

Эксперимент

Пассивный эксперимент - информация об исследуемом объекте накапливается путем пассивного наблюдения, то есть информацию получают в условиях обычного функционирования объекта.

Активный эксперимент предусматривает активное вмешательство в исследуемый процесс, изменяя его по заранее разработанному экспериментатором плану.

Пассивный эксперимент

Задачи при планировании:

- выбор количества и частоты измерений;
- выбор метода обработки результатов измерений.

Наиболее часто целью пассивного эксперимента является построение математической модели объекта, которая может рассматриваться либо как хорошо, либо как плохо организованный объект.

Активный эксперимент

К основным преимуществам активного эксперимента можно отнести следующие:

- планирование эксперимента дает четкую последовательную логическую схему построения всего процесса исследования, т. е. известно, что, когда и как надо делать;
- внедрение активного планирования позволяет повысить эффективность исследований, извлечь наибольшее количество сведений об изучаемых процессах при ограниченных затратах, сократить объем экспериментальных исследований, повысить надежность и четкость интерпретации полученных результатов;
- обработка результатов эксперимента осуществляется стандартными приемами, позволяющими формализовать процесс построения модели и сопоставить материалы различных исследований.

При планировании эксперимента исследователь должен:

- обеспечить высокую надежность и четкость интерпретации результатов экспериментальных исследований;
- составить четкую и последовательную логическую схему построения всего процесса исследования;
- максимально формализовать процесс разработки модели и сопоставления экспериментальных данных различных опытов одного и того же объекта исследований с целью широкого применения электронно-вычислительных средств.

Статистические методы планирования активного эксперимента являются одним из эмпирических способов получения математического описания статистики сложных объектов исследования, то есть уравнения связи отклика объекта и независимых управляемых входных переменных (факторов). При этом математическое описание представляется в виде **полинома**

$$Y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i \neq j}^k b_{ij} X_i X_j + \sum_{i=1}^k b_{ii} X_i^2 + \dots,$$

где Y – функция отклика;

X_1, X_2, \dots, X_k – факторы исследуемого процесса.

Первый этап исследования – составление плана эксперимента

Определяется расположение экспериментальных точек в k -мерном факторном пространстве, иначе говоря, **условия для всех опытов**, которые необходимо провести.

План эксперимента задается в виде **матрицы планирования**, каждая строка которой определяет условия опыта, а каждый столбец – значения контролируемых и управляемых параметров в исследуемом процессе, то есть значения факторов, соответствующих условию опыта. В последний столбец матрицы заносят значения функции отклика, полученные экспериментальным путем в каждом опыте.

Первый шаг – *выбор центра плана*, то есть точки, соответствующей начальному значению всех используемых в эксперименте факторов ($x_{10}, x_{20}, \dots, x_{k0}$), в окрестностях которой в дальнейшем ставится серия планируемых опытов. Начальным значениям факторов будет соответствовать начальное значение функции отклика y_0 . Центр плана обычно выбирается на основе априорных сведений о процессе. Если же их нет, то обычно в качестве центра плана принимается центр исследуемой области.

Второй шаг – *задание интервала варьирования*. Значения факторов в каждом опыте, в случае применения матрицы планирования эксперимента, отличается от начального их значения x_{i0} на величину интервала Δx . Одним из важнейших предварительных условий успешного проведения эксперимента с целью разработки математической модели, адекватной исследуемому процессу, является выбор оптимальной величины Δx . Обычно интервал варьирования выбирают в пределах 0,05 ... 0,3 от диапазона варьирования исследуемого фактора.

Третий шаг – для удобства обработки результатов опытов, проводится *преобразование значений управляемых переменных* (учитываемых в эксперименте факторов x_i) к *безразмерным величинам*

$$x_{i\bar{0}} = (x_i - x_{i0})/\Delta x_i \quad (6.2)$$

где x_{i0} – базовое или начальное значение i -го фактора в центре плана;
 Δx_i – значение интервала варьирования по i -му фактору;
 x_i – текущее значение i -го фактора.

