

Ю. М. ЖУЧЕНКО

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
В БИОЛОГИИ И ХИМИИ**

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

**Гомель
УО «ГГУ им. Ф. Скорины»
2010**

**Министерство образования Республики Беларусь
Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»**

Ю. М. ЖУЧЕНКО

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
В БИОЛОГИИ И ХИМИИ**

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Рекомендовано УМО вузов Республики Беларусь по образованию в области информатики и радиоэлектроники в качестве учебно-методического пособия для студентов учреждений, обеспечивающих получение высшего образования по специальности 1-31 01 01 “Биология”

**Гомель
УО «ГГУ им. Ф. Скорины»
2010**

УДК 57; 004.9+54; 004.9 (075.8)

ББК 28с51+24с51я73

Ж 834

Рецензенты:

С. В. Буга профессор, доктор биол. наук,
заведующий кафедрой зоологии Белорусского государственного
университета;

А. К. Баев доктор хим. наук, профессор кафедры биохимии и
биофизики учреждения образования «Международный
государственный экологический университет им. А. Д. Сахарова

Рекомендовано к изданию научно-методическим советом
учреждения образования «Гомельский государственный
университет имени Франциска Скорины»

Жученко, Ю. М.

Ж 834 Информационные технологии в биологии и химии:
лабораторный практикум для студентов вузов по специальности 1-
31 01 01 «Биология» / Ю. М. Жученко; М-во образования РБ,
Гомельский гос. ун-т им. Ф. Скорины. – Гомель: ГГУ
им. Ф. Скорины, 2010.– 148 с.

ISBN

Целью лабораторного практикума является оказание помощи
студентам в усвоении основ курса информационных технологий в
биологии и химии, в частности математической статистики,
обработки результатов экспериментов с применением
возможностей персональных компьютеров.

Лабораторный практикум адресован студентам специальности 1-
31 01 01 «Биология».

УДК 57; 004.9+54; 004.9 (075.8)

ББК 28с51+24с51я73

ISBN

© Жученко Ю. М., 2010

© УО «Гомельский государственный
университет им Ф. Скорины», 2010

СОДЕРЖАНИЕ

Введение	Ошибка! Закладка не определена.
Лабораторная работа 1	
Дискретные распределения	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 2	
Основные модели теоретических распределений (Statistica 6)	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 3	
Распределение параметров выборки	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 4	
Разнообразие значений признака (MS Excel)	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 5	
Разнообразие значений признака (Statistica 6)	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 6	
Парная корреляция (MS Excel) ...	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 7	
Парная корреляция (Statistica 6) .	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 8	
Частная и множественная линейные корреляции и регрессия (Statistica 6)	Ошибка! Закладка не определена.
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.

Лабораторная работа 9	
Криволинейная корреляция и регрессия (MS Excel)	Ошибка! Закладка не о
Краткие теоретические сведения ...	Ошибка! Закладка не определена.
Задания для выполнения	Ошибка! Закладка не определена.
Лабораторная работа 10	
Криволинейная корреляция и регрессия (Statistica 6).....	6
Краткие теоретические сведения	6
Задания для выполнения	12
Лабораторная работа 11	
Однофакторный дисперсионный анализ (однофакторный комплекс в MS Excel).....	13
Краткие теоретические сведения	13
Задания для выполнения	16
Лабораторная работа 12	
Однофакторный дисперсионный анализ (однофакторный комплекс в Statistica 6).....	17
Краткие теоретические сведения	17
Задания для выполнения	23
Лабораторная работа 13	
Многофакторный дисперсионный анализ (двухфакторный комплекс в MS Excel).....	24
Краткие теоретические сведения	24
Задания для выполнения	28
Лабораторная работа 14	
Многофакторный дисперсионный анализ (многофакторный комплекс в Statistica 6).....	29
Краткие теоретические сведения	29
Задания для выполнения	37
Лабораторная работа 15	
Классификация (дискриминантный анализ в Statistica 6)	38
Краткие теоретические сведения	38
Задания для выполнения	47
Лабораторная работа 16	
Классификация (кластерный анализ в Statistica 6).....	48
Краткие теоретические сведения	48
Задания для выполнения	64
Литература	66

Приложение А (обязательное).....	67
Приложение Б (обязательное)	68
Приложение В (обязательное).....	69
Приложение Г (обязательное)	70
Приложение Д (обязательное).....	71
Приложение Е (обязательное).....	73
Приложение Ж (обязательное).....	75
Приложение И (обязательное).....	76

Лабораторная работа 1

Криволинейная корреляция и регрессия (Statistica 6)

Цель работы: научиться выполнять корреляционный и регрессионный анализ в случае предполагаемой криволинейной функциональной связи в программном продукте **Statistica 6**.

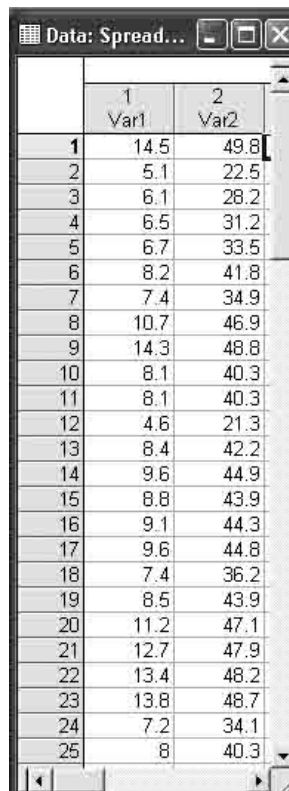
Краткие теоретические сведения

Краткие теоретические сведения изложены в соответствующем разделе лабораторной работы 9.

Проведение анализа

Условия задачи такие же, как в соответствующем разделе лабораторной работы 9.

Ввести исходные данные, как показано на рисунке 10.1.



	1 Var1	2 Var2
1	14.5	49.8
2	5.1	22.5
3	6.1	28.2
4	6.5	31.2
5	6.7	33.5
6	8.2	41.8
7	7.4	34.9
8	10.7	46.9
9	14.3	48.8
10	8.1	40.3
11	8.1	40.3
12	4.6	21.3
13	8.4	42.2
14	9.6	44.9
15	8.8	43.9
16	9.1	44.3
17	9.6	44.8
18	7.4	36.2
19	8.5	43.9
20	11.2	47.1
21	12.7	47.9
22	13.4	48.2
23	13.8	48.7
24	7.2	34.1
25	8	40.3

Рисунок 10.1 – Исходные данные
Var1 – независимая переменная – X ;
Var2 – зависимая переменная – Y .

Проведем анализ в модуле **Nonlinear estimation** (Нелинейная оценка).
Шаг 1. Из Переключателя модулей **Statistica** откройте модуль

Nonlinear estimation (Нелинейная оценка). Высветите название модуля и далее щелкните мышью по названию модуля **Nonlinear estimation** (рисунок 10.2).

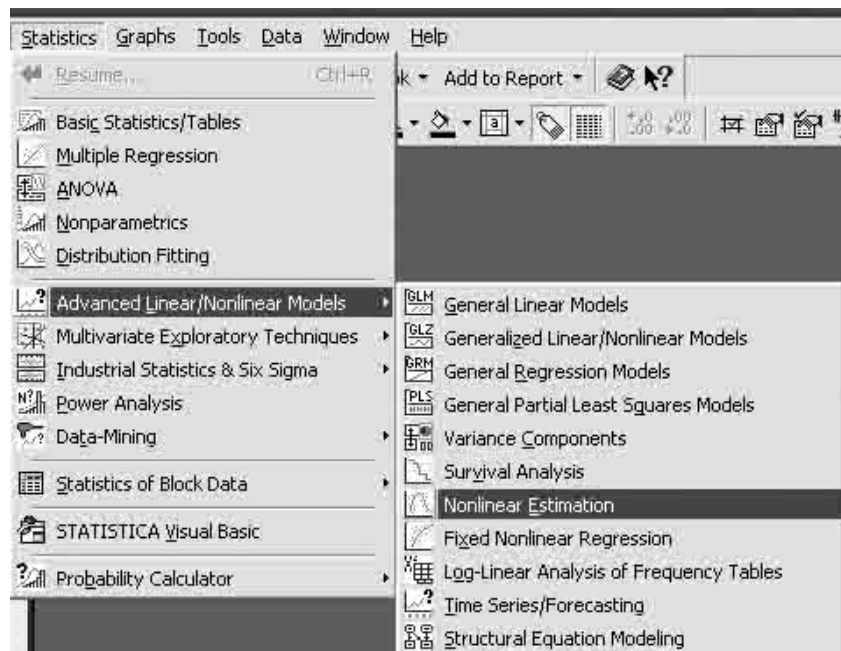


Рисунок 10.2 – Запуск модуля Nonlinear estimation

Шаг 2. На экране появится стартовая панель модуля. Выберите опцию **User specified regression, least squares** (Метод наименьших квадратов) и далее щелкните мышью по названию модуля (рисунок 10.3).

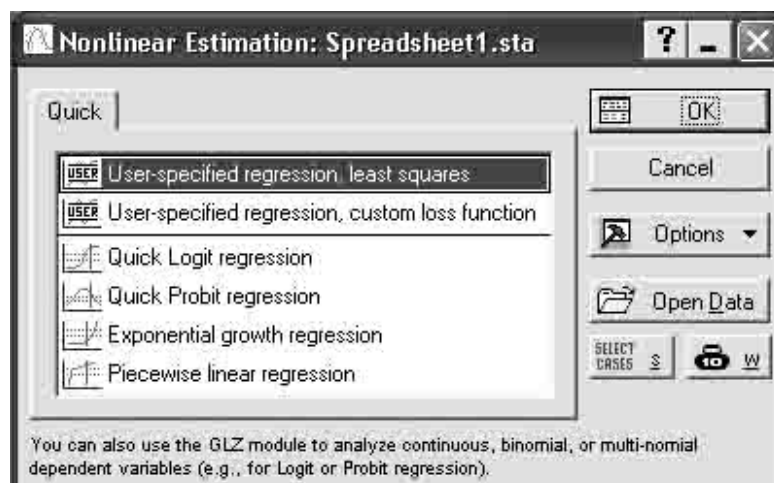


Рисунок 10.3 – Стартовая панель модуля Nonlinear estimation

Шаг 3. В появившемся окне щелкните мышью по кнопке **Function of estimated** (Предполагаемая функция) (рисунок 10.4).

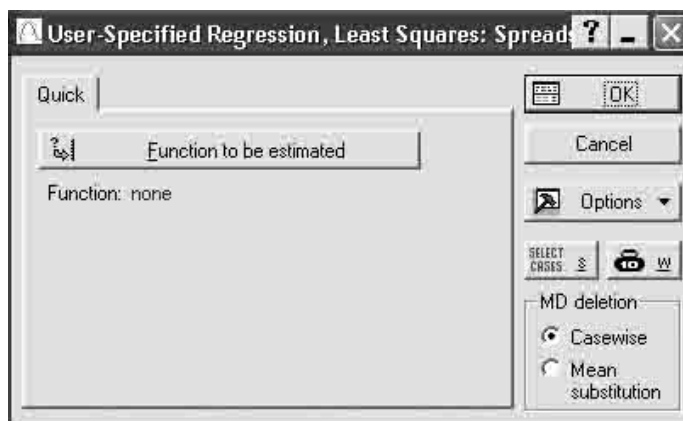


Рисунок 10.4 – Панель ввода функции

Шаг 4. В окне с клавиатуры введите предполагаемую функцию. В отличие от подобной операции в табличном редакторе **MS Excel** в **Statistica 6** вы можете ввести любую формулу, связывающую зависимую и независимую переменные.

В данном случае предполагается, что наиболее подходящей функцией является полином второй степени типа:

$Y = a + b \cdot X + c \cdot X^2$ или в конкретном случае в соответствии с таблицей исходных данных таблицы 10.1:

$$Var2 = a + b * Var1 + c * Var1 ** 2 \text{ (рисунок 10.5)}$$

В нижней части рисунка приведен перечень алгебраических и функциональных символов, которые воспринимаются программой.

Нажмите **ОК**. Затем еще раз **ОК**.

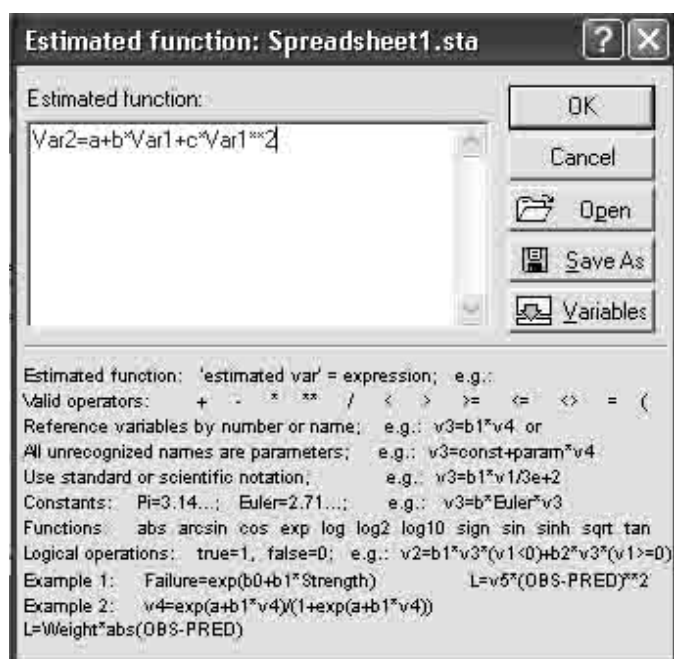


Рисунок 10.5 – Ввод функции

Верхняя часть окна информирует о модели, методе, количестве взятых в анализ пар. В середине окна выберите метод аппроксимации. Например: **Gauss–Newton** (рисунок 10.6). Нажмите **OK**.

Шаг 5. В верхней части появившегося окна результатов (рисунок 10.7) показаны значения корреляционного отношения и его квадрата, 0,99 и 0,98. Это указывает на сильную корреляционную связь между переменными.

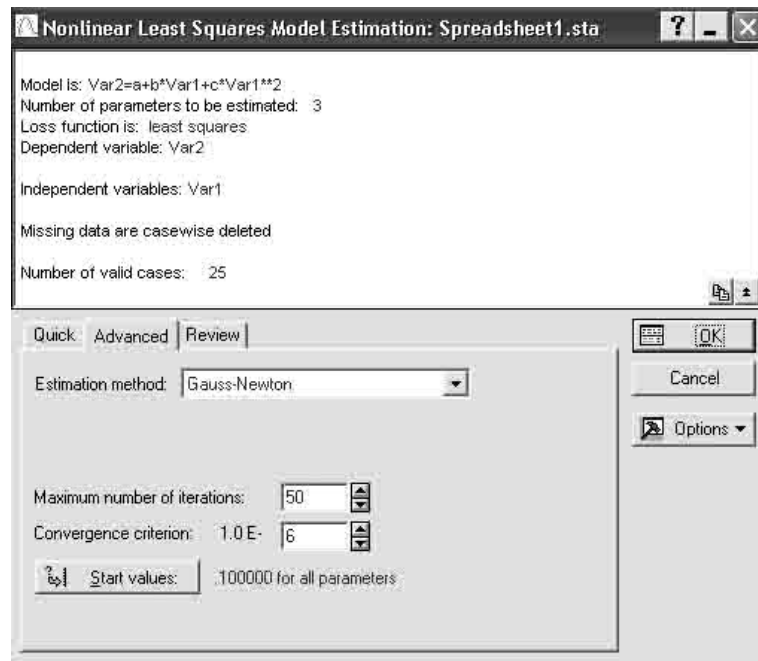


Рисунок 10.6 – Панель пуска аппроксимации

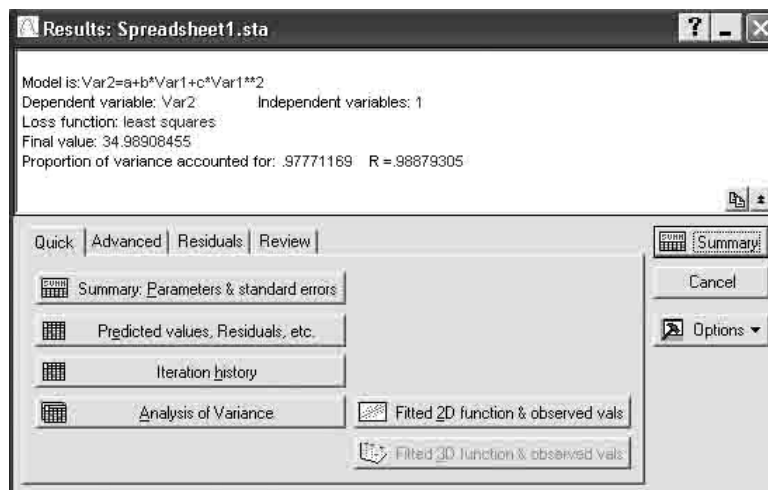


Рисунок 10.7 – Окно результатов

Шаг 6. В окне результатов щелкните мышью по кнопке **Summary Parameters & standard errors** (Итоговые параметры и стандартные ошибки). Полученные результаты (рисунок 10.8) подкрашены красным

цветом, что свидетельствует о достоверности аппроксимации функцией:

$$Y = -0,45 \cdot X^2 + 11,46 \cdot X - 23,41$$

В столбце **Estimate** (Оценка) показаны значения коэффициентов: a , b , c . Далее указаны стандартные ошибки, **t-критерий** при 22 степенях свободы, уровень значимости меньше 0,05, верхний и нижний пределы достоверности.

Model is: Var2=a+b*Var1+c*Var1**2 (Spreadsheet1.sta)
 Dep. Var.: Var2
 Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 22	p-level	Lo. Conf Limit	Up. Conf Limit
a	-23.4071	3.010055	-7.7763	0.000000	-29.6496	-17.1646
b	11.4636	0.641511	17.8697	0.000000	10.1332	12.7940
c	-0.4507	0.032066	-14.0558	0.000000	-0.5172	-0.3842

Рисунок 10.8 – Результаты аппроксимации

Шаг 6. Щелкните мышью по кнопке **Fitted 2D function & observed vals** (Подогнанная функция). На рисунке 10.9 вы увидите графическую интерпретацию корреляционной связи исходных массивов в виде заданной функции: $Y = -0,45 \cdot X^2 + 11,46 \cdot X - 23,41$

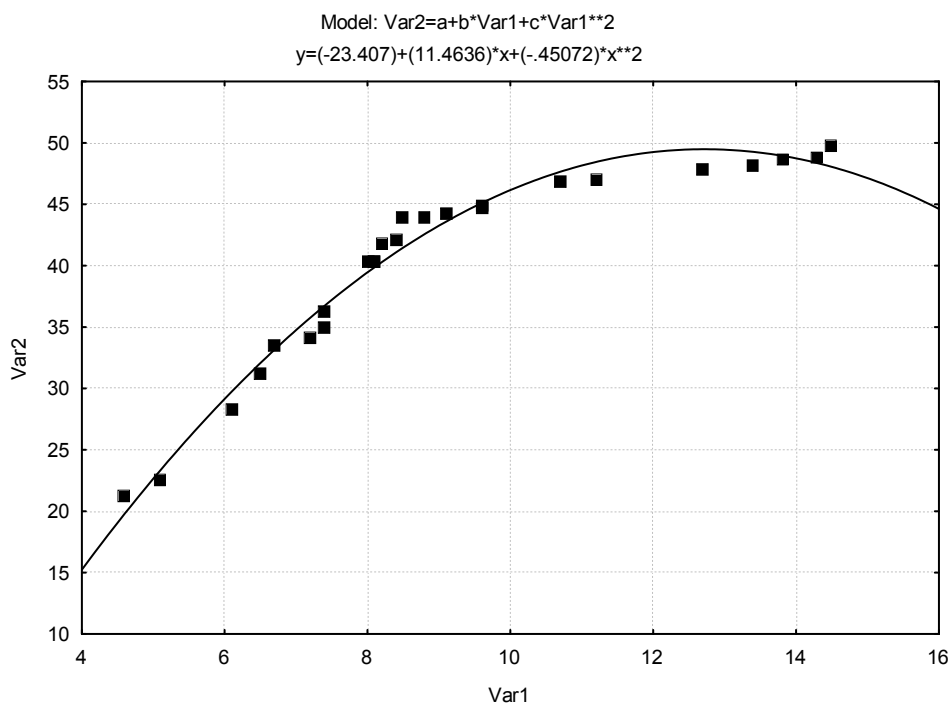
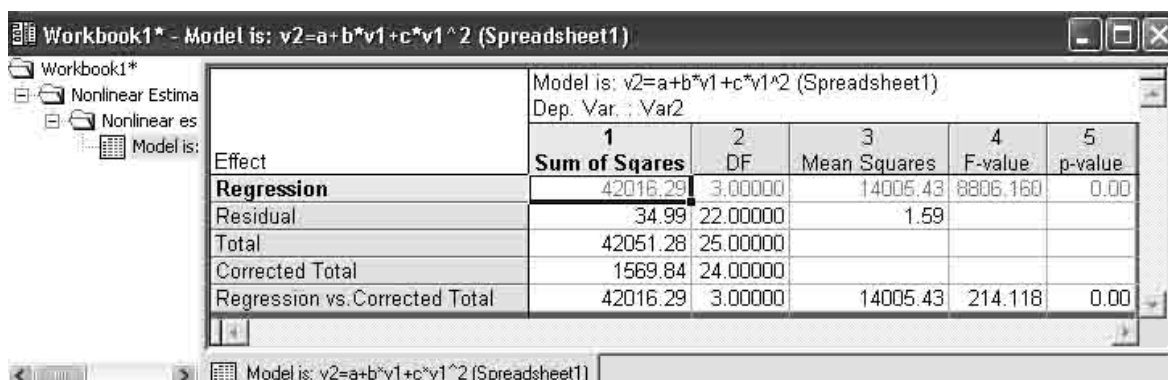


Рисунок 10.9 – Графическая интерпретация корреляционной связи

Шаг 7. В окне результатов (рисунок 10.7) в режиме **Quick** нажмите

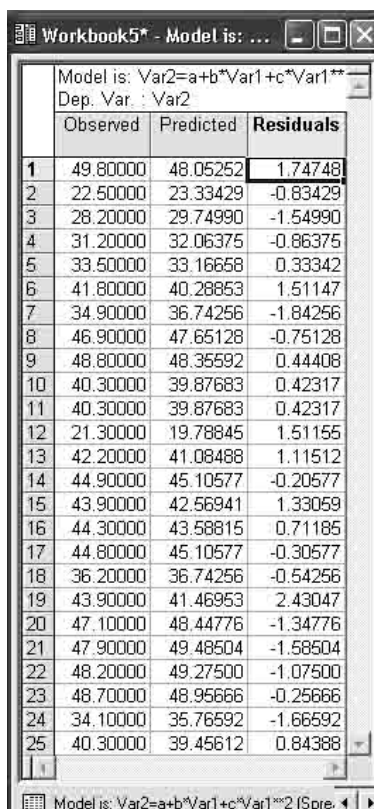
кнопку **Analysis of Variance** (Анализ вариантов). Результат выполненной операции представлен на рисунке 10.10, который свидетельствует о достоверности регрессии ($F = 8806,2$ при $p < 0,00..$).



Effect	1 Sum of Squares	2 DF	3 Mean Squares	4 F-value	5 p-value
Regression	42016.29	3.00000	14005.43	8806.160	0.00
Residual	34.99	22.00000	1.59		
Total	42051.28	25.00000			
Corrected Total	1569.84	24.00000			
Regression vs. Corrected Total	42016.29	3.00000	14005.43	214.118	0.00

Рисунок 10.10 – Результат анализа вариантов

Шаг 8. В окне рисунка 10.7 перейдите в режим просмотра результатов **Residuals** (Остатки). Щелкните мышью по кнопке **Observed, predicted, residual vals** (Наблюдаемый, предсказанный, остаточный). Результаты выполненной операции представлены на рисунке 10.11.



	Observed	Predicted	Residuals
1	49.80000	48.05252	1.74748
2	22.50000	23.33429	-0.83429
3	28.20000	29.74990	-1.54990
4	31.20000	32.06375	-0.86375
5	33.50000	33.16658	0.33342
6	41.80000	40.28853	1.51147
7	34.90000	36.74256	-1.84256
8	46.90000	47.65128	-0.75128
9	48.80000	48.35592	0.44408
10	40.30000	39.87683	0.42317
11	40.30000	39.87683	0.42317
12	21.30000	19.78845	1.51155
13	42.20000	41.08498	1.11512
14	44.90000	45.10577	-0.20577
15	43.90000	42.56941	1.33059
16	44.30000	43.58815	0.71185
17	44.80000	45.10577	-0.30577
18	36.20000	36.74256	-0.54256
19	43.90000	41.46953	2.43047
20	47.10000	48.44776	-1.34776
21	47.90000	49.48504	-1.58504
22	48.20000	49.27500	-1.07500
23	48.70000	48.95666	-0.25666
24	34.10000	35.76592	-1.66592
25	40.30000	39.45612	0.84388

Рисунок 10.11 – Наблюдаемые и аппроксимированные

значения функции

Шаг 9. Для оценки адекватности модели щелкните мышкой по кнопке **Observed vs. Predicted** (Наблюдаемые против предсказанных) (рисунок 10.12).

Из рисунка видно, массивы наблюдаемых и предсказанных значений описываются линейной функцией $Y = k \cdot X$. При этом $k = 1$ и коэффициент парной корреляции близок к 1.

