

## Блочный рандомизированный эксперимент

[Ранее](#) был рассмотрен  $F$ -критерий однофакторного дисперсионного анализа для оценки разностей между математическими ожиданиями с групп. Этот критерий применяется в ситуациях, когда  $n$  однородных элементов (так называемых экспериментальных объектов) случайным образом распределяются по  $s$  уровням исследуемого фактора (так называемые группы условий факторного эксперимента). Такие эксперименты называются полностью рандомизированными.<sup>1</sup>

Кроме того, в заметке [Проверка гипотез: двухвыборочные критерии](#) был описан  $t$ -критерий для оценки разностей между математическими ожиданиями, который используется в ситуациях, связанных с повторяющимися измерениями или согласованными выборками. Этот критерий позволяет оценить различия между условиями проведения двух экспериментов. Предположим, что мы исследуем несколько групп условий или уровней исследуемого фактора. В таких ситуациях совокупности неоднородных объектов или индивидуумов, подлежащих сравнению (или повторным измерениям), называются блоками. Допустим, мы получили числовые результаты измерений для каждой группы условий и комбинаций блоков.

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

Для сравнения полностью рандомизированных и блочных рандомизированных экспериментов вспомним сценарий, посвященный компании Perfect Parachute Company («Идеальный парашют»), использованный в [предыдущей](#) заметке. Парашюты изготавливаются из синтетических волокон, поставляемых четырьмя разными поставщиками. На фабрике используется два вида ткацких станков: Jetta и Turk. Одной из основных характеристик парашюта является его прочность. Предположим, что в полностью рандомизированном эксперименте используются 12 наблюдений — по одному виду ткани на каждую из 12 смен. Любая изменчивость результатов испытаний становится частью экспериментальной ошибки, и, следовательно, различия между четырьмя поставщиками труднее уловить. Для того чтобы уменьшить экспериментальную ошибку, разработаем блочный рандомизированный эксперимент, в котором исследуются три смены, в течение каждой из которых ткется четыре парашюта (один парашют — из волокон, полученных от первого поставщика, второй — из волокон, полученных от второго поставщика, и т.д.). Три смены рассматриваются как блоки, а условием факторного эксперимента является поставщик. Преимущество блочного рандомизированного эксперимента заключается в том, что из экспериментальной ошибки исключается изменчивость между тремя сменами. Следовательно, этот эксперимент часто обеспечивает более точные оценки различий между четырьмя поставщиками.

### Критерии для оценки эффектов условий факторного эксперимента и блоков

Напомним, что в полностью рандомизированном эксперименте полная вариация (SST) подразделяется на межгрупповую (SSA) и внутригрупповую (SSW). Внутригрупповая вариация считается экспериментальной ошибкой, а межгрупповая вариация возникает вследствие различий между условиями факторного эксперимента. Для того чтобы отделить эффект блокировки от экспериментальной ошибки блочного рандомизированного эксперимента, необходимо подразделить внутригрупповую вариацию на межблочную вариацию (SSBL) и случайную ошибку (SSE) (рис. 1). Следовательно, в блочном рандомизированном эксперименте полная вариация результатов измерений представляет собой сумму межгрупповой вариации (SSA), межблочной вариации (SSBL) и случайной ошибки (SSE).

---

<sup>1</sup> Используются материалы книги Левин и др. Статистика для менеджеров. — М.: Вильямс, 2004. — с. 681–693

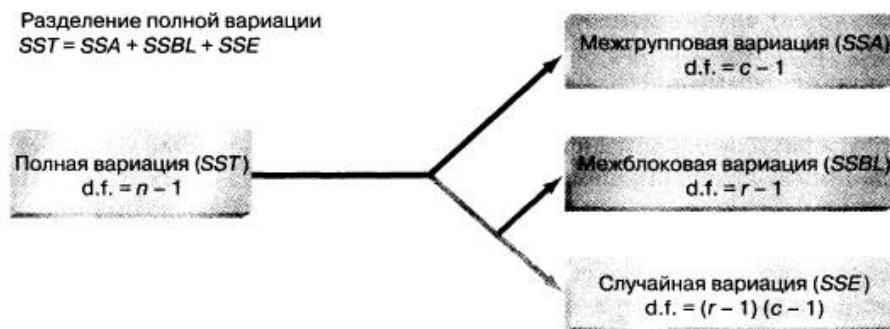


Рис. 1. Разделение полной вариации в блочном рандомизированном эксперименте

Для того чтобы разработать процедуру дисперсионного анализа для блочного рандомизированного эксперимента, введем следующие обозначения:  $r$  — количество блоков,  $c$  — количество групп или уровней фактора,  $n$  — общее количество наблюдений ( $n = rc$ ),  $X_{ij}$  — величина в  $i$ -м блоке и  $j$ -группе,  $\bar{X}_i$  — среднее всех величин из  $i$ -го блока,  $\bar{X}_j$  — среднее всех величин из  $j$ -й группы. Основные параметры блочного рандомизированного эксперимента представлены на рис. 2.

