



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»  
(ДГТУ)**

Кафедра «Информационные системы в строительстве»

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ  
по дисциплине  
МОДЕЛИ И МЕТОДЫ ПЛАНИРОВАНИЯ  
ЭКСПЕРИМЕНТОВ**

Ростов-на-Дону  
ДГТУ  
2024

УДК 004.942

Составитель: Н.Ю. Батурина

Методические указания для выполнения лабораторных работ, контрольной работы по дисциплине «Модели и методы планирования экспериментов» / сост. Н.Ю. Батурина – Ростов-на-Дону : Донской государственный технический университет, 2024. – 41 с.

Содержат краткую теорию, контрольные вопросы по теории, перечень заданий лабораторных (контрольных) работ, требования по выполнению и оформлению заданий, примеры выполнения заданий, перечень информационных ресурсов.

Предназначены для обучающихся направления подготовки 09.04.02

УДК 004.942

Ответственный за выпуск: зав. кафедрой «Информационные системы в строительстве» д-р ф.-м. наук, профессор А.А. Ляпин

---

Издательский центр ДГТУ

Адрес университета и полиграфического предприятия:  
344000, г. Ростов-на-Дону, пл. Гагарина, 1

© Донской государственный технический университет, 2024

## СОДЕРЖАНИЕ

|   |    |
|---|----|
| ВВЕДЕНИЕ.....   | 4  |
| 1. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ .....   | 4  |
| 1.2. Выборки, корреляционный, дисперсионный и регрессионный анализ.....                   | 4  |
| данных.....   | 4  |
| 1.2. Компонентный и кластерный анализ .....   | 7  |
| 1.3. Математические методы планирования экспериментов.....                                | 11 |
| 1.4. Полный факторный план.....   | 15 |
| 1.5. Ортогональный центральный композиционный план .....                                  | 16 |
| 2. КОНТРОЛЬНЫЕ ВОПРОСЫ ПО ДИСЦИПЛИНЕ<br>«Модели и методы планирования экспериментов»..... | 18 |
| 3. ЗАДАНИЯ ЛАБОРАТОРНЫХ (КОНТРОЛЬНЫХ) РАБОТ И<br>ТРЕБОВАНИЯ ПО ИХ ВЫПОЛНЕНИЮ .....        | 19 |
| 4. ПРИМЕРЫ ВЫПОЛНЕНИЯ ЗАДАНИЙ .....   | 20 |
| Задание 1. Корреляционный и регрессионный анализ.....                                     | 20 |
| Задание 2. Автоматизация анализа данных.....  | 23 |
| Задание 3. Компонентный анализ .....  | 26 |
| Задание 4. Кластерный анализ.....   | 30 |
| Задание 5. Планирование эксперимента с использованием ОЦКП.....                           | 34 |
| ЛИТЕРАТУРА .....  | 40 |

## ВВЕДЕНИЕ

Дисциплина «Модели и методы планирования экспериментов» направлена на подготовку магистрантов к самостоятельной деятельности в области организации и проведения научных исследований и имеет цель дать необходимые для практической деятельности знания по многомерному анализу данных и математическим методам планирования экспериментов. Материалы для изучения дисциплины и выполнения лабораторных работ (контрольных заданий) представлены в курсе на портале do.scif [1]. Дополнительные материалы можно найти в источниках [2-8].

В результате освоения дисциплины «Модели и методы планирования экспериментов» у обучающегося должны сформироваться определенные учебным планом компетенции, и магистрант должен:

Знать: модели предметных областей информационных систем и методы планирования экспериментов.

Уметь: разрабатывать планы экспериментов и синтезировать модели процессов и объектов предметных областей; оценивать эффективность экспериментов и качество моделей; профессионально эксплуатировать современное оборудование и приборы.

Владеть: навыками разработки и исследования теоретических и экспериментальных моделей объектов профессиональной деятельности; работой с пакетами прикладных программ математической статистики, Matlab, Mathcad, MS Excel.

Для достижения поставленных целей изучения дисциплины учебным планом предусмотрены лекции, лабораторные занятия, контрольная работа (заочная форма обучения), самостоятельная работа и оценка результатов обучения по дисциплине.

## 1. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ

### 1.2. Выборки, корреляционный, дисперсионный и регрессионный анализ данных

**Математическая статистика** – наука, изучающая методы исследования закономерностей в массовых случайных явлениях и процессах по данным, полученным из конечного числа наблюдений за ними, с целью получения вероятностно-статистических моделей случайных явлений.

При проведении различного рода исследований приходится изучать отдельные признаки больших совокупностей однородных объектов. **Генеральной совокупностью** называется совокупность всех возможных значений признака исследуемой случайной величины (СВ). **Случайной выборкой объема  $n$**  называется множество значений признака СВ, полученное в результате  $n$  наблюдений за этой СВ.

Выборка называется **репрезентативной** (представительной), если она правильно отражает интересующие нас свойства генеральной совокупности. Наибольшую вероятность получить по-настоящему представительную выборку мы имеем тогда, когда отбор объектов исследования из генеральной совокупности мы производим случайным образом.

Если элементов выборки много, то их группируют, при этом возможны два случая задания законов распределения признака по выборке (дискретный и интервальный вариационные ряды), которые задаются в виде таблиц или графиков.

Графической иллюстрацией дискретного вариационного ряда служит полигон частот – ломаная, соединяющая отрезками прямых точки с координатами  $(x_i, \omega_i)$ .

Графической иллюстрацией интервального вариационного ряда служит гистограмма – ступенчатая фигура с высотой ступеньки над каждым интервалом  $(a_{i-1}, a_i)$ , равной  $\omega_i$ .

Для приближенного нахождения параметров распределения (математического ожидания, дисперсии) по выборке применяются точечные и интервальные оценки. **Точечной**

**оценкой** параметра теоретического распределения называется статистическая оценка, задаваемая одним числом. Точечные оценки мало информативны, поскольку это случайные величины, и они могут заметно отличаться от оцениваемого параметра. Для повышения информативности используют **интервальные оценки** (доверительные интервалы).

Для того чтобы точечные статистические оценки обеспечивали “хорошие” приближения неизвестных параметров, они должны быть несмещенными, эффективными и состоятельными.

**Несмещенной** называют точечную оценку, математическое ожидание которой совпадает со значением оцениваемого параметра генеральной совокупности (не зависит от объема выборки).

**Выборочной средней** называется точечная оценка математического ожидания признака, которая вычисляется по формуле среднего арифметического по выборке

$$\bar{x}_s = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n}, \text{ где } n_1 + n_2 + \dots + n_k = n.$$

**Выборочной дисперсией** называется точечная оценка дисперсии признака, которая вычисляется по формуле  $s_B^2 = \overline{x_B^2} - \bar{x}_B^2$ ,  $s_B$  - выборочное среднееквадратическое отклонение, которое может обозначаться  $\sigma_{xB}$ .

Выборочная средняя является несмещенной оценкой генеральной средней, а выборочная дисперсия – нет, поэтому вместо нее используется **исправленная выборочная дисперсия**, определяемая формулой  $\bar{s}_B^2 = \frac{n}{n-1} \cdot s_B^2$ , которая уже является несмещенной оценкой генеральной дисперсии.

Несмещенная оценка называется **эффективной**, если она имеет минимальную дисперсию по сравнению с другими выборочными оценками. В практических исследованиях это означает возможность перехода от точечного оценивания к интервальному.

Оценка называется **состоятельной**, если при увеличении объема выборки (т.е. если  $n \rightarrow \infty$ ) она стремится к оцениваемому параметру. Состоятельность оценок характеризует увеличение их точности с увеличением объема выборки.

**Корреляционный и регрессионный анализ** – это методы математической статистики, позволяющие выявлять связь между величинами, значения которых получают в результате статистических наблюдений. Основная задача корреляционного анализа – оценивание тесноты связи между величинами, а регрессионного – установление ее вида.

Когда каждому значению одной переменной соответствует вполне определенное значение другой, то речь идет о функциональной зависимости  $y = f(x)$ . Очень часто между переменными величинами существуют такие зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а множество значений другой переменной (или определенное условное распределение другой переменной). Такая зависимость получила название статистической.

Возникновение такой связи обусловлено тем, что зависимая переменная  $y$  подвержена влиянию неконтролируемых или неучтенных факторов, а также случайным ошибкам.

Когда каждому значению  $x$  одной переменной  $X$  соответствует определенное условное математическое ожидание (среднее значение)  $M_x(Y)$  другой переменной  $Y$ , то такая зависимость называется **корреляционной**:  $M_x(Y) = f(x)$ . Это уравнение называется модельным уравнением парной регрессии или просто **уравнением регрессии**, а её график – линией регрессии. Переменная  $x$  называется независимой переменной, **фактором** или **признаком**, переменная  $y$  – зависимой переменной, функцией отклика или **результативным признаком**.

Для точного описания уравнения регрессии необходимо знать условный закон

распределения переменной  $Y$  при условии, что переменная  $X$  примет значение  $x$ . В статистической практике, как правило, такой информации получить не удастся, а обычно имеется выборка значений заданного объема.

Пусть имеется выборка объема  $m$ :  $(x_1, y_1)$  или  $(x_{i1}, x_{i2}, \dots, x_{in}, y_i)$ ,  $i=1, 2, \dots, m$ , если есть  $n$  факторов. В этом случае речь может идти о приближенном выражении (оценке, аппроксимации) по выборке. Такой оценкой является выборочное **уравнение множественной регрессии**:

$$\hat{y} = f(x_1, x_2, x_3, \dots, x_n, a, b_1, \dots, b_q)$$

где  $x_1, x_2, x_3, \dots, x_n$  – множество факторов,  $n$  – число факторов;

$a, b_1, \dots, b_q$  – параметры (коэффициенты) уравнения, подлежащие определению.

В частном случае, если  $n=1$ , получается уравнение **парной** регрессии  $\hat{y} = f(x, a, b_1, \dots, b_q)$ .

Отклонением (остатком) в  $i$ -ой точке называется выражение  $|\hat{y}_i - y_i| = \varepsilon_i$ .

Нахождение уравнения регрессии включает несколько этапов:

- 1 шаг: спецификация модели, т.е. выбор вида уравнения регрессии;
- 2 шаг: определение наилучших значений параметров;
- 3 шаг: оценка качества модели по критериям.

Рассмотрим некоторые виды уравнения регрессии:

$\hat{y} = a + b_1 x_1 + \dots + b_n x_n$  – линейная зависимость;

$\hat{y} = a \cdot x_1^{b_1} \cdot x_2^{b_2} \dots x_n^{b_n}$  – степенная зависимость;

$\hat{y} = e^{a+b_1 x_1 + \dots + b_n x_n}$  – показательная зависимость.

Для получения уравнения регрессии, используя математические пакеты, требуется выполнять линеаризацию уравнения путем введения вспомогательных переменных. Как пример рассмотрим линеаризацию степенного уравнения:

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \dots x_n^{b_n} \Rightarrow \ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2 \dots + b_n \ln x_n.$$

Обозначим  $Y = \ln y$ ,  $A = \ln a$ ,  $X_i = \ln x_i$ ,  $i = \overline{1, n}$

тогда  $Y = A + b_1 X_1 + \dots + b_n X_n$  – линеаризованное уравнение и  $y = \exp(Y)$ .

Применительно к множественной регрессии необходимо до получения уравнения регрессии произвести отбор факторов  $x_1, x_2, x_3, \dots, x_n$ . Факторы, включаемые в модель, должны быть количественно измеримы и не должны быть интеркоррелированы (коллинеарны). Коллинеарность факторов может привести к тому, что система уравнений для определения параметров может оказаться плохо обусловленной и повлечь за собой неустойчивость и ненадежность оценок параметров уравнения регрессии.

Для исследования коллинеарности факторов можно пользоваться определитель корреляционной матрицы. Если факторы не коррелированы между собой, то  $\text{Det}(R)=1$  (идеальный случай). Предпочтительно, чтобы  $\text{Det}(R)$  был близок к единице. Если факторы **коррелированы (коллинеарны)** между собой, то  $\text{Det}(R)$  близок к 0.

**Корреляционная матрица** и ее определитель находятся по формулам:

$$r_{x_i x_j} = \frac{\overline{x_i x_j} - \overline{x_i} \cdot \overline{x_j}}{\sigma_{x_i} \sigma_{x_j}}, \quad \text{Det}|R| = \begin{vmatrix} r_{x_1 x_1} & r_{x_1 x_2} & r_{x_1 x_3} & \dots \\ r_{x_2 x_1} & r_{x_2 x_2} & r_{x_2 x_3} & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix},$$

где  $\bar{x}_i$  – оценка математического ожидания переменной  $X_i$ ,  $\sigma_{x_i}$  – ее среднеквадратическое отклонение. Корреляционная матрица всегда симметрична, т.е.  $r_{x_i x_j} = r_{x_j x_i}$ .

При отсутствии коллинеарности факторов в определителе нет элементов, отличных от диагональных, близких к единице. Если какой-либо элемент определителя при  $i \neq j$   $r_{x_i x_j} = r_{x_j x_i} \approx 1$ , это означает наличие линейной связи между факторами  $X_i$  и  $X_j$ , т.е. их коллинеарность, что может привести к незначимости (ненадежности) уравнения регрессии.

Значимость уравнения регрессии в целом и значимость его отдельных коэффициентов определяется в результате дисперсионного анализа, при этом используется формула, связывающая общую, факторную и остаточную суммы квадратов отклонений или остатков:

$$\sum_{i=1}^m (y_i - \bar{y})^2 = \sum_{i=1}^m (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$\begin{matrix} \text{общая} & \text{факторная} & \text{остаточная} \\ TSS & = & ESS + RSS \end{matrix}$$

Для проверки значимости уравнения регрессии в целом используются коэффициент детерминации, критерий Фишера, средняя ошибка аппроксимации.

Коэффициент детерминации  $R^2$  определяется, как единица минус доля необъяснённой (остаточной) дисперсии в общей дисперсии результативного признака  $Y$ :

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}.$$

Для значимости уравнения регрессии коэффициент детерминации  $R^2$  должен быть больше 0,95.

Общая, факторная и остаточная дисперсии, которые используются в критерии Фишера, определяются, как суммы квадратов соответствующих отклонений, отнесенные к соответствующим степеням свободы<sup>1</sup>:

$$\frac{TSS}{\underbrace{m-1}_{k_{общ}}} = D_{общ}, \quad \frac{ESS}{\underbrace{n}_{k_{факт}}} = D_{факт}, \quad \frac{RSS}{\underbrace{m-n-1}_{k_{остат}}} = D_{остат}$$

Расчетное значение критерия Фишера  $F$  для значимости уравнения должно быть больше табличного значения  $F_{табл}$  [7] при заданных уровне значимости и степенях свободы, т.е.

$$F = \frac{D_{факт}}{D_{остат}} > F_{табл}(\alpha, k_{факт}, k_{остат}).$$

Средняя ошибка аппроксимации COA - наиболее чувствительный критерий, желательно, чтобы она была больше 8 процентов, т.е.

$$COA = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% > 8\%$$

Входящие в формулы критериев величины могут быть рассчитаны с помощью различных программных средств, в частности, в Excel с использованием надстройки «Анализ данных».

## 1.2. Компонентный и кластерный анализ

<sup>1</sup> число степеней свободы — это количество значений, используемых при расчете статистической характеристики, которые могут свободно изменяться

**Метод главных компонент (PCA - Principal Component Analysis)** или **компонентный анализ** – это технология многомерного статистического анализа, используемая для сокращения размерности пространства признаков с минимальной потерей полезной информации.

Метод главных компонент включается в состав большинства аналитических платформ и широко используется для снижения размерности входных данных на этапе их преобработки.

С математической точки зрения метод главных компонент представляет собой ортогональное линейное преобразование, которое отображает данные из исходного пространства признаков в новое пространство меньшей размерности.

При этом первая ось новой системы координат строится таким образом, чтобы дисперсия данных вдоль неё была бы максимальной. Вторая ось строится ортогонально первой так, чтобы дисперсия данных вдоль неё, была бы максимальной из оставшихся возможных и т.д. Первая ось называется первой главной компонентой, вторая – второй и т.д.

Компонентный анализ позволяет определить структурную зависимость между факторами. Идея метода заключается в замене коррелированных переменных (факторов) новыми переменными (главными компонентами), между которыми корреляция отсутствует. В результате его использования получается сжатое описание меньшего объема, несущее почти всю информацию, содержащуюся в исходных данных. Главные компоненты, получаются из исходных переменных как линейные комбинации исходных переменных.

#### **Алгоритм компонентного анализа:**

- Отыскание собственных значений корреляционной матрицы данных, центрированных и нормированных по столбцам.
- Выбор среди собственных значений существенных ( $>1$ ).
- Отыскание соответствующих собственных векторов (главных компонент) и матрицы нагрузок, столбцами которой являются собственные векторы.
- Отыскание матрицы счетов (весов), строки которой – ортогональные проекции строк исходных данных на главные компоненты.

Матрица счетов является искомой матрицей сжатых данных.

Рассмотрим расчетные формулы для проведения компонентного анализа.

Пусть есть выборка данных. Например, выборка может быть по людям, растениям, документам, изображениям, звуковым файлом или по чему-то еще, что можно описать с помощью набора числовых данных. Строки всегда соответствуют отдельным объектам, столбцы соответствуют признакам или факторам, например, цвет, длина, вес и т.д. Значения признаков обычно представляют собой вещественные числа, но в некоторых случаях они могут быть булевыми или иметь дискретные значения.

Пусть  $X$  – числовая матрица данных размерности  $(m \times n)$ , элементы которой  $x_{ij}$ . Таковую матрицу называют также матрицей признаков или факторов.

$X_0$  – **стандартизованная матрица** исходных данных, центрированных и нормированных по столбцам. Ее элементы вычисляются по формулам  $x_{0ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{xj}}$ .

