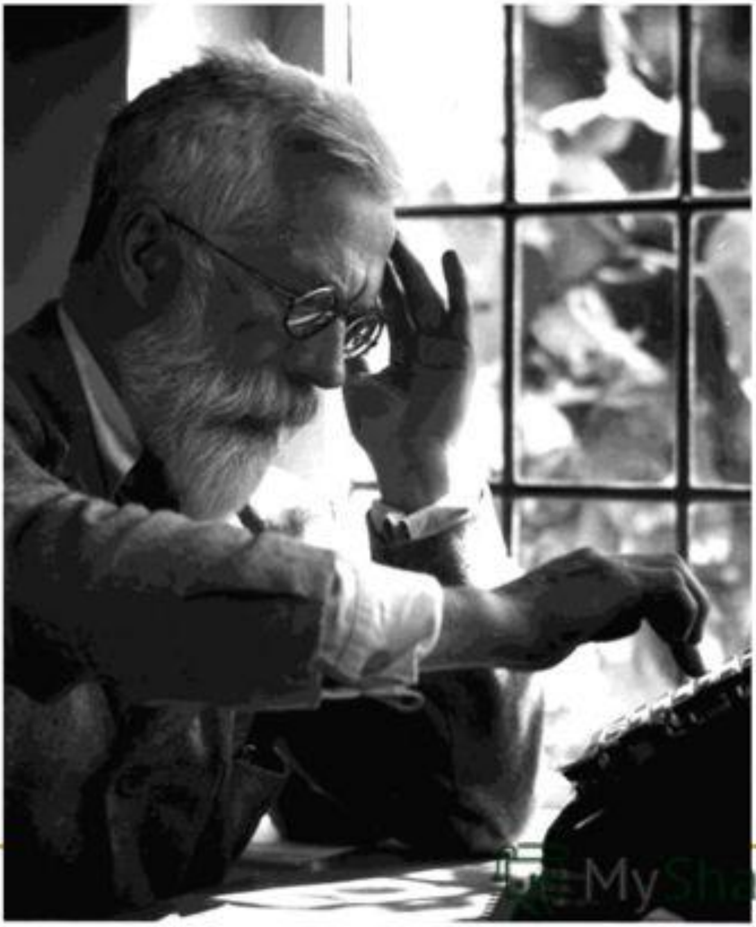
ANOVA

Сравнение **ДВУХ** И **БОЛЕЕ** групп

Дисперсионный анализ ANOVA (analysis of variance)



Sir Ronald Aylmer **FISHER**

Дисперсионный анализ

Рассмотренный ранее *t*-критерий-критерий Стьюдента (равно как и его непараметрические аналоги) предназначен для сравнения исключительно двух совокупностей. Однако часто он неверно используется для попарного сравнения большего количества групп (рис. 1), что вызывает т.н. *эффект множественных сравнений*

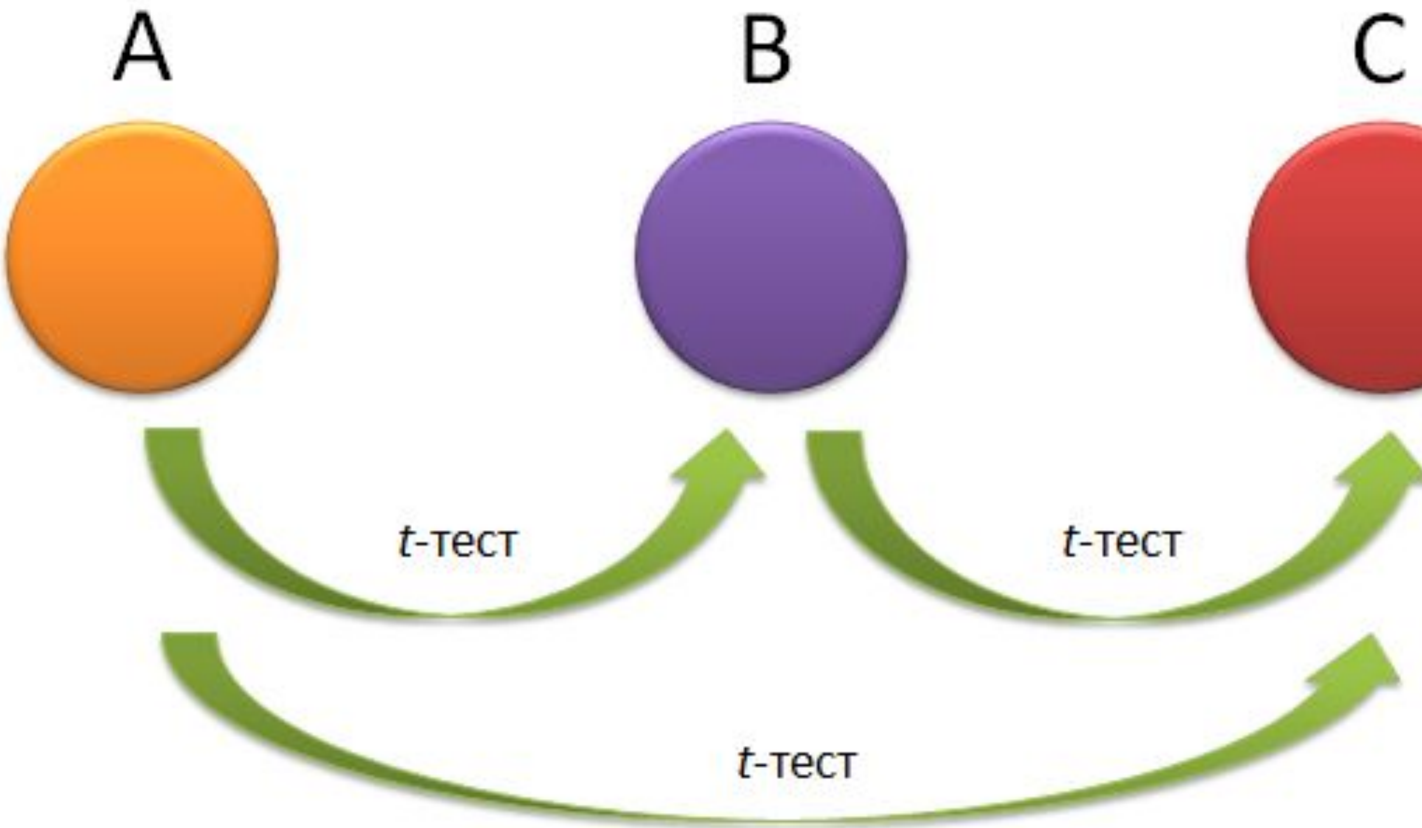


Рис. Пример неверного использования критерия Стьюдента для попарных сравнений трех групп - А, В и С.

