

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Емельянов Сергей Геннадьевич
Должность: ректор
Дата подписания: 16.12.2021 20:54:40
Уникальный программный ключ:
9ba7d3e34c012eba476ffd2d064cf2781953be730df2374d16f3c0ce536f0fc6

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
« 1 » _____ 2018 г.



МЕТОДЫ ОБРАБОТКИ МНОГОМЕРНЫХ СИГНАЛОВ И ДААННЫХ

Методические рекомендации по организации и выполнению
практической работы для аспирантов направлений подготовки
09.06.01

Курск 2018

УДК 004.93:61

Составитель: С.А. Филист.

Рецензент

Доктор технических наук, профессор А.Ф. Рыбочкин

Методы обработки многомерных сигналов и данных: Методические рекомендации по организации и выполнению практической работы / Юго-Зап. гос. ун-т; сост.: С.А. Филист. - Курск, 2018. - 119 с.

Методические указания по структуре, содержанию и стилю изложения материала соответствуют требованиям, предъявляемым к учебным и методическим пособиям.

Предназначены для аспирантов направлений подготовки 09.06.01 «Информатика и вычислительная техника (Системный анализ, управление и обработка информации (технические и медицинские системы))»

Текст печатается в авторской редакции

Подписано в печать 1.03.18. Формат 60x84 1/16.

Усл.печ.л. 6,92. Уч.-изд.л. 6,26 Тираж 100 экз. Заказ: 1434. Бесплатно.

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

ПРАКТИЧЕСКАЯ РАБОТА №1

ИССЛЕДОВАНИЕ МЕТОДОВ ФОРМИРОВАНИЯ ФАЙЛОВ ДАННЫХ С ЦИФРОВЫМИ ОТСЧЕТАМИ СИГНАЛОВ

1.1 Краткие теоретические сведения

Во многих случаях непрерывные сигналы, поступающие с выхода измерительного устройства, не обрабатываются непосредственно, а сначала подвергаются дискретизации, то есть наблюдаются только в определенные моменты времени. В общем случае наблюдения производятся периодически через постоянный промежуток времени T – шаг дискретизации. Следовательно, для физической реализации процесса дискретизации необходимо осуществить свертку дискретизируемой функции $x(\tau)$ с импульсом Дирака:

$$x(\tau) = \int_{-\infty}^{\infty} x(t)\delta(t - \tau)dt \quad (1.1)$$

при $\tau = 0, T, 2T, 3T \dots$

Перейдем от периода дискретизации T к частоте дискретизации $F = \frac{1}{T}$ и перепишем равенство (1.1) в дискретной форме:

$$x(\tau) = x(t) \sum_{k=-\infty}^{\infty} \delta\left(t - \frac{k}{F_c}\right). \quad (1.2)$$

Учитывая формулу Пуассона, согласно которой $\sum_{k=-\infty}^{\infty} \delta\left(t - \frac{k}{F_c}\right) \Leftrightarrow F_c \sum_{n=-\infty}^{\infty} \delta(v - nF_c)$, и теорему свертки, согласно которой умножение оригиналов соответствует свертка Фурье-образов, из (1.2) получим:

$$x(\tau) \Leftrightarrow F_c X(v) \cdot \sum_{n=-\infty}^{\infty} \delta(v - nF_c). \quad (1.3)$$

Из выражения (1.3) следует, что спектр дискретного сигнала $x(\tau)$ представляет собой «периодическую» функцию с периодом F_c , показанную на рисунке 1.1.

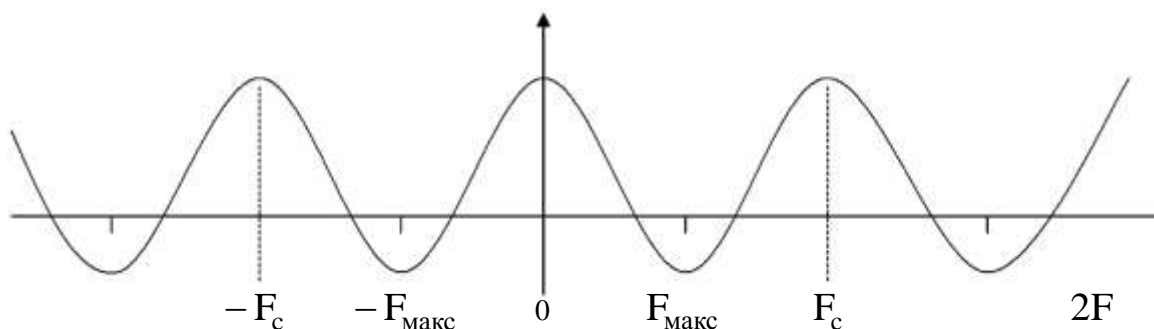


Рисунок 1.1 – Спектр дискретного сигнала

В данном случае спектр сигнала $x(t)$ расположен в интервале $(-F_{\text{макс}}, F_{\text{макс}})$ и, согласно теореме Шеннона, для того, чтобы периодическое повторение спектра, вызванное дискретизацией сигнала, не изменяло повторяемый спектр, необходимо и достаточно выполнения неравенства $F_c \geq 2F_{\text{макс}}$.

В этом случае сигнал может быть восстановлен по дискретным значениям по интерполяционной формуле

$$x(t) = \sum_{k=-\infty}^{\infty} x\left(\frac{k}{F_c}\right) \cdot \frac{\sin \pi F_c (t - k/F_c)}{\pi F_c (t - k/F_c)}. \quad (1.4)$$

Для ответа на вопрос «Как следует на практике производить дискретизацию сигнала?» – необходимо знать, с какой целью осуществляется эта операция: в целях дальнейших вычислений или в целях восстановления сигнала.

Пусть мы дискретизируем сигнал $x(t)$, в целях дальнейших вычислений, который имеет спектр Фурье $X(\nu)$, где ν лежит в интервале $[-F; F]$. Согласно теореме Шеннона, шаг дискретизации T должен удовлетворять неравенству $T \leq \frac{1}{2} \cdot F$. Если дискретизация функции $x(t)$ произведена при выполнении этих условий, то по

значениям $x_k = x(kT)$ функции $x(t)$ в точках дискретизации можно восстановить непрерывный сигнал $x(t)$ по интерполяционной формуле Шеннона (1.4).

Таким образом, вычисление любых значений функции $x(t)$ сводится к вычислению ее значений $x(kT)$ в точках дискретизации, то есть вся информация о функции $x(t)$ содержится в точках дискретизации.

Пусть мы дискретизируем сигнал с целью восстановления непрерывного сигнала.

Хотя интерполяционная формула Шеннона (1.4) теоретически обоснована, ее практическое применение, особенно при приближенных вычислениях, затруднительно.

Обозначим через $S(t)$ и $S_1(t)$ точный и восстановленный сигналы. Пусть $\varepsilon = \Delta S_M(t) / S_M(t) = (S_M(t) - S_1(t)) / S_M(t)$, где $S_M(t)$ – максимальное значение $S(t)$. Доказано, что для синусоидального сигнала с частотой F , частота дискретизации F_c должна удовлетворять неравенству:

$$F_c \geq F \left(\frac{2,2}{\sqrt{\varepsilon}} \right),$$

где ε – допустимая ошибка.

Если $\varepsilon = 0,01$, то $F_c \geq 22 \cdot F$, т.е. частота дискретизации должна быть приблизительно в 10 раз больше частоты Шеннона.

Пусть спектр сигнала обрезан максимальной частотой F_c . Доказано, что в этом случае:

$$F_c \geq F_c \left(2,2 / \sqrt{\varepsilon} \right).$$

Например, если ошибка $\varepsilon \leq 0,01$, то должно быть $F_c \geq 22 \cdot F_c$.

После дискретизации из непрерывного сигнала получается дискретный сигнал. Ось времени t разбивается на равные интервалы τ . Получают, как показано на рис. 1.2, а, $N - 1$ интервала дискретизации. При этом на оси времени отмечают N точек, координатами которых являются следующие значения t : $0; \tau; 2\tau; \dots, (N - 1)\tau$.

Для каждой из этих точек берется отвечающее ей значение x : производится отсчет исходной функции в данной точке.

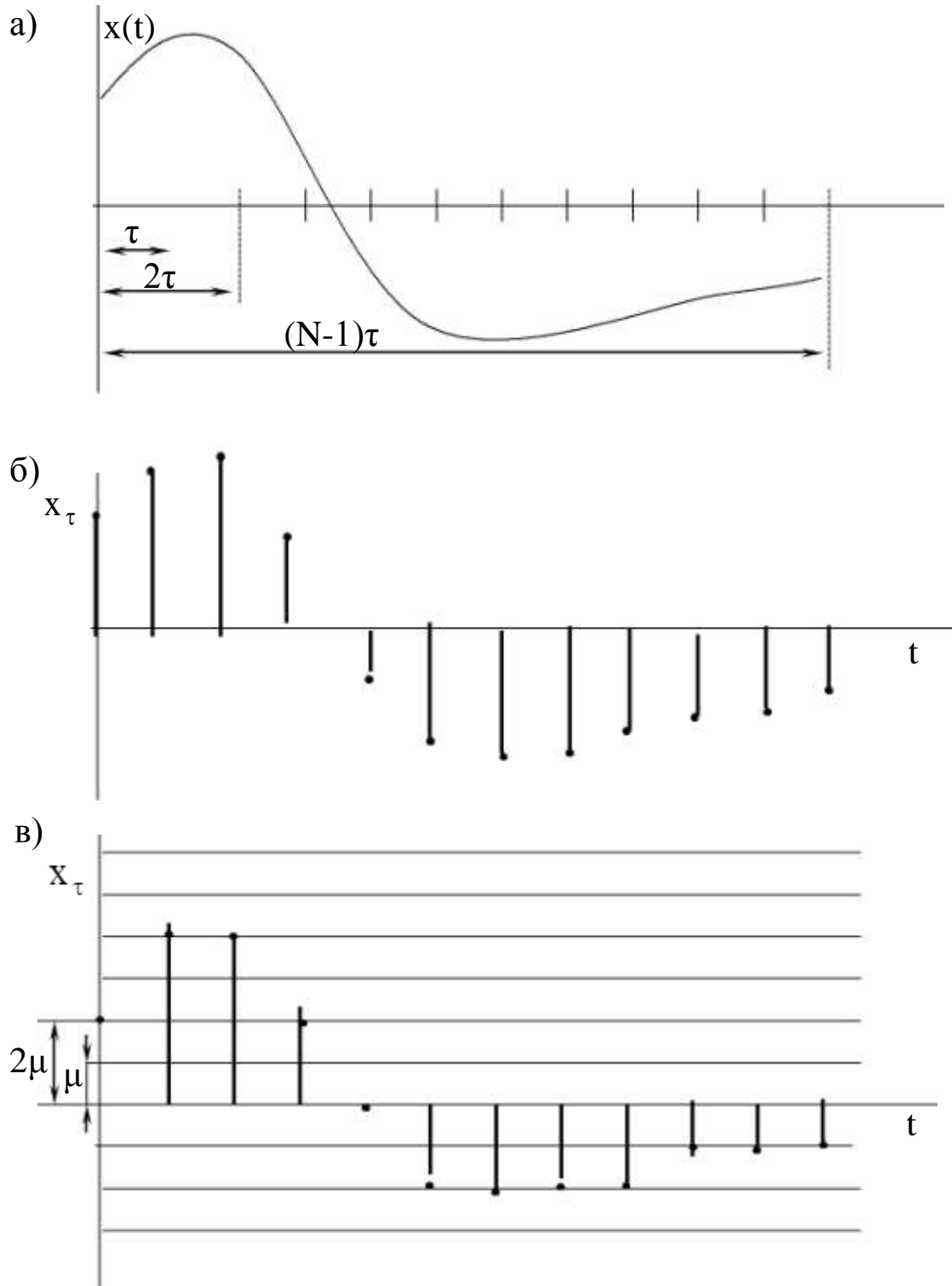


Рисунок 1.2 – а) исходный непрерывный сигнал; б) дискретизированный сигнал; в) цифровой сигнал

Принято изображать полученное дискретное значение переменной величины, опуская из каждой из указанных точек перпендикуляр на ось t . Таким образом, получается, так называемая, решетчатая функция в виде дискретных значений x , для которых примем обозначение x_t (рисунок 1.2, б).

Для того, чтобы дискретизированные указанным образом величины могли быть подвергнуты обработке с помощью цифровых устройств, необходимо из дискретного сигнала получить цифровой сигнал. Для нахождения его производится квантование по уровню: ось x_t разбивается, как показано на рисунке 1.2, в, на ряд интервалов и с точностью до единицы отсчета μ определяется величина каждой из дискрет-ординат x_t . Полученные значения ординат кодируются и в таком виде вводятся в цифровое устройство.

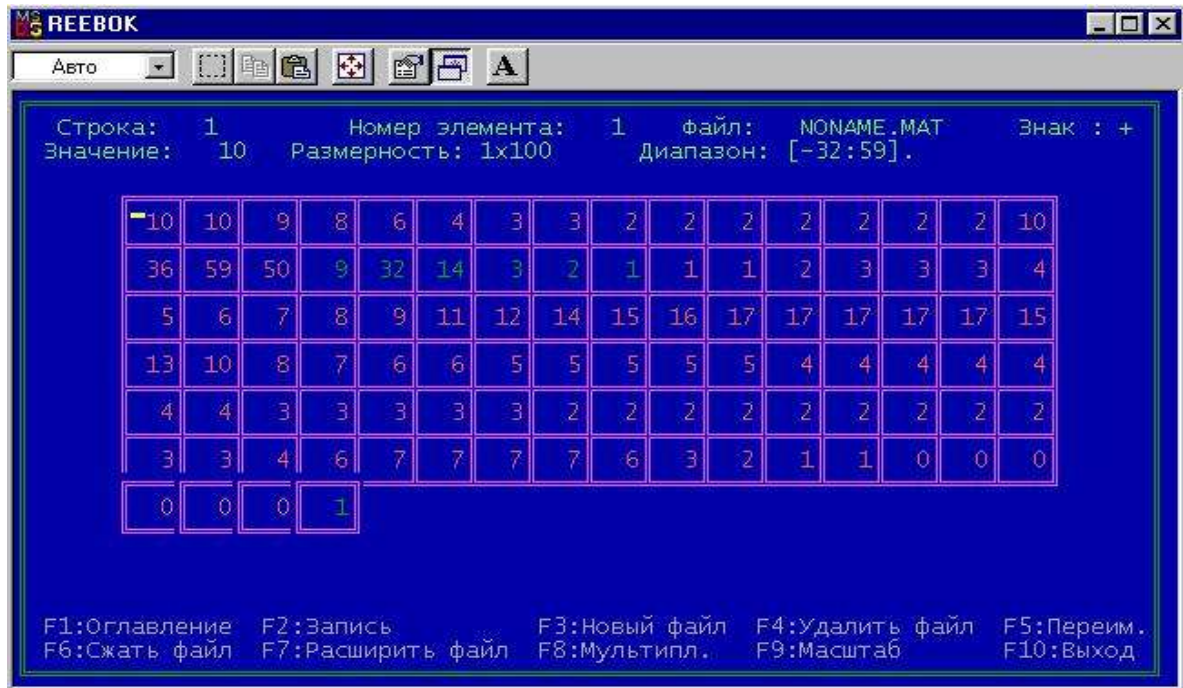
1.2. Объект исследования

Объектом исследования в данной лабораторной работе является электрокардиосигнал с различными патологиями, взятый из атласа клинических электрокардиограмм. Рекомендуется использовать электрокардиограммы из атласа [3]. Выбранный электрокардиосигнал дискретизируется и квантуется. Необходимо сначала получить ксерокопию выбранного электрокардиосигнала, а затем с помощью программы REEВОК сохранить полученные отсчеты в файле данных.

Разрешение, с которым электрокардиосигнал представлен в атласе [3], составляет 1 мм как по вертикали, так и по горизонтали. Масштабы сигналов, представленных в атласе, составляют 25 мм/с и 10 мм/мВ. Реально, на 1 мм мы можем получить не более двух отсчетов, следовательно, частота дискретизации составит не более 50 Гц. Однако это вполне достаточно для исследуемого сигнала, так как он получен с помощью электромеханического самописца, который является фильтром нижних частот и имеет большую постоянную времени.

Оцифровка электрокардиосигнала осуществляется с помощью программы REEВОК, которая имеет два окна: окно графического редактора и окно текстового редактора. Возможные виды этих окон показаны на рисунке 1.3, а и 1.3, б.

а)



б)



Рисунок 1.3 – Окна программы REEBOK: а) окно графического редактора, б) окно текстового редактора

В каждом окне имеется оконное меню, управляемое посредством функциональных клавиш. Процесс создания файла данных

начинается с нажатия клавиши F3 – создать файл. Создаваемый файл имеет заголовок, после которого идут данные по умолчанию представляемые в виде матрицы целочисленных данных, которая может иметь произвольную размерность. Максимальная размерность матрицы файла 128x128 элементов. Если векторный сигнал имеет длину более 128 (точнее более 144) отсчетов, то его целесообразно сегментировать и представить в виде матрицы, которую записать в файле с помощью программы REEВOK. Если данные необходимо обрабатывать в векторной, а не в матричной форме, то после выборки данных из файла, полученная матрица разворачивается по строкам в вектор.

Формат файла данных, получаемый с помощью этой программы, иллюстрирует Листинг 1. Как видно из него перед размещением данных в файле вначале записывается заголовок длиной в 30 байт. Таким образом, файл данных является нетипизированным и запись и считывание из него ведется с помощью процедур работы с нетипизированными файлами, ознакомиться с которыми можно в [2].

Листинг 1.

```
PFHeader = ^FHeader;
```

```
FHeader = record           { заголовок файла данных }
    DataType :string[4];   { тип данных REAL или INTG }

    Cols :longint;         { число столбцов }
    Lines :byte;           { число строк }
    Frequency :longint;    { частота дискретизации }
    Time :longint;         { время наблюдения }
    Period :longint;       { период наблюдения }
    Pause :longint;        { период паузы }

    Min,Max :integer;      { минимальный и максимальный }
                           { элементы }
end;                       { всего 30 байт }
```

В программе REEВOK из элементов заголовка, представленных в Листинге 1, указываются только число столбцов, число строк, максимальный элемент массива и минимальный элемент массива. Необходимо помнить, что хотя целочисленные переменные могут принимать значения в диапазон $(-1024...+1024)$, динамический диапазон сигнала не должен превышать 200. Это объясняется тем, что не все режимы мониторов отображают более 200 точек по вертикали.

Данные целесообразно вводить в текстовом редакторе, а корректировать в графическом редакторе. При этом в окне текстового редактора отображаются элементы только одной строки. Переход по строкам, если их больше одной, осуществляется клавишами PageUp – PageDown. Выбор элемента строки осуществляется клавишами "→", "←", "↓", "↑". После ввода числа в ячейку необходимо подтвердить ввод клавишей "Enter". В противном случае в ячейке останутся старые данные.

Подсказки и параметры ввода расположены в верхней части окна. Меню располагается в нижней части окна. После ввода данных их необходимо записать с помощью клавиши F2. Имя файла вводится после команды создать файл (F3). Расширение файла данных .mat формируется автоматически.

1.3. Цель работы

Изучить методы получения цифровых сигналов и программное обеспечение для формирования файлов данных с цифровыми отсчетами сигналов.

1.4. Порядок выполнения работы

1. Изучите раздел «Дискретизация непрерывных сигналов» [1].
2. Из атласа [3] отсканируйте страницу с электрокардиосигналами выбранной патологии. Каждый студент получает свой индивидуальный вариант (номер страницы атласа с индивидуальным вариантом записывается в журнал преподавателя). Выберите три непрерывных сигнала из разных отведений (по консультации с

преподавателем). В указанных трех сигналах обязательно должен присутствовать сигнал второго отведения. Изобразите теоретический спектр электрокардиосигнала. Определите теоретическую частоту дискретизации сигналов.

3. Изучите программу моделирования данных REEВОК и ее графический и символьный интерфейсы, обратите особое внимание на структуру файла и его заголовков.

4. Выбрав частоту дискретизации сигналов, полученных в п. 2, и задавшись остальными параметрами заголовка файла данных, получите с помощью программы REEВОК файлы данных согласно п. 1.

5. Сосчитайте заголовки файлов данных и сами данные с помощью программы, имеющейся в базе данных или разработанной самостоятельно. Сделайте необходимые выводы.

1.5. Содержание отчета

1. Ксерокопия исследуемого сигнала.
2. Распечатки окон графического редактора с исследуемыми сигналами.
3. Предполагаемый вид спектра исследуемого сигнала и расчет теоретических значений его частоты дискретизации.
4. Распечатка листингов созданных программ.
5. Распечатка заголовков созданных файлов данных.
6. Электронная форма на компакт-диске.
7. Выводы.

1.6. Контрольные вопросы

1. Чем отличается дискретный сигнал от цифрового?
2. Дайте определение Найквистовской частоты дискретизации.
3. Как изменится спектр сигнала, если он дискретизирован с частотой, меньшей, чем Найквистовская?
4. С какой целью перед дискретизацией аналоговый сигнал подвергают низкочастотной фильтрации? Как выбирается частота среза этого фильтра?

5. Нарисуйте структурную схему дискретизатора. Как в ней реализуется соотношение (1.1)?

6. Какие искажения имеют место при переходе от цифрового сигнала к непрерывному? Как реализуется этот переход?

7. Какие искажения дискретного сигнала вызывает отличие дискретизирующего импульса от δ -импульса Дирака?

8. Нарисуйте частотную характеристику усилительного тракта электрокардиосигнала.

9. С чем связаны искажения сигнала при его квантовании? Как изменится спектр функции в результате квантования?

10. С чем связаны искажения сигнала при его дискретизации? Как изменится спектр функции в результате дискретизации?

11. Объясните, почему в кино колесо отправляющего поезда сначала медленно вращается вперед, потом останавливается, а затем начинается вращаться назад?

12. Пусть мы имеем сигнал вида $\cos(8\pi/3t - \pi/3)$. Какова самая низкая искажаемая дискретизацией частота, если шаг дискретизации равен единице?

13. Как изменится спектр функции в результате квантования? Покажите, что это преобразование нелинейное.

14. Представьте алгоритм равномерного квантования функции $f(t)$.

15. Пусть мы дискретизируем функцию $\cos(8\pi/3t - \pi/3)$ с шагом дискретизации единица. Трансформируется ли при этом частота, если да, то в какую?

16. К каким сигналам: дискретным или непрерывным может быть применено поэлементное квантование?

17. Представьте алгоритм равномерного квантования функции $f(t)$ в логарифмическом масштабе.

18. Чем определяются погрешности квантования?

19. Пусть мы имеем сигнал вида $\sum_{n=-\infty}^{n=\infty} \cos(2\pi n x / 9 + \pi / 3)$. Какова самая низкая искажаемая дискретизацией частота, если дискретизация ведется в точках, соответствующих целым значениям x ?

20. Чем принципиально отличается спектр непрерывной и спектр дискретной функции?

21. Колесо велосипеда вращается с частотой 100 Гц. Какова кажущая частота вращения колеса, если стробоскоп дает вспышки с частотой 99 вспышек в 1 с?

22. Используя простые тригонометрические соотношения показать, что в точках дискретизации любая синусоида произвольной частоты f равнозначна синусоиде, лежащей в интервале $[0, 1/(2f)]$. Дискретизация ведется с найквистовской частотой в целые моменты t .

23. Приведите случаи, когда доказательство теоремы отсчетов будет некорректно.

24. Перечислите случаи, когда необходимо уменьшать шаг дискретизации по сравнению с расчетным. Почему рекомендуется это делать всегда.

25. Как изменится спектр сигнала при его дискретизации? Покажите, что это преобразование линейное.

1.7. Порядок защиты работы

Работа может быть зачтена, если студент представил отчет согласно пункта 1.5, исследуемые в работе сигналы соответствуют индивидуальному варианту, электронная форма соответствует представленному тексту, и студент дал исчерпывающие ответы на 10 произвольных вопросов из пункта 1.6.

Список литературы

1. Сергиенко, А. Б. Цифровая обработка сигналов: учебное пособие / А.Б. Сергиенко. - 2-е изд. - СПб.: Питер, 2006. - 751 с.

2. Зубов В.С. Программирование на языке TURBO PASCAL (версии 6.0 и 7.0). – М.: Информационно-издательский дом «Филинь», 1997. – 304 с.

3. Циммерман Франклин. Клиническая электрокардиография. – М.: «Издательство БИНОМ», 1997. – 448 с.

ПРАКТИЧЕСКАЯ РАБОТА №2 ИССЛЕДОВАНИЕ МЕТОДОВ ИМИТАЦИОННОГО МОДЕ- ЛИРОВАНИЯ МОДУЛИРОВАННЫХ СИГНАЛОВ

2.1. Краткие теоретические сведения

При модуляции параметров немодулированного (несущего) колебания:

$$u_{\omega} = U_0 \cos(\omega_0 t + \varphi), \quad (2.1)$$

формируется модулированное колебание, которое в общем случае можно записать в виде:

$$u = U(t) \cos[\omega_0 t + \varphi(t)] = U(t) \cos \Phi(t), \quad (2.2)$$

где $(\omega_0 -$ несущая частота; $\Phi(t) = (\omega_0 t + \varphi(t)) -$ мгновенная фаза колебания.

Амплитуда $U(t)$ и фазовый сдвиг $\varphi(t)$ колебания (2.2) изменяются в зависимости от модулирующего сигнала $s(t)$.

Спектр модулированного колебания (2.2) обычно оказывается шире спектра модулирующего сигнала.

2.1.1. Амплитудно-модулированные колебания

При амплитудной модуляции происходит изменение амплитуды $U(t)$ модулированного колебания при постоянной начальной фазе φ . Выражение (2.2) принимает вид:

$$u_{AM} = U(t) \cos(\omega_0 t + \varphi). \quad (2.3)$$

Огибающая (амплитуда) модулированного колебания изменяется вокруг среднего значения U_0 и связана с модулирующим сигналом $s(t)$ линейным соотношением:

$$U(t) = U_0 + k_A s(t), \quad (2.4)$$

где k_A – коэффициент пропорциональности, такой, что при любых t обеспечивается $U(t) \geq 0$.

Модуляция гармоническим сигналом (тональная модуляция). Если модулирующий сигнал $s(t)$ является гармоническим колебанием, т. е.

$$s(t) = S \cos(\Omega t + \Psi), \quad (2.5)$$

то, согласно (2.4), огибающая модулированного колебания имеет вид:

$$U(t) = U_0 + \Delta U \cos(\Omega t + \Psi), \quad (2.6)$$

где Ω – частота модуляции ($\Omega \ll \omega_0$); ψ – начальная фаза огибающей; $\Delta U = k_A S$ – наибольшее отклонение амплитуды от среднего значения. Мгновенное значение подобного тонально-модулированного колебания выражается как:

$$u_{AM} = U_0 [1 + M \cos(\Omega t + \Psi)] \cos(\omega_0 t + \varphi), \quad (2.7)$$

где отношение $M = \Delta U / U_0$, называемое коэффициентом модуляции, является одним из основных параметров АМ колебания.

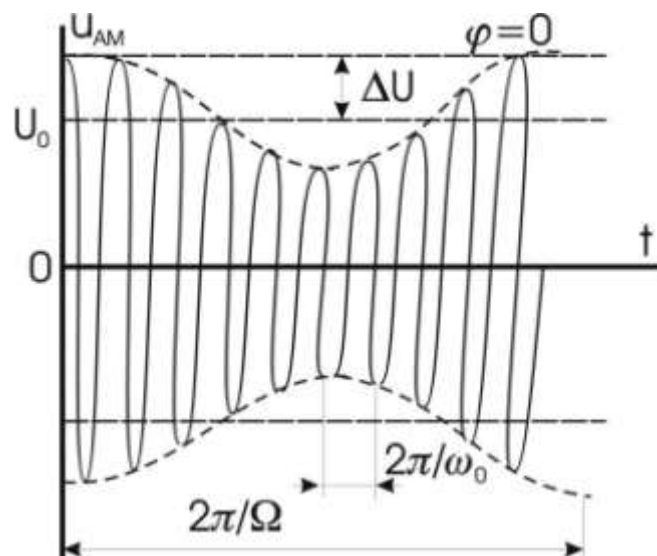


Рисунок 2.1 – Временная диаграмма колебания (2.7)

Колебание u_{AM} имеет дискретный спектр, состоящий из трех высокочастотных спектральных составляющих. Первое слагаемое правой части представляет собой исходное немодулированное колебание с частотой ω_0 и начальной фазой φ . Второе и третье слагаемые называют соответственно верхней и нижней боковыми составляющими. Они расположены симметрично относительно несущего колебания на верхней ($\omega_0 + \Omega$) и нижней ($\omega_0 - \Omega$) боковых частотах модуляции и имеют соответственно начальной фазы $(\varphi + \Psi)$ и $(\varphi - \Psi)$. Амплитуды боковых составляющих равны $U_0 M / 2$. При неискаженной модуляции $U_0 M / 2 \leq U_0 / 2$ ($M \leq 1$). Рисунок 2.2 иллюстрирует формирование спектра АМ колебания. На рисунке 2.2, а изображен спектр колебаний до модуляции. При модуляции (в результате перемножения функций $\cos \Omega t - \cos \omega_0 t$) происходит сдвиг спектра модулирующего сигнала на величину ω_0 и расширение спектра (рисунок 2.2, б). Ширина полосы частот $2\Delta\omega_{AM}$, занимаемой АМ колебанием, определяется удвоенной частотой модулирующего сигнала.

$$2\Delta\omega_{AM} = 2\Omega.$$

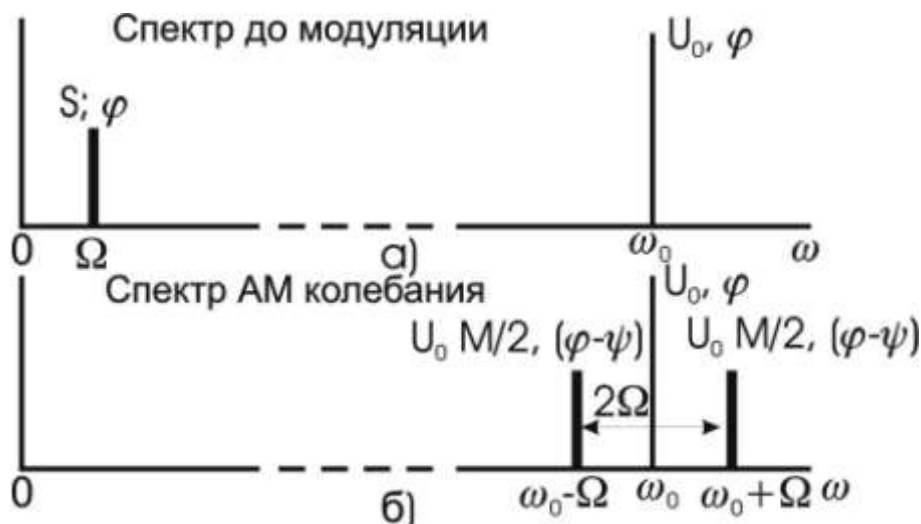


Рисунок 2,2 – Спектр колебаний до модуляции

Модуляция произвольным периодическим сигналом. Периодический модулирующий сигнал $s_n(t)$ может быть представлен рядом Фурье вида (2.8):

$$s_n(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\Omega t + \Psi_n), \quad (2.8)$$

где Ω – основная частота модулирующего сигнала.