**Корреляционная матрица** размерности  $(n \times n)$ , центрированных и нормированных по столбцам исходных данных, вычисляется по формуле  $R(X_0) = \frac{X_0^T \cdot X_0}{m-1}$ .

Пусть  $\lambda_1, \dots, \lambda_n$  – **собственные значения** матрицы  $R(X_0)$ . Они определяются из условия  $\text{Det}(R - \lambda E) = 0$ . Соответствующие собственные векторы  $p$  находятся как решения однородной системы линейных алгебраических уравнений:  $(R - \lambda E)p = 0$ .

Матрица  $P$  размерности  $(n \times k)$ , столбцами которой являются  $k$  собственных векторов, соответствующих  $k$  существенным собственным значениям ( $\lambda_i > 1$ ), называется **матрицей нагрузок** или матрицей главных компонент.



**Матрица счетов**, определяемая формулой  $T = X_0 \cdot P$ , размерности  $(m \times k)$  – это сжатая матрица данных. Строки матрицы счетов – проекции строк  $X_0$  на векторы главных компонент.

Собственные значения  $\lambda_1, \dots, \lambda_p$  матрицы  $R(X_0)$  равны **дисперсиям столбцов** матрицы счетов и показывают вклад соответствующей компоненты в суммарную дисперсию. Данные, спроецированные на главные компоненты, характеризуются наибольшими дисперсиями (рассеяниями), поэтому позволяют сохранить максимум информации по исходным данным, при этом уменьшив размерность задачи.

Для оценки погрешности, допускаемой при уменьшении размерности исходных данных методом PCA, используется **матрица ошибок**  $E = X_0 - T \cdot P^T$  – матрица ошибок размерности  $(m \times p)$ , которая должна быть близка к 0-матрице, если нет потери данных при их сжатии.

**Кластерный анализ** предназначен для разбиения исходных данных на группы, таким образом, чтобы элементы, входящие в одну группу, были максимально «схожи», а элементы из разных групп были максимально «отличными» друг от друга.

Кластерный анализ – группа методов, используемых для классификации объектов или событий в относительно гомогенные (однородные) группы, которые называют **кластерами** (clusters).

В компонентном анализе группируются столбцы, т.е. цель – анализ структуры множества факторов и выявление обобщенных факторов. В кластерном анализе группируются строки, т.е. цель – анализ структуры множества объектов.

Допустима геометрическая интерпретация кластеризации: каждый объект (респондент) – точка в пространстве признаков. Задача кластерного анализа – выделение «сгущений» точек, разбиение совокупности на однородные подмножества объектов. На практике трудно выделить границу кластеров (рис. 1.1).

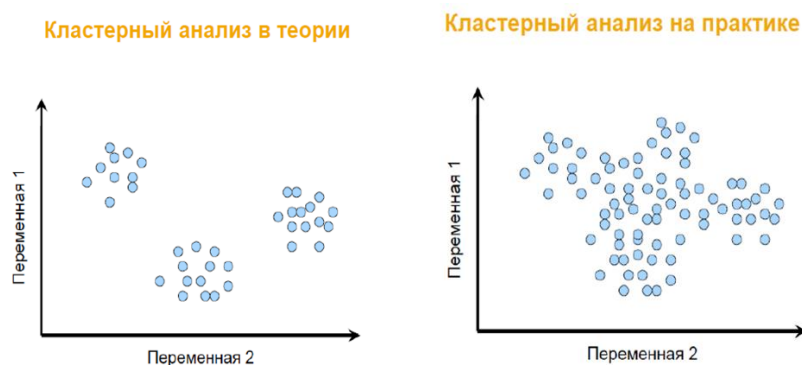


Рисунок 1.1. Кластеризация в теории и на практике

Существуют различные методы кластеризации. Можно выделить два основных типа методов, алгоритмы которых несколько различны (рис. 1.2).



Рисунок 1.2. Виды кластеризации

### Алгоритм иерархической кластеризации

- Первоначально каждый объект образует свой отдельный кластер.
- На каждом последующем шаге два соседних (близких по выбранной мере) кластера объединяются в один.
- Этот процесс продолжается до тех пор, пока не останется только один кластер либо расстояние между кластерами начинает увеличиваться скачкообразно, тогда процесс объединения останавливают.

Мера их близости между отдельными объектами или кластерами – расстояние, для определения которого применяются различные подходы.

Для определения расстояния между объектами  $X_i$  и  $X_j$  чаще используется евклидово расстояние  $d = \sqrt{\sum_{k=1}^n (x_{ki} - x_{kj})^2}$ , где  $n$  – количество факторов или признаков.

Для определения расстояния между кластерами применяются различные подходы:

- Межгрупповая связь (Between-groups linkage) – этот метод в программах обычно устанавливается по умолчанию.
- Ближайший сосед (Nearest neighbor).
- Самый дальний сосед (Furthest neighbor).
- Центроидная кластеризация (Centroid clustering).
- Медианная кластеризация (Median clustering).

Прежде чем проводить кластеризацию обычно выполняется стандартизация данных, в частности применяется Z-стандартизация, рассматриваемая при компонентном анализе

$$x_{oj} = \frac{x_{ij} - \bar{x}_j}{\sigma_{xj}}.$$

### Алгоритм кластеризации методом к-средних

- Случайным образом создаются  $k$  точек - центров кластеров.
- Каждому объекту из матрицы данных ставится в соответствии ближайший к ней центр кластера.
- Вычисляются средние арифметические точек, принадлежащих к определённому кластеру. Эти значения становятся новыми центрами кластеров.
- Шаги 2 и 3 повторяются до тех пор, пока пересчёт центров кластеров будет иметь смысл. Как только высчитанные центры кластеров практически совпадут с предыдущими, алгоритм будет окончен.

### Сравнение кластеризации методом к-средних и иерархической кластеризации

Недостатки кластеризации методом к-средних:

- Чувствительность к выбросам.
- Необходимо заранее задавать количество кластеров, а не как в иерархическом анализе, получать это в качестве результата.

Достоинства:

- Применение при большом объеме данных.
- Простота использования: в качестве метрики используется евклидово расстояние.
- Возможность наглядной интерпретации кластеров с использованием графика «Средних значений в кластерах».

Выполнение кластеризации, даже при относительно небольших объемах данных, требует большого объема вычислений, в математических пакетах существуют встроенные процедуры для этого, однако для понимания алгоритма и трактования результатов, получаемых с помощью этих процедур, лучше изначально выполнить кластеризацию вручную на малом объеме данных, что и предлагается сделать в задании 4.5.

### 1.3. Математические методы планирования экспериментов

**Эксперимент (опыт)** – это совокупность операций, совершаемых над объектом, с целью получения информации о его свойствах. Теория планирования эксперимента тесно связана с многомерными статистическими методами. Эта теория включает математические методы планирования экспериментов, которые дают возможность управлять условиями проведения эксперимента с целью повышения его эффективности. Эксперименты, основывающиеся методах планирования экспериментов, принято называть **активными**. Эксперименты, которые ставятся для решения оптимизационных задач, называются **экстремальными**.

**Цели планирования экспериментов:**

- нахождение условий проведения экспериментов, при которых удастся получить достоверную информацию об исследуемом объекте с наименьшими затратами (при минимальном количестве опытов);
- представление информации в удобной форме для ее извлечения и обработки.

**Задачи планирования экспериментов:**

- построение математической модели изучаемого объекта, включающей показатели качества объекта, зависящие от параметров (факторов), для оценки его свойств;
- нахождение такой комбинации факторов, при которой достигаются экстремальные значения выбранных показателей качества;
- автоматизация решения задач управления экспериментом (планирования, оценки погрешностей, прогнозирования).

Методы и подходы к планируемым экспериментам достаточно хорошо разработаны в научно-прикладных исследованиях. В исследованиях используются различные вычислительные инструменты и инструменты анализа от стандартных математических и статистических пакетов общего назначения (MATHCAD, MATLAB, STATISTICA, SPSS, STATGRAPHICS и др.) до специализированных авторских пакетов, как правило, ориентированных на выбранную предметную область.

Данное пособие поможет приобрести навыки использования моделей и методов планирования экспериментов при решении практических задач из профессиональной области.

#### Математическая модель планирования эксперимента

При планировании эксперимента, следуя кибернетическому подходу, исследуемый объект представляется «черным ящиком», входные параметры которого  $X_1, X_2, \dots, X_n$  – **факторы**, а выходной параметр (или параметры)  $y$  – **отклик**. Выходной параметр является искомым показателем качества исследуемого объекта.

Зависимость выходного параметра или выходных параметров, если их несколько, от входных выполняются либо с помощью экспериментов на моделях (аналитических, численных, имитационных и др.), либо путем натурных экспериментов.

В математической теории планирования экспериментов выходной параметр объекта представляется функцией от входных параметров, которая определяется в виде некоторой регрессионной зависимости.

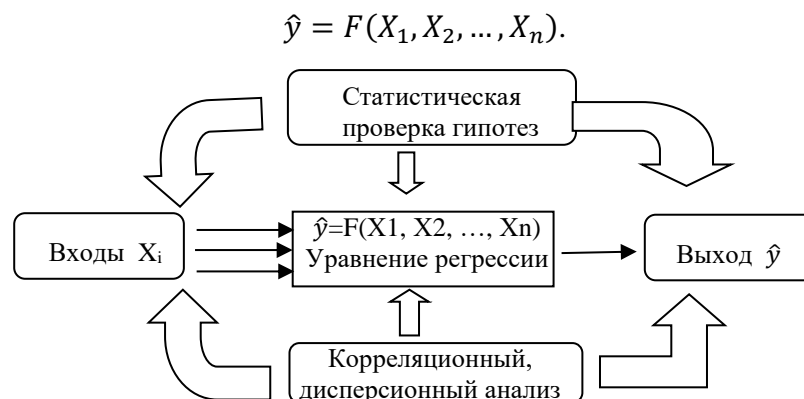


Рисунок 1.3. Математическая модель планирования эксперимента

При этом для получения более полной и достоверной картины зависимости отклика от факторов регрессионный анализ дополняется статистическими исследованиями: корреляционным, дисперсионным анализом, статистической проверкой гипотез (рис. 1.3).

Для построения регрессионной модели используются **планы** (наборы значений входных параметров), которые позволяют с наименьшими затратами (при наименьшем количестве экспериментов) определить регрессионную зависимость и оценить ее значимость.

Рассмотрим основные понятия и теоретические положения теории планирования экспериментов.

**Уровнями факторов**  $X_i$  называется набор значений, которые этот фактор принимает, и которые используются при проведении экспериментов. Для построения эффективной математической модели целесообразно предварительно провести анализ значимости факторов, их возможной коллинеарности, исключить малозначимые факторы.

Допустим известны диапазоны изменения факторов  $[X_{imin}, X_{imax}]$ ,  $i = \overline{1, n}$ . Областью планирования или **факторным пространством** называется многомерный параллелепипед, ограниченный диапазонами изменения факторов.

**Экспериментом, с точки зрения математической модели**, называется такое сочетание значений различных факторов, которое соответствует условиям проведения опыта. Формально заданию условий эксперимента соответствует задание  $n$ -мерной точки  $M^j(X_1^j, X_2^j, \dots, X_n^j)$  в факторном пространстве, координаты которой – значения факторов:

$$X_i^j \in [X_{imin}, X_{imax}], i = \overline{1, n},$$

$X_i^j$  – значение  $i$ -го фактора в натуральном измерении для  $j$ -го эксперимента.

Обозначим количество точек в факторном пространстве или количество экспериментов через  $N$ , тогда  $j = \overline{1, N}$ .

Как правило, точки не выходят за пределы факторного пространства, исключение составляют композиционные планы, о которых будет говориться ниже. Совокупность всех точек эксперимента называется **планом эксперимента**.

#### Кодирование переменных и уравнение регрессии

Изначально значения факторов  $X_i^j$  рассматриваются в натуральном измерении, затем переходят к кодированным (безразмерным, приведенным) значениям факторов. Будем

обозначать их строчными буквами  $x_i^j$ . Значения кодированных переменных для некоординатных планов изменяются в пределах<sup>2</sup> от -1 до +1:

$$x_i^j \in [-1, 1], i = \overline{1, n}, j = \overline{1, N},$$

где  $x_i^j$  – значение кодированного фактора  $x_i$  в j-ой точке плана.

**Формулы перехода к кодированным переменным** имеют вид

$$x_i^j = \frac{X_i^j - X_{iavg}}{X_{iavg} - X_{imin}}, \quad X_{iavg} = (X_{imin} + X_{imax})/2,$$

Координаты  $X_i^j$  точек плана в натуральном измерении (обратный переход) вычисляются по формулам

$$X_i^j = (X_{iavg} - X_{imin})x_i^j + X_{iavg},$$

где  $i = \overline{1, n}, j = \overline{1, N}$ ;

**Центром эксперимента** называется точка, в которой все  $x_i^j = 0$ .

В качестве функции отклика (регрессионной модели)  $f(x_1, x_2, \dots, x_n)$  от кодированных переменных наиболее часто используются **полиномиальные модели первой и второй степени**:

$$\hat{y} = f(x_1, x_2, \dots, x_n) = b_0 + \sum_{i=1}^n b_i x_i + \sum_{\substack{i,j=1 \\ i \neq j}}^n b_{ij} x_i x_j + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^n b_{ijk} x_i x_j x_k + \sum_{i=1}^n b_{ii} x_i^2, \quad ,$$

где  $x_i$  – кодированные переменные полинома.

Введем дополнительные переменные для линеаризации уравнения:

$$x_{n+1} = x_1 x_2, \quad x_{n+2} = x_1 x_3, \dots,$$

тогда полином запишется в линеаризованном виде

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i,$$

где  $k$  – количество коэффициентов при неизвестных в полиноме.

### Определение коэффициентов функции отклика

**Основной матрицей** плана называется матрица размерности N на n

$$X_0 = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^N & x_2^N & \dots & x_n^N \end{bmatrix}.$$

Количество строк матрицы совпадает с количеством экспериментов (точек плана), а количество столбцов – с количеством факторов.

Для удобства расчетов используется также расширенная матрица плана или матрица смешанных факторов, часть столбцов которой являются столбцами основной матрицы плана.

**Расширенной матрицей плана в кодированных переменных** называется матрица X размерности N на  $(k+1)$ , элементы которой  $x_i^j$ , где  $k$  – количество коэффициентов при неизвестных линеаризованного уравнения регрессии. Первый столбец расширенной матрицы состоит из единиц, следующие n столбцов соответствуют основной матрице плана, а остальные столбцы получаются в результате произведений столбцов матрицы основного плана, соответствующих дополнительным переменным  $x_{n+1}, x_{n+2}, \dots, x_k$ :

<sup>2</sup> Значения кодированных переменных могут также изменяться от 0 до 1, но в данном пособии эти случаи не рассматриваются.

$$X = \begin{bmatrix} x_0^1 & x_1^1 & x_2^1 & \dots & x_n^1 \dots x_k^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_0^N & x_1^N & x_2^N & \dots & x_n^N \dots x_k^N \end{bmatrix}.$$

Обозначим  $Y$  – матрицу-столбец значений рассматриваемого показателя качества по результатам эксперимента в  $N$  точках,  $B$  – матрицу-столбец неизвестных коэффициентов полинома:

$$Y^T = (y^{1exp}, y^{2exp}, \dots, y^{Nexp}),$$

$$B^T = (b_0, b_1, \dots, b_k),$$

где  $T$  означает операцию транспонирования матрицы.

Нахождение столбца коэффициентов полинома выполняется следующим образом:

$$Y = X \cdot B \Rightarrow B = (X^T X)^{-1} (X^T Y),$$

что соответствует формулам, полученным по методу наименьших квадратов.

Обозначим  $\hat{Y}$  – матрицу значений показателя качества в точках плана, найденных по уравнению регрессии, т.е.  $\hat{Y} = X \cdot B$ , тогда  $\hat{Y}^T = (y^1, y^2, \dots, y^N)$ .

Значения  $y^j$  ( $j = \overline{1, N}$ ) выходного параметра в точках плана в случае ненасыщенного плана, т.е. при  $N > k+1$ , не совпадают с  $y^{jexp}$ . Качество полученной регрессионной модели оценивается различными методами: с помощью критерия Фишера, по средней ошибке аппроксимации и др. При расчете критерия Фишера используются остаточная и факторная дисперсии, отнесенные к числу степеней свободы (см. раздел 1.1, стр. 7). При расчете **средней ошибки аппроксимации** используется формула

$$COA = \frac{\sum_{i=1}^N |(y^i - y^{iexp}) / y^{iexp}|}{N}.$$

### Виды планов и их характеристика

Планы различаются по своим характеристикам. Основными характеристиками являются насыщенность, ротатабельность, ортогональность, нормированность, симметричность, композиционность.

План называется **насыщенным**, если число экспериментальных точек  $N$  в нем совпадает с числом коэффициентов полинома  $k+1$ . Насыщенный план не позволяет усреднить ошибки эксперимента, если они есть, и определить их величины. Избыточное число опытов при  $N > k+1$  позволяет это сделать.

План называется **ротатабельным**, если дисперсия функции отклика  $f(x_1, x_2, \dots, x_n)$  одинакова на одинаковом расстоянии от центра плана, т.е. точки плана лежат на концентрических гипербсферах.

План называется **ортогональным**, если столбцы матрицы планирования попарно ортогональны

$$\sum_{U=1}^N x_i^U x_j^U = 0, \quad i, j = \overline{0, k}, i \neq j.$$

Для ортогонального плана матрица  $C = X^T X$  является диагональной, поэтому коэффициенты полинома определяются независимо друг от друга.