Общие принципы дисперсионного анализа

- В 1920 г. английский математик **Рональд Фишер** предложил концепцию **дисперсионного анализа.**

от латинского Dispersio – рассеивание,
Analysis of variance (ANOVA) – анализ варианс

- Применяется для исследования влияния одной или нескольких качественных переменных (факторов) на одну зависимую количественную переменную (отклик).

Откуда произошло название **дисперсионный анализ**?

- При исследовании статистической значимости различия между средними двух (или нескольких) групп **сравниваются (анализируются) выборочные дисперсии**.

- Метод применялся для оценки экспериментов в растениеводстве.

- В дальнейшем выяснилась общенаучная значимость дисперсионного анализа для экспериментов в
  - психологии,
  - педагогике,
  - медицине и др.

Основные понятия

- **Факторы - независимые переменные**.

- это те признаки, которые влияют на изучаемое явление.
в эксперименте исследователь имеет возможность варьировать ими и анализировать получающийся результат.

Зависимая переменная - результативные признаки

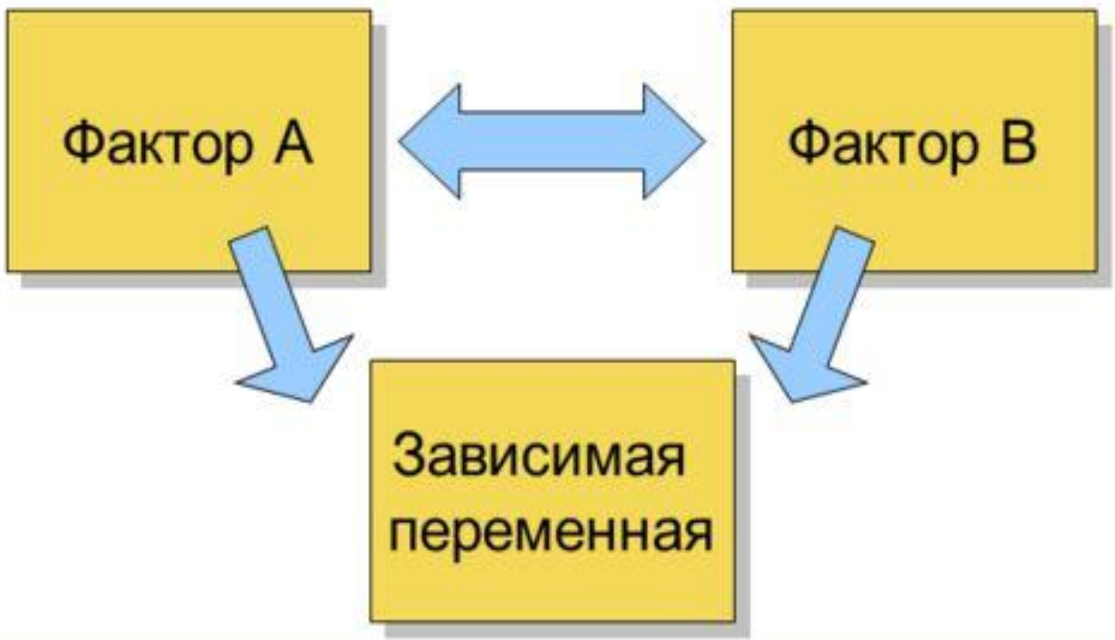
- это те признаки, которые изменяются под влиянием факторных признаков.

Однофакторный и двухфакторный анализ

Дисперсионный анализ, который рассматривает только одну переменную называется **однофакторным дисперсионным анализом** (One-Way ANOVA). Дисперсионный анализ может также применяться в случае двух переменных - это **двухфакторный дисперсионный анализ** (Two-Way ANOVA).

Постановка проблемы

При применении двухфакторного дисперсионного анализа исследователь проверяет влияние двух независимых переменных (факторов) на зависимую переменную. Может быть изучен также эффект взаимодействия двух переменных.



- Основная цель **дисперсионного анализа** (ANOVA):

\- является исследование значимости различия между средними с помощью сравнения (анализа) дисперсий.

Сущность дисперсионного анализа

- Разложение общей дисперсии изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак.

- Это достигается посредством установления значимости различия между выборочными средними.

		A	B	C
		50	70	20
Внутригрупповая		40	80	15
		60	90	25
Средние		50	80	20
			Межгрупповая	
			Общая средняя	50

Разделение **общей дисперсии** на несколько источников, позволяет сравнить дисперсию, вызванную **различием между группами**, с дисперсией, вызванной **внутригрупповой изменчивостью.**

Суммы квадратов отклонений

Межгрупповая сумма квадратов отклонений:

$$SS_b = \sum n_i (\bar{x}_i - \bar{\bar{x}})^2$$

Sum **S**quare **B**etween Groups

Внутригрупповая сумма квадратов отклонений:

$$SS_w = \sum (x - \bar{x}_i)^2$$

Sum **S**quare **W**ithin Groups

Общая сумма квадратов отклонений:

$$SS = \sum (x - \bar{x})^2 = SS_b + SS_w$$

Sum **S**quare

- При истинности нулевой гипотезы (о равенстве средних в нескольких группах наблюдений, выбранных из генеральной совокупности), оценка дисперсии, связанной с внутригрупповой изменчивостью, должна быть близкой к оценке межгрупповой дисперсии.