Если отнести постоянную составляющую $A_0/2$ ряда Фурье к амплитуде несущего колебания U_0 , то огибающую $U(t)$ модулированного колебания можно записать так:

$$U(t) = U_0 + \sum_{n=1}^{\infty} \Delta U_n \cos(n\Omega t + \psi_n), \quad (2.9)$$

где $\Delta U_n = k_A A_n$.

Каждая спектральная составляющая модулирующего сигнала (2.8) так же, как при тональной модуляции, создает две боковые частоты в спектре модулированного колебания.

Если ширину спектра $\Delta\Omega$ модулирующего сигнала определить как:

$$\Delta\Omega = \Omega_{\max} = m\Omega,$$

где Ω_{\max} – максимальная частота, которая учитывается в спектре этого сигнала, соответствующая m -й гармонике, то ширина спектра АМ колебания равна:

$$2\Delta\omega_{AM} = 2\Omega_{\max}.$$

Модуляция непериодическим сигналом. В случае модуляции непериодическим сигналом $s(t)$ со спектральной плотностью $G(\omega)$ огибающую $U(t)$ и АМ колебание $u_{AM}(t)$ можно записать в виде:

$$U(t) = U_0 + k_A s(t), \quad (2.10)$$

$$u_{AM}(t) = [U_0 + k_A s(t)] \cos(\omega_0 t + \varphi), \quad (2.11)$$

спектральная плотность огибающей, сосредоточенная в области нижних частот, «раздваивается» и переносится в область высоких частот, смещаясь на $\pm\omega_0$. спектральная плотность $G_{AM}(\omega)$ АМ колебания имеет две составляющие: $1/2G_u(\omega - \omega_0)$ и $1/2G_u(\omega + \omega_0)$, сконцентрированные вблизи частот ω_0 и $-\omega_0$ соответственно. Ширина спектра $2\Delta\omega_{AM}$ АМ колебания вдвое превышает ширину спектра $\Delta\omega_U$ огибающей.

2.1.2. Колебания с угловой модуляцией

Общие понятия. При угловой модуляции под действием модулирующего сигнала происходит изменение фазового сдвига $\varphi(t)$ несущего колебания (2.1) при постоянной амплитуде U_0 . Модулированное колебание (2.2) принимает вид

$$u(t) = U_0 \cos[\omega_0 t + \varphi] = U_0 \cos \Phi(t). \quad (2.12)$$

Изменение фазового сдвига $\varphi(t)$ происходит как при модуляции мгновенной частоты $\omega(t)$, так и при модуляции непосредственно фазового сдвига $\omega(t)$ колебания (2.12). Поэтому различают два вида угловой модуляции: частотную модуляцию (ЧМ) и фазовую модуляцию (ФМ).

При частотной модуляции имеет место линейная связь между мгновенной частотой $\omega(t)$ колебания (2.12) и модулирующим сигналом $s(t)$:

$$\omega(t) = \omega_0 + \omega_D s(t)^*, \quad (2.13)$$

где ω_D – девиация частоты, т. е. максимальное отклонение частоты $\omega(t)$ от ω_0 .

Полная фаза $\Phi(t)$ ЧМ колебания определяется как интеграл от мгновенной частоты (2.13):

$$\begin{aligned}\Phi(t) &= \int \omega(t)dt + \varphi = \int [\omega_0 + \omega_D s(t)] dt + \varphi = \\ &= \omega_0 t + \omega_D \int s(t)dt + \varphi = \omega_0 t + \varphi(t),\end{aligned}\quad (2.14)$$

где фазовый сдвиг:

$$\varphi(t) = \omega_D \int s(t)dt + \varphi. \quad (2.15)$$

При этом ЧМ колебание принимает вид:

$$u_{\text{ЧМ}} = U_0 \cos[\omega_0 t + \omega_D \int s(t)dt + \varphi]. \quad (2.16)$$

Таким образом, фазовый сдвиг $\varphi(t)$ ЧМ колебания изменяется по закону интеграла от модулирующего сигнала $s(t)$.

Частотно-модулированный сигнал можно получить на выходе частотного модулятора, в котором модулирующий сигнал $s(t)$ либо непосредственно управляет угловой частотой $\omega(t)$ несущего колебания в соответствии с выражением (2.13), либо после интегрирования управляет фазовым сдвигом $\varphi(t)$ несущего колебания согласно выражению (2.15).

При фазовой модуляции в соответствии с модулирующим сигналом $s(t)$ изменяется фазовый сдвиг:

$$\varphi(t) = \varphi + m s(t), \quad (2.17)$$

где φ – начальная фаза немодулированного колебания; m – индекс угловой модуляции, т. е. максимальное отклонение фазы колебания (2.12) от начальной фазы φ . Учитывая (2.17), полную фазу $\Phi(t)$ модулированного колебания (2.12) можно определить по формуле

$$\Phi(t) = \omega_0 t + m s(t) + \varphi. \quad (2.18)$$

При этом фазомодулированное колебание принимает вид:

$$u_{\text{ФМ}} = U_0 \cos[\omega_0 t + m s(t) + \varphi]. \quad (2.19)$$

Мгновенная частота ФМ колебания, равная:

$$\omega(t) = \frac{d\Phi(t)}{dt} = \omega_0 + m \frac{ds(t)}{dt}, \quad (2.20)$$

изменяется по закону производной от модулирующего сигнала $s(t)$.

Следовательно, ФМ сигнал можно получить на выходе фазового модулятора, в котором модулирующий сигнал $s(t)$ либо непосредственно управляет фазовым сдвигом $\varphi(t)$ несущего колебания согласно выражению (2.17), либо после дифференцирования управляет угловой частотой $\omega(t)$ несущего колебания в соответствии с (2.20).

Ограничимся рассмотрением ЧМ и ФМ колебаний при модуляции гармоническим сигналом:

$$s(t) = S \cos \Omega t. \quad (2.21)$$

Сравним основные характеристики сигналов с частотной и фазовой модуляцией.

ЧМ колебание.

Согласно (2.13)

$$\omega(t) = \omega_0 + \omega_d \cos \Omega t. \quad (2.22)$$

Здесь $\omega_d = k_{\text{ЧМ}} S$, т.е. девиация частоты ЧМ колебания пропорциональна амплитуде S модулирующего сигнала и не зависит от частоты модуляции Ω . Тогда:

$$\varphi(t) = m \sin \Omega t + \varphi, \quad (2.23)$$

где

$$m = \omega_d / \Omega. \quad (2.24)$$

То есть индекс угловой модуляции m ЧМК пропорционален амплитуде модулирующего сигнала и обратно пропорционален частоте модуляции.

Таким образом:

$$u_{\text{ЧМ}} = U_0 \cos(\omega_0 t + m \sin \Omega t + \varphi). \quad (2.25)$$

ФМ колебаие

Согласно (2.17)

$$\varphi(t) = \varphi + m \cos \Omega t. \quad (2.26)$$

Здесь $m = k_{\text{ФМ}} S$, т.е. индекс угловой модуляции ФМ колебания пропорционален амплитуде S модулирующего сигнала и не зависит от частоты модуляции ω . Тогда:

$$\Phi(t) = \omega_0 t + \varphi(t) = \omega_0 t + m \cos \Omega t + \varphi, \quad (2.27)$$

$$\omega(t) = \frac{d}{dt} [\omega_0 t + \varphi(t)] = \omega_0 - m \Omega \sin \Omega t = \omega_0 - \omega_{\text{д}} \sin \Omega t. \quad (2.28)$$

где

$$\omega_{\text{д}} = m \Omega. \quad (2.29)$$

Следовательно, девиация частоты $\omega_{\text{д}}$ ФМК пропорциональна амплитуде и частоте модулирующего сигнала.

Таким образом имеем:

$$u_{\text{ФМ}} = U_0 \cos(\omega_0 t + m \cos \Omega t + \varphi). \quad (2.30)$$

Из выражений (2.25) и (2.30) и приведенных графиков видно, что при тональной модуляции нельзя определить, является ли сигнал частотно- или фазомодулированным. Различие между этими видами модуляции проявляется только при изменении частоты мо-

дуляции Ω . При ЧМ с увеличением Ω девиация частоты ω_D остается постоянной, а индекс модуляции m в соответствии с (2.24) уменьшается по закону гиперболы (рисунок 2.3, а), в то время как при ФМ постоянным остается индекс модуляции m , а девиация частоты ω_D в соответствии с (2.29) растет по линейному закону (рисунок 2.3, б).

Если модулирующий сигнал $s(t)$ негармонический, то ЧМ и ФМ колебания различаются по характеру изменения параметров $\omega(t)$ и $\varphi(t)$. На рисунке 2.4 приведены графики мгновенной частоты $\omega(t)$ и фазового сдвига $\varphi(t)$ ЧМ и ФМ колебаний для случая, когда модулирующий сигнал $s(t)$ имеет вид пилообразной функции. Как следует из рисунка 2.4, б, е, угловая частота ЧМ сигнала изменяется монотонно в соответствии с модулирующим сигналом $s(t)$, тогда как частота $\omega(t)$ ФМ сигнала изменяется скачками.

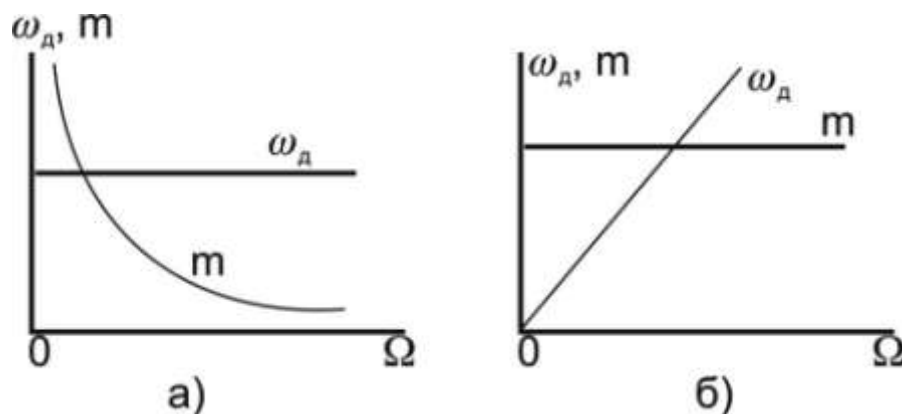


Рисунок 2.3 – График изменения модуляции по закон гиперболы (а) и по линейному закону (б)

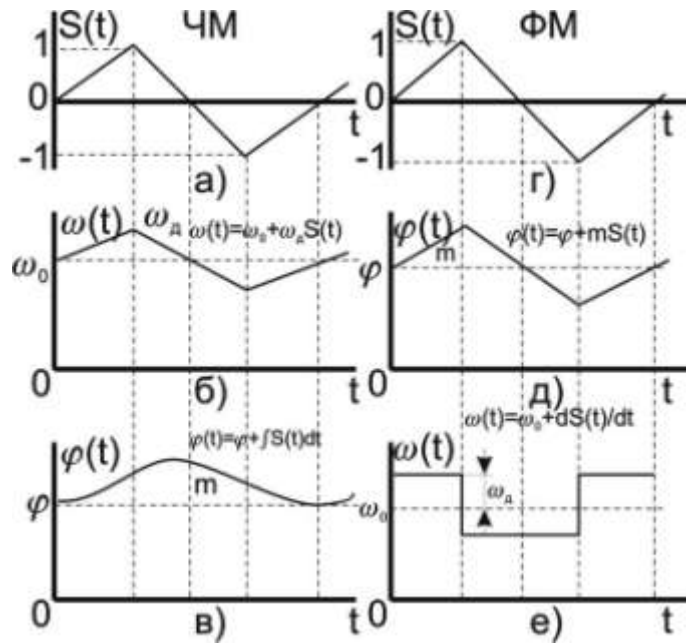


Рисунок 2.4 – Графики мгновенной частоты $\omega(t)$ и фазового сдвига $\varphi(t)$ ЧМ и ФМ колебаний

Спектр колебания при гармонической угловой модуляции. Как уже указывалось, модулированное колебание, представленное в виде:

$$u = U_0 \cos[\omega_0 t + m \sin(\Omega t + \Psi) + \varphi] \quad (2.31)$$

в равной степени соответствует ЧМ и ФМ колебаниям.

Спектр колебания, изображенный на рисунке 2.5, состоит из бес-конечного числа боковых составляющих, расположенных по-парно симметрично относительно несущей частоты ω_0 и имеющих частоты $(\omega_0 + n\Omega)$.

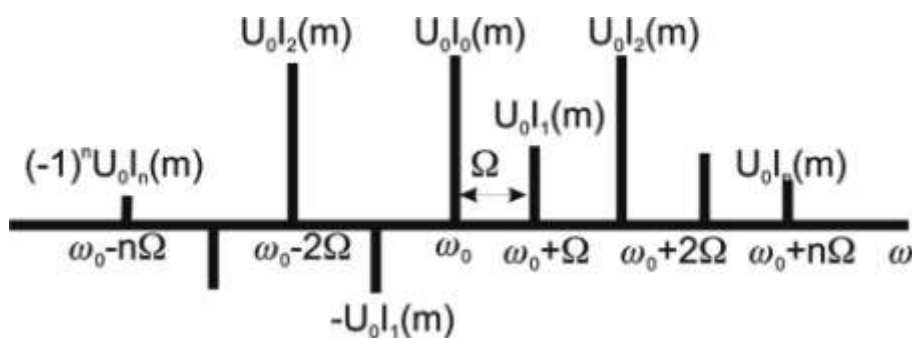


Рисунок 2.5 – Спектр колебания

Теоретически колебание с угловой модуляцией занимает бесконечную полосу частот. Однако для заданного индекса модуляции m практически можно не учитывать боковые составляющие порядка $n = m + 2$ и выше из-за малости их амплитуд.

Ширина спектра $2\Delta\omega$ колебания, ограниченного $(m + 1)$ парой боковых составляющих, выражается приближенным равенством:

$$2\Delta\omega \approx 2\Omega(m + 1) \quad (2.32)$$

а при $m \gg 1$ примерно равна удвоенному значению девиации частоты ω_d .

$$2\Delta\omega \approx 2m\Omega = 2\omega_d . \quad (2.33)$$

При изменении амплитуды S модулирующего сигнала ($\Omega = \text{const}$) в ЧМ и ФМ колебаниях изменяется индекс модуляции m , и, следовательно, изменяются число боковых составляющих, ширина спектра $2\Delta\omega = 2\Omega(m + 1)$ и амплитуды $U_0 J_n(m)$ составляющих спектра.

При изменении частоты модуляции $\Omega(S = \text{const})$ в случае ЧМ изменяется согласно (2.37) индекс модуляции m , следовательно, число и амплитуды составляющих спектра. Ширина спектра $2\Delta\omega_{\text{ЧМ}}$, практически не меняется. При ФМ амплитуды составляющих и их число остаются без изменения (так как $m = \text{const}$), а ширина спектра $2\Delta\omega_{\text{ФМ}}$ изменяется пропорционально частоте модуляции.

3.2. Объект исследования

Объектом исследования является спектр модулированного сигнала. В этих целях предлагается использовать модулированный сигнал, полученный с помощью программы MODEL. Программа MODEL позволяет синтезировать сигнал, состоящий из десяти гармоник, причем каждая гармоника может модулироваться отдельно моногармоническим сигналом. Предусмотрена возмож-

ность как амплитудной, так и фазовой модуляции. Окно программы MODEL показано на рисунке 2.6.

N	Гармоники несущей			Модуляция		
	АМПЛИТУДА	ЧАСТОТА	НАЧ. ФАЗА	ГЛУБИНАмод	ЧАСТмод-ая	НАЧ. ФАЗАмод
1	100.000000	1.00000000	0.00000000	100.000000	0.01000000	0.00000000
2	0.00000000	2.00000000	0.00000000	0.00000000	0.01000000	0.00000000
3	0.00000000	3.00000000	0.00000000	0.00000000	0.01000000	0.00000000
4	0.00000000	4.00000000	0.00000000	0.00000000	0.01000000	0.00000000
5	0.00000000	5.00000000	0.00000000	0.00000000	0.01000000	0.00000000
6	0.00000000	6.00000000	0.00000000	0.00000000	0.01000000	0.00000000
7	0.00000000	7.00000000	0.00000000	0.00000000	0.01000000	0.00000000
8	0.00000000	8.00000000	0.00000000	0.00000000	0.01000000	0.00000000
9	0.00000000	9.00000000	0.00000000	0.00000000	0.01000000	0.00000000
10	0.00000000	10.00000000	0.00000000	0.00000000	0.01000000	0.00000000

AM

F1 – ПОМОЩЬ F2 – ФОРМИРОВАНИЕ ФАЙЛА F3 – КОНФИГУРАЦИЯ ESC – ВЫХОД

Рисунок 2.6 – Окно программы MODEL

С помощью клавиш перемещения курсора можно выбрать любую комбинацию из десяти несущих частот (исключение несущей частоты из континуума осуществляется путем задания нулевой амплитуды). Используя данные лабораторной работы №1 нужно выбрать такие частоты, которые вошли в спектр электрокардиосигнала. В работе используется только амплитудная модуляция. Модулирующие частоты выбираются на основе анализа медленных волн, модулирующих электрокардиосигнал.

После заполнения таблицы, показанной на рисунке 2.6, нажимается клавиша F2, и появляется окно, показанное на рисунке 2.7.

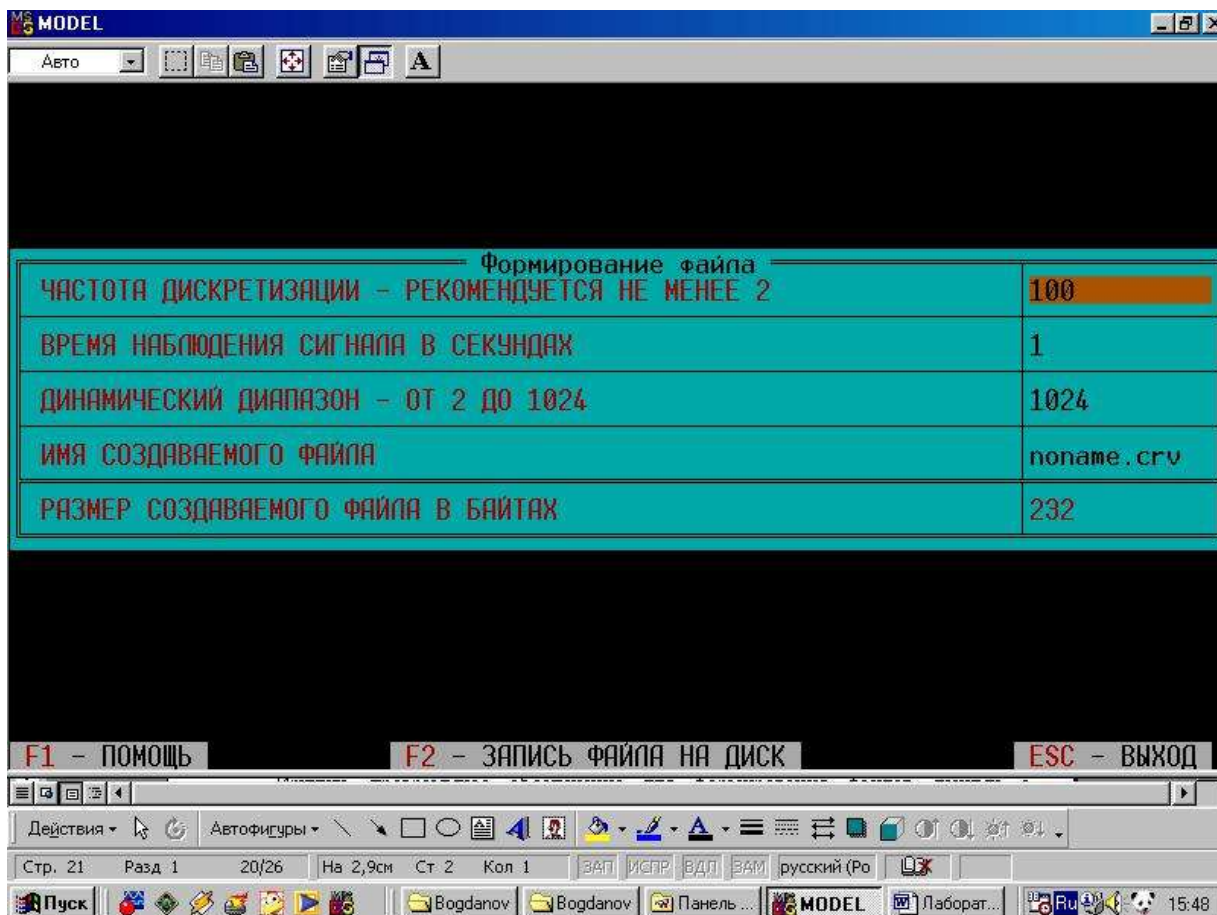


Рисунок 2.7 – Окно формирования файла данных программы MODEL

Частота дискретизации сигнала выбирается на основе соображений, изложенных в материалах лабораторной работы №2. Время наблюдения сигнала и его динамический диапазон выбирается близкими к параметрам сигнала, файл которого был получен при выполнении лабораторной работы №1. При нажатии клавиши F2 формируется файл, структура которого описана в материалах лаборатории №1.

2.3. Цель работы

Изучить методы модуляции сигналов и программное обеспечение для формирования файлов данных с цифровыми отсчетами модулированных сигналов.

2.4. Содержание работы и порядок ее выполнения

Изучите раздел «Модуляция сигналов» [1].

Изучите программу моделирования данных MODEL, обращая особое внимание на структуру файла данных и его заголовок.

Запишите математические выражения амплитудно-модулированного и частотно-модулированного сигналов. Зарисуйте их спектры.

Изучите спектр электрокардиосигнала, полученный при выполнении лабораторной работы №1. Выберите 5...10 значимых гармоник и задайте их значения и величину, согласно полученным спектральным характеристикам реального электрокардиосигнала.

Выберите модулирующие частоты и глубину модуляции для каждой несущей гармоники исходя из свойств реальных сигналов, которые модулируют электрокардиосигнал.

Исходя из пункта 4, задайтесь частотой дискретизации синтезированного сигнала и с помощью программы MODEL создайте файл данных с этим сигналом.

Сосчитайте заголовок файла данных и сами данные с помощью программы, имеющейся в базе данных или разработанной самостоятельно при выполнении лабораторной работы №1 и №2. Сделайте необходимые выводы.

2.5. Содержание отчета

1. Спектр исходного немодулированного сигнала.
2. Представление рядом Фурье модели немодулированного электрокардиосигнала.
3. Спектр медленных волн, модулирующих гармоники электрокардиосигнала
4. Распечатка графика модели электрокардиосигнала, полученного в результате выполнения пункта 2.
5. Распечатка графика модели электрокардиосигнала, полученного в результате выполнения пункта 2 и пункта 3.
6. Электронная форма на компакт-диске с файлом .CRV.
7. Выводы.

2.6. Вопросы для самопроверки

1. Что понимается под несущим и модулирующим колебаниями? В чем заключается процесс модуляции?
2. Как записать АМ колебание при модуляции:
 - а) гармоническим колебанием;
 - б) произвольным периодическим колебанием;
 - в) непериодическим сигналом?
3. Как определяется коэффициент модуляции АМ колебаний?
4. Как формируется спектр АМ колебания при модуляции:
 - а) гармоническим колебанием;
 - б) произвольным периодическим колебанием;
 - в) непериодическим сигналом?
5. Дайте графическое представление спектра.
6. Запишите выражение для колебания с угловой модуляцией. Какими соотношениями связаны полная фаза и мгновенная частота колебания?
7. Как определяются и чем отличаются ЧМ и ФМ колебания?
8. Какой физический смысл имеют понятия «девиация частоты» ω_d и «индекс модуляции» m ? Как они определяются при частотной и фазовой модуляции гармоническим сигналом?
9. По каким приближенным формулам можно определить ширину спектра ЧМ и ФМ колебаний при гармонической модуляции случаях $m \ll 1$ и $m > 1$?
10. От каких параметров модулирующего гармонического сигнала и как зависят спектры ЧМ, ФМ и АМ колебаний?
11. Чем отличаются спектральные и диаграммы АМ и ЧМ колебаний при $m \ll 1$?
12. Дайте определение АЧМ сигнала.
13. Какой вид имеют амплитудный и фазовый спектры АЧМ сигнала при большой базе m сигнала?
14. Запишите в общем виде выражение для узкополосного сигнала. Как устраняется неоднозначность в определении огибающей, фазы и мгновенной частоты узкополосного сигнала?
15. Как представить в комплексной форме произвольный узкополосный сигнал? Что понимается под комплексной огибающей сигнала?
16. Как определить огибающую и фазу узкополосного сигнала, если известна его спектральная плотность?

17. Как осуществляется дискретизация по времени АМ и ЧМ колебания?

2.7. Порядок защиты работы

Работа может быть зачтена, если студент представил отчет согласно пункту 5, исследуемые в работе сигналы соответствуют индивидуальному варианту, электронная форма соответствует представленному тексту, и студент дал исчерпывающие ответы на 10 произвольных вопросов из пункта 6.

Список литературы

1. Баскаков, С. И. Радиотехнические цепи и сигналы. Руководство к решению задач : учебное пособие / С. И. Баскаков. - 2-е изд., перераб. и доп. - М. : Высшая школа, 2002. - 214 с.

ПРАКТИЧЕСКАЯ РАБОТА №4 ДИСКРИМИНАНТНЫЙ АНАЛИЗ В СИСТЕМЕ STATISTICA

Цель работы: освоить дискриминантный анализ данных в системе STATISTICA

Постановка задачи: Для предметной области и соответствующего признакового описания объектов выполнить дискриминантный анализ.

Задание 1. Выполнить дискриминантный анализ цветов ириса.

1.1. Проверить, что измеряемые характеристики объектов (цветов ириса) имеют нормальное распределение, т.е. провести специальное тестирование на нормальность, имеющееся в модуле ANOVA/MANOVA

1.2. Прodelать 6 шагов дискриминантного анализа, описанных в п.2.4.

Задание 2. Выполнить дискриминантный анализ для исходных данных в соответствии с вариантом

2.1. Подготовить исходные данные. В качестве исходных данных взять исходную информацию о состоянии крови и группе антропометрических данных в файлах ИД-оперир и ИД-здоров в соответствии с вариантом

Таблица 3.1 – Подготовка исходных данных

Вариант	Группирующая переменная – Grouping variable	Независимые переменные – Independent variables
1	2	3
1	Столбец Y (Объем удаления щит железы)	Столбцы C-J
2	Столбец Y (Объем удаления щит железы)	Столбцы C, H-P
3	Столбец Y (Объем удаления щит железы)	Столбцы I-P

Продолжение таблицы 3.1.

1	2	3
4	Столбец Y (Объем удаления щит железы)	Столбцы С-J
5	Столбец Y (Объем удаления щит железы)	Столбцы Q-U
6	Столбец Y (Объем удаления щит железы)	Столбцы I-T
7	Столбец Y (Объем удаления щит железы)	Столбцы С-Н
8	Столбец Y (Объем удаления щит железы)	Столбцы С,Н- J, Q
9	Столбец Y (Объем удаления щит железы)	Столбцы С,Н, К, R
10	Столбец Y (Объем удаления щит железы)	Столбцы С,Н, L, S
11	Столбец Y (Объем удаления щит железы)	Столбцы С,Н, N-P, T
12	Столбец Y (Объем удаления щит железы)	Столбцы С,Н, I, U
13	Столбец Y (Объем удаления щит железы)	Столбцы Н- Q
14	Столбец Y (Объем удаления щит железы)	Столбцы Н-P, R
15	Столбец Y (Объем удаления щит железы)	Столбцы Н, I-P

2.2. Проверить, что измеряемые характеристики заданных объектов имеют нормальное распределение, т.е. провести специальное тестирование на нормальность, имеющееся в модуле ANOVA/MANOVA

2.3. Прodelать 7 шагов дискриминантного анализа, описанных в пункте 2.4.

Задание 3. Оформить отчет.

3.1. Краткие теоретические сведения

3.1.1. Введение в дискриминантный анализ

Дискриминантный анализ является одним из методов многомерного статистического анализа. Цель дискриминантного анализа состоит в том, чтобы на основе измерения различных характеристик (признаков, параметров) объекта классифицировать его, то есть отнести к одной из нескольких групп (классов) некоторым оптимальным способом. Под оптимальным способом понимается либо минимум математического ожидания потерь, либо минимум вероятности ложной классификации. Этот вид анализа является многомерным, так как измеряется несколько параметров объекта, по крайней мере, больше одного, например, температура, влажность в технологическом процессе, давление, состав крови, температура больного и т.д.