План называется **нормированным**, если сумма квадратов элементов любого столбца равна числу опытов  $N$ :

$$\sum_{U=1}^N (x_i^U)^2 = N, \quad i = \overline{0, k}.$$

План называется **симметричным**, если сумма элементов любого столбца, кроме первого, равна 0:

$$\sum_{U=1}^N x_i^U = 0, \quad i = \overline{1, k}.$$

Если план позволяет определить коэффициенты полиномов только при первых степенях  $x_i$  и смешанных произведениях факторов, то такой план называется **планом первого порядка**.

План называется **планом второго порядка**, если он позволяет найти коэффициенты полиномов при квадратах переменных  $x_i^2$ . Для таких планов характерно, что факторы варьируются, как минимум, на трех уровнях.

План называется **композиционным**, если он включает дополнительные точки к планам первого порядка.

#### 1.4. Полный факторный план

Формула, по которой рассчитывается число опытов при полном факторном эксперименте (ПФЭ), выглядит следующим образом:

$$N = p^n,$$

где  $N$  – число опытов;  
 $n$  – число факторов;  
 $p$  – число уровней варьирования факторов.

Если число уровней каждого фактора равно двум, то мы имеем полный факторный эксперимент типа ПФЭ  $2^n$

В ПФЭ  $2^n$  используются кодированные значения факторов: +1 и -1, соответствующие наибольшему и наименьшему значениям факторов. Условия эксперимента записываются в виде таблицы, где строки соответствуют различным опытам, а столбцы – значениям факторов. Такая таблица соответствует матрице планирования эксперимента. План ПФЭ обладает свойствами симметричности, нормированности, ортогональности, ротатабельности (для линейной модели).

На рисунке 1.4 приведена матрица расширенного плана ПФЭ  $2^3$ , учитывающая смешанные произведения (взаимодействия) факторов.

| $i$                   | 0     | 1     | 2     | 3     | 4        | 5        | 6        | 7           | 8     |
|-----------------------|-------|-------|-------|-------|----------|----------|----------|-------------|-------|
| $U$                   | $x_0$ | $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_1x_3$ | $x_2x_3$ | $x_1x_2x_3$ | $Y$   |
| 1                     | 1     | -1    | -1    | -1    | +1       | +1       | +1       | -1          | $Y_1$ |
| 2                     | 1     | +1    | -1    | -1    | -1       | -1       | +1       | +1          | $Y_2$ |
| 3                     | 1     | -1    | +1    | -1    | -1       | +1       | -1       | +1          | $Y_3$ |
| 4                     | 1     | +1    | +1    | -1    | +1       | -1       | -1       | -1          | $Y_4$ |
| 5                     | 1     | -1    | -1    | +1    | +1       | -1       | -1       | +1          | $Y_5$ |
| 6                     | 1     | +1    | -1    | +1    | -1       | +1       | -1       | -1          | $Y_6$ |
| 7                     | 1     | -1    | +1    | +1    | -1       | -1       | +1       | -1          | $Y_7$ |
| 8                     | 1     | +1    | +1    | +1    | +1       | +1       | +1       | +1          | $Y_8$ |
| $\sum_{U=1}^N x_{iU}$ | 8     | 0     | 0     | 0     | 0        | 0        | 0        | 0           |       |

Рисунок 1.4. Расширенная матрица плана ПФЭ  $2^3$

Ортогональность плана означает, что матрица  $X^T X$  диагональна, тогда коэффициенты полинома определяются независимо друг от друга по формулам

$$b_i = (\sum_{j=1}^N y^{jexp} x_i^j) / N, i = \overline{0, k}.$$

План ПФЭ  $2^n$  относится к планам первого порядка. Количество коэффициентов полинома от  $n$  факторов, содержащего свободный член, первые степени неизвестных и их смешанные произведения, равно  $C_n^0 + C_n^1 + C_n^2 + \dots + C_n^n = 2^n$ . Оно всегда совпадает с числом точек плана ПФЭ  $2^n$ , поэтому этот план является насыщенным.

Так, при  $n=3$  количество коэффициентов и число точек плана одинаково и равно 8. **Регрессионный полином ПФЭ  $2^3$**  для функции отклика имеет вид:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1x_2 + b_5x_1x_3 + b_6x_2x_3 + b_7x_1x_2x_3,$$

На практике при неколлинеарных факторах коэффициенты при смешанных произведениях близки к нулю, поэтому имеет смысл учитывать только линейные члены. В этом случае при  $n > 2$  количество опытов превышает количество коэффициентов при первых степенях факторов. С одной стороны, это позволяет усреднить данные эксперимента  $y^{jexp}$  и оценить ошибки аппроксимации. С другой стороны, в ПФЭ при увеличении  $n$  разность между числом опытов  $2^n$  и числом коэффициентов  $n+1$  линейной части, увеличивается, план становится избыточным по количеству опытов, возникает проблема уменьшения количества опытов.

### Дробный факторный план

Для того, чтобы уменьшить число опытов, применяются дробные факторные планы.

Дробным факторным экспериментом называется система экспериментов, представляющих собой часть ПФЭ, позволяющая рассчитывать коэффициенты уравнения регрессии и сократить объем экспериментальных данных. Такие эксперименты обладают меньшей информативностью, но позволяют значительно сократить количество опытов. В ДФЭ количество экспериментальных точек  $N = 2^{n-m}$ , где  $m$  – показатель дробности ПФЭ. Если  $m = 1$ , то  $N$  в два раза меньше, чем в ПФЭ. Такие планы называются полурепликами.

### 1.5. Ортогональный центральный композиционный план

Если характер функции отклика нелинейный, то следует учитывать в уравнении регрессии слагаемые, содержащие квадраты факторов. Для этого, как минимум, каждый фактор должен варьироваться на трех уровнях. Использование ПФЭ в этом случае приводит к неоправданному увеличению числа экспериментальных точек. Другой подход, основанный на композиционных планах, предложен Боксом и Уилсоном. Этот подход используется и при построении планов второго порядка.

Обозначим ОЦКП – центральный симметричный ортогональный композиционный план.

В ОЦКП входят:

- ядро – план ПФЭ с  $N_0 = 2^n$  точками плана;
- $n_0$  – центральная точка плана;
- по две «звездные» точки для каждого фактора.

Общее количество точек в плане ОЦКП вычисляется по формуле:

$$N = 2^n + 2n + n_0, \text{ где } n_0 = 1.$$

В ОЦКП каждый кодированный фактор варьируется на пяти уровнях, т.е.  $x_i^j$  ( $-\alpha, -1, 0, 1, \alpha$ ),  $i = \overline{1, n}, j = \overline{1, N}$ , где  $\alpha$  – плечо звездных точек.

Ортогональность плана обеспечивается преобразованием столбцов, соответствующих квадратам факторов, и соответствующим выбором значений координат звездных точек. Расчетные формулы для параметров, обеспечивающих ортогональность, имеют вид

$$a = \sqrt{2^n/N}, \quad \alpha = \sqrt{(\sqrt{N} \cdot 2^n - 2^n)/2}.$$

Вычисленные значения параметров приведены на рисунке 1.5



|          |       |       |       |       |       |       |       |
|----------|-------|-------|-------|-------|-------|-------|-------|
| $n$      | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
| $\alpha$ | 1     | 1,215 | 1,414 | 1,596 | 1,761 | 1,909 | 2,045 |
| $a$      | 0,667 | 0,73  | 0,8   | 0,86  | 0,91  | 0,946 | 0,968 |
| $N$      | 9     | 15    | 25    | 43    | 77    | 143   | 273   |

Рисунок 1.5. Значения параметров плана ОЦКП

На рисунке 1.6 приведена матрица плана ОЦКП при  $n = 3$  в общем виде, учитывающая смешанные произведения (взаимодействия) факторов, их квадраты. На следующем рисунке 1.7 представлена эта матрица, рассчитанная в Excel.

|   | $U$ | $x_0$ | $x_1$             | $x_2$     | $x_3$     | $x_1x_2$ | $x_1x_3$ | $x_2x_3$ | $x_1x_2x_3$ | $x_4 = x_1^2 - a$  | $x_5 = x_2^2 - a$ | $x_6 = x_3^2 - a$ | $Y$      |
|---|-----|-------|-------------------|-----------|-----------|----------|----------|----------|-------------|--|-------------------|-------------------|----------|
| Точки плана ПФЭ $2^3$<br>( $N=2^n$ точек) | 1   | +1    | -1                | -1        | -1        | +1       | +1       | +1       | -1          | $1-a$  | $1-a$             | $1-a$             | $Y_1$    |
|   | 2   | +1    | +1                | -1        | -1        | -1       | -1       | +1       | +1          | $1-a$  | $1-a$             | $1-a$             | $Y_2$    |
|   | 3   | +1    | -1                | +1        | -1        | -1       | +1       | -1       | +1          | $1-a$  | $1-a$             | $1-a$             | $Y_3$    |
|   | 4   | +1    | +1                | +1        | -1        | +1       | -1       | -1       | -1          | $1-a$  | $1-a$             | $1-a$             | $Y_4$    |
|   | 5   | +1    | -1                | -1        | +1        | +1       | -1       | -1       | +1          | $1-a$  | $1-a$             | $1-a$             | $Y_5$    |
|   | 6   | +1    | +1                | -1        | +1        | -1       | +1       | -1       | -1          | $1-a$  | $1-a$             | $1-a$             | $Y_6$    |
|   | 7   | +1    | -1                | +1        | +1        | -1       | -1       | +1       | -1          | $1-a$  | $1-a$             | $1-a$             | $Y_7$    |
|   | 8   | +1    | +1                | +1        | +1        | +1       | +1       | +1       | +1          | $1-a$  | $1-a$             | $1-a$             | $Y_8$    |
| Звездные точки<br>( $2n$ точек)           | 9   | +1    | $-\alpha$         | 0         | 0         | 0        | 0        | 0        | 0           | $\alpha^2 - a$   | $-a$              | $-a$              | $Y_9$    |
|   | 10  | +1    | $+\alpha$         | 0         | 0         | 0        | 0        | 0        | 0           | $\alpha^2 - a$   | $-a$              | $-a$              | $Y_{10}$ |
|   | 11  | +1    | 0                 | $-\alpha$ | 0         | 0        | 0        | 0        | 0           | $-a$   | $\alpha^2 - a$    | $-a$              | $Y_{11}$ |
|   | 12  | +1    | 0                 | $+\alpha$ | 0         | 0        | 0        | 0        | 0           | $-a$   | $\alpha^2 - a$    | $-a$              | $Y_{12}$ |
|   | 13  | +1    | 0                 | 0         | $-\alpha$ | 0        | 0        | 0        | 0           | $-a$   | $-a$              | $\alpha^2 - a$    | $Y_{13}$ |
|   | 14  | +1    | 0                 | 0         | $+\alpha$ | 0        | 0        | 0        | 0           | $-a$   | $-a$              | $\alpha^2 - a$    | $Y_{14}$ |
| Нулевая точка                             | 15  | +1    | 0                 | 0         | 0         | 0        | 0        | 0        | 0           | $-a$   | $-a$              | $-a$              | $Y_{15}$ |
| $\sum_{i=1}^N x_{iU}$                     | -   | $N$   | 0                 | 0         | 0         | 0        | 0        | 0        | 0           | 0  | 0                 | 0                 |          |
| $\sum_{i=1}^N x_{iU}^2$                   | -   | $N$   | $2^* + 2\alpha^2$ |           |           | $2^n$    |          |          |             | $2^* (1-a)^2 + 2(\alpha^2 - a)^2 + a^2 (2n - 2) + n_0 a^2$ |                   |                   |          |

Рисунок 1.6. Расширенная матрица плана ОЦКП при  $n=3$

| №  | $x_0$ | $x_1$  | $x_2$  | $x_3$  | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|----|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|----------|
| 1  | 1     | -1     | -1     | -1     | 1     | 1     | 1     | -1    | 0,27  | 0,27  | 0,27     |
| 2  | 1     | 1      | -1     | -1     | -1    | -1    | -1    | 1     | 0,27  | 0,27  | 0,27     |
| 3  | 1     | -1     | 1      | -1     | -1    | 1     | -1    | 1     | 0,27  | 0,27  | 0,27     |
| 4  | 1     | 1      | 1      | -1     | 1     | -1    | -1    | -1    | 0,27  | 0,27  | 0,27     |
| 5  | 1     | -1     | -1     | 1      | 1     | -1    | -1    | 1     | 0,27  | 0,27  | 0,27     |
| 6  | 1     | 1      | -1     | 1      | 1     | -1    | 1     | -1    | 0,27  | 0,27  | 0,27     |
| 7  | 1     | -1     | 1      | 1      | -1    | -1    | 1     | -1    | 0,27  | 0,27  | 0,27     |
| 8  | 1     | 1      | 1      | 1      | 1     | 1     | 1     | 1     | 0,27  | 0,27  | 0,27     |
| 9  | 1     | -1,215 | 0      | 0      | 0     | 0     | 0     | 0     | 0,75  | -0,73 | -0,73    |
| 10 | 1     | 1,215  | 0      | 0      | 0     | 0     | 0     | 0     | 0,75  | -0,73 | -0,73    |
| 11 | 1     | 0      | -1,215 | 0      | 0     | 0     | 0     | 0     | -0,73 | 0,75  | -0,73    |
| 12 | 1     | 0      | 1,215  | 0      | 0     | 0     | 0     | 0     | -0,73 | 0,75  | -0,73    |
| 13 | 1     | 0      | 0      | -1,215 | 0     | 0     | 0     | 0     | -0,73 | -0,73 | 0,75     |
| 14 | 1     | 0      | 0      | 1,215  | 0     | 0     | 0     | 0     | -0,73 | -0,73 | 0,75     |
| 15 | 1     | 0      | 0      | 0      | 0     | 0     | 0     | 0     | -0,73 | -0,73 | -0,73    |

Рисунок 1.7. Расширенная матрица плана ОЦКП при  $n=3$ , вычисленная в Excel

Вид линейризованного полинома функции отклика:

$$y = b_0 + \sum_{i=1}^k b_i x_i,$$

где  $k = C_n^1 + C_n^2 + \dots + C_n^n + n = 2^n + n - 1$ ,  $k$  – количество коэффициентов при неизвестных;

$$x_{n+1} = x_1 x_2, \quad x_{n+2} = x_1 x_3, \dots, \quad x_{k-1} = x_{n-1}^2 - a, \quad x_k = x_n^2 - a,$$

$x_1, x_2, \dots, x_n$  – кодированные переменные;

Коэффициенты полинома вычисляются по формулам:

$$b_0 = \frac{1}{N} \sum_{j=1}^N y^{jexp}, \quad b_i = \frac{(\sum_{j=1}^N y^{jexp} x_i^j)}{(\sum_{j=1}^N (x_i^j)^2)}, \quad i = \overline{1, k},$$

где  $x_i^j$  – значение фактора  $x_i$  в  $j$ -ой точке плана;

$y^{jexp}$  – экспериментальное значение критерия в  $j$ -ой точке плана.

При  $n = 3$  число точек плана  $N=15$ , количество коэффициентов полинома равно 11. Для звездных точек  $\alpha=1,25$ . Эти точки незначительно выходят за пределы параллелепипеда факторного пространства,

**Регрессионный полином** для функции отклика в ОЦКП имеет вид

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_1 x_2 + b_5 x_1 x_3 + b_6 x_2 x_3 + b_7 x_1 x_2 x_3 + b_8 (x_1^2 - a) + b_9 (x_2^2 - a) + b_{10} (x_3^2 - a)$$

## 2. КОНТРОЛЬНЫЕ ВОПРОСЫ ПО ДИСЦИПЛИНЕ

### «Модели и методы планирования экспериментов»

1. Статистический анализ. Генеральная совокупность и выборка. Репрезентативность выборки.
2. Дискретный и интервальный вариационные ряды распределения.
3. Точечные и интервальные оценки характеристик распределения.
4. Несмещенность, эффективность, состоятельность оценок.
5. Корреляционный и регрессионный анализ. Корреляционная зависимость.
6. Уравнение парной и множественной регрессии.
7. Виды уравнений нелинейной регрессии. Линеаризация уравнения.
8. Корреляционная матрица. Коллинеарность факторов.
9. Дисперсионный анализ. Общая, факторная и остаточная суммы, связь между ними.
10. Критерии значимости уравнения регрессии в целом (коэффициент детерминации, критерий Фишера).
11. Критерий средней ошибки аппроксимации.
12. Инструментальные средства для проведения регрессионного, корреляционного, дисперсионного анализа.
13. Компонентный анализ, его суть и алгоритм.
14. Корреляционная матрица стандартизованных данных.
15. Матрица нагрузок при компонентном анализе.
16. Матрицы счетов и ошибок при компонентном анализе.
17. Кластерный анализ, виды кластеризации.
18. Иерархическая кластеризация, ее алгоритм. Методы нахождения расстояния между объектами и кластерами.
19. Кластеризация методом  $k$ -средних, ее алгоритм. Сравнительный анализ методов кластеризации.
20. Инструментальные средства для проведения компонентного и кластерного анализа.
21. Активные эксперименты. Цели и задачи планирования экспериментов.
22. Математическая модель планирования эксперимента.
23. Уровни факторов, факторное пространство. Эксперимент с точки зрения математической модели.
24. Кодирование переменных.
25. Регрессионная модель в виде полинома.
26. Расширенная и основная матрица плана.

27. Оценка качества регрессионной модели по средней ошибке аппроксимации.
28. Основные характеристики планов.
29. Полный факторный план ПФЭ  $2^n$ , его матрица.
30. Регрессионный полином при ПФЭ  $2^3$  для функции отклика.
31. Дробный факторный план.
32. Ортогональный центральный композиционный план, его матрица.
33. Регрессионный полином при ОЦКП для функции отклика.