Постановка задачи

- Имеются данные о весе томатов (все растение целиком (weight, в кг), которые выращивали в течение 2 месяцев при трех разных экспериментальных условиях (trt, от *treatment*)
- на воде (water),
- в среде с добавлением удобрения (nutrient),
- а также в среде с добавлением удобрения и гербицида 2,4-D (nutrient+24D):

- Рассматриваемый пример соответствует случаю *однофакторного* дисперсионного анализа:
- изучается действие одного фактора - условий выращивания (с тремя уровнями - Water, Nutrient иNutrient+24D)
- на интересующую нас переменную-отклик - вес растений.

Результаты

trt

- 1.5, 1.9, 1.3, 1.5, 2.4, 1.5, # *water*
- 1.5, 1.2, 1.2, 2.1, 2.9, 1.6, # *nutrient*
- 1.9, 1.6, 0.8, 1.15, 0.9, 1.6 # *nutrient+24D*

Переменная **trt** представляет собой **фактор** с тремя уровнями.

Визуализация данных при помощи одномерной диаграммы рассеяния

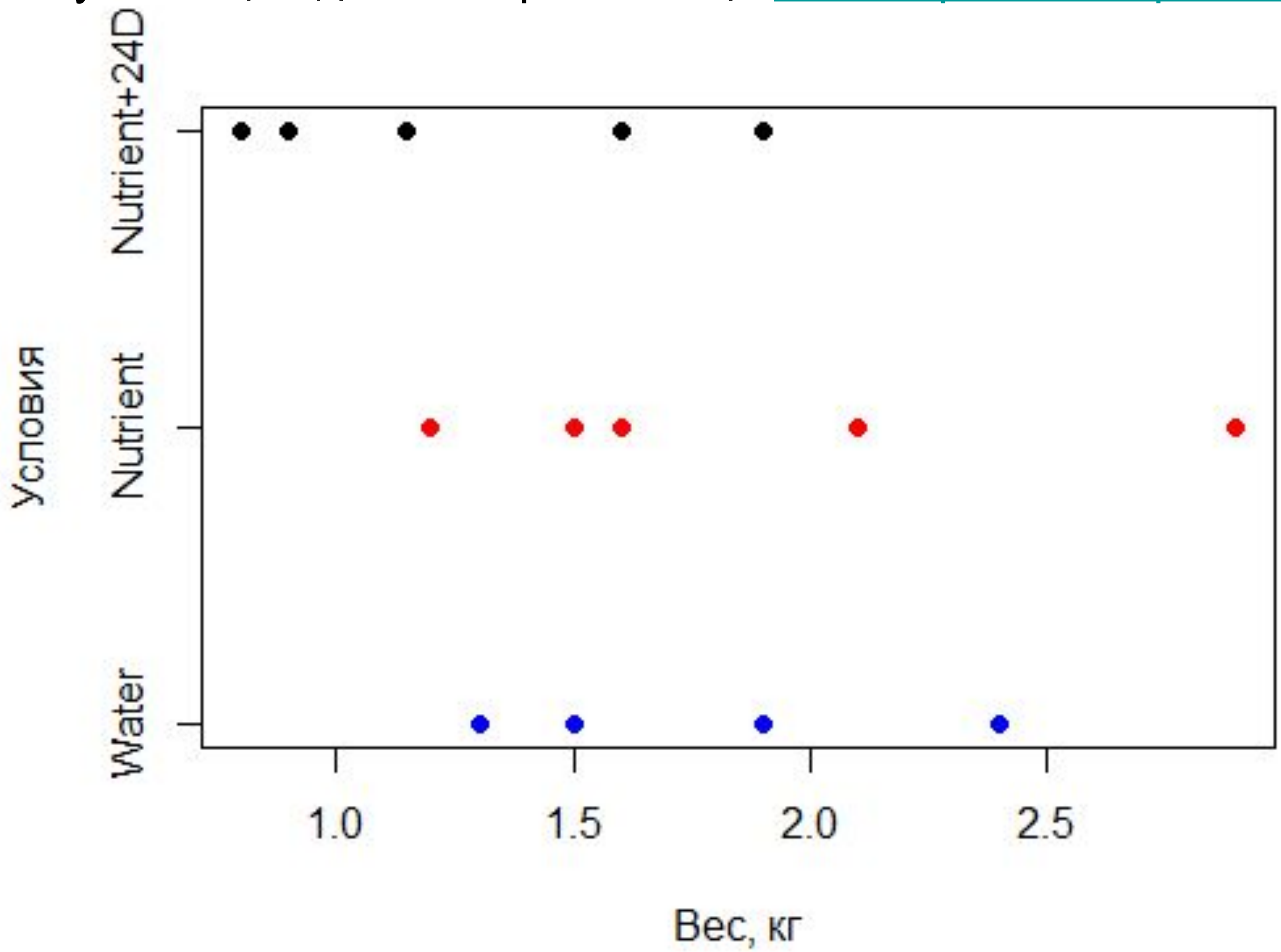


Рис. 2. Результаты измерений веса растений томатов, выращенных при разных экспериментальных условиях.

Значения веса растений достаточно близки для всех трех экспериментальных условий, хотя и есть некоторая тенденция к снижению веса в группе "Nutrient+24D".

- Water 1.683333
- Nutrient 1.750000
- Nutrient+24D 1.325000

- Подлежащую проверке *нулевую гипотезу* можно сформулировать так: исследованные условия выращивания растений не оказывают никакого влияния на вес последних.
- Другими словами, нулевая гипотеза утверждает, что *наблюдаемые различия между групповыми средними несущественны и вызваны влиянием случайных факторов*(т.е. в действительности все полученные измерения веса растений происходят из одной нормально распределенной генеральной совокупности):

Гипотезы

Для выявления различия между тремя и более средними, выдвигаются следующие гипотезы:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_m$$

H_1 : не все средние равны

- К сожалению, исследователь почти никогда не имеет возможности изучить всю генеральную совокупность.
- Как же узнать, верна ли приведенная выше нулевая гипотеза, располагая только выборочными данными?
- Мы можем сформулировать этот вопрос иначе: *какова вероятность получить наблюдаемые различия между групповыми средними, извлекая случайные выборки из одной нормально распределенной генеральной совокупности*?