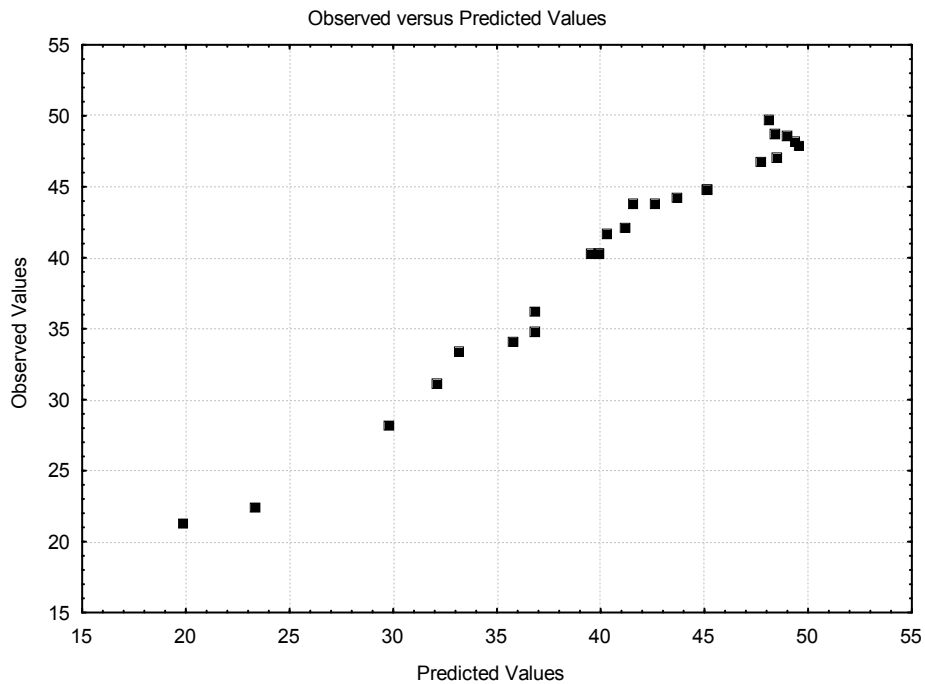


Рисунок 10.12 – Визуализация результатов анализа

К заключению, сделанному в предыдущем разделе 9, следует добавить, что корреляция и регрессия достоверны, так как $F = 8806,2$ (рисунок 10.10) и $t_a = 7,8$, $t_b = 17,9$ и $t_c = 14,1$ (рисунок 10.8), что существенно выше критических значений при $p < 0,00..$

Задания для выполнения

- 1 Введите в таблицу **MS Excel** исходные данные из Приложения Г (таблица Г1).
- 2 Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе.
- 3 Получите результат и сделайте заключение.

Лабораторная работа 2

Однофакторный дисперсионный анализ (однофакторный комплекс в MS Excel)

Цель работы: научиться выполнять однофакторный дисперсионный анализ в программном продукте MS Excel.

Краткие теоретические сведения

При дисперсионном анализе одновременно обрабатывают данные нескольких выборок (вариантов), составляющих единый статистический комплекс, оформленный в виде специальной рабочей таблицы. Структура статистического комплекса и его последующий анализ определяются схемой и методикой эксперимента.

Сущностью дисперсионного анализа является расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части – компоненты, соответствующие структуре эксперимента, и оценка значимости действия и взаимодействия изучаемых факторов по **F–критерию**.

Если обрабатывают однофакторные статистические комплексы, состоящие из нескольких независимых выборок, например l – вариантов в вегетационном опыте, то общая изменчивость результативного признака, измеряемая общей суммой квадратов C_Y , расчленяется на два компонента: варьирование между выборками (вариантами) C_V и внутри выборок C_Z . Следовательно, в общей форме изменчивость признака может быть представлена выражением:

$$C_Y = C_V + C_Z.$$

Здесь вариация между выборками (вариантами) представляет ту часть общей дисперсии, которая обусловлена действием изучаемых факторов, а дисперсия внутри выборок характеризует случайное варьирование изучаемого признака, т. е. ошибку эксперимента.

Общее число степеней свободы $(N - l)$ также расчленяется на две части – степени свободы для вариантов $(l - 1)$ и для случайного варьирования $(N - l)$.

Суммы квадратов отклонений по данным эксперимента – статистического комплекса с l -вариантами – обычно находят в такой последовательности. При общем числе наблюдений N определяются суммы по вариантам V и общая сумма всех наблюдений ΣX . Затем вычисляются:

- корректирующий фактор (поправка) $C = \frac{(\sum X)^2}{N}$;
- общая сумма квадратов $C_Y = \sum X^2 - C$;
- сумма квадратов для вариантов $C_V = \sum \frac{V^2}{l} - C$;
- сумма квадратов для ошибки (остаток) $C_Z = C_Y - C_V$.

Две последние суммы квадратов C_Y и C_Z делятся на соответствующие им степени свободы, т. е. приводятся к сравниваемому виду – одной степени свободы вариации. В результате получают два средних квадрата:

- вариантов $\sigma_V^2 = \frac{C_V}{l-1}$;
- ошибки $\sigma_Z^2 = \frac{C_Z}{N-l-1}$.

Эти средние квадраты и используют в дисперсионном анализе для оценки значимости действия изучаемых факторов. Оценка проводится путем сравнения дисперсии вариантов σ_V^2 с дисперсией ошибки σ_Z^2 по критерию Фишера:

$$F = \frac{\sigma_V^2}{\sigma_Z^2}.$$

Таким образом, за базу – единицу сравнения принимают средний квадрат случайной дисперсии, которая определяет случайную ошибку эксперимента. При этом проверяемой нулевой гипотезой служит предположение: все выборочные средние являются оценками одной генеральной средней и, следовательно, различия между ними несущественны. Если $F_{\text{факт}} < F_{\text{теор}}$, то нулевая гипотеза $H_0: d = 0$ не отвергается; между всеми выборочными средними нет существенных различий и на этом проверка заканчивается. Нулевая гипотеза отвергается, когда $F_{\text{факт}} \geq F_{\text{теор}}$.

В этом случае дополнительно проводят оценку, между какими средними имеются значимые различия.

Проведение анализа

При испытании снотворного средства изучалось его действие для двух факторов: однократная доза (группа *a*) и двойная доза (группа *b*). В каждой группе находилось по 10 подопытных объектов. Результативным признаком является продолжительность сна (в часах).

Исходные данные для выполнения однофакторного дисперсионного анализа представлены в таблице 11.1.

Таблица 11.1 – Исходные данные

a	b
2	6
3	9
1	9
4	7
3	6
6	6
3	3
5	6
6	5
4	6

Шаг 1. Откройте модуль **Анализ данных** и выберите опцию **Однофакторный дисперсионный анализ**, после чего щелкните мышкой **ОК**.

Шаг 2. В появившемся окне выполнить операции и установки, как показано на рисунке 11.1.

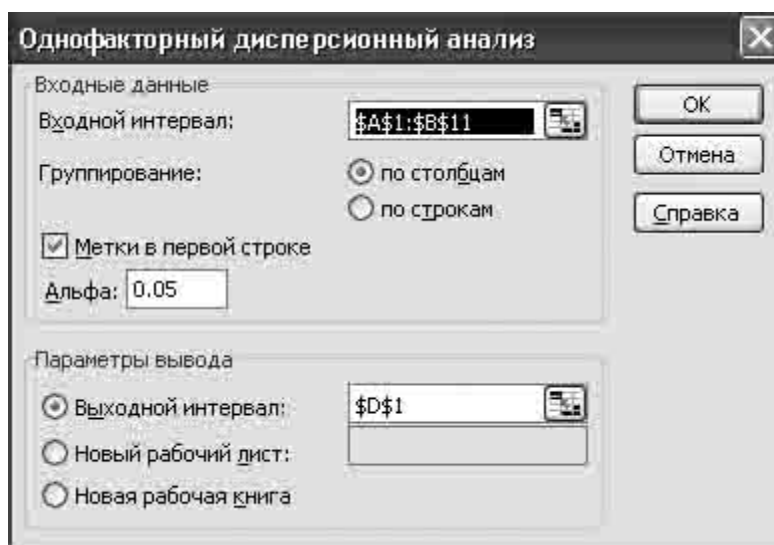


Рисунок 11.1 – Стартовая панель

Шаг 3. Щелкните мышкой **ОК**. Результат обработки появится в указанном поле (выходной интервал \$D\$1, таблицы 11.2, 11.3).

Таблица 11.2 – Статистические параметры

Группы	Счет	Сумма	Среднее	Дисперсия
a	10	37	3,7	2,68
b	10	63	6,3	3,12

Таблица 11.3 – Результаты дисперсионного анализа

Источник вариации	SS	df	MS	F	P-значение	F критическое	Сила влияния, %
Между группами	33,8	1	33,8	11,7	0,003	4,4	39,3
Внутри групп	52,2	18	2,9				60,7
Итого	86	19					100

В рассмотренном примере эмпирический **F-критерий** (критерий Фишера) показывает, что различие между средними статистически значимо (значимо на уровне $p=0,003$, то есть меньше, чем критическое значение $0,05$). Поскольку различие между средними значениями значимо, нулевая гипотеза о равенстве средних отвергается и принимается альтернативная гипотеза о существовании различия между средними.

Таким образом, действие снотворного в однократной и двойной дозах статистически достоверно отличаются.

Задания для выполнения

1 Введите в таблицу **MS Excel** исходные данные из Приложения Д (таблица Д1).

2 Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе.

3 Получите результат и сделайте заключение.

Лабораторная работа 3

Однофакторный дисперсионный анализ (однофакторный комплекс в Statistica 6)

Цель работы: научиться выполнять однофакторный дисперсионный анализ в программном продукте **Statistica 6**.

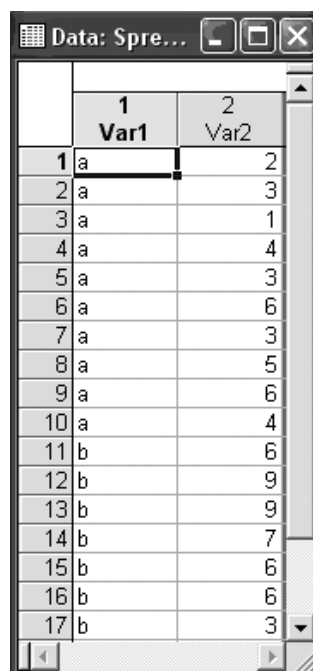
Краткие теоретические сведения

Краткие теоретические сведения изложены в соответствующем разделе лабораторной работы 11.

Проведение анализа

Условия задачи такие же, как в соответствующем разделе лабораторной работы 11.

Введите исходные данные из таблицы 11.1 в созданную таблицу в формате **Statistica 6**, как показано на рисунке 12.1.



	1 Var1	2 Var2
1	a	2
2	a	3
3	a	1
4	a	4
5	a	3
6	a	6
7	a	3
8	a	5
9	a	6
10	a	4
11	b	6
12	b	9
13	b	9
14	b	7
15	b	6
16	b	6
17	b	3

Рисунок 12.1 – Исходные данные
Var1 – факторы;
Var2 – независимая переменная.

Проведем анализ в модуле **ANOVA** (Дисперсионный анализ).

Шаг 1. Из Переключателя модулей **Statistica** откройте модуль **ANOVA**. Высветите название модуля и далее щелкните мышью по названию модуля: **ANOVA** (рисунок 12.2).

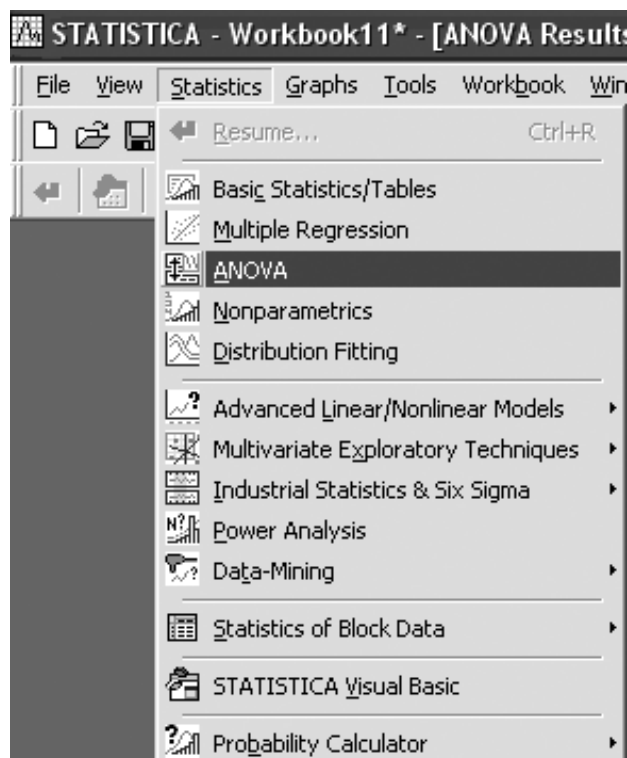


Рисунок 12.2 – Основное меню

Шаг 2. На экране появится стартовая панель модуля (рисунок 12.3). Выполнить установки, как показано на рисунке 12.3.

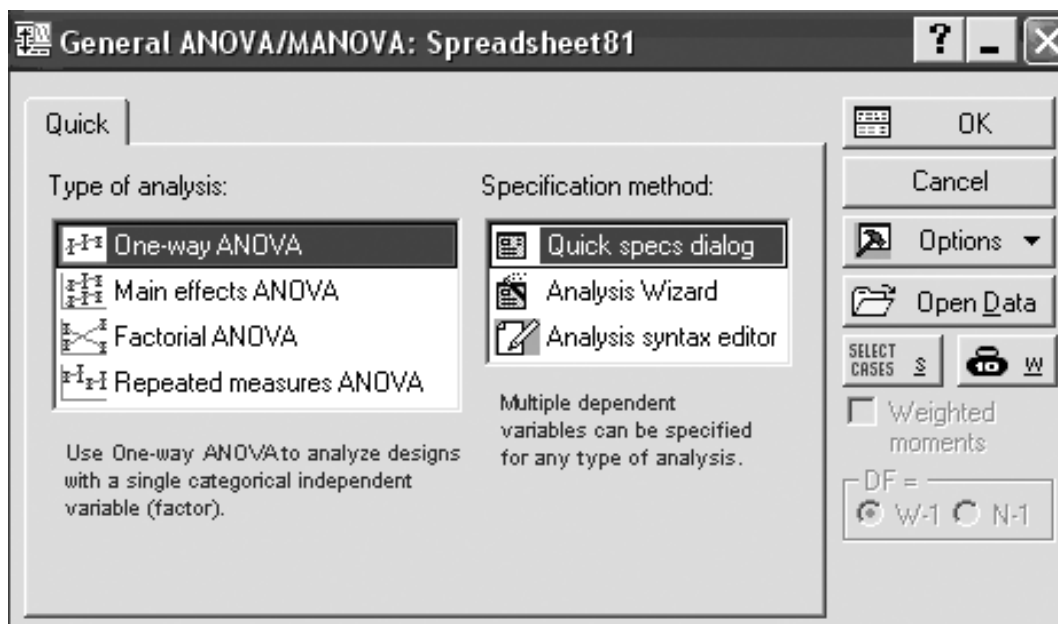


Рисунок 12.3 – Стартовая панель модуля

Шаг 3. После нажатия кнопки **ОК** в появившемся окне выберите переменные для анализа (рисунок 12.4).

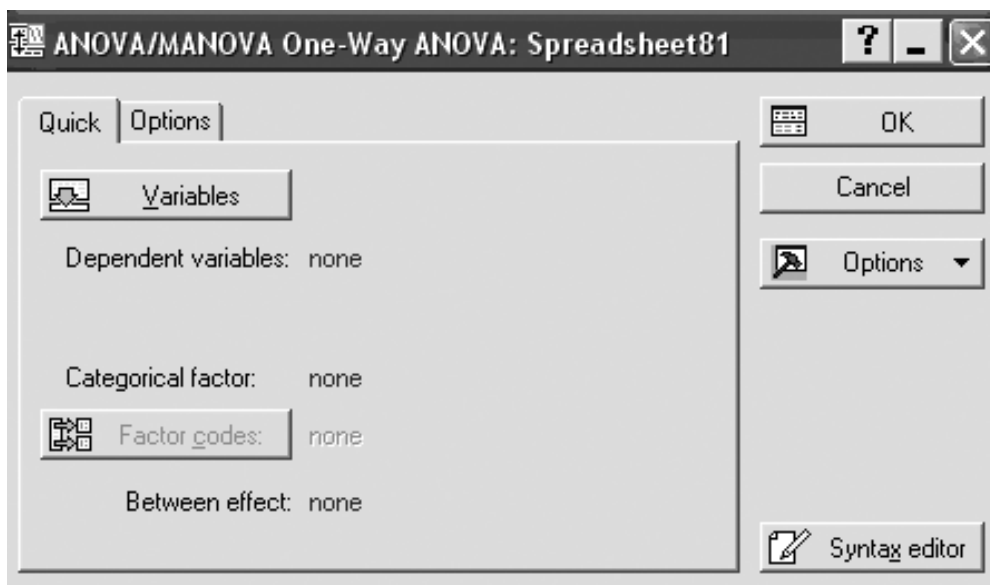


Рисунок 12.4 – Выбор переменных

Выбор переменных осуществляется с помощью кнопки **Variables** (Переменные), находящейся в левом верхнем углу панели. После того как кнопка будет нажата, диалоговое окно **Select dependent variables and categorical predictor (factor)** (Выбрать списки зависимых переменных и факторов) появится на вашем экране (рисунок 12.5).

Шаг 4. В левой части окна имя переменной выберите зависимую переменную, а в правой – фактор. После нажатия кнопки **OK** в появившемся окне выберите **All**, нажав кнопку **Factor codes** (рисунок 12.6.).

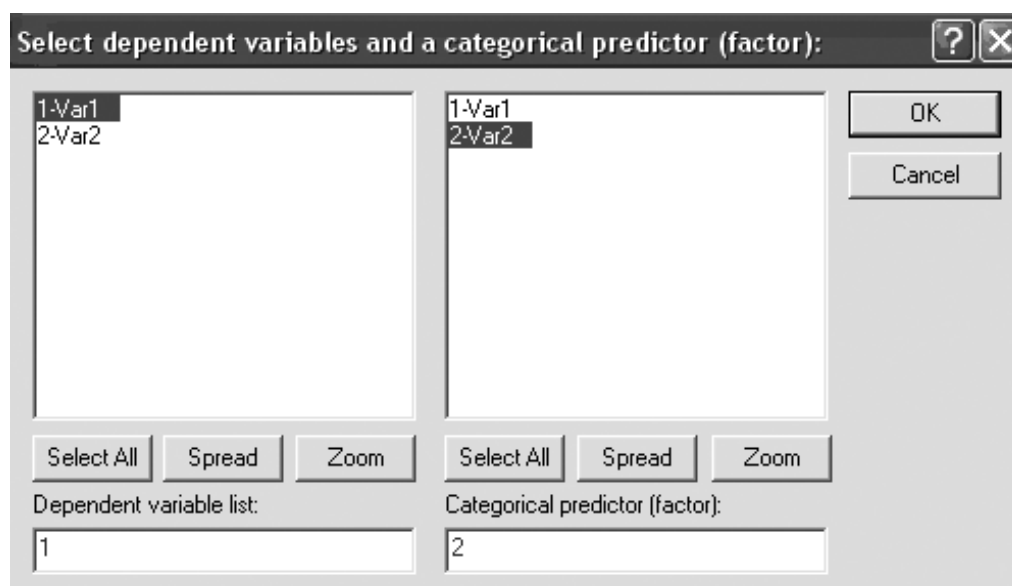


Рисунок 12.5 – Окно выбора переменных для анализа

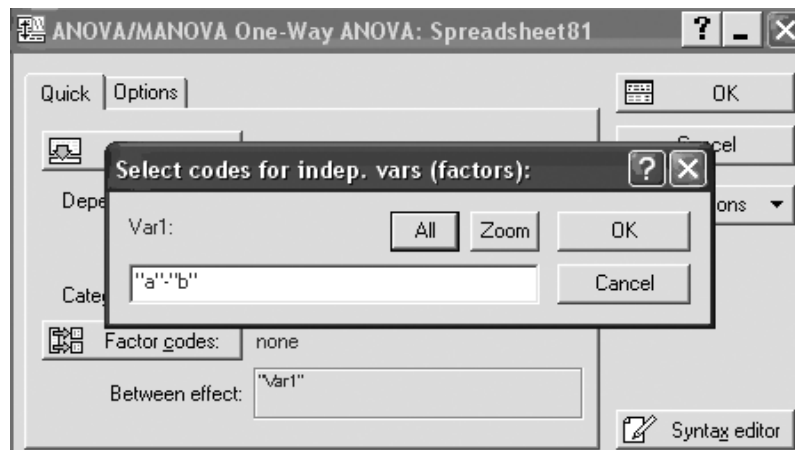


Рисунок 12.6 – Окно выбора факторов

Нажмите кнопку **OK** в правом углу стартовой панели.

Шаг 5. На экране перед вами появится диалоговое окно **Anova Results** (Результаты) (рисунок 12.7). В данном окне выберите **Univariate Results** (Результат дисперсионного анализа). Далее нажмите кнопку **OK**.

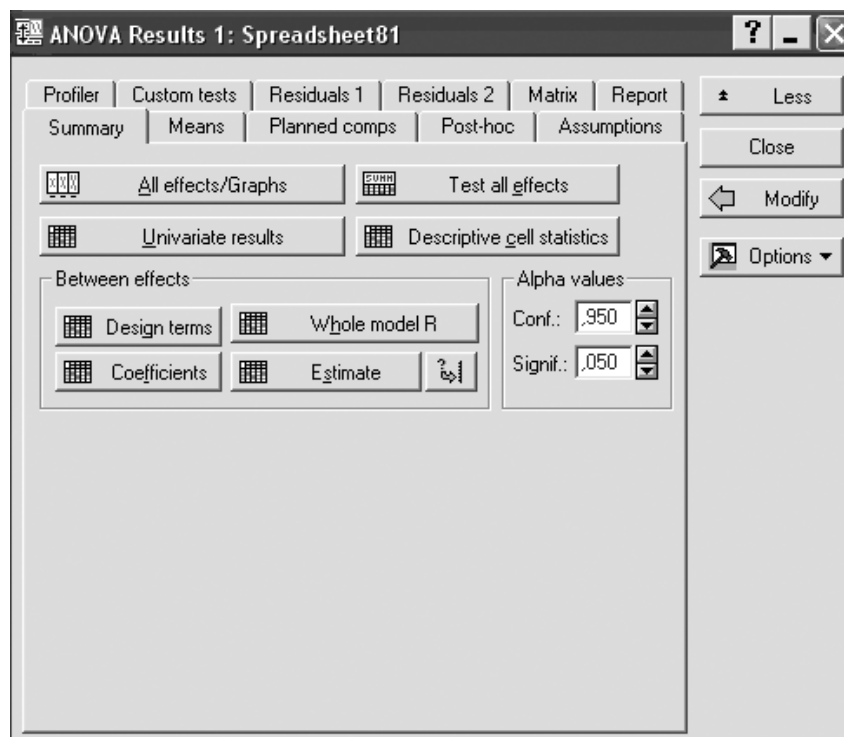


Рисунок 12.7 – Диалоговое окно результатов

В окне результатов рисунка 12.8 представлены результаты дисперсионного анализа:

- между группами – **Var1**;
- внутри групп – **Error**.

Workbook11* - Univariate Results for Each DV (S... [min] [max] [close]

Univariate Results for Each DV (Spreadsheet81)
Sigma-restricted parameterization
Effective hypothesis decomposition

GENERAL Effect	Degr. of Freedom	Var2 SS	Var2 MS	Var2 F	Var2 p
Intercept	1	500,0000	500,0000	172,4138	0,000000
"Var1"	1	33,8000	33,8000	11,6552	0,003094
Error	18	52,2000	2,9000		
Total	19	86,0000			

Univariate Results for Each DV (Spreadsheet81) "Var1"; Unweighte [left] [right]

Рисунок 12.8 – Результаты дисперсионного анализа

Шаг 6. В диалоговом окне результатов (рисунок 12.7) нажмите кнопку **All effects / Grafts**, затем в появившемся окне **ОК**. Результат представлен на рисунке 12.9.

В рассмотренном примере **F-критерий** показывает, что различие между средними статистически значимо (значимо на уровне 0,003, то есть меньше, чем критическое значение 0,05). Поскольку различие между средними значениями значимо, нулевая гипотеза отвергается и принимается альтернативная гипотеза о существовании различия между средними (результат в строке: между группами – **Var1** подсвечивается красным цветом).