Типичные области применения дискриминантного анализа — медицина, управление производством, экономика, геология, контроль качества. В медицине объектом исследования является пациент, когда по результатам измерений различных параметров, проведения диагностических тестов врач определяет, например, необходимо ли хирургическое вмешательство при лечении. В управлении производством принимается решение по отнесению поступающего сырья или продукции к одному из нескольких типов. В экономике важно решение по отнесению клиента к определенному классу при выдаче кредита. Чрезвычайно интересно применение дискриминантного анализа в борьбе с терроризмом. Последствия

действий против террористов могут привести к одному из двух результатов: успешное освобождение заложников и случаи, когда заложники пострадали. Здесь имеются следующие дискриминантные переменные: число террористов, степень поддержки, количество оружия и так далее. Очевидно, лицам, ведущим переговоры, важно классифицировать возникшую ситуацию, с тем чтобы определить свои действия. Широкий круг задач, возникающих на практике и связанных с классификацией, можно решить методами дискриминантного анализа.

В модуле Discriminant analysis (Дискриминантный анализ) системы STATISTICA имеется широкий набор средств, обеспечивающих проведение дискриминантного анализа данных, визуализации и интерпретации результатов.

3.1.2. Постановка задачи дискриминантного анализа

Предположим, имеется n объектов с m характеристиками. В результате измерений каждый объект характеризуется вектором x_1, \dots, x_m , $m > 1$. Задача состоит в том, чтобы по результатам измерений отнести объект к одной из нескольких групп (классов) G_1, \dots, G_k , $k \geq 2$. Иными словами, нужно построить решающее правило, позволяющее по результатам измерений параметров объекта указать группу, к которой он принадлежит. Число групп заранее известно, также известно, что объект заведомо принадлежит к определенной группе.

Пусть X — пространство значений вектора измерений. Решающее правило называется нерандомизированным, если пространство X разбито на k непересекающихся областей; при попадании измерения параметров объекта в k -ю область объект относится к k -й группе. Решающее правило называется рандомизированным, если для каждого вектора наблюдений x задана вероятность $p_i(x)$, с которой объект принадлежит i -й группе, $p_i(x) \geq 0$, $p_1(x) + \dots + p_k(x) = 1$, $i=1, \dots, k$.

Очевидно, при использовании решающего правила возникают потери, вызванные тем, что объект неправильно классифицирован — отнесен к классу i , когда в действительности он принадлежит классу j (i не равно j). Если можно измерить убыток $r(i, j)$ при не-

правильной классификации объекта, то вводят средние потери, к которым приводит применение данного правила, и пытаются найти правило, минимизирующее эти средние потери. Если значение потерь трудно оценить численно, то при построении оптимального правила используют критерий минимальной вероятности ложной классификации.

В дискриминантном анализе можно задать априорные вероятности принадлежности объекта к определенному классу. На практике эти вероятности оцениваются из массива экспериментальных данных.

Так как массив экспериментальных данных накапливается, то эти оценки постепенно уточняются. При этом можно учесть различные факторы, влияющие на принадлежность объекта к определенному классу, например, если поступает мука в хлебное производство, то можно учесть сезонные факторы: вероятность того, что мука будет лучшего качества осенью выше той же вероятности весной.

В случае двух групп объектов дискриминантный анализ эквивалентен множественной регрессии (зависимой переменной является номер группы). Независимые переменные с наибольшими стандартизированными коэффициентами регрессии дают наибольший вклад в предсказание принадлежности объекта к группе. В модуле *Discriminant analysis* (Дискриминантный анализ) реализовано два общих метода дискриминантного анализа: стандартный и пошаговый (включения и исключения). Данные методы дискриминантного анализа аналогичны методам множественной регрессии. В случае двух групп методом наименьших квадратов строится регрессионная прямая (зависимая переменная — номер группы, все остальные переменные — независимые). Если групп несколько, то можно представить себе, что вначале строится дискриминация между группами 1 и 2, затем между 2 и 3, и так далее.

В пошаговом методе модель строится последовательно по шагам. Для метода включения STATISTICA на каждом шаге оценивает вклад в функцию дискриминации не включенных в модель переменных. Переменная, дающая наибольший вклад, включается в модель, далее система переходит к следующему шагу. Если применяется так называемый пошаговый метод исключения, то вначале в

модель включаются все переменные, затем производится их последовательное исключение.

Близкими к методам дискриминантного анализа являются методы дисперсионного анализа, кластерного и факторного анализов и методы множественной регрессии. Отличие кластерного анализа от дискриминантного в том, что в нем заранее не фиксировано число групп (кластеров).

3.1.3. Предположения и ограничения в дискриминантном анализе

Дискриминантный анализ «работает» при выполнении ряда предположений.

Предположение о том, что наблюдаемые величины — измеряемые характеристики объекта — имеют нормальное распределение. Это предположение следует проверять. В модуле имеются специальные опции, позволяющие быстро построить гистограммы и графики на вероятностной бумаге. Специальные тесты на нормальность имеются в модуле ANOVA/ MANOVA. Следует заметить, что умеренные отклонения от этого предположения не являются фатальными.

Предположение об однородности дисперсий и ковариаций наблюдаемых переменных в разных классах (отличие между классами имеется только в средних). Умеренные отклонения от этого предположения также допустимы.

Широкий набор статистик и опций для тестирования различных предположений дискриминантного анализа, в частности так называемый М-критерия Бокса, содержится также в модуле ANOVA/MANOVA — Дисперсионный анализ.

Методы, реализованные в модуле, являются линейными. Функции классификации и дискриминантные функции являются линейными комбинациями наблюдаемых величин.

Важное замечание о проверке предположений анализа. Дискриминантный анализ может быть проведен и когда основные предположения не выполняются (предположение о нормальности и равенстве ковариационных матриц). Задача состоит в интерпретации результатов. В конечном счете, наиболее важным критерием

правильности построенного классификатора является практика. И если окажется, что в результате построен классификатор, «работающий» на практике, то это будет достижением.

3.1.4. Дискриминантный анализ цветов ириса в STATISTICA

Знакомство с возможностями проведения дискриминантного анализа в системе STATISTICA лучше всего начать с разбора апробированного примера. Таким примером является классический пример Фишера — анализа цветков ириса. Задача состоит в том, чтобы по результатам измерения длины и ширины чашелистиков и лепестков цветков ириса отнести ирис к одному из трех типов: SETOSA, VERSICOL, VIRGINIC. Данные для этого примера имеются в файле Irisdat.sta (C:\ ProgramFile\Statistica6\ Examples\Datasets\Irisdat.sta). В файле содержатся результаты измерений 150 цветков ириса, по 50 каждого типа.

Шаг 1. Из Переключателя модулей STATISTICA открываем стартовую панель модуля Discriminant function analysis (Дискриминантный функциональный анализ).

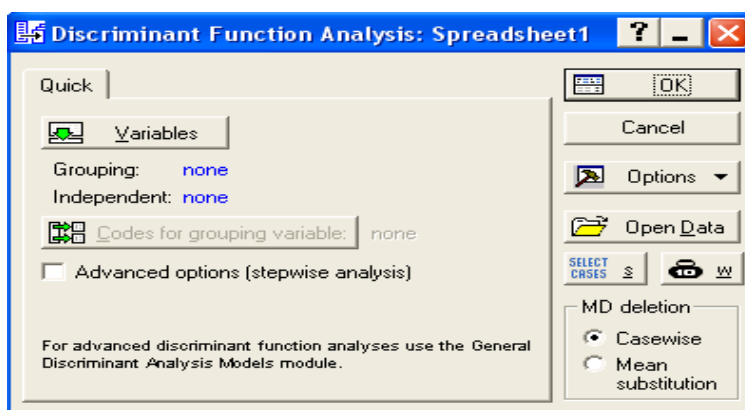


Рисунок 3.1 – Стартовая панель модуля Дискриминантный анализ

Шаг 2. Нажимаем кнопку Open Data (Открыть данные) и открываем файл данных Irisdat.sia из каталога *Examples*. Следующий файл данных появится на экране:

	1	2	3	4	5
	SEPALLEN	SEPALWID	PETALLEN	PETALWID	IRISTYPE
1	5	3,3	1,4	0,2	SETOSA
2	6,4	2,8	5,6	2,2	VIRGINIC
3	6,5	2,8	4,6	1,5	VERSICOL
4	6,7	3,1	5,6	2,4	VIRGINIC
5	6,3	2,8	5,1	1,5	VIRGINIC
6	4,6	3,4	1,4	0,3	SETOSA
7	6,9	3,1	5,1	2,3	VIRGINIC
8	6,2	2,2	4,5	1,5	VERSICOL
9	5,9	3,2	4,8	1,8	VERSICOL
10	4,6	3,6	1	0,2	SETOSA
11	6,1	3	4,6	1,4	VERSICOL
12	6	2,7	5,1	1,6	VERSICOL
13	6,5	3	5,2	2	VIRGINIC
14	5,6	2,5	3,9	1,1	VERSICOL
15	6,5	3	5,5	1,8	VIRGINIC
16	5,8	2,7	5,1	1,9	VIRGINIC

Рисунок 3.2 – Файл данных Iris.sta

Шаг 3. Нажимаем кнопку Variables (Переменные) и выбираем переменные, для анализа.

В качестве Группирующей переменной – Grouping variable выбираем переменную IRISTYPE – ТИПИРИСА.

В качестве Независимых переменных – Independent variables выбираем переменные SEPALLEN - ДЛИНА ЧАШЕЛИСТИКА, SEPALWD - ШИРИНАЧАШЕЛИСТИКА, PETALLEN -ДЛИНА ПЕСТИКА, PETALWD - ШИРИНА ПЕСТИКА.

Шаг 4. Нажимаем кнопку ОК и открываем диалоговое окно Model Definition (Определение модели).

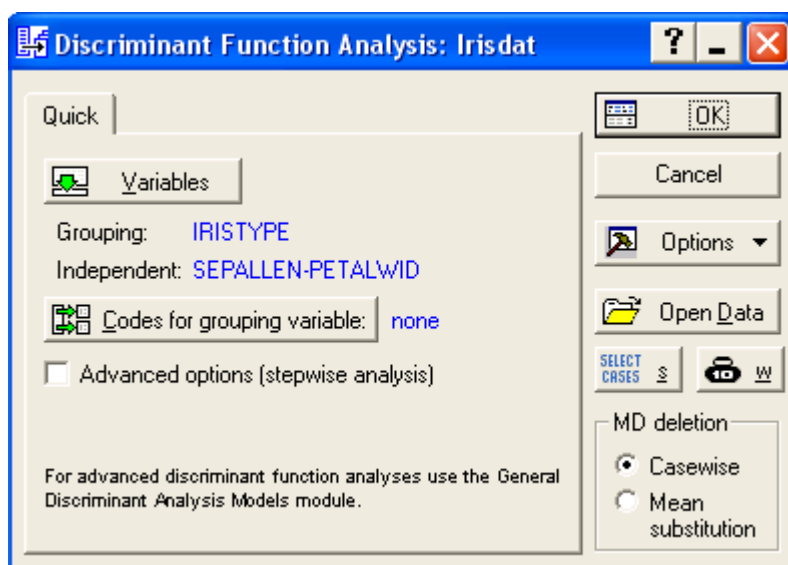


Рисунок 3.3 – Окно определения модели дискриминантного анализа

Делаем установки, как показано на рисунке 3.3. Нажимаем кнопку ОК и запускаем вычислительную процедуру, реализующую пошаговый метод включения.

Шаг 2. Рассмотрим итоги, приведенные в диалоговом окне Результаты дискриминантного анализа.

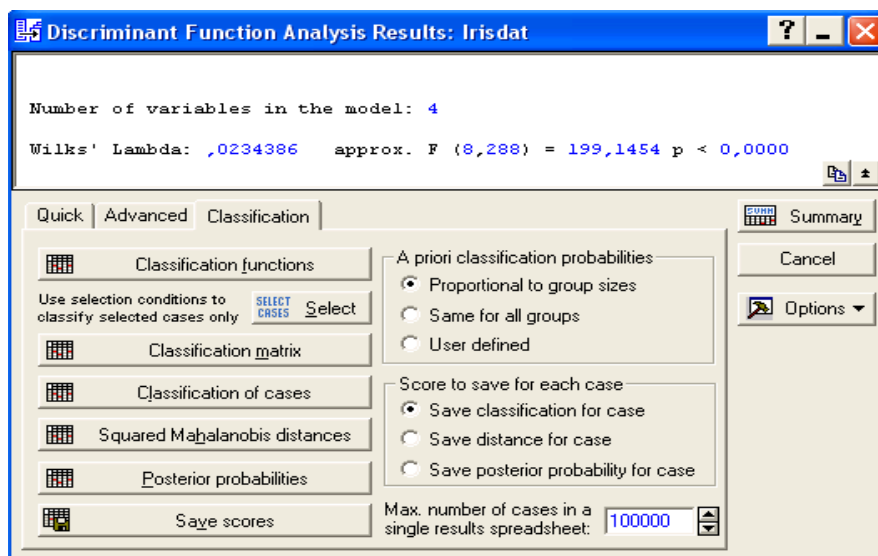


Рисунок 3.4 – Окно результатов дискриминантного анализа данных из файла Iris.sta

Информационная часть окна сообщает, что использован:
Number of variables in the model — Число переменных в модели: 4;
Wilks lambda — Значение лямбды Уилкса: 0.0234386;
арргох. $F(8,288) = 199.1454$ — Приближенное значение F-статистики, связанной с лямбдой Уилкса;
p — уровень значимости F-критерия для значения 199.1454.

Значения статистики лямбда Уилкса лежат в интервале [0,1]. Значения статистики Уилкса, лежащие около 0, свидетельствуют о хорошей дискриминации. Значения статистики Уилкса, лежащие около 1, свидетельствуют о плохой дискриминации. Иными словами, это можно выразить следующим образом: если значения лямбда Уилкса близки к 0, то мощность дискриминации (мощность = 1 — вероятность ошибки) близка к 1, если лямбда Уилкса близка к 1, то мощность близка к 0.

Нажимаем кнопку Variables in the model (Переменные, включенные в модель).

Discriminant Function Analysis Summary (Irisdat)						
No. of vars in model: 4; Grouping: IRISTYPE (3 grps)						
Wilks' Lambda: ,02344 approx. F (8,288)=199,15 p<0,0000						
N=150	Wilks' Lambda	Partial Lambda	F-remove (2,144)	p-level	Toler.	1-Toler. (R-Sqr.)
SEPALLEN	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007
SEPALWID	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALLEN	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
PETALWID	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686

Рисунок 3.5 - Итоговая таблица анализа данных из файла Iris.sta

Посмотрим разделение групп на графике. Для этого инициируем кнопку Canonical analysis & graphs (Канонический анализ и графики). В появившемся диалоговом окне Canonical Analysis (Канонический анализ) нажимаем кнопку Scatterplot of canonical scores (Диаграмма рассеяния канонических значений). На экране появится следующий график:

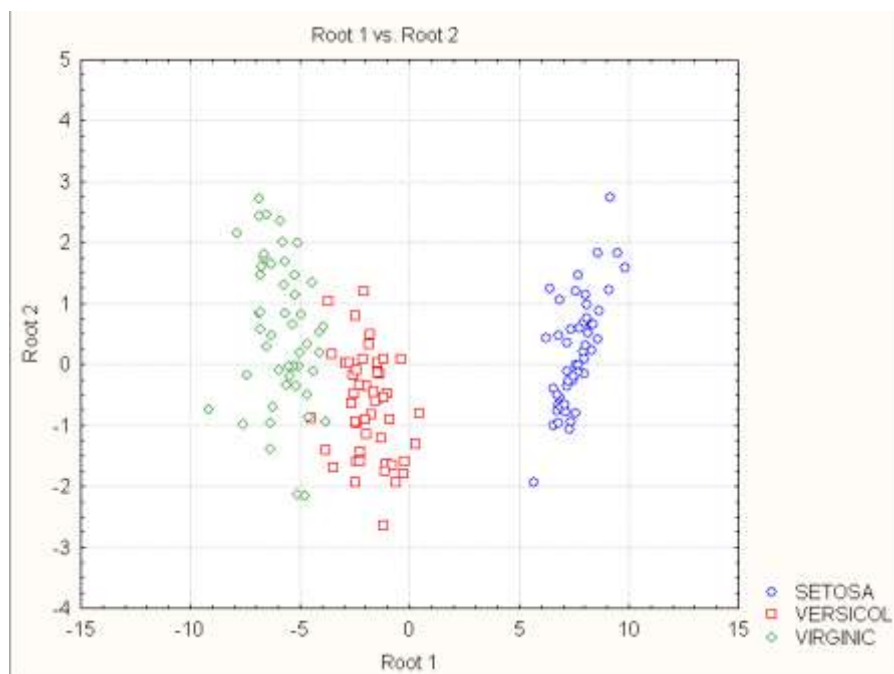


Рисунок 3.6 – Разделение трех типов ириса

На рисунке 3.6 приведена визуализация данных по 150 измерениям длины и ширины чашелистиков и лепестков цветков ириса трех типов (SETOSA, VERSICOL, VIRGINIC) на плоскости (в сис-

теме координат по двум построенным факторам Root1 и Root2 с помощью факторного анализа).

Просмотрим функции классификации. В диалоговом окне Результаты (рисунок 3.4) дискриминантного анализа нажимаем кнопку Classification functions (Функции классификации).

Classification Functions; grouping: IRISTYPE (Irisdat)					
Variable	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333		
SEPALLEN	23,5442	15,6982	12,446		
SEPALWID	23,5879	7,0725	3,685		
PETALLEN	-16,4306	5,2115	12,767		
PETALWID	-17,3984	6,4342	21,079		
Constant	-86,3085	-72,8526	-104,368		

Рисунок 3.7 – Функции классификации, построенные пошаговым методом вперед

С помощью этих функций можно вычислить классификационные значения (метки) для вновь наблюдаемых цветков по формулам:

$$\text{SETOSA} = -16.43 \cdot \text{ДЛЛЕПЕСТ} + 23.59 \cdot \text{ШИРЧАШЕЛ} - 17.4 \cdot \text{ШИРЛЕПЕС} + 23.54 \cdot \text{ДЛЧАШЕЛИ} - 86.31$$

$$\text{VERSICOL} = 2.21 \cdot \text{ДЛЛЕПЕСТ} + 7.07 \cdot \text{ШИРЧАШЕЛ} - 6.43 \cdot \text{ШИРЛЕПЕС} + 12.70 \cdot \text{ДЛЧАШЕЛИ} - 72.85$$

$$\text{VIRGINIC} = 12.76 \cdot \text{ДЛЛЕПЕСТ} + 3.69 \cdot \text{ШИРЧАШЕЛ} - 21.08 \cdot \text{ШИРЛЕПЕС} + 12.5 \cdot \text{ДЛЧАШЕЛИ} - 104.37$$

Пусть мы имеем новый цветок со значениями: ДЛЛЕПЕСТ*, ШИРЧАШЕЛ*, ШИРЛЕПЕС*, ДЛЧАШЕЛИ*. Формально следует подставить эти значения в приведенные выше формулы и вычислить классификационные значения SETOSA*, VERSICOL*, VIRGINIC*. Новый цветок относится к тому классу, для которого классификационное значение максимально.

Конечно, построенные классификационные функции могут быть определены в электронных таблицах как формулы, и для каждого добавленного случая по ним могут быть вычислены класси-

фикационные метки. Таким образом, каждый новый объект автоматически относится к определенному классу.

Расстояния Махаланобиса. Нажимаем кнопку Squared Mahalanobis distance (Квадрат расстояния Махаланобиса) (рисунок 3.4) и видим таблицу с квадратами расстояния Махаланобиса от точек (случаев) до центров групп:

Squared Mahalanobis Distances from Group Centroids (Irisdat)						
Incorrect classifications are marked with *						
Case	Observed Classif.	SETOSA p=.33333	VERSICOL p=.33333	VIRGINIC p=.33333		
1	SETOSA	0,2419	90,6602	181,5587		
2	VIRGINIC	208,5713	27,3188	1,8944		
3	VERSICOL	105,2663	2,2329	13,0720		
4	VIRGINIC	207,9180	31,7492	4,4506		
* 5	VIRGINIC	133,0668	5,2529	7,2359		
6	SETOSA	1,3337	84,0118	170,0569		
7	VIRGINIC	173,1838	26,5620	11,0484		
8	VERSICOL	131,6617	8,4307	14,7647		
* 9	VERSICOL	130,8624	8,6697	6,5068		
10	SETOSA	2,2864	113,6509	210,0239		
11	VERSICOL	99,2338	1,2963	13,8174		
* 12	VERSICOL	149,0303	8,4393	4,8645		

Рисунок 3.8 – Расстояния Махаланобиса для данных из файла iris.sta

Апостериорные вероятности. Рассмотрим группу опций внизу диалогового окна Результаты дискриминантного анализа: A priori classifications probabilities (Априорные вероятности классификации). До анализа мы задаем для каждого случая (в данном примере цветка) вероятность, с какой он принадлежит к определенному классу. После того как анализ выполнен, можно пересчитать эти вероятности и получить апостериорные вероятности классификации. Нажав кнопку Posterior probabilities (Апостериорные вероятности) (рисунок 3.4), мы увидим таблицу с апостериорными вероятностями принадлежности объекта к определенному классу.

Posterior Probabilities (Irisdat)				
Incorrect classifications are marked with *				
Case	Observed Classif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
1	SETOSA	1,000000	0,000000	0,000000
2	VIRGINIC	0,000000	0,000003	0,999997
3	VERSICOL	0,000000	0,995590	0,004410
4	VIRGINIC	0,000000	0,000001	0,999999
* 5	VIRGINIC	0,000000	0,729388	0,270612
6	SETOSA	1,000000	0,000000	0,000000
7	VIRGINIC	0,000000	0,000428	0,999572
8	VERSICOL	0,000000	0,959573	0,040427
* 9	VERSICOL	0,000000	0,253228	0,746772
10	SETOSA	1,000000	0,000000	0,000000
11	VERSICOL	0,000000	0,998093	0,001907
* 12	VERSICOL	0,000000	0,143392	0,856608
13	VIRGINIC	0,000000	0,003146	0,996854
14	VERSICOL	0,000000	0,999997	0,000003
15	VIRGINIC	0,000000	0,006084	0,993916
16	VIRGINIC	0,000000	0,001078	0,998922
17	VIRGINIC	0,000000	0,000001	0,999999

Рисунок 3.9 – Таблица апостериорных вероятностей

В первом столбце таблицы, приведенной на рисунке 3.9, указан тип ириса для каждого случая. Во втором, третьем, четвертом столбце даны апостериорные вероятности отнесения каждого цветка к соответствующему типу. Цветок относится к группе с максимальной апостериорной вероятностью. Знаком * отмечаются неправильно классифицированные при использовании данного правила случаи.

Шаг 6. Классификация новых случаев. Не закрывая диалога Результаты дискриминантного анализа (рисунок 3.4), добавим в таблицу исходных данных новый случай, например, так:

151	5,3	3,1	2,7	0,5
-----	-----	-----	-----	-----

Рисунок 3.10 – Новое наблюдение в данных iris.sta

Для того чтобы понять, к какому классу относится этот объект, нажмите кнопку Posterior probabilities (Апостериорные вероятности) (рисунок 3.4), мы увидим ту же таблицу с апостериорными вероятностями, к которой будет добавлена строка:

151	---	0,999874	0,000126	0,000000
-----	-----	----------	----------	----------

Рисунок 3.11 – Классификация нового наблюдения

Итак, новое наблюдение с вероятностью 0.999874 можно отнести к типу SETOSA.

Шаг 7. Оценка точности полученных решающих правил методом скользящего экзамена.

По аналогии, как в шаге 6 на рисунке 3.10 группа значений удаляется при построении решающих правил и по ним принимают решение о принадлежности к классу. Но по этим объектам есть информация о принадлежности к классам. Находится средневзвешенная доля неверно отнесенных к классу объектов. В качестве весов берутся априорные вероятности классов.

3.2 Контрольные вопросы

1. В чем заключается дискриминантного анализа?
2. Какова постановка задачи дискриминантного анализа?
3. Что такое априорные и апостериорные вероятности?
4. Какие существуют предположения и ограничения в дискриминантном анализе?

3.3 Список литературы

1. Боровиков В.П. Популярное введение в программу STATISTICA . -М.: Финансы и статистика. 1999. -239с.

ПРАКТИЧЕСКАЯ РАБОТА №4 ДЕЦИМАЦИЯ И ИНТЕРПОЛЯЦИЯ СИГНАЛОВ

4.1. Краткие теоретические сведения

Наиболее удобной в обращении на практике функцией является алгебраический многочлен. Чтобы задать многочлен, нужно задать только конечное число его коэффициентов. Значения многочлена легко вычисляется, его легко про дифференцировать, проинтегрировать и т. д. Поэтому алгебраические многочлены нашли широкое применение для приближения (аппроксимации) функций. Наряду с алгебраическими многочленами применяются также тригонометрические многочлены, которые являются более естественными для приближения периодических функций.

Интерполяция функций по формуле Лагранжа

Пусть известны значения некоторой функции f в $n + 1$ различных точках x_0, x_1, \dots, x_n , которые обозначим следующим образом:

$$f_i = f(x_i), i = 0, 1, \dots, n.$$

Например, эти значения получены из эксперимента или найдены с помощью достаточно сложных вычислений. Возникает задача приближенного восстановления функции f в произвольной точке x . Часто для решения этой задачи строится алгебраический многочлен $L_n(x)$ степени n , который в точках x_i принимает заданные значения, т. е.

$$L_n(x_i) = f_i, \quad i = 0, 1, \dots, n, \quad (4.1)$$

и называется интерполяционным. Точки $x_i; i = 0, 1, \dots, n$ называются узлами интерполяции.

Приближенное восстановление функции f по формуле:

$$f(x) = L_n(x) \quad (4.2)$$

называется интерполяцией функции f с помощью алгебраического многочлена.

Существует теорема, согласно которой имеется только один интерполяционный многочлен n -й степени, удовлетворяющий условию (1).

Интерполяционный многочлен, представленный в виде:

$$L_n(x) = \sum_{i=0}^n P_{ni}(x) f_i, \quad (4.3)$$

где

$$P_{ni}(x) = \frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} \quad (4.4)$$
$$x \in [ab]$$

называется интерполяционным многочленом Лагранжа, а функции (4.4) – Лагранжевыми коэффициентами.

Погрешность интерполяции (экстраполяции) в текущей точке оценивается по формуле:

$$|f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |w_n(x)|, \quad (4.5)$$

$$M_{n+1} = \max_{[a,b]} |f^{(n+1)}(x)| \quad (4.6)$$

где

$$w_n(x) = (x - x_0)(x - x_1) \dots (x - x_n). \quad (4.7)$$

Максимальная погрешность интерполяции на всем отрезке $[a,b]$:

$$\max_{[a,b]} |f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \max_{[a,b]} |w_n(x)|. \quad (4.8)$$

4.1.1 Линейная интерполяция

Интерполяция по формуле (4.2) при $n = 1$, т. е. с помощью линейной функции (4.3) называется линейной.

Если ввести обозначения $h = x_1 - x_0$, $q = (x - x_0)/h$, то формула линейной интерполяции может быть записана в следующем виде:

$$f(x) \approx L_1(x) = L_1(x_0 + qh) = (1 - q)f_0 + qf_1. \quad (4.9)$$

Величина q называется фазой интерполяции, которая изменяется в пределах от 0 до 1, когда x пробегает значение от x_0 до x_1 .

Геометрическая линейная интерполяция означает замену графика функции на отрезке $[x_0, x_1]$ хордой, соединяющей точки (x_0, f_0) , (x_1, f_1) , как показано на нижеприведенном рисунке.

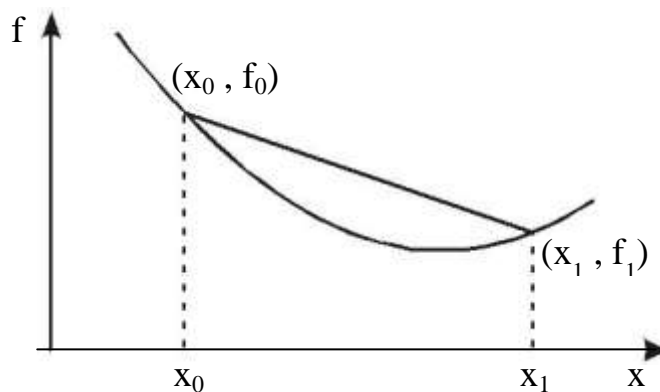


Рисунок 4.1 – Геометрическая линейная интерполяция

4.1.2 Сплайны

Пусть отрезок $[a, b]$ разбит на N равных частичных отрезков $[x_i, x_{i+1}]$, где $x_j = a + ih$, $i = 0, 1, \dots, N - 1$; $x_N = b$, $h = (b - a)/N$.

Сплайном называется функция, которая вместе с несколькими производными непрерывна на всем заданном отрезке $[a, b]$, а на каждом частичном отрезке $[x_i, x_{i+1}]$ в отдельности является некоторым алгебраическим многочленом.

Максимальная по всем частичным отрезкам степень многочленов называется степенью сплайна, а разность между степенью сплайна и порядком наивысшей непрерывной на $[a, b]$ производной – дефектом сплайна.

На практике наиболее широко распространение получили кубические сплайны – сплайны третьей степени, имеющие на $[a, b]$ непрерывную, по крайней мере, первую производную. Величина $m_i = S_3'(x_i)$ называется наклоном сплайна в точке (узле) x_i .

$$S_3(x) = \frac{(x_{i+1} - x)^2(2(x - x_i) + h)}{h^3} \cdot f_i + \frac{(x - x_i)^2(2(x_{i+1} - x) + h)}{h^3} \cdot f_{i+1} + \frac{(x_{i+1} - x)^2(x - x_i)m_i}{h^2} + \frac{(x - x_i)^2(x - x_{i+1})^2 m_{i+1}}{h^2} \quad (4.10)$$

Нетрудно убедиться, что кубический сплайн $S_3(x)$, принимающий в узлах x_i, x_{i+1} , соответственно значения f_i, f_{i+1} , имеет на частичном отрезке $[x_i, x_{i+1}]$ вид.