Вид величины	Количество степеней свободы	Суммы квадратов	Дисперсии	F-статистика
Межгрупповая	$c-1$	SSA	$MSA=SSA/(c-1)$	$F=MSA/MSW$
Межблоковая	$r-1$	SSBL	$MSBL=SSBL/(r-1)$	$F=MSBL/MSE$
Ошибка	$(r-1)(c-1)$	SSE	$MSE=SSE/(r-1)(c-1)$	
Полная	$rc-1$	SST		

Рис. 2. Сводная таблица дисперсионного анализа для блочного рандомизированного эксперимента

Проиллюстрируем блочный рандомизированный эксперимент следующим примером. Предположим, что сеть ресторанов быстрого питания, имеющая четыре подразделения в определенном географическом регионе, желает оценить качество обслуживания в этих ресторанах. Для этой цели директор нанял шесть экспертов, имеющих разный опыт. Чтобы уменьшить эффект вариации между экспертами, был разработан блочный рандомизированный эксперимент, в которых блоками считались эксперты. В свою очередь, четыре ресторана образовали группы условий факторного эксперимента. Шесть экспертов в случайном порядке инспектировали каждый из четырех ресторанов. Для оценки использовалась шкала баллов от 0 (низшая оценка) до 100 (рис. 3).

	A	B	C	D	E	F	G
1	Рестораны						
2	Эксперты	A	Б	В	Г	Всего	Средние
3	1	70	61	82	74	287	71,75
4	2	77	75	88	76	316	79,00
5	3	76	67	90	80	313	78,25
6	4	80	63	96	76	315	78,75
7	5	84	66	92	84	326	81,50
8	6	78	68	98	86	330	82,50
9	Всего	465	400	546	476	1887	
10	Средние	77,50	66,67	91,00	79,33	78,63	
11							

Рис. 3. Рейтинги четырех ресторанов быстрого питания

Как следует из рис. 3,  $r = 6$ ,  $c = 4$ ,  $n = rc = 24$ . Для проведения анализа перейдите по меню *Данные* → *Анализ данных* и в открывшемся окне выберите строку *Двухфакторный дисперсионный анализ без повторений* (рис. 4). Результаты анализа представлены на рис. 5.

	A	B	C	D	E
1	Рестораны				
2	Эксперты	A	Б	В	Г
3	1	70	61	82	74
4	2	77	75	88	76
5	3	76	67	90	80
6	4	80	63	96	76
7	5	84	66	92	84
8	6	78	68	98	86
9	Всего	465	400	546	476
10	Средние	77,50	66,67	91,00	79,33

**Двухфакторный дисперсионный анализ без повторений**

Входные данные  
 Входной интервал:

Метки  
 Альфа:

Параметры вывода  
 Выходной интервал:      
 Новый рабочий диск:  
 Новая рабочая книга

Рис. 4. Использование Пакета анализа для блочного рандомизированного эксперимента

Двухфакторный дисперсионный анализ без повторений						
ИТОГИ	Счет	Сумма	Среднее	Дисперсия		
Эксперт 1	4	287	71,75	76,25		
Эксперт 2	4	316	79	36,667		
Эксперт 3	4	313	78,25	90,917		
Эксперт 4	4	315	78,75	184,917		
Эксперт 5	4	326	81,5	121		
Эксперт 6	4	330	82,5	161		
Ресторан А	6	465	77,5	21,5		
Ресторан Б	6	400	66,667	23,467		
Ресторан В	6	546	91	33,2		
Ресторан Г	6	476	79,333	23,467		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Строки	283,375	5	56,675	3,782	0,020	2,901
Столбцы	1787,458	3	595,819	39,758	0,000	3,287
Погрешность	224,792	15	14,986			
Итого	2295,625	23				

Рис. 5. Результаты инспекции сети ресторанов быстрого питания, полученные с помощью Пакета анализа Excel