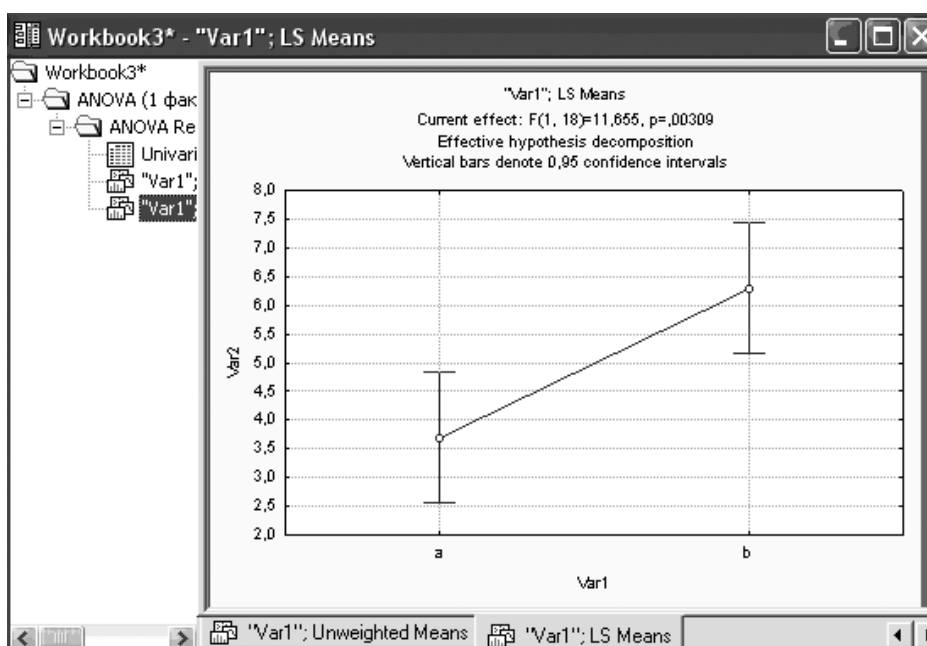


Рисунок 12.9 – Результат анализа

Шаг 7. Из основного меню (рисунок 12.10) выберете модуль **Graphs** (Графика), выберете опцию **Box plots** и щелкните мышкой. Затем выполните установки, как это показано на рисунке 12.11.

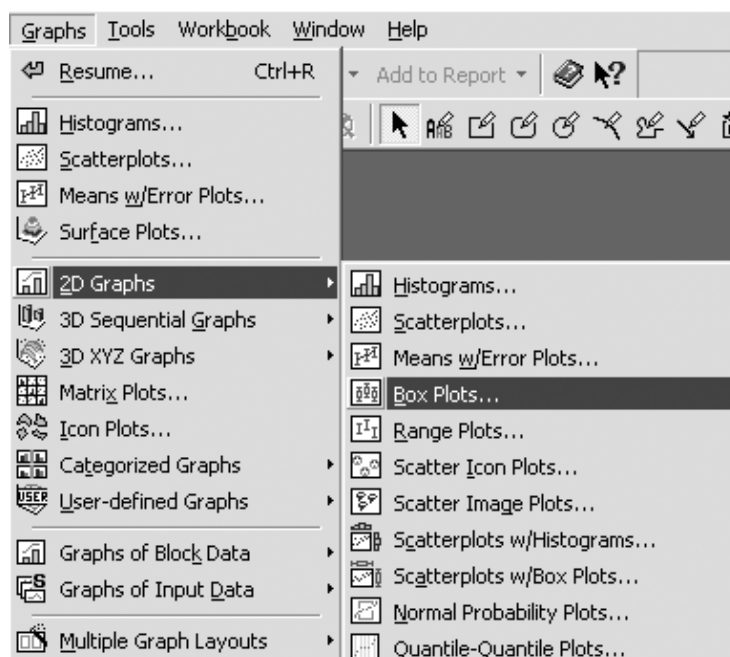


Рисунок 12.10 – Меню запуска

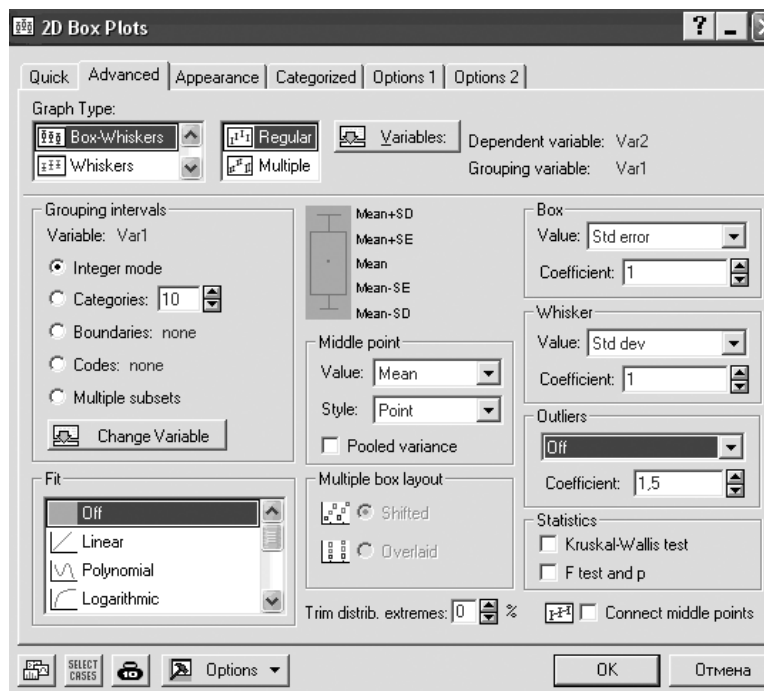


Рисунок 12.11 – Предварительная установка

Нажмите кнопку **ОК** и вы получите визуализацию статистических параметров при воздействии факторов a и b (рисунок 12.12), для

каждого из которых показаны:

- среднее;
- стандартное отклонение;
- стандартная ошибка.

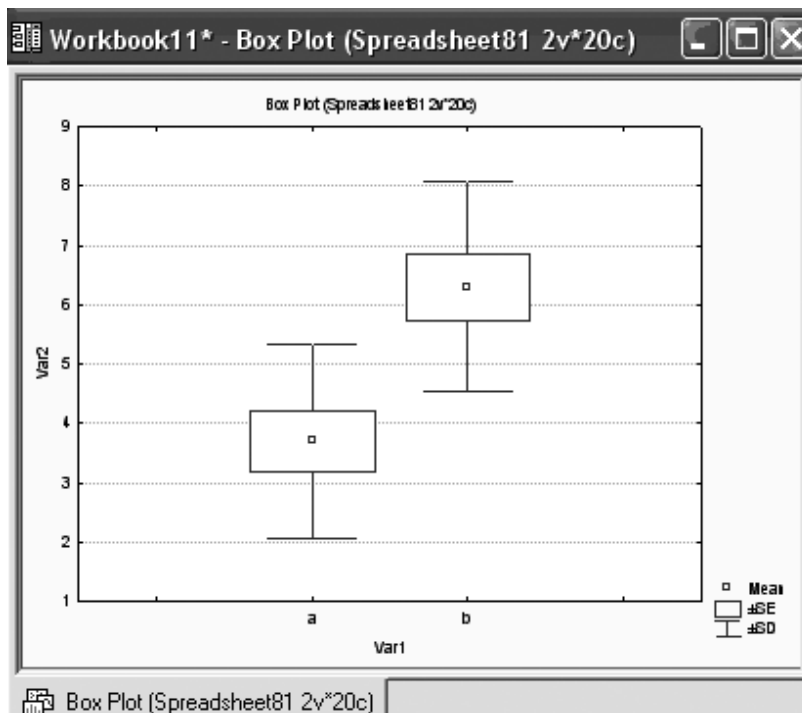


Рисунок 12.12 – График результатов анализа

Результаты, полученные в программном продукте **Statistica 6** точно такие же, как и в табличном редакторе **MS Excel**. Поэтому вывод о действии снотворного аналогичен. Однако в программном продукте **Statistica 6** имеется возможность визуализации результатов, что, несомненно, его делает более привлекательным при представлении результатов исследований.

Задания для выполнения

- 1 Введите в таблицу программного продукта **Statistica 6** исходные данные из Приложения Д (таблица Д1).
- 2 Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе.
- 3 Получите результат и сделайте заключение.

Лабораторная работа 4

Многофакторный дисперсионный анализ (двухфакторный комплекс в MS Excel)

Цель работы: научиться выполнять многофакторный дисперсионный анализ в программном продукте MS Excel.

Краткие теоретические сведения

Многофакторный дисперсионный комплекс – это совокупность исходных наблюдений, позволяющих статистически оценить действие и взаимодействие нескольких изучаемых факторов на изменчивость результативного признака. Эффект взаимодействия составляет ту часть общего варьирования, которая вызвана различным действием одного фактора при разных градациях другого. Специфическое действие сочетаний в эксперименте выявляется тогда, когда при одной градации первого фактора второй действует слабо или угнетающе, а при другой градации он проявляется сильно и стимулирует развитие результативного признака.

Дисперсионный анализ данных многофакторного комплекса проводится в два этапа. Первый этап – разложение общей вариации результативного признака на варьирование вариантов и остаточное: $C_Y = C_V + C_Z$. На втором этапе сумма квадратов отклонения для вариантов разлагается на компоненты, соответствующие источникам варьирования – главные эффекты изучаемых факторов и их взаимодействия.

В двухфакторном опыте:

$$C_V = C_A + C_B + C_{AB}.$$

В трехфакторном:

$$C_V = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}.$$

Дисперсионный анализ двухфакторного анализа по изучению градаций фактора A (число вариантов l_A) и градаций фактора B (число вариантов l_B), проведенного в n повторностях, осуществляется в следующие этапы:

1 Определяются суммы и средние по вариантам, общая сумма и средний урожай по опыту.

2 Вычисляются общая сумма квадратов отклонений, сумма квадратов для вариантов и остатка: $N = l_A \cdot l_B \cdot n$;

$$C = \frac{(\sum X)^2}{N};$$

$$C_Y = \sum X^2 - C;$$

$$C_V = \sum \frac{V^2}{n} - C;$$

$$C_Z = C_Y - C_V.$$

3 Для вычисления сумм квадратов по факторам A , B и взаимодействию AB составляется вспомогательная таблица, в которую записываются суммы по вариантам. Суммируя цифры, находятся суммы A , суммы B и вычисляются суммы квадратов отклонений для главных эффектов и взаимодействия.

Сумма квадратов для фактора A :

$$C_A = \sum \frac{A^2}{l_B \cdot n} - C \text{ при } (l_A - 1) \text{ степенях свободы.}$$

Сумма квадратов для фактора B :

$$C_B = \sum \frac{B^2}{l_A \cdot n} - C \text{ при } (l_B - 1) \text{ степенях свободы.}$$

Сумма квадратов для взаимодействия AB находится по разности:

$$C_{AB} = C_V - C_A - C_B \text{ при } (l_A - 1) \cdot (l_B - 1) \text{ степенях свободы.}$$

При дисперсионном анализе ортогональных комплексов используются аддитивные свойства частных дисперсий (сумм квадратов центральных отклонений).

Если взять отношения частных сумм квадратов к общей:

$$\frac{C_V}{C_Y} + \frac{C_Z}{C_Y} = \frac{C_Y}{C_Y} = 1;$$

$$\frac{C_A}{C_Y} + \frac{C_B}{C_Y} + \frac{C_{AB}}{C_Y} + \frac{C_Z}{C_Y} = \frac{C_Y}{C_Y} = 1.$$

Каждое из этих отношений будет показывать долю участия отдельной частной дисперсии в образовании общей дисперсии. А так как каждая частная сумма квадратов соответствует одному из частных влияний, то отношение частной суммы квадратов к общей измеряет долю данного влияния в общем суммарном статистическом влиянии всех факторов, определяющих развитие данного результативного признака.

Поэтому доля (выраженная в относительных единицах или в процентах) каждой частной дисперсии в общей их сумме может быть

принята за показатель силы влияния, того влияния, которое характеризуется данной частной дисперсией – или одной из факториальных или случайной.

Таким образом, сила влияния фактора (факторов) в дисперсионном анализе измеряется отношением дисперсий частных к общей:

$$\eta_i^2 = \frac{C_i}{C_Y}.$$

Так как этот показатель отражает основной закон разложения общих дисперсий и основное аддитивное свойство частных дисперсий, а также составлен из основных элементов дисперсионного анализа, то отношение одной из факториальных дисперсий (C_V , C_A , C_B , C_{AB}) или случайной дисперсии (C_Z) к общей (C_Y) можно назвать основным показателем силы влияний факторов – организованных и неорганизованных.

Проведение анализа

На 12 опытных делянках проводились экспериментальные работы с посевом кормовых злаковых трав. Факторы a и b отражают объективную ситуацию в процессе проведения опыта (a – освещенность и b – увлажнение) или фактор среды (неорганизованный фактор). Факторы o и p – факторы влияния: o – фоновые, без внесения удобрений, p – с внесением (организованный фактор). Результативным признаком является урожайность. Исходные данные, представленные в таблице 13.1 использовались для выполнения двухфакторного дисперсионного анализа.

Следует отметить, что в проведении анализа в табличном редакторе **MS Excel** количество повторений (факторы o и p) должно быть одинаковым. В данном случае 3. Для выполнения подобного анализа в программном продукте Statistica 6 количество повторений может быть разновеликим.

Таблица 13.1 – Исходные данные

	a	b
o	58	49
o	84	55
o	39	48
p	72	74
p	72	74
p	64	85

Шаг 1. Откройте модуль **Анализ данных** выберите опцию

Двухфакторный дисперсионный анализ с повторениями, после чего щелкните мышкой **ОК**.

Шаг 2. В появившемся окне выполнить операции и установки, как показано на рисунке 13.1. Щелкнуть мышкой **ОК**.

Шаг 3. Результат обработки появится в указанном поле (выходной интервал \$D\$1, таблицы 13.2, 13.3).

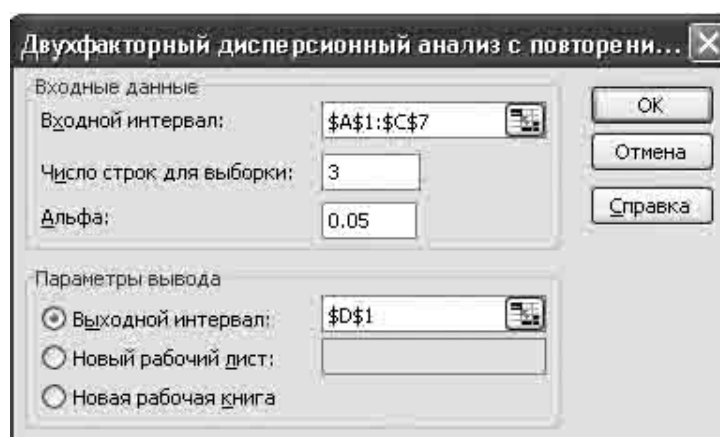


Рисунок 13.1 – Стартовая панель

Таблица 13.2 – Статистические параметры

	a	b	Итого
o			
Счет	3	3	6
Сумма	181	152	333
Среднее	60,33	50,67	55,5
Дисперсия	510,33	14,33	237,9
p			
Счет	3	3	6
Сумма	208	233	441
Среднее	69,33	77,67	73,5
Дисперсия	21,33	40,33	45,5
Итого			
Счет	6	6	
Сумма	389	385	
Среднее	64,83	64,17	
Дисперсия	236,97	240,57	

Таблица 13.3 – Дисперсионный анализ

Источник вариации	SS	df	MS	F	P-значение	F крит.	Сила влияния, %
Выборка (2 фактор)	972	1	972	6,63	0,035	5,32	40,7
Столбцы (1 фактор)	1,33	1	1,33	0,0091	0,93	5,32	0,1
Взаимодействие	243	1	243	1,658	0,23	5,32	10,2

Внутри	1172,67	8	146,58				49,0
Итого	2389	11					100

В рассмотренном примере **F-критерий** показывает, что нулевая гипотеза отвергается и различие между средними статистически значимо за счет влияния второго фактора (значимо на уровне 0,033, что не превышает критического уровня 0,05). Сила влияния этого фактора составляет около 41%. В свою очередь, по первому фактору и взаимодействию обоих факторов нулевая гипотеза о равенстве средних не отвергается, поскольку критерий Фишера меньше табличного значения и уровень значимости $p > 0,05$. Поэтому, в данном случае прибавка к урожаю обусловлена только организованным фактором.

Задания для выполнения

- 1 Введите в таблицу **MS Excel** исходные данные из Приложения Е (таблица Е1).
- 2 Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе.
- 3 Получите результат и сделайте заключение.

Лабораторная работа 5

Многофакторный дисперсионный анализ (многофакторный комплекс в Statistica 6)

Цель работы: научиться выполнять многофакторный дисперсионный анализ в программном продукте **Statistica 6**.

Краткие теоретические сведения

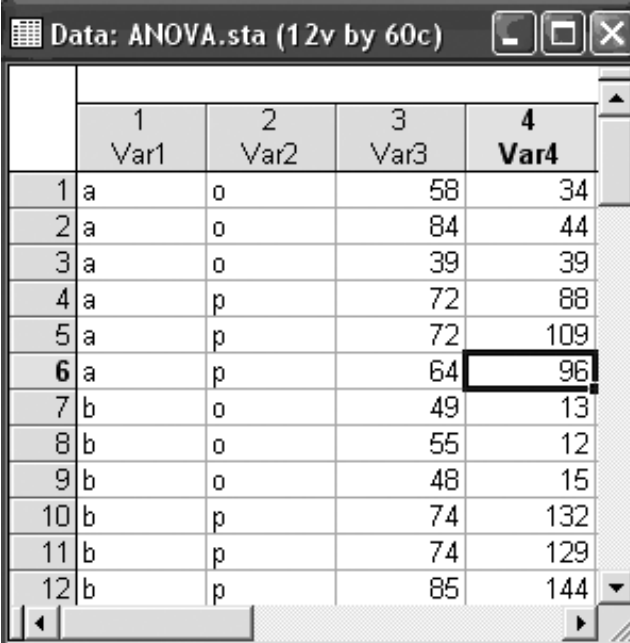
Краткие теоретические сведения изложены в соответствующем разделе лабораторной работы 13.

Проведение анализа

Условия задачи такие же, как в соответствующем разделе лабораторной работы 13.

Введите исходные данные из таблицы 13.1 в созданную таблицу в формате **Statistica 6**, как показано на рисунке 14.1.

Проведем анализ в модуле **ANOVA** (Дисперсионный анализ).



	1 Var1	2 Var2	3 Var3	4 Var4
1	a	o	58	34
2	a	o	84	44
3	a	o	39	39
4	a	p	72	88
5	a	p	72	109
6	a	p	64	96
7	b	o	49	13
8	b	o	55	12
9	b	o	48	15
10	b	p	74	132
11	b	p	74	129
12	b	p	85	144

Рисунок 14.1 – Исходные данные
Var3 – независимая переменная;
Var1, Var2 – факторы.

Шаг 1. Из Переключателя модулей **Statistica** откройте модуль **ANOVA** (Дисперсионный анализ). Высветите название модуля и далее щелкните мышью по названию модуля: **ANOVA** (рисунок 14.2).

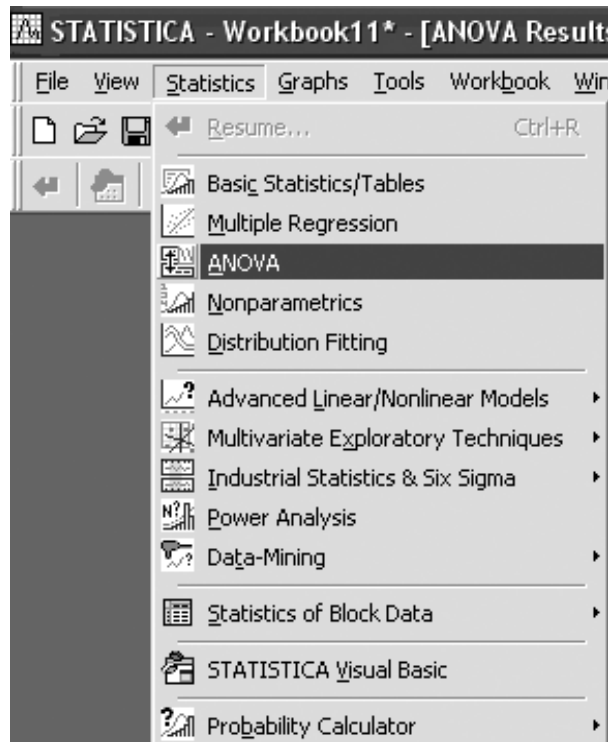


Рисунок 14.2 – Основное меню

Шаг 2. На экране появится стартовая панель модуля (рисунок 14.3). Нажмите **ОК**.

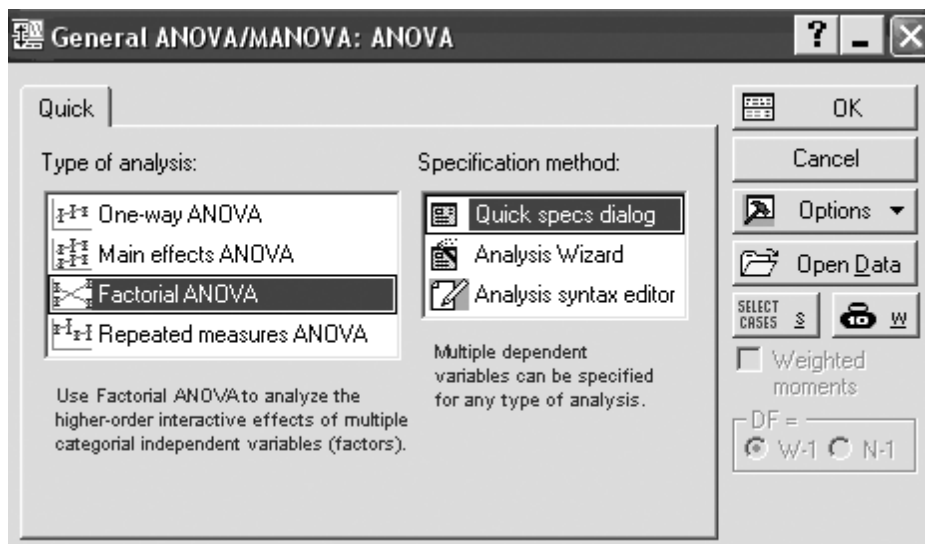


Рисунок 14.3 – Стартовая панель модуля

Шаг 3. После нажатия кнопки **ОК** в появившемся окне выберите переменные для анализа (рисунок 14.4). Выбор переменных осуществляется с помощью кнопки **Variables** (Переменные), находящейся в левом верхнем углу панели. После того как кнопка будет нажата, диалоговое окно **Select dependent variables and categorical**

predictor (factor) (Выбрать списки зависимых переменных и факторов) появится на вашем экране (рисунок 14.5). В левой части окна **имя переменной** выберите зависимую переменную, а в правой – **факторы**.

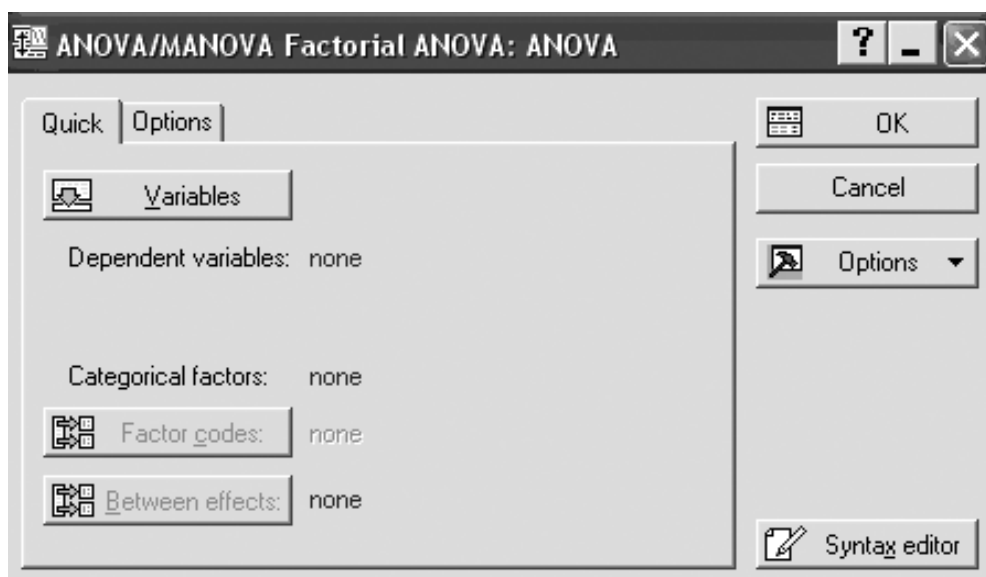


Рисунок 14.4 – Окно выбора переменных

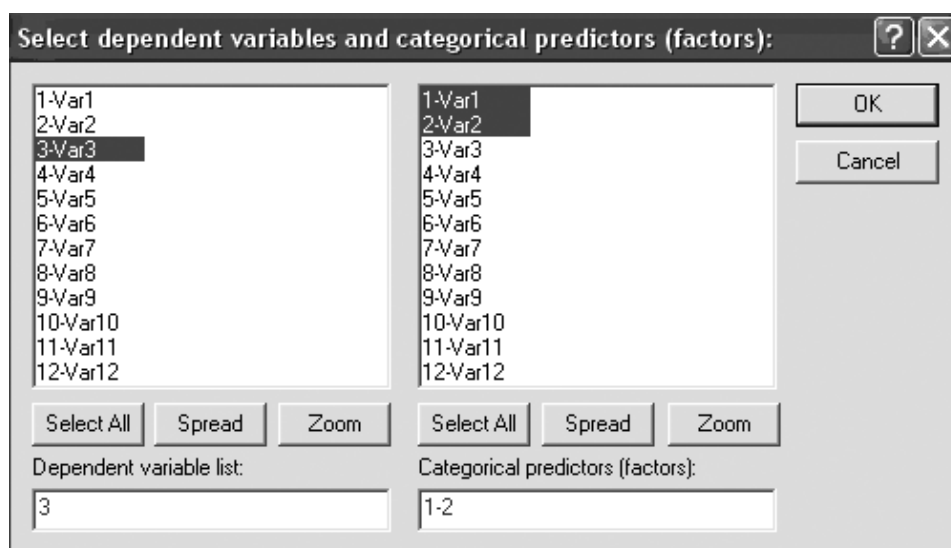


Рисунок 14.5 – Выбор переменных

Шаг 4. После нажатия кнопки **ОК** в появившемся окне выберите **All**, нажав кнопку **Factor codes** (рисунок 14.6).