Таким образом, в безразмерной системе координат верхний уровень фактора при проведении эксперимента равен $+1$, а нижний -1 . Координаты центра плана равны нулю и совпадают с началом координат. При составлении матрицы планирования эксперимента верхний и нижний уровни переменных для упрощения записи можно заменять символами $(+)$ и $(-)$.

Второй этап исследования

Разработку модели процесса следует проводить по принципу «от простого – к более сложному». В соответствии с этим принципом, планирование эксперимента начинают с предположения, что имитируемая модель исследуемого процесса является линейной и в соответствии с (1) имеет вид полинома 1-го порядка

$$Y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i \neq j}^k b_{ij} X_i X_j.$$

Если после обработки и анализа результатов эксперимента выяснится, что сделанное предположение о линейности модели является ошибочным, то переходят к планированию эксперимента из предположения, что эта модель может быть представлена полиномом 2-го порядка и так далее до тех пор, пока не будет разработана адекватная исследуемому процессу математическая модель.

Полным факторным экспериментом

называется эксперимент, реализующий все возможные неповторяющиеся комбинации уровней n независимых управляемых факторов, каждый из которых варьируют на двух уровнях. В этом случае учитывается влияние на функцию отклика исследуемого процесса не только каждого рассматриваемого в эксперименте фактора в отдельности, но и их взаимодействий.

Рассмотрим случай воздействия на функцию отклика Y двух факторов X_1 и X_2 . В соответствии с принципом «от простого к более сложному» предположим, что модель исследуемого процесса является линейной и в соответствии с (3) имеет вид:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2,$$

где b_0 – значение функции отклика Y в центре плана;
 b_1 , b_2 – характеризуют степень влияния факторов X_1 , X_2 на функцию отклика Y (чем он больше по сравнению с другими коэффициентами, тем более весомый вклад в изменение функции отклика вносит данный фактор);
 b_{12} – характеризует весомость влияния взаимодействия 1-го и 2-го факторов на функцию отклика исследуемого процесса.

Все возможные комбинации для двух факторов ($k=2$), варьируемых на двух уровнях, будут исчерпаны, если мы поставим четыре опыта. Опытные точки расположатся в вершинах квадрата, центр которого совпадает с центром плана. Каждому из этих четырех опытов будет соответствовать свое значение функции отклика в зависимости от четырех различных сочетаний двух значений варьируемых в данном эксперименте факторов.

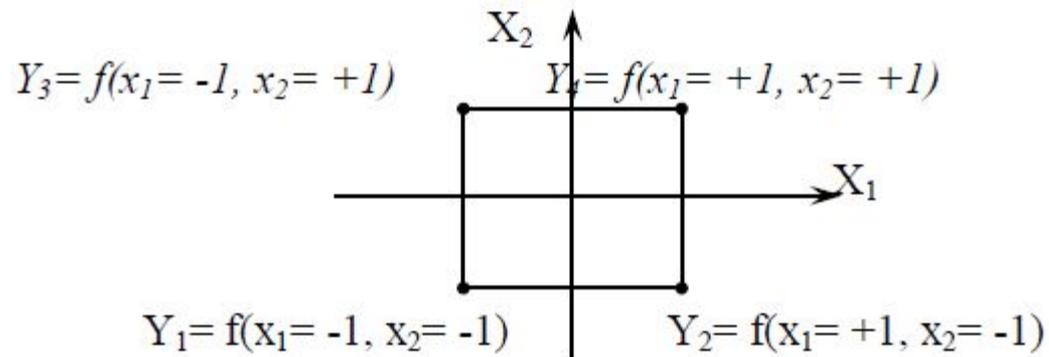


Рисунок 1 – Расположение экспериментальных точек для двух независимых факторов, варьируемых на двух уровнях

Первый столбец матрицы представляет собой нумерацию опытов. Нумерация факторов осуществляется произвольно и в каждом конкретном случае определяется самим исследователем.

Во втором столбце приводятся значения фиктивной переменной $x_0=+1$, соответствующей коэффициенту b_0 .

В последующих столбцах приводятся безразмерные символы, соответствующие верхнему и нижнему уровням варьирования факторов и их взаимодействиям.