Если установить уровень значимости критерия для проверки гипотезы о существовании различий между подразделениями сети ресторанов равным 0,05, решающее правило примет следующий вид: нулевая гипотеза  $H_0 (\mu_1 = \mu_2 = \mu_3 = \mu_4)$  отклоняется, если  $F > 3,287$ . Число 3,287 представляет собой верхнее критическое значение  $F$ -распределения, имеющего три степени свободы в числителе и 15 степеней свободы в знаменателе (строка *Столбцы* колонка *F критическое* таблицы *Дисперсионный анализ* на рис. 5). Поскольку  $F = 39,758 > F_U = 3,29$  (кроме того,  $p = 0,000 < 0,05$ ), мы можем отклонить гипотезу  $H_0$  и утверждать, что средние рейтинги ресторанов статистически значимо различаются между собой. Чрезвычайно малое значение  $p$  означает, что, если бы средние рейтинги четырех

подразделений были одинаковыми, вероятность обнаружить разности между их выборочными средними, была бы крайне малой. Итак, нулевая гипотеза практически невероятна. Следовательно, альтернативную гипотезу можно считать корректной.

Для проверки эффективности блокировки, можно проверить разность между экспертами. При 5%-ном уровне значимости решающее правило можно сформулировать следующим образом: нулевая гипотеза  $H_0 (\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6)$  отклоняется, если вычисленная статистика  $F > 2,901$ . Число 2,901 представляет собой верхнее критическое значение  $F$ -распределения, имеющего пять степеней свободы в числителе и 15 степеней свободы в знаменателе. Поскольку  $F = 3,782 > F_U = 2,901$  (а  $p = 0,02 < 0,05$ ), мы можем отклонить гипотезу  $H_0$  и утверждать, что средние рейтинги ресторанов статистически значимо различаются между собой. Итак, применение блоков уменьшает экспериментальную ошибку.

Кроме обычных ограничений, принятых в однофакторном дисперсионном анализе, необходимо также предположить, что между условиями факторного эксперимента и блоками нет взаимодействия. Иначе говоря, необходимо, чтобы все различия между условиями эксперимента (ресторанами) были согласованы со всеми блоками (отмечались всеми экспертами). После разработки схемы блочного рандомизированного эксперимента и анализа данных о рейтингах ресторанов возникает вопрос: какой эффект оказывает блокирование на дисперсионный анализ? Иначе говоря, получаем ли мы более точные результаты, применяя блокирование при анализе разных групп условий факторного эксперимента? Для того чтобы ответить на этот вопрос, следует вычислить оценку относительной эффективности блочного рандомизированного эксперимента по сравнению с полностью рандомизированным экспериментом.

Оценка относительной эффективности

$$RE = \frac{(r-1)MSBL + r(c-1)MSE}{(rc-1)MSE}$$

Подставляя наши данные, получаем

$$RE = \frac{5 \times 56,675 + 6 \times 3 \times 14,986}{23 \times 14,986} = 1,60$$

Это означает, что для получения такой же точности при сравнении средних по группам в рамках однофакторного дисперсионного анализа нам понадобилось бы в 1,6 раза увеличить количество наблюдений в каждой группе.

### Множественные сравнения: процедура Тьюки

Как и в полностью рандомизированном эксперименте, отклонив нулевую гипотезу о равенстве всех средних по группам, мы можем определить, какая группа условий значительно отличается от остальных. Для блочного рандомизированного эксперимента такая процедура была разработана Джоном Тьюки. Критический размах в процедуре Тьюки вычисляется по формуле:

$$\text{Критический размах} = Q_U \sqrt{\frac{MSE}{r}}$$

где статистика  $Q_U$  представляет собой верхнее критическое значение распределения стьюдентизированного размаха, имеющего  $s$  степеней свободы в числителе и  $(r-1)(c-1)$  степеней свободы в знаменателе. К сожалению, для вычисления стьюдентизированного размаха нет функций в Excel (по крайней мере, я ее не нашел). Приходится пользоваться таблицами. Каждая из  $s(c-1)/2$  пар средних сравнивается с одним критическим размахом. Пара, например, группа  $j$ — группа  $j'$  объявляется статистически значимо разными, если модуль разности между выборочными средними  $|X_j - X_{j'}|$  превышает критический размах.

Продemonстрируем применение процедуры Тьюки на примере анализа сети ресторанов. Поскольку проверке подвергаются четыре ресторана, в процедуре Тьюки будет выполнено  $4(4-1)/2 = 6$  попарных сравнений (рис. 6).