Нажмите кнопку **ОК** в правом углу стартовой панели.

Шаг 5. На экране перед вами появится диалоговое окно **Anova Results** (Результаты) (рисунок 14.7). В данном окне выберите **Univariate Results** (Результат дисперсионного анализа). Далее нажмите кнопку **ОК**.

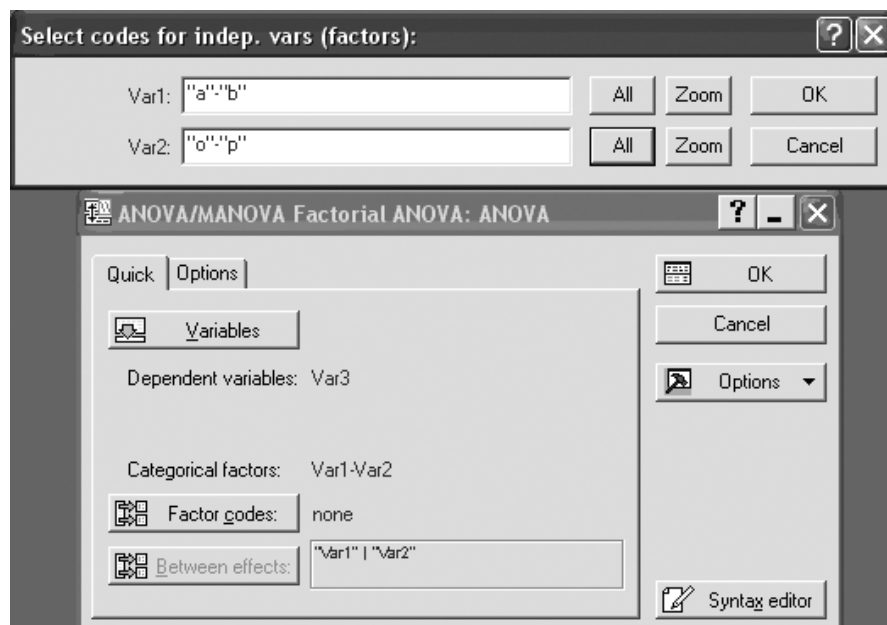


Рисунок 14.6– Окно выбора факторов

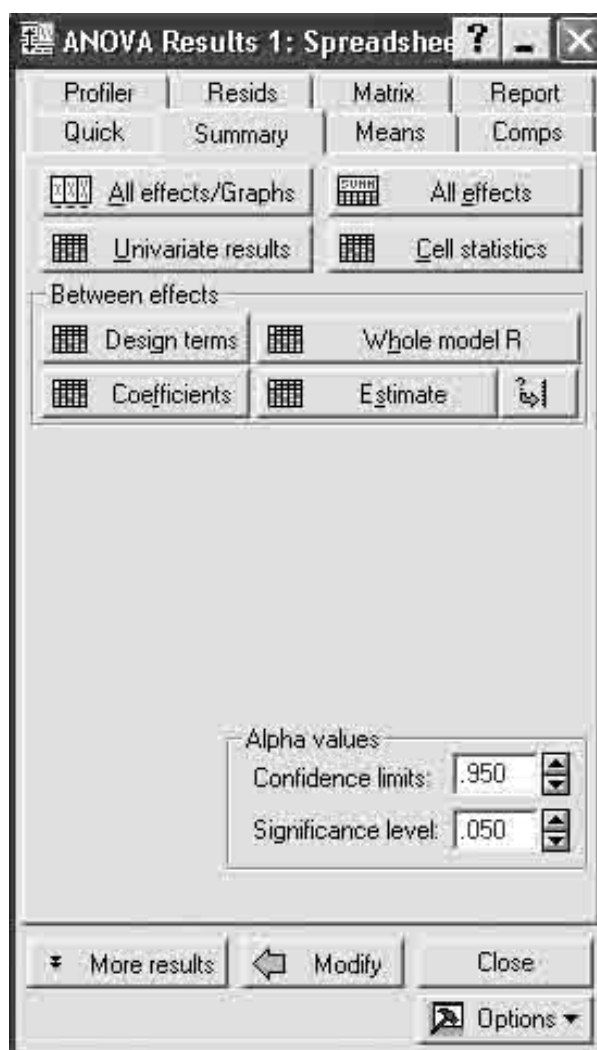


Рисунок 14.7 – Диалоговое окно результатов

В окне результатов (рисунок 14.8) представлены результаты дисперсионного анализа:

- между группами, фактор 1 – **Var1**;
- между группами, фактор 2 – **Var2**;
- взаимодействие – **Var1*Var2**;
- внутри групп – **Error**.

В рассмотренном примере **F-критерий** показывает, что различие между средними статистически значимо за счет влияния второго фактора (значимо на уровне 0,033). Сила влияния этого фактора составляет около 40%. (Результат в строке: между группами **Var2** – фактор 2 подсвечивается красным цветом).

GENERAL Effect	Degr. of Freedom	Var3 SS	Var3 MS	Var3 F	Var3 p
Intercept	1	49923,00	49923,00	340,5776	0,000000
"Var1"	1	1,33	1,33	0,0091	0,926364
"Var2"	1	972,00	972,00	6,6310	0,032867
"Var1""Var2"	1	243,00	243,00	1,6578	0,233902
Error	8	1172,67	146,58		
Total	11	2389,00			

Рисунок 14.8 – Результаты дисперсионного анализа

Визуализация результатов дисперсионного анализа представлена на рисунках 14.9, 14.12 и 14.13.

Шаг 6. В диалоговом окне результатов (рисунок 14.7) нажмите кнопку **All effects / Grafts**, затем в появившемся окне **OK**. Результат представлен на рисунке 14.9 а, b, с.

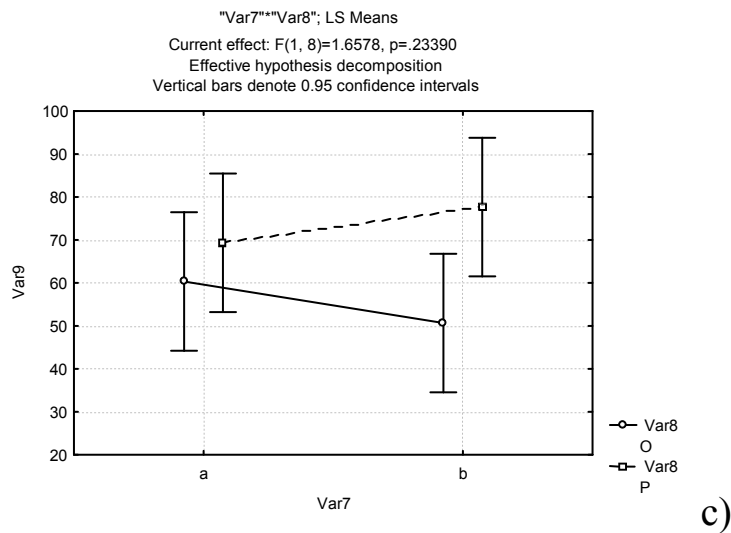
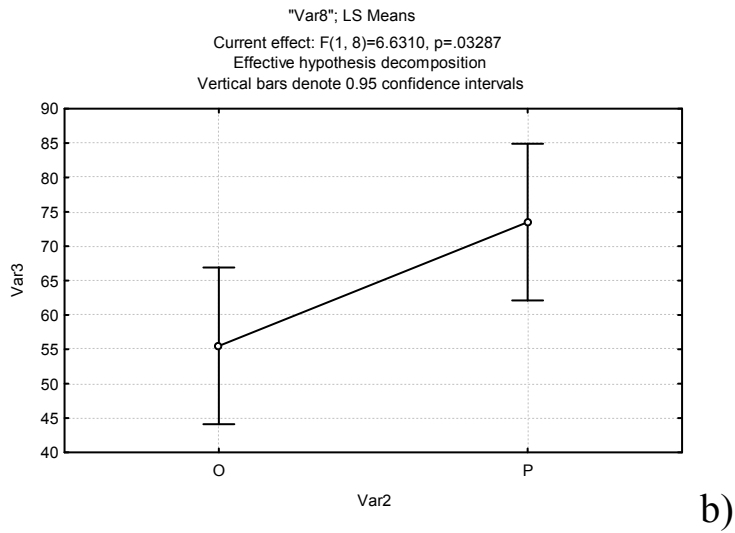
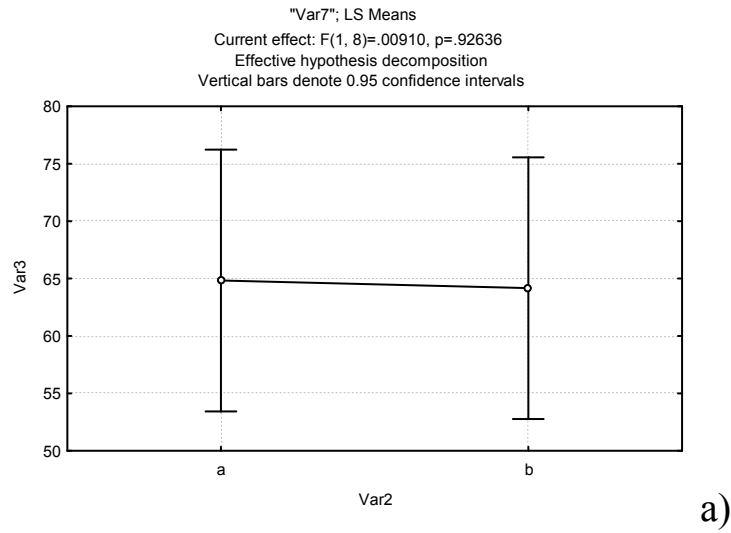


Рисунок 14.9 – Графики дисперсионного анализа: а) – влияние 1-го фактора; б) – влияние 2-го фактора; с) – взаимодействие
 Шаг 7. Из основного меню (рисунок 14.2) выберите модуль **Graphs**

(Графика), выберите опцию **Box plots** (рисунок 14.10) и щелкните мышкой. Затем выполните установки, как это показано на рисунке 14.11.

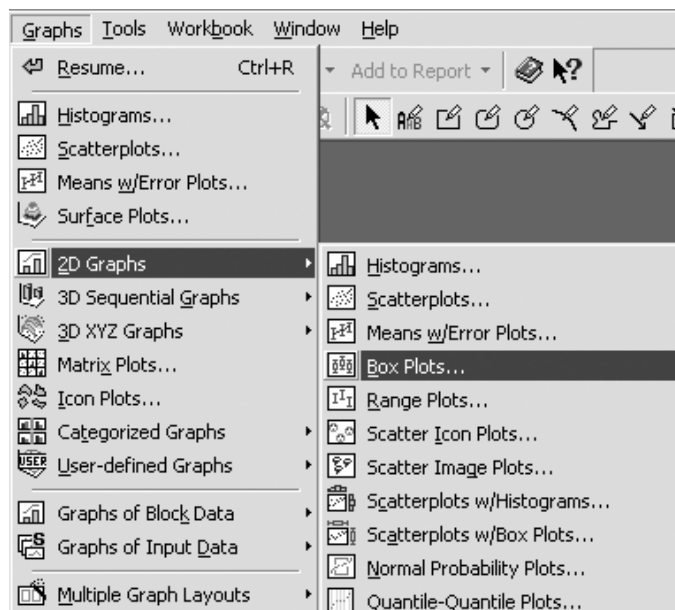


Рисунок 14.10 – Стартовая панель модуля

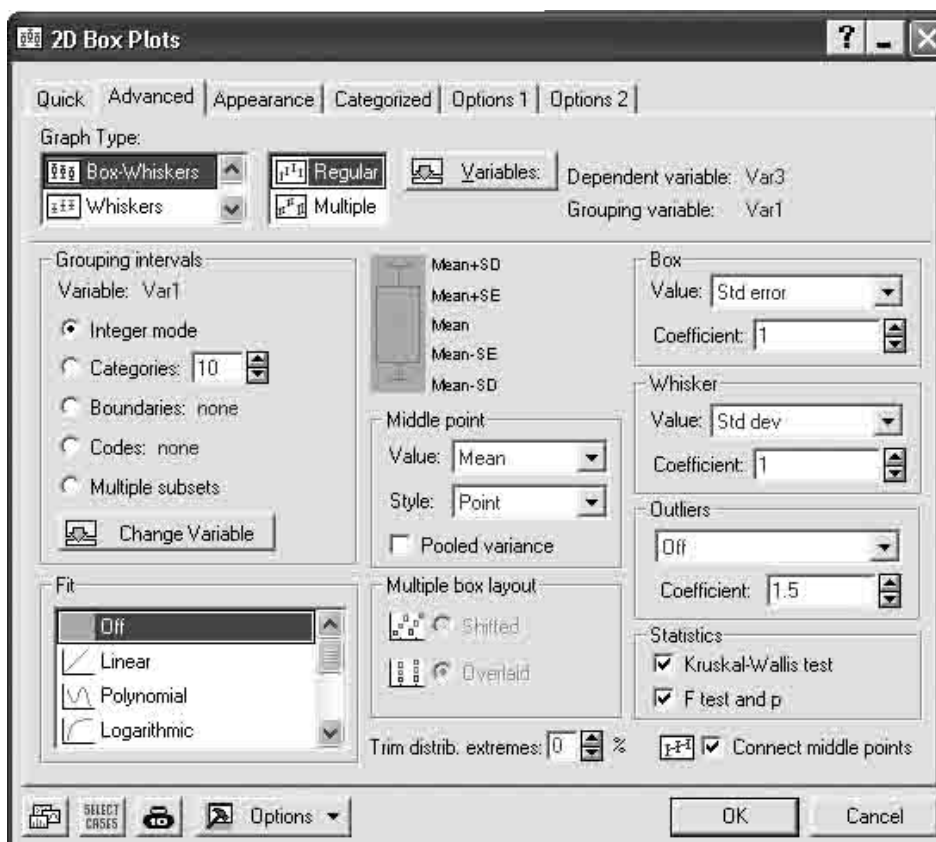


Рисунок 14.11 – Панель установок

Нажмите кнопку **ОК** и вы получите визуализацию статистических параметров при воздействии факторов (рисунки 14.12, 14.13) для каждого из которых показаны:

- среднее;
- стандартное отклонение;
- стандартная ошибка.

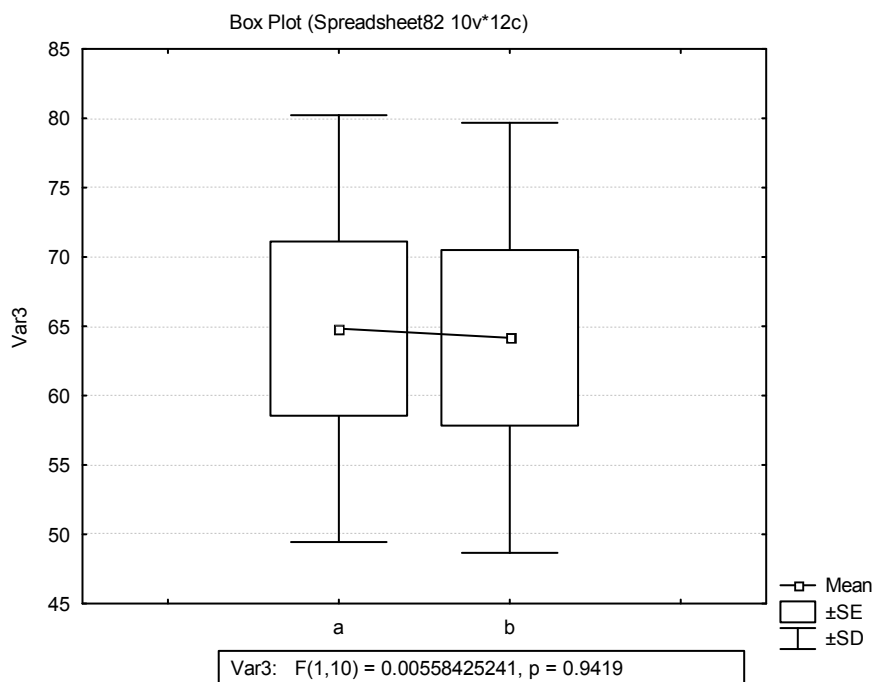


Рисунок 14.12 – График влияния первого фактора (*a* и *b*)

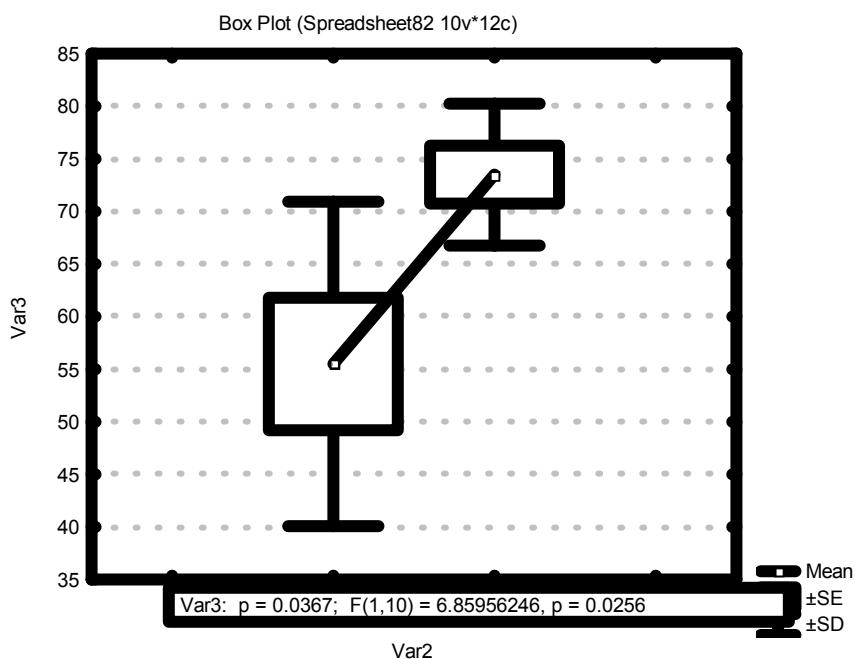


Рисунок 14.13 – График влияния второго фактора (σ и p)

Результаты, полученные в программном продукте **Statistica 6** точно такие же, как и в табличном редакторе **MS Excel**. Поэтому вывод о прибавке к урожаю аналогичен. Однако в программном продукте **Statistica 6** имеется возможность визуализации результатов, что, несомненно, его делает более привлекательным при представлении результатов исследований.

Задания для выполнения

- 1 Таблицу исходных данных из Приложения Е (таблица Е1) в табличном редакторе **Microsoft Excel** преобразуйте в формат **Statistica 6**.
- 2 Введите в таблицу **Statistica 6** преобразованные данные.
- 3 Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе.
- 4 Получите результат и сделайте заключение.

Лабораторная работа 6

Классификация

(дискриминантный анализ в Statistica 6)

Цель работы: научиться на основе измерения различных характеристик (признаков, параметров) объекта классифицировать его, то есть отнести к одной из нескольких групп (классов) некоторым оптимальным способом, используя модуль **Discriminant analysis** (Дискриминантный анализ) **Statistica 6**.

Краткие теоретические сведения

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 6** 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, а так же, как уже говорилось, методы множественной регрессии. Отличие кластерного анализа от дискриминантного в том, что в нем заранее не фиксировано число групп (кластеров).

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. Следует заметить, что умеренные отклонения от этого предположения не являются фатальными.

Предполагается, что дисперсии и ковариации наблюдаемых переменных в разных классах однородны (отличие между классами имеется только в средних). Умеренные отклонения от этого предположения также допустимы.

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

Сделаем 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.

В связи с этим рекомендуется проводить проверку с разумной степенью точности, сосредоточив основные усилия на построении модели.

Классификация цветов ириса

Знакомство с возможностями проведения дискриминантного анализа в системе **Statistica 6** 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**. В файле содержатся результаты измерений 150 цветков ириса, по 50 каждого типа.

Шаг 1. Нажмите кнопку **Open Data** (Открыть данные) и откройте файл данных **Irisdat.sta** из каталога **Examples** (Примеры на диске C). Следующий файл данных появится на экране (рисунок 15.1).

	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

Рисунок 15.1 – Файл данных Iris.sta

Шаг 2. Из Переключателя модулей **Statistica** откройте стартовую панель модуля **Discriminate function analysis** (Дискриминантный функциональный анализ) (рисунки 15.2, 15.3).

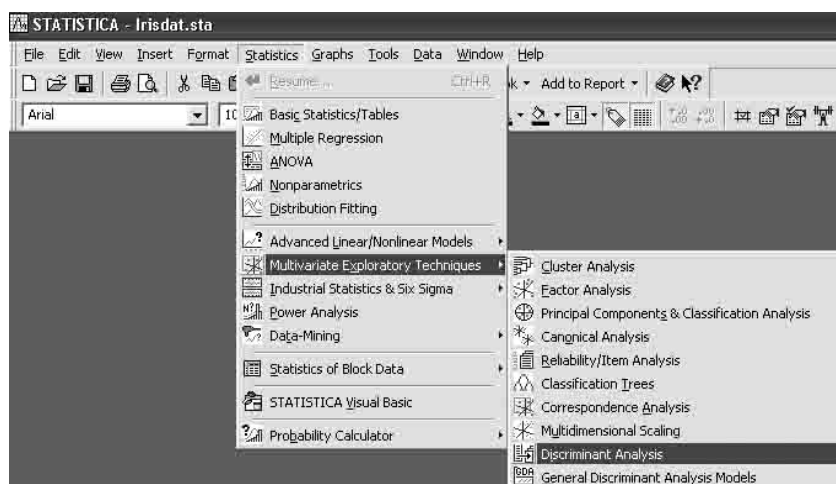


Рисунок 15.2 – Запуск модуля Дискриминантный анализ

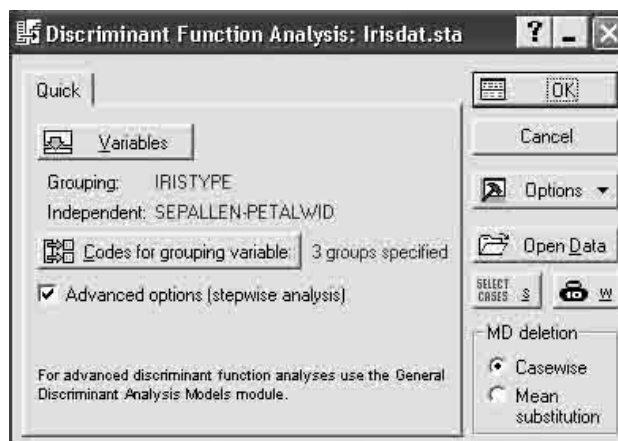


Рисунок 15.3 – Стартовая панель модуля Дискриминантный анализ

Шаг 3. Сделайте установки, как показано на рисунке 15.3. Для этого нажмите кнопку **Variables** (Переменные) и выберите переменные для анализа.

В качестве **Grouping variable** (Группирующая переменная) выберите переменную **Iristype** (Тип ириса) (рисунок 15.4).

В качестве **Independent variables** (Независимые переменные) выберите переменные **Sepallen, Sepalwid, Petallen, Petalwid** (Длина чашелистника, ширина чашелистника, длина лепестка, ширина лепестка) (рисунок 15.4). Нажмите кнопку **OK**.

Выполните установку **Codes for grouping variable** (Коды для группировки переменных) (рисунок 15.5). Нажмите кнопку **OK**.

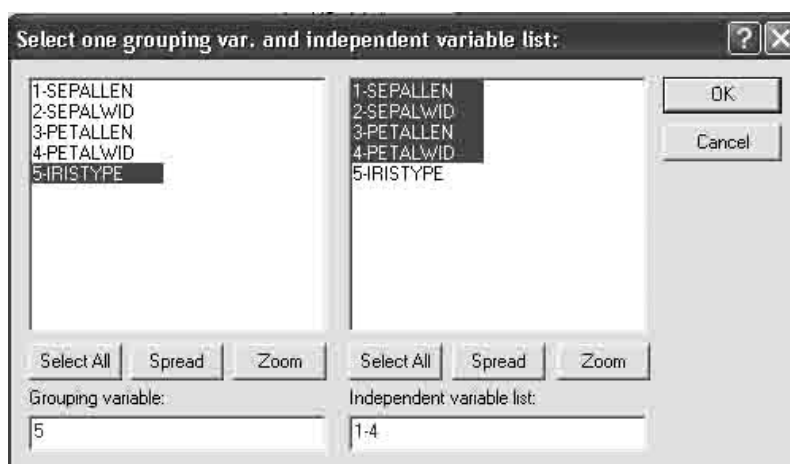


Рисунок 15.4 – Установка переменных (Variables)

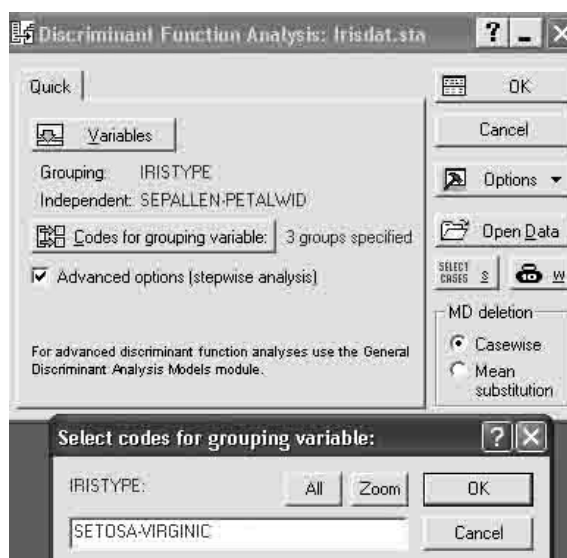


Рисунок 15.5 – Установка кодов для группировки переменных

Шаг 4. Нажмите кнопку **OK** и откройте диалоговое окно **Model**

Definition (Определение модели) (рисунок 15.6).