При построении матрицы планирования ПФЭ существует следующее правило: ***первая строка матрицы в столбцах, соответствующих рассматриваемым в эксперименте факторам, заполняется безразмерным символом, соответствующим нижнему уровню значений фактора в эксперименте, то есть символом (-); продолжение заполнения столбца, соответствующего первому по порядку фактору, проводится последовательным чередованием противоположных знаков (безразмерных значений уровней варьирования фактора); все последующие столбцы, соответствующие другим пронумерованным по порядку факторам, заполняются с частотой смены знака вдвое меньшей, чем для предыдущего столбца.***

Заполнение столбцов, учитывающих взаимодействие факторов, производится как результат перемножения знаков соответствующих факторов в каждой строке.

В последний столбец матрицы заносятся экспериментальные значения функции отклика, полученные в результате проведения каждого опыта.

Матрица планирования ПФЭ типа 2^2

Номер опыта	X_{06}	X_{16}	X_{26}	$X_{16}X_{26}$	Y
1	+	-	-	+	Y_1
2	+	+	-	-	Y_2
3	+	-	+	-	Y_3
4	+	+	+	+	Y_4

Если в эксперименте используются три фактора, а предполагаемая математическая модель линейна, то она соответствует виду:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_{12} X_1 X_2 + b_{13} X_1 X_3 + b_{23} X_2 X_3 + b_{123} X_1 X_2 X_3$$

При варьировании каждым из трех факторов ($k=3$) на двух уровнях число опытов N будет составлять $N=2^3=8$. В этом случае опытные точки располагаются в вершинах куба, центр которого находится в начале координат $(0,0,0)$ (рисунок 2).

Матрица планирования ПФЭ составляется по описанным ранее правилам, и будет иметь следующий вид (таблица 3).

Руководствуясь приведенным ранее правилом можно построить матрицу и для большего числа рассматриваемых в эксперименте факторов, число опытов в которой равно

$$N = 2^k, \quad (6)$$

где k – число учитываемых в эксперименте факторов.

Но выражение (6) справедливо только для линейной модели, соответствующей полиному 1-го порядка, когда варьирование по каждому фактору достаточно проводить на двух уровнях.

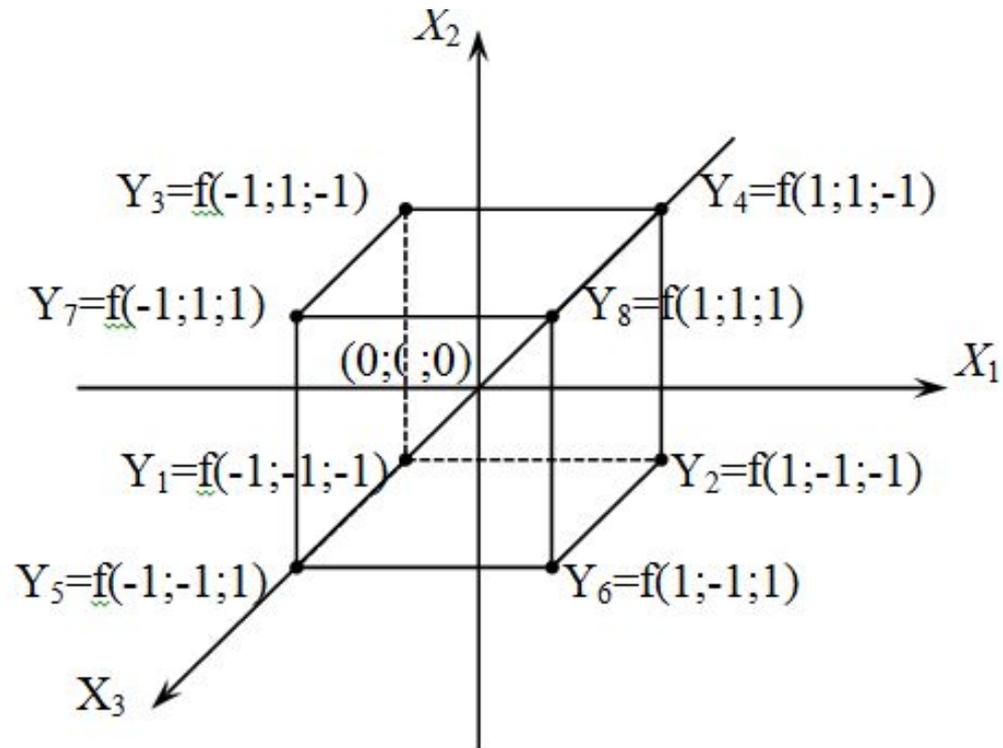


Рисунок 2 – Расположение экспериментальных точек в плане, соответствующем полиному 1-го порядка для трех независимых переменных