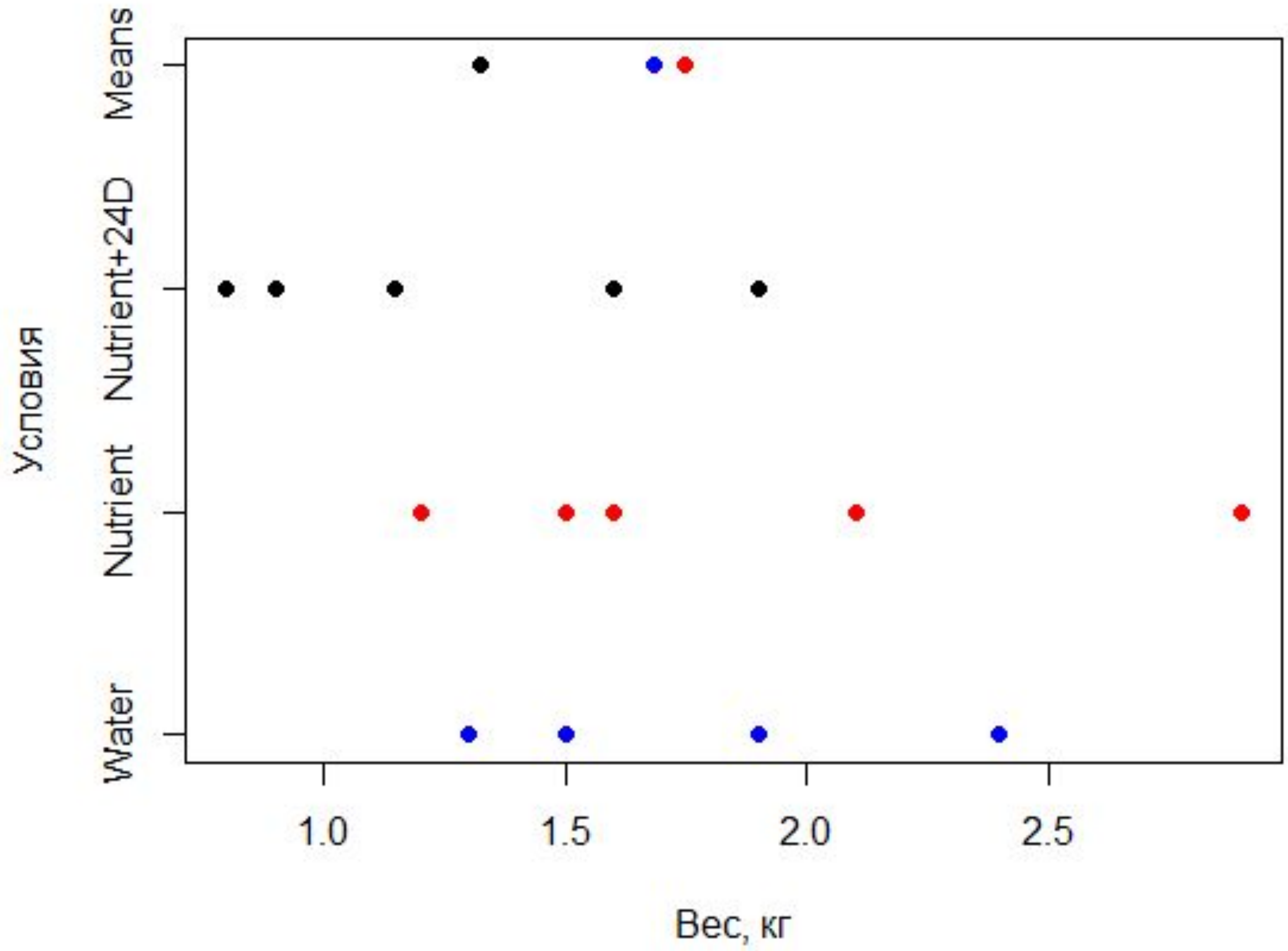
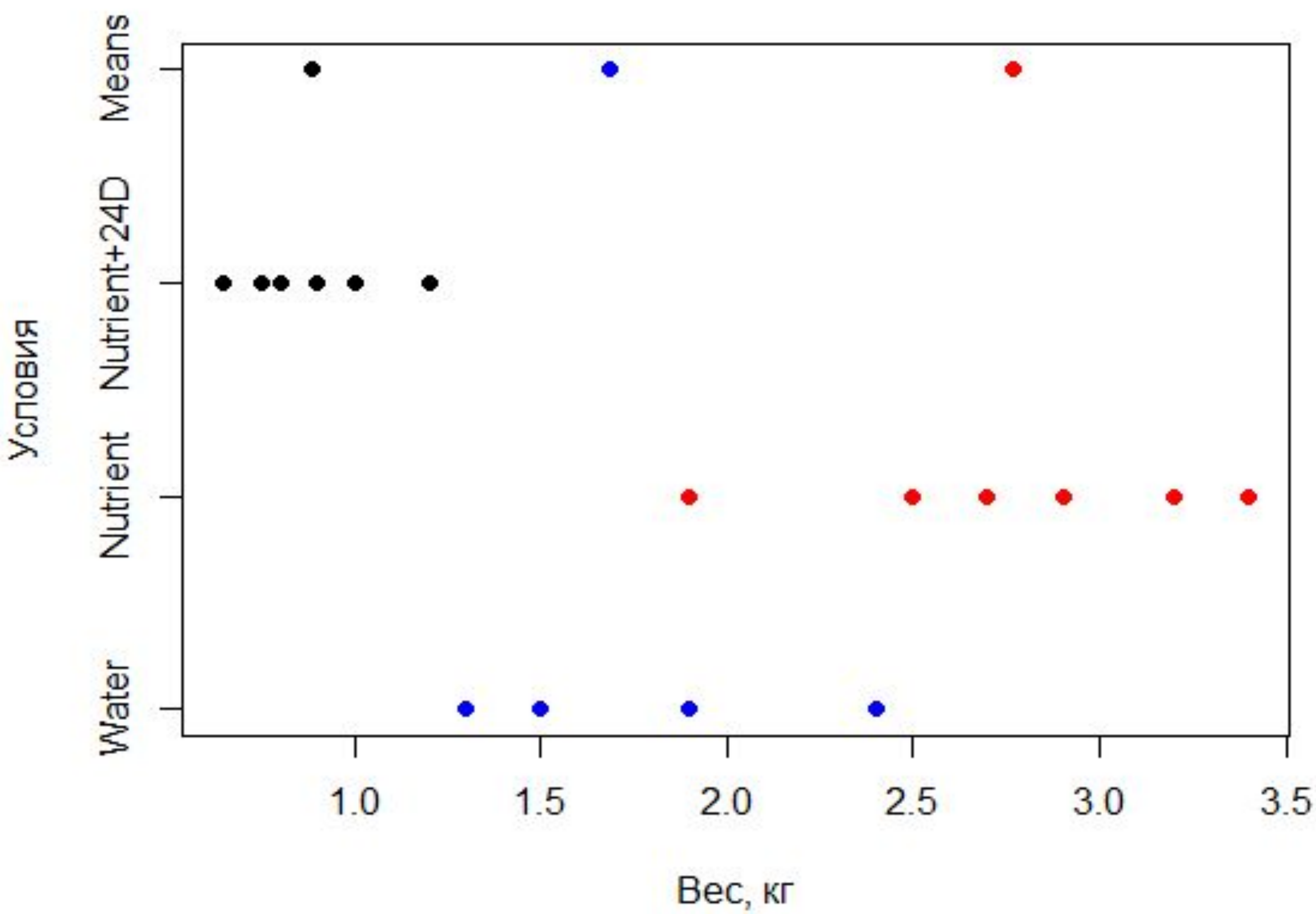