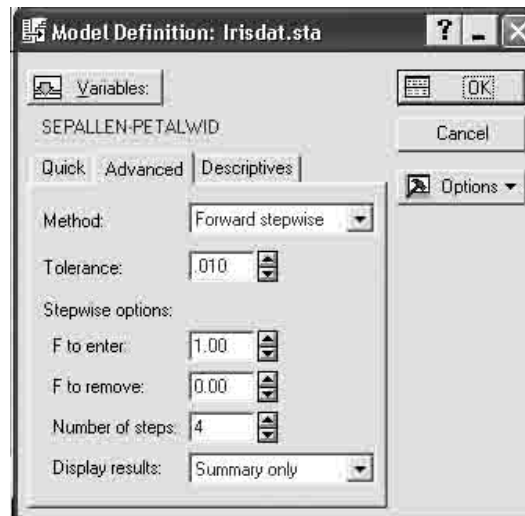


Рисунок 15.6 – Окно определения модели дискриминантного анализа

Шаг 5. Сделайте установки, как показано на рисунке 15.6. Нажмите кнопку **ОК** и запустите вычислительную процедуру, реализующую пошаговый метод включения.

Шаг 6. Всесторонне просмотрите итоги в диалоговом окне **Discriminant Function Analysis Results** (Результаты дискриминантного анализа) (рисунок 15.7).

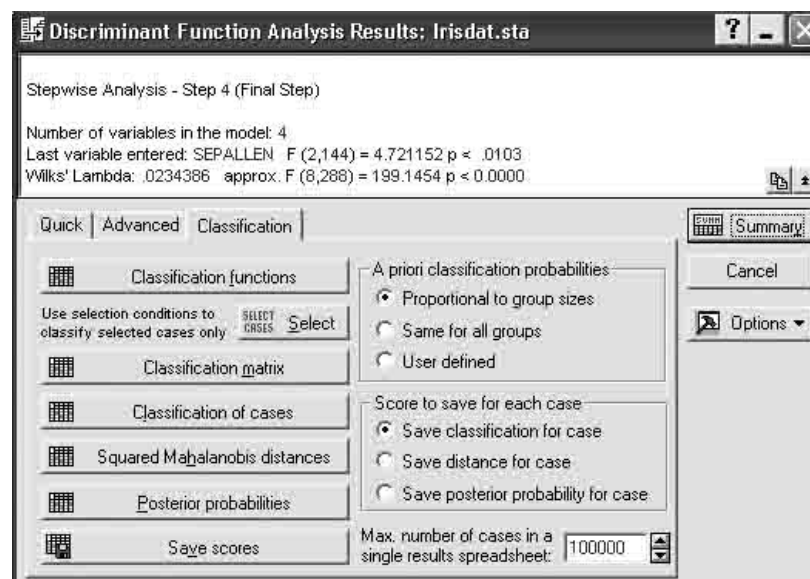


Рисунок 15.7 – Окно результатов дискриминантного анализа данных из файла Iris.sta

Информационная часть окна сообщает, что использовано:

– **Stepwise analysis** (Пошаговый анализ), **Step 4 Final step**

(4 шага);

- **Number of variables in the model** (Число переменных в модели): 4;
- **Last variable entered** (Последняя включенная переменная):

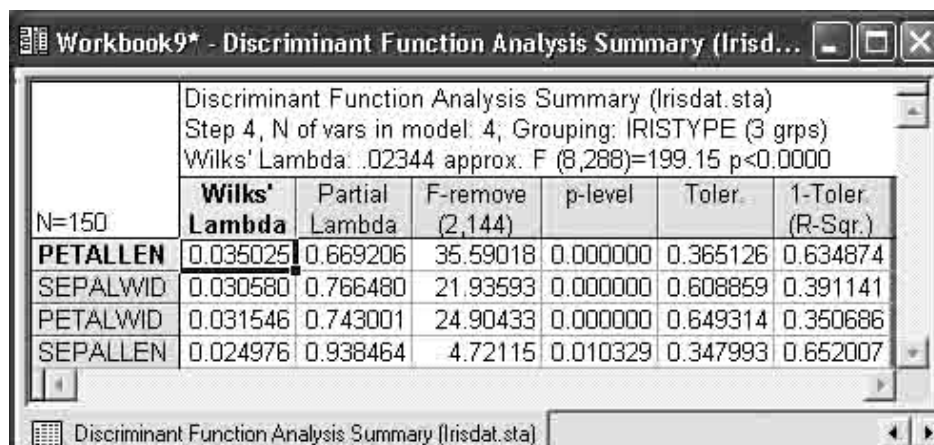
Sepallen, соответствующее значение статистики **F-критерия** ($F(2, 144) = 4,72$), уровень значимости $p < 0,01$;

- **Wilks lambda** (Значение лямбды Уилкса): 0,02;
- **approx. F (4,292) = 199,14** (Приближенное значение F-статистики), связанной с лямбдой Уилкса;
- **p – уровень значимости F-критерия** для значения 199,14;
- **значения статистики лямбда Уилкса** лежат в интервале 0 – 1.

Значения **статистики Уилкса**, лежащие около нуля, свидетельствуют о хорошей дискриминации. Значения **статистики Уилкса**, лежащие около единицы, свидетельствуют о плохой дискриминации.

Иными словами, это можно выразить следующим образом: если значения **лямбды Уилкса** близки к нулю, то мощность дискриминации (мощность = 1 – вероятность ошибки) близка к 1, если **лямбда Уилкса** близка к единицы, то мощность близка к нулю.

Шаг 7. Нажмите кнопку **Variables in the model** (Переменные, включенные в модель). На экране появится итоговая таблица анализа (рисунок 15.8).



Discriminant Function Analysis Summary (Irisdat.sta)						
Step 4, N 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.)
PETALLEN	0.035025	0.669206	35.59018	0.000000	0.365126	0.634874
SEPALWID	0.030580	0.766480	21.93593	0.000000	0.608859	0.391141
PETALWID	0.031546	0.743001	24.90433	0.000000	0.649314	0.350686
SEPALLEN	0.024976	0.938464	4.72115	0.010329	0.347993	0.652007

Рисунок 15.8 – Итоговая таблица анализа данных из файла Iris.sta

Шаг 8. Просмотрите разделение групп на графике. Для этого иницируйте кнопку **Perform Canonical analysis** (Канонический анализ). В появившемся диалоговом окне **Canonical Analysis** (канонический анализ) нажмите кнопку **Scatterplot of canonical scores** (Диаграмма рассеяния канонических значений). На экране появится следующий график (рисунок 15.9).

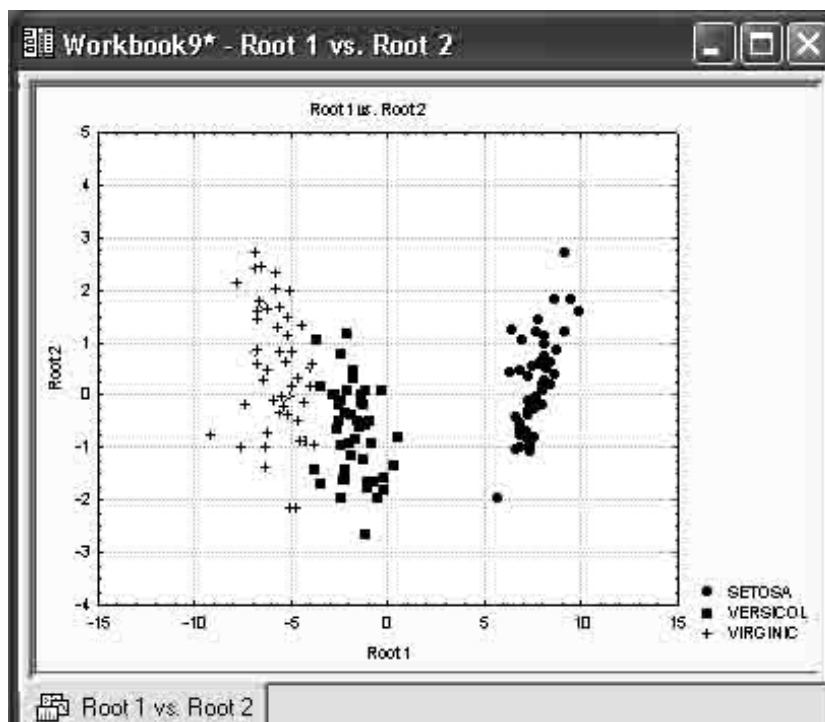


Рисунок 15.9 – Разделение трех типов ириса

Шаг 9. Просмотрите функции классификации. В диалоговом окне Результаты дискриминантного анализа нажмите кнопку **Classification functions** (Функции классификации) (рисунок 15.10).

Variable	Classification Functions; grouping:		
	SETOSA p=.33333	VERSICOL p=.33333	VIRGINIC p=.33333
PETALLEN	-16.4306	5.2115	12.767
SEPALWID	23.5879	7.0725	3.685
PETALWID	-17.3984	6.4342	21.079
SEPALLEN	23.5442	15.6982	12.446
Constant	-86.3085	-72.8526	-104.368

Рисунок 15.10 – Функции классификации, построенные пошаговым Forward stepwise (методом вперед)

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

$$\text{SETOSA} = -16,43 \cdot \text{SI} + 23,69 \cdot \text{Sw} - 17,4 \cdot \text{PI} + 23,54 \cdot \text{Pw} - 86,31;$$

$$\text{VERSICOL} = 5,21 \cdot \text{SI} + 7,07 \cdot \text{Sw} - 6,43 \cdot \text{PI} + 15,70 \cdot \text{Pw} - 72,85;$$

$$\text{VIRGINIC} = 12,76 \cdot \text{SI} + 3,69 \cdot \text{Sw} - 21,08 \cdot \text{PI} + 12,5 \cdot \text{Pw} - 104,37,$$

где:

– SI – Sepallen;

- Sw – Sepalwid;
- Pl – Petallen;
- Pw – Petalwid.

Пусть вы имеете новый цветок со значениями: **Sepallen, Sepalwid, Petallen, Petalwid.**

К какому типу ириса его отнести? 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 классификационные метки. Таким образом, каждый новый объект автоматически относится к определенному классу.

Шаг 10. Нажмите кнопку **Squared Mahalanobis distance** (Квадрат расстояния Махаланобиса) и вы увидите таблицу с квадратами расстояния Махаланобиса от точек (случаев) до центров групп (рисунок 15.11).

Case	Observed Classif.	SETOSA	VERSICOL	VIRGINIC
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

Рисунок 15.11 – Расстояния Махаланобиса для данных из файла Iris.sta

Случай относится к группе, до которой **расстояние Махаланобиса** минимально.

Шаг 11. **Апостериорные вероятности.**

Рассмотрите группу опций внизу диалогового окна **Результаты дискриминантного анализа: A priori classifications probabilities** (Априорные вероятности классификации). До анализа вы задаете для каждого объекта (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), вы увидите таблицу с апостериорными вероятностями принадлежности объекта к определенному классу (рисунок 15.12).

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

Рисунок 15.12 – Таблица апостериорных вероятностей

Интерпретация данной таблицы очень проста. В первом столбце указан тип ириса для каждого случая. Во втором, третьем, четвертом столбцах даны апостериорные вероятности отнесения каждого цветка к определенному типу.

Цветок относится к группе с максимальной апостериорной вероятностью.

Знаком * отмечаются неправильно классифицированные при использовании данного правила случаи (5, 9, 12). В исходной таблице (рисунок 15.1) необходимо внести изменения, а затем заново выполнить процедуру дискриминантного анализа.

Шаг 12. Классификация новых случаев.

Не закрывая диалога **Результаты дискриминантного анализа**, добавьте в таблицу исходных данных новый случай (151 строка на рисунке 15.13).

	1	2	3	4	5
	SEPALLEN	SEPALWID	PETALLEN	PETALWID	IRISTYPE
147	6.7	3	5	1.7	VERSICOL
148	6.3	3.3	6	2.5	VIRGINIC
149	5.3	3.7	1.5	0.2	SETOSA
150	5	2.3	3.3	1	VERSICOL
151	5.3	3.1	2.7	0.5	

Рисунок 15.13 – Новое наблюдение в данных Iris.sta

Шаг 13. Запустите процедуру дискриминантного анализа. Для того чтобы понять, к какому классу относится этот объект, нажмите кнопку **Posterior probabilities** (Апостериорные вероятности), вы увидите ту же таблицу с апостериорными вероятностями, к которой будет добавлена строка (рисунок 15.14).

Case	Observed Classif.	SETOSA p= .33333	VERSICOL p= .32667	VIRGINIC p= .34000
147	VERSICOL	0.000000	0.715864	0.284136
148	VIRGINIC	0.000000	0.000000	1.000000
149	SETOSA	1.000000	0.000000	0.000000
150	VERSICOL	0.000000	1.000000	0.000000
151	---	0.999886	0.000114	0.000000

Рисунок 15.14 – Классификация нового наблюдения

Итак, новое наблюдение с вероятностью 0,999 можно отнести к типу **SETOSA**.

Задания для выполнения

- 1 Создайте пустую электронную таблицу **Spreadsheet.sta**.
- 2 Внесите данные для выполнения расчетов, находящиеся в Приложении Ж (таблица Ж1).
- 3 Выполните процедуры дискриминантного анализа в соответствии с порядком операций, выполненных в настоящем разделе.
- 4 Дайте объяснение полученным результатам.
- 5 Из таблицы 1.1 внесите новые данные.
- 6 Запустите процедуру дискриминантного анализа.
- 7 Внесите поправки в исходную таблицу, пополнив обучающую выборку новой информацией.

Таблица 15.1 – Классификация новых случаев

Класс	1 задание				Класс	2 задание		Класс	3 задание	
	1 признак.	2 признак.	3 признак.	4 признак.		1 признак.	2 признак.		1 признак.	2 признак.
	1,14	1,26	0,99	2,06		0,738	0,658		36,63	31,29
	0,79	0,84	1,17	2,72		0,612	0,243		24,84	19,63
	1,01	1,16	1,06	1,4		0,774	0,233		17,78	13,00
	0,97	1,11	0,73	0,98		0,933	0,271		5,17	1,92

Лабораторная работа 7

Классификация (кластерный анализ в Statistica 6)

Цель работы: научиться на основе применения метода **k-means clustering** (k-средних) **Statistica 6** исходную совокупность объектов разделить на кластеры или группы (классы) схожих между собой объектов.

Краткие теоретические сведения

В **Statistica 6** реализованы следующие методы кластеризации – агломеративные методы: **joining (tree clustering)**, **two-way joining**, а также метод **k-means clustering** (k-средних).

Обычно перед началом классификации данные стандартизуются (вычитается среднее и производится деление на корень квадратный из дисперсии). Полученные в результате стандартизации переменные имеют нулевое среднее и единичную дисперсию. Эту операцию успешно можно проводить в программе **MS Excel**. Рассматриваемые далее данные уже стандартизованы.

Рассмотрим итеративный метод группировки **k-means clustering** (k-средних). Данный метод работает непосредственно с объектами, а не с матрицей сходства.

В методе **k-средних** объект относится к тому классу, расстояние до которого минимально. Расстояние понимается как евклидово расстояние, то есть объекты рассматриваются как точки евклидова пространства.

Всем известно как определить евклидово расстояние. Но как определить расстояние от объекта до совокупности объектов? Оказывается, это можно сделать следующим способом: каждый класс объектов имеет центр тяжести (рассмотрите, как и ранее, простейший случай – представьте, что объект имеет только два параметра, тогда его можно изобразить точкой на плоскости, а группа объектов – это просто группа точек).

Расстояние между объектом и классом есть расстояние между объектом и центром класса. Но как вычислить центр класса? Например, взять средние по каждому параметру. Тогда расстояние между объектом и группой объектов вполне определено и алгоритм может работать.

Представьте, что число объектов в группе равно 2. Соедините эти

точки отрезком прямой и найдите его середину. Это и будет центр тяжести группы, состоящей из двух точек. Расстояние от этого центра до исходной точки будет искомым расстоянием.

Принципиально метод **k-средних** «работает» следующим образом:

- вначале задается некоторое разбиение данных на кластеры (число кластеров определяется пользователем);
- вычисляются центры тяжести кластеров;
- происходит перемещение точек: каждая точка помещается в ближайший к ней кластер;
- вычисляются центры тяжести новых кластеров;
- шаги 2, 3 повторяются, пока не будет найдена стабильная конфигурация (то есть кластеры перестанут изменяться) или число итераций не превысит заданное пользователем. Итоговая конфигурация и является искомой.

В классическом примере рассматриваются автомобили разных марок, которые различаются ценой, расходом горючего и некоторыми техническими характеристиками, например, разгоном – временем, необходимым для того, чтобы достичь скорости 60 миль в час. Однако кластерный анализ с успехом применим и в любых биологических, химических и экологических исследованиях.

Кластеризация автомобилей в модуле Cluster Analysis (Кластерный анализ)

Шаг 1. В рабочем окне **Statistica** выберете название модуля – **Cluster Analysis** (Кластерный анализ), высветите его имя и щелкните на его имени (рисунок 16.1).

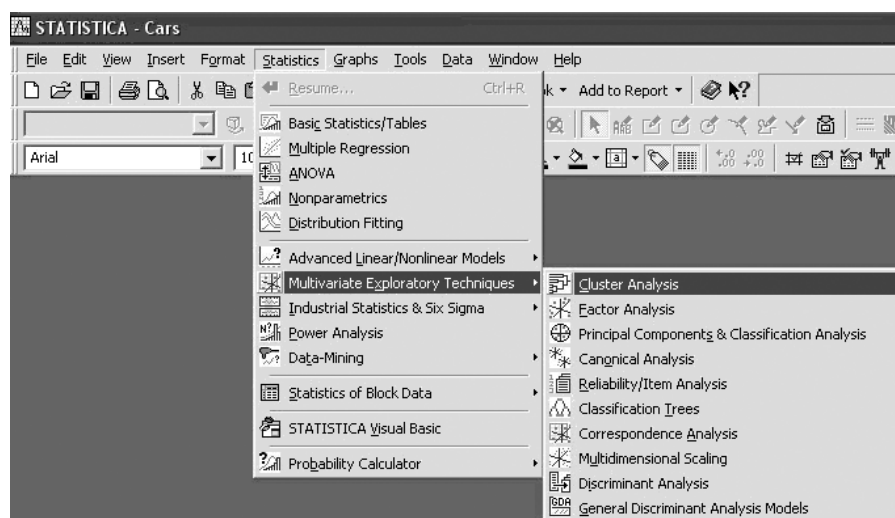


Рисунок 16.1 – Запуск модуля Кластерный анализ

На экране появится стартовая панель модуля **Cluster Analysis** (Кластерный анализ) (рисунок 16.2).

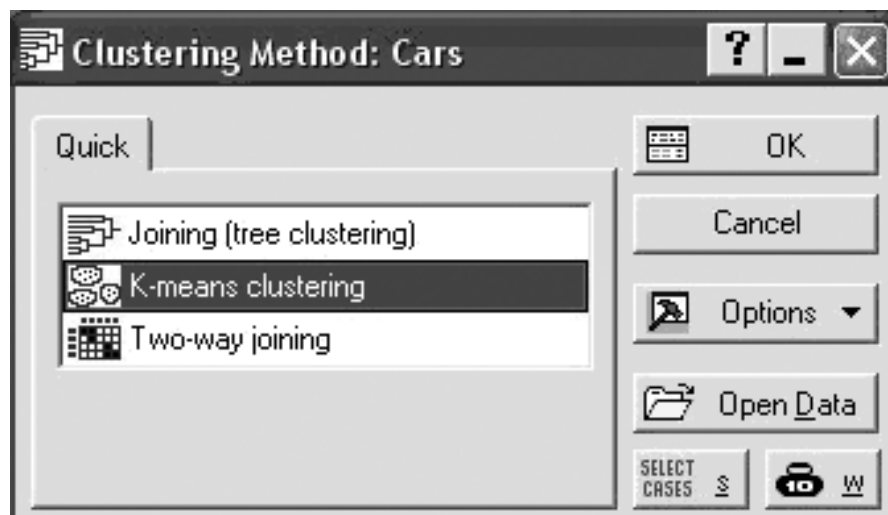


Рисунок 16.2 – Стартовая панель модуля Кластерный анализ

Шаг 2. Стандартным образом, нажав кнопку **Open Data**, откройте окно выбора файла (рисунок 16.3).

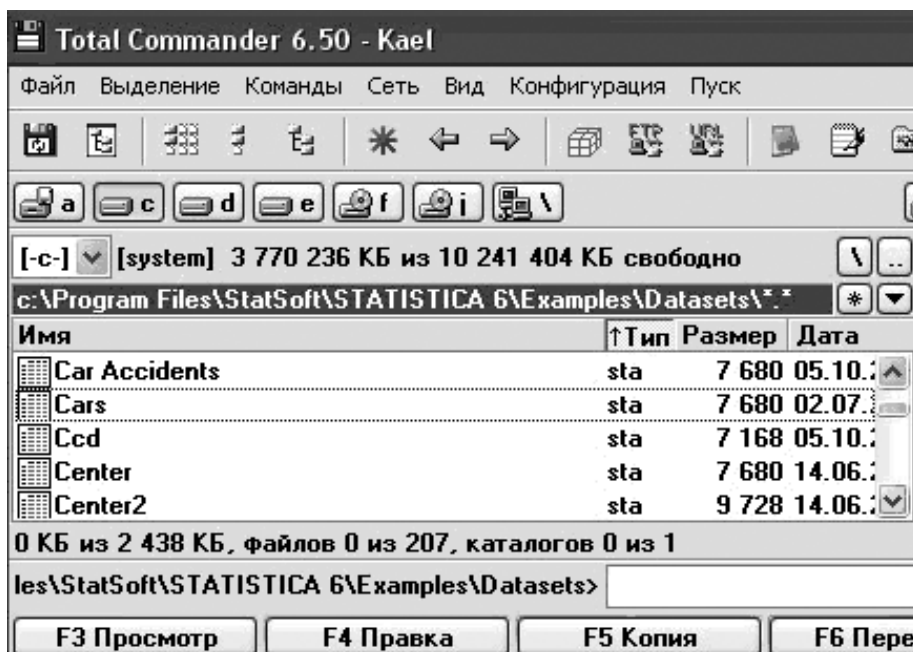


Рисунок 16.3 – Выбор файла с данными об автомобилях

Шаг 3. Выберите в этом окне файл **Cars.sta**, как показано на рисунке, и два раза нажмите левую кнопку мышки. Файл выбран, и вы вернетесь обратно, в стартовую панель модуля.

В рабочем окне, сзади стартовой панели, вы видите открытый файл с данными (рисунок 16.4).

	1	2	3	4	5
	PRICE	ACCELERATION	BRAKING	HANDLING	MILEAGE
Acura	-0,521072363	0,477252671	-0,00657103855	0,381619066	2,07875356
Audi	0,865652474	0,208033216	0,31869537	-0,0913735792	-0,677061608
BMW	0,495859184	-0,801539742	0,192202878	-0,0913735792	-0,153805564
Buick	-0,613520685	1,68874022	0,933087475	-0,20962174	-0,153805564
Corvette	1,23544576	-1,8111127	-0,494470651	0,972859872	-0,677061608
Chrysler	-0,613520685	0,0734234878	0,427117506	-0,20962174	-0,153805564
Dodge	-0,705969008	-0,195795968	0,481328574	0,145122743	-0,153805564
Eagle	-0,613520685	1,21760617	-4,19889364	-0,20962174	-0,677061608
Ford	-0,705969008	-1,54189324	0,987298543	0,145122743	-1,7235737
Honda	-0,42862404	0,409947807	-0,00657103855	0,0268745821	0,369450479
Isuzu	-0,79841733	0,409947807	-0,0607821066	-4,23005922	1,0671252
Mazda	0,126065894	0,679167263	-0,133063531	0,499867227	-1,7235737
Mercedes	1,05054912	0,00611862399	0,119921454	-0,0913735792	-0,153805564
Mitsub.	-0,613520685	-1,00345433	0,0837807415	0,381619066	0,718287842
Nissan	-0,42862404	0,0734234878	-0,00657103855	0,263370905	0,997357732
Olds	-0,613520685	-0,734234878	0,40904715	0,381619066	2,11363729
Pontiac	-0,613520685	0,679167263	0,535539642	0,145122743	0,195031798
Porsche	3,4542055	-2,21494188	-0,295696735	0,618115388	-1,02589897
Saab	0,588307506	0,679167263	0,246413946	0,263370905	0,0206131169
Toyota	-0,0588307506	1,21760617	0,22834359	0,73636355	-0,851480289
VW	-0,705969008	-0,128491104	0,101851098	0,381619066	0,195031798
Volvo	0,218514217	0,611862399	0,13799181	-0,20962174	0,369450479

Рисунок 16.4 – Файл Cars.sta с данными автомобилей разных марок

Из информации в верхней части окна вы видите, что в файле **Cars.sta** записаны цена автомобиля, технические характеристики, количество миль, пройденных на одном галлоне бензина.