### 3. ЗАДАНИЯ ЛАБОРАТОРНЫХ (КОНТРОЛЬНЫХ) РАБОТ И ТРЕБОВАНИЯ ПО ИХ ВЫПОЛНЕНИЮ

Отчеты по лабораторным работам (контрольная работа ЗФО) оформляется в печатном виде согласно требованиям к оформлению письменных работ обучающихся в ДГТУ, в соответствии со своим вариантом (порядковый номер в электронной ведомости) и включает:

- титульный лист;
- постановку задач исследования (формулировку заданий);
- по каждому заданию описание хода работы (проводимых расчетов и исследований с указанием используемых формул, алгоритмов, методов и инструментальных средств);
- результаты и выводы по каждому заданию;
- перечень используемых источников литературы и интернет-источников.

*Замечание.* Кроме сформулированных заданий и числовых данных по вариантам желательно сформулировать собственную задачу (индивидуальное задание), связанное с методами планирования экспериментов из интересующей предметной области, самостоятельно выбрать числовые данные к ней.

В примерах выполнения заданий лабораторных работ, приведенных в данном пособии, используются Excel, Matlab, Colab. При выполнении лабораторных работ, контрольной работы и индивидуального задания допускается применение алгоритмических языков, инструментариев, а также специализированных программ для обработки статистических данных, отличных от упомянутых, используемых для решения задач математического планирования экспериментов.

#### *Постановка задачи 1*

Результативный признак  $y$  (время выполнения запроса) при работе с базой данных зависит от трех комплексных факторов  $x_1, x_2, x_3$ . Фактор  $x_1$  определяет размер, количество файлов, которые нужно соединить для выполнения запроса; фактор  $x_2$  определяет структуру меню базы (число его уровней, их состав и расположения по иерархии); фактор  $x_3$  определяется характеристиками оператора-пользователя. На основе статистических измерений получена таблица зависимости значений результативного признака  $y$  от значений факторов  $x_1, x_2, x_3$ .

Требуется разработать математическую модель, позволяющую оценивать эффективность работы информационной системы (базы данных), влияние различных факторов на время выполнения запроса.

Разработка математической включает следующие этапы ее решения:

- Корреляционный и регрессионный анализ (задание 1).
- Автоматизация анализа данных (задание 2).
- Компонентный анализ (задание 3).
- Кластерный анализ (задание 4).

#### *Постановка задачи 2*

Известно, что результативный признак  $Y$  зависит от трех факторов  $X_1, X_2, X_3$ , для которых заданы диапазоны изменения  $X_i \in [X_{imin}, X_{imax}]$ ,  $i = \overline{1,3}$ . На основании ортогонального центрального композиционного плана (ОЦКП) проведены 15 экспериментов, в результате которых найдены значения результативного признака  $Y_{exp}$ .

Требуется построить регрессионный полином для функции отклика по ОЦКП и создать программу для прогнозирования значений функции отклика по регрессионному полиному (задание 5).

Исходные данные по вариантам представлены Excel-файле в курсе [1] и в гугл-таблице [8] на листе.

Контрольная работа (ЗФО) включает три задания по разным темам дисциплины (задания 1, 4а-б, 5). Номером варианта служит число  $N$ , равное порядковому номеру в электронном журнале.

#### 4. ПРИМЕРЫ ВЫПОЛНЕНИЯ ЗАДАНИЙ

Целесообразно все этапы анализа данных, проводимого в Excel, выполнять на разных листах, дав им соответствующие названия.

##### Задание 1. Корреляционный и регрессионный анализ

*Выполнение задания согласно постановке задачи 1*

Разместим на листе Excel данные в соответствии со своим вариантом. Числовые данные для ЗФО с листа «ЗФО Задание КР 1». Определим корреляционную матрицу для факторов  $x_1, x_2, x_3$ , используя в Excel надстройку «Анализ данных», выбрав инструмент «Корреляция».

Расчеты показали, что факторы  $x_1$  и  $x_2$  коллинеарны, определитель корреляционной матрицы практически равен 0 (рис. 4.1). Изменим значения одного из этих факторов, в нашем случае  $x_2$ , так, чтобы значение определителя корреляционной матрицы было больше 0,8 (рис. 4.2).

| : X ✓ fx {=МОПРЕД(J12:L14)} |        |      |       |      |                                   |              |          |          |          |
|-----------------------------|--------|------|-------|------|-----------------------------------|--------------|----------|----------|----------|
| В                           | С      | Д    | Е     | Ф    | Н                                 | И            | Ж        | З        | Л        |
| №                           | y      | x1   | x2    | x3   |                                   |              | x1       | x2       | x3       |
| 1                           | 57,48  | 2,58 | 9,75  | 1,78 |                                   | x1           | 1        |          |          |
| 2                           | 110,18 | 3,72 | 13,15 | 2,18 |                                   | x2           | 1        | 1        |          |
| 3                           | 203,89 | 5,16 | 17,47 | 2,30 |                                   | x3           | 0,179845 | 0,179845 | 1        |
| 4                           | 85,25  | 2,80 | 10,39 | 2,66 |                                   |              |          |          |          |
| 5                           | 134,51 | 3,61 | 12,83 | 2,57 |                                   |              |          |          |          |
| 6                           | 99,87  | 2,81 | 10,44 | 3,12 |                                   |              | x1       | x2       | x3       |
| 7                           | 195,88 | 4,22 | 14,66 | 3,41 |                                   | x1           | 1        | 1        | 0,179845 |
| 8                           | 134,39 | 3,16 | 11,48 | 3,95 |                                   | x2           | 1        | 1        | 0,179845 |
| 9                           | 223,84 | 4,26 | 14,79 | 4,62 |                                   | x3           | 0,179845 | 0,179845 | 1        |
| 10                          | 289,68 | 4,85 | 16,55 | 4,73 |                                   |              |          |          |          |
| 11                          | 275,38 | 4,55 | 15,65 | 4,92 | Вывод: факторы x1, x2 коллинеарны |              |          |          |          |
| 12                          | 254,14 | 4,19 | 14,56 | 5,55 |                                   |              |          |          |          |
| 13                          | 161,75 | 2,99 | 10,97 | 5,58 | Det                               | -4,09232E-21 |          |          |          |
| 14                          | 329,68 | 4,67 | 16,01 | 5,57 | Определитель близок к нулю        |              |          |          |          |
| 15                          | 292,14 | 4,20 | 14,61 | 5,64 |                                   |              |          |          |          |

Рисунок 4.1. Корреляционная матрица при коллинеарных (коррелированных) факторах

|    |        |      |       |      |                               |          |          |          |
|----|--------|------|-------|------|-------------------------------|----------|----------|----------|
| 14 | 15     | 16   | 17    | 18   | 21                            | 22       | 23       | 24       |
| №  | y      | x1   | x2    | x3   |                               | x1       | x2       | x3       |
| 1  | 57,48  | 2,58 | 9,75  | 1,78 | x1                            | 1        |          |          |
| 2  | 110,18 | 3,72 | 11,15 | 2,18 | x2                            | 0,279238 | 1        |          |
| 3  | 203,89 | 5,16 | 12,56 | 2,30 | x3                            | 0,179845 | -0,24106 | 1        |
| 4  | 85,25  | 2,80 | 13,97 | 2,66 |                               |          |          |          |
| 5  | 134,51 | 3,61 | 15,37 | 2,57 |                               |          |          |          |
| 6  | 99,87  | 2,81 | 16,78 | 3,12 |                               | x1       | x2       | x3       |
| 7  | 195,88 | 4,22 | 18,18 | 3,41 | x1                            | 1        | 0,279238 | 0,179845 |
| 8  | 134,39 | 3,16 | 19,59 | 3,95 | x2                            | 0,279238 | 1        | -0,24106 |
| 9  | 223,84 | 4,26 | 21,00 | 4,62 | x3                            | 0,179845 | -0,24106 | 1        |
| 10 | 289,68 | 4,85 | 22,40 | 4,73 |                               |          |          |          |
| 11 | 275,38 | 4,55 | 20,81 | 4,92 | Вывод: факторы не коллинеарны |          |          |          |
| 12 | 254,14 | 4,19 | 19,22 | 5,55 |                               |          |          |          |
| 13 | 161,75 | 2,99 | 17,62 | 5,58 | Det                           | 0,807358 |          |          |
| 14 | 329,68 | 4,67 | 16,03 | 5,57 | Определитель близок к единице |          |          |          |

Рисунок 4.2. Корреляционная матрица при неколлинеарных (некоррелированных) факторах

По данным варианта найдем коэффициенты уравнения линейной регрессии, используя в Excel надстройку «Анализ данных», выбрав инструмент «Регрессия» (рис. 4.3).

Решение с помощью пакета анализа в Excel

Вывод итогов

Регрессионная статистика

Множественный R0,97

R-квадрат0,95

Нормированный R0,941760

Стандартная ошибка23,43102771

Наблюдения20

Коэффициент детерминации

сумма квадратов отклонений (факторная, остаточная и общая)

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

общая

факторная

остаточная

TSS

=

RSS

+

ESS

дисперсия (сумма квадратов отклонений, отнесенная к числу степеней свободы)

$$\frac{TSS}{m-1} = D_{\text{общ}}, \quad \frac{RSS}{n} = D_{\text{факт}}, \quad \frac{ESS}{m-n-1} = D_{\text{остат}}$$

Расчетное значение критерия Фишера

Дисперсионный анализ

|                   | df | SS         | MS       | F         | Значимость F |
|-------------------|----|------------|----------|-----------|--------------|
| Регрессия(факторн | 3  | 170342,817 | 56780,94 | 103,42366 | 1,09194E-10  |
| Остаток(остаточн) | 16 | 8784,20895 | 549,0131 |           |              |
| Итого(общая)      | 19 | 179127,026 |          |           |              |

Рисунок 4.3. Дисперсионный анализ

Получены следующие результаты. Коэффициент детерминации  $R^2=0,95$ . Расчетное (эмпирическое) значение критерия Фишера примерно  $F=103,4$ . Табличное значение критерия Фишера  $F_{\text{табл}}=3,13$  найдем по таблице [7] (при степенях свободы  $k_1=df_{\text{фактор}}=3$  и  $k_2=df_{\text{остат}}=19$  и уровне значимости  $\alpha=0,05$ ) меньше расчетного, что говорит о значимости уравнения регрессии в целом.

Коэффициенты линейного уравнения регрессии приведены на рисунке 4.4.



|  |          |  |          |   |              |           |          |  |  |
|--|----------|--|----------|---|--------------|-----------|----------|--|--|
| Решение с помощью пакета анализа в Excel |          |  |          |   |              |           |          |  |  |
| Вывод итогов                             |          |  |          |   |              |           |          |  |  |
| Регрессионная статистика                 |          |  |          |   |              |           |          |  |  |
| Множественный F                          | 0,997016 | сумма квадратов отклонений (факторная, остаточная и общая) |          | дисперсия (сумма квадратов отклонений), отнесенная к числу степеней |              |           |          |  |  |
| R-квадрат                                | 0,994041 |  |          |   |              |           |          |  |  |
| Нормированный F                          | 0,992924 |  |          |   |              |           |          |  |  |
| Стандартная ошибка                       | 0,044361 |  |          |   |              |           |          |  |  |
| Наблюдения                               | 20       |  |          |   |              |           |          |  |  |
| Дисперсионный анализ                     |          |  |          |   |              |           |          |  |  |
|  | df       | SS   | MS       | F   | Значимость F |           |          |  |  |
| Регрессия                                | 3        | 5,252445224  | 1,750815 | 889,67688   | 5,29E-18     |           |          |  |  |
| Остаток                                  | 16       | 0,031486759  | 0,001968 |   |              |           |          |  |  |
| Итого                                    | 19       | 5,283931983  |          |   |              |           |          |  |  |
| Коэффициенты регрессии                   |          |  |          |   |              |           |          |  |  |
| Y-пересечение                            | 2,360109 | 0,093955889  | 25,11933 | 2,778E-14   | 2,160932     | 2,5592869 | 2,160932 |  |  |
| ln(x1)                                   | 1,522294 | 0,049669576  | 30,64843 | 1,225E-15   | 1,417        | 1,6275891 | 1,417    |  |  |
| ln(x2)                                   | -0,0633  | 0,032352353  | -1,9565  | 0,0680924   | -0,13188     | 0,0052864 | -0,13188 |  |  |
| ln(x3)                                   | 0,719535 | 0,024089564  | 29,86916 | 1,837E-15   | 0,668467     | 0,7706027 | 0,668467 |  |  |

Рисунок 4.7. Степенная регрессия

## Задание 2. Автоматизация анализа данных

### Выполнение задания согласно постановке задачи 1

В этом задании нужно средствами VBA в Excel создать формы для проведения корреляционного, дисперсионного и регрессионного анализа.

Подключим в надстройках Excel пакет анализа VBA и добавим в Настройках ленты вкладку Разработчик. Зайдем в редактор VBA через Разработчик, Visual Basic либо сочетанием клавиш ALT+F11.

В окончательном виде созданные формы имеют вид, показанный на рисунке 4.8. На рисунке показаны три формы: стартовая форма, форма корреляции и линейной регрессии.

Рисунок 4.8. Формы стартовая, корреляции и линейной регрессии



Для создания форм в редакторе VBA использовались инструменты Insert, UserForm, для добавления объектов на формы (поля ввода, подписи, кнопки) использовались инструменты Toolbox, свойства объектов Name, Caption настраивались в окне Properties (рис. 4.9).

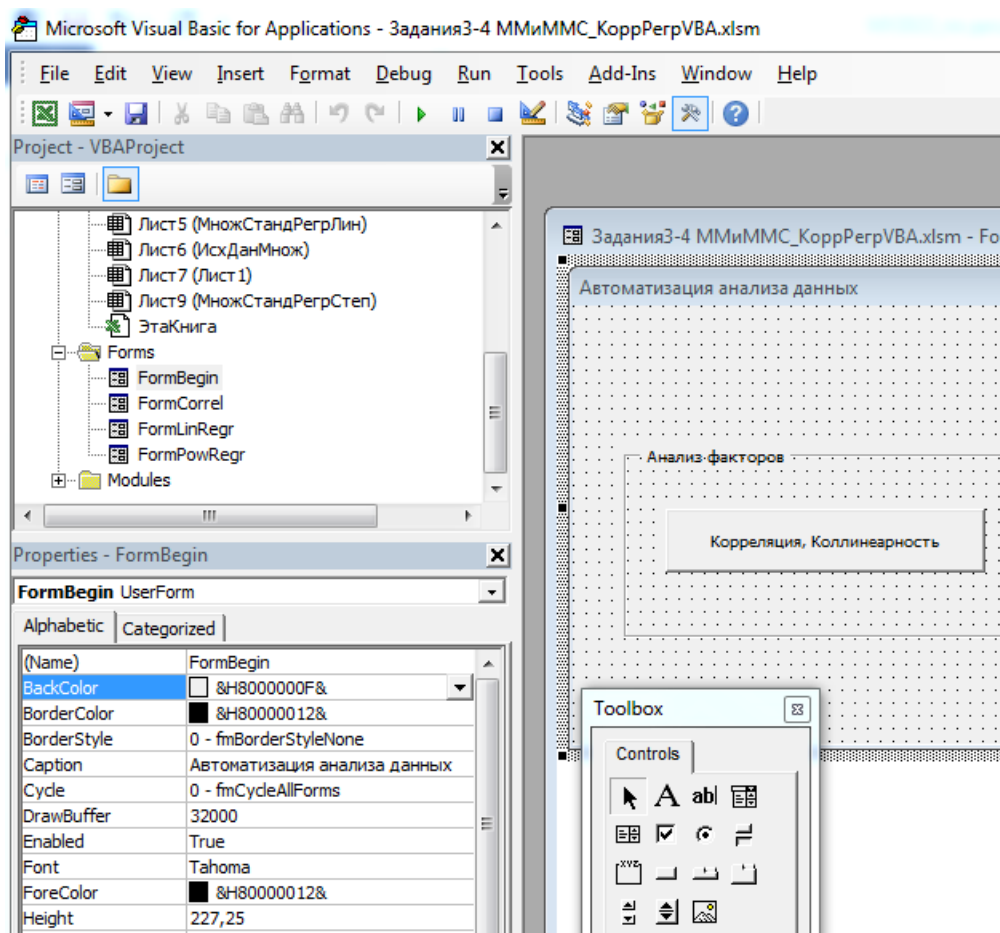


Рисунок 4.9. Настройка свойств стартовой формы

Стартовая форма FormBegin запускается макросом, записанном в стандартном модуле Module1 (рис. 4.10). Для каждой кнопки в модулях форм написана процедура обработки события. Процедуры кнопок модуля FormBegin аналогичны и представлены рисунке 4.10.

Для кнопок Корреляция, Линейная, Степенная регрессия использовалась автоматическая запись макроса. На рисунке 4.11 представлены процедуры модуля FormCorrel. Изначально макрос для выполнения корреляции записывается в стандартном модуле Module, затем код переносится в процедуру модуля формы FormCorrel. При этом следует заменить в автоматически созданном коде везде название листа ActivSheet на название листа Worksheets("Корреляция(VBA)"), на котором выполняется корреляция.