Рис. 3. То же, что рис. 2, но с добавлением точек, отражающих средние значения в каждой экспериментальной группе (Means).

Теперь (*исключительно с целью продемонстрировать принцип*!) несколько изменим исходные данные



- Группы точек, отражающих экспериментальные данные, оказались значительно раздвинутыми вдоль оси X. Результатом этого стало также расхождение групповых средних (Means). Теперь, глядя на рис. 4, почти любой скажет, что экспериментальные группы различаются по весу растений. Почему? Сравните разброс значений *внутри* экспериментальных групп с *разбросом трех групповых средних:*разброс групповых средних на рис. 4 в целом превышает разброс значений в экспериментальных группах (тогда как на рис. 3 мы имели обратную ситуацию).

  Следовательно, для оценки различий между группами следует каким-то образом сравнить разброс групповых средних с разбросом значений внутри групп. Это ключевая идея дисперсионного анализа, уяснив которую, вы не будете испытывать трудности с пониманием излагаемого ниже материала.

-

- Итак, чем больше разброс выборочных средних и чем меньше разброс значений внутри групп, тем меньше вероятность того, что наши группы являются случайными выборками из одной совокупности. Дисперсию генеральной совокупности можно оценить двумя способами. С одной стороны, оценкой дисперсии генеральной совокупностью будет дисперсия, вычисленная *для каждой группы*. Такая оценка не будет зависеть от различий групповых средних. С другой стороны, при верной нулевой гипотезе (см. выше) *разброс групповых средних* тоже позволит оценить дисперсию генеральной совокупности. Очевидно, что такая оценка уже будет зависеть от различий между группами.

- Если экспериментальные группы - это случайные выборки из одной и той же нормально распределенной генеральной совокупности, то оба способа оценки генеральной дисперсии должны давать примерно одинаковые результаты. Соответственно, если эти оценки действительно оказываются близки, то мы не можем отвергнуть нулевую гипотезу. И наоборот: если разница между этими оценками оказывается существенной, мы можем принять альтернативную гипотезу: *маловероятно, что мы получили бы наблюдаемые различия между группами, если бы они были просто случайными выборками из одной нормально распределенной генеральной совокупности*.

- Сравнивая компоненты дисперсии друг с другом посредством F—критерия Фишера, можно определить, какая доля общей вариативности результативного признака обусловлена действием регулируемых факторов.

Суммы квадратов отклонений

Межгрупповая сумма квадратов отклонений:

$$SS_b = \sum n_i(\bar{x}_i - \bar{\bar{x}})^2$$

Sum **S**quare **B**etween Groups

Внутригрупповая сумма квадратов отклонений:

$$SS_w = \sum (x - \bar{x}_i)^2$$

Sum **S**quare **W**ithin Groups

Общая сумма квадратов отклонений:

$$SS = \sum (x - \bar{x})^2 = SS_b + SS_w$$

Sum **S**quare

Факторная и остаточная дисперсия. Критерий

Межгрупповая (факторная) дисперсия:

$$MS_B = \frac{SS_B}{k-1}$$

Mean **S**quare
Between Groups

Внутригрупповая (остаточная) дисперсия:

$$MS_W = \frac{SS_W}{N-k}$$

Mean **S**quare
Within Groups

F-критерий:

$$F = \frac{MS_B}{MS_W}$$

Нулевая гипотеза

- Если **F>Fкр** (при вероятности P=0,95), то влияние фактора **существенно**.
- Если **F<Fкр** (вероятность P < 0,95), фактор **не влияет** на изучаемый признак.

	Группа 1	Группа 2
Наблюдение 1 Наблюдение 2 Наблюдение 3	2 3 1	6 7 5
Среднее Сумма квадратов (СК)	2 2	6 2
Общее среднее Общая сумма квадратов	4 28	

ГЛАВНЫЙ ЭФФЕКТ

	SS	ст.св.	MS	F	p
Эффект	24.0	1	24.0	24.0	.008
Ошибка	4.0	4	1.0		

Однофакторный дисперсионный анализ

ИТОГИ

Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
I группа (контр.)	5	1673	334,6	56,8
II группа	5	1812	362,4	220,8
III группа	5	1885	377	276,5

ANOVA

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	4640	2	2319,8	12,55983	0,0011415	3,885290312
Внутри групп	2216	12	184,7			
Итого	6856	14				

однофакт_дисп_ан / Лист1 / Лист2 / Лист3 / Лист4 / Лист5

Этапы дисперсионного анализа:

- Представить данные в виде таблицы.