Всего в файле содержатся данные о 22 машинах разных марок. Марки машин – это случаи.

Переменные в этом файле:

- **PRICE** – цена;
- **ACCELE – HANDLI** – технические характеристики;
- **MILAGE** – расход горючего (количество миль, пройденных на одном галлоне бензина).

Все характеристики машин уже стандартизованы: например, из значений переменной **price** вычтена средняя цена и разность поделена на корень квадратный из дисперсии. То же относится к другим переменным.

Задача состоит в том, чтобы разбить автомобили на несколько групп, в которых автомобили мало отличаются друг от друга (существенно меньше, чем в целом в совокупности).

Задача эта сложна, так как сравниваются машины не по какому-то

одному параметру, а по нескольким параметрам одновременно.

Вы видите, что по одним характеристикам одни машины близки друг к другу, по другим – нет. В конечном итоге разбиение на группы – тоже не самоцель. Конечно, число параметров можно увеличить. Очевидно, разбив машины на группы, можно лучше в целом представить их совокупность, с тем, чтобы затем более обоснованно принимать решение, например при покупке или обмене одной машины на другую.

Если бы машины сравнивались по одному параметру, например по расходу горючего, то, наверное, следовало бы выбрать машину с меньшим расходом топлива на одну милю. Все машины были бы упорядочены в одну линию, и задача не представляла бы проблем.

Однако параметров несколько, и ситуация существенно усложняется.

Посмотрите на стартовую панель. В главной ее части находится список методов кластерного анализа, реализованных в **Statistica**.

Шаг 4. В списке методов выделите **k-means** (к-средних) (рисунок 16.2) и нажмите кнопку **ОК** в правом верхнем углу панели.

Диалоговое окно метода **k-means** появится на экране (рисунок 16.5).

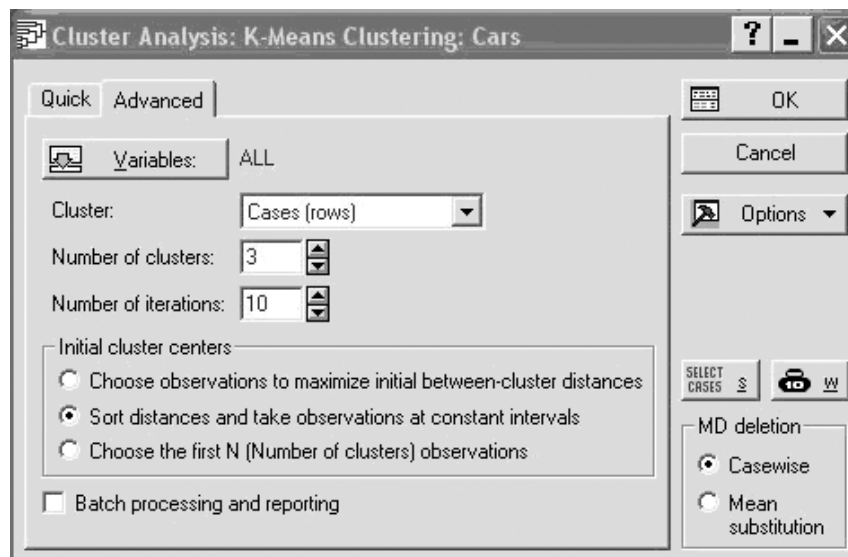


Рисунок 16.5 – Диалоговое окно метода k-means

Шаг 5. Начните работать в данном окне. Прежде всего, выберите переменные для анализа.

Нажмите кнопку **Variables** (Переменные) в левом верхнем углу текущего окна и откройте диалоговое окно: **Select variable for the analysis** (Выбрать переменные для анализа) (рисунок 16.6).

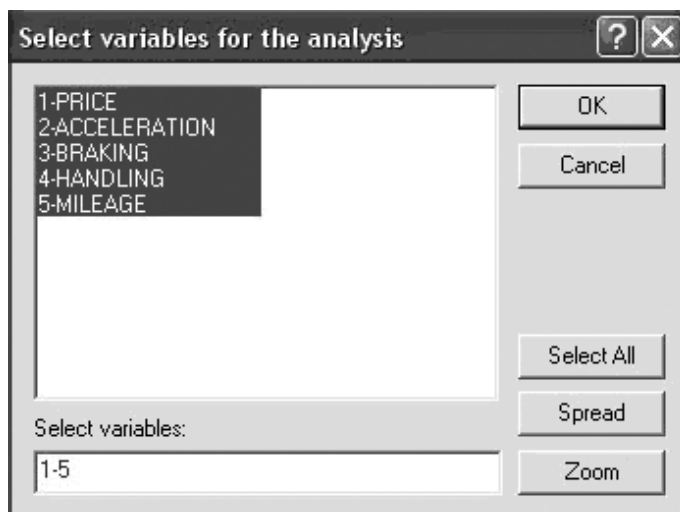


Рисунок 16.6 – Выбор переменных для кластерного анализа

Так как машины разбиты на группы и учитываются все параметры, то нажмите вначале кнопку **Select All** (Выбрать все), а затем нажмите кнопку **OK**.

Шаг 6. Посмотрите далее на поле **Cluster** (Кластер), находящееся ниже кнопки **Variables** (Переменные). Нажав на стрелку в этом поле, выберите пункт меню **Cases** (Случаи). Альтернативный выбор был бы **Variables** (Переменные). Так следует поступить, если нужно кластеризовать переменные.

В данном примере кластеризируются машины, которые являются случаями в исходном файле данных, поэтому выбирается пункт **Cases**.

Шаг 7. В поле **Number of clusters** (Число кластеров) нужно определить число групп, на которые необходимо разбить автомобили. Запишите в это поле число 3.

Таким образом, машины разбиваются на 3 кластера.

Шаг 8. В строке **Number of iterations** (Число итераций) задается максимальное число итераций, используемых при построении классов. Задайте, например, число 11.

Шаг 9. В строке **Missing data** задается способ обработки пропущенных значений в данных (например, для какой-то машины отсутствует значение некоторого параметра). В данном примере пропусков в данных нет и обработки пропущенных значений не происходит.

Группа опций **Initial cluster centers** позволяет задать начальные центры кластеров.

Сделайте установки, как показано на рисунке 16.5.

Шаг 10. После того как все установки сделаны, нажмите кнопку **OK** в верхнем правом углу окна **k-means Clustering** и запустите вычислительную процедуру.

Шаг 11. Спустя несколько секунд после нажатия кнопки **OK** в **k-means Clustering** окно результатов появится на экране (рисунок 16.7).

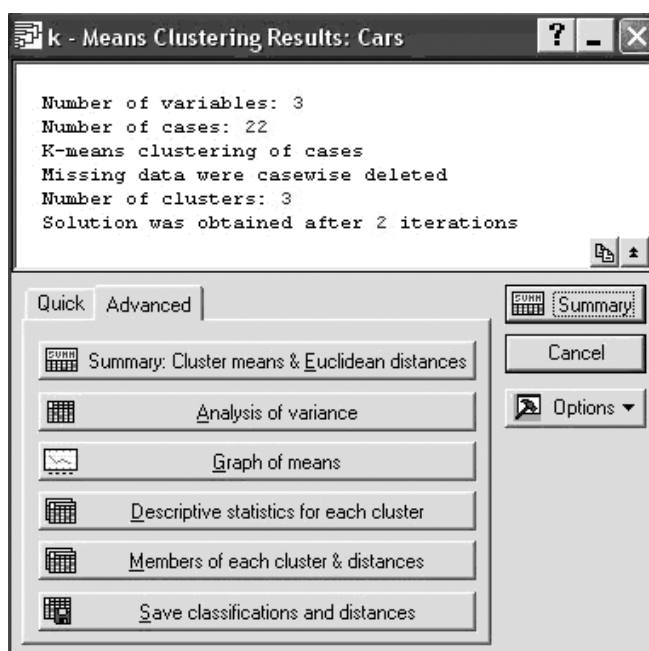


Рисунок 16.7 – Окно результатов кластеризации машин по методу средних

В верхней части окна записана информация: число переменных, число случаев, метод кластеризации, число кластеров, а также сообщение о том, после скольких итераций найдено решение: **Solution was obtained after 3 iterations** (решение найдено после 3 итераций).

Кнопки в нижней части окна позволяют провести анализ результатов кластеризации.

Кнопка **Analysis of variation** (Дисперсионный анализ) позволяет просмотреть таблицу дисперсионного анализа.

Кнопка **Cluster Means&Euclidean Distances** позволяет вывести таблицы, в первой из которых указаны средние для каждого кластера (усреднение производится внутри кластера), во второй указаны евклидовы расстояния и квадраты евклидовых расстояний между кластерами.

Кнопка **Graph of means** позволяет посмотреть средние значения для

каждого кластера на линейном графике.

Кнопка **Descriptive Statistics for each clusters** открывает электронную таблицу с описательными статистиками для каждого кластера (среднее, дисперсия и т. д.)

Кнопка **Save classifications and distances** позволяет сохранить результаты классификации в файле **Statistica** для дальнейшего исследования.

Шаг 12. Следует посмотреть, как распределились машины по кластерам. Нажмите для этого кнопку **Member of each cluster&distances**.

На экране появятся 3 электронные таблицы с названиями машин, отнесенных к определенным кластерам (рисунки 16.8 — 16.10).

	Acura	Buick	Chrysler	Dodge	Honda	Mitsub.	Nissan	Olds	Pontiac	Saab	Toyota	VW	Volvo
Distance	0,754166	0,766466	0,356816	0,384616	0,158199	0,614239	0,297823	0,889882	0,255611	0,508612	0,766000	0,284704	0,362700

Рисунок 16.8 – Первый кластер

	Audi	BMW	Corvette	Ford	Mazda	Mercedes	Porsche
Distance	0,492166	0,414537	0,636028	0,978105	0,849100	0,509201	1,319699

Рисунок 16.9 – Второй кластер

	Eagle	Isuzu
Distance	1,360452	1,360452

Рисунок 16.10 – Третий кластер

В строках таблиц указано расстояние от каждой машины до центра кластера.

Шаг 13. Нажмите на кнопку **Cluster means&Euclidean distances**. На

экране появится таблица, в которой даны евклидовы расстояния между средними кластеров (по каждому из параметров внутри кластера вычисляется среднее, получается 3 точки в пятимерном пространстве, и между ними находится расстояние) (рисунок 16.11).

Cluster Number	No. 1	No. 2	No. 3
No. 1	0,000000	0,938877	2,403320
No. 2	0,968957	0,000000	3,519362
No. 3	1,550264	1,875996	0,000000

Рисунок 16.11 – Расстояния между кластерами

Из таблицы вы видите, что расстояние между первым и вторым кластером 0,969, а например, между вторым и третьим – 1,876.

Над диагональю в таблице даны квадраты расстояний между кластерами.

Шаг 14. С помощью кнопки **Graph of means** (График средних) строятся следующие графики средних значений характеристик машин для каждого кластера (рисунок 16.12).

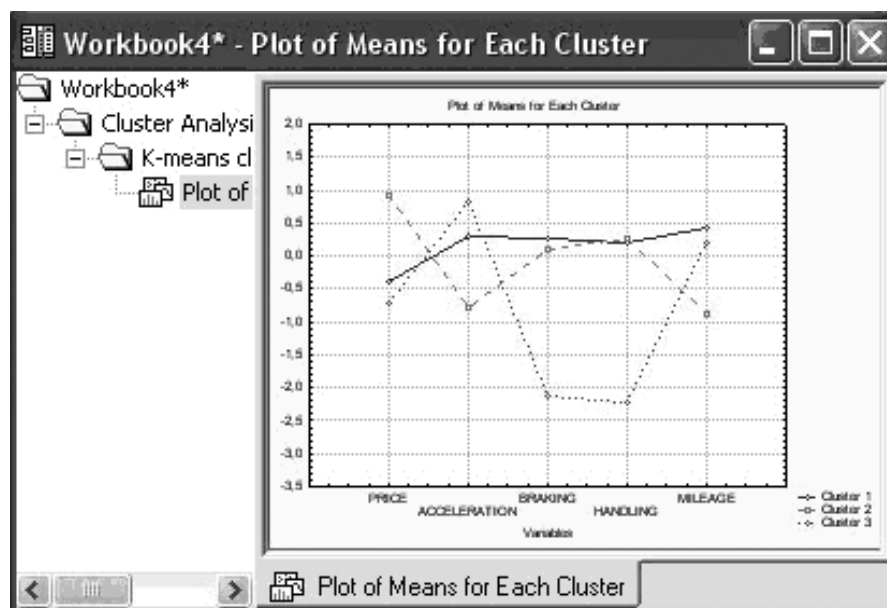


Рисунок 16.12 – График средних для каждого кластера

Закроем окно результатов и вернемся в начальное окно метода

к–средних.

Изменим переменные для анализа.

Шаг 15. Нажмите кнопку **Variables** (Переменные) в левом верхнем углу текущего окна и откройте диалоговое окно: **Select variables for the analysis**. Сделайте в нем установки, как показано на рисунке 16.13 (выберем теперь только 3 параметра, характеризующих машины).

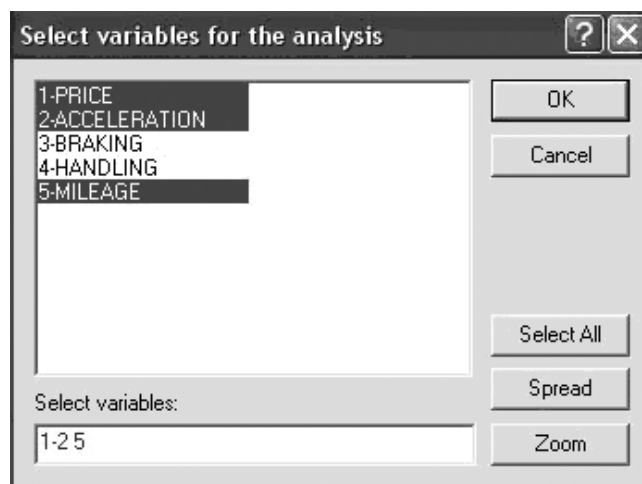


Рисунок 16.13 – Выбор части переменных для кластерного анализа методом к-средних

Шаг 16. Повторите действия, описанные ранее. Нажмите кнопку **Graph of means** (График средних), постройте следующие графики средних значений характеристик машин для каждого кластера (рисунок 16.14).

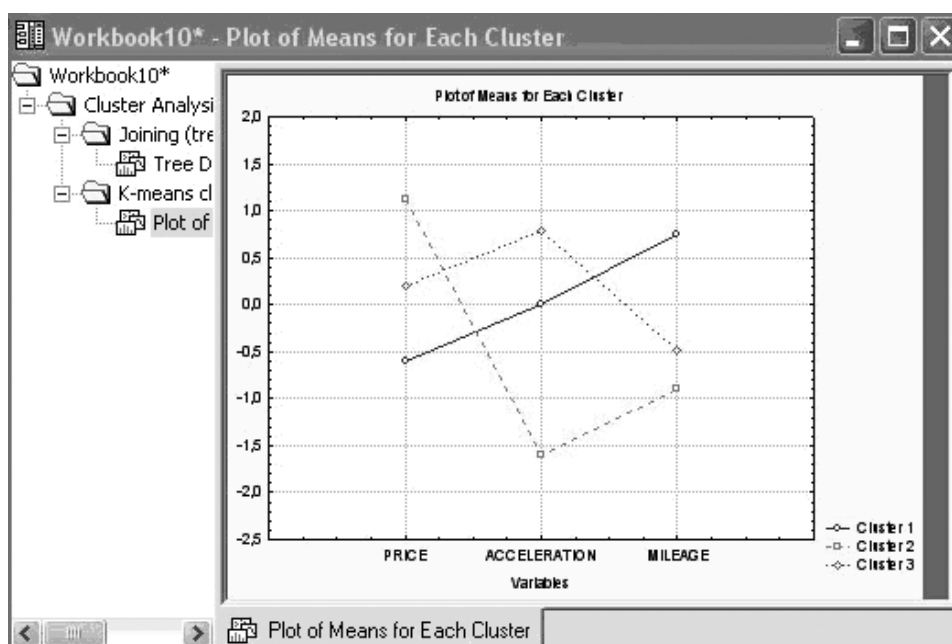


Рисунок 16.14 – График средних для новых кластеров

Заметьте, что состав групп изменился. Теперь машины более отчетливо группируются. Пожертвовав размерностью, сократили число параметров и получили более отчетливо выраженные группы (сравните с рисунком 16.12).

Поэкспериментируйте с этими данными. Возможно, вам удастся найти оптимальную кластеризацию.

После того как вы поработаете с этим примером, обязательно попробуйте расклассифицировать другие свои собственные данные.

В системе реализованы также и другие методы кластеризации, в частности, так называемый **two-way joining**, в котором кластеризируются случаи и переменные одновременно.

Шаг 17. Если вы воспользуетесь **Joining (tree clustering)**, то сможете увидеть дендрограмму, или дерево объединения (рисунок 16.15), о котором говорилось вначале.

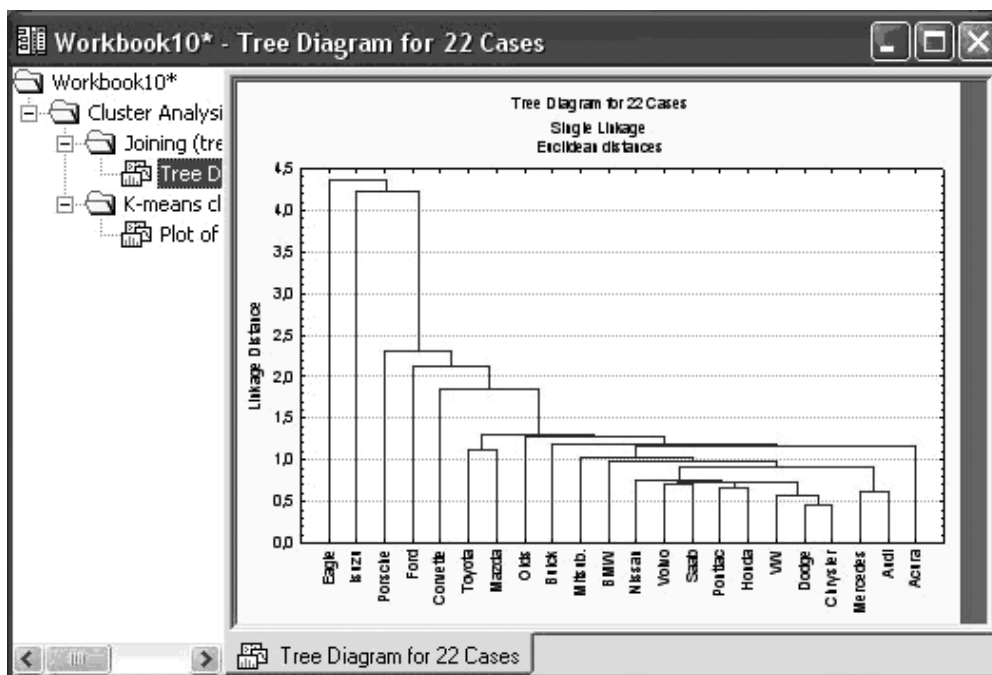


Рисунок 16.15 – Дерево объединения машин разных марок в кластер методом одиночной связи

Классификация населенных пунктов, расположенных в зоне радиоактивного загрязнения

Для классификации населенных пунктов по степени первоочередности проведения мероприятий радиационного или социального характера были учтены следующие факторы:

- демографический (численность населения, возрастная структура населения i -го населенного пункта);
- хозяйственный (отношение числа жителей к числу коров);
- радиационный (средние значения суммарной годовой эффективной индивидуальной дозы и удельной активности молока по i -му населенному пункту).

На основе этой информации для каждого населенного пункта были рассчитаны социально-экономические и радиологические показатели, которые затем были отнормированы на максимальное значение. Соответствующие значения в баллах, присваиваемые i -му населенному пункту приведены в таблице 16.1. Максимальное значение – 1 балл.

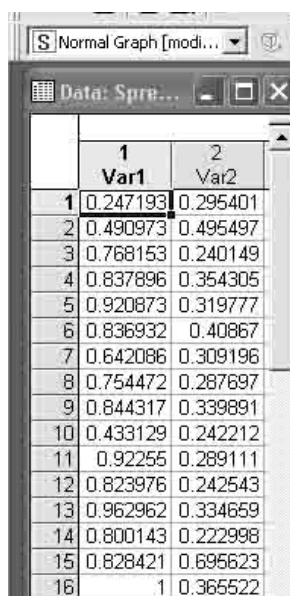
Таблица 16.1 – Исходная информация по населенным пунктам

№ НП	Соц.-экон.	Радиолог.	№ НП	Соц.-экон.	Радиолог.
1	0,247193289	0,295400607	24	0,763286867	0,77716795
2	0,49097333	0,495497008	25	0,451861817	0,339317409
3	0,768153128	0,240149383	26	0,46782343	0,428489112
4	0,83789641	0,354305136	27	0,849075263	0,298006857
5	0,920873425	0,319777147	28	0,760966155	0,251519518
6	0,836931986	0,408670359	29	0,298413449	0,201345166
7	0,642086133	0,30919644	30	0,820472659	0,438164981
8	0,75447239	0,287696776	31	0,886396124	0,251515902
9	0,844316587	0,339890945	32	0,760527189	0,533589347
10	0,43312923	0,242211888	33	0,8274442	0,361130457
11	0,922549745	0,289111366	34	0,894219018	1
12	0,823975995	0,242543293	35	0,38097797	0,458738374
13	0,962962191	0,334658522	36	0,7502056	0,30138691
14	0,800143159	0,222998021	37	0,613001789	0,512266066
15	0,828420843	0,695622825	38	0,442903759	0,369355736
16	1	0,365521838	39	0,852623864	0,230988926
17	0,715452937	0,354404142	40	0,789770272	0,397828781
18	0,885349202	0,572964656	41	0,74822401	0,526758725
19	0,658321153	0,277728731	42	0,892492356	0,369276556
20	0,738969511	0,632279179	43	0,737587506	0,657796926
21	0,683658507	0,357111189	44	0,612246087	0,242905587
22	0,683658507	0,591685683	45	0,773756059	0,23284717
23	0,511566822	0,428064385	46	0,932885374	0,270838401
Примечание – Проводить процедуру стандартизации, которая рекомендуется перед проведением кластерного анализа нет необходимости, так как нормировка уже выполнена,					

Шаг 1. Скопируйте два столбца исходных данных таблицы 16.1 в

созданную пустую электронную таблицу **Spreadsheet.sta** **Statistica 6** (рисунок 16.1).

В таблице 16.1 левый информационный столбец показывает номер населенного пункта. Первый столбец (**Var1**) содержит данные о социально-экономической ситуации, второй (**Var2**) – о радиационно-экологической обстановке.



	1 Var1	2 Var2
1	0.247193	0.295401
2	0.490973	0.495497
3	0.768153	0.240149
4	0.837896	0.354305
5	0.920873	0.319777
6	0.836932	0.40867
7	0.642086	0.309196
8	0.754472	0.287697
9	0.844317	0.339891
10	0.433129	0.242212
11	0.92255	0.289111
12	0.823976	0.242543
13	0.962962	0.334659
14	0.800143	0.222998
15	0.828421	0.695623
16	1	0.365522

Рисунок 16.16 – Исходная информация

Шаг 2. Запустите модуль **Кластерный анализ**.

Шаг 3. В списке методов выделите **k-means** (к-средних) (рисунок 16.2) и нажмите кнопку **ОК** в правом верхнем углу панели.

Диалоговое окно метода **k-means** появится на экране. Выполните установки как показано на рисунке 16.17.

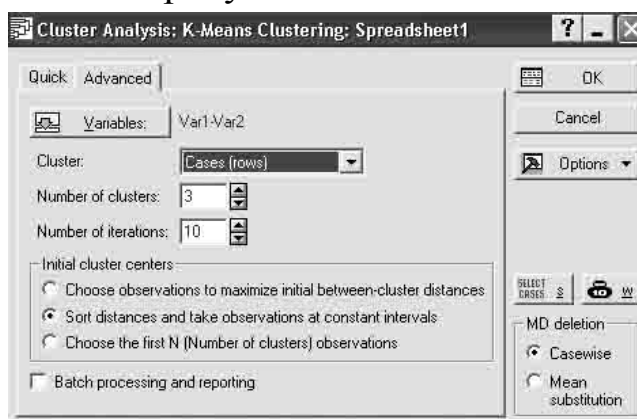


Рисунок 16.17 – Диалоговое окно метода k-means

Шаг 4. После того как все установки сделаны, нажмите кнопку **ОК** в верхнем правом углу окна **k-means Clustering** и запустите

вычислительную процедуру.

Шаг 5. После нажатия кнопки **ОК** в **k-means Clustering** окно результатов появится на экране (рисунок 16.18).