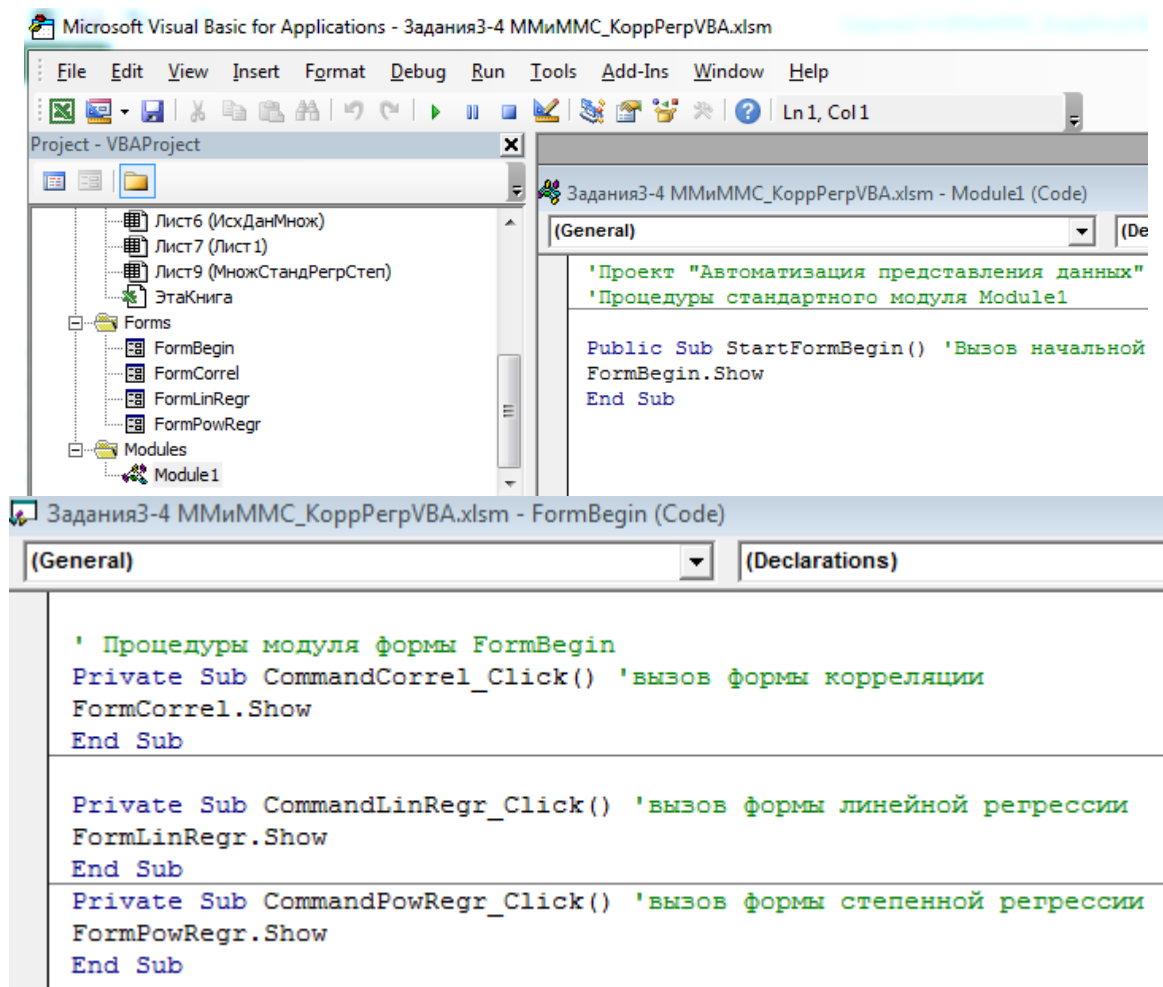


Рисунок 4.10. Процедуры кнопок модуля FormBegin

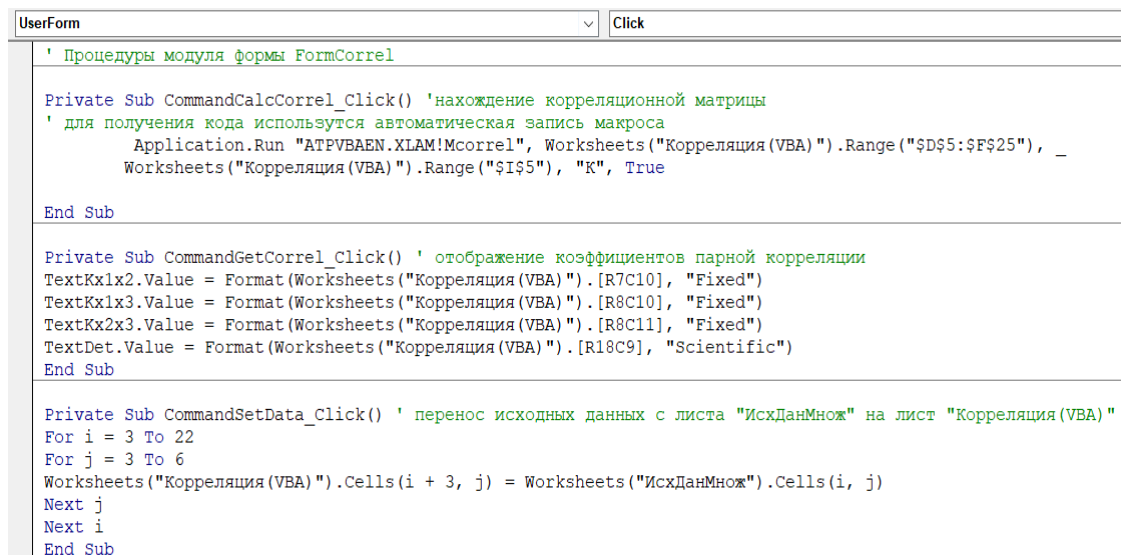


Рисунок 4.11. Процедуры модуля FormCorrel

Результаты работы приложения представлены на Рисунок 4.12.

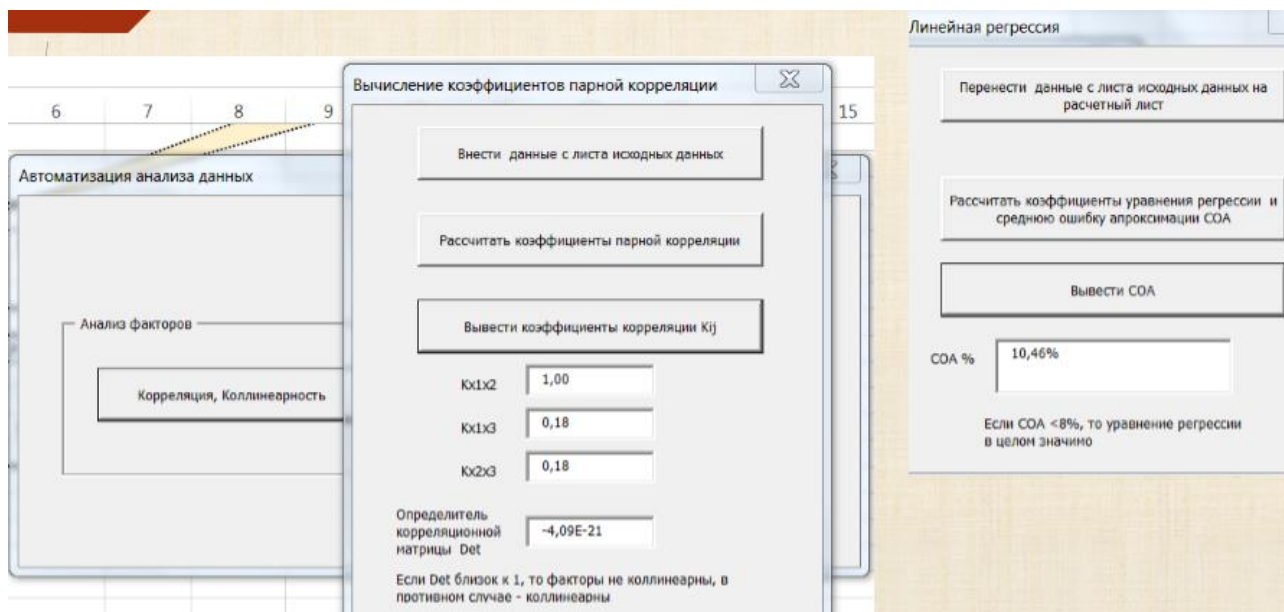


Рисунок 4.12. Результаты работы приложения

### Задание 3. Компонентный анализ

*Выполнение задания согласно постановке задачи 1*

*Этапы решения задачи*

- Создать m-файл: New, script, имя, например, CompAnCorrFam.
- Внести исходные данные: предварительно создать в рабочем каталоге текстовый файл с исходными данными, дав имя, например, lr\_Corr\_XY\_N.
- Найти  $X_0$  и  $R(X_0)$  (ф-и mean, std, corrcoef).
- Найти собственные значения и векторы (ф-я eig).
- Найти матрицу нагрузок  $P$  (главных компонент).
- Найти матрицу счетов  $T$  (сжатой матрицы данных).
- Найти матрицу ошибок  $E$  для оценки потерь.
- Найти матрицу вклада компонент  $cs$ .

Создадим файл в Matlab для выполнения компонентного анализа.

Следует рассмотреть случаи коллинеарных (коррелированных) и неколлинеарных факторов и проанализировать возможность сжатия данных в каждом из случаев.

```

1  clc, clear
2
3  load lr_Corr_XY_N.txt % загрузка коллинеарных данных из текстового файла
4  X = lr_Corr_XY_N(:,2:4)
5
6  [m n] = size(X) %нахождение размерности матрицы
7  XMean = mean(X) %нахождение среднего значения по столбцам по умолчанию
8  XStd = std(X) %нахождение выборочной дисперсии по столбцам, делится на m-1,
9  % а в генеральной на m
10 X0 = (X - repmat(XMean,m,1))./ repmat(XStd,m,1)% центрирование и нормировка
11 %матрицы, применение функции повторения матрицы, m строк, 1 столбец
12 X0Mean = mean(X0) %проверка центрирования
13 X0Std = std(X0)
14 RX0=corrcoef(X0) %вычисление корреляционной матрицы

```

**Случай  
коллинеарных  
факторов**

Рисунок 4.13. Код для нахождения стандартизованной матрицы  $X_0$  исходных данных

|                |         |         |           |         |         |
|----------------|---------|---------|-----------|---------|---------|
| Command Window |         |         | -0.5121   | -0.5120 | -1.1310 |
| m =            |         |         | -1.2800   | -1.2766 | -0.8380 |
| 20             |         |         | 0.4259    | 0.4266  | -0.6831 |
|                |         |         | -0.8565   | -0.8568 | -0.3947 |
|                |         |         | 0.4742    | 0.4791  | -0.0369 |
|                |         |         | 1.1880    | 1.1894  | 0.0219  |
|                |         |         | 0.8251    | 0.8262  | 0.1234  |
| n =            |         |         | 0.3896    | 0.3862  | 0.4599  |
| 3              |         |         | -1.0622   | -1.0627 | 0.4759  |
|                |         |         | 0.9703    | 0.9715  | 0.4706  |
|                |         |         | 0.4017    | 0.4064  | 0.5079  |
|                |         |         | 1.2364    | 1.2338  | 1.0901  |
| XMean =        |         |         | -0.2395   | -0.2434 | 0.8551  |
| 3.8675         | 13.6025 | 4.7382  | 0.7041    | 0.7010  | 1.0260  |
|                |         |         | 0.3049    | 0.3096  | 1.5975  |
|                |         |         | -1.7034   | -1.7044 | 1.7097  |
| XStd =         |         |         | X0Mean =  |         |         |
| 0.8261         | 2.4782  | 1.9408  | 1.0e-15 * |         |         |
|                |         |         | 0.1332    | 0.1332  | -0.5218 |
| X0 =           |         |         | X0Std =   |         |         |
| -1.5560        | -1.5560 | -1.5220 | 1.0000    | 1.0000  | 1.0000  |
| -0.1815        | -0.1815 | -1.3194 |           |         |         |
| 1.5606         | 1.5606  | -1.2556 |           |         |         |
| -1.2969        | -1.2969 | -1.0689 |           |         |         |
| -0.3119        | -0.3119 | -1.1196 |           |         |         |
| -1.2764        | -1.2764 | -0.8333 |           |         |         |
| 0.4252         | 0.4252  | -0.6826 |           |         |         |
| -0.8574        | -0.8574 | -0.4057 |           |         |         |
| 0.4772         | 0.4772  | -0.0607 |           |         |         |
|                |         |         | RX0 =     |         |         |
|                |         |         | 1.0000    | 1.0000  | 0.1802  |
|                |         |         | 1.0000    | 1.0000  | 0.1804  |
|                |         |         | 0.1802    | 0.1804  | 1.0000  |

Рисунок 4.14. Результаты расчетов

```

15 disp(' Матрицы собств векторов и собственных знач по возрастанию собств значений')
16 [U D] = eig(RX0) %нахождение ортогональной матрицы координат собственных векторов
17 % (матрицы %нагрузок) U и преобразованной корреляционной матрицы D (диагональной)
18 % в базисе из собственных векторов (собственные значения по возрастанию)
19 disp(' Матрица нагрузок по убыванию собственных значений')
20 P=[U(:,3) U(:,2)] %выбор существенных компонент с наибольшими собственными
21 % значениями (дисперсиями)
22 PT=P.' % транспонирование матрицы нагрузок (PT*P=E)- ортогональность P
23 PTP=PT*P
24 disp(' Матрица счетов')
25 T=X0*P %нахождение матрицы счетов (новый вектор данных)
26 disp(' Матрица ошибок') |
27 E=X0-T*P.' % проверка - получение исходной матрицы

```

Рисунок 4.15. Код для вычисления матриц нагрузок P, счетов T, ошибок E

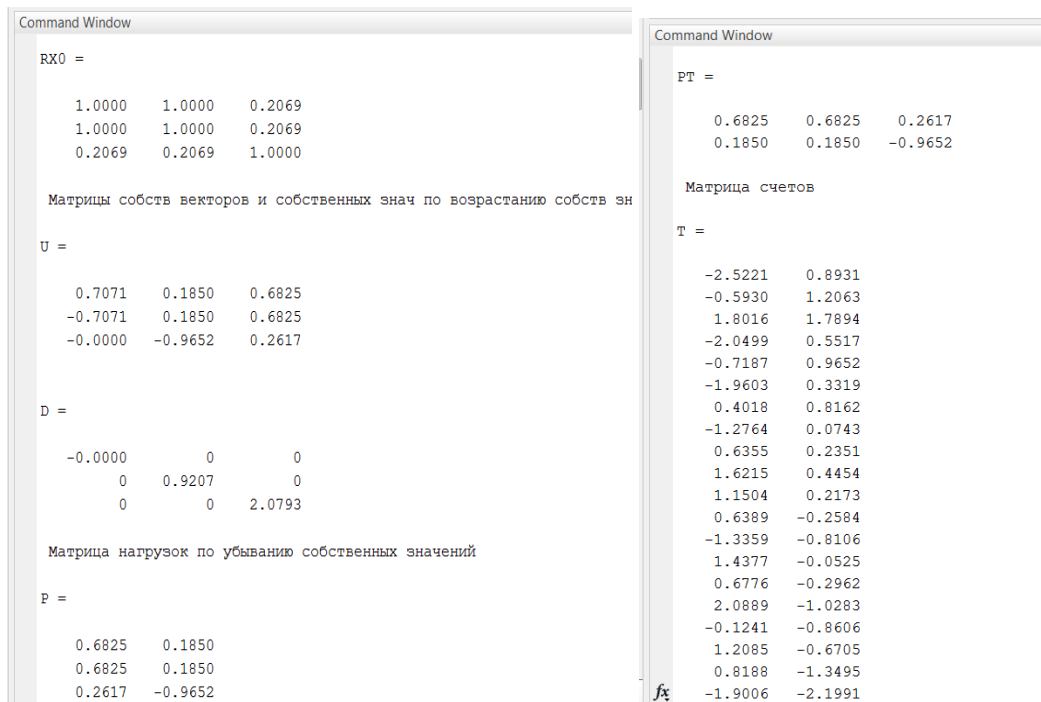


Рисунок 4.16. Результаты расчета матрицы счетов T (сжатых данных)

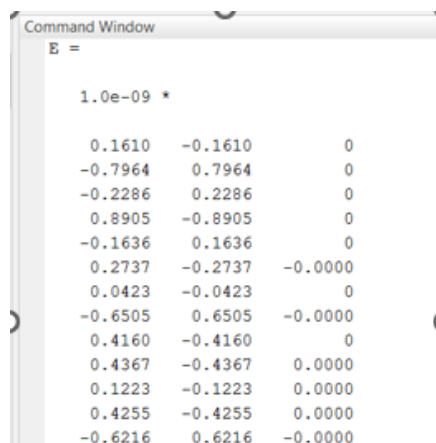


Рисунок 4.17. Вычисление матрицы ошибок E

Матрица ошибок мала (рис. 4.17), что свидетельствует о возможности сжатия данных.

Проведем оценку вклада компонент.

```

28 disp(' создание вектора по диагональным элементам и расположение элементов по убыванию ')
29 Dfli=flipud(diag(D))%создание вектора по диагональным элементам и
30 %зеркальное отображение относительно середины, чтобы собств. значения по убыванию были
31 disp(' вклад компонент одной, двух, трех, начиная с существенных')
32 cs=cumsum(Dfli) / sum(Dfli) %определение вклада каждой компоненты в общую дисперсию
33 %компоненты, использование функции flipud, которая изменяет порядок
34 %следования элементов на противоположный
35 disp('корреляционная матрица сжатых данных')
36 RT=corrcoef(T) %вычисление корреляционной матрицы преобразованных данных
37 covT=cov(T) %вычисление ковариационной матрицы преобразованных данных
38
  
```

Рисунок 4.18. Код для оценки вклада компонент

```

Dfli =

    2.0613
    0.9387
    0.0000

вклад компонент одной, двух, трех, начиная с существенных

cs =

    0.6871
    1.0000
    1.0000

корреляционная матрица сжатых данных

RT =

    1.0000    -0.0000
   -0.0000     1.0000

covT =

    2.0613    -0.0000
   -0.0000     0.9387

```

Рисунок 4.19. Результаты расчета вклада компонент при коллинеарных данных

Значения переменной «cs» показывают, что две компоненты достаточны для сохранения исходной информации, так как вклад третьей компоненты нулевой.