Номер наблюдения (j)	Уровни фактора(i)					
	1	2	3	. . .	I	a
1	x_{11}	x_{21}	x_{31}		x_{i1}	x_{a1}
2	x_{12}	x_{22}	x_{32}		x_{i2}	x_{a2}
. . .						
j	x_{1j}	x_{2j}	x_{3j}		x_{ij}	x_{aj}
n	x_{1n}	x_{2n}	x_{3n}		x_{in}	x_{an}
Суммы по группам:	$\sum x_1$	$\sum x_2$	$\sum x_3$		$\sum x_i$	$\sum x_a$
Средние по группам:	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$		$\overline{X}_i$	$\overline{X}_a$

i – индекс уровня фактора (от 1 до а);

j – индекс варианты (от 1 до n).

- Общее варьирование всех вариант (x_{ij}), независимо от того, в какой группе они находятся, вокруг общей средней $\overline{X}$ характеризуется дисперсией $\mathbf{D}_{общ.}$

$$D_{общ} = \frac{\sum_{ij}(x_{ij} - \overline{x})^2}{N-1}$$

где **N=a·n**– число всех вариант;

$\mathbf{df}_{общ.}$**= N–1** – число степеней свободы.

- Варьирование групповых средних $\bar{X}_i$ или средних каждого уровня данного изучаемого фактора вокруг общей средней $\bar{X}$, характеризуется факторной дисперсией **D$_{факт}$**.

$$D_{факт} = \frac{\sum\limits_{i} n_i(\bar{x}_i - \bar{x})^2}{a-1}$$

df$_{факт}$= a – 1– число степеней свободы.

n_i – среднее число вариант в каждой группе,

n – если число вариант в группах одинаково.

- Варьирование вариант x_{ij} внутри каждой группы вокруг каждой групповой средней $\bar{x}_i$ характеризует случайная или остаточная дисперсия **$D_{случ}$**.

$$D_{случ} = \frac{\sum_i \left[\sum_j (x_{ij} - \bar{x}_i)^2 \right]}{N - a}$$

$df_{случ}$ = N - a – число степеней свободы.

Причем: **(N – a) + (a – 1) = N – 1**

Формулы для однофакторного дисперсионного анализа

Источник варьирования	Сумма квадратов SS (числитель)	Число степеней свободы df (знаменатель)	Формулы для дисперсии MS
Общее (все варианты)	$\sum_{ij}(x_{ij}-\bar{x})^2$	**N – 1**	$\frac{1}{N-1}\sum_{ij}(x_{ij}-\bar{x})^2$
Групповые средние (фактор А)	$\sum_{i} n_i(\bar{x}_i-\bar{x})^2$	**a –1**	$\frac{1}{a-1}\sum_{i} n_i(\bar{x}_i-\bar{x})^2$
Варианты внутри групп (случайные отклонения)	$\sum_{i}\left[\sum_{j}(x_{ij}-\bar{x}_i)^2\right]$	**N – a**	$\frac{1}{N-a}\sum_{i}\left[\sum_{j}(x_{ij}-\bar{x}_i)^2\right]$

Пример. Провести однофакторный дисперсионный анализ для выяснения влияния реагентов на синтез лекарственного препарата (выход-усл.ед).

№	F_1	F_2	F_3	F_4
	x_{1j}	x_{2j}	x_{3j}	x_{4j}
1	5	4	7	8
2	3	3	6	7
3	4	4	5	9
4	4	2	6	7
$\bar{x}_i =$	4	3,25	6	7,75

Предположения:

- **В генеральной совокупности выборки подчиняются нормальному закону распределения**
- **Выборки однородны (дисперсии равны)**

Таблица 1

№	F_1	F_2	F_3	F_4
	x_{1j}	x_{2j}	x_{3j}	x_{4j}
1	5	4	7	8
2	3	3	6	7
3	4	4	5	9
4	4	2	6	7
$\overline{x}_i =$	4	3,25	6	7,75
$\sum x_i =$	16	13	24	31
$\left(\sum x_i\right)^2 =$	256	169	576	961

Таблица 2

№	F_1	F_2	F_3	F_4
	x^2_{1j}	x^2_{2j}	x^2_{3j}	x^2_{4j}
1	25	16	49	64
2	9	9	36	49
3	16	16	25	81
4	16	4	36	49
$\sum x_i^2 =$	66	45	146	243

$$\sum x_{ij}^2 = 500$$

$$\sum x_{ij} = 84$$

$$\left(\sum x_{ij}\right)^2 = 7056$$

Вычисления:

- Сумма квадратов $SS_{общ}$ для общей вариации:

$$SS_{\text{îáù}} = \sum \tilde{\delta}_{ij}^{2} - \frac{\left(\sum x_{ij}\right)^{2}}{N} = 500 - \frac{7056}{16} = 59$$

- Сумма квадратов $SS_{факт}$ для вариации между группами:

$$SS_{фак} = \frac{1}{n}\sum\left(\sum x_{i}\right)^{2} - \frac{\left(\sum x_{ij}\right)^{2}}{N} = \frac{1}{4}\ 1962 - \frac{7056}{16} = 49.5$$

Средний квадрат, характеризующий факторную дисперсию $MS_{факт}$.:

$$MS_{фак} = \frac{SS_{фак}}{df_{фак}} = \frac{49{,}5}{3} = 16{,}5$$

- Сумма квадратов $SS_{случ}$ для вариации внутри групп:

$$SS_{сл}=SS_{общ}-SS_{фак}=59-49{,}5=9{,}5$$

- Сумма квадратов $SS_{случ}$ для вариации внутри групп:

$$MS_{сл}=\frac{SS_{сл}}{df_{сл}}=\frac{9{,}5}{12}=0{,}79$$

т.к $MS_{случ} < MS_{фак,}$

$$F=\frac{MS_{\text{ðàê}}}{MS_{\text{ñë}}}=\frac{16{,}5}{0{,}79}=20{,}84$$

а $F_{табл}=3{,}49$ для $P=0{,}95$ и $df_{сл}=12$ и $df_{фак}=3$

ВЛИЯНИЕ ФАКТОРА ДОСТОВЕРНО!