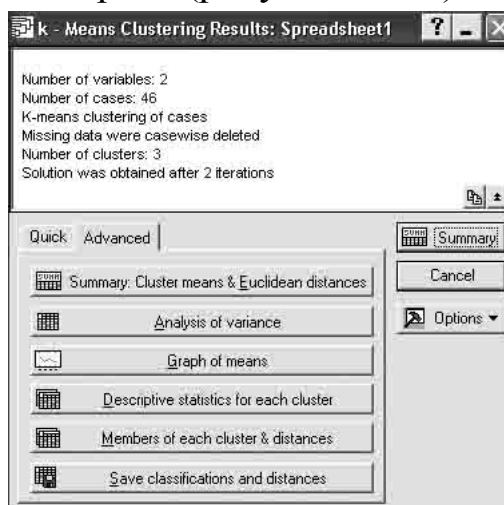


Рисунок 16.18 – Окно результатов кластеризации населенных пунктов

Шаг 6. Нажав кнопку **Member of each cluster&distances**, вы получите 3 таблицы о распределении населенных пунктов по кластерам (рисунки 16.19 – 16.21).

Members of Cluster Number 1 (Spreadsheet1) and Distances from Respective Cluster Center Cluster contains 9 cases									
Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.
C 1	C 2	C 10	C 23	C 25	C 26	C 29	C 35	C 38	
Distance	0.126932	0.108981	0.085823	0.083375	0.031304	0.060520	0.139921	0.072219	0.021170

Рисунок 16.19 – Первый кластер

Members of Cluster Number 2 (Spreadsheet1) and Distances from Respective Cluster Center Cluster contains 27 cases									
Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.
C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 11	C 12	
Distance	0.057823	0.036787	0.077444	0.072492	0.120046	0.043348	0.031556	0.079559	0.047952

Рисунок 16.20 – Второй кластер

Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	
C 15	C 18	C 20	C 22	C 24	C 32	C 34	C 37	C 41	
Distance	0.055052	0.100852	0.022462	0.070963	0.089924	0.082394	0.263728	0.145218	0.087989

Рисунок 16.21 – Третий кластер

В строках таблиц указано расстояние от каждого населенного пункта до центра кластера.

Шаг 7. Нажмите на кнопку **Cluster means&Euclidean distances**. На экране появится таблица, в которой даны евклидовы расстояния между средними кластеров (рисунок 16.22).

Cluster Number	No. 1	No. 2	No. 3
No. 1	0.000000	0.080589	0.103222
No. 2	0.283882	0.000000	0.059137
No. 3	0.321282	0.243181	0.000000

Рисунок 16.22 – Расстояния между кластерами

Над диагональю в таблице даны квадраты расстояний между кластерами.

Шаг 8. С помощью кнопки **Graph of means** (График средних) строятся следующие графики средних значений характеристик населенных пунктов для каждого кластера (рисунок 16.23).

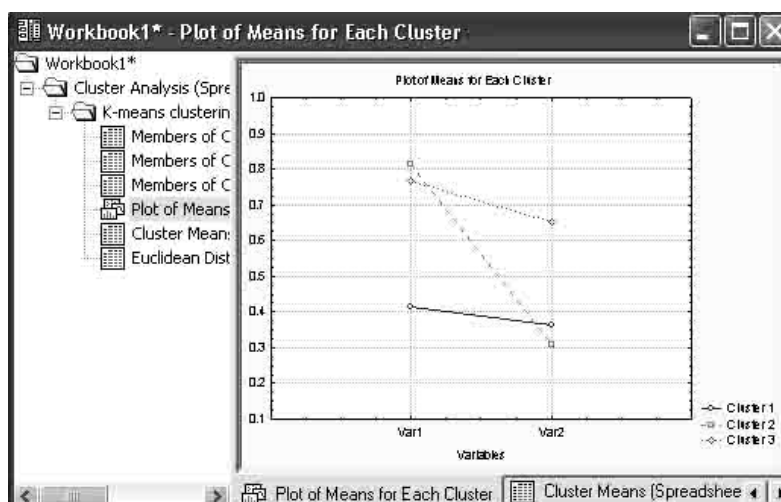


Рисунок 16.23 – График средних для каждого кластера

Полученную таблицу классификации населенных пунктов, предварительно отредактировав (таблица 16.2), переместите ее в модуль **Discriminant analysis** (Дискриминантный анализ).

Выполните процедуры дискриминантного анализа, описанные в разделе 13, и просмотрите разделение групп на графике (*Шаг 8* раздела 13). На экране появится график (рисунок 16.24).

Таблица 16.2 – Таблица результатов классификации

1 кластер			2 кластер			3 кластер		
№ НП	Соц-экон.	Радиолог.	№ НП	Соц-экон.	Радиолог.	№ НП	Соц-экон.	Радиолог.
1	0,25	0,30	3	0,77	0,24	15	0,83	0,70
2	0,49	0,50	4	0,84	0,35	18	0,89	0,57
10	0,43	0,24	5	0,92	0,32	20	0,74	0,63
23	0,51	0,43	6	0,84	0,41	22	0,68	0,59
25	0,45	0,34	7	0,64	0,31	24	0,76	0,78
26	0,47	0,43	8	0,75	0,29	32	0,76	0,53
29	0,30	0,20	9	0,84	0,34	34	0,89	1,00
35	0,38	0,46	11	0,92	0,29	37	0,61	0,51
38	0,44	0,37	12	0,82	0,24	41	0,75	0,53
			13	0,96	0,33	43	0,74	0,66
			14	0,80	0,22			
			16	1,00	0,37			
			17	0,72	0,35			
			19	0,66	0,28			
			21	0,68	0,36			
			27	0,85	0,30			
			28	0,76	0,25			
			30	0,82	0,44			
			31	0,89	0,25			
			33	0,83	0,36			
			36	0,75	0,30			
			39	0,85	0,23			
			40	0,79	0,40			
			42	0,89	0,37			
			44	0,61	0,24			
			45	0,77	0,23			
			46	0,93	0,27			

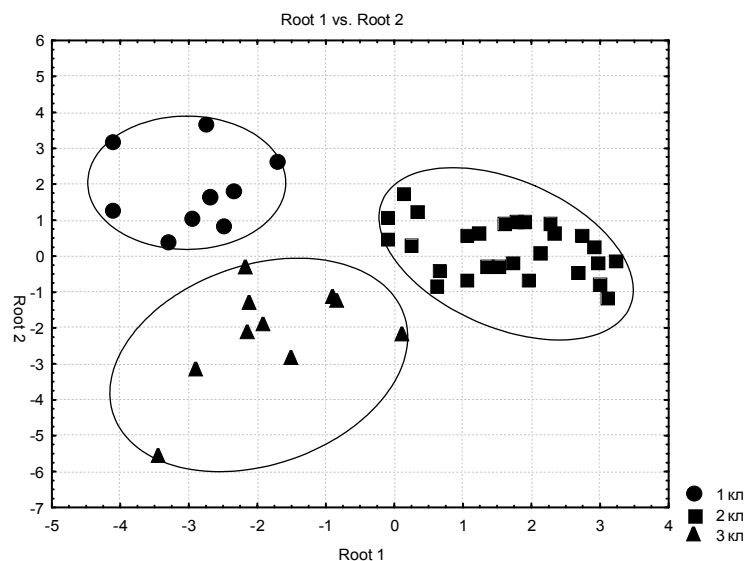


Рисунок 16.24 – Диаграмма рассеяния канонических значений

Из рисунка 16.24 видно четкое разделение на три класса и совместно с анализом результатов таблицы 2 можно рекомендовать:

1 Населенные пункты, объединенные в первый кластер, имеют крайне низкий демографический потенциал, а значит и социально-экономический фактор, т.е. «умирающие» населенные пункты. В данных населенных пунктах нецелесообразно развивать социально-экономическую структуру.

2 Населенные пункты, находящиеся во втором кластере приоритетны в распределении инвестиций, направленных на их социально-экономическое развитие. При этом наиболее перспективными являются населенные пункты, расположенные в правой части этого кластера. Эти пункты имеют стабильный социально-экономический фактор, характеризуются стабильной радиационной обстановкой и в них нецелесообразно проводить защитные мероприятия.

3 Населенные пункты третьего кластера характеризуются повышенными значениями активности ^{137}Cs в молоке и суммарной годовой дозой. Поэтому в этих населенных пунктах необходимо рекомендовать проведение контрмер. При этом мероприятия должны быть первоочередными в населенных пунктах, расположенных в правой и верхней части первого кластера.

Задания для выполнения

- 1 Создайте пустую электронную таблицу **Spreadsheet.sta**.

2 Внесите данные для выполнения расчетов, находящиеся в 8 Приложении И (таблица И1).

3 Выполните процедуры кластерного анализа в соответствии с порядком операций, выполненных в настоящем разделе.

4 Дайте объяснение полученным результатам.

Литература

- 1 Лакин, Г. Ф. Биометрия: учебн. пос. / Г. Ф. Лакин. – М. : «Высшая школа», 1990. – 142 с.
- 2 Зайцев, Г. Н. Математическая статистика в экспериментальной ботанике: учебн. пос. / Г. Н. Зайцев. – М. : «Наука», 1984. – 424 с.
- 3 Плохинский, Н. А. Биометрия: 2-е издание учебн. пос. / Н. А. Плохинский. – М. : «МГУ», 1970. – 368 с.
- 4 Свалов, Н. Н. Вариационная статистика / Н. Н. Свалов. – М. : «Лесная промышленность», 1977. – 177 с.
- 5 Рокитский, П. Ф. Биологическая статистика: изд. 3-е испр. учебн. пос. / П. Ф. Рокитский. – Мн. : «Высшая школа», 1973. – 320 с.
- 6 Боровиков, В. П. Популярное введение в программу Statistica / В. П. Боровиков. – М. : «КомпьютерПресс», 1998. – 69 с.
- 7 Жученко, Ю. М. Статистическая обработка информации с применением персональных компьютеров: практическое руководство для студентов 5 курса / Ю. М. Жученко. – Гомель, УО ГГУ им. Ф. Скорины, 2007. – 101 с.
- 8 Мюллер, П. Таблицы по математической статистике: справочник / П. Мюллер [и др.]. – М. : «Финансы и статистика», 1982. – 64 с.
- 9 Павловский, З. Введение в математическую статистику: перевод с польск. / З. Павловский. – М. : «Статистика», 1967. – 285 с.
- 10 Карасев, А. И. Теория вероятностей и математическая статистика: учебн. пос. / А. И. Карасев. – М. : «Статистика», 1979. – 279 с.
- 11 Бейли, Н. Математика в биологии и медицине: перевод с англ. / Н. Бейли. – М. : «Мир», 1970. – 167 с.
- 12 Урбах, В. Ю. Статистический анализ в биологических и медицинских исследованиях: учебн. пос. / В. Ю. Урбах. – М. : «Медицина», 1975. – 321 с.
- 13 Лапач, С. Н. Статистические методы в медико-биологических исследованиях с использованием Excel / С. Н. Лапач [и др.]. – К. : «МОРИОН», 2000. – 196 с.
- 14 Реброва, О. Ю. Статистический анализ медицинских данных: применение пакета прикладных программ STATISTICA / О. Ю. Реброва. – М.: «МедиаСфера», 2002. – 84 с.

Приложение А (обязательное)

Таблица А1 Задания к лабораторным работам 4 и 5

Варианты заданий														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
13,45	81,92	3,00	20,15	5,48	8	12	6	39	27	1	20	3	29	30
7,35	147,02	2,17	8,79	9,35	2	22	12	49	32	19	20	14	55	71
8,53	109,30	2,53	12,56	3,94	13	1	12	51	22	3	22	11	39	34
16,58	91,55	1,73	5,65	9,94	4	9	15	48	13	16	25	6	46	40
12,86	79,85	2,74	15,44	1,52	22	8	14	102	19	14	22	28	54	51
10,11	80,34	0,78	2,18	13,31	19	11	6	61	12	20	2	15	89	80
17,22	95,98	2,01	7,49	1,67	5	1	1	73	16	12	1	20	71	44
8,72	95,22	1,66	5,26	5,67	11	5	11	63	27	20	11	14	40	39
20,76	150,85	2,73	15,38	12,35	12	12	12	30	1	11	12	3	55	23
5,30	108,46	1,48	4,38	10,36	23	5	14	107	28	18	19	12	69	65
91,55	43,99	1,99	7,30	11,69	6	17	7	37	10	32	14	1	29	52
79,85	106,15	2,21	9,11	5,00	14	7	1	43	17	7	23	1	67	47
80,34	135,03	2,95	19,13	1,84	12	6	18	21	35	9	10	47	69	71
95,98	119,52	2,39	10,90	7,19	6	14	12	44	21	18	2	21	41	40
95,22	111,39	2,28	9,80	2,25	12	7	12	40	13	23	4	26	84	51
4,48	87,89	2,84	17,15	5,44	1	20	4	45	3	17	13	24	20	47
4,35	77,79	2,28	9,76	1,90	10	8	18	44	10	22	22	2	63	44
5,00	147,68	2,26	9,54	7,90	5	7	8	42	5	11	13	8	52	74
4,85	127,11	1,77	5,84	1,98	11	18	11	63	9	13	11	1	74	61
4,99	146,63	2,26	9,62	14,39	21	7	2	33	21	36	5	24	75	64
4,86	129,39	1,07	2,92	9,06	11	7	12	44	31	42	14	16	33	51
4,49	89,21	2,85	17,21	14,61	11	6	8	86	19	18	36	11	52	58
	132,30	1,65	5,19	3,67	18	11	12	49	2	16	2	20	59	48
	133,92	1,72	5,56	6,23	1	3	16	38	2	28	18	26	35	70
	111,96	1,20	3,31	18,84	1	15	12	74	46	23	1	20	63	50
	127,69	2,15	8,59	42,65	11	5	8	17	19	22	11	1	33	63
	118,32	1,99	7,32	3,92	10	1	16	53	17	1	5	12	48	39
	55,86	1,75	5,75	11,07	9	6	12	30	9	16	9	18	61	79
	155,82	0,15	1,16	6,22	23	8	20	73	32	6	13	2	13	9
	54,09	1,50	4,49		14	15	11	30	5	22	18	15	29	21
	69,98	2,55	12,75										34	59
	40,91	2,57	13,08											
	117,10	2,87	17,57											
	80,15	2,32	10,13											
	59,97	2,56	12,90											
	104,08	2,00	7,38											
	77,97	2,16	8,66											
	134,53	2,56	12,89											
	47,15	1,02	2,78											

Приложение Б (обязательное)

Таблица Б1 Задания к лабораторным работам 6 и 7

Варианты заданий													
1		2		3		4		5		6		7	
X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14,7	171,0	5,2	17,5	5,2	17,5	12,0	3,6	1,4	28,4	6,0	3,9	26,8	5,1
13,7	202,2	5,4	17,6	5,4	17,6	12,0	3,7	1,8	28,4	8,2	2,9	47,3	5,0
15,2	124,6	5,7	17,6	5,7	17,6	12,0	4,0	1,9	28,5	6,3	2,1	49,6	4,7
16,3	119,0	6,2	17,6	6,2	17,6	11,9	4,6	2,1	28,7	18,1	1,9	58,0	4,6
16,2	124,3	6,7	17,8	6,7	17,8	11,9	4,9	2,3	29,0	13,9	3,8	59,3	4,5
14,8	175,2	7,1	18	7,1	18	11,7	5,0	2,4	29,0	4,3	2,8	59,7	4,4
16,1	161,7	7,5	18	7,5	18	11,3	5,0	2,4	29,2	11,2	2,9	63,3	4,3
13,9	169,4	7,7	18,1	7,7	18,1	11,2	5,6	2,5	29,3	11,3	2,4	68,5	4,3
14,3	199,9	7,8	18,1	7,8	18,1	10,3	5,9	2,5	29,3	2,9	3,4	73,0	4,2
13,3	115,6	7,9	18,6	7,9	18,6	10,0	6,2	2,7	29,4	20,5	4,0	73,2	4,2
14,0	155,0	8	19,7	8	19,7	9,6	6,3	2,7	29,5	9,4	2,2	74,3	4,2
14,2	141,8	8,3	19,7	8,3	19,7	9,5	6,3	2,9	29,5	15,5	2,4	76,9	4,1
14,4	186,0	8,4	19,8	8,4	19,8	9,4	6,3	2,9	29,5	7,7	1,9	78,1	3,8
14,6	159,4	8,5	19,9	8,5	19,9	9,3	6,4	2,9	29,6	11,3	2,0	80,3	3,7
15,1	122,0	8,8	19,9	8,8	19,9	9,2	6,4	3,0	29,6	7,1	3,7	83,0	3,5
14,6	72,2	8,9	20,1	8,9	20,1	9,1	6,5	3,0	29,7	-0,2	2,4	84,0	3,4
14,7	123,8	8,9	20,1	8,9	20,1	9,0	6,6	3,1	29,7	17,1	4,3	88,2	3,4
14,6	166,5	8,9	20,1	8,9	20,1	8,9	7,1	3,2	29,9	12,9	2,8	92,7	3,4
16,3	117,0	8,9	20,1	8,9	20,1	8,6	7,1	3,3	29,9	7,7	3,8	93,6	3,4
14,9	174,4	9	20,2	9	20,2	8,2	7,1	3,3	29,9	10,5	3,7	95,6	3,3
14,8	186,8	9	20,3	9	20,3	8,2	7,2	3,4	30,0	16,5	3,6	95,9	3,2
14,5	131,1	9,1	20,3	9,1	20,3	8,2	7,4	3,4	30,1	12,1	4,8	96,4	3,0
15,9	199,6	9,3	20,5	9,3	20,5	8,2	7,5	3,5	30,1	9,3	1,9	97,2	3,0
14,3	93,3	9,4	20,6	9,4	20,6	8,2	7,6	3,5	30,1	8,8	3,0	102,1	2,9
13,4	135,6	9,7	20,9	9,7	20,9	7,9	8,0	3,5	30,1	7,6	2,8	102,2	2,9
15,5	121,8	9,7	21	9,7	21	7,7	8,1	3,5	30,2	8,7	3,1	106,1	2,8
15,9	127,6	9,9	21,1	9,9	21,1	7,5	8,2	3,6	30,3	7,3	3,5	106,4	2,8
14,9	149,5	10,1	21,1	10,1	21,1	7,3	8,3	3,6	30,4	1,9	3,7	109,7	2,8
14,5	131,7	10,4	21,2	10,4	21,2	7,3	8,3	3,6	30,4	4,6	3,5	119,4	2,7
15,7	156,5	10,7	21,3	10,7	21,3	7,2	8,8	3,6	30,5	10,7	3,2	131,0	2,6
14,6	185,5	11,6	21,3	11,6	21,3	7,0	9,1	3,7	30,6	5,8	3,0	133,7	2,6
15,8	129,3	12	21,8	12	21,8	6,9	9,3	3,8	30,6	10,6	3,4	134,8	2,4
13,6	223,6	12,3	22	12,3	22	6,3	9,6	3,9	30,6	13,1	3,2	136,1	2,4
14,2	198,6	12,6	22,1	12,6	22,1	6,1	9,6	3,9	30,7	12,9	3,3	138,8	2,4
13,5	180,8	12,7	22,4	12,7	22,4	0,7	29,5	4,0	30,8	8,7	1,6	139,1	2,4
14,6	132,9	13,3	22,7	13,3	22,7	35,9	1,1	4,0	31,0	11,8	2,3	139,4	2,3
15,0	172,4	13,6	23,5	8,8	44	1,1	43,0	4,0	31,1	9,0	5,0	161,0	2,0

Приложение В (обязательное)

Таблица В1 Задания к лабораторной работе 8

Варианты заданий														
1			2			3			4			5		
Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y
2419	2,69	0,69	5,8	0,057	2	31,4	0,04	2	0,021	38	2	2,69	7,25	0,69
2309	3,83	0,83	7,7	0,059	2,3	31,4	0,045	2,3	0,025	34,7	2,3	3,83	14,69	0,83
2205	4,99	0,99	8,5	0,061	2,7	30,4	0,048	2,7	0,044	31,5	2,7	4,99	24,93	0,99
2110	6,06	1,06	9,9	0,068	2,9	27,1	0,062	2,9	0,045	30,6	2,9	6,06	36,78	1,06
2018	7,46	1,46	10,1	0,072	4,3	26,6	0,063	4,3	0,046	26,9	4,3	7,46	55,63	1,46
1919	8,61	1,61	11,2	0,073	5	26,5	0,066	5	0,05	26,8	5	8,61	74,12	1,61
1831	9,77	1,77	12,8	0,074	5,9	25,2	0,066	5,9	0,05	26,4	5,9	9,77	95,55	1,77
1744	10,81	1,81	12,9	0,078	6,1	24,8	0,068	6,1	0,056	24,7	6,1	10,81	116,82	1,81
1661	11,89	1,89	14	0,078	6,6	24,6	0,073	6,6	0,061	24	6,6	11,89	141,30	1,89
1578	12,96	1,96	14,2	0,078	7,1	24,5	0,079	7,1	0,063	23,7	7,1	12,96	167,96	1,96
1497	13,97	1,97	14,6	0,079	7,2	24,1	0,081	7,2	0,068	23,3	7,2	13,97	195,27	1,97
1420	15,09	2,09	15,4	0,081	8,1	23,7	0,083	8,1	0,069	23,1	8,1	15,09	227,76	2,09
1338	16,12	2,12	16,1	0,082	8,3	23,7	0,084	8,3	0,071	23	8,3	16,12	259,73	2,12
1263	17,12	2,12	16,2	0,083	8,3	23,3	0,084	8,3	0,072	22,9	8,3	17,12	292,97	2,12
1191	18,13	2,13	16,8	0,085	8,4	22,5	0,085	8,4	0,072	22,8	8,4	18,13	328,63	2,13
1120	19,14	2,14	17,1	0,086	8,5	22,3	0,092	8,5	0,074	22,7	8,5	19,14	366,34	2,14
1054	20,17	2,17	17,1	0,088	8,8	22,1	0,096	8,8	0,075	22,5	8,8	20,17	407,02	2,17
988	21,17	2,17	17,3	0,088	8,8	22	0,098	8,8	0,078	22,4	8,8	21,17	448,37	2,17
925	22,20	2,20	17,7	0,09	9	21,8	0,098	9	0,085	21,9	9	22,20	492,72	2,20
865	23,22	2,22	17,7	0,092	9,2	20,8	0,099	9,2	0,088	21,4	9,2	23,22	539,13	2,22
805	24,24	2,24	17,8	0,092	9,4	20	0,1	9,4	0,089	19,9	9,4	24,24	587,61	2,24
748	25,25	2,25	18,7	0,093	9,5	20	0,102	9,5	0,09	19,7	9,5	25,25	637,63	2,25
691	26,30	2,30	19,5	0,096	10	19,9	0,102	10	0,092	19,7	10	26,30	691,83	2,30
637	27,35	2,35	19,7	0,099	10,5	19,8	0,103	10,5	0,092	19,6	10,5	27,35	748,10	2,35
587	28,38	2,38	20,4	0,099	10,8	19,7	0,104	10,8	0,094	19,4	10,8	28,38	805,40	2,38
539	29,42	2,42	20,7	0,106	11,2	19,6	0,109	11,2	0,095	18,4	11,2	29,42	865,30	2,42
492	30,42	2,42	21	0,107	11,3	18,7	0,109	11,3	0,096	18	11,3	30,42	925,67	2,42
448	31,43	2,43	21,7	0,108	11,4	18,3	0,11	11,4	0,097	17,7	11,4	31,43	988,07	2,43
407	32,48	2,48	22,5	0,108	11,9	18,2	0,115	11,9	0,102	17,3	11,9	32,48	1054,73	2,48
366	33,48	2,48	22,5	0,11	11,9	18	0,115	11,9	0,103	16,7	11,9	33,48	1120,68	2,48
328	34,52	2,52	22,6	0,112	12,4	17,2	0,116	12,4	0,106	16,7	12,4	34,52	1191,47	2,52
292	35,55	2,55	22,8	0,112	12,8	16,8	0,117	12,8	0,106	15,9	12,8	35,55	1263,76	2,55
259	36,58	2,58	22,9	0,112	13,2	16,8	0,118	13,2	0,106	15,8	13,2	36,58	1338,11	2,58
227	37,69	2,69	23,5	0,113	14,8	15,8	0,122	14,8	0,107	15,6	14,8	37,69	1420,88	2,69
195	38,70	2,70	23,5	0,113	14,9	15,7	0,123	14,9	0,113	15,4	14,9	38,70	1497,80	2,70
167	39,73	2,73	23,9	0,12	15,3	15,6	0,124	15,3	0,122	14,6	15,3	39,73	1578,30	2,73
141	40,76	2,76	24,5	0,12	15,8	15	0,125	15,8	0,124	14,2	15,8	40,76	1661,38	2,76

Приложение Г (обязательное)