Таблица 3 – Матрица планирования ПФЭ типа 2^3

<u>Номер опыта</u>	X_{06}	X_{16}	X_{26}	X_{36}	$X_{16}X_2$ 6	$X_{16}X_3$ 6	$X_{26}X_3$ 6	$X_{16}X_{26}X_{36}$	Y_ξ
1	+	-	-	-	+	+	+	-	Y_1
2	+	+	-	-	-	-	+	+	Y_2
3	+	-	+	-	-	+	-	+	Y_3
4	+	+	+	-	+	-	-	-	Y_4
5	+	-	-	+	+	-	-	+	Y_5
6	+	+	-	+	-	+	-	-	Y_6
7	+	-	+	+	-	-	+	-	Y_7
8	+	+	+	+	+	+	+	+	Y_8

Если анализ результатов эксперимента показывает, что линейная модель, соответствующая полиному первого не адекватна исследуемому процессу, то переходят к планированию и проведению следующего эксперимента исходя уже из предположения, что математическая модель соответствует полиному следующего порядка и так далее. Но при планировании эксперимента, основанного на математической модели, например, соответствующей полиному второго порядка.

$$Y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i \neq j}^k b_{ij} X_i X_j + \sum_{i=1}^k b_{ii} X_i^2 + \sum_{i \neq j}^k b_{\bar{ij}\bar{j}} X_i^2 X_j^2$$

Необходимо обеспечить варьирование по каждому из факторов уже на трех уровнях. Тогда необходимое число опытов, которое нужно провести в эксперименте, должно быть не менее $N=3^k$, для полинома третьего порядка $N=4^k$ и так далее.

Достоинства многофакторного планирования ПФЭ:

- – Опытные точки находятся в оптимальном положении, то есть математическое описание исследуемого процесса оказывается более точным, чем при проведении опытов в точках, расположенных каким-либо другим образом.
- – Планирование и проведение ПФЭ сравнительно просто, что объясняет его широкое применение на практике.
- – Все факторы и соответственно коэффициенты полинома оцениваются независимо друг от друга, что обеспечивается независимостью и ортогональностью столбцов матрицы планирования.

Проведение эксперимента

Оно должно обеспечить сведение к минимуму влияния случайных параметров исследуемого процесса на функцию отклика.

С целью уменьшения их влияния на конечный результат эксперимента, необходимо придерживаться следующих требований:

- предусмотреть проведение нескольких параллельных опытов при одних и тех же условиях, предусмотренных соответствующей строкой матрицы планирования (номером опыта);
- необходимо рандомизировать неконтролируемые параметры процесса, то есть обеспечить их взаимную компенсацию.

Для выполнения первого требования должно быть предусмотрено проведение не менее двух параллельных опытов ($n = 2$), а для более высокой достоверности результатов их число увеличивают. В этом случае результаты n параллельных опытов для каждой строки матрицы планирования усредняют и при анализе результатов эксперимента используют именно усредненное значение функции отклика, соответствующие условиям опыта и подсчитываемое по следующей формуле:

$$Y_{\xi} = \sum_{i=1}^n Y_{\xi i} / n$$

где $\xi = 1, N$ – номер опыта по порядку, установленному первым столбцом матрицы;

i – номер параллельного опыта в ее строке;

$Y_{\xi i}$ – значение функции отклика, соответствующее i -му параллельному опыту в ξ -м номере опыта;

n – число параллельных опытов.

Задача статистического исследования зависимостей

При описании характера или структуры взаимосвязей (зависимостей), существующих между изучаемыми явлениями или показателями, в случае, если эти зависимости выявляются на основании статистического наблюдения за анализируемыми событиями или переменными, осуществляемого по выборке из интересующей генеральной совокупности, имеет место проблема *статистического исследования зависимостей*.