Действительно, легко видеть, что $S_3(x_i) = f_i, S_3(x_{i+1}) = f_{i+1}$. Кроме того, простые вычисления показывают, что $S_3'(x_i) = m_i, S_3'(x_{i+1}) = m_{i+1}$. Можно доказать, что любой алгебраический многочлен третьей степени, принимающий в точках x_i, x_{i+1} значения, равные соответственно f_i, f_{i+1} соответственно равную m_i, m_{i+1} , тождественно совпадает с многочленом (10).

Итак, чтобы задать кубический сплайн $S_3(x)$ на всем отрезке $[a, b]$, нужно задать в $N + 1$ узлах x_i его значения f_i и наклоны или касательные $m_i, i = 0, 1, \dots, N$.

Кубический сплайн, принимающий в узлах x_i те же значения, что и некоторая функция f называется интерполяционным. Он служит для аппроксимации функции f на отрезке $[a, b]$ вместе с несколькими производными.

Способы задания наклонов интерполяционного кубического сплайна.

1) Упрощенный способ.

$$m_i = \frac{f_{i+1} - f_{i-1}}{2h} \cdot 2h, \quad i = 1, 2, \dots, N-1, \quad (4.11)$$

$$m_0 = \frac{4 \cdot f_1 - f_2 - 3 \cdot f_0}{2h}, \quad m_n = \frac{3 \cdot f_n + f_{n-2} - 4 \cdot f_{n-1}}{2h}.$$

2) Если известны значения f'_i производной f' в узлах x_i , то полагаем $m_i = f'_i, i = 0, 1, \dots, N$.

Способы 1 и 2 – локальные, так как с их помощью сплайн строится отдельно на каждом частичном отрезке $[x_i, x_{i+1}]$.

3) Глобальный способ.

Обозначаем через $S''_3(x_i + 0)$ значение $S''_3(x_i)$ в узле x_i справа, найденное непосредственно из выражения (10), а через $S''_3(x_i - 0)$ значение $S''_3(x_i)$ в узле x_i слева, т. е. найденное из соответствующего выражения $S_3(x)$ на частичном отрезке $[x_i, x_{i+1}]$, которое получается из (10) заменой i на $i - 1$.

Имеем

$$S''_3(x_i + 0) = \frac{4 \cdot m_i}{h} - \frac{2 \cdot m_{i+1}}{h} + 6 \cdot \frac{f_{i+1} - f_i}{h^2},$$

$$S''_3(x_i - 0) = \frac{2 \cdot m_{i-1}}{h} - \frac{4 \cdot m_i}{h} - 6 \cdot \frac{f_i - f_{i-1}}{h^2}.$$

Требуем непрерывности $S''(x)$ в узлах:

$$S''_3(x_i - 0) = S''_3(x_i + 0), \quad i = 1, 2, \dots, N-1 \quad (4.12)$$

приходим к следующей системе линейных алгебраических уравнений относительно наклонов:

$$m_{i-1} + 4 \cdot m_i + m_{i+1} = \frac{3 \cdot (f_{i+1} - f_{i-1})}{h}, \quad i = 1, 2, \dots, N-1.$$

Поскольку неизвестных $N + 1$, то необходимо задать еще два условия, которые называются краевыми (они обычно связаны с крайними значениями m_0 , и m_N). Дадим три варианта краевых условий.

а) Если известны $f'_0 = f'(a)$, $f'_N = f'(b)$, то задать $m_0 = f'_0$, $m_N = f'_N$.

б) Производные f'_0 и f'_N аппроксимируем формулами численного дифференцирования третьего порядка точности:

$$m_0 = \frac{1}{6h} \cdot (-11 \cdot f_0 + 18 \cdot f_1 - 9 \cdot f_2 + 2 \cdot f_3); \quad (4.13)$$

$$m_n = \frac{1}{6h} \cdot (11 \cdot f_N - 18 \cdot f_{N-1} - 9 \cdot f_{N-2} - 2 \cdot f_{N-3}).$$

в) В некоторых случаях бывают известны значения f'' на концах отрезка $[a, b]$, т. е. величины $f''_0 = f''(a)$, $f''_N = f''(b)$. Тогда требование $S''_3(a) = f''_0$, $S''_3(b) = f''_N$ приводит к краевым условиям:

$$m_0 = -\frac{m_1}{2} + \frac{3}{2} \cdot \frac{f_1 - f_0}{h} - \frac{h}{4} \cdot f''_0, \quad (4.14)$$

$$m_N = -\frac{m_{N-1}}{2} + \frac{3}{2} \cdot \frac{f_N - f_{N-1}}{h} + \frac{h}{4} \cdot f''_N.$$

Система (12) при всех рассматриваемых краевых условиях имеет единственное решение. Решая систему (4.12) при выбранных краевых условиях, находим наклоны m_i , $i = 0, 1, \dots, N$, во всех узлах. Затем по формуле (4.10) задаем сплайн на каждом частичном отрезке $[x_i, x_{i+1}]$, $i = 0, 1, \dots, N-1$.

Построенный данным глобальным способом сплайн $S_3(x)$ имеет дефект не больше единицы, т. к. этот сплайн обладает на отрезке $[a, b]$ непрерывной второй производной.

$$\max_{[x_i, x_{i+1}]} |f^{(m)}(x) - S_3^{(m)}(x)| \leq ch^{k+1-m} \max_{[a, b]} |f^{(k+1)}(x)|,$$

Интерполяционный сплайн $S_3(x)$ с наклоном, заданным способом 2 или 3, удовлетворяет неравенству:

$$i = 0, 1, \dots, N-1; \quad m = 0, 1, \dots, k; \quad 0 \leq k \leq 3; \quad (4.15)$$

где c – независимая от h , i , f – постоянная.

Точность аппроксимации функции f сплайном $S_3(x)$ управляется выбором N , т. е. шагом $h = (b - a)/N$.

4.2 Аппроксимация функций по методу наименьших квадратов

Интерполяция на практике хороша лишь для таких функций, значения которых не искажены шумом. Случайные ошибки в значениях функции сильно искажают интерполяционное многочлены высоких степеней, а при интерполяции многочленами низких степеней теряется существенная информация. Поэтому, в этом случае, целесообразно применять «сглаживающую» аппроксимацию с минимизацией взвешенной средней квадратической ошибки аппроксимации. Это значит, что для данной функции $f(x)$ требуется построить функцию $F(x)$ вида

$$F(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \dots + a_n\varphi_m(x) \quad (4.16)$$

так, чтобы минимизировать взвешенную среднюю квадратическую ошибку на интервале $[a, b]$:

$$\sigma^2 = \int_a^b \gamma(x)[F(x) - f(x)]^2 dx \quad (4.17)$$

где $\gamma(x)$ – заданная весовая неотрицательная функция.

Если функции $\varphi(x)$ действительны и попарно ортогональны с весом $\gamma(x)$ на интервале $[a, b]$, то есть если

$$\int_a^b \gamma(x) \varphi_i(x) \varphi_0(x) dx = 0; \quad \text{при } i \neq j, \quad (4.18)$$

то искомые коэффициенты определяются по формуле:

$$a_i = \frac{\int_a^b \gamma(x) f(x) \varphi_i(x) dx}{\int_a^b \gamma(x) \varphi_i^2(x) dx}, \quad (i = 0, 1, 2, \dots). \quad (4.19)$$

Аппроксимация ортогональными функциями, например, ортогональными многочленами или тригонометрическими полиномами имеет то замечательное преимущество, что улучшение аппроксимации путем добавления нового члена $a_{n+1} \varphi_{n+1}(x)$ не меняет ранее вычисленные коэффициенты $a_0, a_1, a_2, \dots, a_n$.

Таким образом, для аппроксимации функции $f(x)$ необходимо задать класс приближающих функций или n -мерное пространство, где n – число заданных значений функции $f(x)$, и норму в этом пространстве. При приближении функций многочленами на дискретном множестве точек норма имеет вид:

$$\sigma^2 = \sum_{k=0}^m \gamma_k [F(x_k) - f(x_k)]^2, \quad (4.20)$$

где γ_k заданные положительные веса; $m + 1$ – дискретное множество точек.

Согласно условию ортогональности (6.18):

$$\sum_{k=0}^m \gamma_k \varphi_i(x_k) \varphi_j(x_k) = 0, \quad (i \neq j). \quad (4.21)$$

и на основании (4.19) имеем:

$$a_i = \frac{\sum_{k=0}^m \gamma_k f(x_k) \phi_i(x_k)}{\sum_{k=0}^m \gamma_k \phi_i^2(x_k)}, \quad (i = 0, 1, 2, \dots, n, \quad n \leq m). \quad (4.22)$$

Отметим, что можно использовать другую норму (4.20), тогда получим другое приближение, которое может значительно отличаться от предыдущего.

Приведем пример аппроксимации функций тригонометрическим многочленом:

$$F_m(x) = \alpha_0 + \sum_{p=1}^{m/2} (\alpha_p \cos px + \beta_p \sin px). \quad (4.23)$$

Коэффициенты этого многочлена при учете условия (4.20) находятся согласно формулам:

$$\alpha_0 = \frac{1}{n+1} \cdot \sum_{i=0}^n f \cdot \left(\frac{2\pi i}{n+1} \right), \quad \alpha_p = \frac{2}{n-1} \cdot \sum_{i=0}^n f \cdot \left(\frac{2\pi i}{n+1} \right) \cos \frac{2\pi i}{n+1} p, \quad (4.24)$$

$$\beta_p = \frac{2}{n+1} \cdot \sum_{i=0}^n f \cdot \left(\frac{2\pi i}{n+1} \right) \sin p \frac{2\pi i}{n+1}, \quad p = 1, 2, \dots, m,$$

где

$$x_i = \frac{2\pi i}{n+1}, \quad i = 0, 1, \dots, n.$$

4.3 Цель работы

Получить практические навыки децимации и интерполяции сигналов

4.4 Порядок выполнения работы

1. Запишите математическое выражение полигармонического сигнала с трендом. Нарисуйте его спектр.

2. Вычислите частоту дискретизации сигнала, полученного в пункте 1, и с помощью программы REEВOK создайте файл данных с этим сигналом. Вычислите спектр сигнала, записанного в полученном файле данных, и сравните его с теоретическим, полученным в пункте 1.

3. Составьте схему алгоритма и программный модуль для уменьшения в два раза числа отсчетов сигнала, полученного в пункте 1, не изменяя интервал наблюдения сигнала (апертуру).

4. Запустите программный модуль, полученный в пункте 3, подав на его вход сигнал, файл данных которого был получен в пункте 2. Зарисуйте сигнал на выходе программного модуля. Получите его спектр. Сравните его со спектром исходного сигнала.

5. Составьте схему алгоритма и программный модуль для увеличения в два раза числа отсчетов сигнала, полученного в пункте 1, не изменяя интервал наблюдения сигнала.

6. Запустите программный модуль, полученный в пункте 5, подав на его вход сигнал, файл данных которого был получен в пункте 2. Зарисуйте сигнал на выходе программного модуля. Получите его спектр. Сравните его со спектром исходного сигнала.

4.5 Контрольные вопросы

1. Что такое «сплайн»?
2. Перечислите способы задания наклонов интерполяционного кубического сплайна.
3. Какой график имеет геометрическая линейная интерполяция?
4. Что такое «децимация»?
5. В чем заключается отличие децимации сигнала от интерполяции?
6. Какие этапы имеет децимация цифрового сигнала с целым коэффициентом?
7. Для чего необходим первый этап децимации сигнала?
8. В чем заключается стандартный алгоритм интерполяции сигнала с целым коэффициентом?
9. Как осуществляется прореживание отсчетов сигнала?

10. К чему приводит выполнение процедуры децимации сигнала?

11. Для чего применяют «сглаживающую» аппроксимацию с минимизацией взвешенной средней квадратической ошибки аппроксимации?

12. Как осуществляется интерполяция функций по методу Лагранжа?

13. Какие сплайны на практике наиболее широко распространены?

14. Что Вы понимаете под термином «степень сплайна»?

15. Какая функция наиболее удобна в обращении на практике?

4.6 Список литературы

1. Интерполяция и аппроксимация кривых: методические указания к лабораторной работе / Курск. политехн. ин-т; сост. С.А. Филист. – Курск. – 1993. – 13 с.

ПРАКТИЧЕСКАЯ РАБОТА № 5 ПРИМЕР РЕШЕНИЯ ПРАКТИЧЕСКОЙ ЗАДАЧИ КЛАСТЕРНЫМ АНАЛИЗОМ В ПАКЕТЕ STATISTICA

5.1 Краткие теоретические сведения

Система STATISTICA является одним из наиболее известных в мировой практике пакетов статистической обработки в среде Windows.

Разработчиком пакета является фирма Stat Soft, Inc. (США).

Пакет STATISTICA отличается от большинства других программных продуктов для Windows тем, что состоит из отдельных программ-модулей, каждый из которых содержит конкретный метод обработки данных, например кластерный анализ, регрессионный анализ и т.д. Каждый такой модуль можно рассматривать как самостоятельную программу, независимую от остальных. Но такие операции, как ввод, корректировка данных, осуществляется в любом из модулей по желанию пользователя.

Таким образом, если перед пользователем стоит конкретная задача, к примеру, провести классификацию данных, то достаточно вызвать модуль *Cluster Analysis (Кластерный анализ)*, чтобы провести полноценную работу: ввод данных, их корректировку, построить различные диаграммы, а также осуществить непосредственно заданную классификацию.

Если же обработка данных предполагает использование нескольких групп методов, то есть требуется использование ряда модулей, то каждый модуль может быть вызван из другого, при этом автоматически осуществляется переход файла с обрабатываемыми данными в новый модуль.

Для запуска пакета требуется войти в среду Windows, нажать на кнопку *Пуск*, в строке *Программы* выбрать *STATISTICA*. Далее либо подвести курсор к строке *Основные статистики*, либо щелкнуть левой кнопкой мыши на слове *STATISTICA*. В последнем случае появится меню *STATISTICA Module Switcher (Переключатель модулей системы STATISTICA)*, предлагающее выбрать конкретный статистический модуль (рисунок 5.1).

Далее следует привести курсор на тот модуль, в котором собираемся работать, и щелкнуть дважды левой кнопкой мыши. Операцию вызова также можно осуществить через кнопку *Switch to...* (Переключиться в...). Для отмены вызова меню предназначена клавиша *Cancel* (Отмена).

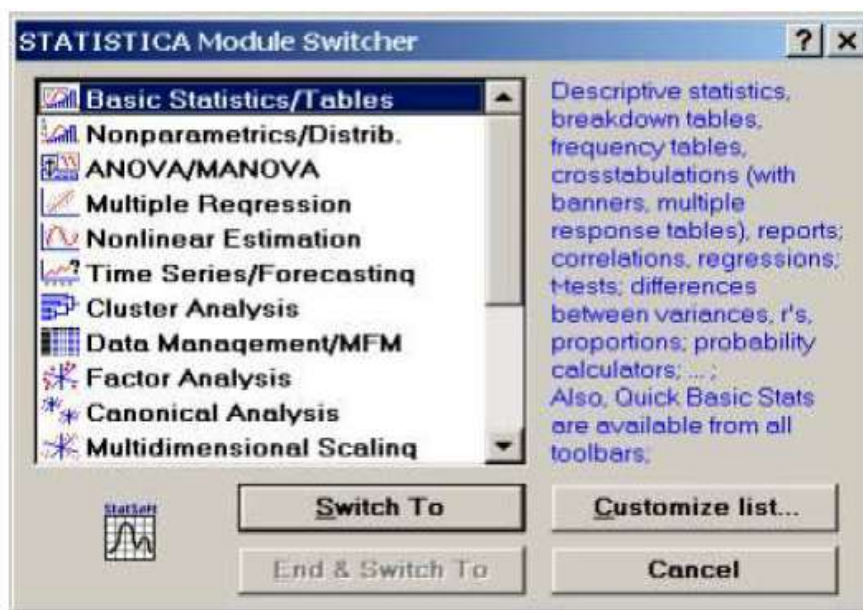


Рисунок 5.1 – Модули системы STATISTICA

Рассмотрим основные этапы проведения кластерного анализа в системе STATISTICA на следующем примере.

В файле (рисунок 5.2) содержатся данные по 20 странам (импорт товаров за 2000–2002 г). Задача состоит в том, чтобы распределить объекты по однородным группам и установить качественные взаимосвязи между группами стран с близкими значениями показателей.



	1 2000	2 2001	3 2002
Австрия	2206	6660	5777
Аргентина	30	441	603
Бельгия	1656	4346	5303
Болгария	2008	7119	9286
Бразилия	639	1114	1245
Великобритания	151	608	2412
Дания	206	91	1586
Германия	4473	6959	11436
Дания	1241	25	751
Индия	83	39	84
Испания	4724	4203	3331
Италия	5979	5671	6611
Канада	81	7	207
Китай	102	129	330
США	386	2297	36120
Турция	6	46	101
Франция	2696	6437	7999
Швеция	438	1027	532
Япония	1265	1238	2139
Южная Корея	162	166	1803

Рисунок 5.2 – Кластерный анализ

5.1.1 Алгоритм выполнения

Предварительно нормируем исходные данные в целях устранения различий в единицах измерения показателей. После вызова строки *Standardize Variables* (Стандартизация значений) на экране компьютера появится диалоговое окно задания параметров нормирования данных (рисунок 5.3).

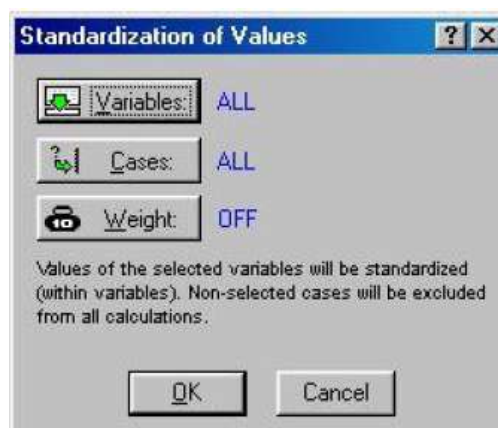


Рисунок 5.3 – Стандартизация значений

Кнопка *Variables* (Переменные) позволяет отобрать те показатели, которые будут нормированы. В моей задаче выбираются все.

С помощью кнопки *Cases* (Наблюдения), можно отобразить лишь часть наблюдений для стандартизации, по умолчанию выбираются все. Также по умолчанию все наблюдения вносят одинаковый вклад в вычисляемые средние значения и стандартные отклонения.

Нажав кнопку *Weight* (Вес), возможно указать “весовую” переменную в диалоговом окне *Define Weight* (Задание веса).

После процедуры нормирования данные выглядят следующим образом (рисунок 5.4):

	1 2000	2 2001	3 2002
Австрия	0,435865107	1,4866889	0,110383
Аргентина	-0,787615895	-0,72109166	-0,52889
Бельгия	0,12774581	0,687861292	0,051819
Болгария	0,324537332	1,68838025	0,543934
Бразилия	-0,445196648	-0,47826827	-0,44956
Великобритания	-0,719582255	-0,86083682	-0,30538
Венгрия	-0,643676953	-0,84737425	-0,40743
Германия	1,7105119	1,63065106	0,809575
Дания	-0,106717231	-0,87190915	-0,5103
Индия	-0,757816036	-0,86613623	-0,59301
Испания	1,85163954	0,665130425	-0,19183
Италия	2,55727771	1,2380926	0,213427
Канада	-0,758940559	-0,87768207	-0,57781
Китай	-0,747133068	-0,83366357	-0,56163
США	-0,587450804	-0,0514331	3,860372
Турция	-0,801110171	-0,86361058	-0,59091
Франция	0,71249776	1,44230959	0,36492
Швейцария	-0,558213207	-0,50965851	-0,53766
ЮАР	-0,093222955	-0,43713622	-0,33911
Япония	-0,713387378	-0,82031369	-0,38062

Рисунок 5.5 – Нормирование данных

Для проведения кластеризации вызовем модуль *Cluster Analysis* (Кластерный анализ) с помощью переключателя модулей *STATISTICA Module Switcher* (рисунок 5.6).

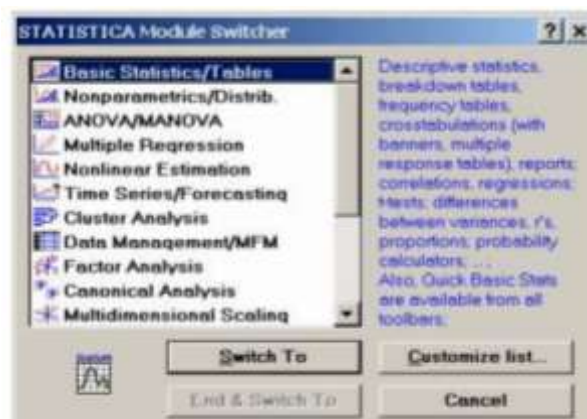


Рисунок 5.6 – Кластеризация

Или через меню *Статистика\Многомерные исследовательские методы\Анализ кластера*.

Появившееся диалоговое окно содержит следующие методы (рисунок 5.7):

- Joining (tree clustering) – Объединение (двевовидная кластеризация)
- K – means clustering – Кластеризация методом К– средних
- Two-way joining – Двухходовое объединение



Рисунок 5.7 – Методы анализа кластера

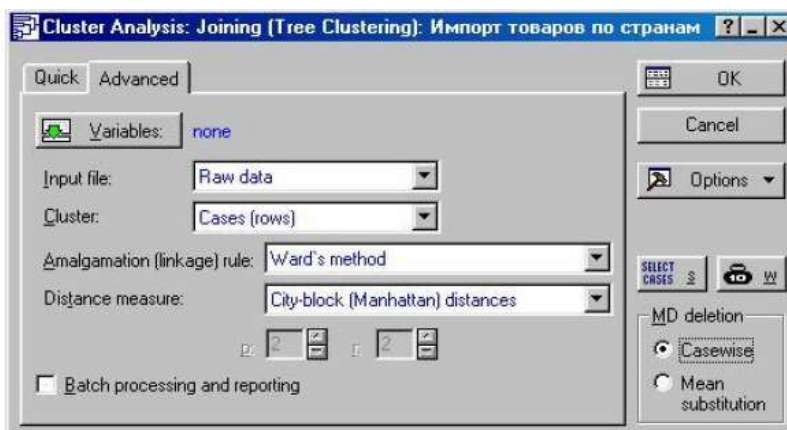


Рисунок 5.8 – Объединение (древовидная кластеризация) – Joining (tree clustering)

Щелкнув по кнопке *Variables* (Переменные), выберем показатели для анализа с помощью мыши или же в белом нижнем окошке через тире наберем их номера с клавиатуры,

если они идут подряд, иначе через пробел (рисунок 5.9). Кнопка *Select All* (Выбрать все) отбирает весь список.

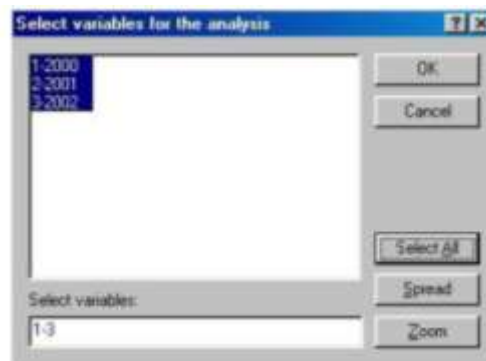


Рисунок 5.9 – Переменные

Input (Исходные данные) (рисунок 5.8) представляет собой раскрывающееся меню. Строка *Distance matrix* (Матрица расстояний) предусмотрена на тот случай, если входная информация представлена в виде мер сходства. В данной задаче выбираем второй тип – *Raw data* (Исходные данные).

В поле *Cluster* (Кластер) (рисунок 5.8) задается направление классификации. При кластеризации самих переменных помечаются *Variables [Columns]* (Переменные [столбцы]), в данной задаче *Cases [rows]* (Наблюдения [строки]).

Строка *Amalgamation [linkage] rule* (Правило объединения [связи]) содержит установки для выбора следующих мер сходства:

- *Single Linkage* (Метод одиночной связи “принцип ближайшего соседа”);
- *Complete Linkage* (Метод полной связи “принцип дальнего соседа”);
- *Unweighted pair-group average* (Невзвешенное попарное среднее);
- *Weighted pair-group average* (Взвешенное попарное среднее);
- *Unweighted pair-group centroid* (Невзвешенный центроидный метод);
- *Weighted pair-group centroid* (Взвешенный центроидный метод);

- *Ward's method* (Метод Варда).

Для решения данной задачи выбираем метод Варда.

В окошке *Distance measure* (Мера расстояния) (рисунок 1.11) предлагаются различные виды расстояний:

- *Squared Euclidean distances* (квадрат Евклидова расстояния)
- *Euclidean distances* (Евклидово расстояние)
- *City-block (Manhattan) distance* (Расстояние городских кварталов (Манхэттенское расстояние))
- *Chebyshev distance metric* (Расстояние Чебышева)
- *Power: SUM(ABS(x-y)**p)**1/r* (Степенное расстояние)
- *Percent disagreement* (Процент несогласия)

Выберем *City-block (Manhattan) distance* (Манхэттенское расстояние [расстояние городских кварталов]).

Перейдем к строке *Missing data [MD deletion]* (Пропущенные данные) (рисунок 5.8).

В системе STATISTICA в модуле *Cluster Analysis* (Кластерный анализ) предусмотрено два способа обработки некомплектных наблюдений, содержащих пропуски хотя бы одной переменной:

- *Casewise deleted* (Построчное удаление) - обработки некомплектные наблюдения полностью исключаются из дальнейшего анализа. Однако подобный метод приводит к смещенности и несостоятельности полученных статистических оценок, а также к искажению эмпирического распределения.

- *Substituted by means* (Замена средними значениями) – пропущенные данные заменяются средними значениями показателя, полученным по комплектным (полным) данным, что также имеет свои недостатки.

В данном примере пропуски отсутствуют, поэтому выбор того или иного метода не актуален.

После установки всех необходимых параметров для проведения кластеризации щелкнем на *Ok* и рассмотрим окно с результатами классификации (рисунок 5.10).

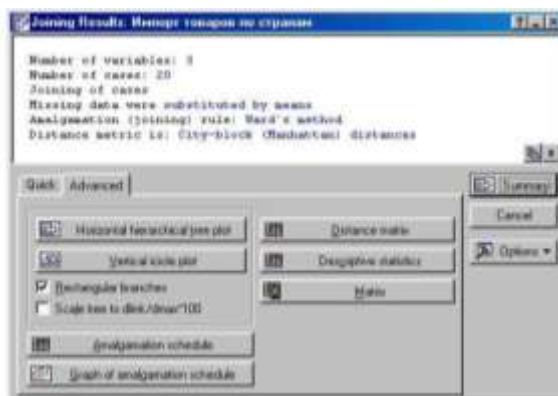


Рисунок 5.10 – Результаты классификации

Рассмотрим *Vertical icicle plot* (Вертикальную древовидную дендограмму).

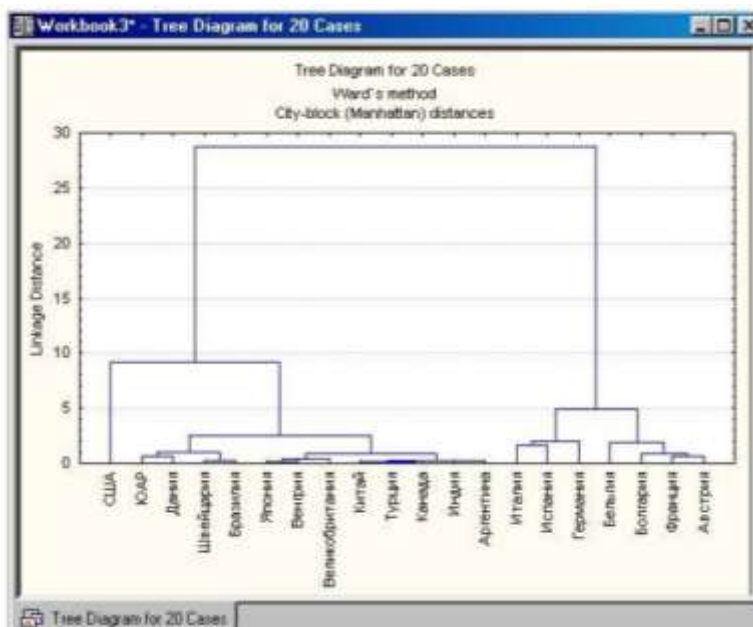


Рисунок 5.11 – Вертикальная древовидная дендограмма

На этой дендограмме (рисунок 5.11) горизонтальная ось представляет наблюдения, вертикальная – расстояние объединения. Таким образом, на первом шаге были объединены Индия и Канада, как имеющие минимальное расстояние, а на последнем – все, уже объединенные в какие-либо кластеры. На основе анализа диаграммы посмотрим, на каком шаге кластеризации следует принять полученную классификацию как окончательную. При числе кластеров равным двум $K=2$: в первый войдут: США, ЮАР, Дания, Швейцария, Бразилия, Япония, Венгрия, Ве-

Великобритания, Китай, Турция, Канада, Индия, Аргентина, а во второй – Австрия, Франция, Болгария, Бельгия, Германия, Испания, Италия; при $K=3$ первый кластер будет состоять из США, второй - ЮАР, Дания, Швейцария, Бразилия, Япония, Венгрия, Великобритания, Китай, Турция, Канада, Индия, Аргентина, третий – все остальные; при $K=4$ первый кластер будет состоять из США, второй - ЮАР, Дания, Швейцария, Бразилия, Япония, Венгрия, Великобритания, Китай, Турция, Канада, Индия, Аргентина, третий - Италия, Испания, Германия, четвертый - Бельгия, Болгария, Франция, Австрия; при $K=5$ первый кластер будет состоять из США, второй - ЮАР, Дания, Швейцария, Бразилия, третий - Япония, Венгрия, Великобритания, Китай, Турция, Канада, Индия, Аргентина, четвертый – Германия, Испания, Италия, пятый – Австрия, Франция, Болгария, Бельгия. Предпочтительнее взять количество кластеров равное 4 или 5, так как при большем их числе теряется наглядность классификации.