	A	B	C	D	E	F	G
1	Фрагмент таблицы "Двухфакторный дисперсионный анализ без повторений"						
2							
3	<b>ИТОГИ</b>	<b>Счет</b>	<b>Сумма</b>	<b>Среднее</b>	<b>Дисперсия</b>		
4	Ресторан А	6	465	77,5	21,5		
5	Ресторан Б	6	400	66,667	23,467		
6	Ресторан В	6	546	91	33,2		
7	Ресторан Г	6	476	79,333	23,467		
8							
9	Попарные сравнения выборочных средних						
10	Пары ресторанов	$X_j$	$X_j$	$ X_j - X_j $			
11	А и Б	77,5	66,667	10,833			
12	А и В	77,5	91,000	13,500			
13	А и Г	77,5	79,333	1,833			
14	Б и В	66,667	91	24,333			
15	Б и Г	66,667	79,333	12,667			
16	В и Г	91	79,333	11,667			
17							

Рис. 6. Попарные сравнения средних значений по выборкам отдельных ресторанов

Для того чтобы вычислить критический размах, заметим (см. рис. 5):  $MSE = 14,986$  и  $r = 6$ . Для  $\alpha = 0,05$ ,  $c = 4$  и  $(r - 1)(c - 1) = 15$  по таблице (рис. 7) находим, что величина  $Q_U$  — верхнее критическое значение тестовой статистики, имеющей четыре степени свободы в числителе и 15 степеней свободы в знаменателе — равна 4,08.

Верхние 5% значений ( $\alpha=0,05$ )											
Числитель, df,											
Знаменатель, df,	2	3	4	5	6	7	8	9	10	11	12
15	3,01	3,67	4,08	4,37	4,60	4,78	4,94	5,08	5,20	5,31	5,40
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35
17	2,98	3,63	4,02	4,30	4,52	4,71	4,86	4,99	5,11	5,21	5,31
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20

Рис. 7. Критическое значение стьюдентизированного размаха  $Q_U$ ; к сожалению, в Excel нет функции, рассчитывающей такой размах

Таким образом:

$$\text{Критический размах} = Q_U \sqrt{\frac{MSE}{r}} = 4,08 \sqrt{\frac{14,986}{6}} = 6,448$$

В этой и двух предыдущих заметках (см. [Однофакторный дисперсионный анализ](#) и [Двухфакторный дисперсионный анализ](#)) описана концепция полностью рандомизированного эксперимента. Приведены различные процедуры анализа эффекта одного и двух факторов. Рассмотрен практический пример, иллюстрирующий применение критериев одно- и двухфакторного дисперсионного анализа. Детально описаны условия выполнения статистических процедур. Напомним, что проверка необходимых условий является неотъемлемой частью статистического анализа и позволяет правильно выбирать критерии. Как показано на структурной схеме (рис. 8), существует несколько подходов к сравнению групп числовых данных на основе экспериментов.

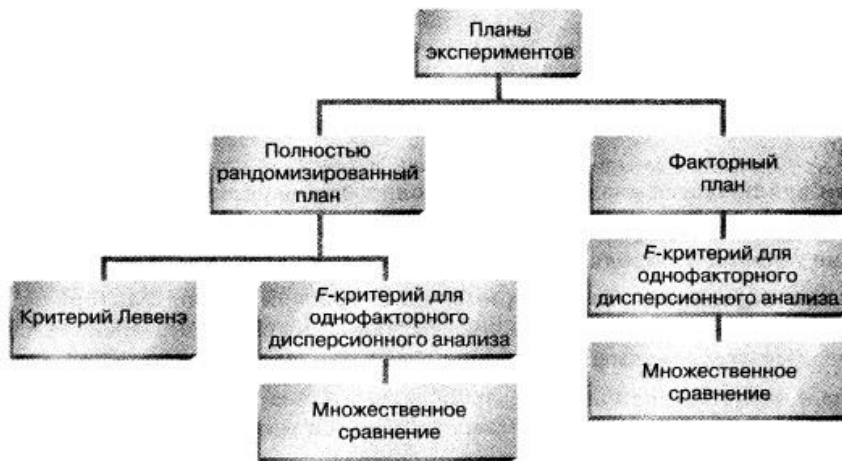


Рис. 8. Структурная схема трех заметок по теме дисперсионного анализа

Предыдущая заметка [Однофакторный дисперсионный анализ](#)

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel](#)