Введем обозначения:

$x^{(1)}, x^{(2)}, \dots, x^{(p)}$ - так называемые «*входные*» переменные, описывающие условия функционирования; в соответствующих математических моделях их называют независимыми, факторами-аргументами, экзогенными, объясняющими;

$y^{(1)}, y^{(2)}, \dots, y^{(m)}$ - *выходные* переменные, характеризующие поведение или результат (эффективность) функционирования, в математических моделях их называют зависимыми, эндогенными, результирующими или объясняемыми;

$\varepsilon^{(1)}, \varepsilon^{(2)}, \dots, \varepsilon^{(m)}$ - латентные случайные «остаточные» компоненты, отражающие влияние (соответственно на $y^{(1)}, y^{(2)}, \dots, y^{(m)}$) не учтенных «на входе» факторов, а также случайные ошибки в измерении анализируемых показателей.

Тогда общая задача статистического исследования зависимостей может быть сформулирована следующим образом:

по результатам n измерений

$$\left\{ \left(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)} ; y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(m)} \right) \right\}_{i=\overline{1, n}} \quad (1.1)$$

исследуемых переменных на объектах (системах, процессах) анализируемой совокупности построить такую функцию

$$f(x^{(1)}, x^{(2)}, \dots, x^{(p)}) = \begin{pmatrix} f^{(1)}(x^{(1)}, \dots, x^{(p)}) \\ f^{(2)}(x^{(1)}, \dots, x^{(p)}) \\ \dots\dots\dots \\ f^{(m)}(x^{(1)}, \dots, x^{(p)}) \end{pmatrix}, \quad (1.2)$$

которая позволила бы наилучшим (в определенном смысле) образом восстанавливать значения результирующих (прогнозируемых) переменных

$Y = (y^{(1)}, y^{(2)}, \dots, y^{(m)})^T$ по заданным значениям объясняющих переменных

$X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$.

Основные этапы статистического исследования зависимостей

Весь процесс статистического исследования зависимостей удобно разложить на основные этапы:

Этап 1 (*постановочный*). Прежде всего, исследователь должен определить:

1) элементарную единицу статистического обследования, или элементарный объект исследования O ;

2) набор показателей $(x^{(1)}, x^{(2)}, \dots, x^{(p)}; y^{(1)}, y^{(2)}, \dots, y^{(m)})$, регистрируемых на каждом из статистически обследованных объектов, с подразделением на *объясняющие* и *результатирующие*;

3) конечные прикладные цели исследования, тип исследуемых зависимостей;

Этап 2 (*информационный*). Он состоит в проведении сбора необходимой статистической информации вида (1.1).

Этап 3 (*корреляционный анализ*). Этот этап позволяет ответить на вопросы, имеется ли вообще какая-либо связь между исследуемыми переменными, какова структура этих связей и как измерить их тесноту?

Этап 4 (*определение класса допустимых решений*). Главной целью исследователя на этом этапе является определение общего вида, структуры искомой связи между Y и X , или, другими словами, описание класса функций F , в рамках которого он будет производить дальнейший поиск конкретного вида интересующей его зависимости.

Этап 5 (анализ мультиколлинеарности предсказывающих переменных и отбор наиболее информативных из них.) Под явлением мультиколлинеарности в регрессионном анализе понимается наличие тесных статистических связей между предсказывающими переменными $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Мультиколлинеарность создает трудности и неудобства при статистическом исследовании зависимостей, поэтому исследователь старается перейти к такой новой системе предсказывающих переменных (отобранных из числа исходных переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ или представленных в виде некоторых их комбинаций), в которой эффект мультиколлинеарности уже не имел бы места.

Этап 6 (вычисление оценок неизвестных параметров, входящих в исследуемое уравнение статистической связи).

Этап 7 (анализ точности полученных уравнений связи).