Щелкнув по кнопке **Amalgamation schedule** (Схема объединения), можно выбрать таблицу результатов со схемой объединения (рисунок 5.12). Первый столбец таблицы содержит расстояния для соответствующих кластеров. Каждая строка показывает состав кластера на данном шаге классификации. Например, на первом шаге (1 строка) объединились Индия и Канада, на втором (2 строка) – Индия, Канада и Турция, на третьем (3 строка) - Индия, Канада, Турция и Китай, на четвертом (4 строка) – Венгрия и Япония, на пятом - Бразилия и Швейцария, и т.д.

linkage distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6
0278675	Индия	Канада				
0688829	Индия	Канада	Турция			
1056906	Индия	Канада	Турция	Китай		
1236922	Венгрия	Япония				
2324989	Бразилия	Швейцария				
3103102	Аргентина	Индия	Канада	Турция	Китай	
3624056	Великобритания	Венгрия	Япония			
5965491	Австрия	Франция				
6197602	Дания	ЮАР				
8278940	Австрия	Франция	Болгария	Турция		
9005987	Аргентина	Индия	Канада	Турция	Китай	Великобритания
1.010666	Бразилия	Швейцария	Дания	ЮАР		

Рисунок 5.12 – Таблица результатов со схемой объединения

Щелкнув по кнопке *Graph of amalgamation schedule* (График схемы объединения) (рисунок 5.13), просмотрим результаты древовидной кластеризации в графическом виде.

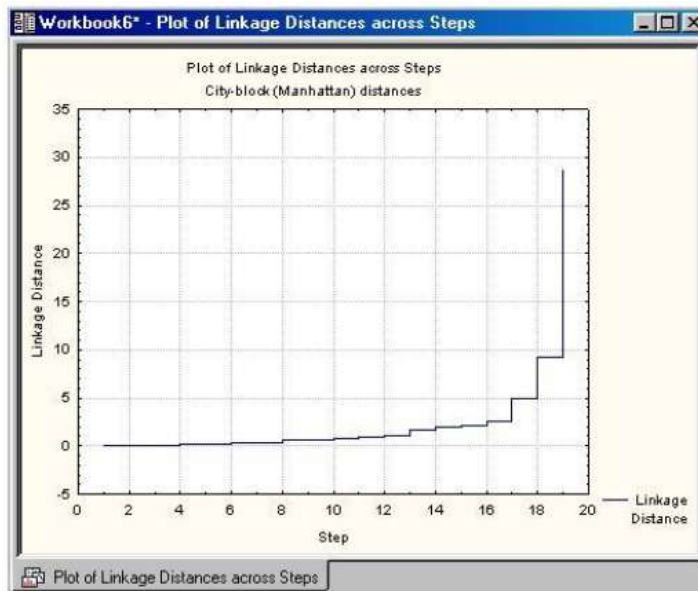


Рисунок 5.13 – График схемы объединения

Просмотр матрицы расстояний осуществляется через кнопку *Distance matrix* (Матрица расстояний) (рисунок 5.14), которая затем сохраняется с помощью опции *Save distance matrix* (Сохранить матрицу расстояний).

Case No	City-block (Manhattan) distances (Импорт товаров по странам)					
	Австрия	Аргентина	Бельгия	Болгария	Бразилия	Вел
Австрия	0,00	4,07	1,17	0,75	3,41	
Аргентина	4,07	0,00	2,91	4,59	0,66	
Бельгия	1,17	2,91	0,00	1,69	2,24	
Болгария	0,75	4,59	1,69	0,00	3,93	
Бразилия	3,41	0,66	2,24	3,93	0,00	
Великобритания	3,72	0,35	2,55	4,24	0,60	
Венгрия	3,93	0,39	2,77	4,46	0,61	
Германия	2,12	6,19	3,28	1,71	5,52	
Дания	3,52	0,85	2,36	4,05	0,79	
Индия	4,25	0,24	3,08	4,77	0,84	
Испания	2,54	4,36	1,99	3,29	3,70	
Италия	2,47	6,05	3,14	3,01	5,38	
Канада	4,25	0,23	3,08	4,77	0,84	
Китай	4,18	0,19	3,01	4,70	0,77	
США	6,31	5,26	5,26	5,97	4,88	

Рисунок 5.14 – Матрица расстояний

Строка *Descriptive statistics* (Описательные характеристики) открывает таблицу результатов со средними значениями и стандартными отклонениями для каждого объекта, включенного в кластерный анализ, т.е. для каждого наблюдения и переменной, в зависимости от установки, выбранной в списке *Cluster* (Кластер) стартовой панели (рисунок 5.15).

Case No	Means and Standard D	
	Mean	Std.Dev.
Австрия	0,677646	0,719304
Аргентина	-0,679198	0,134356
Бельгия	0,289142	0,347382
Болгария	0,852284	0,732343
Бразилия	-0,457677	0,017966
Великобритания	-0,561932	0,224117
Венгрия	-0,632828	0,220172
Германия	1,383579	0,498703
Дания	-0,496409	0,382793
Индия	-0,738987	0,137533
Испания	0,774980	1,026154
Италия	1,336266	1,175005
Канада	-0,738145	0,151012
Китай	-0,714141	0,138987
США	1,073830	2,428054
Турция	-0,751877	0,142862

Рисунок 5.15 – Список cluster в стартовой панели

5.1.3 Кластеризация методом К – средних – K – means clustering

Этот метод кластеризации существенно отличается от иерархических агломеративных методов. Он применяется, если пользователь уже имеет представление относительно числа кластеров, на которые необходимо разбить наблюдения. Тогда метод k – средних строит ровно k различных кластеров, расположенных на возможно больших расстояниях друг от друга.

Рассмотрим работу метода k-средних на данных нашего примера.

Щелчком по строке – *K – means clustering* (Кластеризация методом k-средних) стартовой панели модуля *Cluster analysis* (Кластерный анализ) (рисунок 5.7). На экране появится окно настройки параметров кластеризации (рисунок 5.16).

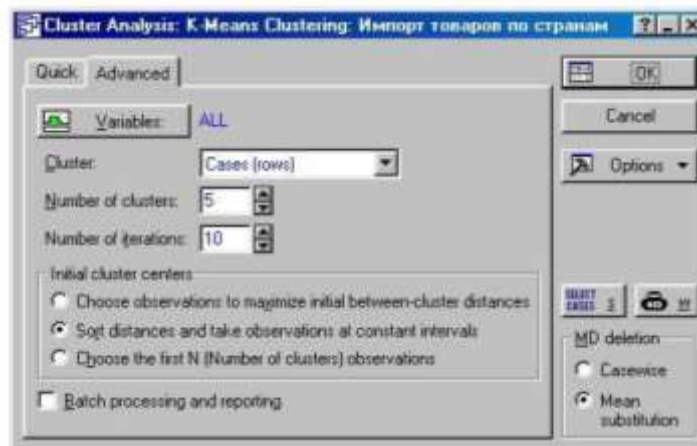


Рисунок 5.16 – Кластерный анализ

С помощью кнопки *Variables* (Переменные) выберем показатели, по которым будет происходить кластеризация. В строке *Cluster* (Кластер) укажем объекты для классификации *Cases [rows]* (Наблюдения [строки]).

Поле *Number of clusters* (Число кластеров) позволяет ввести желаемое число кластеров, которое должно быть больше 1 и меньше чем количество объектов.

Метод k-средних является итерационной процедурой, в результате которой на каждой итерации объекты перемещаются в различные кластеры. Поле *Number of iterations* (Число итераций) предназначено для указания их максимального числа.

Важным моментом при настройке параметров является выбор *Initial cluster centers* (Начальных центров кластеров), так как конечные результаты зависят от начальной конфигурации.

Опция *Choose observations to maximize initial between-cluster distances* (Выбрать набл, максимиз. начальные расстояния между кластерами) выбирает первые k в соответствии с количеством кластеров, наблюдений, которые служат центрами кластеров. Последующие наблюдения заменяют ранее выбранные центры в том случае, если наименьшее расстояние до любого из них больше, чем наименьшее расстояние между кластерами. В результате этой процедуры начальные расстояния между кластерами максимизируются.

Если выбрана опция *Sort distances and take observations at constant intervals* (Сортировать расстояния и выбрать на-

блюдения на постоянных интервалах), то сначала сортируются расстояния между всеми объектами, а затем в качестве начальных центров кластеров выбираются наблюдения на постоянных интервалах.

Choose the first N (Number of cluster) (Выбрать первые N [количество кластеров] наблюдений). Эта опция берет первые N (количество кластеров) наблюдений в качестве начальных центров кластеров.

Для нашего примера сделаем установку - *Sort distances and take observations at constant intervals* (Сортировать расстояния и выбрать наблюдения на постоянных интервалах).

Опция *MD deletion* (Пропущенные данные) устанавливает режим работы с теми наблюдениями (или переменными, если установлен режим *Variables (columns)*) в строке *Cluster*, в которых пропущены данные. По умолчанию установлен режим *Casewise* (Случай удаления). Тогда наблюдения просто исключаются из рассмотрения. Если установить режим *Mean substitution* (Заменять на среднее), то вместо пропущенного числа будет использовано среднее по этой переменной (или наблюдению).

Если установлена опция *Batch processing and reporting* (Пакетная обработка и сообщение), тогда система STATISTICA автоматически осуществит полный анализ и представит результаты в соответствии с установками.

После соответствующего выбора нажмем кнопку *OK*. STATISTICA произведет вычисления и появится новое окно: "*K - Means Clustering Results*" (рисунок 5.17).



Рисунок 5.17 – К - Means Clustering Results

5.1.4 Вывод результатов и их анализ

Появившееся окно с результатами классификации условно разделено на две части. В верхней части содержатся значения параметров, по которым проводится анализ, а в нижней – кнопки для вывода результатов.

В верхней части окна (в том же порядке, как они идут на экране):

- Количество переменных - 3;
- Количество наблюдений - 20;
- Классификация наблюдений (или переменных, зависит от установки в предыдущем окне в строке *Cluster*) методом К - средних;
- Наблюдения с пропущенными данными удаляются или изменяются средними значениями. Зависит от установки в предыдущем окне в строке *MD deletion*.
- Количество кластеров - 5;
- Решение достигнуто после: 3 итераций.

В нижней части окна расположены кнопки для вывода различной информации по кластерам.

Analysis of Variance (Дисперсионный анализ). После нажатия появляется таблица, в которой приведена межгрупповая и внутригрупповая дисперсии (рисунок 5.18). Где строки -

переменные (наблюдения), столбцы - показатели для каждой переменной: дисперсия между кластерами, число степеней свободы для межклассовой дисперсии, дисперсия внутри кластеров, число степеней свободы для внутриклассовой дисперсии, F - критерий, для проверки гипотезы о неравенстве дисперсий. Проверка данной гипотезы похожа на проверку гипотезы в дисперсионном анализе, когда делается предположение о том, что уровни фактора не влияют на результат.

Variable	Between SS	df	Within SS	df	F	signif. p
2000	18,22463	4	0,775366	15	88,14211	0,000000
2001	17,78698	4	1,213024	15	54,98750	0,000000
2002	18,21998	4	0,780024	15	87,59339	0,000000

Рисунок 5.18 – Внутригрупповая и межгрупповая дисперсия

Cluster Means & Euclidean Distances (средние значения в кластерах и евклидово расстояние). Выводятся две таблицы. В первой (рисунок 5.19) указаны средние величины класса по всем переменным (наблюдениям). По вертикали указаны номера классов, а по горизонтали переменные (наблюдения).

Variable	Cluster No. 1	Cluster No. 2	Cluster No. 3	Cluster No. 4	Cluster No. 5
2000	-0,587451	0,400162	-0,300838	-0,741159	2,039810
2001	-0,051433	1,326310	-0,574243	-0,811339	1,177958
2002	3,860373	0,272764	-0,459232	-0,493209	0,277058

Рисунок 5.19 – Cluster Means & Euclidean Distances

Во второй таблице (рисунок 5.20) приведены расстояния между классами. И по вертикали и по горизонтали указаны номера кластеров. Таким образом, при пересечении строк и столбцов указаны расстояния между соответствующими классами.

Причем выше диагонали (на которой стоят нули) указаны квадраты, а ниже просто евклидово расстояние.

Cluster Number	No. 1	No. 2	No. 3	No. 4	No. 5
No. 1	0,000000	5,248163	6,338153	6,518252	7,084682
No. 2	2,290887	0,000000	1,546440	2,152956	0,903491
No. 3	2,517569	1,243559	0,000000	0,083750	3,030321
No. 4	2,553087	1,467296	0,289397	0,000000	4,094800
No. 5	2,661706	0,950522	1,740782	2,023561	0,000000

Рисунок 5.20 – расстояния между классами

Щелкнув по кнопке *Graph of means* (График средних), можно получить графическое изображение информации содержащейся в таблице, выводимой при нажатии на кнопку *Analysis of Variance* (Дисперсионный анализ). На графике показаны средние значения переменных для каждого кластера (рисунок 5.21).

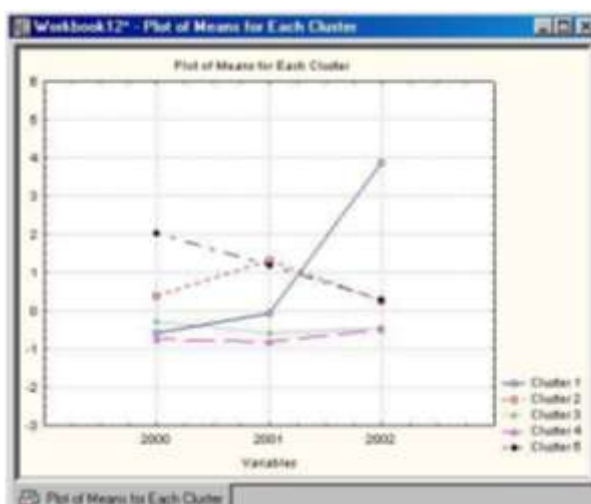


Рисунок 5.21 – График средних

По горизонтали отложены участвующие в классификации переменные, а по вертикали – средние значения переменных в разрезе получаемых кластеров.

Descriptive Statistics for each cluster (Описательная статистика для каждого кластера). После нажатия этой кнопки выводятся окна, количество которых равно количеству кластеров (рисунок

5.22). В каждом таком окне в строках указаны переменные (наблюдения), а по горизонтали их характеристики, рассчитанные для данного класса: среднее, несмещенное среднеквадратичское отклонение, несмещенная дисперсия.

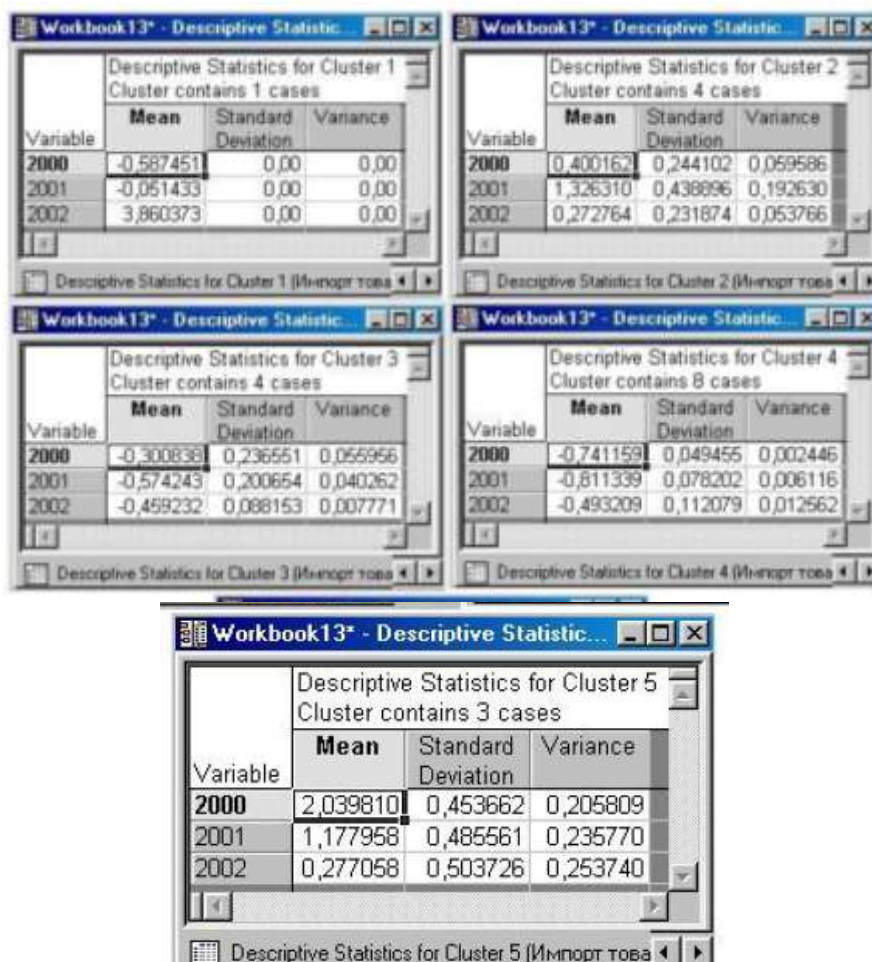


Рисунок 5.22 – Описательная статистика классов

Members for each cluster & distances (Члены каждой группы и расстояния). Выводится столько окон, сколько задано классов (рисунок 5.23). В каждом окне указывается общее число элементов, отнесенных к этому кластеру, в верхней строке указан номер наблюдения (переменной), отнесенной к данному классу и евклидово расстояние от центра класса до этого наблюдения (переменной). Центр класса - средние величины по всем переменным (наблюдениям) для этого класса.

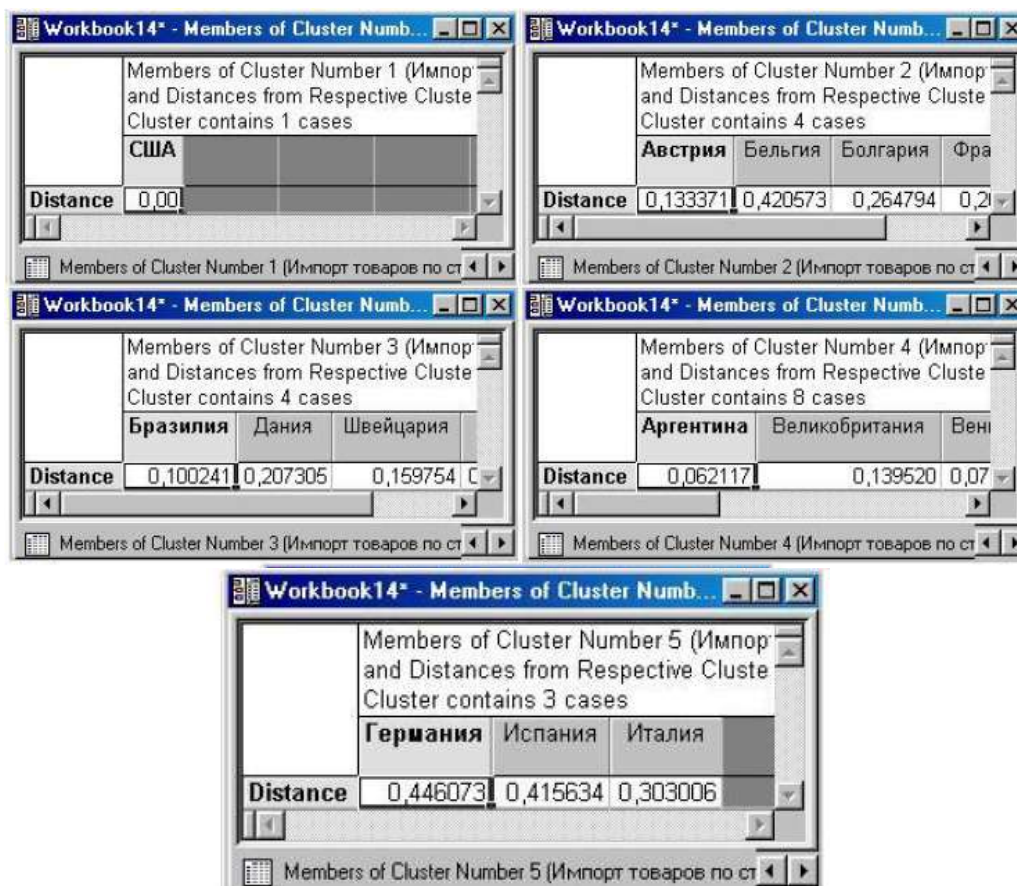


Рисунок 5.23 – Члены каждой группы и расстояния

Save classifications and distances. Позволяет сохранить в формате программы статистика таблицу, в которой содержатся значения всех переменных, их порядковые номера, номера кластеров к которым они отнесены, и евклидовы расстояния от центра кластера до наблюдения. Записанная таблица может быть вызвана любым блоком или подвергнута дальнейшей обработке.

Двухходовое объединение - Two-way joining.

Идея метода *Two-way joining* (Двухходовое объединение) состоит в том, чтобы одновременно классифицировать как наблюдения, так и переменные.

Трудность с интерпретацией полученных результатов возникает вследствие того, что получающиеся кластеры являются по своей природе часто неоднородными.

Вызовем исходное меню *Cluster analysis* (Кластерный анализ) (рисунок 5.7) и выберем строку *Two-way joining* (Двухходовое объединение). В появившемся диалоговом окне (рисунок

5.24) через кнопку *Variables* (Переменные) укажем переменные для анализа.

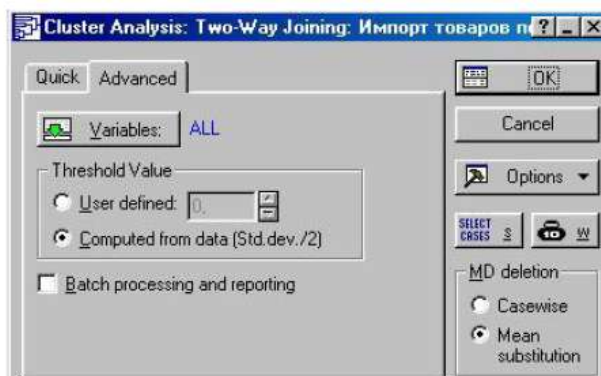


Рисунок 5.24 – Двухходовое объединение

Группа операций *Threshold Value* (Значение порога) содержит два режима: *User defined* (Заданное пользователем) и *Computed from data (Std.Dev./2)* (Вычисленное по данным).

Пороговый параметр определяет принадлежность элементов матрицы данных к формируемым кластерам. Если эта величина слишком велика по сравнению со значениями элементов в матрице, то будет сформирован только один кластер; если очень мала, то кластером будет являться каждая точка данных.

Для большинства случаев берут пороговое значение, равное половине величины общего стандартного отклонения (режим *Computed from data (Std.Dev./2)* (Вычисленное по данным)). После задания всех параметров нажмем *Ok*.

Окно с результатами вычислений имеет следующий вид:

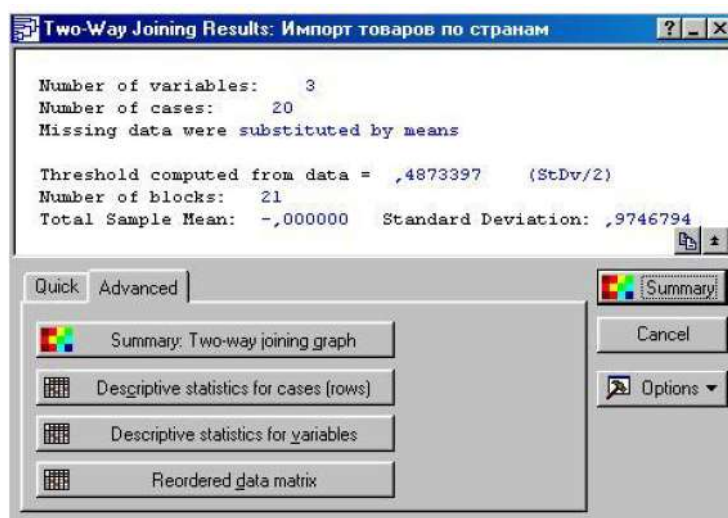


Рисунок 5.25 – Пороговый параметр

Две средние кнопки – *Descriptive statistics for cases (rows)* (Описательные статистики [строк]) и *Descriptive statistics for variables* (Описательные статистики для переменных) – выводят на экран таблицы со статистическими характеристиками данных: средними значениями и стандартными отклонениями (рисунок 5.26).

Variable	Mean	Std. Dev.
2000	0,000000	1,000000
2001	0,000000	1,000000
2002	0,000000	1,000000

Case ID	Mean	Std. Dev.
Австрия	0,677646	0,719304
Аргентина	-0,679198	0,134357
Бельгия	0,289142	0,347382
Болгария	0,852284	0,732343

Рисунок 5.26 – Средние значения и стандартные отклонения

Опция *Reordered statistics for variables* (Переупорядоченная матрица данных) формирует таблицу с наблюдениями в соответствии с результатами двухвходового объединения (рисунок 5.27).

Case ID	2000	2001	2002
Австрия	0,435865	1,486689	0,110383
Франция	0,712498	1,442310	0,384920
Болгария	0,324537	1,688380	0,543934
Германия	1,710512	1,630651	0,809576
Италия	2,557278	1,238093	0,213427
Бельгия	0,127746	0,687861	0,051819
Испания	1,851640	0,865130	-0,191830
Аргентина	-0,787616	-0,721092	-0,528886
Индия	-0,757816	-0,866136	-0,593010
Канада	-0,758941	-0,877682	-0,577813
Турция	-0,801110	-0,863611	-0,590910
Китай	-0,747133	-0,833664	-0,561627
Великобритания	-0,719582	-0,660837	-0,305376

Рисунок 5.27 – Таблица наблюдений

Кнопка *Summary: Two-way joining graph* выводит графическое изображение результатов двухвходового объединения.

В данном случае переупорядоченная матрица данных отображается в виде карты линий уровня (рисунок 5.28).

На графике по горизонтали отложены участвующие в классификации переменные, а по вертикали – наблюдения. Цвета ячеек, находящихся на пересечении, указывают на принадлежность элементов матрицы к определенному кластеру.

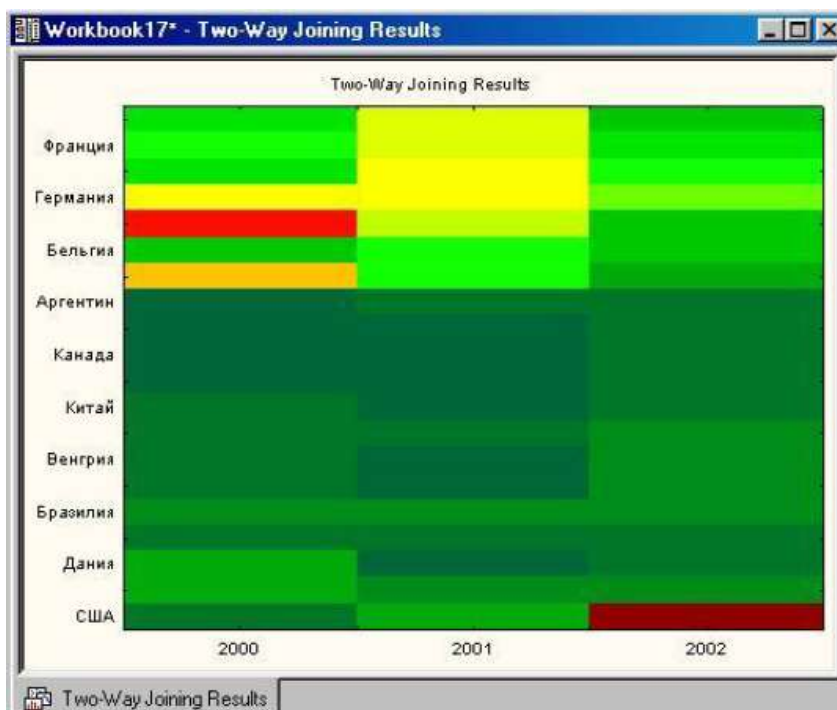


Рисунок 5.28 – Карта линий уровня

Такая сложная интерпретация результатов и спорная их практическая ценность делает рассматриваемую процедуру *Two-way joining* (Двухвходового объединения) не столь привлекательной.

5.2 Задания для самостоятельной работы

Задание 1.

По данным Таблицы 5.1 и алгоритму кластерного анализа провести классификацию объектов иерархическим методом (древовидная кластеризация).

Таблица 5.1 – Данные

№ п/п	Страны	Число врачей на 10000 населения	Смертность на 100000 населения	ВВП по паритету покупательной способности, в % к США	Расходы на здравоохранение, в % к США
		X1	X2	X3	X4
1	2	3	4	5	6
1	Россия	44.5	84.98	20.4	3.2
2	Австралия	32.5	30.58	71.4	8.5
3	Австрия	33.9	38.42	78.7	9.2
4	Азербай-	38.8	60.34	12.1	3.3
5	Армения	34.4	60.22	10.9	3.2

Продолжение таблицы 5.1.

1	2	3	4	5	6
6	Беларусь	43.6	60.79	20.4	5.4
7	Бельгия	41	29.82	79.7	8.3
	Болгария	36.4	70.57	17.3	5.4
9	Великобритания	17.9	34.51	69.7	7.1
10	Венгрия	32.1	64.73	24.5	6
11	Германия	38.1	36.63	76.2	8.6
12	Греция	41.5	32.84	44.4	5.7
13	Грузия	55	62.64	11.3	3.5
14	Дания	36.7	34.07	79.2	6.7
15	Ирландия	15.8	39.27	57	6.7
16	Испания	40.9	28.46	54.8	7.3
17	Италия	49.4	30.27	72.1	8.5
18	Казахстан	38.1	69.04	13.4	3.3
19	Канада	27.6	25.42	79.9	10.2
20	Киргизия	33.2	53.13	11.2	3.4

В качестве расстояния между объектами принять “обычное евклидово расстояние”, а расстояния между кластерами измерять по принципу: “ближайшего соседа”.

Исходные данные не нормировать.

Номер варианта соответствует номеру строки исключаемой из таблицы данных. Т.е. исследования проводятся для всех стран, кроме той, номер строки которой соответствует вашему варианту.