Сила влияния фактора

- Сила влияния фактора η^2_A определяется:

$$\eta^2_A = \frac{D_{факт.}}{D_{факт.} + D_{случ.}}$$

где

$$D_{факт} = \frac{MS_{факт.} - MS_{случ.}}{n} = \frac{MS_{факт.} - D_{случ.}}{n}$$

В нашем случае

$$D_{\hat{o}\grave{a}\hat{e}\grave{o}} = \frac{16{,}5 - 0{,}79}{4} = 3{,}93$$

$$\eta^2_A = \frac{D_{факт.}}{D_{факт.} + D_{случ.}} = \frac{3{,}93}{3{,}93 + 0{,}79} = 0{,}83 = 83\%$$

Достоверность влияния фактора

- Дисперсионный анализ позволяет установить, существуют ли **достоверные различия** между отдельными **уровнями** фактора.

$$S_d = \sqrt{\frac{D_{случ.}}{n}}$$

n – число вариант в каждой группе.

- Отношение разницы d к ее ошибке S_d , т.е. t= $\frac{d}{S_d}$, должно быть таким, чтобы оно гарантировало достоверность не менее чем при P=0,95.

- Коэффициент Q, рассчитан для разного количества **групп а** и степеней свободы $\mathbf{df_{случ}}$.

$$S_d = \sqrt{\frac{D_{случ.}}{n}} = \sqrt{\frac{0,79}{4}} = 0,2$$

$d_{12}=4-3,25=0,7$; $d_{23}=6-3,25=2,75$;

$$t_{12} = \frac{d_{12}}{S_d} = \frac{0,7}{0,2} = 3,8; \qquad t_{23} = \frac{d_{23}}{S_d} = \frac{2,75}{0,2} = 13,9;$$

$Q=4,2$ для $df_{случ}=12$ и $a=4$;

$t_{12}< Q$, разница **не достоверна**!

$t_{23}> Q$, разница **достоверна**!

Вывод:

Вид реагента достоверно влияет на выход лекарственного препарата. Наибольшую эффективность имеет фактор (реагент), градация которого равна F_4.

Номер типа почвы	Результаты измерений урожайности Номер эксперимента				
	1	2	3	…	n
1	x_{11}	X_{12}	X_{13}	…	X_{1n}
2	X_{21}	X_{22}	X_{23}	…	X_{2n}
3	X_{31}	X_{32}	X_{33}	…	X_{3n}
…	…	…	…	…	…
m	X_{m1}	X_{m2}	X_{m3}	…	x_{nm}

Результаты измерения урожайности в относительных единицах

Номер типа почвы	Номер эксперимента					Выборочное среднее
	1	2	3	4	N=5	
i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	$\bar{x} = \frac{1}{5}\sum_{k=1}^{5} x_{ik}$
1	12	15	17	13	16	$\bar{x}_1 =$ 14.6
2	20	17	16	25	14	$\bar{x}_2 =$ 18.4
m=3	10	12	11	13	8	$\bar{x}_3 =$ 10.8

Схема однофакторного дисперсионного анализа

Компонента дисперсии	Сумма квадратов	Число степеней свободы	Выборочная дисперсия
Между типами почвы	$\sum_{i,k}\left(\bar{x}_i - \bar{x}\right)^2$	$m-1$	s_1^2
Внутри типов почвы	$\sum_{i,k}\left(x_{i,k} - \bar{x}_i\right)^2$	$mn-m$	s_2^2
Полная (общая)	$\sum_{i,k}\left(x_{ik} - \bar{x}\right)^2$	$mn-1$	s^2

$$\bar{x} = \frac{1}{15}\sum_{i=1}^{3}\sum_{k=1}^{5} x_{ik} = 14.6$$

$$Q_1 = 5\sum_{i=1}^{3}\left(\bar{x}_i - \bar{x}\right)^2 = 137$$

$$Q_2 = \sum_{i=1}^{3}\sum_{k=1}^{5}\left(x_{ik} - \bar{x}_i\right)^2 = 102.2$$

$$Q = \frac{1}{15}\sum_{i=1}^{3}\sum_{k=1}^{5}\left(x_{ik} - \bar{x}\right)^2 = 239.2$$

$$s_2^2 = \frac{Q_2}{12} = 8.5 \qquad s_1^2 = \frac{Q_1}{2} = 68.5 \qquad s^2 = \frac{Q}{14} = 17.1$$

Для нашего примера таблица однофакторного анализа будет иметь следующий вид

Дисперсионный анализ урожайности на различных типах почвы

Компонента дисперсии	Сумма квадратов	Число степеней свободы	Выборочная дисперсия
Между типами почвы	Q_1=137	2	$s_1^2 = 68.5$
Внутри типов почвы	Q2=102.2	12	$s_2^2 = 8.5$
Полная (общая)	Q_3=239.2	14	$s^2 = 17.1$

- Произведя теперь проверку нулевой гипотезы (4) с помощью распределения, находим

$$F = \frac{s_1^2}{s_2^2} = \frac{68.5}{8.5} = 8.06$$

- При двух степенях свободы большей дисперсии (k1 = 2) и 12 е свободы меньшей дисперсии (k2 = 12) по табл. в приложении II находим критические границы для F, равные при 5%-м уровне рзначимости и 3.88 и 1%-м уровне — 6.93.
- Полученное нами из наблюдений значение превышает указанные границы, и потому нулевая гипотеза должна быть отвергнута, т.е. урожайность на рассматриваемых типах почвы неодинакова.