Проведение расчетов с неколлинеарными данными (код программы аналогичен приведенному выше, меняются только исходные данные) показало отсутствие возможности их сжатия без потери информации, так как матрица ошибок не мала, т.е. две компоненты не передадут всю информацию. На рисунке 4.20 приведены матрица ошибок E в случае неколлинеарных данных и матрица вклада компонент cs.

**Матрица ошибок**

E =

|         |         |         |
|---------|---------|---------|
| 0.3599  | -0.3874 | -0.3371 |
| 0.6394  | -0.6884 | -0.5989 |
| 1.0844  | -1.1675 | -1.0157 |
| -0.0248 | 0.0267  | 0.0232  |
| 0.2075  | -0.2234 | -0.1943 |
| -0.3101 | 0.3339  | 0.2905  |
| 0.0938  | -0.1009 | -0.0878 |
| -0.5256 | 0.5658  | 0.4923  |
| -0.3048 | 0.3281  | 0.2855  |
| -0.1963 | 0.2114  | 0.1839  |
| -0.2243 | 0.2415  | 0.2101  |
| -0.3495 | 0.3763  | 0.3274  |

создание вектора по диагональным элементам и расположение

Dfli =

|        |
|--------|
| 1.2918 |
| 1.1766 |
| 0.5315 |

вклад компонент одной, двух, трех, начиная с существенных

cs =

|        |
|--------|
| 0.4306 |
| 0.8228 |
| 1.0000 |

корреляционная матрица сжатых данных

RT =

|         |         |
|---------|---------|
| 1.0000  | -0.0000 |
| -0.0000 | 1.0000  |

covT =

|         |         |
|---------|---------|
| 1.2918  | -0.0000 |
| -0.0000 | 1.1766  |

Рисунок 4.20. Расчеты в случае неколлинеарных факторов

Выполнение задания согласно постановке задачи 1

Для понимания сути иерархической кластеризации выполним ее пошагово для шести объектов. Исходные данные приведены в следующей таблице. Числовые данные для ЗФО с листа «ЗФО Задание КР 4а-б».

| номер объекта | 1 | 2 | 3 | 4 | 5 | 6  |
|---------------|---|---|---|---|---|----|
| xi            | 1 |   | 1 | 1 | 1 |    |
| 1             | 5 | 6 | 5 | 0 | 1 | 10 |
| xi            | 1 | 1 | 1 |   |   |    |
| 2             | 0 | 2 | 3 | 9 | 9 | 7  |

Д23

**Таблица 1: Исходные данные**

|    | A               | B                   | C  | D  | E  | F  | G  | H  | I | J               | K                   | L  | M  | N  | O  | P  | Q  |
|----|-----------------|---------------------|----|----|----|----|----|----|---|-----------------|---------------------|----|----|----|----|----|----|
| 1  |                 |                     |    |    |    |    |    |    |   |                 |                     |    |    |    |    |    |    |
| 2  |                 |                     |    |    |    |    |    |    |   |                 |                     |    |    |    |    |    |    |
| 3  | $i \setminus j$ |                     | 1  | 2  | 3  | 4  | 5  | 6  |   | $i \setminus j$ |                     | 1  | 2  | 3  | 4  | 5  | 6  |
| 4  |                 | $x1i \setminus x1j$ | 15 | 6  | 15 | 10 | 11 | 10 |   |                 | $x2i \setminus x2j$ | 10 | 12 | 13 | 9  | 9  | 7  |
| 5  | 1               | 15                  | 0  | -9 | 0  | -5 | -4 | -5 |   | 1               | 10                  | 0  | 2  | 3  | -1 | -1 | -3 |
| 6  | 2               | 6                   | 9  | 0  | 9  | 4  | 5  | 4  |   | 2               | 12                  | -2 | 0  | 1  | -3 | -3 | -5 |
| 7  | 3               | 15                  | 0  | -9 | 0  | -5 | -4 | -5 |   | 3               | 13                  | -3 | -1 | 0  | -4 | -4 | -6 |
| 8  | 4               | 10                  | 5  | -4 | 5  | 0  | 1  | 0  |   | 4               | 9                   | 1  | 3  | 4  | 0  | 0  | -2 |
| 9  | 5               | 11                  | 4  | -5 | 4  | -1 | 0  | -1 |   | 5               | 9                   | 1  | 3  | 4  | 0  | 0  | -2 |
| 10 | 6               | 10                  | 5  | -4 | 5  | 0  | 1  | 0  |   | 6               | 7                   | 3  | 5  | 6  | 2  | 2  | 0  |

**Таблица 2: Матрица Евклидовых расстояний**

|    | A               | B      | C    | D    | E    | F    | G    | H    | I | J   | K | L | M | N | O | P | Q |
|----|-----------------|--------|------|------|------|------|------|------|---|---|---|---|---|---|---|---|---|
| 1  |                 |        |      |      |      |      |      |      |   |   |   |   |   |   |   |   |   |
| 2  |                 |        |      |      |      |      |      |      |   |   |   |   |   |   |   |   |   |
| 3  | $i \setminus j$ |        | 1    | 2    | 3    | 4    | 5    | 6    |   | матрица Евклидовых расстояний между объектами |   |   |   |   |   |   |   |
| 4  |                 | Rij(1) |      |      |      |      |      |      |   |   |   |   |   |   |   |   |   |
| 5  | 1               |        | 0,00 | 9,22 | 3,00 | 5,10 | 4,12 | 5,83 |   |   |   |   |   |   |   |   |   |
| 6  | 2               |        | 9,22 | 0,00 | 9,06 | 5,00 | 5,83 | 6,40 |   |   |   |   |   |   |   |   |   |
| 7  | 3               |        | 3,00 | 9,06 | 0,00 | 6,40 | 5,66 | 7,81 |   |   |   |   |   |   |   |   |   |
| 8  | 4               |        | 5,10 | 5,00 | 6,40 | 0,00 | 1,00 | 2,00 |   |   |   |   |   |   |   |   |   |
| 9  | 5               |        | 4,12 | 5,83 | 5,66 | 1,00 | 0,00 | 2,24 |   |   |   |   |   |   |   |   |   |
| 10 | 6               |        | 5,83 | 6,40 | 7,81 | 2,00 | 2,24 | 0,00 |   |   |   |   |   |   |   |   |   |

Рисунок 4.21. Матрица евклидовых расстояний между исходными объектами

Далее продолжим образование кластеров 7,8, 9, 10 по тому же принципу. На каждом шаге образуется один новый кластер (рис. 4.22). В результате получилось два кластера (1,2,3) (4,5,6) с номерами 9 и 10, между которыми расстояние равно 5. Если мы объединим кластеры 9 и 10, то получится 11 кластер, объединяющий все изначально заданные объекты. При иерархической кластеризации номер последнего кластера всегда  $2n-1$ , где  $n$  - количество изначально заданных объектов.



| i\j |        | 1    | 2    | 3    | 4,5   | 6    |  | i\j   |        | 0    | 1     | 2,3   | 4,5  | 6 |
|-----|--------|------|------|------|-------|------|--|-------|--------|------|-------|-------|------|---|
|     | Rij(2) |      |      |      |       |      |  |       | Rij(3) |      |       |       |      |   |
| 1   |        | 0,00 | 2,24 | 3,00 | 4,12  | 5,83 |  | 1     |        | 0,00 | 2,24  | 5,10  | 5,83 |   |
| 2   |        | 2,24 | 0,00 | 1,41 | 5,00  | 6,40 |  | 2,3   |        | 2,24 | 0,00  | 5,00  | 6,40 |   |
| 3   |        | 3,00 | 1,41 | 0,00 | 5,66  | 7,81 |  | 4,5   |        | 5,10 | 5,00  | 0,00  | 2,00 |   |
| 4,5 |        | 5,10 | 5,00 | 6,40 | 0,00  | 2,00 |  | 6     |        | 5,83 | 6,40  | 2,00  | 0,00 |   |
| 6   |        | 5,83 | 6,40 | 7,81 | 2,00  | 0,00 |  |       |        |      |       |       |      |   |
|     |        |      |      |      |       |      |  |       |        |      |       |       |      |   |
| i\j |        | 0    | 1    | 2,3  | 4,5,6 |      |  | i\j   |        | 0    | 1,2,3 | 4,5,6 |      |   |
|     | Rij(4) |      |      |      |       |      |  |       | Rij(5) |      |       |       |      |   |
| 1   |        | 0,00 | 2,24 | 5,10 |       |      |  | 1,2,3 |        | 0,00 | 5,00  |       |      |   |
| 2,3 |        | 2,24 | 0,00 | 5,00 |       |      |  | 4,5,6 |        | 5,00 | 0,00  |       |      |   |
| 5,6 |        | 5,10 | 5,00 | 0,00 |       |      |  |       |        |      |       |       |      |   |

Рисунок 4.22. Иерархическая кластеризация в Excel

Построим иерархическое дерево бинарных кластеров (дендрограмму) и запишем его матрицу (рис 4.23). В первом столбце матрицы нумерация вновь создаваемых кластеров, начиная с 7 до 11. Во втором и третьем столбцах номера объединяемых кластеров, в четвертом – расстояние между ними. Кластеры с номерами 1-6 – это исходные объекты.

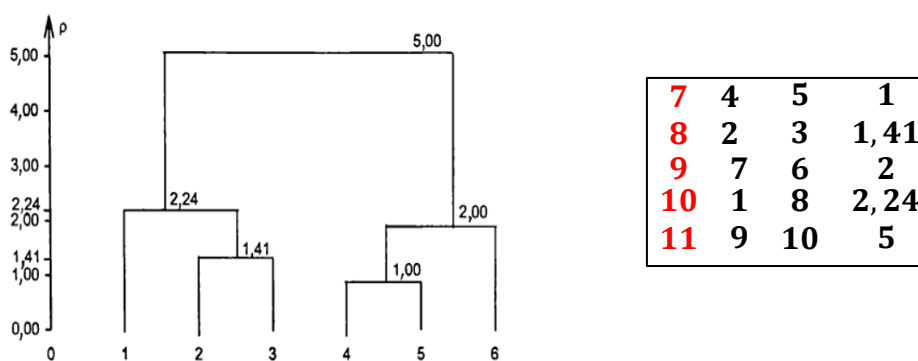


Рисунок 4.23. Дендрограмма и матрица иерархического дерева бинарных кластеров

Выполним иерархическую кластеризацию и кластеризацию методом к-средних. Это можно сделать, используя различные математические пакеты, например, в MatLab, библиотеки Python и др. Числовые данные для ЗФО с листа «ЗФО Задания КР 4.4а-б».

```

1 clc, clear
2 load lr_NotCorr_XY_N.txt % загрузка коллинеарных данных из текстового файла
3 disp('исходная матрица данных')%вывод комментариев в командном окне
4 X = lr_NotCorr_XY_N(:,2:4)
5 disp('нахождение размерности матрицы данных')% вывод в окно заголовка
6 [m n] = size(X) % выводятся в командное окно
7 Y=pdist(X); % нахождение расстояний m*(m-1)/2 различных расстояний
8 % ; означает в командное окно не выводить
9 disp('матрица иерархического дерева бинарных кластеров размерности(m-1)*3')
10 Z=linkage(Y)% Z выдается в командное окно
11 disp(' m+i - индекс кластера, где i - номер строки кластера в матрице Z')
12 disp('первый и второй столбцы - индексы кластеров, которые объединяются')
13 disp('третий столбец - расстояние между объединяющимися кластерами')
14 dendrogram(Z) % построение дендрограммы в отдельном окне

```

Рисунок 4.24. Код для иерархической кластеризации в Matlab

```

16 disp('кластеризация по межгрупповым средним, функция kmeans')
17 NUM=1:m; %задание массива-строки порядковых номеров
18 NUMT=NUM.'; % '.' - транспонирование, ; означает блокировку вывода в окно
19 [IDX,C,SUMDIST]=kmeans(X,5);
20 disp('clust - номера кластеров и номера объектов')
21 clust=[NUMT IDX]
22 disp('C - центры кластеров и SUMDIST - суммарные расстояния до центров')
23 C
24 SUMDIST
25 X1=X(:,1); X2=X(:,2); X3=X(:,3); % массивы столбцов данных
26 color=['r' 'g' 'b' 'c' 'm']; % массив разных цветов для разных кластеров
27 figure % создает новое графическое окно, старое графическое окно остается
28 % построение трехмерного графика кластеров в отдельном окне
29 for i=1:1:m % цикл по объектам
30     switch IDX(i) % переключатель
31     case 1
32         plot3(X1(i),X2(i),X3(i),'Color',color(1),'Marker','o')
33     case 2
34         plot3(X1(i),X2(i),X3(i),'Color',color(2),'Marker','o')
35     case 3
36         plot3(X1(i),X2(i),X3(i),'Color',color(3),'Marker','o')
37     case 4
38         plot3(X1(i),X2(i),X3(i),'Color',color(4),'Marker','o')
39     case 5
40         plot3(X1(i),X2(i),X3(i),'Color',color(5),'Marker','o')
41     end
42     hold on % сохранение текущего графика и добавление нового графика в
43     % текущем окне
44 end
45 grid on % построение координатной сетки

```

Рисунок 4.25. Код для кластеризации методом k-средних в Matlab

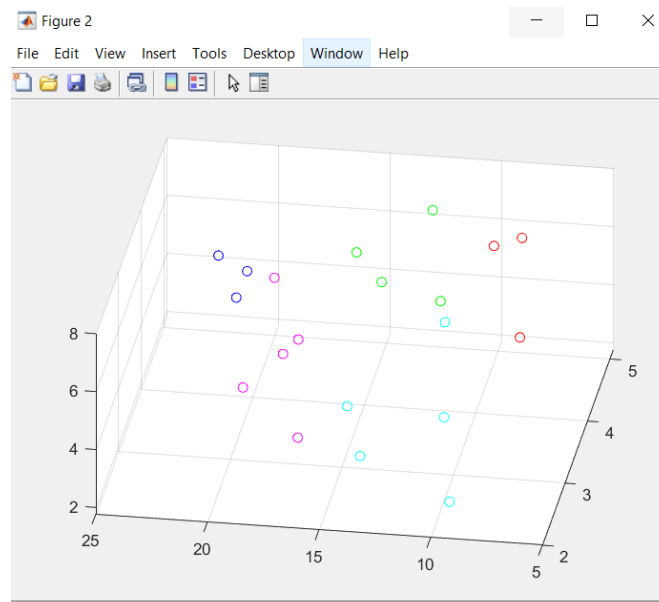


Рисунок 4.26. Визуализация кластеризации в Matlab



Приведем пример кода, созданного на Python в интерактивной среде Colab [6] (рис. 4.27), и результаты кластеризации (рис. 4.28) в соответствии с данными варианта. Исходные данные размещены в файле dataCluster.xlsx и предварительно загружены в сессионное хранилище.

```
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d
import numpy as np
from sklearn.cluster import KMeans
X = np.array(pd.read_excel('dataCluster.xlsx'))
print(X)
print('-'*100)
kmeans = KMeans(n_clusters = 5, n_init=20)
kmeans.fit(X)
print('Центроиды кластеров')
print(kmeans.cluster_centers_)
print('Метки кластеров')
print(kmeans.labels_)
ax=plt.axes(projection="3d")
ax.scatter(kmeans.cluster_centers_[0],kmeans.cluster_centers_[1],kmeans.cluster_centers_[2], color='black')
ax.scatter(X[:,0],X[:,1],X[:,2], c=kmeans.labels_, cmap='rainbow')
plt.xlabel("x")
plt.ylabel("y")
```

Рисунок 4.27. Код для проведения кластеризации на Python

```
Центроиды кластеров
[[ 2.83768961 10.51306883 2.87996849]
 [ 4.38462184 15.15386553 5.83158079]
 [ 2.7252141 10.17564229 6.73520128]
 [ 5.15665052 17.46995157 2.30127113]
 [ 3.84872874 13.54618622 2.71869115]]
Метки кластеров
[0 4 3 0 4 0 1 1 1 1 2 1 1 1 1 2]
Text(0.5, 0.5, 'y')
```

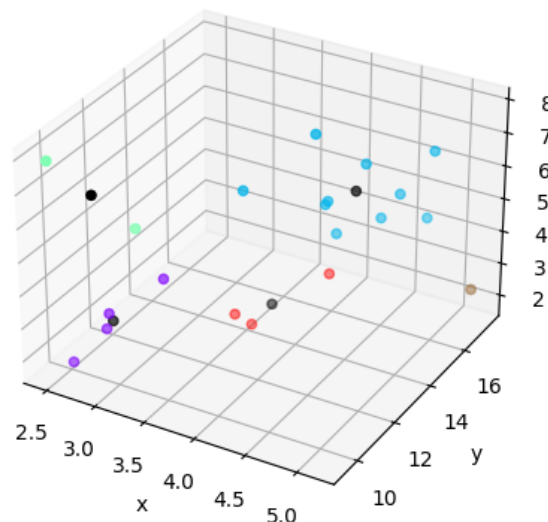


Рисунок 4.28. Визуализация кластеризации на Python

Рассматривалось 5 кластеров. Объекты, принадлежащие одному кластеру, имеют один цвет. С помощью программы найдены центры кластеров (на графике изображены черным цветом), метки каждого объекта, по которым можно определить, к какому кластеру объект принадлежит.

## Задание 5. Планирование эксперимента с использованием ОЦКП

### Постановка задачи 2

Известно, что результативный признак  $Y$  зависит от трех факторов  $X_1, X_2, X_3$ , для которых известны диапазоны изменения  $X_i \in [X_{imin}, X_{imax}]$ ,  $i = \overline{1,3}$ . На основании ортогонального центрального композиционного плана (ОЦКП) проведены 15 экспериментов, в результате которых найдены значения результативного признака  $Y_{exp}$ .