Часть исследования, объединяющая этапы 4, 5, 6 и 7, является *регрессионным анализом*

Показатели корреляционной связи, статистические критерии

1. Выборочное корреляционное отношение η_{yx} – это отношение межгруппового среднеквадратического отклонения к общему среднеквадратическому отклонению, т.е. $\eta_{yx} = \sqrt{\frac{D_m}{D_o}}$.

Выборочное корреляционное отношение служит для оценки тесноты корреляционной связи любого вида.

Свойства выборочного корреляционного отношения:

1. Корреляционное отношение изменяется от 0 до 1, т.е. $0 \leq \eta_{yx} \leq 1$.
2. Если $\eta_{yx} = 0$, то корреляционная зависимость между признаками Y и X отсутствует.
3. Если $\eta_{yx} = 1$, то между признаками Y и X существует функциональная зависимость.
4. Корреляционное отношение связано с линейным коэффициентом корреляции следующим неравенством: $|r_{yx}| \leq \eta_{yx}$.

2. **Линейный коэффициент корреляции r_{yx}** – показатель, измеряющий силу линейной взаимосвязи между признаками X и Y.

Свойства выборочного корреляционного отношения:

1. Линейный коэффициент корреляции изменяется от 0 до 1, т.е. $-1 \leq r_{yx} \leq 1$.
2. Если $r_{yx} = 0$, то линейная корреляционная зависимость между признаками Y и X отсутствует.
3. Если $r_{yx} = \pm 1$, то между признаками Y и X существует функциональная линейная зависимость.

Выборочный линейный коэффициент корреляции r_{yx} (вычисленный по данным выборки, произведенной из генеральной совокупности) является оценкой коэффициента корреляции генеральной совокупности. Если r_{yx} оказался отличным от нуля, то, поскольку выборка произведена случайно, еще нельзя утверждать, что и коэффициент корреляции генеральной совокупности также отличен от нуля. Поэтому возникает необходимость при заданном уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции.

Если нулевая гипотеза отвергается, то это означает, что выборочный линейный коэффициент корреляции r_{yx} значимо отличается от нуля, а Y и X связаны линейной зависимостью.

В качестве критерия проверки нулевой гипотезы принимают случайную величину t , которая при справедливости нулевой гипотезы имеет распределение Стьюдента с $k = n - 2$ степенями свободы. Если $|t_{\text{набл}}| > t_{\text{кр}}$, нулевую гипотезу отвергают.

Выборочный линейный коэффициент корреляции r_{yx} (вычисленный по данным выборки, произведенной из генеральной совокупности) является оценкой коэффициента корреляции генеральной совокупности. Если r_{yx} оказался отличным от нуля, то, поскольку выборка произведена случайно, еще нельзя утверждать, что и коэффициент корреляции генеральной совокупности также отличен от нуля. Поэтому возникает необходимость при заданном уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции.

Если нулевая гипотеза отвергается, то это означает, что выборочный линейный коэффициент корреляции r_{yx} значительно отличается от нуля, а Y и X связаны линейной зависимостью.

В качестве критерия проверки нулевой гипотезы принимают случайную величину t , которая при справедливости нулевой гипотезы имеет распределение Стьюдента с $k = n - 2$ степенями свободы. Если $|t_{\text{набл}}| > t_{\text{кр}}$, нулевую гипотезу отвергают.

5. Множественная регрессия – модель, в которой непрерывная результирующая переменная y зависит от некоторого числа объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$.

6. Множественный коэффициент корреляции – показатель корреляции между наблюдаемыми значениями зависимой переменной во множественной регрессии и значениями, определенными по оцененному уравнению регрессии.

7. Множественный коэффициент детерминации – отношение части вариации результативного признака, объясняемой за счет вариации входящих в уравнение факторов, к общей вариации результативного признака за счет всех факторов.

8. Оценка значимости построенного уравнения регрессии в целом дается с помощью F -критерия Фишера. При этом выдвигается гипотеза, что коэффициент при неизвестной в уравнении равен нулю, и, следовательно, фактор не оказывает влияния на результат.

Оценка значимости уравнения регрессии обычно дается в виде таблицы дисперсионного анализа. Если нулевая гипотеза справедлива, то дисперсия, обусловленная регрессией, и остаточная дисперсия не отличаются друг от друга.

