Полный Блочный Дизайн)

# Пример эксперимента

- Семена разные селекционных популяций были получены от селекционера, который ушел на пенсию
- Популяции были восстановлены в теплице и самоопылены
- S2 линии были пересажены в поле, где они были скрещены между собой (outcross)
- Недостаточно семян для проведения рандомизированных испытаний полученного потомства.
- **Цель** - создать широкую базу для рекуррентного отбора:
  - скрининг S2-тесткроссов
  - рекомбинация выбранных S2 родителей.



# Пример эксперимента

- Расположение блоков в одном направлении
- 3 контрольных сорта (с=3)
- 6 блоков (r=6)
- 50 новых образцов (n=50) – 50 семейств S2-тесткроссов
- Количество делянок =  $n + rc = 50 + 63 = 68$
- Количество делянок на блок =  $68/6 = 11,3 \rightarrow 12$  (с=3, n=9)
- В последнем блоке только 8 делянок (с=3, n=5)
  - Неравные количества новых образцов на блок допустимы.

Block 6												<div>168 17</div> <div>167 33</div> <div>166 26</div> <div>165 C2</div> <div>164 46</div> <div>163 25</div> <div>162 C1</div> <div>161 C3</div>
Block 5	<div>149 6</div> <div>150 12</div> <div>151 C1</div> <div>152 43</div> <div>153 39</div> <div>154 29</div> <div>155 38</div> <div>156 37</div> <div>157 C3</div> <div>158 10</div> <div>159 21</div> <div>160 C2</div>											
Block 4	<div>148 C3</div> <div>147 1</div> <div>146 23</div> <div>145 15</div> <div>144 C2</div> <div>143 40</div> <div>142 48</div> <div>141 C1</div> <div>140 22</div> <div>139 16</div> <div>138 32</div> <div>137 47</div>											
Block 3	<div>125 24</div> <div>126 34</div> <div>127 50</div> <div>128 49</div> <div>129 4</div> <div>130 C1</div> <div>131 20</div> <div>132 2</div> <div>133 35</div> <div>134 19</div> <div>135 C2</div> <div>136 C3</div>											
Block 2	<div>124 C1</div> <div>123 11</div> <div>122 C3</div> <div>121 7</div> <div>120 13</div> <div>119 5</div> <div>118 30</div> <div>117 44</div> <div>116 36</div> <div>115 C2</div> <div>114 3</div> <div>113 31</div>											
Block 1	<div>101 8</div> <div>102 45</div> <div>103 C2</div> <div>104 9</div> <div>105 27</div> <div>106 28</div> <div>107 C1</div> <div>108 42</div> <div>109 18</div> <div>110 14</div> <div>111 41</div> <div>112 C3</div>											

# Как составить план эксперимента?

- Необходимо включить достаточное количество контрольных образцов и их повторностей, чтобы обеспечить хорошую оценку экспериментальной ошибки и адекватную мощность для выявления различий между сортами.
- **Рекомендация по количеству делянок контрольных образцов: 5–10% от всех вариантов**
- Расположите блоки **вдоль градиента поля**, чтобы максимизировать вариацию между блоками и минимизировать вариацию внутри блоков.
- Произвольно назначьте каждый из контрольных образцов каждому блоку.
- Произвольно назначьте варианты опыта оставшимся делянкам
- Или используйте инструмент для рандомизации: FielDHub 😊  
(<https://github.com/DidierMurilloF/FielDHub>)

# Линейная модель

$$Y_{ij} = \mu + \beta_i + c_j + \tau_{k(i)} + \varepsilon_{ij}$$

среднее + блоки + контрольные образцы + новые образцы + ошибка

## А будет ли таблица ANOVA? (нет)

**Для анализа таких сложных схем и используются линейные смешанные модели**

- По сравнению с традиционной фиксированной линейной моделью (ANOVA), линейные смешанные модели имеют ряд преимуществ:
- Они позволяют одновременно оценивать фиксированные эффекты и прогнозировать случайные эффекты.
- Могут обрабатывать несбалансированные (неполные) данные, когда не все генотипы тестируются во всех местах/годах или тестируются с разным числом повторов.

# Фиксированные vs случайные эффекты

- Уровни фактора выбираются намеренно, потому что нас интересуют именно эти уровни (и только они!)

**Пример:** мы занимались селекцией на устойчивость к корневым болезням и в результате разработали 25 различных потенциально новых сортов. Какой (какие) из них мы должны выпустить в качестве коммерческого сорта?

- Выбираем эти 25 генотипов, которые нас интересуют, и тестируем их на устойчивость.
- **"Генотип" является фиксированным фактором**

- Уровни фактора случайным образом выбираются из множества возможных уровней - нас не интересуют конкретные уровни. Задача исследователя заключается в том, чтобы распространить выводы, основанные на выборке, на ВСЕХ представителей этой популяции.