Задание 2.

Решить Задание 1, предварительно нормировав исходные данные.

Задание 3.

Решить Задание 1 при условии, что расстояния между кластерами измеряются по принципу “дальнего соседа”, предварительно нормируя исходные данные.

Задание 4.

Решить Задание 1, но в качестве расстояния между объектами принять “расстояние городских кварталов (Манхэттенское расстояние)”, а расстояния между кластерами измерять по методу Варда. Не нормируя предварительно исходные данные.

Задание 5.

Решить Задание 1 методом К-средних. Предварительно нормируя исходные данные.

Задание 6.

По данным Таблицы 5.2 и алгоритму кластерного анализа провести классификацию объектов иерархическим методом (древовидная кластеризация).

Таблица 5.2 - Данные

№ П/ П	Страны	Продукты					
		Мясо кг	Масло животное кг	Са- харкг	Алкоголь л	Фрукты, кг	Хлебопродукты, кг
		X1	X2	X3	X4	X5	X6
1	2	3	4	5	6	7	8
1	Россия	55	3,9	30	5	28	124
2	Австралия	100	2,6	47	8,2	121	87
3	Австрия	93	5,3	37	12	146	74
4	Азербайджан	20	4,1	12,4	7,9	52	141
5	Армения	20	3,7	4,3	6,5	72	134
6	Беларусь	72	3,6	28	5,4	38	120
7	Бельгия	85	6,9	48	11	83	72

8	Болгария	65	3	18	9,5	92	156
9	Великобрита-	67	3,5	39	8,8	91	91
10	Венгрия	73	1,7	40	10,9	73	106
11	Германия	88	6,8	35	8,1	138	73
12	Греция	83	1	24	8,8	99	108
13	Грузия	21	3,8	36	9,8	55	140
14	Дания	98	5	38	10,3	89	77
15	Ирландия	99	3,3	31	9,6	87	102
16	Испания	89	0,4	26	8,95	103	72
17	Италия	84	2,2	27	9,6	169	118
18	Казахстан	61	4,2	19,2	7,2	10	191
19	Канада	98	3,1	44	7,4	123	77
20	Киргизия	46	4,1	23,5	6,7	20	134

В качестве расстояния между объектами принять “обычное евклидово расстояние”, а расстояния между кластерами измерять по принципу: «ближайшего соседа».

Исходные данные не нормировать.

Номер варианта соответствует номеру строки исключаемой из таблицы данных. Т.е. исследования проводятся для всех стран, кроме той, номер строки которой соответствует вашему варианту.

Задание 7.

Решить Задание 6, предварительно нормировав исходные данные.

Задание 8.

Решить Задание 6 при условии, что расстояния между кластерами измеряются по принципу “дальнего соседа”, предварительно нормируя исходные данные.

Задание 9.

Решить Задание 6, но в качестве расстояния между объектами принять Расстояние городских кварталов (Манхэттенское расстояние), а расстояния между кластерами измерять по методу Варда. Не нормируя предварительно исходные данные.

ПРАКТИЧЕСКАЯ РАБОТА №6 ПРИМЕР РЕШЕНИЯ ПРАКТИЧЕСКОЙ ЗАДАЧИ ФАКТОР- НЫМ АНАЛИЗОМ В СИСТЕМЕ STATISTICA

6.1 Краткие теоретические сведения

Рассмотрим основные этапы проведения кластерного анализа в системе STATISTICA на следующем примере.

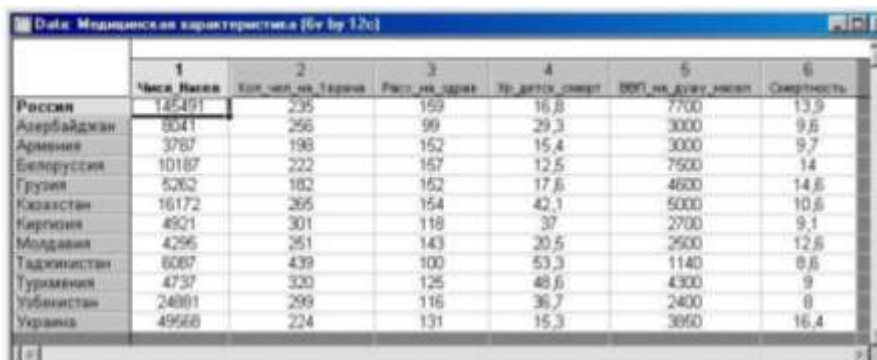
Исходными показателями послужили: X1 – численность населения (тыс.)

X2 – количество человек, приходящихся на одного врача X3 - Расходы на здравоохранение на душу населения (\$) X4 – Уровень детской смертности

X5 – ВВП, рассчитанный по паритету покупательной способности на душу населения (млн. \$)

X6 - Смертность на 1000 человек

В файле (рисунок 6.1) содержатся данные по 12 странам (по медицинской характеристике).



	1	2	3	4	5	6
	Числ. Насел.	Кол. чел. на 1000	Расх. на здрав.	Ур. детск. смерт.	ВВП на душу насел.	Смертность
Россия	125491	235	159	16,8	7700	13,9
Азербайджан	8041	256	99	29,3	3000	9,6
Армения	3767	198	152	15,4	3000	9,7
Белоруссия	10187	222	157	12,5	7500	14
Грузия	5262	182	152	17,6	4600	14,6
Казахстан	16172	265	154	42,1	5000	10,6
Киргизия	4921	301	118	37	2700	9,1
Молдавия	4256	251	143	20,5	2500	12,6
Таджикистан	6067	439	100	53,3	1140	8,6
Туркменистан	4737	320	125	48,6	4300	9
Узбекистан	24881	299	116	36,7	2400	8
Украина	49568	224	131	15,3	3950	16,4

Рисунок 6.1 – Медицинская характеристика

Задачей факторного анализа является объединение большого количества показателей, признаков, которыми характеризуется экономический процесс или объект, в меньшее количество искусственно построенных на их основе факторов, чтобы полученная в итоге система факторов (столь же хорошо описывающая выборочные данные, что и исходная) была наиболее удобна с точки зрения содержательной интерпретации.

6.1.1 Алгоритм выполнения

Модуль *Factor Analysis* (Факторный анализ) содержит широкий набор методов, снабжающих пользователя исчерпывающими средствами выделения факторов и представления результатов. Для вызова данного модуля можно использовать *STATISTICA Module Switcher* (Переключатель модулей), который содержит список всех доступных модулей (рисунок 6.2)

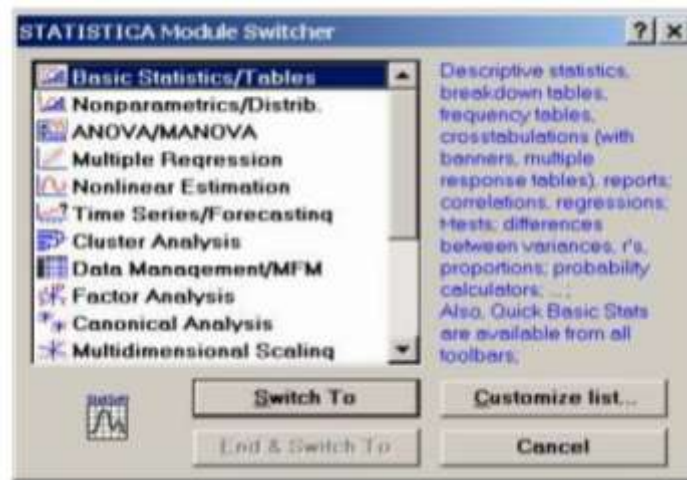


Рисунок 6.2 – Переключатель модулей

Или через меню *Статистика\Многомерные исследовательские методы \ Анализ фактора*.

На экране появится диалоговое окно (рисунок 6.3) *Factor Analysis*:

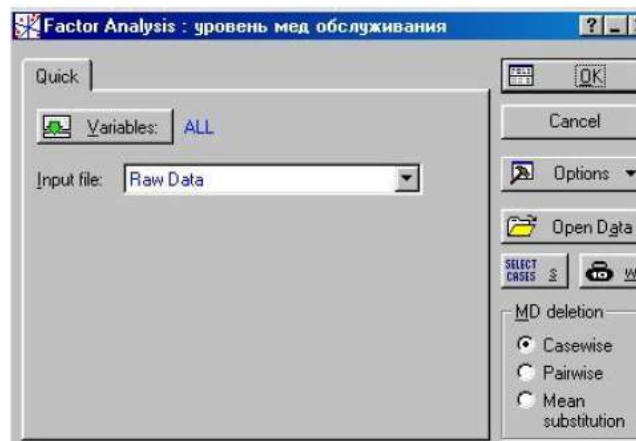


Рисунок 6.3 – Factor Analysis

Кнопка *Variables* (Переменные) позволяет отобрать все переменные из файла данных, которые должны быть включены в факторный анализ (рисунок 6.4). Если при анализе будут использованы все переменные, то можно воспользоваться кнопкой *Select All* (Выбрать все).

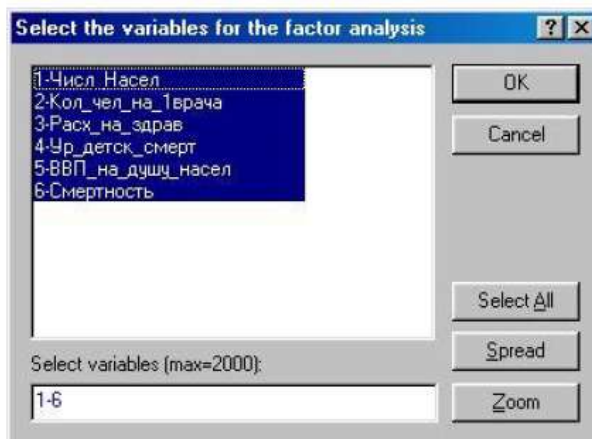


Рисунок 6.4 – Выбор переменных

Input File (Входной файл) (рисунок 6.3) должен содержать или необработанные данные или матрицу корреляций, и быть предварительно созданным в модуле *Factor Analysis* (Факторный анализ) или другом статистическом модуле.

В модуле возможны следующие типы исходных данных:

1. *Correlation Matrix* (Корреляционная матрица)
2. *Raw Data* (Исходные данные)

Выберите, например, *Raw Data*. Это обычный файл данных, где по строкам записаны значения переменных.

1. *MD deletion* (Замена пропущенных переменных) (рис. 2.3). Способ обработки пропущенных значений:

2. *Casewise* (способ исключения пропущенных случаев) - состоит в том, что в электронной таблице, содержащей данные, игнорируются все строки (случаи), в которых имеется хотя бы одно пропущенное значение. Это относится ко всем переменным. В таблице остаются только случаи, в которых нет ни одного пропуска.

3. *Pairwise* (Парный способ исключения пропущенных значений) - игнорируются пропущенные случаи не для всех переменных, а лишь для выбранной пары. Все случаи, в которых нет

пропусков, используются в обработке, например, при поэлементном вычислении корреляционной матрицы, когда последовательно рассматриваются все пары переменных. Очевидно, в способе Pairwise остается больше наблюдений для обработки, чем в способе Casewise. Тонкость, однако, состоит в том, что в способе Pairwise оценки различных коэффициентов корреляции строятся по разному числу наблюдений.

4. *Mean Substitution* (подстановка среднего вместо пропущенных значений). Щелкнув в стартовом окне модуля на кнопку ОК, вы начнете анализ выбранных переменных.

STATISTICA обработает пропущенные значения тем способом, какой вы ей указали, вычислит корреляционную матрицу и предложит на выбор несколько методов факторного анализа.

Вычисление корреляционной матрицы (если она не задается сразу) – первый этап факторного анализа.

После щелчка по Ок можно перейти к следующему диалоговому окну.

Define Method of Factor Extraction (Определить метод выделения факторов) (рисунок 6.5).

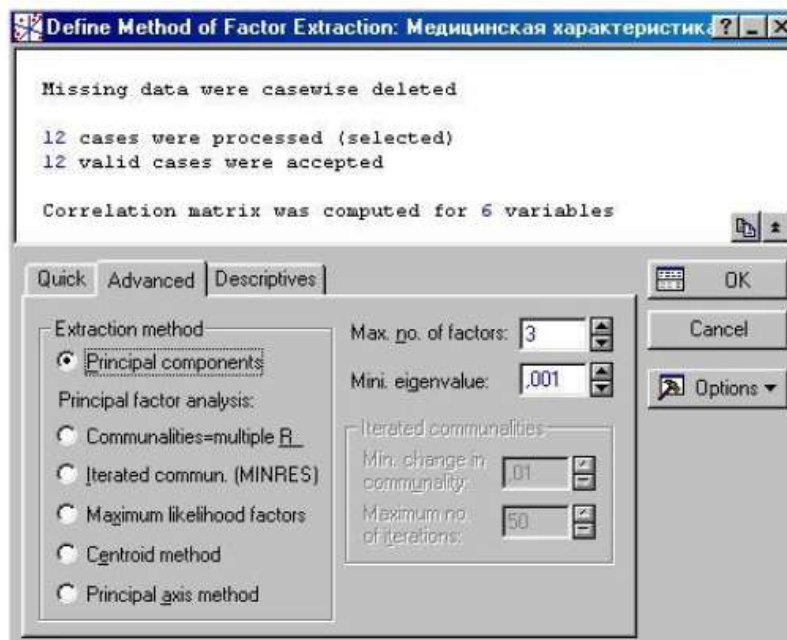


Рисунок 6.5 – Определение метода выделения факторов

Данное окно имеет следующую структуру.

Верхняя часть окна является информационной: здесь сообщается, что пропущенные значения обработаны методом Casewise. Обработано 12 случаев и 12 случаев приняты для дальнейших вычислений. Корреляционная матрица вычислена для 6 переменных.

Группа опций, объединенных под заголовком *Extraction method* (Методы выделения факторов) – позволяет выбрать метод обработки:

1. *Principal components* (метод главных компонент) – позволяет выделить компоненты, работая с первоначальной матрицей корреляций.

2. *Communalities=multiple R* (общности как множественный R-квадрат) – на диагонали матрицы корреляций будут находиться оценки квадрата коэффициента множественной корреляции- R^2 (соответствующей переменной со всеми другими переменными).

3. *Iterated communalities (MINRES)* (метод минимальных остатков) – выполняется в два этапа. Сначала оценки квадрата коэффициента множественной корреляции- R^2 используются для определения общностей, как в предыдущем методе. После первоначального выделения факторов метод корректирует их нагрузки с помощью метода наименьших квадратов с целью минимизировать остаточные суммы квадратов.

4. *Maximum likelihood factors* (метод максимального правдоподобия) – в этом методе считается заранее известным число факторов (оно устанавливается в поле ввода максимального числа факторов, см. ниже). STATISTICA оценит нагрузки и общности, которые максимизируют вероятность наблюдаемой в таком случае матрицы корреляций. В диалоговом окне результатов анализа доступен R-квадрат тест для проверки справедливости принятой гипотезы о числе общих факторов.

5. *Centroid method* (центроидный метод) – основан на геометрическом подходе.

6. *Principal axis method* (метод главных осей) – основан на итеративной процедуре вычисления общностей по текущим собственным значениям и собственным векторам. Итерации продолжаются до тех пор, пока не превышено максимальное число

итераций или минимальное изменение в общностях больше, чем это определено в соответствующем поле.

7. *Max. no. of factors* (Максимальное число факторов). Заданное в этом поле число определяет, сколько факторов может быть выделено при работе рассмотренных выше методов. Это поле работает вместе с полем *Min. eigenvalue* (Минимальное собственное значение). Часто при заполнении этого поля руководствуются критерием Кайзера, который рекомендует использовать лишь те факторы, для которых собственные значения не меньше 1.

Остальные поля доступны только при выбранном методе *Centroid method* (Центроидный метод) или *Principal axis method* (Метод главных осей), и определяют необходимые для успешного выполнения последовательных итераций параметры минимального изменения в общностях и максимального числа итераций.

В окне *Define Method of Factor Extraction* (Определить метод выделения факторов) (рисунок 6.6) щелкните на кнопке *Review correlations, means, standart deviations* (Просмотреть корреляции /средние/стандартные отклонения).

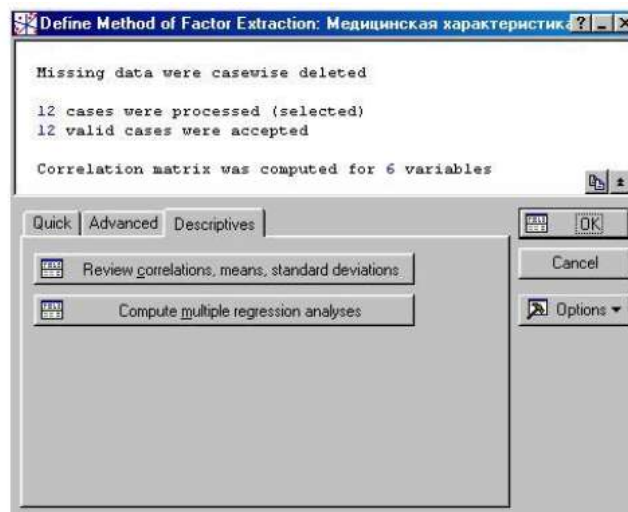


Рисунок 6.6 – Определение метода выделения факторов

Перед вами появилось окно просмотра описательных статистик для анализируемых данных (рисунок 6.7), где можно посмотреть средние, стандартные отклонения, корреляции, ковариации, построить различные графики. Здесь можно провести

дополнительный анализ текущих данных, проверить соответствие выборочных переменных нормальному закону распределения и существование линейной корреляции между переменными.

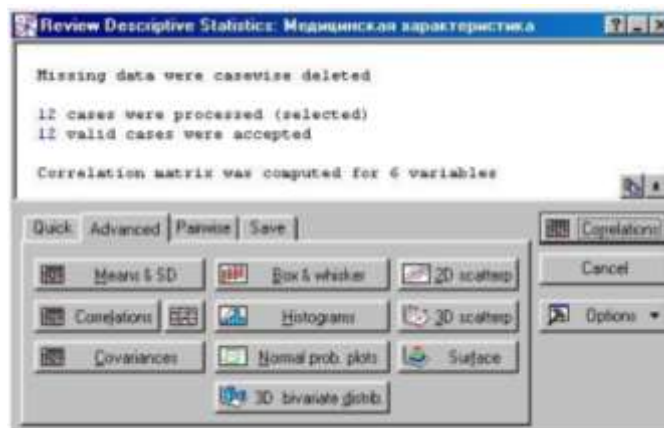


Рисунок 6.7 – Окно просмотра описательных статистик

Щелкните на кнопке *Correlations (Корреляции)* (рисунок 6.8). Вы увидите на экране корреляционную матрицу выбранных ранее переменных.

Variable	Числ_Насел	Кол_чел_на_1_врача	Расх_на_здрав	Ур_детск_смерт	ВВП_на_душу_насел	Смертность
Числ_Насел	1,00	-0,18	0,33	-0,31	0,57	0,41
Кол_чел_на_1_врача	-0,18	1,00	-0,71	0,87	-0,52	-0,64
Расх_на_здрав	0,33	-0,71	1,00	-0,63	0,73	0,60
Ур_детск_смерт	-0,31	0,87	-0,63	1,00	-0,49	-0,76
ВВП_на_душу_насел	0,57	-0,52	0,73	-0,49	1,00	0,58
Смертность	0,41	-0,64	0,60	-0,76	0,58	1,00

Рисунок 6.8 – Корреляции

Выберите опцию *Principal components* (Главные компоненты) и щелкните по кнопке ОК.

Система быстро произведет вычисления, и на экране появится окно *Factor Analysis Results* (Результаты факторного анализа).

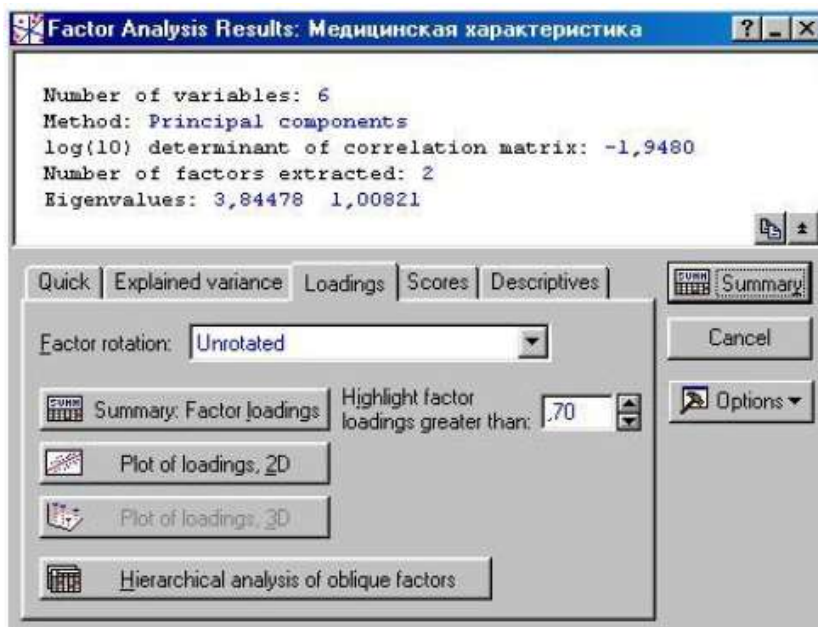


Рисунок 6.9 – Результаты факторного анализа

В верхней части окна *Результаты факторного анализа* дается информационное сообщение:

Number of variables (число анализируемых переменных) – 6;
Method (метод анализа) – главные компоненты;

log(10) determination of correlation matrix (десятичный логарифм детерминанта корреляционной матрицы) – -1,9480;

Number of Factor extraction (число выделенных факторов) – 3;
Eigenvalues (собственные значения) – 3,84478; 1,00821.

В нижней части окна находятся подразделы, позволяющие всесторонне просмотреть результаты анализа численно и графически.

Plot of loadings, 2D и *Plot of loadings, 3D* (Графики нагрузок) – эти опции построят графики факторных нагрузок в проекции на плоскость любых двух выбранных факторов (рисунок 6.10) и в проекции в пространство трех выбранных факторов (для чего необходимо наличие как минимум трех выделенных факторов).

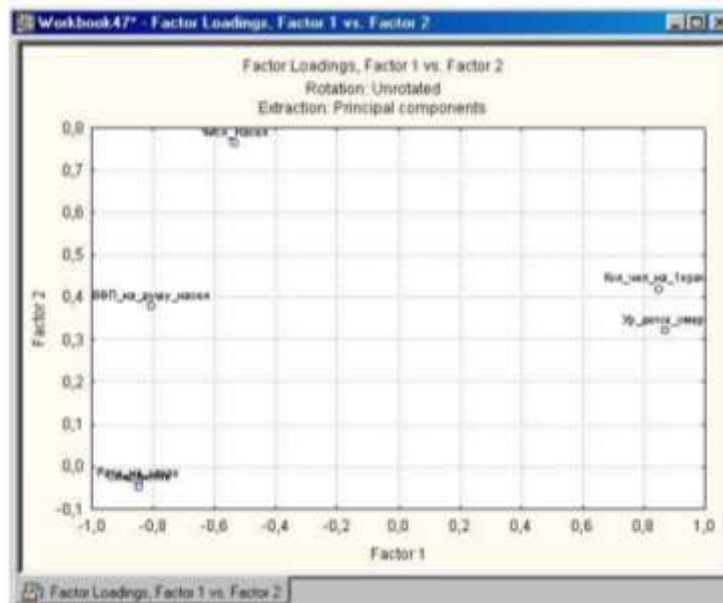


Рисунок 6.10 – Графики нагрузок

Summary. Factor loadings (Факторные нагрузки). Эта опция вызывает таблицу с текущими факторными нагрузками (рисунок 6.11), т.е. вычисленными для данного метода вращения факторов, который указан справа от соответствующей кнопки. В этой таблице факторам соответствуют столбцы, а переменным – строки и для каждого фактора указывается нагрузка каждой исходной переменной, показывающая относительную величину проекции переменной на факторную координатную ось. Факторные нагрузки могут интерпретироваться как корреляции между соответствующими переменными и факторами – чем выше нагрузка по модулю, тем больше близость фактора к исходной переменной; т.о., они представляют наиболее важную информацию для интерпретации полученных факторов. В сгенерированной таблице для облегчения трактовки будут выделены факторные нагрузки по абсолютной величине больше 0,7.

Variable	Factor Loadings (Unrotated)	
	Factor 1	Factor 2
Числ_Насел	-0,533777	0,762876
Кол_чел_на_1_врача	0,850247	0,418680
Расх_на_здрав	-0,847750	-0,041021
Ур_детск_смерт	0,871452	0,321421
ВВП_на_душу_насел	-0,803570	0,378952
Смертность	-0,844457	-0,048387
Expl. Var	3,844777	1,008213
Prp. Totl	0,640796	0,168036

Рисунок 6.11 – Текущие факторные нагрузки

По-видимому, первый фактор более коррелирует с переменными, чем второй.

Их трудно проинтерпретировать, возникает вопрос, какой смысл придать второму фактору. В этом случае целесообразно прибегнуть к повороту осей, надеясь получить решение, которое можно интерпретировать в предметной области.

Щелкните по меню *Factor rotation* (Вращение факторов) (рисунок 6.12).

Цель вращения – получение простой структуры, при которой большинство наблюдений находится вблизи осей координат. При случайной конфигурации наблюдений невозможно получить простую структуру.

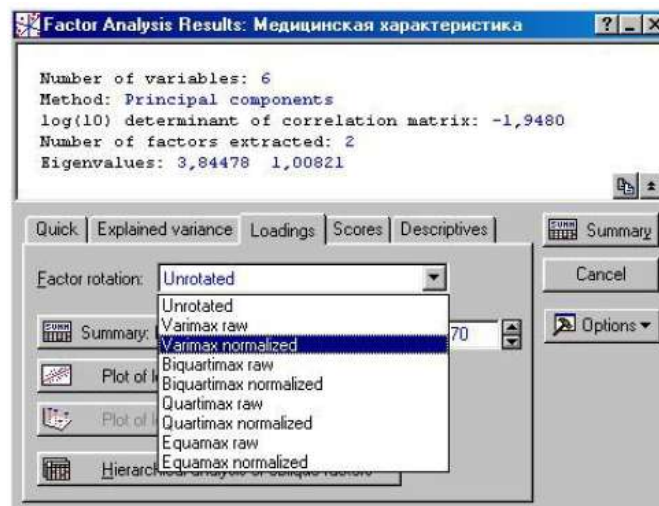


Рисунок 6.12 – Вращение факторов

В данном раскрывающемся меню вы можете выбрать различные повороты оси. Окно предлагает несколько возможностей оценить и найти нужный поворот следующими методами:

Varimax – Варимакс;

Biquartimax – Биквартимакс;

Quartimax – Квартимакс;

Equamax – Эквимакс.

Дополнительный термин в названии методов – *normalized* (нормализованные) – указывает на то, что факторные нагрузки в процедуре нормализуются, т.е. делятся на корень квадратный из соответствующей дисперсии. Термин *raw* (исходные) показывает, что вращаемые нагрузки не нормализованы.

Иницируйте кнопку *Varimax normalized* (Варимакс нормализованный).

Система произведет вращение факторов методом нормализованного варимакса, и окно *Factor Analysis Results* (Результаты факторного анализа) снова появится на мониторе. Вновь иницируйте в этом окне кнопку *Plot of Loadings 2D* (Двумерный график нагрузок). Вы опять увидите график нагрузок (рисунок 6.13).

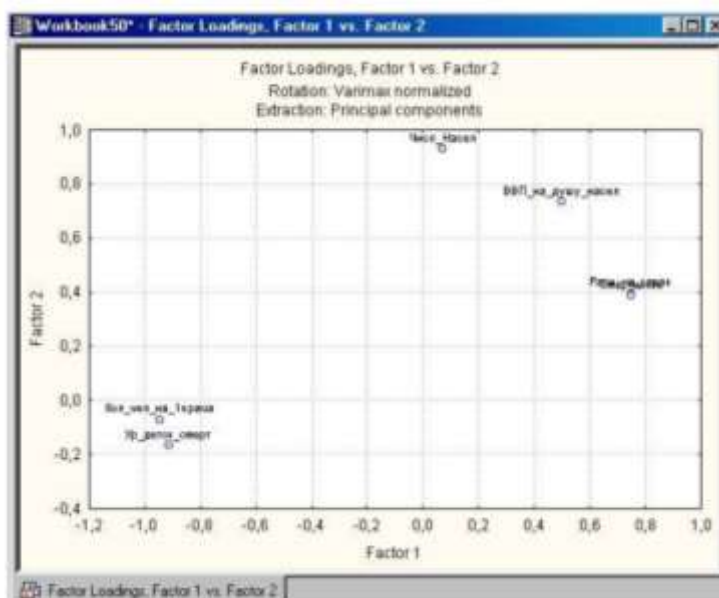


Рисунок 6.13 – Двумерный график нагрузок

Конечно, этот график немного отличается от предыдущего. Посмотрим еще нагрузки численно, иницировав кнопку *Фак-*

торные нагрузки (Factor loadings). Щелкните на кнопке *Summary. Factor loadings*, вы откроете окно:

Variable	Factor 1	Factor 2
Числ_Насел	0,072180	0,928272
Кол_чел_на_1врача	-0,945052	-0,071344
Расх_на_здрав	0,751033	0,395365
Ур_детск_смерт	-0,913904	-0,165890
ВВП_на_душу_насел	0,499612	0,734654
Смертность	0,751938	0,387347
Expl.Var	3,112625	1,740365
Prp.Totl	0,518771	0,290061

Рисунок 6.14 – Factor loadings

Теперь найденное решение уже можно интерпретировать. Факторы чаще интерпретируют по нагрузкам. Первый фактор теснее всего связан с X2 , X3, X4, X6. Второй фактор – X1 и X5. Таким образом, произвели классификацию переменных на две группы. Возникает вопрос: сколькими же факторами следует ограничиваться на практике? Для этого в программном пакете STATISTICA существует критерий *Scree plot* (Критерий каменистой осыпи). В окне *Factor Analysis Results* нажмите кнопку *Scree plot* получите следующий график собственных значений.