Требуется построить регрессионный полином для функции отклика по ОЦКП и создать программу для прогнозирования значений функции отклика по регрессионному полиному.

### Выполнение задания

Данные варианта вставим на лист рабочей книги. Числовые данные для ЗФО с листа «ЗФО Задания КР 5».

Рассчитаем средние значения по формулам  $X_{iavg} = (X_{imin} + X_{imax})/2$  (рис. 4.29). Для удобства в дальнейших расчетах можно дать имена (alf, a, N, Ncoef) ячейкам в соответствии со значениями параметров.

|    | A        | B | C        | D     | E      | F     | G     | H     | I     | J |
|----|----------|---|----------|-------|--------|-------|-------|-------|-------|---|
| 1  | Yexp     |   |          |       |        |       |       |       |       |   |
| 2  | 5,397447 |   |          |       |        |       |       |       |       |   |
| 3  | 7,630355 |   |          |       |        |       |       |       |       |   |
| 4  | 5,508893 |   |          |       |        |       |       |       |       |   |
| 5  | 7,6002   |   |          |       |        |       |       |       |       |   |
| 6  | 6,115107 |   |          |       |        |       |       |       |       |   |
| 7  | 7,116865 |   |          |       |        |       |       |       |       |   |
| 8  | 6,438498 |   |          |       |        |       |       |       |       |   |
| 9  | 8,507912 |   |          |       |        |       |       |       |       |   |
| 10 | 5,824821 |   |          |       |        |       |       |       |       |   |
| 11 | 7,161622 |   |          |       |        |       |       |       |       |   |
| 12 | 6,215063 |   |          |       | alf=   | 1,215 |       |       |       |   |
| 13 | 7,122938 |   |          |       | a=     | 0,73  |       |       |       |   |
| 14 | 7,361126 |   |          |       | N=     | 15    |       |       |       |   |
| 15 | 7,095445 |   |          |       | Ncoef= | 10    |       |       |       |   |
| 16 | 6,747072 |   |          |       |        |       |       |       |       |   |
| 17 |          |   | Номер    |       |        |       |       |       |       |   |
|    |          |   | Варианта | X1MIN | X1MAX  | X2MIN | X2MAX | X3MIN | X3MAX |   |
| 18 |          |   | 0        | 5     | 6      | 4     | 5     | 8     | 11    |   |
| 19 |          |   |          | X1AVG |        | X2AVG |       | X3AVG |       |   |
| 20 |          |   |          | 5,5   |        | 4,5   |       | 9,5   |       |   |

Рисунок 4.29. Исходные данные

В задании будем использовать ортогональный центральный композиционный план (ОЦКП) (см раздел 1.5). Сформируем на листе матрицу основного плана ОЦКП (рис. 4.30) в кодированных переменных и рассчитаем матрицу плана в натуральных переменных, используя формулы

$$X_i^j = (X_{iavg} - X_{imin})x_i^j + X_{iavg}.$$

Значения их этой матрицы должны использоваться при проведении экспериментов для получения столбца значений  $Y_{exp}$ . В учебном примере эти значения уже заданы.

|    |               |   |                          |        |        |   |                         |        |         |   |          |
|----|---------------|---|--------------------------|--------|--------|---|-------------------------|--------|---------|---|----------|
| 22 | A             | B | C                        | D      | E      | F | G                       | H      | I       | J | K        |
| 23 |               |   | КОДИРОВАННЫЕ ЗН ФАКТОРОВ |        |        |   | НАТУРАЛЬНЫЕ ЗН ФАКТОРОВ |        |         |   |          |
| 24 | Номер<br>эксп |   | x1                       | x2     | x3     |   | X1                      | X2     | X3      |   | Yexp     |
| 25 | 1             |   | -1                       | -1     | -1     |   | 5                       | 4      | 8       |   | 5,397447 |
| 26 | 2             |   | 1                        | -1     | -1     |   | 6                       | 4      | 8       |   | 7,630355 |
| 27 | 3             |   | -1                       | 1      | -1     |   | 5                       | 5      | 8       |   | 5,508893 |
| 28 | 4             |   | 1                        | 1      | -1     |   | 6                       | 5      | 8       |   | 7,6002   |
| 29 | 5             |   | -1                       | -1     | 1      |   | 5                       | 4      | 11      |   | 6,115107 |
| 30 | 6             |   | 1                        | -1     | 1      |   | 6                       | 4      | 11      |   | 7,116865 |
| 31 | 7             |   | -1                       | 1      | 1      |   | 5                       | 5      | 11      |   | 6,438498 |
| 32 | 8             |   | 1                        | 1      | 1      |   | 6                       | 5      | 11      |   | 8,507912 |
| 33 | 9             |   | -1,215                   | 0      | 0      |   | 4,8925                  | 4,5    | 9,5     |   | 5,824821 |
| 34 | 10            |   | 1,215                    | 0      | 0      |   | 6,1075                  | 4,5    | 9,5     |   | 7,161622 |
| 35 | 11            |   | 0                        | -1,215 | 0      |   | 5,5                     | 3,8925 | 9,5     |   | 6,215063 |
| 36 | 12            |   | 0                        | 1,215  | 0      |   | 5,5                     | 5,1075 | 9,5     |   | 7,122938 |
| 37 | 13            |   | 0                        | 0      | -1,215 |   | 5,5                     | 4,5    | 7,6775  |   | 7,361126 |
| 38 | 14            |   | 0                        | 0      | 1,215  |   | 5,5                     | 4,5    | 11,3225 |   | 7,095445 |
| 39 | 15            |   | 0                        | 0      | 0      |   | 5,5                     | 4,5    | 9,5     |   | 6,747072 |
| 40 |               |   |                          |        |        |   |                         |        |         |   |          |

Рисунок 4.10. Значения факторов в точках плана и значения уexp

Сформируем на листе расширенную матрицу плана (рис. 4.31), представленную в общем виде, справа припишем столбец значений Yexp.

|    |               |    |        |        |        |    |    |    |    |       |       |       |          |
|----|---------------|----|--------|--------|--------|----|----|----|----|-------|-------|-------|----------|
| 42 | A             | B  | C      | D      | E      | F  | G  | H  | I  | J     | K     | L     | M        |
| 43 | номер<br>эксп | x0 | x1     | x2     | x3     | x4 | x5 | x6 | x7 | x8    | x9    | x10   | Yexp     |
| 44 | 1             | 1  | -1     | -1     | -1     | 1  | 1  | 1  | -1 | 0,27  | 0,27  | 0,27  | 5,397447 |
| 45 | 2             | 1  | 1      | -1     | -1     | -1 | -1 | -1 | 1  | 0,27  | 0,27  | 0,27  | 7,630355 |
| 46 | 3             | 1  | -1     | 1      | -1     | -1 | 1  | -1 | 1  | 0,27  | 0,27  | 0,27  | 5,508893 |
| 47 | 4             | 1  | 1      | 1      | -1     | 1  | -1 | -1 | -1 | 0,27  | 0,27  | 0,27  | 7,6002   |
| 48 | 5             | 1  | -1     | -1     | 1      | 1  | -1 | -1 | 1  | 0,27  | 0,27  | 0,27  | 6,115107 |
| 49 | 6             | 1  | 1      | -1     | 1      | -1 | 1  | -1 | -1 | 0,27  | 0,27  | 0,27  | 7,116865 |
| 50 | 7             | 1  | -1     | 1      | 1      | -1 | -1 | 1  | -1 | 0,27  | 0,27  | 0,27  | 6,438498 |
| 51 | 8             | 1  | 1      | 1      | 1      | 1  | 1  | 1  | 1  | 0,27  | 0,27  | 0,27  | 8,507912 |
| 52 | 9             | 1  | -1,215 | 0      | 0      | 0  | 0  | 0  | 0  | 0,75  | -0,73 | -0,73 | 5,824821 |
| 53 | 10            | 1  | 1,215  | 0      | 0      | 0  | 0  | 0  | 0  | 0,75  | -0,73 | -0,73 | 7,161622 |
| 54 | 11            | 1  | 0      | -1,215 | 0      | 0  | 0  | 0  | 0  | -0,73 | 0,75  | -0,73 | 6,215063 |
| 55 | 12            | 1  | 0      | 1,215  | 0      | 0  | 0  | 0  | 0  | -0,73 | 0,75  | -0,73 | 7,122938 |
| 56 | 13            | 1  | 0      | 0      | -1,215 | 0  | 0  | 0  | 0  | -0,73 | -0,73 | 0,75  | 7,361126 |
| 57 | 14            | 1  | 0      | 0      | 1,215  | 0  | 0  | 0  | 0  | -0,73 | -0,73 | 0,75  | 7,095445 |
| 58 | 15            | 1  | 0      | 0      | 0      | 0  | 0  | 0  | 0  | -0,73 | -0,73 | -0,73 | 6,747072 |

Рисунок 2.31

Далее следует вычислить коэффициенты  $b_i$  полинома регрессии, используя формулы из раздела 1.5

$$b_i = (\sum_{j=1}^N y^{jexp} x_i^j) / (\sum_{j=1}^N (x_i^j)^2), \quad i=0,1, 2, \dots, 10. \quad N=15.$$

Для этого предварительно рассчитаем в ячейках [B59:L59] суммы квадратов значений соответствующих столбцов. На рисунке (4.32) показан фрагмент таблицы с формулами.

|    | A          | B                | C                | D                | E                | F                | G                |
|----|------------|------------------|------------------|------------------|------------------|------------------|------------------|
| 42 |            |                  |                  |                  |                  | x1x2             | x1x3             |
| 43 | номер эксп | x0               | x1               | x2               | x3               | x4               | x5               |
| 44 | 1          | 1                | -1               | -1               | -1               | =C44*D44         | =C44*E44         |
| 45 | 2          | 1                | 1                | -1               | -1               | =C45*D45         | =C45*E45         |
| 46 | 3          | 1                | -1               | 1                | -1               | =C46*D46         | =C46*E46         |
| 47 | 4          | 1                | 1                | 1                | -1               | =C47*D47         | =C47*E47         |
| 48 | 5          | 1                | -1               | -1               | 1                | =C48*D48         | =C48*E48         |
| 49 | 6          | 1                | 1                | -1               | 1                | =C49*D49         | =C49*E49         |
| 50 | 7          | 1                | -1               | 1                | 1                | =C50*D50         | =C50*E50         |
| 51 | 8          | 1                | 1                | 1                | 1                | =C51*D51         | =C51*E51         |
| 52 | 9          | 1                | =-alf            | 0                | 0                | =C52*D52         | =C52*E52         |
| 53 | 10         | 1                | =alf             | 0                | 0                | =C53*D53         | =C53*E53         |
| 54 | 11         | 1                | 0                | =-alf            | 0                | =C54*D54         | =C54*E54         |
| 55 | 12         | 1                | 0                | =alf             | 0                | =C55*D55         | =C55*E55         |
| 56 | 13         | 1                | 0                | 0                | =-alf            | =C56*D56         | =C56*E56         |
| 57 | 14         | 1                | 0                | 0                | =alf             | =C57*D57         | =C57*E57         |
| 58 | 15         | 1                | 0                | 0                | 0                | =C58*D58         | =C58*E58         |
| 59 | суммкв     | =СУММКВ(B44:B58) | =СУММКВ(C44:C58) | =СУММКВ(D44:D58) | =СУММКВ(E44:E58) | =СУММКВ(F44:F58) | =СУММКВ(G44:G58) |

Рисунок 4.32. Вычисление сумм квадратов  $\sum_{j=1}^N (x_i^j)^2$ ,  $i=0,1,2,\dots,10$ ,  $N=15$

Добавим к таблице [A42:M59] дополнительные столбцы N-X, в которых вычисляются произведения  $y^{jexp} x_i^j$  и затем вычисляются  $b_i$  [N59:X59] путем деления этих произведений на соответствующие суммы квадратов (рис. 4. 33). Столбцы С-К на рисунке скрыты.

|    | A          | B  | L        | M    | N         | O         | P        | Q        | R        | S        | T        | U        | V         | W         | X         | Y     |
|----|------------|----|----------|------|-----------|-----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-------|
| 40 |            |    |          |      |           |           |          |          |          |          |          |          |           |           |           |       |
| 41 |            |    |          |      |           |           |          |          |          |          |          |          |           |           |           |       |
| 42 |            |    | x3^2-a   |      |           |           |          |          |          |          |          |          |           |           |           |       |
| 43 | номер эксп | x0 | x10      | Yexp | x0Y       | x1Y       | x2Y      | x3Y      | x4Y      | x5Y      | x6Y      | x7Y      | x8Y       | x9Y       | x10Y      | Yregr |
| 44 | 1          | 1  | 0,27     | 5,40 | 5,3974473 | -5,397447 | -5,39745 | -5,39745 | 5,397447 | 5,397447 | 5,397447 | -5,39745 | 1,457311  | 1,4573108 | 1,4573108 | 5,56  |
| 45 | 2          | 1  | 0,27     | 7,63 | 7,6303547 | 7,630355  | -7,63035 | -7,63035 | -7,63035 | 7,630355 | 7,630355 | 7,630355 | 2,060196  | 2,0601958 | 2,0601958 | 7,59  |
| 46 | 3          | 1  | 0,27     | 5,51 | 5,5088931 | -5,508893 | 5,508893 | -5,50889 | -5,50889 | 5,508893 | -5,50889 | 5,508893 | 1,487401  | 1,4874011 | 1,4874011 | 5,76  |
| 47 | 4          | 1  | 0,27     | 7,60 | 7,6001996 | 7,6002    | 7,6002   | -7,6002  | 7,6002   | -7,6002  | -7,6002  | -7,6002  | 2,052054  | 2,0520539 | 2,0520539 | 7,64  |
| 48 | 5          | 1  | 0,27     | 6,12 | 6,1151069 | -6,115107 | -6,11511 | 6,115107 | 6,115107 | -6,11511 | -6,11511 | 6,115107 | 1,651079  | 1,6510789 | 1,6510789 | 6,08  |
| 49 | 6          | 1  | 0,27     | 7,12 | 7,1168649 | 7,116865  | -7,11686 | 7,116865 | -7,11686 | 7,116865 | -7,11686 | -7,11686 | 1,921554  | 1,9215535 | 1,9215535 | 6,88  |
| 50 | 7          | 1  | 0,27     | 6,44 | 6,4384985 | -6,438498 | 6,438498 | 6,438498 | -6,4385  | -6,4385  | 6,438498 | -6,4385  | 1,738395  | 1,7383946 | 1,7383946 | 6,49  |
| 51 | 8          | 1  | 0,27     | 8,51 | 8,5079118 | 8,507912  | 8,507912 | 8,507912 | 8,507912 | 8,507912 | 8,507912 | 8,507912 | 2,297136  | 2,2971362 | 2,2971362 | 8,36  |
| 52 | 9          | 1  | -0,73    | 5,82 | 5,8248207 | -7,077157 | 0        | 0        | 0        | 0        | 0        | 0        | 4,346627  | -4,252119 | -4,252119 | 5,48  |
| 53 | 10         | 1  | -0,73    | 7,16 | 7,1616223 | 8,701371  | 0        | 0        | 0        | 0        | 0        | 0        | 5,344182  | -5,227984 | -5,227984 | 7,48  |
| 54 | 11         | 1  | -0,73    | 6,22 | 6,2150634 | 0         | -7,5513  | 0        | 0        | 0        | 0        | 0        | -4,536996 | 4,6378357 | -4,536996 | 6,33  |
| 55 | 12         | 1  | -0,73    | 7,12 | 7,1229377 | 0         | 8,654369 | 0        | 0        | 0        | 0        | 0        | -5,199745 | 5,3153142 | -5,199745 | 6,98  |
| 56 | 13         | 1  | 0,75     | 7,36 | 7,3611263 | 0         | 0        | -8,94377 | 0        | 0        | 0        | 0        | -5,373622 | -5,373622 | 5,4930565 | 7,02  |
| 57 | 14         | 1  | 0,75     | 7,10 | 7,0954454 | 0         | 0        | 8,620966 | 0        | 0        | 0        | 0        | -5,179675 | -5,179675 | 5,2947988 | 7,41  |
| 58 | 15         | 1  | -0,73    | 6,75 | 6,7470717 | 0         | 0        | 0        | 0        | 0        | 0        | 0        | -4,925362 | -4,925362 | -4,925362 | 6,77  |
| 59 | суммкв     | 15 | 4,361404 |      | 6,79      | 0,82      | 0,26     | 0,16     | 0,12     | -0,16    | 0,20     | 0,15     | -0,20     | -0,08     | 0,30      |       |
| 60 |            |    |          |      | v0        | v1        | v2       | v3       | v4       | v5       | v6       | v7       | v8        | v9        | v10       |       |
| 61 |            |    |          |      |           |           |          |          |          |          |          |          |           |           |           |       |

Рисунок 4.33. Вычисление коэффициентов полинома регрессии

Далее вычислим Yregr. Используется формула  $y = b_0 + \sum_{i=1}^k b_i x_i$ , которая для ячейки [Y44] приобретает вид

[Y44]=N\$59\*B44+O\$59\*C44+P\$59\*D44+Q\$59\*E44+R\$59\*F44+S\$59\*G44+T\$59\*H44+U\$59\*I44+V\$59\*J44+W\$59\*K44+X\$59\*L44

В результате расчетов получен следующий регрессионный полином в кодированных переменных

$$\hat{y} = 6,79 + 0,82x_1 + 0,26x_2 + 0,16x_3 + 0,12x_1x_2 - 0,16x_1x_3 + 0,20x_2x_3 + 0,15x_1x_2x_3 - 0,20(x_1^2 - 0,73) - 0,08(x_2^2 - 0,73) + 0,3(x_3^2 - 0,73) .$$

**Элементы автоматизации планирования эксперимента с использованием VBA+Excel**

Добавим в документ, используя средства VBA, форму для прогнозирования значения результирующего признака  $y$  в зависимости от введенных значений факторов.