**Пример:** мы хотим определить, насколько большая разница (вариация) среди сортов, показывающих устойчивость к корневым болезням, для борьбы с которыми мы проводили отбор. Мы берем случайную выборку из 25 сортов из всех существующих устойчивых сортов, которая отражает истинное среднее значение популяции и может быть использована для оценки дисперсии популяции.

- **"Генотип" является случайным фактором**

# Фиксированные vs случайные эффекты

## Различные цели

- Фиксированные: цель состоит в том, чтобы сравнить конкретные варианты и проверить гипотезу о том, что эффекты вариантов одинаковы.
- Случайные: цель состоит в том, чтобы оценить компоненты дисперсии.

## Процедуры выборки различны

- Фиксированные: варианты выбираются преднамеренно; если эксперимент повторяется, варианты должны быть теми же, меняется только ошибка
- Случайные: варианты выбираются случайно; дисперсия в популяции обработок вносит вклад в общую сумму квадратов
  - эксперимент может быть повторен с другой выборкой
  - эффекты вариантов и ошибки могут измениться.



# Фиксированные или случайные эффекты в augmented design

Обычно предполагается, **что блоки будут случайным эффектом**

- они представляют собой часть большего набора потенциальных блоков
- мы хотим сделать выводы, выходящие за рамки конкретной выборки блоков в эксперименте.

## Генотипы

- Контрольные образцы считаются **фиксированными эффектами**, поскольку мы хотим сравнить их с конкретными сортами (новые варианты).
- Новые варианты (генотипы, которые испытываем) могут рассматриваться как фиксированные или как случайные эффекты – *возможно два варианта анализа.*
  - В данном контексте они могут быть фиксированными, так как цель - выбрать наилучшие.
  - Однако чаще они рассматриваются как случайные, особенно в генетических исследованиях.

# Анализ смешанных линейных моделей в R

- `install.packages("lme4")`
- `install.packages("lmerTest")`
- `install.packages("Matrix")`

# Что там с популяциями, которые достались от селекционера?

Были собраны данные веса 1000 семян (TSW) для всех 50 новых образцов и 3 вариантов контроля

Entry – все варианты опыта (новые образы + контроль)

Для модели, где "Entry" рассматривается как фиксированный эффект:

- предположим, мы хотим найти взаимосвязь между TSW, блоком и Entry. Эта модель подразумевает, что мы рассматриваем "Entry" как фиксированный эффект и "Block" как случайный эффект. Таким образом, нашу модель можно представить следующим образом:
- **$TSW \sim \text{Entry} + (1|\text{Block}) + \text{error}$**  (для R уберем error)
- Ошибка представляет отклонения модели, которые мы не можем контролировать или моделировать экспериментально.

```

> ad.entry = lmer(TSW ~ Entry +(1|Block), data=ad)
> anovaOut <- anova(ad.entry)
> print (anovaOut)
Type III Analysis of Variance Table with Satterthwaite's method
      Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
Entry 27.243  0.52391     52 10.181  7.5053 0.0007433 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
> modelSum <- summary(ad.entry,correlation=TRUE)
> print (modelSum)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: TSW ~ Entry + (1 | Block)
Data: ad
```

REML criterion at convergence: 17.7

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.205	0.000	0.000	0.000	1.628

Random effects:

Groups	Name	Variance	Std.Dev.
Block	(Intercept)	0.13808	0.3716
	Residual	0.06981	0.2642

Number of obs: 68, groups: Block, 6

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	10.55788	0.33078	14.74497	31.918	5.08e-15	***
Entry2	-0.97410	0.42360	10.75971	-2.300	0.042557	*
Entry3	0.04870	0.42360	10.75971	0.115	0.910583	
Entry4	-0.52411	0.42360	10.75971	-1.237	0.242311	
Entry5	0.14870	0.42360	10.75971	0.351	0.732331	
Entry6	-0.86612	0.42360	10.75971	-2.045	0.066143	.
Entry7	-1.09130	0.42360	10.75971	-2.576	0.026168	*
Entry8	-1.44433	0.42360	10.75971	-3.410	0.006010	**
Entry9	-1.67433	0.42360	10.75971	-3.953	0.002360	**
Entry10	-0.41612	0.42360	10.75971	-0.982	0.347508	
Entry11	1.07870	0.42360	10.75971	2.547	0.027579	*
Entry12	0.06388	0.42360	10.75971	0.151	0.882915	
Entry13	-0.06130	0.42360	10.75971	-0.145	0.887619	
Entry14	0.27567	0.42360	10.75971	0.651	0.528836	
Entry15	-0.88000	0.37365	10.00000	-2.355	0.040284	*
Entry16	-0.37000	0.37365	10.00000	-0.990	0.345410	
Entry17	-1.94141	0.42360	10.75971	-4.583	0.000832	***