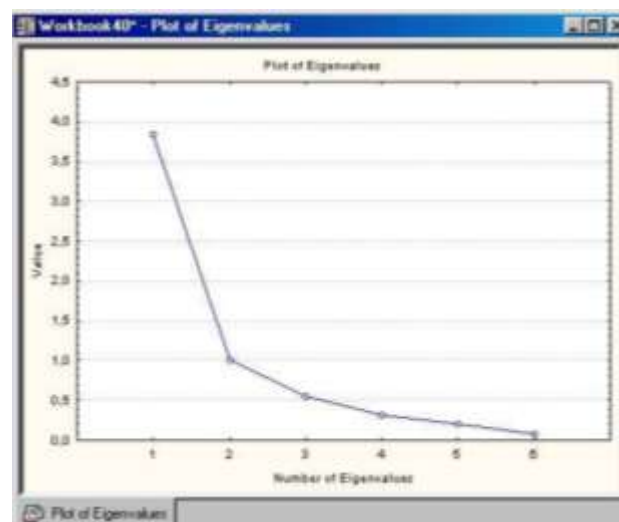


Рисунок 6.15 – График собственных значений

В точках с координатами 1, 2 осыпание замедляется наиболее существенно, следовательно, теоретически можно ограничиваться двумя факторами.

6.2 Задания для самостоятельной работы

Задание 1.

По исходным данным Приложения 3 за исключением переменных таблицы 6.1 согласно номеров вариантов провести факторный анализ. Не совершая поворот факторов.

Получить матрицу парных коэффициентов корреляции.

Получить первые три общие фактора и дать им экономическую интерпретацию по матрице факторных нагрузок.

Таблица 6.1 – Данные

№ варианта	Переменные	№ варианта	Переменные	№ варианта	Переменные
1	Y1, Y2, X4, X5	6	Y1, Y2, X14, X15	11	Y2, Y3, X10, X11
2	Y1, Y2, X6, X7	7	Y1, Y2, X16, X17	12	Y2, Y3, X12, X13
3	Y1, Y2, X8, X9	8	Y2, Y3, X4, X5	13	Y2, Y3, X14, X15
4	Y1, Y2, X10, X11	9	Y2, Y3, X6, X7	14	Y2, Y3, X16, X17
5	Y1, Y2, X12, X13	10	Y2, Y3, X8, X9	15	Y1, Y3, X4, X5

где Y1 – производительность труда;

Y2 – индекс снижения себестоимости продукции;

Y3 – рентабельность;

X4 – трудоемкость единицы продукции;

X5 – удельный вес рабочих в составе ППП;

X6 – удельный вес покупных изделий;

X7 - коэффициент сменности оборудования;

X8 – премии и вознаграждения на одного работника;

X9 – удельный вес потерь от брака;

X10 – фондоотдача;

X11 – среднегодовая численность ППП;

X12 – среднегодовая стоимость ОПФ;

X13 – среднегодовой фонд заработной платы;

X14 – фондовооруженность труда;
X15 – оборачиваемость нормированных оборотных средств;
X16 – оборачиваемость ненормированных оборотных средств;
X17 – непроизводственные расходы.

Задание 2.

На основе решения полученного в Задании 1 произвести поворот факторов методом Варимакс нормализованный.

Получить первые три общие фактора и дать им экономическую интерпретацию по матрице факторных нагрузок.

Задание 3.

На основе решения полученного в Задании 1 произвести поворот факторов методом Варимакс исходный. Получить первые три общие фактора и дать им экономическую интерпретацию по матрице факторных нагрузок.

ПРАКТИЧЕСКАЯ РАБОТА №7

ДИСКРИМИНАНТНЫЙ АНАЛИЗ И ЕГО РЕАЛИЗАЦИЯ В ПАКЕТЕ STATISTICA

7.1 Краткие теоретические сведения

Рассмотрим основные этапы проведения дискриминантного анализа в системе STATISTICA на следующем примере.

Исходными показателями послужили:

X1 – количество человек, приходящихся на одного врача

X2 - Смертность на 1000 человек

X3 – ВВП, рассчитанный по паритету покупательной способности на душу населения (млн. \$)

X4 - Расходы на здравоохранение на душу населения (\$)

В файле (рисунок 7.1) содержатся данные по 10 странам, которые были выбраны и отнесены к соответствующим группам экспертным способом (по уровню медицинского обслуживания):

- страны с высокими уровнем медицинского обслуживания
- страны со средним (удовлетворительным) уровнем медицинского обслуживания
- страны с низким уровнем медицинского обслуживания

Задача состоит в том, чтобы на основе аналогичных показателей классифицировать страны: Молдавия, Украина.



	1	2	3	4	5
	Кол_чел_на_1врача	Расх_на_здрав	ВВП_на_душу_насел	Смертность	Class
Азербайджан	256	99	3000	9,6	низкий
Армения	198	152	3000	9,7	низкий
Белоруссия	222	157	7500	14	высокий
Грузия	182	152	4600	14,6	удовлетв
Казахстан	265	154	5000	10,6	удовлетв
Киргизия	301	118	2700	9,1	низкий
Россия	235	159	7700	13,9	высокий
Таджикистан	439	100	1140	8,6	низкий
Туркмения	320	125	4300	9	удовлетв
Узбекистан	299	116	2400	8	низкий

Рисунок 7.1 – показатели уровня мед обслуживания

7.1.1 Алгоритм выполнения

Для вызова данного модуля можно использовать *Module Switcher* (Переключатель модулей) (рисунок 7.2), который содержит список всех доступных модулей. Высветите название модуля *Discriminant Analysis* (Дискриминантный Анализ) и нажмите кнопку *Switch to* (Переключиться в), или через меню *Статистика\Многомерные исследовательские методы\ Дискриминантный анализ*.

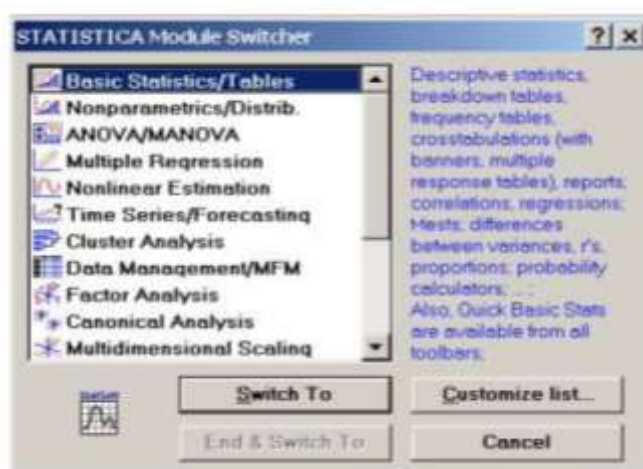


Рисунок 7.2 – Переключатель модулей

На экране появится стартовая панель модуля *Discriminant Function Analysis* (Анализ дискриминантных функций) (рисунок 7.3), в котором:

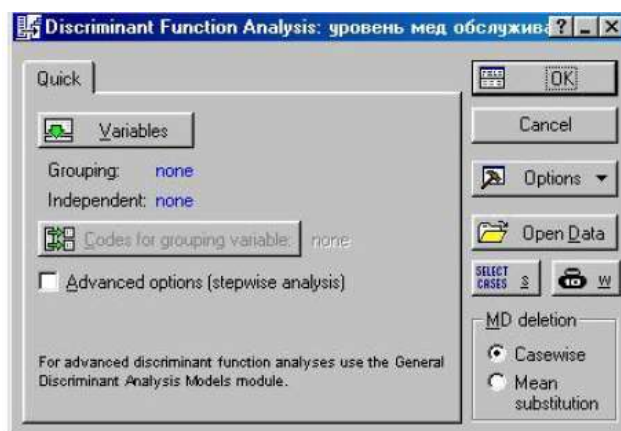


Рисунок 7.3 – Анализ дискриминантных функций

- *Variables* позволяет выбрать *Grouping* (Группируемую переменную) и *Independent* (Независимые переменные).
- *Codes for grouping variable* (Коды для групп переменной) указывают количество анализируемых групп объектов.
- *Missing data* (пропущенные переменные) позволяет выбрать *Case wise* (построчное удаление переменных из списка), либо *Mean substitution* (заменить их на средние значения).
- *Open Data* – открывает файл с данными.
- Можно указать условия выбора наблюдений из базы данных – кнопку *Select Cases* и веса переменных, выбрав их из списка – кнопку *W*.
- При нажатии кнопки *Variables* открывается диалоговое окно выбора переменных (рисунок 7.4).

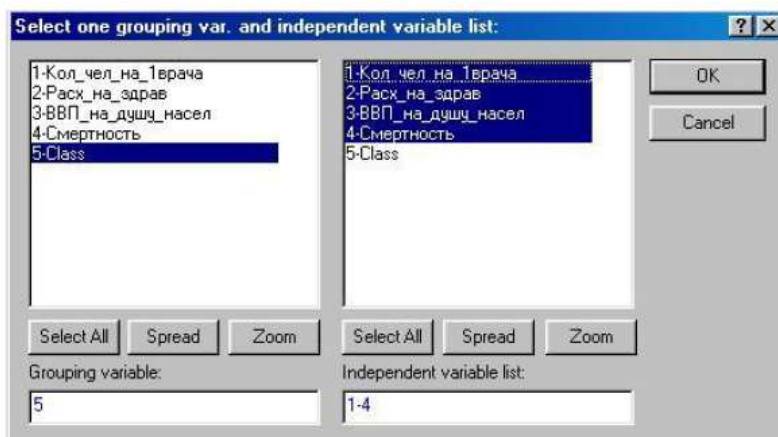


Рисунок 7.4 – Диалоговое окно выбора переменных

В левой части выбирается группирующие переменные, в правой – независимые переменные. Имена переменных в левой и правой части не должны пересекаться. В данном примере в качестве группирующей переменной выбрана переменная *CLASS*, а в качестве группирующих переменных *X1–X4*.

Select All (Выделить все) выделяет все переменные, *Spread* (Подробности) – для просмотра длинного имени, *Zoom* (Информация о переменной) позволяет просмотреть информацию о переменной: ее имя, формат числового значения, описательные статистики: номер в группе, среднее значение, статистическое отклонение.

После нажатия кнопки ОК откроется диалоговое окно *Model Definition* (Определение модели) (рисунок 7.5).



Рисунок 7.5 – Определение модели

В диалоговом окне *Model Definition* предложен выбор метода выбора значимых переменных. *Method* может быть задан *Standart* (Стандартный), *Forward stepwise* (Пошаговый с включением) и *Backward stepwise* (Пошаговый с исключением).

Если выбран *Standart* (Стандартный метод), то все переменные будут одновременно включены в модель.

В методе *Forward stepwise* (Пошаговый с включением) на каждом шаге в модель выбирается переменная с наибольшим F – значением. Процедура заканчивается, когда все переменные, имеющие F – значение больше значения, указанного в поле *F to enter* (F -включить), вошли в модель.

Если выбран метод *Backward stepwise* (Пошаговый с исключением), то в уравнение будут включены все выбранные пользователем переменные, которые затем удаляются в зависимости от величины F – значения. Шаги заканчиваются, когда нет переменных, имеющих F – значение меньше определенного пользователем в поле *F to remove* (F - исключить).

Если при проведении анализа пользователь хочет включить все переменные, то следует установить в поле *F to enter* (F - включить) очень маленькую величину, например 0.0001, а в поле *F to remove* (F – исключить) – 0.0.

Если же требуется исключить все переменные, то в поле *F to enter* (F - включить) следует установить большое значение, например 0.9999, а в поле *F to remove* (F – исключить) – 9998.

Поле *Number of steps* (число шагов) определяет максимальное число шагов анализа, по достижении которых процедура заканчивается.

Поле *Tolerance* (толерантность) позволяет исключить из модели неинформационные переменные. Значение толерантности вычисляется как $1-R^2$ переменной со всеми другими переменными в модели. Если толерантность имеет значение меньшее, чем значение по умолчанию 0.01 (или установленное специально пользователем), то эта переменная признается не информативной и не включается в модель.

В отличие от стандартного метода для пошаговых процедур предусмотрено два режима *Display of results* (вывода результатов) анализа:

- *At each step* (на каждом шаге) – программа выводит на экран диалоговое окно полных результатов на каждом шаге, начиная с нулевого.

- *Summary only* (на заключительном шаге) выводит окно с результатами только на последнем шаге, однако оно содержит опцию для просмотра основных итоговых статистик и для пошаговой процедуры.

- *Descriptives\Review Descriptive Statistics* (Обзор описательных статистик) позволяет получить описательные статистики для выбранных переменных:

- *Pooled within-groups covariances & correlations* (объединенные внутригрупповые ковариации и корреляции);

- *Total covariances & correlations* (полные ковариации и корреляции);

- *Graph* (графики корреляционных функций для всех переменных);

- *Means & number of cases* (средние значения для каждой переменной);

- *Box & wh* (диаграммы размаха);

- *Standart deviations* (стандартные отклонения переменных в каждой группе);
- *Categorized histogram (by group)* (категоризованные гистограммы по группам для каждой переменной);
- *Box & whisker plot (by group)* (диаграммы размаха по группам – категоризованную диаграмму рассеяния (по группам));
- *Categorized scatterplot (by group)* (для двух любых переменных);
- *Categorized normal probability plot (by group)* (категоризованный нормальный график для любой переменной по группам).

Выберем в качестве метода (*Method*) – *Standard* и нажмем ОК. В ходе вычислений системой получены результаты, которые представлены в окне *Discriminant Function Analysis Results* (Результаты анализа дискриминантных функций) (рисунок 7.6).



Рисунок 7.6 – Результаты анализа дискриминантных функций

7.1.2 Вывод результатов и их анализ:

Информационная часть диалогового окна *Discriminant Function Analysis Results* (Результаты Анализа Дискриминантных Функций) сообщает, что:

- *Number of variables in the model* (число переменных в модели) - 4;
- *Wilks' Lambda* (значение лямбды Уилкса) - 0,0086739;

- *Approx. F* (8, 8) (приближенное значение F – статистики, связанной с лямбдой Уилкса) - 9,737242;
- $p < 0.0021$ – уровень значимости F – критерия для значения 9,737242.

Значение статистики Уилкса лежит в интервале [0,1]. Значения статистики Уилкса, лежащие около 0, свидетельствуют о хорошей дискриминации, а значения, лежащие около 1, свидетельствуют о плохой дискриминации.

По данным показателя *Wilks' Lambda* (значение лямбды Уилкса), равного 0,0086739 и по значению F – критерия равного 9,737242, можно сделать вывод, что данная классификация корректная.

В качестве проверки корректности обучающих выборок посмотрим результаты классификационной матрицы, нажав кнопку *Classification matrix* (Классификационная матрица) (рисунок 7.7), предварительно выбрав *Same for all groups* (см. ниже) в правой части окна *Discriminant Function Analysis Results* (рисунок 7.6).

Group	Percent Correct	Predicted classifications		
		удовлетв p=,30000	низкий p=,50000	высокий p=,20000
удовлетв	100,0000	3	0	0
низкий	100,0000	0	5	0
высокий	100,0000	0	0	2
Total	100,0000	3	5	2

Рисунок 7.7 – Классификационная матрица

Из классификационной матрицы можно сделать вывод, что объекты были правильно отнесены экспертным способом к выделенным группам. Если есть предприятия, неправильно отнесенные к соответствующим группам, можно посмотреть *Classification of cases* (Классификация случаев) (рисунок 7.8).

Case	Observed Classif.	1 p=30000	2 p=50000	3 p=20000
Азербайджан	низкий	низкий	удовлетв	высокий
Армения	низкий	низкий	удовлетв	высокий
Белоруссия	высокий	высокий	удовлетв	низкий
Грузия	удовлетв	удовлетв	низкий	высокий
Казахстан	удовлетв	удовлетв	низкий	высокий
Киргизия	низкий	низкий	удовлетв	высокий
Россия	высокий	высокий	удовлетв	низкий
Таджикистан	низкий	низкий	удовлетв	высокий
Туркменистан	удовлетв	удовлетв	низкий	высокий
Узбекистан	низкий	низкий	удовлетв	высокий

Рисунок 7.8 – Классификация случаев

В таблице классификации случаев некорректно отнесенные объекты помечаются звездочкой (*). Таким образом, задача получения корректных обучающих выборок состоит в том, чтобы исключить из обучающих выборок те объекты, которые по своим показателям не соответствуют большинству объектов, образующих однородную группу.

Для этого с помощью метрики Махаланобиса определяются расстояние от всех n объектов до центра тяжести каждой группы (вектор средних), определяемых по обучающей выборке. Отнесение экспертом i -го объекта в j -ю группу считается ошибочным, если расстояние Махаланобиса от объекта до центра его группы значительно выше, чем от него до центра других групп, а апостериорная вероятность попадания в свою группу ниже критического значения. В этом случае объект считается некорректно отнесенным и должен быть исключен из выборки.

Процедура исключения объекта из обучающих выборок состоит в том, что в таблице исходных данных у объекта, который должен быть исключен из выборки (он помечен "*"), убирается номер принадлежности к этой группе, после чего процесс тестирования повторяется. По предположению, сначала убирается тот объект, который наиболее не подходит к определенной группе, т.е. у которого наибольшее расстояние Махаланобиса и наименьшая апостериорная вероятность.

При удалении очередного объекта из группы нужно помнить, что при этом смещается центр тяжести группы (вектор средних), так как он определяется по оставшимся наблюдениям.

После удаления очередного объекта из списка обучающих выборок не исключено, что появятся новые некорректно отнесенные объекты, которые до удаления были учтены как правильно отнесенные. Поэтому данную процедуру нужно проводить, удаляя на каждом шаге лишь по одному объекту и возвращая его обратно в обучающие выборки.

Процедура исключения наблюдений продолжается до тех пор, пока общий коэффициент корректности в классификационной матрице достигнет 100%, т.е. все наблюдения обучающих выборок будут правильно отнесены к соответствующим группам.

Результаты полученных обучающих выборок, представлены в окне *Discriminant Function Analysis Results* (Результаты Анализа Дискриминантных Функций). В результате проведенного анализа общий коэффициент корректности обучающих выборок должен быть равен 100%.

7.1.3 Классификация объектов

На основе полученных обучающих выборок можно проводить повторную классификацию тех объектов, которые не попали в обучающие выборки, и любых других объектов, подлежащих группировке. Для решения данной задачи, существуют два варианта: первый – провести классификацию на основе дискриминантных функций, второй – на основе классификационных функций.

В первом случае необходимо, не закрывая диалогового окна *Discriminant Function Analysis Results*, добавить в таблицу исходных скорректированных данных новые случаи. Для того чтобы понять, к какому классу относится этот объект, нажмите кнопку *Posterior probabilities* (Апостериорные вероятности). В системе имеется три способа задания априорной вероятности:

- *Proportional to group sizes* (Пропорциональные размерам групп);
- *Same for all groups* (Одинаковые для всех групп);
- *User defined* (Заданные пользователем).

После этого вы увидите таблицу с апостериорными вероятностями. К тем группам (классам), которые будут иметь максимальные вероятности, можно отнести новые случаи.

Во втором варианте необходимо в окне диалогового окна *Discriminant Function Analysis Results* нажать кнопку *Classification functions* (Классификационные функции). Появится окно (рисунок 7.9), из которого можно выписать классификационные функции для каждого класса.

Variable	удовлетв p=,30000	низкий p=,50000	высокий p=,20000
Кол_чел_на_1врача	1,834	1,455	2,35
Расх_на_здрав	1,718	1,455	1,98
ВВП_на_душу_насел	0,153	0,116	0,20
Смертность	36,637	29,066	46,93
Constant	-921,497	-576,414	-1526,02

Рисунок 7.9 – Классификационные функции

$$\text{высокий} = -1526,02 + 2,35 * \text{КолЧел_на_врача} + 1,98 * \text{Расх_здрав} + 0,20 * \text{ВВП} + 46,93 * \text{Смертность}$$

$$\text{удовлетв} = -921,497 + 1,834 * \text{КолЧел_на_врача} + 1,718 * \text{Расх_здрав} + 0,153 * \text{ВВП} + 36,637 * \text{Сметрность}$$

$$\text{низкий} = -576,414 + 1,455 * \text{КолЧел_на_врача} + 1,455 * \text{Расх_здрав} + 0,116 * \text{ВВП} + 29,066 * \text{Смертность}$$

С помощью этих функций можно будет в дальнейшем классифицировать новые случаи. Новые случаи будут относиться к тому классу, для которого классифицированное значение будет максимальное. Выбор метода окончательной классификации зависит от количества новых объектов, подлежащих классификации. Если количество новых случаев невелико, можно применить метод, основанный на статистических критериях. Если же количество новых случаев велико, то рациональнее по обучающим

выборкам получить классификационные функции и затем, настроить формулы и провести окончательную классификацию.

Определим принадлежность стран Молдавия и Украина, подставив значения соответствующих показателей в формулы (Таблица 7.1).

Таблица 7.1 – Данные

Страна	Кол-во человек на 1 врача	Расходы на здравоохранение	ВВП на душу населения	Смертность	Высокий	Удовлетворительный	Низкий	Class
Молдавия	251	143	2500	12,6	438,288	628,6372	653,876	Низкий
Украина	224	131	3850	16,4	880,232	904,2738	863,3934	Удовл

Дополнительно можно посмотреть результаты канонического анализа, нажав кнопку *Perform canonical analysis* (выполнить канонический анализ) (рисунок 7.10), если для анализа были выбраны, по крайней мере, три группы и есть хотя бы две переменные в модели.

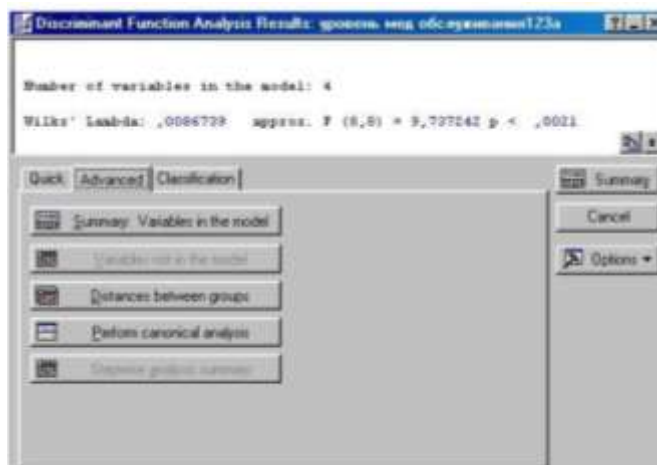


Рисунок 7.10 – Канонический анализ

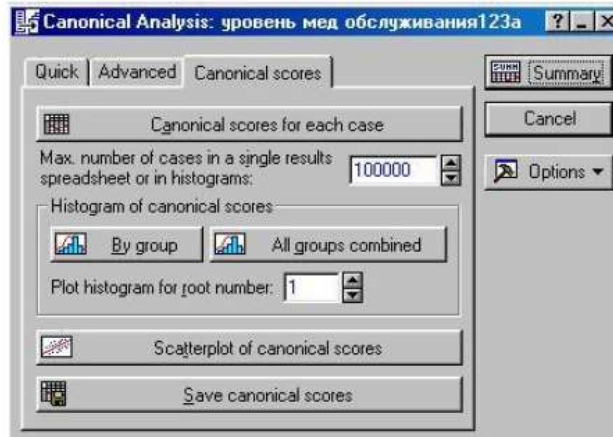


Рисунок 7.11 – Диаграмма рассеяния

В выводном на экран окне *Canonical analysis* (Канонический анализ) (рисунок 7.11) опция *Scatterplot of canonical scores* (Диаграмма рассеяния для значений) строит график рассеяния канонических значений для канонических корней (рисунок 7.12). С его помощью можно определить вклад, который вносит каждая дискриминантная функция в разделение между группами.

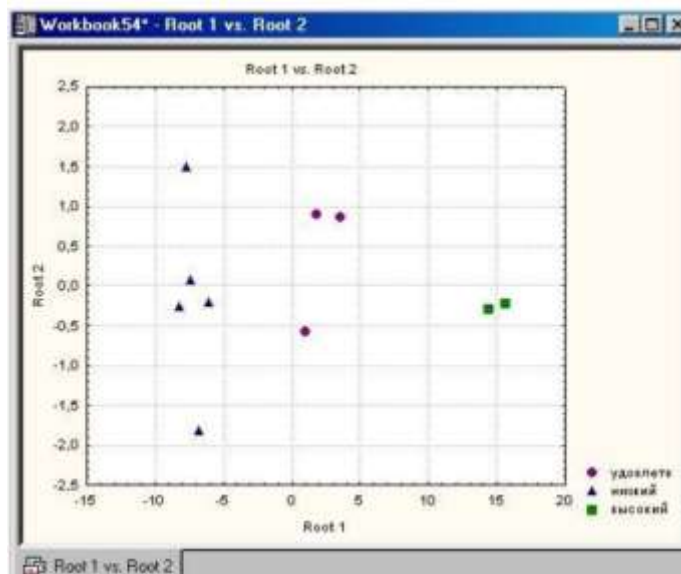


Рисунок 7.12 – График рассеяния канонических корней

7.2 Задания для самостоятельной работы

Задание 1.

В таблице 7.2 представлены 9 машиностроительных предприятий, которые были выбраны и отнесены к соответст-

вующим группам экспертным способом (с высоким и низким уровнем организации управления производством). Требуется на основе аналогичных показателей классифицировать предприятия 10-12 (Таблица 7.3).

Таблица 7.2 – Данные

№ предприятия	Рентабельность %	Производительность труда, млн руб/чел	Уровень организации управления производством
1	23,4	9,1	Высокий
2	19,1	6,6	Высокий
3	17,5	5,2	Высокий
4	17,2	10,1	Высокий
5	5,4	4,3	Низкий
6	6,6	5,5	Низкий
7	8,0	5,7	Низкий
8	9,7	5,5	Низкий
9	9,1	6,6	Низкий

Таблица 7.3 – Данные

№ предприятия	Рентабельность, %	Производительность труда, млн руб/чел
10	9,9	7,4
11	14,2	9,4
12	12,9	6,7

Задание 2.

В таблице 7.4 представлены 9 регионов, которые были выбраны и отнесены к соответствующим группам экспертным способом (с высоким и низким уровнем использования земли). Требуется с помощью дискриминантного анализа провести классификацию 10-12 районов (Таблица 7.5) по показателям объема реализованной продукции растениеводства и животноводство с 1 га посевной площади.

Таблица 7.4 – Данные

№ рай- она	Объем реализованной продукции		Уровень исполь- зования земли
	Растениеводство	Животноводство	
1	0,25	0,41	Низкий
2	0,51	0,51	Низкий
3	0,27	0,42	Низкий
4	0,33	0,56	Низкий
5	1,17	0,28	Высокий
6	4,99	0,67	Высокий
7	5,18	0,45	Высокий
8	2,49	0,38	Высокий
9	2,73	0,33	Высокий

Таблица 7.5 – Данные

№ района	Объем реализованной продукции	
	Растениеводство	Животноводство
10	0,32	0,45
11	0,67	0,32
12	4,6	0,56

Задание 3.

Эффективность деятельности 9 фермерских хозяйств оценивалась по трем показателям: производительность труда (млн. руб./чел.), объем реализованной продукции растениеводства и животноводство с 1 га посевной площади (млн. руб./га). В результате были выделены хозяйства с высоким и низким уровнем использования земли (таблица 7.6).

Необходимо с помощью дискриминантного анализа произвести классификацию 10-12 хозяйств (Таблица 7.7).

Таблице 7.6

№	Производительность труда	Объем реализованной продукции с 1 га посевной площади		Уровень ис- пользования земли
		Растениеводство	Животноводство	
1	8,22	0,25	0,41	Высокий
2	6,43	0,51	0,51	Высокий
3	6,39	0,27	0,42	Высокий
4	5,84	0,33	0,56	Высокий
5	3,14	1,17	0,28	Низкий
6	4,02	4,99	0,67	Низкий
7	4,19	5,18	0,45	Низкий
8	4,33	2,49	0,38	Низкий
9	4,58	2,73	0,33	Низкий

Таблице 7.7 – Данные

№	Производительность труда	Объем реализованной продукции	
		Растениеводство	Животноводство
10	4,92	0,32	0,45
11	5,22	0,67	0,32
12	5,62	4,60	0,56

Задание 4.

В Приложении 2 представлены данные по 65 предприятиям ($n=65$) и указаны номера групп, к которым они были отнесены экспертным способом. Задача состоит в том, чтобы проверить корректность экспертного отнесения предприятий к группам и получить корректные обучающие выборки (коэффициент корректности в классификационной матрице должен быть равен 100%). Выпишите линейные классификационные функции.

Задание 4.

На основе корректных обучающих выборок и классификационных функций, полученных в результате выполнения задания 4, провести группировку предприятий по пяти группам

на основе данных, представленных в Приложении 3, и соответствующих вариантов в таблице 7.7.

Таблице 7.7 – Данные

Номер варианта	Номера предприятий	Номер варианта	Номера предприятий	Номер варианта	Номера предприятий
1	1-50	10	46-95	19	91-140
2	6-55	11	51-100	20	96-145
3	11-60	12	56-105	21	101-150
4	16-65	13	61-110	22	106-155
5	21-70	14	66-115	23	110-160
6	26-75	15	71-120	24	116-165
7	31-80	16	76-125	25	121-170
8	36-85	17	81-130		
9	41-90	18	86-135		

X1 – прибыль (тыс. р.);

X2 – валовая продукция на 1 работника, занятого в сельском хозяйстве (тыс. р.);

X3 – валовая продукция на 1 га сельхозугодий (тыс. р.);

X4 – производство молока на 1 га сельхозугодий (кг); X5 – производство мяса на 1 га сельхозугодий (кг);

X6 – выручка от реализации продукции на 1 работника (тыс. р.);

X7 – выручка на 1 га сельхозугодий (тыс. р.).

7.3 Список литературы

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.:Статистика, 1974. – 240 с.

2. Айвазян С.А., Мхитарян В.С., Прикладная статистика и основы эконометрики. – М.:ЮНИТИ, 1998.