Двухфакторный дисперсионный анализ с повторениями

В таблице. приведены суточные привесы (г) собранных для исследования 18 поросят в зависимости от метода удержания поросят (фактор А) и качества их кормления (фактор В).

Таблица 11.6

Количество голов в группе (фактор А)	*Содержание протеина в корме, г (фактор В)*	
	$B_1=80$	$B_2=100$
$A_1=30$	530, 540, 550	600, 620, 580
$A_2=100$	490, 510, 520	550, 540, 560
$A_3=300$	430, 420, 450	470, 460, 430

Формируем таблицу, сочетая в каждом варианте опыта уровни каждого из факторов:

Повторности	**B1**		**B2**		**B3**		**a=3**
	A1	**A2**	**A1**	**A2**	**A1**	**A2**	**b=2**

Таблица результатов

Результаты вычислений представляют в виде следующей таблицы:

	Сумма квадратов	***df***	**Среднее квадратичное**	***F***
Фактор A	SS_A	$a - 1$	MS_A	F_A
Фактор B	SS_B	$b - 1$	MS_B	F_B
Взаимодействие, AxB	SS_{AxB}	$(a - 1)(b - 1)$	MS_{AxB}	F_{AxB}
Ошибка	SS_{error}	$ab(n - 1)$	MS_{error}	
ИТОГО	…	…	…	

Обозначения

SS_A – сумма квадратов для фактора A

SS_B – сумма квадратов для фактора B

SS_{AxB} – сумма квадратов для взаимодействия факторов

SS_{error} – сумма квадратов для ошибки

a – количество уровней фактора A

b – количество уровней фактора B

n – количество объектов в каждой группе

Формулы для вычислений

$$MS_A = \frac{SS_A}{a-1}; MS_B = \frac{SS_B}{b-1}; MS_{A\times B} = \frac{SS_{A\times B}}{(a-1)(b-1)}; MS_{внутри} = \frac{SS_{внутри}}{ab(n-1)}$$

$$F_B = \frac{MS_B}{MS_{внутри}}$$

df.N. $= b - 1$
df.D. $= ab(n - 1)$

$$F_{A\times B} = \frac{MS_{A\times B}}{MS_{внутри}}$$

df.N. $= (a - 1)(b - 1)$
df.D. $= ab(n - 1)$

$$F_A = \frac{MS_A}{MS_{внутри}}$$

df.N. $= a - 1$
df.D. $= ab(n - 1)$

Взаимодействие

Эффекты факторов, накладываясь друг на друга в разных сочетаниях, приводят к разным последствиям.

Например, если уровень B2 повышает значение признака на 20% в первой строке данных (т.е. в сочетании с уровнем A1), то во второй строке он может его не изменять или даже уменьшать.

- **В фиксированной модели** проверка нулевой гипотезы (определение критерия Фишера) производится так же, как и в однофакторном анализе, т.е. **сравнением среднего квадрата каждого фактора со случайным средним квадратом.**

- **В случайной модели** приходится делить **средний квадрат фактора на средний квадрат взаимодействия.**

- **Необходимость заранее определять, с какой моделью мы имеем дело**.

23	Счет	9	9				
24	Сумма	4440	4810				
25	Среднее	493,33333	534,44444				
26	Дисперсия	2375	4402,7778				
27							
28							
29	Дисперсионный анализ						
30	*точник вариац*	*SS*	*df*	*MS*	F	P-Значение	F критическое
31	Выборка	50011,111	2	25005,55556	100,02	3,285E-08	3,885293835
32	Столбцы	7605,5556	1	7605,555556	30,422	0,0001329	4,747225336
33	Взаимодейств	1211,1111	2	605,5555556	2,4222	0,1307216	3,885293835
34	Внутри	3000	12	250			
35							
36	Итого	61827,778	17				

Очевидно, данные факторы имеют фиксированные уровни, т.е. мы находимся в рамках модели I. Поэтому для проверки существенности влияния факторов А, В и их взаимодействия АВ необходимо найти отношения

$$F_A = \frac{s_1^2}{s_4^2} = \frac{25005,5}{250,0} = 100,0; \quad F_B = \frac{s_2^2}{s_4^2} = \frac{7605,6}{250,0} = 30,4; \quad F_{AB} = \frac{s_3^2}{s_4^2} = \frac{605,6}{250,0} = 2,42$$

Двухфакторный дисперсионный анализ без повторений

H	I	J	K	L	N	P
	Фактор В тип удобрения	Фактор А -сорт пшеницы				yi-
		A1	A2	A3	A4	
	B1	19,00	25,00	17,00	21,00	20,50
	B2	22,00	19,00	19,00	18,00	19,50
	B3	26,00	23,00	22,00	25,00	24,00
	B4	18,00	26,00	20,00	23,00	21,75
	B5	21,00	22,00	21,00	24,00	22,00
	y-j	21,20	23,00	19,80	22,20	

Рисунок 2

Задача : Необходимо определить, влияет ли сорт и тип удобрения на урожайность пшеницы.

	A	B	C	D	E	F	G
1	Двухфакторный дисперсионный анализ без повторений						
2							
3	ИТОГИ	Счет	Сумма	Среднее	Дисперсия		
4	Строка 1	4	82	20,5	11,66667		
5	Строка 2	4	78	19,5	3		
6	Строка 3	4	96	24	3,333333		
7	Строка 4	4	87	21,75	12,25		
8	Строка 5	4	88	22	2		
9							
10	Столбец 1	5	106	21,2	9,7		
11	Столбец 2	5	115	23	7,5		
12	Столбец 3	5	99	19,8	3,7		
13	Столбец 4	5	111	22,2	7,7		
14							
15							
16	Дисперсионный анализ						
17	сточник вариаци	SS	df	MS	F	P-Значение	F критическое
18	Строки	46,2	4	11,55	2,032258	0,1536621	3,259166727
19	Столбцы	28,55	3	9,516667	1,674487	0,22510611	3,490294821
20	Погрешность	68,2	12	5,683333			
21							
22	Итого	142,95	19				