Используя вкладку (рис. 4.34) «РАЗРАБОТЧИК», вставить элемент ActiveX «Кнопка» с заголовком «Прогнозирование» на рабочий лист. В нашем случае название листа PLExp\_00. Это имя будет использоваться в дальнейшем.

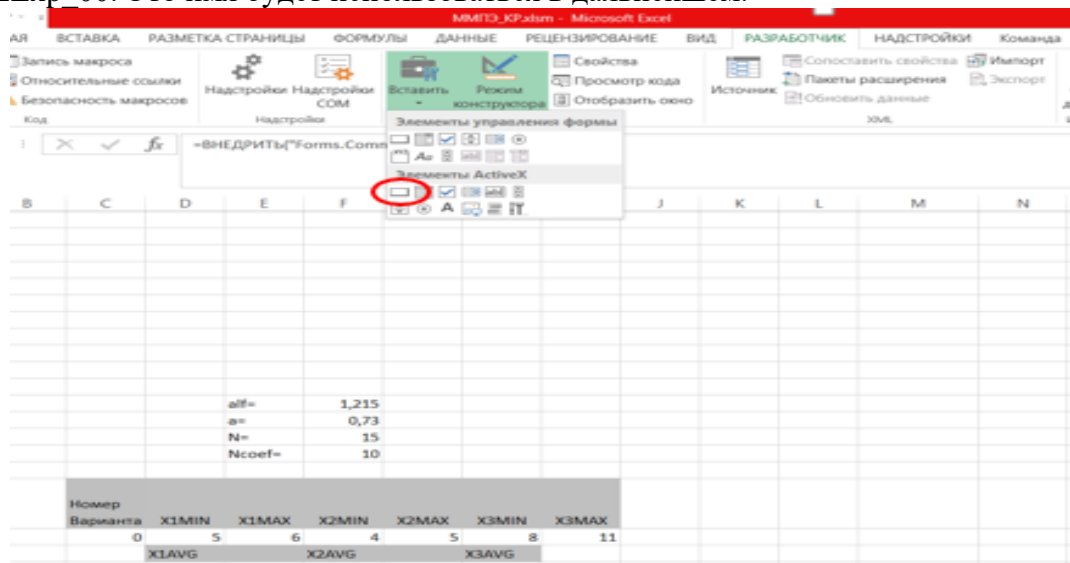


Рисунок 4.34. Добавление кнопки «Прогнозирование»

В режиме конструктора (двойной щелчок по кнопке) (рис. 4.35) настроить ее свойства в окне Properties (изменить Caption на «Прогнозирование») и написать процедуру обработки события (рис. 4.36): по нажатию кнопки вызывается форма прогнозирования Form-Progn.

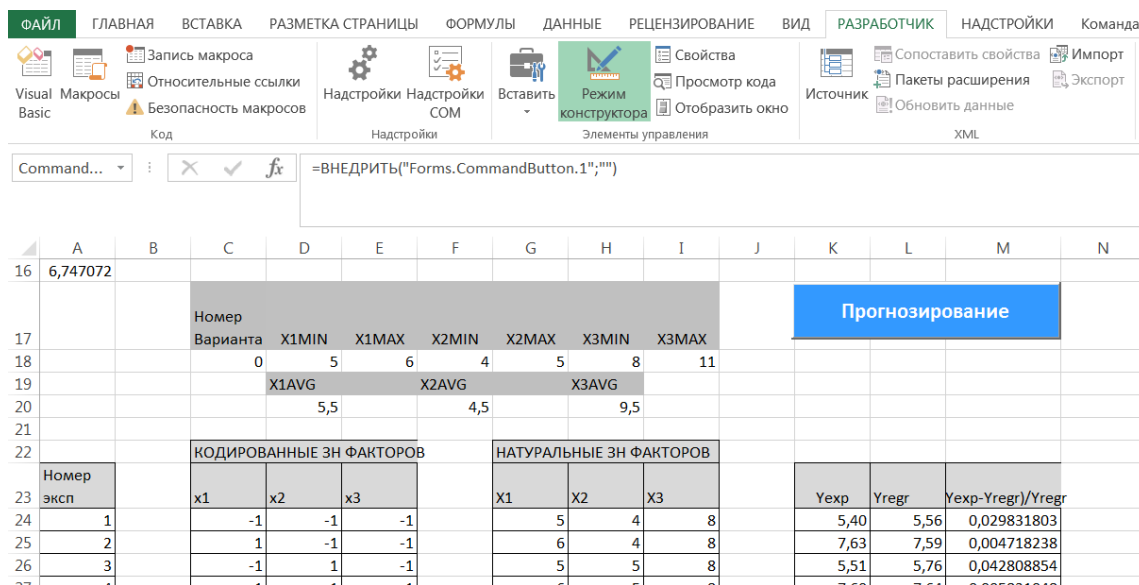


Рисунок 4.35

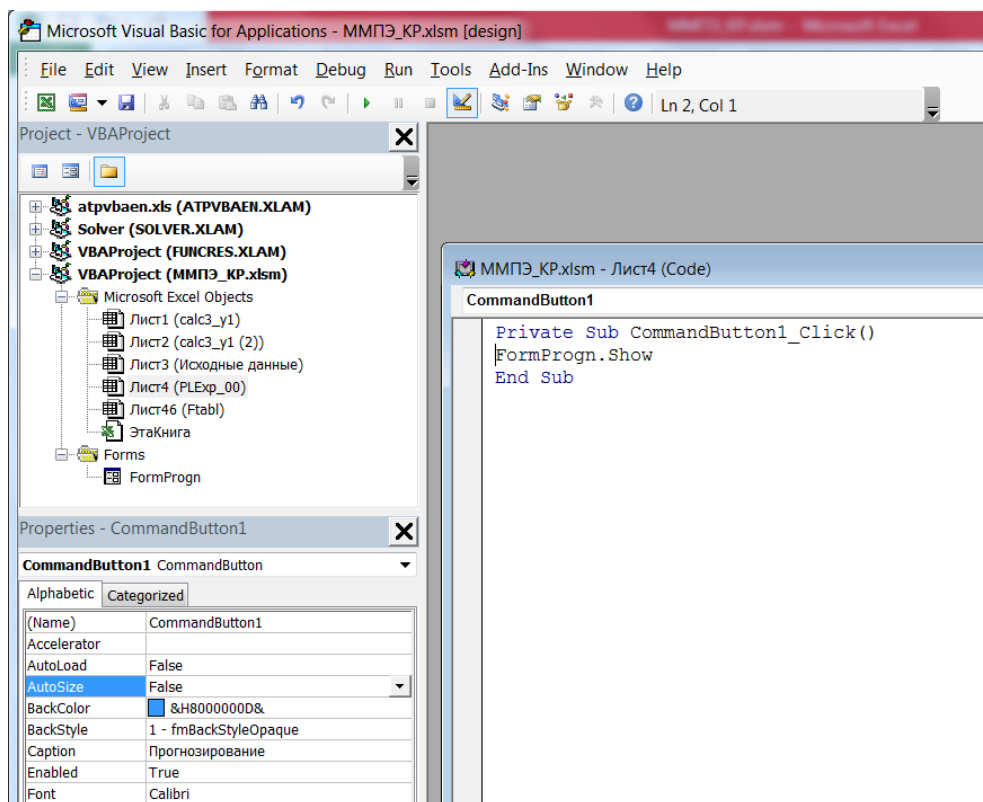


Рисунок 4.36

В редакторе VBA (переход ALT+F11 или с вкладки РАЗРАБОТЧИК) создать форму FormProgn (Insert→UserForm). Разместить на ней объекты (TextBox, CommandButton, Label) с панели Toolbox, как показано на рисунке 4.37.

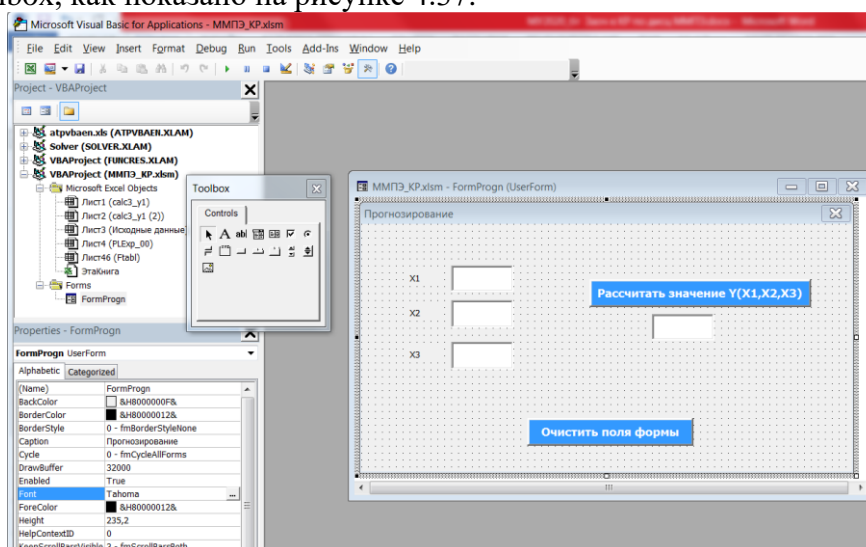


Рисунок 4.37. Объекты формы «Прогнозирование»

Настроить свойства объектов, используя панель свойств Properties в левом нижнем углу. Свойство Name для каждого элемента будет использоваться в дальнейшем в коде при описании процедур обработки событий.

В модуле формы FormProgn (окно кода формы вызывается двойным щелчком по форме или кнопкой View Code в окне проектов Project) написать две процедуры обработки событий, соответствующие нажатию командных кнопок на форме.

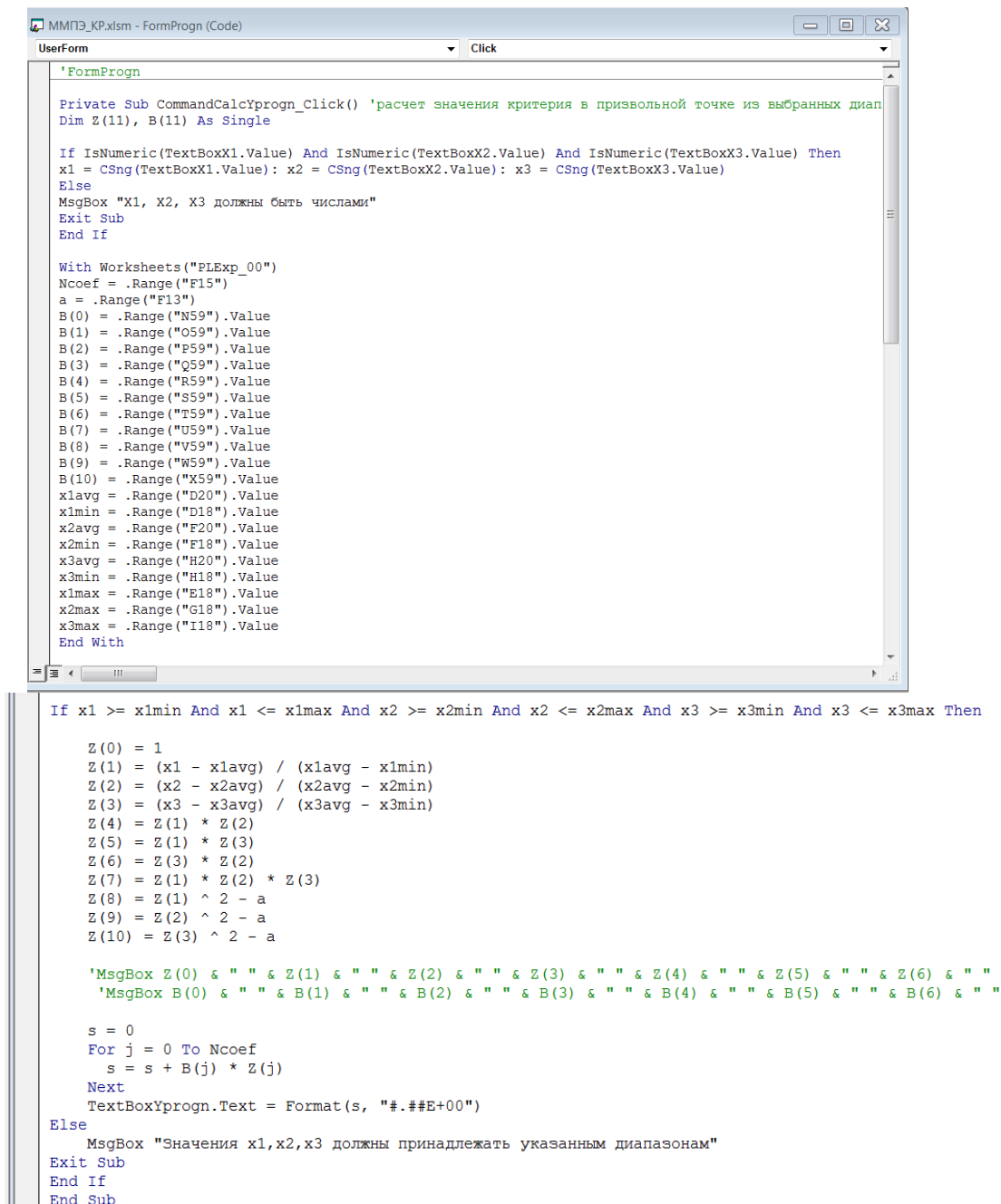


Рисунок 4.38. Процедура обработки события: нажатия кнопки  
«Рассчитать значения Y»

Расчетный лист Excel, имя которого используется в коде, имеет название PLExp\_00.

```

Private Sub CommandClean_Click() 'очистка текстовых полей

TextBoxX1.Value = ""
TextBoxX2.Value = ""
TextBoxX3.Value = ""
TextBoxYprogn.Value = ""
End Sub

```

Рисунок 4.39. Процедура обработки события: нажатия кнопки «Очистить поля формы»

Проверим работу приложения. Переключимся на лист PLExp\_00 и нажмем кнопку «Прогнозирование» (режим Конструктора должен быть отключен). В результате появится

созданная форма. Заполним текстовые поля X1,X2,X3. Программа контролирует, чтобы задаваемые значения были корректными и принадлежали указанным диапазонам. Далее нажимаем кнопку рассчитать и получаем результат (рис. 4.40)  $Y=6,88$ .

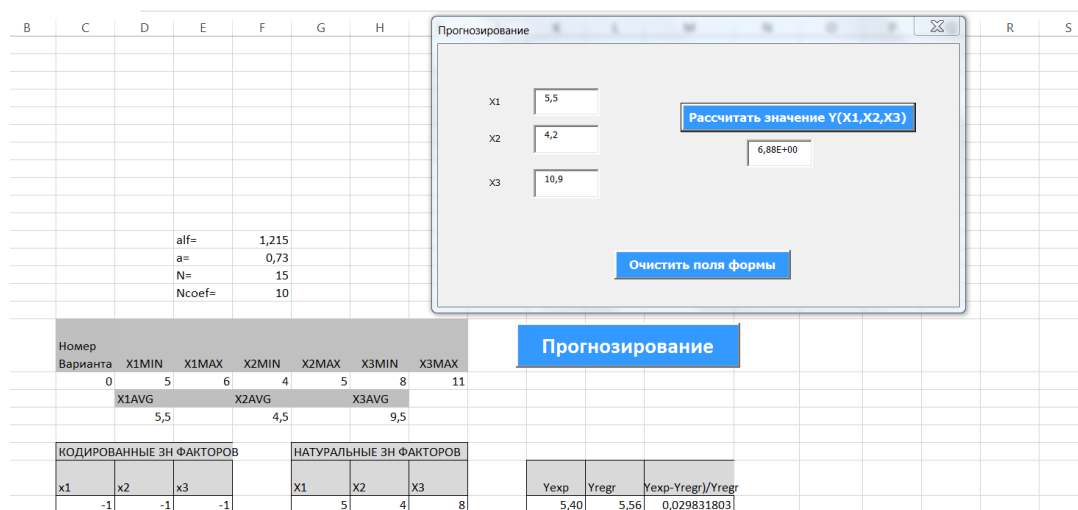


Рисунок 4.40. Результаты работы приложения «Прогнозирование»

## ЛИТЕРАТУРА

1. Онлайн-курс по дисциплине «Модели и методы планирования экспериментов». [Электронный ресурс]. – URL: <https://do.skif.donstu.ru/course/view.php?id=781>
2. Чураков Е.П. Введение в многомерные статистические методы / Е.П. Чураков. – Санкт-Петербург : Лань, 2022. - 148 с.
3. Статистический анализ и теория планирования эксперимента: учебное пособие / Н.И. Сидняев. – Москва : Издательство МГТУ им. Н.Э. Баумана, 2017. – 25 с.
4. Дубров А. М. Многомерные статистические методы: учебник / А. М. Дубров А., В. С. Мхитарян, Л. И. Трошин Л.И. – М.: Финансы и статистика, 2011. – 349 с.
5. Волчёнков Н.Г. Основы программирования на языке Visual Basic для офисных приложений: учебное пособие / Н.Г. Волчёнков. – М.: НИЯУ МИФИ, 2018. – 166 с.
6. Интерактивная среда для программирования на Python в браузере. [Электронный ресурс]. – URL: <https://colab.research.google.com/>
7. Таблица значений критерия Фишера. [Электронный ресурс]. – URL: <https://math.semestr.ru/corel/table-fisher.php>
8. URL : <https://docs.google.com/spreadsheets/d/1zar6xAafiPuZ80XvDw2yi6Rq7Cf6M6QSjWCZQSP1Ebg/edit?usp=sharing>