3. Айвазян С.А., Мхитарян В.С. Прикладная статистика: классификация и снижение размерности. – М.: Финансы и статистика, 1989.

4. Салин В.Н., Чурилова Э.Ю. Практикум по курсу “статистика” (в системе STATISTICA), М., 2002.

5. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988.

6. Дуброва Т.А., Павлов Д.Э., Осипова Н.П. Факторный анализ с использованием пакета “STATISTICA”. Учебное пособие / МГУ экономики, статистики и информатики, - М., 2002.

7. Дуброва Т.А., Бажин А.Г., Бакуменко Л.П. Методы многомерной классификации. Дискриминантный анализ в системе STATISTICA. Учебное пособие / МГУ экономики, статистики и информатики; М., 2002.

8. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. – М.: Финансы и статистика, 2000.

Приложение 1

Таблица 1 – Данные

	Y1	Y2	Y3	X4	X5	X6	X7	X8
1	2	3	4	5	6	7	8	9
1	9.26	204.20	13.26	0.89	0.34	1.73	0.31	0.28
2	9.44	209.60	10.16	0.93	0.33	0.99	0.15	0.25
3	12.11	223.54	13.72	1.33	0.17	1.73	0.14	0.47
4	10.81	236.70	12.83	0.68	0.32	0.47	0.18	1.53
5	9.33	62.00	10.63	0.89	0.36	1.73	0.31	0.21
6	9.87	53.10	9.12	1.53	0.33	1.33	0.17	0.13
7	8.17	172.10	25.95	1.12	0.15	0.97	0.26	0.38
8	9.12	56.50	23.39	0.99	0.32	1.82	0.29	0.38
9	5.88	52.60	14.68	1.65	0.31	0.68	0.26	0.20
10	6.30	46.60	10.05	0.56	0.15	1.80	0.28	0.35
11	6.19	53.20	13.89	0.58	0.17	1.19	0.25	0.20
12	5.46	30.10	9.68	1.53	0.15	0.97	0.49	0.20
13	6.50	146.40	10.03	0.70	0.16	1.15	0.26	0.17
14	6.61	18.10	9.13	1.77	0.15	0.02	0.28	0.25
15	4.32	13.60	5.37	0.74	0.17	0.06	0.17	0.16
16	7.37	89.80	9.86	1.08	0.34	1.39	0.17	0.21
17	7.02	62.50	12.62	1.15	0.34	0.08	0.31	0.19
18	8.25	46.30	5.02	0.97	0.34	0.77	0.18	1.24
19	8.15	103.47	21.18	1.12	0.19	0.77	0.31	0.43
20	8.72	73.30	25.17	0.99	0.19	1.08	0.18	0.14
21	6.64	76.60	19.40	0.58	0.34	0.93	0.31	0.29
22	8.10	73.01	21.00	1.03	0.34	0.10	0.15	0.43
23	5.52	32.30	6.57	1.24	0.15	0.11	0.28	0.17
24	9.37	198.54	14.19	0.89	0.19	1.44	0.18	0.21
25	13.17	598.12	15.81	0.68	0.34	0.48	0.14	0.42
26	6.67	71.69	5.20	1.03	0.19	1.24	0.18	1.19
27	5.68	90.63	7.96	0.73	0.32	0.77	0.29	1.87
28	5.19	82.10	17.50	0.73	0.19	0.93	0.30	0.15
29	10.02	76.20	17.16	0.85	0.33	0.13	0.27	0.03
30	8.16	119.47	14.54	1.03	0.34	1.73	0.29	0.24
31	3.78	21.83	6.21	0.47	0.36	0.77	0.14	0.93
32	6.45	48.40	12.08	0.56	0.33	0.16	0.29	0.13
33	10.38	173.50	9.39	0.89	0.32	0.74	0.44	0.27
34	7.65	74.10	9.28	0.99	0.15	1.95	0.14	0.17
35	8.77	68.60	11.44	1.95	0.16	0.58	0.29	0.24
36	7.00	60.80	10.31	1.03	0.16	1.77	0.18	0.19
37	11.06	355.60	8.65	0.01	0.20	0.70	0.44	0.29
38	9.02	264.81	10.88	0.02	0.15	0.74	0.31	0.25
39	13.28	526.62	9.87	0.60	0.33	1.15	0.18	0.36
40	9.27	118.60	6.14	0.97	0.33	1.19	0.14	0.17
41	6.73	37.10	12.99	1.12	0.19	1.03	0.31	0.23
42	6.72	57.97	9.78	1.77	0.15	1.08	0.18	0.17
43	9.44	51.84	13.22	0.93	0.32	0.13	0.27	0.24
44	7.21	64.70	17.29	1.12	0.16	0.74	0.26	0.26

Продолжение таблицы 1.

1	2	3	4	5	6	7	8	9
45	5.39	48.30	7.11	0.74	0.31	0.99	0.49	0.13
46	5.61	15.00	22.49	0.47	0.32	0.64	0.28	0.28
47	5.59	87.47	12.14	1.12	0.15	1.87	0.31	0.34
49	6.54	267.31	31.34	0.64	0.20	0.08	0.40	4.47
50	4.20	34.20	11.56	0.60	0.19	0.17	0.26	0.25
51	5.19	26.92	30.14	1.53	0.20	0.03	0.44	2.13
52	18.00	43.84	19.71	1.73	0.32	0.02	0.30	0.27
53	11.03	72.00	23.56	1.12	0.15	0.85	0.27	2.20
48	6.57	108.40	15.25	0.58	0.16	1.12	0.26	0.26

Таблица 2 - Данные

	X9	X10	X11	X12	X13	X14	X15	X16	X17
1	2	3	4	5	6	7	8	9	10
1	0.89	0.14	112216.00	166.19	9889.98	6.40	167.29	10.08	17.72
2	1.80	0.30	37631.94	186.10	22123.47	7.80	92.88	14.76	18.39
3	1.53	0.31	45178.00	220.45	10787.25	9.76	159.01	6.45	26.46
4	0.60	0.18	76688.00	169.30	10272.25	7.90	93.96	21.83	22.37
5	1.39	0.37	7361.00	39.93	55268.00	5.35	173.88	11.94	28.13
6	1.24	0.19	84496.00	40.41	45322.00	9.96	162.30	12.60	17.55
7	1.77	0.41	114132.00	102.96	12657.25	4.50	88.56	11.52	21.79
8	0.09	0.36	7801.00	37.02	57564.00	4.88	101.16	8.28	19.52
9	0.52	0.41	84504.00	45.94	118239.88	3.46	167.29	11.52	23.85
10	0.80	2.06	35852.00	40.07	64362.00	3.62	140.76	32.40	21.88
11	0.74	0.41	43244.00	45.44	69647.88	3.56	128.52	11.52	25.68
12	0.05	0.24	6358.00	41.08	49844.00	5.65	177.84	17.28	18.13
13	1.03	0.40	47378.00	136.14	22497.50	4.28	114.48	16.20	25.74
14	1.48	0.21	4210.00	42.39	6920.00	8.83	93.24	13.36	21.21
15	0.73	0.36	3572.50	37.39	5736.00	8.52	126.72	17.28	22.86
16	0.36	0.49	54544.00	101.78	47266.00	7.22	91.27	9.72	16.38
17	0.13	0.43	91264.00	47.91	72080.00	4.82	69.12	16.20	13.21
18	0.46	0.44	5975.00	32.61	83704.00	5.47	66.24	24.88	14.41
19	0.29	0.18	64044.00	103.73	107636.00	6.23	67.16	14.76	13.44
20	1.87	2.24	34328.00	38.95	67592.00	4.25	50.40	7.56	13.69
21	0.47	0.30	58424.00	81.32	99812.00	5.38	70.89	8.64	16.66
22	0.34	0.15	83240.00	67.75	75680.00	5.88	72.00	8.64	15.06
23	0.27	0.17	6462.00	59.66	44196.00	9.27	97.20	9.00	20.09
24	0.80	2.30	114896.00	107.81	20898.50	4.36	80.28	14.76	15.91
25	0.97	0.31	21791.47	512.62	28946.00	10.31	51.48	10.08	18.27
26	1.39	0.44	83568.00	53.53	74687.88	4.72	105.12	14.76	14.44
27	0.16	0.18	68976.00	80.83	8631.25	4.18	128.52	10.38	22.88
28	0.15	0.39	67663.88	59.42	31314.00	3.13	94.68	14.76	15.50
29	1.15	2.60	34428.00	36.96	64752.00	4.02	85.32	20.52	19.35
30	0.21	0.45	127256.00	91.88	8206.25	5.20	76.32	14.46	16.95
31	0.89	0.45	6265.00	17.16	44676.00	2.72	153.00	24.88	30.53
32	1.15	2.25	33192.00	27.29	65188.00	3.12	107.34	11.16	17.78
33	0.13	0.49	127983.88	184.33	22697.00	10.38	90.72	6.45	22.09
34	0.33	0.14	41368.00	58.42	68104.00	5.65	82.44	9.72	18.29

Продолжение таблицы 2.

1	2	3	4	5	6	7	8	9	10
35	0.64	0.18	33556.00	59.31	65616.00	6.67	79.12	3.24	26.05
36	0.93	0.29	124560.00	49.87	127344.00	5.93	120.96	6.45	26.20
37	0.14	0.50	110548.00	391.27	7919.00	11.89	84.60	5.40	17.26
38	0.13	0.26	95968.00	258.61	14314.75	8.30	85.32	6.12	18.95
39	0.16	0.21	21182.50	75.14	9277.13	0.18	101.52	8.64	19.66
40	0.49	0.49	53284.00	123.16	122072.00	8.88	107.34	11.94	16.97
41	1.80	0.28	6338.00	37.21	85792.00	5.82	85.32	7.92	14.63
42	0.15	0.15	44460.00	53.37	79631.88	4.80	131.76	10.08	22.17
43	1.99	2.03	6555.00	32.87	40516.00	5.01	116.64	18.72	22.62
44	0.10	0.40	44340.00	45.87	72580.00	4.12	138.24	13.68	26.44
45	0.17	0.27	35888.00	48.41	72296.00	5.10	156.96	16.56	22.26
46	0.21	0.38	3947.50	13.58	7612.00	3.47	137.52	14.76	19.13
47	1.65	0.45	59356.00	63.69	49072.00	4.22	134.06	7.92	18.28
48	0.64	0.14	75304.00	104.55	22896.00	5.01	155.52	18.36	28.23
49	0.20	0.43	68712.00	223.70	10949.73	11.38	48.60	8.28	12.44
50	1.92	0.14	3351.00	25.88	6235.00	7.67	42.84	14.04	11.64
51	1.65	0.17	6369.00	29.52	84000.00	4.68	143.64	16.79	8.62
52	0.09	0.18	90336.00	41.69	65050.00	4.30	145.80	11.16	20.10
53	0.19	0.17	38988.00	78.11	80708.00	6.62	120.52	14.76	19.41

Приложение 2

Таблица 3 – Данные

	X1	X2	X3	X4	X5	X6	X7	CLASSI
1	2	3	4	5	6	7	8	9
1	2174.000	9658.000	466.000	386.000	35.000	36.400	1756.000	1
2	274.000	10477.000	2321.000	767.000	56.000	35.600	7884.000	1
3	-146.000	6567.000	713.000	581.000	74.000	13.800	1501.000	1
4	-338.000	10282.000	499.000	764.000	51.000	30.200	1466.000	1
5	-716.000	9316.000	677.000	533.000	109.000	20.500	1486.000	1
6	892.800	6425.000	944.000	1390.000	78.000	13.200	1936.000	1
7	191.000	5367.000	786.000	819.000	104.000	13.700	2011.000	1
8	0.000	6342.000	486.000	261.000	52.000	24.100	1841.000	1
9	-107.000	5868.000	531.000	450.000	63.000	22.300	1608.000	1
10	-903.000	6330.000	636.000	401.000	69.000	17.600	1768.000	1
11	-765.000	12573.000	669.000	713.000	47.000	33.900	1806.000	1
12	326.000	4110.000	600.000	373.000	74.000	8.000	1172.000	2
13	150.000	7832.000	288.000	336.000	49.000	20.000	736.000	2
14	-18.000	6793.000	620.000	487.000	104.000	19.400	1775.000	2
15	1.300	4731.000	447.000	405.000	64.000	10.400	979.000	2
16	-380.000	5564.000	565.000	400.000	48.000	14.900	1517.000	2
17	-790.000	5470.000	432.000	509.000	85.000	11.800	935.000	2
18	-666.800	3988.000	364.000	213.000	35.000	10.300	943.000	2
19	-204.500	5121.000	495.000	628.000	77.000	16.700	1616.000	2
20	-094.000	3900.000	420.000	359.000	53.000	9.600	1034.000	2
21	-034.000	5871.000	495.000	353.000	92.000	14.300	1206.000	2
22	-784.000	4352.000	429.000	197.000	62.000	10.900	1070.000	2
23	-403.800	4635.000	378.000	221.000	46.000	10.500	856.000	3
24	-717.000	6056.000	247.700	150.000	24.100	15.600	640.000	3
25	-458.000	5180.000	433.600	429.000	44.400	10.500	880.000	3
26	-908.000	6295.000	206.000	127.000	17.000	22.800	743.000	3
27	-514.000	5340.000	364.000	411.000	17.000	14.400	984.000	3
28	-205.000	5357.000	583.000	716.000	87.000	14.800	1606.600	3
29	403.100	2969.000	382.000	274.000	29.000	5.700	728.000	3
30	-205.000	4924.000	284.000	292.000	35.000	17.500	1010.000	3
31	-256.000	7622.000	342.000	223.000	26.000	14.100	634.000	3
32	-314.000	4394.000	471.000	396.000	68.000	9.900	1065.000	3
33	-027.000	3312.000	284.000	229.000	39.000	11.100	948.000	3
34	1779.000	5001.000	304.400	286.000	37.600	12.000	732.000	3
35	-842.000	4247.000	233.000	189.000	28.000	12.800	757.000	3
36	-542.000	4025.000	199.300	145.000	14.400	12.000	596.000	4
37	-298.000	3429.000	184.000	105.000	18.000	6.700	357.300	4
38	-446.000	3047.000	310.000	244.000	47.000	5.500	560.000	4
39	-236.000	3410.000	181.000	147.000	20.000	10.900	576.000	4
40	-493.000	4551.000	212.000	169.000	22.000	13.800	645.000	4
41	-900.000	4573.000	284.000	254.000	37.000	11.300	698.000	4
42	-586.000	3924.000	212.000	154.000	17.000	13.000	704.000	4
43	-634.000	3751.000	212.000	125.000	17.000	5.400	303.000	4
44	-142.000	4318.000	257.000	151.000	33.000	16.500	985.000	4

Продолжение таблицы 3.

1	2	3	4	5	6	7	8	9
45	-394.000	3140.000	218.000	241.000	47.000	8.500	592.000	4
46	-571.000	4617.000	171.000	137.000	13.000	13.100	484.000	4
47	-728.300	5448.000	348.000	215.000	28.000	5.700	367.000	4
48	-796.000	2902.000	161.000	182.000	22.000	11.400	631.000	4
49	-955.200	3634.000	334.000	361.000	59.000	10.100	925.000	4
50	-294.000	3499.000	204.000	129.000	27.000	6.800	398.000	4
51	-500.000	6368.000	288.000	169.000	27.000	13.300	601.000	4
52	-961.000	4194.000	328.000	312.000	44.000	9.500	744.000	4
53	-934.000	6322.000	510.000	548.000	41.000	14.700	1187.000	4
54	-161.600	3196.000	288.000	149.000	55.000	7.600	684.000	5
55	-004.000	3666.000	168.000	131.000	19.000	8.300	382.000	5
56	-879.000	3058.000	169.000	86.000	23.000	5.600	307.000	5
57	-197.000	5110.000	82.000	57.000	11.000	1.100	174.000	5
58	-310.700	4166.000	207.000	183.000	32.000	9.800	487.000	5
59	-437.000	5168.000	151.000	96.000	8.000	10.700	359.000	5
60	-482.000	2061.000	78.000	47.000	4.000	2.900	110.300	5
61	-855.000	3483.000	109.000	90.000	16.000	7.600	237.000	5
62	-892.200	1917.000	98.000	64.000	9.000	4.000	174.000	5
63	-766.000	2001.000	95.000	87.000	18.000	5.000	239.000	5
64	-950.000	1728.000	87.000	75.000	13.000	3.400	172.300	5
65	-369.000	1094.000	38.000	1.200	3.200	3.300	114.000	5

Приложение 3

Таблица 4 – Данные

	X1	X2	X3	X4	X5	X6	X7
1	2	3	4	5	6	7	8
1	-790.000	5470.000	432.000	509.000	85.000	11.800	935.000
2	-1205.000	3698.000	187.500	156.000	21.600	10.000	507.000
3	-751.000	3448.000	278.000	206.000	25.000	7.400	596.000
4	-107.000	5868.000	531.000	450.000	63.000	22.300	1608.000
5	-903.000	6330.000	636.000	401.000	69.000	17.600	1768.000
6	-1204.500	5121.000	495.000	628.000	77.000	16.700	1616.000
7	-717.000	6056.000	247.700	150.000	24.100	15.600	640.000
8	-1027.000	3312.000	284.000	229.000	39.000	11.100	948.000
9	-625.000	2806.000	201.000	207.000	29.000	6.000	429.000
10	-1784.000	4352.000	429.000	197.000	62.000	10.900	1070.000
11	-1147.000	1849.000	158.000	45.000	13.000	3.100	266.600
12	-614.000	2168.000	191.000	168.000	22.000	9.100	804.000
13	-1080.000	3130.000	213.000	236.000	33.000	11.800	807.000
14	403.100	2969.000	382.000	274.000	29.000	5.700	728.000
15	139.000	3264.000	340.000	316.000	28.000	6.800	711.000
16	-444.000	3920.000	139.000	120.000	9.000	13.100	464.000
17	-833.000	5563.000	271.000	148.000	15.700	8.700	426.000
18	-380.000	5564.000	565.000	400.000	48.000	14.900	1517.000
19	-634.000	3751.000	212.000	125.000	17.000	5.400	303.000
20	-482.000	2061.000	78.000	47.000	4.000	2.900	110.300
21	-1892.000	4823.000	254.000	269.000	33.000	12.800	673.000
22	-256.000	7622.000	342.000	223.000	26.000	14.100	634.000
23	-987.000	4412.000	249.400	236.000	26.400	10.300	582.000
24	-255.000	2806.000	262.000	147.000	11.000	5.000	468.000
25	-908.000	6295.000	206.000	127.000	17.000	22.800	743.000
26	-1012.000	3912.000	102.000	87.000	9.000	11.200	290.000
27	-1500.000	6368.000	288.000	169.000	27.000	13.300	601.000
28	-321.000	1339.000	71.000	50.000	11.000	3.100	161.600
29	-1872.000	3883.000	262.000	230.000	26.000	8.300	557.000
30	-1161.600	3196.000	288.000	149.000	55.000	7.600	684.000
31	307.700	16355.000	939.000	83.000	25.000	8.700	501.000
32	-1146.000	6567.000	713.000	581.000	74.000	13.800	1501.000
33	-718.000	2027.000	126.000	105.000	12.000	7.700	474.000
34	-1197.000	5300.000	485.000	412.000	51.000	10.000	924.000
35	-497.800	5056.000	445.000	536.000	60.000	15.400	1388.100
36	-314.000	4394.000	471.000	396.000	68.000	9.900	1065.000
37	-373.000	5216.000	321.000	311.000	43.000	11.900	734.000
38	-854.300	5602.000	856.000	999.000	120.000	15.000	2285.000
39	-2004.000	3666.000	168.000	131.000	19.000	8.300	382.000
40	-900.000	4573.000	284.000	254.000	37.000	11.300	698.000
41	-219.800	3619.000	333.000	358.000	31.000	7.400	679.800
42	-523.600	3586.000	383.000	442.000	41.000	11.700	1252.300
43	-1731.300	2785.000	388.000	366.000	48.000	5.500	771.000
44	-205.000	5357.000	583.000	716.000	87.000	14.800	1606.600
45	-867.000	3495.000	150.700	98.000	14.200	9.700	419.000

Продолжение таблицы 4.

1	2	3	4	5	6	7	8
46	-950.000	1728.000	87.000	75.000	13.000	3.400	172.300
47	-2408.000	3714.000	253.000	229.000	24.000	7.600	518.000
48	-226.200	2863.000	252.000	203.000	31.000	6.600	582.600
49	-1597.000	2721.000	157.000	133.000	19.000	6.700	388.000
50	-1148.000	5126.000	468.000	340.000	50.000	12.200	1116.000
51	-504.600	3624.000	250.000	235.000	28.000	15.500	1071.000
52	-1796.000	2902.000	161.000	182.000	22.000	11.400	631.000
53	-1879.000	3058.000	169.000	86.000	23.000	5.600	307.000
54	-503.000	4634.000	336.000	271.000	31.000	11.100	802.000
55	-1394.000	3140.000	218.000	241.000	47.000	8.500	592.000
56	653.000	4791.000	245.000	202.000	37.000	17.300	881.000
57	-1094.000	3900.000	420.000	359.000	53.000	9.600	1034.000
58	-1949.000	3294.000	322.000	347.000	84.000	10.600	1034.000
59	-1064.000	2610.000	203.000	237.000	43.000	7.400	578.000
60	-1571.000	4617.000	171.000	137.000	13.000	13.100	484.000
61	-447.000	3390.000	282.000	258.000	33.000	7.100	592.000
62	-768.000	2845.000	332.000	241.000	43.000	5.500	637.000
63	-848.000	5197.000	269.200	163.000	14.200	9.900	513.000
64	-205.000	4924.000	284.000	292.000	35.000	17.500	1010.000
65	-1961.000	4194.000	328.000	312.000	44.000	9.500	744.000
66	-767.200	5863.000	622.200	277.000	95.000	13.800	1779.000
67	-1294.000	3499.000	204.000	129.000	27.000	6.800	398.000
68	-375.000	4291.000	187.000	165.000	18.000	11.300	493.000
69	-18.000	6793.000	620.000	487.000	104.000	19.400	1775.000
70	-526.000	3338.000	247.200	216.000	15.700	6.800	501.000
71	-745.000	3262.000	214.000	103.000	29.000	5.600	370.000
72	-113.400	7444.000	506.000	669.000	50.000	16.200	1100.200
73	-1332.000	4455.000	184.000	325.000	54.000	10.500	825.000
74	-1493.000	4551.000	212.000	169.000	22.000	13.800	645.000
75	-446.000	3047.000	310.000	244.000	47.000	5.500	560.000
76	714.600	7655.000	687.000	226.000	36.000	16.400	1474.000
77	179.000	4097.000	474.000	345.000	46.000	8.000	536.000
78	274.000	10477.000	2321.000	767.000	56.000	35.600	7884.000
79	-361.100	2754.000	385.000	499.000	62.000	10.700	1493.000
80	-1765.000	2977.000	281.000	249.000	34.000	6.500	612.000
81	1.300	4731.000	447.000	405.000	64.000	10.400	979.000
82	-3636.000	4129.000	133.000	57.000	14.000	13.900	446.000
83	-1016.000	4028.000	217.000	236.000	23.000	10.600	568.000
84	-2766.000	2001.000	95.000	87.000	18.000	5.000	239.000
85	-1008.000	2954.000	179.000	133.000	17.000	7.600	461.000
86	-452.000	2034.000	251.000	185.000	48.000	3.700	457.000
87	-419.500	1982.000	237.000	242.000	66.000	11.100	1327.000
88	-1306.800	2400.000	270.000	212.000	29.000	4.800	538.000
89	-809.000	4590.000	180.000	147.000	23.000	12.500	488.000
90	-1765.000	12573.000	669.000	713.000	47.000	33.900	1806.000
91	24.000	4666.000	268.000	344.000	16.000	8.900	512.000
92	-454.000	4799.000	433.000	421.000	41.000	9.200	831.000
93	-599.300	3535.000	287.000	287.000	40.000	9.000	752.000

Продолжение таблицы 4.

1	2	3	4	5	6	7	8
94	-2142.000	4318.000	257.000	151.000	33.000	16.500	985.000
95	-1514.000	5340.000	364.000	411.000	17.000	14.400	984.000
96	-1458.000	5180.000	433.600	429.000	44.400	10.500	880.000
97	-337.000	3017.000	187.000	191.000	29.000	6.600	411.000
98	-279.000	4211.000	294.000	259.000	41.000	6.400	451.000
99	-1376.000	4961.000	292.000	294.000	25.000	17.000	999.000
100	-403.800	4635.000	378.000	221.000	46.000	10.500	856.000
101	-2716.000	9316.000	677.000	533.000	109.000	20.500	1486.000
102	-298.000	3429.000	184.000	105.000	18.000	6.700	357.300
103	191.000	5367.000	786.000	819.000	104.000	13.700	2011.000
104	-728.300	5448.000	348.000	215.000	28.000	5.700	367.000
105	-404.000	2988.000	161.000	97.000	20.000	7.900	426.200
106	-473.700	5850.000	454.000	644.000	68.000	17.600	1365.500
107	-444.200	2351.000	297.000	183.000	35.000	4.500	564.000
108	-1236.000	3410.000	181.000	147.000	20.000	10.900	576.000
109	-715.000	2980.000	199.000	182.000	42.000	8.900	606.000
110	-1166.000	3798.000	273.000	267.000	42.000	8.600	619.000
111	-1141.000	3395.000	123.000	105.000	18.000	12.400	449.000
112	-486.000	1304.000	167.000	109.000	20.000	2.300	289.000
113	-667.000	2935.000	154.000	136.000	20.000	5.800	302.000
114	1365.000	5833.000	272.000	145.000	17.000	9.300	434.000
115	1553.000	2922.000	214.000	190.000	47.000	7.900	582.000
116	-706.000	3878.000	230.300	121.000	31.800	7.000	417.000
117	-1615.100	3779.000	216.000	264.000	18.000	13.300	763.000
118	-916.000	2362.000	173.000	144.000	18.000	6.900	505.000
119	-854.000	4532.000	278.000	233.000	26.000	11.600	714.000
120	-314.000	2788.000	191.000	105.000	21.000	5.900	406.000
121	-2128.000	3904.000	203.000	152.000	13.000	8.500	440.000
122	1779.000	5001.000	304.400	286.000	37.600	12.000	732.000
123	360.000	5490.000	315.000	214.000	30.000	14.900	854.000
124	2174.000	9658.000	466.000	386.000	35.000	36.400	1756.000
125	-662.000	3257.000	180.000	94.000	12.000	5.500	303.000
126	-288.300	3678.000	399.000	485.000	43.000	9.900	1074.000
127	1173.000	3426.000	202.000	214.000	32.000	8.500	502.000
128	-374.200	2823.000	216.000	62.000	13.000	4.400	337.000
129	-1527.000	3273.000	164.000	136.000	16.000	14.300	719.000
130	326.000	4110.000	600.000	373.000	74.000	8.000	1172.000
131	-542.000	4025.000	199.300	145.000	14.400	12.000	596.000
132	-745.000	2529.000	151.400	117.000	12.000	7.100	427.000
133	-409.600	2442.000	191.000	113.000	19.000	5.300	413.000
134	-892.200	1917.000	98.000	64.000	9.000	4.000	174.000
135	-2253.000	6344.000	359.000	321.000	62.000	22.800	1292.000
136	-563.700	4248.000	343.000	260.000	42.000	10.800	868.900
137	-101.000	4962.000	523.000	537.000	66.000	11.300	1190.000
138	37.000	4725.000	148.000	115.000	17.000	11.900	371.200
139	-393.300	5373.000	325.000	135.000	22.000	7.600	462.000
140	-1050.000	4004.000	189.000	187.000	43.000	11.600	544.000
141	-478.000	1709.000	165.000	132.000	21.000	4.800	465.000

Продолжение таблицы 4.

1	2	3	4	5	6	7	8
142	-1279.100	3469.000	337.000	375.000	74.000	9.000	871.300
143	-2034.000	5871.000	495.000	353.000	92.000	14.300	1206.000
144	-453.000	5507.000	296.000	281.000	36.000	17.200	925.000
145	0.000	6342.000	486.000	261.000	52.000	24.100	1841.000
146	-2033.000	2908.000	289.000	254.000	28.000	9.500	941.000
147	-1738.000	4906.000	169.200	149.000	27.900	13.400	464.000
148	-1900.000	5603.000	246.600	181.000	16.300	14.700	648.000
149	-1437.000	5168.000	151.000	96.000	8.000	10.700	359.000
150	-1978.000	3354.000	442.000	400.000	59.000	7.500	1098.000
151	-3161.000	3641.000	260.400	201.000	37.300	9.900	711.000
152	-1369.000	1094.000	38.000	1.200	3.200	3.300	114.000
153	-531.000	3714.000	211.600	189.000	20.800	10.900	622.000
154	-884.300	2866.000	237.000	146.000	30.000	6.800	566.000
155	-1955.200	3634.000	334.000	361.000	59.000	10.100	925.000
156	-1004.000	4948.000	230.000	144.000	32.000	11.800	548.000
157	-1855.000	3483.000	109.000	90.000	16.000	7.600	237.000
158	-1338.000	10282.000	499.000	764.000	51.000	30.200	1466.000
159	-878.000	4395.000	275.000	208.000	20.000	7.200	447.000
160	-1078.000	4367.000	216.000	176.000	25.000	11.500	567.000
161	-666.800	3988.000	364.000	213.000	35.000	10.300	943.000
162	-277.600	1744.000	202.000	222.000	31.000	5.700	665.000
163	-350.200	2549.000	246.000	161.000	34.000	6.300	610.000
164	-412.000	4331.000	416.000	253.000	64.000	10.900	1045.000
165	-573.000	5163.000	407.000	285.000	43.000	12.900	1017.000
166	-197.000	5110.000	82.000	57.000	11.000	1.100	174.000
167	-1842.000	4247.000	233.000	189.000	28.000	12.800	757.000
168	-666.000	4222.000	284.700	198.000	44.300	14.000	944.000
169	892.800	6425.000	944.000	1390.000	78.000	13.200	1936.000
170	-1597.000	3396.000	162.000	146.000	23.000	10.600	506.000