Таблица Г1 Задания к лабораторным работам 9 и 10

Варианты заданий													
1		2		3		4		5		6		7	
X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
14,5	49,8	33,3	0,03	9,06	2,13	0,9	66,9	0,97	0,4	1,5	0,7	0,87	0,15
5,1	22,5	7,1	0,2	8,95	2,12	0,01	0,01	1,21	2,79	1,47	0,69	0,98	0,11
6,1	28,2	11,2	0,13	5,58	1,61	-0,8	-46,4	0,83	0,09	1,58	0,73	1,32	0,04
6,5	31,2	14,3	0,08	11,26	2,38	1,2	165,1	1,16	1,83	1,57	0,73	1,22	0,05
6,7	33,5	17,2	0,06	9,14	2,14	1,5	314,8	0,95	0,31	1,45	0,68	0,81	0,18
8,2	41,8	22,2	0,05	2,87	0,83	1,1	141,1	1,53	22,9	1,45	0,69	1,47	0,02
7,4	34,9	34,7	0,03	12,01	2,42	-0,1	-0,1	1,09	1,13	1,02	0,51	0,96	0,11
10,7	46,9	11,2	0,1	6,86	1,81	0,4	5,8	0,48	0	1,18	0,58	0,77	0,2
14,3	48,8	3,7	0,37	7	1,89	0,8	60,5	1,1	1,15	0,92	0,46	0,93	0,12
8,1	40,3	33,8	0,03	2,86	0,99	-0,1	-0,3	1,15	1,73	0,96	0,48	0,88	0,14
8,1	40,3	10,9	0,12	7,59	1,97	1,2	175	1,04	0,73	1,78	0,79	0,96	0,11
4,6	21,3	13,1	0,08	10,47	2,25	2,6	1845,6	1,23	3,15	0,55	0,2	0,73	0,23
8,4	42,2	18,1	0,07	6,78	1,77	1,6	421,1	0,75	0,04	1,21	0,6	1,56	0,02
9,6	44,9	20,1	0,06	2,57	0,69	1,2	192,9	1,02	0,62	0,86	0,43	1,25	0,05
8,8	43,9	23,4	0,05	7,59	1,96	0,7	42,4	1,11	1,3	0,61	0,26	1,27	0,04
9,1	44,3	30,1	0,04	9,59	2,17	-0,01	-0,01	1,14	1,63	1,09	0,54	0,68	0,26
9,6	44,8	27	0,04	11,08	2,35	0,02	0,02	1,35	7,22	0,86	0,43	1,12	0,07
7,4	36,2	25,4	0,04	12,33	2,43	0,2	1,1	1,11	1,31	1,12	0,56	1,53	0,02
8,5	43,9	25,6	0,04	11,76	2,42	1,6	443,8	1,78	88,7	0,53	0,19	0,72	0,23
11,2	47,1	30,1	0,04	9,49	2,2	0,4	8,4	0,79	0,06	1,16	0,58	1,13	0,07
12,7	47,9	27,7	0,04	12,2	2,48	-0,3	-2,8	0,84	0,1	1,05	0,52	1,12	0,07
13,4	48,2	26	0,05	9,62	2,22	-0,5	-13,1	0,83	0,1	0,57	0,22	0,94	0,12
13,8	48,7	23,2	0,05	12,53	2,48	0,8	50	1,21	2,76	0,9	0,45	1,32	0,04
7,2	34,1	19	0,06	3,51	1,06	0,3	4,3	1,2	2,61	0,75	0,36	0,95	0,12
8	40,3	32,1	0,03	4,76	1,46	0,6	19,8	1,31	5,75	0,86	0,43	1,2	0,06
		16,8	0,07	10,41	2,3	0,2	1,1	1,67	49,21	0,77	0,37	0,72	0,23
		12,1	0,09	8,71	2,09	0,6	23,2	1,18	2,29	1,61	0,74	0,69	0,25
		9,3	0,17	8,7	2,12	1,4	255,9	0,65	0,01	0,68	0,31	0,96	0,11
		6,7	0,26	9,68	2,17	0,8	44,8	0,89	0,17	1,01	0,51	0,9	0,14
		21,5	0,05	10,04	2,24	1,5	326,4	0,68	0,01	1,36	0,65	1,67	0,01

Приложение Д (обязательное)

Таблица Д1 Задания к лабораторным работам 11 и 12

Варианты заданий											
1				2				3			
a	b	c	d	a	b	c	d	a	b	c	d
9,4	21,1	12,5	21,0	1,95	6,87	1,47	1,94	24,4	21,5	17,3	33,0
8,4	8,2	12,1	5,0	2,87	3,34	2,77	2,05	15,8	23,2	22,8	14,2
4,9	16,6	12,1	24,1	1,89	3,32	3,06	1,69	19,5	16,8	11,0	18,3
13,4	10,5	15,4	14,8	2,97	6,19	2,58	2,34	16,5	20,5	17,8	27,2
11,1	18,4	2,0	29,1	1,70	5,75	2,38	2,41	16,5	8,2	23,5	5,6
3,6	11,3	9,5	20,1	2,05	5,61	3,25	1,21	15,0	19,4	11,9	22,6
15,1	18,3	12,1	11,2	3,02	7,33	3,32	2,83	21,4	28,6	15,0	31,4
13,2	11,2	11,0	19,5	1,57	4,43	3,02	1,56	17,5	21,3	24,6	26,4
12,0	19,4	5,8	20,5	1,49	4,45	2,10	1,81	17,2	19,8	19,9	33,9
11,6	10,0	16,0	22,9	2,28	5,84	3,25	2,61	11,5	12,7	20,8	26,6
18,2	11,2	12,1	17,7	3,29	5,20	2,57	2,43	11,5	23,5	22,8	14,1
12,4	11,1	16,6	13,8	1,91	3,68	2,86	1,36	12,0	20,2	29,2	20,7
15,2	6,3	15,5	9,3	3,32	4,09	3,54	2,25	15,3	19,1	25,4	25,8
9,8	16,3	11,5	20,4	2,21	7,21	2,76	1,60	30,2	29,8	19,8	21,7
8,0	18,8	9,3	16,7	4,32	3,28	3,32	1,03	33,3	25,9	25,6	29,4
12,0	10,1	14,4	18,3	1,57	6,31	2,89	1,53	15,4	22,5	10,8	27,0
13,3	17,6	12,2	29,1	1,07	4,69	2,51	1,84	32,1	31,9	13,1	23,2
12,0	6,6	13,2	13,1	2,19	5,98	2,61	1,34	40,1	11,1	12,3	29,9
16,7	14,7	6,2	19,5	0,02	4,00	3,30	2,35	22,3	21,3	14,3	17,8
7,3	18,2	8,1	20,0	2,06	5,27	1,97	1,88	25,0	14,4	24,2	38,9
8,7	5,5	13,9	12,4								
13,1	17,7	12,2	14,3								
7,6	11,9	5,3	15,5								
11,2	5,5	15,7	10,4								
8,5	10,4	14,2	31,7								
19,4	9,2	10,5	11,5								
10,6	10,7	13,6	21,2								
7,3	10,8	10,1	15,5								
17,9	13,8	13,0	15,5								
20,3	19,1	11,4	26,1								
17,2	15,6	13,3	15,7								
13,3	12,9	12,6	24,0								
10,2	14,7	12,5	20,7								
15,9	15,2	9,8	13,7								
15,0	14,8	17,6	15,0								
10,1	9,0	11,0	9,0								

Окончание таблицы Д1

Варианты заданий												
4				5				6				
a	b	c	d	a	b	c	d	a	b	c	d	e
5,8	9,5	31,3	1,1	9,4	21,1	12,5	21,0	2,30	4,72	0,70	1,00	0,21
7,8	10,6	10,8	0,7	8,4	8,2	12,1	5,0	1,20	4,55	1,37	0,94	0,43
12,4	3,1	27,8	1,6	4,9	16,6	12,1	24,1	1,55	1,64	0,91	1,06	0,31
8,5	24,9	19,8	0,9	13,4	10,5	15,4	14,8	2,17	4,01	1,50	1,20	0,52
5,7	10,0	18,1	1,1	11,1	18,4	2,0	29,1	2,21	5,34	1,31	1,24	0,34
10,7	19,0	28,2	0,7	3,6	11,3	9,5	20,1	1,77	3,47	1,07	1,20	0,02
11,3	16,2	43,7	1,6	15,1	18,3	12,1	11,2	0,76	6,03	1,39	0,97	0,34
12,5	20,9	19,1	0,7	13,2	11,2	11,0	19,5	2,23	5,06	1,15	0,95	0,24
4,7	2,1	28,0	1,0	12,0	19,4	5,8	20,5	2,44	4,25	1,19	0,83	0,46
11,8	24,5	20,1	1,1	11,6	10,0	16,0	22,9	2,04	3,61	0,98	0,73	0,60
5,4	17,2	15,0	1,1	18,2	11,2	12,1	17,7	2,38	3,55	1,08	0,95	0,31
10,3	13,5	20,7	0,8	12,4	11,1	16,6	13,8	2,84	4,54	1,27	0,97	0,28
6,5	14,7	12,6	0,9	15,2	6,3	15,5	9,3	2,46	3,19	1,22	1,14	0,62
10,1	7,2	11,5	1,1	9,8	16,3	11,5	20,4	1,53	4,23	1,44	1,13	0,11
14,2	16,0	16,6	1,0	8,0	18,8	9,3		1,68	4,84	0,80	0,83	0,81
9,8	4,9	22,9	1,1	12,0	10,1	14,4		3,20	3,97	1,04	0,82	
10,0	19,4	12,0	1,3	13,3	17,6	12,2		1,45	5,10	0,68	0,98	
12,4	11,6	21,0	1,4	12,0	6,6	13,2		1,72	4,87	1,18	1,08	
6,7	15,8	15,1	0,7	16,7	14,7	6,2		2,30	2,06	0,99		
5,7	7,8	26,2	0,7	7,3	18,2	8,1		1,74	5,21	0,94		
12,1	16,8	24,0	1,1	8,7	5,5	13,9		2,35	3,50			
8,0	19,3	26,2	0,9		17,7	12,2		2,43	5,31			
10,4	19,1	8,0	1,6		11,9	5,3		1,16	3,81			
9,8	17,1	35,8	0,9		5,5	15,7		2,30	4,61			
					10,4	14,2		1,66	3,55			
						10,5			3,17			
						13,6			3,93			
						10,1			4,82			
						13,0			3,91			
						11,4			5,68			
						13,3						
						12,6						
						12,5						
						9,8						
						17,6						
						11,0						

Приложение Е (обязательное)

Таблица Е1 Задания к лабораторным работам 13 и 14

Варианты заданий														
1					2					3				
	a	b	c	d		a	b	c	d		a	b	c	d
A	0,02	3,28	1,47	1,03	A	14,2	24,9	43,7	1,1	A	3,6	21,1	2,0	5,0
A	1,07	3,32	1,97	1,21	A	12,5	24,5	35,8	0,7	A	4,9	19,4	5,3	9,0
A	1,49	3,34	2,10	1,34	A	12,4	20,9	31,3	1,6	A	7,3	19,1	5,8	9,3
A	1,57	3,68	2,38	1,36	A	12,4	19,4	28,2	0,9	A	7,3	18,8	6,2	10,4
A	1,57	4,00	2,51	1,53	A	12,1	19,3	28,0	1,1	A	7,6	18,4	8,1	11,2
A	1,70	4,09	2,57	1,56	A	11,8	19,1	27,8	0,7	A	8,0	18,3	9,3	11,5
A	1,89	4,43	2,58	1,60	A	11,3	19,0	26,2	1,6	A	8,4	18,2	9,5	12,4
A	1,91	4,45	2,61	1,69	A	10,7	17,2	26,2	0,7	A	8,5	17,7	9,8	13,1
A	1,95	4,69	2,76	1,81	B	10,4	17,1	24,0	1,0	A	8,7	17,6	10,1	13,7
A	2,05	5,20	2,77	1,84	B	10,3	16,8	22,9	1,1	A	9,4	16,6	10,5	13,8
B	2,06	5,27	2,86	1,88	B	10,1	16,2	21,0	1,1	A	9,8	16,3	11,0	14,3
B	2,19	5,61	2,89	1,94	B	10,0	16,0	20,7	0,8	A	10,1	15,6	11,0	14,8
B	2,21	5,75	3,02	2,05	B	9,8	15,8	20,1	0,9	B	10,2	15,2	11,4	15,0
B	2,28	5,84	3,06	2,25	B	9,8	14,7	19,8	1,1	B	10,6	14,8	11,5	15,5
B	2,87	5,98	3,25	2,34	B	8,5	13,5	19,1	1,0	B	11,1	14,7	12,1	15,5
B	2,97	6,19	3,25	2,35	B	8,0	11,6	18,1	1,1	B	11,2	14,7	12,1	15,5
B	3,02	6,31	3,30	2,41	C	7,8	10,6	16,6	1,3	B	11,6	13,8	12,1	15,7
B	3,29	6,87	3,32	2,43	C	6,7	10,0	15,1	1,4	B	12,0	12,9	12,1	16,7
B	3,32	7,21	3,32	2,61	C	6,5	9,5	15,0	0,7	B	12,0	11,9	12,2	17,7
B	4,32	7,33	3,54	2,83	C	5,8	7,8	12,6	0,7	B	12,0	11,3	12,2	18,3
					C	5,7	7,2	12,0	1,1	B	12,4	11,2	12,5	19,5
					C	5,7	4,9	11,5	0,9	B	13,1	11,2	12,5	19,5
					C	5,4	3,1	10,8	1,6	B	13,2	11,1	12,6	20,0
					C	4,7	2,1	8,0	0,9	B	13,3	10,8	13,0	20,1
										C	13,3	10,7	13,2	20,4
										C	13,4	10,5	13,3	20,5
										C	15,0	10,4	13,6	20,7
										C	15,1	10,1	13,9	21,0
										C	15,2	10,0	14,2	21,2
										C	15,9	9,2	14,4	22,9
										C	16,7	9,0	15,4	24,0
										C	17,2	8,2	15,5	24,1
										C	17,9	6,6	15,7	26,1
										C	18,2	6,3	16,0	29,1
										C	19,4	5,5	16,6	29,1
										C	20,3	5,5	17,6	31,7

Окончание таблицы Е1

Варианты заданий																
4				5			6			7						
	a	b	c	d		a	b	c		a	b		a	b	c	d
A	15,5	18,8	59,9	0,8	A	0,55	0,62	2,12	A	9,4	21,1	A	9,4	21,1	12,5	21,0
A	21,2	23,9	62,7	33,9	A	1,32	2,68	4,77	A	8,4	8,2	A	8,4	8,2	12,1	5,0
A	15,5	25,3	56,9	67,4	A	1,12	2,36	2,34	A	4,9	16,6	A	4,9	16,6	12,1	24,1
A	19,0	26,4	46,2	87,5	A	0,93	1,40	2,79	A	13,4	10,5	A	13,4	10,5	15,4	14,8
A	12,4	26,4	86,2	88,2	A	0,95	1,66	2,85	A	11,1	18,4	A	11,1	18,4	2,0	29,1
A	11,6	27,2	34,5	98,0	A	0,87	0,87	1,05	A	3,6	11,3	A	3,6	11,3	9,5	20,1
A	15,8	27,8	44,6	108,1	A	0,51	0,02	1,12	B	20,3	19,1	B	15,1	18,3	12,1	11,2
B	9,0	28,0	52,7	109,0	B	0,99	1,50	4,85	B	11,2	5,5	B	13,2	11,2	11,0	19,5
B	11,9	29,6	46,3	117,6	B	1,44	2,05	4,54	B	15,1	18,3	B	12,0	19,4	5,8	20,5
B	24,1	30,7	65,0	135,7	B	1,44	3,63	4,75	B	13,2	11,2	B	11,6	10,0	16,0	22,9
B	13,5	31,0	51,5	135,8	B	0,14	2,45	4,59	B	12,0	19,4	B	18,2	11,2	12,1	17,7
B	18,4	31,6	52,5	142,5	B	1,55	1,76	1,41	B	11,6	10,0	B	12,4	11,1	16,6	13,8
B	21,2	32,2	71,7	146,0	B	0,72	0,33	4,29	C	18,2	11,2	C	15,2	6,3	15,5	9,3
B	7,1	32,2	68,5	148,3	B	1,02	1,29	3,42	C	12,4	11,1	C	9,8	16,3	11,5	20,4
C	9,9	32,8	56,6	159,0	C	0,92	3,08	4,87	C	15,2	6,3	C	8,0	18,8	9,3	16,7
C	15,1	33,1	65,8	161,1	C	0,92	1,16	5,40	C	7,6	11,9	C	12,0	10,1	14,4	18,3
C	20,0	33,2	34,2	174,4	C	2,09	4,45	5,12	C	9,8	16,3	C	13,3	17,6	12,2	29,1
C	15,0	35,3	70,4	198,5	C	1,71	1,09	4,23	C	8,0	18,8	C	12,0	6,6	13,2	13,1
C	15,9	36,0	71,9	200,6	C	0,85	1,30	3,87	D	12,0	10,1	D	16,7	14,7	6,2	19,5
C	18,7	36,6	45,4	235,4	C	1,27	2,40	4,43	D	13,3	17,6	D	7,3	18,2	8,1	20,0
C	4,1	39,3	45,6	261,8	C	0,46	0,72	4,80	D	12,0	6,6	D	8,7	5,5	13,9	12,4
					D	0,76	0,01	1,48	D	16,7	14,7	D	13,1	17,7	12,2	14,3
					D	0,59	3,39	2,65	D	7,3	18,2	D	7,6	11,9	5,3	15,5
					D	1,01	3,20	2,94	D	8,7	5,5	D	11,2	5,5	15,7	10,4
					D	0,54	1,79	4,30	E	13,1	17,7	E	8,5	10,4	14,2	31,7
					D	1,40	1,36	5,88	E	8,5	10,4	E	19,4	9,2	10,5	11,5
					D	1,12	2,24	4,06	E	19,4	9,2	E	10,6	10,7	13,6	21,2
					D	0,41	3,35	3,67	E	10,6	10,7	E	7,3	10,8	10,1	15,5
									E	7,3	10,8	E	17,9	13,8	13,0	15,5
									E	17,9	13,8	E	20,3	19,1	11,4	26,1
												F	17,2	15,6	13,3	15,7
												F	13,3	12,9	12,6	24,0
												F	10,2	14,7	12,5	20,7
												F	15,9	15,2	9,8	13,7
												F	15,0	14,8	17,6	15,0
												F	10,1	9,0	11,0	9,0

Приложение Ж (обязательное)

Таблица Ж1 Задания к лабораторной работе 15

1 задание					2 задание			3 задание		
Класс	1 призн.	2 призн.	3 призн.	4 призн.	Класс	1 призн.	2 призн.	Класс	1 призн.	2 призн.
a	1,01	0,48	0,87	1,12	d	0,247	0,295	a	9,23	5,26
a	1,14	1,11	1,38	1,17	d	0,491	0,495	c	19,82	14,00
a	1,22	1,44	1,02	0,79	c	0,768	0,240	a	11,59	8,06
a	1,09	0,94	0,91	1,1	c	0,838	0,354	a	11,06	6,99
a	0,65	0,89	1,09	1,12	c	0,921	0,320	a	7,58	2,83
a	1,21	1,22	1,11	1,4	c	0,837	0,409	a	9,35	5,73
a	0,95	1,04	0,76	1,15	b	0,642	0,309	a	10,27	6,24
b	2,01	1,24	0,78	0,8	b	0,754	0,288	d	25,95	20,80
b	1,64	0,93	0,96	1,48	c	0,844	0,340	a	11,04	6,83
b	2,13	1,38	1,15	0,93	d	0,433	0,242	b	16,03	11,53
a	1,41	0,85	1,15	0,73	c	0,923	0,289	e	31,41	26,44
a	1,22	1,03	1,1	0,87	c	0,824	0,243	b	16,92	11,96
b	2,81	1,13	1	0,65	c	0,963	0,335	b	16,02	11,47
b	2,05	1,25	0,91	0,88	c	0,800	0,223	b	17,74	12,96
b	2,31	0,87	1,14	1,37	a	0,828	0,696	b	17,00	11,96
b	1,8	0,84	0,94	0,89	c	1,000	0,366	b	18,36	13,36
a	1,15	1,27	1,25	1,16	b	0,715	0,354	b	15,73	11,25
a	0,74	2	0,86	0,85	a	0,885	0,573	b	17,98	13,18
a	0,89	2,05	0,9	0,87	b	0,658	0,278	b	17,33	12,11
a	0,68	1,39	0,71	0,99	a	0,739	0,632	a	12,62	8,28
a	0,94	2,07	1,09	0,78	b	0,684	0,357	c	21,71	16,34
a	1,06	1,92	1,01	0,82	a	0,684	0,592	c	22,13	16,88
a	0,76	2,27	1,09	1,17	d	0,512	0,428	e	30,27	26,10
a	1,12	1,86	0,9	1,26	a	0,763	0,777	c	20,92	15,50
a	1,11	1,9	1,33	0,93	d	0,452	0,339	e	30,17	25,93
a	1,03	1,25	0,91	1,02	d	0,468	0,428	c	22,83	17,16
a	1,57	2,1	0,94	1,09	c	0,849	0,298	c	21,02	15,63
a	1,05	0,56	1,82	0,95	c	0,761	0,252	c	21,84	16,83
a	0,92	0,93	2,17	1,17	d	0,298	0,201	e	31,82	26,62
a	1,11	0,9	2,36	1,03	c	0,820	0,438	d	25,06	19,97
a	1,44	1,29	2,11	0,95	c	0,886	0,252	d	26,06	21,15
a	0,81	1,03	2,36	1,12	a	0,761	0,534	c	20,60	15,17
a	0,69	1,29	1,79	0,92	c	0,827	0,361	d	24,51	19,51
a	1,04	0,83	2,76	1,26	a	0,894	1,000	e	28,94	23,35
a	0,98	1,17	1,62	0,93	d	0,381	0,459	d	26,12	21,25
a	1,23	0,88	2,69	0,63	b	0,750	0,301	d	23,90	18,94
a	1,25	1,11	2,52	1,38	a	0,613	0,512	d	23,74	18,34
c	1,26	0,95	0,66	2,44	d	0,443	0,369	d	25,30	20,05
c	1,03	1,27	1,39	2,15	c	0,853	0,231	c	20,51	15,11

Приложение И (обязательное)

Таблица И1 Задания к лабораторной работе 16

1 задание (4 класса.)		2 задание (3 класса)				3 задание (5 классов)	
1 признак	2 признак	1 признак	2 признак	3 признак	4 признак	1 признак	2 признак
0,25	0,30	24,4	38,0	2,6	7,0	13,16	8,54
0,49	0,50	19,6	30,2	2,0	6,2	10,40	6,32
0,77	0,24	27,7	40,7	2,7	7,4	10,44	6,47
0,84	0,35	23,8	35,8	2,4	6,7	9,89	5,97
0,92	0,32	19,8	30,6	2,1	6,2	12,81	8,39
0,84	0,41	28,3	41,8	2,9	7,5	7,80	3,41
0,64	0,31	22,2	33,1	2,3	6,3	12,18	8,07
0,75	0,29	21,1	31,5	2,2	6,2	13,27	8,75
0,84	0,34	30,4	44,1	3,0	7,8	8,17	4,48
0,43	0,24	22,1	32,8	2,2	6,3	18,36	13,35
0,92	0,29	30,7	44,6	3,1	8,0	24,73	19,53
0,82	0,24	59,1	83,4	5,5	12,9	15,54	11,24
0,96	0,33	6,2	15,6	0,2	3,6	17,65	12,60
0,80	0,22	39,5	61,2	3,8	9,7	17,41	12,19
0,83	0,70	46,0	67,4	4,3	10,8	17,90	13,17
1,00	0,37	46,3	67,7	4,4	10,8	16,41	11,74
0,72	0,35	36,9	56,7	3,6	9,1	15,74	11,28
0,89	0,57	41,2	61,9	4,0	9,9	16,83	11,82
0,66	0,28	47,7	69,2	4,4	11,1	20,40	14,73
0,74	0,63	47,0	68,8	4,4	11,0	20,60	15,14
0,68	0,36	46,7	68,7	4,4	11,0	20,72	15,45
0,68	0,59	46,6	67,9	4,4	10,9	19,93	14,38
0,51	0,43	45,7	67,3	4,3	10,7	22,09	16,87
0,76	0,78	38,7	59,4	3,8	9,4	20,94	15,53
0,45	0,34	43,7	64,7	4,2	10,5	21,40	16,02
0,47	0,43	43,5	63,7	4,2	10,4	16,33	11,73
0,85	0,30	42,7	62,9	4,1	10,3	21,34	16,00
0,76	0,25	60,8	86,9	5,7	13,5	21,67	16,27
0,30	0,20	60,1	85,1	5,6	13,3	21,00	15,58
0,82	0,44	58,1	81,5	5,4	12,7	30,33	26,10
0,89	0,25	54,0	78,3	5,2	12,0	25,62	20,30
0,76	0,53	65,3	93,2	6,1	14,5	26,03	21,01
0,83	0,36	66,6	93,8	6,2	14,6	24,12	19,06
0,89	1,00	54,0	78,4	5,2	12,0	25,35	20,11
0,38	0,46	64,2	92,4	6,0	14,1	24,47	19,23
0,75	0,30	59,7	84,9	5,6	13,3	24,17	19,19
0,61	0,51	56,5	81,0	5,4	12,6	25,80	20,43
0,44	0,37	65,2	93,2	6,1	14,3	23,85	18,81
0,85	0,23	60,5	86,1	5,6	13,5	29,33	24,94

Учебное издание

ЖУЧЕНКО Юрий Михайлович

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
В БИОЛОГИИ И ХИМИИ
Лабораторный практикум**

Учебно-методическое пособие

Редактор В. И. Шкредова
Корректор В. В. Калугина

Подписано в печать 2010. Формат 60×84 $\frac{1}{16}$
Бумага офсетная. Ризография. Усл. печ. л.
Уч. – изд. л. Тираж **100** экз. Заказ №

Издатель и полиграфическое исполнение:
учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»
ЛИ № 02330/0549481 от 14. 05. 2009.
ЛП № 02330/0150450 от 03. 02. 2009.
Ул. Советская, 104, 246019, г. Гомель